Biometrics 69, 184–196
March 2013

DOI: 10.1111/j.1541-0420.2012.01819.x

A Wavelet-Based Bayesian Approach to Regression Models with Long Memory Errors and Its Application to fMRI Data

Jaesik Jeong,[1] Marina Vannucci[2,*] and Kyungduk Ko[3]

[1]Department of Biostatistics, Indiana University, Indianapolis, Indiana 46202, U.S.A.
[2]Department of Statistics, Rice University, Houston, Texas 77005, U.S.A.
[3]Department of Mathematics, Boise State University, Boise, Idaho 83725, U.S.A.
[*] *email:* marina@rice.edu

Summary. This article considers linear regression models with long memory errors. These models have been proven useful for application in many areas, such as medical imaging, signal processing, and econometrics. Wavelets, being self-similar, have a strong connection to long memory data. Here we employ discrete wavelet transforms as whitening filters to simplify the dense variance–covariance matrix of the data. We then adopt a Bayesian approach for the estimation of the model parameters. Our inferential procedure uses exact wavelet coefficients variances and leads to accurate estimates of the model parameters. We explore performances on simulated data and present an application to an fMRI data set. In the application we produce posterior probability maps (PPMs) that aid interpretation by identifying voxels that are likely activated with a given confidence.

Key words: Bayesian inference; fMRI Data; Long memory errors; Regression models; Wavelets.

1. Introduction

In this article we deal with a linear regression model of the type

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon, \tag{1}$$

where the error terms are strongly correlated, and specifically long memory. Data from long memory processes have the distinctive feature that the correlation between distant observations is not negligible. This characteristic leads to dense variance–covariance matrices that make existing inferential methods computationally expensive. Best linear unbiased estimates (BLUE) of the parameters are usually more efficient than the ordinary least squares (OLS) estimates, see Beran (1994). Their computation, however, is prohibitive due to the necessary iterative estimation procedures of dense variance–covariance matrices. Approximate maximum likelihood methods are therefore often employed (Fox and Taqqu, 1986; Li and McLeod, 1986).

Here we employ discrete wavelet transforms (DWTs) as whitening filters that allow to simplify the variance–covariance structure of the data. Wavelets, being self-similar, have a strong connection to long memory processes and have proven to be a powerful tool for the analysis and synthesis of data from such processes, see Wornell and Oppenheim (1992), McCoy and Walden (1996), Jensen (2000), Abry et al. (2003), Craigmile, Guttorp, and Percival (2005), and Stoev and Taqqu (2005), among others. The ability of wavelets to simultaneously localize a process in the time and scale domains results in representing many dense matrices in a sparse form. When transforming measurements from a long memory process, wavelet coefficients are approximately uncorrelated, in contrast with the dense long memory covariance structure of the data, see Tewfik and Kim (1992) and Craigmile and Percival (2005), among others. Ko and Vannucci (2006) describe a wavelet-based Bayesian estimation procedure to estimate the parameters of a general Gaussian ARFIMA (autoregressive fractionally integrated moving average) model, with unknown autoregressive and moving average parameters. Here we exploit their approach for the estimation of the parameters of the error term in (1). By employing the decorrelation properties of the wavelet transforms we write a relatively simple model in the wavelet domain, where estimation of the model parameters is carried out via a Bayesian approach. This inferential procedure uses exact wavelet coefficient variances and leads to accurate parameter estimates.

Regression models with correlated errors have found useful applications in many areas, such as medical imaging, signal processing, and econometrics. In this article we consider imaging data, and in particular functional magnetic resonance imaging (fMRI) data. Statistical methods play an important role in the analysis of fMRI data and have generated a growing literature, see Lindquist (2008) for a recent review. Zarahn, Aguirre, and D'Esposito (1997) and Aguirre, Zarahn, and D'Esposito (1997) first suggested modeling the noise in fMRI data using an $1/f$ process. Fadili and Bullmore (2002) adopted models of type (1), assuming fractional Brownian motion (fBm) as the error term. They used DWTs to derive a near Karhunen-Lóeve-type expansion of the variance–covariance structure of the long memory error and derived approximate

maximum likelihood estimates (MLE) of the model parameters. Meyer (2003) applied generalized linear model with drifts and errors contaminated by long-range dependencies. See also Bullmore et al. (2004) for a nice review of wavelet-based methods for fMRI data. Here we first show performances of our method via a thorough simulation study and then investigate their use in the analysis of fMRI data by direct application to a benchmark data set. In the application we employ a more general fractal structure for the error term. In addition, we exploit our Bayesian approach to produce PPMs that aid interpretation, as one can easily identify voxels that are likely activated with a given confidence. Bayesian approaches have recently found successful applications to fMRI, see Woolrich et al. (2004), Bowman et al. (2008), and Guo, Bowman, and Kilts (2008), among others.

The remainder of this article is organized as follows: In Section 2, we introduce the model and the necessary basic concepts on long memory processes. We also describe wavelet transforms and exemplify how these are applied to the data. Finally, we describe the Bayesian model in the wavelet domain that leads to the estimation of the long memory and model parameters. We also assess performances on simulated data. In Section 3 we focus on applications on imaging data and report results from a simulation study and an application to fMRI data. Section 4 concludes the article.

2. Methods

2.1 *Regression Models with Correlated Errors*

In this article we consider regression models of type (1), where $\boldsymbol{y}$ is the $(N \times 1)$ vector of response data, $\boldsymbol{X}$ is the $(N \times p)$ design matrix consisting of the $(N \times 1)$ covariate vectors $\boldsymbol{x}_i, i = 1, 2, \ldots, p$, and $\boldsymbol{\beta}$ is the $(p \times 1)$ regression coefficient vector. Without loss of generality, we take the data to be centered at 0, so an intercept is not needed in the model. We then assume correlated errors by modeling ε as a $\frac{1}{f}$ process, that is, a long memory process. A long memory process is characterized by a slow decay in its autocovariance, that is

$$\gamma(h) \sim Ch^{-\alpha}, \tag{2}$$

where C is a positive constant depending on the process, $0 < \alpha < 1$ and h is large. Some heuristic approaches exist to identify the presence of long memory in time series data, such as the variance ratio and the R/S statistic, among others. Beran (1994) provides a in-depth discussion of these methods and of more general issues related to the choice of particular long memory parametrizations. Regression models of type (1) have found numerous applications in medical imaging. We give a representative application to fMRI data in Section 3.

A class of long memory processes we will consider later in the applications is the fractionally integrated Gaussian processes with parameter d, or $I(d)$, first introduced by Granger and Joyeux (1980) and Hosking (1981). The autocovariance function of an $I(d)$ process is

$$\gamma(h) = \sigma^2 \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(d)\gamma(1-d)\Gamma(1-d+h)}. \tag{3}$$

I(d) processes are stationary and invertible if $-.5 < d < .5$. They are long memory for $0 < d < .5$, have "antipersistent" behavior for $-.5 < d < 0$ and have no memory for $d = 0$.

Early contributions to the estimation of long memory parameters used approximate maximum likelihood methods, see for example, Li and McLeod (1986) and Fox and Taqqu (1986). Beran (1994) investigated asymptotic sampling theory properties of exact and approximate maximum likelihood methods.

2.2 *Wavelets and DWTs.*

Wavelets are families of orthonormal basis functions that can be used to parsimoniously represent other functions. For example, in $\mathbb{L}_2(\mathbb{R})$ an orthogonal wavelet basis is obtained by dilating and translating a *mother wavelet* $\psi(x)$, a squared integrable zero mean function, as

$$\psi_{m,n}(x) = 2^{m/2}\psi(2^m x - n), \tag{4}$$

with $m, n \in \mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\}$. A generic function $f(t) \in \mathbb{L}_2(\mathbb{R})$ can then be represented by a wavelet expansion of the type

$$f(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} z_{m,n}\psi_{m,n}(t), \tag{5}$$

with wavelet coefficients $z_{m,n} = \int f(t)\psi_{m,n}(t)dt$ describing features of the function $f(t)$ at spatial locations indexed by n and scales indexed by m.

