Paradigm Free Mapping With Sparse Regression Automatically Detects Single-Trial Functional Magnetic Resonance Imaging Blood Oxygenation Level Dependent Responses

César Caballero Gaudes,[1,2]* **Natalia Petridou,**[1,3] **Susan T. Francis,**[1] **Ian L. Dryden,**[4] **and Penny A. Gowland**[1]

[1]*Sir Peter Mansfield Magnetic Resonance Centre, School of Physics and Astronomy, University of Nottingham, Nottingham, United Kingdom*
[2]*Department of Radiology and Medical Informatics, Hôpitaux Universitaire de Genève (HUG), Universite de Genève, Genève, Switzerland*
[3]*Department of Radiology, Rudolf Magnus Institute, University Medical Centre Utrecht, Utrecht, Netherlands*
[4]*Department of Statistics, University of South Carolina, Columbia, South Carolina*

◆ ——————————————————————————————————— ◆

Abstract: The ability to detect single trial responses in functional magnetic resonance imaging (fMRI) studies is essential, particularly if investigating learning or adaptation processes or unpredictable events. We recently introduced paradigm free mapping (PFM), an analysis method that detects single trial blood oxygenation level dependent (BOLD) responses without specifying prior information on the timing of the events. PFM is based on the deconvolution of the fMRI signal using a linear hemodynamic convolution model. Our previous PFM method (Caballero-Gaudes et al., 2011: Hum Brain Mapp) used the ridge regression estimator for signal deconvolution and required a baseline signal period for statistical inference. In this work, we investigate the application of sparse regression techniques in PFM. In particular, a novel PFM approach is developed using the Dantzig selector estimator, solved via an efficient homotopy procedure, along with statistical model selection criteria. Simulation results demonstrated that, using the Bayesian information criterion to select the regularization parameter, this method obtains high detection rates of the BOLD responses, comparable with a model-based analysis, but requiring no information on the timing of the events and being robust against hemodynamic response function variability. The practical operation of this sparse PFM method was assessed with single-trial fMRI data acquired at 7T, where it automatically detected all task-related events, and was an improvement on our previous PFM method, as it does not require the definition of a baseline state and amplitude thresholding and does not compromise on specificity and sensitivity. *Hum Brain Mapp 34:501–518, 2013.*   © 2011 **Wiley Periodicals, Inc.**

◆ === ◆

INTRODUCTION

Measuring the brain's response to single trial events with blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI) [Menon et al., 1998; Richter et al., 1997] opens up the possibility of investigating learning or adaptation effects [Grill-Spector et al., 2001]. In recent work, we introduced a novel method to detect and characterize the hemodynamic response to single-trial events without prior information about their timing: paradigm free mapping (PFM) [Caballero-Gaudes et al., in press]. This allows the investigation of unpredictable events, such as interictal events in epilepsy [Bagshaw et al., 2005; Vulliemoz et al., 2010] or signal changes in pharmacological fMRI [Wise and Tracey, 2006], and provides a new method of studying activity in the resting state when there is no interaction with the subject [Petridou et al., 2011]. The PFM method does not require definition of the onset and duration of the event as is usually required for standard model-based analyses [Friston et al., 1998a]. Our previous PFM approach was based on the linear deconvolution of the BOLD fMRI signal assuming a particular hemodynamic response function (HRF), the use of the L_2-norm regularized estimator of ridge regression, and statistical assessment against a baseline state [Caballero-Gaudes et al., 2011]. In this work, we refine the PFM method by incorporating sparse regression techniques, specifically the Dantzig selector (DS) [Candes and Tao, 2007] in combination with model selection criteria [Zou et al., 2007], to take account of the fact that individual single-trial events occur sparsely in time. The proposed method, named sparse PFM (SPFM), avoids the definition of a baseline period and the need for amplitude thresholding of the deconvolved signal, thereby enhancing the detection of significant changes of the underlying signal driving the BOLD response with no prior information on timing.

PFM approaches assume a linear hemodynamic model. Linear deconvolution of the underlying signal was initially proposed in Gitelman et al. [2003] to enhance the sensitivity of fMRI to psychophysiological interactions rather than hemodynamic responses. Substantial advances have recently been made in using dynamic filtering methods for the deconvolution of the unknown neuronal-related signal and the estimation of the physiological variables and parameters governing the BOLD signal [Friston et al., 2008b; 2010; Havlicek et al., 2011; Riera et al., 2004] based on a nonlinear hemodynamic model, such as the Balloon–Windkessel model [Buxton et al., 1998; Friston et al., 2000]. However, here SPFM uses a linear model to gain computational speed and detect the BOLD events at a single-voxel level.

Sparse regression estimates are found by imposing an L_1-norm regularization penalty on the magnitude of the weights of the model regressors or features so that the weights of irrelevant regressors or features are reduced to zero [Bruckstein et al., 2009; Park and Casella, 2008; Tibshirani, 1996; Tipping, 2001]. Sparse regression methods have been shown to be superior to L_2-norm regularized techniques in a wide range of fMRI, magnetoencephalography (MEG), and electroencephalography (EEG) applications due to their improved model interpretability and estimation accuracy. To date, they have mainly been used for multivariate classification purposes where a classifier is trained (e.g., using LASSO, sparse Bayesian learning, sparse logistic regression or Elastic Net) to select those voxels or features that better discriminate between experimental conditions or states [Carroll et al., 2009; De Martino et al., 2008; Friston et al., 2008a; Grosenick et al., 2008; Liu et al., 2009; Michel et al., 2010; Raizada et al., 2010; Ryali et al., 2010; Valente et al., 2011; Van Gerven et al., 2009; 2010]. Analysis approaches promoting sparse solutions, either based on variational Bayesian frameworks with sparsifying priors or L_1-norm regularization, have been proposed for spatiotemporal fMRI models [Flandin and Penny, 2007; Long et al., 2004; Van De Ville et al., 2007] or MEG and EEG inverse problems [Friston et al., 2008c; Gramfort and Kowalski, 2009; Ou et al., 2009; Valdes-Sosa et al., 2009; Wipf and Nagarajan, 2009], but only in the spatial dimension to enhance the spatial localization of cortical activations. To our knowledge, little work has used temporal sparse models to detect the activations either in fMRI [Khalidov et al., 2011] or MEG and EEG [Bolstad et al., 2009].

In this article, SPFM is first evaluated with simulated fMRI data, and compared with standard general linear model (GLM) analysis and the empirical Bayes estimation (EBE) approach for hemodynamic deconvolution proposed in Gitelman et al. [2003]. Its feasibility and usefulness is then demonstrated with the same experimental data from a visuomotor paradigm analyzed previously in Caballero-Gaudes et al. [2011] to allow direct comparison between both PFM approaches.

THEORY

In BOLD fMRI, the signal from a voxel is usually modeled as a signal $x(t)$ resulting from the convolution of an "input" signal related to neuronal activity $s(t)$ and a regional HRF $h(t)$ [Boynton et al., 1996; Glover et al., 1999], summed with an additional term $e(t)$, representing noise arising from instrumental noise, pulsatile cardiac and respiratory fluctuations, other endogenous hemodynamic fluctuations of non-neuronal origin, and motion-related effects [Lund et al., 2006]. In fMRI, these continuous signals are sampled every TR seconds ($t = n$TR), so that the measured signal in fMRI acquisitions can be defined in discrete time as follows:

$$y(n) = x(n) + e(n) = \sum_{i=0}^{L-1} h(i)s(n-i) + e(n) \qquad (1)$$

where L is the discrete-time length of the HRF and $n = 1, \ldots, N$, where N is the number of observations in the fMRI time series. This model can be rewritten as

$$\mathbf{y} = \mathbf{Hs} + \mathbf{e}, \qquad (2)$$

where $\mathbf{y}$, $\mathbf{s}$, and $\mathbf{e}$ are column vectors of length N representing the voxel signal time series, the input signal and the noise, respectively. $\mathbf{H}$ is the Toeplitz convolution matrix of dimension $N \times N$ defined from the shape of the HRF [Caballero-Gaudes et al., 2011; Gitelman et al., 2003].

Sparse Regression With the DS

With Gaussian noise that is identically distributed and independent at each time point, the maximum likelihood estimate of the hemodynamic response to the input signal is obtained using the ordinary least squares (OLS) estimator that minimizes the residual sum of squares (RSS) between the modeled ($\mathbf{Hs}$) and measured ($\mathbf{y}$) time series [Hastie et al., 2001]. Because the number of HRF-shaped regressors in $\mathbf{H}$ is equal to the number of observations N, it is appropriate to regularize the OLS estimator with a penalization term based on the L_p-norm of the vector $\mathbf{s}$, $\|\mathbf{s}\|_p = (\sum_{i=1}^{N} |s_i|^p)^{1/p}$, such that the estimate ($\hat{\mathbf{s}}$) is given by

$$\hat{\mathbf{s}} = \min_{\mathbf{s}} \|\mathbf{y} - \mathbf{Hs}\|_2^2 \text{ subject to } \|\mathbf{s}\|_p \leq \alpha, \qquad (3)$$

where the L_0-norm of $\mathbf{s}$ gives the number of nonzero coefficients and the L_∞-norm gives $\max(|s_i|)$.

Previously, the ridge regression estimate with $P = 2$ (L_2-norm) was used for PFM [Caballero-Gaudes et al., in press]. However, if it is assumed that single trial BOLD responses are generated by brief (on the fMRI time scale) bursts of neuronal activation, then in Eq. (2) the neuronal-related signal $\mathbf{s}$ is a sparse vector with few coefficients, whose amplitudes are significantly different from zero. We consider that the vector $\mathbf{s}$ is sparse if its L_0-norm $\|\mathbf{s}\|_0 \ll N$ [Bruckstein et al., 2009] and sparse estimates of $\mathbf{s}$ can be obtained by solving Eq. (3) with $P = 0$. Unfortunately, this problem is nondeterministic polynomial-time-hard for the convolution model defined in Eq. (2) and solving it requires an exhaustive search for the optimal solution across all possible combinations of the columns of $\mathbf{H}$ [Bruckstein et al., 2009]. A practical solution is to solve Eq. (3) with $p = 1$, known as the least absolute shrinkage and selection operator (LASSO) [Tibshirani, 1996] or basis pursuit denoising (BPDN) [Chen et al., 1998], as the L_1-norm is a convex function and therefore it allows the use of efficient convex optimization solvers with rapid convergence

to the global solution [Bach et al., in press; Tropp and Wright, 2010].

The DS [Candes and Tao, 2007] is an alternative to the LASSO or BPDN estimators that computes an estimate of $\mathbf{s}$ by solving the following optimization problem:

$$\hat{\mathbf{s}} = \min_{\mathbf{s}} \|\mathbf{s}\|_1 \text{ subject to } \left\|\mathbf{H}^T(\mathbf{y} - \mathbf{Hs})\right\|_\infty \leq \delta, \qquad (4)$$

which can be equivalently rewritten in the same form as Eq. (3) as

$$\hat{\mathbf{s}} = \min_{\mathbf{s}} \left\|\mathbf{H}^T(\mathbf{y} - \mathbf{Hs})\right\|_\infty \text{ subject to } \|\mathbf{s}\|_1 \leq \alpha, \qquad (5)$$

with one-to-one correspondence between the non-negative regularization parameters α and δ. Importantly, the L_∞-norm term used in the DS in Eq. (4) or (5) is equivalent to the differentiation of the RSS (L_2-norm term used in Eq. (3)) with respect to $\mathbf{s}$. This highlights the theoretical and practical similarity of DS to LASSO or BPDN [Bickel et al., 2009; Bickel, 2007; James et al., 2009].

