# Wavelet-Based Parameter Estimation for Trend Contaminated Fractionally Differenced Processes

Peter F. Craigmile  Peter Guttorp  Donald B. Percival

NRCSE

Technical Report Series

NRCSE-TRS No. 077

February 4, 2004

# Wavelet-Based Parameter Estimation for Trend Contaminated Fractionally Differenced Processes

**Peter F. Craigmile**[1], **Peter Guttorp**[2] **and Donald B. Percival**[3].

[1] Department of Statistics, 1958 Neil Avenue, The Ohio State University, Columbus, OH 43210.

[2] Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195–4322.

[3] Applied Physics Laboratory, Box 355640, University of Washington, Seattle, WA 98195–5640.

**Abstract:** A common problem in the analysis of time series is how to deal with a possible trend component, which is usually thought of as large scale (or low frequency) variations or patterns in the series that might be best modeled separately from the rest of the series. Trend is often confounded with low frequency stochastic fluctuations, particularly in the case of models such as fractionally differenced (FD) processes, that can account for long memory dependence (slowly decaying autocorrelation) and can be extended to encompass nonstationary processes exhibiting quite significant low frequency components. In this paper we assume a model of polynomial trend plus FD noise and apply the discrete wavelet transform (DWT) to separate a time series into pieces that can be used to estimate both the FD parameters and the trend. The estimation of the FD parameters is based on an approximate maximum likelihood approach that is made possible by the fact that the DWT decorrelates FD processes approximately. We demonstrate our methodology by applying it to a popular climate dataset.

**Keywords:** fractionally differenced processes; discrete wavelet transform; trend; approximate Gaussian likelihood; confidence intervals.

## 1 Introduction

In recent years long memory processes have been used to model natural phenomena in areas such as atmospheric science, geophysics and hydrology. Such processes are characterized by slowly decaying autocorrelations that can be hard to model using standard models such as the autoregressive moving average (ARMA) processes (Box et al. 1994). One common example of a long memory process, the fractionally differenced (FD) process (Granger and Joyeux 1980; Hosking 1981), extends existing (integer) integrated processes. The mathematical tractability of FD processes allows for a varied range of estimation methods.

Considering the FD process singly, a common method of parameter estimation involves calculating the exact likelihood and maximizing with respect to the parameters. Beran (1994) gives a review and evaluation of this. He concludes that the two factors hampering this method in practice are (1) slow computations (particularly for large $N$) and (2) inaccuracies due to a large number of computations (the matrix calculations are $O(N^2)$). Various approximate likelihood methods have been proposed to overcome this (Beran 1994). Some of these methods exploit fast transforms of the data such as the fast Fourier transform (Robinson (1994) – see Moulines and Soulier (2000) for an in depth analysis of this estimator) or wavelet transforms (Wornell (1995) and McCoy and Walden (1996) consider likelihood approaches; Abry et al. (1993), Abry et al. (1995), Veitch and Abry (1999) and Bardet et al. (2000) study least square methods). Vannucci and Corradi (1999) consider Bayesian estimation schemes for long memory processes, and Jensen (2000) examines a wavelet-based likelihood method for the estimation of ARFIMA processes.

There is less literature in the case of such a process contaminated by a trend component. The topic of long range dependence and trends is dealt with in Smith (1989a), Smith (1989b) and Smith (1993). Teverovsky and Taqqu (1997) consider tests for long memory dependence in the presence of two types of trend (shifting means and slowly decaying trend). Percival and Bruce (1998) extend the wavelet-based approximate likelihood estimates of McCoy and Walden (1996) to work in the presence of polynomial trends. Deo and Hurvich (1998) consider linear trends with fractionally integrated errors. Hurvich and Chen (2000) provide a spectral estimation method that can handle some nonstationary ARFIMA processes with a low order polynomial trend component. Giraitis et al. (2001) consider families of tests for long memory, observed in the presence of deterministic trends. Leipus and Viano (2003) extend the work of the previous paper to the case of stochastic trends. Beran and Feng (2002) use variable bandwidth smoothing to estimate such processes with additive trend.

In this paper we consider estimation of the parameters of a trend contaminated FD process using the discrete wavelet transform (DWT). Wavelet transforms of such time series are useful for the following reasons.

1. They approximately decorrelate FD and related processes. We will show the resulting wavelet coefficients form a near independent Gaussian sequence, simplifying the statistics significantly.

2. Wavelets can separate certain nonlinear trends from noise, thus allowing us to analyze dependent time series with a trend.

3. Wavelets have excellent time and frequency localization, which can be useful for investigating local deviations from a statistical model.

We concentrate on polynomial trends, but it is easy using the final two properties to extend these results to other smooth and nonsmooth trends (see Craigmile, Percival, and Guttorp (2003) for further details). By using the wavelet coefficients of the transform in a multivariate Gaussian model (with an assumed simplified correlation structure for the coefficients), we can estimate the parameters using maximum likelihood. In particular we consider two models:

1. White noise wavelet model – we assume the wavelet coefficients are independent both within and between wavelet scales;

2. AR(1) wavelet model. We show that there is often a small lag one autocorrelation between wavelet coefficients on a specific scale. As a model for this, we assume independence between scales, and an AR(1) model within each scale.

