WAVELET-BASED fMRI STATISTICAL ANALYSIS AND SPATIAL INTERPRETATION: A UNIFYING APPROACH

Dimitri Van De Ville, Thierry Blu, Michael Unser

Biomedical Imaging Group
Swiss Federal Institute of Technology Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

ABSTRACT

Wavelet-based statistical analysis methods for fMRI are able to detect brain activity without smoothing the data. Typically, the statistical inference is performed in the wavelet domain by testing the t-values of each wavelet coefficient; subsequently, an activity map is reconstructed from the significant coefficients. The limitation of this approach is that there is no direct statistical interpretation of the reconstructed map. In this paper, we propose a new methodology that takes advantage of wavelet processing but keeps the statistical meaning in the spatial domain. We derive a spatial threshold with a proper non-stationary component and determine optimal threshold values by minimizing an approximation error. The sensitivity of our method is comparable to SPM's (Statistical Parametric Mapping).

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) has become a key modality in neuroscience. The detection of a weak and noisy activity signal requires statistical analysis. One of the widely deployed and recognized methods for fMRI analysis is SPM [1]. In SPM, one typically applies a Gaussian prefilter to reduce the effect of the measurement noise. Statistical inference is based on the theory of continuous Gaussian random fields.

The "wavelet community" has proposed an alternative approach where the Gaussian prefilter is replaced by a spatial wavelet transform [2, 3]. The detection of functional activity is performed in the wavelet domain by applying a coefficient-wise t-test. An activity map is reconstructed from the coefficients that are marked as significant. This reconstructed map is very useful for visualization purposes, but is not directly interpretable in statistical terms because each activated wavelet coefficients contributes to many voxels. This problem is often dealt with by applying an ad-hoc spatial threshold to the reconstructed image; researchers have also proposed to re-test in the spatial domain [4], or to recursively reconstruct the activity map by controlling the false discovery rate [5].

In this paper, we re-examine the initial wavelet-based scheme from a different point of view. Our basic philosophy is to reconstruct an intermediate activity map from *processed* (but not tested!) wavelet coefficients, and to test it afterwards in the spatial domain. For this goal, we derive a proper non-stationary spatial threshold. From the point of view of implementation, the difference with respect to the standard approach is minimal. However, the statistical interpretation in the spatial domain adds an important advantage to the current method.

We have implemented our new method as a toolbox for SPM. With an example, we show that its sensitivity is comparable to that of SPM.

2. CLASSICAL WAVELET-BASED METHOD INCORPORATING THE GLM

We briefly review the wavelet-based method, which is extended with the general linear model (GLM) to easily set up experiments and incorporate the effect of the hemodynamic response function (HRF) [6, 7, 8].

Consider an MRI dataset $v_{\mathbf{n}}(t)$, $\mathbf{n} \in \mathbb{Z}^3$, $t \in \mathbb{Z}$, where $\mathbf{n}$ and $t = 1, \ldots, N_t$ are the 3D-spatial and temporal indexes, respectively. For each volume, we compute the non-redundant 3D spatial discrete wavelet transform, yielding the coefficients $w_{\mathbf{k}}(t)$, where the index $\mathbf{k}$ addresses all subbands and orientations. For the ease of notation, we write the spatial wavelet decomposition of our data set as

$$v_{\mathbf{n}}(t) = \sum_{\mathbf{k}} w_{\mathbf{k}}(t)\psi_{\mathbf{k}}(\mathbf{n}), \tag{1}$$

where the wavelet basis functions $\psi_{\mathbf{k}}$ are shifted and dilated versions of some prototype (tensor product of wavelets or scaling functions).

