fMRI Detection via Variational EM Approach

Wanmei Ou

Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA

Abstract. In this paper, we study Gaussian Random Fields (GRFs) as spatial smoothing priors in Functional Magnetic Resonance Imaging (fMRI) detection, and we solve GRFs using the variational Expectation-Maximization (EM) algorithm. Relatively high noise in fMRI images presents a serious challenge for the detection algorithms, creating a need for spatial regularization of the signals. Spatial regularization is usually employed before or after detection, forming a two-process detector. In a two-process detector, defects produced in the first process usually interfere with the performance of the second process. Among all two-process detectors, the Gaussian-smoothing-based detector, which performs detection on spatially smoothed signals, is the most popular. Gaussian filters, traditionally employed to boost the signal-to-noise ratio, often remove small activation regions. This is mainly caused by applying a fixed, but arbitrary, Gaussian filter uniformly over the entire image. In this work, we propose the EM-GRF-based detector that iterates between these two processes, so that parameters of the Gaussian filter are readjusted according to the feedback from detection, and the parameters for detection are readjusted according to the feedback from the filtered results. In addition, we compare the performance of this detector with the Gaussian-smoothing-based detector through ROC analysis on simulated data, and demonstrate their applications in a real fMRI study.

1 Introduction

Functional magnetic resonance imaging (fMRI) has provided researchers with a non-invasive dynamic method for studying brain activation by capturing the change in blood oxygenation levels. Most fMRI detection algorithms operate by comparing the time course of each voxel with the experimental protocol, labelling the voxels whose time courses correlate significantly with the protocol as "active." The commonly used General Linear Model (GLM) [6] further assumes that the fMRI signal possesses linear characteristics with respect to the stimulus and that the temporal noise is white. Furthermore, converting the GLM estimates of each voxel into confidence interval statistics, such as P-value and Z-score, forms a so-called Statistical Parametric Map (SPM). A confidence level is selected to reject the null hypothesis, which is non-activation in this application. The signal from those voxels whose P-value or Z-score above a selected confidence level is unlikely to be explained by the null hypothesis. Therefore, these voxels are labelled as active voxels, and a binary map of active areas is created. In some applications, the sign of the GLM estimate is augmented into the confidence statistic, producing a trinary activation map indicating both positive and negative activations. However, because of a low signal-to-noise ratio (SNR), the binary maps typically contain many small false positive islands, creating a need for spatial regularization.

A common approach to reducing such false detections employs a Gaussian filter to smooth the fMRI signal prior to applying the GLM detector, and we refer this detector as the Gaussian-smoothing-based detector. Unfortunately, Gaussian smoothing, though intended to combat low SNR, leads to overly smoothed SPMs and a loss of detail in the resulting activation maps. In addition, the window size of the Gaussian filter is entirely arbitrary, which makes comparison of results difficult. A number of alternative approaches have explicitly incorporated spatial and temporal correlations into the estimation procedure. Examples include autoregressive spatio-temporal models [2, 15], Markov Random Fields (MRFs) [3, 5, 4], Bayesian models inferring hidden psychological states [7], adaptive thresholding methods that adjust statistical significance of active regions according to their size. Similarly to the Gaussian-smoothing-based detector, most of these detectors[1] involve two processes: spatial regularization and detection. Therefore, defects

[1] Cosman's MRF-based detector [3] combines detection and spatial regularization into one process, but it is restricted to binary activation configuration.

produced in the first process will be carried over to the second one. For example, mild MRF regularization usually cannot remove false positive voxels which acquire strong activation statistics in the detection process.

We are interested in a real-number representation of brain activation, rather than discrete activation state representation. By employing the two-gamma function [8][2] as a protocol-dependent regressor in the GLM model, we are able to quantify activation using a real number. This enables us to distinguish positive from negative activations, as well as activation strength. Gaussian Random Fields (GRF) is employed for modelling spatial coherency, rather than a discrete spatial model, because of the real-number representation of activation. We assume that, given the activation strength of each voxel, the time courses of different voxels are conditionally independent. Spatial coherency is a model of the activation, not the observed time courses. Different from the approaches mentioned in the previous paragraph, our approach unifies spatial regularization and detection into an iterative process. Due to the fact that neither the parameters in the signal model nor the parameters in the GRF are known, we adopt the variational Expectation-Maximization (EM) approach. Parameters estimated in one process can be readjusted given the feedback from the other process until convergence occurs. We compare the performance of our detector with the conventional Gaussian-smoothing-based detector.

In the next section, we present the background of GLM and briefly outline how the GLM detector can be augmented with a GRF spatial prior. Section 3 reviews the variational EM algorithm followed by empirical evaluation of the detectors on simulated data and real data sets in Section 4. Conclusions and discussions are in Section 5.

2 General Linear Model Augmented with a GRF Spatial Prior

2.1 Background

An fMRI scan contains a time course for each voxel. GLM models the fMRI signal as a linear combination of the protocol-dependent signals and the protocol-independent signals, such as cardiopulmonary factors. GLM assumes the brain behaves as a Linear Time Invariant (LTI) system. The presence of the protocol-dependent signal indicates that the corresponding voxel is active due to the stimulus.

