False Discovery Rate for Wavelet-Based Statistical Parametric Mapping

Dimitri Van De Ville, *Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*—Model-based statistical analysis of functional magnetic resonance imaging (fMRI) data relies on the general linear model and statistical hypothesis testing. Due to the large number of intracranial voxels, it is important to deal with the multiple comparisons problem. Many fMRI analysis tools utilize Gaussian random field theory to obtain a more sensitive thresholding; this typically involves Gaussian smoothing as a preprocessing step. Wavelet-based statistical parametric mapping (WSPM) is an alternative method to obtain parametric maps from non-smoothed data. It relies on adaptive thresholding of the parametric maps in the wavelet domain, followed by voxel-wise statistical testing. The procedure is conservative; it uses Bonferroni correction for strong type I error control. Yet, its sensitivity is close to SPM's due to the excellent denoising properties of the wavelet transform. Here, we adapt the false discovery rate (FDR) principle to the WSPM framework. Although explicit p-values cannot be obtained, we show that it is possible to retrieve the FDR threshold by a simple iterative scheme. We then validate the approach with an event-related visual stimulation task. Our results show better sensitivity with preservation of spatial resolution; i.e., activation clusters align well with the gray matter structures in the visual cortex. The spatial resolution of the activation maps is even high enough to easily identify a voxel that is very likely to be caused by the draining-vein effect.**

***Index Terms*—Bonferroni correction, false discovery rate, functional magnetic resonance imaging, general linear model, parametric hypothesis-driven statistical test, wavelet transform.**

I. Introduction

FUNCTIONAL magnetic resonance imaging (fMRI) has developed into a most versatile technique to map brain activity, with many applications in both clinical and fundamental neurosciences. The backbone of a typical fMRI experiment is the stimulation protocol, the design of which needs to be adapted to the neurological question to answer. The subsequent fMRI data analysis is hypothesis driven and directed towards finding evidence in the spatio-temporal dataset for the presence of a stimulus-related response [1]. Typical fMRI analysis tools deploy general linear models to extract parametric maps that are then subjected to statistical hypothesis testing. In order to be sound statistically, the hypothesis testing procedure has to be corrected for the fact one is performing multiple comparisons. This issue is important due to the abundance of voxels to be tested. In order to control the global (family-wise) false positives rate, one can apply the so-called Bonferroni correction. This procedure is foolproof statistically, but it is often too conservative because it does not take into account the spatial correlation that is present in these data.

A popular alternative method for dealing with the multiple comparisons problem is thresholding based on Gaussian random field theory (GRFT) [2], [3], which is implemented in many software tools; e.g., the statistical parametric mapping (SPM) approach [4], [5]. The key component of GRFT is the preprocessing of the data by smoothing. It has the advantage of improving sensitivity, albeit at the risk of losing spatial resolution.

Data-driven methods that do not (or only partially) rely on prior knowledge have been proposed as well. The most popular ones are subspace methods, such as principal components analysis [6] and independent component analysis [7]–[9]. Next to revealing unmodeled trends, they can also be used to denoise fMRI data by stripping components.

The wavelet transform [10] is a powerful tool that has also found its way into biomedical signal and image processing, including fMRI data analysis; several overviews are available [11]–[15]. Within the context of parametric mapping, the multiresolution and energy compaction properties of the wavelet transform are advantageously exploited to "statistically denoise" the activation maps by thresholding the wavelet coefficients [16]–[18]. One of the difficulties associated with such probabilistic shrinkage is how to map the statistics back into the spatial domain in order to be able to declare a particular *voxel* active. 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 [16]) or re-testing in the spatial domain without taking into account the effect of the initial test in the wavelet domain [19].

In our recent work [20], [21], we have proposed the wavelet-based statistical parametric mapping (WSPM) framework that combines adaptive denoising of the parametric maps in the wavelet domain with statistical testing at the voxel-level. The type I error rate is controlled by Bonferroni correction, which guarantees high specificity. At the same time, we reported comparable sensitivity and lesser spatial bias than SPM [21].

False discovery rate (FDR) is another appealing solution to the multiple comparisons problem [22], [23]. The basic principle is to control the expected *proportion* of false positives; i.e., to become more permissive on the number of false positives

Manuscript received March 16, 2008; revised September 22, 2008. Current version published January 23, 2009. This work was supported by the Swiss National Science Foundation (MU) under Grant 200020-109415, the Center for Biomedical Imaging (DVDV) of the Geneva-Lausanne Universities, the EPFL, and by the foundations Leenaards and Louis-Jeantet. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Z. Jane Wang.

The authors are with the Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2008.2007762

1932-4553/$25.00 © 2008 IEEE

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

when more voxels are declared active. This method has enjoyed a large success in a variety of areas, including neuroimaging [24]. In principle, it can be combined with any of the aforementioned methodologies. Most notably, Raz [25], [26] deployed the FDR principle for the coefficient-wise testing in the wavelet domain. Srikanth *et al.* [27] estimated the FDR based on statistical resampling [28]; their method aims at a spatial-domain threshold after wavelet denoising.

As we shall see in this paper, we can also apply the FDR principle to the WSPM framework, but it requires some specific adaptation. For instance, we need to take into account the fact that statistical testing in WSPM is based on a bound for the null hypothesis rejection probability, meaning that we do not have direct access to the voxels' p-values. The main difficulty is that WSPM relies on the interplay of two thresholds (in the wavelet and spatial domains, respectively), the optimal values of which depend on the desired significance level. We show that these obstacles can be overcome and that it is theoretically possible to control FDR; in practice, this can be implemented by a simple (and fast) iterative scheme. Our experiments demonstrate that the FDR procedure enhances the sensitivity of WSPM, while retaining its intrinisic high-spatial resolution.

The paper is organized as follows. In Section II, we briefly review the discrete wavelet transform (DWT). Next, in Section III, we highlight the main elements of wavelet-based methods: the general linear model, coefficient-wise thresholding in the wavelet domain, and the spatio-wavelet framework. In Section IV, we briefly revisit the FDR principle and we adapt this strategy to WSPM. Finally, we demonstrate the feasibility of our approach in Section V for experimental data of an event-related visual stimulation task.

II. The Discrete Wavelet Transform

The DWT is a powerful tool for multiresolution signal analysis [10], [29]. The transform of a signal $v(x)$ is a basis decomposition into a sum of shifted and dyadically-scaled versions of the (bandpass) wavelet functions ψ and the scaling function φ. Mathematically, we write

$$v(x) = 2^{-\frac{J_w}{2}} \sum_k c_{-J_w}[k]\varphi(2^{-J_w}x - k) + \sum_{j=-J_w}^{-1} 2^{\frac{j}{2}} \sum_k d_j[k]\psi(2^j x - k)$$

for a decomposition of J_w levels. The residual after J_w decomposition levels is represented by the integer shift of the (low-pass) scaling function $\varphi(2^{-J_w}x - k)$. The low-pass coefficients and detail (or wavelet) coefficients are given by

$$c_{-J_w}[k] = \left\langle v(\cdot), 2^{-\frac{J_w}{2}}\tilde{\varphi}(2^{-J_w}\cdot - k)\right\rangle \quad (1)$$

$$d_j[k] = \left\langle v(\cdot), 2^{\frac{j}{2}}\tilde{\psi}(2^j\cdot - k)\right\rangle \quad (2)$$

respectively. The analysis functions $\tilde{\varphi}$ and $\tilde{\psi}$ are the duals of φ and ψ; they satisfy the biorthogonality property $\langle \psi(2^j\cdot - k_1), \tilde{\psi}(2^j\cdot - k_2)\rangle = \delta_{k_1-k_2}$. In practice, the coefficients are calculated using a fast iterated filterbank algorithm.

