Wavelet-based Estimation of Linear Regression Models with an Application to fMRI Image Data

Kyungduk Ko** and Marina Vannucci *

May 15, 2006

To appear in IEEE Transactions on Signal Processing

SUMMARY

Linear regression models with long memory noise have been proven useful for application in many areas, such as medical imaging, signal processing, and econometrics. In this paper we analyze a linear regression model with two error components, a long memory and a white noise. We employ discrete wavelet transforms in order to simplify the dense variance-covariance matrix of the additive error structure. We then adopt an EM algorithm for the estimation of the model parameters. We first evaluate performances on simulated data and then show how the proposed model can be used for the analysis of real fMRI data.

KEYWORDS: Discrete Wavelet Transforms, EM Algorithm, Functional Magnetic Resonance Imaging, Long Memory

**Department of Mathematics, Boise State University, Boise, ID 83725-1555. *Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA. Phone: 979-845-3141, Fax: 979-845-3144. KO@MATH.BOISESTATE.EDU and MVANNUCCI@STAT.TAMU.EDU
1 Introduction

Data from long memory processes have the distinctive feature that the correlation between distant observations is not negligible. Fractional Brownian motion (fBm) and fractional Gaussian noise (fGn) (Mandelbrot and van Ness, 1968) are continuous models with long memory behavior, while autoregressive fractionally integrated moving-average (ARFIMA) models, first introduced by Granger and Joyeaux (1980) and Hosking (1981), are typical examples of discrete-time long memory processes. The main characteristic of long memory processes is that the spectral density function goes to infinity as the frequency goes to zero. This characteristic leads to dense variance-covariance matrices that make existing inferential methods computationally expensive. 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). Sowell (1992) derived the exact covariance matrix to compute the likelihood function under the assumption that the roots of the autoregressive polynomial are simple. Beran (1994) investigated asymptotic sampling theory properties of exact and approximate maximum likelihood methods.

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. 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). Many authors have studied wavelet-based estimation of long memory models and have investigated applications in various fields, such as signal processing and financial time series. Wornell and Oppenheim (1992) proposed an estimation procedure of the parameters of a fGn signal with additive white noise that uses the Expectation-Maximization (EM) algorithm of Laird et al. (1977). For ARFIMA models, Jensen (2000) also applied an EM method for the estimation of the long memory parameter and, recently, Ko and Vannucci (2006) have proposed wavelet-based Bayesian estimation procedures for the long memory and the other model parameters.
In this paper we deal with a linear regression model where the error terms are strongly correlated, and specifically long memory. For such model the best linear unbiased estimates (BLUE) of the parameters are usually more efficient than the ordinary least squares (OLS) estimates. Their computation, however, is prohibitive because of the necessary iterative estimation procedure of dense variance-covariance matrices. Beran (1994) asymptotically assessed the loss when ordinary least squares (OLS) estimates are used instead of the BLUE. Regression models with correlated errors have been used in the literature of medical imaging for the analysis of functional magnetic resonance imaging (fMRI) data. Fadili and Bullmore (2002) analyzed a simple linear regression model with fractional Brownian motion (fBm) as its error term. They used discrete wavelet transforms (DWT) to derive a near Karhunen-Loève-type expansion of the variance-covariance structure of the long memory error and adopted the golden section search method, a variation of the bisection method, in order to get approximate maximum likelihood estimates (MLE) of the model parameters. Meyer (2003) applied generalized linear model with drifts and errors contaminated by long-range dependencies to fMRI image data.

The particular model we adopt in this paper is motivated by Zarahn et al. (1997), who argued that fMRI time series are often dominated by both long-range dependencies and by white noise. We therefore consider a linear regression model with a long memory noise embedded in a white noise and propose a wavelet-based EM algorithm for the estimation of the model parameters. For the long memory error we assume an ARFIMA(0,d,0) process, the simplest form of a fractional ARIMA, also known as I(d) process. We use discrete wavelet transforms in order to simplify the variance-covariance structure of the response variable and propose an extension of the EM algorithm of Wornell and Oppenheim (1992) and Jensen (2000) for the estimation of the model parameters. The novel model we propose is quite flexible and includes as special cases both the widely adopted long memory regression model (without the white noise component) and the model without the trend parameters adopted by Wornell and Oppenheim (1992) and by Jensen (2000).

The remainder of this paper is organized as follows: In Section 2, we introduce the
necessary basic concepts on long memory processes and on discrete wavelet transforms. In Section 3 we describe our wavelet-based linear regression model with additive $I(d)$ and white noise and illustrate the EM method we implemented for the parameters estimation. In Section 4 we report the results from a simulation study and from a real application to functional magnetic resonance imaging data. Details of the proposed EM algorithm are given in the Appendix.

2 Preliminaries

2.1 Long memory processes

Long memory processes, also known as $1/f$-like processes, have a spectral density function (SDF) of the form

$$S(f) \sim \frac{\sigma_L^2}{|f|^\gamma},$$

where $f$ indicates frequency and where $\gamma$ is a spectral parameter. In this paper we consider fractionally integrated long-memory processes $I(d)$ of the type

$$(1 - B)^d X_t = \epsilon_t$$

with $d > -0.5$, $B$ the backshift operator, $d$ the long memory parameter and $\epsilon_t$ a zero mean white noise with an innovation variance $\sigma_L^2$. This process is also known as the simplest autoregressive fractionally integrated moving-average process, or ARFIMA$(0, d, 0)$. The fractional difference operator $\Delta^d = (1 - B)^d$ can be defined as a power series in $B$ for $d \neq 0$. By using the binomial expansion we have, for any scalar $d$,

$$(1 - B)^d \equiv \sum_{j=0}^{\infty} \pi_j B^j = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j,$$

where $\binom{d}{j} = d! / j! (d - j)! = \Gamma(d + 1) / \Gamma(j + 1) \Gamma(d - j + 1)$. The autocovariance function of an $I(d)$ process is given by

$$\gamma(\tau) = \sigma_L^2 \frac{\Gamma(1 - 2d)\Gamma(d + \tau)}{\Gamma(1 - d + \tau)\Gamma(1 - d)\Gamma(d)}, \quad \tau = 1, 2, 3, \ldots,$$
and its autocorrelation function by

\[ \rho(\tau) = \frac{\Gamma(d + \tau)\Gamma(1 - d)}{\Gamma(1 - d + \tau)\Gamma(d)}. \]  

The long memory parameter \( d \) is related to more well known Hurst exponent \( H \) by the relationship

\[ d = H - \frac{1}{2}. \]  

\( I(d) \) processes are stationary and invertible if \(-0.5 < d < 0.5\). They are long memory for \( 0 < d < 0.5 \), short memory for \(-0.5 < d < 0 \) and have no memory for \( d = 0 \) (Hosking, 1981).