Since we focus on using a real number to represent activation strength in this work, we employ one of the most commonly used Hemodynamic Response Functions (HRF), a two-gamma function, to construct the model of the protocol-dependent signal, $\boldsymbol{\gamma}$. $\boldsymbol{\gamma}$ is a column vector, resulting from a convolution of the HRF and the experimental protocol. x_i indicates the strength of the protocol-dependent signal. The protocol-independent signal is unknown, but it is of a lower frequency compared with the protocol-dependent signal. One of the most common approaches is to model the protocol-independent signal as a constant (DC signal); other approaches include low-frequency Fourier components or low-order polynomials. In this paper, we adopt the former approach. Therefore, we remove the DC component of the original fMRI signal in the preprocessing step. The zero-mean fMRI signal is $\boldsymbol{y_i} \in \mathbb{R}^{N_T}$ $(i = 1, ..., N_V)$, where N_T is the number of time samples and N_V is the number of voxels in the scan. According to the GLM model, the signal is a noisy version of the protocol-dependent signal:

$$\boldsymbol{y_i} = x_i\boldsymbol{\gamma} + \boldsymbol{\epsilon}_i \tag{1}$$

for $i = 1, ..., N_V$. For white temporal noise, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \sigma_i^2 I)$. The least squares estimate of the activation strength of voxel i, x_i, is found through a linear regression:

$$\hat{x}_i = (\boldsymbol{\gamma}^T\boldsymbol{\gamma})^{-1}\boldsymbol{\gamma}^T\boldsymbol{y}_i, \tag{2}$$

[2] A two-gamma function captures the fact that there is a small dip after the HRF has return to zero: $h(t) = (t/d_1)^{a_1}\exp(\frac{-(t-d_1)}{b_1}) - c(t/d_2)^{a_2}\exp(\frac{-(t-d_2)}{b_2})$ where, $d_j = a_j b_j$ is the time to the peak, and $a_1 = 6$, $a_2 = 12$, $b_1 = b_2 = 0.9s$, and $c = 0.35$

and the corresponding F-statistic is given by $F_i = \hat{x}_i \hat{\Sigma}_{x_i}^{-1} \hat{x}_i / 1$, where the denominator corresponds to the number of regression coefficients in x_i, and it is 1 in our case. $\hat{\Sigma}_{x_i}$ is the estimated variance. Jezzard [8] shows the general derivation of F_i with multiple regression coefficients in x_i. A conventional GLM detector constructs SPM using P-value, converted from F_i. In contrast, we use x_i to represent activations in this work.

Let random variable $\boldsymbol{X} = [X_1, ..., X_{N_V}]$ represent an activation configuration of all voxels in the volume. Continuous random variable X_i represents activation strength of voxel i. $\boldsymbol{x} = [x_1, \cdots, x_{N_V}]$ is one possible configuration, which is also called the SPM. Fig. 1 depicts the graphical representation of the model. The conditional probability of obtaining an fMRI scan, y, given the activation strength of all voxels, $\boldsymbol{x}$, can be formulated as:

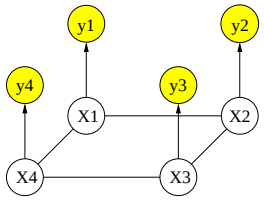

Fig. 1. *Graphical model for* $P_{\boldsymbol{X},Y}$.

$$P_{Y|\boldsymbol{X},\boldsymbol{\theta}}(y|\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{N_V} p_{\boldsymbol{Y}_i|X_i}(\boldsymbol{y}_i|x_i,\boldsymbol{\theta}) = \prod_{i=1}^{N_V} \mathcal{N}(\boldsymbol{y}_i; x_i\boldsymbol{\gamma}, \sigma_i^2 I) \tag{3}$$

where y denotes signals of all voxels in an fMRI volume. The second equality is obtained based on the assumption that the noisy signal of each voxel is independent given the activation strength of the voxels.

2.2 A Gaussian Random Field (GRF) Prior

Choosing a GRF prior to model brain activation corresponds to our interest in a real-number representation of brain activation. GRF is able to capture the biological fact that adjacent voxels tend to have similar activation level. It is more appropriate to quantify brain activation using a real number rather than quantizing it into binary or trinary states, since some regions of the brain should be more active than others associated with a particular task. Additionally, according to neuroscience literatures [1, 13, 10], some regions in the brain show negative activations when a subject performs certain tasks. For example, Schwartz [12] found that pre-frontal cortex regions usually shows negative activation during rapid eye movement sleep. Negative activation is caused by decreased glucose metabolism, but a further interpretation for negative activation is still an active research area.

GRF is formed by multivariate Gaussian variables. It is characterized by mean, $\boldsymbol{w}$, and covariance matrix, Σ. Estimating the covariance matrix is difficult given a limited amount of data, which is usually the case in fMRI studies. Since GRF is used to model spatial coherency, we pre-define the covariance based on the distance of the image grid.

$$\Sigma_{i,j} = 2 - \frac{2}{1 + e^{-D_{i,j}}} \tag{4}$$

where, D is the distance matrix of any pair of voxels. For example, $D_{i,j}$ stores the distance between voxel i and j. With the logistic function, the elements in Σ is normalized between zero and one. The structure of Σ agrees with our model that voxels appearing close should have a higher correlation. In contrast, we assume $\boldsymbol{w}$ is unknown in our model. Therefore, the prior distribution of the activation strength, $\boldsymbol{X}$ is

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{w}, \Sigma). \tag{5}$$

In this paper, we only use the proposed GRF to model voxels within each slice because most fMRI data is highly anisotropic with respect to the third spatial dimension. However, the extension to a three dimensional volume of data is straightforward but requires more computation time.

