

# Classification of fMRI Time Series in a Low-Dimensional Subspace with a Spatial Prior

François G. Meyer, *Senior Member, IEEE*, and Xilin Shen

**Abstract**—We propose a new method for detecting activation in fMRI data. We project the fMRI time series on a low dimensional subspace spanned by wavelet packets in order to create projections that are as non Gaussian as possible. Our approach achieves two goals: it reduces the dimensionality of the problem by explicitly constructing a sparse approximation to the dataset and it also creates meaningful clusters allowing the separation of the activated regions from the clutter formed by the background time series. We use a mixture of Gaussian densities to model the distribution of the wavelet packet coefficients. We expect activated areas that are connected, and impose a spatial prior in the form of a Markov random field. Our approach was validated with *in vivo* data and realistic synthetic data, where it outperformed a linear model equipped with the knowledge of the true hemodynamic response.

**Index Terms**—fMRI, functional MRI, wavelet packets, mixture of Gaussian densities

## I. INTRODUCTION

**F**UNCTIONAL Magnetic Resonance Imaging (fMRI) can quantify hemodynamic changes induced by neuronal activity during a sensory or cognitive stimulations. The goal of the analysis is to detect the “activated” voxels where the dynamic changes in the fMRI signal can be explained by the stimulus time series. The fMRI data can be considered as a multivariate spatio-temporal process, where each scan is a four dimensional volume dataset  $s_v(t)$ , indexed by time  $t$ , and by the three-dimensional voxel position  $v$ . Most methods of analysis consider temporal and spatial dimensions separately. Univariate methods disregard the spatial correlation between adjacent time-series. The most frequently used technique is the General Linear Model (GLM). This model assumes that the fMRI signal can be represented as a linear combination of covariates that include the stimulus time-series, as well as physical or physiological processes (e.g. a drift [1], [2]). The statistical inference is complicated by the fact the noise  $\nu_v(t)$  in real fMRI data is correlated with a  $1/f$  spectral behavior associated with long memory processes [3], [4], [2]. In reality, the relationship between the fMRI signal and the stimulus time-series is known to be nonlinear [5], [6], [7].

Multivariate methods exploit the intrinsic spatio-temporal correlations that exist in the data. The large hyperoxygenation [8] which is at the origin of the BOLD contrast is spatially diffused, and therefore we expect neighboring time-series to be spatially correlated. Other sources of correlation in the

data include the spatial correlation introduced by the interpolation filters during the reconstruction algorithms [9]. Principal Components Analysis (PCA) has been used to find the set of orthogonal components that can best explain the variance in the observations [10]. Because multivariate methods analyze all voxels in the brain at once, the large number of voxels (in comparison to the number of scans) requires that the dimensionality of the problem be reduced. In contrast to PCA, Independent Component Analysis (ICA) can decompose the observations into components that are statistically independent [11], [12]. Unfortunately, if the observations are noisy, then it becomes impossible to separate the components from the noise [13]; in practice, if the analysis is performed on real data, then the components are not even approximatively independent [13]. A common problem associated with the usage of PCA or ICA is the interpretation of the components. The interpretation usually relies on post hoc heuristics such as visual inspection of the similarity between the component and the stimulus time-series [12]. Global clustering methods [14] partition the time-series according to their similarity. Again, the dimensionality of the problem is overwhelming. ICA and clustering do not directly lend themselves to statistical inference, and are mostly used for exploratory analysis.

We propose in this paper a new method for detecting activation in fMRI data. We show that we can construct projections of the fMRI signal on low dimensional subspaces that reveal the presence of activated time series. These projections are selected from large libraries of wavelet packets by studying the probability distribution of the wavelet packet coefficients. We model this distribution with a finite mixture of multivariate Gaussian densities. We develop methods for estimating the number of components in the mixture, their associated parameters, and interpret their physiological roles. This paper is organized as follows. In the next section we provide a general overview of the approach. In section III-A we review some basic concepts about wavelet packets. In section IV-A we provide some experimental evidence for the statistical model that we propose in section IV-B to describe the probability densities of wavelet packet coefficients of fMRI time-series. Section V describes the estimation of the activation maps. Results of experiments conducted on synthetic and real data are presented in Section VI.

## II. GENERAL OVERVIEW OF THE APPROACH

The goal of this work is to detect the regions of the brain that are involved in the neuronal activity triggered by the stimulus. In the absence of any specific model of the fMRI signal, we

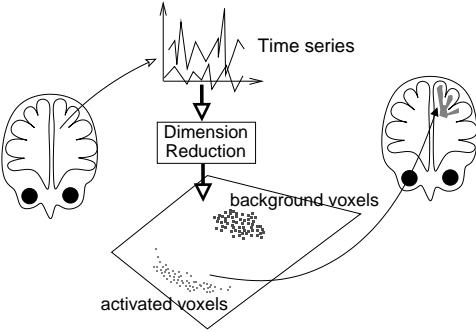


Fig. 1. We project fMRI time series on a low dimensional subspace that is carefully chosen for its ability to split the time series into two distinct clusters. Activated time series are identified and labeled as activated in activation maps.

favor an exploratory approach. The problem is formulated as a segmentation problem : the brain is divided into homogeneous functional areas wherein the fMRI signal is approximatively uniform. We make the following two assumptions : (1) the fMRI time-series will be similar for all activated voxels in the same functional areas; (2) the fMRI time-series from activated areas will be different from the background (non activated) time-series. Formally, we seek to cluster the time-series into two classes: activated and non activated. We consider each time series to be a vector of  $\mathbb{R}^T$ , where  $T$  is the number of time sample (typically of the order of 100-200). The clustering of vectors in high dimension is hard, and we propose therefore to reduce the dimensionality of the time-series by projecting the vectors  $s_v$  on a low dimensional subspace (see Fig 1). The choice of the subspace is critical. As explained in section IV-A, the projections on almost any wavelet packets of the activated and background time-series are lumped together into a Gaussian distribution, making the subsequent clustering impossible. We need therefore to find the few wavelet packets that create two well separated clusters of projections (see Fig. 1).

### III. REDUCTION OF DIMENSIONALITY

In our search for interesting projections we can be guided by some of the experimental properties of the data. It has been observed [3], [4] that data collected under the null-hypothesis condition exhibit the  $1/\omega$  spectrum associated with long memory processes. This  $1/\omega$  spectral behavior can also be created by many physiological signals that have a self-similar structure with a power spectrum of the form  $|\omega|^{-2H+1}$ . Fadili and Bullmore [4], [15] have proposed to model the fMRI noise with fractional Brownian motion and have estimated Hurst exponents  $H$  greater than  $1/2$  (see also [16] for a recent and detailed analysis of the values of the Hurst exponent for different populations, and Ciuciuc et al. [17] for a multifractal analysis). The long memory structure of the fMRI signal constitutes a basic hurdle to any statistical analysis of the data: any analysis will be tracking the slowly varying changes in the signal instead of the effects of the stimulus. Fortunately, the wavelet transform provides an approximation to the Karhunen-Loeve transform for the long memory processes [18]. As shown in [18], the correlation between two wavelet coefficients

$\alpha(j, k)$  and  $\alpha(j', k')$  decreases asymptotically as

$$\text{corr}(\alpha(j, k), \alpha(j', k')) \sim O(|2^{-j}k - 2^{-j'}k'|^{-2(p-H+1)}),$$

when the distance between the coefficients is large ( $|k - k'| \rightarrow \infty$ ). The parameters  $p$  is the number of vanishing moments of the wavelets, and it should be larger than  $H$ . In this case, we can assume that the wavelet coefficients  $\alpha(j, k)$  and  $\alpha(j', k')$  are approximatively uncorrelated.

Under these circumstances it is therefore natural to focus our analysis on projections constructed from wavelets, which will perform a multiscale analysis of the time dynamics of the signals [15], [4]. Several studies indicate that one finds dynamic changes of the fMRI signal in time, frequency, and in space [19], [15]. A more precise analysis, that can optimize the time-scale representation is given by wavelet packets that provide a rich library of waveforms both localized in time and in frequency.

#### A. Multiscale analysis and wavelet packets

We introduce in this section the notations associated with wavelet packets. These notations will be used in the sequel of the paper. Let  $\psi^0(t)$  be the scaling function and let  $\psi^1(t)$  be the wavelet associated with a multiresolution analysis [20]. Let  $\{h_n\}$  and  $\{g_n\}$  be the lowpass and highpass filter associated with this wavelet. We define the basic wavelet packets for  $t = 0, \dots, T$  by

$$\begin{aligned}\psi^{2n}(t) &= \sum_k h_k \psi^n(2t - k) \\ \psi^{2n+1}(t) &= \sum_k g_k \psi^n(2t - k).\end{aligned}\quad (1)$$

For the sake of simplicity we assume that  $T = 2^{J_0}$  (the situation where  $T$  is not a power of two can be easily handled as explained in [21]). A multiscale wavelet packet is given by

$$\psi_{j,k,l}(t) = 2^{-j/2} \psi^k(2^{-j}t - l). \quad (2)$$

A wavelet packet is identified by an index  $\gamma = (j, k, l)$ , where

- $j = 0, \dots, J$  represents the scale :  $\psi_{j,k,l}$  has a support of size  $2^j$ .  $J$  is the maximum scale, and  $J \leq J_0$ .
- $k = 0, \dots, 2^j - 1$  represent the frequency index at a given scale  $j$  :  $\psi_{j,k,l}$  has roughly  $k$  oscillations.
- $l = 0, \dots, 2^{J_0-j} - 1$  represents the translation index within a node  $(j, k)$  :  $\psi_{j,k,l}$  is located at  $l2^{-j}$ .