Wavelets have been extremely successful as a tool for the analysis and synthesis of discrete data. Suppose we observe a time series as a realization of a random process $\{Y_t, t = 0, \pm 1, \pm 2, ...\}$ and let us indicate the time series as $\mathbf{y} = (y_1, y_2, \ldots, y_N)^T$. We assume $N = 2^J$ with J a positive integer denoting the scale of the data. This is not a real restriction and methods allowing wavelet transforms to be applied to any length of data exist, see Taswell and McGill (1994). A DWT, Mallat (1989), can be used to reduce the data to a set of wavelet coefficients. Although it operates via recursive applications of filters, for practical purposes a DWT of order $r < J$ is often represented in matrix form as

$$\mathbf{z} = \mathcal{W}\mathbf{y}, \tag{6}$$

with $\mathcal{W}$ an $N \times N$ real-valued and orthogonal matrix of the form

$$\mathcal{W} = \left[W_1^T, W_2^T, \ldots, W_r^T, V_r^T\right]^T, \tag{7}$$

with $W_m, m = 1, 2, \ldots, r$ and V_r of dimension $N/2^m \times N$ and $N/2^r \times N$, respectively, and $\mathbf{z}$ a vector of coefficients

$$\mathbf{z} = \left[\mathbf{z}_1^T, \mathbf{z}_2^T, \ldots, \mathbf{z}_r^T, \mathbf{s}_r^T\right]^T. \tag{8}$$

Here larger values of m indicate coarser approximations. The vector $\mathbf{z}$ consists of $(r + 1)$ subvectors, with $\mathbf{z}_m = W_m\mathbf{y}$ of dimension $N(m) = N/2^m$, for $m = 1, \ldots, r$, and $\mathbf{s}_r = V_r\mathbf{y}$ of dimension $N/2^r$, with $N = N' + N/2^r$ and $N' = \sum_{m=1}^{r} N(m)$. The elements of the vector $\mathbf{s}_r$ are called scaling coefficients and represent a coarser approximation of the data, whereas the vectors $\mathbf{z}_m$ contain the wavelet coefficients $z_{m,n}, n = 1, \ldots, N(m)$ representing local features of the data at resolution scale m and location n. An inverse transformation exists to reconstruct the data from its wavelet decomposition.

2.2.1 *Covariance matrix of wavelet coefficients.* Let us consider equation (6). The form of the autocovariance function of the process $\{Y_t\}$ determines the covariance matrix of the

data $\mathbf{y}$ which, in turn, allows us to calculate the covariance matrix of the corresponding wavelet coefficients $\mathbf{z}$ as

$$\mathbf{\Sigma}_z = \mathcal{W}\mathbf{\Sigma}_y\mathcal{W}^T, \tag{9}$$

with $\mathcal{W}$ the orthogonal matrix representing the wavelet transform and with $\mathbf{\Sigma}_y$ calculated as $\mathbf{\Sigma}_y(i,j) = [\gamma(|i-j|)]$, where $\gamma(h)$ is the autocovariance function of the process $\{Y_t\}$ generating the data.

Although algebraic expressions for $\mathcal{W}$ are available, the matrix form (6) of the DWT is mainly used for illustrative purposes. Also, matrix calculations of $\mathbf{\Sigma}_z$ as in (9) are impractical for long time series. Vannucci and Corradi (1999) proposed a method to compute variances and covariances of wavelet coefficients that uses the recursive filters of the DWT. Their method can be used to compute the matrix $\mathbf{\Sigma}_z$ in a very efficient way. First the two-dimensional DWT is applied to the matrix $\mathbf{\Sigma}_y$. The diagonal blocks of the resulting matrix provide the "within-scale" variances and covariances of wavelet coefficients at same levels. The one-dimensional DWT is then applied to the rows of the off-diagonal blocks to obtain "across-scale" variances and covariances of wavelet coefficients belonging to different levels.

2.3 *Model in the Wavelet Domain*

In our approach we employ the DWT as a tool to simplify the likelihood. Long memory data have, in fact, a dense covariance structure that makes the exact likelihood of the data difficult to handle, Beran (1994). Applying the DWT to both sides of model (1) leads to

$$\mathbf{y}_w = \mathbf{X}_w\boldsymbol{\beta} + \varepsilon_w, \tag{10}$$

where $\mathbf{y}_w = \mathcal{W}\mathbf{y}$, $\mathbf{X}_w = \mathcal{W}\mathbf{X}$, and $\varepsilon_w = \mathcal{W}\varepsilon$. Here $\varepsilon_w \sim N(0, \mathbf{\Sigma}_{\varepsilon_w})$, where $\mathbf{\Sigma}_{\varepsilon_w} = \sigma^2\mathbf{\Sigma}_\Psi$ is the $(N \times N)$ diagonal matrix with elements $\sigma^2\sigma^2_{m\,n}$ indicating the variance of the nth wavelet coefficient at the mth scale. Exact variances of wavelet and scaling coefficients can be computed from (9) following the algorithm of Vannucci and Corradi (1999). Notice how this construction leads to separate variance parameters for the individual coefficients.

It has been emphasized in the literature that wavelet transforms tend to "whiten" the data, meaning that wavelet coefficients tend to be less correlated than the original data. For long memory processes, such decorrelation properties are well documented. Wavelet transforms, being band-pass filters, balance the divergence of the spectrum of long memory data at frequencies close to zero, therefore whitening the data, see Tewfik and Kim (1992), Craigmile and Percival (2005), and Ko and Vannucci (2006), among others. Tewfik and Kim (1992), in particular, proved that the correlation between wavelet coefficients decreases exponentially fast across scales and hyperbolically fast along time. Ko and Vannucci (2006) focus on ARFIMA models (a large class of long memory processes that includes the $I(d)$ processes) and investigate the effect of the approximately uncorrelated property by comparing estimates of the long memory parameter obtained by using the diagonalized structure in (10) with those obtained by using the exact form, that is, the full variance and covariance matrix of the wavelet coefficients. Their results confirm that the approximation to uncorrelated wavelet coefficients is reasonable with data from a long memory process.

2.4 *Bayesian Inference*

We describe our inferential procedure by assuming a model of type (1) and $I(d)$ errors, meaning that the matrix $\mathbf{\Sigma}_y$ in (9) is constructed based on an autocovariance function of type (3) that depends on the two parameters σ^2 and d. We notice that our modeling approach can be applied to error models with other long memory parametrizations and to more general processes, including fractal processes of type (2). In such cases, construction (9) leads to wavelet coefficient variances that depend on the unknown characteristic parameters of the chosen parameterization/process.

Since $\mathbf{y}$ is multivariate normal and the DWT is a linear transformation, then $\mathbf{y}_w$ is also multivariate normal. Our likelihood is therefore

$$L(\mathbf{\Theta}|\mathbf{y}_w, \mathbf{X}) = \frac{(\sigma^2)^{-N/2}|\mathbf{\Sigma}_\Psi|^{-1/2}}{(\sqrt{2\pi})^N}\exp\left\{-\frac{1}{2\sigma^2}Q_w\right\},$$

where $\mathbf{\Theta} = (\beta, \sigma^2, d)$ and $Q_w = (\mathbf{y}_w - \mathbf{X}_w\boldsymbol{\beta})'\mathbf{\Sigma}_\Psi^{-1}(\mathbf{y}_w - \mathbf{X}_w\boldsymbol{\beta})$. We assume $\pi(\boldsymbol{\beta}, \sigma^2, d) = \pi(\boldsymbol{\beta}|\sigma^2)\pi(\sigma^2)\pi(d)$, choose a normal inverse gamma prior for $(\boldsymbol{\beta}, \sigma^2)$,

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2) = \pi(\boldsymbol{\beta}|\sigma^2)\pi(\sigma^2) &= (\sigma^2)^{-1} \\ &\times \exp\left\{-\frac{1}{\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} \\ &\times \frac{1}{(\sigma^2)^{\delta_0/1+1}}\exp\left\{-\frac{\gamma_0}{2\sigma^2}\right\}\end{aligned}$$

and a beta distribution prior for the long memory parameter d,

$$\pi(2d) = \frac{\Gamma(\eta+\nu)}{\Gamma(\eta)\Gamma(\nu)}(2d)^{\eta-1}(1-2d)^{\nu-1}, 0 < d < 1/2.$$

In all applications of this article we will use noninformative priors, that is, uniform priors on β and d and an improper prior, $p(\sigma^2) \approx 1/\sigma^2$, on σ^2. With our posterior distribution being proper, we find that convergence and mixing of the MCMC are not affected by the particular prior choice. The posterior distribution of $\mathbf{\Theta}$ given $(\mathbf{y}_w, \mathbf{X}_w)$ is

$$\begin{aligned}\pi(\boldsymbol{\beta}, \sigma^2, d|\mathbf{y}_w, \mathbf{X}_w) &\propto L(\mathbf{y}_w|\mathbf{X}_w, \mathbf{\Theta})\pi(\mathbf{\Theta}) \\ &\propto (\sigma^2)^{-(\frac{N+\delta_0+2}{2}+1)}|\mathbf{\Sigma}_\Psi|^{-\frac{1}{2}}(2d)^{\eta-1}(1-2d)^{\nu-1} \\ &\quad\times \exp\left\{-\frac{1}{2\sigma^2}\left[\gamma_0 + Q_w\right.\right. \\ &\quad\left.\left. + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right]\right\},\end{aligned}$$

from which we can easily get the full conditional distributions of each parameter:

- The full conditional distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta}|\sigma^2, d, \mathbf{y}_w, \mathbf{X}_w \sim N\left(R_w^{-1}({\mathbf{X}_w}^{*'}{\mathbf{y}_w}^* + \boldsymbol{\beta}_0), \sigma^2\mathbf{R}_w^{-1}\right),$$

  where $\mathbf{R}_w = ({\mathbf{X}_w}^{*'}{\mathbf{X}_w}^* + \mathbf{I})$, ${\mathbf{X}_w}^* = {\mathbf{\Sigma}_\Psi}^{-1/2}\mathbf{X}_w$ and ${\mathbf{y}_w}^* = {\mathbf{\Sigma}_\Psi}^{-1/2}\mathbf{y}_w$.
- The full conditional distribution of σ^2 is

$$\begin{aligned}\sigma^2|\boldsymbol{\beta}, d, \mathbf{y}_w, \mathbf{X}_w \sim IG\Bigg(&\frac{N+\delta_0+2}{2}, \frac{1}{2}\left[\gamma_0\right. \\ &\left. + Q_w + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right]\Bigg).\end{aligned}$$

Table 1

Simulated data: MSEs of the estimates of d averaged over 100 replicates, obtained with the GPH, SR, HR, fEXP methods (see text), and with our wavelet-based Bayesian approach (WB). The last two columns refer to the estimates of σ^2 and β obtained with our method. Results refer to the case $\sigma^2 = 1$ and $\beta = .01$

d	N	GPH(d)	SR(d)	HR(d)	fEXP(d)	WB(d)	WB(σ^2)	WB(β)
	256	.0685	.0505	.0028	.0355	.0023	.0082	4.78×10^{-6}
.1	512	.0331	.0311	.0019	.0075	.0012	.0034	7.80×10^{-7}
	1024	.0178	.0155	.0006	.0039	.0006	.0014	1.42×10^{-7}
	256	.0438	.0368	.0043	.0217	.0036	.0048	1.15×10^{-6}
.25	512	.0340	.0280	.0016	.0075	.0015	.0034	2.28×10^{-6}
	1024	.0267	.0175	.0007	.0048	.0007	.0025	3.94×10^{-7}
	256	.0608	.0526	.0047	.0240	.0036	.0065	3.14×10^{-5}
.4	512	.0358	.0276	.0013	.0159	.0012	.0033	5.80×10^{-6}
	1024	.0293	.0210	.0009	.0088	.0008	.0019	9.87×10^{-7}

- The full conditional distribution of d is

$$\pi(d|\boldsymbol{\beta}, \sigma^2, \mathbf{y}_w, \mathbf{X}_w) \propto |\boldsymbol{\Sigma}_\Psi|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} Q_w\right\} \times (2d)^{\eta-1}(1-2d)^{\nu-1}. \quad (11)$$

Since distribution (11) is not in closed form we implement a Metropolis step using a Normal proposal for d within Gibbs steps for $\boldsymbol{\beta}$ and σ^2. In all applications of this article we used a proposal distribution centered in the previous value and with a standard deviation of .5, a value that gave us acceptance rates in the range 35–65%. The acceptance probability of a candidate point d_{new} in the Metropolis step is

$$\alpha = \min\left\{\frac{\pi(d_{\text{new}}|\boldsymbol{\beta}, \sigma^2, \mathbf{y}_w, \mathbf{X}_w)}{\pi(d_{\text{old}}|\boldsymbol{\beta}, \sigma^2, \mathbf{y}_w, \mathbf{X}_w)}, 1\right\}.$$

2.5 *Simulation Studies*

A computationally simple method to generate a time series that exhibits long memory properties was proposed by McLeod and Hipel (1978) and involves the Cholesky decomposition of the correlation matrix $\mathbf{R}_\varepsilon(i,j) = [\rho(|i-j|)]$. Given $\mathbf{R}_\varepsilon = \mathbf{M}\mathbf{M}'$ with $\mathbf{M} = [m_{i,j}]$ a lower triangular matrix, and given $\epsilon_t, t = 1, \ldots, N$, a Gaussian white noise series with zero mean and unit variance, then the series

$$\varepsilon_t = \gamma(0)^{1/2} \sum_{i=1}^{t} m_{t,i}\epsilon_i \quad (12)$$

will have autocorrelation $\rho(\cdot)$. Using this method we performed a first simulation from model (1) fixing $\sigma^2 = 1$, $\beta = .01$, and $x = [1, \ldots, N]$ and choosing the autocovariance function of type (3) from an $I(d)$ process.

We used three different values of d and N, that is $d = .1, .25, .4$, and $N = 256, 512, 1024$. We applied DWTs with Daubechies minimum phase wavelets (Daubechies, 1992), with four vanishing moments. In our previous work we found that wavelets with higher degrees of regularity produce slightly better estimates of the long memory parameter for large sample sizes (Ko and Vannucci, 2006), as they ensure wavelet coefficients to be approximately uncorrelated. On the other hand, the support of the wavelets increases with the regularity and boundary effects may arise in the DWT, so that a trade-off is often necessary. In this article, only the wavelet coefficients were used in the estimation procedure, that is, scaling coefficients were removed. We ran MCMC chains with 5000 iterations, discarding the first 1000 as burn-in. For each case under study we generated 100 time series. We report MSEs of all estimates in Table 1. As expected, when the sample size increases the estimates of d get closer to the true values, and their variances become smaller. Our approach also produces estimates of the innovation variance σ^2 and of the slope β. We found similar results with other choices of the parameters σ^2 and β (results not shown).

For comparison, we looked into some of the most commonly used methods for the estimation of the long memory parameter. In particular, we considered the semiparametric estimator (GPH) proposed by Geweke and Porter-Hudak (1983), which is based on a regression on periodogram ordinates, the smoothed version of the GPH estimator (SR), the efficient approximate MLE estimator for Gaussian data proposed by Haslett and Raftery (1989, HR) and the fractional EXP method based on the estimate of the slope of the periodogram on log–log coordinates (fEXP), see Beran (1994). While our method estimates both the slope and the long memory parameter simultaneously, these methods produce an estimate of d only. Therefore, in order to compare results with our method, following Beran (1994), we first fitted a linear trend $g(x_t) = \beta_0 + \beta_1 t$ to the data using least squares and then estimated d with the above methods by fitting an I(d) model to the residuals. Table 1 reports MSEs of the estimates of the long memory parameter, showing excellent performances of our method with respect to the competing ones.

3. Applications to fMRI Data

fMRI is a common tool for detecting changes in neuronal activity. It measures blood oxygenation level-dependent (BOLD) contrast that depends on changes in regional cerebral blood-flow (rCBF). The complete relationship between the neuronal activity and the vascular response is not fully known yet. A widely used model looks at an fMRI signal as the convolution of regional cerebral blood-flow response to stimulus with an hemodynamic response function (HRF). The basic idea is that an fMRI signal gets delayed hemodynamically in measuring the change in the metabolism of BOLD contrast by stimulation (Buxton and Frank, 1997). Suppose that $\mathbf{x}_R$ represents the neuronal activity from a cerebral