The rationale behind the DS is to ensure that the model residuals are only weakly linearly dependent on the columns of the model matrix $\mathbf{H}$, regardless of the probability distribution of the noise, that is, the correlation between the residuals and the shape of the HRF at any time point is below a certain threshold imposed by the regularization parameter δ. A nonzero coefficient in $\hat{\mathbf{s}}$ implies that an activation event is detected in the fMRI time series because the corresponding HRF-shaped regressor in $\mathbf{H}$ explains an important component of the variability of the voxel time series. Minimization of the L_1-norm forces the least informative coefficients of $\hat{\mathbf{s}}$ to exactly zero as δ increases, so that only the most relevant regressors of the model remain at the estimate. From Eq. (4), it can be seen that the null estimate, that is, $\hat{\mathbf{s}} = \mathbf{0}$, becomes a valid solution when $\delta \geq \|\mathbf{H}^T\mathbf{y}\|_\infty$.

Selection of the Regularization Parameter

Often, an appropriate value of δ is not known in advance and is chosen from a set of candidate values. Homotopy continuation procedures for sparse regression can be useful for that purpose [Efron et al., 2004; James et al., 2009; Osborne et al., 2000] because they enable the computation of the complete set of possible estimates for the L_1-norm regularization problem in Eq. (4), also known as the regularization path. In particular, the homotopy continuation algorithm developed in Asif and Romberg [2009] for the DS was used to obtain our results. The use of homotopy procedures along with model selection criteria [Zou et al., 2007] was found useful to choose an appropriate value of δ.

Hereafter, we use the subscript δ to denote that a variable depends on the regularization parameter. For a given δ, let $\hat{\mathbf{s}}_\delta = \{\hat{s}_i \neq 0\}$ be the vector containing the nonzero coefficients of the estimate $\hat{\mathbf{s}}$, also known as the support

set of $\hat{\mathbf{s}}$, and let $\mathbf{H}_\delta$ be the matrix including the subset of columns of $\mathbf{H}$ corresponding to these nonzero coefficients. The homotopy algorithm is initialized with $\delta = \|\mathbf{H}^T\mathbf{y}\|_\infty$, which corresponds to the null estimate of $\hat{\mathbf{s}}$. As δ decreases toward zero, the procedure alternatively updates the solution to the DS problem in Eq. (4) (the primal problem) and its dual [Asif and Romberg, 2009]. The algorithm can be divided in two main phases: the primal update and the dual update. In the primal phase, the solution to the DS problem in Eq. (4) (the primal solution) is updated using the conditions established by the DS constraint and the current dual solution (i.e., the vector of Lagrange multipliers of the DS problem). In the second phase, the dual solution is updated based on the current primal solution and constraints in the dual variables. Specifically, at critical values of δ one coefficient of the DS solution is either included or removed from the support set (i.e., coefficient becomes different or equal to zero). The amplitudes and signs of the nonzero coefficients are then adapted accordingly to optimize fit to data [Asif and Romberg, 2009]. Since more HRF-shaped regressors are fitted to the fMRI voxel time series as δ decreases, the number of degrees of freedom to fit the time series also increases. This could result in overfitting of the voxel time series at very low values of δ.

To avoid overfitting, the choice of δ was based on model selection criteria that accounted for the effective degrees of freedom used to fit the voxel time series and the estimated residuals. An analytical expression for the effective degrees of freedom for the DS is not obvious because it is a nonlinear subset selection procedure [Hastie et al., 2001]. Mimicking the definition of the effective degrees of freedom for linear estimators, such as least-squares or ridge regression [Hastie et al., 2001], we propose approximating the degrees of freedom of the DS as

$$\mathrm{df}_\delta = \mathrm{Tr}\Big(\mathbf{H}_\delta\big(\mathbf{H}_\delta^T\mathbf{H}_\delta\big)^{-1}\mathbf{H}_\delta^T\Big), \qquad (6)$$

based on numerical simulations and building on theoretical results for the closely related LASSO technique [Bickel et al., 2009; Candes and Tao, 2007; James et al., 2009; Zou et al., 2007]. The rank of the matrix in Eq. (6) is, at most, the number of columns in $\mathbf{H}_\delta$, that is, the number of nonzero coefficients of the DS estimate. The effective number of degrees of freedom df_δ is computed with Eq. (6) for all estimates in the regularization path. Subsequently, the optimal value of δ (δ^*) is selected via adaptive model selection criteria [Zou et al., 2007]:

$$\delta^* = \min_\delta Ln\Big(\|\mathbf{y} - \mathbf{H}_\delta\hat{\mathbf{s}}_\delta\|_2^2\Big) + \frac{K}{N}\mathrm{df}_\delta. \qquad (7)$$

In this work, we compared two traditional model selection criteria: the Akaike information criterion (AIC) with $K = 2$ [Akaike, 1974] and the Bayesian information criterion (BIC) or minimum description length with $K = Ln(N)$ [Schwarz, 1978]. Once δ is chosen, the corresponding opti-

mal estimate of the DS is automatically selected from the regularization path.

Debiasing

Once an estimate is obtained with the DS for each voxel time series (selected with either BIC or AIC in this study), a debiasing step was performed to overcome the tendency of sparse estimators to underestimate the true value of the nonzero coefficients [Candes and Tao, 2007; James et al., 2009]. Before debiasing, additional effects of noninterest that can explain some variability of the voxel time series (e.g., the realignment parameters accounting for movement-effects) were included into the model in the form of a matrix $\mathbf{X}$. A new signal model can be written conditional on the DS estimate as $\mathbf{y} = \mathbf{H}_D\mathbf{s}_D + \mathbf{e}$, where the debiasing design matrix is $\mathbf{H}_D = [\mathbf{H}_\delta \ \mathbf{X}]$ and the weights are $\mathbf{s}_D = [\hat{\mathbf{s}}_\delta^T \mathbf{s}_X^T]^T$. The coefficients of this model were computed using simple OLS, that is, $\hat{\mathbf{s}}_D = \big(\mathbf{H}_D^T\mathbf{H}_D^T\big)^{-1}\mathbf{H}_D^T\mathbf{y}$. The zero coefficients that were not included in the support set of the selected DS estimate remained as zero. Figure 1 shows an example of the operation of the SPFM method for a simulated fMRI time series.

METHODS

Simulated and experimental fMRI data were used to evaluate the performance of the DS algorithm for SPFM and to investigate the differences between using AIC or BIC to select the regularization parameter δ. In all cases, the convolution matrix $\mathbf{H}$ was defined using a two Γ-variate canonical HRF with standard SPM parameters including a time-to-peak of 5 s and duration 32 s [Friston et al., 1998a], sampled at the corresponding TR. The discrete length of the HRF was equal to $L = 17$ samples.

Simulated Data

Three simulation experiments were performed using simulated fMRI time series $y(t)$ created according to Eq. (1). Each time series was characterized by the number of events in the neuronal-related input signal $s(t)$, the shape of the HRF with which $s(t)$ is convolved $\tilde{h}(t)$ (and corresponding convolution matrix $\tilde{\mathbf{H}}$), which may be different from the HRF used for the deconvolution ($h(t)$ and convolution matrix $\mathbf{H}$) when investigating the effect of mismatches between the simulated and deconvolved HRF, and the temporal signal-to-noise ratio of the time series. The duration of the simulated fMRI time series was 256 s, and $s(t)$ and $\tilde{h}(t)$ were initially created at 200 ms intervals. For the input signal $s(t)$, data sets containing zero events (no events) to 10 events, each of duration 2 s and constant amplitude but random polarity were generated. The onsets of the events were randomly positioned along the time series, without enforcing a minimal time interval between simulated events. Varying the number of events in the input signal allows an evaluation of the performance of the SPFM method with AIC and BIC criteria with variable
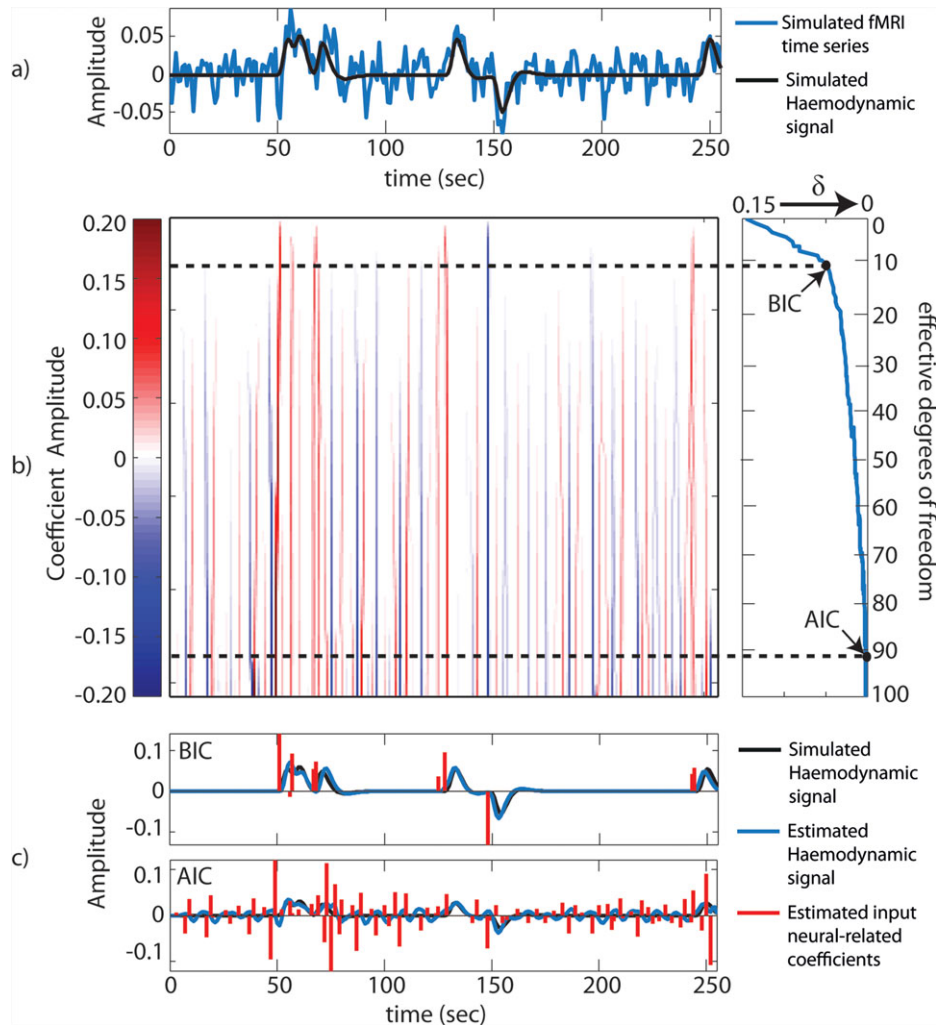
Figure 1.

Illustration of the SPFM method. (**a**) For each fMRI voxel time series (simulated here with six activation events of duration 2 s convolved with the canonical HRF with time-to-peak of 5 s, *tSNR* of 50, TR 2 s and 128 time points), the regularization path of possible DS estimates of the neuronal-related signal is computed by iteratively solving the DS algorithm with decreasing values of δ (**b**). As the size of the support set of nonzero coefficients increases with decreasing δ, the effective degrees of freedom also increase because more canonical HRF are fitted to the fMRI time series. (**c**) The appropriate estimate of the neuronal-related coefficients (red) and the neuronal-related haemodynamic signal (blue) are chosen according to model selection criteria. The BIC correctly detected the six activation events, whereas the AIC completely fails to estimate the stimulus signal and overfits the fMRI time series. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

levels of sparsity in the vector **s**. Three different shapes were simulated for the hemodynamic response $\tilde{h}(t)$, based on the canonical HRF [Friston et al., 1998a] with variable time-to-peak of the first gamma function ($a_1 = 3$ s, 5 s (identical to the HRF model used for the DS fitting), and 8 s). This allowed assessment of the robustness of the SPFM method to HRF variability.

