© DIGITAL STOCK

Surfing the Brain

An Overview of Wavelet-Based Techniques for fMRI Data Analysis

BY DIMITRI VAN DE VILLE, THIERRY BLU, AND MICHAEL UNSER

The measurement of brain activity in a noninvasive way is an essential element in modern neurosciences. Modalities such as electroencephalography (EEG) and magnetoencephalography (MEG) recently gained interest, but two classical techniques remain predominant. One of them is positron emission tomography (PET), which is costly and lacks temporal resolution but allows the design of tracers for specific tasks; the other main one is functional magnetic resonance imaging (fMRI), which is more affordable than PET from a technical, financial, and ethical point of view, but which suffers from poor contrast and low signal-to-noise ratio (SNR). For this reason, advanced methods have been devised to perform the statistical analysis of fMRI data.

The blood-oxygen-level-dependent (BOLD) signal, discovered by [1] in the 1990s and later elucidated in [2], has allowed fMRI to evolve into a prominent tool to perform noninvasive studies of the function of the brain. In T2*-weighted magnetic resonance (MR) images, the BOLD signal exhibits a weak and noisy contrast. The aim of fMRI data analysis is to detect this signal in a time series of acquisitions. The purpose of this article is to give a unifying overview of techniques that deploy the wavelet transform to perform this analysis.

The wavelet transform is a powerful tool [3], [4]. Unlike the Fourier sinusoids, which provide a sharp frequency characterization of a given signal but are unable to identify transient events, wavelets achieve a balance between localization in space or time and localization in the frequency domain. This balance is intrinsic to multiresolution, which allows the analysis to deal with image features at any scale. As the discrete wavelet transform (DWT) corresponds to a basis decomposition, it provides a nonredundant and unique representation of the signal. These fundamental properties are key to the efficient decomposition of the nonstationary processes typical of fMRI experimental settings. Consequently, wavelets have received a large recognition in biomedical signal and image processing; several overviews are available [5]–[7], including work that is tailored to fMRI [8].

The first application of wavelets in fMRI was pioneered by Ruttimann et al. [9], [10]. After computing the wavelet transform of each volume, the parameter for an on/off type activation is extracted, followed by a coefficient-wise statistical test for this parameter. Such a procedure takes advantage of two properties of the wavelet transform. First, wavelets allow us to obtain a sparse representation of the activation map, in the sense that only a few wavelet coefficients are needed to efficiently encode the spatial activation patterns. Consequently, the SNR of signal-carrying coefficients has increased with respect to the original voxels, thus improving the potential sensitivity of detecting activation patterns burried in large noise. Second, the wavelet transform approximately acts as a decorrelator. Therefore, the use of simple techniques to deal with the multiple-testing problem, such as Bonferroni correction, is appropriate since the coefficients are nearly decorrelated. The power of the statistical test in the wavelet domain has been increased by proposing other error rates than the Type I error (i.e., the number of false positives). Hilton et al. [11] introduced recursive testing (change-point detection) in fMRI analysis, which consists of altering the hypotheses of the test procedure in the wavelet domain. Raz and Turetsky [12], [13] applied the principle of false discovery rate (FDR).

The wavelet transform has also been deployed along the temporal dimension. At the same time, LaConte et al. [14] and Alexander et al. [15] proposed a temporal denoising preprocessing step. Serial correlations in fMRI data are common due to head-motion artifacts, background neuronal processes, and acquisitions effects. Bullmore et al. [16] pioneered bootstrapping techniques in the wavelet domain to deal with the colored noise structure of fMRI data. Bootstrapping techniques rely on the whitening property of the wavelet transform to generate surrogate data that are used to build an empirical statistical measure under the null hypothesis [17]–[19]. Tscharner et al. [20] proposed the use of the continuous wavelet transform in a nonparametric detection scheme. Fadili and Bullmore [21] exploited the whitening property of the discrete transform to obtain a best linear unbiased estimate for the parameters of the linear model (LM). Hossein-Zadeh et al. [22] deployed a redundant wavelet transform for nonparametric detection, while Meyer [23] proposed them as a tool to estimate semiparametric models in fMRI. Finally, Shimizu et al. [24] and Maxim et al. [25] obtained spectral characteristics of fMRI time series using the wavelet transform.

We give an overview here of techniques that deploy the wavelet transform in the spatial domain. To this aim, we consider the coefficient-wise statistical test procedure in

0739-5175/06/$20.00©2006IEEE

the wavelet domain as the basic scheme. Most methods have concentrated on relaxing the Bonferroni correction of such a coefficient-wise test. For each of them, we clearly specify the modifications applied, and we highlight some interesting unifications. For example, FDR (based on an alternative error rate) and recursive testing (based on modified hypotheses) seem unrelated at first sight, still, they can be interpreted very similarly. Additionally, the statistical test procedure based on the posterior probability density function of the Bayesian framework can be, under relatively mild assumptions, interpreted in an FDR-like way. All these methods require a prespecified level α to determine the significance of the test procedure. We point out the weak Type I error control of these methods and note that it could be used as the common ground to relate the various prespecified levels. Finally, we include recent advances such as a spatio-wavelet approach, which puts back the statistical test procedure in the spatial domain while maintaining powerful wavelet processing.

The article is organized as follows. In the following section, we introduce the basics of fMRI data analysis, together with a purely spatial statistical test. We also briefly highlight the main features of statistical parametric mapping (SPM), a standard method for fMRI analysis. The overview of wavelet-based methods is presented, and finally, examples and experimental results are given.

The Basics of fMRI Data Analysis

We introduce here basic elements for a parametric, hypothesis-driven approach to the statistical analysis of fMRI. We first describe a voxelwise testing procedure that does not exploit the spatial correlation properties of the data. Next, we briefly mention some features of SPM, a popular methodology with an associated software package.

A Purely Spatial Approach

An fMRI dataset $v[\mathbf{n}; t]$ consists of a sequence of slices or volumes, where $\mathbf{n} \in \mathbb{Z}^3$ and $t = 1, \ldots, N$, are the three-dimensional (3-D) spatial and temporal indices, respectively. For each voxel, we also introduce a time-series vector of length N: $\mathbf{v}[\mathbf{n}] = [v[\mathbf{n}; 1] \ldots v[\mathbf{n}; N]]^{\mathrm{T}}$.

In a parametric approach, the temporal behavior of a voxel can be explained by a LM that is then fitted to the data. Specifically, if one wants to identify L regression variables, one writes the model as

$$\mathbf{v}[\mathbf{n}] = \mathbf{X}\mathbf{y}[\mathbf{n}] + \mathbf{e}[\mathbf{n}], \quad \text{for each } \mathbf{n}, \tag{1}$$

where $\mathbf{X}$ is the $N \times L$ design matrix, $\mathbf{y}[\mathbf{n}]$ is the $L \times 1$ parameter vector, and $\mathbf{e}[\mathbf{n}]$ is the residual error. In a simple block-based paradigm, $\mathbf{X}$ could consist of just two columns: one for the on-off stimulus, eventually convolved with the hæmodynamic response function (HRF), and one for the background signal. The regression variables in (1) are easy to estimate if we assume that the residual errors $\mathbf{e}[\mathbf{n}]$ are (temporally) independently and identically Gaussian-distributed. In such a case, the optimal unbiased estimate of $\mathbf{y}[\mathbf{n}]$ is the least squares solution $\bar{\mathbf{y}}[\mathbf{n}] = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{v}[\mathbf{n}]$. The residual error is then $\bar{\mathbf{e}}[\mathbf{n}] = \mathbf{v}[\mathbf{n}] - \mathbf{X}\bar{\mathbf{y}}[\mathbf{n}]$. Typically, one is only interested in a subset of the fitted parameters, for example, the parameter for the first column of the design matrix for the simple on/off experiment. The contrast vector $\mathbf{c}$ extracts the parameter(s) of interest into the two important measures

$$u[\mathbf{n}] = \mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}}[\mathbf{n}], \tag{2}$$

$$s^2[\mathbf{n}] = \bar{\mathbf{e}}^{\mathrm{T}}[\mathbf{n}]\bar{\mathbf{e}}[\mathbf{n}]\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}. \tag{3}$$

Assuming that the model holds and that the noise is independent identically distributed (i.i.d.), these are known to follow a Gaussian and a χ^2-distribution, respectively, and are independent from one another.

At this point, hypothesis testing is performed to determine whether or not the voxel $\mathbf{n}$ is activated, i.e., if the mean value is zero or not:

$$\begin{aligned} \mathcal{H}_0 &: \mathcal{E}[u[\mathbf{n}]] = 0, \\ \mathcal{H}_1 &: \mathcal{E}[u[\mathbf{n}]] > 0. \end{aligned} \tag{4}$$

Testing the nonvalidity of the null hypothesis for a given significance level α is achieved by thresholding a test statistic built out of $u[\mathbf{n}]$ and $s^2[\mathbf{n}]$. Specifically, the value

$$t[\mathbf{n}] = \frac{u[\mathbf{n}]}{\sqrt{s^2[\mathbf{n}]/J}}, \quad \text{with } J = N - \mathrm{rank}(\mathbf{X}), \tag{5}$$

follows a normalized Student t distribution with J degrees of freedom. The test procedure can then be summarized as checking whether $t[\mathbf{n}] \geq \tau$, with $p = \mathrm{Prob}[\mathfrak{t} \geq \tau]$, where $\mathfrak{t}$ is a reference random variable that follows the Student t distribution. Notice that the test is one-sided since we are interested in detecting activation (as opposed to deactivation). The detected parameter map is

$$d[\mathbf{n}] = H(t[\mathbf{n}] - \tau)u[\mathbf{n}], \tag{6}$$

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

$$H(t) = \begin{cases} 0, & \text{when } t < 0, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

In other words, the term $H(t[\mathbf{n}] - \tau)$ in (6) acts as a weight that is equal to 1 for $t[\mathbf{n}] \geq \tau$ and 0 otherwise. To summarize, we show a schematic overview of the spatial approach in Figure 1.