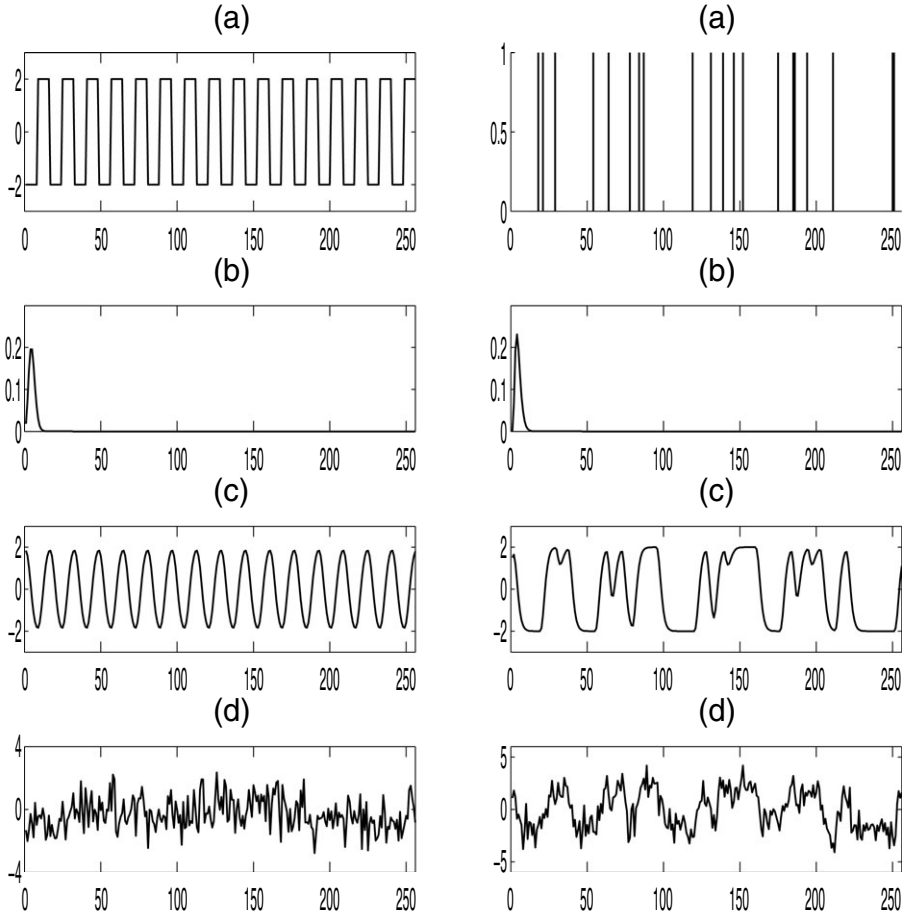

Figure 1. Simulated fMRI data with $n = 256$. Plot (a), first column, refers to the blocked design (period 16), plot (a), second column, to the event-related design. Plots (b) show a Poisson HRF with $\lambda = 4$. Plots (c) show the convolved signal of (a) with (b), and plots (d) the simulated fMRI signal with an I(.2) added noise. Parameters (β, σ^2) were set to (.01,1).

region and that $\mathbf{X}_B$ is a basis of neural stimula. A projection of $\mathbf{x}_R$ onto the basis is given by

$$\mathbf{x}_R = \mathbf{X}_B \boldsymbol{\beta}, \tag{13}$$

where $\boldsymbol{\beta}$ is a parameter vector used in expanding $\mathbf{x}_R$ into $\mathbf{X}_B$. Convolution of both sides with an HRF $\mathbf{H}$ gives a linear model with additive noise of the type

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{14}$$

where $\mathbf{y} = \mathbf{H}\mathbf{x}_R$ is an observed fMRI signal at a single voxel and $\mathbf{X} = \mathbf{H}\mathbf{X}_B$ is the design matrix representing the convolved stimulus. Several choices are available for the response function $\mathbf{H}$. Glover (1999) adopted the hemodynamic function

$$h(t) = c_1 t^{n_1} \exp\left(-\frac{t}{t_1}\right) - a_2 c_2 t^{n_2} \exp\left(-\frac{t}{t_2}\right), \tag{15}$$

with $c_1 = \max(t^{n_i} e^{-t/t_2})$ and where $n_1 = 5.0$, $t_1 = 1.1$ seconds, $n_2 = 12.0$, $t_2 = .9$ seconds, and $a_2 = .4$. Meyer (2003) used the same HRF as in (15) but with different parameters. Fadili and Bullmore (2002) used a Poisson response function with $\lambda = 4$,

$$h_\lambda(t) = \frac{e^{-\lambda}\lambda^t}{t!}. \tag{16}$$

Residuals of model (14) are typically assumed autocorrelated and are due to instrumental noise, such as head movement in the scanner. Two types of autocorrelation structures have been assumed in the literature of fMRI signals: first-order autoregressive and $1/f$ long memory structures (Smith et al., 1999). A white noise component is sometimes added to the chosen fMRI noise model, see also our discussion in the Conclusion section. Estimation procedures for fMRI time series with autocorrelated residuals are quite complicated and computationally expensive. Fadili and Bullmore (2002) applied DWTs and estimated the model parameters in the wavelet domain using an iterative maximum likelihood estimation method. Meyer (2003) proposed an estimation method in the wavelet domain for a generalized linear model where a drift is used to explain polynomial trends in the data. Turkheimer et al. (2003) developed analysis of variance (ANOVA) methods for multiresolution analysis of multisubject fMRIs. Here we apply our wavelet-based Bayesian approach, first to simulated fMRI signals and then to real data. In the real data application we also produce PPMs that aid interpretation by identifying voxels that are likely activated with a given confidence.

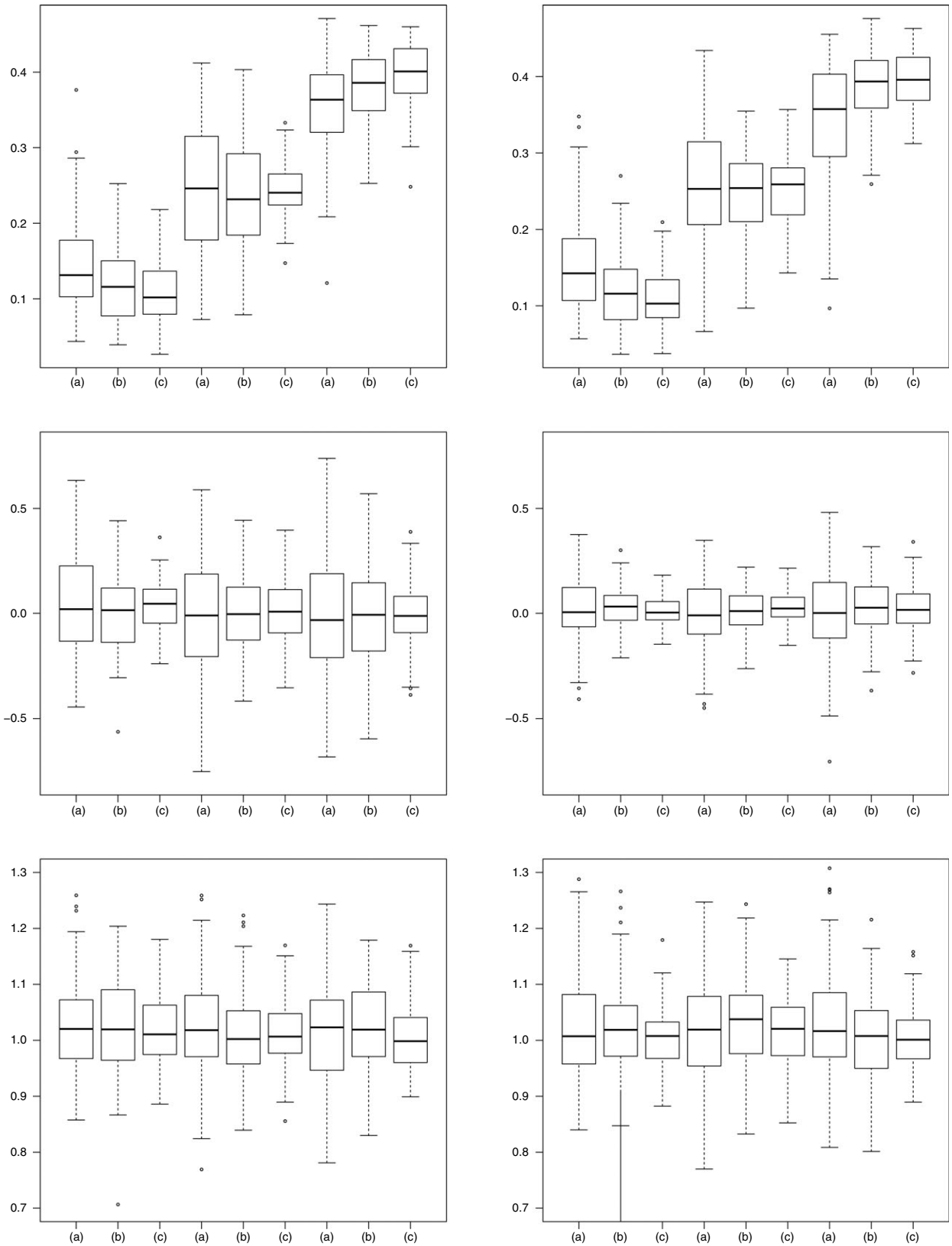

Figure 2. Simulated fMRI data: Boxplots of the estimates of d, β, and σ^2, respectively, for $d = [.1, .25, .4]$. On the x-axis, (a) N=128, (b) N=256, (c) N=512. Results are for the case SNR = .5. Plots on the first column refer to the blocked design, those on the second column to the event-related design.