2.3 MAP Estimate

Presented in Eq. (3), the fMRI signal of each voxel is conditionally independent, given the activation strength. We seek the MAP estimate of $\boldsymbol{x}$ given the signals,

$$\begin{aligned}
\boldsymbol{x}^* &= \arg\max p_{\boldsymbol{X}|Y}(\boldsymbol{x}|y;\boldsymbol{\theta}) && (6)\\
&= \arg\max p_{\boldsymbol{X}|Y}(y|\boldsymbol{x};\boldsymbol{\theta}) p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) && (7)\\
&= \arg\max \prod_{i=1}^{N_V} \mathcal{N}(\boldsymbol{y}_i; x_i\boldsymbol{\gamma}, \sigma_i^2 I)\mathcal{N}(\boldsymbol{x};\boldsymbol{w},\Sigma) && (8)
\end{aligned}$$

where, $\boldsymbol{\theta}$ includes all of the parameters in the likelihood and prior models, $\{\sigma_i\ \forall i\}$, $\boldsymbol{w}$, and Σ. If all of the parameters are known, we can obtain a close form solution for $\boldsymbol{x}$. In practise, the parameters are unknown except that we treat Σ as a known parameter, and we approximate this MAP problem via the variational EM approach. The variational EM algorithm allows us to find a distribution, $q_{\boldsymbol{X}}(\boldsymbol{x})$, which closely approximates $p_{\boldsymbol{X}|Y}(\boldsymbol{x}|y;\boldsymbol{\theta})$. Then, we approximate $\boldsymbol{x}^*$ by $\hat{\boldsymbol{x}}^* = \arg\max q_{\boldsymbol{X}}(\boldsymbol{x})$.

3 Variational EM

To formulate our problem into the EM framework, we consider Y and $\boldsymbol{X}$ as the observed and hidden variables. In the EM algorithm, we would like to maximize the logarithm of the probability of the observed data, $L = \log P_Y(y;\boldsymbol{\theta}) = \log \int_{\boldsymbol{x}} P_{\boldsymbol{X},Y}(\boldsymbol{x},y;\boldsymbol{\theta})$. We cannot optimize L directly because it involves unknown functions $P_Y(y;\boldsymbol{\theta})$ or $P_{\boldsymbol{X},Y}(\boldsymbol{x},y;\boldsymbol{\theta})$. We maximize its lower bound $G(q_{\boldsymbol{X}},\boldsymbol{\theta})$ instead. The following is the derivation of obtaining this lower bound:

$$\begin{aligned}
L(\boldsymbol{\theta}) &= \log \int_{\boldsymbol{x}} P_{\boldsymbol{X},Y}(\boldsymbol{x},Y;\boldsymbol{\theta}) && (9)\\
&= \log \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \frac{P_{\boldsymbol{X},Y}(\boldsymbol{x},Y;\boldsymbol{\theta})}{q_{\boldsymbol{X}}(\boldsymbol{x})} && (10)\\
&\geq \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \log \frac{P_{\boldsymbol{X},Y}(\boldsymbol{x},Y;\boldsymbol{\theta})}{q_{\boldsymbol{X}}(\boldsymbol{x})} && (11)\\
&= E_q\left[\log \frac{P_{\boldsymbol{X},Y}(\boldsymbol{x},Y;\boldsymbol{\theta})}{q_{\boldsymbol{X}}(\boldsymbol{x})}\right] && (12)\\
&= G(q_{\boldsymbol{X}},\boldsymbol{\theta}) && (13)\\
&= -D(q_{\boldsymbol{X}}(\boldsymbol{x})||P_{\boldsymbol{X}|Y}(\boldsymbol{x}|y,\boldsymbol{\theta})) + \log p_Y(y;\boldsymbol{\theta}) && (14)
\end{aligned}$$

where, $q_{\boldsymbol{X}}(\boldsymbol{x})$ represents the set of admissible probability density functions of X. Eq.(11) is obtained according to Jensen's Inequality. $G(q_{\boldsymbol{X}},\boldsymbol{\theta})$ is less or equal to $L(\boldsymbol{\theta})$ because $D(\cdot)$, the KL-Divergence of two distributions, is a non-negative quantity. We can further expand the lower bound as the following:

$$\begin{aligned}
G(q_{\boldsymbol{X}},\boldsymbol{\theta}) &= \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \log \frac{p_{Y,\boldsymbol{X}}(y,\boldsymbol{x};\boldsymbol{\theta})}{q_{\boldsymbol{X}}(\boldsymbol{x})} && (15)\\
&= \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \log \frac{p_{Y|\boldsymbol{X},\boldsymbol{\theta}}(y|\boldsymbol{x},\boldsymbol{\theta}) P_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{q_{\boldsymbol{X}}(\boldsymbol{x})}\\
&= \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \log p_{Y|\boldsymbol{X},\boldsymbol{\theta}}(y|\boldsymbol{x},\boldsymbol{\theta}) + \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \log P_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) - \int_{\boldsymbol{x}} q_{\boldsymbol{X}}(\boldsymbol{x}) \log q_{\boldsymbol{X}}(\boldsymbol{x}). && (16)
\end{aligned}$$

According to the EM algorithm, we should set $q_{\boldsymbol{X}}(\boldsymbol{x}) = P_{\boldsymbol{X}|Y}(\boldsymbol{x}|y, \boldsymbol{\theta})$ at each E-step. At each M-step, we maximize $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$ over $\boldsymbol{\theta}$ while fixing $q_{\boldsymbol{X}}$. Even though we can obtain a close form solution for $P_{\boldsymbol{X}|Y}(\boldsymbol{x}|y, \boldsymbol{\theta})$ at the E-step in our problem, the algorithm requires marginalization of $P_{\boldsymbol{X}|Y}(\boldsymbol{x}|y, \boldsymbol{\theta})$ at the M-step, and it is highly computation intensive. For instance, when we update σ_i^2 at each M-step, we need to marginalize $P_{\boldsymbol{X}|Y}(\boldsymbol{x}|y, \boldsymbol{\theta})$ over $\boldsymbol{X}$ except X_i. In contrast, if $\{x_1,\ x_2, \cdots,\ x_{N_v}\}$ are independently distributed, the proposed MAP problem is similar to finding the MAP labels in a mixture of Gaussian problem, which can be solved through the EM algorithm. In the mixture of Gaussian problem, we can explicitly derive the *posteriori* distribution of X_i, $P_{X_i|\boldsymbol{Y}}(x_i|\boldsymbol{y}_i, \boldsymbol{\theta})$.

