

STATISTICAL PARAMETRIC MAPPING OF FUNCTIONAL MRI DATA USING WAVELETS ADAPTED TO THE CEREBRAL CORTEX

Hamid Behjat^{1,2}, Nora Leonardi^{2,3}, Dimitri Van De Ville^{2,3}

¹ Signal Processing Group, Department of Electrical and Information Technology, Lund University, SE-22100 Lund, Sweden

² Medical Image Processing Lab (MIPLAB), École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

³ Department of Radiology and Medical Informatics, University of Geneva, CH-1211 Geneva, Switzerland

ABSTRACT

Wavelet approaches have been successfully applied to the detection of brain activity in fMRI data. Spatial activation patterns have a compact representation in the wavelet domain. However, classical wavelets designed for regular Euclidean spaces are not optimal for the topologically complicated gray-matter (GM) domain where activation is expected. We hypothesized that wavelet bases that are adapted to the structure of the GM, would be more powerful in detecting brain activity. We therefore combine (1) a GM-based graph wavelet transform as an advanced spatial transformation for fMRI data with (2) the wavelet-based statistical parametric mapping framework (WSPM). We introduce suitable design choices for the graph wavelet transform and evaluate the performance of the proposed approach both on simulated and real fMRI data. Compared to SPM and conventional WSPM, the graph-based WSPM shows improved detection of finely 3D-structured brain activity.

Index Terms— Statistical testing, functional MRI, spectral graph theory, graph wavelet transform, wavelet thresholding

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a powerful technique to study brain activity non-invasively. Typically, a method called statistical parametric mapping (SPM) is applied to noisy fMRI data to identify differences in brain activity in individual voxels [1]. The multiple comparison problem is dealt with using the Gaussian Random Field theory [2], which however requires spatial smoothing of the data. Although this step improves the signal-to-noise ratio it also leads to a loss in spatial specificity.

Since we expect brain activity to be highly localized, we can efficiently represent it using wavelets [3], whose strength lie in their power to efficiently represent transient patterns. Wavelet-based statistical parametric mapping (WSPM) is a unified framework that has been proposed to integrate wavelet processing and spatial statistical testing of fMRI data [4, 5]. However, both SPM and WSPM assume the data to be defined on a regular Euclidean space, while brain activity lies on the convoluted structure of the cerebral cortex, also referred to as gray matter (GM). The analysis of fMRI data with classical wavelets has shown great promise, but we expect wavelets that are adapted to the anatomical domain to improve the detection of activation patterns.

In related work, Ozkaya et al. used the lifting scheme to construct wavelet bases adapted to the cerebral cortex [6]. The promising results reported, in spite of shift variance, further motivated our

study. Our proposed framework interprets the cerebral cortex as a graph, where each GM voxel becomes a vertex, and uses a recently proposed graph wavelet transform to define wavelets for this irregular domain [7].

We have integrated the recently developed spectral graph wavelet transform as an advanced transform for fMRI brain mapping into the WSPM framework, and thus provide a novel method to enable statistical processing of fMRI data by taking into account a priori knowledge of the true signal domain. In what follows, we will first have a brief recapitulation on WSPM (Sect. 2.1) and an introduction to spectral graph wavelets (Sect. 2.2). Next, we describe the design of the brain graph (Sect. 2.3) and optimal graph wavelet filters (Sect. 2.4). We present results from an application to both simulated and real fMRI data and compare these to the standard SPM and WSPM (Sect. 3). Finally, we draw some conclusions (Sect. 4).

2. GRAPH WAVELET-BASED ACTIVITY DETECTION

2.1. WSPM Using Spectral Graph Wavelets

Unlike the conventional SPM approach in which a Gaussian filter is initially applied to the data as a means to reduce noise, in wavelet approaches noise is reduced by thresholding the wavelet coefficients. The unique feature of the WSPM framework is to consider thresholding in the wavelet domain as a denoising step only, followed by statistical thresholding of the reconstruction such that the null-hypothesis rejection probability (i.e., false positives) in the spatial domain is properly controlled. We denote the wavelet domain and spatial domain threshold parameters as τ_w and τ_s , respectively.

