Wavelet-Based Estimation of a Semiparametric Generalized Linear Model of fMRI Time-Series

François G. Meyer, *Member, IEEE*

***Abstract*—This paper addresses the problem of detecting significant changes in fMRI time series that are correlated to a stimulus time course. This paper provides a new approach to estimate the parameters of a semiparametric generalized linear model of fMRI time series. The fMRI signal is described as the sum of two effects: a smooth trend and the response to the stimulus. The trend belongs to a subspace spanned by large scale wavelets. The wavelet transform provides an approximation to the Karhunen–Loève transform for the long memory noise and we have developed a scale space regression that permits to carry out the regression in the wavelet domain while omitting the scales that are contaminated by the trend. In order to demonstrate that our approach outperforms the state-of-the art detrending technique, we evaluated our method against a smoothing spline approach. Experiments with simulated data and experimental fMRI data, demonstrate that our approach can infer and remove drifts that cannot be adequately represented with splines.**

I. Introduction

FUNCTIONAL brain imaging utilizes the coupling between local electrical activity and regional changes in blood flow and blood oxygenation level. Blood oxygenation level-dependent (BOLD) fMRI uses deoxyhemoglobin as contrast agent: deoxygenated hemoglobin induces a difference in magnetic susceptibility relative to the surrounding. The cascade of physiological events that trigger the changes in the BOLD signal remains an area of active research [1], [2]. It appears that after the onset of stimulation the oxygen concentration rapidly drops in response to the local energy metabolism triggered by the stimulation; the dip [3] is then followed by a large hyperoxygenation [1]. This imbalance between oxygen metabolism and oxygen supply is at the origin of the BOLD contrast. Unfortunately, changes in the fMRI signal are only of the order of a few percents. The detection of changes in the BOLD signal is further complicated by the presence of a large number of instrumental and physiological noises that contaminate the fMRI signal [4]. Long-term physiological drifts and instrumental instability contribute to a systematic increase or decrease in the signal with time. While the exact cause for the drift of the baseline signal is not completely understood [5], this structured trend constitutes a basic hurdle to any statistical analysis of the data. Indeed, the aim of the standard analysis techniques is to estimate the correlation between the stimulus time course $x(t)$ and the fMRI time series. The simplest version of these methods is based on the linear model associated with the two-sided Student's t-test and assumes that the fMRI signal $y_{\mathbf{M}}(t)$ at a point $\mathbf{M}$ in the brain is given by

$$y_{\mathbf{M}}(t) = \beta_{\mathbf{M}} x(t) + n(t) \tag{1}$$

where $n(t)$ is a stationary Gaussian white noise. $\beta_{\mathbf{M}}$ is a scalar that measures the strength of the response at the voxel $\mathbf{M}$. It is obvious that if the baseline drift is not removed, any analysis based on the model (1) will be tracking the large variation in the signal instead of the effects of the stimulus. In order to obtain a baseline from which one can estimate the effect of the stimulus it is thus essential to infer and remove the systematic drift, or trend, in the data. Baseline drifts in fMRI data have been described by linear [6]–[8], polynomials [9], and splines. A global polynomial cannot track as well as a spline function the variation of the baseline (unless we increase the order of the polynomial, and we then run into the problem of undesired oscillations). One could, therefore, use smoothing splines [10] to model and estimate the baseline drift and we compared our method to a smoothing spline approach in the experiments. A common problem with smoothing splines is the difficulty to find the optimal value of the smoothing parameter [10]. While several techniques are available [10], they all involve an optimization problem. Traditional baseline removal techniques are merely preprocessing step. Low-pass filters, autoregressive filters, and Kalman filters have been proposed to estimate the baseline drift [11]. All these techniques estimate a drift and the estimation of $\beta_{\mathbf{M}}$ is then performed on the detrended data. Because the drift removal and the detection of activation are decoupled, one cannot gain any understanding on the effect of the detrending on the estimation of $\beta_{\mathbf{M}}$.

In this paper, we take a different point of view. We propose to include the baseline as part of our model. The model becomes a partially linear model [10]

$$y_{i,\mathbf{M}} = \theta_{i,\mathbf{M}} + \beta_{\mathbf{M}} x_i + \nu_{i,\mathbf{M}},\ i = 0, \ldots, N-1 \tag{2}$$

where $\boldsymbol{\theta}_{\mathbf{M}} = [\theta_{0,\mathbf{M}}, \ldots, \theta_{N-1,\mathbf{M}}]^t$ is a baseline drift. The noise $\nu_{i,\mathbf{M}}$ is correlated with a $1/f$ spectral behavior associated with long memory processes [12], [13]. The first contribution of this work is a new model of the drift that belongs to a subspace spanned by large scale wavelets. The complexity of the trend (determined by its scale in the wavelet domain) is estimated with an improved version of Akaike's information criterion (AIC) [14]. Furthermore, the trend is directly part of our model and we can interpret the estimation of $\beta_{\mathbf{M}}$ in terms of the time-scale properties of the data. The wavelet transform provides an approximation to the Karhunen–Loève transform for the long memory noise $\nu_{i,\mathbf{M}}$ and $\beta_{\mathbf{M}}$ can be estimated in the

Manuscript received October 25, 2001; revised September 5, 2002. This work was supported by a Whitaker Foundation Biomedical Engineering Research Grant.

The author is with the Department of Electrical Engineering, University of Colorado, Boulder, CO 80302 USA (e-mail: francois.meyer@colorado.edu.).

Digital Object Identifier 10.1109/TMI.2003.809587

wavelet domain while omitting the scales that are contaminated by the trend. The second contribution of the work is a new method to estimate the drift $\boldsymbol{\theta}_{\mathbf{M}}$ and to test for the significance of the response $\beta_{\mathbf{M}}$ to the stimulus. We note that our approach is not based on the "denoising" of the fMRI time-series. Wavelet-based denoising has been applied to fMRI data for the removal of random acquisition noise [15]. Our approach does not exploit a wavelet decomposition in the spatial domain of each MR image, as proposed in [16] and [17]. This paper is organized as follows. In Section II, we review some wavelet notations. In Section III, we describe our partially linear model. The estimation of the parameters of the model are presented in Section IV. Results of experiments are presented in Section V.