Now we introduce the time-series vector of length N_t for a wavelet coefficient $\mathbf{k}$; i.e., $\mathbf{w_k} = [w_{\mathbf{k}}(1) \ldots w_{\mathbf{k}}(N_t)]^{\mathrm{T}}$. This vector is modeled by the GLM as $\mathbf{w_k} = \mathbf{X}\mathbf{y_k} + \mathbf{e_k}$, where $\mathbf{X}$ is the $N_t \times L$ design matrix, and where $\mathbf{e_k}$ is the residual error. For a simple block-based experiment, $\mathbf{X}$ contains two columns: one for the on-off stimulus (eventually

0-7803-8388-5/04/$20.00 ©2004 IEEE

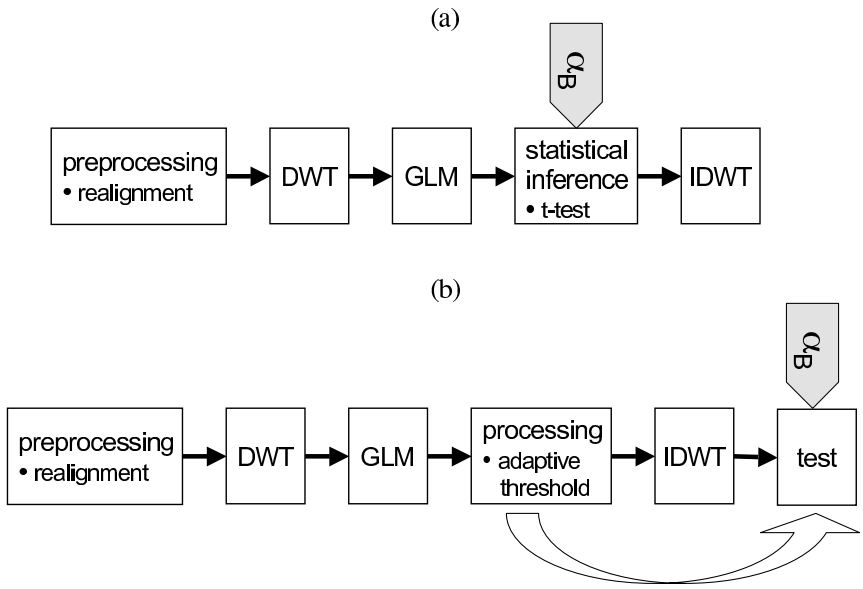

Fig. 1. (a) The classical wavelet-based approach. The desired significance level α_B is fed to the statistical inference stage in the wavelet domain. (b) The proposed approach treats the wavelet coefficients with an adaptive threshold. After reconstruction, a statistical test is performed in the spatial domain, taking into account the wavelet processing.

convolved with the HRF), and one for the background. We assume the elements of $\mathbf{e_k}$ to be independently and identically Gaussian distributed. The unbiased least-squares estimate of $\mathbf{y_k}$ is given by $\hat{\mathbf{y}}_\mathbf{k} = (\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{X}^\mathrm{T}\mathbf{w_k}$. Usually, one is only interested in the information of $\hat{\mathbf{y}}_\mathbf{k}$ that is extracted by a so-called contrast vector $\mathbf{c}$ (e.g., $\mathbf{c} = [1\ 0]^\mathrm{T}$ in our example). Therefore, for each wavelet coefficient $\mathbf{k}$, we obtain two scalar values:

$$g_\mathbf{k} = \mathbf{c}^\mathrm{T}\hat{\mathbf{y}}_\mathbf{k}, \quad (2)$$

$$s^2_\mathbf{k} = \hat{\mathbf{e}}^\mathrm{T}_\mathbf{k}\hat{\mathbf{e}}_\mathbf{k}\mathbf{c}^\mathrm{T}(\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{c}, \quad (3)$$

where $g_\mathbf{k}$ and $s^2_\mathbf{k}$ follow a Gaussian and a Chi-squared distribution, respectively. One then builds a t-value for each wavelet coefficient,

$$t_\mathbf{k} = \frac{g_\mathbf{k}}{\sqrt{s^2_\mathbf{k}/J}}, \quad \text{with } J = N_t - \mathrm{rank}(\mathbf{X}), \quad (4)$$

which can be tested against a threshold τ_w. The value τ_w is given by the desired significance level α (e.g., 5%) for a two-sided[1] t-test, corrected for multiple testing using Bonferroni-correction; in other words, the significance level for each test is $\alpha_B = \alpha/N_c$ where N_c is the number of intracranial wavelet coefficients.

