fMRI Data Analysis With Nonstationary Noise Models: A Bayesian Approach

Huaien Luo and Sadasivan Puthusserypady*, *Senior Member, IEEE*

***Abstract*—The assumption of noise stationarity in the functional magnetic resonance imaging (fMRI) data analysis may lead to the loss of crucial dynamic features of the data and thus result in inaccurate activation detection. In this paper, a Bayesian approach is proposed to analyze the fMRI data with two nonstationary noise models (the time-varying variance noise model and the fractional noise model). The covariance matrices of the time-varying variance noise and the fractional noise after wavelet transform are diagonal matrices. This property is investigated under the Bayesian framework. The Bayesian estimator not only gives an accurate estimate of the weights in general linear model, but also provides posterior probability of activation in a voxel and, hence, avoids the limitations (i.e., using only hypothesis testing) in the classical methods. The performance of the proposed Bayesian methods (under the assumption of different noise models) are compared with the ordinary least squares (OLS) and the weighted least squares (WLS) methods. Results from the simulation studies validate the superiority of the proposed approach to the OLS and WLS methods considering the complex noise structures in the fMRI data.**

***Index Terms*—Bayesian estimator, functional magnetic resonance imaging (fMRI), fractional noise, general linear model (GLM), receiver operating characteristic (ROC) curve, wavelet transform.**

I. Introduction

FUNCTIONAL magnetic resonance imaging (fMRI) measures the change of the blood oxygenation level dependent (BOLD) signals due to different brain functions [1]. It is an important noninvasive technique to understand the brain functions. The fMRI data is obtained by producing a sequence of 3-D magnetic resonance (MR) images, covering the whole brain repeatedly with a repetition time (TR). In these images, the BOLD signal is dominated by strong interfering noises, resulting in low signal-to-noise ratios (SNRs) [2], [3]. The sources of these noises may be genuine random physiological noise (such as aliased cardio-respiratory pulsation), scanner-induced noise, movements of the subjects (such as uncorrected head movements, lower jaw movements), and so on. The analysis methods also may induce residual noise due to imperfect models.

The objective of the fMRI data analysis is to detect the weak BOLD signal from the noisy data and determine the regions of the brain which are activated by external stimuli. However, the noisy nature of the fMRI data makes the detection of the BOLD effect difficult and challenging. Many methods have been proposed to analyze the fMRI data, such as the general linear model (GLM) [4] implemented in the statistical parametric mapping (SPM) [5], the correlation method [6], the blind source separation (BSS) [7] method, the clustering analysis method [8], and so on. In these methods, the noise is generally assumed to be stationary, that is, the underlying noise model is time-invariant. However, since the fMRI noises are inherently time-varying, the stationary assumption may not be appropriate considering the complexity of the data. In [9], a spatially nonstationary but temporally stationary spatio-temporal noise model was developed to fit the fMRI time series spatiotemporally. However, the nonstationarities also exist temporally. Several factors may induce the nonstationarities in the fMRI time series. Neurophysiological processes, such as the change of the number of neurons involved in a specific activity at different time points and the extraneous auditory and visual stimuli or background memory processes, may cause the variance of the noise to change [10]. Besides, the abrupt movements from the subjects are another source of the nonstationarity of the noise in fMRI [11]. For example, the abrupt movements of the lower jaw of subjects may affect only a few consecutive fMRI images, causing the variance of these images to be high and violating the stationary assumption of the noise. Moreover, the $1/f$-like noise or nonstationary fractional noises in the fMRI data are reported and investigated by many researchers recently [12]–[14]. These factors show that stationarity assumption of the noise may not be realistic to cope with the complex fMRI data. The general nonstationary noise model to incorporate the time-varying stochastic properties of fMRI noise (such as transients and sudden changes in the fMRI data) is needed for a more accurate fMRI data analysis.

In this paper, two noise models (time-varying variance model [11] and fractional noise model [13]) are investigated to capture the nonstationarities in the fMRI data. Different from the assumption (spatially nonstationary but temporally stationary) used in [9], these two noise models are temporally nonstationary. The covariance matrices of the time-varying variance noise and the fractional noise after wavelet transform are diagonal matrices. This property is investigated under the Bayesian framework. The Bayesian methods are utilized both in single-subject (first-level) analysis [15], [16] and group

Manuscript received December 14, 2005; revised December 20, 2006. This work was supported by the Faculty Research Grant R-263-000-360-112, National University of Singapore, Singapore. *Asterisk indicates corresponding author.*

H. Luo is with the Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576 (e-mail: g0305766@nus.edu.sg).

*S. Puthusserypady is with the Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576 (e-mail: elespk@nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2007.902591

studies (multilevel experiments) [17], [18]. In [15], a variational Bayesian method was proposed to infer on the GLM for fMRI data. The haemodynamic response basis functions in the GLM using variational Bayesian are then constrained to sensible HRF shape in [16]. In [17] and [18], the Bayesian approach was utilized to deal with the inference problem on the hierarchical linear models for fMRI group analysis. These works show the advantages of the Bayesian analysis for fMRI data. The classical approaches, such as SPM, working under the mechanism of rejecting or accepting null hypothesis, have some limitations. For example, the p values in the SPM are the probabilities of the effects under null hypothesis which states there is no activation in each voxel. These p values do not reflect the probabilities of the actual effects when the voxel is truly activated. Bayesian approaches, on the contrary, can give the probability that a voxel is activated or the probability that the effect is greater than some threshold value. Hence, these limitations of the classical approaches could be overcome using the Bayesian methods.

This paper is organized as follows. Section II briefly describes the most commonly used GLM in fMRI data analysis. Two nonstationary noise models are discussed in Section III. One is the time-varying variance noise model and the other is the fractional noise model. The orthonormal wavelet decomposition as an approximate Karhunen–Loéve transform (KLT) to the fractional noise process is also discussed in this section. The Bayesian estimator of the proposed linear regression models is presented in Section IV. In Section V, the experimental examples and results are discussed in detail, followed by some concluding remarks in Section VI.

II. General Linear Model

GLM is one of the most widely used methods to analyze the fMRI data along with the well-known software—SPM [5]. This method assumes a linear model to the time series of each voxel. Suppose there are N voxels and the data is measured at acquisition time $t = 1, 2, \ldots, T$. Let $\mathbf{y}_n$ denote the vector (of dimension $T \times 1$) of observed/measured intensity changes in the nth voxel of the fMRI dataset. GLM assumes that the measured time series at each voxel is the weighted linear combination of different regressors in the design matrix. This can be mathematically represented as

$$\mathbf{y}_n = \mathbf{\Phi}\mathbf{w}_n + \boldsymbol{\epsilon}_n \tag{1}$$

where $\mathbf{w}_n$ is a vector (of dimension $M \times 1$) of weighting parameters which could be found by the least square methods; $\boldsymbol{\epsilon}_n$ is the $T \times 1$ dimension vector of noise terms in the nth voxel; the design matrix $\mathbf{\Phi}$ has the dimension $T \times M$ with each row corresponding to one time point (scan) of the regressors, and the columns correspond to the different explanatory variables or regressors in the model.