2.1.1. Wavelet Domain Model Fitting

SPM is a mass univariate approach where a general linear model (GLM) is fitted to the time evolution of each voxel and a statistic (e.g. a t value) is calculated for each voxel. WSPM computes the discrete wavelet transform (DWT) for each scan and the model fitting is done on the temporal behavior of each of the wavelet coefficients using a GLM with the same N_r experimental regressors as in conventional SPM, as $\mathbf{y}_w^k = X\boldsymbol{\beta}_w^k + \boldsymbol{\epsilon}_w^k$, where \mathbf{y}_w^k is a $N_t \times 1$ vector which represents the temporal change in intensity of a wavelet coefficient, X is a design matrix of size $N_t \times N_r$, $\boldsymbol{\beta}_w^k$ is a $N_r \times 1$ vector of regression parameters, $\boldsymbol{\epsilon}_w^k$ is a $N_t \times 1$ vector of residual errors and k indexes the coefficients of the scaling function and wavelets at all scales and orientations.

2.1.2. Wavelet Domain Processing

Next, the data is denoised by thresholding the wavelet domain t-values, which at the k-th wavelet coefficient are computed as

$$t_w^k = \frac{\mathbf{c}^T \boldsymbol{\beta}_w^k}{\boldsymbol{\sigma}_w^k \sqrt{\mathbf{c}^T (X^T X)^{-1} \mathbf{c}}} \quad (1)$$

$$\boldsymbol{\sigma}_w^k = \sqrt{\frac{\boldsymbol{\epsilon}_w^{k,T} \boldsymbol{\epsilon}_w^k}{\text{tr}(I - X X^+)}} \quad (2)$$

where \mathbf{c} is an appropriate $N_r \times 1$ contrast vector, $\boldsymbol{\sigma}_w^k$ is the estimate of the standard deviation at the k-th wavelet coefficient, and X^+ is the pseudo-inverse of the design matrix.

By thresholding the wavelet domain t-values by τ_w , and using the inverse DWT, the reconstructed *processed* parameter map after bias reduction, $\tilde{\mathbf{u}}$, is computed as

$$\tilde{\mathbf{u}} = \min \left(\sum_{k=1}^{N_c} \mathbf{u}_w[k] \boldsymbol{\psi}_k, \sum_{k=1}^{N_c} H(|t_w^k| - \tau_w) \mathbf{u}_w[k] \boldsymbol{\psi}_k \right) \quad (3)$$

where \mathbf{u}_w are the wavelet coefficients and H is the Heaviside step function [4].

The final parameter map of detected activations results from a second thresholding step applied to the spatial domain t-values as

$$\bar{\mathbf{u}} = H \left(\frac{\tilde{\mathbf{u}}}{\sum_{k=1}^{N_c} \boldsymbol{\sigma}_w^k |\boldsymbol{\psi}_k|} - \tau_s \right) \tilde{\mathbf{u}}. \quad (4)$$

The optimal combination of τ_w and τ_s minimizes the worst-case error between the unprocessed and the detected parameter map [4].

2.2. Spectral Graph Wavelet Transform

2.2.1. Graphs and their Eigenspace

An undirected graph $G = (V, E)$ is described by the set of vertices V and edges E , where each edge is denoted by a pair of unordered indices, (i, j) . For a weighted graph with N_v vertices and no loops, the adjacency matrix, A , and the diagonal matrix, D , of the graph are given by

$$A_{i,j} = \begin{cases} e_{i,j} & \text{if } (i, j) \text{ is an edge,} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

$$D_{i,i} = \sum_j A_{i,j}, \quad (6)$$

respectively, where $e_{i,j}$ indicate the edge weights. Note that both A and D are symmetric $N_v \times N_v$ matrices. The normalized graph Laplacian matrix, \mathcal{L} , is then computed as $\mathcal{L} = I - D^{-1/2} A D^{-1/2}$. \mathcal{L} is symmetric and positive semi-definite, and its eigenvalue decomposition $\mathcal{L} \chi_l = \lambda_l \chi_l$ leads to a set of N_v real, non-negative eigenvalues, λ_l , and a corresponding set of eigenvectors, $\{\chi_l\}$, which form a complete set of orthonormal vectors [8].