Table 2.
Simulated fMRI data: Means and standard errors of the estimates of d, β, and σ^2, respectively, for $N = [128, 256, 512]$ and $d = [.1, .25, .4]$, for blocked (upper) and event-related (lower) designs. Results are for the cases $SNR = 5$ and $SNR = 10$. True $\beta = .01$ and true $\sigma^2 = 1$

	Blocked		$\hat{d}$(SE)	$\hat{\beta}$(SE)	$\hat{\sigma}^2$(SE)
		$d = .1$	.14004(.00630)	.01489(.00945)	1.03602(.00975)
	$n = 128$	$d = .25$	.26527(.00697)	.01578(.01062)	1.02133(.00894)
		$d = .4$	.37552(.00565)	−.00071(.01219)	1.01087(.01044)
		$d = .1$	.11932(.00490)	.01325(.00661)	1.00631(.00777)
SNR=5	$n = 256$	$d = .25$	.24712(.00638)	.00796(.00756)	1.01432(.00790)
		$d = .4$	.38564(.00463)	.00705(.00860)	1.00241(.00708)
		$d = .1$	.10089(.00365)	.01452(.00421)	1.01607(.00581)
	$n = 512$	$d = .25$	.25162(.00429)	.00721(.00517)	.99202(.00572)
		$d = .4$	.39609(.00428)	.01133(.00625)	1.00076(.00614)
		$d = .1$	.13055(.00552)	.00983(.00259)	1.03340(.00897)
	$n = 128$	$d = .25$	.24567(.00858)	.00878(.00333)	1.01986(.01047)
		$d = .4$	.35216(.00736)	.01172(.00393)	.99719(.01036)
		$d = .1$	.11878(.00435)	.00908(.00203)	1.01186(.00726)
SNR=10	$n = 256$	$d = .25$	.24958(.00606)	.00850(.00247)	1.01229(.00852)
		$d = .4$	.38397(.00477)	.00776(.00264)	1.02245(.00732)
		$d = .1$	.10596(.00366)	.01036(.00145)	1.00493(.00602)
	$n = 512$	$d = .25$	.24586(.00437)	.00742(.00164)	1.00582(.00568)
		$d = .4$	.39811(.00348)	.00825(.00207)	1.01562(.00544)
	Event-related		$\hat{d}$(SE)	$\hat{\beta}$(SE)	$\hat{\sigma}^2$(SE)
		$d = .1$	.13909(.00507)	.00985(.00475)	1.04570(.01005)
	$n = 128$	$d = .25$	.24166(.00831)	.01354(.00617)	1.02494(.00864)
		$d = .4$	.34774(.00683)	.02298(.00713)	1.01983(.00888)
		$d = .1$	.11986(.00483)	.00656(.00363)	1.01277(.00742)
SNR=5	$n = 256$	$d = .25$	.23882(.00616)	.00642(.00416)	1.00610(.00797)
		$d = .4$	.38640(.00505)	.01408(.00529)	1.01579(.00711)
		$d = .1$	.10418(.00340)	.01216(.00208)	1.01372(.00620)
	$n = 512$	$d = .25$	.24660(.00366)	.00877(.00259)	1.01058(.00550)
		$d = .4$	.39906(.00331)	.01164(.00397)	1.00220(.00618)
		$d = .1$	.15798(.00675)	.01101(.00162)	1.04193(.00924)
	$n = 128$	$d = .25$	.24235(.00700)	.00943(.00167)	1.03319(.01120)
		$d = .4$	.35394(.00608)	.01102(.00233)	1.00481(.00903)
		$d = .1$	.11853(.00419)	.00903(.00111)	1.01699(.00796)
SNR=10	$n = 256$	$d = .25$	.23925(.00627)	.00858(.00107)	1.00886(.00815)
		$d = .4$	.37895(.00470)	.01176(.00152)	1.01226(.00834)
		$d = .1$	.10922(.00397)	.01017(.00062)	.99415(.00676)
	$n = 512$	$d = .25$	.24226(.00446)	.01001(.00084)	1.00384(.00561)
		$d = .4$	.39869(.00383)	.01042(.00125)	.99829(.00542)

3.1 *A Simulation Study for fMRI Data*

There are two major types of experimental designs for fMRI studies: blocked and event-related. In the blocked design, a condition is presented continuously for an extended time interval (block) and different task conditions are usually alternating in time. The increase in fMRI signal in response to a stimulus is additive. This means that the amplitude of the hemodynamic response (HR) increases when multiple stimuli are presented in rapid succession. When each block is alternated with a rest condition in which the HR has enough time to return to baseline, a maximum amount of variability is introduced in the signal. Because of this, blocked designs offer considerable statistical power (Friston et al., 1999). Event-related fMRI refers to a technique for detecting the brain's response to brief stimuli or "events." In the event-related design, discrete and short-duration events are presented with timing and order that may be randomized. Event-related designs allow more real world testing, however, the statistical power of event-related designs is inherently low, because the signal change in the fMRI signal following a single stimulus presentation is typically small, see Tie et al. (2009) and references therein.

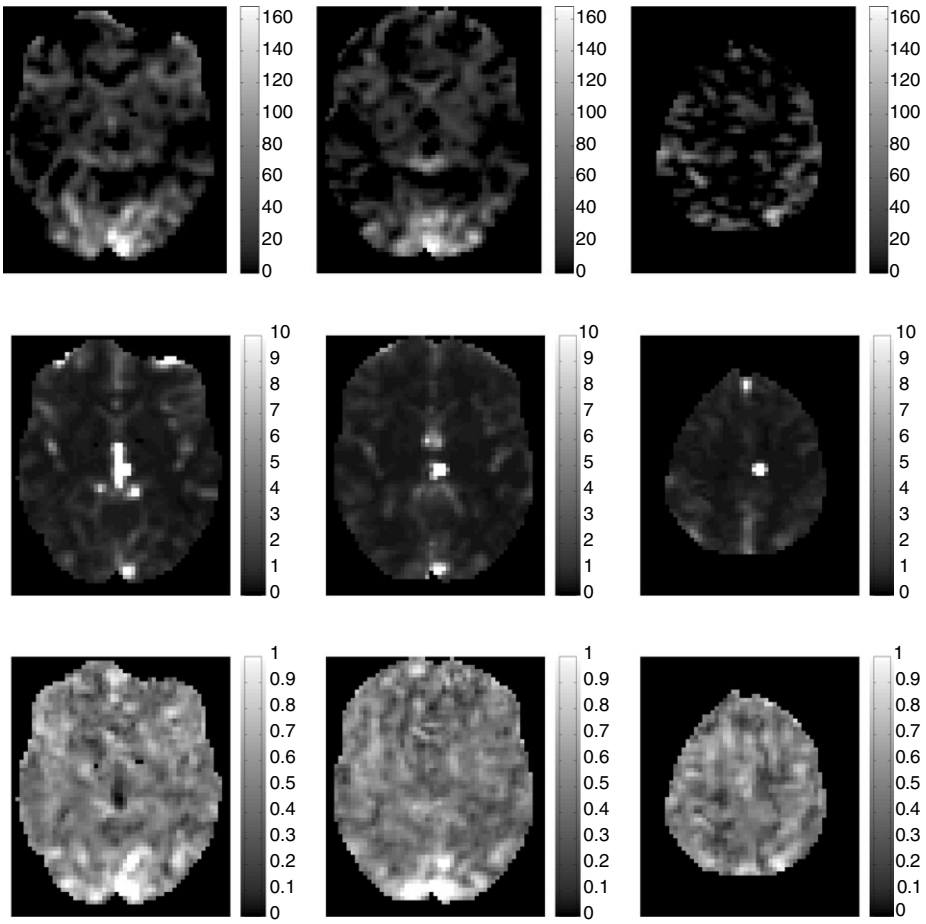

Figure 3. fMRI data. Activation maps obtained by mapping the estimates of the regression parameter β (first row), the innovation variance σ^2 (second row), and the long memory parameter d (third row), at each voxel of single slices of image data, obtained with our method. Images in the first column refer to activations on the primary visual cortex (V1), those in the second column to activations on the motion-selective cortical area (V5), and those in the third column to activations on the posterior parietal cortex (PP).

For the blocked design, we generated artificial fMRI signals by generating a square wave signal as

$$x(t) = A \sum_{-\infty}^{\infty} g(t - kP), \tag{17}$$

with A and P the amplitude and fundamental period of the signal, respectively, and with

$$g(t) = \begin{cases} 1, & 0 \le t < P/2 \\ -1, & P/2 \le t < P \\ 0, & \text{otherwise,} \end{cases} \tag{18}$$

and by convolving the signal with a Poisson HRF (16) with $\lambda = 4$. For the event-related design we simply simulated a series of delta functions at irregular time points. The convolved signal was then embedded with an I(d) error component. Figure 1 illustrates this process.