If $\boldsymbol{\epsilon}_n$ in (1) is assumed as white noise with the corresponding covariance matrix, $Cov(\boldsymbol{\epsilon}_n) = \mathbf{I}$ (where $\mathbf{I}$ is an identity matrix of dimension $T \times T$), and if the design matrix $\mathbf{\Phi}$ is of full column rank (and, consequently, $\mathbf{\Phi}^T\mathbf{\Phi}$ invertible), the least squares solution of the parameter $\mathbf{w}_n$ is the ordinary least squares (OLS) estimator, as shown in (2)

$$\hat{\mathbf{w}}_{n(\mathrm{OLS})} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{y}_n. \tag{2}$$

Clearly, the independent and identically distributed (i.i.d.) white noise assumption in OLS is inappropriate considering the temporal auto-correlations and the nonstationary nature of the fMRI signals. To deal with the temporal auto-correlations in the fMRI data, coloring and prewhitening methods are proposed [19], [20]. Many nonstationary noise models [9], [10], [13] are also proposed to model the nonstationary properties of the fMRI data. Among these proposed noise models, two nonstationary noises models are considered in this paper under the Bayesian framework. The first model (time-varying variance model) examines the variance variations in different time points or scans. The second model (fractional noise model) investigates the nonstationary fractional noise in the fMRI data. The covariance matrices of the time-varying variance noise model and the fractional noise model (after wavelet transform) are similar diagonal matrices and can be considered with the same estimation algorithms.

III. Noise Models

A. Time-Varying Variance Model

The abrupt physical movements of the subjects in the fMRI experiment, such as the lower jaw movements, may affect a few fMRI images, causing the variances of these fMRI images to rise and breaking the stationary assumption. In [11], the time-varying variance noise model is introduced and it is reported that such increase in noise variance is multiplicative. That is, the noise interferences affect the variances in the fMRI images multiplicatively, but not additively. Hence, the variance at each time point is modeled as a scaled version of the overall variance in that voxel. This time-dependent noise is modeled as a Gaussian process $\boldsymbol{\epsilon}_n \sim \mathcal{N}(0, \mathbf{B}_n^{-1})$ with the precision matrix $\mathbf{B}_n$ (i.e, the inverse of the covariance matrix) defined as

$$\mathbf{B}_n = \begin{pmatrix} s_1 & 0 & \ldots & 0 \\ 0 & s_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & s_T \end{pmatrix} \beta_n = \mathbf{S}\beta_n \tag{3}$$

where $s_1, s_2, \ldots, s_T$ are the scaling parameters, $\mathbf{S}$ is a $T \times T$ scaling matrix and β_n is a scalar representing the overall noise precision in the nth voxel. This noise precision matrix shows that the precisions (or inverse of variance) of the nth voxel at different time points are scaled versions of the overall precision β_n observed in that voxel. In [11], it is testified that the scaling change of the precision has a spatial uniformity. This shows that if the noises of the voxels in some portion of the image changes, all other voxels in the same image will also change to a similar degree. Thus, although the overall variance is different in different voxels (i.e., different β_n in different voxels), the scaling parameters are assumed to be the same across the whole

image (i.e., the scaling matrix $\mathbf{S}$ is assumed to be the same for all the voxels).

Under the assumption that the precision matrix of noise $\boldsymbol{\epsilon}_n$ is $\mathbf{B}_n = \mathbf{S}\beta_n$, the maximum likelihood estimate of $\mathbf{w}_n$ in (1) is a weighted least squares (WLS) estimate as follows:

$$\hat{\mathbf{w}}_{n(\mathrm{WLS})} = (\boldsymbol{\Phi}^T\mathbf{S}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{S}\mathbf{y}_n. \tag{4}$$

However, the WLS estimate requires us to have an accurate estimate of the scaling matrix $\mathbf{S}$.

Traditional method to estimate the scaling matrix $\mathbf{S}$ makes use of the residuals $\mathbf{r}_n$ of the OLS estimates, where the residuals is defined as

$$\mathbf{r}_n = \mathbf{y}_n - \boldsymbol{\Phi}\hat{\mathbf{w}}_{n(\mathrm{OLS})} = \mathbf{y}_n - \boldsymbol{\Phi}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{y}_n. \tag{5}$$

The overall precision β_n is estimated by [21]

$$\hat{\beta}_n = \frac{T - \mathrm{rank}(\boldsymbol{\Phi})}{\mathbf{r}_n^T\mathbf{r}_n} \tag{6}$$

where $\mathbf{r}_n^T\mathbf{r}_n$ is the sum of the squares of the residuals and $(T - \mathrm{rank}(\boldsymbol{\Phi}))$ is the appropriate degrees of freedom.

Considering that $\mathbf{r}_n\mathbf{r}_n^T$ is the estimate of the covariance matrix of the noise [i.e., $\mathbf{B}_n^{-1}$ as shown in (3)], the inverse of the scaling parameters $\mathbf{s}_{inv} = \mathrm{diag}(\mathbf{S}^{-1}) = [s_1^{-1}, s_2^{-1}, \ldots, s_T^{-1}]^T$ is estimated by averaging these variance estimates weighted by β_n over the N voxels

$$\hat{\mathbf{s}}_{inv} = \frac{\sum_{n=1}^{N}\mathrm{diag}\left(\beta_n\mathbf{r}_n\mathbf{r}_n^T\right)}{N} \tag{7}$$

where the operator $\mathrm{diag}(\bullet)$ transforms the diagonals of a square matrix into a column vector. N is the total number of voxels considered due to the spatial uniformity of the scaling parameters.

Although this classical method is simple to implement, it is a biased estimate of the scaling parameters $\mathbf{s}$ as shown in [11] that the precision matrix of $\mathbf{r}_n$ is not equal to $\mathbf{S}\beta_n$. This bias comes from estimation of $\mathbf{s}$ using the residuals of the OLS estimates which assume the covariance matrix of the residuals to be an identity matrix. This bias may cause the t-test value in this voxel invalid and, hence, incorrect inference about the activation status of this voxel if the noise covariance matrix is not an identity matrix. In Section IV, a Bayesian estimator to accurately estimate the scaling parameters $\mathbf{s}$ is proposed.

B. Fractional Noise Model

It is reported that the noise in the fMRI time series obtained under the resting or null conditions exhibits long-range autocorrelation in time and $1/f$-like spectral properties [22]. This means that the spectral density $S(|f|) \propto |f|^{\lambda}$ with the spectral exponent $\lambda < 0$. One of the $1/f$-like processes is the fractional Brownian motion (fBm). It is a zero mean, nonstationary, and nondifferentiable process with the auto-covariance (r) between time t_1 and t_2 defined by the Hurst exponent (H) [23]

$$r(t_1, t_2) = \frac{1}{2}\sigma^2\left(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H}\right) \tag{8}$$

where

$$\sigma^2 = \Gamma(1 - 2H)\frac{\cos(\pi H)}{\pi H}. \tag{9}$$

Let the length of the fMRI time series be $T = 2^p$, where p is an integer. Applying the discrete wavelet transform (DWT) to both sides of (1), we get a GLM in the wavelet domain [26] as follows:

$$\mathbf{y}_n^W = \boldsymbol{\Phi}^W\mathbf{w}_n + \boldsymbol{\epsilon}_n^W \tag{10}$$

where $\mathbf{y}_n^W$ and $\boldsymbol{\epsilon}_n^W$ are the results of applying the DWT respectively to the data $\mathbf{y}_n$ and noise $\boldsymbol{\epsilon}_n$ up to the maximum scale p, and $\boldsymbol{\Phi}^W$ is the wavelet transform applying to each columns of the design matrix $\boldsymbol{\Phi}$.

The wavelet transform of noise $\boldsymbol{\epsilon}_n$ up to the maximum scale p is denoted as

$$\boldsymbol{\epsilon}_n^W = [a_{p,1}, d_{p,1}, d_{p-1,1}, d_{p-1,2}, \ldots, d_{1,1}, \ldots, d_{1,2^{p-1}}]^T \tag{11}$$

where $a_{p,1}$ is the scaling coefficient (or approximation coefficient) at level p, and $d_{m,k}$ $(k = 1, \ldots, 2^{p-m})$ is the wavelet coefficients (or detail coefficients) at level $m(m = 1, \ldots, p)$.

