Computational Statistics and Data Analysis 52 (2008) 4859–4871

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Efficient modeling and inference for event-related fMRI data

Chunming Zhang [a,*], Yuefeng Lu [b], Tom Johnstone [c], Terry Oakes [c], Richard J. Davidson [c]

[a] *Department of Statistics, University of Wisconsin, Madison, WI 53706, USA*

[b] *Eli Lilly and Company, USA*

[c] *Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin, Madison, WI 53705, USA*

ARTICLE INFO

Article history:
Received 15 November 2007
Received in revised form 31 March 2008
Accepted 31 March 2008
Available online 8 April 2008

ABSTRACT

Event-related functional magnetic resonance imaging (efMRI) has emerged as a powerful technique for detecting brains' responses to presented stimuli. A primary goal in efMRI data analysis is to estimate the Hemodynamic Response Function (HRF) and to locate activated regions in human brains when specific tasks are performed. This paper develops new methodologies that are important improvements not only to parametric but also to nonparametric estimation and hypothesis testing of the HRF. First, an effective and computationally fast scheme for estimating the error covariance matrix for efMRI is proposed. Second, methodologies for estimation and hypothesis testing of the HRF are developed. Simulations support the effectiveness of our proposed methods. When applied to an efMRI dataset from an emotional control study, our method reveals more meaningful findings than the popular methods offered by AFNI and FSL.

1. Introduction

Functional magnetic resonance imaging (fMRI) is a recent and exciting method that allows investigators to determine which areas of the brain are involved in a particular task. In an fMRI experiment, subjects are asked to perform certain motor, sensory or cognitive tasks, and a time series of three-dimensional MR images is recorded. The fMRI signal is a result of the blood oxygen level dependent (BOLD) contrast. One important task in analyzing fMRI data is to determine regions of activation.

Traditional fMRI experiments are block-designed and constrained by successive stimuli, which are of the same type. Event-related fMRI (efMRI) is a more recently developed technique that is used to detect brain responses to brief stimuli and offers more flexibility over traditional block-designed fMRI techniques. For example, in a neuroimage study (Gobbini et al., 2004), subjects viewed a random series of face photos. The study showed that subjects have a stronger response when viewing faces of personally familiar individuals than faces of famous individuals or strangers, which as interpreted by Gobbini et al. (2004) may "reflect the spontaneous activation of social knowledge about the personality and attitudes of close friends and relatives and the less guarded attitude one has around these people". For a more complete review of efMRI and related statistical issues, refer to Rosen et al. (1998) and Josephs and Henson (1999).

The analysis of massive spatio-temporal data sets generated by efMRI requires adequate statistical modeling. Let $\{y(t_i)\}_{i=1}^n$ be the observed efMRI time series on a specific voxel (for *volume element* in three dimensional space). Let $s(\cdot)$ be the observed input stimulus function which assumes values of 1 and 0, indicating presence and absence of a stimulus in the experimental

* Corresponding address: Department of Statistics, 1300 University Avenue, University of Wisconsin, Madison, WI 53706, USA. Tel.: +1 608 262 0084; fax: +1 608 262 0032.

E-mail addresses: cmzhang@stat.wisc.edu (C. Zhang), LU_YUEFENG@Lilly.com (Y. Lu), itjohnstone@wisc.edu (T. Johnstone), troakes@wisc.edu (T. Oakes), rjdavids@wisc.edu (R.J. Davidson).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.03.033

design. Denote by $*$ the convolution operator. A convolution model popularly used in fMRI study is captured by

$$y(t) = s * h(t) + d(t) + \epsilon(t), \quad t = t_1, \ldots, t_n, \tag{1.1}$$

where h, d and ϵ represent the unknown hemodynamic response function (HRF), drift function and noise process respectively. Most existing methods model $h(\cdot)$ as the difference of two gamma functions or a linear combination of gamma functions or a linear combination of a gamma function and its Taylor expansion (Worsley et al., 2002; Lange and Zeger, 1997; Josephs and Henson, 1999). Genovese (2000) constructed $h(\cdot)$ as a "bell" function with cubic splines. The temporal drift $d(\cdot)$ is usually approximated by a quadratic or higher-order polynomial (Worsley et al., 2002) or polynomial splines (Genovese, 2000). Note that restrictive assumptions on the HRF and drift may produce biased estimates of the true hemodynamic responses. Goutte et al. (2000) estimated $h(\cdot)$ using smooth FIR filters and reported that some subtle details of the HRF can be revealed by the filters but not by previous approaches based on gamma functions.

The noise components in (1.1) are well-known to be temporally auto-correlated. In practice, for any parametric or nonparametric estimation method for the HRF and drift, an effective and computationally fast method for estimating the noise covariance matrix is needed. Genovese (2000) assumed independent errors for computational convenience. Other assumptions like the AR(p) structure, most commonly AR(1), are used in Worsley et al. (2002).

Since an fMRI dataset contains time-course measurements over voxels, the number of which is typically of the order of 10^4–10^5, a multiple testing procedure is also needed. In the seminal work, Worsley et al. (2002) proposed a Gaussian random field method which approximates the family-wise error by modeling test statistics over the entire brain as a Gaussian random field. It has been found to be conservative in some cases (Nichols and Hayasaka, 2003). Nichols and Hayasaka (2003) also discussed the use of permutation tests and their simulation studies show that permutation tests tend to be more sensitive in finding activated regions. The false discovery rate (FDR) has recently become popular. Genovese et al. (2002) discussed the use of FDR in fMRI. Pacifico et al. (2004) extended FDR control to random fields and illustrated its use for fMRI data.

This paper aims to develop new methodologies which are important not only to parametric, but also to nonparametric, estimation and hypothesis testing of the HRF. First, an effective and computationally fast scheme for estimating the error covariance matrix is proposed. Second, methodologies for estimation and hypothesis testing of the HRF are developed. The rest of the article is organized as follows. Section 2 introduces the background of efMRI. Section 3 proposes a method for estimating the noise covariance matrix. Section 4 presents methodologies for estimation and hypothesis testing of the HRF. Simulation evaluation and real data analysis are illustrated in Sections 5 and 6 respectively. The paper concludes with a brief discussion in Section 7.

2. Background of efMRI data

2.1. Probabilistic aspects of efMRI design

Following Josephs and Henson (1999, Sec. 6) and Friston et al. (1999), the efMRI designs can be categorized into two groups, deterministic and stochastic. In deterministic designs, the stimulus transition probabilities are either 0 or 1. Stochastic designs refer to those in which the probability is strictly less than 1, and a special case of stochastic designs is fully randomized designs, which are desirable in most situations. Stochastic designs may be stationary, where the probability is constant over time, or dynamic, where the probabilities vary with time. Rapid-presentation event-related (RPER) design is itself a fully randomized design and includes other fully randomized designs (like the fixed-interval event-related design) as special cases. Software is also available for automatically implementing the RPER design.

Suppose that there are r event types. At any time, the stationary stochastic design randomly chooses an event type or "no event" with fixed probabilities. Furthermore, for the RPER design, not only are the probabilities fixed, but the choice of the current event type is also independent of the choices of the previous event types. For stimulus functions $s_1(t), \ldots, s_r(t)$ associated with r event types, the stationary stochastic design can be characterized by $P\{s_j(t) = 1\} = p_j, j = 1, \ldots, r$, i.e. at any time t, $(s_1(t), \ldots, s_r(t), 1 - \sum_{j=1}^{r} s_j(t))$ follows a multinomial distribution $(1; p_1, \ldots, p_r, 1 - \sum_{j=1}^{r} p_j)$. Furthermore, if we assume that $s_{j_1}(t_u)$ and $s_{j_2}(t_v)$ are independent at any $t_u \neq t_v$, then the design is RPER.