### 2.2 Discrete wavelet transforms

Wavelets are families of orthonormal basis functions that can be used to parsimoniously represent other functions. For example, in \( L_2(\mathbb{R}) \) an orthogonal wavelet basis is obtained by dilating and translating a mother wavelet \( \psi(t) \), which is squared integrable and satisfies the admissibility condition

\[ \int \psi(t) dt = 0. \]  

The wavelet family is defined as

\[ \psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n), \]  

with \( m, n \in \mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\} \), that is as a collection of dilated and translated versions of \( \psi(t) \). The smoothness (or regularity) of the wavelet functions depends on their number of vanishing moments, i.e., \( \int x^k \psi(t) dt = 0, \ k = 0, 1, 2, \ldots, k_0 - 1 \). Any function \( x(t) \in L_2(\mathbb{R}) \) can be represented by a wavelet expansion of the type

\[ x(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \omega_{m,n} \psi_{m,n}(t) \]  

with wavelet coefficients \( \omega_{m,n} = \int x(t) \psi_{m,n}(t) dt \) describing features of the function \( x(t) \) 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 and let us indicate
the series as \( \mathbf{X} = (x_1, \ldots, x_N) \) with \( N = 2^J \) and \( J \) a positive integer denoting the scale of the data. Using \( 2^J \) points, with \( J \) an integer, is not a real restriction and methods exist to overcome this limitation allowing wavelet transforms to be applied to any length of data. A discrete wavelet transform (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 \( r \) is often represented in matrix form as \( \mathbf{Z} = \mathbf{W} \mathbf{X} \), with \( \mathbf{W} \) an \( N \times N \) orthogonal matrix of the form

\[
\mathbf{W} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \ldots, \mathbf{W}_r^T, \mathbf{V}_r^T]^T, \tag{10}
\]

that decomposes the data into sets of coefficients

\[
\mathbf{Z} = [\mathbf{z}_1^T, \mathbf{z}_2^T, \ldots, \mathbf{z}_r^T, \mathbf{x}_r^T]^T. \tag{11}
\]

Here \( m \in M = \{1, 2, \ldots, r\} \) denotes the level of the DWT, with larger values of \( m \) indicating coarser approximation and with \( r \) being the maximum level of the transform. The vector \( \mathbf{z}_m = \mathbf{W}_m \mathbf{X} \) is of dimension \( N(m) = N/2^m, m = 1, 2, \ldots, r \) and \( \mathbf{x}_r = \mathbf{V}_r \mathbf{X} \) is of dimension \( N/2^r \), with \( N = N' + N/2^r \) and \( N' = \sum_{m=1}^r N(m) \). Coefficients \( \mathbf{x}_r \) are called scaling coefficients and represent a coarser approximation of the data, while coefficients \( \mathbf{z}_1, \ldots, \mathbf{z}_r \) are wavelet coefficients representing local features of the data at different resolution scales. An inverse transformation exists to reconstruct the data from its wavelet decomposition.

### 3 Methods

#### 3.1 The Model

In this paper we consider regression models with additive Gaussian fractionally integrated and white noise error components. The model can be written as

\[
\mathbf{y}_t = \mathbf{X}_t \beta + \mathbf{v}_t + \mathbf{w}_t, \quad t = 1, \ldots, N, \tag{12}
\]

where \( \mathbf{y}_t \) is the \( (N \times 1) \) response data, \( \mathbf{X}_t \) is the \( (N \times p) \) design matrix, \( \beta \) is the \( (p \times 1) \) regression coefficient vector, \( \mathbf{v}_t \) is a zero-mean Gaussian \( 1/f \)-type noise, and \( \mathbf{w}_t \) is a zero-
mean Gaussian white noise. We assume that $v_t$ and $w_t$ are independent of each other.

Regression models with correlated errors have been used in the literature of medical imaging for the analysis of functional magnetic resonance imaging (fMRI) data. The particular model we adopt is motivated by Zarahn et al. (1997), who argued that fMRI time series are often dominated by both long-range dependencies and white noise. Model (12), in particular, allows voxel-wise analysis of fMRI data, that is, $y_t$ is the fMRI time series at a given voxel, $X_t$ is the design matrix at time $t$, and $\beta$ denotes the strength of the response at the voxel under consideration.

In our approach we take advantage of the decorrelation properties of the wavelets and apply a discrete wavelet transform (DWT) to both sides of model (12). The DWT serves 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, see for example Beran (1994). Simpler models, instead, can be used for wavelet coefficients. Decorrelation properties of the wavelet transforms for long memory processes are well documented in the literature. Tewfik and Kim (1992) proved that the correlation between wavelet coefficients decreases exponentially fast across scales and hyperbolically fast along time. In the case of ARFIMA models, Jensen (2000) and Ko and Vannucci (2006) provide evidence that these rates of decay allow the DWT to do a credible job at decorrelating the highly autocorrelated long memory structure. These results, of course, strictly depend on the long-memory structure and do not apply to other processes. Percival et al. (2001) suggest to look at wavelet packets as a way to decorrelate processes for which the standard DWT’s fail, such as for short-memory processes. See also Gabbanini et al. (2004).

In the context of our model (12) applying a DWT to the data allows us to get an approximately diagonalized variance-covariance matrix of the $1/f$-type noise component $v_t$. The model can be expressed in the wavelet domain as

$$y_w = X_w \beta + v_w + w_w,$$

where $y_w$, $v_w$ and $w_w$ are column vectors of wavelet coefficients resulting from applying the DWT to $y_t$, $v_t$ and $w_t$, respectively, and where $X_w$ is a matrix resulting from applying the
DWT to the columns of $X_t$. Hereafter we omit the time index $t$.

We use results from Wornell and Oppenheim (1992) and McCoy and Walden (1996) and assume that $v_{m,n}$, the element of the vector $\mathbf{v}_w$ corresponding to the $n$th wavelet coefficient at level $m$, is normally distributed with mean 0 and variance

$$\text{var}(v_{m,n}) = \sigma^2(2^{2d})^{-m},$$

where $\sigma^2 = (2\pi)^{-2d}\sigma_L^22^{2Jd}[2 - 2^{2d}]/[1 - 2d]$ with $d$ the long memory parameter and $\sigma_L^2$ the innovation variance. Here $n \in \{1, 2, \ldots, N/2^m\}$. Wornell and Oppenheim (1992) proposed formula (14) for computing the variances of the wavelet coefficients in the context of the estimation of a $1/f$-like signal with additive white noise. McCoy and Walden (1996) adopted the same idea in the context of a wavelet-based maximum likelihood estimation method for the long memory parameter of an I($d$) process. Jensen (2000) extended their approach to the estimation of ARFIMA($p,d,q$) models.

For estimation purposes we re-parameterize the variance of the process via

$$\eta = 2^{2d}.$$  \hspace{1cm} (15)

The maximum likelihood estimate of the long-memory parameter $d$ can be easily calculated from the MLE of $\eta$ using its invariant property, that is

$$\hat{d} = \frac{1}{2} \log_2 \hat{\eta}.$$

Note that the first two moments of $\mathbf{y}_w$ are $E(\mathbf{y}_w) = \mathbf{X}_w\beta$ and $\text{Cov}(\mathbf{y}_w) = \Sigma = \Sigma_v + \Sigma_w$ where $\Sigma_v$ is the $(N' \times N')$ diagonalized covariance matrix of the long memory noise $\mathbf{v}_w$ with diagonal elements $\sigma^2\eta^{-m}$, $m = 1, 2, \ldots, r$ and $\Sigma_w = \sigma^2_w I$ is the $(N' \times N')$ diagonal covariance matrix of the white noise $\mathbf{w}_w$. Since $\mathbf{y}$ is multivariate normal and the DWT is a linear transformation, it follows that $\mathbf{y}_w$ is also normally distributed.