In section 2 we define the DWT. We define the fractionally differenced process in section 3 and demonstrate the statistical properties of the DWT of these process (with and without trend) in section

4. We outline the approximate maximum likelihood scheme for the white noise wavelet model in section 5, and for the AR(1) wavelet model in section 6. We provide theory for these estimators, under the assumptions that the approximating models are true, in section 7, and obtain approximate confidence intervals for the model parameters. In section 8, Monte Carlo simulations are used to assess these methods in practice. We also compare our methods to those of Hurvich and Chen (2000). In section 9 we apply our theory to a northern hemisphere temperature dataset obtained from the Climate Research Unit, University of East Anglia, UK. We close with a summary and discussion in section 10 (proofs of the results presented in this paper can be downloaded from `http://www.stat.ohio-state.edu/~pfc/`).

## 2 The discrete wavelet transform

Suppose $\{X_t : t = 0, \ldots, N-1\}$ is our observed time series with $N$ divisible by $2^J$ for some positive integer $J$. For an even positive integer $L$, let $\{h_l : l = 0, \ldots, L-1\}$ denote the Daubechies *wavelet filter* of unit $l_2$ norm. The squared gain function for the wavelet filter is given by

$$\mathcal{H}_{1,L}(f) \quad = \quad 2 \, \sin^L(\pi f) \sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f). \qquad (1)$$

For a particular choice of $L$ there are multiple filters, $\{h_l\}$, that share this squared gain function. This is because the transfer function, $H_{1,L}(f) = \sum_{l=0}^{L-1} h_l e^{-i2\pi f l}$, associated with the squared gain function via $\mathcal{H}_{1,L}(f) = |H_{1,L}(f)|^2$, is not unique. Daubechies (1992) distinguishes between two (of the possible) choices:

1. the extremal phase, D($L$), filters are the ones that exhibit the smallest delay (have maximum cumulative energy) over other choices of scaling filter;

2. the least asymmetric, LA($L$), filters (which differs from the D($L$) filters when $L$=8, 10, ...) are the closest approximations to linear phase filters.

We now define the level $j$ wavelet coefficients in terms of a filtering of our data, $\{X_t\}$ (in practice, we can calculate the wavelet coefficients efficiently using a cascade algorithm rather than filtering the data directly (Mallat 1989; Percival and Walden 2000)). For $L_j = (2^j - 1)(L - 1) + 1$, the level $j$ wavelet filter, $\{h_{j,l} : l = 0, \ldots, L_j - 1\}$ can be defined as the inverse Fourier transform of its transfer function,

$$H_{j,L}(f) \;=\; e^{-i2\pi(L_{j-1}-1)f}\, H_{1,L}(2^{j-1}f) \prod_{k=0}^{j-2} H_{1,L}(1/2 - 2^k f), \tag{2}$$

which in turn defines the $j$th level squared gain function $\mathcal{H}_{j,L}(f) = |H_{j,L}(f)|^2$. This filter is an approximate bandpass filter with a passband given by $|f| \in [1/2^{j+1}, 1/2^j]$. Then for $N_j = N/2^j$, the level $j$ wavelet coefficients are

$$W_{j,k} \;=\; \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(k+1)-1-l \bmod N}, \quad k = 0, \ldots, N_j - 1. \tag{3}$$

These coefficients are associated with changes in averages on scale $2^{j-1}$ and with times spaced $2^j$ apart. The first $B_j = \min(\lceil (L-2)(1-2^{-j}) \rceil, N_j)$ wavelet coefficients are affected by circularly filtering data; that is, the coefficients $\{W_{j,k} : k = 0, \ldots, B_j - 1\}$ combine data from the start and end of the sequence. We refer to these as the boundary coefficients. The remaining $M_j = N_j - B_j$ are unaffected by boundaries and we call them the nonboundary (nb) coefficients, $\{(W_{nb})_{j,k} \equiv W_{j,B_j+k} : j = 1, \ldots, J;\; k = 0, \ldots, M_j - 1\}$. The statistical properties of the boundary coefficients can be quite different from those of the nonboundary coefficients.

# 3 Fractionally differenced processes

The FD process is a long memory dependence model that has become popular in recent years, mainly due to its tractable mathematical properties. The process was originally proposed by Granger and Joyeux (1980) and Hosking (1981) as an extension to ARIMA$(0, d, 0)$ models to allow for fractional values of $d$.