2.2.2. Spectral Graph Wavelets

For the classical wavelet transform, wavelets are formed by shifting and scaling a mother wavelet ψ , as $\psi_{s,t}(x) = \frac{1}{s} \psi\left(\frac{x-t}{s}\right)$, where $\psi_{s,t}(x)$ indicates the wavelet at scale s and location t . It is, however,

not straightforward to generalize these operations to graphs. Hammond et al. [7] have tackled this issue by solving the problem in the Fourier domain. The classical continuous wavelet in the Fourier domain is given by

$$\psi_{s,t}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\psi}(s\omega) e^{-j\omega t}(\omega) e^{j\omega x} d\omega \quad (7)$$

where the scaling parameter s solely affects the argument of $\hat{\psi}$. An interesting observation is that the complex exponentials, $e^{j\omega x}$, which are the basis functions of the Fourier transform, can also be seen as the eigenfunctions of the 1-D Laplacian operator, i.e. $\frac{d^2}{dx^2}$. The analogue of the Fourier domain for graphs is then defined as the eigenspace of the graph Laplacian matrix [8]. The graph Fourier transform of a graph signal f is given by $\hat{f} = \langle f, \chi_\ell \rangle$.

The spectral graph wavelets with J scales and the scaling function are then obtained as

$$\psi_{j,n}(m) = \sum_{l=0}^{N_v-1} g(t_j \lambda_l) \chi_l^*(n) \chi_l(m), \quad j = 1, \dots, J \quad (8)$$

$$\phi_n(m) = \sum_{l=0}^{N_v-1} h(t_J \lambda_l) \chi_l^*(n) \chi_l(m) \quad (9)$$

where g is a suitable wavelet generating kernel, t_j indicates the wavelet scale, n the vertex and h the scaling function. Finally, the spectral graph wavelet and scaling coefficients are obtained from the inner product of the wavelets ϕ and scaling function ψ with f as

$$w_\psi(j, n) = \sum_{l=0}^{N_v-1} g(t_j \lambda_l) \hat{f}(l) \chi_l(m), \quad j = 1, \dots, J \quad (10)$$

$$w_\phi(n) = \sum_{l=0}^{N_v-1} h(t_J \lambda_l) \hat{f}(l) \chi_l(m). \quad (11)$$

We refer to [7] for further details and an efficient approximation for large graphs with several thousands of nodes.

2.3. Brain Graph Design

In order to obtain wavelets adapted to the cerebral cortex, we need to define an appropriate set of graph vertices and edges. We propose the following procedure: first, the brain is segmented into gray matter (GM), white matter and cerebrospinal fluid, which are then normalized to the standard Montreal Neurological Institute (MNI) space with a voxel size of $2 \times 2 \times 2 \text{ mm}^3$. Since the GM is where we expect the brain activity to be localized, we construct a brain graph by thresholding the GM probability map, typically at 50% (i.e., indicating that the voxel is at least 50% likely to be GM). Moreover, within the resulting mask, isolated regions smaller than 20 voxels, which can stem from the normalization and thresholding steps, are removed. We treat the remaining voxels as graph vertices, and compute three-dimensional connections between adjacent voxels assuming a 26-connected neighborhood. The edges of the brain graph were then weighted according to the GM probabilities of the vertices at each end. The weight of the edge between vertices i and j was defined as $e_{i,j} = (P_i P_j)^5$. Empirically, we found that this choice led to realistic activation patterns as it heavily penalizes connections between vertices which have a GM probability $\leq 90\%$, and on the other hand, enhances the diffusion of the wavelets along regions of GM with high probability.

