HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2018 February 15.

Published in final edited form as:

Neuroimage. 2017 February 15; 147: 390–408. doi:10.1016/j.neuroimage.2016.12.045.

Comparison of fMRI analysis methods for heterogeneous BOLD responses in block design studies

Jia Liu[#1], **Ben A Duffy**[#1], **David Bernal-Casas**[1], **Zhongnan Fang**[1,2], and **Jin Hyung Lee**[1,2,3,4,*]

[1]Department of Neurology & Neurological Sciences, Stanford University, Stanford, CA 94305, USA

[2]Department of Electrical Engineering, Stanford University, Stanford, CA 94305

[3]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

[4]Department of Neurosurgery, Stanford University, Stanford, CA 94305, USA

[#] These authors contributed equally to this work.

Abstract

A large number of fMRI studies have shown that the temporal dynamics of evoked BOLD responses can be highly heterogeneous. Failing to model heterogeneous responses in statistical analysis can lead to significant errors in signal detection and characterization and alter the neurobiological interpretation. However, to date it is not clear that, out of a large number of options, which methods are robust against variability in the temporal dynamics of BOLD responses in block-design studies. Here, we used rodent optogenetic fMRI data with heterogeneous BOLD responses and simulations guided by experimental data as a means to investigate different analysis methods' performance against heterogeneous BOLD responses. Evaluations are carried out within the general linear model (GLM) framework and consist of standard basis sets as well as independent component analysis (ICA). Analyses show that, in the presence of heterogeneous BOLD responses, conventionally used GLM with a canonical basis set leads to considerable errors in the detection and characterization of BOLD responses. Our results suggest that the 3rd and 4th order gamma basis sets, the 7th to 9th order finite impulse response (FIR) basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets are optimal for good balance between detection and characterization, while the 1st order Fourier basis set (coherence analysis) used in our earlier studies show good detection capability. ICA has mostly good detection and characterization capabilities, but detects a large volume of spurious activation with the control fMRI data.

Keywords

hemodynamic response; block design; fMRI data analysis; optogenetic fMRI

[*]Corresponding Author: Jin Hyung Lee, PhD, ljinhy@stanford.edu, 1201 Welch Road, #P206, Stanford, CA 94305.

1. Introduction

Reliable detection of evoked blood oxygenation level dependent (BOLD) responses is critical to estimate the brain activation maps in fMRI studies. In addition, there has been an increasing interest in characterizing temporal features such as onset and duration to investigate activation timing of BOLD responses across brain regions and experimental conditions (Byers et al., 2015; Handwerker et al., 2012; Lindquist et al., 2009; Liu et al., 2015; Weitz et al., 2014). However, accurate detection and characterization remain challenging in scenarios where BOLD responses exhibit a large variability in the temporal dynamics (Aguirre et al., 1998; Gonzalez-Castillo et al., 2012; Handwerker et al., 2004), such as in studies of disease states (Amemiya et al., 2012; Matthews et al., 2006), and in small animal studies with anesthesia (Schlegel et al., 2015; Schroeter et al., 2014; Williams et al., 2010). In these cases, commonly used general linear model (GLM) (Friston et al., 1994) with a canonical hemodynamic response function (HRF) is often not the best choice. For example, in an fMRI study of motor control in human ischemic patients, GLM with a canonical HRF failed to detect motor cortex activation (Amemiya et al., 2012). It also failed to estimate temporal features of the BOLD responses (Calhoun et al., 2004; Lindquist et al., 2009). In these studies, onset and duration differences between experimental conditions were misinterpreted as differences in the amplitudes of evoked BOLD responses. These substantial detection and characterization errors stress the importance of proper choice of analysis methods.

Nevertheless, it is currently not clear which methods are optimal in scenarios of heterogeneous BOLD responses. This is partially due to the large set of analysis approaches available, yet few comprehensive evaluations have been conducted, especially in block-design studies. Over the past decades, dozens of methods have been proposed. Among the most accessible ones are those implemented in widely available software packages, such as GLM with the canonical basis set (Calhoun et al., 2004; Friston et al., 1998; Henson et al., 2002; Steffener et al., 2010), the gamma basis set, the Fourier basis set, the finite impulse response (FIR) basis set, and the B-spline basis set (Genovese, 2000). Likewise, optimized methods for specific datasets have been considered. For example, colleagues have developed specific basis sets to estimate onset delays (Liao et al., 2002), implemented transient plus sustained models to detect transient responses in block-design experiments (Giraud et al., 2000; Harms and Melcher, 2002; Seifritz et al., 2002), and designed basis sets that incorporate prior information of BOLD responses (Woolrich et al., 2004). Additionally, data-driven methods are employed as they place few assumptions on the hemodynamic responses. Commonly used methods include independent component analysis (ICA) (Beckmann and Smith, 2004; Esposito et al., 2002; McKeown et al., 1998a; McKeown et al., 1998b), principal component analysis (PCA) (Backfrieder et al., 1996; Sychra et al., 1994), and fuzzy clustering analysis (Baumgartner et al., 2000; Chuang et al., 1999; Wismüller et al., 2002).

In block-design studies, only data-driven methods, such as ICA, PCA, and unsupervised clustering, have been compared on their detection and characterization performance (Baumgartner et al., 2000; Erhardt et al., 2011; Meyer-Baese et al., 2004), but not the more widely-used model-based approaches. In contrast, another study assessed several HRF

models' ability to estimate HRF parameters from a block-design experiment, but did not examine detection performance (Shan et al., 2014). More often, comparisons were not conducted as the main purpose of the study, but to support the introduction of new approaches to analyze fMRI data (Calhoun et al., 2001a; Harms and Melcher, 2003; McKeown et al., 1998b; Moritz et al., 2003), or to highlight the heterogeneity of the observed BOLD responses (Amemiya et al., 2012; Gonzalez-Castillo et al., 2012; Pujol et al., 2009; Schlegel et al., 2015). As a result, it is difficult to derive a comprehensive evaluation from the literature, due to the limited range of statistical methods employed and/or assessment conducted in each study.

Here, we investigate the robustness of six widely available methods against heterogeneous BOLD responses in block-design studies. Given the fact that the vast majority of methods already incorporate information about the shape of evoked hemodynamic responses during the detection stage, we focused not only on each method's detection performance, but also on their characterization power (Degras and Lindquist, 2014; Makni et al., 2008). A detailed comparison of state-of-the-art methods for analyses of heterogeneous BOLD responses is presented. Evaluations are carried out in the GLM framework and include standard basis sets as well as ICA. In order to evaluate each methods' performance against fMRI data with heterogeneous BOLD responses, we use data from a recently published optogenetic fMRI (ofMRI) study of dynamic control of forebrain by central thalamus (Liu et al., 2015). To further validate each method's performance, we also use simulated data with varying temporal dynamics. Advantages and shortcomings of each approach are quantified using receiver operating characteristic (ROC) analysis and root-mean-square error (RMSE) of fit. Together, our results aim to provide practical recommendations on proper methods selection for analyzing block-design fMRI data with heterogeneous BOLD responses.

2. Methods

2.1 fMRI analysis methods

In this study, a set of six different approaches including model-based and data-driven methods was evaluated. The same block-design paradigm was used across methods. It consisted of 30 s baseline, followed by six 60 s cycles, each consisting of 20 s stimulation and 40 s rest, unless otherwise noted. To enable comparison across methods, a single statistical analysis platform is needed. Therefore, the linear regression platform in Statistical Parametric Mapping (SPM, Wellcome Trust Center for Neuroimaging) was employed for statistical analysis. All methods were evaluated by using different sets of regressors within the same GLM framework. The detailed description of each method is included as follows:

(i) The canonical basis set was selected from the SPM toolbox as one of the most commonly used methods. Model orders up to 3 were included in the evaluation. In the present study, GLM with a single canonical HRF as basis function is referred to as the 1^{st} order canonical basis set. GLM with a canonical HRF and its temporal derivative as basis functions is referred to as the 2^{nd} order canonical basis set. GLM with a canonical HRF and its temporal and dispersion derivatives as basis functions is referred to as the 3^{rd} order canonical basis set. The canonical basis functions were first convolved with the stimulation paradigm before being used as regressors for the canonical basis set.

(ii) The gamma basis set was selected from the SPM toolbox as another widely available method. Model orders up to 4 were investigated. Each order includes a set of K gamma functions of increasing dispersions as basis functions, where K denotes the model order. Similar with the canonical basis set, the gamma basis functions were first convolved with the stimulation paradigm before being used as regressors for the gamma basis set.

(iii) The FIR basis set was included as one of the most flexible basis sets. The model order of 3 to 10 was investigated. Each order includes a set of K contiguous boxcar functions, in which the bin width of each boxcar function equals T/K, where K denotes the model order, and T represents the length of each stimulation cycle (60 seconds). For simplicity, only results from the odd numbers (e.g., model order of 3, 5, 7, and 9) are shown in figures. Additionally, we investigated the model order of 20, in which the bin width of each boxcar function equals our image acquisition interval (3 seconds), a common practice when employing the FIR basis set. Unlike the canonical and gamma basis sets, the FIR basis set was not convolved with the stimulation paradigm before being used as regressors.

(iv) The B-spline basis set was selected as another popular analysis method (Genovese, 2000; Schlegel et al., 2015). The model order of 3 to 10 was included in the evaluation. Each order includes a set of K cubic spline functions created using the program 3dDeconvolve in the AFNI software package (Cox, 1996; Ward, 2006), where K denotes the model order. Similar with the FIR basis set, only results from the odd numbers are shown for simplicity (e.g., model order of 3, 5, 7, and 9), and the B-spline basis set was not convolved with the stimulation paradigm before being used as regressors.

(v) The Fourier basis set was selected due to its capability to exploit the periodic nature of the experimental paradigm and evoked responses (Bullmore et al., 1996; Pinto et al., 2016). Model orders up to 5 were investigated. Each order includes a set of K sine and K cosine functions at harmonic frequencies: f_1, 2 f_1, …, K f_1 Hz, where K denotes the model order, and f_1 represents the frequency of repeated stimulation cycles (1/60 Hz). Similar with the FIR and B-spline basis sets, the Fourier basis set was not convolved with the stimulation paradigm before being used as regressors.

It is worth noting that, GLM with the 1st order Fourier basis set is mathematically equivalent with coherence analysis, a frequency-domain analysis method that is widely used in periodic block-design studies (Amemiya et al., 2012; Bandettini et al., 1993; Engel et al., 1997; Lee et al., 2010), including the ofMRI datasets we utilized in the present study (Liu et al., 2015). A coherence value was defined as a ratio of the magnitude of each time series' Fourier transform (F) at the frequency of repeated stimulation cycles (f_1, 1/60 Hz) and the total energy of all frequency components:

$$\text{coherence} = \frac{|F(f_1)|}{\sqrt{\sum_f |F(f)|^2}} \quad (1)$$

According to Engel et al., the coherence value is equivalent to the Pearson's correlation coefficient of the target time series with the best fitted sinusoid waveform at f_1 in the least-

squares sense (Engel et al., 1997). Therefore, coherence analysis is equivalent to GLM with the 1st order Fourier basis set, according to:

$$\beta_0 \cdot \sin(2\pi f_1 t + \theta) = \beta_1 \cdot \sin(2\pi f_1 t) + \beta_2 \cdot cos(2\pi f_1 t) \quad (2)$$

when $\beta_1 = \beta_0 \cdot \cos(\theta)$ and $\beta_2 = \beta_0 \cdot \sin(\theta)$. In Eq. 2, β_0, β_1, and β_2 are the coefficients of the model, t denotes time, and θ represents the phase shift of the best fitted sinusoidal waveform.

(vi) Spatial ICA was chosen as one of the most commonly used data-driven approaches (Calhoun et al., 2001b). GIFT ICA algorithm (Calhoun et al., 2001a) with the Infomax approach (Bell and Sejnowski, 1995) was used to extract the spatially independent components. Since there has been no consensus on the optimal method for estimating the number of independent components, the default setting in the GIFT software package (20 components) was used. After ICA decomposition, the independent components representing the signal of interest, which we refer to as signal components, were selected. In the present study, we assumed that their time courses share similar periodicity as the stimulation paradigm. Selection was achieved using the following two steps.

First, we ranked all components' associated time courses based on their power spectrum. In the study by Moritz et al., the frequency power spectrum of each independent component time courses were ranked by their magnitude contributions at the frequency of repeated stimulation cycles (Moritz et al., 2003). Here, we quantified this ranking by calculating the coherence value for the time series of each component. In addition, we incorporated the time series' Fourier transform magnitude at the second harmonic (f_2, 1/30 Hz) to maximize the separation between signal and noise components (Ngan et al., 2009). Here we refer to the modified coherence value as coherence_m:

$$\text{coherence}_m = \sqrt{\frac{|F(f_1)|^2 + |F(f_2)|^2}{\sum_f |F(f)|^2}} \quad (3)$$

Then, we used hierarchical agglomerative clustering (Johnson, 1967) to separate the extracted components into two groups: one group with signal components and the other group with noise components. Here, we assumed that the signal components exhibited distinctly higher coherence_m values than the noise components. Therefore, if we separated the extracted signal components into two groups based on their coherence_m values, the cluster with higher coherence_m values should predominately contain signal components, while the other cluster with lower coherence_m values should primarily contain noise components. Based on these assumptions, we used hierarchical agglomerative clustering to group the obtained coherence_m values into a hierarchical cluster tree, as shown in the dendrogram in Figure S1. We then cut the hierarchical tree to yield two clusters that have the largest inter-cluster distance. The cluster with higher coherence_m values was used as the group with signal components (Fig. S1). The time series of all the signal components were

employed as a set of regressors, resembling a unified GLM-ICA approach (Hu et al., 2005; Pujol et al., 2009).

