



Wavelet-generalised least squares: a new BLU estimator of linear regression models with $1/f$ errors

Jalal M. Fadili, E. T. Bullmore

► To cite this version:

Jalal M. Fadili, E. T. Bullmore. Wavelet-generalised least squares: a new BLU estimator of linear regression models with $1/f$ errors. *NeuroImage*, 2002, 15, pp.217-232. 10.1006/nimg.2001.0955 . hal-01018183

HAL Id: hal-01018183

<https://hal.science/hal-01018183>

Submitted on 3 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Wavelet-Generalized Least Squares: A New BLU Estimator of Linear Regression Models with $1/f$ Errors

M. J. Fadili^{*†} and E. T. Bullmore[†]

^{*}Centre for Speech and Language, University of Cambridge, United Kingdom; and [†]Brain Mapping Unit and Wolfson Brain Imaging Centre, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom

Long-memory noise is common to many areas of signal processing and can seriously confound estimation of linear regression model parameters and their standard errors. Classical autoregressive moving average (ARMA) methods can adequately address the problem of linear time invariant, short-memory errors but may be inefficient and/or insufficient to secure type 1 error control in the context of fractal or scale invariant noise with a more slowly decaying autocorrelation function. Here we introduce a novel method, called wavelet-generalized least squares (WLS), which is (to a good approximation) the best linear unbiased (BLU) estimator of regression model parameters in the context of long-memory errors. The method also provides maximum likelihood (ML) estimates of the Hurst exponent (which can be readily translated to the fractal dimension or spectral exponent) characterizing the correlational structure of the errors, and the error variance. The algorithm exploits the whitening or Karhunen–Loève-type property of the discrete wavelet transform to diagonalize the covariance matrix of the errors generated by an iterative fitting procedure after both data and design matrix have been transformed to the wavelet domain. Properties of this estimator, including its Cramér–Rao bounds, are derived theoretically and compared to its empirical performance on a range of simulated data. Compared to ordinary least squares and ARMA-based estimators, WLS is shown to be more efficient and to give excellent type 1 error control. The method is also applied to some real (neurophysiological) data acquired by functional magnetic resonance imaging (fMRI) of the human brain. We conclude that wavelet-generalized least squares may be a generally useful estimator of regression models in data complicated by long-memory or fractal noise.

Key Words: $1/f$ noise; wavelets; maximum likelihood; Monte Carlo simulation; fMRI; fractal.

INTRODUCTION

Suppose that we observe a data vector y or $\{y_i\}_{i=1, 2, 3, \dots, N}$ together with corresponding explanatory variables x_k $k = 1, 2, 3, \dots, p$. Define X to be the $N \times p$ design matrix whose columns are the vectors x_k . The standard regression model offers a framework where the observation y is related to the explanatory factors or covariates x_k via an additive linear model. This simple relationship can be written

$$y = X\beta + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma) \quad (1)$$

where ϵ is a $N \times 1$ zero mean random vector with covariance matrix Σ , β is the p -dimensional model parameter vector with a coefficient corresponding to each factor or covariate, and the observation vector y has finite energy in \mathbb{R}^N .

This equation divides the total variance of the observations into two main components independent from each other: one deterministic and one stochastic. The deterministic part is a linear function of the regressors through the model coefficients. The stochastic part models the uncertainty on the measurements. Standard assumptions are that the errors ϵ are independent and normally distributed. The effect of deviations from normality has been one of the main topics of robust statistics (see, e.g., Beran, 1994; Huber, 1981; Hampel et al., 1986; Koul, 1993). There is also an extended literature on deviations from independence, which have been modeled by autoregressive and other short-memory processes (e.g., ARMA, ARIMA) (Box and Jenkins, 1976; Pollock, 1999), as well as by state-space approaches (Jones, 1993). Here, we consider the linear regression model where ϵ is a wide sense stationary long-memory process.

The problem of estimating regression model parameters in the presence of contaminating errors with

[†]To whom correspondence and reprint requests should be addressed at present address: GREYC-ISMRA Image CNRS UMR 6072 6, Bd du Maréchal Juin, 14000 Caen Cedex, France. Fax: +33-(0)31-45-26-98. E-mail: Jalal.Fadili@greyc.ismra.fr.

strong correlation between observations far apart from each other is now widespread in many diverse fields, simply because natural signals very commonly exhibit long-range dependence (see examples in physics, Cas-sandro and Jona-Lasinio, 1978), geophysics (Foufoula-Georgiou and Kumar, 1994; Torrence and Compo, 1998; Whitcher et al., 2000), electronics (Voss, 1979; Ziel, 1986), econometrics (Jensen and Whitcher, 2000; Jensen, 1994), electrophysiology (Goldberger et al., 1990; Raz et al., 1999), and imaging (Mallat, 1989; Lundahl et al., 1986; Krueger et al., 1996). In particular, there is some evidence to suggest that the error structure of functional magnetic resonance imaging (fMRI) time series may be complicated by long-range dependencies or, to put it another way, may be characterized by a $1/f$ -like power spectrum (Zarahn et al., 1997).

The statistical problems of regression in the context of long-memory or self-similar errors have already attracted the attention of several investigators (Beran, 1994; Koul, 1993, 1994; Robinson and Hidalgo, 1997). In particular, Beran investigated relationships between the design matrix X , long-range dependence in the errors, and the distribution of the regression model parameter estimates. He also quantified the asymptotic loss when using the ordinary least squares (OLS) estimator instead of the best linear unbiased (BLU) estimator through several useful theorems (Beran, 1994, pp. 172–194) and showed that the variance of both OLS and BLU estimators converged more slowly towards zero than in the case of short-memory errors, the asymptotic efficiency depending on both the complexity of the design matrix and the decaying speed of the autocorrelation function. Although, as expected, the BLU estimator was consistently more efficient than OLS, it requires iterative estimation of the covariance matrix Σ , which is generally unknown a priori, entailing a computationally demanding and numerically unstable inversion of $\hat{\Sigma}$ at each step of the algorithm (Beran, 1994; Azzalini, 1996).

The purpose of this paper is to propose a novel wavelet-based estimator for the parameters of a linear regression model with long-memory errors. We first provide a theoretical introduction to the class of long-memory processes and to the principles of time-scale analysis by wavelets (more comprehensive accounts of wavelets are provided elsewhere (Mallat, 1998; Wickerhauser, 1994; Chui, 1994; Daubechies, 1992)). We then prove the usefulness of the discrete wavelet transform in the context of regression models with long-memory errors. The key idea is that the orthonormal wavelet decomposition provides us with a near Karhunen–Loève-type expansion (Wornell, 1990, 1996; Flandrin, 1992; Tewfik and Kim, 1992) and will therefore be a good whitening filter for self-similar processes. This allows us to assume that the off-diagonal elements of the error covariance matrix Σ are zero—

which makes the inversion Σ^{-1} much faster and more stable to compute. An iterative algorithm is proposed which approximates the maximum likelihood (BLU) estimator of both model and noise parameters, this approximation being as good as the approximation of the discrete wavelet transform to the ideal Karhunen–Loève expansion. Monte-Carlo simulations are reported to validate theoretically anticipated properties of this wavelet-generalized least squares estimator and to compare it, in terms of sensitivity, efficiency and inferential validity (nominal type 1 error control), to the well known OLS estimator and an ARMA-based generalized least squares procedure, denoted ARLS, which assumes a short-memory autoregressive AR(q) model for the errors. Finally, the method is illustratively applied to analysis of fMRI data acquired at rest and during simple visual stimulation.

THEORETICAL BACKGROUND

Stationary Long-Memory Processes

One well-known model of long-memory processes, proposed by Mandelbrot and van Ness (1966), is fractional Brownian motion, of which classical Brownian motion is a special case. Fractional Brownian motion is self-similar or fractal and characterized by a single scalar self-similarity parameter $0 < H < 1$, called the Hurst exponent (Hurst, 1951) ($H = \frac{1}{2}$ for classical Brownian motion). Although fractional Brownian motion is not stationary in the usual sense (Beran, 1994), the associated increment process is stationary. For classical Brownian motion, the increment process is white Gaussian noise. Two more general examples of this class are fractional Gaussian noise (fGn) and autoregressive fractionally integrated moving average (ARFIMA) processes. We shall show that both fGn and ARFIMA processes have $1/f$ spectral properties, i.e., no single characteristic frequency but disproportionate power at low frequencies, suggesting that wavelets will provide a natural basis for their decomposition (Flandrin, 1992; Wornell, 1996).

Fractional Gaussian noise. Fractional Gaussian noise was first defined as the increments between successive values of a fractional Brownian motion parameterized by Hurst exponent $0 < H < 1$ (Mandelbrot and Ness, 1968; Beran, 1994; Deriche and Tewfik, 1993; Wornell, 1996). More formally, for $H \in]0, 1[$, there exists exactly one Gaussian process G_t that is the stationary increment of a self-similar process F_t . This process is called a fractional Gaussian noise (fGn) and the corresponding self-similar process is called fractional Brownian motion (fBm). The autocovariance structure of the fGn $G_t = F_t - F_{t-1}$ is given by (Beran, 1994; Wornell, 1996):

$$R_G(\tau) = \frac{\sigma^2}{2} [(\tau + 1)^{2H} - 2\tau^{2H} + (\tau - 1)^{2H}] \quad (2)$$

$$\underset{\tau \rightarrow \infty}{\approx} \sigma^2 H(2H - 1)\tau^{2H-2}$$

for lag $\tau \geq 0$ and $R_G(\tau) = R_G(-\tau)$ for $\tau < 0$. For $\frac{1}{2} < H < 1$, the process has long-range dependence or persistence; for $H = \frac{1}{2}$ the process is white; and for $0 < H < \frac{1}{2}$, the process has short-range dependence or anti-persistence and the correlations sum to zero (Beran, 1994). Under the constraint that $0 < H < 1$, the spectral density of G_t can be derived by Taylor expansion (Beran, 1994):

$$S(f) \underset{f \rightarrow 0}{\approx} \frac{\sigma_H^2}{f^{2H-1}} \quad (3)$$

$$\sigma_H^2 = \sigma^2(2\pi)^{-2H} \sin(\pi H) \Gamma(2H + 1).$$

In fact, this approximation is very good even for relatively large frequencies and will be assumed for $f \in [-1/2, 1/2]$. As shown by Eq. (3), the fGn has a $1/f$ spectral density with an exponent $-1 < \gamma = 2H - 1 < 1$.