2.2. Convolution model for efMRI

We start with a single event type. Typically, the peak value of the HRF $h(\cdot)$ is reached after a short delay of the stimulus and drops quickly to zero. We thus suppose that $h(t) = 0$ for $t > t_m$ and focus on estimating the first m values of $h(t_i)$, where m is far below n, the length of the fMRI time series. To facilitate discussion, we assume that $y(\cdot)$ and $s(\cdot)$ have equal time resolution of 1 s. Letting $\mathbf{h} = (h(t_1), \ldots, h(t_m))^{\mathrm{T}}$ and $\mathbf{s}_i = (s(t_i - t_1), \ldots, s(t_i - t_m))^{\mathrm{T}}$ where $s(t) = 0$ for $t < 0$, model (1.1) can be rewritten as $y(t_i) = \mathbf{s}_i^{\mathrm{T}}\mathbf{h} + d(t_i) + \epsilon(t_i)$ and re-expressed as

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{d} + \boldsymbol{\epsilon}, \tag{2.1}$$

where $\mathbf{y} = (y(t_1), \ldots, y(t_n))^{\mathrm{T}}$, $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_n]^{\mathrm{T}}$, $\mathbf{d} = (d(t_1), \ldots, d(t_n))^{\mathrm{T}}$ and $\boldsymbol{\epsilon} = (\epsilon(t_1), \ldots, \epsilon(t_n))^{\mathrm{T}}$.

For multiple types of events, model (1.1) can be extended to the general form,

$$y(t) = s_1 * h_1(t) + \cdots + s_r * h_r(t) + d(t) + \epsilon(t), \quad t = t_1, \ldots, t_n, \tag{2.2}$$

in which we assume that $h_1(\cdot), \ldots, h_r(\cdot)$ and $d(\cdot)$ are smooth functions. Let $\mathbf{S}_j$ be the design matrix and $\mathbf{h}_j$ be the vector of the HRF corresponding to the jth event type. Then the form of (2.1) continues to hold, since the matrix form of (2.2) becomes

$$\mathbf{y} = \mathbf{S}_1\mathbf{h}_1 + \cdots + \mathbf{S}_r\mathbf{h}_r + \mathbf{d} + \boldsymbol{\epsilon} \equiv \mathbf{S}\mathbf{h} + \mathbf{d} + \boldsymbol{\epsilon}, \tag{2.3}$$

where $\mathbf{S} = [\mathbf{S}_1, \ldots, \mathbf{S}_r]$ and $\mathbf{h} = [\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_r^{\mathrm{T}}]^{\mathrm{T}}$.

3. Estimation of noise covariance matrix for efMRI

In fMRI data analysis, not all commonly used assumptions on the error process are computationally feasible in estimating the error covariance matrix. Since the temporal correlation decreases significantly as the time interval between measurements increases, we assume that $\{\epsilon(t_i)\}$ is a zero-mean stationary g-dependent sequence (Sen, 1968, p. 1724) with variance equal to σ^2, where g is a non-negative integer. Namely, the random vectors $(\epsilon(t_1), \ldots, \epsilon(t_i))$ and $(\epsilon(t_j), \epsilon(t_{j+1}), \ldots)$ are stochastically independent if $j - i > g$. Denote by $\rho(j)$ and $\gamma(j)$ the auto-correlation and auto-covariance of $\{\epsilon(t_i)\}$ respectively, with time lag $j \geq 0$. It is readily seen that

$$\rho(0) = 1, \quad \text{and} \quad \rho(|j|) = 0 \text{ for } |j| > g, \tag{3.1}$$

$$\gamma(j) = \sigma^2\rho(j). \tag{3.2}$$

Denote by R_n the true correlation matrix of $\boldsymbol{\epsilon}$, i.e., $\mathrm{cov}(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}) = \sigma^2 R_n$, where

$$R_n = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(n-1) \\ \rho(1) & 1 & \cdots & \rho(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(n-1) & \rho(n-2) & \cdots & 1 \end{bmatrix}.$$

To reduce the computational cost in analyzing massive fMRI data sets, we propose a difference-based approach for estimating $\sigma^2 R_n$. Unlike some smoothing-based approaches, the difference-based method enjoys computational simplicity and good practical performance. Define by $\mathcal{D}\{y\}(t_i) = y(t_i) - y(t_{i-1})$ the operator for the first-order difference of $\{y(t_i)\}$ and by $\mathcal{D}^k = \mathcal{D}\mathcal{D}^{k-1}$ the kth-order difference. Concerning the order of difference, analogous to the asymptotic study (Fan and Zhang, 2003) performed in diffusion models for financial time series, the higher-order difference may decrease the estimation bias but may also exponentially escalade the variance of an estimator. In the convolution model (2.2) for fMRI data, our simulation studies indicate that the second-order difference outperforms the first-order difference in estimating the covariance matrix and thus the second-order difference is adopted. Denote by

$$e(t_i) \equiv \mathcal{D}^2\left(y - \sum_{j=1}^{r} s_j * h_j\right)(t_i) \tag{3.3}$$

the transformed data after applying the second-order difference. Let $\gamma_e(j)$ be the auto-covariance of $\{e(t_i)\}$, with time lag $j \geq 0$. Since $e(t_i) = \mathcal{D}^2 d(t_i) + \epsilon(t_i) - 2\epsilon(t_{i-1}) + \epsilon(t_{i-2})$, direct calculations give that

$$\gamma_e(j) = \gamma(j-2) - 4\gamma(j-1) + 6\gamma(j) - 4\gamma(j+1) + \gamma(j+2), \quad j = 0, 1, \ldots.$$

This, along with (3.1)–(3.2), leads to the identities,

$$\begin{bmatrix} \gamma_e(0) \\ \gamma_e(1) \\ \gamma_e(2) \\ \vdots \\ \gamma_e(g-1) \\ \gamma_e(g) \end{bmatrix} = \begin{bmatrix} 6 & -8 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -4 & 7 & -4 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 & -4 & 6 & -4 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -4 & 6 \end{bmatrix} \begin{bmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(g-1) \\ \gamma(g) \end{bmatrix}. \tag{3.4}$$

Substituting $\{\gamma_e(j)\}$ with their empirical moment estimates and solving Eq. (3.4), we can obtain the estimates $\{\hat{\gamma}(j)\}$ of $\{\gamma(j)\}$. Utilizing (3.1)–(3.2), the noise variance and auto-correlation can be estimated by $\hat{\sigma}^2 = \hat{\gamma}(0)$ and $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$ respectively. It is then immediate to acquire an estimate $\hat{R}_n$ of R_n.

In practice, we need a preliminary estimate of the HRF in (3.3). Again, this initial estimate $\hat{\mathbf{h}}_{\mathbf{DBE}}$ can be obtained via a difference-based method: *apply the first-order difference to* $\mathbf{y}$ *and* $\mathbf{S}$ *simultaneously, ignore the drift term in* (2.3)*, followed by an ordinary least-squares estimation of the HRF*. To illustrate the idea, let us take the difference of (2.3). Under the smoothness assumption on the drift function $d(\cdot)$, $d(t_i) - d(t_{i-1}) = d'(t_{i-1})n^{-1} + O(n^{-2}) = O(n^{-1})$. Hence the difference of (2.3) yields

$$y(t_i) - y(t_{i-1}) = (\boldsymbol{e}_i - \boldsymbol{e}_{i-1})^{\mathrm{T}}\mathbf{S}\mathbf{h} + \epsilon(t_i) - \epsilon(t_{i-1}) + O(n^{-1}), \quad i = 2, \ldots, n,$$

where $\boldsymbol{e}_j$ denotes the jth column of an identity matrix. As the sequence length n grows, the magnitude of the difference (and higher-order difference) of $d(\cdot)$ will be negligible. This example lends support to the root-n consistency of $\hat{\mathbf{h}}_{\mathbf{DBE}}$ under mild conditions.

4. Methodology for inference of the HRF

In this section, we first develop methodologies for the flexible estimation of the HRF and then conduct hypothesis testing for the HRF. The local-linear smoothing (Fan and Gijbels, 1996) technique will be focused on. Extensions to local-polynomial smoothing and other nonparametric curve-fitting methods can similarly be made.

4.1. Two stage estimation of the HRF

Let S_d be an $n \times n$ local-linear smoothing matrix (see Zhang (2003)), associated with the design points $\{t_1, \ldots, t_n\}$, where the subscript d emphasizes the use for estimating the drift. Throughout the paper, $t_i = i/n$ is assumed. Then the vector of smoothed responses, via local-linear smoothing to a vector $\mathbf{y}$ of responses, is given by $S_d\mathbf{y}$. Note that S_d carries information about the design points, a kernel function K and a bandwidth parameter b, but does not rely on the configuration of the response variables. Refer to Zhang (2003, Sec. 2.3) for further discussion on finite-sample and asymptotic properties of the smoothing matrix.