The above separation was conducted without a predefined number of signal components and was solely based on each component's coherence_m value. In other words, the number of signal components was determined by the inherent structure of the data, thus avoiding biasing the results by using fixed numbers of signal components. We also did not use a predefined coherence_m threshold during the separation, as such threshold may only be selected in an arbitrary fashion.

2.2 Performance metrics

We evaluated the aforementioned methods based on their detection and characterization performance. For the ofMRI data, since the ground truth is unknown, detection volume and modified ROC curves (Nandy and Cordes, 2003) were used as detection metrics. The modified ROC curve was created by plotting the fraction of detected voxels in each subject with ChR2-containing virus injection (experimental group) against the fraction of detected voxels averaged across subjects with saline injection (control group) at varying thresholds. The fraction was calculated as a ratio of the number of detected voxels and the number of brain-masked voxels in each dataset. To quantify the modified ROC curve, area under the ROC curve (AUC) was calculated using a small fraction of the modified ROC curve (fraction of control positive < 0.05) instead of the entire curve, since this region is more relevant for fMRI analysis (Nandy and Cordes, 2003). For the simulated data, where the ground truth is known, we used true positive rate (TPR), false positive rate (FPR), and AUC as detection metrics. TPR was defined as the percentage of ground truth positive voxels that were correctly detected as activated. FPR was defined as the percentage of simulated noise voxels that were incorrectly detected as activated. ROC was used to characterize TPR and FPR at varying thresholds (Skudlarski et al., 1999). As before, AUC was calculated using a small fraction of the ROC curve (FPR < 0.05) instead of the entire curve.

The characterization performance metrics included temporal parameter estimation accuracy and RMSE of fit. Two standard parameters, onset and duration, were used for temporal parameter estimation. Onset was defined as the time to half-peak (Hunter et al., 2003; Weilke et al., 2001), and duration was defined as full-width at half-peak (Lindquist et al., 2009). These parameters were used to characterize time series without assuming any specific shape. The estimated value was calculated from a single period. For the model-based methods, the temporal structure of the fitted time series is the same over each stimulation cycle. However, it is not the case for ICA, therefore we averaged fitted time series across cycles and estimated values from the averaged period. With the ofMRI data, since the ground truth is not available, we calculated the temporal parameter estimation error as the difference between the estimated value and the value measured from the observed time course. As before, since the temporal structure of the observed time series is not the same over each stimulation cycle, we averaged the observed time series across cycles and measured onset and duration from the averaged period. With the simulated data, since the ground truth is available, we calculated the temporal parameter estimation error as the difference between the estimated value and the ground truth. In addition to the temporal

parameter estimation, RMSE of fit was used to evaluate each method's capability to estimate the time course of the BOLD responses for the ofMRI and simulated data:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (4)$$

where $\hat{y}_i$ is the fitted data and n is the number of time points. With the ofMRI data, since the ground truth is not available, we used observed data for y_i. For the simulated data, we used the ground truth signal for y_i instead.

As described above, AUC and RMSE were calculated differently for the ofMRI datasets and the simulated datasets. Specifically, for the ofMRI datasets, AUC was calculated using the modified ROC curve that plots the fractions of detected voxels with the experimental group against those with the control group, while for the simulated datasets, AUC was calculated using the ROC curve that plots TPR against FPR. Similarly, RMSE was calculated relative to the observed data for the ofMRI datasets, but was calculated relative to the ground truth for the simulated datasets. With such differences, the values of the corresponding performance metrics can be very different in the ofMRI datasets and the simulated datasets, especially for AUC, where up to one order of magnitude difference was observed (Table 1). Therefore, to enable a direct comparison between the performance of each method across ofMRI and simulated datasets, we standardized the AUC and RMSE values according to the following formulas:

$$\text{Standardized } \mathrm{AUC}_i = \frac{\mathrm{AUC}_i - \mathrm{AUC}_{\text{second lowest}}}{\mathrm{AUC}_{\text{second highest}} - \mathrm{AUC}_{\text{second lowest}}} \quad (5)$$

$$\text{Standardized } \mathrm{RMSE}_i = \frac{\mathrm{RMSE}_i - \mathrm{RMSE}_{\text{second lowest}}}{\mathrm{RMSE}_{\text{second highest}} - \mathrm{RMSE}_{\text{second lowest}}} \quad (6)$$

To eliminate the influence of extreme values, the second highest value and the second lowest value were used in the above formulas. In this case, the method with the second highest AUC or RMSE was assigned a value of 1, while the method with the second lowest AUC or RMSE was assigned a value of 0. The method with the highest AUC or RMSE was assigned a value of larger than 1, while the method with the lowest AUC or RMSE was assigned a negative value.

2.3 Image analysis

For the ofMRI datasets, custom written software in MATLAB (MathWorks, Inc.) was used for image reconstruction, motion correction (Fang and Lee, 2013), and registration. The acquired 4D fMRI images were manually registered to a common space using a six degree-of-freedom rigid body transformation. Low-frequency drift was removed by temporal high

pass filtering with a cut-off frequency of 1/128 Hz as implemented in SPM. 5 or 6 ofMRI acquisitions were collected for each subject. After preprocessing, the 4D fMRI images for each subject were normalized to the same scale to account for differences in mean and variance. All the ofMRI acquisitions were then averaged for each subject before statistical analysis. Both single-subject and group-level analyses were conducted during statistical analysis.

For the single-subject analysis, we show results that either do not involve smoothing in the preprocessing step to preserve the distinct hemodynamic responses and enable comparisons at the single-voxel level (D'Esposito et al., 1997; Gazzola and Keysers, 2009), or involve spatial smoothing with a 0.5 mm FWHM Gaussian kernel to increase the signal-to-noise ratio. Throughout the present study, results with the single-subject analysis refer to those that do not involve smoothing in the preprocessing step, unless otherwise noted. Prior to statistical analysis, the correlations within each set of regressors were removed using the SPM implementation of Gram-Schmidt orthogonalization. After generating statistical parametric maps using the linear regression platform in SPM, a threshold was applied to define activated voxels. Commonly used voxel-wise threshold settings were employed. For the non-smoothed data, we applied $p < 0.05$ with Bonferroni correction to control the family-wise error rate (FWER), uncorrected $p < 0.001$, and false discovery rate (FDR) < 0.05 (Benjamini and Hochberg, 1995; Genovese et al., 2002) on all brain-masked voxels. For the spatially-smoothed data, we used $p < 0.05$ with random field theory correction to control the FWER on all brain-masked voxels. Note that here we used voxel-wise inference instead of cluster-wise inference during statistical analysis. This is because with our imaging acquisition parameters, SPM's cluster-wise inferences may yield inflated false positive rate at the first-level analysis (Eklund et al., 2012).

For the group-level analysis, fixed-effects analysis and random-effects analysis were conducted using SPM to show the detected activation maps at the group level. Experimental ofMRI data from 10 subjects were included. We applied a 0.5 mm FWHM Gaussian kernel to spatially smooth the raw data in order to increase the signal-to-noise ratio and ameliorate differences in the inter-subject localization. With the fixed-effects analysis, the time series from each subject were temporally concatenated across different subjects before entering into a first-level analysis using SPM. With the random-effects analysis, regression coefficient estimates from the single-subject analysis for each subject were entered into a second-level analysis, and a full factorial design was used for each method and model order at the second level analysis. The potentially unequal variance of the regression residuals across subjects was accounted for by using the correction algorithm implemented in SPM. With both the fixed-effects and random-effects analyses, because the data was spatially smoothed, the statistical threshold was set to $p < 0.05$ with random field theory correction to control the FWER (Worsley et al., 1992). Similar to the single-subject analysis, here we used voxel-wise inference, as SPM's cluster-wise inference may yield inflated false-positive rates at the group level (Eklund et al., 2016).

Since the group-level analysis requires each method to share the same regressors across different subjects, group ICA was conducted to select a common set of signal components among different subjects. Group ICA was performed based on the self-organizing clustering

method (Esposito et al., 2005). In short, 20 independent components were extracted for each subject using GIFT software package. For a total of 10 subjects in each group, 200 independent components were extracted. Then, a similarity matrix (SM) value was calculated for each pair of the 200 total independent components, resulting in a 200 × 200 matrix. Each value in the similarity matrix was calculated using a weighted sum of the correlation coefficients of each component pair's spatial component maps (CC_s) and their associated time courses (CC_t):

$$\mathrm{SM}(i,j) = \lambda \cdot \mathrm{CC_s}(i,j) + (1 - \lambda) \cdot \mathrm{CC_t}(i,j) \quad (7)$$

In Eq. 7, i and j each represent an independent component. λ was set to 0.5, so that the spatial and temporal correlations were equally weighted. The similarity matrix was then converted into a dissimilarity matrix (DM) according to:

$$\mathrm{DM}(i,j) = \sqrt{1 - \mathrm{SM}(i,j)} \quad (8)$$

Based on the resulting DM matrix, we invoked a supervised hierarchical clustering algorithm, which links components to each other only if they were from different subjects. In this way, similar components in different subjects were clustered into the same group, yielding a total of 20 groups, where each group contains 10 components, and each component originated from a different subject. The mean time courses from each group were ranked based on their $\mathrm{coherence_m}$ values, and signal groups were selected using hierarchical agglomerative clustering as described earlier. The mean time series from each of the signal groups were employed as a set of regressors for the fixed-effects and random-effects analyses.

For the simulated datasets, the analysis was conducted similarly to the single-subject analysis for the ofMRI datasets as described earlier. Due to the space limit and scope of the present work, spatial smoothing was not involved in the preprocessing step and group-level analyses were not conducted with the simulated data. As before, prior to statistical analysis, the correlations within each set of regressors were removed using the SPM implementation of Gram-Schmidt orthogonalization. The statistical parametric maps were generated using the linear regression platform implemented in SPM. Commonly used voxel-wise threshold settings were employed to define activated voxels, which include $p < 0.05$ with Bonferroni correction to control the FWER, uncorrected $p < 0.001$, and FDR < 0.05 (Benjamini and Hochberg, 1995; Genovese et al., 2002) on all brain-masked voxels.

2.4 ofMRI data

As mentioned above, we used data from a recently published in vivo ofMRI study (Liu et al., 2015) to evaluate the analysis methods. ofMRI is a novel technique that combines optogenetics with fMRI readouts (Abe et al., 2012; Desai et al., 2011; Kahn et al., 2013; Lee et al., 2016; Lee et al., 2010; Weitz and Lee, 2013). Compared to conventional electrical stimulation or sensory stimulation, ofMRI allows visualization of the causal effects of

specific neuronal populations. It directly stimulates neurons, which eliminates the confounding effects of stimulating many cell types at the same time. The high level of variability in the evoked BOLD responses reported in recent ofMRI studies (Byers et al., 2015; Desai et al., 2011; Duffy et al., 2015; Lee et al., 2016; Lee et al., 2010; Liu et al., 2015; Takata et al., 2015; Weitz et al., 2014) offers an excellent opportunity to assess the robustness of different methods to cope with heterogeneous BOLD responses.

The ofMRI data used in the present study were acquired using a 7 T Bruker Biospec small animal MRI system at UCLA and a 7 T Agilent MR901 horizontal bore scanner at Stanford as described previously (Liu et al., 2015). Briefly, adult female Sprague-Dawley rats (> 11 weeks old) were used. Gradient recalled echo BOLD (TR/TE = 750 ms/12 ms) with a four-interleave spiral readout was used to acquire 23 coronal slices at a 3 s temporal resolution. The in-plane field of view was 35 × 35 mm^2 and the slice direction coverage was 11.5 mm. The data was then reconstructed to a 128 × 128 × 23 matrix.

In the present study, three groups of subjects were included in the ofMRI datasets. (i-ii) The first two groups of subjects are the experimental groups. In these subjects, adeno associated viruses that were engineered to express channelrhodopsin-2 (ChR2) were stereotaxically injected into the central thalamus of each subject. A fiber optic cannula was subsequently implanted for light delivery. (i) In the first experimental group, data from 10 subjects with 10 Hz or 40 Hz optical stimulation in the central thalamus was used. The stimulation paradigm consisted of 30 s baseline, followed by six 60 s cycles, each consisting of 20 s stimulation and 40 s rest. Throughout the present study, the experimental ofMRI data refer to data from the first experimental group, unless otherwise noted. (ii) In the second experimental group, data from one subject with 100 Hz optical stimulation in the central thalamus was used. Here we employed a slightly different stimulation paradigm compared to the first experimental group, in which 10 s of stimulation was applied in each of the six cycles instead of 20 s (Fig. S9A). (iii) The third group of subjects is the control group. In this case, saline was injected into the central thalamus of each subject and a fiber optic cannula was implanted for light delivery. Data from two subjects with 40 Hz optical stimulation in the central thalamus was used. The stimulation paradigm was the same as in the first experimental group.

2.5 Simulated data

Simulated datasets were generated to utilize data with known ground truth. We assumed that a diverse range of signal shapes was evoked using the same six-cycle block design as in the experimental ofMRI data. Each simulated slice was based on a single imaging slice from the experimental ofMRI data during the baseline period. Random, non-physiological system disturbances were modeled by additive Gaussian noise and were added into all the brain-masked voxels. In each slice, signals with the same shape were added into two “active” regions in the cortex and striatum. The activation signal was created by convolving the canonical HRF used in SPM with a boxcar function with a varying onset and duration. The onset shift of the boxcar function was set to vary between 0 and 20 s, time locked to the 20 s stimulation block. The duration range was set to vary between 5 and 50 s, to reflect the transient and prolonged BOLD responses observed in previous studies (Duffy et al., 2015;

Gonzalez-Castillo et al., 2012; Weitz et al., 2014). After convolving each boxcar function with the canonical HRF used in SPM, the resulting time series that did not return to below 50 % maximum amplitude at the end of each cycle were excluded, resulting in 85 shapes, each with a distinct onset and duration. To ensure the generalizability of the results, three different contrast-to-noise ratios (CNR), 1, 1.5, and 2, were used. CNR was calculated as a ratio of the signal amplitude and the standard deviation of the underlying noise in the time domain.