For $d \in [-1/2, 1/2)$ and $\sigma^2 > 0$, the stationary Gaussian process $\{X_t : t \in \mathbb{Z}\}$ is an FD$(d)$ (or ARFIMA$(0, d, 0)$) process if it has a spectrum

$$S(f) \;=\; \sigma^2\, |2\sin(\pi f)|^{-2d}, \qquad |f| \leq 1/2. \tag{4}$$

Here $d$ is known as the fractional difference parameter and $\sigma^2$ is the innovations variance. For $d \in (-1/2, 1/2)$ the process is stationary and invertible, and is a white noise (i.e., uncorrelated) process for $d = 0$. For $d = -1/2$ the process is stationary, but noninvertible. We can extend this model by letting $d \geq 1/2$ in equation (4), and obtain a class of non-stationary processes that become stationary after differencing $\lfloor d + 1/2 \rfloor$ times (Yaglom 1958). Taking differences of the process, we can let $d \leq -1/2$ to obtain a stationary, but noninvertible, process. For $d \in [-1/2, 1/2)$ the covariance sequence can be shown to be (Beran (1994), Hosking (1981) for the $d = -1/2$ case)

$$s_k \;=\; \sigma^2 \, \frac{(-1)^k \, \Gamma(1-2d)}{\Gamma(1-d+k)\,\Gamma(1-d-k)}. \tag{5}$$

Fast simulation of FD processes is possible using the Davies–Harte algorithm (Davies and Harte 1987; Wood and Chan 1994; Craigmile 2003). Further properties along with an extensive history of the FD process can be found in, e.g., Beran (1994) and Samorodnitsky and Taqqu (1994), sections 7.13 and 14.7.

## 4 The nonboundary wavelet coefficients of an FD process

Suppose we observe a realization of a Gaussian FD($d$) process, $\{X_t : t = 0, \ldots, N-1\}$. By the linearity of the DWT the wavelet coefficients of the realization are Gaussian. By definition of the level $j$ wavelet filter, $\{h_{j,l}\}$, $\sum_l h_{j,l} = 0$ (e.g., Percival and Walden (2000), Table 154), and it follows that the wavelet coefficients have zero expectation. We now investigate the second moment properties of the nonboundary (nb) wavelet coefficients. By equation (348a) of Percival and Walden (2000) we have

$$\mathrm{cov}((W_{nb})_{j,k}, (W_{nb})_{j',k'}) \;=\; \int_{-1/2}^{1/2} e^{i2\pi[2^{j'}(k'+1)-2^{j}(k+1)]f} \, H_{j,L}(f)\, H^*_{j',L}(f)\, \sigma^2 (2\sin(\pi f))^{-2d} \, df,$$

where $^*$ denotes the complex conjugation operator. Between scales the DWT acts as a whitening transform for an FD process; that is, for $j \neq j'$, $\mathrm{cov}(W_{j,k}, W_{j',k'}) \approx 0$. This approximation improves with increasing $L$. In fact as $L \to \infty$ the covariance tends to zero, as the next result states.

**Theorem 4.1** *Let $(W_{nb})_{j,k}$ and $(W_{nb})_{j',k'}$ be the level $j$ and $j'$ wavelet coefficients for an FD process, $\{X_t\}$, based upon a wavelet filter $\{h_l\}$ of width $L$. Then* $\mathrm{cov}((W_{nb})_{j,k}, (W_{nb})_{j',k'}) \to 0$ *as $L \to \infty$ when $j \neq j$.*

Thus for sufficiently long wavelet filters we can bound the covariance between different wavelet levels by some small $\epsilon$. In practice we would like to use longer wavelet filters to decorrelate between wavelet scales, but this also has the effect of decreasing the number of nonboundary wavelet coefficients. This results ignores what effect longer wavelet filters will have upon within scale correlations, which is the subject of the next theorem.

**Theorem 4.2** *When $d < (L+1)/2$ the nonboundary wavelet coefficients within a given level $j$ are a portion of a zero mean stationary process with autocovariance sequence given by*

$$\sigma^2 \, s_{j,\tau}(d) = \int_{-1/2}^{1/2} e^{i2\pi f\tau} S_j(f)\, df, \tag{6}$$

*where* $S_j(f) = 2^{-j} \sum_{k=0}^{2^j-1} \mathcal{H}_{j,L}(2^{-j}(f+k))(2\sin(\pi 2^{-j}(f+k)))^{-2d}$.

Thus within a particular wavelet scale the nonboundary wavelet coefficients of an FD process are also approximately uncorrelated if $S_j(\cdot)$ is close to the spectrum for a white noise process; that is, $S_j(\cdot)$ is approximately flat. Figure 1 illustrates that this is a good approximation for an FD(0.45) process with $\sigma^2 = 1$ analyzed using a LA(8) wavelet filter. The top left panel shows the spectrum of the process along with the approximate passbands that correspond to the first five wavelet levels. The top right panel shows $S_j(\cdot)$ for $j = 1, \ldots, 5$. The lower panels illustrate the approximations to these spectra used in the paper. If we assume that the wavelet coefficients are uncorrelated per each wavelet level, we obtain the flat spectra given in the lower left panel. Clearly the lower right panel show spectra that better model the true spectra of the wavelet coefficients. In this case we assume that the wavelet coefficients on each level follow an AR(1) model, where the AR parameters are given by $\phi_j(d) = s_{j,1}(d)/s_{j,0}(d)$ with variance $\eta_j^2(d) = \sigma^2(1 - \phi_j^2(d))$ and hence depend on $d$ alone.

