



# Wavelet-based approaches for multiple hypothesis testing in activation mapping of functional magnetic resonance images of the human brain

Jalal M. Fadili, Edward T. Bullmore

## ► To cite this version:

Jalal M. Fadili, Edward T. Bullmore. Wavelet-based approaches for multiple hypothesis testing in activation mapping of functional magnetic resonance images of the human brain. SPIE-Mathematical Imaging: Wavelet Applications in Signal and Image Processing X, 2003, San Diego, United States. pp.405-416, 10.1117/12.503377 . hal-01165650

**HAL Id: hal-01165650**

**<https://hal.science/hal-01165650>**

Submitted on 19 Jun 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Wavelet-based approaches for multiple hypothesis testing in activation mapping of functional magnetic resonance images of the human brain.

Jalal M. FADILI<sup>a</sup> and Edward T. BULLMORE<sup>b</sup>

<sup>a</sup>GREYC CNRS UMR 6072, ENSICAEN 6, Bd du Maréchal Juin, 14050 Caen, France

<sup>b</sup>Brain Mapping Unit, Addenbrooke's Hospital, University of Cambridge, CB2 2QQ, UK

## ABSTRACT

Wavelet-based methods for multiple hypothesis testing are described and their potential for activation mapping of human functional magnetic resonance imaging (fMRI) data is investigated. In this approach, we emphasize convergence between methods of wavelet thresholding or shrinkage and the problem of multiple hypothesis testing in both classical and Bayesian contexts. Specifically, our interest will be focused on ensuring a trade off between type I probability error control and power dissipation. We describe a technique for controlling the false discovery rate at an arbitrary level of type 1 error in testing multiple wavelet coefficients generated by a 2D discrete wavelet transform (DWT) of spatial maps of fMRI time series statistics. We also describe and apply recursive testing methods that can be used to define a threshold unique to each level and orientation of the 2D-DWT. Bayesian methods, incorporating a formal model for the anticipated sparseness of wavelet coefficients representing the signal or true image, are also tractable. These methods are comparatively evaluated by analysis of "null" images (acquired with the subject at rest), in which case the number of positive tests should be exactly as predicted under the null hypothesis, and an experimental dataset acquired from 5 normal volunteers during an event-related finger movement task. We show that all three wavelet-based methods of multiple hypothesis testing have good type 1 error control (the FDR method being most conservative) and generate plausible brain activation maps.

**Keywords:** Wavelets, Multiple hypothesis testing, Brain, Bayesian, Neuroimaging

## 1. INTRODUCTION

Nonparametric wavelet-based regression has been a fundamental tool in data analysis over the past two decades and is still an expanding area of ongoing research. The goal is to recover an unknown image, say  $g$ , based on sampled data that are contaminated with noise. Only very general assumptions about  $g$  are made such as that it belongs to a certain class of functions (e.g. Besov space). Nonparametric regression (or denoising) techniques provide a very effective and simple way of finding structure in data sets without the imposition of a parametric regression model. During the 1990s, the nonparametric regression literature was arguably dominated by nonlinear wavelet shrinkage and wavelet thresholding estimators.<sup>1-3</sup> These estimators are a new subset of an old class of nonparametric regression estimators, namely orthogonal series methods. Moreover, these estimators are easily implemented through fast algorithms so they are very appealing in practical situations.<sup>4</sup>

Since the seminal papers by Donoho & Johnstone,<sup>1</sup> the image processing literature has been inundated by hundreds of papers applying or proposing modifications of the original algorithm in image processing problems. Several data adaptive wavelet thresholding estimators have been developed; see for example<sup>5,6</sup> and references therein.

Various Bayesian approaches for nonlinear wavelet thresholding and nonlinear wavelet shrinkage estimators have also recently been proposed. These estimators have been shown to be effective and it is argued that they are less *ad hoc* than the classical methods discussed above. In the Bayesian approach a prior distribution is

---

Further author information: (Send correspondence to M.J.FADILI)

E-mail: Jalal.Fadili@greyc.ismra.fr, Telephone: +33 (0)2 31 45 29 20 Address: GREYC CNRS UMR 6072, ENSICAEN 6, Bd du Maréchal Juin, 14050 Caen, France. This Human Brain Project/Neuroinformatics research is funded by the National Institute of Biomedical Imaging and Bioengineering and the National Institute of Mental Health.

imposed on the wavelet coefficients. The prior model is designed to capture the sparseness of wavelet expansions. Then the image is estimated by applying a suitable Bayesian decision rule to the resulting posterior distribution of the wavelet coefficients. Such wavelet estimators have been discussed in several recent papers.<sup>7–18</sup> Moreover, it has been shown that Bayesian wavelet estimators outperform classical thresholding estimators, which operate on wavelet coefficients one at a time, in terms of mean squared error in finite sample situations. A detailed study involving recent classical and Bayesian methods in the development towards high-performance wavelet estimators and their finite sample properties is reported in.<sup>19</sup>

Functional MRI is a relatively new technique for measuring the neural correlates of cognitive processes. Many difficulties must be addressed when processing fMRI data such as the weakness of the signal enhancement and the multiple sources of artifacts. In order to extract functional information and detect activated regions using fMRI, the most widely adopted procedures are generally based on linear (or, less often, non-linear) regression theory. The desired effect time courses are included in a design matrix which also incorporates some other "nuisance" variables, e.g., low frequency drift, head movements etc. Then the time course at each pixel  $(m, n)$  can be expressed as a linear combination of some covariates  $\mathbf{x}_k$  via an additive linear model:

$$\mathbf{y}_{mn} = X\beta_{mn} + \varepsilon_{mn}, \quad \varepsilon \sim \mathcal{N}(0, \Sigma_{mn}) \quad (1)$$

where  $\varepsilon_{mn}$  is a zero mean random vector with covariance matrix  $\Sigma_{mn}$ . Standard assumptions are that the errors  $\varepsilon_{mn}$  are independent and normally distributed. There is also an extended literature on modelling  $\varepsilon_{mn}$  as short-memory processes (AR, ARMA, ARIMA); see<sup>20</sup> and references therein. Recently, we have proposed a new wavelet-based estimator of such linear models in the presence of long-memory  $(1/f)$  error processes.<sup>20, 21</sup> Owing to its superior efficiency and robustness in this context, the wavelet-generalized least-squares (WLS) algorithm described in<sup>21</sup> is used here for estimation of linear model statistics from the fMRI time series observed at each voxel of all fMRI datasets. This yields a time series statistic map in which the linear model parameter of interest (divided by its standard error) is represented at each voxel in the image. Various statistical maps, including Student T-maps, Fisher-Snedecor F-maps or Gaussian Z-maps, have been reported in the neuroimaging literature. In the following, we will consider Gaussian statistical maps. This is not restrictive since any other standardised statistic map can be transformed to a Z-map using appropriate integral transformations.