### 3.2 Estimation of the model parameters

We want to obtain maximum likelihood estimate (MLE) of the parameter vector $\Theta = (\beta, \sigma^2, \eta, \sigma^2_w)$ of the transformed model (13). In that model $\mathbf{y} = (\mathbf{y}_w, \mathbf{X}_w)$ is the observed
(incomplete) data while $\mathbb{V} = (\mathbf{v}_w)$ is unobserved. Then $(\mathbb{Y}, \mathbb{V})$ consists of our complete data. The likelihood $L(\Theta)$ and log-likelihood $\ell(\Theta)$ of the incomplete data can be written as

$$L(\Theta) = \frac{|\Sigma|^{-1/2}}{(\sqrt{2\pi})^{N'}} \exp \left\{ -\frac{1}{2} (\mathbf{y}_w - \mathbf{X}_w\beta)'\Sigma^{-1}(\mathbf{y}_w - \mathbf{X}_w\beta) \right\}$$

and

$$\ell(\Theta) = -\frac{1}{2} \left\{ N' \log(2\pi) + \log |\Sigma| + (\mathbf{y}_w - \mathbf{X}_w\beta)'\Sigma^{-1}(\mathbf{y}_w - \mathbf{X}_w\beta) \right\},$$

respectively. We need to differentiate $\ell(\Theta)$ with respect to $\beta$, $\sigma^2$, $\eta$ and $\sigma_w^2$ in order to get their MLE estimates. However the resulting normal equations, whose solutions give the stationary points of $\ell(\Theta)$, are not easy to solve. We therefore resort to the use of an Expectation-Maximization (EM) algorithm (Laird et al., 1977), an iterative algorithm often used to get MLE estimates. An EM algorithm consists of two steps, which are called the expectation step (E-step) and the maximization step (M-step), respectively. We report details of the derivations of our EM algorithm in the Appendix. We deal directly with the log-likelihood of the complete data $(\mathbb{Y}, \mathbb{V})$, which is

$$Q(\Theta|\hat{\Theta}) = -\frac{1}{2} \left\{ N' \log(2\pi) + N' \log \sigma_w^2 + \sigma_w^{-2}(\mathbf{y}_w - \mathbf{X}_w\beta)'(\mathbf{y}_w - \mathbf{X}_w\beta) \right\}$$

$$-\frac{1}{2} \left\{ \left[ \sigma_w^{-2}\text{tr}(A(\hat{\Theta})) + \sigma_w^{-2}B(\hat{\Theta})'B(\hat{\Theta}) \right] - 2\sigma_w^{-2}B(\hat{\Theta})'(\mathbf{y}_w - \mathbf{X}_w\beta) \right\}$$

$$-\frac{1}{2} \left\{ N' \log(2\pi) + \log \sigma^2 + \log |\Sigma_\eta| + \sigma^{-2}[\text{tr}(\Sigma_\eta^{-1}A(\hat{\Theta})) + B(\hat{\Theta})'\Sigma_\eta^{-1}B(\hat{\Theta})] \right\},$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix. Let $\hat{\Theta} = (\hat{\beta}, \hat{\sigma}^2, \hat{\eta}, \hat{\sigma}_w^2)$ denote the estimates of $\Theta = (\beta, \sigma^2, \eta, \sigma_w^2)$ at a given iteration. Given the current estimate $\hat{\Theta}$, the E-step updates the diagonal matrices $A(\Theta)$ and $B(\Theta)$ as

$$A(\Theta) = (\Sigma_v + \Sigma_w)^{-1}$$

and

$$B(\Theta) = (\Sigma_v + \Sigma_w)^{-1}\Sigma_w(\mathbf{y}_w - \mathbf{X}_w\beta).$$
In the M-step we evaluate the new parameter estimates \( \hat{\Theta} = (\hat{\beta}, \hat{\sigma}^2, \hat{\eta}, \hat{\sigma}_w^2) \) using \( A(\hat{\Theta}) \) and \( B(\hat{\Theta}) \) from the E-step. The resulting estimates are

\[
\hat{\eta} \leftarrow \sum_{m=1}^{r} T_m K_m(\hat{\Theta}) \hat{\eta}^m = 0 \tag{22}
\]

\[
\hat{\sigma}^2 = \frac{\sum_{m=1}^{r} K_m(\hat{\Theta}) \hat{\eta}^m}{N'} \tag{23}
\]

\[
\hat{\beta} = (X_w'X_w)^{-1} \left[ X_w'y_w - B(\hat{\Theta})'X_w \right] \tag{24}
\]

\[
\hat{\sigma}_w^2 = \frac{(y_w - X_w\hat{\beta})'(y_w - X_w\hat{\beta}) + \text{tr}(A(\hat{\Theta})) + B(\hat{\Theta})'[B(\hat{\Theta}) - 2(y_w - X_w\hat{\beta})]}{N'}
\]

where \( T_m = \sum_{m=1}^{r} mN(m) - mN' \) and \( K_m(\hat{\Theta}) = \sum_{n=1}^{N'/m} [A_n^m(\hat{\Theta}) + B_n^m(\hat{\Theta})^2] \) with \( A_n^m(\hat{\Theta}) \) the diagonal element of the matrix \( A(\hat{\Theta}) \) corresponding to the \( n \)th location at level \( m \) and \( B_n^m(\hat{\Theta}) \) the diagonal element of the matrix \( B(\hat{\Theta}) \) corresponding to the \( n \)th location at level \( m \). E-step and M-step are iterated and repeated until a desired tolerance is reached. Note that the resulting estimates are biased.

The complete-data log-likelihood, \( Q(\Theta|\hat{\Theta}) \) given by (19), is continuous in \( \Theta \) and \( \hat{\Theta} \) and therefore the estimates converge to stationary points of the incomplete-data log-likelihood \( \ell(\Theta) \) given by (18) (Wu, 1983). The algorithm iterates towards the maximum likelihood estimates of \( \beta, \sigma^2, \eta \) and \( \sigma_w^2 \). In the simulation study presented below we tried several starting values of the parameters to empirically ensure that the resulting estimates are global maximum values.

### 4 Applications

#### 4.1 Simulation study

There are various ways to generate a time series that exhibits long memory properties. In this paper we use a computationally simple method that was proposed by McLeod and Hipel (1978) and that involves the Cholesky decomposition of the correlation matrix \( R_\varepsilon(i,j) = [\rho(|i-j|)] = [\rho(\tau)] \), where \( \tau = |i-j| = 1, 2, 3, \ldots, N-1 \). Given \( R_\varepsilon = MM' \) with \( M = [m_{i,j}] \)
a lower triangular matrix, let \( \epsilon_t, t = 1, \ldots, n \) be a Gaussian white noise series with zero mean and unit variance, then the series

\[
epsilon_t = \gamma(0)^{1/2} \sum_{i=1}^{t} m_{t,i} \epsilon_i
\]  

has autocorrelation function \( \rho(\tau) \).