First consider a single event type. For given $\mathbf{h}$, applying S_d to $\mathbf{y} - \mathbf{Sh}$ in (2.1) supplies an estimate, $S_d(\mathbf{y} - \mathbf{Sh})$, of the vector $\mathbf{d}$ of drift component. Putting this estimate to (2.1), we obtain

$$\widetilde{\mathbf{y}} \doteq \widetilde{\mathbf{S}}\mathbf{h} + \boldsymbol{\epsilon}, \tag{4.1}$$

where $\widetilde{\mathbf{y}} = (\mathbf{I} - S_d)\mathbf{y}$, $\widetilde{\mathbf{S}} = (\mathbf{I} - S_d)\mathbf{S}$ and $\mathbf{I}$ denotes an identity matrix. Notice that (4.1) can be regarded as a general linear model. Suppose that R_n, the true correlation matrix of $\boldsymbol{\epsilon}$, is known. (For unknown R_n, refer to the preceding Section 3 for the estimation of R_n.) By the weighted least squares method, a rough estimate of $\mathbf{h}$ is produced by

$$\hat{\mathbf{h}}_{\text{rough}} = (\widetilde{\mathbf{S}}^{\mathrm{T}} R_n^{-1} \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^{\mathrm{T}} R_n^{-1} \widetilde{\mathbf{y}}. \tag{4.2}$$

This offers the first-stage estimate of the HRF. The second-stage estimate of the HRF smooths $\hat{\mathbf{h}}_{\text{rough}}$ to produce a refined estimate,

$$\hat{\mathbf{h}}_{\text{smooth}} = S_h \hat{\mathbf{h}}_{\text{rough}},$$

where S_h denotes an $m \times m$ local-linear smoothing matrix associated with the design points $\{t_1, \ldots, t_m\}$ and the subscript h emphasizes the use for estimating the HRF $\mathbf{h}$ in (2.1).

For multiple types of events, suppose that $\mathbf{h}_j, j = 1, \ldots, r$, in (2.3) are of common length m. We can similarly obtain the first- and second-stage estimates of the HRF,

$$\hat{\mathbf{h}}_{\text{rough}} = (\widetilde{\mathbf{S}}^{\mathrm{T}} R_n^{-1} \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^{\mathrm{T}} R_n^{-1} \widetilde{\mathbf{y}}, \quad \text{and} \quad \hat{\mathbf{h}}_{\text{smooth}} = S_h \hat{\mathbf{h}}_{\text{rough}}, \tag{4.3}$$

where $S_h = \text{diag}\{S_{h_1}, \ldots, S_{h_r}\}$ (in which the diag operator is used to define S_h as a block-diagonal matrix) and each S_{h_j} is an $m \times m$ local-linear smoothing matrix. Our simulation studies in Section 5.1 demonstrate that the efficiency gain of $\hat{\mathbf{h}}_{\text{smooth}}$ over $\hat{\mathbf{h}}_{\text{rough}}$ in estimating the HRF is negative when the noise level is small and is negligible when the noise level is large.

4.2. Selection of bandwidth

Bandwidth selection in the first-stage smoothing: The bandwidth selection used in S_d for $\hat{\mathbf{h}}_{\text{rough}}$ serves as bandwidth selection for the parametric component in model (2.3). Define $V_n = R_n^{-1}$ and $\widetilde{\mathbf{d}} = (\mathbf{I} - S_d)\mathbf{d}$. From (4.3), the mean squared error (MSE) of $\hat{\mathbf{h}}_{\text{rough}}$ is

$$E[\|\hat{\mathbf{h}}_{\text{rough}} - \mathbf{h}\|^2 | \{s_k(t_i - t_j) : 1 \le k \le r, 1 \le i \le n, 1 \le j \le \min(i, m)\}] = I_1 + I_2, \tag{4.4}$$

where $I_1 = \|(\widetilde{\mathbf{S}}^{\mathrm{T}} V_n \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^{\mathrm{T}} V_n \widetilde{\mathbf{d}}\|^2$ reflects the squared bias, and $I_2 = \text{trace}\{(\widetilde{\mathbf{S}}^{\mathrm{T}} V_n \widetilde{\mathbf{S}})^{-1} \widetilde{\mathbf{S}}^{\mathrm{T}} V_n (\mathbf{I} - S_d) R_n (\mathbf{I} - S_d^{\mathrm{T}}) V_n \widetilde{\mathbf{S}} (\widetilde{\mathbf{S}}^{\mathrm{T}} V_n \widetilde{\mathbf{S}})^{-1}\} \sigma^2$ reflects the variance. To balance the trade off between the bias and variance of $\hat{\mathbf{h}}_{\text{rough}}$, the data-driven bandwidth can be chosen to minimize the plug-in estimate of (4.4). That is, substituting the unknown quantities σ^2 and R_n by the estimates developed in Section 3; replacing $\mathbf{d}$ by an estimate obtained from local-linear smoothing to $\mathbf{y} - \mathbf{S}\hat{\mathbf{h}}_{\mathtt{DBE}}$ (where the bandwidth parameter can be selected via minimizing the generalized cross-validation criterion).

Bandwidth selection in the second-stage smoothing: The second-stage smoothing, applied to the first-stage estimate $\hat{\mathbf{h}}_{\text{rough}}$, intends to produce a smoother estimate. Nevertheless, care is needed in choosing the amount of smoothing. For a small noise level, the excessive bias introduced by the second-stage smoothing may deteriorate $\hat{\mathbf{h}}_{\text{rough}}$. In contrast, for a large noise level, since $\hat{\mathbf{h}}_{\text{rough}}$ tends to be very rough, an appropriate smoothing may increase the estimation efficiency. We propose a data-driven bandwidth selector used in S_h for the second-stage smoothing to minimize Mallow's C_p criterion,

$$C_p = \|\hat{\mathbf{h}}_{\text{rough}} - S_h \hat{\mathbf{h}}_{\text{rough}}\|^2 + 2\hat{\sigma}^2 \text{trace}\{(\widetilde{\mathbf{S}}^{\mathrm{T}} \hat{R}_n^{-1} \widetilde{\mathbf{S}})^{-1} S_h\}, \tag{4.5}$$

where $(\widetilde{\mathbf{S}}^{\mathrm{T}} \hat{R}_n^{-1} \widetilde{\mathbf{S}})^{-1} \hat{\sigma}^2$ estimates $\text{var}(\hat{\mathbf{h}}_{\text{rough}} | \{s_k(t_i - t_j) : 1 \le k \le r, 1 \le i \le n, 1 \le j \le \min(i, m)\})$.

4.3. Hypothesis testing for the HRF

Besides estimating the HRF, neuro-scientists are particularly interested in learning whether a specific voxel is activated by the stimuli, and whether HRFs activated by different types of events really differ. They correspond to testing the following hypotheses,

$$H_0 : \mathbf{h} = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{h} \neq \mathbf{0}, \tag{4.6}$$

and $H_0 : \mathbf{h}_{j_1} = \mathbf{h}_{j_2}$ versus $H_1 : \mathbf{h}_{j_1} \neq \mathbf{h}_{j_2}$, where $j_1 \neq j_2$. These testing problems can be rephrased in a more general form,

$$H_0 : A\mathbf{h} = \mathbf{0} \quad \text{versus} \quad H_1 : A\mathbf{h} \neq \mathbf{0}, \tag{4.7}$$

where A is a full row rank matrix.

We now describe a pseudo nonparametric F-type test statistic. Recall that approximation (4.1) takes the form of a general linear model. Let $\mathsf{k} = \text{rank}(A)$. For an HRF estimator $\hat{\mathbf{h}}$, the pseudo F-test statistic is formed by

$$\frac{(A\hat{\mathbf{h}})^{\mathrm{T}}\{A(\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{S}})^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\mathbf{h}})/\mathsf{k}}{\widetilde{\sigma}^2},$$

where $\hat{\mathbf{h}} = \hat{\mathbf{h}}_{\text{rough}}$ and $\widetilde{\sigma}^2 = \hat{\mathbf{r}}^{\mathrm{T}}\hat{R}_n^{-1}\hat{\mathbf{r}}/(n - rm)$ with $\hat{\mathbf{r}} = \widetilde{\mathbf{y}} - \widetilde{\mathbf{S}}\hat{\mathbf{h}}$. Our simulations in Fig. 4 lend numerical support that the sampling distribution of the pseudo F-test statistic, under the null hypothesis in (4.7), can be well approximated by an F distribution with degrees of freedom k and $n - rm$. Insights into the distributional approximation can be gained from Fan et al. (2001).