The significance level of a statistical test corresponds to the Type I error rate, i.e., the probability p of a false positive detection. To carry out the analysis of fMRI data, we have to repeat the same

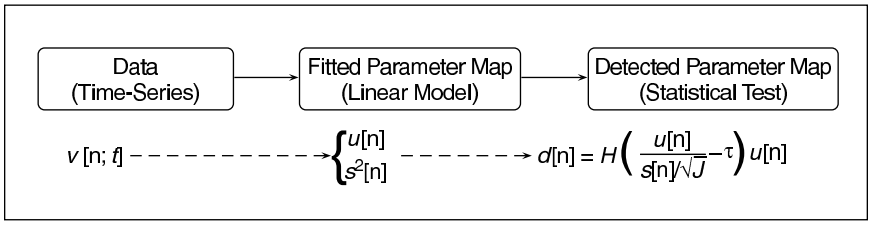

Fig. 1. A schematic overview of the purely spatial approach.

test for many voxels, which increases the global Type I error rate. In particular, for V tests, we expect to obtain pV false positives, which usually become too high since V is large. Achieving a global Type I error rate that remains below a desired significance level α requires that

$$\text{Prob}\left[\bigvee_{\mathbf{n}} t[\mathbf{n}] \geq \tau\right] \leq \alpha, \tag{8}$$

where the $\bigvee$-symbol denotes the disjunction operator. The simplest way of keeping the Type I error rate below α is by Bonferroni correction. We give an upper bound to the probability as

$$\text{Prob}\left[\bigvee_{\mathbf{n}} t[\mathbf{n}] \geq \tau\right] \leq \sum_{\mathbf{n}} \text{Prob}[t \geq \tau] = pV, \tag{9}$$

where V is the number of voxels in the volume or, more precisely, the number of intracranial voxels. As a result, utilizing the significance level $p = \alpha/V$ for the individual tests guarantees the desired global level. Unfortunately, while the Bonferroni correction maintains a strong Type I error control, it also highly reduces the sensitivity. The main reason for the suboptimality of Bonferroni correction lies in the fact that it does not take into account the spatial correlation between voxels [26].

Statistical Parametric Mapping

We now briefly mention the SPM approach [27], [28], which is probably the most popular parametric hypothesis-driven method for the analysis of fMRI data. To control the multiple-hypothesis testing problem, SPM considers the data as a lattice representation of a continuous Gaussian random field. For this to be valid, SPM needs a smoothing preprocessing step [29], [30]. Specifically, the volumes $v[\mathbf{n}; t]$ are filtered by a Gaussian, characterized by its full-width-at-half-maximum (FWHM), to obtain the volumes $v_g[\mathbf{n}; t]$. The hypotheses of the statistical test of SPM are

$$\begin{aligned} \mathcal{H}_0 &: \mathcal{E}[u_g[\mathbf{n}]] = 0, \\ \mathcal{H}_1 &: \mathcal{E}[u_g[\mathbf{n}]] > 0, \end{aligned} \tag{10}$$

where $u_g[\mathbf{n}]$ is the parameter map derived by the LM from the smoothed data. Correction for multiple testing is done based on an approximation of (8), which takes into account the local expected Euler characteristic.

SPM is implemented as a software package for MATLAB. This package has evolved a lot over the years. In its current version (SPM2), it allows the development of third-party add-ons as toolboxes. The main advantage of such an open software architecture is clearly the possibility of extending the already wide functionality of SPM and to tailor it to specific applications.

Wavelets in fMRI—Surfing the Brain

A Primer to the Wavelet Transform

The DWT of a one-dimensional signal $v(x)$—typically represented by its samples $v[n]$, $n \in Z$—is a decomposition into a sum of basis functions. These functions are shifted and dilated versions of a (bandpass) wavelet function $\psi(x)$ and shifted versions of a (low-pass) scaling function $\varphi(x)$. Mathematically, we write

$$\begin{aligned} v(x) =& 2^{J_w/2} \sum_k c_{J_w}[k]\varphi(x/2^{J_w} - k) \\ &+ \sum_{j=1}^{J_w} 2^{j/2} \sum_k w_j[k]\psi(x/2^j - k) \end{aligned} \tag{11}$$

for a decomposition of J_w iterations. The low-pass coefficients and detail (wavelet) coefficients are given by

$$c_{J_w}[k] = \langle v(\cdot), 2^{-J_w/2}\tilde{\varphi}(\cdot/2^{J_w} - k)\rangle, \tag{12}$$

$$w_j[k] = \langle v(\cdot), 2^{-j/2}\tilde{\psi}(\cdot/2^j - k)\rangle, \tag{13}$$

respectively [3], [4]. The functions $\tilde{\varphi}$ and $\tilde{\psi}$ are related to φ and ψ.

In practice, the calculation of the coefficients is performed by a fast iterated filterbank algorithm. Specifically, at the jth iteration, the scaling coefficients $c_j[k]$ and the wavelet coefficients $w_j[k]$ are obtained by digital filtering and subsampling with an analysis scaling filter $\tilde{h}$ and with an analysis wavelet filter $\tilde{g}$, respectively:

$$c_j[k] = (c_{j-1} * \tilde{h}^{\mathrm{T}})[2k], \tag{14}$$

$$w_j[k] = (c_{j-1} * \tilde{g}^{\mathrm{T}})[2k]. \tag{15}$$

The wavelet coefficients (eventually after modification) can then be used to reconstruct the signal using the inverse scheme. The coefficients at the jth iteration are recovered by

$$c_j[k] = (c_{j+1} \uparrow 2 * h)[k] + (w_{j+1} \uparrow 2 * g)[k], \tag{16}$$

where h and g are synthesis scaling and wavelet filters, respectively, and where $\uparrow 2$ corresponds to an upsampling operation by a factor 2. The reconstruction is repeated until the finest level is reached, where the $c_0[k]$ represent $v(x)$ by integer-shifts of $\varphi(x)$.

The wavelet transform can be extended to multiple dimensions in a straightforward way using the tensor product. In two dimensions, for example, this construction leads to one scaling function $\varphi(x_1)\varphi(x_2)$ and three wavelets $\varphi(x_1)\psi(x_2)$, $\psi(x_1)\varphi(x_2)$, $\psi(x_1)\psi(x_2)$. It leads to one scaling function and seven wavelets in 3-D.

In the remainder of this article, we write the multidimensional version of the decomposition (11) as

$$v(\mathbf{n}) = \sum_{\mathbf{k}} v_w[\mathbf{k}]\psi_{\mathbf{k}}(\mathbf{n}), \tag{17}$$

where the coefficients $v_w[\mathbf{k}]$ run over all scales and orientations of the decomposition (including the low-pass), while the functions $\psi_{\mathbf{k}}$ correspond to the associated scaled, shifted, and dilated version of the scaling function or of the wavelet, including the normalization factor. The subscript w is used consistently to denote the wavelet coefficients, while the indexes $\mathbf{n}$ and $\mathbf{k}$ are used in the spatial domain and in the wavelet domain, respectively.

Wavelet Processing as Denoising

Wavelets were primarily applied in medical imaging for denoising, a technique that was pioneered by Weaver et al. [31] in the context of MRI. In a general denoising approach, the LM of (1) is replaced by a nonparametric regression where the data is modeled as the superposition of an unknown signal and a noise component:

$$v[\mathbf{n}; t] = y[\mathbf{n}; t] + e[\mathbf{n}; t]. \tag{18}$$

The spatial wavelet transform of each of the volumes $v[\mathbf{n}; t]$ provides us with the wavelet coefficients $v_w[\mathbf{k}; t]$, on which we can then apply a hard or soft thresholding rule. The denoised volumes are reconstructed from the modified wavelet coefficients $\hat{v}_w[\mathbf{k}; t]$, as indicated in Figure 2.

The well-known technique of wavelet shrinkage has been systematized by Donoho and Johnstone [32], [33] and has been applied to fMRI data as a preprocessing step [34]–[36]. In a recent study by Wink and Roerdink [37], an extensive comparison of various threshold-selection schemes suggests that the denoised volumes show higher SNRs and that they retain activation patterns well. Another promising indication was made by Aston et al. [38], who shows that the residual in the spatial domain, after wavelet shrinkage, would decrease significantly. Nevertheless, suitable statistical treatments of the denoised volumes, such as performing a test, are made difficult by the nonlinear operation in the wavelet domain.

Wavelet Processing and Nonparametric Methods

The denoising strategies make relatively mild statistical assumptions. Several alternative methods have been proposed to take into account the specificity of the fMRI spatio-temporal data. Among the nonparametric methods, we like to briefly mention the few that deploy the transform in the spatial domain.

Long et al. [39] proposed to optimize a spatio-temporal likelihood function, with a spatially varying threshold in the wavelet domain as a regularization for the activation map. Breakspear et al. [40] applied nonparametric tests based on bootstrapping methods in the spatial and spatio-temporal domain. The wavelet transform has a great advantage since it allows the application of bootstrapping on a spatially constrained region, thus having a great potential to investigate functional connectivity in the brain. In a recent work, Whitcher [41] applied a wavelet bootstrap based on random processes of the Matern class.

In the remaining of this article, we primarily focus on parametric hypothesis-driven approaches.

Wavelet Processing as Probabilistic Shrinkage

We consider the LM (1) and translate it into the wavelet domain:

$$\mathbf{v}_w[\mathbf{k}] = \mathbf{X}\mathbf{y}_w[\mathbf{k}] + \mathbf{e}_w[\mathbf{k}], \text{ for each } \mathbf{k}, \tag{19}$$

where the vector $\mathbf{v}_w[\mathbf{k}]$ is formed out of the wavelet coefficients $v_w[\mathbf{k}; t]$. Under the Gaussian i.i.d. assumption for the noise, the estimates of the parameters can be determined as $\bar{\mathbf{y}}_w\mathbf{k} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{v}_w[\mathbf{k}]$ and the residual error as $\bar{\mathbf{e}}_w[\mathbf{k}] = \mathbf{v}_w[\mathbf{k}] - \mathbf{X}\bar{\mathbf{y}}_w[\mathbf{k}]$. Consequently, for the parameter of interest, we obtain

$$u_w[\mathbf{k}] = \mathbf{c}^T\bar{\mathbf{y}}_w[\mathbf{k}], \tag{20}$$

$$s_w^2[\mathbf{k}] = \bar{\mathbf{e}}_w^T[\mathbf{k}]\bar{\mathbf{e}}_w[\mathbf{k}]\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}. \tag{21}$$

We now have all the elements to explain the various methods of probabilistic shrinkage.

Coefficient-Wise Testing

The translation of hypothesis testing of the parameter map into the wavelet domain was first proposed by Ruttimann et al. [9], [10] and was later followed and extended by other researchers [42]–[45]. The original hypotheses (4) are modified into

$$\mathcal{H}_0 : \mathcal{E}[u_w[\mathbf{k}]] = 0,$$
$$\mathcal{H}_1 : \mathcal{E}[u_w[\mathbf{k}]] \neq 0. \tag{22}$$

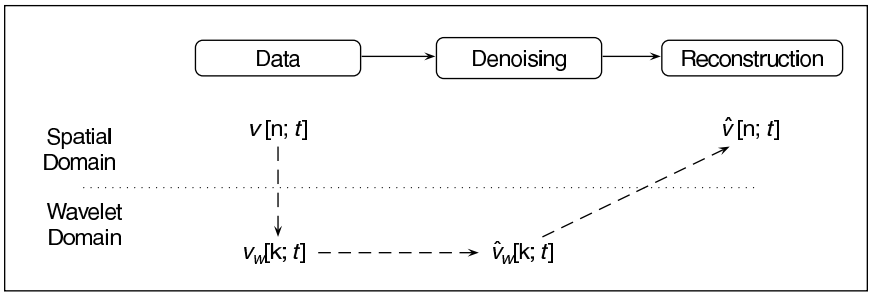

Fig. 2. A schematic overview of a wavelet-based denoising approach.