A noise term, $e(t)$, was added to the simulated signal time series. Noise was created as the sum of uncorrelated Gaussian noise and sinusoidal signals to simulate a realis-

tic noise model with both thermal noise and cardiac and respiratory physiological fluctuations, respectively. The sinusoidal term was generated as

$$\sum_{i=1}^{4} \frac{1}{2^{i-1}}\left(\sin\left(2\pi f_{r,i}t + \phi_{r,i}\right) + \sin\left(2\pi f_{c,i}t + \phi_{c,i}\right)\right) \quad (8)$$

with up to fourth-order harmonics per cardiac and respiratory component. Each harmonic $f_{r,i}$ and $f_{c,i}$ was randomly generated following Normal distributions with variance 0.04 and mean if_r and if_c, for $i = 1, \ldots, 4$, and where the

fundamental frequencies were $f_r = 0.3$ Hz for the respiratory component [Birn et al., 2006] and $f_c = 1.1$ Hz for the cardiac component [Shmueli et al., 2007]. The phases of each harmonic ϕ were randomly selected from a Uniform distribution between 0 and 2π radians. A range of temporal signal-to-noise ratios (tSNR) between 30 and 80 was simulated, which corresponds to a range of contrast-to-noise ratios between 1.8 and 4.8 assuming a BOLD signal change of 6% typically observed at 7T (as used in the experimental work described here) [van der Zwaag et al., 2009a]. To simulate physiological noise that is proportional to BOLD signal change, a variable ratio between the physiological noise (σ_P) and the thermal (σ_0) was modeled as $\sigma_P/\sigma_0 = a(tSNR)^b + c$, where $a = 5.01 \times 10^{-6}$, $b = 2.81$, and $c = 0.397$. This physiological-thermal noise model was extracted following the experimental measures of the physiological-to-thermal noise ratio at 7T in Table 3 in Triantafyllou et al. [2005]). The noise term, $e(t)$, was then added to the simulated time series, $s(t)$, to create an fMRI time series, $y(t)$, which was then downsampled to a TR of 2 s ($N = 128$ time points).

Each voxel time series was analyzed with the SPFM procedure described in Theory section. For comparison, the simulated time series were also analyzed with the EBE method introduced in Gitelman et al. [2003], which also enables deconvolution of the underlying signal from the fMRI BOLD time series without prior information on the timing of the events. This approach is based on a two-stage hierarchical model that imposes Gaussian priors on the coefficients of the underlying signal. Posterior estimates are iteratively computed via EBE and restricted maximum likelihood [Gitelman et al., 2003]. In our simulations we used the functions spm_peb_ppi.m and spm_PEB.m implemented in SPM8 (FIL, UCL, UK). The time resolution used for the EBE deconvolution was set equal to the TR, as it was done for SPFM.

We also analyzed these data with a standard GLM fitted with OLS. The GLM included one regressor per nonzero coefficient in the simulated input signal. Each regressor was created as the convolution of Dirac impulses at the time of the nonzero coefficients in the simulated input signal with a canonical HRF with time-to-peak of 5 s as used in the model for SPFM. The GLM analysis allowed comparison with the results that could be achieved if perfect knowledge of the timing of the events were available a priori (i.e., in a non paradigm free scenario). This information is required neither for SPFM nor for EBE, where the analysis is carried out blind to onsets and duration of neural stimuli.

Simulation 1

The first simulation was designed to evaluate the ability of SPFM to detect active voxels. A pattern of 36×36 voxels was created comprising nine squares of 12×12 voxels distributed in three rows and three columns. This layout is shown on the left of Figure 2. In the centre of each 12×12 square, a smaller 4×4 square was simulated as having "active" voxels and the surrounding pixels were simulated as "non active" with zero events. The active pixels in each 4×4 square were characterized by the number of events (2, 6, or 10) simulated in the neuronal-related input signal and the HRF shape (canonical HRF with time-to-peaks (a_1) of 3, 5, and 8 s). A noise term with random Gaussian noise and sinusoidal noise was added to each time series as described in Eq. (8). We generated 1,000 repetitions of this pattern at three different tSNRs (30, 55, and 75) with random noise realizations across repetitions. For the GLM analysis, the information about the onsets and number of events that is required to define the regressors was provided accordingly. We computed the average rate that each pixel was labeled as active (detection rate) across the 1,000 repetitions, where a pixel was labeled as active based on an F-statistic ($P < 10^{-4}$, uncorrected) of the hypothesized model for the GLM analysis or estimated model for the SPFM analysis. As each square comprised 12×12 pixels this threshold would correspond to a P value of 1.44×10^{-2} with Bonferroni correction.

To label a pixel as active for the EBE approach, we computed the posterior probability of the EBE model and compared it with the posterior probability of a "null model" comprising a single column of ones (NULL). The log-evidence of the model m for a time series y (a measure which indicates the preference shown by the data for the given model) is $L(m) = \log P(y|m)$, which can be approximated by the "free energy" (a measure that expresses the uncertainty of the data averaged over instances of the generative model and the complexity of the model) after convergence of the EBE estimation [Penny et al., 2007]. Assuming that both models are equiprobable, it can be shown that the posterior probability of the EBE model given the time series y is given by (see Eq. (16) in Penny et al. [2007])

$$p(\text{EBE}|y) = \frac{\exp(L(\text{EBE}))}{\exp(L(\text{EBE})) + \exp(L(\text{NULL}))} \quad (9)$$

where the model evidences, $L(\text{EBE})$ and $L(\text{Null})$, were computed using spm_PEB.m. A pixel was labeled as active when $P(\text{EBE}|y) > 0.99$, that is, when the difference between both model evidences was $L(\text{EBE}) - L(\text{NULL}) > 4.6$ [Penny et al., 2007].

Simulation 2

The second experiment evaluated the sensitivity and specificity of the SPFM method to detect the exact location in time of the simulated events, that is, the time of the nonzero coefficients in the simulated neuronal-related input signal. We generated 1,000 fMRI time series for each simulation scenario (number of events in the simulated neuronal-related signal (2, 6, and 10 events of duration 2s), HRF shape (canonical HRF with time-to-peaks (a_1) of 5 and 8 s), and tSNR (30–80)). A false positive (FP) event was defined as occurring when a nonzero coefficient in
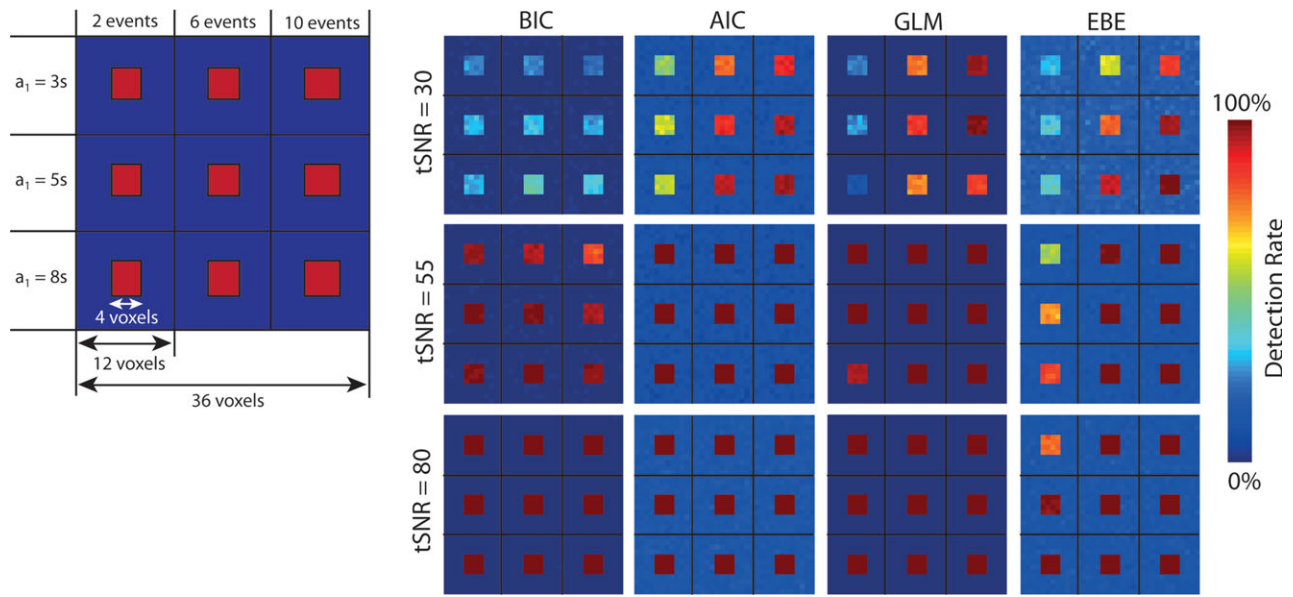
Figure 2.

Detection rates of Simulation 1. A pattern of 36 × 36 pixels was created comprising nine squares of 12 × 12 pixels distributed in three rows and three columns, as shown at the left of the figure. One thousand repetitions of this pattern were generated at three different *tSNRs* (30, 55, 80). The information about the onsets and number of events for the GLM-based analysis was adapted accordingly within each 12 × 12 square. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the estimated signal $\hat{s}$ did not correspond to a nonzero coefficient in the simulated input signal s after subsampling it to TR, and a false negative (FN) event as when a nonzero coefficient in s did not correspond with a nonzero coefficient in $\hat{s}$. The FP rate (FPR = number of false positives/number of positives (nonzero coefficients in $\hat{s}$)) and FN rate (FNR = number of false negatives/number of negatives (zero coefficients in $\hat{s}$)) were used to calculate the temporal sensitivity (1-FNR) and specificity (1-FPR). The tradeoff between temporal sensitivity and specificity was summarized in terms of Receiver Operating Characteristics (ROC) curves.

For the EBE approach, the assumption of Gaussian priors in addition to the use of cosine bases to expand the representation of the underlying signal (Gitelman et al., 2003) cause the estimated signal to be smooth and those nonrelevant coefficients to not be reduced to exactly zero as achieved with L_1-norm regularization. Therefore, some type of thresholding procedure is required to decide whether an event has occurred at a given time point i. Assuming a prior expectation equal to 0, we computed the marginal posterior probability that the coefficient $s_{i,EBE}$ of the posterior mean computed with EBE at time point i exceeds a threshold γ. This posterior probability is given by $P(s_{i,EBE}|y) = 1-\Phi((\gamma-s_{i,EBE})/\sqrt{C_{ii}})$, where C_{ii} is the i-th coefficient of the main diagonal of the posterior covariance matrix of the estimate (see Friston et al. [2002] for analytical expressions), and $\Phi(.)$ is the cumulative distribution function of the Normal distribution with zero mean and unit variance. Note that a time series of posterior probabilities is computed for each simulated time series, giving rise to "temporal" posterior probability maps [Friston and Penny, 2003]. The threshold γ, which determines what is a relevant estimate, was set at one standard deviation of the posterior residuals. An event was detected at time point i if $P(s_{i,EBE}|y) > 0.9$.