The ARFIMA process. The fractional ARIMA model is a natural extension of the classical ARIMA(p, d, q) model (Granger and Joyeux, 1980). The ARFIMA model captures the slowly decaying autocovariance structure of a long-memory process with fractional difference parameter $d = H - \frac{1}{2}$, where the parameter d is allowed to take non-integer values. By definition, an ARFIMA process is a zero-mean stationary long-memory process A_t (Beran, 1994) such that:

$$P(z^{-1})(1 - z^{-1})^d A_t = Q(z^{-1})\epsilon_t \left\{ \begin{array}{l} |d| < 1/2 \\ \epsilon_t \text{ i.i.d. } \sim \mathcal{N}(0, \sigma_\epsilon^2) \end{array} \right. \quad (4)$$

where z^{-1} is the usual lag operator, and P and Q are the short-memory polynomials of order p and q with roots outside the unit circle. Note that the case $d > \frac{1}{2}$ can be reduced to the case $-\frac{1}{2} < d \leq \frac{1}{2}$ by appropriate differences (Beran, 1994). However, only the range specified in Eq. (4) is of interest in most situations.

Ignoring the short-memory terms, one can prove that the covariance structure of such a process is (Beran, 1994)

$$R_A(\tau) = \frac{\sigma_\epsilon^2 \Gamma(1 - 2d) \Gamma(\tau + d)}{\Gamma(d) \Gamma(1 - d) \Gamma(\tau + 1 - d)} \underset{\tau \rightarrow \infty}{\propto} |\tau|^{2d-1}, \quad (5)$$

and its spectral density function is

$$S(f) = \sigma_\epsilon^2 |2 \sin(\pi f)|^{-2d} \underset{f \rightarrow 0}{\propto} \frac{\sigma_\epsilon^2}{f^{2d}}. \quad (6)$$

Thus the ARFIMA(0, d , 0) process, or simply ARFIMA process for ease of notation, has a nearly $1/f$ power spectrum with $\gamma = 2d = 2H - 1$. To avoid any pathological behaviour, the ARFIMA process is defined for $|d| < \frac{1}{2}$; hence, the spectral exponent γ is in the range $]-1, 1[$ (Beran, 1994).

Discrete Wavelet Transform

The dyadic orthonormal wavelet transform of a finite energy signal f is defined as the inner product (Mallat, 1989, 1998):

$$d_{j,k} = \langle f, \psi_{j,k} \rangle = 2^{-j/2} \int f(t) \psi(2^{-j}t - k) dt, \quad (j, k) \in \mathbb{Z}^2. \quad (7)$$

The coefficient $d_{j,k}$ is the detail coefficient (or the wavelet coefficient) at scale j and position k . The function ψ is the wavelet function, whose collection of dilations j and translations k form an orthonormal basis in the Hilbertian space $L^2(\mathbb{R})$. Any continuous function qualifies if the admissibility conditions are satisfied: ψ is well-localized around zero (rapid decay to zero as $t \rightarrow \infty$) and oscillates ($\int \psi(t) dt = 0$). These conditions can be strengthened to include more vanishing moments (up to an order R) and/or higher order continuous derivatives (Daubechies, 1992). From a filter-bank point of view, the wavelet $\psi_{j,k}$ can be viewed as an octave band-pass filter in $[-\pi/2^j, -\pi/2^{j+1}] \cup [\pi/2^{j+1}, \pi/2^j]$. In a multiresolution framework, the wavelet coefficient at scale j can be interpreted as the difference of information between two approximations to f at scales $j - 1$ and j .

The multiresolutional approximation of the signal, i.e., the smooth component of the data, at scale j is the inner product (Mallat, 1989, 1998):

$$a_{j,k} = \langle f, \phi_{j,k} \rangle = 2^{-j/2} \int f(t) \phi(2^{-j}t - k) dt, \quad (8)$$

$$(j, k) \in \mathbb{Z}^2.$$

The coefficient $a_{j,k}$ is the approximation coefficient at scale j and position k . The function ϕ is the scaling function ($\int \phi(t) dt \neq 0$), whose collection of translations k form an orthonormal basis. From a filter-bank point of view, the scaling function $\phi_{j,k}$ can be viewed as a low-pass filter in the octave $]-\pi/2^{j+1}, \pi/2^{j+1}[$. A signal analyzed up to a scale J can be perfectly reconstructed from its detail coefficients $d_{j,k}$ ($j \in \{1, \dots, J\}$) and the remaining approximation coefficient(s) at the scale J (Mallat, 1989).

The wavelet transform has become very popular in digital signal processing owing to the pyramidal of S .

Mallat (Mallat, 1989), routinely used to calculate the discrete wavelet transform (DWT) of a sampled signal, which requires only $O(N)$ operations and is thus faster than the fast Fourier transform, which requires $O(N \log N)$ operations. In practical applications to finite time series ($N = 2^J$ without loss of generality), we must choose the regularity and type of wavelet, and the method of boundary correction adopted to estimate wavelet coefficients at the limits of the data.

It has been shown that the wavelet coefficients of a long-range stationary $1/f$ process have a correlation structure whose magnitude decays as (Wornell, 1990, 1996; Tewfik and Kim, 1992):

$$|\rho_{j,j'}^{k,k'}| \underset{2^j k - 2^{j'} k' \rightarrow \infty}{\sim} O(|2^j k - 2^{j'} k'|^{2(H-1)-2R}), \quad (9)$$

where R is the number of vanishing moments. Therefore, for $0 < H < 1$, the intercoefficient correlations, both within and between scales of the decomposition, can be reasonably ignored for any wavelet with sufficient number of vanishing moments (i.e., with $R \geq 1$). For the fBm process, Flandrin (Flandrin, 1992) has shown this property even for the simplest Haar wavelet ($R = 1$) for $0 < H < 1/2$. It should be noted that, when dealing with finite length data (as is the case of course in fMRI time series analysis), increasing the regularity of the wavelet will theoretically diminish the magnitude of correlation between wavelet coefficients but also increases the risk of artifactually large intercoefficient correlations arising due to boundary effects. A simulation illustrating this fact is given in (McCoy, 1994). Hence, a compromise has to be negotiated in choosing a wavelet with reasonable regularity. Here we have used a Daubechies wavelet with four vanishing moments. This wavelet has a large enough number of vanishing moments to ensure rapid decay of the wavelet coefficient autocorrelation function in Eq. (9). It is also the most compactly supported wavelet for any number of vanishing moments and this compactness mitigates the extent of intercoefficient correlations introduced by periodic boundary correction in computing wavelet coefficients at the limits (0 and $N - 1$) of a finite times series (Mallat, 1998).

WAVELET-GENERALIZED LEAST SQUARES

Model Formulation

As before we write a linear regression model in the time domain

$$y = X\beta + \epsilon \quad (10)$$

but now we specify that the random errors have a long-memory covariance structure parameterized by

one of a number of simply related fractal noise parameters, H , γ or d , i.e., $\epsilon \sim \mathcal{N}(0, \Sigma(H))$. The power-law spectrum of such processes can be generally written as $S(f) \underset{|f| \rightarrow 0}{\sim} (\sigma^2 c_\gamma / |f|^\gamma)$, where c_γ is a dimensionless function of the spectral exponent γ (e.g., see Eq. (3)).

Taking the discrete wavelet transform of Eq. (10) gives us the following linear model in the wavelet domain

$$y_w = X_w \beta + \epsilon_w, \quad (11)$$

where X_w is the column-wise DWT of the original design matrix X , y_w , and ϵ_w are, respectively, the wavelet transforms of the observed data and the noise process up to the maximum scale J . The transform coefficients can be expressed at each scale j and position k :

$$y_{j,k} = \sum_{m=1}^p x_{jk}^m \beta_m + \epsilon_{jk} \quad \forall (j, k) \in \mathcal{R} \quad (12)$$

$$\mathcal{R} = \{j \in \{1, \dots, J\}, k \in \{0, \dots, N/2^j - 1\}\}.$$

The orthonormal wavelet basis decomposition acts like a Karhunen–Loève expansion of $1/f$ processes. To a good approximation, the variance-covariance matrix of the noise is therefore diagonalized by the wavelet transform to take the following form:

$$\Sigma_J = \begin{pmatrix} S_{a_j} & & & \\ & S_{d_j} & & \\ & & \ddots & \\ & & & \begin{pmatrix} S_{d_2} & \dots & S_{d_2} \end{pmatrix} \\ & & & & \begin{pmatrix} S_{d_1} & \dots & S_{d_1} \end{pmatrix} \end{pmatrix}, \quad (13)$$

where S_{a_j} and S_{d_j} , $j = 1, 2, \dots, J$, are the variances of the scaling and wavelet coefficients as defined in the next section.

Fractal Noise Parameter Estimation

The wavelet coefficients of a $1/f$ Gaussian process at level j are a set of stationary i.i.d. variables (Flandrin, 1992; Tewfik and Kim, 1992) $\epsilon_{j,k} \sim \mathcal{N}(0, S_{d_j})$. The smooth component at scale J is also a normally distributed variable with variance S_{a_J} .