Now let $Y_t = T_t + X_t$ and perform a DWT on these data (Craigmile et al. 2003). By the linearity of the transform the wavelet coefficients are Gaussian. Because a Daubechies wavelet filter of order $L$ has $L/2$ embedded differencing operations we can zero out a trend of polynomial order $K$ in the nonboundary wavelet coefficients if $K \leq L/2 - 1$; that is, only the boundary wavelet coefficients will contain the trend component. The above results apply and the nonboundary wavelet coefficients can be regarded approximately as either uncorrelated or following an AR(1) model on each level.

# 5 The white noise model

We now consider the simplest model for estimating the parameters of the FD process using the wavelet coefficients (the next section explores the refinement given by the AR(1) model). Assume that the nonboundary wavelet coefficients $\{(W_{nb})_{j,k} : j = 1, \ldots, J,\ k = 0, \ldots, M_j - 1\}$ form an independent sample with $(W_{nb})_{j,k} \sim N(0, s_{j,0}(d)\sigma^2)$. The likelihood function for this model is

$$L_M(d, \sigma^2 | (W_{nb})_{j,k}) = \prod_{j=1}^{J} \prod_{k=0}^{M_j-1} (2\pi s_{j,0}(d)\sigma^2)^{-1/2} \exp\left( -\frac{(W_{nb})_{j,k}^2}{2 s_{j,0}(d)\sigma^2} \right).$$

If we let $R_j = \sum_{k=0}^{M_j-1} (W_{nb})_{j,k}^2$ denote the sum of squares of the level j nonboundary wavelet coefficients and $M = \sum_{j=1}^{J} M_j$, then maximizing the likelihood is equivalent to minimizing twice the negative of the log likelihood; that is,

$$-2\, l_M(d, \sigma^2 | (W_{nb})_{j,k}) = M \log(2\pi\sigma^2) + \sum_{j=1}^{J} M_j \log(s_{j,0}(d)) + \sum_{j=1}^{J} \frac{R_j}{s_{j,0}(d)\sigma^2}. \tag{7}$$

For a given $d$, the above is a function of $\sigma^2$ that is minimized when

$$\widehat{\sigma}_M^2(d) = \frac{1}{M} \sum_{j=1}^{J} \frac{R_j}{s_{j,0}(d)}. \tag{8}$$

Substituting this estimator into equation (7), we obtain a function of $d$ alone, known as the profile log likelihood (McCullagh and Nelder 1989):

$$-2\, l_M(d, \widehat{\sigma}_M^2(d) | (W_{nb})_{j,k}) = M \left( \log(2\pi\widehat{\sigma}_M^2(d)) + 1 \right) + \sum_{j=1}^{J} M_j \log(s_{j,0}(d)). \tag{9}$$

Maximizing with respect to $d$ yields the maximum likelihood estimator, $\widehat{d}_M$.

## 6 The first order autoregressive model

In Figure 1 we illustrated that within scales, a good approximation to the spectrum of the nonboundary wavelet coefficients is to assume an AR(1) model per scale. We now investigate this in further detail. We assume $\{(W_{nb})_{j,k} : k = 0 \ldots M_j - 1\}$ is a portion of an AR(1) process; that is,

$$(W_{nb})_{j,k} \quad = \quad \phi_j(d)(W_{nb})_{j,k-1} + (Z_{nb})_{j,k} \tag{10}$$

where $\{(Z_{nb})_{j,k} \sim \text{i.i.d. } N(0, \eta_j(d)\sigma^2) : j = 1 \ldots J,\ k = 0 \ldots M_j - 1\}$. The parameters of the AR(1) process on each wavelet level $j$ depend on just the FD parameter $d$ – we do not fit a separate AR process to each scale. For any given level j, the Yule–Walker equations (see e.g. Box, Jenkins, and Reinsel (1994)) yield

$$\phi_j(d) \quad = \quad s_{j,1}(d)/s_{j,0}(d) \quad \text{and} \quad \eta_j(d) \ = \ s_{j,0}(d)(1 - \phi_j^2(d)). \tag{11}$$

Assuming again independence between coefficients on different scales, it follows from Box, Jenkins, and Reinsel (1994) that the equivalent of Equation (7) is

$$\begin{aligned} -2\ l_M(d, \sigma^2 | (W_{nb})_{j,k}) \quad &= \quad M \log(2\pi\sigma^2) + \sum_{j=1}^{J} \left[ M_j \log(\eta_j(d)) - \log(1 - \phi_j^2(d)) \right] \\ &\quad + \sum_{j=1}^{J} \left[ \frac{(W_{nb})_{j,0}^2 (1 - \phi_j^2(d)) + \sum_{k=1}^{M_j - 1} \{(W_{nb})_{j,k} - \phi_j(d)(W_{nb})_{j,k-1}\}^2}{\sigma^2 \eta_j(d)} \right]. \end{aligned} \tag{12}$$

If we minimize with respect to $\sigma^2$, we obtain

$$\widehat{\sigma}_M^2(d) \quad = \quad M^{-1} \sum_{j=1}^{J} \left[ \frac{(W_{nb})_{j,0}^2 (1 - \phi_j^2(d)) + \sum_{k=1}^{M_j - 1} \{(W_{nb})_{j,k} - \phi_j(d)(W_{nb})_{j,k-1}\}^2}{\eta_j(d)} \right],$$

and the profile log likelihood is proportional to

$$-2\ l_M(d, \widehat{\sigma}_M^2(d) | (W_{nb})_{j,k}) \quad = \quad M \left[ \log(2\pi\widehat{\sigma}_M^2(d)) + 1 \right] + \sum_{j=1}^{J} \left[ M_j \log(\eta_j(d)) - \log(1 - \phi_j^2(d)) \right].$$

## 7 Properties of the wavelet-based estimators

In this section we provide theory for the estimators under the models discussed in Sections 5 and 6. In particular this theory provides approximate confidence intervals for the FD parameter. These results give an illustration of the large sample properties of what we can think of as "wavelet-based models for long memory" (Li and Oh 2002). We examine further properties of these estimators by simulation in section 8.