In the meantime, our simulations reveal that the approximation may be less accurate when the noise level is small. The lack of accuracy is caused by the approximation (4.1) which ignores the $\widetilde{\mathbf{d}}$ term. This term is negligible when the noise level is large, but becomes a significant source of bias when the noise level is small. To remedy this, we use a bias corrected pseudo F-test statistic,

$$\frac{(A\hat{\mathbf{h}}_{\text{bc}})^{\mathrm{T}}\{A(\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{S}})^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\mathbf{h}}_{\text{bc}})/\mathsf{k}}{\widetilde{\sigma}^2_{\text{bc}}},$$

where $\hat{\mathbf{h}}_{\text{bc}} = \hat{\mathbf{h}} - (\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{S}})^{-1}\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\hat{\mathbf{d}}}$ and $\widetilde{\sigma}^2_{\text{bc}} = \hat{\mathbf{r}}_{\text{bc}}^{\mathrm{T}}\hat{R}_n^{-1}\hat{\mathbf{r}}_{\text{bc}}/(n - rm)$, with $\hat{\mathbf{r}}_{\text{bc}} = \widetilde{\mathbf{y}} - \widetilde{\mathbf{S}}\hat{\mathbf{h}} - \widetilde{\hat{\mathbf{d}}}$, $\widetilde{\hat{\mathbf{d}}} = (\mathbf{I} - S_d)\hat{\mathbf{d}}$, $\hat{\mathbf{d}} = S_d(\mathbf{y} - \mathbf{S}\hat{\mathbf{h}})$ and $\hat{\mathbf{h}} = \hat{\mathbf{h}}_{\text{rough}}$. Note that for a large noise level, $\widetilde{\mathbf{d}}$ is negligible and thus the bias corrected and un-corrected versions are similar.

5. Simulation study

5.1. Simulation over a single voxel

We simulate an fMRI experiment with a single run and a single event type. In the simulation, $n = 200$, $t_i = i/n$, $i = 1, \ldots, n$, $m = 18$ and 1000 realizations are conducted. The stimuli are generated from independent Bernoulli trials such that $P\{s(t) = 1\} = .5$. Following Glover (1999), the HRF is $h(t) = g_1(1.5(200t-1))/a_1 - g_2(1.5(200t-1))/a_2$, $t = t_1, \ldots, t_{18}$, where $g_1(t) = (t-5.5)^5 \exp\{-(t-5.5)/.9\}$ and $g_2(t) = .4(t-5.5)^{12}\exp\{-(t-5.5)/.7\}$, $a_1 = \max\{g_1(t)\}$ and $a_2 = \max\{g_2(t)\}$. The drift function is $d(t) = 10\sin\{\pi(t - .21)\}$, $t = t_1, \ldots, t_{200}$. The noise process is $\epsilon = \epsilon_1 + \epsilon_2$, with independent $\{\epsilon_1\}$ and $\{\epsilon_2\}$ (see Purdon et al. (2001)); $\{\epsilon_1(t)\}$ are i.i.d. normal with mean zero and variance $.5216^2$, $.3689^2$, $.2608^2$ and $.1844^2$, respectively; ϵ_2 is AR(1), i.e. $\epsilon_2(t_i) = \rho\epsilon_2(t_{i-1}) + z(t_i)$ with $\rho = .638$ and $z(t_i)$ follows the normal distribution with mean zero and variance $.5216^2$, $.3689^2$, $.2608^2$ and $.1844^2$ respectively such that the noise lag-one auto-correlation equals .4 and the signal-to-noise ratio (SNR) is about 1, 2, 4 and 8, where $\text{SNR} = \text{var}(\mathbf{Sh})/\text{var}(\boldsymbol{\epsilon})$. Throughout the paper, the Epanechnikov kernel function is used in the implementation and $g = 2$ is used in the covariance matrix estimation.

First, we compare the sums of squared errors (SSE) of six types of HRF estimates.

WPLt: $\hat{\mathbf{h}} = (\widetilde{\mathbf{S}}^{\mathrm{T}}R_n^{-1}\widetilde{\mathbf{S}})^{-1}\widetilde{\mathbf{S}}^{\mathrm{T}}R_n^{-1}\widetilde{\mathbf{y}}$ where the bandwidth b in S_d minimizes $\|\hat{\mathbf{h}} - \mathbf{h}\|^2$.

WPLe: $\hat{\mathbf{h}} = (\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{S}})^{-1}\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{y}}$ where b in S_d minimizes $\|\hat{\mathbf{h}} - \mathbf{h}\|^2$.

pWPL: $\hat{\mathbf{h}}_{\text{rough}} = (\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{S}})^{-1}\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{y}}$ where b in S_d minimizes the estimate of (4.4).

2WPL: $\hat{\mathbf{h}}_{\text{smooth}}$, where b in S_h minimizes the C_p criterion (4.5).

nWPL: $\hat{\mathbf{h}} = (\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{S}})^{-1}\widetilde{\mathbf{S}}^{\mathrm{T}}\hat{R}_n^{-1}\widetilde{\mathbf{y}}$ where b in S_d minimizes the generalized cross-validation.

DBE: $\hat{\mathbf{h}}_{\mathtt{DBE}}$ described at the end of Section 3.

In practice, since the true R_n is unknown, WPLt and WPLe are infeasible but theoretically offer benchmarks for evaluating the performances of the estimated covariance matrix and the HRF estimator. Conclusions can be drawn from Fig. 1. First, the SSEs of WPLt, WPLe, pWPL, nWPL and DBE are ranked in decreasing order. Second, the SSEs of WPLt and WPLe are nearly indistinguishable. This reinforces the effectiveness of the estimated covariance matrix in Section 3. Third, 2WPL slightly outperforms pWPL when the noise level is large, but has adverse effects when the noise level is small. Fourth, the pWPL estimate of the HRF achieves a smaller SSE than nWPL and thus will be adopted in the rest of the paper.

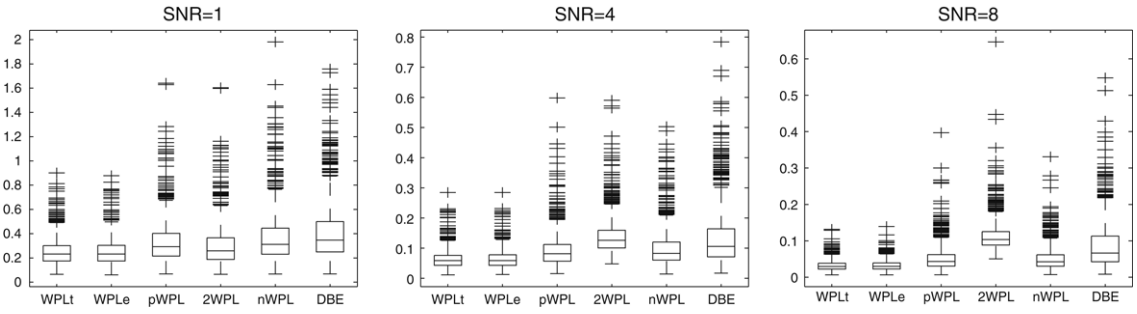

Fig. 1. Boxplots of $\|\hat{\mathbf{h}} - \mathbf{h}\|^2$ for six types of estimates $\hat{\mathbf{h}}$ listed in Section 5.1.