For each method, we summarized the analysis results that were obtained from the simulated datasets with different signal shapes. This was conducted by averaging the analysis results across all signal shapes assuming a uniform distribution, or by calculating weighted average based on the bivariate probability distribution of onset and duration for the 10 Hz and 40 Hz stimulation ofMRI data. The probability distribution was estimated using the onset and duration measured from the observed raw time series in the ofMRI datasets. The superset of voxels detected by each method across all subjects were included. The threshold was set to $p < 0.05$ with Bonferroni correction. To minimize the effect of outliers, the onset and duration range between the 5^{th} and 95^{th} percentile was included. Based on these onset and duration values, a histogram approach was used to estimate the probability distributions. Specifically, we generated a 2D histogram using the following steps. First, we divided the onset and duration values from the ofMRI data into a series of consecutive and non-overlapping bins. The bins were specified based on the onset and duration of each simulated signal shape. Then, we counted how many values fell into each bin. The ratio of the number of values in each bin and the total number of values across all bins was obtained as the probability for the corresponding simulated signal shape.

Concerning ICA, before computing the summary statistics, we defined a set of specific regressors for each distribution of onset and duration. This was different from model-based methods, where the set of regressors was pre-determined. In ICA, to obtain the distribution-specific regressors, additional datasets were generated to depict different distributions. For example, we used spatial concatenation of simulated datasets with different signal shapes to generate the uniform distribution. In this case, each signal shape was present in the same number of active voxels. Conversely, to simulate the 10 Hz and 40 Hz stimulation ofMRI data, we generated datasets with the same data size and “active” regions as those for the uniform distribution; however, unlike the uniform distribution, here each signal shape was present in a different number of active voxels. In particular, the number of active voxels that contained each signal shape was calculated as the product of the total number of active voxels in the datasets and the probability of the signal shape in the 10 Hz or 40 Hz stimulation ofMRI data distribution. As a result, three datasets were generated, each of them depicting a different distribution of onset and duration. In each dataset, we extracted independent components using the GIFT software package. We then selected signal components using coherence_m value plus hierarchical agglomerative clustering, as described earlier in the manuscript. Finally, the signal components were used as a set of distribution-specific regressors. From here, the next steps of the statistical analysis were set to be the same for ICA as for model-based methods with the goal of fairly compare the performance across different methods.

3. Results

3.1 GLM with the 1st order canonical basis set leads to detection and characterization errors with experimental ofMRI datasets

Figure 1 shows the detection and characterization results by GLM with the 1st order canonical basis set using recently published experimental ofMRI data (Liu et al., 2015). Based on previous anatomical and electrophysiological studies, where widespread projections from the stimulation site (i.e., the central thalamus) to the forebrain have been demonstrated (Deschenes et al., 1996; Steriade and Glenn, 1982; Van der Werf et al., 2002), we would expect a large volume of forebrain activities to be detected with the ofMRI data. However, the 1st order canonical basis set detects cortical and thalamic responses with the 10 Hz stimulation ofMRI data, but detects a small volume of responses with the 40 Hz stimulation ofMRI data, in contrast to what was detected using coherence analysis in our previous study (Liu et al., 2015). We then take a close look at the observed BOLD responses detected by the 1st order canonical basis set. There, we see variations in their temporal dynamics across different stimulation frequencies (Fig. 1E-H) and brain regions (Fig. S2A, B). Specifically, at 10 Hz stimulation, the observed BOLD responses (Fig. 1E, G), especially those in the cortex (Fig. S2A), show similar onset and duration as the convolution of the canonical HRF with the experimental paradigm, which henceforth we refer to as canonical response. While at 40 Hz stimulation, the observed BOLD responses show delayed onset and extended duration compared to the canonical response (Fig. 1F, H), especially in the thalamus and striatum (Fig. S2B). However, the 1st order canonical basis set is not able to characterize these diverse temporal features in the ofMRI data (Fig. 1I-L), as its shape is fixed and only its amplitude is allowed to vary (Worsley and Friston, 1995).

The above results obtained by the 1st order canonical basis set demonstrate its inability to cope with highly variable responses. Here, we sought to understand the proper choice of methods in these scenarios. Specifically, we conducted a systematic evaluation to assess a set of six standard methods' capabilities to detect and characterize heterogeneous BOLD responses. These include GLM with the canonical, gamma, FIR, B-spline, and Fourier basis sets, as well as ICA (Fig. 2).

3.2 GLM with the 2nd and 3rd order canonical basis sets, the 2nd to 4th order gamma basis sets, the 5th to 20th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets show good detection performance with the ofMRI datasets

We first examined the detection performance across different methods with the experimental ofMRI data. With the 10 Hz stimulation data, the 2nd and 3rd order canonical basis sets, the 1st to 4th order gamma basis sets, the 5th to 7th order B-spline basis sets, the 2nd order Fourier basis set, and ICA detect significantly greater volumes compared to the 1st order canonical basis set (Fig. 3A). Among the different methods, ICA detects the largest volume. Nonetheless, each method detects similar activations at the stimulation site (i.e., the thalamus) and the downstream brain regions (Fig. 3B). The observed BOLD responses detected by different methods also share similar onset and duration as the canonical response (Fig. 3C, D), especially in the cortex (Fig. S2C). In contrast, with the 40 Hz stimulation data,

the majority of the methods yield significantly larger volumes compared to the 1st order canonical basis set, including the 2nd and 3rd order canonical basis sets, the 2nd to 4th order gamma basis sets, the 3rd to 20th order FIR basis sets, the 3rd to 9th order B-spline basis sets, the 1st to 5th order Fourier basis sets, and ICA (Fig. 3E). Among the different methods, ICA detects the largest volume. Unlike the detection results with the 10 Hz stimulation ofMRI data, not all methods are able to detect responses at the stimulation site and the downstream brain regions during 40 Hz stimulation (Fig. 3F). Specifically, the 1st order canonical basis set detects a very small volume in the thalamus, cortex, and striatum, while most of the other methods detect a large volume in these regions. In addition, with 40 Hz stimulation, the observed BOLD responses, especially those in the cortex, show much higher variations in their temporal dynamics compared to the 10 Hz stimulation ofMRI datasets (Fig. 3G, H, S2D). Most of the methods are able to detect BOLD responses with substantial onset and duration deviations from the canonical response, except for the 1st order canonical basis set (Fig. 3H).

Next, we examined the consistency of these detection results across different conditions. There, detection results with similar trends within and across methods are obtained when: (i) different threshold settings are used (Fig. S3); (ii) raw data is spatially smoothed to increase the signal-to-noise ratio (Fig. S4); and (iii) fixed-effects analysis at the group level is used (Fig. S5). By invoking random-effects analysis at the group level, the brain regions detected at the subject level are similarly detected, although with a smaller volume (Fig. S6). Notably, with random-effects analysis, increasing the model order within each method leads to a larger detection volume with the 10 Hz and 40 Hz stimulation ofMRI datasets. This is because methods with larger numbers of regressors take more contrast images per subject into the second level analysis (Ashburner et al., 2008), resulting in a greater number of total degrees of freedom and very sensitive statistical tests. Specifically, in the case of n subjects and k regressors in the model, the total degrees of freedom are $n \times k - 1$ at the second level analysis. This is different from the single-subject and fixed-effects analyses, where the total degrees of freedom are fixed at $m - 1$ for fMRI data with m time frames. As a result, with random-effects analysis, the differences in the detection results across methods are largely governed by the differences in the total degrees of freedom, rather than each method's capabilities to handle heterogeneous BOLD responses.

We then examine each method's detection performance with the control ofMRI data. In the group of control subjects, saline was injected into the brain instead of ChR2-expressing virus. Therefore, we do not expect any optogenetically-evoked neuronal activity, and assume all detected responses with the control dataset to be false positive. In other words, methods with smaller detection volumes are preferred. As shown in Figure 4A and B, most of the methods detect similar volumes as the 1st order canonical basis set. However, ICA detects a considerably larger volume of spurious activations than any other methods. Similar results are obtained when different threshold settings are used (Fig. S8A-F), and when the control data is spatially smoothed to increase the signal-to-noise ratio (Fig. S8G, H).

To summarize the above-described detection results with experimental and control ofMRI datasets at varying thresholds, we generate modified ROC curves and calculate AUC values for each method (Fig. 4C-F). ROC curves with large areas under the curve and hence high

AUC values are preferred as they indicate the method's ability in yielding a large detection volume with the experimental ofMRI data while maintaining a small amount of spurious detections with the control ofMRI data across different threshold settings. With both 10 Hz and 40 Hz stimulation ofMRI datasets, most of the methods yield significantly higher AUC values compared to the 1st order canonical basis set (Fig. 4C, D). Compared to the detection volume results with the experimental ofMRI data, we observe similar trends within and across methods except with ICA. ICA detects the largest volume with the experimental ofMRI data, but yields one of the lowest AUC values. This is consistent with previous results that ICA detects a larger volume of false positive activations than any other methods with the control ofMRI data. Similar detection results are obtained when using an ofMRI dataset with a different block-design paradigm (Fig. S9E and F). With the ofMRI data, we show that the 2nd and 3rd order canonical basis sets, the 2nd to 4th order gamma basis sets, the 5th to 20th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets yield high AUC values and good detection capabilities (Figure 4C-F, Table 1).

3.3 GLM with the 4th order gamma basis set, the 20th order FIR basis set, the 7th to 9th order B-spline basis sets, and the 3rd to 5th order Fourier basis sets show good characterization performance with the ofMRI datasets

The characterization results of the experimental ofMRI data are shown in Figure 5. First, we examine the characterization errors of each method. With the 10 Hz and 40 Hz stimulation ofMRI datasets, the 4th order gamma basis set, the 20th order FIR basis set, the 7th to 9th order B-spline basis sets, and the 3rd to 5th order Fourier basis sets consistently yield significantly lower onset errors, duration errors, and RMSE compared to the 1st order canonical basis set (Fig. 5A-C), suggesting good characterization performance. As shown in Figure 5C, increasing the model order within each method leads to a decrease in the RMSE. This is because with the ofMRI data, RMSE is calculated relative to the observed time courses, and any increase in the model order within each method naturally leads to a better fit to the observed data. Similar characterization results are obtained when analyzing an ofMRI dataset with a different block-design paradigm (Fig. S9G).

Next, we take a closer look at the onset and duration estimated by different methods at each detected voxel. As shown in Figure 5D, the onset and duration estimated by the 4th order gamma basis set, the 9th order B-spline basis set, and the 5th order Fourier basis set show similar patterns as those measured from the observed data, while the values estimated by the canonical basis set show a clear deviation, suggesting biased estimations. Since each regressor in the 20th order FIR basis set has a bin width of the image acquisition interval (3 seconds), their onset and duration estimates are therefore fixed at integer multiples of 3 seconds and do not show a continuous pattern in Figure 5D. As shown in Figure S1, ICA extracts two signal components in most of the subjects. With these subjects, ICA yields a fixed duration value for each estimated onset value (Fig. 5D), and cannot characterize the differences among BOLD responses that share the same onset but different duration. This is the same with other methods that have two regressors, such as the 2nd order canonical basis set, the 2nd order gamma basis set, and the 1st order Fourier basis set (coherence analysis) (data not shown).

3.4 GLM with the 2nd to 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 1st to 5th order Fourier basis sets, and ICA show good detection performance with the simulated datasets

To further validate the above results obtained with ofMRI data, we conducted the assessment using simulated data with varying onset and duration (Fig. 6A). Figure 6B shows three probability distributions of onset and duration used in the calculation of summarized results. The 10 Hz and 40 Hz stimulation ofMRI data distributions are based on the bivariate probability distributions of onset and duration estimated from the experimental ofMRI data. The uniform distribution is equivalent to averaging across the analysis results from the simulated datasets with different signal shapes.

The detection results of simulated data are summarized in Figure 7. First, we examine the TPR metric (Fig. 7A, B). Unlike other methods, the 1st to 3rd order canonical basis sets and the 1st and 2nd order gamma basis sets can only detect small onset and duration deviations from the canonical response. Across different distributions of onset and duration, the 3rd and 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 1st to 5th order Fourier basis sets, and ICA similarly yield high TPR (Fig. 7B).

Next, we examined the FPR metric. As shown in Figure 7C and D, in each method, false positives are not detected in the majority of the simulation. ICA detects a larger volume of spurious activations than any other methods with the control ofMRI data (Fig. 4A, B), but yields similar FPR as other methods with the simulated data across different threshold settings (Fig. 7C, D, S11A).

To summarize the TPR and FPR metrics at varying thresholds, we generate ROC curves and calculate AUC values for each method. Across different distributions of onset and duration, the 2nd to 4th order gamma basis sets, the 5th to 20th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 1st to 5th order Fourier basis sets, and ICA similarly yield the highest AUC (Fig. 7E, F). Detection results with similar trends within and across methods are obtained when different threshold settings are applied (Fig. S11) and under different CNR levels (Fig. S12), with the exception that at lower CNR levels, the 20th order FIR basis set underperforms relative to other methods (Fig. S12C, Table 1).