We set $\beta = .01$ and $\sigma^2 = 1$. We considered three sample sizes, $N = 128, 256$, and 512 and three different values of the long memory parameter, $d = .1, .25$, and $.4$. While we fixed $P = 16$ in the blocked design, in the event-related design we varied the number of stimula (10 for $n = 128$, 20 for $n = 256$, and 40 for $n = 512$), keeping the duration of each stimulus to 10. In both designs, we chose different values of the amplitude parameter A in order to obtain different values of the signal-to-noise ratio (SNR $= .5, 5, 10$), calculated as $\text{SNR} = 10 \log_{10}(A/\sigma^2)$.

We again applied DWTs with Daubechies wavelets and four vanishing moments. MCMC chains were run with 5000 iterations discarding the first 1000 iterations as burn-in. We ran 100 replications for each scenario ($N = 128, 256, 512$ and $d = .1, .25, .4$). Our Matlab code performed 1000 MCMC iterations in 20 seconds for $N = 128$, 35 seconds for $N = 256$, and 2 minutes for $N = 512$, on a laptop computer with CPU Duo 2.33 GHZ and 2GB of RAM. Figure 2 displays boxplots of the estimates of d, β and σ^2, for the different sample sizes and long memory parameter values and $SNR = .5$. Mean squares errors of the estimates (not shown) decreased consistently as the sample size increased, for fixed d. The estimates of d and

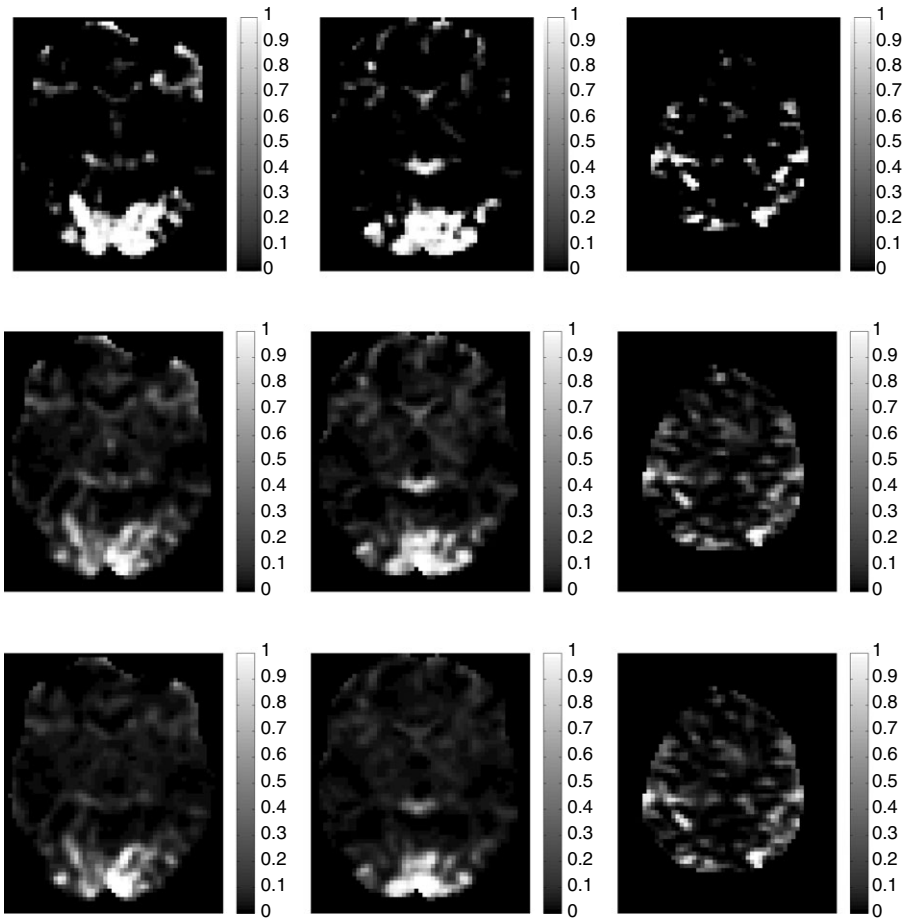

Figure 4. fMRI data. Activation maps obtained by mapping probabilities of the estimates of the linear regression parameter β at each voxel of single slices of image data, thresholded according to the procedure described in the article. First row refers to our method, second row to our implementation of an MLE procedure, and third row to a model with AR(1) noise. Images in the first column refer to activations on the primary visual cortex (V1), those in the second column to activations on the motion-selective cortical area (V5), and those in the third column to activations on the posterior parietal cortex (PP).

σ^2 had a slight negative bias in almost all cases, indicating underestimation. Also, for fixed N, in most cases the MSEs of $\hat{d}$ increased as the value of d approached its boundary region (0 or .5). This behavior was also noticed by Jensen (2000), in his estimation of a model for long memory data contaminated by white noise, and by Cheung and Diebold (1994). We also looked at the results for SNR = 5 and SNR = 10. Table 2 reports means and standard errors for these cases, calculated over 100 replications. Overall, as expected, we noticed that the MSEs were smaller for larger SNR values, that is that the variability of the estimates decreased as the amplitude of the signals increased.

3.2 *Case Study on fMRI Data*

We exemplify our method on the epoch data set provided by the *Wellcome Department of Imaging Neuroscience* and available at http://www.fil.ion.ucl.ac.uk/spm/data/. The data set was originally collected by Büchel and Friston (1997). Below we give some description of how the data set was acquired and refer readers to the original article for more details.

- **Image Acquisition:** The experiment, consisting of four runs, each lasting 5 minutes 22 seconds, was performed on a single subject using a 2 Telsa Magnetom VISION whole body MRI system (Siemens, Erlangen, Germany) equipped with a head volume coil: TE=40 ms, TR=3.22 seconds, and 64×64 imaging matrix [19.2×19.2 cm]. Four hundred T_2^*-weighted fMRI images (100 images for each run) were originally acquired at each of 32 contiguous multislices (slice thickness 3 mm, giving 9.6 cm vertical field of view) in the whole brain, except for the lower half of the cerebellum and the most inferior part of the temporal lobes. The first 10 scans in each run were discarded in order to eliminate magnetic saturation effects. Thus 360 image volumes were used in our estimation procedure.
- **Experimental Design:** The original experiment was performed under four different conditions: "Fixation," "attention," "no attention," and "stationary." The subject was asked to look at a fixation point in the middle of a transparent screen. Under visual motion conditions,

250 white dots moved rapidly from the fixation point in random directions, at a constant speed, toward the border of the screen where they vanished. Before scanning, five 30-second trials of the visual stimulus were given with five different speeds of the moving dots. During the "attention" and "no attention" conditions the subject fixated centrally while the white dots emerged from the fixation point toward the edge of the screen. In the "attention" condition the subject was instructed to detect changes of speed. In the "no attention" condition the instruction was to just look at the moving points. During the "fixation" condition the subject saw only dark screen except for the visible fixated dot. In the "stationary" condition the fixation point and 250 stationary dots were presented to the subject.

We applied our wavelet-based Bayesian model to single voxels on the 32 contiguous slices. For each voxel we had $N = 360$ image volumes. We grouped the four conditions into two categories and defined a vector with elements set to 1 for the images acquired in the "attention" condition and to zero otherwise. We then convolved this vector with the Poisson hemodynamic function with $\lambda = 4$ and used the resulting signal to form the covariate X in model (1). Analyses were carried out in MATLAB. The estimated parameters were mapped on transverse images using the MRI toolbox 2.0 by Darren Weber.