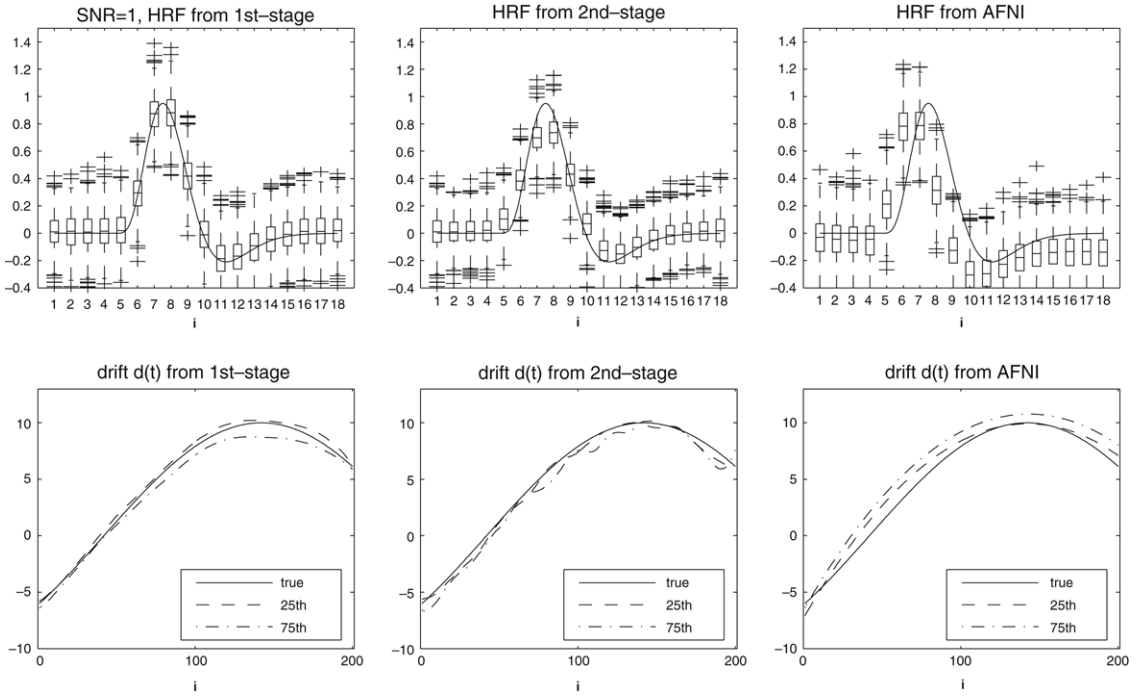

Fig. 2. Comparison of first-stage, second-stage and AFNI estimates of the HRF and drift. Top panel: boxplots of $\hat{\mathbf{h}}_{\text{rough}}$ and $\hat{\mathbf{h}}_{\text{smooth}}$; the solid curves are the true function $h(t)$. Bottom panel: solid curves denote the true drift $d(t)$, and the estimated curves from two typical samples are presented corresponding to the 25th (the dashed curve) and the 75th (the dash-dotted curve) percentiles among the ASE–ranked values, where $\text{ASE} = \sum_{i=1}^{n}\{\hat{d}(i) - d(i)\}^2/n$.

Second, we compare the first-stage and second-stage estimates of the HRF and temporal drift. Fig. 2 (with SNR $= 1$) and Fig. 3 (with SNR $= 8$) are displayed, where the first- and second-stage estimates of drifts are $\hat{\mathbf{d}}_{\text{rough}} = S_d(\mathbf{y} - \mathbf{S}\mathbf{h}_{\text{rough}})$ and $\hat{\mathbf{d}}_{\text{smooth}} = S_d(\mathbf{y} - \mathbf{S}\mathbf{h}_{\text{smooth}})$ respectively. Evidently, compared with $\hat{\mathbf{h}}_{\text{smooth}}$, $\hat{\mathbf{h}}_{\text{rough}}$ are more centered around the actual values of the HRF. On the other hand, the drift function estimates are less affected by first-stage and second-stage estimation of the HRF. Besides, the HRF and drift estimates by our method outperform those by AFNI.

Next, we perform the hypothesis testing of (4.6). The fMRI data are simulated in the same way as above except that $\mathbf{h} = \mathbf{0}$. For the sake of presentation, only the cases of SNR equal to 1 and 8 are studied; the former is the "large noise level" case whereas the latter is the "small noise level" case. To focus on the performance of the proposed F-test, we use the true covariance matrix and fix the smoothing parameters at their theoretically optimal values (minimizing MSE of estimators) for estimating the HRF and drift in each simulation. The QQ plots of the (1st up to 99th) percentiles of the bias corrected pseudo F-test statistic against those of the $F_{m,n-m}$ distribution are presented in Fig. 4. For either large or small noise levels, we observe that the sampling distribution of the bias corrected F-test statistic nearly perfectly agrees with the F distribution.

5.2. Simulation over an entire brain

We simulate a whole brain fMRI dataset, in which the experiment design, timings and size are identical to those of the real fMRI dataset in Section 6. An HRF profile is extracted from a voxel which shows the strongest responses in the real dataset.

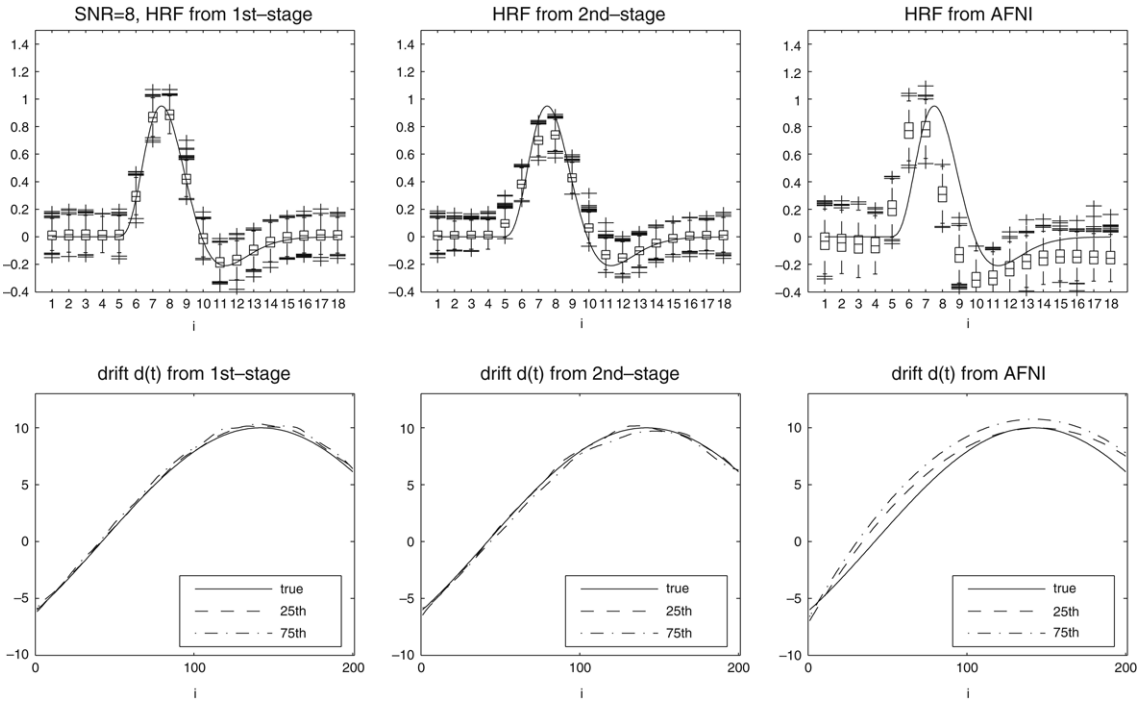

Fig. 3. The captions are the same as those in Fig. 2, except that SNR = 8.

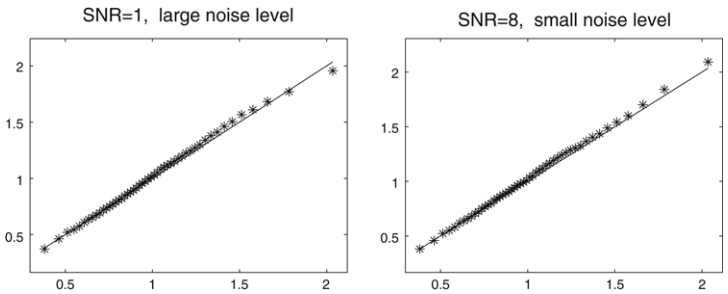

Fig. 4. Empirical quantiles (star) of the bias corrected pseudo F-test statistic versus quantiles (solid line) of $F_{m,n-m}$ distribution.