Simulation 3

Using the same simulated fMRI data as in Simulation 2, we computed the mean square error (MSE) between the simulated hemodynamic signal $\tilde{H}s$ at TR resolution and its estimate $H\hat{s}$:

$$\text{MSE} = \frac{1}{N_b}\sum_{i=1}^{N_b}\left\|\tilde{H}s - H\hat{s}\right\|_2^2. \tag{10}$$

Therefore, this third simulation investigated the performance of SPFM, EBE and GLM in estimating the hemodynamic events that were simulated in the fMRI time series in case of a perfect HRF model or HRF model mismatches. For SPFM, mismatches in the HRF model may lead to FPs or FNs in the detection of the nonzero coefficients of the simulated input signal. For example, if the simulated HRF exhibits a shorter time to peak than the

HRF used in the model, the simulated BOLD events could still be detected by the SPFM method and the voxel be labeled as active as evaluated in Simulation 1, but with an earlier onset than simulated, which could result in a FP and FN as evaluated in Simulation 2.

Computational cost

To compare the computational cost of SPFM, EBE, and GLM, we measured the time in seconds required to fit a single voxel time series as a function of the number of scans (N), where N ranged between 64 and 1,024 scans. All algorithms were implemented in Matlab R2009b and the simulations were done using a 2.8-Ghz Intel Core 2 Duo MacBook Pro with 4GB RAM.

Experimental Data

The SPFM method was evaluated on the same five individual fMRI datasets that were previously used to evaluate the PFM approach [Caballero-Gaudes et al., 2011]. The five subjects provided informed consent under the approval of the University of Nottingham ethics committee.

Data acquisition and paradigm

fMRI data were acquired on a 7T Philips scanner (Best, Netherlands) using a 16-channel head coil (Nova Medical, MA) and a single-shot gradient-echo EPI sequence (2 mm isotropic resolution, SENSE factor = 1.5, TE = 30 ms, TR = 2 s, flip angle = 80°, N = 342 time points). Twenty oblique slices were acquired at approximately +15° to the canto-meatal line above the corpus callosum extending from the superior frontal to occipital cortices. A T2*-weighted scan was also acquired as anatomical reference (three-dimensional spoiled FLASH, 1 mm isotropic resolution). Subjects' heads were secured in place using foam pads inside the head coil. Cardiac and respiratory data were recorded using a respiratory belt and a pulse oximeter for physiological noise correction. Electromyography (EMG) measurements were acquired for both hands (left extensor (LE), right extensor and right flexor (RF) muscles) [Caballero-Gaudes et al., 2011].

The fMRI experimental paradigm lasted 684 s and consisted of single-trial finger-opposition tapping events interleaved with three periods of rest. After an initial rest period of 140 s, the subject was visually cued to perform finger tapping at 140 and 180 s (trial duration: 4 s). A second rest period of 200 s followed in the time interval from 184 to 384 s. At 384 s, a message ("TAP at will") was projected onto a screen instructing subjects about the beginning of a period of 300 s when they had to perform two self-paced finger tapping trials of 4 s, at a time of their choosing, until the end of the run. Subjects were asked to fixate on a cross during rest periods. The visual instructions were projected from an LCD projector onto a screen located inside the scanner room, which subjects viewed through prism glasses with angled mirrors. Subjects were instructed on the paradigm prior to the scanning session. The EMG recordings captured that Subject A performed an additional self-paced tapping at the end of the run, and Subject B performed four self-paced finger tapping events instead of the two events instructed. Both cases were also confirmed by personal questionnaire after the MRI session.

SPFM data analysis

fMRI datasets were initially corrected for motion, physiological cardiac and respiratory fluctuations with RETROICOR [Glover et al., 2000], and detrended via deconvolution for sine and cosine waveforms with period equal to the scan duration and up to fourth-order Legendre polynomials, reducing the serial correlations of the time series. Finally, voxel time series were normalized to give percent signal change. These steps were performed using AFNI (NIMH/NIH, Cox, [1996]).

The preprocessed time series were then analyzed with the SPFM method described above that was implemented using in-house programs written in Matlab (The Mathworks, Natick, MA). The BIC criterion was used to select the regularization parameter of the DS because the simulation results showed that the SPFM method performed better with BIC rather than with AIC (see Results). To reduce computational cost, the homotopy procedure was stopped when the regularization parameter was below the maximum absolute deviance (MAD) estimate of the noise standard deviation [Donoho and Johnstone, 1994], or when the number of nonzero coefficients in the estimated input signal was more than half the fMRI time points. The six rotation and translation realignment parameters estimated during motion correction were also included as nuisance regressors ($\mathbf{X}$) in the debiasing model $\mathbf{H}_D$. Spatial clustering with a minimum cluster size of two contiguous voxels with nonzero coefficients (no amplitude threshold) of the estimate (extent-threshold) was applied at each time point to reduce possible isolated FPs.

As for PFM [Caballero-Gaudes et al., 2011], the outcome of the SPFM analysis is a sequence of brain activation maps that depicts the spatiotemporal dynamics of the estimated neuronal-related signal at each time point. To reduce the dimensionality of the four-dimensional results and to assist in identifying of periods when significant brain activation occurs, two activation time series (ATS) were created for each dataset, counting the number of voxels with positive and negative nonzero coefficients at each time point [Caballero-Gaudes et al., 2011].

For each peak in the ATS corresponding to a finger tapping event, a statistical map was created to illustrate the spatial relevance of the peaks. We first computed t-statistics for each of the detected events (nonzero coefficients in the input signal) estimated by the DS (i.e., the regressors in $\mathbf{H}_\delta$ based on the debiasing model) and assumed zero t-statistics for the zero coefficients. Next, t-statistics were converted to z-scores because the number of detected

events differed between voxels and consequently the degrees of freedom of the *t*-statistics varied between voxels. To condense the whole sequence of maps corresponding to the finger tapping event into a single map, the value in each voxel was set to the maximum of the *z*-scores during the period around a finger tapping event when more than 100 voxels showed positive or negative activation in the ATS.

GLM analysis

The preprocessed datasets were also analyzed with a GLM-based approach. For each event, two regressors were created from the convolution of impulse events with the SPM-canonical HRF and its temporal derivative [Friston et al., 1998a], with onset corresponding to the start of tapping as recorded in the EMG. Two regressors were similarly created at the time of the visual cue "TAP AT WILL". Additionally, the six translation and rotation parameters estimated during rigid-body realignment for motion correction were included as nuisance regressors. This GLM was fitted without modeling noise serial correlations (three-dimensional-Deconvolve function in AFNI) in concordance with the SPFM model. To capture the polarity of the BOLD signal change, statistical GLM maps were created for each event depicting the T-statistic of the canonical HRF regressor. Each map was thresholded according to an F-test of the two regressors ($P < 0.01$) after FDR correction for multiple comparisons [Benjamini et al., 2006; Benjamini and Hochberg, 1995]. A minimum cluster size of two contiguous voxels was used for spatial clustering.

To quantify the agreement between the SPFM and GLM maps, for each tapping event, we measured the number of voxels showing activations in each map (|GLM| and |SPFM|) and the number of overlapping voxels (|SPFM ∩ GLM|). Then, we computed the Dice measure (D = 2* |SPFM ∩ GLM|/|SPFM| + |GLM|).

RESULTS

Simulated Data

Figure 1 shows an example of the SPFM method for a simulated fMRI time series with six events each of duration 2 s (Fig. 1a). The corresponding regularization path is depicted in Figure 1b. As the regularization parameter δ decreases from $\|\mathbf{H}^T\mathbf{y}\|_\infty = 0.15$ to 0, the number of nonzero elements in the support set increases (left graph in Fig. 1b), more HRFs are fitted to the time series and more degrees of freedom are used to fit the data (right graph in Fig. 1b). As shown in Figure 1c, selecting δ according to BIC results in the correct detection of the six simulated activation events therefore obtaining an accurate deconvolution of the original simulated hemodynamic component of the signal, whereas the use of AIC leads to overfitting

of the fMRI time series since an excessive number of hemodynamic events are detected.

Simulation 1

The results of Simulation 1 are shown in Figure 2 which maps the average rate of labeling a voxel as active over 1,000 repetitions (*F*-test, $P < 10^{-4}$ uncorrected) for SPFM (BIC and AIC) and EBE ($P(\text{EBE}|y) > 0.99$) without prior knowledge of the timing of any events, and GLM with timing of events given in the model, for various *tSNR* values and simulated HRF shapes.

With simulated parameters typical of 7T, little difference in the detection rate is observed between those squares for which the simulated and fitted HRF were the same (middle row in each block) and those with an HRF mismatch (top and bottom rows), which illustrates the robustness of the method to HRF mismatches. The pattern of activations revealed with the SPFM method using BIC at *tSNR* of 55 and 80 resemble the ideal pattern, and the detection rate obtained for SPFM with BIC and GLM are similar. SPFM with BIC and GLM showed high specificity since, on average, those voxels in the background with no events were not declared as activations. In contrast, larger detection rates can be observed in the background for the patterns computed with AIC and the EBE approach, indicating that both EBE and SPFM with AIC feature lower specificity than SPFM with BIC. Very low detection rates are achieved with BIC at a *tSNR* of 30 in the true activated voxels. The use of AIC increases the sensitivity in detecting true positives at a low *tSNR* of 30, but this is achieved at cost to decreasing specificity, that is, augmenting FPs in the background compared with BIC. The patterns obtained with SPFM AIC illustrate higher sensitivity than those computed with EBE without decrease in specificity. With GLM, knowing the timing of the events improves the detection rate of the true active voxels at *tSNR* of 30 when the time series have six or 10 events, but not with two events. This shows that at low *tSNR*, the large level of noise existing in the true active voxels limits the power of the F-statistics of the model estimated with SPFM using BIC so that the voxel is not labeled as active, even though the detected events have a high probability of being true events (see ROC values of SPFM with BIC at low *tSNR* in Simulation 2).

Simulation 2

The temporal specificity and sensitivity results of Simulation 2 are shown in the ROC curves of Figure 3. The SPFM method using BIC detects the nonzero coefficients of the activation events with FPRs lower than 4% (specificity > 0.96) in a nomismatch scenario, that is, when the simulated HRF is identical to the fitted HRF. If AIC is adopted, the accuracy in detecting the time of the events decreases to values of FPR around 10% with little increase in sensitivity. As the *tSNR* of the time series decreases, the
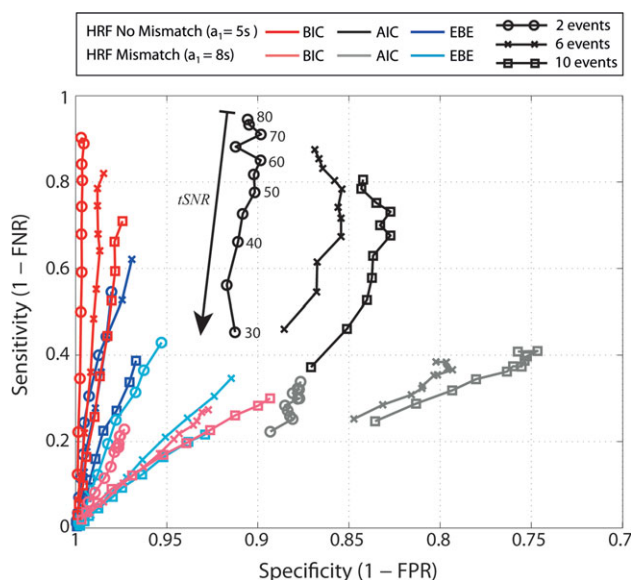
Figure 3.