II. Some Wavelet Notations

We introduce in this section the notations associated with a discrete wavelet transform. These notations will be used in the sequel of this paper. Let $\Psi(t)$ be the wavelet and let $\Phi(t)$ be the scaling function associated with a multiresolution analysis [18]. Let $\{h_n\}$ be the lowpass filter and let $\{g_n\}$ be the highpass filter associated with this wavelet transform. Let $\mathbf{x} = \{x_n\}$, $n = 0, \ldots, N-1$ be a discrete signal. For the simplicity of the presentation we assume that $N = 2^J$. The wavelet coefficients of $\mathbf{x}$ are defined by the following recursions:

$$sx_k^0 = x_k \quad k = 0, \ldots, N-1 \tag{3}$$

$$sx_k^{j+1} = \sum_n g_{n-2k} sx_n^j \quad k = 0, \ldots, 2^{-j-1}N - 1 \tag{4}$$

$$dx_k^{j+1} = \sum_n h_{n-2k} sx_n^j, \qquad k = 0, \ldots, 2^{-j-1}N - 1. \tag{5}$$

The wavelet transform $\mathbf{W}$ at scale J is a linear operator that maps $\mathbf{x}$ to $\mathbf{Wx}$ given by

$$\mathbf{Wx} = \Big[sx_0^J, dx_0^J, dx_0^{J-1}, dx_1^{J-1}, \ldots, dx_0^j, \ldots dx_{2^{-j}N-1}^j, \ldots, dx_0^1, \ldots\ldots, dx_{2^{-1}N-1}^1\Big]^t. \tag{6}$$

We also require that the wavelet ψ have p vanishing moments. As a consequence, polynomials of degree $p-1$ will have a very sparse representation in such a wavelet basis: all the d_k^j are equal to zero, except for the coefficients located at the border of the dyadic subdivision ($k = 0, 1, 2, 4, \ldots, 2^{J-1}$).

III. Model of the fMRI Signal

As explained in Section I, work we propose to describe the fMRI signal $y_{i,\mathbf{M}}$, measured at time i inside the voxel M, with the following model:

$$y_{i,\mathbf{M}} = \theta_{i,\mathbf{M}} + \beta_{\mathbf{M}} x_i + \nu_{i,\mathbf{M}}, \; i = 0, \ldots, N-1 \tag{7}$$

where $\boldsymbol{\theta}_{\mathbf{M}} = [\theta_{0,\mathbf{M}}, \ldots, \theta_{N-1,\mathbf{M}}]^t$ is a baseline drift. In the sequel, we drop the $\mathbf{M}$ subscript to ease readability; however, one should keep in mind that our method effectively estimates a different trend for each voxel.

A. Drift Model

Our understanding of the origin of the drift [5] does not provide us with a specific function, or even a parametric model of the drift. Our assumption is that the trend is a superposition of physical and physiological phenomena that occur at different time scale, but that do not vary greatly over a short interval of time. An appropriate model for the trend is thus provided by a linear combination of large scale wavelets

$$\theta(t) = s\theta_0^J \Phi\left(2^{-J}t\right) + \sum_{j=J_0}^{J} \sum_{k=0}^{2^{-j}N-1} d\theta_k^j \psi\left(2^{-j}t - k\right). \tag{8}$$

This model assumes that all the fine scale coefficients, $d\beta_k^j$, $1 \leq j \leq J_0 - 1$ of $\boldsymbol{\beta}$ are zero. The finest scale J_0 characterizes the complexity of the trend. Because the wavelet has p vanishing moments, $\theta(t)$ can approximate polynomials of degree $p-1$ with a small error. This nonparametric model of the trend not only describes the low-frequency fluctuations in the signal (encoded by the wavelet coefficients s_k^J, $d_k^J, \ldots$), but it also describes rapid localized changes, as is shown in Fig. 6. It is important to note that a model based on a Fourier expansion of the trend would miserably fail to describe localized changes, unless almost all of the harmonics were included in the model.

B. Noise Model

It has been noticed by several authors [12], [13] that data collected under the null-hypothesis condition exhibit the $1/f$ spectrum associated with long memory processes. An example of a $1/f$ noise process is the fractional Brownian motion (fBm) [19]. By definition, a stochastic process $\nu(t)$ is a fractional Brownian motion of Hurst exponent H if it is a zero-mean Gaussian nonstationary process with an autocovariance of the form

$$r_\nu(t,s) = \frac{1}{2}\sigma^2\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \tag{9}$$

where

$$\sigma^2 = r_\nu(1,1) = \Gamma(1-2H)\frac{\cos \pi H}{\pi H} \tag{10}$$

with $0 < H < 1$. Because the process in nonstationary, one can only define an *average spectrum*. The average spectrum has the form

$$\frac{\sigma^2}{|\omega|^{2H+1}}. \tag{11}$$

Fractional Brownian motion processes have been used successfully to describe many physiological signals that have a self-similar structure [20]. After this manuscript was submitted, Fadili and Bullmore [13] (see also [21]) proposed to model the fMRI noise with fractional Brownian motion and have estimated Hurst exponents H greater than $1/2$.

We model the noise process as a fractional Brownian motion process. We note that several authors [12] have argued that the fMRI signal was also contaminated by a white noise in addition to the long memory noise. In this paper, we only consider the fractional Brownian motion noise, but we are currently investigating the addition of a white noise process to our model.

The wavelet transform of a fBm process is composed of stationary Gaussian random coefficients. The wavelet transform approximatively uncorrelate the fractional Brownian motion process [22]–[24]. As shown in [22], the correlation of the wavelet coefficients $d\nu_k^j$ and $d\nu_{k'}^{j'}$ of an fBm process decays

like $O(|2^{-j}k - 2^{-j'}k'|^{2(H-p)})$, where p is the number of vanishing moments. Consequently, if p is large enough we will assume that the coefficients $d\nu_k^j$ are uncorrelated Gaussian variables with variance σ_j^2. The variance can be computed [23] and depends on the scale j and the parameter H. In practice, the variance can be estimated at each scale j with a robust estimator such as the median absolute deviation of the coefficients at that scale [25].