For a wavelet filter of width $L$, let $\Theta_L \equiv \{\boldsymbol{\theta} = (d, \sigma^2)^T \in \mathbb{R}^2 : d < (L+1)/2 \text{ and } \sigma^2 > 0\}$ denote the parameter space of interest. Suppose that $\boldsymbol{\theta}_0 = (d_0, \sigma_0^2)^T \in \Theta_L$ denotes the true values of the parameters, which are estimated by $\widehat{\boldsymbol{\theta}}_M = (\widehat{d}_M, \widehat{\sigma}^2_M(\widehat{d}_M))^T$ under the white noise or AR(1) wavelet model. Also, let $m_j = \lim_{M\to\infty}(M_j/M)$ and for any differentiable function, $g$, define the operator

$$\Delta_1(g(x)) = \frac{\frac{d}{dx}g(x)}{g(x)}. \tag{13}$$

The following two theorems provide the large sample properties of the estimators under the white noise and AR(1) wavelet models respectively.

**Theorem 7.1** *Suppose that the white noise model is the true model for the nonboundary wavelet coefficients within each level. Then the following holds.*

*(a) (Consistency) With probability converging to one there exist solutions,* $\widehat{\boldsymbol{\theta}}_M$*, of the likelihood equation such that* $\widehat{\boldsymbol{\theta}}_M \to_p \boldsymbol{\theta}_0$*, as* $M \to \infty$*.*

*(b) (Joint asymptotic normality)* $\sqrt{M}(\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0) \to_d N(0, \Sigma_0^{-1}(\boldsymbol{\theta}_0))$*, as* $M \to \infty$*, where*

$$\Sigma_0(\boldsymbol{\theta}) = \frac{1}{2}\begin{bmatrix} \sum_{j=1}^J m_j\, \Delta_1^2(s_{j,0}(d)) & \sigma_\epsilon^{-2}\sum_{j=1}^J m_j\, \Delta_1(s_{j,0}(d)) \\ \sigma_\epsilon^{-2}\sum_{j=1}^J m_j\, \Delta_1(s_{j,0}(d)) & \sigma_\epsilon^{-4} \end{bmatrix}.$$

*(c) (Marginal asymptotic normality of* $\widehat{d}_M$*)* $\sqrt{M}(\widehat{d}_M - d_0) \to_d N(0, \psi_0^2(d_0))$*, as* $M \to \infty$*, where*

$$\psi_0^2(d) = 2\left[\Big(\sum_{j=1}^J m_j\, \Delta_1^2(s_{j,0}(d))\Big) - \Big(\sum_{j=1}^J m_j\, \Delta_1(s_{j,0}(d))\Big)^2\right]^{-1}.$$

(d) *(Exact distribution of $\widehat{\sigma}^2_M(d_0)$) $\widehat{\sigma}^2_M(d_0) =_d \sigma^2\chi^2_M/M$, where $\chi^2_M$ denotes a chisquared random variable with $M$ degrees of freedom.*

To calculate $\Delta_1(s_{j,0}(d))$ in the above theorem, we need to know $\frac{d}{dd}s_{j,\tau}(d)$. We have that

$$s'_{j,\tau}(d) = \frac{d}{dd}s_{j,\tau}(d) = -4\int_0^{1/2} [\log\sin(\pi f)]\,\mathcal{H}_{j,L}(f)\cos(2^{j+1}\pi f\tau)(2\sin(\pi f))^{-2d}\,df. \quad (14)$$

This derivative is obtained via Leibnitz's rule that allows us to interchange differentiation and integration.

**Theorem 7.2** *Suppose that the AR(1) model is the true model for the nonboundary wavelet coefficients within each wavelet level. Then the following holds.*

(a) *(Consistency) With probability converging to one there exist solutions, $\widehat{\boldsymbol{\theta}}_M$, of the likelihood equation such that $\widehat{\boldsymbol{\theta}}_M \to_p \boldsymbol{\theta}_0$, as $M \to \infty$.*

(b) *(Joint asymptotic normality) $\sqrt{M}(\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0) \to_d N(0, \Sigma_1^{-1}(\boldsymbol{\theta}_0))$, as $M \to \infty$, where*

$$\Sigma_1(\boldsymbol{\theta}) = \frac{1}{2}\begin{bmatrix} \sum_{j=1}^J m_j\,\Delta_1^2(\eta_j(d)) & \sigma_\epsilon^{-2}\sum_{j=1}^J m_j\,\Delta_1(\eta_j(d)) \\ \sigma_\epsilon^{-2}\sum_{j=1}^J m_j\,\Delta_1(\eta_j(d)) & \sigma_\epsilon^{-4} \end{bmatrix}.$$