For each voxel, the simulated drift is obtained from an adequate smoothing to the time series for the corresponding voxel of the real dataset. The simulated noise variance profile is determined from a variance map, which is made by a $5 \times 5 \times 5$ spatial median smoothing on median values of squared residuals of the real time series subtracting the simulated drift profile as mentioned before. The noise process $\epsilon(t_i)$ is generated in a fashion similar to that of Section 5.1. Specifically, the variances of $\epsilon_1(t_i)$ and $z(t_i)$ are set equal such that $\text{var}\{\epsilon(t_i)\}$ is one fifth of the variance map. The HRF profile, in accordance with the stimuli in the experiment, are added to two regions which are postulated to be truly active. In those zones, the HRFs have been re-scaled to about 17% and 12% of the amplitude of the original HRF profiles so that the estimation of the HRF is more challenging.

We compare the activated regions detected using our method with popular imaging tools AFNI (at http://afni.nimh.nih.gov/afni/) (Cox, 1996) and FSL (at http://www.fmrib.ox.ac.uk/fsl/) (Smith et al., 2004; Woolrich et al., 2001). Particularly, we specify the HRF in FSL as the difference of two gamma functions, the canonical form implemented in FSL, and specify the drift term in AFNI as a quadratic polynomial. The FDR level .05 is used for performing the multiple comparison. Fig. 5 highlights the true activated regions. From Figs. 6–8, both AFNI and FSL fail to locate an activated brain area, and that the other region, though correctly detected, has appreciably reduced size relative to the actual size. Our method has the best detection of both activated regions.

6. Real data application

In an emotional control study, subjects saw a series of negative or positive emotional images, and were asked to either suppress or enhance their emotional response to the image, or to simply attend to the image. So there are 6 types of

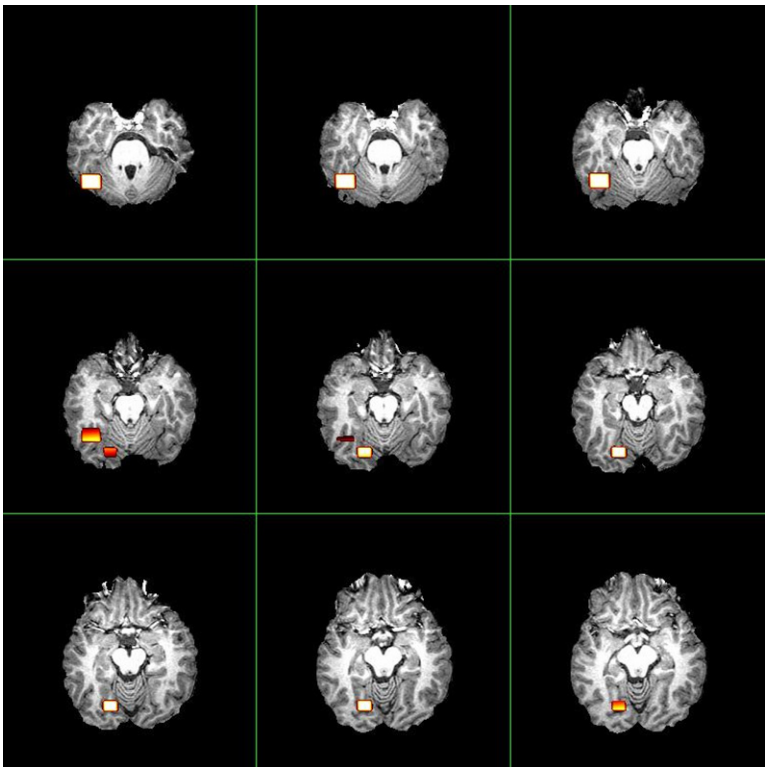

Fig. 5. True activated regions (denoted by hot color) for the simulated fMRI dataset.

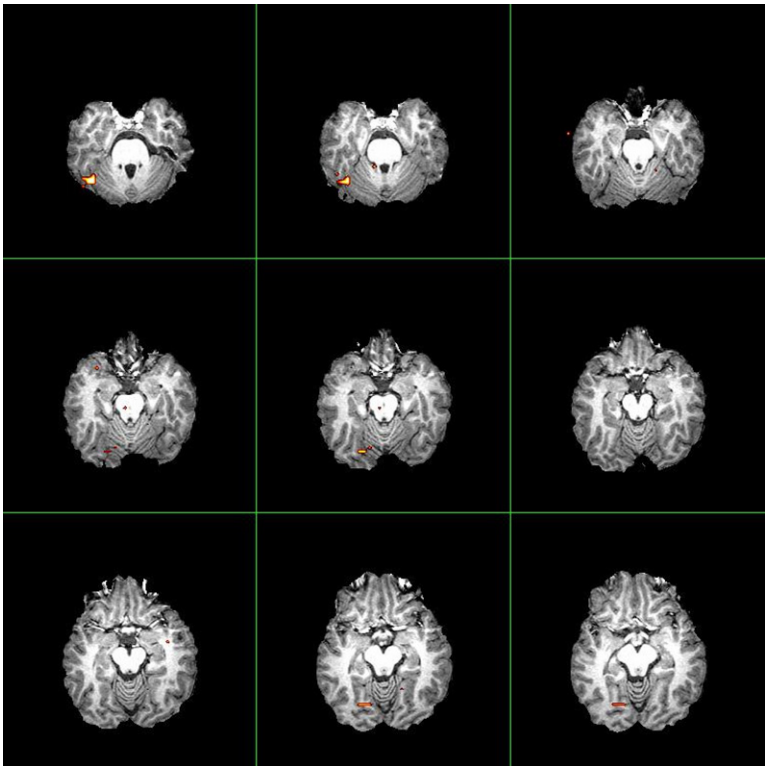

Fig. 6. Activated regions discovered by AFNI for the simulated fMRI dataset.

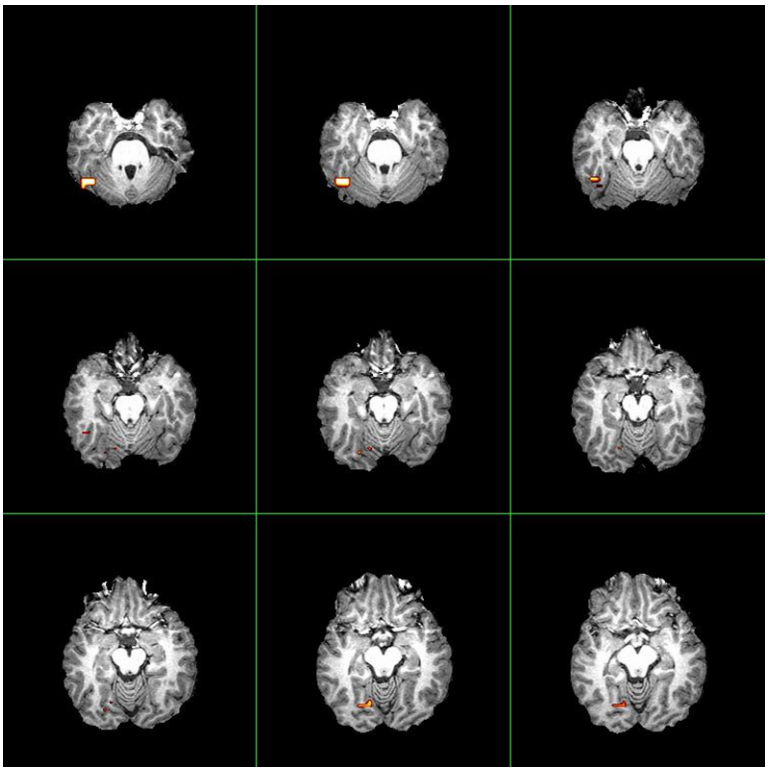

Fig. 7. Activated regions discovered by FSL for the simulated fMRI dataset.