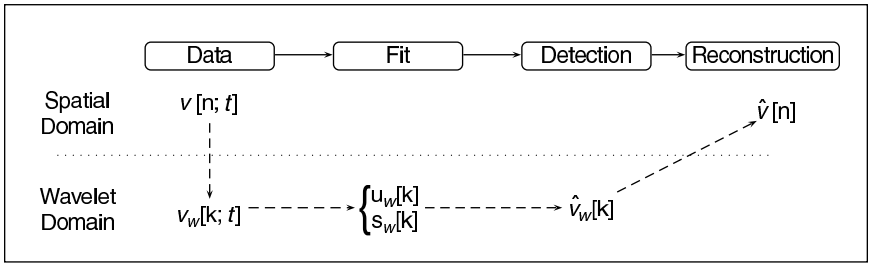

Fig. 3. A schematic overview of the probabilistic shrinkage approaches.

Similar to the spatial-domain voxelwise test, we can construct a test statistic for each wavelet coefficient,

$$t_w[\mathbf{k}] = \frac{u_w[\mathbf{k}]}{s_w[\mathbf{k}]/\sqrt{J}}, \tag{23}$$

which follows a Student t distribution with J degrees of freedom. The test procedure itself needs to be modified into a two-sided test since all wavelet coefficients could contribute to positive values of the parameter map, even when these coefficients are negative. Therefore, we test $|t_w[\mathbf{k}]| > \tau$, with $p = \text{Prob}[|t| \geq \tau]$. The wavelet coefficients $u_w[\mathbf{k}]$ for which the t-values survive the test are included in the reconstruction of a parameter map $\hat{u}[\mathbf{n}]$. See Figure 3.

Hypothesis testing in the wavelet domain still suffers from the multiple-

testing problem. Once more, we can use Bonferroni correction to have $pV = \alpha$, where α is the desired global significance level and where V corresponds to the number of intracranial voxels. (The latter is approximately equal to the number of intracranial wavelet coefficients.) Because of the decorrelating properties of the wavelet transform, the application of the Bonferroni correction in the wavelet domain should be more optimal than in the spatial domain. The total number of tests can also be further reduced by performing an omnibus test first at each decomposition level [10].

False Discovery Rate

Although controlling the global Type I error rate has an easy interpretation (i.e., we have a chance α of finding a false positive in the whole volume), it is sometimes considered too conservative. Therefore, alternative error rates have been proposed to increase the power of detecting true positives. The FDR, or the expected proportion of false positives to total positives, was proposed by Benjamini, Abramovich, and Hochberg [46]–[48], and afterwards applied to neuroimaging by Raz and others [12], [13], [49], [50]. Mathematically, FDR is defined as

$$\mathrm{FDR} = \mathcal{E}\left[\frac{D_F}{D}\right], \tag{24}$$

where D_F is the number of false positives and where D is the total number of positives. For $D = 0$, the FDR is defined as 0. Keeping the FDR at a desired fraction α corresponds to allowing, on average, for αD false positives.

The FDR principle can be translated into a practical algorithm as follows: first, we sort the t-values $t_w[\mathbf{k}]$ of the wavelet coefficients as $|t_{(1)}| \geq |t_{(2)}| \geq \cdots \geq |t_{(V)}|$. Consequently, the corresponding p-values $p_{(i)} = \mathrm{Prob}[|t| \geq |t_{(i)}|]$ are increasing. Then, the FDR corresponding to $D = i$ total positives can be rewritten as

$$\mathcal{E}\left[\frac{D_F}{D}\middle| D = i\right] = \frac{\mathcal{E}[D_F | D = i]}{i} = \frac{V}{i}\mathrm{prob}[|t| \geq |t_{(i)}|]. \tag{25}$$

Therefore, keeping FDR $\leq \alpha$ requires that

$$\mathrm{Prob}[|t| \geq |t_{(i)}|] \leq \alpha\frac{i}{V}. \tag{26}$$

This insight leads to the so-called step-up procedure for controlling the FDR.

- The wavelet coefficients are sorted according to increasing p-values.
- The index i is determined by the largest p-value $p_{(i)}$ for which $p_{(i)} \leq \alpha i / V$.
- The wavelet coefficients $u_{(1)}, \ldots, u_{(i)}$ are retained for the reconstruction of a parameter map $\hat{u}[\mathbf{n}]$.

The FDR approach assumes (at least) positively dependent test statistics [51]. Further on, it is important to recognize that FDR only offers a weak Type I error control; i.e., the effective Type I error rate is equal to α only for the omnibus null-hypothesis test.

Recursive Hypothesis Testing or Change-Point Detection

Ogden et al. [11], [52] proposed the idea of recursive testing (change-point detection) to retain the maximal number of wavelet coefficients with strong evidence, while testing the null hypothesis on the remaining coefficients. This technique, which could be implemented as a recursive method, has also been extended to hypothesis testing in the context of fMRI by Fadili and Bullmore [50]. Here, we show that recursive testing can be implemented in a way very similar to FDR.

The starting point consists of adjusting the hypotheses by introducing a set of indices S such that

$$\begin{aligned} \mathcal{H}_0 &: \mathcal{E}[u_w[\mathbf{k}]] = 0, \mathbf{k} \in S, \\ \mathcal{H}_1 &: \mathcal{E}[u_w[\mathbf{k}]] \neq 0, \mathbf{k} \in S, \text{ and} \\ &\quad \mathcal{E}[u_w[\mathbf{k}]] = 0, \mathbf{k} \notin S. \end{aligned} \tag{27}$$

Expressing the Type I error rate under the null hypothesis, assuming independent test statistics, results into

$$\mathrm{Prob}\left[\bigvee_{\mathbf{k}\in\mathbf{S}} |t_w[\mathbf{k}]| > \tau\right] = 1 - \mathrm{Prob}\left[\bigwedge_{\mathbf{k}\in\mathbf{S}} |t_w[\mathbf{k}]| \leq \tau\right] \tag{28}$$

$$= 1 - \mathrm{Prob}\left[\max_{\mathbf{k}} |t_w[\mathbf{k}]| \leq \tau\right] \tag{29}$$

$$\leq 1 - \mathrm{Prob}[|t| \leq \tau]^{\#S} \tag{30}$$

$$= 1 - (1 - p)^{\#S}, \tag{31}$$

where #S is the cardinal of the set S. Imposing a global significance level α thus results into

$$p \leq 1 - (1 - \alpha)^{1/\#S}. \tag{32}$$

Therefore, recursive testing can be translated into a procedure that is very similar to FDR:

- The wavelet coefficients are sorted according to increasing p-values.
- The index i is determined by the largest $p_{(i)}$ for which $p_{(i)} \leq 1 - (1 - \alpha)^{1/(V-i)}$.
- The wavelet coefficients $u_{(1)}, \ldots, u_{(i)}$ are retained and used for the reconstruction of $\hat{u}[\mathbf{n}]$.

Once again, recursive testing provides only a weak Type I error control with a significance level α. The method can be implemented in a subband-by-subband way, but in that case, one should divide α by the number of subbands in order to keep the same weak Type I error control.

The underlying ideas of FDR and recursive testing are still subjects of active research. Shen et al. [53] recently proposed enhanced FDR as a combination of both principles (i.e., alternative error rate with modified hypotheses). Some promising preliminary results for synthetic fMRI datasets were published by Pavlicová et al. [54].

Bayesian Framework

The Bayesian framework has been applied to further refine the statistics in the wavelet domain by Vidakovic and Abramovich [55], [56]. Fadili and Bullmore [50], [57] further developed this framework in the context of fMRI. The basic ingredient of the Bayesian approach is the prior proba-

bility density function of the wavelet coefficients, which consists of the mixture model

$$f_{u_w[\mathbf{k}]}(u) = m_{\vartheta(\mathbf{k})} f_{\mathcal{H}_1}(u; \vartheta(\mathbf{k})) + (1 - m_{\vartheta(\mathbf{k})}) f_{\mathcal{H}_0}(u), \quad (33)$$

where $\vartheta(\mathbf{k})$ returns the scale and orientation of the wavelet coefficient with index $\mathbf{k}$, where $m_{\vartheta(\mathbf{k})}$ is the probability mass of having an active coefficient at scale and orientation $\vartheta(\mathbf{k})$, and where $f_{\mathcal{H}_1}$ and $f_{\mathcal{H}_0}$ are the probability density functions under the alternative and null hypothesis, respectively. Traditionally, the ratio test between the posterior probability of having the alternative versus the null hypothesis has been proposed.

An alternative approach, which is more closely related to the p-value interpretation of the previously presented methods, consists of applying a threshold for the posterior probability of the null hypothesis [50], [58], [59]. Specifically, we rewrite the posterior probability of the null hypothesis when $|t_w[\mathbf{k}]| \geq \tau$, using Bayes' rule, as

$$\begin{aligned} \text{Prob}\left[\mathcal{H}_0 \middle| |t_w[\mathbf{k}]| \geq \tau\right] &= \frac{\text{Prob}[\mathcal{H}_0]\text{Prob}[|t_w[\mathbf{k}]| \geq \tau | \mathcal{H}_0]}{\text{Prob}[|t_w[\mathbf{k}]| \geq \tau]} \\ &= \frac{(1 - m_{\vartheta(\mathbf{k})}) f_{\mathcal{H}_0}(t_w[\mathbf{k}] \geq \tau)}{f(t_w[\mathbf{k}] \geq \tau)}. \quad (34) \end{aligned}$$

The threshold τ is then chosen such that this probability is below a desired value α. In practice, the Bayesian framework requires us to estimate the hyperparameters by an expectation-maximization algorithm [60].

Storey [59], [61] has shown that the posterior probability of the Bayesian framework is—assuming independent test statistics $t_w[\mathbf{k}]$—equivalent to the positive FDR, defined as

$$\begin{aligned} \text{pFDR} &= \frac{\text{FDR}}{\text{Prob}[\text{D} > 0]} = \mathcal{E}\left[\frac{D_F}{D} \middle| D > 0\right] \\ &\equiv \text{Prob}\left[\mathcal{H}_0 \middle| |t_w[\mathbf{k}]| \geq \tau\right]. \quad (35) \end{aligned}$$

Curiously, the Bayesian framework becomes equivalent to the FDR strategy when assuming $\text{Prob}[\mathcal{H}_0] = 1$. Indeed, assuming that the p-value $p_{(i)}$ corresponds to a threshold $\tau_{(i)}$, we can rewrite (34) using $p_{(i)} = \text{Prob}[|t_{(i)} \geq \tau_{(i)}|\mathcal{H}_0]$ and $\text{Prob}[|t| \geq \tau_{(i)}] = i/V$ as $Vp_{(i)}/i$. This asymptotic behavior provides us with a useful insight into the interpretation of α for the Bayesian framework: it shows that controlling the posterior probability of the null hypothesis roughly corresponds to controlling the positive FDR.