In multiple dimensions, the transform can be applied in a separable way, leading to $2^{N_d} - 1$ different tensor-product wavelets in N_d dimensions. We use the shorthand notation

$$v(\mathbf{x}) = \sum_{\mathbf{k}} v_w[\mathbf{k}]\psi_{\mathbf{k}}(\mathbf{x}) \quad (3)$$

where $v_w[\mathbf{k}]$ can be both low-pass or wavelet coefficients at all different scales and orientations, and where $\psi_{\mathbf{k}}$ denote the corresponding basis functions, irrespective of their type.

III. Wavelet-Based fMRI Data Analysis

We review the various elements for wavelet-based hypothesis-driven analysis of fMRI data. First, we describe the general linear model (GLM) and voxel-wise statistical testing. Next, we summarize the essential properties of the wavelet-based analysis methods.

A. The General Linear Model

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 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 the parametric approach, the temporal behavior of a voxel is described by a GLM that is fitted to the data [4]. Specifically, one describes the model as

$$\mathbf{v}[\mathbf{n}] = \mathbf{X}\mathbf{y}[\mathbf{n}] + \mathbf{e}[\mathbf{n}], \quad \text{for each } \mathbf{n} \quad (4)$$

where $\mathbf{X}$ is the $N \times L$ design matrix that contains the L regressors, $\mathbf{y}$ is the $L \times 1$ parameter vector, and $\mathbf{e}[\mathbf{n}]$ is the (random) error.

Given the observed data $\mathbf{v}[\mathbf{n}]$, the classical least-squares estimate of the parameters of the GLM is given by $\bar{\mathbf{y}}[\mathbf{n}] = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{v}[\mathbf{n}]$. This estimate is optimal provided that the error component is independently and identically Gaussian distributed. The corresponding residual is $\bar{\mathbf{e}}[\mathbf{n}] = \mathbf{v}[\mathbf{n}] - \mathbf{X}\bar{\mathbf{y}}[\mathbf{n}]$. Next, the information of interest is extracted from $\bar{\mathbf{y}}[\mathbf{n}]$ by a contrast vector $\mathbf{c}$. At this stage, we obtain two scalar values for each voxel, as follows:

$$u[\mathbf{n}] = \mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}}[\mathbf{n}] \quad (5)$$

$$s^2[\mathbf{n}] = \bar{\mathbf{e}}[\mathbf{n}]^{\mathrm{T}}\bar{\mathbf{e}}[\mathbf{n}]\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}. \quad (6)$$

Under the null hypothesis (only noise), the contrast variable $u[\mathbf{n}]$ and its estimated variance $s^2[\mathbf{n}]/J$, follow a Gaussian and a χ^2 distribution (with $J = N_t - \mathrm{rank}(\mathbf{X})$ degrees of freedom), respectively [1], [30].

Now, hypothesis testing is performed to determine whether or not the contrast of interest for the voxel $\mathbf{n}$ is significant; that is, if the mean value is zero or above:

$$\mathcal{H}_0 : E[u[\mathbf{n}]] = 0$$
$$\mathcal{H}_1 : E[u[\mathbf{n}]] > 0. \quad (7)$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

Under the null hypothesis, the test statistic, derived from $u[\mathbf{n}]$ and $s^2[\mathbf{n}]$

$$t[\mathbf{n}] = \frac{u[\mathbf{n}]}{\sqrt{s^2[\mathbf{n}]/J}}, \tag{8}$$

follows a normalized Student t distribution with J degrees of freedom. We control the type I error (false positive rate) at $\alpha = \mathrm{Prob}[t \geq \tau]$, which leads to the null hypothesis rejection (and thus detection of activation) when $t[\mathbf{n}] \geq \tau$.

In fMRI data analysis, hypothesis testing is applied to a large number of voxels, which increases the global type I error. An easy, but very conservative, solution is Bonferroni correction: the significance level of an individual test is maintained at $\alpha_B = \alpha/V$, where α is now the desired global type I error, and V is the number of tests (typically, the number of intracranial voxels). This correction guarantees strong type I error control, but it often reduces sensitivity as it does not take into account the spatial correlation between voxels, which is potentially present in fMRI data [31].

B. Wavelet Processing by Coefficient-Wise Testing

Following the lead of Ruttimann *et al.* [16], [32], the complete hypothesis testing framework can be conveniently transposed into the *spatial* wavelet domain[1]. Since then, the method has been refined and extended by many other researchers [17], [33]–[35]. In the present setting, we apply the linear model (4) to the time-course of every coefficient (after applying the spatial wavelet transform to each volume $v[\mathbf{n}; t]$). The coefficient-wise GLM then reads

$$\mathbf{v}_w[\mathbf{k}] = \mathbf{X}\mathbf{y}_w[\mathbf{k}] + \mathbf{e}_w[\mathbf{k}], \quad \text{for each } \mathbf{k} \tag{9}$$

where the vector $\mathbf{v}_w[\mathbf{k}]$ is formed out of the wavelet coefficients $v_w[\mathbf{k}; t]$. The estimation process is essentially the same as before: regression weights and residual are obtained as $\bar{\mathbf{y}}_w[\mathbf{k}] = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{v}_w[\mathbf{k}]$ and $\bar{\mathbf{e}}_w\mathbf{k} = \mathbf{v}_w[\mathbf{k}] - \mathbf{X}\bar{\mathbf{y}}_w[\mathbf{k}]$, respectively. The contrast of interest is then extracted as

$$u_w[\mathbf{k}] = \mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}}_w[\mathbf{k}], \tag{10}$$

$$s_w^2[\mathbf{k}] = \bar{\mathbf{e}}_w^{\mathrm{T}}[\mathbf{k}]\bar{\mathbf{e}}_w[\mathbf{k}]\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}. \tag{11}$$

The hypotheses of (7) are modified into

$$\mathcal{H}_0 : E[u_w[\mathbf{k}]] = 0$$
$$\mathcal{H}_1 : E[u_w[\mathbf{k}]] \neq 0. \tag{12}$$

The test is two-sided because the contribution of a wavelet coefficient to the reconstruction of a given voxel can be either positive or negative. The coefficient-wise statistical testing then evaluates the t-value of each wavelet coefficient

$$t_w[\mathbf{k}] = \frac{u_w[\mathbf{k}]}{s_w[\mathbf{k}]/\sqrt{J}}. \tag{13}$$

[1]The wavelet transform is performed along the spatial dimensions.