The library of wavelet packets is constructed iteratively starting from the scaling function  $\psi^0$ , and the wavelet  $\psi^1$ . Figure 2-A shows a sketch of a wavelet packet tree for 7 scales. Each node  $(j, k[])$  of the tree contains  $\psi_{j,k,l}$ ,  $l = 0, \dots, 2^{6-j}$  wavelet packets  $\psi_{j,k,l}$ . These functions are shown for two particular subtrees in Fig. 2-B and 2-C. Each row in the box B or C corresponds to a set of nodes of the wavelet packet tree A. For instance, in B we have: row 1 =  $(3, 0, []), (4, 0, []), (4, 1, []), (5, 0, []), (5, 1, []), (5, 2, []), (5, 3, [])$ , and row 4 =  $(6, 0, []), \dots, (6, 7, [])$ . The effect of the scale  $j$  can be understood by looking at the support of the wavelet packets in each subtree. The functions on the top rows are well localized in time (small scale), but poorly localized in frequencies. The functions on the last rows have lost their

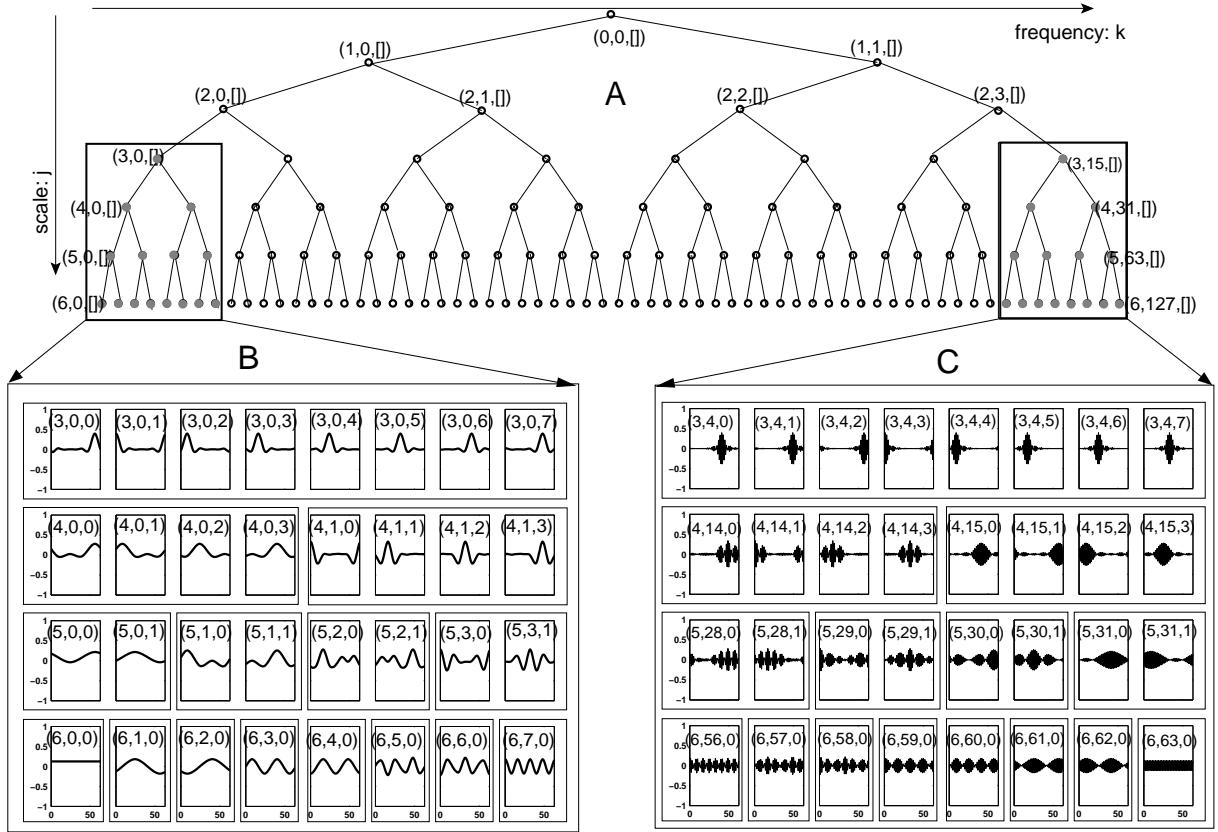


Fig. 2. Two subtrees (B,C) within the wavelet packet tree (A). Each node  $(j, k, [])$  in the wavelet packet tree (A) corresponds to a subspace of wavelet packets  $\psi_{j,k,l}$ ,  $l = 0, \dots, 2^{6-j}$ . As we move down the tree (larger scale  $j$ ), each node contains fewer and fewer functions, so that at each level (scale  $j$ ), the number of wavelet packets is constant. The subtrees B and C are shown in detail with the corresponding wavelet packet functions. Each row in the box B or C corresponds to a set of nodes of the wavelet packet tree A. For instance, in B we have: first row =  $\{(3, 0, [])\}$ , second row =  $\{(4, 0, []), (4, 1, [])\}$ , and last row =  $\{(6, 0, []), \dots, (6, 7, [])\}$ .

temporal localization (very large scale), but are precisely located in the frequency domain : they roughly behave as sinusoidal functions oscillating at low (Fig. 2-B) and high (Fig. 2-C) frequencies. In the remaining of the paper, we consider  $\psi_\gamma = \psi_{j,k,l}$  to be a  $T \times 1$  vector. The projection of the signal  $s_v$  on  $\psi_\gamma$  is called a wavelet packet coefficient, and is given by  $\alpha_v(\gamma) = \psi_\gamma^T s_v$ . We also require that the wavelet  $\psi$  have  $p$  vanishing moments. As a consequence, polynomials of degree  $p - 1$  will have a very sparse representation : all the wavelet packets coefficients  $\alpha(j, k, l)$  with  $k \geq 1$  are equal to zero, except for the coefficients located at the border of the dyadic subdivision.

### B. Searching for good projections

We reduce the dimensionality of our analysis by projecting the fMRI time series on a low dimensional subspace spanned by some carefully selected wavelet packets. We select these wavelet packets according to their ability to reveal the organization of the data in terms of two clusters: activated and non activated time series. Unfortunately, for (almost) any given wavelet packet, the projections of activated and non activated time-series on that wavelet packet are lumped together into a Gaussian distribution (see Fig. 3-A for a pictorial interpretation, and section IV-A for results of experiments). Of course, we shouldn't be surprised since we know from a general

result of statistics [22] that the distribution of nearly all one-dimensional projections of high-dimensional datasets appear Gaussian. In order to disentangle the activated time series from the background time series, we need to find the projections that yield the least Gaussian distribution of the coefficients. Indeed, as shown in section IV-A, a non Gaussian distribution of the coefficients can be interpreted as follows: small coefficients correspond to background time-series, and large coefficients correspond to activated response highly correlated to the wavelet packet. This criterion can be implemented in practice by maximizing the skewness of the distribution of the projections (other measures of non-Gaussianity include negentropy and kurtosis [23]). We consider the empirical distribution of the wavelet packet  $\alpha_v(\gamma)$  coefficients computed over all voxels  $v$ , and estimate the skewness.

## IV. PROBABILITY DISTRIBUTIONS OF THE WAVELET PACKET COEFFICIENTS

As explained in Section IV-B, we use a mixture of Gaussian densities to model the distribution of the wavelet packet coefficients. Rather than arbitrarily imposing this model, we provide in section IV-A a careful justification for the mixture model. Our justification is experimental and rely on statistical tests of Gaussianity from *in vivo* fMRI data. Section IV-A is not required for understanding the remaining of the paper: we

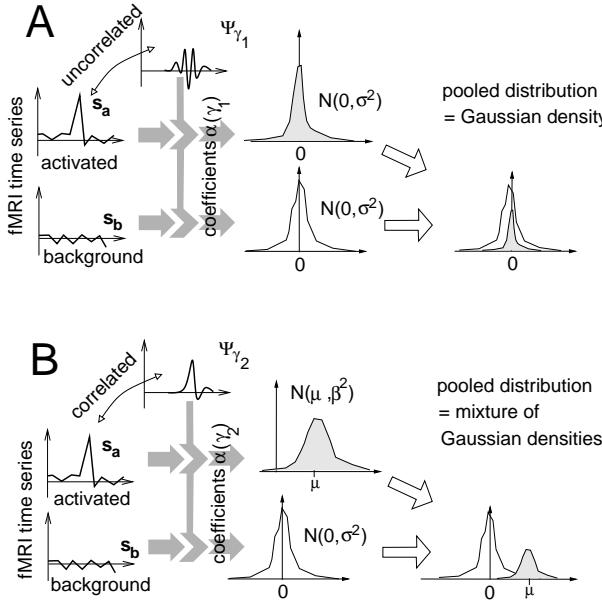


Fig. 3. Distribution of the wavelet packet coefficients. **A:** the wavelet packet has no correlation with the fMRI response then the coefficients are very small. The variance of the distribution of the coefficients is the same for background or activated voxels. **B:** the wavelet packet is correlated to the response at activated voxels. The variances of the two distributions (background and activated voxels) are now different.

have summarized below the main experimental findings that provide support for the mixture model.