In our simulation study we generated artificial fMRI signals by generating a square wave signal as

\[
x(t) = A \sum_{-\infty}^{\infty} g(t - kP),
\]

where \( A \) and \( P \) are the amplitude and fundamental period of the signal, respectively, and where

\[
g(t) = \begin{cases} 
1, & 0 \leq t < P/2 \\
-1, & P/2 \leq t < P \\
0, & \text{otherwise},
\end{cases}
\]

and by convolving the signal with a Poisson hemodynamic response function (HRF),

\[
HRF_{\text{Poisson}}(t) = \frac{\lambda^t \exp(-\lambda)}{t!}, \quad t = 0, 1, 2, \ldots,
\]

with \( \lambda = 4 \). The convolved signal was then embedded with the error components, a Gaussian \( I(d) \) error and a white noise. Figure 1 illustrates this process: Plot (a) shows the simulated square wave signal with \( N = 256 \), plot (b) the Poisson HRF with \( \lambda = 4 \). In plot (c) the square wave signal is convolved with the Poisson HRF. Finally, plot (d) shows the simulated fMRI signal with additive long memory and white noise errors.

We considered different values of the amplitude of the square wave signal and of the error variance parameters \( \sigma_L^2 \) and \( \sigma_W^2 \). Here we present results for two different sets of values obtained by imposing the signal-to-noise ratio (SNR),

\[
SNR = 10 \log_{10} \left( \frac{A}{\sigma_L^2 + \sigma_W^2} \right),
\]
to be 0.5 and 2.5, respectively. We also looked at the performances of the method for different sample sizes, \( N = 2^6, 2^7, 2^8 \), different values of the long memory parameter, \( d = 0.05, 0.15, 0.25, 0.35, 0.45 \), and different regression coefficient values, \( \beta = 0.5, 1, 1.5 \). We applied discrete wavelet transforms with Daubechies minimum phase wavelets with 4 vanishing moments. In our previous work we found that wavelets with high degrees of regularity produce slightly better estimates of the long memory parameter for large sample sizes, Ko and Vannucci (2006). Wavelets with higher numbers of vanishing moments ensure wavelet coefficients 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 paper, only the wavelet coefficients were used in the EM estimation procedure, that is, scaling coefficients were removed.

We report results in terms of biases and root mean squares errors (RMSE) of the estimates computed based on fifty replications for each combination of the parameters and of the sample size. Tables 1, 2 and 3 summarize results for the case SNR=0.5 and \( \beta = 0.5, 1, 1.5 \), respectively. Tables 4, 5 and 6 show results for the case SNR=2.5. Biases and root mean squares errors of the estimates of \( \beta \) decrease consistently as the sample size increases, for both SNR scenarios and for fixed \( d \). Overall we notice that the RMSE’s of \( \hat{\beta} \) are smaller for the larger SNR value, which means that the variability of the estimates of \( \beta \) decreases as the amplitude of the signals increases. The estimates of \( d \) and \( \sigma^2_L \), which govern the long-range dependent behavior of the signal, have negative biases in almost all cases and for both SNRs, indicating a slight underestimation. This phenomenon is also pointed out in Jensen (2000) when considering the model with long memory data contaminated by white noise. Biases and RMSE’s of \( \hat{d} \) decrease as the sample size increases. For fixed \( N \), in most cases the RMSE’s of \( \hat{d} \) tend to increase as the value of \( d \) approaches its boundary region (0 or 0.5). This behavior of \( \hat{d} \) was also noticed by Jensen (2000) and by Cheung and Diebold (1994). Finally, we notice that the RMSE’s of \( d \) and \( \hat{\sigma}^2_L \) are bigger for larger SNR’s. The larger the SNR the smaller the RMSE of \( \hat{\beta} \) and the larger the RMSE of \( \hat{d} \) and \( \hat{\sigma}^2_L \).
4.2 Application to real fMRI data

We exemplify our method on the epoch dataset provided by the Wellcome Department of Imaging Neuroscience (http://www.fil.ion.ucl.ac.uk/spm/data/). The dataset was originally collected by Büchel and Friston (1997). We provide some description on how the dataset was acquired and refer readers to the Büchel and Friston paper for more details.

**Image Acquisition:** The experiment, consisting of four runs, each lasting 5 min 22 s, was performed on a single subject using a 2 Telsa Magnetom VISION whole body MRI system (Siemens, Erlangen) equipped with a head volume coil: TE=40 ms, TR=3.22 s, and 64 × 64 imaging matrix [19.2 × 19.2 cm]. 400 $T_2^*$-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 ten scans in each run were discarded in order to eliminate magnetic saturation effects. Thus 360 image volumes were used for our wavelet based EM estimation.

**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, towards the border of the screen where they vanished. Before scanning, five 30 s 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 towards 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.

**Results:** We applied our wavelet-based EM method 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 $\lambda = 4$. The resulting signal was used to form the covariate $X_i$ in our model (12). The analysis was carried out in MATLAB 7.1. The estimated parameters were mapped on transverse image using MRIcro 1.39 by Chris Rorden.

Neuroimaging studies have shown that a stimulus with visual motion, like the rapidly moving dots in this experiment, activates the primary visual cortex (V1), the motion-selective cortical area (V5), and the posterior parietal cortex (PP) (Bushnell et al., 1981; Mountcastle et al., 1981; Treue and Maunsell, 1996). Figure 2 and 3 show the activations detected by mapping the estimates of the model parameters on a transverse view. Brighter colors on the images denote higher estimated values of the corresponding parameter, that is stronger activations (or cerebral responses) to the given visual stimulus. The rectangles in Figure 2 represent the spatial coordinates corresponding to the V1, V5, and PP areas of the subject, which were identified by Büchel and Friston (1997). As expected, these areas are activated by the given ‘attentional’ visual stimulus. Figure 2 shows the activation maps obtained by the estimates of the trend parameter $\beta$. The top transverse view shows the estimated activations on the primary visual cortex (V1), the middle transverse view shows the activations on the motion-selective cortical area (V5), and the bottom transverse view shows the estimated activations on the posterior parietal cortex (PP). Our wavelet-based EM method correctly detects activations in the designated area of the posterior parietal cortex, with activations that occur quite broadly around the corresponding coordinates of V1 and V5.

In addition to the estimate of $\beta$, our method provides estimates of the long memory parameter, $d$, the innovation variance, $\sigma^2_L$, of the long memory error and the white noise variance, $\sigma^2_W$, at each voxel of the image. Figure 3 shows the activations obtained by mapping the estimates of $d$, $\sigma^2_L$, and $\sigma^2_W$. The top transverse view shows the activations estimated by the long memory parameter $d$. We see that the fMRI data are indeed contaminated by long memory errors and that some of the cerebral responses have large values of $d$. The middle transverse view denotes the activations obtained by mapping the $\hat{\sigma}^2_L$ estimates
and demonstrates that the innovation variance of the long memory error is not uniformly distributed over the brain but that it has relatively large values in very small parts around the center area. The bottom transverse view is the activated image obtained by mapping the estimates of the white noise variance $\sigma^2_w$. This shows that it has large values 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 hypothesised too. It is still an open question whether fMRI noise may indeed reveal information about the brain activity, such as long range dependencies in the cerebral response to stimuli.

5 Conclusions

In this paper we have proposed a wavelet-based method to estimate the parameters in a regression model with two error components, a long memory error and a white noise. 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 designed an EM algorithm for the estimation of the parameters. The proposed method is very flexible and includes, as simpler cases, some models previously studied by other authors. For example, with a slight change of our algorithm we could analyze the model without the trend term, which is the model treated by Wornell and Oppenheim (1992) and by Jensen (2000), or the model without the white noise component, which is studied by Fadili and Bullmore (2002). We have evaluated performances of our method on simulated data and also showed an application to real fMRI data. Our results confirm that the wavelet-based EM method we propose is very suitable for applications to fMRI data and that it also provides additional information when compared to existing methods.