In Figure 4, we show the maximum p-value included in the detected parameter map as a function of the total number of detections D. All methods obtain at least a weak Type I error rate of α. Bonferroni correction is clearly most conservative, followed by the recursive testing method and FDR. From this point of view, the Bayesian approach should be comparable to FDR.

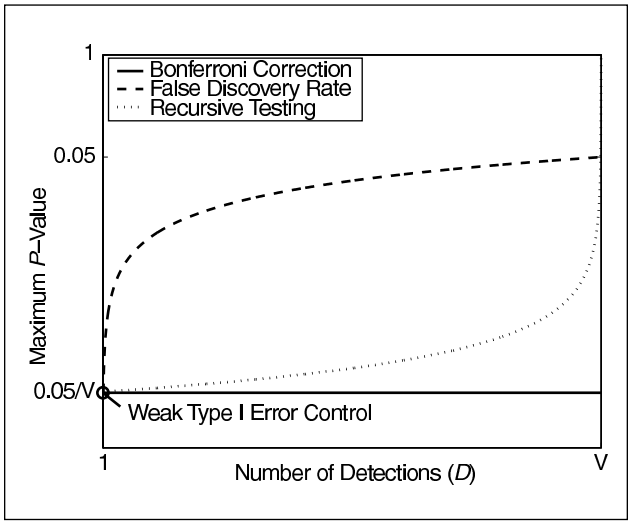

Fig. 4. The maximum p-value as a function of the total number of detections D. The p-values are shown in logarithmic scale. The curves correspond to $\alpha = 5\%$ and $V = 1{,}000$.

Spatio-Wavelet Approach: Combining Wavelet Processing with Spatial Testing

Probabilistic shrinkage techniques concentrate on improving the sensitivity of detecting wavelet coefficients, which is a proper aim given the hypotheses (22). These wavelet-based approaches perform at the same time approximation (i.e., improving the SNR of the parameter map as an effect of wavelet shrinkage) and detection (i.e., deciding whether the null hypothesis has been violated for a wavelet coefficient). However, these approaches have difficulties to map the statistics back into the spatial domain. In other words, how can we determine whether a particular voxel of the reconstructed parameter map $\hat{u}[\mathbf{n}]$ is activated? Some proposed solutions include the application of an ad-hoc threshold (e.g., a percentage of the maximal signal level or of the estimated noise level [10]) or retesting in the spatial domain without taking into account the effect of the initial test in the wavelet domain [62]).

The spatio-wavelet approach that we recently proposed [63], still relies on the good properties of the wavelet transform to process the parameter map in the wavelet domain but removes the statistical interpretation; i.e., there is no statistical hypothesis test involved (see Figure 5). (This approach has been implemented as a toolbox for SPM, called *WSPM: Wavelet-based SPM*, and is available at http:// bigwww.epfl.ch/wspm.) Therefore, in some sense, the spatio-wavelet approach is comparable to wavelet processing as a denoising step. However, a statistical detection

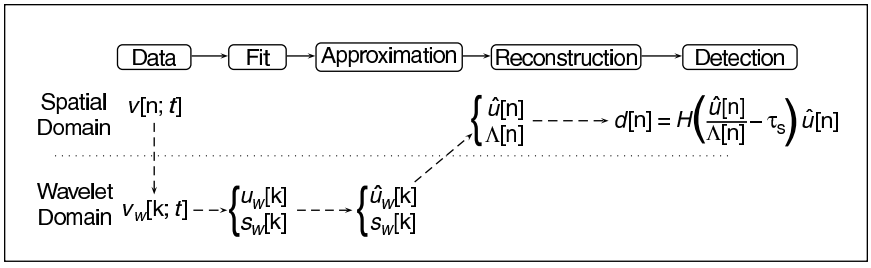

Fig. 5. A schematic overview of the spatio-wavelet approach.

Wavelets achieve a balance between localization in space or time and localization in the frequency domain.

procedure is proposed in the spatial domain that takes into account the effect of the wavelet processing. In its current form, the spatio-wavelet framework uses Bonferroni correction for multiple hypothesis testing, thus ensuring a strong Type I error control. We now explain the two steps of the spatio-wavelet approach (i.e., wavelet processing and detection in the spatial domain).

Wavelet Processing

Many results (e.g., Desco et al. [44]) indicate that a simple coefficient-wise t-test in the wavelet domain, even with Bonferroni correction, seems to achieve sufficient sensitivity. Therefore, we get inspiration from a scheme where the approximation step is as follows: the wavelet coefficients $u_w[\mathbf{k}]$ that survive the thresholding $|t_w[\mathbf{k}]| \geq \tau_w$ are kept and reconstructed as

$$\hat{u}[\mathbf{n}] = \sum_{\mathbf{k}} \underbrace{H(|t_w[\mathbf{k}]| - \tau_w) u_w[\mathbf{k}]}_{\hat{u}_w[\mathbf{k}]} \psi_{\mathbf{k}}(\mathbf{n}). \quad (36)$$

While the initial parameter map $u[\mathbf{n}]$ can be considered as the best linear estimate (it is even the best one among all estimators under the Gaussian assumption), the approximated map $\hat{u}[\mathbf{n}]$ is expected to be improved by the nonlinear thresholding operation in the wavelet domain, which exploits the spatial correlation. Since the approximation step has now lost its statistical interpretation, the threshold value τ_w becomes a parameter of the framework, to be determined later.

Detection in the Spatial Domain

In a second phase, we want to test whether a particular voxel of the reconstructed parameter map can be considered as activated or not. Therefore, we propose the hypotheses

$$\begin{aligned} \mathcal{H}_0 &: \mathcal{E}[\hat{u}[\mathbf{n}]] = 0, \\ \mathcal{H}_1 &: \mathcal{E}[\hat{u}[\mathbf{n}]] > 0, \end{aligned} \quad (37)$$

which are expressed in the spatial domain. The test procedure relies on a theorem [63] that states that

$$\text{Prob}\,[\hat{u}[\mathbf{n}] \geq \tau_S \Lambda[\mathbf{n}]] \leq \Upsilon(\tau_W, \tau_S), \quad (38)$$

where $\Lambda[\mathbf{n}]$ is a special reconstruction of the standard deviations $s_w[\mathbf{k}]$ that makes use of the absolute-valued wavelets, given by

$$\Lambda[\mathbf{n}] = \sum_{\mathbf{k}} \frac{s_W[\mathbf{k}]}{\sqrt{J}} |\psi_{\mathbf{k}}(\mathbf{n})|. \quad (39)$$

The function $\Upsilon(\tau_w, \tau_s)$ is data independent and is a function of the parameters τ_w and τ_s. It is given by

$$\Upsilon(\tau_w, \tau_s) = \min_{a>0} \mathcal{E}[(1 + a(\xi - \tau_s \varsigma))_+], \quad (40)$$

where ξ and ς are the reference random variables

$$\xi = H\left(\left|\frac{\mathfrak{u}}{\mathfrak{s}/\sqrt{J}}\right| - \tau_w\right)\mathfrak{u}, \quad (41)$$

$$\varsigma = \mathfrak{s}/\sqrt{J}. \quad (42)$$

There, $\mathfrak{u}$ follows a normalized Gaussian distribution, and $\mathfrak{s}$ follows a normalized χ^2-distribution with J degrees of freedom and is statistically independent of $\mathfrak{u}$. The derivation of (40) requires no assumptions on the spatial correlation of the data.

Let us assume for a moment that τ_w and τ_s are given; then, $\Upsilon(\tau_w, \tau_s)$ gives us a probability p. If we find that $\hat{u}[\mathbf{n}] \geq \tau_s \Lambda[\mathbf{n}]$, where $\hat{u}$ has been processed in the wavelet domain based on the threshold τ_w, then we classify the voxel $\mathbf{n}$ of the parameter map as activated with a Type I error probability (considering only this single test) of p. The global Type I error rate corresponds, due to multiple testing, to pV.

Finally, we denote the detected parameter map as

$$d[\mathbf{n}] = H(\hat{u}[\mathbf{n}] - \tau_s \Lambda[\mathbf{n}])\hat{u}[\mathbf{n}]. \quad (43)$$

Parameter Selection

In practice, we want to pick a desired significance level α that corresponds to the global Type I error rate. However, there is an infinity of combinations (τ_w, τ_s) that provide the same probability $\Upsilon(\tau_w, \tau_s) = \alpha/V$. The proposed solution in [63] consists of preferring the most faithful detected parameter map with respect to the unprocessed one. In a mathematical way, we want to minimize the worst-case error between the unprocessed parameter map $u[\mathbf{n}]$ and the detected one $d[\mathbf{n}]$. A sharp (i.e., attainable) bound for this error can be found as

$$|u[\mathbf{n}] - d[\mathbf{n}]| \leq (\tau_w + \tau_s)\Lambda[\mathbf{n}]. \quad (44)$$

So, we should look for the parameters

$$(\tau_w, \tau_s) = \arg\min_{\tau_w, \tau_s} \left\{\tau_w + \tau_s, \text{ subject to } \Upsilon(\tau_w, \tau_s) = \frac{\alpha}{V}\right\}. \quad (45)$$

In the case where we assume $s_w[\mathbf{k}]/\sqrt{J}$ to be the true standard deviation $\sigma_w[\mathbf{k}]$, the optimal values (τ_w, τ_s) correspond to

Only a few wavelet coefficients are needed to efficiently encode the spatial activation patterns.

$$\tau_w = \sqrt{-W_{-1}\left(-2\pi\frac{\alpha^2}{V^2}\right)}, \quad (46)$$

$$\tau_s = \frac{1}{\tau_w}, \quad (47)$$

where $W_{-1}(\cdot)$ is the (-1)-branch of the Lambert W-function; it is the inverse of the function $f(W) = W\exp(W)$. In the general case, the calculation of (40), with its optimization on a, needs to be done by a numerical procedure.

Compensating for Temporal Correlation

In practice, taking into account the temporal correlation of fMRI data plays an important role [64]–[66]. Since the presented techniques deploy the wavelet transform in the spatial domain, we can easily incorporate a stationary model to deal with serial correlations in the wavelet domain. We used the autoregressive model of SPM, for which the parameters are estimated by a restricted maximum-likelihood method [66], which is then incorporated into the estimation of the LM to prewhiten the data. The equivalent degrees of freedom are estimated by the Satterthwaite approximation [64].