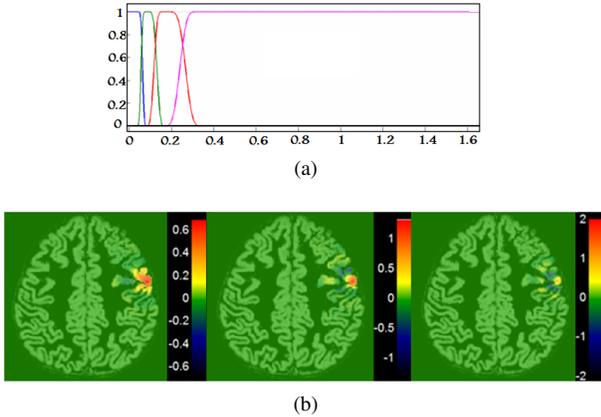


Fig. 1: (a) The partitioning of the graph spectrum by the scaling function (blue curve) and wavelet generating kernels for $J = 3$ scales. (b) Corresponding anatomically adapted scaling function (left) and wavelet kernels at the two lower spectrum scales, localised and overlaid on the cerebral cortex.

2.4. Brain Graph Wavelets

We constructed wavelets adapted to the underlying structure of the cerebral cortex using the normalized graph Laplacian matrix \mathcal{L} , and a tight frame design, which conserves energy in the wavelet domain [9]. However, we optimized the design of the wavelet and scaling function filters for the application at hand. Our brain graphs are very large (N_v typically $> 40'000$), and we have found it beneficial to focus the partitioning to approximately the lower 20% of the spectrum as the associated eigenvectors carry the most relevant information; in other words, the upper 80% of the spectrum is captured by the first scale wavelet, and the wavelets at the remaining scales and the scaling function partition the lower end of the spectrum (Fig. 1).

The computationally most costly part of the implementation is the absolute-value wavelet reconstruction needed for the spatial thresholding in (4). Using the EPFL Greedy cluster, we obtained results overnight by using up to 30 computation nodes (approximately $30 \times 8 = 240$ hours). Conventional WSPM, using 3-D wavelets, takes up to an hour on a single node (i.e. a modern desktop computer). The computational burden of sgWSPM can be reduced in multi-subject studies by precomputing the absolute-value-wavelets for a fixed GM graph in some normalized space (e.g. MNI).

3. EXPERIMENTAL RESULTS

3.1. Synthetic Data

A synthetic dataset was constructed from a real T1-weighted anatomical image, which was first segmented and then normalized to the MNI space with a voxel size of $2 \times 2 \times 2 \text{ mm}^3$. The GM was then thresholded at 70%, binarized, and isolated regions were removed, resulting in an unweighted graph. An activation pattern \mathbf{y} that diffuses from multiple center points along the domain of the graph was created by consecutive applications of the adjacency matrix A to an activation vector, \mathbf{x} , which contained two activations centers in the central slices: $\mathbf{y} = R(10^{-i} A^i \mathbf{x})$, where i defines the extent of the diffusion and R is a nonlinear transformation applied to suppress increasingly high activation values and limit the values to the range $[0, 1]$ (Fig. 2a). Next, 60 “functional” volumes were cre-

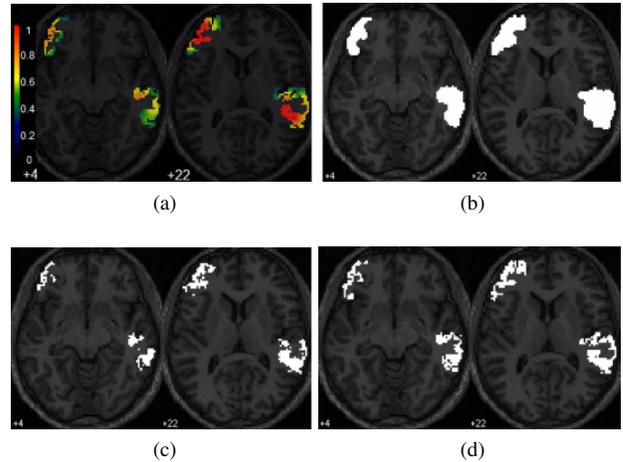


Fig. 2: Synthetic data set. a) Ground truth, b)-d) detected parameter maps using SPM, WSPM and sgWSPM, respectively.