ROC curves of temporal sensitivity and specificity to detect the onset of the simulated activation events (Simulation 2) for SPFM using the BIC and AIC criteria and for EBE. ROC curves were computed with no HRF mismatch where the simulated BOLD events were generated with the same HRF as the model (canonical HRF with time-to-peak of 5 s), whereas those corresponding to a mismatch scenario where the simulated BOLD events were generated with an HRF different from the model (canonical HRF with time-to-peak of 8 s). Line markers indicate number of activation events: 2 (circles), 6 (crosses), and 10 (squares). Decreasing sensitivity in each curve corresponds to *tSNR* decreasing from 80 to 30, illustrated by vertical arrow on the graph. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

curves show reduced temporal sensitivity without reduced temporal specificity. In particular, using the BIC criterion at *tSNR* < 35, the sensitivity drops whilst the specificity remains fairly constant, suggesting strict control of FP detections with BIC. However, if the simulated HRF has a time-to-peak of 8 s compared with a fitted HRF with time-to-peak of 5 s, considerable degradation in both temporal sensitivity and specificity occurs, as expected because the time points of the simulated events (nonzero coefficients) cannot be exactly determined in this situation. With a difference in the simulated and fitted time-to-peaks of 3 s and a TR of 2 s, it is probable that any detected event will apparently be at a time point which is shifted with respect to its actual timing. Nevertheless, FPR values lower than 5% can be achieved with BIC when there is a mismatch in the simulated and fitted HRF. As expected in sparse regression techniques, a comparison between the ROC curves for varying number of events indicates SPFM provides better specificity–sensitivity when the input signal has fewer events (i.e., is more sparse). In case of perfect

knowledge of the HRF, the EBE approach shows considerably lower temporal sensitivity than both SPFM using BIC and AIC. EBE demonstrates higher temporal specificity than SPFM with AIC in both HRF situations. However, the temporal specificity obtained with EBE is generally inferior to that obtained with SPFM with BIC at all *tSNR* values particularly for smaller number of events.

Simulation 3

Figure 4 plots the MSE curves computed in Simulation 3. The similarity between the MSE curves computed with and without mismatches between the simulated and fitted HRF events also demonstrates that SPFM is robust against HRF variability and produces accurate estimates of the neuronal-related hemodynamic component of the signal.

For few simulated fMRI time series, we observed the EBE algorithm did not converge to sensible estimates of the simulated signals generating outliers in the distribution of MSE values. Hence, the EBE curves plot the trimmed mean: generally this involved excluding the highest and lowest 0.5% MSE values and five values were excluded, but when no events were simulated in the time series the highest and lower 5% MSE values and fifty values were excluded because outliers were more frequently observed.

As anticipated from the ROC curves in Figure 3, the SPFM method better detects a small number of events with the BIC criterion, compared with the AIC criterion (Fig. 4a,b,d,f). When a mismatch between the simulated and fitted HRF exists, SPFM obtains a better estimate of the simulated hemodynamic signal than a GLM analysis (fitted via OLS) at high *tSNR* and particularly for low number of activations (Fig. 4d,f). The MSE curves also highlight that if numerous events are expected, it may be beneficial to use the AIC criterion to select δ in very high tSNR scenarios (*tSNR* > 80) (Fig. 4e,f). The estimates obtained with EBE are less accurate than those obtained with SPFM using BIC in all scenarios except at *tSNR* < 40 with more than six events. As expected, the EBE approach always estimates a hemodynamic signal different from zero although the time series includes no events. Whereas this undesired behaviour is similar to the one found with SPFM using AIC at low *tSNR*, its operation deteriorates remarkably at high *tSNR* (Fig. 4a). At very high *tSNR* the MSE curves of the SPFM method with BIC converge to the GLM curves obtained with perfect knowledge of the onsets of the events and HRF shape (Fig. 4c,e).

Experimental Data

SPFM detected all finger tapping events for all subjects, which were all evident in the EMG recordings, including those BOLD signal changes associated to the visual cue "TAP AT WILL". Figure 5 plots the positive (black) and negative (red) ATS for the five datasets, along with the corresponding EMG recordings of the RF and LE muscles. SPFM also revealed transient activations during periods
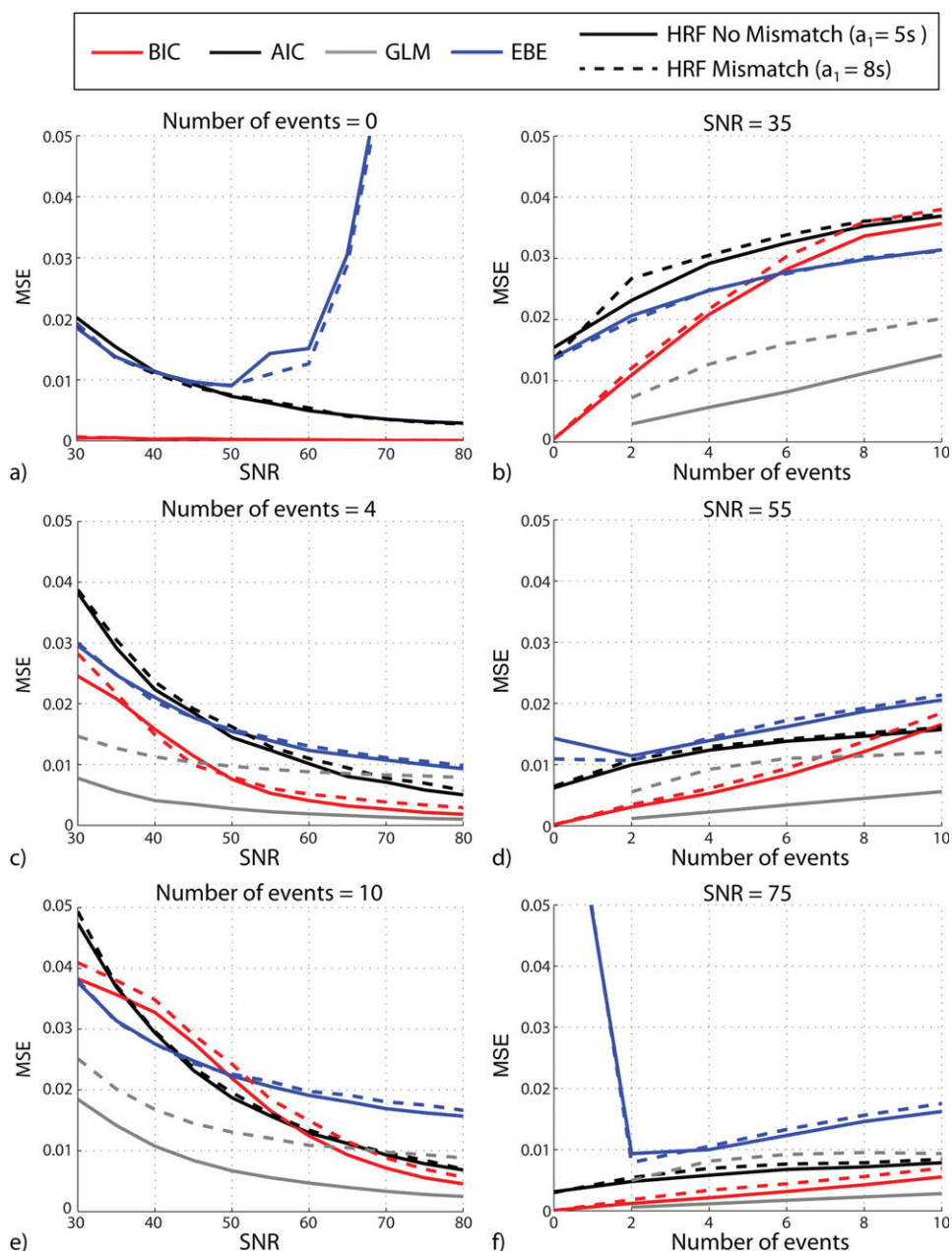
Figure 4.

MSE between the simulated and estimated hemodynamic component of the signal (Simulation 3) obtained for SPFM using the BIC and AIC criteria, EBE and GLM estimated via OLS, as a function of *tSNR* for (**a**) 0, (**c**) 4, and (**e**) 10 activation events; and as a function of the number of events for *tSNR* of (**b**) 35, (**d**) 55, and (**f**) 75. Dashed lines correspond to HRF Mismatch (HRF time to peak of 8s), whereas solid lines correspond to no HRF mismatch. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

when the subject was instructed to remain at rest. The spatial maps of these events were not random, but formed spatial clusters in areas of sensorimotor processing, in some cases simultaneous to muscle activity recorded in the EMG. Furthermore, these activation patterns encompassed cortical regions typically associated with the visual,

default-mode and dorsal attention networks [Petridou et al., 2011].

Figure 6 shows the SPFM with BIC and GLM maps for the finger tapping events of Subject E for three representative slices out of 20 slices, overlaid over the corresponding T2*-weighted anatomical image. The maps for the rest of
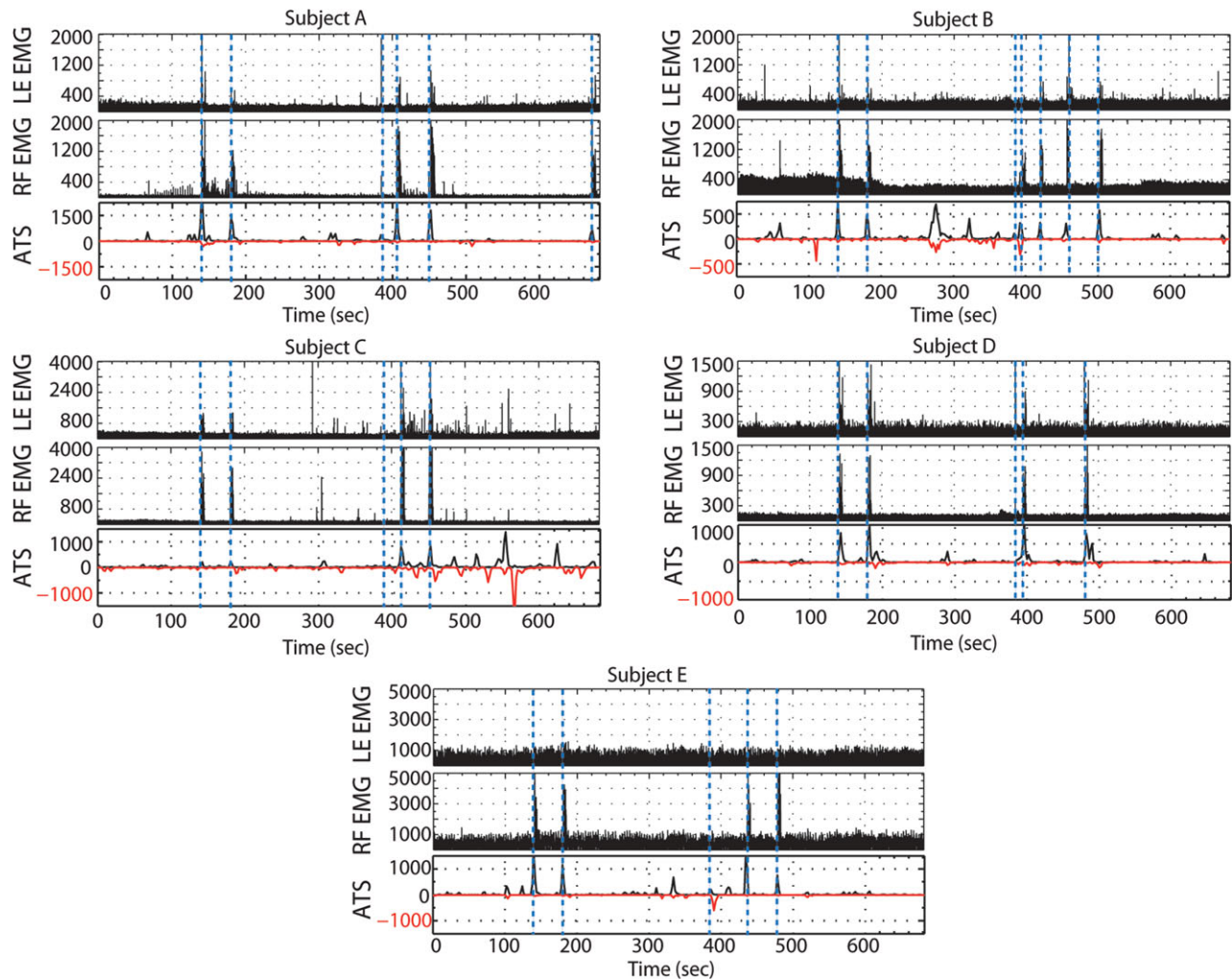
Figure 5.

