WSPM OR HOW TO OBTAIN STATISTICAL PARAMETRIC MAPS USING SHIFT-INVARIANT WAVELET PROCESSING

Dimitri Van De Ville, Thierry Blu, Michael Unser

Biomedical Imaging Group
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

Recently, we have proposed a new framework for detecting brain activity from fMRI data, which is based on the spatial discrete wavelet transform. The standard wavelet-based approach performs a statistical test in the wavelet domain, and therefore fails to provide a rigorous statistical interpretation in the spatial domain. The new framework provides an "integrated" approach: the data is processed in the wavelet domain (by thresholding wavelet coefficients), and a suitable statistical testing procedure is applied afterwards in the spatial domain. This method is based on conservative assumptions only and has a strong type-I error control by construction. At the same time, it has a sensitivity comparable to that of SPM. Here, we discuss the extension of our algorithm to the redundant discrete wavelet transform, which provides a shift-invariant detection scheme. The key features of our technique are illustrated with experimental results. An implementation of our framework is available as a toolbox (WSPM) for the SPM2 software.

1. INTRODUCTION

Parametric hypothesis-driven methods are commonly used for the analysis of fMRI data. SPM [1, 2] is one of the most popular software packages for this type of analysis. Its main characteristics are the preprocessing of the data by a fixed Gaussian filter and the application of Gaussian random field theory to control the multiple hypothesis testing problem.

The "wavelet community" has come up with an alternative which is to replace the Gaussian prefilter by the spatial discrete wavelet transform (DWT) [3]. The basic idea is to apply a statistical test to the wavelet representation of the parameter map, which is obtained after modelling the time-courses with a linear model (LM). Thanks to the sparsifying and decorrelating properties of the DWT, the sensitivity of such a test is higher than in the spatial domain. Unfortunately, there is a fundamental problem to map back the statistical significance in the spatial domain and determine whether a particular voxel can be considered as activated.

Recently, we proposed a novel framework [4] where we determined the statistical significance (hence testing) in the spatial domain, taking into account the influence of prior wavelet processing (=denoising). The method is briefly summarized in Sect. 2.

One of the major disadvantages of the discrete wavelet transform (DWT) is its shift variance. In this paper, we investigate the use of the redundant DWT and its influence on the framework. We also consider the combination of multiple non-redundant DWTs. The feasibility of the proposed approach is experimentally demonstrated. We report an increased sensitivity of a factor 2 while maintaining strong type-I error control.

2. THE INTEGRATED FRAMEWORK

We explain the integrated framework for wavelet-based statistical analysis step-by-step, as shown in Fig. 1 (a). Consider the fMRI dataset $v[\mathbf{n};t]$, where $\mathbf{n}$ is the spatial index and $t = 1,\ldots,N_t$ the temporal one. We assume that every volume has been realigned and eventually normalized, but not smoothed.

2.1. Spatial representation using the DWT

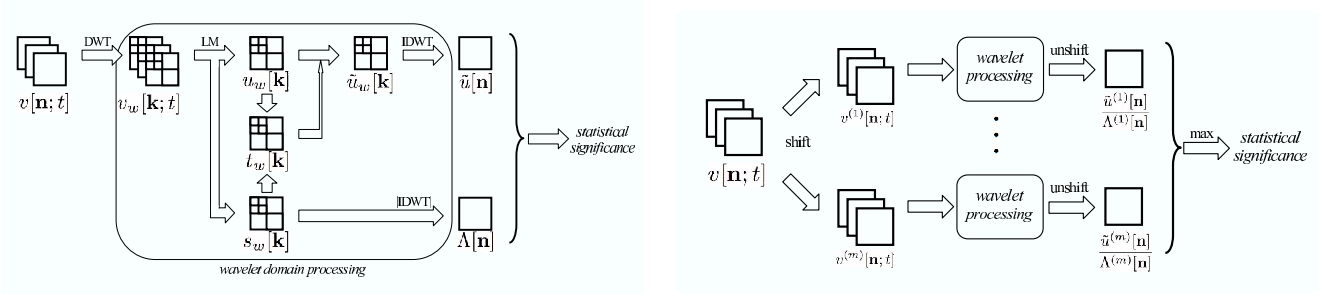
First, the DWT is applied along the spatial dimensions to each volume. The wavelet representation of the signal is

$$v[\mathbf{n};t] = \sum_{\mathbf{k}} v_w[\mathbf{k};t]\psi_{\mathbf{k}}[\mathbf{n}], \tag{1}$$

where $v_w[\mathbf{k};t]$ denotes the wavelet coefficients and where $\psi_{\mathbf{k}}$ is the corresponding wavelet or scaling function; the index $\mathbf{k}$ runs over all subbands and orientations. The DWT is a decomposition into a basis, providing non-redundancy and perfect reconstruction. The DWT is implemented as a fast iterated filterbank.