IV. Maximum Likelihood Estimation

We describe here a method to estimate the vector $\boldsymbol{\theta}$ and the scalar β from a single time series $\mathbf{y}$. The estimation is done in the wavelet domain and a Student's t-test can be performed directly on the wavelet coefficients to assess the significance of β. Applying the wavelet transform $\mathbf{W}$ on both sides of (7), yields

$$\mathbf{W}\mathbf{y} = \mathbf{W}\boldsymbol{\theta} + \beta\mathbf{W}\mathbf{x} + \mathbf{W}\boldsymbol{\nu}. \tag{12}$$

As explained in Section III-B, we assume that the wavelet coefficients of the noise are uncorrelated Gaussian random variables with a covariance matrix given by

$$\boldsymbol{\Sigma} = \operatorname{diag}\left\{\sigma_J^2, \sigma_J^2, \ldots, \sigma_1^2\right\}. \tag{13}$$

Let $n_0 = 2^{-J_0+1}N$ be the number of coefficients that describe the trend. Our model of the trend is

$$\mathbf{W}\boldsymbol{\theta} = \left[s\theta_0^J, d\theta_0^J, \ldots, d\theta_{2^{-J_0}N-1}^{J_0}, 0, \ldots\ldots, 0\right]. \tag{14}$$

Equation (12) can be written as a standard regression problem

$$\mathbf{W}\mathbf{y} = \mathbf{A}\boldsymbol{\xi} + \mathbf{W}\boldsymbol{\nu} \tag{15}$$

where the $n_0 + 1$ vector of unknown parameters $\boldsymbol{\xi}$ is defined by

$$\boldsymbol{\xi} = \left[s\theta_0^J, d\theta_0^J, \ldots, d\theta_{2^{-J_0}N-1}^{J_0}, \beta\right] \tag{16}$$

and the $N \times (n_0 + 1)$ matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 1 & & & 0 & sx_0^J \\ & 1 & & 0 & dx_0^J \\ & & \backslash & | & \vdots \\ & & 1 & 0 & dx_{2^{-J_0}N-2}^{J_0} \\ & & & 1 & dx_{2^{-J_0}N-1}^{J_0} \\ & & & 0 & dx_0^{J_0-1} \\ & & & | & \vdots \\ & & & 0 & dx_{2^{-1}N-1}^{1} \end{bmatrix}. \tag{17}$$

We note that the variances of the noise at different scales are not equal. This is a case of heteroscedasticity where the ordinary least squares needs to be replaced with a weighted least squares. The maximum likelihood estimate of $\boldsymbol{\xi}$ is given by

$$\hat{\boldsymbol{\xi}} = \left[\mathbf{A}^t\boldsymbol{\Sigma}^{-1}\mathbf{A}\right]^{-1}\mathbf{A}^t\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}. \tag{18}$$

Let $\mathcal{P}_{J_0}$ the projector onto the first $2^{-J_0+1}N$ coordinates

$$(\mathcal{P}_{J_0}\mathbf{x})_i = \begin{cases} x_i, & \text{if } i = 0, \ldots, 2^{-J_0+1}N - 1 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

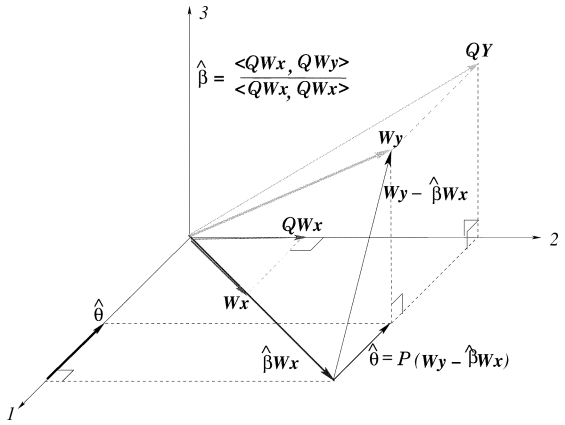

Fig. 1. Geometric interpretation of the estimation of β and $\boldsymbol{\theta}$. See the details in the text.

and let $\mathcal{Q}_{J_0}$ be the projector onto the last $N - 2^{-J_0+1}N$ coordinates ($\mathcal{P}_{J_0} + \mathcal{Q}_{J_0} = \mathbf{I}$). It can be shown that the maximum likelihood estimates are given by

$$\hat{\beta} = \frac{\left\langle \mathcal{Q}_{J_0}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{x}, \mathcal{Q}_{J_0}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{y}\right\rangle}{\left\langle \mathcal{Q}_{J_0}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{x}, \mathcal{Q}_{J_0}\boldsymbol{\Sigma}^{-1}\mathbf{W}\mathbf{x}\right\rangle} \tag{20}$$

$$\hat{\boldsymbol{\theta}} = \mathbf{W}^{-1}\mathcal{P}_{J_0}(\mathbf{W}\mathbf{y} - \hat{\beta}\mathbf{W}\mathbf{x}). \tag{21}$$

This scale space regression permits to carry out the regression in the wavelet domain while omitting the scales that are contaminated by the trend. Because $\hat{\beta}$ is a Gaussian random variable, one can test its significance with a Student t-test with $N - n_0 - 1$ degrees of freedom. The number of degrees of freedom of the Student t-test decreases as we increase the complexity n_0 of the trend. Fig. 1 provides a geometric interpretation of the estimation of β and $\boldsymbol{\theta}$. The trend $\boldsymbol{\theta}$ belongs to the subspace spanned by the axis (1). In order to estimate β, one needs to project $\mathbf{W}\mathbf{x}$ and $\mathbf{W}\mathbf{y}$ onto the plane (2,3) which is perpendicular to the axis (1). Within that subspace we can perform the regression and estimate the activation β. The trend is the projection on the axis (1) of the residual error $\mathbf{W}\mathbf{y} - \hat{\beta}\mathbf{W}\mathbf{x}$.

A. Model Selection for the Trend

The number of degrees of freedom, $n_0 = 2^{-J_0+1}N$, of the baseline drift $\boldsymbol{\theta}$ is solely determined by the scale J_0. We first notice that as we decrease the scale (a finer resolution of the trend), we increase n_0 and the number of degrees of freedom, $N - n_0 - 1$, of the Student t-test, decreases. A naive solution to the selection of J_0 consists in picking the scale that provides the best fit, where the fit can be measured by the sum of squared residuals, or the P-value. Unfortunately, if the trend contains many parameters that are irrelevant, then the variance of the estimate $\hat{\boldsymbol{\beta}}$ will be increased. Decreasing the complexity of the trend can introduce a bias in the estimation of $\boldsymbol{\beta}$, but will decrease the variance of $\hat{\boldsymbol{\beta}}$.