The variances of the detail coefficients at scale j and the approximation at scale J are given by:

$$S_{d_j} = \frac{2^j}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma^2 c_\gamma}{\omega^\gamma} |\Psi(2^j \omega)|^2 d\omega$$

$$S_{a_j} = \frac{2^j}{2\pi} \int_{-\infty}^{+\infty} \frac{\sigma^2 c_\gamma}{\omega^\gamma} |\Phi(2^j \omega)|^2 d\omega, \quad (14)$$

where Ψ and Φ are, respectively, the Fourier transforms of the wavelet and the scaling functions. This general expression can be simplified when the wavelet function is approximated by an ideal bandpass filter, and the scaling function by an ideal low-pass filter; in this case, Eq. (14) simplifies to:

$$S_{d_j} \approx \frac{2^{j+1}}{(2\pi)^\gamma} \int_{2^{-(j+1)}}^{2^{-j}} \frac{\sigma^2 c_\gamma}{f^\gamma} df \text{ for } j \in \{1, \dots, J\}$$

$$S_{a_j} \approx \frac{2^{J+1}}{(2\pi)^\gamma} \int_0^{2^{-(J+1)}} \frac{\sigma^2 c_\gamma}{f^\gamma} df. \quad (15)$$

Integration gives the following expressions for $-1 < \gamma < 1$:

$$S_{d_j} \approx \frac{\sigma^2 c_\gamma (2^j)^\gamma}{(2\pi)^\gamma [1 - \gamma]} [2 - 2^\gamma] = \sigma^2 S_{d_j}(\gamma). \quad (16)$$

Likewise, the variance of the scaling coefficient at level J is

$$S_{a_J} \approx \frac{\sigma^2 c_\gamma (2^{J+1})^\gamma}{(2\pi)^\gamma [1 - \gamma]} = \sigma^2 S_{a_J}(\gamma). \quad (17)$$

Equations (16) and (17) are not valid for the limiting value $\gamma = 1$ ($H = 1$), although the result in this case remains easy to calculate. From Eq. (16), it can be seen that there exists a linear relationship in log-log space between the detail variance at each scale and the scale index:

$$\log_2(S_{d_j}) = \gamma j + \kappa. \quad (18)$$

A similar relationship can be shown even for higher order cumulants (Pesquet-Popescu, 1999) and this property has been exploited to synthesize 1/f-like processes. It has also been extensively used to characterize self-similar processes and to estimate the self-similarity parameter in the context of the classical DWT (Flandrin, 1992; Mallat, 1989; Jensen, 1994, 1999; Percival, 1992) or using the stationary DWT (also called the maximum overlap DWT) (Jensen and Whitcher, 2000; Whitcher and Jensen, 2000). However, despite its simplicity and flexibility, this estimator requires a

robust estimate of the variance of the coefficients at each level, which naturally becomes more problematic when bigger detail levels (with fewer coefficients) are involved and/or when the time series is not very long (Percival, 1995).

Approximately Maximum Likelihood Estimation in the Wavelet Domain

Maximum likelihood (ML) estimators provide an alternative approach to estimation, classification, discrimination, or prediction of long-memory of 1/f errors and have been developed in the wavelet domain by several authors (Deriche and Tewfik, 1993; Wornell, 1992, 1996; Wornell and Oppenheim, 1992a, b; McCoy, 1994; McCoy and Walden, 1996; Taquq et al., 1995; Jensen, 1994, 2000; see Wornell, 1996, for review). The ML estimator of the fractal noise parameter H described below is therefore not original; but, to the best of our knowledge, this wavelet-generalized least squares (WLS) algorithm for combined maximum likelihood estimation of both H and the linear model parameter vector β (and the error variance σ^2) is distinctly novel. However, it is worth noting that we will assume the off-diagonal elements of the projected covariance matrix Σ to be zero (Wornell, 1990; McCoy and Walden, 1996) in the calculation of the likelihood function. This is a reasonable approximation, given the well-known Karhunen-Loève-type properties of the wavelet transform applied to a 1/f process, but it is still an approximation. (We later show by Monte Carlo studies that some empirical properties of the estimator coincide closely with the expected performance of the maximum likelihood estimator, further suggesting that diagonalization of the error covariance matrix is an acceptable approximation in the wavelet domain).

The likelihood function for the model in the wavelet domain is

$$L(\theta) = \frac{|\Sigma_J^{-1}|}{(2\pi)^{N/2}} e^{-((y_w - X_w \beta)^t \Sigma_J^{-1} (y_w - X_w \beta)/2)}, \quad (19)$$

$$\theta^t = (\beta^t, \gamma, \sigma^2),$$

which can be rearranged to give the log-likelihood function,

$$LL(\theta) = -\frac{1}{2} [N \log(2\pi\sigma^2) + \log(S_{a_J})$$

$$+ \sum_j N(j) \log(S_{d_j})]$$

$$- \frac{1}{2\sigma^2} \left[\frac{(y_{J,0} - \sum_{l=1}^p x_{J,0}^l \beta_l)^2}{S_{a_J}} \right. \quad (20)$$

$$\left. + \sum_{j,k} \frac{(y_{j,k} - \sum_{l=1}^p x_{j,k}^l \beta_l)^2}{S_{d_j}} \right],$$

where $N(j) = N/2^j$, S_{d_j} and S_{a_j} are, respectively, defined in Eqs. (16) and (17). The ML parameter estimates are those that maximize the log-likelihood, i.e., $\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} (LL(\theta))$.

The WLS algorithm. The wavelet-generalized least squares algorithm we propose for solving this equation can now be fully specified as follows:

1. Define the wavelet basis to be used, the method for boundary correction, and the maximum scale of decomposition J .

2. Initialize by an ordinary least squares fit of the regression model in the time domain.

3. Take the DWT of each column of the design matrix X and of the data vector y .

4. Let $d_{j,k} = y_{j,k} - \sum_{m=1}^p x_{j,k}^m \beta_m$ be the detail residuals at scale j and location k , and $a_{j,0} = y_{j,0} - \sum_{m=1}^p x_{j,0}^m \beta_m$ be the approximation residual at scale J and location 0. Differentiating Eq. (20) with respect to σ^2 gives directly the ML estimate of the residual variance

$$\hat{\sigma}_{ML}^2 = \frac{1}{N} \left[\frac{a_{j,0}^2}{S_{a_j}} + \sum_{j,k} \frac{d_{j,k}^2}{S_{d_j}} \right]. \quad (21)$$

5. The long-memory parameter γ is then estimated by substituting σ^2 in Eq. (20) to minimize the concentrated log-likelihood, either directly (if the noise has an exactly $1/f$ spectrum (Wornell, 1992, 1996) or by numerical methods. We used the golden section search method, which will generally work well provided we start with an initial interval containing γ (Press et al., 1992). Since we here expect $0 < H < 1$, it follows that an appropriate initial interval is $-1 < \gamma < 1$. (An additional advantage of numerical minimization is that it allows us to consider nearly- $1/f$ processes where the octave bandpass variances are only known numerically). In any case, since γ is related to H via an

invertible transformation, its ML estimator corresponds with that of H .

6. Finally, using backsubstitution, the parameter vector β is estimated in matrix form by

$$\hat{\beta} = (X_w^t \hat{\Sigma}_J^{-1} X_w)^{-1} X_w^t \hat{\Sigma}_J^{-1} y_w. \quad (22)$$

It is obvious that since Σ_J is diagonal, its inverse is easily accessible by taking the reciprocals of the diagonal elements.

7. Goto step 4 and iterate until the change in successive parameter estimates is arbitrarily small, i.e., $\leq 10^{-2}$. Typically the algorithm converges quickly, requiring three iterations or less to find stable parameter estimates.

This approach is similar in flavor to the Box–Jenkins methodology for ARMA model fitting (Box and Jenkins, 1976). In fact, it can be regarded as a generalization of ARMA model estimation to the classically intractable case of error processes with infinite-range dependence.

Properties of the WLS Estimator

It is well known that under regularity conditions, the ML estimator exists, is efficient and consistent in both simple and mean-square senses (at least asymptotically) (Azzalini, 1996). The Gauss–Markov theorem provides a formal proof that the ML estimate of β is asymptotically unbiased with minimal variance, i.e., it is the best linear unbiased estimator. In addition to these properties, asymptotically the ML estimator is Normal and the variances of the parameter estimates are well-approximated by the Cramér–Rao lower bounds (CRLB) derived from the expected Fisher information matrix.

Derivation of Cramér–Rao bounds for WLS estimates. First, we define the Fisher information matrix as follows:

$$I^{-1} = \begin{pmatrix} \frac{N}{2\Delta\sigma^4} & \frac{-A}{\Delta} & 0 & \cdots & 0 \\ \frac{-A}{\Delta} & \frac{1}{2\Delta} [(\log(S_{a_j}))'^2 + \sum_j N(j)(\log(S_{d_j}))'^2] & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & & \\ 0 & 0 & & & ((X_w^t \Sigma_J^{-1} X_w)^{-1}) \end{pmatrix}, \quad (23)$$

where $A = (1/2\sigma^2)[\log(S_{a_j})' + \sum_j N(j)\log(S_{d_j})']$ and $\Delta = (N[(\log(S_{a_j})')^2 + \sum_j N(j)(\log(S_{d_j})')^2]/4\sigma^4) - A^2$.

The Cramér–Rao lower bounds are then given by the diagonal elements of I^{-1} :

$$\text{Var}(\hat{\gamma}) \geq \frac{2N}{N[(\log(S_{a_j})')^2 + \sum_j N(j)(\log(S_{d_j})')^2] - [\log(S_{a_j})' + \sum_j N(j)\log(S_{d_j})']^2}$$

$$\begin{aligned} \text{Var}(\hat{\sigma}^2/\sigma^2) &\geq \frac{2[(\log(S_{a_j})')^2 + \sum_j N(j)(\log(S_{d_j})')^2]}{N[(\log(S_{a_j})')^2 + \sum_j N(j)(\log(S_{d_j})')^2] - [\log(S_{a_j})' + \sum_j N(j)\log(S_{d_j})']^2} \quad (24) \end{aligned}$$

$$\text{Var}(\hat{\beta}) \geq \text{diag}((X_w^T \Sigma_J^{-1} X_w)^{-1}).$$

In this equation, the symbol $'$ stands for the first derivative with respect to γ . The bound on the Hurst exponent H can be directly deduced from Eq. (24) ($\text{Var}(\hat{H}) = \text{Var}(\hat{\gamma}/4)$) and clearly this bound depends only on H .