Acknowledgments

15
This work was partially supported by the National Science Foundation.

**Appendix : Derivations of the EM steps**

We show here how to derive the E-step and M-step of the algorithm used in this paper. Let \( \hat{\Theta} \) be an estimate of \( \Theta \) at a given iteration. Since the joint probability distribution \( p(y_w, v_w) \) is easily obtained as the product of the conditional distribution \( p(y_w|v_w) \) and the marginal distribution of \( p(v_w) \), the complete-data log-likelihood, \( Q(\Theta|\hat{\Theta}) \), for the E-step is as follows:

\[
Q(\Theta|\hat{\Theta}) = E[ \log \; p(y_w, v_w; \Theta)|y_w, X_w, \hat{\Theta}] \\
= E[ \log \; p(y_w|v_w; \Theta) + \log \; p(v_w; \Theta)|y_w, X_w, \hat{\Theta}] \\
= E[ \log \; p(v_w|y_w; \Theta)|y_w, X_w, \hat{\Theta}] + E[ \log \; p(v_w; \Theta)|y_w, X_w, \hat{\Theta}]. \tag{30}
\]

The first expectation on the right-hand side of (30) can be written as

\[
E[ \log \; p(y_w|v_w; \Theta)|y_w, X_w, \hat{\Theta}] \\
= -\frac{1}{2} \left[ N' \log(2\pi) + \log |\Sigma_w| + (y_w - X_w\beta)'\Sigma_w^{-1}(y_w - X_w\beta) \right] \\
- \frac{1}{2} \left[ \text{tr}(\Sigma_w A(\hat{\Theta})) + B(\hat{\Theta}')\Sigma_w B(\hat{\Theta}) - 2B(\hat{\Theta})'\Sigma_w (y_w - X_w\beta) \right], \tag{31}
\]

where

\[
A(\Theta) = (\Sigma_v + \Sigma_w)^{-1} \tag{32}
\]

and

\[
B(\Theta) = (\Sigma_v + \Sigma_w)^{-1}\Sigma_w (y_w - X_w\beta). \tag{33}
\]

The second expectation on the right-hand side of (30) becomes

\[
E[ \log \; p(v_w; \Theta)|y_w, X_w, \hat{\Theta}] \\
= -\frac{1}{2} \left[ N' \log(2\pi) + \log |\Sigma_v| + \text{tr}(\Sigma_v A(\hat{\Theta})) + B(\hat{\Theta})'\Sigma_v B(\hat{\Theta}) \right]. \tag{34}
\]

Note that \( \Sigma_w = \sigma^2 w I \) where \( I \) is the \( (N' \times N') \) identity matrix. Let \( \Sigma_v = \sigma^2 \Sigma_{\eta} \) such that the
diagonal elements of $\Sigma_\eta$ consist of $\eta^{-m}$, $m = 1, 2, \ldots, r$. We now have

\[
Q(\Theta|\tilde{\Theta}) = -\frac{1}{2} \left[ N' \log(2\pi) + N' \log \sigma_w^2 + \sigma_w^{-2} (y_w - X_w/\beta)' (y_w - X_w/\beta) \right] \\
- \frac{1}{2} \left\{ \left[ \sigma_w^{-2} \text{tr}(A(\tilde{\Theta})) + \sigma_w^{-2} B(\tilde{\Theta})B(\tilde{\Theta})' \right] - 2 \sigma_w^{-2} B(\tilde{\Theta})' (y_w - X_w/\beta) \right\} \\
- \frac{1}{2} \left\{ N' [\log(2\pi) + \log \sigma^2] + \log |\Sigma_\eta| + \sigma^{-2} [\text{tr}(\Sigma_\eta^{-1} A(\tilde{\Theta})) + B(\tilde{\Theta})' \Sigma_\eta^{-1} B(\tilde{\Theta})] \right\}. \tag{35}
\]

We update $\tilde{\Theta} = (\tilde{\beta}, \tilde{\sigma}^2, \tilde{\eta}, \tilde{\sigma}_w^2)$ from the previous iteration to $\hat{\Theta} = (\hat{\beta}, \hat{\sigma}^2, \hat{\eta}, \hat{\sigma}_w^2)$ via the M-step with (35). Notice that the first two terms on the right-hand side of (35) are only related to $\beta$ and $\sigma_w^2$ and that the last term is related to $\sigma^2$ and $\eta$. We first differentiate $Q(\Theta|\tilde{\Theta})$ with respect to $\beta$ and have

\[
\hat{\beta} = (X_w'X_w)^{-1} \left[ X_w'Y_w - B(\tilde{\Theta})'X_w \right]. \tag{36}
\]

Differentiating (35) with respect to $\sigma_w^2$ gives

\[
\hat{\sigma}_w^2 = \frac{(y_w - X_w\hat{\beta})' (y_w - X_w\hat{\beta}) + \text{tr}(A(\tilde{\Theta})) + B(\tilde{\Theta})' [B(\tilde{\Theta}) - 2(y_w - X_w\hat{\beta})]}{N'}. \tag{37}
\]

As for the estimates of $\sigma^2$ and $\eta$, the last term of (35) can be written as

\[
S = N' \log(2\pi) + N' \log \sigma^2 - \sum_{m=1}^{r} mN(m) \log \eta + \sigma^{-2} \sum_{m=1}^{r} \eta^n \sum_{n=1}^{N/2m} [A^n_m(\tilde{\Theta}) + B^n_m(\tilde{\Theta})^2], \tag{38}
\]

where $A^n_m(\tilde{\Theta})$ is the diagonal element of the matrix $A(\tilde{\Theta})$ corresponding to the $n$th location at level $m$ and $B^n_m(\tilde{\Theta})$ is the diagonal element of the matrix $B(\tilde{\Theta})$ corresponding to the $n$th location at level $m$. We have the following two normal equations with respect to $\sigma^2$ and $\eta$,

\[
\sigma^2 N' = \sum_{m=1}^{r} \eta^m K_m(\tilde{\Theta}) \tag{39}
\]

and

\[
\sum_{m=1}^{r} mN(m) = \sigma^{-2} \sum_{m=1}^{r} m\eta^m K_m(\tilde{\Theta}), \tag{40}
\]

where $K_m(\tilde{\Theta}) = \sum_{n=1}^{N/2m} [A^n_m(\tilde{\Theta}) + B^n_m(\tilde{\Theta})^2]$. Eliminating $\sigma^2$ from (39) and (40) gives

\[
\sum_{m=1}^{r} T_m K_m(\tilde{\Theta}) \eta^m = 0, \tag{41}
\]

17
where $T_m = \left[\sum_{m=1}^{r} mN(m) - mN'\right]$. The polynomial equation (41) of $\eta$ has a unique positive real solution $\hat{\eta}$ if $r \geq 2$ and not all $K_m$ are zero (Wornell and Oppenheim, 1992). Back-substituting the $\hat{\eta}$ from (41) into the eliminated variable gives

$$
\hat{\sigma}^2 = \frac{\sum_{m=1}^{r} K_m(\hat{\Theta})\hat{\eta}^m}{N'}.
$$

(42)