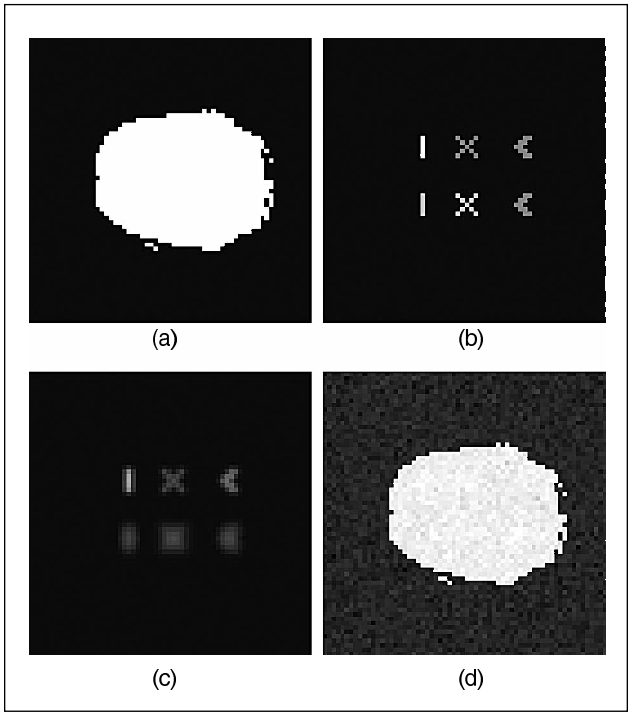

Fig. 6. (a) The software phantom was constructed using the intracranial mask in which activation clusters were embedded. (b) 2-D seed activations were smoothed by a Gaussian filter (FWHM = 1.5 voxel for the upper activations, FWHM = 3 voxel for the lower ones). (c) The parameter map served to generate 80 volumes for which the activated regions followed the time course of Figure 7. (d) An example of a generated slice, corrupted with additive white Gaussian noise.

Examples

As mentioned previously, the wavelet transform can be applied in multiple dimensions. However, while fMRI data consist of a series of 3-D volumes, the transform is mostly applied in two-dimensional (2-D) slice-by-slice to avoid artifacts due to the slice timing effect in the MRI acquisition.

Here, we use the B-spline wavelet transform [67], which can be tuned by the polynomial degree of the scaling function (related to the number of vanishing moments of the wavelet). They also constitute a complete family of semi-orthogonal wavelets, among which the most popular flavors are orthogonal, B-spline (pure B-spline at the synthesis side), and dual B-spline (pure B-spline at the analysis side). The orthogonal transform of degree 1, with two iterations, is denoted as ortho 1/2.

Software Phantom Study

To illustrate some of the differences between the various approaches, including the influence of the choice of the wavelet transform, we perform a software phantom study with known activation clusters. First, we retain from a real dataset a 3-D intracranial mask ($64 \times 64 \times 22$), for which a

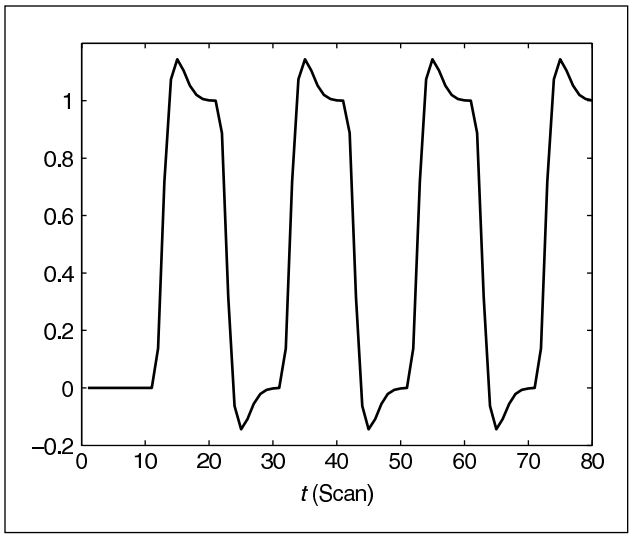

Fig. 7. A time course for the activated voxels of the software phantom (repetition time TR = 3 s, block design with four cycles of ten volumes per epoch, HRF included).

slice is shown in Figure 6(a). This mask serves as a reference background intensity level of 100%. Following Desco et al. [44], we now build a parameter map by first putting six activation seeds, shown in Figure 6(b). The signal levels of these seed activations are (from left to right) 8%, 4%, and 4% for the top ones, and 6%, 6%, and 4%, for the bottom ones. To obtain more realistic activations, the seed parameter map is smoothed by a Gaussian filter. We chose two different filters: the top half was smoothed by a filter with FWHM = 1.5 voxel, while the bottom half was smoothed with FWHM = 3 voxel. This parameter map, shown in Figure 6(c), is then used to produce a time course of 80 volumes, according to a block paradigm of four cycles with ten volumes per epoch, including the HRF as used by SPM (repetition time TR = 3 s), see Figure 7. Each of the volumes is corrupted by additive Gaussian white noise of standard deviation 4%. As an example, the slice containing activation from Scan 15 is shown in Figure 6(d).

We select the desired global significance level to be $\alpha = 5\%$. For the approaches with an error rate different from the Type I error control, α corresponds to the weak Type I error rate. The number of tests taken into account is the number of intracranial voxels (15,923 out of 90,112). The evaluated methods are as follows:

- spatial t-test
- coefficient-wise t-test
- FDR
- recursive testing applied in a subband-by-subband way
- spatio-wavelet method.

Also, each orientation is considered separately. Therefore, the prespecified level for the recursive testing strategy inside each subband and orientation is $\alpha/(3J_w + 1)$.

Let us first take a look at the results of the simple spatial t-test. The parameter map shown in Figure 8(a) corresponds to the voxel-by-voxel least squares solution $u[\mathbf{n}]$ of the LM with the true design matrix. The residual $s[\mathbf{n}]/\sqrt{J}$ of the corresponding parameter is shown in Figure 8(b), while the test statistic $t[\mathbf{n}]$ is depicted in Figure 8(c). Finally, the detected parameter map $d[\mathbf{n}]$ (i.e., voxels that survive the Bonferroni-corrected threshold

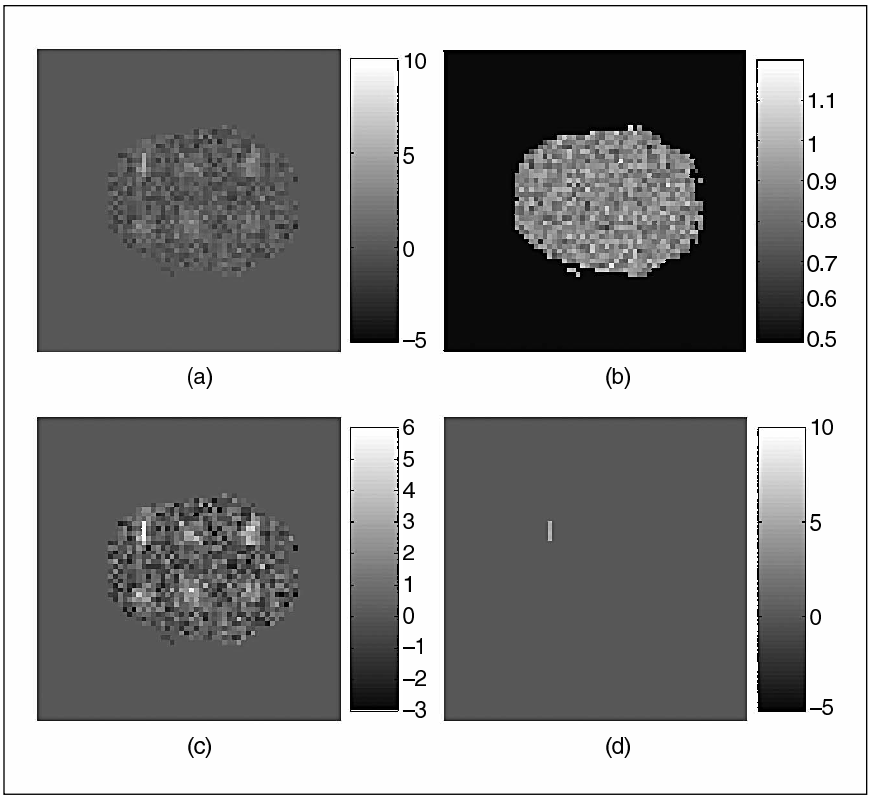

Fig. 8. Results for the software phantom study using the spatial t-test: (a) the estimated parameter map $u(\mathbf{n})$ from the LM; (b) the residual $s[n]/\sqrt{J}$ of the parameter; (c) the test statistic $t(\mathbf{n})$; and (d) the detected parameter map $d(\mathbf{n})$.

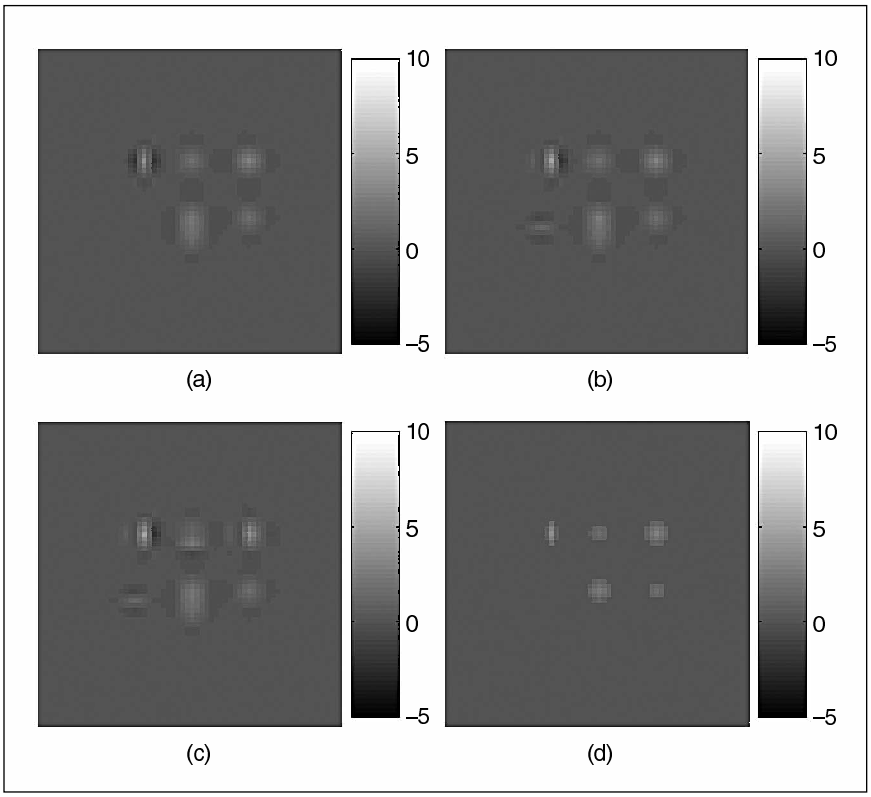

Fig. 9. The resulting parameter map for the software phantom study using wavelet-based methods (orthogonal B-spline wavelet transform, degree 1, two iterations): (a) coefficient-wise t-test; (b) FDR; (c) recursive testing; and (d) spatio-wavelet approach.

$t[\mathbf{n}] > 4.84$) is shown in Figure 8(d). Clearly, the sensitivity of the spatial t-test with Bonferroni correction for multiple testing is very low.

Table 1. An overview of the threshold values for the various methods ($\alpha = 5\%$) in the software phantom study.