The test procedure checks whether $|t_w[\mathbf{k}]| > \tau$, with $\alpha = \mathrm{Prob}[|t| \geq \tau]$. The wavelet coefficients $u_w[\mathbf{k}]$ that survived the test can then be reconstructed as

$$\tilde{u}[\mathbf{n}] = \sum_{\mathbf{k}} \underbrace{H\left(|t_w[\mathbf{k}]| - \tau\right) u_w[\mathbf{k}]}_{\tilde{u}_w[\mathbf{k}]} \psi_{\mathbf{k}}(\mathbf{n}) \tag{14}$$

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

$$H(t) = \begin{cases} 0, & \text{when } t < 0, \\ 1, & \text{otherwise.} \end{cases} \tag{15}$$

The number of coefficients that is tested in the wavelet domain is at least equal to the number of tests in the spatial domain. Again, Bonferroni correction can be applied by setting the false positive rate (for an individual test) to α/V, where V corresponds to the number of intracranial voxels. (The latter is approximately equal to the number of intracranial wavelet coefficients.) Due to the decorrelating properties of the wavelet transform, Bonferroni correction should be closer to optimal.

C. WSPM Approach: Wavelet Processing Followed by Spatial Testing

The specificity of WSPM[2] is to combine wavelet domain adaptive denoising with voxel-wise statistical testing in an integrated fashion [20], [21]. The spatio-wavelet framework uses Bonferroni correction for multiple hypothesis testing, thus ensuring strong type I error control. We give a brief overview of the different steps of this method.

Here too, wavelet coefficients $u_w[\mathbf{k}]$ are hard thresholded based on the comparison $|t_w[\mathbf{k}]| > \tau_w$. As opposed to coefficient-wise testing, the threshold τ_w is not obtained as the result of a statistical test, but kept as a general parameter of the algorithm. After reconstruction of $\tilde{u}[\mathbf{n}]$ from the thresholded coefficients, we perform a hypothesis test in the spatial domain to determine whether the voxel $\mathbf{n}$ is activated and thus has a nonzero mean

$$\mathcal{H}_0 : E[\tilde{u}[\mathbf{n}]] = 0 \tag{16}$$

$$\mathcal{H}_1 : E[\tilde{u}[\mathbf{n}]] > 0. \tag{17}$$

The main result of the integrated framework [20] is that the statistical test amounts to checking whether or not $\tilde{u}[\mathbf{n}] \geq \tau_s \Lambda[\mathbf{n}]$, where τ_s is the threshold in the spatial domain and $\Lambda[\mathbf{n}]$ is the reconstruction of the values $s_w[\mathbf{k}]/\sqrt{J}$ by a modified inverse DWT algorithm:

$$\Lambda[\mathbf{n}] = \sum_{\mathbf{k}} \frac{s_w[\mathbf{k}]}{\sqrt{J}} |\psi_{\mathbf{k}}(\mathbf{n})|. \tag{18}$$

The probability of a false detection under the null hypothesis can be bound as

$$\mathrm{Prob}[\tilde{u}[\mathbf{n}] \geq \tau_s \Lambda[\mathbf{n}]] \leq \Upsilon(\tau_w, \tau_s) = \alpha_B \tag{19}$$

[2]This approach has been implemented as a toolbox for SPM2/SPM5, called "WSPM: Wavelet-Based SPM", and is available at http://bigwww.epfl.ch/wspm.

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

where $\Upsilon(\tau_w, \tau_s) = \min_{a>0} E[(1 + a(\xi_{\tau_w} - \tau_s \varsigma))_+]$ is a data-independent function of the parameters τ_w and τ_s, with $(x)_+ = \max(0, x)$. Here, ξ_{τ_w} follows a truncated normalized Gaussian distribution and ς a χ-distribution with J degrees of freedom—see [20]. The bound can be chosen equal to α_B, the desired significance level after Bonferroni correction. However, there is an infinite number of combinations (τ_w, τ_s) that yields the same probability bound α_B. In [20], we proposed the solution that minimizes the worst-case error between the unprocessed ($u[\mathbf{n}]$) and detected parameter map, to be equivalent to minimizing the sum $\tau_w + \tau_s$, subject to $\Upsilon(\tau_w, \tau_s) = \alpha_B$. For $N_t > 50$, as is typically the case for fMRI, the threshold can be determined as

$$\tau_w = \sqrt{-W_{-1}(-2\pi\alpha_B^2)}, \quad \tau_s = 1/\tau_w \tag{20}$$

where W_{-1} is the -1-branch of the Lambert W-function; it is the inverse of the function $f(W) = W\exp(W)$.

The WSPM toolbox also includes a couple of further extensions [21], such as spatial bias reduction and combining multiple transforms for improved shift-invariance. For simplicity, we consider the original framework as summarized here. However, the results of this paper can readily be adapted to these extensions as well.

IV. False Discovery Rate for WSPM

A. The Principle of FDR

Although controlling the family-wise type I error rate has an easy interpretation (i.e., we have a chance α of finding any false positive in the whole volume), it is often considered as being too conservative. FDR is a popular alternative that controls the proportion of false positives to total positives [22]. One attractive feature is that is maintains the same type I error rate as Bonferroni correction when there is no more than one detection; hence the so-called "weak type I error control" property. Mathematically, FDR is defined as the expectation

$$\mathrm{FDR} = E\left[\frac{D_F}{D}\right] \tag{21}$$

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. In practice, the FDR can be controlled by the observed proportion

$$E\left[\frac{D_F}{D}|D = i\right] = \frac{E[D_F|D = i]}{i} = \frac{p^{(i)}V}{i} \tag{22}$$

where $p^{(i)}$, $i = 1, \ldots, V$, are the sorted p-values, and $D = i$ the total positives. Therefore, keeping $\mathrm{FDR} \leq \alpha$ requires

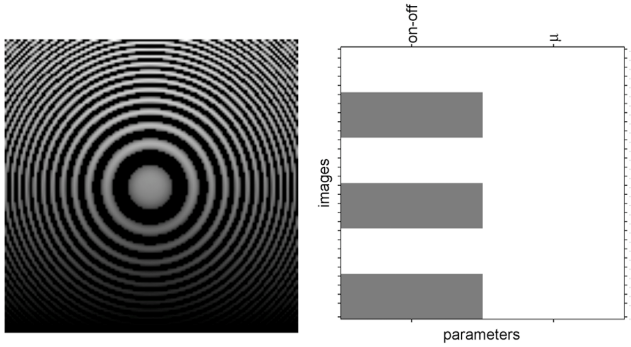
$$p^{(i)} \leq i\frac{\alpha}{V}. \tag{23}$$

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

1) The elements to be tested are sorted according to increasing p-values.
2) The index i is determined by the largest p-value $p^{(i)}$ for which $p^{(i)} \leq i\alpha/V$.
3) The FDR threshold T can be determined by $p^{(i)} = \mathrm{Prob}[t \geq T]$.

The standard FDR approach assumes test statistics that are positively dependent [23]; this is satisfied when noise is Gaussian and nonnegatively correlated—a reasonable assumption for fMRI datasets [24]. In the other case, one can add a factor $(\sum_{k=1}^{V} 1/k)^{-1}$ to the right-hand side of (23) to compensate for the dependency between voxels [23].