(c) *(Marginal asymptotic normality of $\widehat{d}_M$) $\sqrt{M}(\widehat{d}_M - d_0) \to_d N(0, \psi_1^2(d_0))$, as $M \to \infty$, where*

$$\psi_1^2(d) = 2\left[\left(\sum_{j=1}^J m_j\Delta_1^2(\eta_j(d))\right) - \left(\sum_{j=1}^J m_j\Delta_1(\eta_j(d))\right)^2\right]^{-1}.$$

In the above theorem, we calculate $\Delta_1(\eta_j(d)) = \frac{d}{dd}\eta_j(d)/\eta_j(d)$ by taking derivatives of equation (11) with respect to $d$. In particular:

$$\begin{aligned} \frac{d}{dd}\phi_j(d) &= \frac{s'_{j,1}(d)}{s_{j,0}(d)} - \phi_j(d)\Delta_1(s_{j,0}(d)) \\ \frac{d}{dd}\eta_j(d) &= s'_{j,0}(d)(1-\phi_j^2(d)) + 2s_{j,0}(d)\phi_j(d)\left(\frac{d}{dd}\phi_j(d)\right). \end{aligned}$$

Table 1 tabulates $\psi_k^2(d)$ from the above theorems for various widths $L$, under either the white noise ($k = 0$) or the AR(1) model ($k = 1$) within each wavelet scale. For fixed $L$, the asymptotic variance decreases with increasing $d$. It also decreases with increasing $L$ for stationary $d < 1/2$, but increases with $L$ for non-stationary $d \geq 1/2$. As a result, $\psi_k^2(d)$ becomes more uniform across $d$ as $L$ increases.

We can obtain approximate $100(1-\alpha)\%$ confidence intervals (CIs) for $d$ based upon the approximate DWT models and their profile likelihoods via the log likelihood ratio statistic

$$2\log\lambda(d, \widehat{\sigma}_M^2(d)) \quad = \quad 2\left[l_M(\widehat{d}, \widehat{\sigma}_M^2(\widehat{d})) - l_M(d, \widehat{\sigma}_M^2(d))\right],$$

(the Wald or Rao test statistics could also be used to provide a confidence interval). Standard statistical theory (Lehmann 1998) suggests that an approximate $100(1-\alpha)\%$ confidence interval is given by $\{d : 2\log\lambda(d, \widehat{\sigma}_M^2(d)) \leq q_1(1-\alpha)\}$, where here $q_1(1-\alpha)$ denotes the $(1-\alpha)$th quantile of a chisquared random variable with 1 degree of freedom.

# 8 Monte Carlo Studies

Our aim of this section is to investigate how well the estimators perform in practice for the white noise and AR(1) wavelet models. We also compare the AR(1) model estimator to estimators in Hurvich and Chen (2000). All realizations of FD processes are created using the Davies–Harte algorithm with $\sigma^2 = 1$ (there is no lose of generality with this arbitrary choice).

## 8.1 Estimation of the long memory parameter

We first investigated how well wavelet-based estimators of the difference parameter, $d$, performed in practice. We simulated 1024 replications of FD($d$) processes of length $N = 256$, 512 and 1024 for values of $d$ ranging from 0 to 1.50 in steps of 0.25. Each time series was analyzed using the DWT with the Haar, D(4) and LA(8) wavelet filters. The number of levels we analyzed to, $J$, was dependent on the wavelet filter and the sample size $N$, namely $J = \log_2(N) - L/2$. We estimated $d$ via the white noise and AR(1) wavelet models. Each study, carried out in the statistical software package R (R Development Core Team 2003), was performed by minimizing the negative log-profile likelihood for values of $d$ ranging in [−1,3] (an arbitrary choice). We used two methods for calculating $s_{j,\tau}(d)$ (defined in equation (6)):

1. Exact form: Use numerical integration with a Gauss rule, calculating $\mathcal{H}_{j,L}(f)$ using the modulus squared of equation (2).

2. Bandpass approximation: $\mathcal{H}_{j,L}(f)$ is approximated by the squared gain function for a bandpass filter with passband $[1/2^{j+1}, 1/2^j]$, yielding, e.g., $s_{j,0}(d) \approx 2^{j+1} \int_{1/2^{j+1}}^{1/2^j} [2\sin(\pi f)]^{-2d}\, df$.

Figure 2 shows a plot of the root mean square error (RMSE) of the estimates for each case. The standard errors for the RMSEs (calculated using 512 bootstrap samples) are bounded by 0.0025; that is, approximately the height of the plotting symbols. For the Haar case we only plotted results for $d \leq 1.25$ (since the condition $d < (L+1)/2$ does not hold for $d = 1.5$). In all cases of wavelet filter and model we can see that estimation is best for small values of $d$. For $d > 0$ the RMSEs tend to be smaller for the exact $s_{j,\tau}(d)$ calculation compared to the bandpass approximation. This difference increases with $d$, but decreases with increasing wavelet filter order because $\mathcal{H}_{j,L}(\cdot)$ converges to an ideal bandpass filter as $L \to \infty$ (Lai 1995). The empirical value of the RMSE is worse in general for the white noise as compared with the autoregressive model. This is because, as shown in Figure 1, the AR(1) model gives us a better approximation to the correlation structure of the wavelet coefficients than the white noise model does (the white noise approximation deteriorates with increasing $d$). The RMSEs increase with wavelet order, and decrease for longer time series.