The task of choosing a model of adequate complexity using a small number of samples is a difficult statistical problem. Several model selection criteria have been developed [26]; they all measure the quality of the fit (using the mean squared error),

but penalize complex models. In this paper, we use an improved version of AIC, defined by Hurvich and Tsai [14] and given by

$$\mathrm{AIC}_C = \log \hat{\sigma}^2 + \frac{N + n_0}{N - n_0 - 2} \tag{22}$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2 given by

$$\hat{\sigma}^2 = \frac{1}{N} \left\| \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \right\|. \tag{23}$$

The original AIC criterion leads to overfitting when working with small samples. We also computed a second selection criterion, the Schwartz information criterion (SIC) [27], given by

$$\mathrm{SIC} = \log \hat{\sigma}^2 + \frac{n_0 \log n_0}{N}. \tag{24}$$

This criterion results in a much stronger penalty for overfitting (large n_0). We have noticed that the two criteria select the same model most of the time. On some rare occasions, SIC preferred a simpler model than AIC_C, at the expense of increasing the P value.

V. Experiments

A major difficulty with fMRI research concerns the validation of any analysis of experimental *in vivo* data. For this reason, our method was first evaluated using experimental fMRI trends with blended simulated hemodynamic responses. We test the sensitivity of the method by adding progressively a controlled amount of noise. The method was then applied to experimental fMRI data. In order to demonstrate that our approach outperforms the state-of-the art detrending technique we evaluated our method against a smoothing spline approach. As explained in Section I, a smoothing spline provides a very powerful method to estimate the baseline drift. Splines have been used to estimate the trend component of time series [28]. A cubic smoothing spline was fitted to the fMRI data using the MATLAB spline toolbox. The smoothing parameter was optimized using an oracle as explained in the next section. Our algorithms were implemented in C on a Linux platform. In all experiments, we used the 9–7 biorthogonal filters [29] with $p = 4$ vanishing moments. These filters are very close to being orthogonal and because they are symmetric, they have linear phase. They also offer the advantage of not requiring a special treatment at the boundaries of the time interval.

A. Simulated Data

We first report results of experiments on simulated data. The simulated fMRI time series $\mathbf{y} = [y_0, \ldots, y_{N-1}]^t$ were constructed using the following equation:

$$y_i = (h^* x)_i + \theta_i + \nu_i \quad i = 0, \ldots, N-1 \tag{25}$$

where h is a model of the hemodynamic response and $\mathbf{x} = [x_0, \ldots, x_{N-1}]^t$ is the stimulus time course. We used the following model of the hemodynamic response provided by Glover in [9]

$$h(t) = c_1 t^5 e^{-t/t1} - 0.4 c_2 t^1 2 e^{-t/t2} \tag{26}$$

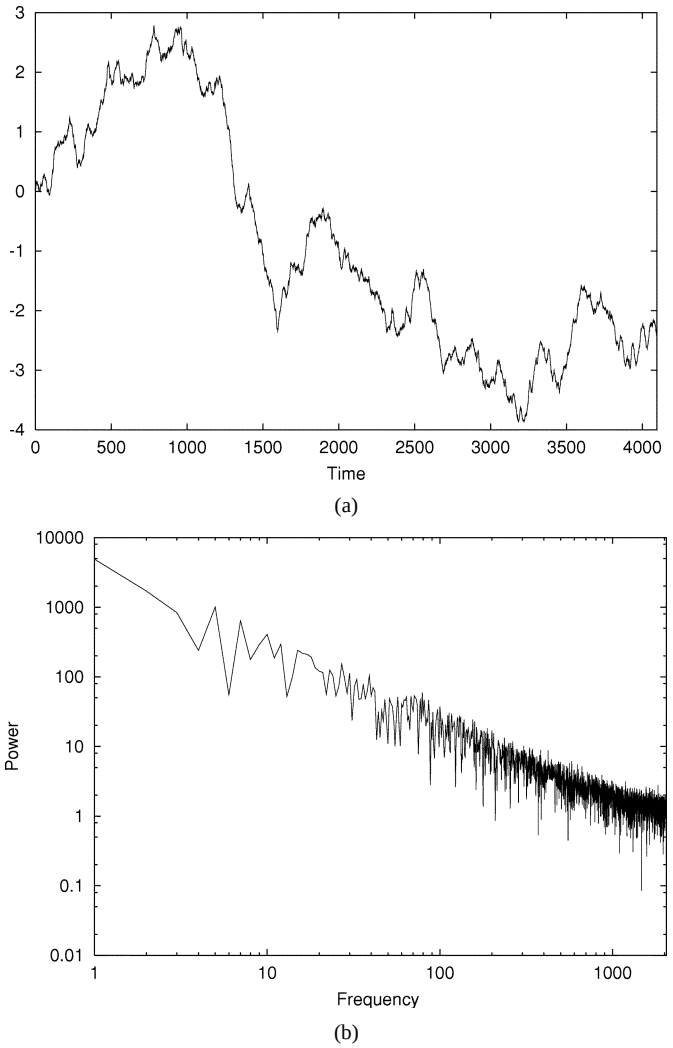

Fig. 2. (a) Realization of a synthetic fBm ($H = 0.8$). (b) Magnitude of the Fourier transform of the time-series on the top in log-scale log-power coordinates.

where $c_1 = 21.056\,082\,2$, $t_1 = 1.1$ s, $c_2 = 5.478\,24e7$ and $t_2 = 0.9$ s. We assumed that images where acquired every 1.5 s (i.e., the sampling period between i and $i+1$ was 1.5 s). Each stimulus time series $\mathbf{x}$ was a boxcar signal composed of 28 alternating segments of 1 and -1. Each segment of 1, or -1, lasted for 6 s. During the first 12 s, there was no stimulus.

The noise ν_i was a synthetic fBm generated with the wavelet-based approach described in [30] (see also [31]). This new construction reproduces the theoretical properties of fBm and makes it possible to control the variance of the noise process. Fig. 2 shows a realization (with 4096 times samples) of the synthetic fBm with $H = 0.8$; the magnitude of the Fourier transform of the time–series, shown in Fig. 2(b), is a straight line in the log-log scale. We used $H = 0.8$ for all our experiments.