3.5 GLM with the 3rd and 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 2nd to 5th order Fourier basis sets, and ICA show good characterization performance with the simulated datasets

The characterization results of the simulated data are shown in Figure 8. First, we examine the estimation errors of onset and duration (Fig. 8A-D). For the canonical basis set, adding the 2nd and 3rd order decreases the onset errors, but at the cost of the duration errors. Across different distributions of onset and duration, the 3rd and 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 2nd to 5th order Fourier basis sets, and ICA similarly yield low onset and duration errors. Note that, four signal components are extracted using ICA with each distribution of onset and duration (Fig. S10A).

Finally, RMSE is used to assess each method's capability to predict the observed BOLD responses. As shown in Fig. 8E and F, the 1st to 3rd order canonical basis sets and the 1st and 2nd order gamma basis sets yield small RMSE near the canonical response, but large RMSE when onset and duration deviate substantially. Across different distributions of onset and duration, the 3rd and 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 1st to 5th order Fourier basis sets, and ICA similarly yield low RMSE. Characterization results with similar trends within and across methods are obtained when analyzing the simulated datasets with different CNR levels (Fig. S13), with the exception that at higher CNR levels, the 1st order Fourier basis set (coherence analysis) underperforms relative to other methods, especially with the 10 Hz stimulation ofMRI data distribution (Table 1).

3.6 GLM with the 3rd and 4th order gamma basis sets, the 7th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets show good balance between detection and characterization

We summarize each method's detection and characterization performance with the ofMRI and simulated datasets. Each method's AUC and RMSE values are averaged across the 10 Hz and 40 Hz stimulation ofMRI datasets for the real data, and across different CNR levels and distributions of onset and duration for the simulated data (Table 1). As shown in Figure 9, most of the methods show superior detection and characterization capabilities compared to conventionally used 1st order canonical basis set. To identify the methods that perform well with both the ofMRI and simulated datasets, we standardized the mean AUC and RMSE of each method shown in Figure 9A and B, and compared the standardized values across the ofMRI and simulated datasets. As shown in Figure 9C, most of the methods exhibit similar detection and characterization performances between the ofMRI and simulated datasets, except for the 20th order FIR basis set and ICA. The 20th order FIR basis set yields low RMSE relative to the observed data with the ofMRI datasets, but exhibits high RMSE relative to the ground truth with the simulated datasets. This is possibly because the 20th order FIR basis set overfits the observed data. ICA yields a high AUC value with the simulated datasets but shows a low AUC value with the ofMRI datasets, as it detects a larger volume of false positive activations than other methods with the control ofMRI data. In summary, with both ofMRI and simulated datasets, we show that the 3rd and 4th order gamma basis sets, the 7th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets exhibit superior detection and characterization performance over other methods.

4. Discussion

The large number of existing analysis approaches necessitates a comprehensive assessment to ease the selection of methods in scenarios with heterogeneous BOLD responses, yet none have been performed thoroughly with a block-design paradigm. In the present work, we address this issue by systematically evaluating a series of standard analysis methods using rodent ofMRI data (Liu et al., 2015) and simulations with a block-design paradigm. We find that, conventionally used GLM with a canonical basis set leads to considerable detection and characterization errors in the presence of heterogeneous BOLD responses. GLM with the 3rd

and 4th order gamma basis sets, the 7th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets are the optimal methods as they offer good balance between detection and characterization. GLM with the 1st order Fourier basis set (coherence analysis) used in our earlier studies shows good detection capability. ICA shows good detection and characterization performance with the simulated data, but detects a large volume of spurious activations with the control ofMRI data.

Our study aims to evaluate the performance of various methods and strives to provide recommendations for method selection primarily in the analysis of animal fMRI data. In small animal fMRI studies, variations in the temporal dynamics of the hemodynamic responses can be more severe than those in human studies. This is likely due to the use of anesthesia to minimize motion and restraint stress (Schlegel et al., 2015; Schroeter et al., 2014; Williams et al., 2010), the use of disease models (Nersesyan et al., 2004; Weber et al., 2008), and the use of direct brain stimulation (Angenstein et al., 2009; Byers et al., 2015; Duffy et al., 2015; Ferenczi et al., 2016; Lee et al., 2016; Lee et al., 2010; Liu et al., 2015; Weitz et al., 2014) in small animal studies. Under these conditions, evoked BOLD responses can differ considerably from the canonical response. For example, in urethane-anesthetized mice, a 20 s electrical stimulation in the hind paw causes the fMRI responses in the thalamus to be delayed by approximately 11 s compared to the canonical response (Schroeter et al., 2014). In another study with isoflurane-anesthetized rats, optogenetic stimulation of the hippocampus for 20 s evokes BOLD responses that last for over 50 s (Weitz et al., 2014). In such cases, employing the analysis approaches recommended in the present study may be beneficial to accurately detect and characterize these heterogeneous responses.

Notably, in animal ofMRI studies, analysis methods that can accurately detect and characterize evoked BOLD responses are particularly useful in the identification of stimulation-related artifacts. During ofMRI experiments, stray light and local tissue heating associated with the optical stimulation can introduce undesirable artifacts to the data (Christie et al., 2013; Schmid et al., 2016), hence it is important to identify contaminated data during the analysis process. Specifically, artifacts caused by stray light are typically detected in vision-related brain regions (Schmid et al., 2016), whereas optogenetically-evoked responses are usually absent there when the stimulation target is outside of the visual pathway (Byers et al., 2015; Desai et al., 2011; Duffy et al., 2015; Lee et al., 2016; Lee et al., 2010; Liu et al., 2015; Takata et al., 2015; Weitz et al., 2014). As a result, accurate detection of BOLD responses in vision-related regions is essential to identify data contaminated with these artifacts. Conversely, at the stimulation site, artifacts caused by local tissue heating usually exhibit both positive and negative fMRI signal changes (Christie et al., 2013), whereas optogenetically-evoked responses usually show only one type of polarity (Duffy et al., 2015; Lee et al., 2010; Liu et al., 2015; Takata et al., 2015; Weitz et al., 2014). As a result, accurate characterization of BOLD responses at the stimulation site is critical to identify data contaminated with heating artifacts. In both cases, employing the methods that show good detection and characterization performance in the present study may be advantageous in the analysis process.

Beyond providing guidelines for animal studies, our results may also be useful for analyzing block-design fMRI data in human studies, especially those with highly variable responses.

For example, in ischemic patients, a 24 s hand-grasping task causes the BOLD responses in the primary motor cortex to be delayed by up to 24 s compared to those in the cerebellum (Amemiya et al., 2012). In patients with major arterial stenosis, a 10 s handball squeeze task causes the BOLD responses in the primary motor cortex to be delayed by approximately 3 s compared to those in healthy individuals (Roc et al., 2006). Occasionally, heterogeneous BOLD responses can also be observed in healthy human subjects. For example, a 9 s painful mechanical stimuli causes the BOLD responses in the somatosensory cortex to persist for approximately 18 s (Pujol et al., 2009). After averaging a large number of trials for each subject, a 20 s visual stimulation plus attention control task leads to highly variable BOLD responses throughout the brain (Gonzalez-Castillo et al., 2012). In such cases, adopting the methods recommended in the present work may also be useful to reduce errors in the detection and characterization tasks.

Notably, our results, which are based on rodent ofMRI data and simulated data guided by experimental data, are in good agreement with published data on human block-design studies. For example, GLM with the 1st order canonical basis set detected a smaller volume of hemodynamic responses compared to ICA in patients with fibromyalgia when painful stimuli were applied (Pujol et al., 2009). In a like manner, during psychomotor tasks in healthy subjects, GLM with the 1st order canonical basis set did not uncover BOLD responses in the frontal regions that were detected by ICA (McKeown et al., 1998b). Consistently with our results, when the elicited BOLD responses were less heterogeneous during a simple visual paradigm in healthy subjects, GLM with the 1st order canonical basis set and ICA delivered comparable detection results (Calhoun et al., 2001a).

Nonetheless, it should be noted that there exist several differences in regards to the characterization performed here compared to the commonly conducted ones in event-related studies. First, instead of characterizing impulse response functions, we focused on characterizing responses evoked by blocks of stimulation, which we refer to as block responses, although the impulse response function characterizations are more widely investigated (Lindquist et al., 2009; Shan et al., 2014). While block-design studies are not optimal for modeling impulse response functions (Maus et al., 2012), characterizing block responses can often offer great insights. For example, delayed occipital block responses during face encoding task may serve as an early marker for Alzheimer's disease in human (Rombouts et al., 2005). In rodent ofMRI studies, prolonged block responses could be indicative of seizure-like afterdischarge activities in the hippocampus (Weitz et al., 2014). These variations in the block responses could originate from various sources including the timing of neuronal activity, the impulse response function, and nonlinearities. Nonetheless, accurate detection and characterization of these block response variations provide an important first step towards in-depth investigations of the underlying mechanism. Second, we did not impose constraints on the basis sets, although it is recommended when modeling impulse response functions in order to avoid physiologically ambiguous or implausible shapes (Calhoun et al., 2004; Steffener et al., 2010; Woolrich et al., 2004). This is because a larger degree of variations may be observed in block responses compared to impulse response functions. For example, impulse response functions are usually considered to be physiological meaningful if they only have one peak (Calhoun et al., 2004), or exhibit a sensible range of time-to-peak values (Henson et al., 2002). However, because of the

transient neuronal activities at block transitions (the beginning and end of each task-block), physiological meaningful bimodal block responses have been observed in humans (Marxen et al., 2012). Block responses from the ofMRI datasets used here also exhibit a wide range of onset shifts relative to the canonical response. Therefore, we did not impose constraints on the basis sets.

The present work demonstrates advantages of flexible models in achieving better detection and characterization performances, but the analysis results should be interpreted with caution. First, using flexible models may lead to the detection of physiologically implausible responses. To reduce such undesirable errors, it may be helpful to examine the temporal dynamics of detected BOLD responses in each region of interest. Special attention may be paid to unexpected activation and prior knowledge of anatomical and functional connectivity could be used to assess whether detected responses are physiologically plausible. Electrophysiological experiments, such as in vivo extracellular recordings, may also help to confirm the neural origin of detected responses. Second, the regression coefficients estimated with flexible models may lack interpretability, in contrast to conventionally used 1^{st} order canonical basis set where the regression coefficient typically reflects the amplitude of the BOLD responses. For example, in the case of the Fourier basis set, it is difficult to interpret the regression coefficients of higher harmonics, as they only represent the refinement of model fitting, and are not interpretable features of the BOLD responses, such as amplitude, onset, and duration.

In this study, systematic evaluations of various statistical methods are presented, but a few caveats exist. First, the simulated data was generated by adding Gaussian noise to the ground truth signal, which did not account for any other type of physiological noise that may be present in the real data (e.g., colored noise). Adding physiological noise may affect the TPR, FPR, and AUC results (Welvaert and Rosseel, 2012). Second, the simulated datasets did not include other types of BOLD responses, such as nonstationary or biphasic responses (Fox et al., 2005a; Fox et al., 2005b; Gonzalez-Castillo et al., 2012; Harms and Melcher, 2002; Uludag, 2008). Third, we evaluated a variety of analysis methods on their simultaneous detection and characterization capabilities, but it may be possible to achieve higher accuracies using advanced statistical methods, such as two-gamma-variate fitting (Yu et al., 2016) and exponential fitting (Bosshard et al., 2015). Future comparisons including advanced methods may be useful. Fourth, it is worth noting that, there exist many metrics to compare methods besides those employed in the present work. Among the commonly used ones are the Akaike Information Criterion (Akaike, 1998), the Bayesian Information Criterion (Schwarz, 1978), and the Bayesian model selection (Wasserman, 2000). In this work, we focused on two specific applications of fMRI analysis – detection and characterization of the BOLD responses – utilizing application-specific performance metrics, such as AUC for detection and RMSE of fit for characterization. Fifth, the method employed to select signal components assumes the existence of signals of interest in the datasets. However, if used on data without signals of interest, such as the control ofMRI data, this method would yield a group of ‘signal’ components that are predominantly noise (Fig. S7). Employing such ‘signal’ components during data analysis may lead to undesirable consequences, such as the detection of spurious activations, as shown in Section 3.2. With the purpose of comparing the performance against other methods, we employed these

'signal' components from the control ofMRI data. However, when using ICA for data analysis, time series and spatial maps of the selected signal components should be inspected prior to further analysis to avoid erroneous detections. Finally, the present study mainly focused on periodic block designs, and may not directly apply to non-periodic block-design studies. For example, our signal component selection method assumes that the signal of interest is periodic with respect to time. Therefore, this method is not suitable to extract signal components from data with a non-periodic paradigm. Additionally, the sinusoidal functions in the Fourier basis set was used to model BOLD responses over the entire scan session. In this case, their fundamental frequency was set to the frequency of repeated stimulation cycles, which would only apply to studies with periodic designs. Nonetheless, the Fourier basis set may be used for non-periodic block-design studies if it is used to model the impulse response function instead. In this case, the sinusoidal functions may need to be convolved with the stimulation paradigm before being used as regressors. For example, the Fourier basis set was used to model impulse responses in epilepsy patients and successfully detected activations that may be indicative of propagated epileptiform activity (Lemieux et al., 2008). Therefore, further evaluations may be necessary to understand each method's detection and characterization capabilities in non-periodic block-design studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported through funding from the NIH/NIBIB R00 Award (4R00EB008738), Okawa Foundation Research Grant Award, NIH Director's New Innovator Award (1DP2OD007265), NSF CAREER Award (1056008), Alfred P. Sloan Foundation Research Fellowship, NIH/NINDS R01NS087159, NIH/NINDS R01NS091461, and NIH/NIARF1AG047666. J.H.L. would like to thank Karl Deisseroth for providing the DNA plasmids used for the optogenetic experiments. The authors would also like to thank the entire Lee Lab members for their assistance with the ofMRI experiments.