The orthonormal wavelet decomposition is an approximate whitening filter for fBm [24]. The correlation between the wavelet coefficients within any of the scales and the cross-correlation coefficients between different scales are very small [25] and, hence, can be ignored for any wavelet provided that the number of vanishing moments is sufficiently large (greater than $2H + 1$). It is also shown in [25], [27], and [28] that, for $1/f$-like noises, the wavelet coefficients $d_{m,k}$ at level m and the scale coefficient $a_{p,1}$ at level p are normally distributed with zero mean and denoted as

$$d_{m,k} \sim \mathcal{N}(0, V_{d_m}) \qquad m = 1, \ldots, p;\ \ k = 1, \ldots, 2^{p-m} \tag{12}$$

$$a_{p,1} \sim \mathcal{N}\left(0, V_{a_p}\right) \tag{13}$$

where $V_{d_m} = \mathrm{var}\{d_{m,k}\}$ $(k = 1, \ldots, 2^{p-m})$ is the variance of the wavelet coefficients $d_{m,k}$ at level $m(m = 1, \ldots, p)$ and $V_{a_p} = \mathrm{var}\{a_{p,1}\}$ is the variance of the scaling coefficients $a_{p,1}$ at the level p.

Thus, for $1/f$-like noises, the orthonormal wavelet decomposition behaves like a KLT and the transformed fractional noise $\boldsymbol{\epsilon}_n^W$ could be modeled by Gaussian process $\boldsymbol{\epsilon}_n^W \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^W)$ with the $T \times T$ diagonal covariance matrix defined as (14), shown at bottom of the next page.

The diagonal matrix $\boldsymbol{\Lambda}^W$ is composed of V_{a_p} (the variance of scaling coefficients at level p) and p submatrices representing the covariance matrix of the wavelet coefficients $d_{m,k}$ at level $m(m = p, p-1, \ldots, 1)$. From (12), it is seen that the wavelet coefficients $d_{m,k}$ at level m are with the same variance V_{d_m}. Hence, the submatrix at level m is of identical value V_{d_m} at the diagonal and of dimension $2^{p-m} \times 2^{p-m}$.

Since the wavelet function at different scales are bandpass filters and scaling function is a low-pass filter, it is verified in

[28] that the variance of the wavelet coefficients at level m and the scaling coefficient at level p can be denoted as scaled version of σ^2

$$\begin{aligned} V_{d_m} &= \sigma^2 S_{d_m}; \quad m = 1, 2, \ldots, p \\ V_{a_p} &= \sigma^2 S_{a_p} \end{aligned} \tag{15}$$

where S_{d_m} $(m = 1, 2, \ldots, p)$ and S_{a_p} are scaling constants determined by H [26]. Hence, the diagonal elements of covariance matrix $\mathbf{\Lambda}^W$ are scaled versions of σ^2 and can be represented by

$$\mathbf{\Lambda}^W = \begin{pmatrix} S_{a_p} & & & & & & & \\ & S_{d_p} & & & & & & \\ & & S_{d_{p-1}} & & & & & \\ & & & S_{d_{p-1}} & & & & \\ & & & & \ddots & & & \\ & & & & & S_{d_1} & & \\ & & & & & & \ddots & \\ & & & & & & & S_{d_1} \end{pmatrix} \sigma^2 \tag{16}$$

$$= \operatorname{diag}^{-1}\left\{\left(S_{a_p}, S_{d_p}, S_{d_{p-1}}, S_{d_{p-1}}, \ldots S_{d_1}, \ldots, S_{d_1}\right)\sigma^2\right\} \tag{17}$$

where $\operatorname{diag}^{-1}\{\bullet\}$ transforms a vector into diagonals of a square matrix.

From the above descriptions, it is seen that the covariance matrix of the fractional noise in the wavelet domain (17) has almost the same structure as the covariance matrix of the time-varying variance noise in the time domain [see (3)]. The only difference is that for wavelet-transformed fractional noise, the scaling parameters at the same level are of the same values. Since the covariance matrices have similar structure in both cases, we could put these two cases under the same framework, and, in the next section, we do not distinguish between these two cases unless it is necessary.

IV. Bayesian Estimator

The fMRI data analysis aims at a sensitive and robust detection of activation regions of the brain under different sensory and cognitive experiments. This requires an accurate estimate of the weights $\mathbf{w}_n$ in GLM. For the nonstationary noises presented in the previous section, the WLS and OLS estimators may fail because: i) the WLS estimator requires an accurate estimate of the covariance matrix which is difficult and ii) the OLS estimator does not consider the nonstationary properties of the noise. In this section, this problem is investigated under the Bayesian framework to obtain a robust estimator of the weights in GLM. For simplicity of notation, we drop the subscript n in (1) and represent the GLM model as

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\epsilon} \tag{18}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_k, \ldots, y_T]^T$ is the noise vector, and $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_k, \ldots, \epsilon_T]^T$ is the error vector and $\mathbf{\Phi}$, $\mathbf{w}$ are the same as introduced in Section II. In [29], a Bayesian method to estimate the parameter $\mathbf{w}$ under the white Gaussian noise with fixed variance is provided. Here, we extend the Bayesian method to the situation where the variance of the noise is changing with the time.

The noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{B}^{-1})$ is assumed as a Gaussian noise with zero mean and diagonal precision matrix $\mathbf{B} = \operatorname{diag}^{-1}\{(s_1, s_2, \ldots, s_T)\beta\}$ as introduced in Section III (for both time-varying variance noise and fractional noise after wavelet transform). At time k $(k = 1, 2, \ldots, T)$, the noise ϵ_k are independent samples from zero-mean Gaussian with precision $s_k\beta$ which may be different at each time point k. Thus, the output signal y_k at time point k is also a Gaussian distributed process with varying precision at each time point. With the assumption of independence of the output at different time points, the likelihood of the whole time series $\mathbf{y}$ can be derived as a multivariate Gaussian density function

$$\begin{aligned} p(\mathbf{y}|\mathbf{w}, \mathbf{B}) &= \frac{|\mathbf{B}|^{T/2}}{(2\pi)^{T/2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{\Phi}\mathbf{w})^T \mathbf{B}(\mathbf{y} - \mathbf{\Phi}\mathbf{w})\right\} \\ &\sim \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{\Phi}\mathbf{w})^T \mathbf{B}(\mathbf{y} - \mathbf{\Phi}\mathbf{w})\right\}. \end{aligned} \tag{19}$$

We also treat the weight vector $\mathbf{w}$ as random and assume its prior distribution is Gaussian with zero mean and precision matrix $\mathbf{A}$ [29]

$$p(\mathbf{w}|\mathbf{A}) = \mathcal{N}(0, \mathbf{A}^{-1}) \sim \exp\left\{-\frac{1}{2}\mathbf{w}^T \mathbf{A}\mathbf{w}\right\} \tag{20}$$

$$\begin{aligned} \mathbf{\Lambda}^W &= Cov\left(\boldsymbol{\epsilon}_n^W\right) \\ &= \begin{pmatrix} V_{a_p} & & & & & \\ & V_{d_p} & & & & \\ & & \begin{pmatrix} V_{d_{p-1}} & \\ & V_{d_{p-1}} \end{pmatrix} & & & \\ & & & \ddots & & \\ & & & & \begin{pmatrix} V_{d_1} & \ddots \\ \ddots & V_{d_1} \end{pmatrix} \end{pmatrix} \end{aligned} \tag{14}$$

where $\mathbf{A}$ is a diagonal matrix formed by M hyperparameters as follows:

$$\mathbf{A} = \mathrm{diag}^{-1}(\alpha_1, \alpha_2, \ldots, \alpha_M). \tag{21}$$

This prior means, at this time, the best guess about the value of the ith weight w_i is 0, and α_i^{-1} represents the uncertainty about this guess. Moreover, we specify the hyperpriors over $\mathbf{A}$ and $\mathbf{B}$ to be uniform.