Method	SPM (6 mm)	WSPM		sgWSPM	
		Wavelet scales	1	3	1
Detections	7805	3218	2631	4124	4171
Sensitivity	95%	64%	54%	81%	83%
FPS	3349	208	120	298	270

Table 1: Statistics on the detections of SPM, classical WSPM and sgWSPM for the synthetic data set. Results are presented for a single as well as three wavelet scales. False positives are reported instead of specificity because of the high number of true negatives (approx. $43'000$).

ated by corrupting \mathbf{y} with additive white Gaussian noise ($\sigma = 2$). Without any loss of generality, the GLM consisted of a single (constant) regressor, which makes it equivalent to a one-sample t-test.

We implemented conventional SPM (isotropic Gaussian smoothing with $\text{FWHM} = 6 \text{ mm}^1$) and WSPM using (1) standard orthogonal wavelets (3-D, redundant), and (2) graph wavelets (sgWSPM), both with one and three wavelet scales. Table 1 summarizes the performance of the three frameworks. SPM has the highest sensitivity (95%), at the cost of detecting a large number of false positives (3349 out of 7805 detections). sgWSPM shows a large improvement in sensitivity over WSPM using both a single scale as well as three wavelet scales (81% and 83% vs. 64% and 54%), while also increasing the number of false positives (298 and 270 vs. 208 and 120), which remain, however, a fraction of the total number of detections. In general, the number of false positives is much smaller for both wavelet approaches than for SPM, a result of the non-linear denoising procedure in the wavelet domain. We also note that sgWSPM using three wavelet scales leads to a slightly better sensitivity and specificity compared to using a single wavelet scale, whereas for classical wavelets the sensitivity drops when using three scales, as previously observed [6]. Thus, in what follows, we present the detected maps resulting from using a single scale for the classical wavelets and three scales for the graph wavelets, in analysing both the synthetic data as well as the real data.

Fig. 2 shows the detected maps for a corrected significance level

¹A minimal amount of smoothing is required in SPM for the application of a threshold by Gaussian random field theory.

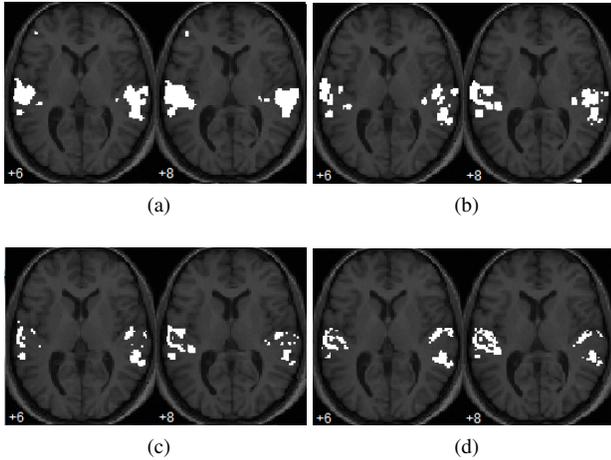


Fig. 3: Detected activations for single-subject auditory fMRI data. Detected parameter maps using (a) SPM, (b) WSPM, (c) WSPM with detections outside GM50 region removed, and (d) sgWSPM.

of 5%. Note that the detection maps are binarized due to the difference in t-values of the different approaches and to minimize visual bias in their interpretation. We observe that WSPM and sgWSPM detect finely structured activation patterns in contrast to SPM. Also, many of the weak activations (green and blue color in Fig. 2a) are well detected by sgWSPM but missed by WSPM.

3.2. Real Data

Next, we applied the proposed approach to single-subject fMRI data with alternating cycles of auditory stimulation and resting [10]. The data were acquired on a 2T scanner with the following parameters: $TR = 7s$, 64 contiguous slices, $3 \times 3 \times 3 \text{ mm}^3$ voxels, 96 volumes.

To evaluate the performance of the three approaches, we introduce an additional measure that we find interesting and intuitive in interpreting the results in the absence of the ground truth: *GM50* indicates the number of detections that lie in a domain with a GM probability of at least 50%. Due to the sparse vascularization and low hemodynamic response efficiency of white matter [11], and the absence of blood vessels within the CSF, it is not expected to observe brain activity within these tissue types. Thus, as an empirical measure and in a loose sense, we can think of detections that lie outside of *GM50* to likely be false positives.