Given such a spatial brain map of time series statistics, the problem at the heart of this paper is how best to estimate the true activation map from its noisy realisation. Almost universally in current practice, the first step taken towards this goal is to reduce the noise in the observed map. This is most often done using a monoresolutional Gaussian filter whose width must be arbitrarily specified. This will generally entail loss of resolution of spatially detailed features in the map and will cost sensitivity to detect other spatial features of the true image which do not conform in size or shape to the Gaussian kernel. Instead we here consider a multiresolutional, wavelet-based solution to the problem, which is known to provide optimal estimates for a wide class of image spaces. This is done in an adaptive data-driven way. The second step is to decide which pixels are activated, for a pre-specified type I probability error  $\alpha$ . This amounts to a binary decision problem where each pixels is compared to the  $\alpha$ -level critical threshold. This decision step is currently implemented at the level of individual voxels or spatial clusters of contiguous suprathreshold voxels surviving a preliminary voxel-level test. At either voxel or cluster level the false positive error may be controlled over multiple tests in terms of the family-wise error rate or the false discovery rate; see<sup>22</sup> for a review.

In this paper, we propose to tackle this latter problem using a fully wavelet-based hypothesis testing framework. More specifically, we suggest that the ideas of wavelet thresholding or shrinkage can be viewed as an approach to multiple hypothesis testing. In the neuroimaging community, some authors have already proposed to use wavelet domain denoising to obtain an estimate of activation maps in fMRI<sup>23</sup> and positron emission tomography (PET).<sup>24, 25</sup> See also the review in.<sup>26</sup> However, these approaches share the drawback of applying the simple universal thresholding approach without assigning any probability risk to their maps. Moreover, the multiple comparisons problem was not explicitly addressed in their approach. To our knowledge, Ruttiman et al.<sup>27–29</sup> was the first to propose a wavelet-based hypothesis testing approach for brain activation mapping using Bonferroni corrected thresholds unique to each level of the 2D-DWT of spatial statistic maps. In the same spirit, Feilner<sup>30, 31</sup> proposed the fractional spline wavelets to analyse time series statistic maps, using the fractional order of the spline to control the smoothness of the reconstructed image. However, all these methods are arguably too

*ad hoc* and restrictive since they are only readily applicable to block designs. Sophisticated (event-related) experimental paradigms elicit brain activation in response to a number of discrete cognitive trials and any multiple hypothesis testing method that is likely to be of real usefulness to the contemporary neuroimaging community needs to be generally adaptable to these and other experimental designs.

We here introduce a coherent statistical framework to this estimation procedure in both the classical and the Bayesian contexts. Several solutions from the statistical theory of wavelets are proposed to address the serious problem of the large number of hypotheses being tested in activation mapping of a single fMRI dataset (typically the search volume or number of voxels tested will be in the order of 20,000).

This paper is organized as follows: in Section 2 we define the nonparametric regression problem and introduce some notational aspects. In 3, which is the core of the paper, we elaborate the multiple hypothesis testing framework in the wavelet domain. Several variants of this general approach are described in classical and Bayesian contexts. Section 4 presents some experimental results on null and experimental fMRI datasets. Finally, conclusions and directions of future work are briefly summarised.

## 2. NONPARAMETRIC WAVELET-BASED REGRESSION

Let  $g_{mn}, m, n = 0, \dots, N-1$  be equally-spaced samples of a real-valued image; without loss of generality  $N$  is considered as a power of 2 ( $N = 2^J$ ). Now consider the standard nonparametric regression model:

$$y_{mn} = g_{mn} + \epsilon_{mn} \quad (2)$$

where  $\epsilon_{mn}$  are iid normal random variables with mean zero and variance  $\sigma^2$  independent of  $g_{mn}$ . The goal is to recover the underlying function or true image  $g$  from the observed noisy data  $y_{mn}$ , without assuming any particular parametric structure for  $g$ . Let  $\mathbf{y}$ ,  $\mathbf{g}$  and  $\epsilon$  denote the matrix representations of the corresponding entities and let  $\mathbf{D} = \mathcal{W}\mathbf{y}$ ,  $\mathbf{S} = \mathcal{W}\mathbf{g}$  and  $\mathbf{V} = \mathcal{W}\epsilon$ , where  $\mathcal{W}$  is the two dimensional dyadic orthonormal wavelet transform (DWT) operator.<sup>32</sup> In a two dimensional setting, the subbands  $HH_j$ ,  $HL_j$  and  $LH_j$ ,  $j = J_c, \dots, J-1$  correspond to the detail coefficients in diagonal, horizontal and vertical orientations, and the subband  $LL_{J_c}$  is the approximation or the smooth component.  $J_c$  is the coarsest scale of the decomposition that will usually be specified as  $J_c = \log_2 \log N + 1$  from asymptotic considerations. Let  $s_{mn}^{oj}$  be the detail coefficient of the true image  $\mathbf{g}$  at location  $(m, n)$ , scale  $j$  and orientation  $o$ , and similarly for  $d_{mn}^{oj}$  and  $v_{mn}^{oj}$ . Due to the orthogonality of the basis,  $v_{mn}^{oj}$ , the wavelet coefficients obtained by the DWT of white noise, will also be independent normal variables with the same variance. It follows from Eq.2 that :

$$d_{mn}^{oj} = s_{mn}^{oj} + v_{mn}^{oj}, j = J_c, \dots, J-1; \quad m, n = 0, \dots, 2^j - 1 \quad (3)$$

The sparseness of the wavelet expansion makes it reasonable to assume that essentially only a few large detail coefficients in  $\mathbf{D}$  contain information about the underlying image  $\mathbf{g}$ , while small values can be attributed to the noise which uniformly contaminates all wavelet coefficients. It is also advisable to keep the approximation coefficients intact because they represent low-frequency terms that usually contain important features about the image  $\mathbf{g}$ . By thresholding or shrinking the detail coefficients and inverting the DWT, one can obtain an estimate of the underlying image  $\mathbf{g}$ . So the resulting three-step wavelet-based estimation procedure can be summarized by the following diagram:

$$\mathbf{y} \xrightarrow{\text{DWT}} \mathbf{D} \xrightarrow{\text{Nonlinear thresholding operator } \eta} \{\hat{\mathbf{S}} = \eta(\mathbf{D})\} \xrightarrow{\text{IDWT}} \hat{\mathbf{g}}$$

where  $\eta$  is a nonlinear (shrinkage or thresholding) operator. Examples of such an operator are the universal threshold and hard and soft thresholding rules introduced by Donoho et al.<sup>1</sup> Following their work, a variety of methods have been proposed in the literature to choose the threshold level and the thresholding rule; see for example.<sup>19</sup>

## 3. WAVELET THRESHOLDING AS A HYPOTHESIS TESTING PROBLEM

### 3.1. Classical approach

We mean by ‘‘classical’’ that no prior distribution is imposed on the true unknown wavelet coefficients, in contrast to the Bayesian approach where a prior density is specified to capture the sparseness of wavelet coefficients of the true image  $g$ .