**Our findings.** The experiments described in section IV-A demonstrate that:

- If  $\mathbf{v}$  is a background voxel, then the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is Gaussian distributed with a zero-mean. This result is independent of the wavelet packets  $\psi_\gamma$ .
- If  $\mathbf{v}$  is an activated voxel, then the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is Gaussian distributed. The mean of the distribution depends on the wavelet packets  $\psi_\gamma$ . There exists a small number of wavelet packets for which the mean of the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is non zero (these wavelet packets have a strong correlation with the activated signal). For all other  $\gamma$  the distribution has a zero-mean (these wavelet packets have no correlation with the fMRI signal).
- In addition, if the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  has a zero-mean, then its variance is very similar to the variance of the wavelet packet coefficients of the background time series.

A direct consequence of these findings, is that the pooled distribution of activated and non activated voxels can be modeled by a mixture of Gaussian density (Fig. 3-A). If  $\gamma$  is chosen such that the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  has a zero-mean for all activated voxels, then the pooled distribution is a mixture of two Gaussian distributions with zero mean and similar variance. In other words, the pooled distributions is Gaussian. This will happen for most wavelet packets  $\psi_\gamma$  (Fig. 3-A). On the other hand, if we choose  $\gamma$  such that  $\psi_\gamma$  has a large correlation with activated time series, then the means of the Gaussian distributions are different, and the mixture is a true mixture (Fig. 3-B).

### A. What can we learn from the projections of fMRI data?

If we consider the fMRI signal  $s_{\mathbf{v}}(t)$  at a given voxel  $\mathbf{v}$  as a stochastic process that varies as a function of time  $t$ , we can estimate the probability distribution of the process from the realizations  $s_{\mathbf{v}}(0), \dots, s_{\mathbf{v}}(T-1)$ . For the rest of the paper, we assume that we have removed any possible systematic trend in the data (as explained in [2], this can be accomplished simply by setting the coarse scale wavelet coefficients to zero). We expect  $s_{\mathbf{v}}(t)$  at a non activated voxel  $\mathbf{v}$  to be approximatively constant (we have removed the drift), and since our wavelet have sufficiently many vanishing moments we should get  $\langle \psi, s_{\mathbf{v}} \rangle \approx 0$ . Most  $\psi_\gamma$  should also be uncorrelated to  $s_{\mathbf{v}}(t)$  at an activated voxel  $\mathbf{v}$ . Indeed,  $\psi_\gamma$  oscillates too fast in comparison to the slow variation of the bold response (see Fig. 3-A), and thus  $\langle \psi, s_{\mathbf{v}} \rangle \approx 0$ . However, there will be a small number of indices  $\gamma$  for which  $\psi_\gamma$  oscillates slowly and is highly correlated to  $s_{\mathbf{v}}(t)$  (see Fig. 3-B). For these  $\gamma$ ,  $\alpha_{\mathbf{v}}(\gamma)$  will be large.

**The experiments.** In [24] we conducted several experiments with real fMRI datasets in order to test the following hypotheses :

- 1) If  $\mathbf{v}$  is a background voxels, then the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is Gaussian distributed with a zero-mean for all wavelet packets  $\psi_\gamma$ .
- 2) If  $\mathbf{v}$  is an activated voxel, then the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is Gaussian distributed.

Furthermore, there exists a small number of wavelet packets  $\psi_\gamma$  for which the mean of the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is non zero (these wavelet packets have a strong correlation with the activated signal). For all other  $\gamma$  the distribution has a zero-mean (these wavelet packets have no correlation with the fMRI signal).

Given a wavelet packet  $\psi_\gamma$ , we test these two hypotheses by testing the empirical distribution of  $\alpha_{\mathbf{v}}(\gamma)$  for normality. We first partition the time series into two classes, (1) activated and (2) background (non activated). The distribution of  $\alpha_{\mathbf{v}}(\gamma)$  is estimated from the values of  $\alpha_{\mathbf{v}}(\gamma)$  for all the voxels  $\mathbf{v}$  inside each class (ergodicity argument). Because we expect variations in the activation strength throughout the activated areas, we cannot assume that the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  will be the same for all activated voxels  $\mathbf{v}$ . Instead of using all the activated voxels to estimate the distribution of  $\alpha_{\mathbf{v}_0}$ , we perform a local analysis inside a small neighborhood  $\mathcal{U}(\mathbf{v}_0)$  centered around each voxel  $\mathbf{v}_0$  (see Fig. 4). We test the hypothesis that the empirical distribution of the coefficients  $\{\alpha_{\mathbf{v}}(\gamma), \mathbf{v} \in \mathcal{U}(\mathbf{v}_0)\}$  is Gaussian. Several test for normality exist, and we use the W Shapiro-Wilk test [25] because of its greater statistical power to detect non Gaussian distributions. As  $\mathbf{v}_0$  is moved throughout the brain we collect many samples of the  $W$  statistic. Of course, we expect the sample mean and sample variance of the distribution of  $\alpha_{\mathbf{v}}(\gamma)$  to vary as a function of the position of the neighborhood  $\mathcal{U}(\mathbf{v}_0)$ . Fortunately, the  $W$  statistic is independent of the mean and variance of the distribution being tested. This allows us to combine  $W$  statistics from different neighborhood  $\mathcal{U}(\mathbf{v}_0)$  for a given wavelet packet  $\psi_\gamma$ . We compute the empirical distribution of  $W$  (over all position of  $\mathcal{U}(\mathbf{v}_0)$ ). In order to

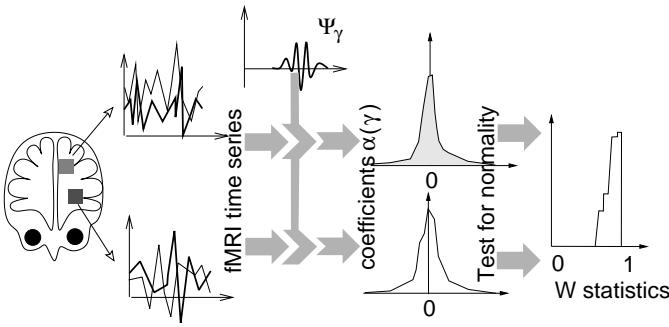


Fig. 4. For each voxels  $v_0$  in the brain, we consider a small neighborhood  $\mathcal{U}(v_0)$  centered around  $v_0$ , and we test the hypothesis that the empirical distribution of the coefficients  $\{\alpha_v(\gamma), v \in \mathcal{U}(v_0)\}$  is Gaussian.

quantify non-normality, we compare this empirical distribution with the distribution of the  $W$  statistic obtained under the normality assumption. The comparison is performed using the  $\chi^2$  distance. The null hypothesis is that the distribution of the  $W$  statistics is similar to a distribution of the  $W$  statistic obtained under the normality assumption. We report here the results of one experiment; more results are available in [24]. We consider an fMRI dataset that demonstrates activation of the visual cortex with a periodic stimulus (see Section VI for details about the data). We used the general linear model [26] to determine the status : activated/non activated of each time-series  $s_v$ . The background time-series had a  $p$ -value greater than 0.1, the activated time-series had a  $p$ -value smaller than  $10^{-3}$ . For every index  $\gamma$ , and for every voxel  $v$ , we computed the wavelet packet coefficient  $\alpha_v(\gamma)$  of the time-series  $s_v$ . Fig. 5-top locally from background time-series only (we use  $\alpha_v$  in the computation of  $W$  only shows the empirical distribution of the  $W$  statistic computed if  $v$  is non activated). Each row corresponds to the corresponding row in the subtree **B** in Fig. 2. To reject the null hypothesis of normality at a significance level of 0.1, the  $\chi^2$  values must be greater than 9.23. Because all the  $\chi^2$  distances are less than 9.23, the hypothesis that  $\alpha_v$  is Gaussian can be accepted. Fig. 5-bottom shows the empirical distribution of the  $W$  statistic computed locally from activated time-series only. Again, each box corresponds to a wavelet packet from the subtree **B** in Fig. 2. Because we have much fewer samples, the shapes of the empirical distributions are different from the background distributions. To reject the null hypothesis of normality at a significance level of 0.001, the  $\chi^2$  values must be greater than 20.51. Because all values of the  $\chi^2$  are smaller than 20.51, the hypothesis that  $\alpha_v$  is Gaussian can be accepted. In summary, we can accept both hypotheses. Other experiments, which are described in [24], resulted in the same conclusion.

**Related results.** It has been shown [27] that the probability distribution of the BOLD signal is approximatively Gaussian if  $v$  is a non activated voxel. In a very general sense, our experiments are also consistent with the result of Diaconis and Freedman [22] about one dimensional projections of high dimensional data being approximatively Gaussian.