Now $\hat{\Theta} = (\hat{\beta}, \hat{\sigma}^2, \hat{\eta}, \hat{\sigma}_{\omega}^2)$ in (35) is updated with $\hat{\Theta} = (\hat{\beta}, \hat{\sigma}^2, \hat{\eta}, \hat{\sigma}_{\omega}^2)$. This completes the derivation of our EM steps.
| $d$ | $n$ | $\hat{\beta}$ & RMSE & BIAS | $\hat{\delta}_d$ & RMSE & BIAS | $\hat{\sigma}_{\|}^2$ & RMSE & BIAS | $\hat{\sigma}_W$ & RMSE & BIAS |
|-----|-----|------------|--------|--------|----------------|--------|--------|----------------|--------|--------|
| 0.05 | $2^6$ | 0.254 & 0.235 & -0.011 & 0.235 & -0.046 & 0.189 & -0.104 & 0.205 & -0.030 |
| 0.15 | $2^6$ | 0.308 & 0.260 & -0.018 & 0.260 & -0.086 & 0.214 & -0.053 & 0.231 & -0.015 |
| 0.25 | $2^6$ | 0.285 & 0.348 & -0.007 & 0.348 & -0.194 & 0.260 & -0.066 & 0.203 & -0.067 |
| 0.35 | $2^6$ | 0.311 & 0.573 & 0.064 & 0.573 & -0.089 & 0.266 & -0.007 & 0.215 & -0.033 |
| 0.45 | $2^6$ | 0.311 & 0.350 & 0.038 & 0.350 & -0.227 & 0.268 & 0.039 & 0.249 & -0.073 |
| 0.05 | $2^7$ | 0.191 & 0.145 & 0.011 & 0.145 & 0.037 & 0.137 & -0.067 & 0.222 & 0.149 |
| 0.15 | $2^7$ | 0.193 & 0.161 & -0.032 & 0.161 & -0.033 & 0.145 & 0.079 & 0.166 & 0.077 |
| 0.25 | $2^7$ | 0.157 & 0.147 & -0.014 & 0.147 & -0.061 & 0.109 & -0.037 & 0.184 & 0.101 |
| 0.35 | $2^7$ | 0.233 & 0.186 & 0.014 & 0.186 & -0.031 & 0.177 & -0.012 & 0.203 & 0.049 |
| 0.45 | $2^7$ | 0.247 & 0.477 & 0.046 & 0.477 & 0.035 & 0.225 & 0.041 & 0.194 & 0.018 |
| 0.05 | $2^8$ | 0.117 & 0.106 & -0.007 & 0.106 & 0.029 & 0.161 & -0.148 & 0.195 & 0.160 |
| 0.15 | $2^8$ | 0.136 & 0.096 & 0.041 & 0.096 & -0.008 & 0.142 & -0.126 & 0.205 & 0.154 |
| 0.25 | $2^8$ | 0.159 & 0.008 & 0.027 & 0.008 & -0.015 & 0.104 & -0.080 & 0.192 & 0.128 |
| 0.35 | $2^8$ | 0.132 & 0.088 & -0.002 & 0.088 & -0.030 & 0.083 & -0.045 & 0.132 & 0.073 |
| 0.45 | $2^8$ | 0.167 & 0.083 & 0.036 & 0.083 & -0.013 & 0.088 & 0.016 & 0.134 & 0.071 |

Table 1: Bias and root mean square error (RMSE) of the estimated parameters when $SNR = 0.5$ and $\beta = 0.5$ for different sample sizes and different values of the long memory parameter $d$. 

19
<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$d$</td>
<td>$n$</td>
<td>$\hat{\beta}$</td>
<td></td>
<td>$\hat{d}$</td>
<td></td>
<td>$\hat{\sigma}_L^2$</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^6$</td>
<td>0.269</td>
<td>0.026</td>
<td>0.248</td>
<td>-0.106</td>
<td>0.240</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^6$</td>
<td>0.270</td>
<td>0.038</td>
<td>0.295</td>
<td>-0.116</td>
<td>0.255</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^6$</td>
<td>0.284</td>
<td>0.028</td>
<td>0.356</td>
<td>-0.100</td>
<td>0.225</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^6$</td>
<td>0.268</td>
<td>0.045</td>
<td>0.641</td>
<td>-0.093</td>
<td>0.297</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^6$</td>
<td>0.269</td>
<td>0.020</td>
<td>0.369</td>
<td>-0.168</td>
<td>0.275</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.197</td>
<td>0.018</td>
<td>0.117</td>
<td>-0.024</td>
<td>0.207</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.185</td>
<td>0.009</td>
<td>0.164</td>
<td>-0.082</td>
<td>0.161</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.198</td>
<td>0.008</td>
<td>0.151</td>
<td>-0.060</td>
<td>0.144</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^7$</td>
<td>0.217</td>
<td>0.038</td>
<td>0.153</td>
<td>-0.039</td>
<td>0.121</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^7$</td>
<td>0.185</td>
<td>-0.023</td>
<td>0.186</td>
<td>-0.058</td>
<td>0.2169</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.136</td>
<td>-0.003</td>
<td>0.131</td>
<td>-0.017</td>
<td>0.195</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.122</td>
<td>0.013</td>
<td>0.092</td>
<td>-0.014</td>
<td>0.160</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.139</td>
<td>-0.007</td>
<td>0.095</td>
<td>-0.019</td>
<td>0.127</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^8$</td>
<td>0.158</td>
<td>-0.005</td>
<td>0.082</td>
<td>0.019</td>
<td>0.082</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^8$</td>
<td>0.151</td>
<td>0.001</td>
<td>0.083</td>
<td>-0.010</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Table 2: Bias and root mean square error (RMSE) of the estimated parameters when $SNR = 0.5$ and $\beta = 1$ for different sample sizes and different values of the long memory parameter $d$. 
<table>
<thead>
<tr>
<th>$d$</th>
<th>$n$</th>
<th>$\hat{\beta}$</th>
<th>$\hat{d}$</th>
<th>$\hat{\sigma}_L^2$</th>
<th>$\hat{\sigma}_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RMSE</td>
<td>BIAS</td>
<td>RMSE</td>
<td>BIAS</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^6$</td>
<td>0.281</td>
<td>-0.019</td>
<td>0.246</td>
<td>-0.062</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^6$</td>
<td>0.276</td>
<td>-0.091</td>
<td>0.442</td>
<td>-0.087</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^6$</td>
<td>0.258</td>
<td>-0.019</td>
<td>0.227</td>
<td>-0.105</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^6$</td>
<td>0.303</td>
<td>-0.013</td>
<td>0.570</td>
<td>-0.006</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^6$</td>
<td>0.298</td>
<td>0.025</td>
<td>0.310</td>
<td>-0.008</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.194</td>
<td>0.002</td>
<td>0.180</td>
<td>-0.040</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.203</td>
<td>-0.022</td>
<td>0.184</td>
<td>-0.020</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.201</td>
<td>0.032</td>
<td>0.179</td>
<td>-0.091</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^7$</td>
<td>0.178</td>
<td>0.013</td>
<td>0.282</td>
<td>-0.049</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^7$</td>
<td>0.240</td>
<td>0.010</td>
<td>0.160</td>
<td>-0.060</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.130</td>
<td>-0.014</td>
<td>0.110</td>
<td>0.006</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.134</td>
<td>-0.013</td>
<td>0.101</td>
<td>-0.025</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.152</td>
<td>0.027</td>
<td>0.085</td>
<td>0.000</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^8$</td>
<td>0.128</td>
<td>0.049</td>
<td>0.083</td>
<td>0.001</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^8$</td>
<td>0.152</td>
<td>0.014</td>
<td>0.079</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Table 3: Bias and root mean square error (RMSE) of the estimated parameters when $SNR = 0.5$ and $\beta = 1.5$ for different sample sizes and different values of the long memory parameter $d$. 
Table 4: Bias and root mean square error (RMSE) of the estimated parameters when $\text{SNR} = 2.5$ and $\beta = 0.5$ for different sample sizes and different values of the long memory parameter $d$. 