The FDR principle has already been applied to the coefficient-wise hypothesis testing scenario of Section III-B by [18], [25], [26], but not yet to WSPM, which requires some special adaptation.



(a) (b)

Fig. 1. (a) Contrast image and (b) design matrix that are used for generating the synthetic dataset.

B. FDR for WSPM

The WSPM framework is based on a bound for the null hypothesis rejection probability, and consequently it does not give direct access to the p-values of the voxels. In addition, the two thresholds—τ_w and τ_s—are distinct and they even vary in inverse proportion. However, the significance level α of the thresholding procedure $\tilde{u}[\mathbf{n}]/\Lambda[\mathbf{n}] \geq \tau_s$ is clearly an upper bound for the true p-value of a detected voxel. Therefore, we define the "critical p_*-value" as the bound when the voxel just survived the procedure, that is, $p_* = \Upsilon(\tau_w, \tau_s)$ for which we have $\tilde{u}[\mathbf{n}]/\Lambda[\mathbf{n}] = \tau_s$.

The FDR procedure can now be based on the critical p_*-values. First, we sort the voxels according to increasing $p_*^{(i)}$, with associated thresholds $(\tau_w^{(i)}, \tau_s^{(i)})$. Controlling the FDR at α requires to find the largest index i for which the following equation holds:

$$p_*^{(i)} = \Upsilon\left(\tau_w^{(i)}, \tau_s^{(i)}\right) \leq i\frac{\alpha}{V}. \tag{24}$$

Note that the FDR is bounded by considering the critical p_*-values, even if the order of the voxels for increasing (unknown) p-values, $p^{(l)}$, would be different from $p_*^{(i)}$. Indeed, $p_*^{(i)}$ also bounds $p^{(l)}$

$$p^{(l)} \leq p_*^{(i)}, \quad l = 1, \ldots, i. \tag{25}$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

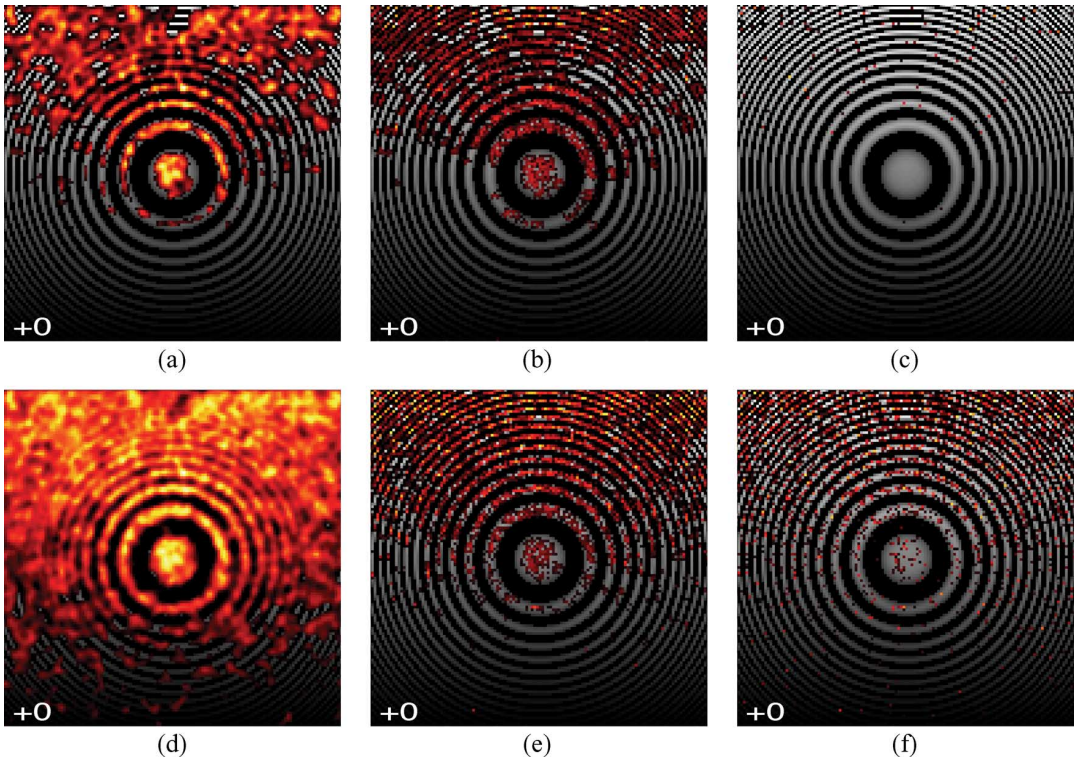


Fig. 2. Results for the synthetic dataset. (a) 5% FWE for SPM, (b) 5% Bonferroni for WSPM, (c) 5% Bonferroni for GLM. (d)–(f) 5% FDR for (d) SPM, (e) WSPM, (f) GLM.