Abbreviations

ofMRI optogenetic functional magnetic resonance imaging

References

Abe Y, Sekino M, Terazono Y, Ohsaki H, Fukazawa Y, Sakai S, Yawo H, Hisatsune T. Opto-fMRI analysis for exploring the neuronal connectivity of the hippocampal formation in rats. Neurosci Res. 2012; 74:248–255. [PubMed: 22982343]

Aguirre GK, Zarahn E, D'Esposito M. The variability of human, BOLD hemodynamic responses. Neuroimage. 1998; 8:360–369. [PubMed: 9811554]

Akaike, H. Selected Papers of Hirotugu Akaike. Springer; 1998. Information theory and an extension of the maximum likelihood principle; p. 199-213.

Amemiya S, Kunimatsu A, Saito N, Ohtomo K. Impaired hemodynamic response in the ischemic brain assessed with BOLD fMRI. Neuroimage. 2012; 61:579–590. [PubMed: 22507231]

Angenstein F, Kammerer E, Scheich H. The BOLD response in the rat hippocampus depends rather on local processing of signals than on the input or output activity. A combined functional MRI and electrophysiological study. The Journal of Neuroscience. 2009; 29:2428–2439. [PubMed: 19244518]

Ashburner J, Barnes G, Chen C, Daunizeau J, Flandin G, Friston K, Gitelman D, Kiebel S, Kilner J, Litvak V. Face Group fMRI Data. SPM8 Manual. 2008:261–280.

Backfrieder W, Baumgartner R, Samal M, Moser E, Bergmann H. Quantification of intensity variations in functional MR images using rotated principal components. Physics in Medicine and Biology. 1996; 41:1425–1438. [PubMed: 8858728]

Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS. Processing strategies for time-course data sets in functional MRI of the human brain. Magn Reson Med. 1993; 30:161–173. [PubMed: 8366797]

Baumgartner R, Ryner L, Richter W, Summers R, Jarmasz M, Somorjai R. Comparison of two exploratory data analysis methods for fMRI: fuzzy clustering vs. principal component analysis. Magn Reson Imaging. 2000; 18:89–94. [PubMed: 10642106]

Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imaging. 2004; 23:137–152. [PubMed: 14964560]

Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 1995; 7:1129–1159. [PubMed: 7584893]

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological). 1995:289–300.

Bosshard SC, Stuker F, von Deuster C, Schroeter A, Rudin M. BOLD fMRI of C-Fiber Mediated Nociceptive Processing in Mouse Brain in Response to Thermal Stimulation of the Forepaws. PloS one. 2015; 10:e0126513. [PubMed: 25950440]

Bullmore E, Brammer M, Williams SC, Rabe-Hesketh S, Janot N, David A, Mellers J, Howard R, Sham P. Statistical methods of estimation and inference for functional MR image analysis. Magn Reson Med. 1996; 35:261–277. [PubMed: 8622592]

Byers B, Lee HJ, Liu J, Weitz AJ, Lin P, Zhang P, Shcheglovitov A, Dolmetsch R, Pera RR, Lee JH. Direct in vivo assessment of human stem cell graft-host neural circuits. Neuroimage. 2015; 114:328–337. [PubMed: 25936696]

Calhoun V, Adali T, Pearlson G, Pekar J. A method for making group inferences from functional MRI data using independent component analysis. Human brain mapping. 2001a; 14:140–151. [PubMed: 11559959]

Calhoun V, Stevens M, Pearlson G, Kiehl K. fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. Neuroimage. 2004; 22:252–257. [PubMed: 15110015]

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Human brain mapping. 2001b; 13:43–53. [PubMed: 11284046]

Christie IN, Wells JA, Southern P, Marina N, Kasparov S, Gourine AV, Lythgoe MF. fMRI response to blue light delivery in the naive brain: implications for combined optogenetic fMRI studies. Neuroimage. 2013; 66:634–641. [PubMed: 23128081]

Chuang K-H, Chiu M-J, Lin C-C, Chen J-H. Model-free functional MRI analysis using Kohonen clustering neural network and fuzzy C-means. IEEE Trans Med Imaging. 1999; 18:1117–1128. [PubMed: 10695525]

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical research. 1996; 29:162–173. [PubMed: 8812068]

D'Esposito M, Detre JA, Aguirre GK, Stallcup M, Alsop DC, Tippet LJ, Farah MJ. A functional MRI study of mental image generation. Neuropsychologia. 1997; 35:725–730. [PubMed: 9153035]

Degras D, Lindquist MA. A hierarchical model for simultaneous detection and estimation in multi-subject fMRI studies. Neuroimage. 2014; 98:61–72. [PubMed: 24793829]

Desai M, Kahn I, Knoblich U, Bernstein J, Atallah H, Yang A, Kopell N, Buckner RL, Graybiel AM, Moore CI. Mapping brain networks in awake mice using combined optical neural control and fMRI. J Neurophysiol. 2011; 105:1393–1405. [PubMed: 21160013]

Deschenes M, Bourassa J, Parent A. Striatal and cortical projections of single neurons from the central lateral thalamic nucleus in the rat. Neuroscience. 1996; 72:679–687. [PubMed: 9157314]

Duffy B, Choy M, Chuapoco M, Madsen M, Lee J. MRI compatible optrodes for simultaneous LFP and optogenetic fMRI investigation of seizure-like afterdischarges. Neuroimage. 2015; 123:173–184. [PubMed: 26208873]

Eklund A, Andersson M, Josephson C, Johannesson M, Knutsson H. Does parametric fMRI analysis with SPM yield valid results?—An empirical study of 1484 rest datasets. Neuroimage. 2012; 61:565–578. [PubMed: 22507229]

Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Sciences. 2016; 113:7900–7905.

Engel SA, Glover GH, Wandell BA. Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex. 1997; 7:181–192. [PubMed: 9087826]

Erhardt EB, Rachakonda S, Bedrick EJ, Allen EA, Adali T, Calhoun VD. Comparison of multi-subject ICA methods for analysis of fMRI data. Human brain mapping. 2011; 32:2075–2095. [PubMed: 21162045]

Esposito F, Formisano E, Seifritz E, Goebel R, Morrone R, Tedeschi G, Di Salle F. Spatial independent component analysis of functional MRI time series: To what extent do results depend on the algorithm used? Human brain mapping. 2002; 16:146–157. [PubMed: 12112768]

Esposito F, Scarabino T, Hyvarinen A, Himberg J, Formisano E, Comani S, Tedeschi G, Goebel R, Seifritz E, Di Salle F. Independent component analysis of fMRI group studies by self-organizing clustering. Neuroimage. 2005; 25:193–205. [PubMed: 15734355]

Fang Z, Lee JH. High-throughput optogenetic functional magnetic resonance imaging with parallel computations. J Neurosci Methods. 2013; 218:184–195. [PubMed: 23747482]

Ferenczi EA, Zalocusky KA, Liston C, Grosenick L, Warden MR, Amatya D, Katovich K, Mehta H, Patenaude B, Ramakrishnan C, Kalanithi P, Etkin A, Knutson B, Glover GH, Deisseroth K. Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. Science. 2016; 351:aac9698. [PubMed: 26722001]

Fox MD, Snyder AZ, Barch DM, Gusnard DA, Raichle ME. Transient BOLD responses at block transitions. Neuroimage. 2005a; 28:956–966. [PubMed: 16043368]

Fox MD, Snyder AZ, McAvoy MP, Barch DM, Raichle ME. The BOLD onset transient: identification of novel functional differences in schizophrenia. Neuroimage. 2005b; 25:771–782. [PubMed: 15808978]

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. Event-related fMRI: characterizing differential responses. Neuroimage. 1998; 7:30–40. [PubMed: 9500830]

Friston KJ, Holmes AP, Worsley KJ, Poline J, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Human brain mapping. 1994; 2:189–210.

Gazzola V, Keysers C. The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. Cerebral Cortex. 2009; 19:1239–1255. [PubMed: 19020203]

Genovese CR. A Bayesian time-course model for functional magnetic resonance imaging data. Journal of the American Statistical Association. 2000; 95:691–703.

Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage. 2002; 15:870–878. [PubMed: 11906227]

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. Representation of the temporal envelope of sounds in the human brain. J Neurophysiol. 2000; 84:1588–1598. [PubMed: 10980029]

Gonzalez-Castillo J, Saad ZS, Handwerker DA, Inati SJ, Brenowitz N, Bandettini PA. Whole-brain, time-locked activation with simple tasks revealed using massive averaging and model-free analysis. Proceedings of the National Academy of Sciences. 2012; 109:5487–5492.

Handwerker DA, Gonzalez-Castillo J, D'Esposito M, Bandettini PA. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. Neuroimage. 2012; 62:1017–1023. [PubMed: 22366081]

Handwerker DA, Ollinger JM, D'Esposito M. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage. 2004; 21:1639–1651. [PubMed: 15050587]

Harms MP, Melcher JR. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. J Neurophysiol. 2002; 88:1433–1450. [PubMed: 12205164]

Harms MP, Melcher JR. Detection and quantification of a wide range of fMRI temporal responses using a physiologically-motivated basis set. Human brain mapping. 2003; 20:168–183. [PubMed: 14601143]

Henson R, Price CJ, Rugg M, Turner R, Friston KJ. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. Neuroimage. 2002; 15:83–97. [PubMed: 11771976]

Hu D, Yan L, Liu Y, Zhou Z, Friston KJ, Tan C, Wu D. Unified SPM-ICA for fMRI analysis. Neuroimage. 2005; 25:746–755. [PubMed: 15808976]

Hunter MD, Farrow TF, Papadakis NG, Wilkinson ID, Woodruff PW, Spence SA. Approaching an ecologically valid functional anatomy of spontaneous "willed" action. Neuroimage. 2003; 20:1264–1269. [PubMed: 14568495]

Johnson SC. Hierarchical clustering schemes. Psychometrika. 1967; 32:241–254. [PubMed: 5234703]

Kahn I, Knoblich U, Desai M, Bernstein J, Graybiel A, Boyden E, Buckner R, Moore C. Optogenetic drive of neocortical pyramidal neurons generates fMRI signals that are correlated with spiking activity. Brain research. 2013; 1511:33–45. [PubMed: 23523914]

Lee HJ, Weitz AJ, Bernal-Casas D, Duffy BA, Choy M, Kravitz AV, Kreitzer AC, Lee JH. Activation of Direct and Indirect Pathway Medium Spiny Neurons Drives Distinct Brain-wide Responses. Neuron. 2016; 91:412–424. [PubMed: 27373834]

Lee JH, Durand R, Gradinaru V, Zhang F, Goshen I, Kim DS, Fenno LE, Ramakrishnan C, Deisseroth K. Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. Nature. 2010; 465:788–792. [PubMed: 20473285]

Lemieux L, Laufs H, Carmichael D, Paul JS, Walker MC, Duncan JS. Noncanonical spike-related BOLD responses in focal epilepsy. Human brain mapping. 2008; 29:329–345. [PubMed: 17510926]

Liao CH, Worsley KJ, Poline JB, Aston JA, Duncan GH, Evans AC. Estimating the delay of the fMRI response. Neuroimage. 2002; 16:593–606. [PubMed: 12169246]

Lindquist MA, Meng Loh J, Atlas LY, Wager TD. Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. Neuroimage. 2009; 45:S187–198. [PubMed: 19084070]

Liu J, Lee HJ, Weitz AJ, Fang Z, Lin P, Choy M, Fisher R, Pinskiy V, Tolpygo A, Mitra P, Schiff N, Lee JH. Frequency-selective control of cortical and subcortical networks by central thalamus. eLife. 2015; 4:e09215. [PubMed: 26652162]

Makni S, Idier J, Vincent T, Thirion B, Dehaene-Lambertz G, Ciuciu P. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. Neuroimage. 2008; 41:941–969. [PubMed: 18439839]

Marxen M, Cassidy RJ, Dawson TL, Ross B, Graham SJ. Transient and sustained components of the sensorimotor BOLD response in fMRI. Magn Reson Imaging. 2012; 30:837–847. [PubMed: 22495237]

Matthews PM, Honey GD, Bullmore ET. Applications of fMRI in translational medicine and clinical practice. Nature Reviews Neuroscience. 2006; 7:732–744. [PubMed: 16924262]

Maus B, van Breukelen GJ, Goebel R, Berger MP. Optimal design for nonlinear estimation of the hemodynamic response function. Human brain mapping. 2012; 33:1253–1267. [PubMed: 21567658]

McKeown MJ, Jung T-P, Makeig S, Brown G, Kindermann SS, Lee T-W, Sejnowski TJ. Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. Proceedings of the National Academy of Sciences. 1998a; 95:803–810.