2.2. Temporal modelling using the LM

The wavelet coefficients are now arranged in a time-course vector as $\mathbf{v}_w[\mathbf{k}] = [v_w[\mathbf{k};1]\ldots v_w[\mathbf{k};N_t]]^{\mathrm{T}}$. For each index



(a) (b)

Fig. 1. An overview of the integrated framework for wavelet-based statistical analysis. (a) The standard method. (b) The extension of the method to improve for shift-invariance by combining multiple shifts of the data.

$\mathbf{k}$, we setup a LM that is intended to explain temporal evolution of $\mathbf{v}_w[\mathbf{k}]$:

$$\mathbf{v}_w[\mathbf{k}] = \mathbf{X}\mathbf{y}_w[\mathbf{k}] + \mathbf{e}_w[\mathbf{k}], \tag{2}$$

where the design matrix $\mathbf{X}$ contains L regressors (i.e., basis functions related to the experiment), $\mathbf{y}_w[\mathbf{k}]$ is a parameter vector of length L, and $\mathbf{e}_w[\mathbf{k}]$ is the residual error. In the basic form of the LM, the residuals are assumed to be (temporally) independently and identically gaussian-distributed, which allows us to estimate $\bar{\mathbf{y}}_w[\mathbf{k}]$ by the pseudo-inverse of $\mathbf{X}$, and the residuals as $\bar{\mathbf{e}}_w[\mathbf{k}] = \mathbf{v}_w[\mathbf{k}] - \mathbf{X}\bar{\mathbf{y}}_w[\mathbf{k}]$. The signal of interest, which will be used to detect activity later on, is extracted by the contrast vector $\mathbf{c}$:

$$u_w[\mathbf{k}] = \mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}}_w[\mathbf{k}], \tag{3}$$
$$s_w^2[\mathbf{k}] = \bar{\mathbf{e}}_w[\mathbf{k}]^{\mathrm{T}}\bar{\mathbf{e}}_w[\mathbf{k}]\mathbf{c}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{c}. \tag{4}$$

Often, the general linear model (GLM) is used, which allows us to deal with temporal correlations of the noise.

2.3. Denoising in the wavelet domain

For each wavelet "location" $\mathbf{k}$, we have an estimated contrast value $u_w[\mathbf{k}]$, and a residual error $s_w[\mathbf{k}]$. In the absence of activation (so the wavelet coefficient only "captures" noise), the proportion

$$t_w[\mathbf{k}] = \frac{u_w[\mathbf{k}]}{\sqrt{s_w^2[\mathbf{k}]/J}}, \text{ with } J = N_t - \mathrm{rank}(\mathbf{X}), \tag{5}$$

should follow a Student t-distribution with J degrees of freedom. We propose to hard-threshold the coefficients $u_w[\mathbf{k}]$ based on this t-value. The reconstruction in the spatial domain can then be written as

$$\tilde{u}[\mathbf{n}] = \sum_{\mathbf{k}} \tilde{u}_w[\mathbf{k}]\psi_{\mathbf{k}}(\mathbf{n}), \text{where} \begin{cases} \tilde{u}_w[\mathbf{k}]{=}u_w[\mathbf{k}], & \text{when } |t_w[\mathbf{k}]|{>}\tau_w, \\ 0, & \text{otherwise,} \end{cases}$$

where τ_w is a threshold parameter to be determined later.