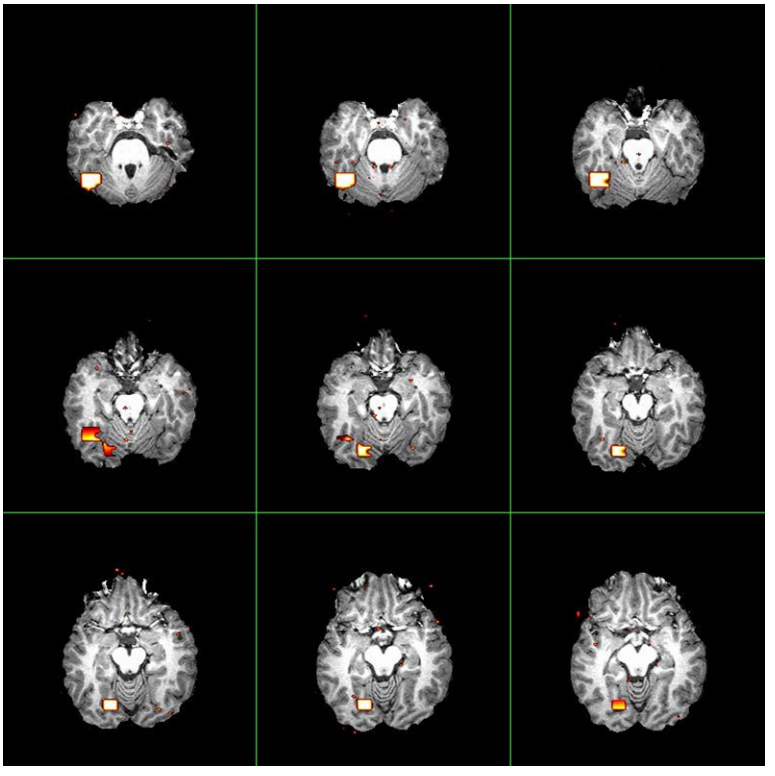

Fig. 8. Activated regions discovered by our Method for the simulated fMRI dataset.

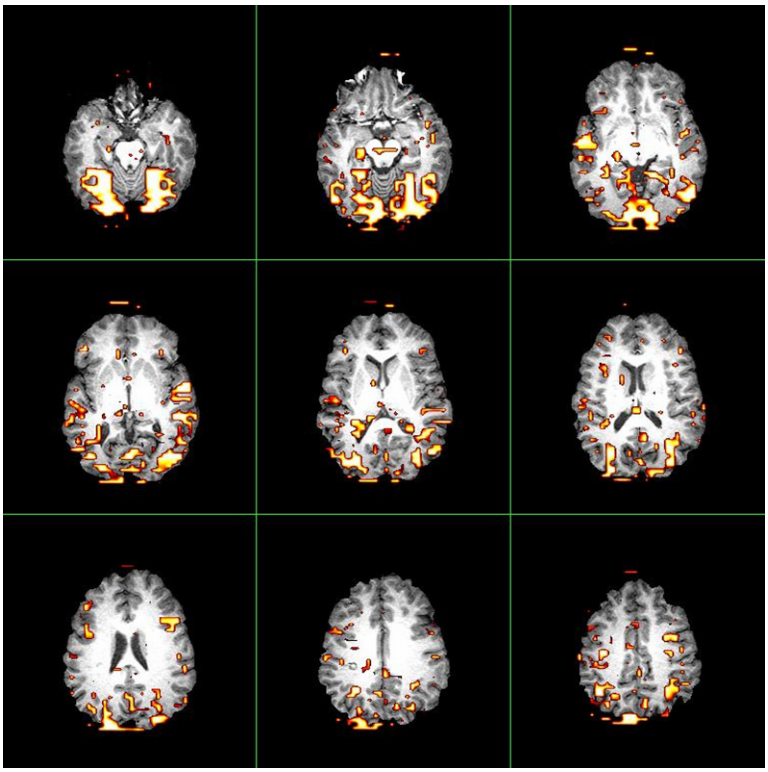

Fig. 9. Activated regions discovered by AFNI for the real fMRI dataset.

events: negative-enhance, negative-attend, negative-suppress, positive-enhance, positive-attend and positive-suppress. The sequence of trials was randomized. The time between successive trials also varied.

The size of the whole brain dataset is $64 \times 64 \times 30$. At each voxel, the time series has 6 runs, each containing 185 observations with a time resolution of 2 s, thus TR $= 2$ s and the total length is 1110. In contrast, the length of stimuli is 2220; the timing of the stimuli has a time resolution of 1 s, and thus each HRF output will also be sampled at 1 s. The study aims to estimate the BOLD response to each of the trial types for 1–18 s following the image onset. We analyze the fMRI dataset containing one subject. The length of the estimated HRF is set equal to 18.

6.1. Description of preprocessing and postprocessing

The pre-processing prior to statistical analysis was conducted on all images using FSL. Motion correction using MCFLIRT (part of FSL) was applied to the images time series. Following this, an automated Brain Extraction Tool (part of FSL) was utilized to remove the skull.

After the statistical analysis, the FLIRT tool (part of FSL) was applied to register the low resolution fMRI image time series onto the high resolution T1-weighted image. The obtained spatial transform was then applied to the image of test statistics.

6.2. Detection of activated regions

Figs. 9–11 compare detected regions using our method with those using AFNI and FSL. Again we specify the HRF in FSL as the difference of two gamma functions and specify the drift in AFNI as a quadratic polynomial. We use FDR level .001 to carry out the multiple comparison. Regions detected by our method are closer to those afforded by AFNI, but our method finds activation in more clustered regions of the brain. AFNI gives more tiny scattered findings, which are more likely to be false discoveries. FSL detects very scattered regions which are difficult to interpret. Additionally, the volumes of the detected regions by FSL are substantially smaller than those of AFNI and our method.

7. Discussion

We have presented new methodologies for the statistical analysis of efMRI data from two challenging aspects. From the modeling aspect, a difference-based approach for estimating the large noise covariance matrix is proposed. This is particularly useful for modeling massive datasets where computational effectiveness and simplicity are highly demanded.

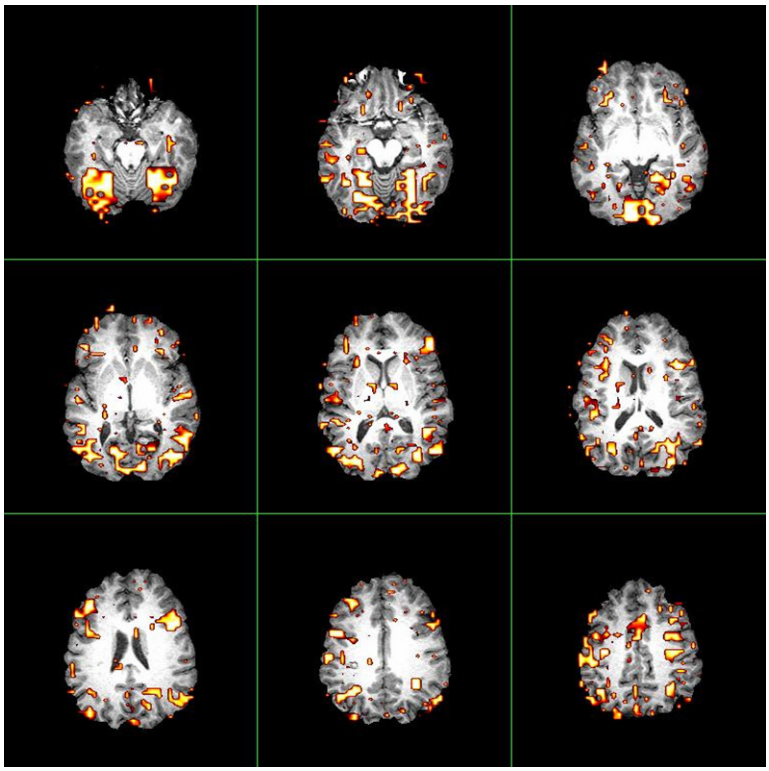

Fig. 10. Activated regions discovered by FSL for the real fMRI dataset.

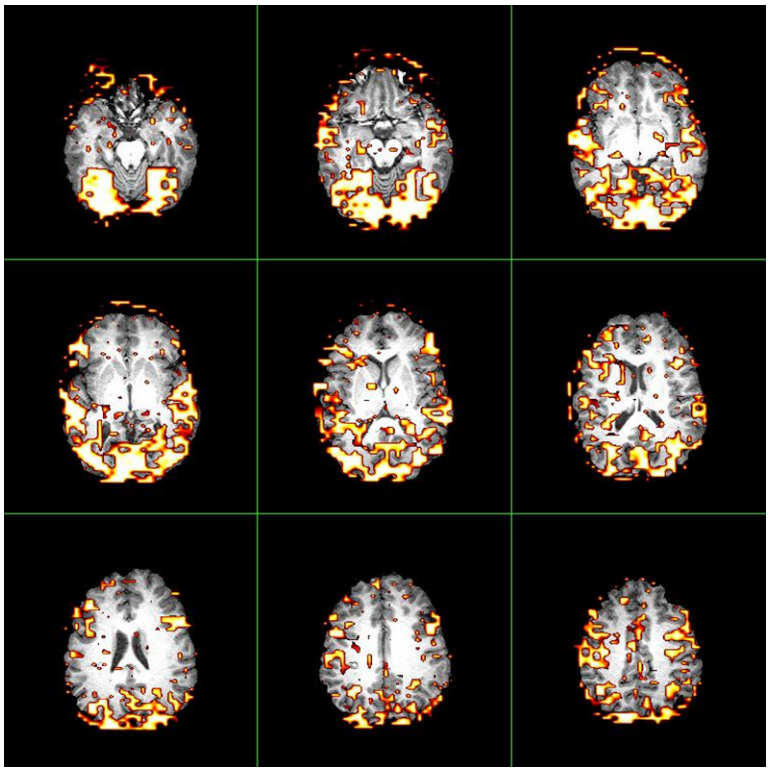

Fig. 11. Activated regions discovered by our method for the real fMRI dataset.