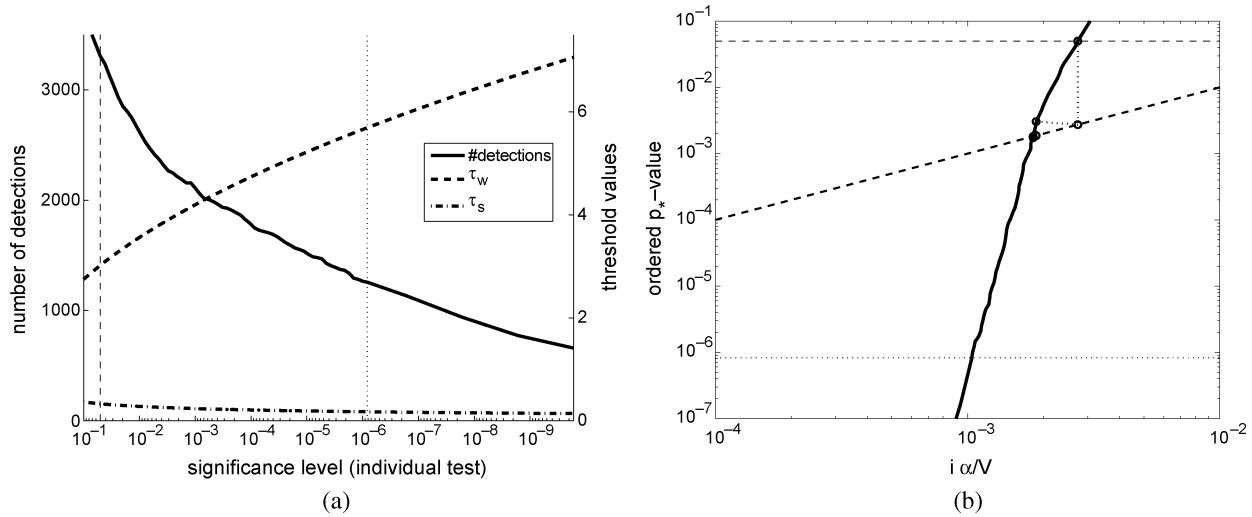


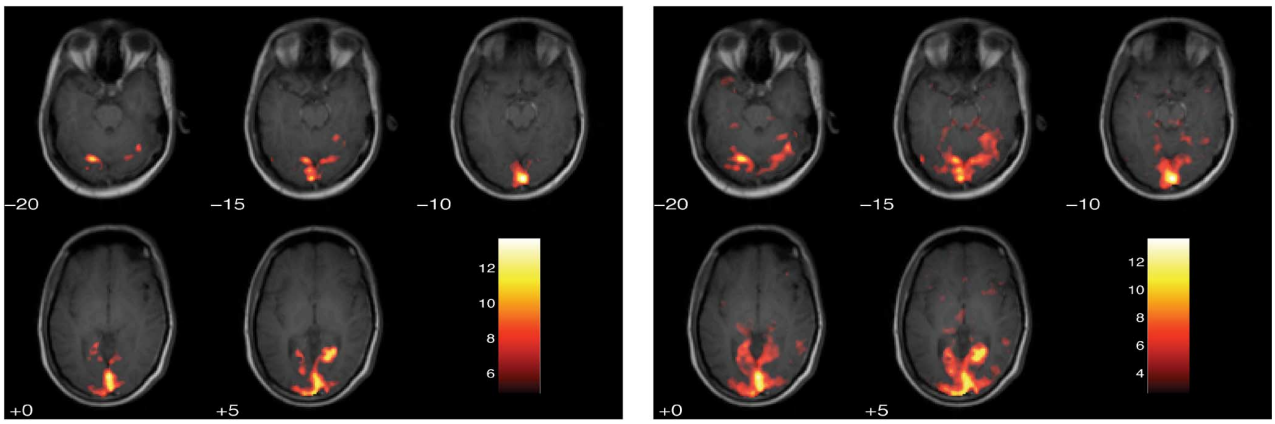
Fig. 3. (a) Number of detections and threshold values as a function of the significance level α of the individual hypothesis test. The number of detections monotonously decreases as the confidence gets larger. The vertical lines indicate the 5% significance level without correction for multiple comparisons ($- - -$) and with Bonferroni correction ($\cdots$), respectively. (b) The $p - p$ plot: the critical p-values as function of the relative number of detections. The dashed line indicates the FDR. The horizontal lines correspond to uncorrected and Bonferroni corrected significance levels. The trajectory of the iterative algorithm to find the FDR threshold is also shown.

Retrieving the critical thresholds $(\tau_w^{(i)}, \tau_s^{(i)})$ for every voxel would be a tedious task. However, there is a monotonous relationship between critical p_*-values and number of voxels that survived testing; i.e., when the significance α increases, the number of detected voxels increases as well.

This observation allows us to deploy the following iterative procedure to find the FDR threshold.

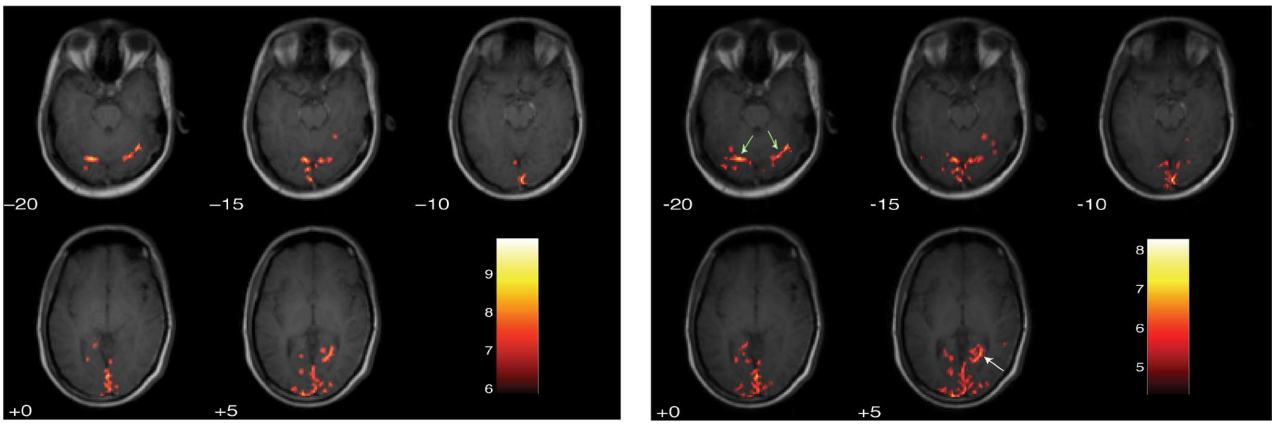
1) The thresholding procedure is applied using the thresholds (τ_w, τ_s) according to the significance level α without Bonferroni correction. This would correspond to the most permissive case of FDR thresholding where every voxel would be detected, $i = V$.
2) The number of detections i is determined to establish the type I error rate $i\alpha/V$ that is equivalent to FDR at α. The

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 4. Results using SPM. (a) 5% FWE. (b) 5% FDR.



(a) (b)

Fig. 5. Results using WSPM. (a) 5% Bonferroni correction. (b) 5% FDR.

updated thresholds (τ_w, τ_s) are obtained and the thresholding procedure is re-applied.

3) Step 2 is repeated until convergence.

V. Experimental Results

A. Synthetic Data

We generated a synthetic fMRI dataset that allowed us to study the behavior of the methods for varying signal-to-noise ratio (SNR) and feature size. To that aim, we generated the 2-D zoneplate image according to the formula

$$\frac{1}{2}\left(1+\cos\left(\frac{\pi}{\sqrt{2}N}\left(x_1^2+x_2^2\right)\right)\right)$$

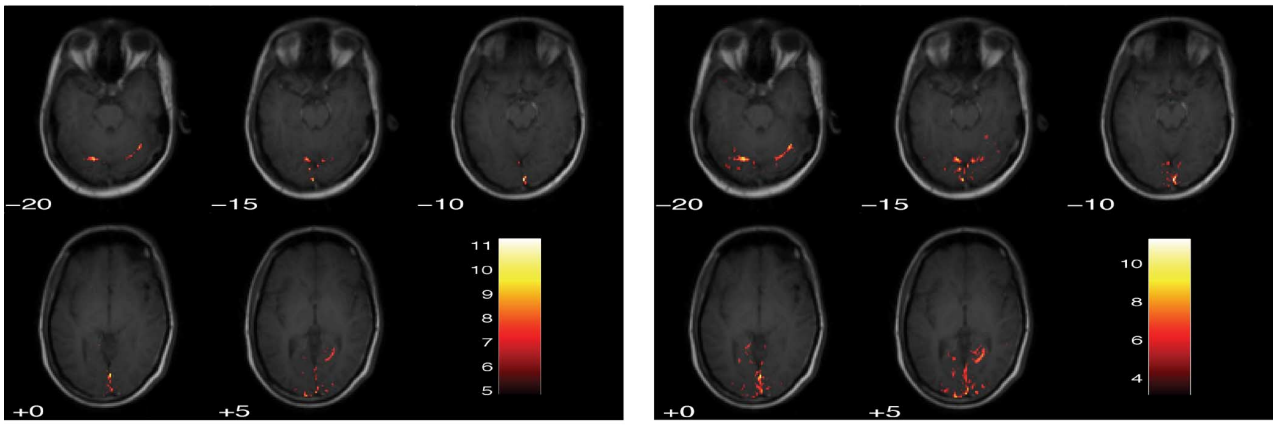
where $x_1, x_2 = -(N/2), -(N/2), \ldots, (N/2)-1$. Next, the image is thresholded at 0.75 to create fine structures. Finally, the contrast's intensity is multiplied in the top-down direction by a linear ramp. The final contrast image is shown in Fig. 1(a), for $N = 128$. We generated 60 images with Gaussian noise of mean 8 and standard deviation 1 and added the contrast image modulated by the block paradigm of Fig. 1(b); i.e., three consecutive on and off blocks of ten scans. This way, we obtained a synthetic 2-D dataset with varying SNR (top to bottom) and feature size (excentricity).