#### B. A statistical model for the coefficients

In the rest of the paper we assume that we have found a small number  $K$  of wavelet packets  $\psi_{\gamma_k}$  for which the em-

pirical joint distribution of  $\alpha_v = [\alpha_v(\gamma_0), \dots, \alpha_v(\gamma_{K-1})]^T$  (computed over all voxels in the brain) is non Gaussian. The

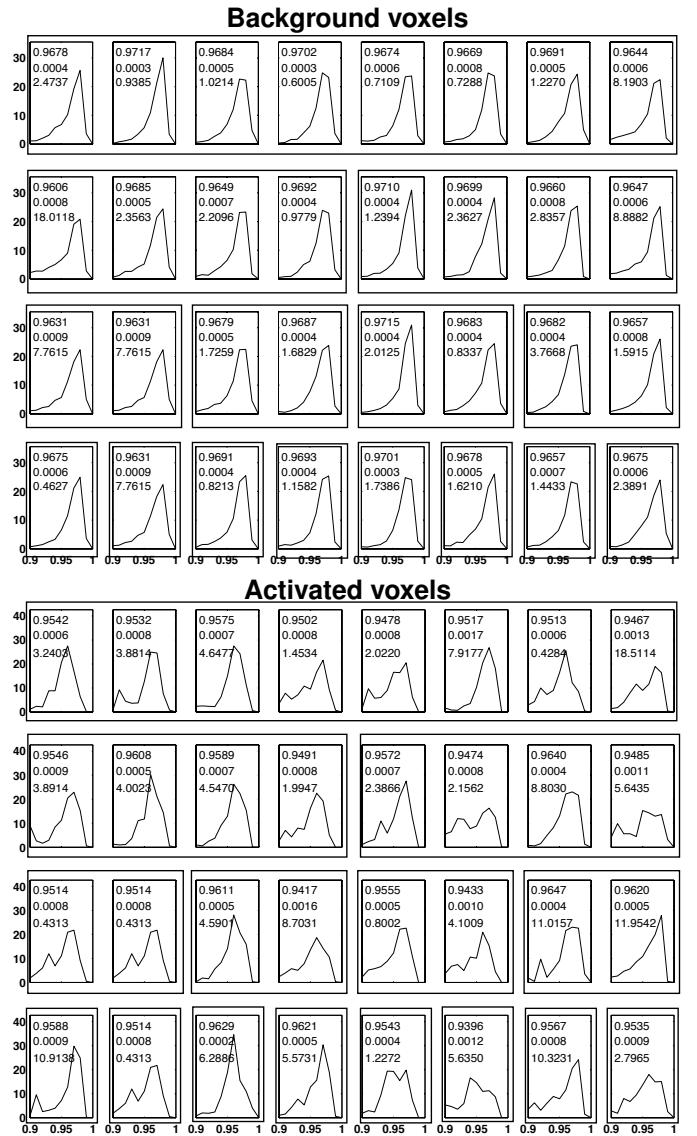


Fig. 5. Histogram of the  $W$  statistic computed from the background (top) and activated (bottom) time series. For each histogram we show the mean, the variance, and the  $\chi^2$  score. Each box corresponds to a wavelet packet node from the subtree **B** in Fig. 2.

wavelet packets  $\psi_{\gamma_k}, k = 0, \dots, K - 1$  are selected using a greedy approach, as explained in section III-B. We note that a best basis search, as described in [28], [29], [30], could also be performed. From here on we will only study the probability distribution of the  $\alpha_v(\gamma_k), k = 0, \dots, K - 1$ . An appropriate model for the joint distribution of the wavelet packet coefficients  $\alpha_v$  is a finite mixture of multivariate Gaussian densities,

$$\Pr(\alpha) = \sum_{m=1}^M \pi_m \phi(\alpha, \mu_m, \Sigma_m). \quad (3)$$

The mixing parameters  $\pi_m$  are positive weights that add up to 1. The density  $\phi$  is the  $K$ -multivariate normal density function. One of the components (chosen arbitrarily to be

$m = 1$ ) models the background voxels. After removing a slowly varying trend, the background time-series  $s_v(t)$  should be approximatively constant, and the wavelet packet coefficients of the background should be zero (or very small). We expect therefore to have  $\mu_1 \simeq 0$ . The other components correspond to activated voxels. We note that the model (3) is global (the dependence on  $v$  is removed), and we need more than one component to describe the different levels of activations. Voxels that are strongly activated will contribute to a component  $m$  with a large  $\mu_m$ , and weakly activated voxels will contribute to a component with a small  $\mu_m$ . Voxels with a similar activation strength are grouped together in the same component (irrespective of their relative spatial proximity). Mixture models have been proposed [31], [32] to describe the distribution of a statistical test performed on individual voxels. Woolrich and colleagues have recently [33] proposed a spatial mixture model (similar to [32]) combined with a spatial regularization based on Markov random fields. Penny and Friston [34] proposed to model the spatial distribution of the signals  $s_v(t)$  as a mixture of several activated areas and one background component.

### C. Maximum likelihood estimate

The maximum likelihood estimates  $\hat{\mu}_m$ ,  $\hat{\Sigma}_m$  and  $\hat{\pi}_m$  of the mixture parameters can be computed from the measurements using the Expectation Minimization (EM) algorithm [35]. The estimation of the number of components  $M$  is often difficult [35] but can be addressed using model selection criteria [36]. The real difficulty here is that we expect only a few percent of the total number of voxels to be activated. If we use all the voxels in the brain to estimate the density of the wavelet packet coefficients, then the small bumps in the density (3) generated by the activated components will always be hidden. Indeed, the mixing proportion  $\pi_1$  for the background component, which is proportional to the relative number of non activated voxels, is close to 1. Whereas the mixing proportions  $\pi_m$  for the activated regions is close to zero. This issue can be addressed by restricting the number of activated components to a single component. We have noticed that the variance of the activated component is usually much larger than the variance of the background. We believe that this is due to the differences in the activation strength. We can use the estimates  $\hat{\mu}_m$ ,  $\hat{\Sigma}_m$  and  $\hat{\pi}_m$  to compute the posterior probability given by

$$\Pr(m|\boldsymbol{\alpha} = \boldsymbol{\alpha}_{v_i}) = \frac{\hat{\pi}_m \phi(\boldsymbol{\alpha}_v, \hat{\mu}_m, \hat{\Sigma}_m)}{\sum_{l=1}^M \hat{\pi}_l \phi(\boldsymbol{\alpha}_v, \hat{\mu}_l, \hat{\Sigma}_l)}, \quad (4)$$

which provides an estimate of the probability that measurement  $\boldsymbol{\alpha}_{v_i}$  at voxel  $v_i$  be generated by component  $m$ .

## V. ACTIVATION MAPS

### A. A spatial prior

The detection of activation can be expressed as the problem of assigning a label  $a_i$  to each voxel  $v_i$ , where

$$a_i = \begin{cases} 1 & \text{if } v_i \text{ is in an activated area,} \\ 0 & \text{if } v_i \text{ is a background voxel.} \end{cases} \quad (5)$$

The label  $a_i$  will be a function of the wavelet packet coefficients  $\boldsymbol{\alpha}_{v_i}$ . The label will also depend on a spatial prior model described below, which can be used to eliminate the large number of unlikely activated patterns. Truly activated voxels tend to be spatially clustered, while falsely activated voxels tend to be scattered. The mixture model (3) does not account for the spatial organization of the pattern of activation. We believe that activated regions should be composed of a small number of connected patterns of activation.

Let  $\mathbf{a} = (a_0, \dots, a_{V-1})^T$  be the vector of labels indicating the status (activated vs non activated) of the  $V$  voxels inside the brain. A convenient model for the activation map  $\mathbf{a}$  is provided by a Markov random field (MRF) [37]. Formally, we define a neighborhood system  $\mathcal{N}$  to be a collection of subsets  $\mathcal{N}_v$  of the original lattice grid that satisfies the following properties

- 1)  $v \notin \mathcal{N}_v$ ,
- 2)  $v \in \mathcal{N}_u \Leftrightarrow u \in \mathcal{N}_v$ .

A subset  $c$  is a clique if every pair of distinct voxels in  $c$  are neighbors. In our application, we are mainly interested in first order (nearest) neighbors system (see Fig. 6). The joint probability distribution  $\Pr(\mathbf{a})$  of an activation pattern can be expressed using a Gibbs distribution, and is given by  $\Pr(\mathbf{a}) = \frac{1}{Z} e^{-\lambda U(\mathbf{a})}$ , where  $Z$  is a normalizing factor,  $\lambda$  is a fixed nonnegative parameter, and  $U(\mathbf{a})$  is an energy function which is defined on the set of all cliques  $\mathcal{C}$  by  $U(\mathbf{a}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{a})$ . Each potential function  $V_c(\mathbf{a})$  is a function of the values of the label field  $\mathbf{a}$  on the particular clique  $c$ . In the case of first-order neighbors, we only model the local interactions between the central voxel  $v_0$  and its four neighbors (see Fig. 6).

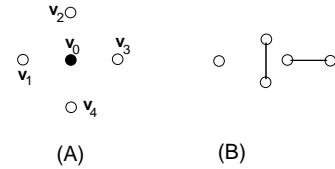


Fig. 6. Neighborhood  $\mathcal{N}_{v_0}$  formed by the four nearest neighbors of  $v_0$  (A), and the corresponding cliques configurations (B).

Let  $c$  be the clique formed by  $v_i$  and  $v_0$  (see Fig. 6), we define the potential function  $V_c(\mathbf{a}) = |a_{v_0} - a_{v_i}|$ . This results in a Hamiltonian that is identical to the standard Potts model for  $q = 2$  [38], which has been used for image segmentation and denoising. Most likely configurations of  $\mathbf{a}$  will be composed of large connected areas where  $\mathbf{a}$  is constant (activated or non activated), and scattered activation will be penalized. The Potts model pulls strength from neighboring voxels without having to smooth the data across the boundaries of the activated regions.