The detected parameter maps for the three methods are shown in Fig. 3. As for the synthetic data set, SPM detects large blobs of activation because of the smoothing applied to the dataset (Fig. 3a). Standard WSPM detects “blocky” activations in the auditory cortex (e.g. we observe little cubes; Fig. 3b), of which many lie outside *GM50* (Fig. 3c), while detections by sgWSPM follow the complex three-dimensional structure of the cerebral cortex (Fig. 3d).

As can be seen in Table 2, the total number of detections using SPM or WSPM is much larger than when using sgWSPM (3229, 3405 and 1688 respectively). However, only around two-thirds of these detections lie within *GM50* (2145/3229 for SPM, and 2024/3405 for WSPM), suggesting a larger number of false positives, and thus reduced specificity, for these two approaches.

Method	SPM (6 mm)	WSPM	sgWSPM
Detections	3239	3405	1688
GM50	2145	2024	1688

Table 2: Statistics on the detections of SPM, classical WSPM (1 scale) and sgWSPM (3 scales) for the real data set. *GM50* denotes the number of detections in voxels with a GM probability of at least 50%.

4. CONCLUSION

Building on the wavelet-based SPM (WSPM) framework, we have proposed a new approach for statistical analysis of fMRI data using graph wavelets, where the graph is designed based on a-priori knowledge of the structure of the cerebral cortex. This approach enables an efficient spatial adaptation to the true anatomical domain where the observed data is expected to lie, as well as to encode brain activity at fine scales. Evaluations on simulated data have shown superior performance in terms of sensitivity (compared to classical WSPM) and good control of type I errors (as opposed to SPM). The application to real fMRI data showed the detection of finely structured activations along the structure of the cerebral cortex. The extension of the framework to multi-subject studies is under investigation and will be presented in future work.

5. ACKNOWLEDGEMENTS

This work was supported in part by the Swiss National Science Foundation (grant PP00P2-123438) and in part by the CIBM.

6. REFERENCES

- [1] 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, 1994.
- [2] J. B. Poline, K. J. Worsley, A. C. Evans, and K. J. Friston, “Combining spatial extent and peak intensity to test for activations in functional imaging,” *Neuroimage*, vol. 5, pp. 83–96, 1997.
- [3] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, “Wavelets and functional magnetic resonance imaging of the human brain,” *Neuroimage*, vol. 23 Suppl 1, pp. S234–S249, 2004.
- [4] D. Van De Ville, T. Blu, and M. Unser, “Integrated wavelet processing and spatial statistical testing of fMRI data,” *Neuroimage*, vol. 23, pp. 1472–1485, 2004.
- [5] D. Van De Ville, M.L. Seghier, F. Lazeyras, T. Blu, and M. Unser, “WSPM: wavelet-based statistical parametric mapping,” *Neuroimage*, vol. 37, pp. 1205–1217, 2007.
- [6] S.G. Ozkaya and D. Van De Ville, “Anatomically adapted wavelets for integrated statistical analysis of fMRI data,” in *Proc. IEEE Int. Symp. Biomedical Imaging*, Chicago, IL, 2011, pp. 469–472.
- [7] D.K. Hammond, P. Vandergheynst, and R. Gribonval, “Wavelets on graphs via spectral graph theory,” *Appl Comput Harmon Anal*, vol. 30, pp. 129–150, 2011.
- [8] F.R.K. Chung, *Spectral graph theory*, AMS, Providence, RI, 1997.
- [9] N. Leonardi and D. Van De Ville, “Wavelet frames on graphs defined by fMRI functional connectivity,” in *Proc. IEEE Int. Symp. Biomedical Imaging*, Chicago, IL, 2011, pp. 2136–2139.
- [10] K. Friston G. Rees, “Single subject epoch (block) auditory fMRI activation data,” <http://www.fil.ion.ucl.ac.uk/spm/data/>, 1999.
- [11] N.K. Logothetis and B.A. Wandell, “Interpreting the BOLD signal,” *Annu Rev Physiol*, vol. 66, pp. 735–769, 2004.