Inspired by the mixture of Gaussian problem, we restrict the family of admissible probability density function, $q_{\boldsymbol{X}}(\boldsymbol{x})$, as independently Gaussian distributions (Eq. (17)). In the case of probability distributions, the appropriate simplification comes from properties of independence. The simplest family of variational distribution is one where all the variables, $\{X_i :\ i = 1, ..., N_V\}$, are independent of each other.

$$q_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{N_V} \mathcal{N}(x_i; \omega_i, \eta_i^2) \tag{17}$$

where, ω_i and η_i^2 are the mean and variance of X_i's distribution. Some later derivations are simplified by writing the mean and variance into vector forms: $\boldsymbol{\omega} = [\omega_1, \cdots, \omega_{N_V}]^T$ and $\boldsymbol{\eta}^2 = [\eta_1^2, \cdots, \omega_{N_V}^2]^T$. The admissible probability density function reflects our interest of a real-number representation of activation because of the Gaussian distribution. We find $q_{\boldsymbol{X}}(\boldsymbol{x})$ among the set of admissible probability density functions that provides the best approximation to $P_{\boldsymbol{X}|Y}(\boldsymbol{x}|y, \boldsymbol{\theta})$. The consequence approach is called the variational EM algorithm.

You may notice that each hidden variable, X_i, is a continuous random variable, there is an infinite rather than a finite number of labels. Nevertheless, the EM algorithm and the variational EM algorithm are general to both continuous and discrete hidden variables. Summation in the conventional EM formulas are replaced by integration.

By substituting Eq. (3), (5) and (17) into Eq. (16), we get

$$\begin{aligned} G(q_{\boldsymbol{X}}, \boldsymbol{\theta}) =& \frac{N_T}{2} \int_{i=1}^{N_V} \log(2\pi\sigma_i^2) - \frac{1}{2} \int_{i=1}^{N_V} [\boldsymbol{y}_i^T \boldsymbol{y}_i / \sigma_i^2 - 2\mu_i \boldsymbol{y}_i^T \boldsymbol{\gamma} / \sigma_i^2 + \boldsymbol{\gamma}^T \boldsymbol{\gamma} (\mu_i^2 + \eta_i^2)] - \frac{1}{2} \log(|\Sigma|) \\ & - \frac{N_V}{2} \log(2\pi) - \frac{1}{2} \left[\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} + \text{diag}(\Sigma^{-1})^T \boldsymbol{\eta}^2 - 2\boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\omega} + \boldsymbol{\omega}^T \Sigma^{-1} \boldsymbol{\omega} \right] \\ & + \frac{1}{2} \int_{i=1}^{N_V} \log(2\pi e \eta_i^2). \end{aligned} \tag{18}$$

The following outlines the E and M steps in the variational EM algorithm:

- E-Step: Maximize $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$ with respect to the admissible function, $q_{\boldsymbol{X}}$. This means to maximize $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$ with respect to the variational parameters, μ_i and η_i^2, $\forall i$.
- M-Step: Maximize $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$ with respect to the model parameters, $\boldsymbol{\omega}$ and σ_i^2, $\forall i$.

It is important to note that the lower bound, $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$, may not touch $L(\boldsymbol{\theta})$ in each E-step in the variational EM algorithm, while the EM algorithm guarantees this touch.

In the E-step, we set the partial derivatives of $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$ to zero with respect to the parameters in the admissible probability density function, μ_i and η_i^2.

$$\begin{aligned} \frac{\partial G(q_{\boldsymbol{X}}, \boldsymbol{\theta})}{\partial \boldsymbol{\mu}} &= S\boldsymbol{y}_r - \boldsymbol{\gamma}^T \boldsymbol{\gamma} S \boldsymbol{\mu} - \Sigma^{-1} \boldsymbol{\mu} + \Sigma^{-1} \boldsymbol{\omega} = 0 \\ &\Rightarrow \boldsymbol{\mu} = (\boldsymbol{\gamma}^T \boldsymbol{\gamma} S + \Sigma^{-1})^{-1} (S\boldsymbol{y}_r + \Sigma^{-1} \boldsymbol{\omega})^{-1} \end{aligned} \tag{19}$$

where, $\boldsymbol{y}_r = [\boldsymbol{y}_1^T\boldsymbol{\gamma},\ \cdots,\ \boldsymbol{y}_N^T\boldsymbol{\gamma}]^T$, $\boldsymbol{\sigma^2}^T = [\sigma_1^2,\ \cdots,\ \sigma_N^2]^T$, and $S = (\boldsymbol{\sigma^2}^T I)^{-1}$.

$$\begin{aligned}\frac{\partial G(q_{\boldsymbol{X}}, \boldsymbol{\theta})}{\partial \eta_i^2} &= -\frac{1}{2\sigma_i^2}\boldsymbol{\gamma}^T\boldsymbol{\gamma} + \frac{1}{2\eta_i^2} - \frac{1}{2}\mathrm{diag}(\Sigma^{-1})_i = 0 \\ &\Rightarrow \eta_i^2 = \frac{1}{\mathrm{diag}(\Sigma^{-1})_i + \boldsymbol{\gamma}^T\boldsymbol{\gamma}/\sigma_i^2}\end{aligned} \tag{20}$$

where, $\mathrm{diag}(\Sigma^{-1})_i$ is the i^{th} element in the vector formed by the diagonal of Σ^{-1}.