McKeown MJ, Makeig S, Brown GG, Jung TP, Kindermann SS, Bell AJ, Sejnowski TJ. Analysis of fMRI data by blind separation into independent spatial components. Human brain mapping. 1998b; 6:160–188. [PubMed: 9673671]

Meyer-Baese A, Wismueller A, Lange O. Comparison of two exploratory data analysis methods for fMRI: unsupervised clustering versus independent component analysis. IEEE Trans Inf Technol Biomed. 2004; 8:387–398. [PubMed: 15484444]

Moritz CH, Rogers BP, Meyerand ME. Power spectrum ranked independent component analysis of a periodic fMRI complex motor paradigm. Human brain mapping. 2003; 18:111–122. [PubMed: 12518291]

Nandy RR, Cordes D. Novel ROC-type method for testing the efficiency of multivariate statistical methods in fMRI. Magn Reson Med. 2003; 49:1152–1162. [PubMed: 12768594]

Nersesyan H, Hyder F, Rothman DL, Blumenfeld H. Dynamic fMRI and EEG recordings during spike-wave seizures and generalized tonic-clonic seizures in WAG/Rij rats. J Cereb Blood Flow Metab. 2004; 24:589–599. [PubMed: 15181366]

Ngan SC, Hu X, Tan LH, Khong PL. Improvement of spectral density-based activation detection of event-related fMRI data. Magn Reson Imaging. 2009; 27:879–894. [PubMed: 19535208]

Pinto J, Jorge J, Sousa I, Vilela P, Figueiredo P. Fourier modeling of the BOLD response to a breath-hold task: Optimization and reproducibility. Neuroimage. 2016; 135:223–231. [PubMed: 26908316]

Pujol J, López-Solà M, Ortiz H, Vilanova JC, Harrison BJ, Yücel M, Soriano-Mas C, Cardoner N, Deus J. Mapping brain response to pain in fibromyalgia patients using temporal analysis of FMRI. PloS one. 2009; 4:e5224. [PubMed: 19381292]

Roc AC, Wang J, Ances BM, Liebeskind DS, Kasner SE, Detre JA. Altered hemodynamics and regional cerebral blood flow in patients with hemodynamically significant stenoses. Stroke. 2006; 37:382–387. [PubMed: 16373653]

Rombouts SA, Goekoop R, Stam CJ, Barkhof F, Scheltens P. Delayed rather than decreased BOLD response as a marker for early Alzheimer's disease. Neuroimage. 2005; 26:1078–1085. [PubMed: 15961047]

Schlegel F, Schroeter A, Rudin M. The hemodynamic response to somatosensory stimulation in mice depends on the anesthetic used: Implications on analysis of mouse fMRI data. Neuroimage. 2015; 116:40–49. [PubMed: 25979665]

Schmid F, Wachsmuth L, Albers F, Schwalm M, Stroh A, Faber C. True and apparent optogenetic BOLD fMRI signals. Magn Reson Med. 2016 In Press.

Schroeter A, Schlegel F, Seuwen A, Grandjean J, Rudin M. Specificity of stimulus-evoked fMRI responses in the mouse: the influence of systemic physiological changes associated with innocuous stimulation under four different anesthetics. Neuroimage. 2014; 94:372–384. [PubMed: 24495809]

Schwarz G. Estimating the dimension of a model. The annals of statistics. 1978; 6:461–464.

Seifritz E, Esposito F, Hennel F, Mustovic H, Neuhoff JG, Bilecen D, Tedeschi G, Scheffler K, Di Salle F. Spatiotemporal pattern of neural processing in the human auditory cortex. Science. 2002; 297:1706–1708. [PubMed: 12215648]

Shan ZY, Wright MJ, Thompson PM, McMahon KL, Blokland GG, de Zubicaray GI, Martin NG, Vinkhuyzen AA, Reutens DC. Modeling of the hemodynamic responses in block design fMRI studies. J Cereb Blood Flow Metab. 2014; 34:316–324. [PubMed: 24252847]

Skudlarski P, Constable RT, Gore JC. ROC analysis of statistical methods used in functional MRI: individual subjects. Neuroimage. 1999; 9:311–329. [PubMed: 10075901]

Steffener J, Tabert M, Reuben A, Stern Y. Investigating hemodynamic response variability at the group level using basis functions. Neuroimage. 2010; 49:2113–2122. [PubMed: 19913625]

Steriade M, Glenn LL. Neocortical and caudate projections of intralaminar thalamic neurons and their synaptic excitation from midbrain reticular core. J Neurophysiol. 1982; 48:352–371. [PubMed: 6288887]

Sychra J, Bandettini P, Bhattacharya N, Lin Q. Synthetic images by subspace transforms I. Principal components images and related filters. Medical Physics. 1994; 21:193–201. [PubMed: 8177152]

Takata N, Yoshida K, Komaki Y, Xu M, Sakai Y, Hikishima K, Mimura M, Okano H, Tanaka KF. Optogenetic Activation of CA1 Pyramidal Neurons at the Dorsal and Ventral Hippocampus Evokes Distinct Brain-Wide Responses Revealed by Mouse fMRI. PloS one. 2015; 10:e0121417. [PubMed: 25793741]

Uludag K. Transient and sustained BOLD responses to sustained visual stimulation. Magn Reson Imaging. 2008; 26:863–869. [PubMed: 18479869]

Van der Werf YD, Witter MP, Groenewegen HJ. The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. Brain Research Reviews. 2002; 39:107–140. [PubMed: 12423763]

Ward BD. Deconvolution analysis of fMRI time series data. AFNI 3dDeconvolve Documentation. 2006:25–26.

Wasserman L. Bayesian Model Selection and Model Averaging. J Math Psychol. 2000; 44:92–107. [PubMed: 10733859]

Weber R, Ramos-Cabrer P, Justicia C, Wiedermann D, Strecker C, Sprenger C, Hoehn M. Early prediction of functional recovery after experimental stroke: functional magnetic resonance imaging, electrophysiology, and behavioral testing in rats. J Neurosci. 2008; 28:1022–1029. [PubMed: 18234880]

Weilke F, Spiegel S, Boecker H, von Einsiedel HG, Conrad B, Schwaiger M, Erhard P. Time-resolved fMRI of activation patterns in M1 and SMA during complex voluntary movement. J Neurophysiol. 2001; 85:1858–1863. [PubMed: 11353002]

Weitz AJ, Fang Z, Lee HJ, Fisher RS, Smith WC, Choy M, Liu J, Lin P, Rosenberg M, Lee JH. Optogenetic fMRI reveals distinct, frequency-dependent networks recruited by dorsal and intermediate hippocampus stimulations. Neuroimage. 2014; 107:229–241. [PubMed: 25462689]

Weitz AJ, Lee JH. Progress with optogenetic functional MRI and its translational implications. Future Neurology. 2013; 8:691–700.

Welvaert M, Rosseel Y. How ignoring physiological noise can bias the conclusions from fMRI simulation results. Journal of neuroscience methods. 2012; 211:125–132. [PubMed: 22960507]

Williams KA, Magnuson M, Majeed W, LaConte SM, Peltier SJ, Hu X, Keilholz SD. Comparison of alpha-chloralose, medetomidine and isoflurane anesthesia for functional connectivity mapping in the rat. Magn Reson Imaging. 2010; 28:995–1003. [PubMed: 20456892]

Wismüller A, Lange O, Dersch DR, Leinsinger GL, Hahn K, Pütz B, Auer D. Cluster analysis of biomedical image time-series. International Journal of Computer Vision. 2002; 46:103–128.

Woolrich MW, Behrens TE, Smith SM. Constrained linear basis sets for HRF modelling using Variational Bayes. Neuroimage. 2004; 21:1748–1761. [PubMed: 15050595]

Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. J Cereb Blood Flow Metab. 1992; 12:900–918. [PubMed: 1400644]

Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited--again. Neuroimage. 1995; 2:173–181. [PubMed: 9343600]

Yu X, He Y, Wang M, Merkle H, Dodd SJ, Silva AC, Koretsky AP. Sensory and optogenetically driven single-vessel fMRI. Nat Methods. 2016; 13:337–340. [PubMed: 26855362]

Highlights

Six fMRI analysis methods are evaluated against heterogeneous BOLD responses.

- Evaluations are conducted using both real and simulated data.
- Conventionally used canonical HRF leads to detection and characterization errors.
- Flexible models show robust detection and characterization performance.
- The gamma, finite impulse response, B-spline, and Fourier basis sets are preferred.

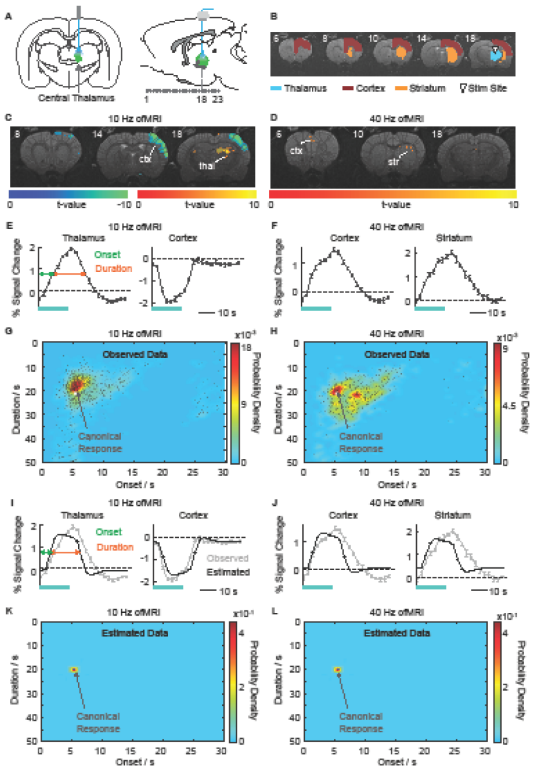

Figure 1.

GLM with 1st order canonical basis set leads to detection and characterization errors with experimental ofMRI datasets. (A) Schematic of the ofMRI experimental design indicating the site of transduced cells in the central thalamus (green), optical fiber location (blue line), and location of acquired coronal fMRI slices (1…23). Slice numbers correspond to those denoted in the activation maps in subsequent figures. (B) Regions of interest used to extract the time series in subsequent figures are shown. (C, D) Activation maps detected by GLM with the 1st order canonical basis set from a representative subject are shown. For group-level activation maps, please see Supplementary Figure S5 and S6. In panel C-L, non-smoothed ofMRI data were used and the threshold was set to $p < 0.05$ with Bonferroni correction. T2-weighted anatomical images are used as underlays. (E, F) Plots show the observed BOLD responses that are detected by GLM with the 1st order canonical basis set. They were generated by first averaging the time series of detected voxels that fell within each region of interest, followed by averaging over six stimulation cycles of the resulting time series. Error bars represent standard error of the mean (SEM) across different stimulation cycles. The superset of voxels detected by GLM with the 1st order canonical basis set in each subject were used in panel E-L (N = 10 subjects). Percent signal change was calculated relative to the baseline period. Horizontal blue bars represent the 20 s period of optical stimulation. (G, H) Plots show the onset and duration measured from the observed data. Each grey dot refers to a detected voxel from a subject (N = 10 subjects). Probability density color map was calculated with a histogram of 50 bins along each axis. Canonical response was generated by convolving the canonical HRF with the experimental paradigm. (I, J) Plots show the fitted BOLD responses that are estimated by GLM with the 1st order canonical basis set. The observed responses shown in panel E and F are overlaid here for

comparison using gray line. (K, L) Plots show the onset and duration estimated by GLM with the 1st order canonical basis set. Abbreviations are as follows: ctx (cortex), str (striatum), and thal (thalamus).

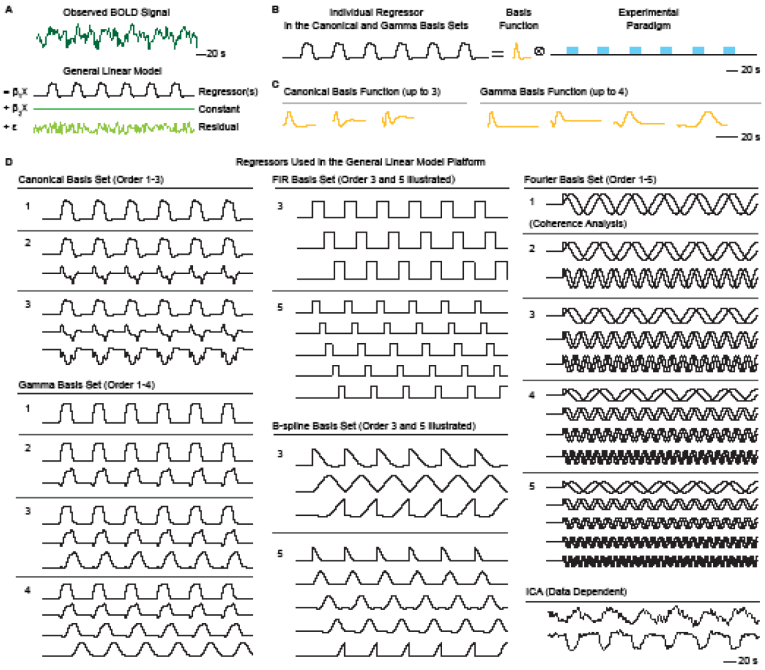

Figure 2.

A GLM platform was used to compare analysis methods. (A) Illustration of the general linear model framework. β_1 and β_2 are the coefficients in the model, and ε represents residual. (B) Each regressor used in the canonical and gamma basis sets is obtained by convolving a basis function with the experimental paradigm. Blue boxcars in the experimental paradigm represent the 20 s period of optical stimulation. (C) Illustration of the basis functions in the canonical and gamma basis sets. (D) Illustration of different sets of regressors used in the general linear model framework. Due to the space constraint, only order 3 and 5 are shown for the FIR and B-spline basis sets.

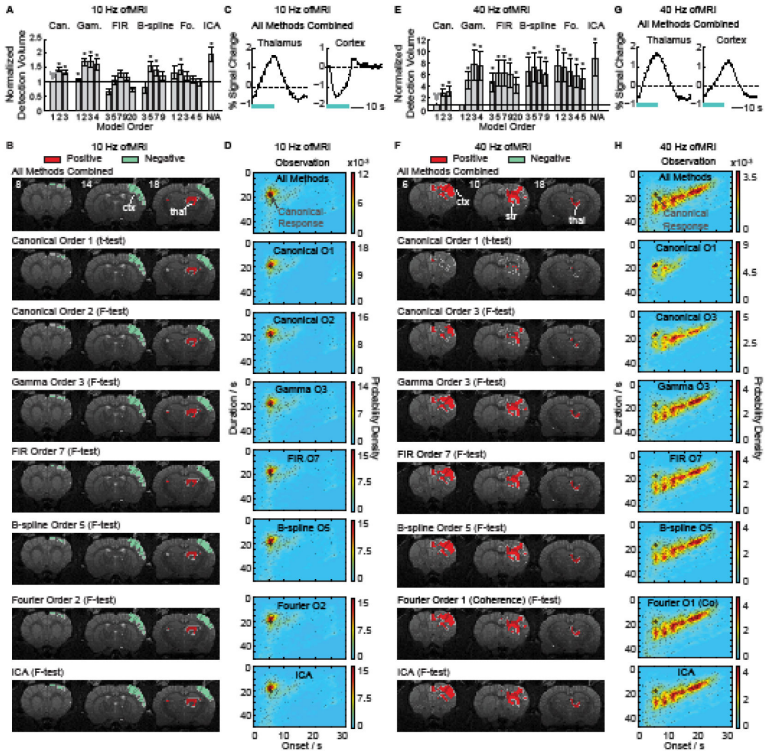

Figure 3.