### 3.1.1. Thresholding as a multiple hypotheses testing problem

The main idea here is simply to reformulate wavelet-thresholding as a multiple hypothesis testing problem. For each observed wavelet coefficient at each scale, orientation and location, we test the following hypothesis:

$$H_0 : s_{mn}^{oj} = 0 \text{ versus } H_1 : s_{mn}^{oj} \neq 0$$

The observed detail coefficient is distributed according to  $d_{mn}^{oj}/s_{mn}^{oj} \sim \mathcal{N}(s_{mn}^{oj}, \sigma^2)$ . This detail coefficient is retained in the reconstruction if  $H_0$  is rejected with a risk  $\alpha$ , otherwise it is discarded. Classical approaches to multiple hypothesis testing face serious problems because of the large number of hypotheses being tested simultaneously. In other words, if the error is controlled at an individual level, the test is too permissive and the chance of erroneously retaining a coefficient is extremely high; whereas if the family-wise error is controlled, the test is too conservative and the chance of falsely discarding a coefficient is extremely high. Abramovich and colleagues<sup>33, 34</sup> have proposed a way to control such dissipation of power based on type 1 error control in terms of the false discovery rate (FDR).

Let  $T$  be the number of observed wavelet coefficients that are retained by the thresholding procedure. From these  $T$  coefficients,  $TP$  (true positives) are correctly retained and  $FP = T - TP$  (false positives) are erroneously retained. The error in such a procedure is expressed in terms of the random variable  $FPF = FP/T$ , i.e., the proportion of the retained wavelet coefficients that should properly have been rejected. Obviously,  $FPF$  is defined as zero when  $T = 0$  since no error of this type can be made when no coefficient is retained. The FDR of empirical wavelet coefficients can now be defined as the expectation of  $FPF$ , i.e., the expected proportion of false positives among the total number of coefficients surviving the threshold.

Following Abramovich et al,<sup>33, 34</sup> we propose maximizing the number of retained wavelet coefficients subject to the constraint  $FPF < \alpha$ , which can be operationalised in terms of the following algorithm to calculate the global threshold of the map:

1. For each of the  $nd = N \times N - 1$  observed wavelet coefficients  $\{d_{mn}^{oj} : j = 0, \dots, J-1; m, n = 0, \dots, 2^j-1; o = HH, HL, LH\}$ , calculate the corresponding double-sided p-value,  $p_{mn}^{oj}$  under  $H_0$ :

$$p_{mn}^{oj} = 2 \left( 1 - \Phi \left( \frac{|d_{mn}^{oj}|}{\hat{\sigma}} \right) \right) \quad (4)$$

where  $\Phi(x)$  is the cumulative normal distribution of a standard normal variable. If one is able to measure  $\hat{\sigma}$ , then this estimate could be used. If not,  $\sigma$  can be estimated from the coefficients in the horizontal  $HH$  orientation at the finest scale using the popular robust estimator<sup>1</sup>:

$$\hat{\sigma} = \frac{\text{MAD}(d_{mn}^{HH_1})}{0.6745} \quad (5)$$

where MAD is the median absolute deviation. We can see that all the  $N^2 - 1$  wavelet detail coefficients of the image are being tested in this step. In the context of fMRI, one is usually interested only in voxels representing brain tissue and not skull, scalp or surrounding air. These non-cerebral elements of the image are therefore discarded in a preliminary pre-processing step; it follows that the number  $nd$  has to be reduced by taking into account only the number of intra-cranial voxels.

2. Sort the  $p_{mn}^{oj}$  in an ascending order,  $p_1 \leq p_2 \leq \dots \leq p_{nd}$ .
3. Find the last index such that  $i_{\text{FDR}} = \max_i (p_i < (i/nd)\alpha)$ .
4. For this index, calculate the critical threshold corresponding to this double-sided p-value:

$$\lambda_{\text{FDR}} = \hat{\sigma} \Phi^{-1} \left( 1 - \frac{p_{i_{\text{FDR}}}}{2} \right) \quad (6)$$

5. Use  $\lambda_{\text{FDR}}$  and apply classical hard (kill or keep) or soft (kill or shrink) thresholding rules to the observed wavelet coefficients up to a coarse level  $J_c$ .

6. Apply the inverse DWT to obtain an estimate of the image  $g$ .

Clearly this is a global thresholding procedure in which the same threshold value  $\lambda_{\text{FDR}}$  is applied to coefficients at all scales and orientations of the decomposition. Indeed, it can be easily shown that the universal threshold value of Donoho ( $\lambda_U = \sigma\sqrt{2\log N^2}$ ) is a special case of the general procedure described here that confers a type 1 probability error  $1/(N^2\sqrt{\pi\log N^2})$  for large  $N$ , which is also equal to its power under the alternative  $H_1$ .

### 3.1.2. Thresholding as a recursive hypothesis testing problem

Here we use prior work by Ogden and colleagues<sup>35</sup> to develop a recursive procedure for multiple hypothesis testing which uses level-dependent thresholds  $\lambda_j$ . Rather than seeking to include as many wavelet coefficients as possible (subject to constraint in terms of type 1 error), the recursive hypothesis testing procedure<sup>35</sup> includes a wavelet coefficient only when there is strong, affirmative evidence that it is needed in the reconstruction.