2.4. Statistical testing in the spatial domain

The statistical hypotheses are formulated in the spatial domain as

$$\mathcal{H}_0 : E[\tilde{u}[\mathbf{n}]] = 0, \quad \mathcal{H}_1 : E[\tilde{u}[\mathbf{n}]] > 0. \tag{6}$$

The null hypothesis is rejected based on the following inequality:

$$\mathrm{Prob}\left[\frac{\tilde{u}[\mathbf{n}]}{\Lambda[\mathbf{n}]} \geq \tau_s\right] \leq \Upsilon(\tau_w, \tau_s), \tag{7}$$

where $\Upsilon(\tau_w, \tau_s)$ is data-independent, and $\Lambda[\mathbf{n}]$ is a "rectified" reconstruction of the residuals:

$$\Lambda[\mathbf{n}] = \sum_{\mathbf{k}} \frac{s_w[\mathbf{k}]}{\sqrt{J}} \left|\psi_{\mathbf{k}}(\mathbf{n})\right|. \tag{8}$$

The threshold values τ_w and τ_s are obtained by choosing a desired global significance level α, which fixes $\Upsilon(\tau_w, \tau_s) = \alpha/N$, where N is the number of intra-cranial voxels to compensate for multiple testing. The proportion $\tilde{u}[\mathbf{n}]/\Lambda[\mathbf{n}]$ can be interpreted as a meaningful statistical parametric map in the spatial domain.

3. EXTENSIONS

One of the major disadvantages of the non-redundant DWT is its shift-variance. A 1-D example is shown to illustrate the main ideas: a small Gaussian-like activation $u_0[n]$ is corrupted with additive white noise ($N_t = 80$ realizations, $\sigma_n = 0.25$). We apply the framework using the orthogonal B-spline wavelet (degree 2, 1 iteration) and fixing the significance level at $\alpha = 0.5\%$, which gives $\tau_w = 4.53$, $\tau_s = 0.22$. Using the non-redundant DWT, the activation is detected depending on its position, as can be seen in Fig. 2 (a) and (b). The shifted input signal of (b) has a less compact representation in the wavelet domain, which changes the result of the thresholding step.

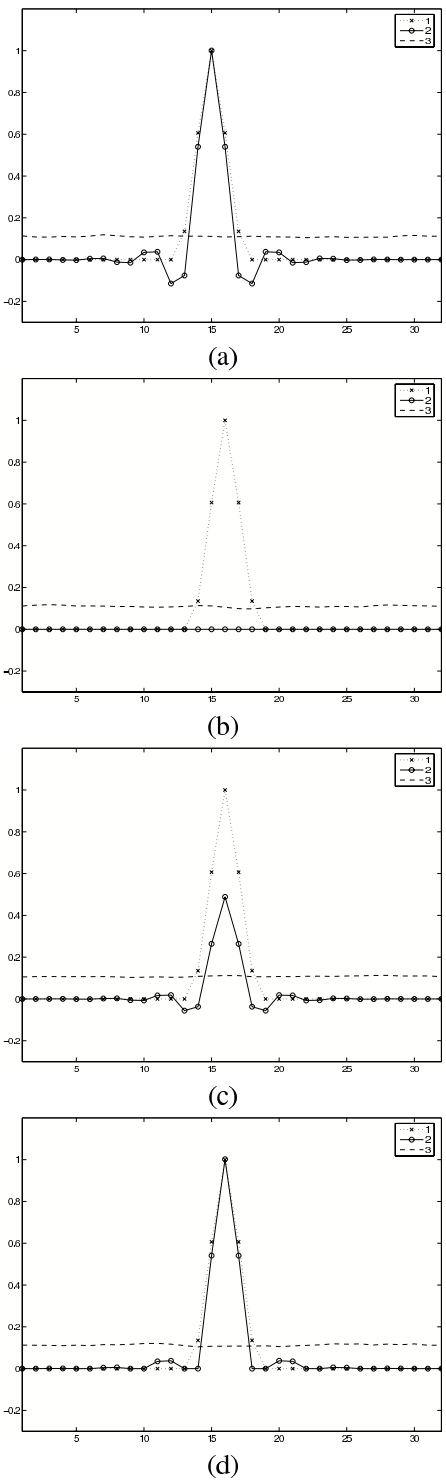


Fig. 2. 1-D example. 1: ground truth $u_0[n]$; 2: adaptive threshold $\tau_s\Lambda[n]$; 3: reconstructed contrast $\tilde{u}[n]$. (a) Non-redundant DWT, activation is placed at an even position. (b) Non-redundant DWT, activation is placed at an odd position. (c) Redundant DWT. (d) Multiple non-redundant DWTs.