MONTE CARLO SIMULATIONS

In this section, we present the results of Monte Carlo studies using simulated data to cross-validate wavelet-generalized squares by comparison to ordinary least squares (OLS), i.e., assuming the model errors are serially independent or white, the ARMA-generalized least squares (ARLS), i.e., assuming a short-memory AR(q) model for the errors. The order of the AR process was selected in an unsupervised way for each realization by minimizing the Schwarz Bayesian criterion (SBC) (Schwarz, 1978).

Simulated Data

Zero-mean periodic signals of various frequencies were generally embedded in a realization of stationary Gaussian $1/f$ noise, i.e., an ARFIMA or fGn process. Signals were simulated as three different frequencies by convolving square waves with a Poisson function ($\lambda = 4$). All simulated series had $N = 256$, $J = 8$. The following parameters were varied in simulated data to investigate their impact on the performance of the three estimators: the long-memory exponent H characterizing the $1/f$ process; the standard deviation σ ; the frequency of signal embedded in the noise; and the amplitude of the signal or the signal-to-noise ratio (SNR) in decibels:

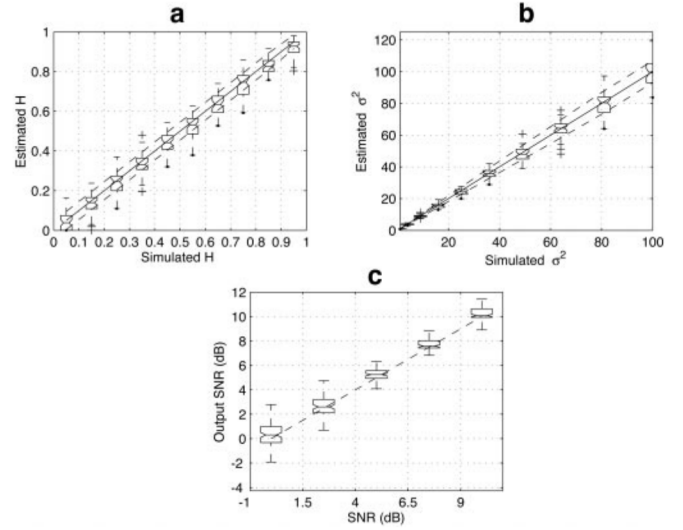


FIG. 1. (a, b) Boxplots of $1/f$ noise parameter H and variance σ^2 estimated by wavelet-generalized least squares (WLS) in simulated time series. In both plots, the solid line $y = x$ denotes perfect correspondence between simulated and estimated parameters, and the dashed lines on either side of it denote \pm approximate CRLB on variances of the parameters. (c) Boxplot of the SNR estimated by WLS (in decibels) as a function of the simulated SNR.

$$\text{SNR} = 20 \log_{10} \left(\frac{\text{Amplitude}}{\sigma_{\text{noise}}} \right). \quad (25)$$

To synthesize $1/f$ noise we could have used a wavelet-based algorithm (Wornell, 1996; McCoy and Walden, 1996; Deriche and Tewfik, 1993). However, we preferred a Fourier-based method (Davies and Harte, 1987) to synthesize both the fGn and ARFIMA processes, on the grounds that using a wavelet-based model for simulation might immediately bias the comparison of estimators in favor of wavelet-generalized least squares.

In experiments where the focus was on the impact of one parameter in particular, the values of all other parameters in the simulation were allowed to vary randomly.

Effects of Varying H , σ^2 , and SNR

For each of 10 different values of the Hurst exponent, and each of 10 different amounts of error variance, 100 time series were realized. The results of estimating H and σ^2 by WLS in these data are shown in Figs. 1a and 1b. We can see that the performance of the estimator is encouraging. The estimated parameters are unbiased over the whole range of values simulated and the variances are not much greater than their Cramér–Rao bounds. It is of course expected that the parameter variances should be somewhat greater than their lower bounds simply due to the finite length of the simulated time series. We also investigated the relationship be-

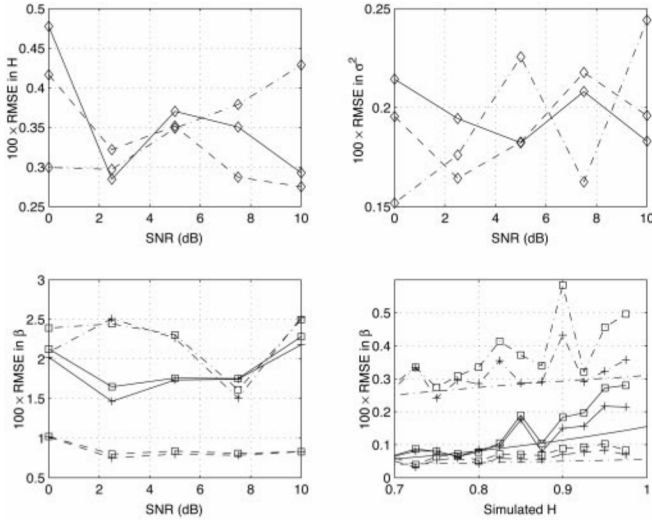


FIG. 2. Root mean square (RMS) errors in estimation of model and noise parameters by wavelet-generalized least squares (WLS) and ordinary least squares (OLS) in simulated data with different frequencies of periodic signal. Top left: RMS error in H versus SNR for three different signal frequencies: solid line—period length = 64; dash-dotted line—period length = 32; dashed line—period length = 16. Top right: RMS error in σ^2 versus SNR for three different signal frequencies—line types coding for frequency as before. Bottom left: RMS error in β versus SNR for three different signal frequencies and two estimators, WLS and OLS. Line types code for signal frequency as before; point markers code for estimator: (+) WLS, (\square) OLS. Bottom right: RMS error in β versus H for three different signal frequencies and two estimators, coded by line type and point markers as before. The lines without any marker correspond with the predicted approximate CRLB given by Eq. (24).

tween estimated and simulated SNR, allowing H to vary randomly. As shown in Fig. 1c, there was very good agreement and the variability tended to decrease as simulated SNR became greater, which is intuitively acceptable.

Effects of Varying Frequency of Signal

We also investigated the effects of varying the frequency of the embedded signal on the efficiency of WLS estimation of H and σ^2 . Efficiency was quantified in terms of the root mean square error between simulated and estimated values of H and σ^2 . Three different frequencies were studied, corresponding to period lengths of 16, 32, or 64 time points. As shown in Fig. 2, there was little evidence for any systematic effects of signal frequency on efficiency of estimation.

Comparison to OLS and ARLS

The alternative estimators do not provide estimates of H but all three methods can be compared in terms of efficiency of estimation of the linear model parameter vector β . These results are shown in Fig. 2 (bottom left). It can be seen that WLS is more efficient than OLS for all three signal frequencies and over the full

range of SNR; although, as expected (Beran, 1994), the advantages of WLS are most salient for the lowest frequency design. As a consequence of the finite length of the observations, these errors are also bounded by the theoretical limit, which depends on the structure of the design matrix (see Eq. (24)).

Differential efficiency of the estimator of β was also assessed in relation to variability in the fractal noise parameter H . As shown in Fig. 2 (bottom right), and as hypothetically expected, the relative merits of the WLS estimator become increasingly salient as H becomes large.

Comparison to Generalized Least Squares (GLS)

In the absence of prior knowledge about the form of the full error covariance matrix Σ , generalized least squares (GLS) is computationally expensive and potentially unstable numerically. However, GLS is the ML estimator in the time domain and we have therefore also compared WLS to GLS, using prior knowledge of H and σ in simulated data to specify the error covariance matrix Σ . This comparison provides a measure of any bias in estimation by WLS due to our assumption that the off-diagonal elements of the error covariance matrix are zero.

A periodic input function (period = 32) with $\beta = 0, 0.2, 0.6, 1, 1.4$ (SNR ≤ 1.5 dB) was added to each of 100 ARFIMA model series generated with $H = 0, 0.1, 0.3, 0.5, 0.7, 0.99$. The root mean square error (RMSE) and bias of the WLS and time-domain ML estimators were calculated; see Table 2.²

In these simulations, there is only slight difference between WLS and GLS in terms of empirical bias and RMSE. There was not a great deal of difference between efficiency of the WLS estimator and the exact Cramér–Rao lower bounds. Furthermore, there is no evidence for any influence of the input signal amplitude on the performance of the WLS estimator. These empirical findings stand in favor of WLS as an unbiased and optimally efficient estimator for regression models with long-range dependent errors. More specifically, they indicate that our assumption of zero off-diagonal elements in the error covariance matrix does not importantly bias WLS estimates with respect to the performance of the GLS estimator which does not share that assumption.

² Monte Carlo experiments using Daubechies wavelets with $R = 4$ vanishing moments (D4) are presented here. To assess the influence of the choice of wavelet on WLS estimation, the DWT of each simulated series was also calculated using the Daubechies wavelet family with increasing regularity $R = \{1, 2, 4, 6, 8, 10\}$ and the Haar wavelet, $R = 1$. Tables reporting the empirical bias and RMSE of WLS based on the other Daubechies wavelets can be obtained upon request from M.J.F. In brief, there was no differential bias of WLS as a function of R .

TABLE 1

Expected and Mean Observed False-Positive Fractions for Wavelet-Generalized Least Squares (WLS), Ordinary Least Squares (OLS), and ARMA-Generalized Least Squares (ARLS) Estimators Applied to Analysis of 7 “Null” fMRI Datasets

Expected	WLS	OLS	ARLS
0.0010	0.0008 (0.0004)	0.0035 (0.0027)	0.0015 (0.0014)
0.0020	0.0017 (0.0007)	0.0069 (0.0053)	0.0030 (0.0027)
0.0030	0.0025 (0.0010)	0.0103 (0.0079)	0.0045 (0.0041)
0.0040	0.0033 (0.0014)	0.0136 (0.0105)	0.0060 (0.0054)
0.0050	0.0042 (0.0017)	0.0170 (0.0131)	0.0075 (0.0067)
0.0060	0.0050 (0.0021)	0.0203 (0.0156)	0.0090 (0.0080)
0.0070	0.0058 (0.0024)	0.0236 (0.0181)	0.0105 (0.0093)
0.0080	0.0067 (0.0027)	0.0269 (0.0206)	0.0119 (0.0106)
0.0090	0.0075 (0.0030)	0.0301 (0.0230)	0.0134 (0.0118)
0.0100	0.0083 (0.0034)	0.0333 (0.0254)	0.0149 (0.0131)

Note. Standard deviations are in parentheses.