Rather than using an analytic model of the drift, we extracted drift signals from experimental fMRI data. We smoothed time series extracted from an fMRI data set with a smoothing kernel. The smooth drift $\boldsymbol{\theta}$ was a slowly varying function $\boldsymbol{\theta}$ that was then added to the synthetic fMRI response according to (25).

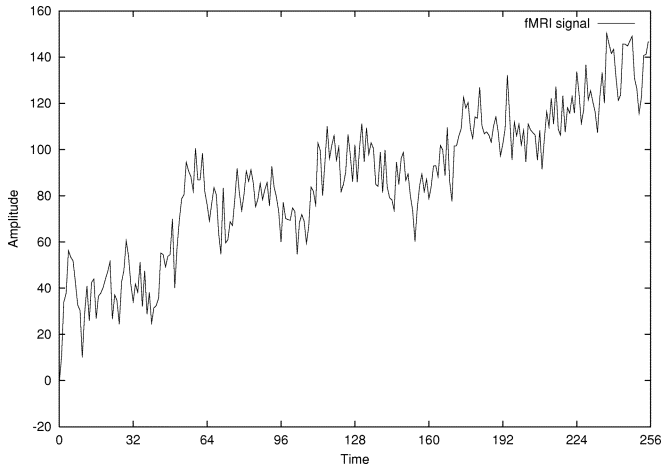

Fig. 3. One realization of a time series for the first data set, with $\sigma = 8$.

TABLE I
SYNTHETIC DATA SET 2, $\sigma = 12$. PERFORMANCE OF THE DETRENDING FOR VARIOUS TREND SCALES

Scale of the trend	P-value	AIC_C	SIC
3	0.0027246	6.6437	6.3099
4	0.00070459	6.3114	5.6889
5	0.00096047	6.2565	5.4047
6	0.0095406	6.616	5.6692
7	0.031598	7.0021	6.0131
8	0.11494	7.4821	6.4644

An example of a simulated fMRI time series is shown in Fig. 3 with $\sigma = 8$.

Table I illustrates the principle of the selection of the scale of the trend. The minimum of the two criteria, AIC_C and SIC, is reached for the same value of the scale, 5. We note that this value does not correspond to the smallest P-value.

We compared the performance of the detrending algorithm with a smoothing spline. The spline $s(t)$ was the solution of the minimization problem

$$\min_{s} \quad \rho \sum_{i=0}^{N-1} |y_i - s(t_i)|^2 + (1-\rho) \int |s''(t)|^2 \, dt. \qquad (27)$$

The solution is a natural cubic spline with knots at the sample points t_i. The number of parameters to describe the spline is $N + 5$. The parameter ρ controls the quality of the fit: if $\rho = 0$, then the solution is the least squares fit and if $\rho = 1$ the solution is in interpolating spline. Several algorithms [32] exist that provide a solution to the optimization of ρ for smoothing some noisy data. Unfortunately, the goal of these algorithms is to recover a signal corrupted by noise and not to estimate the general trend of a time series. We therefore decided to optimize ρ with an oracle: ρ was optimized with the knowledge of the true trend $\boldsymbol{\theta}$. We first searched for the value of the smoothing parameter ρ^0 that minimized

$$\min_{\rho \in [0,1]} \quad \min_{s} \quad \rho \sum_{i=0}^{N-1} |\theta_i - s(t_i)|^2 + (1-\rho) \int |s''(t)|^2 \, dt. \qquad (28)$$

TABLE II
SYNTHETIC DATA SET 1. P-VALUE FOR SEVERAL VALUES OF THE STANDARD DEVIATION σ

Method	$\sigma = 8$	$\sigma = 10$	$\sigma = 12$
No trend	0.078044	0.16956	0.26513
Spline	1.8977e-06	0.002055	0.028673
Wavelet	3.4341e-07	0.0002928	0.0096322

TABLE III
SYNTHETIC DATA SET 2. P-VALUE FOR SEVERAL VALUES OF THE STANDARD DEVIATION σ

Method	$\sigma = 8$	$\sigma = 10$	$\sigma = 12$
No trend	0.029786	0.097294	0.1894
Spline	5.2274e-07	0.00013965	0.0031053
Wavelet	1.0379e-08	1.6578e-05	0.00096047

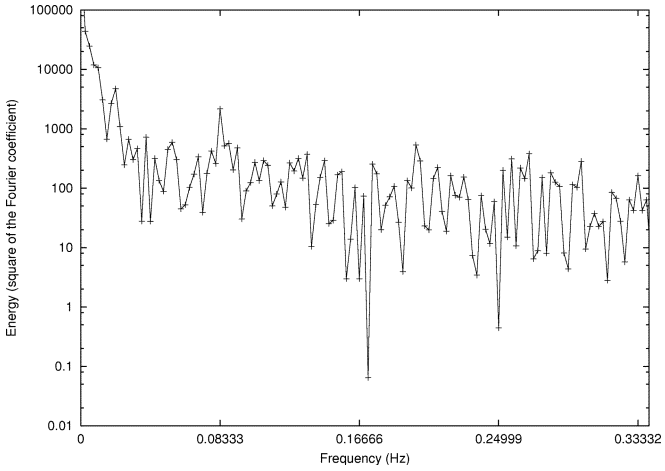

Fig. 4. Power spectrum of the time-series shown in Fig. 3. Most of the energy is located in the low frequency (note the logarithmic scale on the y axis).

This value ρ^0 was then used to estimate the trend from the noisy signals. We report in Tables II and III the mean P-value obtained with different levels of noise. As expected, the P-value increases as the noise level increases. The estimation of the trend with a smoothing spline performed quite well, even at high noise level. Nevertheless, the wavelet-based approach gave systematically better performances than the splines, at all noise levels. We also note that the number of parameters needed to represent a spline is much greater than n_0.