3.1. Redundant DWT

The redundant DWT can be obtained by the "à trous" algorithm, which does not subsample the data but modifies the filters accordingly. For a DWT with N_s subbands and J_w iterations, the redundancy factor is $(N_s-1)\cdot J_w+1$. While the redundant DWT is shift-invariant, the required storage space and computational load quickly increases as the number of iterations goes up, in particular for multi-dimensional data.

The redundant DWT can be plugged into the framework without any modification. This is particularly tempting since the threshold values τ_w and τ_s remain unchanged too. Unfortunately, the energy captured by a single coefficient is less than in the non-redundant case. Consequently, if the same number of coefficients survive thresholding, the reconstructed parameter map $\tilde{u}[\mathbf{n}]$ will have lower values, and thus a higher risk to be undetected in the spatial domain. The 1-D example, in Fig. 2 (c), shows the decrease of $\tilde{u}[n]$ and thus also the significance $\tilde{u}[n]/\Lambda[n]$.

3.2. Multiple non-redundant DWTs

We propose to mitigate this problem by analyzing the data under M different shifts: the data volumes are shifted by $\mathbf{d}^{(m)}$, $m=1,\ldots,M$, each set is analyzed using the non-redundant DWT, and the results unshifted by $-\mathbf{d}^{(m)}$. Thus, for each shift, we still have the null hypothesis condition

$$\text{Prob}\left[\frac{\tilde{u}^{(m)}[\mathbf{n}]}{\Lambda^{(m)}[\mathbf{n}]}\geq\tau_s\right]\leq\Upsilon(\tau_w,\tau_s). \tag{9}$$

We combine these results by selecting the shift that has the highest statistical significance:

$$\begin{aligned}
&\text{Prob}\left[\max_m\left(\frac{\tilde{u}^{(m)}[\mathbf{n}]}{\Lambda^{(m)}[\mathbf{n}]}\right)\geq\tau_s\right]\\
&= \text{Prob}\left[\bigvee_{m=1}^{M}\tilde{u}^{(m)}[\mathbf{n}]\geq\tau_s\Lambda^{(m)}[\mathbf{n}]\right]\\
&\leq \sum_{m=1}^{M}\text{Prob}\left[\tilde{u}^{(m)}[\mathbf{n}]\geq\tau_s\Lambda^{(m)}[\mathbf{n}]\right]\\
&\leq M\Upsilon(\tau_w,\tau_s).
\end{aligned}$$

Clearly, the penalty to pay is the redundancy factor M, which increases the threshold values τ_w and τ_s for a fixed α (to guarantee the same strong type I error control). In practice, M can be kept quite low and the "best" result of each shift is obtained.

For the 1-D example, the redundancy is $M=2$, which fixes the new threshold values at $\tau_w=4.69$, $\tau_s=0.21$. In Fig. 2 (d), the activation is recovered without any loss of energy or significance.

The approach allows us to control the redundancy factor. To be truly shift-invariant, it requires $N_s\cdot J_w$ shifts. For example, 4 shifts in the XY-plane ([0 0], [1 0], [0 1], [1 1]) would

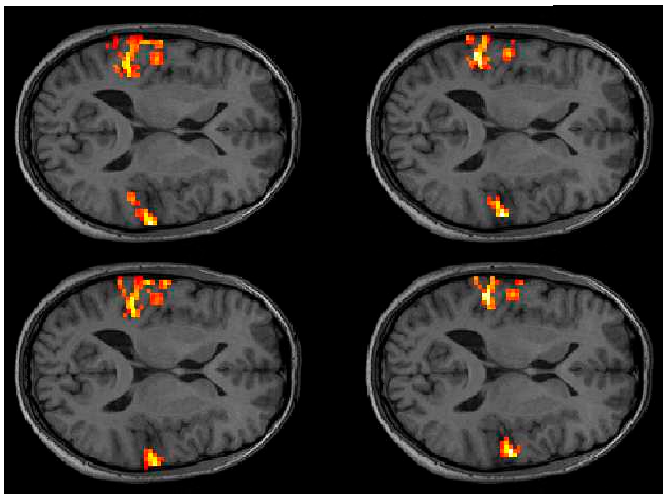


Fig. 3. Wavelet-based parametric maps using the non-redundant DWT. Influence of a spatial shift on the detection results. The number of detected voxels, from top to bottom, left to right: 452, 388, 403, 464.

be shift-invariant for $J_w = 1$. For a higher number of iterations, this redundancy is still satisfactory (giving quasi-shift-invariance), with a reasonable increase of storage space and computational load.