The basic idea of Bayesian estimator is to maximize the posterior probability over the weights $\mathbf{w}$ and the hyperparameters $\mathbf{A}$ and $\mathbf{B}$ given the data $\mathbf{y}$, i.e., maximizing $p(\mathbf{w}, \mathbf{A}, \mathbf{B}|\mathbf{y})$ [29], [30]. This posterior is further decomposed as

$$p(\mathbf{w}, \mathbf{A}, \mathbf{B}|\mathbf{y}) = p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B}) p(\mathbf{A}, \mathbf{B}|\mathbf{y}). \tag{22}$$

Maximizing the left hand side of (22) is equivalent to maximizing the two probabilities on the right hand side, that is, maximizing $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ and $p(\mathbf{A}, \mathbf{B}|\mathbf{y})$.

To maximize the first probability of the right hand side of (22), we rewrite $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ according to Bayes rule

$$\begin{aligned} p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B}) &= \frac{p(\mathbf{w}, \mathbf{y}|\mathbf{A}, \mathbf{B})}{p(\mathbf{y}|\mathbf{A}, \mathbf{B})} \\ &= \frac{p(\mathbf{y}|\mathbf{w}, \mathbf{B}) p(\mathbf{w}|\mathbf{A})}{p(\mathbf{y}|\mathbf{A}, \mathbf{B})}. \end{aligned} \tag{23}$$

Rather than evaluating (23) directly (this needs to evaluate the term $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ which is not easy), we can rearrange (23) and get the probabilities $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ and $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ simultaneously as shown below [using (19) and (20)]

$$\begin{aligned} &p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B}) p(\mathbf{y}|\mathbf{A}, \mathbf{B}) \\ &\quad = p(\mathbf{y}|\mathbf{w}, \mathbf{B}) p(\mathbf{w}|\mathbf{A}) \\ &\quad \sim \exp\left\{ -\frac{1}{2}(\mathbf{y} - \mathbf{\Phi}\mathbf{w})^T \mathbf{B} (\mathbf{y} - \mathbf{\Phi}\mathbf{w}) - \frac{1}{2}\mathbf{w}^T \mathbf{A} \mathbf{w} \right\} \\ &\quad \sim \exp\left\{ -\frac{1}{2}(\mathbf{w} - \mathbf{u})^T \mathbf{\Lambda}^{-1} (\mathbf{w} - \mathbf{u}) \right. \\ &\qquad \left. -\frac{1}{2}\mathbf{y}^T (\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1} \mathbf{y} \right\} \end{aligned} \tag{24}$$

where

$$\mathbf{\Lambda}^{-1} = (\mathbf{A} + \mathbf{\Phi}^T \mathbf{B} \mathbf{\Phi}) \tag{25}$$

and

$$\mathbf{u} = \mathbf{\Lambda} \mathbf{\Phi}^T \mathbf{B} \mathbf{y}. \tag{26}$$

By rearranging (24) into two multivariate Gaussian functions, the following equations are obtained:

$$\begin{aligned} p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B}) &= (2\pi)^{-\frac{M}{2}} |\mathbf{\Lambda}|^{-\frac{1}{2}} \\ &\quad \times \exp\left\{ -\frac{1}{2}(\mathbf{w} - \mathbf{u})^T \mathbf{\Lambda}^{-1} (\mathbf{w} - \mathbf{u}) \right\} \end{aligned} \tag{27}$$

and

$$\begin{aligned} p(\mathbf{y}|\mathbf{A}, \mathbf{B}) &= (2\pi)^{-\frac{T}{2}} |\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T|^{-\frac{1}{2}} \\ &\quad \times \exp\left\{ -\frac{1}{2}\mathbf{y}^T (\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1} \mathbf{y} \right\}. \end{aligned} \tag{28}$$

The maximum of the first probability $p(\mathbf{w}|\mathbf{y}, \mathbf{A}, \mathbf{B})$ at the right hand side of (22) is clearly at the mean, i.e., the best estimate of the weights $\mathbf{w}$ is

$$\hat{\mathbf{w}} = \mathbf{u}. \tag{29}$$

Maximizing second probability of the right hand side of (22) is further decomposed as maximization of $p(\mathbf{A}, \mathbf{B}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{A}, \mathbf{B}) p(\mathbf{A}) p(\mathbf{B})$ with respect to $\mathbf{A}$ and $\mathbf{B}$. For uniform hyperprior distributions, it is equivalent to maximizing $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ with respect to hyperparameters α_i; $i = 1, \ldots, M$ in the matrix $\mathbf{A}$, the scaling parameters s_i; $i = 1, 2, \ldots, T$ and overall precision β in matrix $\mathbf{B}$.

It is convenient if we maximize the logarithm of this quantity $p(\mathbf{y}|\mathbf{A}, \mathbf{B})$ and accordingly the objective function becomes $\mathcal{L} = \log(p(\mathbf{y}|\mathbf{A}, \mathbf{B}))$. Without considering the constant term, the objective function is

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2}\log|\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T| \\ &\quad -\frac{1}{2}\mathbf{y}^T (\mathbf{B}^{-1} + \mathbf{\Phi}\mathbf{A}^{-1}\mathbf{\Phi}^T)^{-1} \mathbf{y} \\ &= -\frac{1}{2}\left[-\log|\mathbf{\Lambda}| - \log|\mathbf{B}| - \log|\mathbf{A}| \right. \\ &\qquad \left. +(\mathbf{y} - \mathbf{\Phi}\mathbf{u})^T \mathbf{B} (\mathbf{y} - \mathbf{\Phi}\mathbf{u}) + \mathbf{u}^T \mathbf{A} \mathbf{u} \right]. \end{aligned} \tag{30}$$

Maximizing the objective function $\mathcal{L}$ with respect to α_i, s_i and β, the following update equations are derived:

$$\alpha_i = \frac{1}{\mathbf{\Lambda}_{ii} + \hat{w}_i^2} \tag{31}$$

$$s_i = \frac{1}{\mathrm{trace}\left(\mathbf{\Lambda} \beta \boldsymbol{\phi}_i^T \boldsymbol{\phi}_i \right) + \beta (\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}})_i^2} \tag{32}$$

$$\beta = \frac{T}{\mathrm{trace}(\mathbf{\Lambda}\mathbf{\Phi}^T \mathbf{S} \mathbf{\Phi}) + (\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}})^T \mathbf{S} (\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}})} \tag{33}$$

where $\mathbf{\Lambda}_{ii}$ is the ith diagonal element of the posterior covariance matrix $\mathbf{\Lambda}$ in (25), $\hat{w}_i$ is the ith element of the posterior mean vector $\hat{\mathbf{w}}$ in (29), $\boldsymbol{\phi}_i$ is the ith row vector of the design matrix $\mathbf{\Phi}$ and $(\mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}})_i$ is the ith element of the estimated error $\mathbf{r}_b = \mathbf{y} - \mathbf{\Phi}\hat{\mathbf{w}}$.

By iterative updating of (31) to (33), together with $\mathbf{\Lambda}$ and $\hat{\mathbf{w}}$ from (25) and (29), this update algorithm converges to the optimum solution. For the time-varying variance noise model, different voxels are assumed to have same variance scaling parameters s_i, and, hence, at the end of each updating cycle, s_i could be averaged over all the relevant voxels. For the fractional noise model, after wavelet transform, the variance scaling parameters of the transformed noise in the wavelet domain are testified to have the same value at the same decomposition level as shown

in (14) and (15); during the updating process, the scaling parameters s_i at the same level are averaged to obtain an accurate estimation.