In the M-step, we set the partial derivatives of $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$ to zero with respect to the modelling parameters, ω_i and σ_i^2.

$$\begin{aligned}\frac{\partial G(q_{\boldsymbol{X}}, \boldsymbol{\theta})}{\partial \boldsymbol{\omega}} &= \Sigma^{-1}\boldsymbol{\mu} - \Sigma^{-1}\boldsymbol{\omega} = 0 \\ &\Rightarrow \boldsymbol{\omega} = \boldsymbol{\mu}\end{aligned} \tag{21}$$

and

$$\begin{aligned}\frac{\partial G(q_{\boldsymbol{X}}, \boldsymbol{\theta})}{\partial \sigma_i^2} &= -\frac{N_T}{2\sigma_i^2} + \frac{1}{2(\sigma_i^2)^2}[\boldsymbol{y}_i^T\boldsymbol{y}_i - 2\boldsymbol{y}_i^T\boldsymbol{\gamma}\mu_i + \boldsymbol{\gamma}^T\boldsymbol{\gamma}(\mu_i^2 + \eta_i^2)] = 0 \\ &\Rightarrow \sigma_i^2 = \frac{1}{N_T}(\boldsymbol{y}_i^T\boldsymbol{y}_i - 2\boldsymbol{y}_i^T\boldsymbol{\gamma}\mu_i + \boldsymbol{\gamma}^T\boldsymbol{\gamma}(\mu_i^2 + \eta_i^2)).\end{aligned} \tag{22}$$

We stop the iteration when the lower bound, $G(q_{\boldsymbol{X}}, \boldsymbol{\theta})$, converges to a local maximum. As mentioned before, we would like to obtain $\hat{\boldsymbol{x}}^* = \arg\max q_{\boldsymbol{X}}(\boldsymbol{x})$. Since $q_{\boldsymbol{X}}(\boldsymbol{x})$ is composed by independently distributed Gassian distributions, $\hat{\boldsymbol{x}}^* = \boldsymbol{\omega}$. The ROC analysis, the confusion matrix analysis, and the thresholded activation maps presented in the next section are created from $\boldsymbol{\omega}$.

4 Results

In this section, we used both synthetic and real fMRI data to compare the performance of the proposed EM-GRF-based detector and the Gaussian-smoothing-based detector.

4.1 Synthetic Data Sets

To quantitatively evaluate the performance of the method, we generated realistic phantom data by applying EM segmentation [9] to an anatomical MRI scan and placing activation areas of variable size (average diameter of 15mm) randomly in the gray matter. We randomly assigned the selected activation regions as either positively or negatively active regions. In other words, all of the voxels in an activation region have an identical activation state; voxels belonging to different activation regions, although they are very close, may be in different activation states. We then downsampled the scan to an fMRI resolution. The gray matter voxels represent 10% of the total number of voxels in a volume, and 7% and 3% of the gray matter voxels are positively and negatively active voxels. We created simulated fMRI scans based on a fixed parametric hemodynamic response function, on an event-related protocol, and on varying levels of noise. Fig. 2 displays the synthetic signals before and after adding noise.

More realistic synthetic data should include variation in activation strength in selected activation regions. However, this type of data prevents us from performing ROC analysis and confusion matrix analysis. Therefore, we chose to have identical activation strength for the selected active voxels in our synthetic data. In other words, the simulated HRF for the negative voxels is the negative version of the HRF for the positive voxels. As mentioned before, the EM-GRF-based detector is flexible enough to

handle variation in activation strength, so it can handle equal activation strength with minor reduction in statistical power. Results of applying this detector to real fMRI data can better illustrate the advantages of this detector.

We used the estimated SNR, $\widehat{\mathrm{SNR}} = -10\log_{10}(|\boldsymbol{\gamma}\hat{x}|^2)/\hat{\sigma^2}$, to determine a realistic level of the simulated noise, as the true SNR is unaccessible for real fMRI scans. Since the signal and noise overlap in some frequency bands, part of the noise energy is assigned to the estimated signal during detection. The estimated SNR is therefore an optimistic approximation of the true SNR, which can still be used as a monotonic upper bound of the true SNR. In our real fMRI studies, the estimated SNR is about -5dB. Here, we illustrate the results of the synthetic data for two levels of true SNR, -6dB and -9dB, which correspond to estimated SNR of -4.3dB and -6.2dB, respectively. Fig 2b and c display the signals from a non-active voxel, a positively active voxel, and a negatively active voxel at two noise levels. As we can see, it is very difficult for humans to differentiate the three types of activations from the signals.

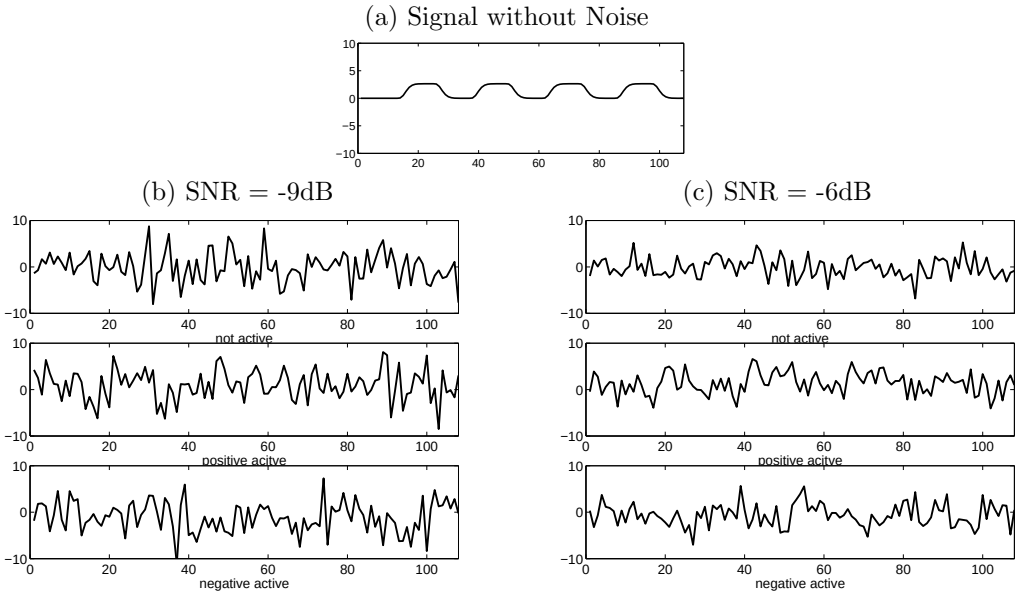

Fig. 2. *Three types of time course signals at two noise levels.*

In all experiments, we used the same GLM detector based on the two-gamma HRF introduced in Section 2.1. To create a baseline for comparison, we applied the GLM detector to the raw signals. On the other hand, in the Gaussian-smoothing-based detector, the GLM detector is applied to the spatially smoothing signals. To evaluate the EM-GRF-based detector, we performed the variational EM algorithm described in Eq. (19) to Eq. (22) on the raw images.