### B. Parameter Estimation

The estimation problem can be formalized using a Bayesian approach. Our goal is to estimate  $\mathbf{a}$  given the measurements  $\boldsymbol{\alpha}_{v_i}, i = 0, \dots, V-1$ . Let  $\Pr(\mathbf{a}|\boldsymbol{\alpha}_{v_0}, \dots, \boldsymbol{\alpha}_{v_{V-1}})$  be the posterior probability distribution of the hidden labels  $\mathbf{a}$ , given the

measurements  $\alpha_{\mathbf{v}_i}, \mathbf{v}_i \in \mathcal{N}(x_0)$ . The most likely activation pattern  $\hat{\mathbf{a}}$  is given by the mode of the posterior probability

$$\hat{\mathbf{a}} = \underset{(a_0, \dots, a_{V-1})}{\operatorname{argmax}} \Pr(\mathbf{a} | \alpha_{\mathbf{v}_0}, \dots, \alpha_{\mathbf{v}_{V-1}}).$$

By Bayes theorem,

$$\Pr(\mathbf{a} | \alpha_{\mathbf{v}_0}, \dots, \alpha_{\mathbf{v}_{V-1}}) \propto \Pr(\alpha_{\mathbf{v}_0}, \dots, \alpha_{\mathbf{v}_{V-1}} | \mathbf{a}) \Pr(\mathbf{a}).$$

The prior probability  $\Pr(\mathbf{a})$  is given by the Gibbs distribution. We assume that once the status (activated or not) of all voxels is given, then the wavelet packet coefficients  $\alpha_{\mathbf{v}_i}$  are independent (see also[32] for a similar hypothesis), and thus

$$\Pr(\alpha_{\mathbf{v}_0}, \dots, \alpha_{\mathbf{v}_{V-1}} | \mathbf{a}) = \prod_{j=0}^{V-1} \Pr(\alpha_{\mathbf{v}_j} | \mathbf{a}). \quad (6)$$

We also assume that the value of the signal  $s_j$  (or equivalently the vector of  $K$  wavelet packet coefficients  $\alpha_{\mathbf{v}_j}$ ) at voxel  $\mathbf{v}_j$ , only depends on the status  $a_j$  of that very voxel. Therefore,

$$\Pr(\alpha_{\mathbf{v}_j} | \mathbf{a}) = \Pr(\alpha_{\mathbf{v}_j} | a_j). \quad (7)$$

The probability  $\Pr(\alpha_{\mathbf{v}_j} | a_j)$  is the probability distribution of the  $K$  wavelet packet coefficients, given the status  $a_j$  of the voxel  $\mathbf{v}_j$ . If  $\mathbf{v}_j$  is non activated ( $a_j = 0$ ), then it belongs to the background component ( $m = 1$  in (3)), and we have from (3)

$$p(\alpha_{\mathbf{v}_j} | a_j = 0) = \phi_1(\alpha_{\mathbf{v}_j}, \hat{\mu}_1, \hat{\sigma}_1). \quad (8)$$

If  $\mathbf{v}_j$  is activated ( $a_j = 1$ ), then it belongs to one of the activated components ( $m = 2, \dots, M$  in (3)), and

$$p(\alpha_{\mathbf{v}_j} | a_j = 1) = \frac{1}{\sum_{m=2}^M \hat{\pi}_m} \sum_{m=2}^M \hat{\pi}_m \phi_m(\alpha_{\mathbf{v}_j}, \hat{\mu}_m, \hat{\sigma}_m). \quad (9)$$

To summarize, the most likely activation pattern  $\hat{\mathbf{a}}$  given the measurements  $\alpha_{\mathbf{v}_i}$  is the solution of

$$\hat{\mathbf{a}} = \underset{(a_0, \dots, a_{V-1})}{\operatorname{argmax}} \prod_{j=0}^{V-1} \Pr(\alpha_{\mathbf{v}_j} | a_j) \exp \left\{ -\lambda \sum_{c \in \mathcal{C}} V_c(\mathbf{a}) \right\}. \quad (10)$$

We use the simulated annealing algorithm [39] to find the global minimum of (10). The unknown parameter  $\lambda$  provides a trade off between the measurement and the spatial prior. Our experiments with synthetic data show that the best performance (both in terms of the false negatives and false positives) is obtained when  $\lambda$  is between 1 and 1.5. We choose  $\lambda$  to be 1 in all experiments. We start the cooling process of the simulated annealing with an initial map provided by (4).

## VI. EXPERIMENTS

In the absence of any ground truth about the strength and the extent of the activation in the human brain, the validation of a new method for the analysis of fMRI data is always problematic. We propose therefore to first validate our approach with realistic simulated data. We then proceed to the evaluation of our method using *in-vivo* data.

**Input:**  $s_{\mathbf{v}_i}(t), t = 0, \dots, T-1; i = 0, \dots, V-1$

**Algorithm:**

- 1) **For all** voxel  $\mathbf{v}_i, i = 0, \dots, V-1$ :
  - compute the undecimated wavelet packet coefficients  $\alpha_{\mathbf{v}_i}(\gamma)$  for all nodes  $\gamma$  of the wavelet packet tree (for most experiment, the knowledge of the stimulus waveform allows us to compute only part of the full tree).
- 2) **For all** nodes  $\gamma$  of the wavelet packet tree:
  - Compute the skewness of the distribution of  $\alpha_{\mathbf{v}}(\gamma)$ .
  - Select the first  $K$  wavelet packets with the largest skewness.
- 3) Estimate the parameters of the mixture of Gaussian model (3) with the EM algorithm from the data  $\alpha_{\mathbf{v}_i}(\gamma_k), k = 0, \dots, K-1; i = 0, \dots, V-1$ .
- 4) Compute the most likely activation pattern  $\hat{\mathbf{a}}$ , given by (10), with the simulated annealing algorithm.

**Output:** activation maps  $a_i, i = 0, \dots, V-1$

Fig. 7. Algorithm for the detection of activated areas.

### A. Data analysis

All the datasets were analyzed according to the algorithm described in Fig. 7. We used Coiflet filters with four vanishing moments. We use  $M = 2$  components in the mixture of Gaussian densities. Additional components can be used to describe the large variation in the activation strength [40]. However, as explained in section IV-C, these components become difficult to estimate. In all our experiments we observed that  $K = 3$  wavelet packets were sufficient to cluster the data into activated and background voxels. For comparison purposes, we analyzed the data using the general linear model [41]. We used only one regressor for this linear model. During the analysis of the *in-vivo* data, the regressor was obtained by convolving the canonical hemodynamic response, defined by (13), with the stimulus time-series. The regressor used for the analysis of the simulated data was the true “activated” signal. In this case the linear model had access to an “oracle” since it was given the perfect knowledge about clean the fMRI signal. We note that an spatially smoothing the raw data may create activated regions that are more connected. However, we believe that this approach will not create crisp activated regions, and could result in false positives. A Student  $t$ -test was applied to the regression coefficient to test its significance. The null hypothesis was that the regression coefficient was not significant. The activation map was obtained by thresholding the  $p$  value from the  $t$ -test.

### B. Synthetic data

Our synthetic datasets were designed by blending activation in non activated time-series that were extracted from a real *in vivo* dataset. The time-series  $s_{\mathbf{v}}(t)$  at each voxel  $\mathbf{v}$  is given by

$$s_{\mathbf{v}}(t) = (\beta_{\mathbf{v}} h) * x(t) + \nu_{\mathbf{v}}(t), \quad (11)$$

where  $\nu_v(t)$  is a non activated time-series extracted from non activated voxels in the checkerboard dataset, which is described in Section VI-C.1. The activation time-series  $f_v(t)$  was obtained by convolving the canonical hemodynamic filter  $h(t)$  used in SPM [26] with the stimulus time-series  $x(t)$ ,

$$f_v(t) = (\beta_v h) * x(t) \quad (12)$$

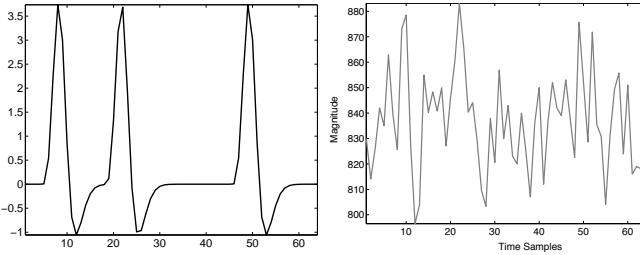


Fig. 8. Left:  $h * x(t)$ . Right: synthetic activated time-series constructed by using Eq. (11), with  $\beta = 1$

Group	1	2	3	4
SNR	0.352	0.441	0.516	0.591

TABLE I

SIGNAL TO NOISE RATIO OF THE SYNTHETIC DATA SET

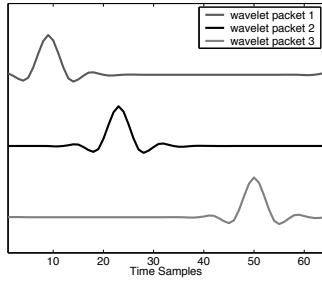


Fig. 9. The three best wavelet packets that give rise to the least Gaussian distribution of the coefficients.

where

$$h(t) = \left( \frac{t}{d_1} \right)^{a_1} e^{-(t-d_1)/b_1} - c \left( \frac{t}{d_2} \right)^{a_2} e^{-(t-d_2)/b_2}, \quad (13)$$

and  $\beta_v$  is the activation strength at voxel  $v$  (if  $\beta_v = 0$ , then  $v$  is a non activated). We use the following values for the parameters  $a_1 = 6$ ,  $a_2 = 12$ ,  $b_1 = 0.9$ ,  $b_2 = 0.9$ ,  $c = 0.35$ ,  $d_1 = a_1 b_1$  is the time to peak, and  $d_2 = a_2 b_2$  is the time to undershoot. Several synthetic datasets were generated with various level of activation strength  $\beta_v$ . The stimulus time-series  $x(t)$  consisted of three randomly placed stimulations (see Fig 8-left). The length of the time-series was  $T = 128$  sec, and  $TR = 2$  sec. Note that because the background time-series are directly extracted from the fMRI data, we have no strict control over their temporal variance. Therefore, even for time-series designed to have the same activation strength, the contrast between activation and background will not be exactly the same. We report here the results of the analysis of 80 synthetic datasets. For each dataset, the activation strength  $\beta_v$  was the same for all activated voxels. Rather than trying to

model the geometry of the cortex, we designed four different slices with activated regions with different shapes (see Fig. 13). The four images had a size of  $20 \times 20$  voxels. There were altogether 108 activated voxels, which was 6.75% of the total number of voxels in each dataset. The signal to noise ratio (ratio of the standard deviation of  $(\beta_v h) * x$

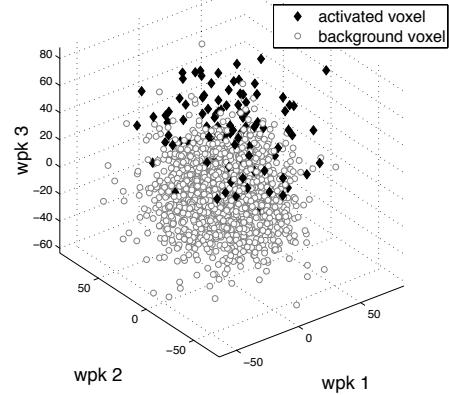


Fig. 10. Each dot is a time-series (black diamond = activated, white circle = background). Labels (activated vs non activated) are based on the ground truth.