From the inference part, we present a two-stage estimation of the HRF, followed by a nonparametric hypothesis testing of the HRF. This is particularly useful for determining whether a voxel is inactivated or not.

In closing, several points bear mentioning. First, for the sake of illustration, we concentrate on the Frequentist approach for modeling and inference of the HRF nonparametrically. Other innovative work using Bayesian non- and semi-parametric methods includes Genovese (2000), Fahrmeir and Gössl (2002), Ciuciu et al. (2003) and Marrelec et al. (2003). Compared with the methodology using the Bayesian paradigm, our approach does not need to choose the prior distribution (which is not always easy to be selected in practice) for model parameters and the computational load is reduced. A thoroughly detailed comparison between the Frequentist and Bayesian approaches for brain mapping is beyond the scope of the current paper and we hope to report in future study. Second, Zhang et al. (2006) proposed a new multiple comparison procedure, called FDRm. Simulation studies suggest that FDRm better accommodates the spatially dependent information of voxel-wise p-values than the conventional FDR procedure. (Following the suggestions of the associate editor and the referee, illustration and application of FDRm to fMRI datasets are removed from the paper.)

Acknowledgements

We thank Andy Alexander, Kjell Doksum, Kam-Wah Tsui and Zhengjun Zhang for helpful comments. The research is supported in part by National Science Foundation grants DMS-03-53941, DMS-07-05209 and Wisconsin Alumni Research Foundation. The authors are grateful to the Editor, the Associate Editor and an anonymous referee for insightful comments and suggestions.

Appendix. List of notations and symbols

- n: the length of the time series.
- m: length of $\mathbf{h}_j$ for the jth stimulus.
- r: the number of types of stimuli.
- $\mathbf{y}$: $n \times 1$ vector for fMRI signals.
- $s(\cdot)$: the input stimulus function.
- $s_j(\cdot)$: the jth input stimulus function, $j = 1, \ldots, r$.
- $\mathbf{S} = [\mathbf{S}_1, \ldots, \mathbf{S}_r]$: $n \times rm$ matrix.
- $\mathbf{h} = [\mathbf{h}_1^{\mathrm{T}}, \ldots, \mathbf{h}_r^{\mathrm{T}}]^{\mathrm{T}}$: $rm \times 1$ vector for HRFs.
- $\mathbf{d}$: $n \times 1$ vector for drift components.
- $t_i = i/n$, $i = 1, \ldots, n$.
- $\gamma(j)$: auto-covariance of $\{\epsilon(t_i)\}$.
  $\rho(j)$: auto-correlation of $\{\epsilon(t_i)\}$.
  $\gamma_e(j)$: auto-covariance of $\{e(t_i)\}$.
- $\boldsymbol{\epsilon}$: $n \times 1$ vector for errors, and $\mathrm{var}(\boldsymbol{\epsilon}) = \sigma^2 R_n$.
- $R_n = (\rho(|i-j|))$: $n \times n$, the true correlation matrix of $\boldsymbol{\epsilon}$.
  $V_n = R_n^{-1}$.
- S_d: $n \times n$, smoothing matrix associated with the design points $\{t_1, \ldots, t_m\}$ for estimating drift vector $\mathbf{d}$.
- S_h: $m \times m$, smoothing matrix associated with the design points $\{t_1, \ldots, t_m\}$ for estimating HRF $\mathbf{h}$.
- $\widetilde{\mathbf{y}} = (\mathbf{I} - S_d)\mathbf{y}$, and $\widetilde{\mathbf{S}} = (\mathbf{I} - S_d)\mathbf{S}$.
- $K(\cdot)$: a kernel function.
- $\mathsf{k} = \mathrm{rank}(A)$: where the matrix A is $\mathsf{k} \times rm$ and $\mathsf{k} \leq rm$.

References

Ciuciu, P., Poline, J.B., Marrelec, G., Idier, J., Pallier, C., Benali, H., 2003. Unsupervised robust nonparametric estimation of the hemodynamic response function for any fMRI experiment. IEEE Trans. Med. Imaging 22, 1235–1251.
Cox, R.W., 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Fahrmeir, L., Gössl, C., 2002. Semiparametric Bayesian models for human brain mapping. Stat. Model. 2, 235–250.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and Its Applications. Chapman and Hall, London.
Fan, J., Zhang, C.M., Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. Ann. Statist. 29, 153–193.
Fan, J., Zhang, C.M., 2003. A reexamination of diffusion estimators with applications to financial model validation. J. Amer. Statist. Assoc. 98, 118–134.
Friston, K.J., Zarahn, E., Josephs, O., Henson, R.N.A., Dale, A.M., 1999. Stochastic designs in event-related fMRI. NeuroImage 607–619.
Genovese, C.R., 2000. A Bayesian time-course model for functional magnetic resonance imaging data (with discussion). J. Amer. Statist. Assoc. 95, 691–703.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.
Gobbini, M.I., Leibenluft, E., Santiago, N., Haxby, J.V., 2004. Social and emotional attachments in the neural representation of faces. NeuroImage 22, 1628–1635.
Goutte, C., Nielsen, F.A., Hansen, L.K., 2000. Modeling the haemodynamic response in fMRI using smooth FIR filters. IEEE Trans. Med. Imaging 19, 1188–1201.
Josephs, O., Henson, R.N.A., 1999. Event-related functional magnetic resonanceimaging: Modelling, inference and optimization. Philosophical Transactions of the Royal Society 354, 1215–1228.
Lange, N., Zeger, S.L., 1997. Non-linear Fourier times series analysis for human brain mapping by functional magnetic resonance imaging. Appl. Statist. 46, 1–29.

Marrelec, G., Benali, H., Ciuciu, P., Pélégrini-Issac, M., Poline, J.-B., 2003. Robust Bayesian estimation of the hemodynamic response function in event-related BOLD MRI using basic physiological information. Hum. Brain Mapp. 19, 1–17.
Nichols, T., Hayasaka, S., 2003. Controlling the familywise error rate in functional neuroimaging: A comparative review. Statistical Methods in Medical Research 12, 419–446.
Pacifico, M.P., Genovese, C., Verdinelli, I., Wasserman, L., 2004. False discovery control for random fields. J. Amer. Statist. Assoc. 99, 1002–1014.
Purdon, P.L., Solo, V., Weissko, R.M., Brown, E., 2001. Locally regularized spatiotemporal modeling and model comparison for functional MRI. NeuroImage 14, 912–923.
Rosen, B.R., Buckner, R.L., Dale, A.M., 1998. Event-related functional MRI: Past, present, and future. Proc. Natl. Acad. Sci. USA 95, 773–780.
Sen, P.K., 1968. Asymptotic normality of sample quantiles for m-dependent processes. Ann. Math. Statist. 39, 1724–1730.
Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208–S219.
Woolrich, M.W., Ripley, B.D., Brady, M., Smith, S.M., 2001. Temporal autocorrelation in univariate linear modelling of FMRI data. NeuroImage 14, 1370–1386.
Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.
Zhang, C.M., 2003. Calibrating the degrees of freedom for automatic data smoothing and effective curve checking. J. Amer. Statist. Assoc. 98, 609–628.
Zhang, C.M., Lu, Y., Johnstone, T., Oakes, T., Davidson, R., 2006. Efficient modeling and inference for event-related fMRI data, Technical report #1125, Department of Statistics, University of Wisconsin, Madison.