We adjust the definitions of false positive and true positive rate for handling three activation types. False positive rate is defined as the percentage of non-active voxels being classified as either positively or negatively active voxels. We define true positive rate as the percentage of the active voxels (including positively and negatively active voxels) being correctly classified. As we can see, misclassification between two activation types are ignored. However, in the range of false positive rates between 0.001% and 0.1%, there is no misclassification between the two activation types in our experiments. We performed the confusion matrix analysis in addition to the ROC analysis to analyze whether the detectors have different detection performance for positive and negative activations.

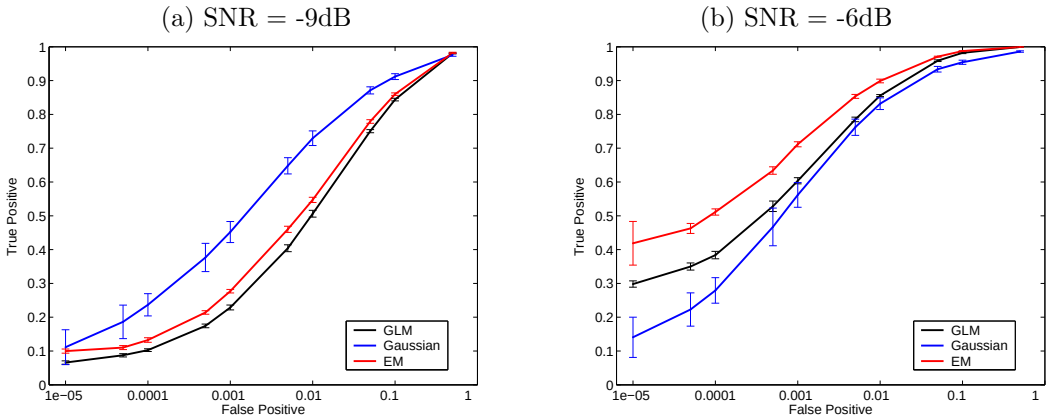

Fig. 3. *ROC curves for different smoothing techniques, at two noise levels. False positive rate is shown on the log scale.*

Fig. 3 shows the ROC curves created by the three detectors by varying the threshold applied to the corresponding statistics of each method. Since the F-statistic, provided by the GLM detector and the Gaussian-smoothing-based detector, is one-sided, we augment it with the sign of $\hat{x}_i$ to distinguish positive activations from negative activations. On the other hand, the mean of the activation statistic from the EM-GRF-based detector has a range from $(-\infty, \infty)$. Due to the large number of voxels in the volume and the relatively small number of active voxels, only very low false positive rates are of interest (we focus on the false positive rates below 0.5%, which correspond to about 7% of the total number of the active voxels, or approximately 150 voxels). The error bars indicate the standard deviation of the true detection rate over 15 different data sets, independently created and processed.

(a) GLM detector, SNR = -9 dB

Ground Truth	*Detection*		
	Negative Active	Not Active	Positive Active
Negative Active	23.46	76.54	0.00
Not Active	0.05	99.90	0.05
Positive Active	0.00	77.55	22.45

(b) GLM detector, SNR = -6 dB

Ground Truth	*Detection*		
	Negative Active	Not Active	Positive Active
Negative Active	60.04	39.96	0.00
Not Active	0.05	99.90	0.05
Positive Active	0.00	39.39	60.61

(c) Gaussian-smoothing-based detector, SNR = -9 dB

Ground Truth	*Detection*		
	Negative Active	Not Active	Positive Active
Negative Active	38.81	61.19	0.00
Not Active	0.04	99.90	0.06
Positive Active	0.00	50.85	49.15

(d) Gaussian-smoothing-based detector, SNR = -6 dB

Ground Truth	*Detection*		
	Negative Active	Not Active	Positive Active
Negative Active	49.59	50.41	0.00
Not Active	0.04	99.90	0.06
Positive Active	0.00	39.83	60.17

(e) EM-GRF-based detector, SNR = -9 dB

Ground Truth	*Detection*		
	Negative Active	Not Active	Positive Active
Negative Active	27.99	72.01	0.00
Not Active	0.05	99.90	0.05
Positive Active	0.00	72.52	27.48

(f) EM-GRF-based detector, SNR = -6 dB

Ground Truth	*Detection*		
	Negative Active	Not Active	Positive Active
Negative Active	71.44	28.56	0.00
Not Active	0.05	99.90	0.05
Positive Active	0.00	29.02	70.98

Table 1. *Detection performance of three selected detectors at two SNR levels.*

As expected, the accuracy of all methods improves with increasing SNR. At high noise levels (low SNR), Gaussian smoothing outperforms EM-GRF. As the simplest smoothing technique, Gaussian smoothing is more robust to noise. At either SNR level, we can notice that the Gaussian-smoothing-based detector has a larger standard derivation. Applying a fixed window size for smoothing causes the detector's high sensitivity to activation configurations. The EM-GRF-based detector, on the other hand, adjusts the smoothing parameters according to the detection results and the detection parameters according to the smoothing outputs. Therefore, it is more robust to different activation configurations. As the SNR increases, EM-GRF provides better regularization of the activation state (for example, at SNR=-6dB, at the false positive rate of 0.01%, the MRF outperforms the Gaussian smoothing by about 12% in true detection accuracy; at 70% true detection, the EM-GRF-based detector approximately halves the false detections compared to the Gaussian-smoothing-based detector). We believe that with the improving scanning technology, the EM-GRF-based detector will become even more helpful in reducing spurious false detection islands.