Let  $\tilde{d}^j_i = \frac{d^j_i}{\sigma}$  be independent random variables  $\mathcal{N}(\tilde{s}^j_i, 1)$ ,  $i = \{1, \dots, n_j\}$  that represent the observed wavelet coefficients at any level  $j$ , where  $n_j = 2^{2j}$  if all detail coefficients at scale  $j$  are tested. For fMRI data, and for the same reason as above,  $n_j$  is the number of intra-cranial voxels at scale  $j$ . Let  $I_{n_j}$  represent a non-empty subset of indices  $\{1, \dots, n_j\}$ . Then the multiple hypothesis testing problem could be expressed as:

$$H_0 : \tilde{s}^j_i = 0, i \in I_{n_j} \text{ versus } H_1 : \tilde{s}^j_i \neq 0, i \in I_{n_j} \text{ and } \tilde{s}^j_i = 0, i \notin I_{n_j} \quad (7)$$

(For readability, the superscript  $o$  (denoting orientation) was not used in this expression). To test the set of hypotheses given in Eq.7 we use the standard likelihood ratio test (LRT).<sup>35</sup> If the cardinality of the set  $I_{n_j}$  is not known, which is the case in practice, the LRT for the above hypotheses would be based on the test statistic  $\sum_{i=1}^{n_j} |\tilde{d}^j_i|^2 \sim \chi^2_{n_j} |H_0|$ . However, this is not the most appropriate test statistic, since only a few of the  $\tilde{s}^j_i$ 's are non-zero, resulting in poor power of detection when  $I_{n_j}$  contains only a few coefficients. If one knows the cardinality  $c$  of  $I_{n_j}$ , then the LRT would be based on the sum of squares of the  $c$  largest  $\tilde{d}^j_i$ 's. However,  $c$  is not known in practice. Therefore, Ogden<sup>35</sup> suggested a recursive testing procedure for  $I_{n_j}$  containing *one* element each time. The LRT statistic would then be the largest  $|\tilde{d}^j_i|^2$ . It has been shown that the corresponding critical threshold at level  $\alpha$  is equal to:

$$|\lambda_{\text{LRT}}^j|^2 = \left\{ \Phi^{-1} \left[ \frac{(1-\alpha)^{1/n_j} + 1}{2} \right] \right\}^2 \quad (8)$$

Ogden and Parzen<sup>35</sup> then proposed the following recursive algorithm for choosing the threshold  $\lambda_j$ , which we extend to the 2D case for each level and orientation:

1. At each level and orientation, calculate  $|\lambda_{\text{LRT}}^j|^2$  using Eq.8 and compare the largest  $|\tilde{d}^j_i|^2$  with this critical value.
2. If the square value of  $\tilde{d}^j_i$  is larger, this indicates that there is still significant signal among wavelet coefficients. Remove this  $\tilde{d}^j_i$ , set  $n_j = n_j - 1$  and return to Step 1.
3. If  $|\tilde{d}^j_i| < |\lambda_{\text{LRT}}^j|$ , then there is no strong evidence of presence of signal in the remaining coefficients. Set the threshold  $\lambda_j$  of the current level and orientation to the largest remaining  $|\tilde{d}^j_i|$ .
4. Apply the soft thresholding scheme using  $\lambda_j$  so that small coefficients (indistinguishable from pure noise) are shrunk to zero and the significant coefficients included in the reconstruction are shrunk toward zero by the maximum absolute value of the small coefficients. This is accomplished in an adaptive data-driven way at each scale and orientation.

In the hypothesis testing procedures described above, the user has some control on the smoothness of the reconstructed image by means of the pre-specified type 1 probability level  $\alpha$ . A compromise must be reached between oversmooth and noisy estimates. Popular choices are  $\alpha = 0.05$  or  $0.01$ .

### 3.2. Bayesian approach

Similar in spirit to the multiple hypothesis testing procedures discussed in Section 3.1.1, a Bayesian method for obtaining a wavelet thresholding estimator was considered by Vidakovic.<sup>12</sup> This method also involves testing the following hypothesis  $H_0 : d_{mn}^j = 0$  versus  $H_1 : d_{mn}^j \neq 0$ . However, the Bayesian framework here imposes a prior which describes the variability of the wavelet coefficients  $s_{mn}^j$  of the true image  $g$ .

This requires a prior distribution that has a point mass component; otherwise the testing is impossible because any continuous prior density will give the prior (and hence the posterior) probability of zero to the precise hypothesis. When including a point mass at zero, and with an appropriate choice of the Bayes rule, it is possible to get a wavelet coefficient estimate which is exactly zero.<sup>12</sup> Then we will choose the prior mixture model according to<sup>9,36</sup>:

$$f_s(s) = p_j \tilde{f}(s) + (1 - p_j) \delta(s) \quad (9)$$

where  $p_j$  is the mixing proportion,  $\delta(s)$  is a point mass at zero and  $\tilde{f}(s)$  describes the behaviour of  $s_{mn}^j$  under  $H_1$  (when  $s_{mn}^j$  is nonzero) which occurs with probability  $p_j$ . Considering  $\tilde{f}(s)$  as a Gaussian pdf with zero mean and variance  $\sigma_j^2$ , Abramovich and colleagues<sup>37</sup> proposed the ratio test (RT) statistic as the Bayes thresholding rule:

$$\eta_{mn}^j = \frac{P(H_1/d_{mn}^j)}{P(H_0/d_{mn}^j)} \underset{H_1}{\overset{H_0}{\leq}} \eta_\alpha \quad (10)$$

In contrast to previous authors, we here introduce an  $\alpha$ -level critical threshold  $\eta_\alpha$ . Again, this will allow us to control the smoothness of the reconstructed map. Furthermore, one can easily show that the marginal pdf of the observed wavelet coefficients is:

$$\begin{aligned} f_d(d_{mn}^j) &= p_j f_{d/H_1}(d_{mn}^j) + (1 - p_j) f_{d/H_0}(d_{mn}^j) \\ &= p_j \phi(d_{mn}^j; \sigma_j^2) * \phi(d_{mn}^j; \sigma^2) + (1 - p_j) \phi(d_{mn}^j; \sigma^2) \end{aligned} \quad (11)$$

where  $\phi(x; \sigma^2)$  is the centered Gaussian pdf with variance  $\sigma^2$ . From this expression it is straightforward to derive the expression of the RT:

$$\eta_{mn}^j = \frac{p_j}{1 - p_j} \frac{\sigma}{\sqrt{\sigma_j^2 + \sigma^2}} e^{\frac{d_{mn}^j{}^2}{2\sigma^2} \left( \frac{\sigma_j^2}{\sigma_j^2 + \sigma^2} \right)} \quad (12)$$

One can clearly see, from Eq.12, that under the null hypothesis, i.e.,  $d_{mn}^j \sim \mathcal{N}(0, \sigma^2)$ , the random variable  $\left( \frac{d_{mn}^j}{\sigma} \right)^2$  is  $\chi^2(1)$  distributed. To test  $H_0$  for a pre-specified  $\alpha$  value, one can easily derive the corresponding critical threshold  $\eta_\alpha$ , and then apply the Bayesian RT-based wavelet estimator which is essentially a hard thresholding rule:

$$\hat{s}_{mn}^j = d_{mn}^j \mathbf{1}(\eta_{mn}^j \geq \eta_\alpha) \quad (13)$$

where  $\mathbf{1}(x)$  is the indicator function.