Inferential Aspects of Linear Model Estimation

Normality of signal parameter estimates. Another important characteristic of an estimator is the distri-

bution of its estimate. In the context of linear models, it is well known that the OLS estimator tends to underestimate the error variance when the contaminating process presents a nonnegligible serial correlation (Box and Jenkins, 1976; Darlington, 1990; Pollock, 1999). As a consequence, standardized statistics constructed as the ratio of estimated linear model parameters to their standard errors will be overestimated resulting in an inflated or uncontrolled probability of type 1 error.

In these simulated data, we can predict that $s_{\beta} = (\hat{\beta} - \beta_{\text{true}})/\sigma_{\beta}$ should have a standard Normal distribution. We checked this assumption, estimating σ_{β} by $\hat{\sigma}(X'X)^{-1/2}$ for OLS and ARLS and by Eq. (24) for WLS.

As shown in Fig. 3, the WLS estimates are graphically indistinguishable from standard Normal for all three signal frequencies. However, OLS estimates are clearly not standard Normal, especially when the signal frequency is lowest. The distribution of ARLS estimates also departs from Normal in the tails, especially in the lowest frequency design, although it performs considerably better than OLS.

TABLE 2

Bias and Root Mean Square Error of Wavelet-Generalized Least Squares (WLS) and GLS

		Bias					
	β	H = 0	H = 0.1	H = 0.3	H = 0.5	H = 0.7	H = 0.99
GLS	0	-0.0035	0.0010	0.0052	-0.0036	0.0097	-0.0028
	0.2	-0.0020	-0.0010	0.0033	-0.0209	-0.0098	0.0082
	0.6	-0.0007	0.0020	0.0031	-0.0056	-0.0089	-0.0148
	1	0.0004	-0.0006	-0.0047	-0.0127	0.0039	0.0048
	1.4	0.0029	0.0019	0.0025	0.0063	-0.0135	-0.0203
WLS	0	-0.0019	-0.0002	0.0061	-0.0038	0.0080	-0.0073
	0.2	-0.0012	-0.0016	0.0036	-0.0207	-0.0125	0.0060
	0.6	-0.0004	0.0011	0.0033	-0.0055	-0.0079	-0.0143
	1	0.0002	-0.0015	-0.0043	-0.0127	0.0032	0.0013
	1.4	0.0025	0.0023	0.0037	0.0062	-0.0144	-0.0175
		Root mean square error					
Exact Cramér–Rao lower bound		0.0330	0.0387	0.0534	0.0735	0.1008	0.1578
	β	H = 0	H = 0.1	H = 0.3	H = 0.5	H = 0.7	H = 0.99
GLS	0	0.0331	0.0432	0.0610	0.0765	0.0935	0.1496
	0.2	0.0320	0.0379	0.0498	0.0779	0.1126	0.1656
	0.6	0.0309	0.0436	0.0571	0.0685	0.0993	0.1436
	1	0.0326	0.0353	0.0541	0.0870	0.0844	0.1415
	1.4	0.0326	0.0421	0.0591	0.0795	0.1024	0.1672
WLS	0	0.0337	0.0451	0.0614	0.0765	0.0943	0.1533
	0.2	0.0323	0.0392	0.0508	0.0779	0.1132	0.1716
	0.6	0.0320	0.0442	0.0574	0.0687	0.1010	0.1501
	1	0.0353	0.0362	0.0544	0.0870	0.0843	0.1463
	1.4	0.0338	0.0427	0.0587	0.0799	0.1017	0.1681

Note. Bias and root mean square of WLS using a Daubechies wavelet with R = 4 vanishing moments, is compared by analysis of simulated data to the performance of generalized least squares (GLS), which is the maximum likelihood estimator in the time domain. The specification of the full error covariance matrix Σ for GLS is informed by prior knowledge of H and σ in these simulated data. These data suggest that the WLS assumption of zero off-diagonal elements in Σ does not cause any differential bias or loss of efficiency in estimation compared to the ML estimator in the time domain.

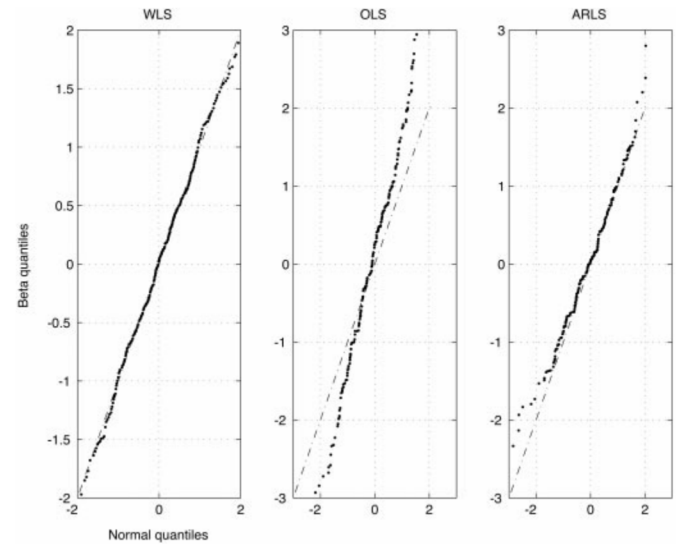
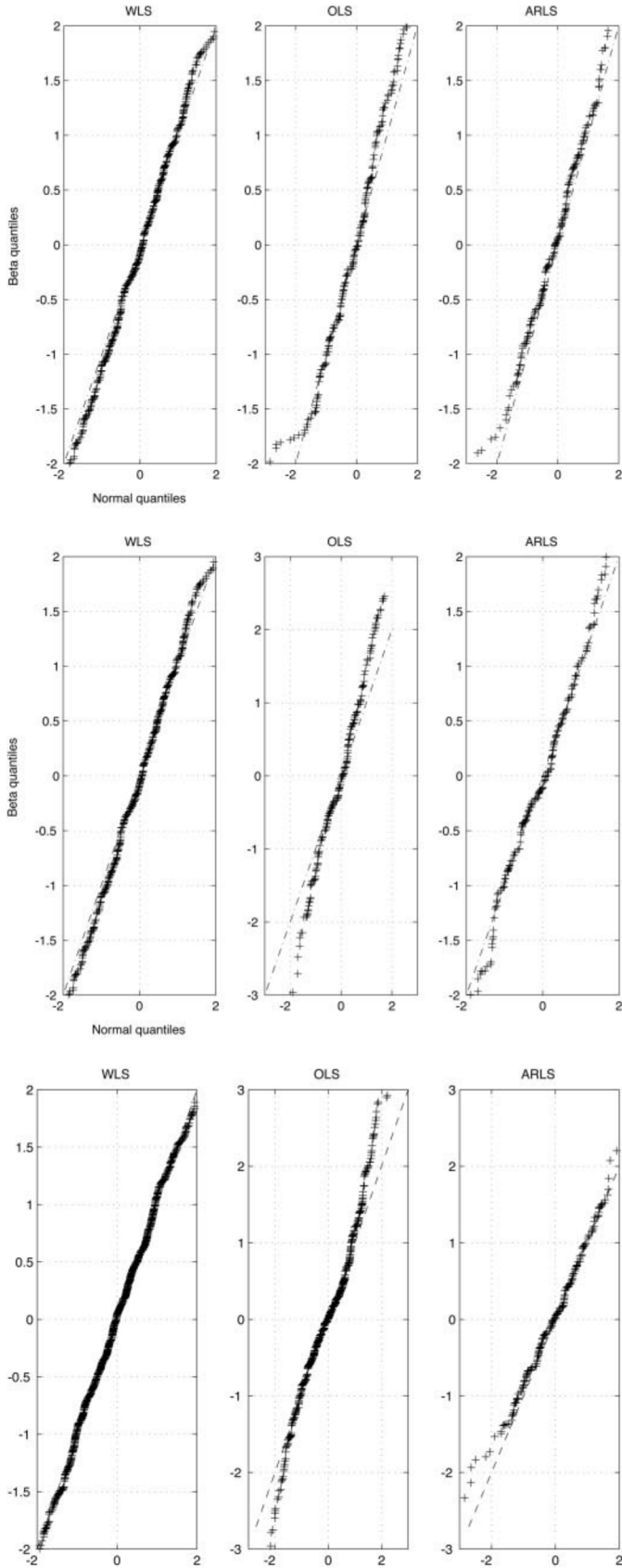


FIG. 4. Normal quantile plots of linear model parameters estimated by wavelet-generalized least squares (WLS), ordinary least squares (OLS), and autoregressive-based least squares (ARLS) in simulated data with signal period = 32. Columns, left to right: WLS, OLS, ARLS. H was randomly varied in the range $[0.9, 0.98]$. In all panels, the dash-dotted line is the line of identity. σ^2 and SNR are randomly generated.

We also considered the relative merits of the estimators as the value of H was manipulated at a given design frequency. Figure 4 shows the results for high values of the Hurst exponent $H \in [0.9, 0.98]$, in the case of the intermediate period length 32. These results can be compared to the middle row in Fig. 3. As one might predict intuitively, the Normality of parameter estimates obtained by OLS and ARLS is particularly compromised when H is large; although WLS continues to yield standard Normal parameter estimates under those circumstances. These results suggest that the WLS estimator is robust vis-à-vis H , despite a slight increase in the estimate error (as shown in Fig. 2); whereas numerical instability problems were posed for the short-memory AR(q) model, which becomes inadequate, when H is large.