ATS computed with SPFM using the DS algorithm with BIC along with the EMG recordings of the left extensor (LE) and right flexor (RF). Each ATS counts the number of voxels with positive activation (black, positive *y*-axis) and negative activation (red, negative *y*-axis) revealed by the SPFM method at each time point. The blue dashed lines mark the times of the task-related events (visually cued finger tapping, self-paced finger tapping, and visual cue "TAP at WILL"). The SPFM method was able to detect all finger tapping events for all subjects, and timely with the EMG recordings. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the datasets are given as Supporting Information. In general, the figures illustrate the clusters of activations found in the SPFM maps were in the same regions as those shown in the GLM maps. Table I lists the Dice coefficients for each finger tapping event. Across all finger tapping events and subjects, we observed an average Dice measure of 50.4% (std: 13.7%). The ratios varied considerably across trials and subjects, and individual trial and subject data showed that some subjects consistently show larger regions for GLM and others consistently show larger regions for SPFM. The main clusters of activations related to the finger tapping events shown both in the SPFM maps and the GLM maps were located in areas of visual and sensorimotor processing such as the supplementary motor area, premotor cortex, primary motor and somatosensory cortices, superior parietal lobule, and occipital cortex.

Computational Cost

Figure 7 plots the median time for 1,000 random simulated time series as a function of the number of scans of the voxel time series (*N*), and also indicates the 5% and 95% at each point. Note that the figure is plotted in logarithmic scale.

It can be seen that SPFM using the primal-dual pursuit algorithm [Asif and Romberg, 2009] to solve the DS is
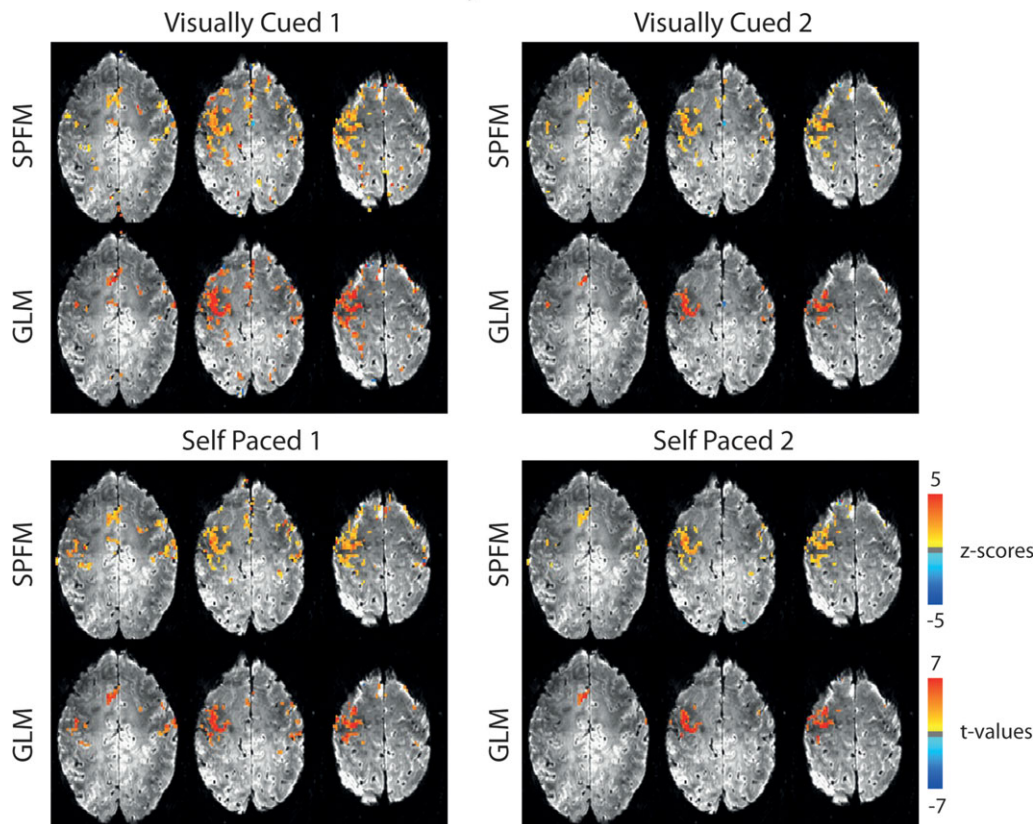
Figure 6.

SPFM and GLM maps for the visually cued and self paced finger tapping events performed by Subject E (see Supporting Information for the rest of subjects), overlaid over the corresponding T2*-weighted anatomical image. SPFM maps display the maximum z-score (normalized t-statistic) of the coefficients esti- mated during each activation event. GLM maps display t-statistics of the canonical HRF regressor for those voxels where the F-test of the informed basis set regressors has a $P < 0.01$, FDR corrected. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

approximately 15 times faster than the EBE approach when N is approximately 200, but this difference rises to 30 times at $N = 1,024$. For 342 scans, as in the experimental datasets, SPFM takes a median time of 0.56 s to compute the deconvolved estimate, whereas the time for EBE is 9.42 s. The computational time of GLM is substantially lower than that of SPFM (118 times faster at $N = 64$ increasing to 813 times faster at $N = 1,024$) as the onset times of the activations and the neuronal related signal (stimulus function) are known in advance and not blindly estimated as in SPFM and EBE.

DISCUSSION

Both simulation and experimental results demonstrate the validity of the SPFM method, using the DS in combination with BIC, to detect single trial BOLD responses without prior information on the timing of the events. SPFM does not require thresholding or the definition of a baseline period, thus enhancing our previous PFM approach [Caballero-Gaudes et al., 2011]. Additionally, spatial averaging was necessary in PFM to improve statistical power. As amplitude thresholding is no longer required, SPFM makes the exploration and interpretation of the results simpler. Our simulations demonstrated that this benefit is obtained without compromising temporal and spatial specificity and sensitivity (Figs. 2 and 3). This resembles the automatic selection features observed in spatial decoding approaches with sparse regression [Yamashita et al., 2008].

SPFM assumes active/nonactive type of underlying activity, so no threshold is required, whereas the EBE approach expands the underlying signal in terms of a set of cosine functions and establishes Gaussian priors on the coefficients, which makes the estimates smoother and not reduced to zero similar to the ridge regression used for PFM [Caballero-Gaudes et al., 2011]. In contrast to SPFM, the EBE approach requires a threshold to be set on the

TABLE I. Dice coefficients between SPFM and GLM maps for each finger tapping event and subjectDice coefficients: 2*|SPFM ∩ GLM| / |GLM| + |SPFM|

Tap/subject	A	B	C	D	E
VC1	0.499	0.568	0.271	0.555	0.671
VC2	0.609	0.642	0.217	0.645	0.471
SP1	0.568	0.469	0.394	0.596	0.636
SP2	0.655	0.437	0.381	0.533	0.626
SP3	0.214	0.451	–	–	–
SP4	–	0.473	–	–	–
Average			0.504		

computed posterior estimates to decide whether an event has occurred at a given time point. This decision was made here adapting the concept of posterior probability maps [Friston and Penny, 2003]. As hinted in [Gitelman et al., 2003], a wavelet basis could instead be used to enhance the time-frequency representation of the deconvolved time series. This option could make the EBE approach more similar to SPFM. However, our simulations show that EBE did not show appropriate operation for voxels with no events. It must be noted that when the EBE approach is used for the analysis of psychophysiological interactions its aim is to deconvolve the underlying neural signal only of time series showing a significant experimental effect and not to detect activation across the whole brain [Gitelman et al., 2003].

All the methods evaluated in this work assume a linear hemodynamic model as this is usually adopted for fMRI analyses [Friston et al., 1998a]. Consequently, when using SPFM, one must be aware that assuming a linear model can lead to inaccuracy in delay and amplitude estimates if nonlinear effects in the BOLD response occur [Birn et al., 2001; de Zwart et al., 2009; Glover, 1999; Riera et al., 2004]. It would be possible to use SPFM to detect the events, and subsequently analyze the form of the HRF for that event in more detail. Another possibility would be to augment the convolution matrix in Eq. (2) with extra regresssors that model nonlinear effects (e.g., considering quadratic [second-order Volterra kernels] or cubic [third-order Volterra kernels] interactions between the linear regressors) [Friston et al., 1998b]. This would be possible due to the use of sparse regression methods where the number of covariates describing the data can be much larger than the number of observations (the $P > N$ problem) [Candes and Tao, 2007].

If a particular nonlinear hemodynamic model was assumed, one could alternatively use methods that implicitly build upon a nonlinear description of the hemodynamic response. Based on the Balloon-Windkessel model [Buxton et al., 1998; Friston et al., 2000], the methods of dynamic expectation maximization [Friston et al., 2008b], generalized filtering [Friston et al., 2010], squared-root cubature Kalman smoother [Havlicek et al., 2011], and the local linearization filter used in Riera et al. [2004] are filter-

ing schemes that also furnish dynamic deconvolution of the input signal, along with time-dependent estimates of physiological signals (vasodilatory signal, blood flow, blood volume, and deoxyhemoglobin content) and hemodynamic parameters. These schemes are more computationally intensive than the EBE approach and thus considerably slower than SPFM solved via homotopy algorithms (Fig. 7). Moreover, high SNR in the fMRI signal is required to allow these techniques to achieve appropriate conditional estimates of the unknown parameters. Consequently, these methods are mainly used for the deconvolution of region-of-interest fMRI time series.

SPFM can also be compared with independent component analysis (ICA) approaches. The analysis of the same datasets with spatial probabilistic ICA (MELODIC, FSL, Oxford, UK); [Beckmann and Smith, 2004] was shown in Caballero-Gaudes et al. [2011]. A fundamental difference between ICA and PFM approaches is that ICA is a multivariate method that blindly decomposes the fMRI data into independent components, either spatial components with a common time course (spatial ICA) or temporal components with a common spatial map (TICA), whereas PFM approaches are based on the univariate, voxelwise hemodynamic deconvolution of the fMRI BOLD signal (in this case assuming a canonical HRF model). Therefore, TICA or SICA only produce static maps unless selection
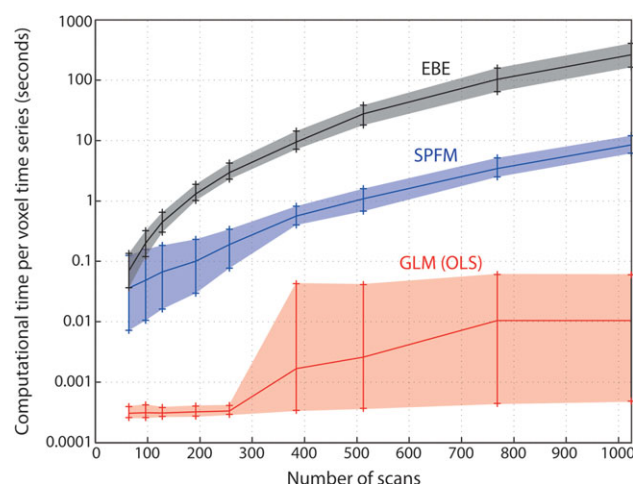

Figure 7.