The estimation bias is not shown in these Figures. In general the bias decreases as we increase the wavelet order, and for $d \neq 0$ is bounded by $\pm 0.01$ (with a maximum standard deviation of 0.003). This is because we obtain better decorrelation between wavelet scales when we use longer wavelet filters. The bias is smaller for the exact $s_{j,\tau}(d)$ calculation compared to the bandpass approximation, and for the AR(1) wavelet model compared to the white noise wavelet model (the biases for the white noise model using an LA(8) filter with a bandpass variance calculation are displayed in Table 2). Because of the relatively low biases the variance largely determines the MSE of these estimators.

By the inherent differencing of the wavelet filter, we obtain the same simulation results if we add a polynomial trend of order $K$ as long as $K \leq L/2 - 1$ as well as $d < (L+1)/2$.

We now evaluate the theory of Section 7 (where we assume an approximating model for the wavelet coefficients) on the basis of the simulation results. In particular, Figure 3 compares the theoretical RMSE for the AR(1) model, with the estimated RMSE obtained using the above simulation. The theoretical RMSE is $\sqrt{\psi_1^2(d)/M}$, and is calculated using the equation in Theorem 7.2. We used exact variance calculations for $s_{j,\tau}(d)$. The theoretical and simulated values are closest for longer filter widths $L$, smaller $d$, and larger $N$. This is as expected since the theoretical RMSE is an asymptotic value that is calculated under the assumption of perfect decorrelation between scales (which by Theorem 4.1 occurs for longer filter widths $L$), using an approximating model within scale, which fits better for values of $d$ closest to zero.

### 8.2 Comparisons with the Hurvich–Chen estimator

We now compare our estimator of the FD parameter $d$ with that of Hurvich and Chen (2000). In their paper the authors propose a complex valued taper which can be used to estimate $d$ in the presence of a lower order polynomial trend when $d \in (-0.5, 1.5)$. This estimator is based on the Gaussian semiparametric estimator due to Künsch (1987), which can be used when $d \in (-0.5, 0.5)$. The method takes a periodogram of the data $\{X_t : t = 0, \ldots, N-1\}$ and estimates $d$ using only a fraction of the Fourier frequencies of the spectral estimate. More precisely for $f_j = j/N$ such that $0 < f_j \leq 1/2$, let

$$F^{(p)}(f_j) \quad = \quad \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t \, e^{-i2\pi f_j t} \tag{15}$$

denote the (orthonormal) discrete Fourier transform of the data (see chapter 3 of Percival and Walden (2000)), and let $S^{(p)}(f_j) = |F(f_j)|^2$ denote the periodogram of the data. Then the Künsch (1987) form of the estimator is to choose some $m < N/2$ and minimize the following with respect to $d \in (-0.5, 0.5)$ and $\sigma^2 > 0$:

$$Q(d, \sigma^2) = \frac{1}{m} \sum_{j=1}^{m} \left[ \log(\sigma^2) + \log(g_j(d)) + \frac{S^{(p)}(f_j)}{\sigma^2 g_j(d)} \right],$$

where $g_j(d) = (2\sin(\pi f_j))^{-2d}$. The minimizer of $Q$ with respect to $\sigma^2$ (which depends on $d$) is

$$\sigma_Q^2(d) = \frac{1}{m} \sum_{j=1}^{m} \frac{S^{(p)}(f_j)}{g_j(d)},$$

and after substitution,

$$Q(d, \sigma_Q^2(d)) = \log(\sigma_Q^2(d)) + \frac{1}{m} \sum_{j=1}^{m} \log(g_j(d)) + 1.$$

Minimizing this with respect to $d$ is equivalent to minimizing

$$Q^*(d, \sigma_Q^2(d)) = \log(\sigma_Q^2(d)) - 2d \left[ \frac{1}{m} \sum_{j=1}^{m} \log(2\sin(\pi j/N)) \right], \tag{16}$$

say, over the range $d \in (-0.5, 0.5)$. The refinement due to Hurvich and Chen (2000) is first to difference the data, thus turning a realization of a process with $d \in (-0.5, 1.5)$ into one with a difference parameter in the range $(-1.5, 0.5)$. This yields the Gaussian semiparametric estimator (GSE). However the estimator is not good for low values of $d$ as we have overdifferenced the time series. To compensate for this, Hurvich and Chen (2000) use a complex spectral taper given by

$$h_t = \frac{1}{2} \left[ 1 - \exp\left( \frac{i2\pi(t + 1/2)}{N} \right) \right],$$

for $t = 0, \ldots, N-1$ (this is basically an extension of Tukey's cosine bell taper to the complex plane). They provide a direct formula to obtain the spectral estimator under this taper from equation (15):

$$F^{(hc)}(f_j) = \frac{1}{\sqrt{2}} \left( F^{(p)}(f_j) - e^{-i\pi/N} F^{(p)}(f_{j+1}) \right).$$