To apply the Bayesian RT-based thresholding rule, the parameters  $p_j$ ,  $\sigma_j$  and  $\sigma$  should be estimated appropriately. Several solutions have been proposed in the literature. For example, one could use the robust estimate Eq.5 for  $\sigma$  and an iterative expectation-minimization (EM) algorithm to get maximum likelihood estimates (MLE) of  $p_j$  and  $\sigma_j$ .<sup>11,13</sup>

In the mixture prior model of Eq.9, other forms for  $\tilde{f}$  have been considered in the literature. In their so-called BAMS method, Vidakovic et al.<sup>38</sup> chose a standard exponential prior on the unknown  $\sigma^2$ , and obtained a double exponential pdf for  $\tilde{f}$ . Their results can be easily exploited to derive a closed-form expression for  $\eta_{mn}^j$ . For robustness reasons, Vidakovic<sup>12</sup> also suggested the use of central Student  $t$  distributions (more heavily tailed than the normal) as a prior. However, no closed form expression is available for the corresponding Bayesian thresholding rule.

In fact, many other prior distributions can be used provided that they are unimodal, centered and peaked at zero, and symmetric. Complicated prior pdfs can become useless in practice, although theoretically powerful, since in general, closed-form expressions are not available for them thus necessitating intensive numerical integration. It is also worth noting that the distribution of the decision RT in Eq.10 quickly becomes a complicated

function of  $d_{mn}^j$  as the complexity of the prior increases. Therefore, its distribution is not easily accessible under  $H_0$ . Alternatively, one must use (computer-intensive) resampling techniques such as the bootstrap to estimate its distribution.

## 4. RESULTS AND DISCUSSION

### 4.1. Experimental designs and data acquisition

Gradient-echo echoplanar imaging data were acquired as follows:

#### 4.1.1. Null datasets at 1.5T

4 normal volunteers were studied (at 1.5 Tesla (T)) while they lay quietly in the scanner with their eyes closed (6 min). 72  $T_2^*$ -weighted images were acquired at each of 26 contiguous slices of data in an oblique axial plane using the GE LX EchoSpeed system (General Electric, Milwaukee WI) at CHU in Caen, France: time to echo (TE) 60 ms, time to repetition (TR) = 5 sec,  $64 \times 64$  voxel slices ( $N = 64, J = 6$ ), inplane resolution  $3.5 \times 3.5$  mm, slice thickness = 5 mm. In each time series, the 4 first volumes were eschewed to ensure magnetization stabilization.

For all of the null experimental data the same general approach to regression model specification was adopted. An  $N_t = 68$ -length boxcar vector was constructed to indicate which images were acquired during presentation of an activation condition and which were acquired during presentation of a baseline condition. This vector was convolved with a Poisson kernel, parameterised by  $\lambda = 4s$ . This Poisson-convolved input function was combined with a unitary constant column vector to form the  $(N_t \times 2)$  design matrix  $\mathbf{X}$ . To assess the influence of the experimental design complexity, high (0.032 Hz), intermediate (0.016 Hz) and low (0.008 Hz) frequency boxcar functions were modelled. For all time series, the design matrix was fit using the WLS algorithm<sup>21</sup> after motion correction of each observed fMRI time series; we did not initially detrend the time series in any way.

#### 4.1.2. Event-related experimental datasets at 1.5T

We studied 5 single male subjects during a discrete-trial or event related (ER) experiment. The task was simply to oppose finger and thumb of the right hand repeatedly. During this experiment, 26 slices of gradient echo echoplanar imaging data were acquired using a 1.5T GE MRI system with acquisition parameters as above. This simple experiment is expected to activate areas of the brain that are important in motor tasks.

## 4.2. Results

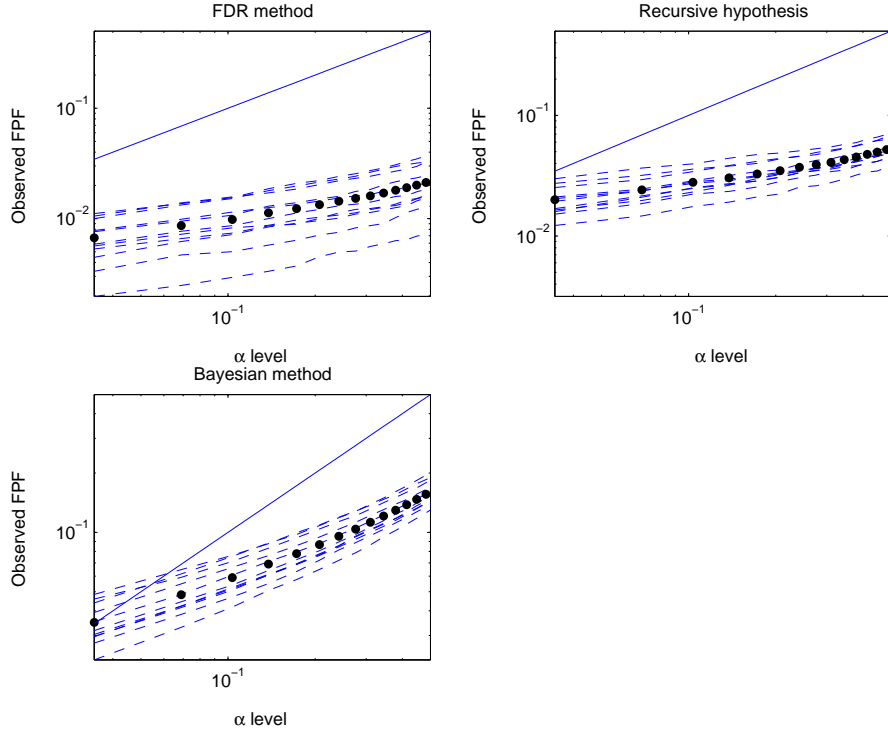
### 4.2.1. Null datasets

We assessed the relative performance of our hypothesis testing procedures in terms of type 1 error control using the null fMRI datasets. We tested the null hypothesis (in the wavelet domain) over a range of critical values corresponding to probabilities of type 1 error  $\alpha$  in the range  $[0, 0.5]$ . For each null dataset and design matrix, the WLS algorithm was applied and the resulting statistical maps were estimated using the hypothesis testing methods described above: the FDR controlling global threshold; the recursive hypothesis testing method using a level and orientation-adaptive threshold; and the Bayesian methods. In all cases, the Daubechies wavelet with 4 vanishing moments was used with a coarsest thresholded level  $J_c = 2$ .