<table>
<thead>
<tr>
<th>$d$</th>
<th>$n$</th>
<th>$\hat{\beta}$</th>
<th>BIAS</th>
<th>$\hat{d}$</th>
<th>RMSE</th>
<th>BIAS</th>
<th>$\hat{\sigma}_L^2$</th>
<th>RMSE</th>
<th>BIAS</th>
<th>$\hat{\sigma}_W$</th>
<th>RMSE</th>
<th>BIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>$2^6$</td>
<td>0.189</td>
<td>0.035</td>
<td>0.314</td>
<td>0.036</td>
<td>0.266</td>
<td>-0.073</td>
<td>0.232</td>
<td>-0.078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>$2^6$</td>
<td>0.208</td>
<td>0.008</td>
<td>0.294</td>
<td>-0.102</td>
<td>0.244</td>
<td>-0.094</td>
<td>0.238</td>
<td>-0.088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>$2^6$</td>
<td>0.246</td>
<td>0.018</td>
<td>0.344</td>
<td>-0.145</td>
<td>0.238</td>
<td>-0.028</td>
<td>0.191</td>
<td>-0.064</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>$2^6$</td>
<td>0.237</td>
<td>-0.011</td>
<td>0.319</td>
<td>-0.098</td>
<td>0.342</td>
<td>0.070</td>
<td>0.253</td>
<td>-0.101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>$2^6$</td>
<td>0.241</td>
<td>0.072</td>
<td>0.313</td>
<td>-0.182</td>
<td>0.294</td>
<td>0.062</td>
<td>0.228</td>
<td>-0.096</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.151</td>
<td>0.014</td>
<td>0.185</td>
<td>-0.059</td>
<td>0.170</td>
<td>-0.124</td>
<td>0.173</td>
<td>0.085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.167</td>
<td>-0.009</td>
<td>0.147</td>
<td>-0.024</td>
<td>0.140</td>
<td>-0.083</td>
<td>0.156</td>
<td>0.051</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.132</td>
<td>-0.027</td>
<td>0.170</td>
<td>-0.065</td>
<td>0.161</td>
<td>-0.055</td>
<td>0.163</td>
<td>0.069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>$2^7$</td>
<td>0.161</td>
<td>0.011</td>
<td>0.161</td>
<td>-0.093</td>
<td>0.130</td>
<td>0.005</td>
<td>0.179</td>
<td>0.049</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>$2^7$</td>
<td>0.229</td>
<td>0.045</td>
<td>0.140</td>
<td>-0.045</td>
<td>0.157</td>
<td>0.060</td>
<td>0.202</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.095</td>
<td>-0.013</td>
<td>0.095</td>
<td>0.016</td>
<td>0.157</td>
<td>-0.136</td>
<td>0.120</td>
<td>0.135</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.103</td>
<td>0.013</td>
<td>0.088</td>
<td>0.002</td>
<td>0.134</td>
<td>-0.108</td>
<td>0.173</td>
<td>0.101</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.098</td>
<td>-0.008</td>
<td>0.087</td>
<td>-0.015</td>
<td>0.111</td>
<td>-0.080</td>
<td>0.124</td>
<td>0.056</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>$2^8$</td>
<td>0.085</td>
<td>0.010</td>
<td>0.085</td>
<td>-0.035</td>
<td>0.074</td>
<td>-0.034</td>
<td>0.116</td>
<td>0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>$2^8$</td>
<td>0.155</td>
<td>0.029</td>
<td>0.094</td>
<td>-0.044</td>
<td>0.085</td>
<td>0.037</td>
<td>0.121</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td>$n$</td>
<td>$\hat{\beta}$ RMSE</td>
<td>$\hat{\beta}$ BIAS</td>
<td>$\hat{d}$ RMSE</td>
<td>$\hat{d}$ BIAS</td>
<td>$\hat{\sigma}_L^2$ RMSE</td>
<td>$\hat{\sigma}_L^2$ BIAS</td>
<td>$\hat{\sigma}_W$ RMSE</td>
<td>$\hat{\sigma}_W$ BIAS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>---------------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>------------------</td>
<td>----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>$2^6$</td>
<td>0.188 0.284</td>
<td>-0.026 -0.111</td>
<td>0.240 0.302</td>
<td>-0.055 -0.175</td>
<td>0.204 0.231</td>
<td>-0.098 -0.056</td>
<td>0.208 0.195</td>
<td>-0.025 -0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>$2^6$</td>
<td>0.218 0.277</td>
<td>-0.032 -0.111</td>
<td>0.209 0.302</td>
<td>-0.111 -0.175</td>
<td>0.223 0.231</td>
<td>-0.063 -0.056</td>
<td>0.208 0.195</td>
<td>-0.023 -0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>$2^6$</td>
<td>0.237 0.302</td>
<td>-0.040 -0.175</td>
<td>0.231 0.302</td>
<td>-0.056 -0.175</td>
<td>0.195 0.231</td>
<td>-0.002 -0.002</td>
<td>0.208 0.195</td>
<td>-0.023 -0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>$2^6$</td>
<td>0.256 0.371</td>
<td>0.009 -0.205</td>
<td>0.211 0.371</td>
<td>-0.060 -0.205</td>
<td>0.208 0.231</td>
<td>-0.100 -0.100</td>
<td>0.208 0.195</td>
<td>-0.100 -0.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>$2^6$</td>
<td>0.228 0.429</td>
<td>0.060 -0.019</td>
<td>0.3447 0.429</td>
<td>0.029 -0.019</td>
<td>0.271 0.3447</td>
<td>-0.043 -0.043</td>
<td>0.208 0.231</td>
<td>-0.100 -0.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.147 0.192</td>
<td>0.004 -0.040</td>
<td>0.200 0.192</td>
<td>-0.150 -0.040</td>
<td>0.177 0.200</td>
<td>-0.071 -0.071</td>
<td>0.208 0.195</td>
<td>-0.071 -0.071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.149 0.155</td>
<td>0.019 -0.037</td>
<td>0.159 0.155</td>
<td>-0.108 -0.037</td>
<td>0.150 0.159</td>
<td>-0.071 -0.071</td>
<td>0.208 0.195</td>
<td>-0.071 -0.071</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.139 0.148</td>
<td>0.013 -0.044</td>
<td>0.131 0.148</td>
<td>-0.049 -0.044</td>
<td>0.160 0.131</td>
<td>-0.070 -0.070</td>
<td>0.208 0.195</td>
<td>-0.070 -0.070</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>$2^7$</td>
<td>0.194 0.155</td>
<td>0.039 -0.056</td>
<td>0.136 0.155</td>
<td>-0.029 -0.056</td>
<td>0.175 0.136</td>
<td>0.067 0.067</td>
<td>0.208 0.195</td>
<td>0.067 0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>$2^7$</td>
<td>0.165 0.239</td>
<td>0.025 -0.031</td>
<td>0.244 0.239</td>
<td>-0.049 -0.031</td>
<td>0.217 0.244</td>
<td>0.059 0.059</td>
<td>0.208 0.195</td>
<td>0.059 0.059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.102 0.117</td>
<td>0.024 0.007</td>
<td>0.189 0.117</td>
<td>-0.176 0.007</td>
<td>0.199 0.189</td>
<td>0.163 0.163</td>
<td>0.208 0.195</td>
<td>0.163 0.163</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.117 0.086</td>
<td>0.025 -0.009</td>
<td>0.149 0.086</td>
<td>-0.135 -0.009</td>
<td>0.197 0.149</td>
<td>0.169 0.169</td>
<td>0.208 0.195</td>
<td>0.169 0.169</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.121 0.010</td>
<td>0.007 0.031</td>
<td>0.141 0.010</td>
<td>-0.099 0.031</td>
<td>0.180 0.141</td>
<td>0.121 0.121</td>
<td>0.208 0.195</td>
<td>0.121 0.121</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.35</td>
<td>$2^8$</td>
<td>0.116 0.090</td>
<td>0.009 0.008</td>
<td>0.146 0.090</td>
<td>-0.048 0.008</td>
<td>0.185 0.146</td>
<td>0.117 0.117</td>
<td>0.208 0.195</td>
<td>0.117 0.117</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45</td>
<td>$2^8$</td>
<td>0.109 0.088</td>
<td>0.008 -0.036</td>
<td>0.076 0.088</td>
<td>0.006 -0.036</td>
<td>0.179 0.076</td>
<td>0.126 0.126</td>
<td>0.208 0.195</td>
<td>0.126 0.126</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Bias and root mean square error (RMSE) of the estimated parameters when $SNR = 2.5$ and $\beta = 1$ for different sample sizes and different values of the long memory parameter $d$. 