Computational time in seconds required by SPFM, EBE, and GLM (estimated using OLS) to calculate the coefficient estimates of one voxel time series as a function of the number of scans (N). The value and range at each point are defined by the 50% (median) and the 5% and 95%, respectively, of 1,000 repetitions. For SPFM, the DS homotopy procedure was stopped when the regularization parameter was below the MAD estimate of the noise standard deviation or the number of elements in the active set of nonzero coefficients exceeded half of the number of scans. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and recombination of the relevant components is performed, whereas SPFM features maps displaying the dynamics of the underlying activity in a straightforward manner. Even though temporal ICA has proven its ability to identify transient events in the fMRI signal [Biswal and Ulmer, 1999; Seifritz et al., 2002] similar to the detection features of SPFM, TICA analyses have been applied only within reduced regions [Biswal and Ulmer, 1999] or after dimensionality reduction, for example, either with spatial ICA or principal component analysis (PCA) [Calhoun et al., 2001; McKeown et al., 2003; Seifritz et al., 2002]. Biswal and Ulmer [1999] used TICA to separate CO_2 global hypercapnia effects from transient task-induced effects of bilateral finger tapping. Also using TICA, Seifritz et al. [2002] showed a complex pattern of activity in the human auditory cortex comprising sustained and transient effects with overlapping spatial maps. Crucially, SPFM allows us to directly study transient BOLD effects.

Methodological Issues and Simulations

The DS was used in this work to introduce the SPFM approach, but alternative L_1-norm regularized estimators could be investigated within the same framework of PFM. In particular, the LASSO [Tibshirani, 1996] or BPDN [Chen et al., 1998] are widely used sparse regression techniques that have a close relationship with the DS [Bickel et al., 2009; James et al., 2009]. We also performed the same simulation experiments using the LASSO (data not shown) and found that both estimators provide relatively equivalent performance with the BIC, with the DS showing some advantage over the LASSO in terms of ROC curves and accuracy of estimation of the simulated hemodynamic signal. With the AIC, however, the DS exhibited a significantly sparser and better operation in the SPFM approach than the LASSO, in agreement with theoretical studies [Bickel et al., 2009; James et al., 2009].

The use of sparse regression is advantageous for sparse (single-trial and single event) paradigms; however, it might be inappropriate for other experimental designs where less sparse patterns of BOLD activations might be expected, such as block paradigms or experimental designs with trains of events [Liu et al., 2001]. We tested the method with simulated signals for which 15% of time points had event-related activations (20 nonzero coefficients out of 128 time points) and the DS gave good results with these less sparse signals, in agreement with the behavior shown by the DS in relatively nonsparse scenarios [Candes and Tao, 2007; James et al., 2009]. Alternative regression techniques combining L_1- and L_2-norm regularization terms, such as Elastic Net [Zou and Hastie, 2005] or L_1-reweighted methods [Wipf and Nagarajan, 2010], or incorporating both spike-like and epoch-like constraints in the regularization cost function could be future methods to be investigated within the PFM framework to enhance the bound on the degree of sparsity of the responses below which the proposed SPFM approach provides satisfactory results.

In this study, the regularization parameter δ was selected according to model selection criteria (BIC and AIC) following the work by Zou et al. [2007]. Our simulation and experimental results demonstrated that the BIC criterion provided stricter FP control and better estimation accuracy than the AIC criterion, which exhibited a tendency for data overfitting by detecting more events than actually existed to minimize prediction error [Zou et al., 2007]. Furthermore, our simulations demonstrated that the SPFM approach is robust against HRF variability [Aguirre et al., 1998; Handwerker et al., 2004] and non-Gaussianity of noise, which included sinusoidal trends representing physiological cardiac and respiratory fluctuations [Birn et al., 2006; Shmueli et al., 2007; Triantafyllou et al., 2005]. Robustness against errors in the temporal characteristics of the assumed HRF model is observed in the estimation of the neuronal-related hemodynamic component of the signal (Fig. 4), and in the spatial specificity and sensitivity of the technique (Fig. 2) despite the fact the temporal specificity and sensitivity to find the times of the activation events accurately can be reduced with HRF mismatches (Fig. 3). This feature suggests that SPFM is a denoising method that is tailored to estimate the HRF activations of the fMRI voxel time series [LaConte et al., 2000]. In addition, other HRF models can be assumed for the hemodynamic deconvolution with SPFM. Furthermore, the SPFM method presented in this work can be extended to include correlated noise provided that serial correlations are accurately estimated [Worsley et al., 2002].

SPFM is more computationally intensive than fitting GLM via OLS, but faster than the EBE approach. This is due to the need to compute the regularization path of the DS to choose the appropriate regularization parameter and its corresponding solution. At each iteration, the computational cost of each iteration of the homotopy procedure can be similar to solving an OLS estimation problem of the same dimensions as $\mathbf{H}_\delta$ [Asif and Romberg, 2009]. Furthermore, the GLM fit can be done simultaneously for multiple voxel time series, or once for the whole brain, because the model matrix is the same for every voxel. On the contrary, SPFM performs the deconvolution on a voxel-by-voxel basis and it builds the most appropriate model fitting the time series at each voxel [Pendse et al., 2010; Razavi et al., 2003]. The computational cost of SPFM and EBE could be easily reduced with parallelizable implementations.

Experimental Results

SPFM successfully detected all single trial finger tapping events without prior knowledge of the timing of the events (Fig. 5), showing improved performance in detecting the single-trial events than PFM using ridge regression which was unable to show 5 out of the 23 events in the

ATS [Caballero-Gaudes et al., 2011]. The SPFM maps showed cortical activations in areas typically involved in processing visually cued and self-paced finger tapping tasks [Witt et al., 2008] and showed large regional concordance with the GLM maps (Fig. 6 and Supporting Information Figure) and the maps obtained with PFM displayed in [Caballero-Gaudes et al., 2011]. Good agreement was found between the SPFM and GLM maps despite the absence of spatial smoothing in our analysis. The variability of the Dice coefficients across trials and subjects is likely to be related to different physiological characteristics for each subject and due to differences in the power of each technique for single-trial analysis, and hence size of the detected ROI [Windischberger et al., 2002]. Furthermore, we note that one could use SPFM as an exploratory approach such that the onset times when coordinated clusters of activations occur throughout the scan can be identified from the peaks of the ATS, and subsequently the activation events could further assessed using other hypothesis-based analysis methods, such as GLM.

This work was performed at 7T. The *tSNR* of our experimental data ranged from 50 to 60 in grey matter for a 2 mm isotropic voxel resolution (8 mm^3), in agreement with the values shown in Table 3 in Triantafyllou et al. [2005]. Assuming a BOLD signal change of 6% [Van Der Zwaag et al., 2009a], the contrast to noise ratio (CNR) would be between 3 and 3.6. To obtain a similar CNR range at 3T assuming a BOLD signal change of 4% [Van Der Zwaag et al., 2009a], the tSNR should be of 75–90. This increase in tSNR can be obtained at 3T by augmenting the voxel volume to 48–75 mm^3 (Table 3 in Triantafyllou et al., [2005]) or 27–45 mm^3 if a 32-channel array coil is used [Triantafyllou et al., 2011].

Benefiting from the higher contrast to noise ratio available at ultra-high MR fields (7T) [Bianciardi et al., 2009; Van Der Zwaag et al., 2009b], SPFM also revealed transient and coordinated clusters of cortical activations during rest periods. The relevance of these spontaneous BOLD signal changes and their link to changes in the functional connectivity of the motor, visual, dorsal attention and default mode networks are investigated in parallel work [Petridou et al., 2011]. These results depend on the robustness of SPFM against HRF variability since one could conjecture that the temporal characteristics of the BOLD response at rest might differ from the canonical HRF assumed here. Using a sliding window correlation analysis, we found that the correlation strength within these networks considerably varies across time, peaking at the time of the spontaneous activations detected with SPFM based on a canonical HRF model. These transient activations cannot be detected by standard model-based approaches that require the onset of activation to be specified, calling for analysis methods that, similar to SPFM, explore the non-stationary dynamics of the resting fMRI signal [Chang and Glover, 2011; Majeed et al., 2010] and go beyond the slow-frequency fluctuations (below 0.1 Hz) and the long-time scales typically investigated in resting state studies [Fox and Raichle, 2007; Cole et al., 2010].

REFERENCES

Aguirre GK, Zarahn E, D'Esposito M (1998): The variability of human, BOLD hemodynamic responses. Neuroimage 8:360–369.

Akaike H (1974): A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723.

Asif MS, Romberg J (2009): Dantzig selector homotopy with dynamic measurements. Proc SPIE Comput Imag VII 7246.

Bach F, Jenatton R, Mairal J, Obozinski G (in press): Convex optimization with sparsity-inducing norms. In Sra S, Nowozin S, Wright SJ, editors, Optimization for Machine Learning, MIT Press.

Bagshaw AP, Hawco C, Bénar CG, Kobayashi E, Aghakhani Y, Dubeau F, Pike GB, Gotman J (2005): Analysis of the EEG-fMRI response to prolonged bursts of interictal epileptiform activity. Neuroimage 24:1099–1112.

Beckmann CF, Smith SM (2004): Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Trans Med Imag 23:137–152.

Benjamini Y, Krieger AM, Yekutieli D (2006): Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93:491–507.

Benjamini Y, Hochberg Y (1995): Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 57:289–300.

Bianciardi M, Fukunaga M, van Gelderen P, Horovitz SG, de Zwart JA, Shmueli K, Duyn JH (2009): Sources of functional magnetic resonance imaging signal fluctuations in the human brain at rest: A 7 T study. Magn Reson Imaging 27:1019–1029.

Bickel P (2007): Discussion: The Dantzig selector: Statistical estimation when p is much larger than n. Ann Stat 35:2352–2357.

Bickel P, Ritov Y, Tsybakov A (2009): Simultaneous analysis of Lasso and Dantzig Selector. Ann Stat 37:1705–1732.

Birn RM, Diamond JB, Smith MA, Bandettini PA (2006): Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. Neuroimage 31:1536–1548.

Birn RM, Saad ZS, Bandettini PA (2001): Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. Neuroimage 14:817–826.

Biswal BB, Ulmer JL (1999): Blind source separation of multiple signal sources of fMRI data sets using independent component analysis. J Comput Assist Tomogr 23:265–271.

Bolstad A, Van Veen B, Nowak R (2009): Space-time event sparse penalization for magneto-/electroencephalography. Neuroimage 46:1066–1081.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Bruckstein AM, Donoho DL, Elad M (2009): From sparse solutions of systems of equations to sparse modeling of signals and images. SIAM Rev 51:34–81.

Buxton RB, Wong EC, Frank LR (1998): Dynamics of blood flow and oxygenation changes during brain activation: The Balloon model. Mag Reson Med 39:855–864.

Caballero-Gaudes C, Petridou N, Dryden IL, Bai L, Francis ST, Gowland PA (2011): Detection and characterization of single-trial fMRI BOLD responses: Paradigm free mapping 32:1400–1418.

Calhoun VD, Adali T, Pearlson GD, Pekar JJ (2001): Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. Hum Brain Mapp 13:43–53.

Candes E, Tao T (2007): The Dantzig selector: Statistical estimation when p is much larger than n. Ann Stat 35:2313–2351.

Carroll MK, Cecchi GA, Rish I, Garg R, Rao AR (2009): Prediction and interpretation of distributed neural activity with sparse models. Neuroimage 44:112–122.

Chang C, Glover GH (2010): Time-frequency dynamics of resting-state brain connectivity measured with fMRI. Neuroimage 50:81–98.