Type 1 error calibration. To assess the relative performance of the estimators in terms of type 1 error control, we simulated a set of data in which the signal amplitude was set to zero. We then tested the null hypothesis $H_0: \beta = 0$ over a range of critical values of the standard Normal distribution corresponding to

FIG. 3. Normal quantile plots of linear model parameters estimated by wavelet-generalized least squares (WLS), ordinary least squares (OLS), and autoregressive-based least squares (ARLS) in simulated data with three different signal frequencies. Columns, left to right: WLS, OLS, ARLS. Rows, top to bottom: signal period = 16, signal period = 32, signal period = 64. H was randomly varied in the range $[0, 0.9]$. In all panels, the dash-dotted line is the line of identity.

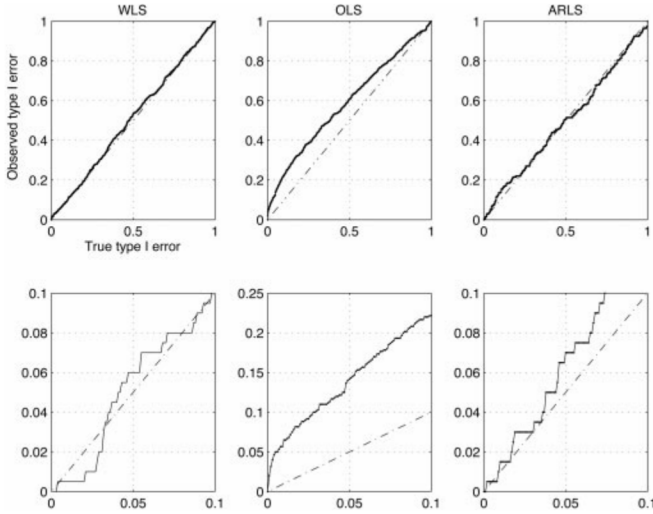


FIG. 5. Type 1 error calibration curves for wavelet-generalized least squares (WLS, left column), ordinary least squares (OLS, middle column), and autoregressive-based least squares (ARLS, right column). Top row: observed versus predicted type 1 error probabilities in the range $[0, 1]$. Bottom row: observed versus predicted type 1 error probabilities in the range $[0, 0.1]$. In all panels, the dash-dotted line is the line of identity.

probabilities of type 1 error α in the range $[0, 0.1]$. Nominal type 1 error control is indicated by the observed number of positive tests (over $M = 500$ realizations) being less than or equal to the expected number $= \alpha M$.

First, to assess the global ability of the estimators in controlling the type 1 error, the noise parameters were allowed to vary randomly. The results are summarized in Fig. 5 for a wide range of α . In the first row are displayed the observed vs simulated type 1 errors for the whole interval $[0, 1]$, while in the second row is a zoom on the range $[0, 0.1]$. It is clear from Fig. 5 that type 1 error control by WLS is excellent over the full range of probability thresholds and that type 1 error control by OLS is unacceptably poor: the number of observed positive tests by OLS, when the nominal double-sided $\alpha = 0.05$, is approximately three times its predicted level. The performance of the ARLS estimator is intermediate. In the second experiment, the value of H is fixed. As shown in Fig. 6, and as might be expected intuitively, the performance of OLS is sensitive to the size of H and type 1 error control becomes worse as H becomes large. It is also shown that ARLS is rather more likely to yield a greater than predicted number of positive tests than WLS, especially at medium or large values of H .

APPLICATION TO REAL fMRI DATA

Neurophysiological time series measured using functional magnetic resonance imaging (fMRI) can be analyzed by linear regression in the context of autocorre-

lated errors (Bullmore et al., 1996). The spectral structure of fMRI noise is $1/f$ -like (Zarahn et al., 1997) and this residual or endogenous autocorrelation is probably generated by both physical and physiological mechanisms. We have previously investigated the noise structure of fMRI data (Fadili et al., 2000) and exploited the whitening property of the discrete wavelet transform in this context to make nonparametric statistical inference by resampling in the wavelet domain (Bullmore et al., 2001). This latter work illustrated the superiority of the wavelet whitening approach over short-memory (first and third order) AR models in dealing with the potentially complex autocorrelational structure of fMRI noise.

Experimental Designs and Data Acquisition

Gradient-echo echoplanar imaging data were acquired as follows:

Null (1.5T and 3T). Seven normal volunteers were studied (5 at 1.5 Tesla (T) and 2 at 3T) while they lay quietly in the scanner with their eyes closed (6 min for the 1.5T data and 17 min for the 3T). For the 1.5 T datasets, 68 T_2^* -weighted images were acquired at each of 26 contiguous slices of data in an oblique axial plane using the GE LX EchoSpeed system (General Electric, Milwaukee, WI) at CHU of Caen, France: time to echo (TE) 60 ms, TR = 5 s, inplane resolution 3.5×3.5 mm, slice thickness = 5 mm, number of excitations = 1. For the 3T datasets, 512 T_2^* -weighted images were acquired at 17 contiguous slices in an oblique plane using

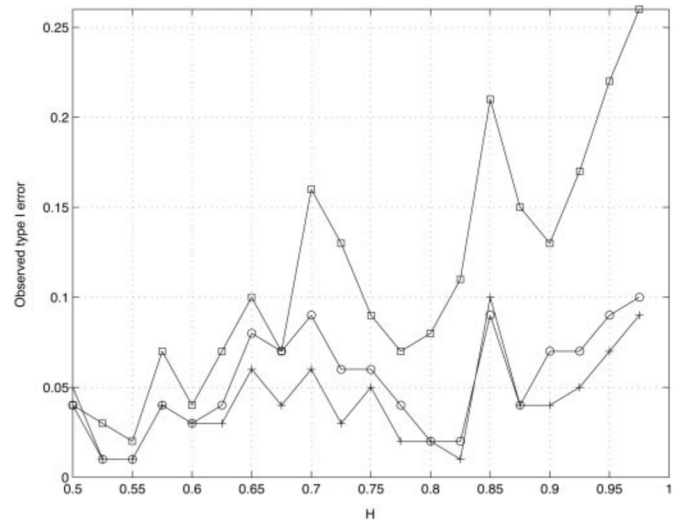


FIG. 6. Observed type 1 error rate versus H for three estimators: wavelet-generalized least squares (WLS, \times), autoregressive-based least squares (ARLS, \circ), and ordinary least squares (OLS, \square). The predicted type 1 error probability is fixed at 0.05. It is clear that type 1 error control by the OLS estimator is increasingly poor as the noise parameter H is increased in the simulated data; and that ARLS is consistently more likely than WLS to yield an excess number of positive tests.

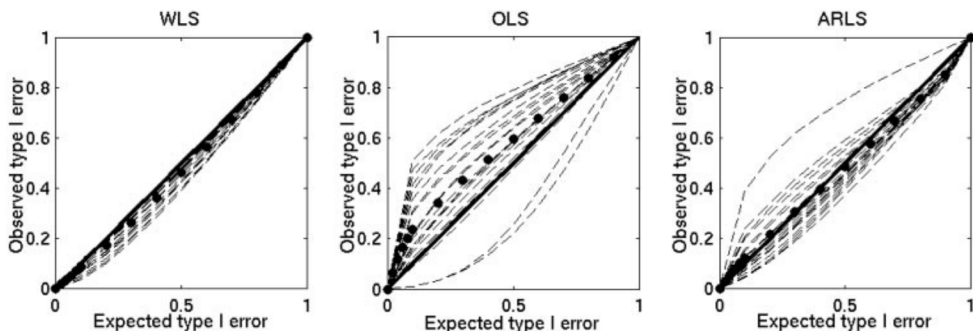


FIG. 7. Type 1 error calibration curves for wavelet and time domain estimation schemes based on WLS.

the Bruker system (Bruker Medical, Etlinger, Germany) at the Wolfson Brain Imaging Centre (WBIC), Cambridge, UK: TE = 40 ms, TR = 2 s, in-plane resolution 1.72×3.4375 mm, slice thickness = 5 mm, number of excitations = 1.

Visual stimulation (3T ER). We studied a single male subject during a discrete-trial or event-related experiment in which a black-and-white checkerboard visual stimulus was flashed for 500 ms every 16.5 s over the course of 4 min, 28 s. Each time the subject saw the checkerboard flash he pressed a button with his right hand. During this experiment, 17 slices of gradient echo echoplanar imaging data were acquired using a 3T Bruker MRI system with acquisition parameters as above. This simple experiment is expected to activate areas of the brain (occipital cortex) that are important in visual perception.

Time Domain Models for Design Matrix

For all of the null experimental data the same general approach to regression model specification was adopted. An N -length boxcar vector was constructed to indicate which images were acquired during presentation of an activation condition and which were acquired during presentation of a baseline condition. This vector was convolved with a Poisson kernel, parameterized by $\lambda = 4$ s. This Poisson-convolved input function was combined with a unitary constant column vector to form the $(N \times 2)$ design matrix X . Again, to assess the influence of the experimental design complexity, high (0.032 Hz), intermediate (0.016 Hz), and low (0.008 Hz) frequency boxcar functions were modelled. The design matrix was fit after motion correction of each observed fMRI time series; we did not initially detrend the time series in any way.

Results

Functional MRI: Null data. Each of the 7 null datasets was analyzed by fitting a periodic input function at each of the 3 frequencies to each of the M intracerebral voxels. For a valid test of size α , the number of positive voxels observed when the null hypothesis (of zero pe-

riodic trend) is true, as presumably it is in these data, should be less than or equal to the expected number of positive voxels = αM . The results are shown graphically in Fig. 7. Table 1 shows the expected and the mean observed type 1 probability errors with their standard deviations across all datasets. These fractions are shown for a low probability threshold range $[0, 0.01]$. The WLS is clearly superior to the OLS and the ARLS estimators. The OLS is not valid for almost all images and performs particularly badly at low design frequencies. Moreover, there is more variability between images in the expected type 1 probability error after ARLS prewhitening and the observed number exceeds the expected number of positive tests in 8 images. This variability increases as the input function frequency decreases, which shows the sensitivity of AR prewhitening to the complexity of the design matrix. The overall poor performance of the OLS and ARLS methods provides circumstantial support for the validity of a $1/f$ power law noise model for fMRI. Again, as for the simulation studies, the type 1 error calibration curves provided by the WLS estimator are valid for all images and there is no evidence for sensitivity of WLS to the input function frequency. In many images with intermediate and high frequency input function, the ARLS prewhitening and WLS methods are evidently somewhat over-conservative, which is consistent with our previous findings with a wavelet domain resampling scheme (Bullmore et al., 2001).