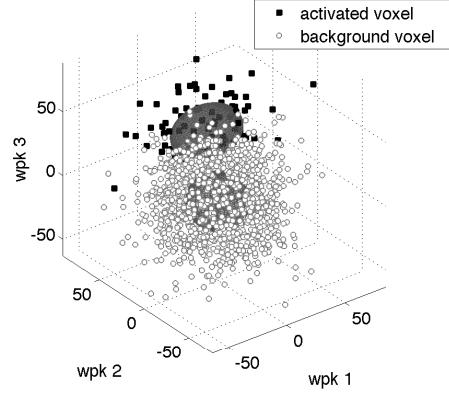


Fig. 11. Each dot is a time-series (black diamond = activated, white circle = background). Labels obtained after the estimation. Ellipsoids describe the isosurface located at one standard deviation away from the mean of each Gaussian density.

over the standard deviation of the detrended noise  $\nu_v$ ) was estimated for each dataset. We used four groups (1, 2, 3 and 4) of 20 datasets each. The level of activation ( $\beta_v$ ) was the same within each group, and we computed the average signal to noise ratio for each group (see Table I). Figure 9 shows the three best wavelet packets that were selected by our method based on the maximization of the skewness. The wavelet packets play the role of second order derivative operator. They are centered around the three stimuli onsets, and will be able to detect the variation in the fMRI signal triggered by the stimulus. We project each time-series  $s_v(t)$  on the three wavelet packets shown in Fig. 9, and compute three coordinates (the three wavelet packet coefficients) for  $s_v(t)$ . Figure 10 shows the projections of all the time-series from one dataset of group 1 in the coordinate system formed

by the three wavelet packets. The labels (red diamond for activated time-series, and blue circle for background time-series) are based on the ground truth. Clearly the two sets of time-series overlap significantly ; this is a consequence of the low activation strength. We expect that the classification will be difficult. The labels shown in figure 11 are now estimated by our method. We have drawn two ellipsoids in Fig. 11. Each ellipsoid corresponds to an isosurface located at one standard deviation away from the mean of the Gaussian distribution. The ellipsoids provide a visual clue about the spread of each component: background and activated voxels.

We compared our approach to a linear model equipped with the perfect knowledge of the hemodynamic response. A Student *t*-test was applied to the regression coefficient to test its significance, and voxels with a *p*-value smaller than a threshold were considered activated. Figure 13 shows several activation maps obtained with one dataset of group 1. The ground truth is overlaid. The activation maps obtained with the linear model are shown in **A** for  $p = 0.001$  and in **B** for  $p = 0.005$ . Clearly the linear model at  $p = 0.001$  is missing many activated voxels, and contains many false positives at  $p = 0.005$ . In **C** we show the activation maps obtained without spatial regularization: the label is based only on the posterior probability in equation (4). Finally, **D** shows the activation map obtained using the MRF model. Even though the SNR is low for this dataset, our approach is able to recover most activated voxels without any false activation. Figure 12 provides a quantitative comparison of our approach with the linear model. The average numbers of false alarms and missed activated voxels are plotted as a function of the group number (the SNR increases monotonically from group 1 to group 4 ; see Table I). We should keep in mind that the linear model has access to an oracle, in the form of the perfect knowledge of the hemodynamic response  $h(t)$ , and should therefore perform very well. In fact, if the noise added in (11) were to be white, we know from the matched filter theorem that the linear model would provide the optimal detector. In this experiment the noise is extracted from the data, and is probably not white [15]. We note that our approach outperforms the linear model, for  $p = 0.005$  and  $p = 0.001$ , in terms of missed activations. Our method is also almost optimal in terms of false alarm rate: only a very conservative choice of the *p*-value outperforms our method. Obviously this choice of *p* results in the worst performance in terms of missed activation. We also note that the the MRF model helps reducing the number of false positives, and increase the number of true detections.

### C. In vivo data

*1) Visual stimuli.* : This dataset (collected by Dr. Tanabe, University of Colorado) demonstrates activation of the visual cortex using a block paradigm. A flashing checkerboard image was presented to a subject for 30 seconds, and a blank image was presented for the next 30 seconds. This alternating cycle was repeated four times. Images were acquired with a 1.5 T Siemens MAGNETOM Vision equipped with a standard quadrature head coil and an echo planar subsystem. Functional images were acquired using a gradient-echo echo planar

sequence (TR = 3s, FOV = $24 \times 24$  cm, slice thickness = 3 mm, imaging matrix=  $128 \times 128$ , voxel size= $1.88 \times 1.88 \times 3$  mm). Eighty images were obtained from each axial slice for a total of 12 contiguous slices. We conducted our experiment using slices 8, 9, 10, and 11. In order to maintain an activation ratio of about 5 to 10 percent,

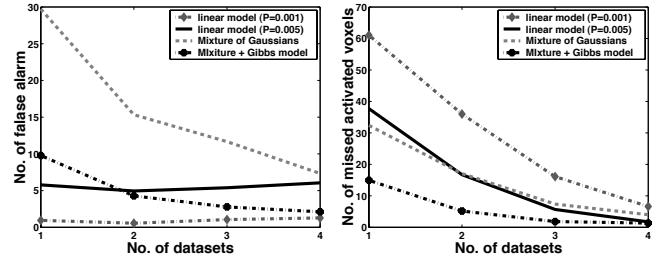


Fig. 12. Average number of false alarms (left) and missed activated voxels (right) as a function of the group number (the SNR increases monotonically from group 1 to group 4 ; see Table I).

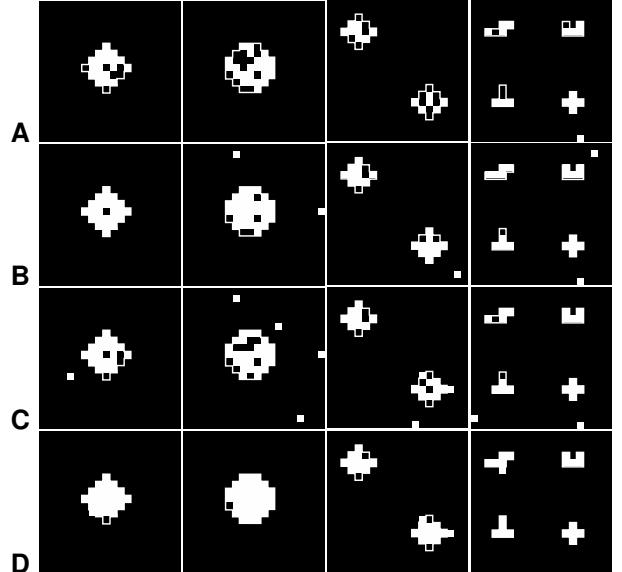


Fig. 13. Activation maps (left to right: slice 1 to slice 4) Maps obtained with the linear model for  $p = 0.001$  (**A**) and for  $p = 0.005$  (**B**). Activation maps obtained without spatial regularization (the label is based on the posterior probability in equation (4)) (**C**). Finally, **D** shows the activation map obtained using the MRF model.

we further selected a window with its upper left corner (43, 19) and lower right corner (85, 40). In total there were 3084 valid voxels inside the window. Fig. 14-left shows the three best wavelet packets that maximized the skewness of the distribution of the wavelet packet coefficients. We note that the second best wavelet packet is almost periodic with a period matching the period of the stimulus. The first and second wavelet packet reinforce the first and last oscillations of the second wavelet packet. Fig. 14-right shows an activated time-series in the visual cortex. Figure 15 shows the projections of all the time series corresponding to voxels inside the analysis window. The coordinate system is formed by the wavelet packets shown in Fig. 14. The labels (red diamond for activated time series, and blue circle for background time-series) are based on our

algorithm. We note that the covariance matrix of the activated component, modeled as an ellipsoid, is much more elongated than the covariance matrix of the background component. This observation could justify using one additional component to model the two levels of activation, as is described in [40]. Figure 18 shows activation maps obtained with the linear model. The regressor was obtained by convolving the

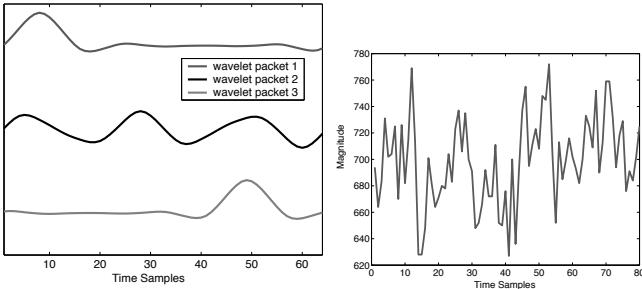


Fig. 14. Left: the three best wavelet packets that give rise to the most interesting projections of the time-series. The top and bottom wavelet packets have a good time localization. The middle one has a good frequency localization. Right: example of an activated time-series.