Chen SS, Donoho DL, Saunders MA (1998): Atomic decomposition by basis pursuit. SIAM J Sci Comput 20:33–61.

Cole DM, Smith SM, Beckmann CF (2010): Advances and pitfalls in the analysis and interpretation of resting-state fMRI data. Front Syst Neurosci 6:4–8.

Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

De Martino F, Valente G, Staeren N, Ashburner J, Goebel R, Formisano E. (2008): Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. Neuroimage 43:44–58.

de Zwart JA, van Gelderen P, Jansma JM, Fukunaga M, Bianciardi M, Duyn JH (2009): Haemodynamic nonlinearities affect BOLD response timing and amplitude. Neuroimage 47:1649–1658.

Donoho DL, Johnstone IM (1994): Ideal spatial adaptation by wavelet shrinkage. Biometrika 81:425–455.

Efron B, Hastie T, Johnstone I, Tibshirani R (2004): Least angle regression. Ann Stat 32:407–499.

Flandin G, Penny WD (2007): Bayesian fMRI data analysis with sparse spatial basis function priors. Neuroimage 34:1108–1125.

Fox MD, Raichle ME (2007): Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8:700–711.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998a): Event-related fMRI: Characterizing differential responses. Neuroimage 7:30–40.

Friston KJ, Josephs O, Rees G, Turner R (1998b): Nonlinear event-related responses in fMRI. Mag Reson Med 39:41–52.

Friston KJ, Mechelli A, Turner R, Price CJ (2000): Nonlinear responses in fMRI: The balloon model, volterra kernels, and other hemodynamics. Neuroimage 12:466–477.

Friston KJ, Penny P, Phillips C, Kiebel S, Hinton G, Ashburner J (2002): Classical and Bayesian inference in neuroimaging: Theory. Neuroimage 16:465–483.

Friston KJ, Penny W (2003): Posterior probability maps and SPMs. Neuroimage 19:1240–1249.

Friston K, Chu C, Mourao-Miranda L, Hulme O, Rees G, Penny W, Ashburner J (2008a): Bayesian decoding of brain images. Neuroimage 39:181–205.

Friston KJ, Trujillo-Barreto N, Daunizeau J. (2008b) DEM: A variational treatment of dynamic systems. Neuroimage 41:849–885.

Friston KJ, Harrison L, Daunizeau J, Kiebel S, Philllips C, Trujillo-Barreto N, Henson R, Flandin G, Mattout J (2008c) Multiple sparse priors for M/EEG inverse problems. Neuroimage 39:1104–1120.

Friston KJ, Stephan KE, Daunizeau J (2010): Generalised filtering. Math Probl Eng 621670.

Gitelman DR, Penny WD, Ashburner J, Friston KJ (2003): Modeling regional and psychophysiologic interactions in fMRI: The importance of hemodynamic deconvolution. Neuroimage 19:200–207.

Glover GH (1999): Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9:416–429.

Glover GH, Li TQ, Ress D (2000): Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med 44:162–167.

Gramfort A, Kowalski M (2009): Improving M/EEG source localization with an inter-condition sparse prior. IEEE Int Symp Biomed Imag 141–144.

Grill-Spector K, Malach R (2001): fMR-Adaptation: A tool for studying the functional properties of human cortical neurons. Acta Psycho (Amst) 107:293–321.

Grosenick L, Greer S, Knutson B (2008): Interpretable classifiers for fMRI improve prediction of purchases. IEEE Trans Neural Syst Rehabil Eng 16:539–548.

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

Hastie T, Tibshirani R, Friedman J (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Ed. New York: Springer.

Havlicek M, Friston KJ, Jiri J, Brazdil M, Calhoun VD (2011): Dynamic modeling of neuronal responses in fMRI using cubature Kalman filtering. Neuroimage 56:2109–2128.

James GM, Radchenko P, Lv J (2009): DASSO: Connections between the Dantzig selector and LASSO. J R Stat Soc B Stat Methodol 71:127–142.

Khalidov I, Fadili MJ, Lazeyras F, Van De Ville D, Unser M (2011): Activelets: Wavelets for sparse representation of hemodynamic responses. Signal processing 19:2810–2821.

LaConte SM, Ngan SC, Hu X (2000): Wavelet transform-based Wiener filtering of event-related fMRI data. Magn Reson Med 44:746–757.

Liu H, Palatucci M, Zhang J (2009): Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp 649–656.

Liu TT, Frank LR, Wong EC, Buxton RB (2001): Detection power, estimation efficiency, and predictability in event-related fMRI. Neuroimage ,13:759–773.

Long C, Brown EN, Manoach D, Solo V (2004): Spatiotemporal wavelet analysis for functional MRI. Neuroimage 23:500–516.

Lund TE, Madsen KH, Sidaros K, Luo WL, Nichols TE (2006): Non-white noise in fMRI: Does modelling have an impact? Neuroimage 29:54–66.

Majeed W, Magnuson M, Hasenkamp W, Schwarb H, Schumacher EH, Barsalu L, Keiholz SD (2011): Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. Neuroimage 54:1140–1150.

McKeown MJ, Hansen LK, Sejnowski TJ (2003): Independent component analysis of functional MRI: What is signal and what is noise? Curr Opin Neurobiol 13:620–629.

Menon RS, Luknowsky DC, Gati JS (1998): Mental chronometry using latency-resolved functional MRI. Proc Nat Acad Sci USA 95:10902–10907.

Michel V, Gramfort A, Varoquaux G, Thirion (2010): Total variation regularization enhances regression-based brain activity

prediction. In: First workshop on brain decoding: Pattern Recognition Challenges in Neuroimaging (WBD). pp 9–12.

Osborne MR, Presnell B, Turlach BA (2000): A new approach to variable selection in least squares problems. IMA J Numer Anal 20:389–403.

Ou W, Hämäläinen M, Golland P (2009): A distributed spatio-temporal EEG/MEG inverse solver. Neuroimage 44:932–946.

Park T, Casella G (2008): The Bayesian Lasso. J Am Stat Assoc 103:681–686.

Pendse GV, Baumgartner R, Schwarz A, Coimbra A, Borsook D, Becerra L (2010): A statistical framework for optimal design matrix generation with application to fMRI. IEEE Trans Med Imag 29:1573–611.

Penny W, Flandin G, Trujillo-Barreto N (2007): Bayesian comparison of spatially regularized general linear models. Hum Brain Mapp 28:275–293.

Petridou N, Caballero-Gaudes C, Dryden IL, Francis ST, Gowland PA (2011): Periods of rest in fMRI contain transient events which are related to slow spontaneous fluctuations. In: 17th Annual Meeting of the Organization for Human Brain Mapping.

Raizada RD, Tsao FM, Liu HM, Holloway ID, Ansari D, Kuhl PK (2010): Linking brain-wide multivoxel activation patterns to behaviour: Examples from language and math. NeuroImage 51:462–471.

Razavi M, Grabowski TJ, Vispoel WP, Monahan P, Mehta S, Eaton B, Bolinger L (2003): Model assessment and model building in fMRI. Hum Brain Mapp 20:227–238.

Richter W, Andersen PM, Georgopoulos AP, Kim SG (1997): Sequential activity in human motor areas during a delayed cued finger movement task studied by time-resolved fMRI. Neuroreport 8:1257–1261.

Riera J, Watanabe J, Kazuki I, Naoki M, Aubert E, Ozaki T, Kawashima R. (2004): A state-space model of the hemodynamic approach: Nonlinear filtering of BOLD signals. Neuroimage 21:547–567.

Ryali S, Supekar K, Abrams DA, Menon V (2010): Sparse logistic regression for whole-brain classification of fMRI data. Neuroimage 51:752–764.

Schwarz G (1978): Estimating the dimension of a model. Ann Stat 6:461–464.

Shmueli K, van Gelderen P, de Zwart JA, Horovitz SG, Fukunaga M, Jansma JM, Duyn JH (2007): Low-frequency fluctuations in the cardiac rate as a source of variance in the resting-state fMRI BOLD signal. Neuroimage 38:306–320.

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

Tibshirani R (1996): Regression shrinkage and selection via the LASSO. J R Stat Soc B Stat Methodol 58:267–288.

Tipping ME (2001): Sparse bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244.

Tropp JA, Wright SJ (2010): Computational methods for sparse solution of linear inverse problems. Proc IEEE 98:948–958.

Triantafyllou C, Hoge RD, Krueger G, Wiggins DJ, Potthast A, Wiggins GC, Wald LL (2005): Comparison of physiological noise at 1.5T, 3T and 7T and optimization of fMRI acquisition parameters. Neuroimage 26:243–250.

Triantafyllou C, Polimeni JR, Wald LL (2011): Physiological noise and signal-to-noise ratio in fMRI with multi-channel array coils. Neuroimage 55:595–606.

Valdes-Sosa PA, Vega-Hernandez M, Sanchez-Bornot JM, Martinez-Montes E, Bobes MA (2009): EEG source imaging with spatio-temporal tomographic nonnegative independent component analysis. Hum Brain Mapp 30:1898–1910.

Valente G, De Martino F, Esposito F, Goebel R, Formisano E (2011): Predicting subject-driven actions and sensory experience in a virtual world with relevance vector machine regression of fMRI data. Neuroimage 56:651–661.

Van De Ville D, Seghier ML, Lazeyras F, Blu T, Unser M (2007): WSPM: Wavelet-based statistical parametric mapping. Neuroimage 37:1205–1217.

Van Der Zwaag W, Francis S, Head K, Peters A, Gowland P, Morris P, Bowtell R (2009a): fMRI at 1.5, 3 and 7 T: Characterising BOLD signal changes. Neuroimage 47:1425–1434.

Van Der Zwaag W, Marques JP, Hergt M, Gruetter R (2009b) Investigation of high-resolution functional magnetic resonance imaging by means of surface and array radiofrequency coils at 7T. Magn Reson Imaging 27:1011–1018.

Van Gerven MA, Cseke B, de Lange FP, Heskes T (2010): Efficient Bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior. Neuroimage 50:150–161.

Van Gerven MA, Hesse C, Jensen O, Heskes T (2009): Interpreting single trial data using groupwise regularization. Neuroimage 46:665–676.

Vulliemoz S, Lemieux L, Dauzineau J, Michel CM, Duncan JS (2010): The combination of EEG source imaging and EEG-correlated functional MRI to map epileptic networks. Epilepsia 51:491–505.

Windischberger C, Lamm C, Bauer H, Moser E (2002): Consistency of inter-trial activation using single-trial fMRI: Assessment of regional differences. Brain Res Cogn Brain Res 13:129:138.

Wipf D, Nagarajan S (2009): A unified Bayesian framework for MEEG/EEG source imaging. Neuroimage 44:947–966.

Wipf D, Nagarajan S (2010): Iterative reweighted L1 and L2 methods for finding sparse solutions. IEEE J Sel Top Signal Proc 4:317–329.

Wise RG, Tracey I (2006): The role of fMRI in drug discovery. J Magn Reson Imaging 23:862–876.

Witt ST, Laird AR, Meyerand ME (2008): Functional neuroimaging correlates of finger-tapping task variations: An ALE meta-analysis. Neuroimage 42:343–356.

Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans AC (2002): A general statistical analysis for fMRI data. Neuroimage 15:1–15.

Yamashita O, Sato M, Yoshioka T, Tong F, Kamitani Y (2008): Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. Neuroimage 42:1414–1429.

Zou H, Hastie T (2005): Regularization and variable selection via the elastic net. J R Stat Soc B Stat Methodol 67:301–320.

Zou H, Hastie T, Tibshirani R (2007): On the degrees of freedom of the LASSO. Ann Stat 35:2173–2192.