After statistical inference, the detected coefficients are reconstructed as

$$r_\mathbf{n} = \sum_\mathbf{k} T_{\tau_w}(t_\mathbf{k}) g_\mathbf{k} \psi_\mathbf{k}(\mathbf{n}), \quad (5)$$

where $T_{\tau_w}(t_\mathbf{k})$ is an indicator function controlled by a two-sided threshold; i.e., $T_{\tau_w}(t_\mathbf{k}) = 1$ for $|t_\mathbf{k}| \geq \tau_w$, and 0 otherwise. As reported before, the wavelet coefficients can be used advantageously to detect activity. However, depending on the support of the wavelet, the volume $r_\mathbf{n}$ contains many non-zero voxels; often, an ad-hoc spatial threshold is applied to obtain "acceptable" detection maps. Additionally, the reconstructed value $r_\mathbf{n}$ does not have a direct statistical interpretation, in contrast with the t-value $t_\mathbf{k}$ in the wavelet domain. This constitutes the main disadvantage of the wavelet approach when compared to detection methods that operate in the spatial domain.

3. JOINT SPATIO-WAVELET STATISTICAL ANALYSIS

The major advantage of wavelet-based methods is their high sensitivity, even when very conservative (e.g., Bonferroni-corrected) testing procedures are used. Here, we propose to still rely on thresholding in the wavelet domain to improve the SNR. However, we now perform the statistical test afterwards in the spatial domain, taking into account the processing that has been done in the wavelet domain (see Fig. 1).

Our method does not require the wavelet coefficients to

[1] Both positive and negative wavelet coefficients can contribute to an increased signal in the spatial domain.

be decorrelated. Our only assumption is that N_t is sufficiently large (say $N_t > 50$) for $s_{\mathbf{k}}/\sqrt{J}$ to be a reliable estimate of the true standard deviation $\sigma_{\mathbf{k}}$ of $g_{\mathbf{k}}$. This hypothesis is fullfilled in most fMRI experiments.

3.1. Adaptive spatial threshold

The procedure here is the same as the classical one explained in Sect. 2 up to the calculation of the t-value in the wavelet domain. We also consider the thresholded values $T_{\tau_w}(t_{\mathbf{k}})g_{\mathbf{k}}$, except that there is no statistical test yet at this stage. τ_w is now considered to be a global threshold parameter to be determined later on. First, we look for a spatially-varying threshold function $q_{\mathbf{n}}$, such that, under the null hypothesis, the probability that the reconstruction of the processed wavelet coefficients contributing to the value of the voxel $\mathbf{n}$ does not exceed the required significance level α_B:

$$P\left[\left|\sum_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})g_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{n})\right| \geq q_{\mathbf{n}}\right] \leq \alpha_B. \qquad (6)$$

To this end, we rewrite the absolute value of the reconstruction term, introducing the true standard deviation $\sigma_{\mathbf{k}}$ of $g_{\mathbf{k}}$:

$$\begin{aligned}\left|\sum_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})g_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{n})\right| &\leq \sum_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})|g_{\mathbf{k}}|\,|\psi_{\mathbf{k}}(\mathbf{n})| \\ &= \sum_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})\frac{|g_{\mathbf{k}}|}{\sigma_{\mathbf{k}}}\sigma_{\mathbf{k}}\,|\psi_{\mathbf{k}}(\mathbf{n})| \\ &= \Lambda_{\mathbf{n}}\sum_{\mathbf{k}} \lambda_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})\frac{|g_{\mathbf{k}}|}{\sigma_{\mathbf{k}}}, \qquad (7)\end{aligned}$$

with $\Lambda_{\mathbf{n}} = \sum_{\mathbf{k}'} \sigma_{\mathbf{k}'}\,|\psi_{\mathbf{k}'}(\mathbf{n})|$ and $\lambda_{\mathbf{k}} = \sigma_{\mathbf{k}}\,|\psi_{\mathbf{k}}(\mathbf{n})|/\Lambda_{\mathbf{n}}$. Due to the property $\sum_{\mathbf{k}} \lambda_{\mathbf{k}} = 1$, we observe that the expected value of (7) is given by

$$\Lambda_{\mathbf{n}}\, E\left[T_{\tau_w}(t)\left|\frac{g}{\sigma}\right|\right], \qquad (8)$$

where t follows a t-distribution with J degrees of freedom, and g/σ is distributed according to a normalized Gaussian. We now have all the elements to apply to (6) the following Bienaymé-Tchebycheff inequality