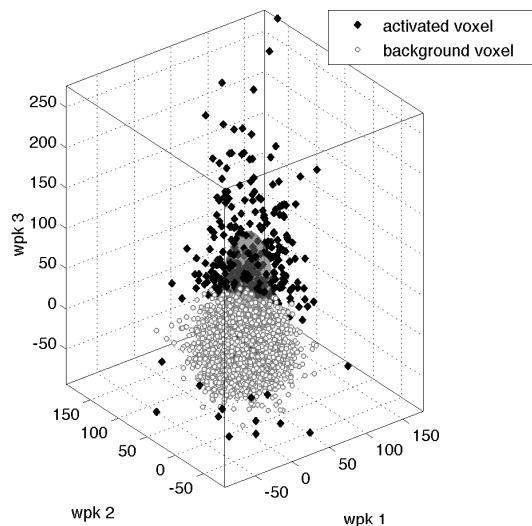


Fig. 15. Each dot is a time-series (black diamond = activated, white circle = background). Labels obtained after the estimation. Ellipsoids describe the isosurface located at one standard deviation away from the mean of each Gaussian density.

hemodynamic response, defined by (13), with the stimulus time series. Voxels with a  $p$ -value smaller than 0.001 were considered activated and color coded red. We note that activated voxels appear disconnected. Figure 19 shows activation maps obtained with our approach. We note that activated voxels are more connected, resulting in large activated areas. The activated areas detected by both methods appear to be in the visual cortex.

2) *Event-related fMRI study of visual-mental imagery* : The original dataset was collected by Dr. Klein and collaborators [42] and was obtained from the fMRI data center [43]. The experiment was designed to measure cerebral activity in the primary visual cortex when subjects where asked to perform mental imagery task with their eyes closed. We only analyze

here the runs associated with a concrete imagery task: a subject first heard the name of an animal and formed a visual image of that animal (e.g. a cat). Fourteen seconds later the subject is asked about the presence or absence of a specific feature (e.g. pointed ears) of the animal's body. Participants reported their answer to the question by pressing a 'yes' or 'no' response key, with one key placed in each hand. Other trials involve abstract features of the animal, and are not discussed here.

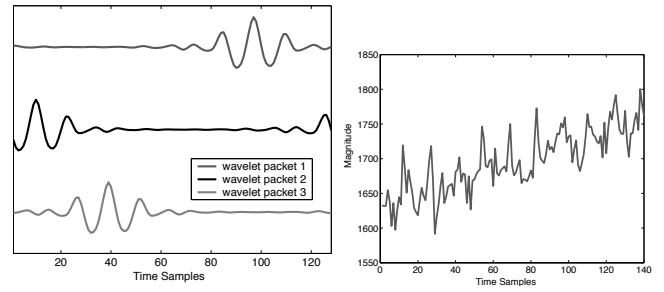


Fig. 16. Left: the three best wavelet packets that give rise to the most interesting projections of the time-series. All wavelet packets are similar, and behave as smooth second order derivatives. Right: example of an activated time-series.

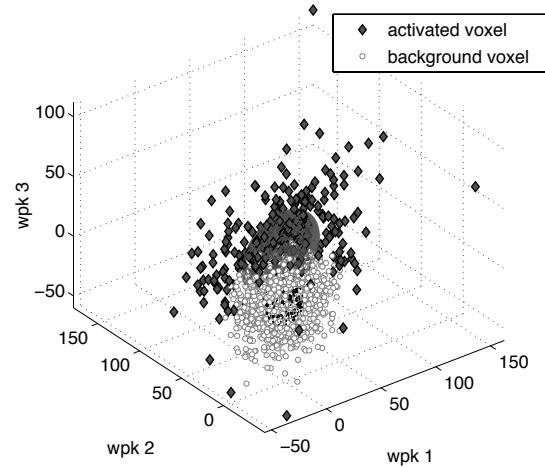


Fig. 17. Each dot is a time-series (black diamond = activated, white circle = background). Labels obtained after the estimation. Ellipsoids describe the isosurface located at one standard deviation away from the mean of each Gaussian density.

Each run consists of 140 images ( $TR = 2$  seconds), and a periodic auditory stimulus meant to evoke a visual response is presented every 14 seconds. Details about the data acquisition and the experiment design can be found in [42]. In this work we analyze the functional data of the first subject. We focus our analysis on the last two slices where we expect to see activation. The average of the three runs is computed at each voxel and used as the input for our analysis.

Fig. 16-left shows the three best wavelet packets that maximized the skewness of the distribution of the wavelet packet coefficients. All wavelet packets are localized and behave as smooth second order derivative. Fig. 16-right shows an activated time-series. The wavelet packet will detect changes in the signal that are triggered by the stimulus. Figure 17

shows the projections of all the time-series corresponding to voxels inside the window of analysis. The coordinate system is formed by the wavelet packets shown in Fig. 16. The labels (red diamond for activated time-series, and blue circle for background time-series) are based on the mixture of Gaussian model. As observed in the previous dataset, we notice that covariance matrix of the activated component, modeled as an ellipsoid, is much more elongated than the covariance matrix of the background component. Figure 20 shows activation

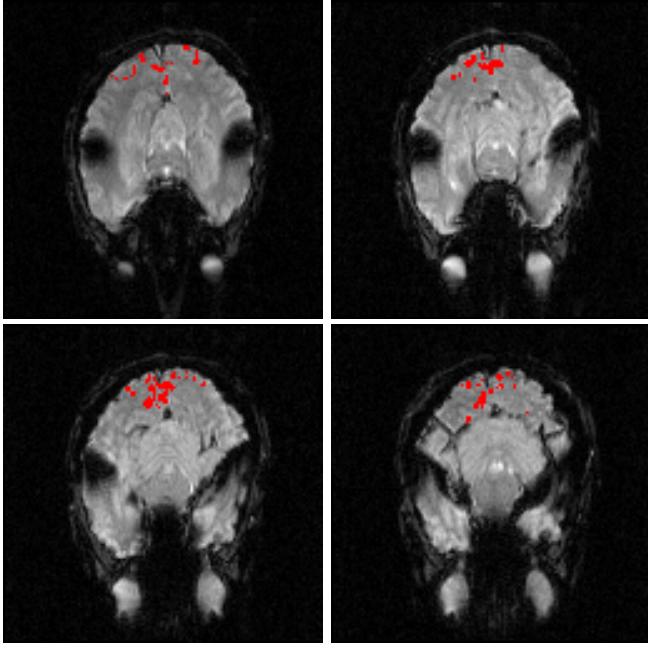


Fig. 18. Activation maps (top to bottom, left to right: slices 8 to 11) obtained with the linear model,  $p = 0.001$ .

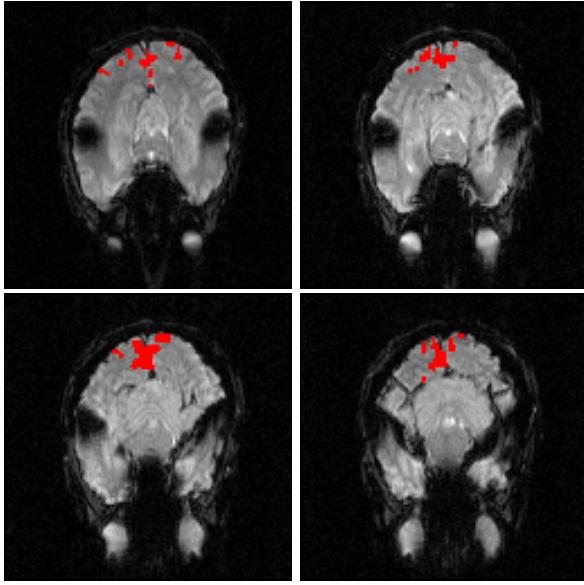


Fig. 19. Activation map obtained by our approach. From top to bottom, left to right are slice 8 to slice 11.

maps obtained with the linear model. The regressor was obtained by convolving the hemodynamic response, defined

by (13), with the stimulus time series. Voxels with a  $p$ -value smaller than 0.001 were considered activated. Activated voxels appear disconnected. Figure 21 shows activation maps obtained with our approach. Activated areas appear to be in the somatosensory (left and right posterior parietal cortex (**A**), pre-motor and motor areas (**B**), supplementary motor area (**C**) [44]. In addition, frontal and supplementary eye fields could also be activated (in regions **B** and **C** respectively. These areas are triggered by the stimulus and the task.

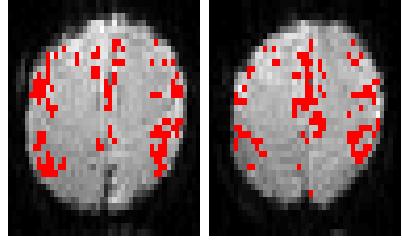


Fig. 20. Activation map obtained with the linear model (left: slice 17, right: slice 18). Voxels with a  $p$ -value smaller than 0.001 were considered activated and colored in red.