For a valid hypothesis testing method with a pre-specified type 1 error probability level  $\alpha$ , the number of intra-cranial positive voxels observed in the wavelet domain when the null hypothesis is true, as presumably it is in these data, should be less than or equal to the expected number of positive voxels =  $\alpha \times \text{Number of intra-cranial voxels in the wavelet domain}$ . The results are shown graphically in Fig.1.

As expected for the FDR method, the observed FPF is always much less than the pre-specified risk  $\alpha$ . This is true for all null datasets and simulated design matrices. On the one hand, the FDR method appears to be somewhat more conservative than the recursive hypothesis testing method at low risk levels although these two methods become comparable at high  $\alpha$  values ( $> 0.5$ ). On the other hand, the Bayesian approach performs well in average but fails in controlling the FPF in some datasets especially at low  $\alpha$ .

A key issue is to find out which wavelet and coarsest thresholded level are best suited for the detection of activation patterns in terms of some measure of quality. For the moment, we only used the FPF rate as a measure

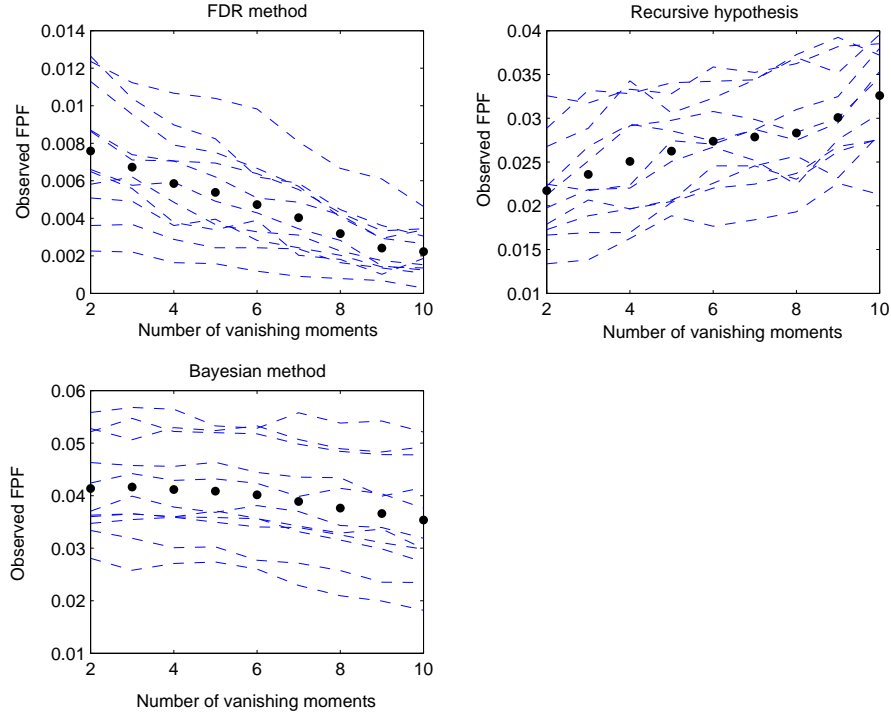


**Figure 1.** Observed FPF rate versus pre-specified risk  $\alpha$ . Each dashed curve corresponds to results of thresholding a time series statistic maps based on analysis of a single null fMRI dataset with an arbitrary design matrix. The black filled circles correspond to the average curve. The solid line is the identity line.

of quality. This index only reflects the performance of the methods under the null hypothesis, and one must ideally also know how the hypothesis testing techniques work under conditions where the alternative hypothesis is true.

**Effect of the wavelet regularity** We studied the influence of the wavelet regularity on the false discovery rate by varying the number of vanishing moments of the Daubechies wavelet; in all cases, the coarsest decomposition level was  $J = 2$ . The results are depicted in Fig.2. The observed FPF for the FDR method tends to decrease as the Daubechies wavelet becomes more regular. The observed false positive fraction of the recursive hypothesis testing method shows a slight increase with the wavelet regularity while no significant change is observed on the FPF given by the Bayesian method. However, from these preliminary results, no firm conclusion can really be drawn concerning the best choice of wavelet order since only the FPF was taken into account as a quality measure. In future, we intend to refine this evaluation by taking into account other measures such as the detection power or the false negative fraction.

**Effect of the coarsest level** In our notation, the resolution levels are numbered from 0 to  $J - 1$ , with 0 being the coarsest level (see 2). From Fig.3, the observed FPF systematically decreases as the decomposition level increases for all the thresholding methods presented. This is not surprising as the number of wavelet coefficients, and thus the number of hypotheses being tested, simultaneously increases as the decomposition level decreases. Again, these plots should be read carefully since only the FPF was used as a quality measure.



**Figure 2.** Observed FPF as a function of the number of vanishing moments of the Daubechies wavelet. The coarsest decomposition level was 2.