Table 1 presents the confusion matrix of average statistics of the three detectors over the same 15 data sets, when false positive rate is fixed at 0.1%. While the GLM detector and the EM-GRF-based detector have similar detection power for both positive and negative activations, the Gaussian-smoothing-based detector has higher detection power for the positive activation. As mentioned before, our synthetic data has approximately twice as many positively active voxels as negatively active voxels. The Gaussian-smoothing-based detector averages signals spatially which leads to overemphasis of the positive activation. Some weak negative signals are suppressed during the averaging process.

Besides the quantitative analysis presented above, we find it useful to visually inspect the resulting activation maps. Fig. 4 illustrates the detection results by showing one axial slice of the estimated activation map. The top image shows the phantom activation areas that were placed in the volume and used to generate the simulated fMRI scan. The bottom four rows show the same slice in the reconstructed volume at two different noise levels. All the reconstructions were performed at 0.1% false positive rate. In other words, each image in Fig. 4 shows one slice in the reconstructed volume that corresponds to a point on the ROC curve of the respective detector at 0.1% false positive rate.

The activation map produced by the basic GLM detector shows very little of the original activation at low SNR (Fig. 4b). Given this map, users would have difficulty inferring the true activation areas and disambiguating them from spurious false detections. The corresponding activation map at a high SNR (Fig. 4e) gives reasonable result but it is still fragmented. The Gaussian-smoothing-based detector (Fig. 4c) leads to an improved estimate compared with Fig. 4b. Activation regions that are square are better captured by the Gaussian-smoothing-based detector. At low SNR (-9dB), the EM-GRF model (Fig. 4d) fills in many of the active pixels that were missed by the GLM detector, but as we saw before, its results are not as accurate as those produced by the Gaussian-smoothing-based detector. At higher SNR (-6dB), EM-GRF produces a relatively accurate result (Fig. 4g). On the other hand, the Gaussian-smoothing-based detector is not able to detect activation regions that have a long, thin pattern. It again depicts the Gaussian-smoothing-based detector's high sensitivity to activation configurations.

4.2 Real fMRI Data Sets

In the real fMRI experiments, since the ground truth is unavailable, constructing ROC analysis is not possible. Instead, we visually compare the resulting activation maps. This approach effectively evaluates the ability of each detector to reconstruct the true activation areas with less evidence on the strength of the signal.

In this fMRI study [14], the original scans were obtained during an auditory "two-back" word experiment. Each experiment consisted of five rest epochs and four task epochs, each epoch 30 seconds long. In the rest condition, the subjects were instructed to concentrate on the noise of the scanner and lie still. In the task condition, the subjects were presented with a series of pre-recorded single-digit numbers, one number every three seconds. The subjects were asked to tap their index finger to the thumb when hearing

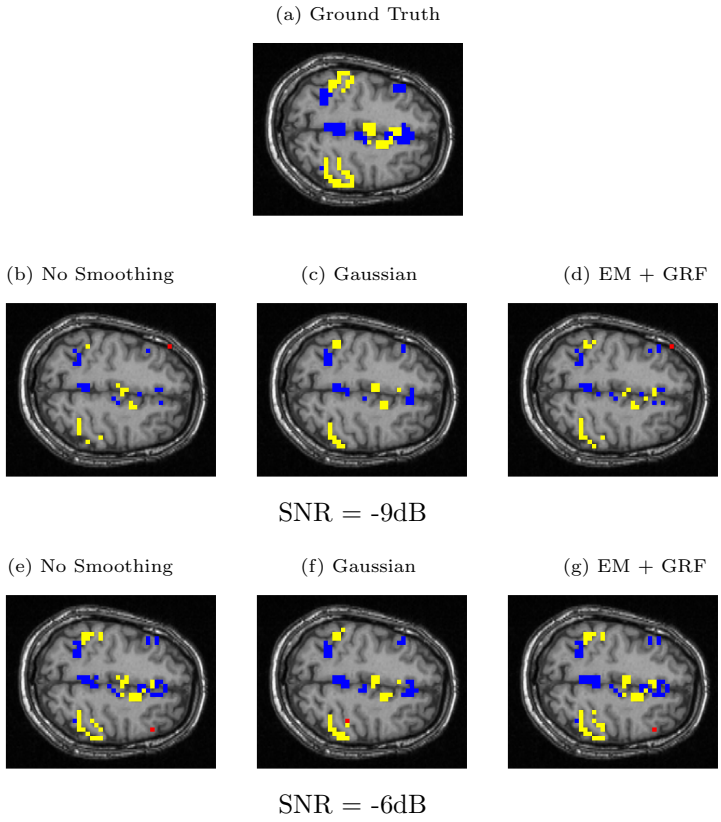

Fig. 4. *One slice from estimated activation maps for the same ground truth at 0.1% false positive rate. True detections are shown in yellow; false positive voxels are shown in red.*

a number that was the same as the one spoken two numbers before. The experiment was repeated ten times for each subject. The anatomical images were acquired on a 1.5 Tesla GE signa clinical MR scanner (T1-weighted SPGR, 256×256, 124 slices, 1.5mm slice thickness). The EPI images were acquired on the same scanner (axial, TR/TE=2500/50msec, FA90, 64×64, 24 slices, 6mm slice thickness, no gap). The original study contains nine subjects, but for the purposes of voxel-by-voxel comparison of the detectors, we present the results for one subject across all detectors. The estimated SNR when averaging over all voxels in the brain was -4.7dB (-2.3dB when averaging voxels in selected Regions of Interest (ROIs) relevant to the task).