$$P\left[|X| \geq Y\right] \leq \frac{E[|X|]}{|Y|}. \qquad (9)$$

In particular, if we select the threshold function to be $q_{\mathbf{n}} = \tau_s\Lambda_{\mathbf{n}}$, with τ_s a positive scalar, we obtain

$$P\left[\frac{\left|\sum_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})g_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{n})\right|}{\sum_{\mathbf{k}} \sigma_{\mathbf{k}}\,|\psi_{\mathbf{k}}(\mathbf{n})|} \geq \tau_s\right] \leq \frac{E\left[T_{\tau_w}(t)\cdot\left|\frac{g}{\sigma}\right|\right]}{\tau_s}. \qquad (10)$$

The right-hand side can be computed independently from the data and for any choice of the thresholds τ_w and τ_s. Thus, our task is now to determine some optimal threshold values that make the probability bound on the right-hand side of (10) equal to α_B.

3.2. Optimal threshold values

Eq. (10) only provides a partial answer to our problem because there is still an infinity of possible combinations of τ_w and τ_s that achieve the desired significance level. Specifically, by using the property that the t-distribution converges to a Gaussian for large N_t, we obtain the relationship between both thresholds for a fixed α_B:

$$\tau_s = \frac{1}{\alpha_B}\sqrt{\frac{1}{\pi}}\exp\left(-\frac{\tau_w^2}{2}\right). \qquad (11)$$

Next, we show how a simple argument from approximation theory helps us to find the optimal threshold values. Consider the following reconstructions:

$$\begin{aligned} u_{\mathbf{n}} &= \sum_{\mathbf{k}} g_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{n}), \\ r_{\mathbf{n}} &= \sum_{\mathbf{k}} T_{\tau_w}(t_{\mathbf{k}})g_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{n}), \\ r'_{\mathbf{n}} &= T_{\tau_s\Lambda_{\mathbf{n}}}(r_{\mathbf{n}})r_{\mathbf{n}}, \end{aligned}$$

corresponding respectively to the reconstruction of the GLM's fitted parameters[2], the reconstructed data after wavelet thresholding, and the final result including spatial thresholding. It is appropriate to minimize the "approximation error" between the non-processed data $u_{\mathbf{n}}$ and the final result $r'_{\mathbf{n}}$:

$$\begin{aligned}|u_{\mathbf{n}} - r'_{\mathbf{n}}| &\leq |u_{\mathbf{n}} - r_{\mathbf{n}} + r_{\mathbf{n}} - r'_{\mathbf{n}}| \\ &\leq |u_{\mathbf{n}} - r_{\mathbf{n}}| + |r_{\mathbf{n}} - r'_{\mathbf{n}}| \\ &\leq (\tau_w + \tau_s)\sum_{\mathbf{k}} \sigma_{\mathbf{k}}\,|\psi_{\mathbf{k}}(\mathbf{n})|\,.\end{aligned}$$

So, we need to minimize the sum of both thresholds. Using (11), we can find the optimal values:

$$\tau_s = 1/\tau_w, \quad \tau_w = \sqrt{-W_{-1}\left(-\frac{\alpha_B^2\pi}{2}\right)}, \qquad (12)$$

where $W_{-1}(\cdot)$ is the -1-branch of the Lambert W-function, which can be evaluated numerically [9].

4. RESULTS

We show an example of a block-based fMRI experiment with auditory stimulation [10], consisting of $N_t = 84$ isotropic volumes (3mm×3mm×3mm). The setup of the design matrix has been done as recommended in SPM and incorporates a model for the HRF. We used Bonferroni-correction for multiple testing that takes into account the number of intracranial voxels (i.e., $N_c = 2.6e6$). We

[2]Notice that the same $u_{\mathbf{n}}$ would be obtained when fitting the GLM directly in the spatial domain.