3.2.1 *Conditional PPMs.* Below we show activations detected by mapping the estimates of the model parameters at each voxel of single slices of imaging data and then thresholding the corresponding conditional posterior probabilities at a specified confidence. The resulting PPMs can be seen as a modified version of the posterior maps produced by the software SPM following the method of Friston et al. (2002) and Friston and Penny (2003). Our procedure, in particular, is entirely based on the MCMC samples. Let $\hat{E}_{\beta|y}$ and $\hat{V}_{\beta|y}$ be the mean and variance of a realization of β from the Markov chain at a single voxel. We define the resulting conditional posterior probability p for that voxel as

$$p = 1 - \Phi\left(\frac{\kappa - \hat{E}_{\beta|y}}{\sqrt{\hat{V}_{\beta|y}}}\right), \tag{19}$$

where Φ is the standard normal distribution function and κ is the thresholding rule set as

$$\kappa = \zeta + z_\alpha \sqrt{\hat{V}_{\beta|y}}, \tag{20}$$

with ζ the standard deviation of $\hat{E}_{\beta|y}$ over all voxels in the slice, which we estimate based on the MCMC samples as the standard deviation of the posterior estimates of β over all voxels, and z_α the upper α percentile of the standard normal distribution. In the applications we set α to .05**,** that is, we used $z_\alpha = 1.64$, to obtain activation maps thresholded at a 95% confidence. In the PPM of Friston et al. (2002) and Friston and Penny (2003), the posterior mean and variance of β are used instead of $\hat{E}_{\beta|y}$ and $\hat{V}_{\beta|y}$, respectively, and ζ in (20) is estimated from the prior. One should be careful when interpreting the maps obtained by thresholded conditional posterior probabilities. In particular, a voxel showing a high value does not indicate activation but rather a high chance of being activated. This aids interpretation as one can easily identify the voxels with posterior chances for being activated with a given confidence.

3.2.2 *Results.* In analyzing fMRI data we followed the approach of Wornell and Oppenheim (1992) and McCoy and Walden (1996) and assumed a variance progression formula for the wavelet coefficients of the type

$$\sigma^2 \sigma_{m\,n}^2 = \sigma^2 (2^\alpha)^{-m}, \tag{21}$$

with $\alpha \in (0, 1)$ the long memory parameter and σ^2 the innovation variance. The reason for this choice is that it results in an error term, which is not restricted to a specific class of long memory processes, like $I(d)$, but rather it encompasses the general fractal process of type (2), including long memory processes. In particular, we are not restricting the long memory parameter to the stationary range, as it is instead in the case of $I(d)$ processes (Long et al., 2005). For inference, we used a very similar MCMC procedure to what described in Section 2.4, with Gibbs's steps for β and σ^2 and a Metropolis step on α with a Gaussian proposal centered in the previous value and a standard deviation of .05, which worked well for us.

Neuroimaging studies have shown that a stimulus with visual motion, like the rapidly moving dots in this experiment, activates the primary visual cortex, the motion-selective cortical area, and the posterior parietal cortex, see Bushnell, Goldberg, and Robinson (1981), Mountcastle, Anderson, and Motter (1981), and Treue and Maunsell (1996). We therefore focused on these regions. Figure 3 shows activations maps obtained by mapping the estimates of the regression parameter β (first row), the innovation variance σ^2 (second row), and the long memory parameter α (third row), at each voxel of single slices of image data, obtained with our method. Images in the first column refer to activations on the primary visual cortex (V1; 2275 voxels), those in the second column to activations on the motion-selective cortical area (V5; 2327 voxels), and those in the third column to activations on the posterior parietal cortex (PP; 1325 voxels). Brighter colors on the images denote higher estimated values of the parameter, that is stronger activations (or cerebral responses) to the given visual stimulus.

In Figure 4 we compare the activation maps obtained by mapping probabilities of the estimates of the linear regression parameter β, as described earlier, with equivalent maps obtained with two competing methods: an MLE method (second row), which we implemented as described in the Appendix, and a model with an AR(1) noise (third row) that ignores the long-memory nature of the data. Columns refer to the V1, V5, and PP regions, respectively. Brighter colors on the images denote high chance of strong activations (or cerebral responses) to the given visual stimulus. The areas most likely activated by the given "attentional" visual stimulus correspond to the spatial coordinates of the V1, V5, and PP areas of the subject, as identified by Büchel and Friston (1997). All methods correctly detect activations in the designated areas, with activations that occur quite broadly around the corresponding

coordinates of V1, V5, and PP. Notice, however, that the detection is much sharper with our Bayesian maps.

Other authors have compared estimation performances of models for fMRI data that make different assumptions on the error structure and have pointed out that models ignoring the long memory autocorrelation in the errors result in standard errors of the estimates that are inflated, see for example, Friston and Penny (2003). Our maps in Figure 3 confirm that fMRI data are indeed contaminated by long memory errors and that some of the cerebral responses have large values of α. Also, they show that the innovation variance of the long memory error is not uniformly distributed over the brain and that it has relatively large values in very small parts around the center area and in outer area of the brain. Several authors have pointed out that fMRI signals are contaminated by instrumental and physiological noises. The sources for such errors are, however, not completely known. Long memory noise, in particular, has been attributed to head movement caused by slow rotation or translation during scanning, as well as to cardiac and respiratory cycle-related pulsations. Neurophysiological sources could be hypothesized too. However, it is still an open question whether fMRI noise may indeed reveal information on brain activity, such as long-range dependencies in response to stimula.

4. Conclusions

In this article we have proposed a wavelet-based method to estimate the parameters of a regression model with an $1/f$ error component. We have carried out estimation in the wavelet domain in order to simplify the treatment of the dense covariance matrix of the long memory error, and have used a Bayesian approach for the estimation of the model parameters. Our inferential procedure uses exact wavelet coefficients variances and leads to accurate estimates of the model parameters. Although we have chosen to mainly focus on the fractionally integrated long memory processes of Hosking (1981) and Granger and Joyeux (1980), our inferential procedure can be easily extended to handle more general classes of long memory processes. For example, in the application of this article we have employed a very general variance progression formulation. Whitening properties of DWTs for long memory processes are well documented (Tewfik and Kim, 1992; Craigmile and Percival, 2005; Ko and Vannucci, 2006, among others). Furthermore, wavelet packets have been investigated as a more general tool that can decorrelate other processes than the long memory, see Percival, Sardy, and Davison (2001) and Gabbanini et al. (2004).

We have evaluated performances of our method on simulated data and also showed an application to fMRI data. Results have confirmed that our wavelet-based approach is suitable for applications to fMRI data. There we have produced PPMs that aid interpretation, as one can easily identify voxels that are likely activated with a given confidence. Single-subject scanning is receiving renewed interest in the fMRI field, due to its use for presurgical purposes. In addition, our approach can be used to produce single-subject posterior maps as a form of meta-analysis for intersubject investigations, similarly in spirit to other Bayesian approaches to fMRI modeling (Bowman et al., 2008).

As a possible extension of the model we have presented, in a basic study on the analysis of fMRI time series (Zarahn et al., 1997) argue that fMRI signals are contaminated by a white noise in addition to long memory noise, and Wornell and Oppenheim (1992) develop an estimation procedure of the model parameters under two additive noises. Other interesting extensions include models accounting for the estimation of the HRF and/or Bayesian spatiotemporal models that incorporate spatial correlation among brain responses via network priors (Gossl, Auer, and Fahrmeir, 2001; Quiros, Montes-Diez, and Gamerman, 2010).

5. Supplementary Materials

Matlab code and fMRI data are available with this article at the Biometrics website on Wiley Online Library.

Acknowledgement

Vannucci is partially supported by NSF/DMS grant DMS-1007871.

References

Abry, P., Flandrin, P., Taqqu, M. S., and Veitch, D. (2003). Self-similarity and long-range dependence through the wavelet lens. In *Theory and Applications of Longrange Dependence*, P. Doukhan, G. Oppenheim, and M. Taqqu (eds), 526–556. Boston: Birkhauser.

Aguirre, G., Zarahn, E., and D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics II. *NeuroImage* **5**, 199–212.

Beran, J. (1994). *Statistics for Long-Memory Processes*. New York: Chapman & Hall.

Bowman, F., Caffo, B., Bassett, S., and Kilts, C. (2008). Bayesian hierarchical framework for spatial modeling of fMRI data. *NeuroImage* **39,** 146–156.

Büchel, C. and Friston, K. (1997). Modulation of connectivity in visual pathways by attention: Cortical interactions evaluated with structural equation modelling and fMRI. *Cerebral Cortex* **7,** 768–778.

Bullmore, E., Fadili, J., Maxim, V., Sendur, L., Whitcher, B., Suckling, J., Brammer, M., and Breakspear, M. (2004). Wavelets and functional magnetic resonance imaging of the human brain. *NeuroImage* **23,** 234–249.

Bushnell, M., Goldberg, M., and Robinson, D. (1981). Behavioral enhancement of visual responses in monkey cerebral in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology* **46,** 755–772.

Buxton, R. and Frank, L. (1997). A model for the coupling between cerebral blood flow and oxygenation metabolism during neural stimulation. *Journal of Cerebral Blood Flow Metabolism* **17,** 64–72.

Cheung, Y. and Diebold, F. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics* **62,** 301–306.

Craigmile, P., Guttorp, P., and Percival, D. (2005). Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes. *IEEE Transactions on Signal Processing* **53,** 3151–3161.