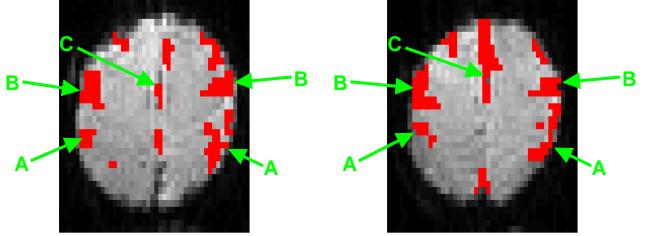


Fig. 21. Activation map obtained with our method (left: slice 17, right: slice 18). **A**: somatosensory, **B** motor and pre-motor , and **C** supplementary motor areas. In addition, frontal and supplementary eye fields could also be activated (in regions **B** and **C** respectively).

## VII. CONCLUSION

We have described in this paper a new method for detecting activation in fMRI data. We take advantage of the implicit low dimensionality of the fMRI dataset in order to project the time series on a low dimensional subspace. The subspace is chosen in order to create projections that are as non Gaussian as possible. Our approach achieves two goals: it reduces the dimensionality of the problem by explicitly constructing a sparse approximation to the dataset and it also create meaningful clusters allowing the separation of the activated regions from the clutter formed by the background time series. We expect activated areas that are connected, and impose a spatial prior in the form of a Markov random field. Our approach was validated with realistic synthetic data and outperformed a linear model equipped with the knowledge of the true hemodynamic response. Experiments with in vivo data indicate that our approach can find activated regions that are similar to those found by a linear model, yet more connected. We are currently extending the paradigm of “dimensionality reduction and clustering” to projections on manifolds instead of linear subspaces [45].

## ACKNOWLEDGMENTS

XS and FGM were supported by the Institute for Pure and Applied Mathematics, UCLA, during the semester on Multiscale Geometry and Analysis in High Dimensions. The authors are very grateful to the reviewers for their comments and suggestions.

## REFERENCES

- [1] F. Meyer and G. McCarthy, "Wavelet based estimation of baseline drifts in fMRI," in *Information Processing in Medical Imaging*, M. F. Insana and R. M. Leahy, Eds. Springer-Verlag, 2001, pp. 232–238.
- [2] F. Meyer, "Wavelet based estimation of a semi parametric generalized linear model of fMRI time-series," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 315–322, March 2003.
- [3] E. Zarahn, G. Aguirre, and M. D'Esposito, "Empirical analysis of fMRI statistics : I. Spatially unsmoothed data collected under null hypothesis conditions," *NeuroImage*, vol. 5, pp. 179–197, 1997.
- [4] J. Fadili and E. Bullmore, "Wavelet-generalized least squares : a new BLU estimator of linear regression models with  $1/f$  errors," *NeuroImage*, vol. 15, pp. 217–232, 2002.
- [5] K. Friston, O. Josephs, G. Rees, and R. Turner, "Nonlinear event-related responses in fMRI," *Magn. Reson. Med.*, vol. 39, pp. 41–52, 1998.
- [6] K. Miller, W. Luh, T. Liu, A. Martinez, T. Obata, E. Wong, L. Frank, and R. Buxton, "Nonlinear temporal dynamics of the cerebral blood flow response," *Human Brain Mapping*, vol. 13, pp. 1–12, 2001.
- [7] A. Vazquez and D. Noll, "Nonlinear aspects of the BOLD response in functional MRI," *Human Brain Mapping*, vol. 7, pp. 108–118, 1998.
- [8] I. Vanzetta and A. Grinvald, "Increased cortical oxidative metabolism due to sensory stimulation: implications for functional brain imaging," *Science*, vol. 286, pp. 1555–8, Nov. 1999.
- [9] M. Brett, I. Johnsrude, and A. Owen, "The problem of functional localization in the human brain," *Nature Reviews, Neuroscience*, vol. 3, pp. 243–249, 2002.
- [10] M. Gabbay, C. Brennan, E. Kaplan, and L. Sirovich, "A principal components-based method for the detection of neuronal activity maps: application to optical imaging," *NeuroImage*, vol. 11, pp. 313–325, 2000.
- [11] B.B. Biswal and J. Ulmer, "Blind source separation of multiple signal sources of fMRI data sets using independent component analysis," *Journal of Computer Assisted Tomography*, vol. 23(3), pp. 265–271, 1999.
- [12] M. McKeown, "Detection of consistently task-related activations in fMRI data with hybrid independent component analysis," *NeuroImage*, no. 11, pp. 24–35, 2000.
- [13] A. Hyvärinen, "Survey on independent component analysis," *Neural Computing Surveys*, pp. 94–128, 1999.
- [14] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, and P. Boesiger, "A new correlation-based fuzzy logic clustering algorithm for fMRI," *Magnetic Resonance in Medicine*, vol. 40, pp. 249–260, 1998.
- [15] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological (fMRI) time series analysis : resampling methods in time and wavelet domain," *Human Brain Mapping*, vol. 78, pp. 61–78, 2001.
- [16] V. Maxim, L. Xendur, J. Fadili, J. Suckling, R. Gould, R. Howard, and E. Bullmore, "Fractional Gaussian noise, functional MRI and alzheimer's disease," *NeuroImage*, vol. 25, pp. 141–158, 2005.
- [17] P. Ciuciu, P. Abry, C. Pallier, and A. Roche, "Log-cumulant wavelet-leaders multifractal analysis for fmri time series," in *submitted to Proc. MMBIA*, New York., June 2006.
- [18] P. Abry, P. Flandrin, M. Taqqu, and D. Veitch, "Self-similarity and long-range dependence through the wavelet lens," in *Theory and applications of long-range dependence*, 2003, pp. 527–556.
- [19] P. Mitra and B. Pesaran, "Analysis of dynamic brain imaging data," *Biophysical Journal*, vol. 76, pp. 691–708, 1999.
- [20] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [21] F. Meyer, A. Averbuch, and J.-O. Strömberg, "Fast adaptive wavelet packet image compression," *IEEE Transactions on Image Processing*, pp. 792–800, May 2000.
- [22] P. Diaconis and D. Freedman, "Asymptotics of graphical projection pursuit," *Annals of Statistics*, vol. 12, no. 3, pp. 793–815, 1984.
- [23] R. Duda, P. Hart, and D. Stark, *Pattern classification*. Wiley, 2001.
- [24] J. Chinrungrueng, "Local clustering of functional magnetic resonance data," Ph.D. dissertation, University of Colorado at Boulder, Dept. of Elec. Engineering, November 2002.
- [25] S. Shapiro and M. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, pp. 591–611, 1965.
- [26] K. Friston, A. Holmes, K. Worsley, J. Poline, C. Frith, and R. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, pp. 189–210, 1995.
- [27] C. Chen, C. Tyler, and H. Baseler, "Statistical properties of BOLD magnetic resonance activity in the human brain," *NeuroImage*, vol. 20, pp. 1069–1109, 2003.
- [28] F. Meyer and J. Chinrungrueng, "Analysis of event-related fMRI data using best clustering bases," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 933–939, Aug. 2003.
- [29] ———, "Analysis of event-related fMRI data using best clustering bases," in *Information Processing in Medical Imaging*, C. Taylor and J. Noble, Eds. Springer-Verlag, 2003, pp. 623–634.
- [30] ———, "Clustering of spatio-temporal signals," in *Proc. IEEE Workshop on Statistical Signal Processing*, 2003, pp. 469–472.
- [31] B. Everitt and E. Bullmore, "Mixture model mapping of brain activation in functional magnetic resonance images," *Human Brain Mapping*, vol. 7, pp. 1–14, 1999.
- [32] N. Hartvig and J. Jensen, "Spatial mixture modeling of fMRI data," *Human Brain Mapping*, vol. 11, pp. 233–248, 2000.
- [33] M. Woolrich, T. Behrens, C. Beckmann, and S. Smith, "Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data," *IEEE Transactions on Medical Imaging*, vol. 24, pp. 1–11, 2005.
- [34] W. Penny and K. Friston, "Mixture of general linear models for functional neuroimaging," *IEEE Trans. on Medical Imaging*, vol. 22, pp. 504–14, 2003.
- [35] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*. Wiley, 1997.
- [36] J. Verbeek, N. Vlassis, and B. Kroese, "Efficient greedy learning of Gaussian mixtures," *Neural Computation*, vol. 15, pp. 469–485, 2003.
- [37] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 6, 6, pp. 721–741, Nov. 1984.
- [38] F. Wu, "The Potts model," *Review of modern physics*, vol. 54, no. 1, pp. 235–270, Jan. 1982.
- [39] S. Kirkpatrick, C. Gelatt, , and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.
- [40] F. Meyer and X. Shen, "Multiscale analysis of fMRI data with mixture of Gaussian densities," in *Proc. IEEE International Symposium on Biomedical Imaging*, 2004, pp. 1175–1178.
- [41] K. Worsley, J. Poline, K. Friston, and A. Evans, "Characterizing the response of fMRI data using multivariate linear models," *NeuroImage*, vol. 2, pp. 305–319, 1997.
- [42] I. Klein, A. Paradis, J. Poline, S. Kosslyn, and D. Bihan, "Transient activity in the human calcarine cortex during visual-mental imagery: An event-related fMRI study," *Journal of Cognitive Neuroscience*, vol. 12, pp. 15–32, 2000.
- [43] "<http://www.fmridc.org/f/fmridc/>"
- [44] T. Scarabino and U. Salvolini, *Atlas of Morphological and Functional Anatomy of the Brain*. Springer, 2006, in collaboration with F.I. Dalle, H. Duvernoy and P. Rabischong.
- [45] X. Shen and F. Meyer, "Analysis of event-related fMRI data using diffusion map," in *Information Processing in Medical Imaging*, G. Christensen and M. Sonka, Eds. Springer-Verlag, LNCS 3565, 2005, pp. 652–663.