	Spatial Threshold	Wavelet Threshold
Spatial t-test	4.84	—
Coefficient-wise t-test	—	4.84
FDR*	—	4.23 (12)
Recursive*	—	3.82 (18)
Spatio-wavelet	0.19	5.25

*The methods indicated by * require a data-dependent threshold. For those, we listed the results for the orthogonal B-spline wavelet transform (degree 1, two iterations) and indicated the total number of retained wavelet coefficients in parentheses.*

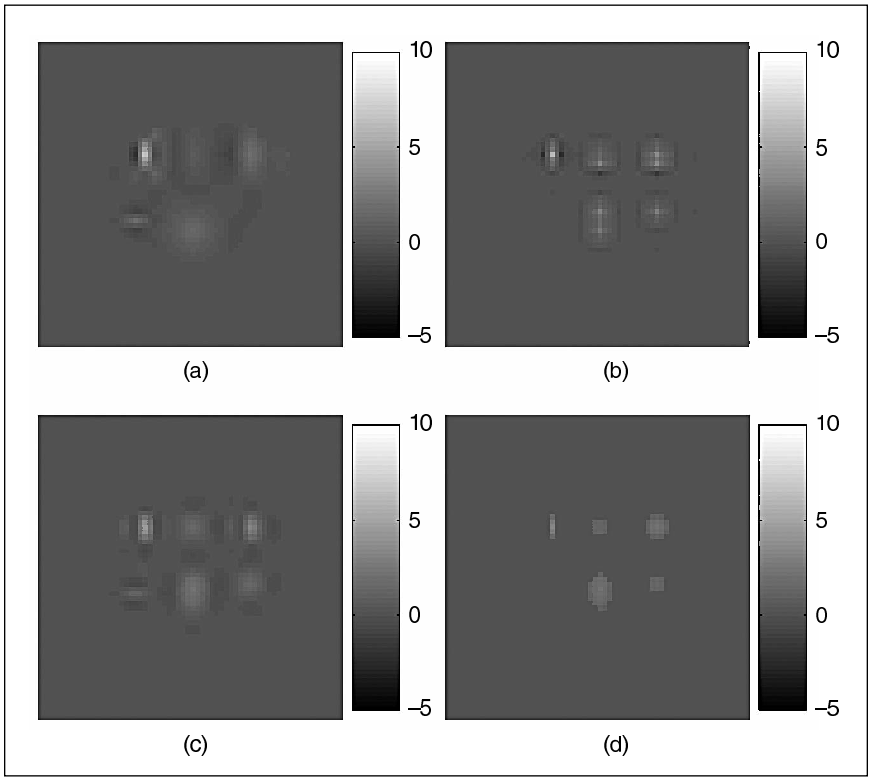

Fig. 10. Resulting parameter map for the software phantom study using wavelet-based methods: (a) FDR (ortho 1/3); (b) FDR (ortho 0/2); (c) FDR (ortho 2/2); and (d) spatio-wavelet approach (ortho 2/2).

We compare the various methods from an SNR point of view. The SNR of the obtained parameter map can be computed since we know the true activations; see Figure 6(c). For techniques with spatial-domain processing, we consider the detected parameter maps as the final result. Specifically, for the spatial t-test, we consider (6), while for the spatio-wavelet method we have (43). For purely wavelet-based techniques, we take the reconstruction from the retained wavelet coefficients.

We show the results for the orthogonal B-spline transform, degree 1, two iterations, in Figure 9. In Table 1, we also list the corresponding threshold values. While the coefficient-wise t-test simply applies to the wavelet domain the same threshold as the spatial t-test, FDR and recursive testing lower this threshold in a data-dependent way. The threshold in the wavelet domain for the spatio-wavelet approach turns out to be the largest, but this method allows for a statistical interpretation in the spatial domain afterwards. FDR and recursive testing, in Figure 9(b) and (c), show activation for the six clusters. The coefficient-wise t-test and the spatio-wavelet approach, in Figure 9(a) and (d), only respond to five clusters. We observe that it does not pay to iterate the transform more than two times. Similar to [44], we find that applying more than two iterations decreases the SNR and the localization of the activation clusters. See for example Figure 10 and Table 2. The degree of the B-spline wavelet transform determines the smoothness of the wavelet and its number of vanishing moments. We compare degree 0 and 2 in Figure 10(b) and (c) for the FDR method. We also include the result for the spatio-wavelet approach for ortho 2/2. The best SNR values are obtained for the FDR and recursive testing methods. Nevertheless, it should be noted that these methods falsely detected for some realizations of the experiment activity elsewhere in the volume.

Another interesting result is the statistical parameter map in the spatial domain, available for some of the techniques, which determines whether a specific voxel can be considered as activated. In Figure 11, we show the statistical parameter map for the spatial t-test, the spatio-wavelet approach, and SPM for two values of the FWHM.

Table 2. SNR values (in db) for the parameter map obtained by the various methods.

	Ortho 1/1	Ortho 1/2	Ortho 1/3	Ortho 0/2	Ortho 2/2
Coefficient-wise t-test	2.31 (5)	2.41 (5)	1.84 (3)	2.38 (5)	3.00 (5)
FDR	1.93 (5)	2.46 (6)	1.82 (5)	2.67 (5)	3.31 (6)
Recursive	1.93 (5)	1.93 (6)	1.16 (6)	2.90 (5)	3.71 (6)
Spatio-wavelet	2.77 (4)	2.36 (5)	1.54 (2)	1.90 (4)	2.74 (5)

The total number of detected activation clusters is listed in parentheses.

In a parametric approach, the temporal behavior of a voxel can be explained by a linear model that is then fitted to the data.

The color maps are scaled according to the range of their respective test statistic. We observe that the results obtained by the spatio-wavelet approach are comparable to those obtained by SPM. Fadili and Bullmore [50] proposed the use of an interesting alternative to evaluate the quality of wavelet techniques in the spatial domain, based on the receiver operating characteristic for a binomial mixture model (which needs to be estimated by an expectation maximization algorithm).

The choice of the wavelet function can also be linked to the Gaussian smoothing and the concept of resolution elements (resels) of SPM. In [68], we proposed the use of the fractional B-spline wavelet transform [69], [70], where we put the pure B-spline at the analysis side of the transform. The close resemblance of the B-spline with the Gaussian function, together with a reinterpretation of resels, provided us with an optimal fractional degree between 1.16 and 1.86, depending on the number of iterations. Also, 2-D wavelet transforms using the quincunx subsampling scheme [71] with fractional orders [72] appear to be interesting candidates to be applied to fMRI data.

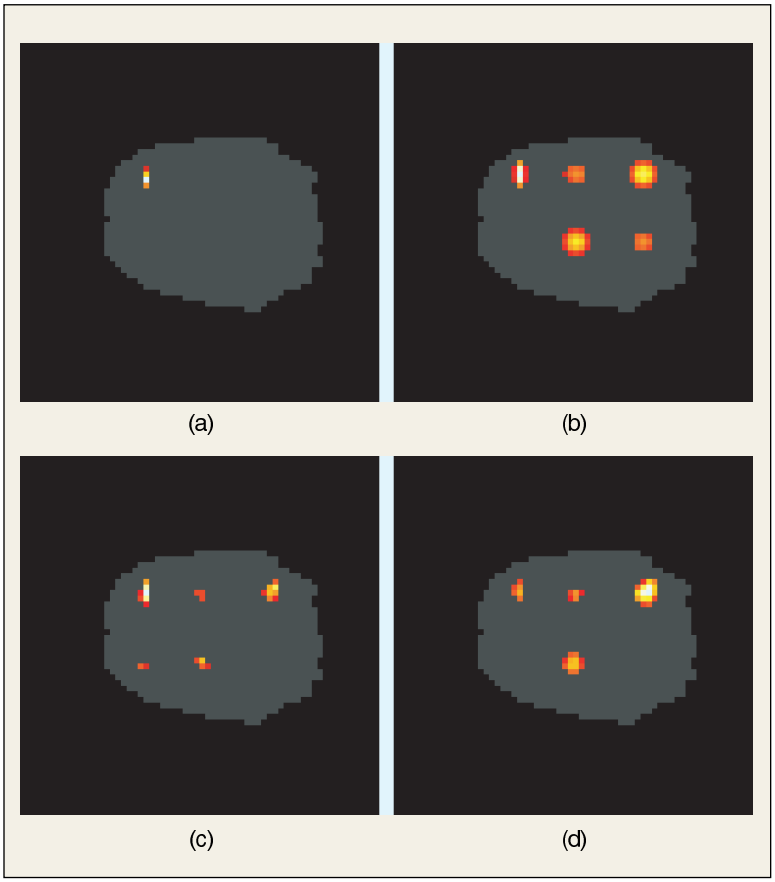

Fig. 11. The statistical parameter map for the software phantom study: (a) voxel-wise spatial *t*-test, (b) spatio-wavelet approach, (c) SPM (FWHM = 1.5 voxel), and (d) SPM (FWHM = 3.0 voxel).

Block-Based Experiment

We also show results for an fMRI experiment with auditory stimulation following a block-based paradigm [73]. Data were acquired on a 2T Siemens Magneton, 7 s repetition time, $64 \times 64 \times 64$ volumes with voxels of physical size 3 mm $\times$ 3 mm $\times$ 3 mm. The total number of volumes was $N = 84$. The setup of the design matrix with hæmodynamic model was done with SPM. The significance level was fixed again at 5%. For all methods, the temporal correlation was taken into account by the weighted least squares solution for the LM proposed by SPM.

For this example, we limited ourselves to the orthogonal B-spline wavelet transform of degree 1.0, two iterations. In Table 3, we listed the threshold values of the various methods. Again, the FDR and recursive testing strategies allow for the lowest threshold values. In Figure 12, we show the parameter map (slice containing the auditory cortex) for the various methods. FDR and recursive testing, in Figure 12(b) and (c), are more sensitive than the coefficient-wise *t*-test and the spatio-wavelet approach, in Figure 12(a) and (d).

The statistical parameter maps are shown in Figure 13. The grayscale background comes from one volume of the original time series. The detected activations are shown in color again.

Table 3. An overview of the threshold values for the various methods ($\alpha = 5\%$) in the block-based experiment.

	Spatial Threshold	Wavelet Threshold
Spatial *t*-test	5.24	—
Coefficient-wise *t*-test	—	5.24
FDR*	—	4.00 (109)
Recursive*	—	3.82 (111)
Spatio-wavelet	0.18	5.53

*The methods indicated by * require a data-dependent threshold. For those, we listed the results for the orthogonal b-spline wavelet transform (degree 1, two iterations) and indicated the total number of retained wavelet coefficients in parentheses.*

Wavelets were primarily applied in medical imaging for denoising in the context of MRI.