Craigmile, P. F. and Percival, D. B. (2005). Asymptotic decorrelation of between-scale wavelet coefficients. *The IEEE Transactions on Information Theory* **51,** 1039–1048.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, Volume **61**: *CBMS-NSF Conference Series*. SIAM.

Fadili, M. and Bullmore, E. (2002). Wavelet-generalised least squares: A new BLU estimator of linear regression models with $1/f$ errors. *NeuroImage* **15,** 217–232.

Fox, R. and Taqqu, M. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* **14,** 517–532.

Friston, K., Holmes, A., Price, C., Buchel, C., and Worsley, K. (1999). Multisubject fMRI studies and conjunction analyses. *NeuroImage* **10,** 385–396.

Friston, K. and Penny, W. (2003). Posterior probability maps and SPMs. *NeuroImage* **19,** 1240–1249.

Friston, K., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Theory. *NeuroImage* **16,** 465–483.

Gabbanini, F., Vannucci, M., Bartoli, G., and Moro, A. (2004). Wavelet packet methods for the analysis of variance of time series with application to crack widths on the Brunelleschi dome. *Journal of Computational and Graphical Statistics* **13,** 639–658.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4,** 221–237.

Glover, G. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage* **9,** 416–429.

Gossl, C., Auer, D., and Fahrmeir, L. (2001). Bayesian modeling of the hemodynamic response function in BOLD fMRI. *Neuroimage* **14,** 140–148.

Granger, C. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* **1,** 15–29.

Guo, Y., Bowman, F., and Kilts, C. (2008). Predicting the brain response to treatment using a Bayesian hierarchical model with application to a study of schizophrenia. *Human Brain Mapping* **29,** 1092–1109.

Haslett, J. and Raftery, A. (1989). Space-time modeling with long-memory dependence: Assessing Ireland's wind power resource. *Journal of Applied Statistics* **38,** 1–50.

Hosking, J. (1981). Fractional differencing. *Biometrika* **68,** 165–176.

Jensen, M. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control* **24**, 361–386.

Ko, K. and Vannucci, M. (2006). Bayesian wavelet analysis of autoregressive fractionally integrated moving-average processes. *Journal of Statistical Planning and Inference* **136,** 3415–3434.

Li, W. and McLeod, A. (1986). Fractional time series modelling. *Biometrika* **73,** 217–221.

Lindquist, M. (2008). The statistical analysis of fMRI data. *Statistical Science* **23,** 439–464.

Long, C., Brown, E., Triantagfyllou, C., Aharon, I., Wald, L., and Solo, V. (2005). Nonstationary noise estimation in functional MRI. *NeuroImage* **28,** 890–903.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11,** 674–693.

McCoy, E. and Walden, A. (1996). Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics* **5,** 26–56.

McLeod, A. and Hipel, K. (1978). Preservation of the rescaled adjusted range. Parts 1, 2 and 3. *Water Resources Research* **14**, 491–512.

Meyer, F. (2003). Wavelet-based estimation of a semiparametric generalized linear model of fMRI time-series. *IEEE Transactions on Medical Imaging* **22,** 315–322.

Mountcastle, V., Anderson, R., and Motter, B. (1981). The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience* **1,** 1218–1225.

Percival, D., Sardy, S., and Davison, A. (2001). Wavestrapping time series: Adaptive wavelet-based bootstraping. In *Nonlinear and Nonstationary Signal Processing*, W. Fitzgerald, R. Smith, A. Walden, and P. Young (eds). Cambridge, England: Cambridge University Press.

Quiros, A., Montes-Diez, R., and Gamerman, D. (2010). Bayesian spatio temporal model of fMRI data. *NeuroImage* **49,** 442–456.

Smith, A., Lewis, B., Ruttimann, U., Ye, F., Sinwell, T., Yang, Y., Duyn, J., and Frank, J. (1999). Investigation of low frequency drift in fMRI signals. *NeuroImage* **9,** 526–533.

Stoev, S. and Taqqu, M. S. (2005). Asymptotic self-similarity and wavelet estimation for long-range dependent fractional autoregressive integrated moving average time series with stable innovations. *Journal of Time Series Analysis* **26,** 211–249.

Taswell, C. and McGill, K. (1994). Wavelet transform algorithms for finite-duration discrete-time signals. *ACM Transactions on Mathematical Software* **20,** 398–412.

Tewfik, A. and Kim, M. (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory* **38,** 904–909.

Tie, Y., Suarez, R., Whalen, S., Radmanesh, A., Nortan, I., and Golby, A. (2009). Comparison of blocked and event-related fMRI designs for pre-surgical language mapping. *NeuroImage* **47,** T107–T115.

Treue, S. and Maunsell, H. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature* **382,** 539–541.

Turkheimer, F., Aston, J., Banati, R., Riddle, C., and Cunningham, V. (2003). A linear wavelet filter for parametric imaging with dynamic. *IEEE Transactions on Medical Imaging* **22,** 289–301.

Vannucci, M. and Corradi, F. (1999). Covariance structure of wavelet coefficients: Theory and models in a Bayesian perspective. *Journal of the Royal Statistical Society, Series B* **61,** 971–986.

Woolrich, M., Jenkinson, M., Brady, J., and Smith, S. (2004). Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging* **23,** 213–231.

Wornell, G. and Oppenheim, A. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Transactions on Signal Processing* **40,** 611–623.

Zarahn, E., Aguirre, G., and D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics I. *NeuroImage* **5,** 179–197.

Received October 2011. *Revised July* 2012.
Accepted July 2012.

Appendix

Derivation of the MLE of $\boldsymbol{\Theta} = (\beta, \sigma^2, \eta)$

The likelihood function of model (10) with one explanatory variable X_w is

$$L(\boldsymbol{\Theta}) = \frac{(\sigma^2)^{-N/2}|\boldsymbol{\Sigma}_\Psi|^{-1/2}}{(\sqrt{2\pi})^N} \exp \times \left[-\frac{1}{2\sigma^2}(\mathbf{y}_w - \mathbf{X}_w\beta)'\boldsymbol{\Sigma}_\Psi^{-1}(\mathbf{y}_w - \mathbf{X}_w\beta)\right],$$

where $\boldsymbol{\Sigma}_\Psi$ is a diagonal matrix whose elements are $(2^\alpha)^{-m}, m = 1, \ldots, r$. We consider only one explanatory variable, but the derivation can be easily extended to multiple explanatory variables. Let $\eta = 2^\alpha$ in the variance progression

(21). Maximizing $L(\mathbf{\Theta})$, we have the following maximum likelihood estimates of $\mathbf{\Theta}$:

$$\hat{\eta} = \arg\max_{\eta} \sum_{m=1}^{r} T_m K_m \eta^m ,$$

$$\hat{\sigma}^2 = \sum_{m=1}^{r} K_m \hat{\eta}^m \Big/ \sum_{m=1}^{r} N(m),$$

$$\hat{\beta} = \frac{\sum_{m=1}^{r} \sum_{n=1}^{N(m)} \hat{\eta}^m y_{m,n} X_{m,n}}{\sum_{m=1}^{r} \sum_{n=1}^{N(m)} \hat{\eta}^m X_{m,n}^2},$$

where $T_m = \sum_{m=1}^{r} mN(m) - m \sum_{m=1}^{r} N(m)$ and $K_m = \sum_{n=1}^{N(m)} (y_{m,n} - \beta X_{m,n})^2$. The estimates $\hat{\eta}$ and $\hat{\sigma}^2$ follow from Wornell and Oppenheim (1992) via the normal equations,

$$\sigma^2 \sum_{m=1}^{r} N(m) = \sum_{m=1}^{r} \eta^m K_m ,$$

$$\sum_{m=1}^{r} mN(m) = \sigma^{-2} \sum_{m=1}^{r} m\eta^m K_m .$$

Figure 4 shows activation maps obtained by mapping probabilities of the estimates of the linear regression parameter β at each voxel of a single slice obtained with the method described earlier. To obtain these maps, we applied a thresholding rule similar to the one we employ in our Bayesian estimation method, that is,

$$p = 1 - \Phi\left(\frac{\vartheta - \hat{\beta}}{\sqrt{\hat{V}(\beta)}} \right),$$

where ϑ is set to $\vartheta = \hat{E}(\beta) + z_\alpha \sqrt{\hat{V}(\beta)}$ with $\hat{E}(\beta)$ and $\sqrt{\hat{V}(\beta)}$ the mean and standard deviation of the estimated β's over all voxels in the slice. Here $\hat{\beta}$ denotes the estimate of β at each voxel. We set $\alpha = 05$, that is, $z_\alpha = 1.64$.