In practice, some of the α_is in (21) will approach infinity, which means the w_i should be zero given the data at hand. Thus, the corresponding regressor functions could be "pruned" and the remaining regressors are kept to construct a suitable design matrix. This is a flexible design matrix determination method in GLM for fMRI data analysis as introduced in [31].

Having estimated the posterior probability density function of the weights $\mathbf{w}$ at each voxel, a map of the activation regions in the brain could be obtained by computing the posterior probability that a voxel is activated or the probability that an effect is greater than some threshold value. Given the effect size γ, the posterior probability is [32]

$$1 - \Psi\left(\frac{\gamma - \mathbf{c}^T\mathbf{u}}{\sqrt{\mathbf{c}^T\mathbf{\Lambda}\mathbf{c}}}\right) \tag{34}$$

where $\mathbf{c}$ is the contrast vector, $\Psi(\cdot)$ is the normal cumulative distribution function (cdf), $\mathbf{u}$ and $\mathbf{\Lambda}$ are defined in (26) and (25), respectively. These posterior probabilities will form posterior probability maps (PPMs) of the fMRI activation detection.

In the classical approach, the point estimate of the weights $\mathbf{w}$ by OLS and WLS methods is used to calculate a t-statistics [2] at each voxel

$$t = \frac{\mathbf{c}^T\hat{\mathbf{w}}}{\sqrt{\mathbf{c}^T\mathbf{\Lambda}_{\hat{\mathbf{w}}}\mathbf{c}}} \tag{35}$$

where $\hat{\mathbf{w}}$ is the least-square estimate of the parameter $\mathbf{w}$ and $\mathbf{\Lambda}_{\hat{\mathbf{w}}}$ is the covariance matrix of the estimate $\hat{\mathbf{w}}$. The statistic t is of Student's t distribution with the effective degree of freedom (df) as derived in [4].

These t-statistic values obtained at each voxel are then mapped to form SPMs for further determination of the activated regions in the brain.

V. Results

In this section, simulation studies are carried out to compare the performance of the proposed approach to the OLS and the WLS estimators introduced in the previous sections. Both simulated as well as real data are examined. We first compare the accuracy of the estimated weights in GLM. The activation detection ability of these methods are then investigated. Last, results from the real fMRI data are given and discussed.

A. Simulated Data

1) Time-Varying Variance Noise: We first show that the proposed Bayesian approach could give more efficient (lower variance) estimate of the weights $\mathbf{w}$ than the OLS and the WLS estimators. A simulated block experimental fMRI design is investigated in this study. The BOLD signal is represented by a square waveform. When the stimulus is applied (ON), the representing waveform has the value 1; while when the stimulus is absent (OFF), the representing waveform has the value 0 [33]. An illustration of a block experimental design and its square waveform representation is shown in Fig. 1.

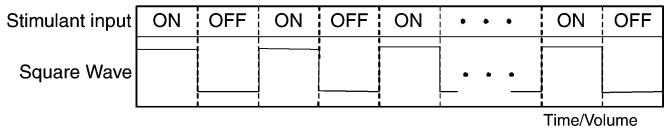

Fig. 1. Illustration of a block experimental design and its square waveform (boxcar function) representation.

TABLE I
Standard Deviation (SD) of Estimated $\hat{w}$ on Simulated Data With Different Weight and Noise

		OLS	WLS	Bayesian
i.i.d. noise	$w = 0$	0.3564	0.3579	0.1673
	$w = 0.5$	0.3558	0.3577	0.3318
	$w = 1$	0.3527	0.3545	0.3472
	$w = 2$	0.3527	0.3546	0.3065
time-varying variance noise	$w = 0$	0.4086	0.1153	0.0321
	$w = 0.5$	0.4100	0.1162	0.0583
	$w = 1$	0.4101	0.1170	0.0578
	$w = 2$	0.4084	0.1159	0.0570

Two hundred realizations of 128-point time series [i.e., $128(T)$ images with 200 (N) voxels in fMRI data sets] under different weighting of the square waveform are generated. Both the i.i.d. noise and time-varying variance noise are then added to the data set to simulate the noisy fMRI data. These data are fitted with a GLM model with a design matrix $\mathbf{\Phi}$ (dimension $128\times$ 2) composed of two regressors: the constant value 1 to model the mean grey level in fMRI voxels and a square waveform to model the experimental design. Since we are only concerned about the weight parameter related to the square waveform regressor, only the standard deviation of this weight w is reported. Table I shows the standard deviation of the estimate $\hat{w}$ under different values of w (in this paper, only the results for $w = 0$, 0.5, 1, 2 are reported) and different noise properties.

From Table I, it is clear that the proposed Bayesian method always results in lower standard deviation (SD) estimate of the weight w for all the values of w tested no matter whether the noise is i.i.d. or time-varying variance. This implies that the Bayesian estimator is more accurate than the OLS and WLS estimators. This improvement of our approach comes from the fact that the Bayesian estimator not only captures the true variance structure of the noise better than OLS and WLS, but also utilizes a flexible design matrix which is more suitable for the actual nonstationary data. Accurate regression weights are desired since these regression weights are fitted with a hierachical linear model for second-level analysis or random effect analysis [33]. Lower SD estimates will give more sensitive results for these higher level analysis.

Next, the detection ability of these estimators are compared. A slice from a real fMRI data set is used as the background image. Time-varying variance noises with same scaling parameters but different overall variances are then added to each voxel of the data to form a 3-D fMRI time series. At some specific regions of the image, simulated BOLD signals are added. The BOLD signal is simulated by assuming the brain and MR-scanner as a linear system. Thus, the BOLD responses are the convolution of the experimental paradigm and the

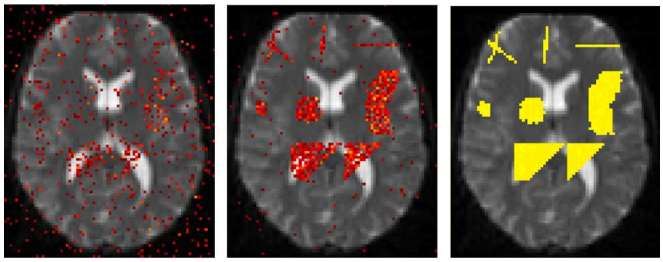

Fig. 2. Detection results of simulated fMRI data using different methods: (a) OLS method with thresholded SPM ($t > 1.7, p < 0.05$); (b) WLS method with thresholded SPM ($t > 1.7, p < 0.05$); (c) Bayesian method with PPM ($P(\text{effect} > 0.4) > 0.9$).

haemodynamic response function (HRF) [33]. For the block fMRI experiment, the experimental paradigm is simulated by a square waveform (or boxcar function) as introduced previously. The HRF is chosen as the difference between two gamma functions [34]. These simulated fMRI data is fitted by the GLM model with the design matrix composed of two regressors: simulated BOLD response and a constant vector of value 1. For the Bayesian method, the PPM is obtained as introduced in Section IV. For the OLS and WLS methods, SPMs can be obtained. Fig. 2 shows the detection results of simulated fMRI data using these methods.