4. RESULTS

We show experimental results for a block-based paradigm [5] (Siemens 2T Magneton, auditory stimulation, repetition time $7s$, spatial dimensions $64 \times 64 \times 64$, voxel size $3 \times 3 \times 3$mm, 84 volumes). Realignment and setup of the GLM (canonical HRF; highpass filter ($1/168$Hz); AR(1) model for serial correlation of the noise) was done using SPM2. We fix the global significance level at 5%. Results of activation maps are shown for a slice that cuts through the auditory cortex, and mapped on a T_1 anatomical scan as background.

The framework was applied using 2-D (slice-by-slice) separable orthogonal B-spline wavelets (degree 1, 1 iteration). WSPM uses the complete GLM setup of SPM2, including the estimation procedure for the AR model and the Satterthwaite approximation for the degrees of freedom. The threshold values for the desired significance level correspond to $\tau_w = 5.47$, $\tau_s = 0.18$. In Fig. 3, we show the activation maps for the non-redundant DWT (top left) and also for different spatial shifts (horizontal, vertical, and diagonal direction). Clearly, the influence of such a shift in not negligible. Not only the detected patterns differ, but the total number of detected voxels also varies between 388 and 464.

The sensitivity drops significantly when we apply the redundant DWT; the total number of detections being only 187.

The proposed combination of multiple non-redundant DWTs (for the 4 different shifts introduced before) fixes the threshold values at $\tau_w = 5.72$ and $\tau_s = 0.17$. Despite the higher threshold in the wavelet domain, the total number of detected voxels attains 812, which is a remarkable improvement compared to the non-redundant approach. The detections seem to follow well the transverse temporal gyrus. As a reference, the result of SPM2 for 5mm smoothing is shown in Fig. 4 (left), corresponding to 430 detected voxels.

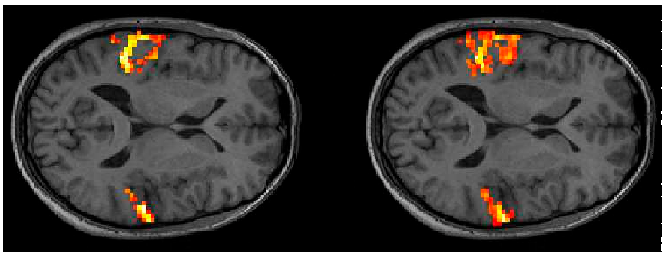


Fig. 4. Left: Parametric map of SPM2 with 5mm smoothing (430 detected voxels). Right: Wavelet-based parametric map using the combination of the 4 non-redundant DWTs of Fig. 3 (812 detected voxels).

For datasets with higher resolution, thus smaller voxel size, it is advantageously to increase the number of iterations of the DWT. At the same time, keeping the same number of spatial shift seems to be a good compromise.

5. CONCLUSION

The redundant DWT presented itself as the ideal candidate to extend the framework for shift-invariant analysis. However, next to the high demands for storage and computation, the redundant DWT results into a significantly lower sensitivity. Therefore, we proposed the combination of multiple non-redundant DWTs, with a controlled degree of redundancy, resulting in higher sensitivity. Future research should also concentrate to improve the directionality of the 2-D DWT, see [6].

The WSPM toolbox is freely available to the research community at `http://bigwww.epfl.ch/wspm`.

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," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.

[2] R. Frackowiak, K. Friston, C. Frith, R. Dolan, and J. Mazziotta, *Human Brain Function*, Academic Press, 1997.

[3] D. Van De Ville, T. Blu, and M. Unser, "Surfing the brain: An overview of wavelet-based techniques for fMRI data analysis," *IEEE Engineering in Medicine and Biology Magazine*, p. in press, 2005.

[4] 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.

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

[6] L. Sendur, V. Maxim, B. Whitcher, and E. Bullmore, "Multiple hypothesis and mapping of functional MRI data in orthogonal and complex wavelet domains," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3413–3426, Sept. 2005.