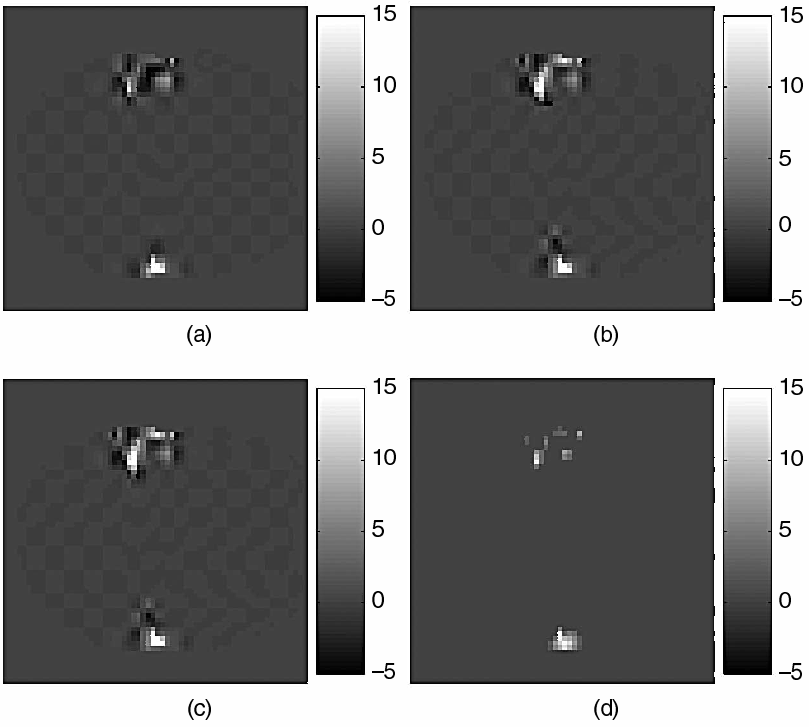

Fig. 12. Resulting parameter map for the block-based experiment using wavelet-based methods (orthogonal B-spline wavelet transform, degree 1, two iterations): (a) coefficient-wise *t*-test; (b) FDR; (c) recursive testing; and (d) spatio-wavelet approach.

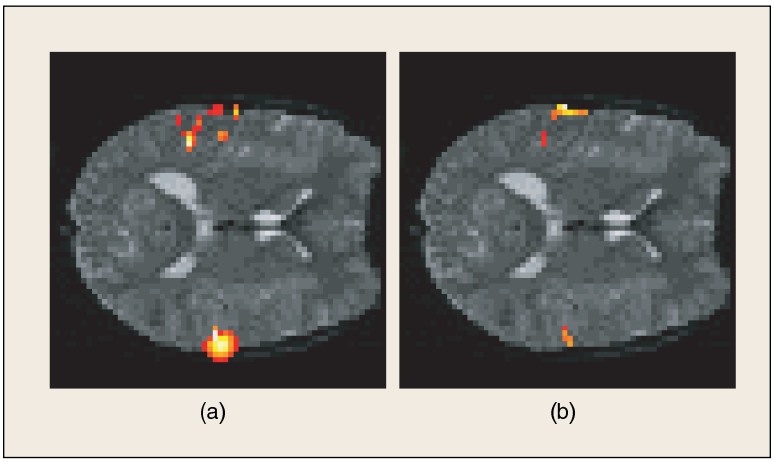

Fig. 13. Statistical parameter map for the block-based experiment: (a) spatio-wavelet approach and (b) SPM (FWHM = 6 mm).

The results of the spatio-wavelet approach can be compared against those obtained by SPM with FWHM = 6 mm. The sensitivity of the wavelet-based approach seems, given the strong Type I error control, a promising result for future research.

Conclusions

Many wavelet-based techniques have been proposed for the analysis of fMRI data. In this article, we gave an overview with unification of the most important techniques that deploy the wavelet transform in the spatial domain. We also included experimental results to illustrate the potential of these techniques. Applying the wavelet transform has two main advantages. First, we obtain a sparse representation of the parameter map, which improves the sensitivity for detecting signal burried in heavy noise. Second, wavelets form a basis, thus the transformation is nonredundant by default. Nonredundancy is a beneficial property to limit the number of tests in the wavelet domain.

Further improvements on wavelet-based techniques can still be expected. For example, recent multidimensional but nonseparable wavelet transforms [74] are potentially interesting to process neuroimaging data. Another important question is how to relax the shift-invariance of the transform (e.g., by deploying the redundant transform) without jeopardizing the statistical significance [75]. Finally, an interesting future research topic is functional connectivity, where wavelets could be deployed jointly in the spatial and temporal domain; e.g., for nonparametric tests using bootstrapping in the wavelet domain [76].

Acknowledgments

This work was supported by the Center for Biomedical Imaging (CIBM) of the Geneva-Lausanne Universities and the EPFL, the foundations Leenaards and Louis-Jeantet as well as by the Swiss National Science Foundation under grant 200020-109415. Finally, we would like to thank Dr. Philippe Thévenaz for his valuable remarks.

Dimitri Van De Ville received the engineering and Ph.D. degrees in computer science from Ghent University, Belgium, in July 1998 and January 2002, respectively. While studying for the Ph.D., he was a research assistant with the Fund for Scientific Research, Flanders, Belgium, at the Medical Image and Signal Processing Group within the Department of Electronics and Information Systems. He currently is with the Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), and the Center for Biomedical Imaging (CIBM). His current research interests include spline and wavelet theory and biomedical signal and imaging applications such as fMRI and microscopy imaging. Since 2002, he is editor of the *Wavelet Digest*, the electronic newsletter of the wavelet community. Van De Ville is associate editor of *IEEE Signal Processing Letters*.

Thierry Blu received the Diplme d'ingnieur from Ecole Polytechnique, France, in 1986, and from Télécom Paris (ENST), France, in 1988. In 1996, he obtained the Ph.D. in electrical engineering from ENST for a study on iterated rational filterbanks, applied to wideband audio coding. He is with the Biomedical Imaging Group at the Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, on leave from France Télécom National Center for Telecommunications Studies (CNET), Issyles-Moulineaux, France. He is currently serving as an associate editor for *IEEE Transactions on Image Processing*. His research interests include (multi)wavelets, multiresolution analysis, multirate filterbanks, approximation and sampling theory, psychoacoustics, optics, and wave propagation.

Michael Unser received the M.S. (summa cum laude) and Ph.D. degrees in electrical engineering in 1981 and 1984, respectively, from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. From 1985–1997, he worked as a scientist with the National Institutes of Health, Bethesda, Maryland. He is now a professor and director of the Biomedical Imaging Group at the EPFL. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He is the author of over 120 published journal papers in these areas. He is the associate editor-in-chief of *IEEE Transactions on Medical Imaging* and the editor-in-chief of the *Wavelet Digest*, the electronic newsletter of the wavelet community. He has acted as associate editor or member of the editorial board for eight other international journals, including *IEEE Signal Processing Magazine*, *IEEE Transactions on Image Processing* (1992–1995), and *IEEE Signal Processing Letters* (1994–1998). He serves as regular chair for SPIE's conference on wavelets, which has been held annually since 1993. He was general cochair for the first IEEE International Symposium on Biomedical Imaging (ISBI'2002), which was held 7–10 July 2002 in Washington, DC. He also chairs the newly created technical committee of the IEEE-Signal Processing Society on Bio Imaging and Signal Processing (BISP). He received the 1995 and 2003 Best Paper Awards and the 2000 Magazine Award from the IEEE Signal Processing Society.

Address for Correspondence: Dimitri Van De Ville, Biomedical Imaging Group, Ecole Polytechnique Fédérale de Lausanne, Biomedical Imaging Group, BM4.140, Station 17, 1015 Lausanne, Switzerland. Phone: +44 21 6935142. Fax: +44 21 6933701. E-mail: dimitri.vandeville@epfl.ch.

References

[1] S. Ogawa and T.M. Lee, "Magnetic resonance imaging of blood vessels at high fields: In vivo and in vitro measurements and image simulation," *Magn. Reson. Med.*, vol. 16, no.1, pp. 9–18, 1990.
[2] N.K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, no. 6843, pp. 150–157, 2001.
[3] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no.7, pp. 674–693, 1989.
[4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.
[5] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, no. 4, pp. 626–638, 1996.
[6] A. Laine, "Wavelets in temporal and spatial processing of biomedical images," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 511–550, 2000.
[7] E. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statistical Methods Med. Res.*, vol. 12, no. 5, pp. 375–399, 2003.
[8] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, "Wavelets and functional magnetic resonance imaging of the human brain," *NeuroImage*, vol. 23, suppl. 1, pp. S234–S249, 2004.
[9] U.E. Ruttimann, N.F. Ramsey, D.W. Hommer, P. Thévenaz, L. Chulhee, and M. Unser, "Analysis of functional magnetic resonance images by wavelet decomposition," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 1995, vol. 1, pp. 633–636.
[10] U. Ruttimann, M. Unser, R. Rawlings, D. Rio, N. Ramsey, V. Mattay, D. Hommer, J. Frank, and D. Weinberger, "Statistical analysis of functional MRI data in the wavelet domain," *IEEE Trans. Med. Imag.*, vol. 17, no. 2, pp. 142–154, 1998.
[11] M. Hilton, T. Ogden, D. Hattery, G. Eden, and B. Jawerth, "Wavelets in biology and medicine," in *Wavelet Denoising of Functional MRI Data*. Boca Raton, FL: CRC Press, 1996, pp. 93–114.
[12] J. Raz and B.I. Turetsky, "Wavelet ANOVA and fMRI," in *Proc. SPIE Conf. Mathematical Imaging: Wavelet Applications Signal Image Processing VII*, San Diego, CA, 1999, vol. 3813, pp. 561–570.
[13] J. Raz, "FDR wavelet shrinkage estimators for inhomogeneous Poisson processed images," in *Proc. SPIE Conf. Wavelets Applications Signal Image Processing VIII*, San Diego (CA), USA, 2000, to be published.
[14] S.M. LaConte, S.-C. Ngan, and X. Hu, "Wavelet transform-based Wiener filtering of event-related fMRI data," *Magn. Reson. Med.*, vol. 44, no. 5, pp. 746–757, 2000.
[15] M.E. Alexander, R. Baumgartner, C. Windischberger, E. Moser, and R.L. Somorjai, "Wavelet domain de-noising of time-courses in MR image sequences," *Magn. Reson. Imag.*, vol. 18, no. 9, pp. 1129–1134, 2000.
[16] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological time series analysis: Resampling methods in time and wavelet domains," *Hum. Brain Map.*, vol. 12, no. 2, pp. 61–78, 2001.
[17] E.T. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M.J. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statistical Methods Med. Res.*, vol. 12, no. 5, pp. 375–399, 2003.
[18] M. Breakspear, M.J. Brammer, and P.A. Robinson, "Construction of multivariate surrogate sets from nonlinear data using the wavelet transform," *Physica D*, vol. 182, no. 1, pp. 1–22, 2003.
[19] O. Friman and C.-F. Westin, "Resampling of fMRI time series," *NeuroImage*, vol. 25, no. 3, pp. 859–867, Apr. 2005.
[20] V. von Tscharner and K.R. Thulborn, "Specified-resolution wavelet analysis of activation patterns from BOLD contrast fMRI," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 704–714, Aug. 2001.
[21] M.J. Fadili and E. Bullmore, "Wavelet-generalised least squares: a new BLU estimator of linear regression models with 1/f errors," *NeuroImage*, vol. 15, no. 1, pp. 217–232, 2002.
[22] G.-A. Hossein-Zadeh, H. Soltanian-Zadeh, and B.A. Ardekani, "Multiresolution fMRI activation detection using translation invariant wavelet transform and statistical analysis based on resampling," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 302–314, Mar. 2003.
[23] F.G. Meyer, "Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 315–322, Mar. 2003.