From Fig. 2, it is clear that the proposed Bayesian method is more robust and sensitive compared to the OLS and WLS methods with more true activations and less false activations detected. This is because the noise structure estimated by the proposed method is closer to the true noise. To have a clearer comparison of the detection ability, the receiver operator characteristic (ROC) analysis [35] is used to investigate the activation maps of Bayesian, OLS and WLS methods. The ROC curve is a plot of true positive ratio (TPR) versus false positive ratio (FPR) under different threshold values. The method that can detect most of the real activations while minimizing the detection of false activations is more desirable. Fig. 3(a) and (b) shows the ROC curves under the i.i.d. noise and time-varying variance noise, respectively. For i.i.d. noise, the three methods have comparable performances which is shown in Fig. 3(a), while for the time-varying variance noise, the ROC curves in Fig. 3(b) indicate that, under the same FPR, the proposed Bayesian method could actually detect more real activations. This clearly shows the superior performance of the proposed method compared to the OLS and WLS methods.

2) Fractional Noise: To investigate the effect of fractional noise or $1/f$-like noise, we first synthesize realizations of fBm noise based on the statistical model of fBm using the method introduced in [12]. The total length of the simulated fBm noise is $T = 128$ with the decomposition level $p = 7$ (since $T = 2^p$). These noises are added to the simulated BOLD signals to form the simulated fMRI data. The design matrix is constructed as the same as the one introduced previously. The OLS method is first applied to the original simulated data in the time domain (denoted as "OLS-time"). After transforming the original simulated data and design matrix in (1) into wavelet domain as shown in (10), the Bayesian method and OLS method (denoted

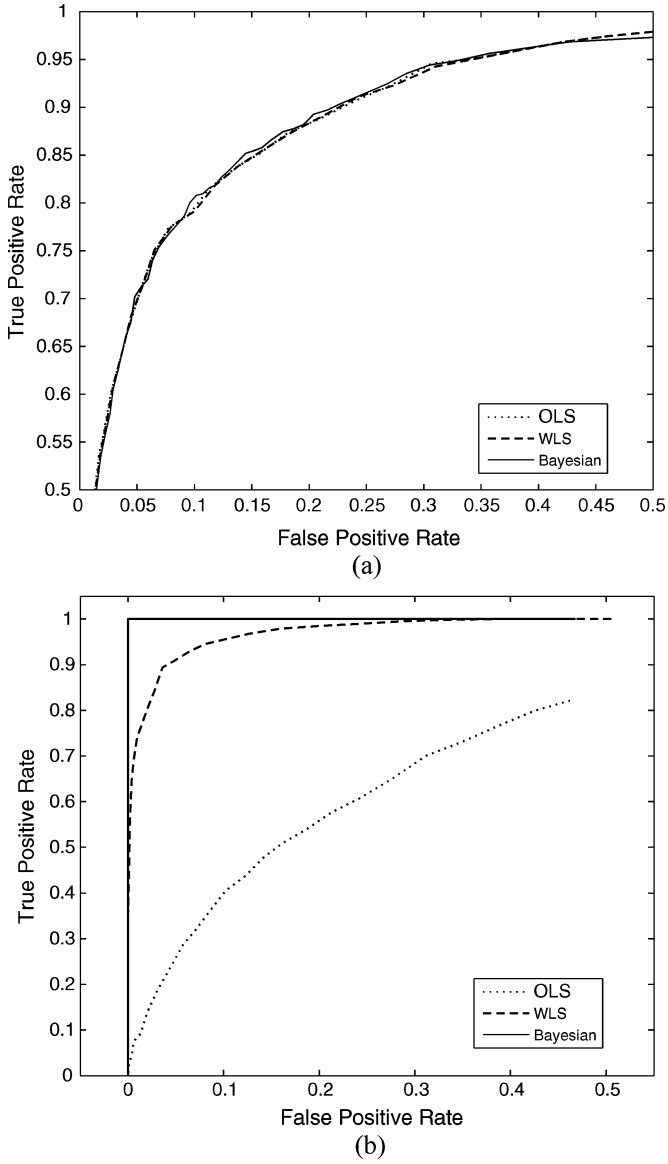

Fig. 3. ROC curves for simulated noisy data: (a) for i.i.d. noise; (b) for time-varying variance noise.

as "OLS-wavelet") are applied to the transformed data in the wavelet domain. Daubechies wavelets with vanishing moments 4 ("db4") are used for DWT since they are the most compactly supported wavelets with sufficient vanishing moments to whiten the fBm noise. To compare the detection results of different methods, generalized least squares (GLS) estimator is also investigated. In the GLS estimator, the noise covariance matrix is known according to the specified value of H and σ in (8) and (9). This GLS method is used as a comparison reference for the efficiency of the Bayesian method in the wavelet domain. Table II shows the SD of estimate $\hat{w}$ under different weight value w and different H using GLS, OLS-time, OLS-wavelet, and Bayesian methods in the wavelet domain.

From Table II, it is seen that the Bayesian method in wavelet domain is a robust and efficient estimator when the noise considered is fBm noise. Both the OLS method in time domain (OLS-time) and in wavelet domain (OLS-wavelet) perform worse when the Hurst exponent H is high since the noise

TABLE II
STANDARD DEVIATION (SD) OF ESTIMATED $\hat{w}$ ON SIMULATED DATA WITH DIFFERENT WEIGHT AND HURST EXPONENT

		GLS	OLS-time	OLS-wavelet	Bayes
$w=0$	$H=0.1$	0.0503	0.0518	0.0537	0.0384
	$H=0.3$	0.0664	0.0881	0.0937	0.0485
	$H=0.5$	0.0667	0.1305	0.1397	0.0484
	$H=0.7$	0.0608	0.2389	0.2433	0.0438
	$H=0.9$	0.0408	0.4499	0.3304	0.0332
$w=0.5$	$H=0.1$	0.0501	0.0523	0.0569	0.0529
	$H=0.3$	0.0678	0.0868	0.1002	0.0725
	$H=0.5$	0.0711	0.1326	0.1389	0.0797
	$H=0.7$	0.0642	0.2262	0.1980	0.0700
	$H=0.9$	0.0414	0.4601	0.3307	0.0491
$w=1$	$H=0.1$	0.0531	0.0556	0.0628	0.0551
	$H=0.3$	0.0685	0.0874	0.0924	0.0701
	$H=0.5$	0.0674	0.1277	0.1550	0.0730
	$H=0.7$	0.0637	0.2179	0.2155	0.0686
	$H=0.9$	0.0390	0.4542	0.3658	0.0442

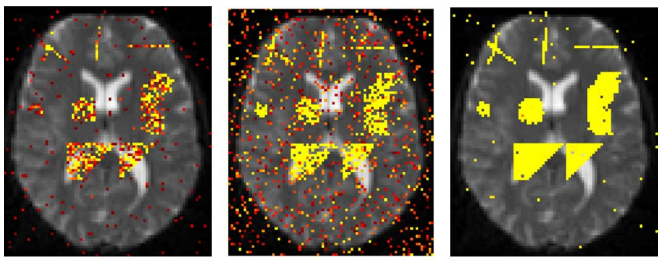

Fig. 4. Detection results of simulated data using fBm noise model: (a) OLS in time domain with thresholded SPM ($t > 3.4, p < 0.001$); (b) OLS in wavelet domain with thresholded SPM ($t > 3.4, p < 0.001$); (c) Bayesian method in wavelet domain with PPM ($P(\text{effect} > 1) > 0.99$).

assumption does not match the simulated data. The SD of Bayesian estimator is close to the GLS method, which shows that the Bayesian estimator is an accurate estimate of the weights $\mathbf{w}$ and could lead to almost the same performance as the GLS method. In the case of $w = 0$, the SD of the Bayesian method is even lower than the GLS method due to the pruning property of the proposed Bayesian method. These results show that wavelet transform actually whitens the noise and the diagonal error covariance matrix is accurately estimated by the Bayesian method.