1) Effect of the Detrending on the Spectrum of the fMRI Signal: In order to gain some understanding into the principle of the algorithm, we estimated the spectrum of the time series shown in Fig. 3 before and after the removal of the trend. A discrete Fourier transform was computed (the signal was multiplied by a Gaussian window in order to enforce periodicity). The square of the Fourier coefficients are shown in Fig. 4. As is obvious from Fig. 4, most of the energy lies at the low frequencies. Fig. 5 shows the spectrum of the detrended time series ($J_0 = 5$). It is important to notice that the low-frequency components, that constitute the trend in our model, have been carefully removed without lowering the amplitude of the stimulus. Indeed one can clearly detect the peak at the frequency of the stimulus (1/12 Hz). Unless the stimulus frequency is extremely low, a wavelet filter can be tuned to carefully "excise"

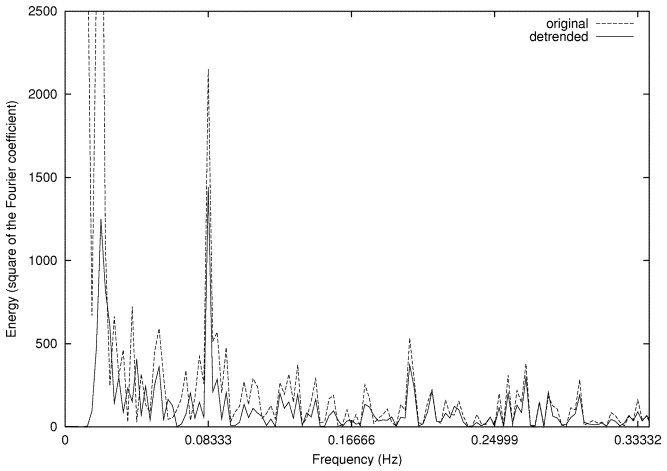

Fig. 5. Power spectrum of the detrended time series. The large scale (low frequencies) contaminants have been removed. The detrending has not removed the energy at the frequency of the stimulus (0.0833 Hz).

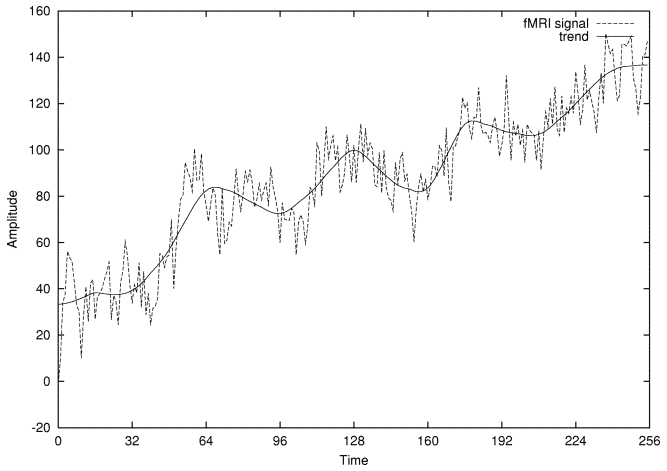

Fig. 6. Time-series shown in Fig. 1 with the optimal $J_0 = 5$ superimposed.

the baseline drift without lowering the Fourier coefficient at the main frequency and the subsequent harmonics of the stimulus.

It is important to realize that because smoothing splines are not filter banks, we cannot control their effect on the spectrum of the detrended signal with the same precision as wavelets. Fig. 6 shows the time-series of Fig. 3 with the optimal trend $(J_0 = 5)$ superimposed. One notes that as the scale of the trend becomes finer (e.g., $J_0 = 3$), the trend starts tracking the variations in the BOLD signal that are due to the stimulus response and will result in a poorer detection performance.

B. Experimental fMRI Data

We illustrate here the principle of the algorithm with some data provided by G. McCarthy (Brain Imaging and Analysis Center, Duke University, Durham, NC), that demonstrate left posterior temporal lobe activation during auditory comprehension [33]. The study involved several subjects who listened passively to alternating sentences spoken in English (their native language) and Turkish (which they did not understand). Each time series was composed of 28 alternating auditory segments of English and Turkish. Each segment lasted for 6 s and images were acquired every 1.5 s. There was a delay of 12 s from

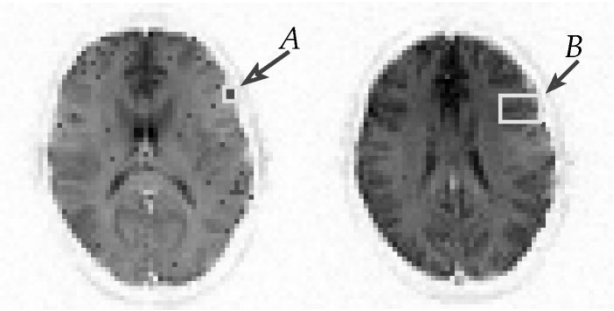

Fig. 7. Turkish-English with wavelet trend. (left) Slice 4. (right) Slice 5. Activation map thresholded at $p = 0.005$ superimposed on the raw EPI images.

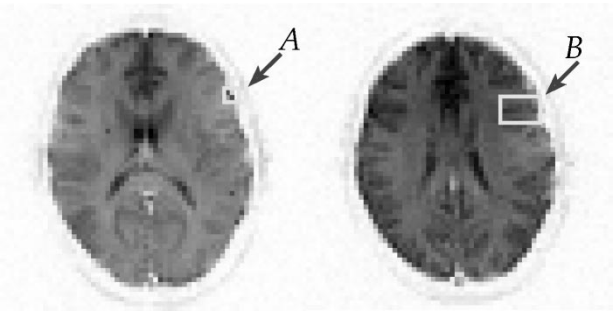

Fig. 8. Turkish-English, no trend correction. (left) Slice 4. (right) Slice 5. Activation map thresholded at $p = 0.005$ superimposed on the raw EPI images.

the first image to the onset of the first sentence. TR $= 1\,500$, slice thickness $= 9$ mm, skip $= 2$ mm, imaging matrix $= 128 \times 64$, voxel size $= 3.2 \times 3.2 \times 9$ mm. More details about the experiments are available in [33]. We analyzed the performance of the detrending with two experiments using two different sets of data from the alternating English–Turkish experiment described in the previous section. A Student t-test was designed to compare the signal under the two conditions: English sentences, or Turkish sentences. Pixels with a P-value less than 0.005 were deemed activated and colored in red in the activation maps.