Functional MRI: Activation Mapping

As shown in Fig. 8, the brain areas which show significantly large values of the model parameter vector β estimated by OLS are very similar to the areas of functional activation identified by thresholding the same parameters estimated by WLS. In fact, the WLS estimator appears to provide a somewhat fuller or more sensitive characterization of the cerebral response. WLS additionally estimates σ^2 at each voxel of the image, demonstrating that the noise variance is not homogeneously distributed but tends to be larger in outer (cortical) areas of the brain. WLS also estimates

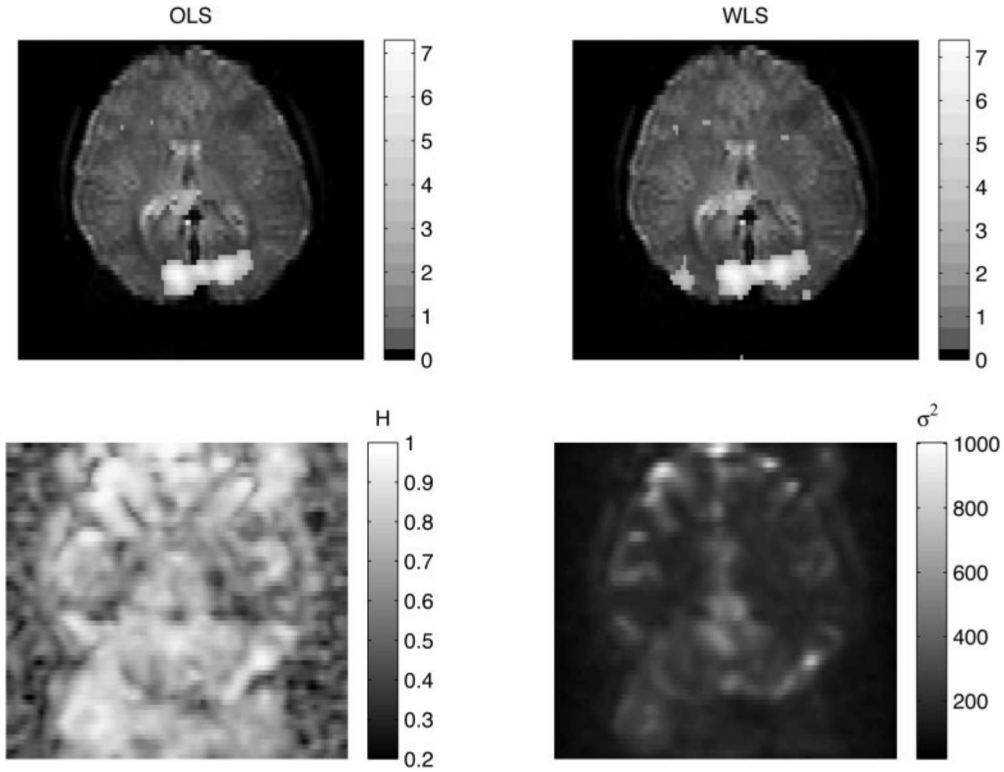


FIG. 8. Statistical maps generated by linear modelling of fMRI time series measured during an event-related visual experiment. Top left: the linear model parameter vector β estimated by ordinary least squares (OLS) is mapped at each voxel of a single slice of imaging data and thresholded. Areas indicated by large parameter values are located in occipital cortex, at the back of the brain. Top right: the linear model parameter vector estimated by WLS is identically mapped and thresholded. The extent of occipital activation is somewhat greater than in the OLS map. Bottom right: the noise variance is mapped at each voxel by WLS. Bottom left: the fractal noise parameter H is mapped at each voxel by WLS. Note that the value of H is large, i.e., >0.5 , at most voxels in the image representing brain tissue.

H at each voxel, demonstrating that fMRI time series in the brain may be contaminated by long-memory processes with large fractal noise parameters, i.e., $\hat{H} \geq 0.5$. These preliminary observations suggest that WLS is feasible for fMRI data analysis and may offer considerable advantages over alternative estimators given the prevalence of long-memory errors in these data.

DISCUSSION

Here we have presented a new linear estimator for signal and noise parameters in the context of long-memory errors. Our approach, which we have called wavelet-generalized least squares (WLS), depends for its validity on some recently established mathematical results concerning the properties of the discrete wavelet transform (DWT). Specifically, it has been shown that the DWT is a near-optimal whitening filter for long-memory or $1/f$ -like noises. To put it another way, the DWT of a $1/f$ process closely approximates the ideal Karhunen–Loève expansion. This result underwrites our key assumption that the covariance matrix of the regression errors can be diagonalized, i.e., the off-diagonal elements can be set to zero, after projection of both

the data and the design matrix into the wavelet domain. Diagonalization simplifies numerical identification of parameter estimates and implies that the WLS estimator is theoretically approximate to the best linear unbiased (BLU) estimator and can provide maximum likelihood estimates of both signal and noise parameters.

It could be argued that these theoretically anticipated and desirable characteristics of the WLS estimator will be compromised if the wavelet expansion of the data does not have ideal whitening properties. Indeed, it is likely that the DWT of a real, finite time series will not be ideal in this respect, at least partly because of artifactual inter-coefficient correlations introduced by boundary correction at the limits of the data. We have therefore used Monte Carlo methods rigorously to evaluate the bias and efficiency of the estimator in analysis of simulated data and have shown, encouragingly, that WLS is unbiased over a wide range of data conditions and its efficiency closely approximates theoretically derived limits. We conclude that the method can be regarded as a good candidate for the best linear unbiased estimator of regression models with $1/f$ errors.

The potential benefits of wavelet-generalized least squares for fMRI data analysis were illustrated by comparison to ordinary least squares and ARMA-generalized least squares. WLS demonstrated consistently superior type 1 error control and also revealed interesting heterogeneities in the spatial distribution of noise properties. Time series recorded in cerebral cortex seem especially liable to complication by large amounts of noise with persistent or long-memory structure ($\hat{H} \geq 0.5$). Our simulations suggest that it is precisely in this context, of highly persistent error autocorrelations, that WLS will most markedly out-perform alternative estimators assuming only a short-memory structure in the errors (ARLS), or simply no autocorrelational structure in the errors at all (OLS). It remains a very interesting question whether the concentration of persistent autocorrelational structures in cortical regions indicates a neurophysiological source for errors of this sort. It is tempting to speculate that persistent fractal noise, of the sort that we have demonstrated by univariate analysis of each voxel, might determine corticocortical correlations identified by multivariate analysis of fMRI data acquired from humans “at rest” (Lowe et al., 2000). In general, the possibility that fMRI “noise” may in fact be informative about the brain probably warrants further investigation and WLS provides a convenient summary of the most important parameters of 1/f-like noises.

Our methods can be compared to work by Craigmile and colleagues (Craigmile et al., 2000), using the DWT to estimate both trend and long-memory noise components of a signal. Detail coefficients at the boundaries of each level and the scale coefficients of the DWT encode information about trends and the within-level boundary-independent wavelet coefficients can be regarded as independent or an auto-regressive process. Wavelet-generalized least squares differs from this approach in that we are using the parameters of a linear model (not a subset of wavelet coefficients) to account for experimentally determined and other trends in the data. However, one approach that may be interesting to consider in future is the use of partly linear models (PLM) of the form

$$y_t = X_t\beta + g(t) + \epsilon_t. \quad (26)$$

Here β is a vector of unknown parameters, $g(t)$ is an arbitrary and unknown (possibly nonlinear) function over \mathbb{R} , X_t are vectors of explanatory variables that are either random i.i.d. or fixed design points, and ϵ_t are zero-mean long-memory error processes with finite variance. PLMs are semiparametric models since they contain both parametric X_t and nonparametric $g(t)$ components. They allow easier interpretation of each variable and may be preferred to a completely nonparametric approach because of the well-known curse

of dimensionality. Partly linear models are also more flexible than the standard linear models, since they combine both parametric and nonparametric components, which may be useful if it is believed that the response depends linearly on some covariates but nonlinearly on others. Wavelets are natural tools for estimating such models and we are currently exploring this application in the context of fMRI.

It is interesting also to compare wavelet-generalized least squares to the wavelet resampling scheme we have previously proposed for computational (nonparametric) inference on fMRI time series (Bullmore et al., 2001). Both WLS and wavelet resampling exploit the same key mathematical result—namely, that the DWT is a good whitening filter for 1/f processes. However, the resampling scheme uses this property principally to justify exchangeability of wavelet coefficients under the null hypothesis; see (Nicholls and Holmes, 2001) for an introduction to exchangeability in relation to nonparametric inference on functional neuroimaging data. This allows us to resample fMRI time series while preserving their unknown autocorrelational structure, so that test statistics estimated in the observed data are not biased with respect to statistics estimated by fitting the same regression model to data reconstituted in the time domain after wavelet resampling. And this in turn allows us to construct a valid test of the null hypothesis by comparing the observed statistics to the critical values of the permutation distribution ascertained by repeated wavelet resampling. However, the estimator in this resampling scheme is ordinary least squares, which will be less than optimally efficient in the context of long-memory errors, whether observed or resampled. Wavelet-generalized least squares can also support valid inference but it has the crucial advantage of greater efficiency, i.e., less variability in estimation of model parameters. It also provides a quantitative summary of the noise properties at each voxel, rather than “blindly” reproducing these properties by resampling of wavelet coefficients.