The simulated 3-D fMRI time series are also investigated. The simulated data is generated as explained before. The difference is that the noises added here are synthesized fBm noises with different H at different voxels. Fig. 4 shows the activation detection results of these methods. It is clear from Fig. 4 that the proposed Bayesian estimator in the wavelet domain performs better than the OLS method (in time domain and wavelet domain) with more true activations and fewer false activations detected. The ROC curves of the OLS method (OLS-time and OLS-wavelet) and the Bayesian method are also given in Fig. 5 with a clear illustration of the better detection ability of the Bayesian method.

B. Real fMRI Data

In the real fMRI experiment, the subjects were asked to judge the line orientation. It is a block visuospatial processing experiment. A total of 100 acquisitions were made (the scan to scan

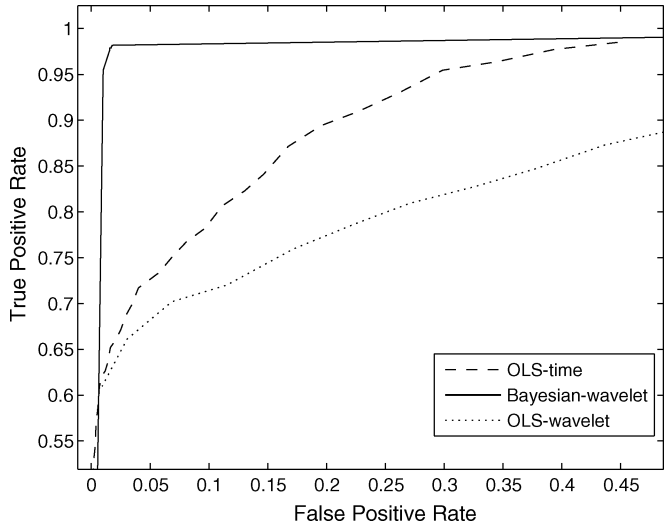

Fig. 5. ROC curves of OLS (in time domain and wavelet domain) and Bayesian (after DWT) methods for simulated fMRI data corrupted with fBm noises.

repetition time $\mathrm{TR} = 3$ s), with the experimental block length and rest block length 10. The condition for successive blocks alternated between rest and visuospatial stimulation, starting with rest. The first few scans are discarded due to $T1$ effects in the initial scans of an fMRI time series. For simplicity, the first ten volumes (i.e., the first block) of the data are discarded. The details of the experiment can be found in [36].

Fig. 6 shows the results of applying OLS, WLS, Bayesian methods assuming time-varying variance noise model and Bayesian methods assuming fractional noise model. For the fractional noise model, the time series are up-sampled in order to make the length of the time series to be a power of 2 and at the same time keep the scaling properties of the fractional noise. From these figures, we could see that the activation of the striate cortex and extrastriatal cortex are detected. Fig. 6(a) and (b) shows, respectively, the detection results of OLS and WLS method with the significance level p uncorrected. It is seen that at this p level, the detection results may include some erratic scattered noises. Considering the multiple comparison problem, a stringent correction using Bonferroni correction [33] is used here as shown in Fig. 6(c) and (d), which illustrate that some potentially activated voxels may fail to be detected. Compared to the OLS and WLS methods, the Bayesian methods perform better [Fig. 6(e) and (f)]. Besides, the Bayesian methods provide the probabilities of the activation effect, avoiding the problems encountered in the SPM to correct or adjust the p levels due to multiple comparison problem. These results show that the Bayesian estimators under these two noise assumptions can provide alternative estimators and better detection results.

VI. CONCLUSION

In the fMRI data analysis, the noise is generally assumed to be stationary. However, it is an inappropriate assumption given the complex nature of the fMRI data. The fMRI data is sensitive to the subjects' movements, resulting in the time-varying variance noise. In addition, the fractional noise or $1/f$-like noise exists in the fMRI data. In this paper, a Bayesian method is proposed to detect the activated voxels under two nonstationary noise

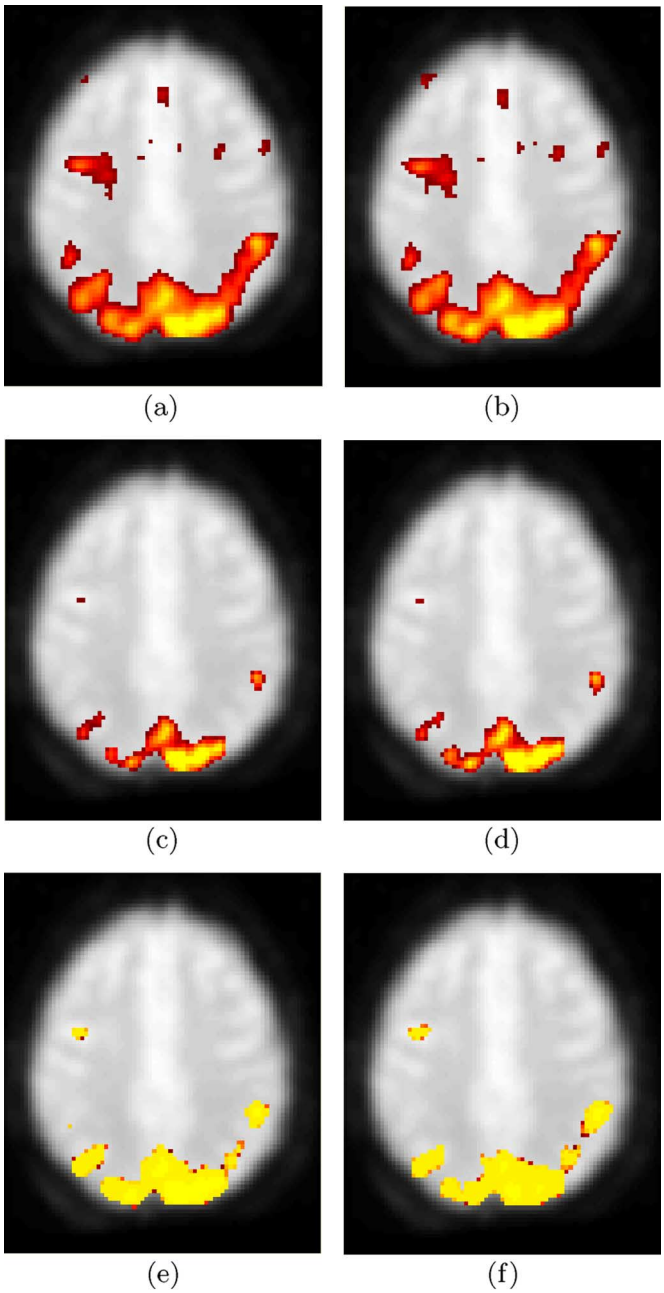

Fig. 6. Results of a block visuospatial processing task fMRI data: (a) thresholded SPM of OLS method ($t > 3.4, p < 0.001$, uncorrected); (b) thresholded SPM of WLS method ($t > 3.4, p < 0.001$, uncorrected); (c) thresholded SPM of OLS method with Bonferroni correction ($t > 7, p < 0.05$, corrected); (d) thresholded SPM of WLS method with Bonferroni correction ($t > 7, p < 0.05$, corrected); (e) PPM of Bayesian method using time-varying variance noise model ($P(\text{effect} > 0.8) > 0.99$); (f) PPM of Bayesian method using fractional noise model ($P(\text{effect} > 0.8) > 0.99$).

structures (time-varying variance noise and fractional noise). For time-varying variance noise, the variance of the noise at each time point is a scaled version of the overall variance in that voxel. For fractional noise, wavelet transform is first applied as a whitening filter. The coefficients of the transformed noise at each decomposition level are a set of i.i.d. variables with the variance a scaled version of the overall variance. Thus, the covariance matrices of these two assumed noise models have similar structures. This similarity leads to similar estimators or updating algorithms.