Data processing was performed using SPM2 software.[3] Smoothing was performed for three pixel units FWHM for GRFT-based analysis. The exact generative model of the block paradigm was entered into the GLM. For WSPM, the 2-D separable linear orthogonal B-spline wavelet transform [36] with two decomposition levels was applied. With WSPM, statistical maps are reported as $\tau_w + \tilde{u}[\mathbf{n}]/\Lambda[\mathbf{n}]$ for those voxels that survived testing [21]. As a reference point, we also applied the voxel-wise GLM approach to non-smoothed data.

In Fig. 2(a), we show the results for SPM at 5% FWE, overlaid on the original contrast image in grayscale. The fine features towards the border of the image are not well resolved due to the effect of smoothing. Similarly, when applying SPM at 5% FDR, in Fig. 2(d), the sensitivity increases but only the large rings at the center are well separated. The results for WSPM are shown

[3]Available at: http://www.fil.ion.ucl.ac.uk/spm/.

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 6. Results using GLM with nonsmoothed data. (a) 5% Bonferroni correction. (b) 5% FDR.

in Fig. 2(b) and (e) for 5% Bonferroni and 5% FDR, respectively. While the lower threshold τ_w for WSPM-FDR increases the number of wavelet coefficients to be included in the reconstruction, close comparison of (b) and (e) reveals that this effect does not only lead to more detected voxels, but also to more refined patterns. Finally, the results for the voxel-wise GLM approach, in Fig. 2(c) and (f), show the beneficial effect of FDR. However, the continuity on the rings is much less conserved because spatial correlation is not exploited.

B. Measured Data

We conducted a single-subject fMRI study (healthy volunteer, right-handed male, 31y) at the Geneva University Hospital on the Siemens Magnetom TrioTim 3T scanner. The anatomical data was acquired by an MPRAGE T1 scan (coronal orientation, TR 2.5s, TI 1.1s, 224 slices, 1 mm thickness, in-plane resolution: 0.9 mm $\times$ 0.9 mm). The functional data was acquired by EPI T2$*$ scans (axial orientation, $\mathrm{TR/TE} = 1\ \mathrm{s}/30\ \mathrm{ms}$, 435 scans, 16 slices, 5-mm thickness, in-plane resolution: 1.9 mm $\times$ 1.9 mm). Visual stimulation was applied in an event-related paradigm: different images were short-flashed during 50 s with fixed inter-stimulus time (17.5 s); during rest a small cross-hair was shown in the middle of the field-of-view to retain the subject's attention.

Again, data processing was performed using SPM2, including realignment and 6-mm FWHM smoothing (for GRFT-based analysis). The GLM contained temporal and dispersion derivatives to cope with subject-dependent variations of the hemodynamic response function. Temporal correlation was taken into account by SPM's autoregressive model. In this study, we only focus on the combined contrast that extracted the regressor related to any image stimulation; in total, there were 24 visual stimuli.

The wavelet transform can be applied in multiple dimensions. However, while fMRI data consist of a series of 3-D volumes, we applied the transform in 2-D slice-by-slice due to the rather large slice thickness. We used the orthogonal B-spline wavelet transform of degree 1, with a single decomposition level. The same GLM and temporal modeling as SPM are deployed by WSPM to extract regressor weights.

We analyzed the behavior of WSPM as a function of the significance level. In Fig. 3(a), we plot the number of detections for various significance levels, together with the associated threshold values τ_w and τ_s. The vertical lines indicate the 5% significance with and without Bonferroni correction. We note that the number of detections decreases monotonously as the significance goes up. The wavelet threshold τ_w, which is applied to wavelet-domain t-values, is most influenced by varying signifance, while the spatial threshold τ_s decreases only slowly. A more interesting way to interpret these results is in the form of the $p-p$ plot of Fig. 3(b). We convert the number of detections i into the equivalent type I error $i\alpha/V$ for FDR, and plot these versus the significance level of the test that should be (very close) to the critical p_*-value. The FDR threshold can now be determined by intersecting this plot with the identity. We also indicate the trajectory of the iterative procedure that we proposed: the initial significance level is chosen α without Bonferroni correction. Using the observed number of activations, we can adapt the significance level for the type I error rate equivalent to FDR α. As can be observed, the procedure quickly converges (typically 2–4 iterations) to the FDR threshold.

Next, we investigated the statistical parameter maps using SPM; the results are shown in Fig. 4 and are overlaid on the anatomical reference slices (left hemisphere on the left according to neurological convention). The effect of smoothing is clearly apparent from the blobbiness of the activation clusters. The smoothing kernel was set at 6 mm FWHM, which is the minimally recommended amount to ensure the validity of the GRFT threshold [2]. Switching to FDR thresholding using SPM clearly improves the sensitivity in the visual cortex, but also brings along many detections in more anterior regions that are less likely to be stimulus-related.

In Fig. 5, we show the results for WSPM. In Fig. 5(a), although we use strong type I error control (Bonferroni correction), the sensitivity is comparable to SPM: almost any cluster detected by SPM can be found also in the WSPM map, but

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

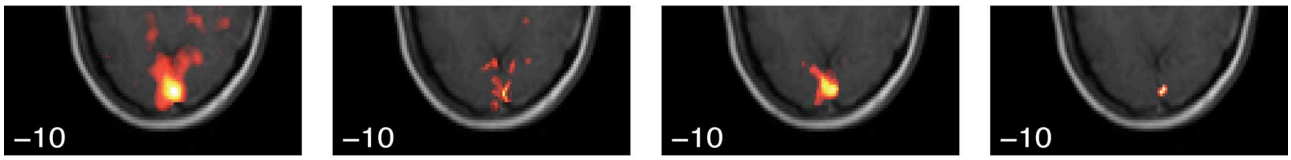


Fig. 7. Results for the posterior part of slice −10 mm. Main effect: (a) SPM (5% FDR); (b) WSPM (5% FDR). Negative dispersion: (c) SPM (5% FDR); (d) WSPM (5% FDR).