### 4.3. Functional MRI: activation mapping

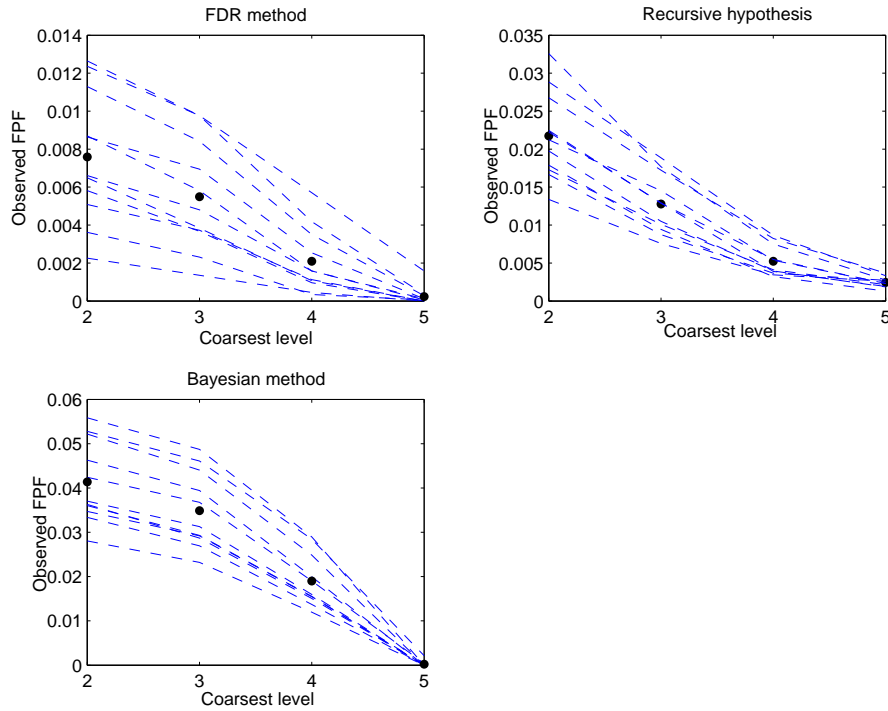
As shown in Fig.4, the brain areas which show significantly large values are located in the expected motor areas (primary and supplementary motor areas, and cerebellum). Because of its soft thresholding rule, the recursive hypothesis testing method gives a smoother activation map. Moreover, the Bayesian methods appear to provide a somewhat fuller or more sensitive characterization of the cerebral response. On the one hand, this seems to confirm the relatively conservative behaviour of the FDR method. On the other hand, one should probably expect a higher power level (sensitivity) for the recursive hypothesis testing and the Bayesian approaches.

## 5. CONCLUSION

In this paper, we proposed a fully wavelet-based hypothesis testing framework for activation mapping based on functional magnetic resonance images of the human brain. Algorithms from the statistical theory of wavelets were adapted to the case of analysing 2D spatial maps of linear model parameters estimated by analysis of the fMRI time series observed at each voxel. Two classical and one Bayesian method for multiple hypothesis testing were presented. Our preliminary results are very promising. Nevertheless, our measure of quality, which was based solely on the false positive fraction throughout this paper, has to be refined by taking into account other measures such as the detection power or the false negative fraction. This is the focus of ongoing work in our group.

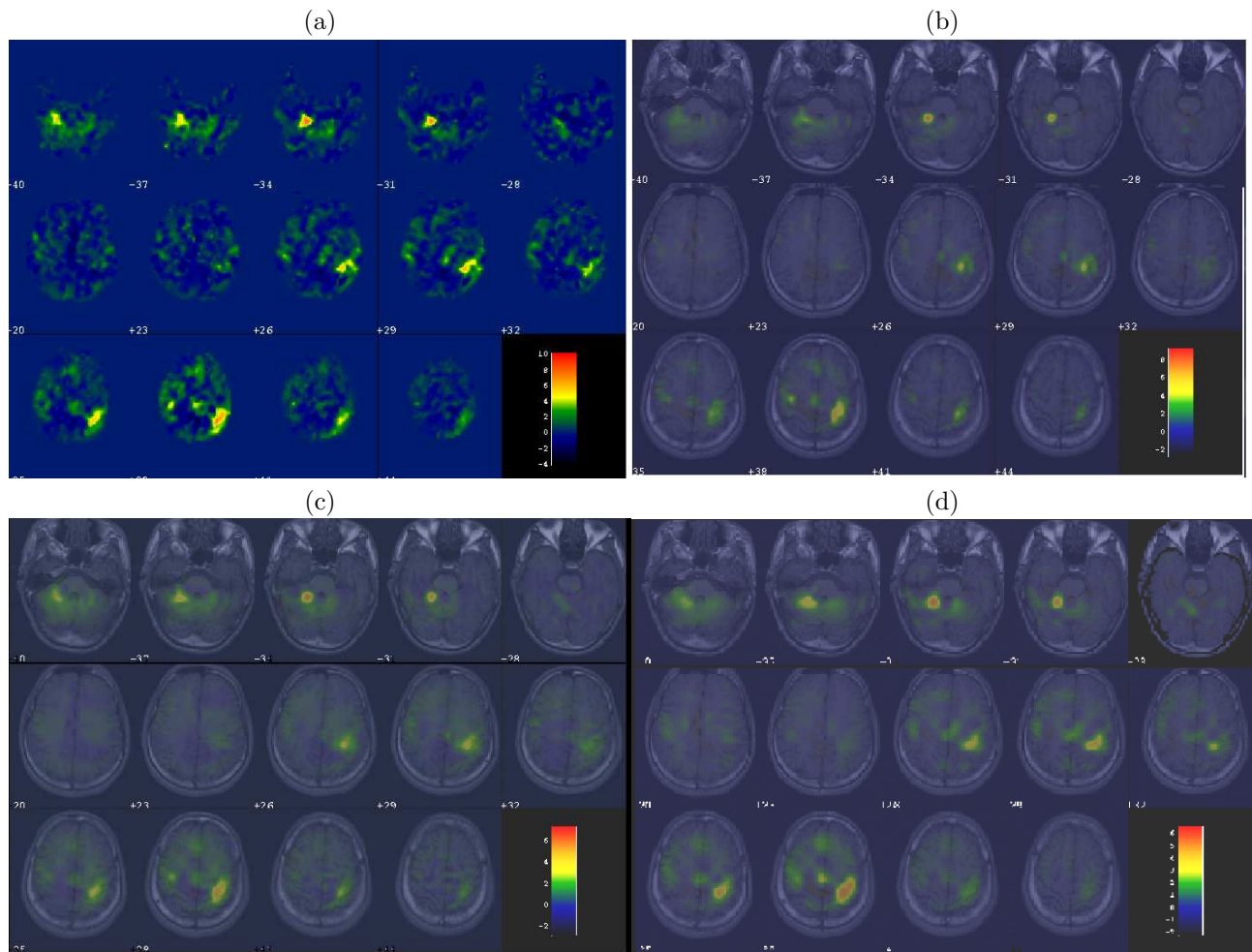
## REFERENCES

1. D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika* **81**(3), pp. 425–455, 1994.
2. D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association* **90**(432), pp. 1200–1224, 1995.



**Figure 3.** Observed FPF as a function of coarsest thresholded scale. The Daubechies wavelet with 4 vanishing moments was used.