Fig. 5a shows one axial slice in the reconstructed activation map using GLM without any spatial smoothing on the full-length fMRI signal (all 9 epochs). The ground truth for this scan is unknown, but we can use this map as a visual reference when evaluating the performance of the detectors on the time course of reduced length. For example, Fig. 5b shows the result of applying the same GLM detector to the first 5 epochs of the time course. This map is more fragmented due to loss in SNR from reducing the length of the signal. The other two images illustrate the results of applying the Gaussian-smoothing-based detector and the EM-GRF-based detector. Although Gaussian smoothing removes most of the single-voxel-activation islands, its activation map (Fig. 5c) is an overestimate compared with Fig. 5a. EM-GRF regularization (Fig. 5d) yields reasonable reconstruction results that are close to the activation map estimated from the full-length signal, but do not look overly smoothed. This highlights the potential benefit of using the EM-GRF-based detector in fMRI detection.

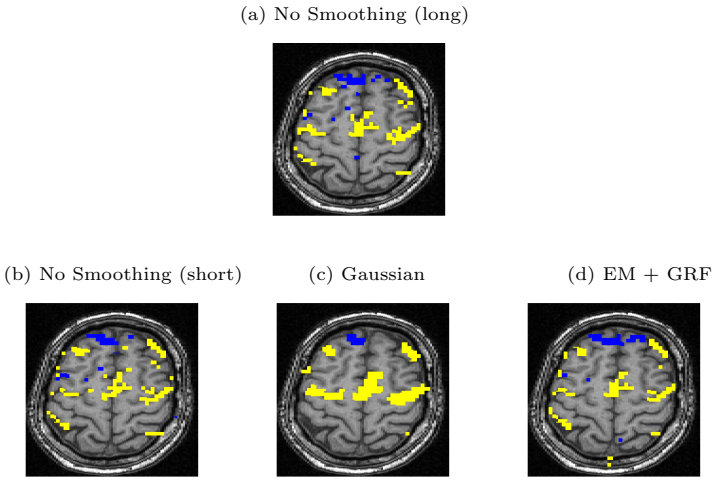

Fig. 5. *Real fMRI study. One slice in the estimated activation map. (a) No spatial smoothing, using the entire time course. (b)-(d) Estimation based on the first five epochs of the time course using different spatial smoothing methods.*

5 Discussion and Conclusions

Different from traditional fMRI detectors that represent activation strength by significance statistics of the regression coefficients, we directly used the regression coefficients to represent activation strength and activation type. Furthermore, we incorporated GRF to provide spatial regularization for the regression coefficients. Our approach is flexible to other spatial regularization methods as long as they reflect the real-number representation of activation.

Our studies confirm the importance of spatial regularization in reducing fragmentation of the activation maps. This paper investigates the improvement of the proposed EM-GRF-based detector in fMRI image analysis. Unlike most current detectors, which consider detection and spatial regularization as two separate processes, the EM-GRF-based detector iteratively adjusts the detection parameters and the spatial regularization parameters through the variational EM algorithm. This iterative detector is similar to a feedback system, and it is able to diminish adversarial effect caused by unsatisfactory performance of the first process in two-process detectors. The experiments on real fMRI data confirm this advantage of the EM-GRF-based detector over the Gaussian-smoothing-based detector in terms of its ability to recover activation from significantly shorter time courses.

Acknowledgement. We thank Dr. L.P. Panych for providing fMRI data.

References

1. Aron, A.R., *et al.* Human Midbrain Sensitivity to Cognitive Feedback and Uncertainty During Classification Learning. *Journal of Neurophysiol*, 92:1144-1152, 2004.
2. Burock, M.A., and Dale, A.M. Estimation and detection of event-related fmri signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, 11:249–260, 2000.
3. Cosman, E.R., Fisher, J. and Wells, W.M. Exact MAP activity detection in fMRI using a GLM with an spatial. *In Proc. MICCAI'04*, 2:703–710, 2004.
4. Descombes, X., Kruggel, F., and Von Cramon, D.Y. fMRI signal restoration using a spatio-temporal Markov random field preserving transitions. *NeuroImage*, 8:340–349, 1998.
5. Descombes, X., Kruggel, F. and Von Cramon, D.Y. Spatio-temporal fMRI analysis using Markov random fields. *IEEE TMI*, 17(6):1028–1039, 1998.

6. Friston, K.J., *et al.* Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
7. Hojen-Sorensen, F., Hansen, L.K. and Rasmussen, C.E. Bayesian modeling of fMRI time series. *Adv. Neuroinform. Processing Syst.*, Vol.12:754–760, 2000.
8. Jezzard, F., *et al.* Functional MIR – An Introduction to Methods. *OXFORD*, 2002.
9. Pohl, K.M., *et al.* Anatomical guided segmentation with non-stationary tissue class distributions in an expectation-maximization framework. *Proc. IEEE ISBI*, 81–84, 2004.
10. Liotti, M., *et al.* Brain responses associated with consciousness of breathlessness (air hunger). *PNAS*, 98:2035-2040, 2001
11. Rajapakse, J.C. and Piyaratna, J. Bayesian modeling of fMRI time series. *IEEE Transactions on Biomedical Engineering*, 48:1186–1194, 2001.
12. Schwartz, S. and Maquet P. Sleep imaging and the neuro-psychological assessment of dreams. *Trends in Cognitive Science*, 6:23-30, 2002.
13. Seghier, M.L., *et al.* Bayesian modeling of fMRI time series. *NeuroImage*, 21:463–472, 2004.
14. Wei, X., *et al.* Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *NeuroImage*, 21:1000–1008, 2004.
15. Woolrich, M.W., *et al.* Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE TMI*, 23(2):213–231, 2004.