TABLE I
ZOOM-UP OF THE CONTRAST AND t-VALUES OF THE NEIGHBORHOOD NEAR THE PEAK VOXEL AT SLICE −10 mm. VOXELS THAT DO NOT SURVIVE STATISTICAL TESTING ARE INDICATED BETWEEN PARENTHESES

		SPM			WSPM		
		1	2	3	1	2	3
main effect	contrast	21.6	19.8	16.1	56.0	18.7	-5.09
	t-value	13.0	12.4	10.2	7.31	5.10	(-3.74)
dispersion	contrast	21.5	19.3	14.5	32.0	14.7	-0.13
	t-value	5.86	5.45	4.16	5.88	5.43	(-5.04)

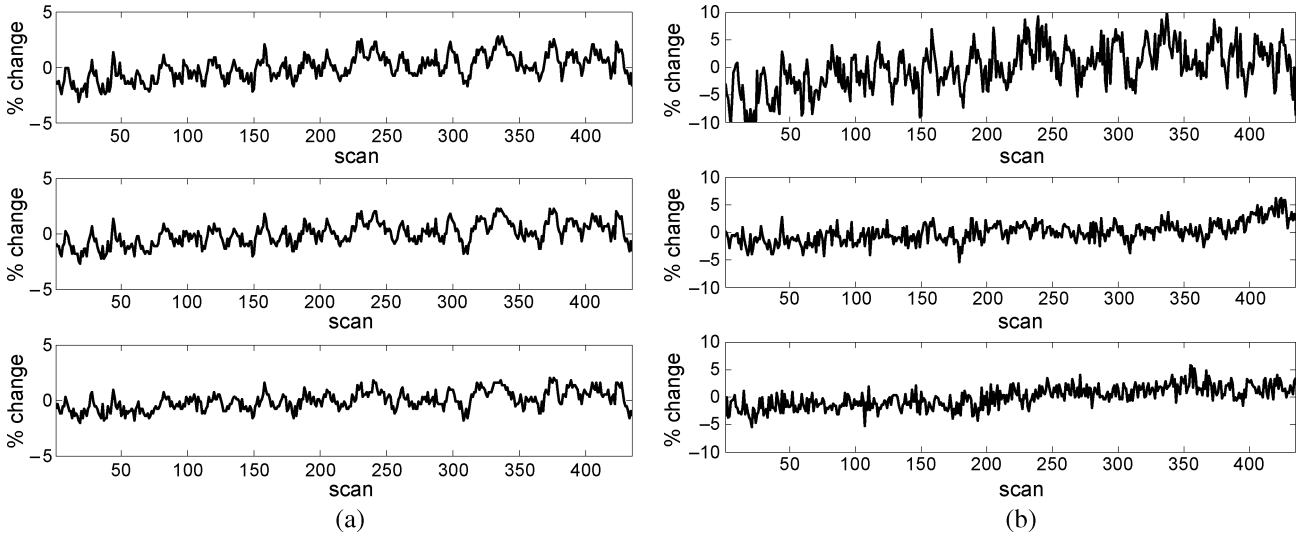


Fig. 8. Time courses of the voxels listed in Table I, extracted from (a) the smoothed data as used by SPM; (b) the unsmoothed data as used by WSPM. The time-courses are plotted relative to the baseline, which is extracted as the signal's average. From top to bottom: voxels 1, 2, 3.

with finer spatial detail. The proposed FDR thresholding significantly improves the sensitivity. Most activation clusters do not simply grow isotropically as in Fig. 4(b); instead they tend to extend along the gray matter structures in the visual cortex up to the anterior part of the calcarine sulcus. Also note the fine-detailed cluster near the lateral ventricle and occipital posterior horn [+5 mm, indicated by the white arrow in Fig. 5(b)], which is smoothed over the tissue borders in the SPM maps. Another interesting activation cluster is the fusiform gyrus [−20 mm, indicated by the green arrows in Fig. 5(b)], which plays an important role in object categorization [37]. As a point of comparison, we also show the result for GLM with non-smoothed data in Fig. 6. The result after Bonferroni correction in Fig. 6(a) shows the reduction in sensitivity. It is interesting to observe that applying FDR thresholding improves sensitivity to a comparable level as WSPM with Bonferroni correction [see Figs. 6(b) and 5(a)].

We now focus on the strong peak in slice −10 mm, which seems to fall in between the hemispheres when observing the anatomical reference (see Fig. 7). While the peak spreads out in the SPM map, WSPM preserves a fairly isolated voxel. Moreover, we also tested the data for the significance of "negative dispersion", which answers the question "does the voxel's timecourse corresponds to a (temporally) more dispersed response?" The SPM and WSPM maps for this contrast are shown in Fig. 7(c) and (d), respectively. The WSPM map shows a small isolated cluster, which might suggest that signal from an underlying draining vein dominates the voxel's timecourse—a known and important artefact in fMRI gradient-echo acquisitions [38]. In Table I, we list the contrast and t-values for the main and dispersion contrasts of three adjacent voxels. At the same time, we also plot the underlying time-courses in Fig. 8. For WSPM, which relies on the unsmoothed data, the BOLD response of voxel 1 is very strong (almost 10% with respect to baseline) and dispersed, but remains well isolated. Notice that voxel 3 even shows a (non-significant) negative response. For SPM, smoothing contaminates the neighboring voxels 2 and 3 and lowers the response's strength of voxel 1. These elements

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

reinforce our suspicion that the voxel captures the response from a draining vein. This effect would be difficult to observe from the smoothed maps and data.

VI. Conclusion

Since wavelet-based statistical parametric mapping is performed on non-smoothed data, it usually produces parametric maps that contain fine spatial resolution. In its original form, WSPM is based on strong type I error control using Bonferroni correction. Here, we introduced the FDR principle for WSPM. Although the p-values cannot be obtained directly, we showed that the FDR threshold can still be retrieved by a simple iterative scheme. We observed that the sensitivity of FDR-thresholded WSPM maps increases significantly. At the same time, the obtained maps allow to resolve thin spatial patterns. In our experimental results, the in-plane spatial resolution was still relatively low (1.9 mm $\times$ 1.9 mm). Consequently, we used only a single decomposition level. Increasing the number of decomposition levels did not improve on the results.

High-spatial-resolution parametric mapping in fMRI is especially relevant for single-subject studies; e.g., for retinotopic mapping or object categorization in the visual cortex and associative regions [37]. Moreover, the advent of high-field MRI scanners and dedicated surface coils also increases the availability of high-resolution data. An important collateral effect of SPM smoothing is the decrease of inter-subject variability, which improves the results on the group level. By contrast, WSPM is more sensitive to individual differences. Therefore, group-level WSPM analysis should pay particular attention to normalization; e.g., using landmark-based registration for the cortical areas under investigation.

The proposed FDR thresholding procedure is included in the latest available version of the WSPM toolbox.

Acknowledgment

The authors thank Dr. F. Lazeyras and Dr. S. Simon for the fMRI acquisition.

References

[1] P. Jezzard, P. M. Matthews, and S. M. Smith, *Functional MRI an Introduction to Methods*. Oxford, U.K.: Oxford Univ. Press, 2001.