From a theoretical perspective, the method could be refined in respect of at least two underlying assumptions. First, we have assumed that the long-memory process in the data is uncontaminated by Gaussian (white) noise. Extension to the case of stationary 1/f processes contaminated with Gaussian white noise could also be handled using a more complicated Expectation–Maximization (EM) type algorithm to partition the total error variance into white and colored components (Wornell and Oppenheim, 1992a; Wornell, 1996). Second, we have assumed that the long-memory parameter H is time invariant. There are a number of ways in which our approach could be developed to obviate this assumption. For example, assuming the process is at least locally stationary, one can simply apply the proposed estimator using the classical fixed-window-length segmentation technique. Finally, it is

worth emphasizing that the goodness of DWT as a whitening filter for real data is considerably influenced by the way in which wavelet coefficients are estimated at the boundaries of the time series. We have here used a simple convolutional method of periodic boundary correction but there are nonconvolutional filters one can apply to estimate coefficients at the edges of each level of detail, and these may offer worthwhile improvements in performance of the estimator by diminishing artifactual inter-coefficient correlations due to periodic boundary correction. In particular, the slightly over-conservative nature of tests based on WLS estimates may be attributable to artifactual inter-coefficient correlations and consequently treatable by use of a nonconvolutional boundary correction algorithm.

CONCLUSION

We conclude that wavelet-generalized least squares is an optimal estimator of linear model and fractal noise parameters that may be of widespread utility given the natural ubiquity of long-memory errors.

ACKNOWLEDGMENTS

We thank Dr. Matthew Brett of the MRC Cognition and Brain Sciences Unit, Cambridge, and colleagues at the Wolfson Brain Imaging Centre, Addenbrooke's Hospital, University of Cambridge for their help in acquiring the fMRI data. We also thank two anonymous reviewers for constructive comments on an earlier draft of this paper.

REFERENCES

Azzalini, A. 1996. *Statistical Inference Based on the Likelihood*. London, Chapman & Hall.

Beran, J. 1994. *Statistics for Long-Memory Processes*. London, Chapman & Hall.

Box, G. E. P., and Jenkins, G. M. 1976. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A. S., Mellers, J. D. C., Howard, R., and Sham, P. 1996. Statistical methods of estimation and inference for functional mr image analysis. *Magn. Reson. Med.* 35: 261–277.

Bullmore, E. T., Long, C., Suckling, J., Fadili, M. J., Calvert, G., Zelaya, F., Carpenter, T. A., and Brammer, M. J. 2001. Color noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Hum. Brain Mapp.* 12(2): 61–78.

Cassandro, M., and Jona-Lasinio, G. 1978. Critical behavior and probability theory. *Adv. Phys.* 27: 913–941.

Chui, C. 1994. *Wavelets: Theory, Algorithms, and Applications*. Academic Press, San Diego/London.

Craigmille, P., Percival, D., and Guttrop, P. 2000. *Wavelet-Based Parameter Estimation for Trend Contaminated Fractionally Differenced Processes* Technical Report, NRCSE, Department of Statistics, University of Washington.

Darlington, R. 1990. *Regression and Linear Models*. McGraw-Hill, New York/London.

Daubechies, I. 1992. *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA.

Davies, R., and Harte, D. 1987. Tests for hurst effect. *Biometrika* 74: 95–102.

Deriche, M., and Tewfik, A. 1993. Maximum likelihood estimation of the parameters of discrete fractionally differenced gaussian noise process. *IEEE Trans. Signal Proc* 41(10): 2977–2989.

Fadili, M., Ruan, S., Bloyet, D., and Mazoyer, B. 2000. A multi-step unsupervised fuzzy clustering analysis of fmri time series. *Hum. Brain Mapp.* 10(4): 160–178.

Flandrin, P. 1992. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Trans. Inf. Theory* 38: 910–917.

Foufoula-Georgiou, E., and Kumar, P. 1994. *Wavelets in Geophysics*. Academic Press, San Diego/London.

Goldberger, A., Rigney, D., and West, B. 1990. Chaos and fractals in human physiology. *Sci. Am.* 46: 42–49.

Granger, C., and Joyeux, R. 1980. An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* 1: 15–29.

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. 1986. *Robust Statistics. The Approach Based on Influence Functions*. Wiley, New York.

Huber, P. 1981. *Robust Statistics*. Wiley, New York.

Hurst, H. 1951. Long-term storage capacity of reservoirs. *Trans. Am. Soc. Civil Eng.* 116: 770–799.

Jensen, J. 1994. *Wavelet Analysis of Fractionally Integrated Processes*. Technical Report, Department of Economics, Washington University.

Jensen, J. 1999. Using wavelets to obtain a consistent ordinary least squares estimator of long memory parameter. *J. Forecast.* 18(1): 17–32.

Jensen, J. 2000. An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *J. Econ. Dynam. Control* 24: 361–387.

Jensen, J., and Whitcher, B. 2000. *Time-Varying Long-Memory in Volatility: Detection and Estimation with Wavelets*. Technical Report, Department of Economics, University of Missouri.

Jones, R. 1993. *Longitudinal Data with Serial Correlation: A State-Space Approach*. Chapman & Hall, London.

Koul, H. 1993. Asymptotics of r -, md - and lad -estimators in linear regression models with long range dependent errors. *Prob. Theory Related Fields* 95: 535–553.

Koul, H. 1994. Regression quantiles and related processes under long range dependent errors. *J. Multivariate Anal.* 51: 318–337.

Krueger, W. M., Jost, S. D., Rossi, K., and Azen, U. 1996. On synthesizing discrete fractional brownian motion with applications to image processing. *Graph. Mod. Imag. Proc.* 58: 334–344.

Lowe, M., Dzemidzic, M., Lurito, J., Mathews, V., and Phillips, M. 2000. Correlations in low-frequency bold fluctuations reflect cortico-cortical connections. *Neuroimage* 12: 582–587.

Lundahl, T., Ohley, W., Kay, S., and Siffert, R. 1986. Fractional brownian motion: A maximum likelihood estimator and its application to image texture. *IEEE Trans. Med. Imag.* 5(3): 152–161.

Mallat, S. G. 1989. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI* 11(7): 674–693.

Mallat, S. G. 1998. *A Wavelet Tour of Signal Processing*. Academic Press.

Mandelbrot, B. B., and van Ness, J. W. 1968. Fractional brownian motion, fractional noises and applications. *SIAM Rev.* 10(4): 422–437.

McCoy, E. J. 1994. *Some New Statistical Approaches to the Analysis of Long-Memory Processes*. Unpublished doctoral dissertation, Imperial College.

- McCoy, E. J., and Walden, A. T. 1996. Wavelet analysis and synthesis of stationary long-memory processes. *J. Comp. Graph. Stat.* 5: 1–31.
- Nicholls, T., and Holmes, A. 2001. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Mapp.*, in press.
- Percival, D. B. 1992. Simulating gaussian random processes with specified spectra. *Comp. Sci. Stat.* 24: 534–538.
- Percival, D. B. 1995. On estimation of the wavelet variance. *Biometrika* 82(3): 619–631.
- Pesquet-Popescu, B. 1999. Statistical properties of the wavelet decomposition of certain coc-gaussian self-similar processes. *Signal Process.* 75: 303–322.
- Pollock, D. S. G. 1999. *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, San Diego/London.
- Press, W. H., Teukolsky, A., Vetterling, W. T., and Flannery, B. 1992. *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge Univ. Press.
- Raz, J., Dickerson, L., and Turetsky, B. 1999. A wavelet packet model of evoked potentials. *Brain Lang.* 66: 61–88.
- Robinson, P. M., and Hidalgo, F. J. 1997. *Time Series Regression with Long Range Dependence*. London, Suntory and Toyota International Centres for Economics and Related Disciplines.
- Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6: 461–464.
- Taqqu, M. S., Teverovsky, V., and Willinger, W. 1995. Estimators for long-range dependence: An empirical study. *Fractals* 3(4): 785–798.
- Tewfik, A. H., and Kim, M. 1992. Correlation structure of the discrete wavelet coefficients of fractional brownian motion. *IEEE Trans. Inf. Theory* 38: 904–909.
- Torrence, C., and Compo, G. P. 1998. A practical guide to wavelet analysis. *Bull. Am. Meteor. Soc.* 79(1): 61–78.
- Van der Ziel, A. 1986. *Noise in Solid State Devices and Circuits*. Wiley, New York/Chichester.
- Voss, R. F. 1979. 1/f (flicker) noise: A brief review. *Proc. Ann. Symp. Freq. Contr.* 40–46.
- Whitcher, B., Guttorp, S. D. P., and Percival, D. B. 2000. Wavelet analysis of covariance with application to atmospheric time series. *J. Geophys. Res.*, in press.
- Whitcher, B., and Jensen, J. M. 2000. Wavelet estimation of a local long-memory parameter. *Explor. Geophys.* 31: 89–98.
- Wickerhauser, W. V. 1994. *Adapted Wavelet Analysis from Theory to Software*. IEEE Press, Wellesley, New York.
- Wornell, G. W. 1990. A karhunen-loeve-like expansion for 1/f process via wavelets. *IEEE Trans. Inf. Theory* 36(4): 859–861.
- Wornell, G. W. 1992. Wavelet-based representations for the 1/f family fractal processes. *Proc. IEEE* 81: 1428–1450.
- Wornell, G. W. 1996. *Signal Processing with Fractals: A Wavelet-Based Approach*. Prentice Hall.
- Wornell, G. W., and Oppenheim, A. V. 1992a. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Proc.* 40(3): 611–623.
- Wornell, G. W., and Oppenheim, A. V. 1992b. Wavelet-based representations for class of self-similar signals with application to modulation. *IEEE Trans. Inf. Theory* 38(2): 785–800.
- Zarahn, E., Aguire, G. K., and D'Esposito, M. 1997. Empirical analyses of bold fmri statistics: I. Spatially unsmoothed data collected under null hypothesis conditions. *NeuroImage* 5: 179–197.