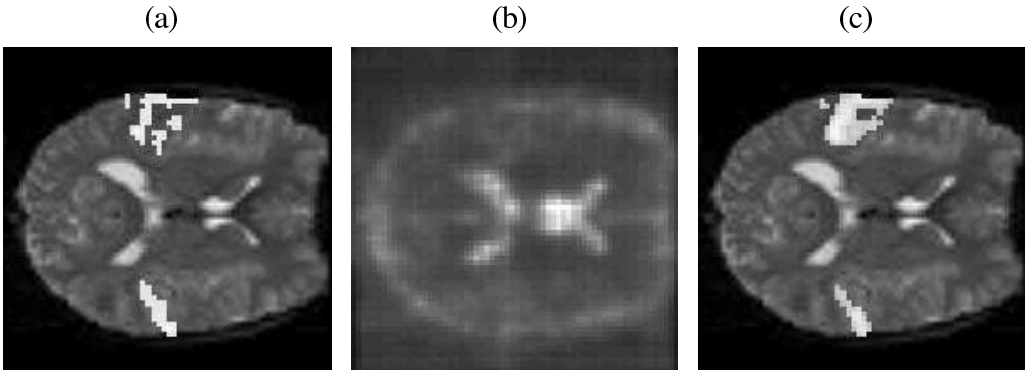

Fig. 2. (a) Map obtained using the proposed approach showing the activation in the auditory cortex. (b) Reconstruction of the "adaptive" part of the spatial threshold; i.e., $\Lambda_{\mathbf{n}} = \sum_{\mathbf{k}} \sigma_{\mathbf{k}} |\psi_{\mathbf{k}}(\mathbf{n})|$. (c) Map obtained using SPM (presmoothed data).

applied the 3D separable symmetric orthonormal B-spline wavelet transform of degree 0.7 and one iteration [7]. The optimal thresholds are $\tau_w = 5.83$ and $\tau_s = 0.17$. The spatial activation map contains those voxels where $r_{\mathbf{n}}/\Lambda_{\mathbf{n}}$ exceeds τ_s. Notice that the value $r_{\mathbf{n}}/\Lambda_{\mathbf{n}}$ can be used to compute a p-value using (11).

In Fig. 2 (a), the slice at the auditory cortex is showing activation and can be compared against the result of SPM in (c). The activated region in the upper part of the activation map appears disconnected, an observation that deserves further examination from the brain's structure in the T1 acquisition. In Fig. 2 (b), the reconstruction $\Lambda_{\mathbf{n}}$ of the $\sigma_{\mathbf{k}}$'s (or at least their estimates $s_{\mathbf{k}}/\sqrt{J}$) is shown, using the "absolute value reconstruction". The sensitivity of our approach is that it yields approximately the same number of activated voxels (550) as SPM.

5. DISCUSSION & CONCLUSION

We proposed a unifying approach that takes advantage of wavelet domain processing, but provides a statistical interpretation in the spatial domain. The main result is derived using Tchebycheff's inequality and the minimization of the approximation error between the unprocessed and final activation map. Although Tchebycheff's inequality is rarely used for computational purposes, the truncated nature of the probability density functions (i.e., after wavelet thresholding with τ_w) renders it an appropriate tool in our case, as confirmed by the reported sensitivity. The proposed method requires only slight modifications to a standard wavelet-based approach implementation; i.e., the computation of τ_w and τ_s, and the reconstruction of $\Lambda_{\mathbf{n}}$.

Currently, N_t is assumed to be large enough for the error of the estimate $\sigma_{\mathbf{k}}$ to be negligible. In future work, we intend to take the variability of this estimate into consideration as well.

6. REFERENCES

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

[2] 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*, vol. 17, no. 2, pp. 142–154, 1998.

[3] 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)*, Tampere, Finland, Sept. 2000, vol. IV, pp. 2013–2016.

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

[5] E. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statistical methods in medical research*, vol. 12, no. 5, pp. 375–399, 2003.

[6] 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.

[7] D. Van De Ville, T. Blu, and M. Unser, "Wavelets versus resels in the context of fMRI: establishing the link with SPM," in *SPIE's Symposium on Optical Science and Technology: Wavelets X*, San Diego CA, USA, Aug. 2003, SPIE, vol. 5207, pp. 417–425.

[8] K. Mueller, G. Lohmann, S. Zysset, and Y. von Carmon, "Wavelet statistics of functional MRI data and the general linear model," *Journal of Magnetic Resonance Imaging*, vol. 17, pp. 20–30, 2003.

[9] R. M. Corless, G. H. Gonnet, E. G. Hare, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathematics*, vol. 5, no. 4, pp. 329–359, 1996.

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