GLM with the 2nd and 3rd order canonical basis sets, the 2nd to 4th order gamma basis sets, the 5th to 7th order B-spline basis sets, the 2nd order Fourier basis set, and ICA yield large detection volumes with the experimental ofMRI datasets. (A-D) Detection capability assessment with the 10 Hz stimulation ofMRI data. (E-H) Detection capability assessment with the 40 Hz stimulation ofMRI data. (A, E) Bar graphs show the active volume detected by each method. The detection volume was first normalized to the active volume detected by GLM with the 1st order canonical basis set (gray arrowhead and dashed horizontal line) for each subject, and then averaged across different subjects. Non-smoothed ofMRI data were used and the threshold was set to $p < 0.05$ with Bonferroni correction in each panel. Error bars represent SEM across different subjects. Asterisk indicates $p < 0.05$ compared with GLM with the 1st order canonical basis set using one-sided Wilcoxon signed-rank test (N = 10 subjects). (B, F) Activation maps from a representative subject are shown. For group-level activation maps, please see Supplementary Figure S5 and S6. Here, for simplicity, we show results from conventionally used GLM with the 1st order canonical basis set, as well as from the model order within each method that yields the largest detection volume in panel A and E. The gray border represents the superset of voxels that are detected by all tested methods. Colored voxels are detected by the method denoted in the figure legend. The evoked response is considered positive if the average percent signal change over entire stimulation cycle is above zero, otherwise is considered negative. T2-weighted anatomical images are used as underlays. (C, G) Plots show the observed BOLD responses that are detected by all methods combined. Error bars represent SEM across different stimulation cycles. Horizontal blue bars represent the 20 s period of optical stimulation. (D, H) Plots show the onset and duration measured from the BOLD responses that are detected by each

method. The onset and duration are measured from the observed time course, not estimated from the fitted time course. Each grey dot refers to a detected voxel from a subject (N = 10 subjects). The probability density color map is overlaid. Note that, the plots for the 1st order canonical basis set are identical to those in Figure 1G and H. Abbreviations are as follows: Can. (canonical), Gam. (gamma), Fo. (Fourier), ctx (cortex), str (striatum), thal (thalamus), and co (coherence analysis).

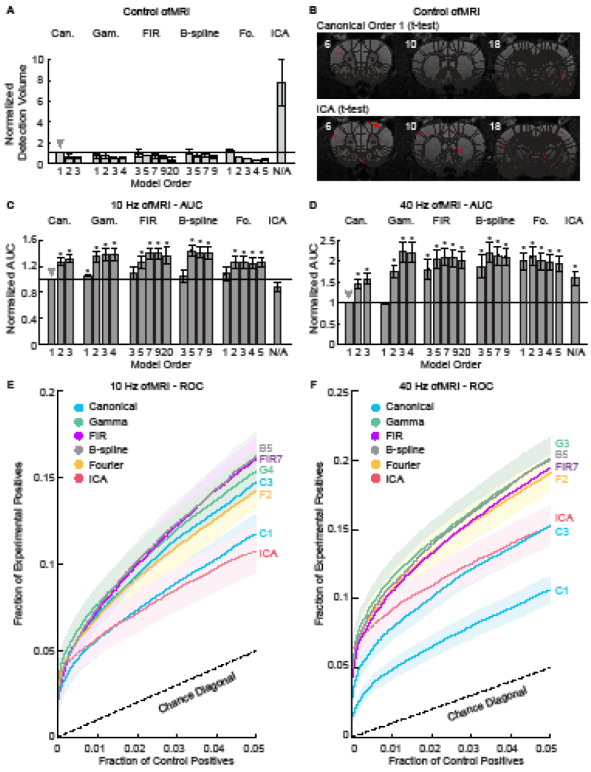

Figure 4.

GLM with the 2nd and 3rd order canonical basis sets, the 2nd to 4th order gamma basis sets, the 5th to 20th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets show good detection performance with the ofMRI datasets. (A) Bar graph shows the normalized detection volume with the control ofMRI data. The detection volume was first normalized to the active volume detected by GLM with the 1st order canonical basis set for each subject (gray arrowhead and dashed horizontal line), and then averaged across different subjects. Error bars represent SEM across different subjects (N = 2 subjects). Non-smoothed ofMRI data are used in each panel. The threshold was set to uncorrected $p < 0.001$ in panel A and B to better show the spurious activations detected by each method. Please see Supplementary Figure S8 and its figure caption for results with other threshold settings. (B) Activation maps from a representative subject are shown. T2-weighted anatomical images are used as underlays. In this subject, one signal component was extracted for ICA and t-test was used during statistical analysis. (C, D) Bar graphs represent normalized AUC values for each method. AUC values were first normalized to the AUC value yielded by GLM with the 1st order canonical basis set (gray arrowhead and dashed horizontal line) for each subject, and then averaged across different subjects. Error bars represent SEM across different experimental subjects. Asterisk indicates $p < 0.05$ compared with GLM with the 1st order canonical basis set using one-sided Wilcoxon signed-rank test (N = 10 subjects). (E, F) Modified ROC curves are shown. For simplicity, we show results from conventionally used GLM with the 1st order canonical basis set, as well as from the model order within each method that yields the largest AUC values in panel C and D. Shaded areas represent SEM across different experimental subjects (N = 10 subjects). C1 to C3, G1 to G4, FIR3 to FIR20, B3 to B9, and F1 to F5 refer to different model orders in the

canonical, gamma, FIR, B-spline, and Fourier basis sets. Abbreviations are as follows: Can. (canonical), Gam. (gamma), and Fo. (Fourier).

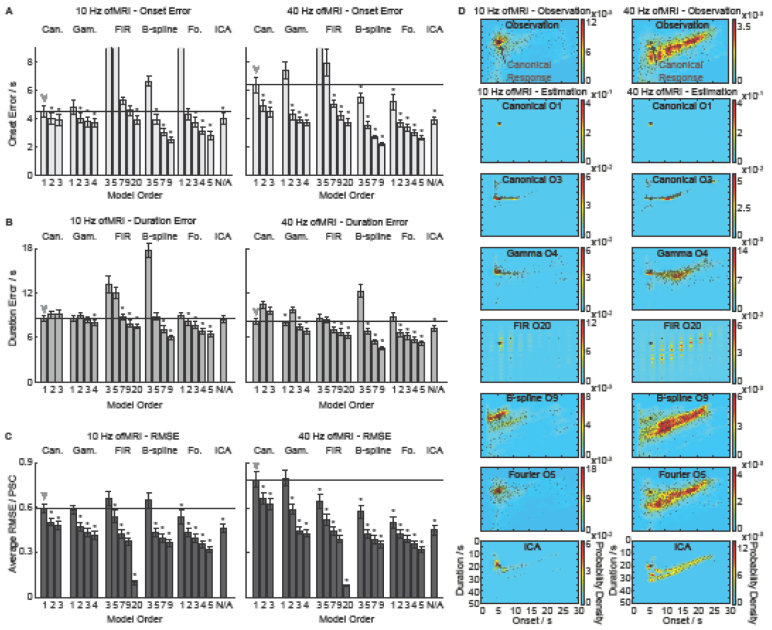

Figure 5.

GLM with the 4th order gamma basis set, the 20th order FIR basis set, the 7th to 9th order B-spline basis sets, and the 3rd to 5th order Fourier basis sets show good characterization performance with the ofMRI datasets. The results in each panel were calculated using the superset of voxels detected by all methods combined, rather than the voxels detected by each method. This way each method's characterization capabilities can be assessed without the influence of their detection performance. Non-smoothed ofMRI data were used and the threshold is set to $p < 0.05$ with Bonferroni correction in each panel. (A-C) Bar graphs show the average onset error (A), duration error (B), and RMSE (C). In panel A, the onset errors of some methods are beyond the upper limit of the plot. Therefore, these values are not shown to better illustrate the onset errors of the rest methods. The methods with extreme onset errors include the 3rd and 5th order FIR basis sets and the 1st order Fourier basis set with the 10 Hz stimulation ofMRI data, and the 3rd order FIR basis set with the 40 Hz stimulation ofMRI data. The onset and duration error was calculated relative to the onset and duration measured from the observed time course and is in the unit of seconds. RMSE was calculated relative to the observed data and is in the unit of percent signal change (PSC). The mean errors in panel A to C were calculated by first averaging the errors across different detected voxels within each subject, and then averaging across different subjects. Error bars represent SEM across different subjects. Asterisk indicates $p < 0.05$ compared with GLM with the 1st order canonical basis set using one-sided Wilcoxon signed-rank test (N = 10 subjects). The results using GLM with the 1st order canonical basis set is denoted with a gray arrowhead and a dashed horizontal line. (D) Plots show the onset and duration measured from the observed data (top panel) and estimated by each method (bottom panels). For simplicity, we show results from conventionally used GLM with the 1st order canonical basis set, as well as from the model order within each method that yields the lowest errors in panel A to C. Each grey dot refers to a detected voxel from a subject (N = 10 subjects). The probability density color map is overlaid. Note that, the plots for the 1st order canonical basis

set are identical to those in Figure 1K and L. Abbreviations are as follows: Can. (canonical), Gam. (gamma), and Fo. (Fourier).

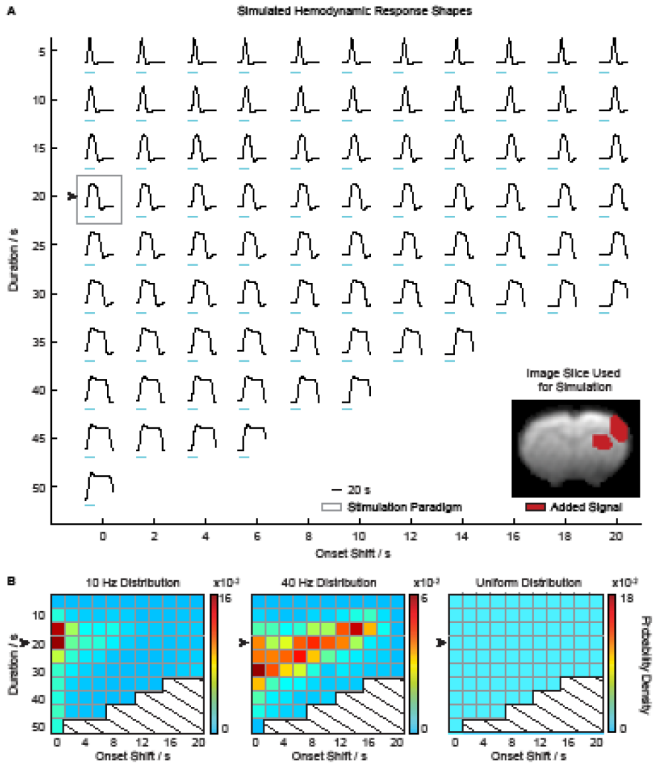

Figure 6.

Simulated datasets were designed to have a wide range of onset and duration. (A) Signals were simulated with varying onset and duration. The simulated shape that matches the canonical response is denoted by a black arrowhead in each panel. The horizontal axis refers to the onset shift of each signal shape relative to the canonical response. The vertical axis refers to the duration of each signal shape. Bottom right shows the spatial activation pattern of the simulated data with the ground truth positive voxels overlaid in red. (B) Three probability distributions of onset and duration are shown. Each square in the grid corresponds to the probability of the onset and duration from a different simulated signal shape.

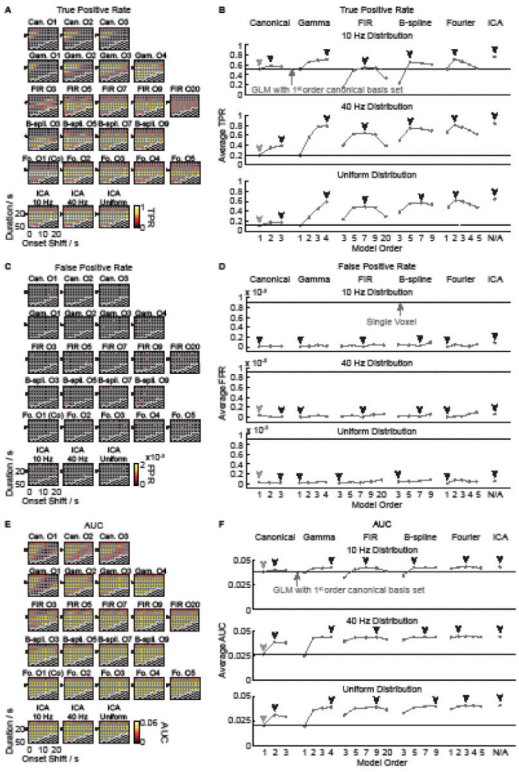

Figure 7.