The weight estimator is investigated under the Bayesian framework. This estimator could provide a probability that a voxel is activated or the probability that the activation is greater than some threshold. This advantage enhances the ability to detect activated regions of the brain and avoids the limitations of the classical methods. The proposed methods are compared to the OLS and the WLS methods both on simulated as well as real fMRI data. The ROC analysis validates that the proposed Bayesian methods are more accurate than the OLS and WLS methods. These results show that the proposed Bayesian methods under the time-varying variance noise model and the fractional noise model are efficient and robust methods for brain activity detection in fMRI data analysis.

REFERENCES

[1] S. Ogawa, T. Lee, A. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain of high magnetic fields," *Magn. Res. Med.*, vol. 14, pp. 68–78, 1990.

[2] P. Jezzard, P. Matthews, and S. Smith, Eds., *Functional Magnetic Resoance Imaging: An Introduction to Methods*. Oxford, U.K.: Oxford Univ. Press, 2001.

[3] M. A. Brown and R. C. Semelka, *MRI: Basic Principles and Applications*, 3rd ed. New York: Wiley-Liss, 2003.

[4] K. Worsley and K. Friston, "Analysis of fMRI time-series revisited-again," *NeuroImage*, vol. 2, pp. 173–181, 1995.

[5] K. Friston and J. Ashburner *et al.*, SPM 97 Course Notes, Wellcome Dept. Cogn. Neurol., Univ. College London, London, U.K., 1997.

[6] O. Friman, J. Carlsson, P. Lundberg, M. Borga, and H. Knutsson, "Detection of neural activity in functional MRI using canonical correlation analysis," *Magn. Res. Med.*, vol. 45, pp. 323–330, 2001.

[7] M. Mckeown, S. Makeig, G. Brown, T. Jung, S. Kindermann, A. Bell, and T. Sejnowski, "Analysis of fMRI data by blind separation into independent spatial components," *Human Brain Mapp.*, vol. 6, pp. 160–188, 1998.

[8] H. Chen, H. Yuan, D. Yao, L. Chen, and W. Chen, "An integrated neighborhood correlation and hierarchical clustering approach of functional MRI," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 452–258, Mar. 2006.

[9] M. Woolrich, M. Jenkinson, J. Brady, and S. M. Smith, "Fully Bayesian spatio-temporal modeling of fMRI data," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 213–231, Feb. 2004.

[10] C. Long, E. Brown, C. Triantafyllou, I. Aharon, L. L. Wald, and V. Solo, "Nonstationary noise estimation in functional MRI," *NeuroImage*, vol. 28, pp. 890–903, 2005.

[11] J. Diedrichsen and R. Shadmehr, "Detecting and adjusting for artifacts in fMRI time series data," *NeuroImage*, vol. 27, pp. 624–634, 2005.

[12] 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 Mapp.*, vol. 12, pp. 61–78, 2001.

[13] E. Bullmore, J. Fadili, V. Maxim, L. Sendur, B. Whitcher, J. Suckling, M. Brammer, and M. Breakspear, "Wavelets and functional magnetic resonance imaging of the human brain," *NeuroImage*, vol. 23, pp. S234–S249, 2004.

[14] F. Meyer, "Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 315–322, Mar. 2003.

[15] W. Penny, S. Kiebel, and K. Friston, "Variational Bayesian inference for fMRI time series," *NeuroImage*, vol. 19, pp. 1477–1491, 2003.

[16] M. Woolrich, T. Behrens, and S. Smith, "Constrained linear basis sets for HRF modelling using variational Bayes," *NeuroImage*, vol. 21, pp. 1748–1761, 2004.

[17] K. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Theory," *NeuroImage*, vol. 16, no. 2, pp. 465–483, 2002.

[18] M. Woolrich, T. Behrens, C. Beckman, M. Jenkinson, and S. Smith, "Multilevel linear modelling for fMRI group analysis using Bayesian inference," *NeuroImage*, vol. 21, pp. 1732–1747, 2004.

[19] M. Burock and A. M. Dale, "Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach," *Human Brain Mapp.*, vol. 11, pp. 249–260, 2000.
[20] M. W. Woolrich, B. D. Ripley, M. Brady, and S. M. Smith, "Temporal autocorrelation in univariate linear modeling of fMRI data," *NeuroImage*, vol. 14, pp. 1370–1386, 2001.
[21] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny, Eds., *Human Brain Function*, 2nd ed. New York: Elsevier, 2004, ch. 37.
[22] E. Bullmore, J. Fadili, M. Breakspear, R. Salvador, J. Suckling, and M. Brammer, "Wavelets and statistical analysis of functional magnetic resonance images of the human brain," *Statist. Meth. Med. Res.*, vol. 12, pp. 375–399, 2003.
[23] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. New York: Academic, 1999.
[24] G. Wornell, "A Karhunen–Loéve-like expansion for 1/f processes via wavelets," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 859–861, Jul. 1990.
[25] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 910–917, Mar. 1992.
[26] 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.
[27] A. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 904–909, Mar/ 1992.
[28] E. McCoy and A. Walden, "Wavelet analysis and synthesis of stationary long-memory processes," *J. Comput. Graph. Statist.*, vol. 5, no. 1, pp. 26–56, 1996.
[29] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
[30] D. Mackay, "Bayesian interpolation," *Neural Comput.*, vol. 4, no. 3, pp. 415–417, 1992.
[31] H. Luo and S. Puthusserypady, "A sparse Bayesian method for determination of flexible design matrix for fMRI data analysis," *IEEE Trans. Circuits Syst I, Reg. Papers*, vol. 52, no. 12, pp. 2699–2706, Dec. 2005.
[32] K. Friston, D. Glaser, R. Henson, S. Kiebel, C. Phillips, and J. Ashburner, "Classical and Bayesian inference in neuroimaging: Applications," *NeuroImage*, vol. 16, no. 2, pp. 484–512, 2002.
[33] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny, Eds., *Human Brain Function*, 2nd ed. New York: Elsevier, 2004.
[34] K. Friston, P. Fletcher, O. Josephs, A. Holmes, M. Rugg, and R. Turner, "Event-related fMRI: Characterising differential responses," *NeuroImage*, vol. 7, pp. 30–40, 1998.
[35] R. Constable, P. Skudlarski, and J. Gore, "An ROC approach for evaluating functional brain MR imaging and postprocessing protocols," *Magn. Res. Med.*, vol. 34, pp. 57–64, 1995.
[36] V. Ng, E. Bullmore, G. de Zubicaray, A. Cooper, J. Suckling, and S. Williams, "Identifying rate-limiting nodes in large-scale cortical networks for visuospatial processing: An illustration using fMRI," *J. Cogn. Neurosci.*, vol. 13, pp. 537–546, 2001.

Huaien Luo received the B.E. and M.E. degrees in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2000 and 2003, respectively. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, National University of Singapore.

His research interests include statistical signal processing, especially biomedical signal processing and data analysis.

Sadasivan Puthusserypady (M'00–SM'05) received the B.Tech. degree in electrical engineering and the M.Tech. degree in instrumentation and control systems engineering from the University of Calicut, India, in 1986 and 1989, respectively, and the Ph.D. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1995.

From 1993 to 1996, he was a Research Associate at the Department of Psychopharmacology, NIMHANS, Bangalore. He was a Postdoctoral Research Fellow at the Communications Research Laboratory, McMaster University, Hamilton, ON, Canada, from 1996 to 1998. From 1998 to 2000, he was a Senior Systems Engineer at Raytheon Systems Canada, Ltd., Waterloo, ON. He is currently an Assistant Professor at the Department of Electrical and Computer Engineering, National University of Singapore. His research interests are in biomedical signal processing, neural networks, multiuser detection, and chaotic communication systems.