[24] Y. Shimizu, M. Barth, C. Windischberger, E. Moser, and S. Thurner, "Wavelet-based multifractal analysis of fMRI time series," *NeuroImage*, vol. 22, no. 3, pp. 1195–1202, July 2004.

[25] V. Maxim, L. Sendur, J. Fadili, J. Suckling, R. Gould, R. Howard, and E. Bullmore, "Fractional Gaussian noise, functional MRI and Alzheimer's disease," *NeuroImage*, vol. 25, no. 1, pp. 141–158, Mar. 2005.

[26] B.R. Logan and D.B. Rowe, "An evaluation of thresholding techniques in fMRI analysis," *NeuroImage*, vol. 22, no. 1, pp. 95–108, 2004.

[27] 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," *Hum. Brain Map.*, vol. 2, no. 4, pp. 189–210, 1995.

[28] R. Frackowiak, K. Friston, C. Frith, R. Dolan, and J. Mazziotta, *Human Brain Function*. New York: Academic, 1997.

[29] K. Worsley, S. Marrett, P. Neelin, and A. Evans, "Searching scale space for activation in PET images," *Hum. Brain Map.*, vol. 4, no. 1, pp. 74–90, 1996.

[30] J. Poline, K. Worsley, A. Evans, and K. Friston, "Combining spatial extent and peak intensity to test for activations in functional imaging," *NeuroImage*, vol. 5, no. 2, pp. 83–96, 1997.

[31] J.B. Weaver, X. Yansun, D.M. Healy, and L.D. Cromwell, "Filtering noise from images with wavelet transforms," *Magn. Reson. Med.*, vol. 21, no. 2, pp. 288–295, 1991.

[32] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[33] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, Mar. 1995.

[34] J.C. Wood and K.M. Johnson, "Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR," *Magn. Reson. Med.*, vol. 41, no. 3, pp. 631–636, 1999.

[35] M.E. Alexander, R. Baumgartner, A.R. Summers, C. Windischberger, M. Klarhoefer, E. Moser, and R.L. Somorjai, "A wavelet-based method for improving signal-to-noise ratio and contrast in MR images," *Magn. Reson. Imag.*, vol. 18, no. 2, pp. 169–180, 2000.

[36] S. Zaroubi and G. Goelman, "Complex denoising of MR data via wavelet analysis: Application for functional MRI," *Magn. Reson. Imag.*, vol. 18, no. 1, pp. 59–58, 2000.

[37] A.M. Wink and J.B.T.M. Roerdink, "Denoising functional MR images: A comparison of wavelet denoising and Gaussian smoothing," *IEEE Trans. Med. Imag.*, vol. 23, no. 3, pp. 374–387, June 2004.

[38] J. Aston, R.N. Gunn, R. Hinz, and F. Turkheimer, "Wavelet variance components in image space for spatio-temporal neuroimaging data," *NeuroImage*, vol. 25, no. 1, pp. 159–168, Mar. 2005.

[39] C. Long, E.N. Brown, D. Manoach, and V. Solo, "Spatiotemporal wavelet analysis for functional MRI," *NeuroImage*, vol. 23, no. 2, pp. 500–516, 2004.

[40] M. Breakspear, M.J. Brammer, E.T. Bullmore, P. Das, and L.M. Williams, "Spatiotemporal wavelet resampling for functional neuroimaging data," *Hum. Brain Map.*, vol. 23, no. 1, pp. 1–25, 2004.

[41] B. Whitcher, "Wavelet-based bootstrapping of spatial patterns on a finite lattice," *Computational Statist. Data Anal.*, to be published.

[42] M. Brammer, "Multidimensional wavelet analysis of functional magnetic resonance images," *Hum. Brain Map.*, vol. 6, no. 5-6, pp. 378–382, 1998.

[43] F.E. Turkheimer, M. Brett, J.A.D. Aston, A.P. Leff, P.A. Sargent, R.J. Wise, P.M. Grasby, and V.J. Cunningham, "Statistical modelling of positron emission tomography images in wavelet space," *J. Cerebral Blood Flow Metabolism*, vol. 20, no. 11, pp. 1610–1618, 2000.

[44] M. Desco, J. Hernandez, A. Santos, and M. Brammer, "Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation," *Hum. Brain Map.*, vol. 14, no. 1, pp. 16–27, 2001.

[45] K. Mueller, G. Lohmann, S. Zysset, and Y. von Carmon, "Wavelet statistics of functional MRI data and the general linear model," *J. Magn. Reson. Imag.*, vol. 57, no. 1, pp. 20–30, 2003.

[46] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Royal Statist. Soc. B*, pp. 289–300, 1995.

[47] F. Abramovich and Y. Benjamini, "Wavelets and statistics," in *Thresholding of Wavelet Coefficients as Multiple Hypotheses Testing Procedure*. New York: Springer-Verlag, 1995, pp. 5–14.

[48] F. Abramovich and Y. Benjamini, "Adaptive thresholding of wavelet coefficients," *Computational Statist. Data Anal.*, vol. 22, no. 4, pp. 351–361, 1996.

[49] C.R. Genovese, N.A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, no. 4, pp. 772–786, 2002.

[50] M.J. Fadili and E.T. Bullmore, "A comparative evaluation of wavelet-based methods for multiple hypothesis testing of brain activation maps," *NeuroImage*, vol. 23, no. 3, pp. 1112–1128, 2004.

[51] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[52] R.T. Ogden and E. Parzen, "Change-point approach to data analytic wavelet thresholding," *Statist. Comput.*, vol. 6, no. 2, pp. 93–99, 1996.

[53] X. Shen, H.-C. Huang, and N. Cressie, "Nonparametric hypothesis testing for a spatial signal," *J. Amer. Statistical Assoc.*, vol. 97, no. 460, pp. 1122–1140, 2002.

[54] M. Pavlicová, N. Cressie, and T. Santner, "Using enhanced FDR for simultaneous thresholding of fMRI data," *Proc. Amer. Statist. Assoc.*, Biometrics Section, pp. 2653–2658, 2002.

[55] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rule and Bayes factors," *J. Amer. Statist. Assoc.*, vol. 93, no. 441, pp. 173–179, 1998.

[56] F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Royal Statistical Soc. B*, vol. 60, pp. 725–749, 1998.

[57] J.M. Fadili and E.T. Bullmore, "Wavelet-based approaches for multiple hypothesis testing in activation mapping of functional magnetic resonance images of the human brain," in *Wavelets: Applications in Signal and Image Processing X*, M. Unser, A. Aldroubi, and A. F. Laine, Eds. Bellingham, WA: SPIE, 2003, vol. 5207, pp. 405–416.

[58] J. Berger, B. Boukai, and Y. Wang, "Unified frequentist and Bayesian testing of a precise hypothesis," *Statistical Sci.*, vol. 12, no. 3, pp. 133–160, 1996.

[59] J.D. Storey, "A direct approach to false discovery rates," *J. Royal Statistical Soc. B.*, vol. 64, no. 3, pp. 479–498, 2002.

[60] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[61] J.D. Storey, "The positive false discovery rate: A Bayesian interpretation and the Q-value," *Ann. Statist.*, vol. 31, no. 6, pp. 2013–2035, 2003.

[62] Z. Fu, Y. Hui, and Z.-P. Liang, "Joint spatiotemporal statistical analysis of functional MRI data," in *Proc. ICIP*, 1998, pp. 709–713.

[63] D. Van De Ville, T. Blu, and M. Unser, "Integrated wavelet processing and spatial statistical testing of fMRI data," *NeuroImage*, vol. 23, no. 4, pp. 1472–1485, Dec. 2004.

[64] K.J. Worsley and K.J. Friston, "Analysis of fMRI time-series revisited—again," *NeuroImage*, vol. 2, no. 3, pp. 173–181, 1995.

[65] K.J. Friston, O. Josephs, E. Zarahn, A.P. Holmes, and J.-B. Poline, "To smooth or not to smooth? Bias and efficiency in fMRI time series analysis," *NeuroImage*, vol. 12, no. 2, pp. 196–208, 2000.

[66] K.J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Theory," *NeuroImage*, vol. 16, no. 2, pp. 465–483, 2002.

[67] M. Unser, A. Aldroubi, and M. Eden, "A family of polynomial spline wavelet transforms," *Signal Process.*, vol. 30, no. 2, pp. 141–162, 1993.

[68] D. Van De Ville, T. Blu, and M. Unser, "Wavelets versus resels in the context of fMRI: Establishing the link with SPM," in *Proc. SPIE's Symp. Optical Science Technology: Wavelets X*, San Diego CA, U.S.A., Aug. 2003, vol. 5207, pp. 417-425.

[69] M. Unser and T. Blu, "Fractional splines and wavelets," *SIAM Rev.*, vol. 42, no. 1, pp. 43–67, 2000.

[70] T. Blu and M. Unser, "A complete family of scaling functions: The (α, τ)-fractional splines," in *Proc. 28th IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP'03)*, Hong Kong, Apr. 2003, vol. 6, pp. 421-424.

[71] J. Kovačević and M. Vetterli, "Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for R^n," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 533–555, Mar. 1992.

[72] M. Feilner, D. Van De Ville, and M. Unser, "An orthogonal family of quincunx wavelets with continuously-adjustable order," *IEEE Trans. Image Processing*, vol. 14, no. 4, pp. 499–510, Apr. 2005.

[73] G. Rees and K. Friston, "Single subject epoch (block) auditory fMRI activation data," 1999 [Online]. Available: http://www.fil.ion.ucl.ac.uk/spm/data/

[74] D. Van De Ville, T. Blu, and M. Unser, "Isotropic polyharmonic B-splines: Scaling functions and wavelets," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1798–1813, Nov. 2005.

[75] D. Van De Ville, T. Blu, and M. Unser, "WSPM or how to obtain statistical parametric maps using shift-invariant wavelet processing," in *Proc. 31th Int. Conf. Acoustics, Speech, Signal Processing (ICASSP'06),* Toulouse, France, to be published.

[76] R.S. Patel, D. Van De Ville, and F.D. Bowman, "Determining significant connectivity by 3D spatiotemporal wavelet packet resampling of functional neuroimaging data," *NeuroImage*, to be published.