1) Result of the First Experiment: Fig. 7 shows the result of the t-test for the slices 4 and 5. The activation maps were thresholded at $P = 0.005$ and are superimposed on the raw EPI data. The left side of the brain is represented on the right side of the image. The maps were generated with two runs of alternating Turkish–English intervals, starting with Turkish. The maps clearly show activated pixels in the left inferior frontal lobe (region A and B). Fig. 8 shows an activation map generated by thresholding a standard t-test without removing any trend. For each slice we selected a region of interest (ROI) that contained strongly activated voxels ($P < 10^{-5}$). The activation in these regions was assumed to be truly caused by the stimulus and not by physiological or random noise. The two ROIs are shown within yellow rectangles that are pointed at by the arrows A and B in slice 4 and 5, respectively. The performance of the detrending in each ROI was quantified using the following factors: 1) the number of activated voxels inside the ROI; 2) the mean P-value for all the voxels inside the ROI; and 3) the smallest P-value inside the ROI. These numbers are reported in Tables IV and V. For both slices, the detrending resulted in a

TABLE IV
TURKISH-ENGLISH, SLICE 4. EFFECT OF THE DETRENDING INSIDE THE ROI A (4 VOXELS)

Trend model	number of activated voxels	mean P-value	minimum P-values
no trend	2	0.05066	7.1008e-05
linear	2	0.00376	1.0726e-05
cubic spline	4	2.2057e-03	5.4347e-07
wavelet	4	4.5326e-04	7.91e-07

TABLE V
TURKISH-ENGLISH, SLICE 5. EFFECT OF THE DETRENDING INSIDE THE ROI B (6 VOXELS)

Trend model	number of activated voxels	mean P-value	minimum P-values
no trend	3	3.7018e-02	8.884e-07
linear	4	0.00460	2.9084e-07
cubic spline	5	1.5877e-03	1.0131e-07
wavelet	6	3.6358e-04	1.268e-08

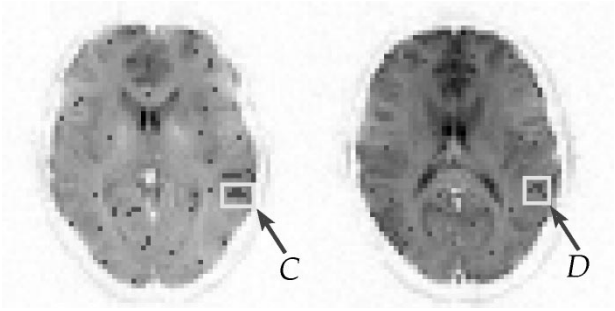

Fig. 9. English-Turkish with wavelet trend. (left) Slice 3. (right) Slice 4. Activation map thresholded at $p = 0.005$ superimposed on the raw EPI images.

noticeable improvement by increasing the number of activated voxels, while decreasing the mean P-value inside the ROIs. The optimal scale was chosen according to AIC_C for each voxel. The performance of the smoothing spline detrending procedure was quantified using the same factors. These numbers are reported in Tables IV and V for the ROI A and B, respectively. We also report the effect of detrending with a linear function. We used the value of ρ that was optimized for the synthetic data sets and the smoothing spline performed quite well.

2) Result of the Second Experiment: A second experiment was conducted with a different data set. Fig. 9 shows the result of the t-test for the slices 3 and 4 after detrending with $J_0 = 4$. The activation maps were thresholded at $P = 0.005$ and are superimposed on the raw EPI data. The maps were generated with two runs of alternating English–Turkish intervals, starting with English. The maps show in red activated pixels in the left posterior temporal lobe (regions C and D). Fig. 10 shows an activation map generated by thresholding a standard t-test without removing any trend. For each slice we again selected a region of interest (ROI) that contained strongly activated voxels ($P < 10^{-3}$). We note that the mean P-value before detrending was not as high as in the previous experiment. The

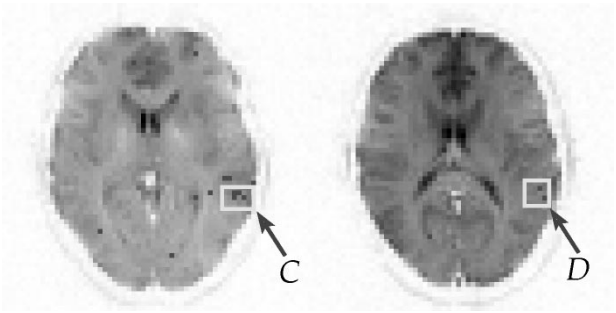

Fig. 10. English-Turkish, no trend correction. (left) Slice 3. (right) Slice 4. Activation map thresholded at $p = 0.005$ superimposed on the raw EPI images.

TABLE VI
ENGLISH-TURKISH, SLICE 3. EFFECT OF THE DETRENDING INSIDE THE ROI C (6 VOXELS)

Trend model	number of activated voxels	mean P-value	minimum P-values
no trend	4	2.991e-03	3.7587e-03
linear	4	2.381e-03	9.5419e-05
cubic spline	6	1.5578e-03	1.1582e-05
wavelet	6	8.4305e-04	7.0537e-07

TABLE VII
ENGLISH-TURKISH, SLICE 4. EFFECT OF THE DETRENDING INSIDE THE ROI D (6 VOXELS)

Trend model	number of activated voxels	mean P-value	minimum P-values
no trend	2	9.4974e-03	1.8923e-03
linear	5	5.9516e-03	1.7691e-03
cubic spline	5	3.5858e-03	8.9608e-05
wavelet	6	1.4488e-03	1.8598e-05

two ROIs are within the yellow rectangles that are pointed at by the arrows C and D in slices 3 and 4, respectively. The performance of the detrending in each ROI was quantified using the same factors as in the previous experiments. These numbers are reported in Tables VI and VII. For both slices, the detrending resulted in a noticeable improvement by increasing the number of activated voxels, while keeping the mean P-value inside the ROIs at the same value. As before, the performance of the smoothing spline detrending procedure was quantified using the same factors. These numbers are reported in Tables VI and VII for the ROI C and D, respectively. As explained in the previous section, the smoothing spline performed quite well. But the wavelet-based approach still outperformed the estimation of the trend using a smoothing spline, while using a much smaller set of parameters.