GLM with the 2nd to 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 1st to 5th order Fourier basis sets, and ICA show good detection performance with the simulated datasets. (A, C, E) Colormaps show true positive rate

(A), false positive rate (C), and AUC (E) for each method. Threshold was set to $p < 0.05$ with Bonferroni correction in panel A-D. CNR is 1.5 in each panel. Results from the simulated shape that matches the canonical response is denoted by black arrowheads. (B, D, F) Plots show average true positive rate (B), false positive rate (D), and AUC (F). The results using GLM with the 1st order canonical basis set is denoted with a gray arrowhead and a dashed horizontal line. The model order that yields the highest true positive rate, the lowest false positive rate, and the highest AUC within each method is denoted by black arrowheads. Abbreviations are as follows: Can. (canonical), Gam. (gamma), B-spli. (B-spline), Fo. (Fourier), and Co (coherence analysis).

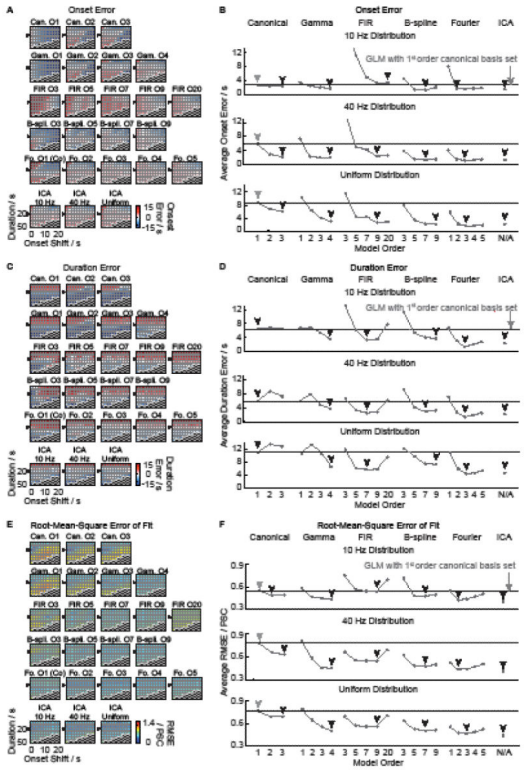

Figure 8.

GLM with the 3rd and 4th order gamma basis sets, the 5th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, the 2nd to 5th order Fourier basis sets, and ICA show good characterization performance with the simulated datasets. Similar to Figure 5, the results in each panel were calculated using the ground truth positive voxels rather than the voxels detected by each method, so that each method's characterization capability can be assessed without the influence of their detection performance. (A, C, E) Colormaps show onset errors (A), duration errors (C), and RMSE of fit (E) for each method. Color bar represents onset and duration errors in the unit of seconds, and RMSE in the unit of percent signal change (PSC). CNR is 1.5 in each panel. (B, D, F) Plots show average onset errors (B), duration errors (D), and RMSE (F). The results using GLM with the 1st order canonical basis set is denoted with a gray arrowhead and a dashed horizontal line. The model order that yields the smallest errors within each method is denoted by black arrowheads. Note that, in panel B, the onset errors of the 3rd order FIR basis set with the 10 Hz and 40 Hz stimulation ofMRI data distributions are beyond the upper limit of the plot and therefore are not shown to better illustrate the onset errors of the rest methods. Abbreviations are as follows: Can. (canonical), Gam. (gamma), B-spli. (B-spline), Fo. (Fourier), and Co (coherence analysis).

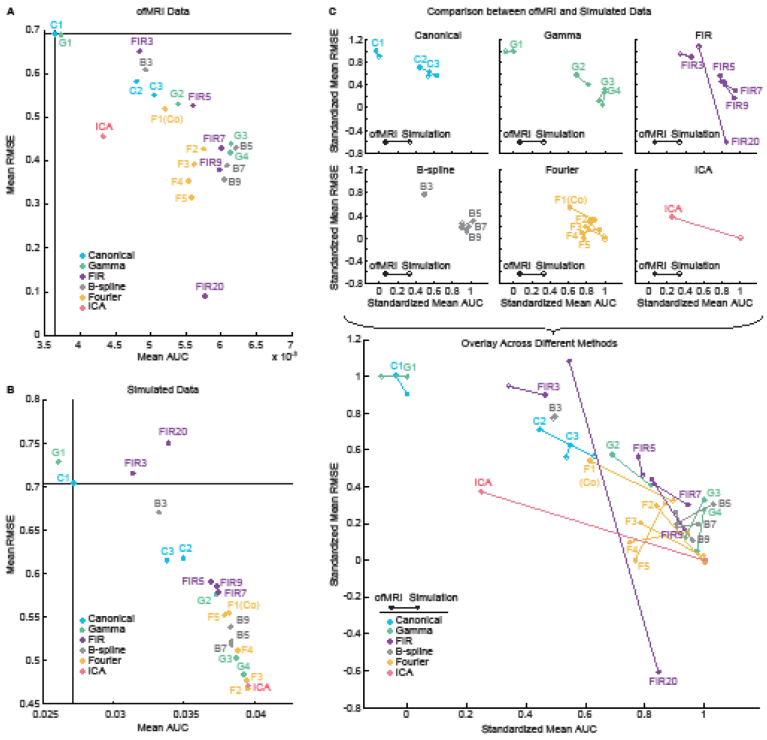

Figure 9.

GLM with the 3rd and 4th order gamma basis sets, the 7th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets show good balance between detection and characterization. (A, B) Plots show the mean AUC and RMSE of each method. Non-smoothed ofMRI data were used when calculating AUC and RMSE. The dashed line represents the results using GLM with the 1st order canonical basis set. C1 to C3, G1 to G4, FIR3 to FIR20, B3 to B9, and F1 to F5 refer to different model orders in the canonical, gamma, FIR, B-spline, and Fourier basis sets. Co stands for coherence analysis. (C) Plots show the standardized mean AUC and RMSE of each method. The bottom panel shows an overlay across different methods. For each method, the standardized AUC and RMSE from the ofMRI datasets (solid circle) is connected to the standardized AUC and RMSE from the simulated datasets (open circle) using a colored dashed line. The standardized AUC and RMSE are calculated using the mean AUC and RMSE of each method shown in panel A and B. See Section 2.2 for details regarding the calculation of standardized AUC and RMSE.

Table 1

GLM with the 3rd and 4th order gamma basis sets, the 7th to 9th order FIR basis sets, the 5th to 9th order B-spline basis sets, and the 2nd to 5th order Fourier basis sets show superior detection and characterization performance over other methods. Color bar represents the relative performance of different methods with each performance metric. Non-smoothed ofMRI data were used when calculating AUC and RMSE. To eliminate the influence of extreme values, the color scale was set based on the value of the second best performing method (red) and the second worst performing method (green) in each column. The best performing method exhibits the highest AUC and the smallest RMSE, and the worst performing method exhibits the lowest AUC and the largest RMSE. C1 to C3, G1 to G4, FIR3 to FIR20, B3 to B9, and F1 to F5 refer to different model orders in the canonical, gamma, FIR, B-spline, and Fourier basis sets. Co stands for coherence analysis.

Metrics		AUC ($\times 10^{-3}$)			RMSE ($\times 10^{-1}$)			AUC ($\times 10^{-2}$)										RMSE ($\times 10^{-1}$)									
Datasets		ofMRI			ofMRI			Simulation										Simulation									
CNR		N/A			N/A			1			1.5			2				1			1.5			2			
Condition		10 Hz	40 Hz	Mean	10 Hz	40 Hz	Mean	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	Mean	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	Mean
Canonical	C1	3.84	3.43	3.64	5.91	7.94	6.93	3.22	1.96	1.43	3.83	2.58	2.14	3.92	2.86	2.52	2.78	5.75	7.97	7.92	5.46	7.82	7.64	5.37	7.75	7.69	7.04
	C2	4.77	4.83	4.80	5.01	6.67	5.84	3.33	2.89	2.14	4.03	3.81	3.21	4.09	4.09	3.85	3.58	5.46	6.85	7.19	4.88	6.56	6.89	4.66	6.36	6.72	6.18
	C3	4.89	5.21	5.05	4.76	6.28	5.52	3.17	2.80	1.97	3.94	3.75	3.00	4.02	4.04	3.73	3.46	5.81	6.86	7.36	4.91	6.31	6.86	4.61	6.03	6.62	6.15
Gamma	G1	4.07	3.38	3.73	5.78	8.03	6.91	3.09	1.77	1.36	3.77	2.41	2.06	3.85	2.72	2.46	2.67	6.06	8.24	8.04	5.78	8.07	7.86	5.70	8.02	7.81	7.29
	G2	5.03	5.74	5.39	4.71	5.93	5.32	3.44	3.29	2.45	4.18	4.20	3.67	4.13	4.20	4.02	3.82	5.32	6.25	6.85	4.69	5.82	6.45	4.49	5.68	6.33	5.77
	G3	5.12	7.15	6.14	4.31	4.50	4.41	3.47	3.70	2.83	4.21	4.33	3.89	4.15	4.20	4.09	3.96	5.33	5.37	6.28	4.45	4.58	5.59	4.10	4.25	5.33	5.03
	G4	5.16	7.11	6.13	4.10	4.30	4.20	3.48	3.63	3.04	4.27	4.34	3.98	4.23	4.22	4.11	4.01	5.47	5.58	5.99	4.37	4.54	5.02	3.90	4.08	4.63	4.84
FIR	FIR 3	4.01	5.67	4.84	6.58	6.48	6.53	1.87	2.70	2.01	3.25	3.89	3.15	3.75	4.06	3.58	3.14	7.98	7.27	7.51	7.45	6.66	6.94	7.31	6.49	6.77	7.15
	FIR 5	4.70	6.50	5.60	5.36	5.21	5.28	2.91	3.21	2.64	4.10	4.27	3.81	4.17	4.18	3.97	3.70	6.89	6.71	6.85	5.70	5.58	5.74	5.26	5.13	5.31	5.91
	FIR 7	5.25	6.76	6.00	4.19	4.43	4.31	3.02	3.15	2.62	4.26	4.29	3.90	4.18	4.20	4.08	3.74	6.98	6.98	7.10	5.42	5.46	5.60	4.75	4.81	4.97	5.78
	FIR 9	5.20	6.75	5.98	3.71	3.90	3.81	2.93	3.04	2.59	4.22	4.31	3.95	4.23	4.22	4.12	3.73	7.31	7.41	7.48	5.36	5.48	5.60	4.51	4.67	4.81	5.85
	FIR 20	4.96	6.57	5.77	0.99	0.82	0.91	2.14	2.22	1.92	3.93	4.12	3.70	4.20	4.20	4.10	3.39	10.17	10.13	10.22	6.93	6.90	7.00	5.33	5.38	5.46	7.50
Spline	S3	3.87	5.99	4.93	6.45	5.76	6.10	2.10	3.00	2.39	3.42	4.05	3.37	3.82	4.12	3.63	3.32	7.68	6.89	6.99	7.10	6.24	6.35	6.93	6.02	6.13	6.70
	S5	5.39	7.03	6.21	4.33	4.30	4.31	3.24	3.49	2.84	4.28	4.33	3.89	4.21	4.22	4.04	3.84	6.12	6.11	6.46	4.80	4.84	5.26	4.29	4.32	4.80	5.22
	S7	5.31	6.86	6.09	3.91	3.90	3.90	3.23	3.39	2.86	4.26	4.34	4.03	4.14	4.18	4.11	3.84	6.47	6.47	6.66	4.77	4.79	5.03	4.02	4.06	4.34	5.18
	S9	5.27	6.81	6.04	3.61	3.56	3.58	3.18	3.26	2.78	4.26	4.32	4.05	4.24	4.24	4.16	3.83	6.98	6.99	7.12	4.95	4.96	5.12	4.02	4.07	4.25	5.39
Fourier	F1 (Co)	3.99	6.42	5.21	5.34	5.06	5.20	3.28	3.59	2.93	4.18	4.30	3.78	4.19	4.20	3.96	3.91	6.03	5.73	6.06	5.50	5.22	5.57	5.34	5.04	5.42	5.55
	F2	4.75	6.75	5.75	4.29	4.27	4.28	3.50	3.65	3.12	4.32	4.38	4.06	4.20	4.21	4.14	4.04	5.44	5.45	5.77	4.26	4.37	4.73	3.82	3.93	4.33	4.68
	F3	4.75	6.49	5.62	3.96	3.92	3.94	3.42	3.57	3.06	4.34	4.38	4.16	4.20	4.22	4.18	4.04	5.94	5.95	6.12	4.38	4.43	4.64	3.71	3.78	4.01	4.77
	F4	4.71	6.36	5.53	3.52	3.56	3.54	3.22	3.37	2.90	4.32	4.37	4.15	4.22	4.22	4.19	3.98	6.58	6.59	6.71	4.70	4.73	4.88	3.85	3.91	4.07	5.11
	F5	4.70	6.45	5.58	3.15	3.19	3.17	2.99	3.14	2.71	4.25	4.33	4.09	4.20	4.20	4.20	3.88	7.24	7.22	7.34	5.09	5.10	5.23	4.08	4.14	4.28	5.52

Metrics		AUC ($\times 10^{-3}$)			RMSE ($\times 10^{-1}$)			AUC ($\times 10^{-2}$)										RMSE ($\times 10^{-1}$)									
Datasets		ofMRI			ofMRI			Simulation										Simulation									
CNR		N/A			N/A			1			1.5			2				1			1.5			2			
Condition		10 Hz	40 Hz	Mean	10 Hz	40 Hz	Mean	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	Mean	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	10 Hz	40 Hz	Uniform	Mean
ICA	N/A	3.46	5.19	4.33	4.60	4.54	4.57	3.48	3.62	3.10	4.33	4.36	4.13	4.23	4.21	4.18	4.04	5.45	5.49	6.45	4.09	4.17	4.75	3.78	3.67	4.55	4.71