23
<table>
<thead>
<tr>
<th>$d$</th>
<th>$n$</th>
<th>$\hat{\beta}$</th>
<th>BIAS</th>
<th>$\hat{d}$</th>
<th>BIAS</th>
<th>$\hat{\sigma}_L^2$</th>
<th>BIAS</th>
<th>$\hat{\sigma}_W$</th>
<th>BIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>$2^6$</td>
<td>0.226</td>
<td>0.002</td>
<td>0.307</td>
<td>-0.057</td>
<td>0.258</td>
<td>-0.107</td>
<td>0.257</td>
<td>0.042</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^6$</td>
<td>0.202</td>
<td>-0.003</td>
<td>0.320</td>
<td>-0.086</td>
<td>0.241</td>
<td>-0.082</td>
<td>0.234</td>
<td>0.035</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^6$</td>
<td>0.263</td>
<td>-0.027</td>
<td>0.750</td>
<td>-0.002</td>
<td>0.298</td>
<td>-0.076</td>
<td>0.234</td>
<td>0.037</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^6$</td>
<td>0.275</td>
<td>0.006</td>
<td>0.577</td>
<td>-0.034</td>
<td>0.230</td>
<td>0.015</td>
<td>0.241</td>
<td>0.073</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^6$</td>
<td>0.257</td>
<td>0.018</td>
<td>0.426</td>
<td>-0.130</td>
<td>0.303</td>
<td>-0.011</td>
<td>0.262</td>
<td>0.023</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^7$</td>
<td>0.164</td>
<td>-0.038</td>
<td>0.164</td>
<td>-0.018</td>
<td>0.212</td>
<td>-0.180</td>
<td>0.187</td>
<td>0.079</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^7$</td>
<td>0.130</td>
<td>-0.006</td>
<td>0.172</td>
<td>-0.050</td>
<td>0.174</td>
<td>-0.126</td>
<td>0.226</td>
<td>0.119</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^7$</td>
<td>0.145</td>
<td>-0.002</td>
<td>0.160</td>
<td>-0.057</td>
<td>0.160</td>
<td>-0.055</td>
<td>0.256</td>
<td>0.129</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^7$</td>
<td>0.182</td>
<td>-0.008</td>
<td>0.236</td>
<td>-0.054</td>
<td>0.226</td>
<td>-0.073</td>
<td>0.224</td>
<td>0.070</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^7$</td>
<td>0.164</td>
<td>0.008</td>
<td>0.543</td>
<td>0.046</td>
<td>0.262</td>
<td>-0.007</td>
<td>0.309</td>
<td>0.161</td>
</tr>
<tr>
<td>0.05</td>
<td>$2^8$</td>
<td>0.093</td>
<td>0.023</td>
<td>0.118</td>
<td>-0.008</td>
<td>0.213</td>
<td>-0.203</td>
<td>0.256</td>
<td>0.222</td>
</tr>
<tr>
<td>0.15</td>
<td>$2^8$</td>
<td>0.084</td>
<td>-0.017</td>
<td>0.088</td>
<td>-0.006</td>
<td>0.171</td>
<td>-0.164</td>
<td>0.234</td>
<td>0.201</td>
</tr>
<tr>
<td>0.25</td>
<td>$2^8$</td>
<td>0.135</td>
<td>0.013</td>
<td>0.083</td>
<td>0.003</td>
<td>0.137</td>
<td>-0.116</td>
<td>0.218</td>
<td>0.163</td>
</tr>
<tr>
<td>0.35</td>
<td>$2^8$</td>
<td>0.118</td>
<td>0.000</td>
<td>0.087</td>
<td>-0.029</td>
<td>0.113</td>
<td>-0.077</td>
<td>0.187</td>
<td>0.139</td>
</tr>
<tr>
<td>0.45</td>
<td>$2^8$</td>
<td>0.145</td>
<td>0.056</td>
<td>0.078</td>
<td>-0.026</td>
<td>0.078</td>
<td>-0.028</td>
<td>0.168</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Table 6: Bias and root mean square error (RMSE) of the estimated parameters when $SNR = 2.5$ and $\beta = 1.5$ for different sample sizes and different values of the long memory parameter $d$. 
Figure 2: Activation maps obtained by mapping the estimates of the linear regression parameter $\beta$, estimated at each voxel using our wavelet-based EM method: (Top plot) activations on the primary visual cortex (V1); (Middle plot) activations on the motion-selective cortical area (V5); (Bottom plot) activations on the posterior parietal cortex (PP). The rectangles denote the spatial coordinates of V1 (top), V5 (middle) and PP (bottom), respectively.
Figure 3: Activation maps obtained by mapping the estimates of the error-related parameters $d, \sigma_L^2$ and $\sigma_W^2$, at each voxel of the image: (Top plot) activations by $\hat{d}$; (Middle plot) activations by $\hat{\sigma}_L^2$; (Bottom plot) activations by $\hat{\sigma}_W^2$. 
References