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

[3] 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.

[4] 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 Mapp.*, vol. 2, pp. 189–210, 1995.

[5] R. Frackowiak, K. Friston, C. Frith, R. Dolan, and J. Mazziotta, *Hum. Brain Func.*. New York: Academic, 1997.

[6] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. J. Frackowiak, "Functional connectivity: The principal component analysis of large (pet) datasets," *J. Cereb. Blood Flow Metab.*, vol. 13, pp. 5–14, 1993.

[7] M. McKeown, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Hum. Brain Mapp.*, vol. 6, pp. 160–188, 1998.

[8] C. Beckmann and S. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 137–152, Feb. 2004.

[9] T. Adali and V. D. Calhoun, "Complex ICA of brain imaging data," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 136–139, Sep. 2007.

[10] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet decomposition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, pp. 674–693, 1989.

[11] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proc. IEEE*, vol. 84, pp. 626–638, 1996.

[12] A. Laine, "Wavelets in temporal and spatial processing of biomedical images," *Annu. Rev. Biomed. Eng.*, vol. 2, pp. 511–550, 2000.

[13] 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," *Statist. Methods Med. Res.*, vol. 12, no. 5, pp. 375–399, 2003.

[14] 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, pp. S234–S249, 2004.

[15] D. Van De Ville, T. Blu, and M. Unser, "Surfing the brain: An overview of wavelet-based techniques for fMRI data analysis," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 2, pp. 65–78, 2006.

[16] 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.

[17] 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. Cereb. Blood Flow and Metab.*, vol. 20, pp. 1610–1618, 2000.

[18] 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.

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

[20] 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.

[21] D. Van De Ville, M. Seghier, F. Lazeyras, T. Blu, and M. Unser, "WSPM: Wavelet-based statistical parametric mapping," *NeuroImage*, vol. 37, no. 4, pp. 1205–1217, Oct. 2007.

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

[23] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, pp. 1165–1188, 2001.

[24] C. R. Genovese, N. A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, pp. 772–786, 2002.

[25] J. Raz and B. I. Turetsky, "Wavelet ANOVA and fMRI," in *Proc. SPIE Conf. Mathematical Imaging: Wavelet Applications in Signal and Image Processing VII*, San Diego, CA, 1999, vol. 3813, pp. 561–570, SPIE.

[26] J. Raz, "FDR wavelet shrinkage estimators for inhomogeneous Poisson processed images," in *SPIE Conf. Wavelets Applications in Signal and Image Processing VIII*, San Diego, CA, 2000, to be published.

[27] R. Srikanth, R. Casanova, P. J. Laurienti, A. M. Peiffer, and J. A. Maldjian, "Estimation of false discovery rate for wavelet-denoised statistical parametric maps," *NeuroImage*, vol. 33, pp. 72–84, 2006.

[28] D. Yekutieli and Y. Benjamini, "Resampling based false discovery rate controlling multiple test procedures for correlated test statistics," *J. Stat. Plan. Inference*, vol. 82, pp. 171–196, 1999.

[29] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic, 1998.

[30] K. Worsley, C. Liao, J. Aston, V. Petre, G. Duncan, and A. Evans, "A general statistical analysis for fMRI data," *NeuroImage*, vol. 15, pp. 1–15, 2002.

[31] B. R. Logan and D. B. Rowe, "An evaluation of thresholding techniques in fMRI analysis," *NeuroImage*, vol. 22, pp. 95–108, 2004.

[32] 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.

[33] M. Brammer, "Multidimensional wavelet analysis of functional magnetic resonance images," *Hum. Brain Mapp.*, vol. 6, pp. 378–382, 1998.

[34] M. Desco, J. Hernandez, A. Santos, and M. Brammer, "Multiresolution analysis in fMRI: Sensitivity and specificity in the detection of brain activation," *Hum. Brain Mapp.*, vol. 14, pp. 16–27, 2001.

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.

[35] 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. 17, pp. 20–30, 2003.
[36] M. Unser, A. Aldroubi, and M. Eden, "A family of polynomial spline wavelet transforms," *Signal Process.*, vol. 30, pp. 141–162, 1993.
[37] H. P. Op de Beeck, J. Haushofer, and N. G. Kanwisher, "Interpreting fMRI data: Maps, modules and dimensions," *Nature Rev. Neurosci.*, vol. 9, pp. 123–135, Feb. 2008.
[38] J. Frahm, K.-D. Merboldt, W. Hänicke, A. Kleinschmidt, and H. Boecker, "Brain or vein-oxygenation of flow? on signal physiology in functional MRI of human brain activation," *NMR in Biomed.*, vol. 7, no. 1–2, pp. 45–53, 1994.

Dimitri Van De Ville (M'02) received the M.S. degree in engineering and computer sciences from Ghent University, Ghent, Belgium, in 1998, and the Ph.D. degree, in 2002.

He obtained a grant as Research Assistant with the Fund for Scientific Research Flanders Belgium (FWO). In 2002, he joined Prof. M. Unser's Biomedical Imaging Group at the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. In December 2005, he became Head of the Signal Processing Unit at the University Hospital of Geneva, Geneva, Switzerland, as part of the Centre d'Imagerie Biomédicale (CIBM). His research interests include wavelets, statistical analysis, multidimensional splines, and applications in biomedical imaging, such as functional magnetic resonance imaging, spectroscopy, electro-encephalography, and microscopy.

Dr. Van De Ville serves as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (since February 2006) and was previously an Associate Editor for IEEE SIGNAL PROCESSING LETTERS (2004–2006). Since 2003, he has also been an Editor and Webmaster of *The Wavelet Digest*. He is a member of the Technical Committee of the IEEE Signal Processing Society on Bio Imaging and Signal Processing (BISP). He has been the General Co-Chair of the biannual SPIE Wavelets Conference since 2007, together with V. Goyal and M. Papadakis.

Michael Unser (M'89–SM'94–F'99) received the M.S. (*summa cum laude*) and Ph.D. degrees in electrical engineering in 1981 and 1984, respectively, from the Ecole Polytechnique Frale de Lausanne (EPFL), Lausanne, Switzerland.

From 1985 to 1997, he was a Scientist with the National Institutes of Health, Bethesda, MD. He is now Full 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 has published over 150 journal papers on those topics, and is one of ISI's Highly Cited authors in Engineering (http://www.isihighlycited.com).

Dr. Unser was Associate Editor-in-Chief (2003–2005) for the IEEE TRANSACTIONS ON MEDICAL IMAGING and has served as Associate Editor for the same journal (1999–2002; 2006–2007), the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992–1995), and the IEEE SIGNAL PROCESSING LETTERS (1994–1998). He is currently a member of the editorial boards of *Foundations and Trends in Signal Processing*, the *SIAM Journal of Imaging Sciences*, and *Sampling Theory in Signal and Image Processing*. He co-organized the first IEEE International Symposium on Biomedical Imaging (ISBI2002) and was the founding Chair of the 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.

Authorized licensed use limited to: IEEE Xplore. Downloaded on February 17, 2009 at 07:23 from IEEE Xplore. Restrictions apply.