3. D. L. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *J. R. Statist. Soc. B* **57**(2), pp. 301–337, 1995.
4. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. PAMI* **11**(7), pp. 674–693, 1989.
5. B. Vidakovic, *Statistical Modeling by Wavelets*, John Wiley & Sons, New York, 1999.
6. D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge press, 2000.
7. E. P. Simoncelli and E. H. Adelson, "Noise removal via bayesian wavelet coring," in *Third Int'l Conf on Image Proc*, **1**, pp. 379–382, IEEE Sig Proc Society, (Lausanne), 1996.
8. H. Chipman, E. Kolaczyk, and R. McCulloch, "Adaptive bayesian wavelet shrinkage," *J. Am. Statist. Ass.* **92**, pp. 1413–1421, 1997.
9. F. Abramovich, T. Sapatinas, and B. Silverman, "Wavelet thresholding via a bayesian approach," *J. R. Statist. Soc. B* **60**, pp. 725–749, 1998.
10. M. Crouse, R. Nowak, and R. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Transactions on Signal Processing* **46**(4), pp. 886–902, 1998.
11. I. Johnstone and B. Silverman, "Empirical bayes approaches to mixture problems and wavelet regression," tech. rep., Department of Mathematics, University of Bristol, UK., 1998.
12. B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *Journal of the American Statistical Association* **93**(441), pp. 173–179, 1998.
13. M. A. Clyde and E. I. George, "Empirical bayes estimation in wavelet nonparametric regression," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, eds., pp. 309–322, Springer-Verlag, New York, 1999.
14. M. Clyde and E. George, "Flexible empirical bayes estimation for wavelets," *J. R. Statist. Soc. B* **62**, pp. 681–698, 2000.
15. M. Vannucci and F. Corradi, "Covariance structure of wavelet coefficients: theory and models in a bayesian perspective," *J. R. Statist. Soc. B* **61**, pp. 971–986, 1999.
16. S. Huang and H. Lu, "Bayesian wavelet shrinkage for nonparametric mixed effects models," *Statist. Sinica* **10**, pp. 1021–1040, 2000.
17. S. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing* **9**(9), pp. 1522–1531, 2000.



**Figure 4.** (a) Example of an original statistical map representing linear model parameters estimated by wavelet-generalized least squares at each voxel of the 2D image. Estimated activation maps with  $\alpha = 0.01$  using: (b) the FDR method, (c) the recursive hypothesis testing method and (d) the Bayesian thresholding estimator. The activation task was repeated right index-thumb opposition which is expected to activate contralateral regions of motor and somatosensory cortex, supplementary motor area and ipsilateral cerebellum. This pattern of activation is perhaps most clearly seen in the map obtained by the Bayesian thresholding operation.

18. A. Achim, A. Bezerianos, and P. Tsakalides, "Novel bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.* **20**, pp. 772–783, 2001.
19. A. Antoniadis, J. Bigot, and T. Sapatinas, "Wavelet estimators in nonparametric regression: A comparative simulation study," *Journal of Statistical Software* **6**(6), 2001.
20. E. Bullmore, C. Long, J. Suckling, M. Fadili, G. Calvert, F. Zelaya, A. Carpenter, and M. Brammer, "Color noise and computational inference in neurophysiological (fmri) time series analysis: Resampling methods in time and wavelet domains," *Human Brain Mapping* **12**(2), pp. 61–78, 2001.
21. M. Fadili and E. Bullmore, "Wavelet-generalised least squares: a new blu estimator of linear regression models with  $1/f$  errors," *NeuroImage* **15**, pp. 217–232, 2001.
22. K. W. S. Marrett, P. Neelin, A. Vandal, and K. F. A. Evans, "A unified statistical approach for determining significant signals in images of cerebral activation," *Human Brain Mapping* **4**, pp. 58–73, 1999.
23. M. Brammer, "Multidimensional wavelet analysis of functional magnetic resonance images," *Hum Brain Mapp* **6**, pp. 378–382, 1998.
24. F. Turkheimer, M. Brett, D. Visvikis, and V. Cunningham, "Multiresolution analysis of emission tomography images in the wavelet domain," *Journal Cereb Blood Flow Metabol* **19**, pp. 1189–1208, 1999.
25. U. Ruttimann, M. Unser, and D. Rio, "Statistical analysis of image differences by wavelet decomposition," in *Proceedings of the Sixteenth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Engineering Advances: New Opportunities for Biomedical Engineers (EMBS'94)*, **I**, pp. A28–A29, (Baltimore MD, USA), November 3–6 1994.
26. E. Bullmore, M. 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 in Medical Research*, p. in press, 2003.
27. U. Ruttimann, N. Ramsey, D. Hommer, P. Thévenaz, C. Lee, and M. Unser, "Analysis of functional magnetic resonance images by wavelet decomposition," in *Proceedings of the 1995 IEEE International Conference on Image Processing (ICIP'95)*, **I**, pp. 633–636, (Washington DC, USA), October 23–26 1995.
28. U. Ruttimann, M. Unser, P. Thévenaz, C. Lee, D. Rio, and D. Hommer, "Statistical analysis of image differences by wavelet decomposition," in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, eds., ch. 5, pp. 115–144, CRC Press, Boca Raton FL, USA, 1996.
29. 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 Transactions on Medical Imaging* **17**(2), pp. 142–154, 1998.
30. M. Feilner, T. Blu, and M. Unser, "Statistical analysis of fMRI data using orthogonal filterbanks," in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing VII*, **3813**, pp. 551–560, (Denver CO, USA), July 19–23 1999.
31. M. Feilner, T. Blu, and M. Unser, "Analysis of fMRI data using spline wavelets," in *Proceedings of the Tenth European Signal Processing Conference (EUSIPCO'00)*, **IV**, pp. 2013–2016, (Tampere, Finland), September 4–8 2000.
32. S. G. Mallat, *A Wavelet tour of signal processing*, Academic Press, New York, 2nd ed., 1999.
33. F. Abramovich and Y. Benjamini, "Thresholding of wavelet coefficients as multiple hypotheses testing procedure," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., pp. 5–14, Springer-Verlag, New York, 1995.
34. F. Abramovich and Y. Benjamini, "Adaptive thresholding of wavelet coefficients," *Computational Statistics and Data Analysis* **22**, pp. 351–361, 1996.
35. R. T. Ogden and E. Parzen, "Change-point approach to data analytic wavelet thresholding," *Statistics and Computing* **6**(2), pp. 93–99, 1996.
36. M. Clyde, G. Parmigiani, and B. Vidakovic, "Multiple shrinkage and subset selection in wavelets," *Biometrika* **85**(2), pp. 391–401, 1998.
37. F. Abramovich and T. Sapatinas, "Bayesian approach to wavelet decomposition and shrinkage," in *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidakovic, eds., pp. 33–50, Springer-Verlag, New York, 1999.
38. B. Vidakovic and F. Ruggeri, "BAMS method: Theory and simulations," 2001.