VI. CONCLUSION

In this paper, we addressed the problem of estimating the parameters of the semiparametric generalized linear model (2). This model allows to describe the fMRI response to a stimulus contaminated by a fractional Brownian motion noise and a systematic baseline drift. The trend belongs to a subspace spanned

by large scale wavelets. We have developed a scale space regression that permits to carry out the regression in the wavelet domain while omitting the scales that are contaminated by the trend. In order to demonstrate that our approach outperforms the state-of-the art detrending technique, we evaluated our method against a smoothing spline approach. Experiments with simulated data and experimental fMRI data, demonstrate that our approach can infer and remove drifts that cannot be adequately represented with splines. Our approach results in a noticeable improvement by reducing the false positive rate and increasing the true positive rate. While periodic stimuli still constitute the majority of fMRI experimental designs, we are currently extending our approach to event-related stimuli.

ACKNOWLEDGMENT

The author would like to thank G. McCarthy, Director of the Brain Imaging and Analysis Center, Duke University, Durham, NC, for making the fMRI data available for this paper.

REFERENCES

[1] I. Vanzetta and A. Grinvald, "Increased cortical oxidative metabolism due to sensory stimulation: Implications for functional brain imaging," *Science*, vol. 286, pp. 1555–1558, Nov. 1999.

[2] P. A. Bandettini, "The temporal resolution of functional MRI," in *Functional MRI*, C. T. W. Moonen and P. A. Bandettini, Eds. New York: Springer-Verlag, 1999, pp. 205–220.

[3] X. Hu, E. Yacoub, T. H. Le, E. R. Cohen, and K. Ugurbil, "Functional MRI decrease at the onset of stimulation," in *Functional MRI*, C. T. W. Moonen and P. A. Bandettini, Eds. New York: Springer-Verlag, 1999, pp. 243–252.

[4] P. Jezzard, "Physiological noise: Strategies for correction," in *Functional MRI*, C. T. W. Moonen and P. A. Bandettini, Eds. New York: Springer-Verlag, 1999, pp. 173–182.

[5] A. M. Smith, B. K. Lewis, U. E. Ruttimann, F. Q. Ye, T. M. Sinnwell, Y. Yang, J. H. Duyn, and J. A. Frank, "Investigation of low frequency drift in fMRI signal," *NeuroImage*, vol. 9, no. 5, pp. 526–533, 1999.

[6] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain," *Magn. Reson. Med.*, vol. 30, pp. 161–173, 1993.

[7] B. B. Biswal, E. A. DeYoe, and J. S. Hyde, "Reduction of physiological fluctuations in fMRI using digital filters," *Magn. Reson. Med.*, vol. 35, pp. 107–113, 1996.

[8] M. J. Lowe and D. P. Russell, "Treatment of baseline drifts in fMRI time series analysis," *J. Comput. Assist. Tomogr.*, vol. 23, no. 3, pp. 463–473, 1999.

[9] G. H. Glover, "Deconvolution of impulse response in event-related bold fMRI," *NeuroImage*, no. 9, pp. 416–429, 1999.

[10] R. L. Eubank, *Non Parametric Regression and Spline Smoothing*. New York: Marcel-Dekker, 1999.

[11] F. Kruggel, D. Y. von Cramon, and X. Descombes, "Comparison of filtering methods for fMRI datasets," *NeuroImage*, no. 10, pp. 530–543, 1999.

[12] E. Zarahn, G. K. Aguire, and M. D'Esposito, "Empirical analysis of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null hypothesis conditions," *NeuroImage*, vol. 5, pp. 179–197, 1997.

[13] J. Fadili and E. Bullmore, "Wavelet-generalized least squares: a new BLU estimator of linear regression models with $1/f$ errors," *NeuroImage*, vol. 15, pp. 217–232, 2002.

[14] K. L. Hurvich and C. L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, pp. 297–307, 1989.

[15] S. C. Ngan, S. M. LaConte, and X. Hu, "Temporal filtering of event-related fMRI data using cross-validation," *NeuroImage*, vol. 11, pp. 797–804, 2000.

[16] U. E. Ruttimann *et al.*, "Statistical analysis of functional MRI in the wavelet domain," *IEEE Trans. Med. Imag.*, vol. 17, pp. 142–154, Feb. 1998.

[17] J. Raz and B. Turetsky, "Wavelet ANOVA and fMRI," *Wavelet Applicat. Signal Image Processing VII*, pp. 561–570, 1999.

[18] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[19] B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, 1968.

[20] P. C. Ivanov, M. G. Rosenblum, C. K. Peng, J. M. S. Havlin, H. E. Stanley, and A. L. Goldberger, "Scaling behavior of heartbeat intervals obtained by wavelet-based time-series analysis," *Nature*, vol. 383, no. 6598, pp. 323–327, 1996.

[21] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter, and M. Brammer, "Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domain," *Human Brain Mapping*, vol. 78, pp. 61–78, 2001.

[22] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 904–900, Mar. 1992.

[23] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, March 1992.

[24] G. W. Wornell, "A Karhunen-Loève-like expansion for $1/f$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, pp. 859–861, Mar. 1998.

[25] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Stat. Soc. B*, vol. 59, no. 2, pp. 319–351, 1997.

[26] A. D. R. McQuarrie and C. L. Tsai, *Regressions and Time Series Model Selection*. Singapore: World Scientific, 1998.

[27] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

[28] S. Efromovich, *Non Parametric Curve Estimation: Methods, Theory and Application*. New York: Springer-Verlag, 1999.

[29] A. Cohen, I. Daubechies, and J. Feauveau, "Bi-orthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485–560, 1992.

[30] P. Abry and F. Sellan, "The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: remarks and fast implementation," *Appl. Comput. Harmon. Anal.*, vol. 3, pp. 377–383, 1996.

[31] Y. Meyer, F. Sellan, and M. Taqqu, "Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion," *J. Fourier Anal. Appl.*, vol. 5, pp. 465–494, 1999.

[32] P. Craven and G. Wahba, "Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation," *Numerische Mathematik*, vol. 31, pp. 377–403, 1979.

[33] M. J. Schlosser, N. Aoyagi, R. K. Fullbright, J. C. Gore, and G. McCarthy, "Functional MRI studies of auditory comprehension," *Human Brain Mapping*, vol. 6, pp. 1–13, 1998.