Letting $S^{(hc)}(f_j) = \left| F^{(hc)}(f_j) \right|^2$, minimization of expression (16) with $S^{(p)}(f_j)$ replaced by $S^{(hc)}(f_j)$ in the expression for $\sigma_Q^2(d)$ yields an estimator denoted as GSET ('T' stands for tapered). The authors indicate this estimator is consistent and has a Gaussian limit distribution under a set of conditions given in their paper. They also indicate that a good choice of $m$ is given by $N^{4/5}/4$.

We now compare GSE and GSET with our wavelet-based estimator. We conducted a simulation study similar to Hurvich and Chen (2000), for which they consider time series of length 513 for an FD($d$) process with $d = 0$ to 1.4 in steps of 0.2. For the wavelet-based estimator, we used the AR(1) model with exact $s_{j,\tau}(d)$ calculations in conjunction with a level $J = 6$ LA(8) DWT of the first 512 point of each simulated series. The bias, variance and RMSEs of the sample of estimates for 500 replications of each method are shown in three parts of Table 2. We see that the wavelet-based estimator clearly outperforms the other two estimators in terms of RMSE, variance and (with the single exception of $d = 0.8$) magnitude of bias. The reason for the disparity in the results is because the GSE and GSET methods use a trim factor $m$. In this simulation study, the estimate of $d$ for these spectral methods are based on 37 periodogram bins. For the wavelet-based method we use 480 wavelet coefficients. If we reduce the number of wavelet levels that we use to estimate $d$, then the results become more comparable. To demonstrate this, the last columns of Table 2 display the bias, variance and RMSE for an AR(1) model fit using only levels $j = 3, \ldots, 6$ (this choice means that the range of frequencies collectively covered by the wavelet coefficients is approximately the same as covered by GSE and GSET). The bias for this subset AR(1) model is smaller than for the GSET method (except again at $d = 0.8$), but the variance is larger (except for $d = 0$). Trimming of periodogram bins or wavelet levels is useful in practice if we want to estimate $d$ in the presence of the short range dependence (see, e.g., Hurvich and Chen (2000) and Bardet et al. (2000)).

## 9 A northern hemisphere temperature series

Figure 4 shows a time series plot of a deseasonalized version of the monthly deviations in the average Northern hemisphere temperature (in units of degrees Celsius) from 1854 to 1998, relative to the monthly average over the period 1961 to 1990. The data come from the Climate Research Unit, University of East Anglia, UK. This is a newly updated version of the dataset that incorporates combinations of grid data (over the sea and land) from 1000 extra sites, new reference periods and an increased resolution. We deseasonalized the time series via a harmonic analysis, fitting the model

$$\begin{aligned} Y_t &= A_1 \cos(0.994 \times 2\pi t\Delta t) + B_1 \sin(0.994 \times 2\pi t\Delta t) \\ &\quad + A_2 \cos(1.001 \times 2\pi t\Delta t) + B_2 \sin(1.001 \times 2\pi t\Delta t) + X_t, \end{aligned}$$

for $t = 0 \ldots N-1$, where $\Delta t = 1/12$ year is the sampling interval, and here $N = 1664$. This least squares model is valid for long memory dependence by Yajima (1988). Visually there is an indication of an upward trend and increased variability at the start of the series.

There has been much interest in earlier versions of this time series. These data only went up to 1989 and were averaged over a different reference period (1950–1979). Smith (1993) illustrates the problem of trying to fit an auto-regressive model to the data. Using an estimate of $d$ based on a spectral estimate, he concludes significant long memory behavior of the series, with $d$ ranging from 0.290 to 0.403 (depending on the choice of two key parameters in the estimator). He finds a significant linear trend, adapting for long memory. Alternatively, using a kernel smoother Beran and Feng (2002) obtain an ML estimate for $d$ of 0.33 with a 95% CI of [0.19,0.46]. At the 5% level there is no significant trend (the authors also analyze the land and sea averages individually and concludes the former has a trend, and the latter does not). We now analyze the updated series using our proposed methodology, which allows us to estimate $d$ even if data is contaminated by a low order polynomial (as might be the case here). We first assess whether an FD($d$) process is reasonable for this series.

Figure 5 shows a periodogram of the data. If we take the log of the spectrum (equation 4) we have

$$\log(S_X(f)) = \log(2\sigma^2) - 2d \log(2\sin(\pi f \Delta t)),$$

for $0 < f < 1/(2\Delta t)$, where $\Delta t = 1/12$ year is the sampling rate. For small $x$, $\sin(x) \approx x$ and thus

$$\log(S_X(f)) \approx \log(2\sigma^2) - 2d \log(2\pi f \Delta t). \tag{17}$$

Hence an FD process is a good model if the log spectrum versus log frequency is approximately a straight line for small $f$, as in this case. By calculating the slope of the line for small enough $f$ we obtain an estimate of $d$. For this dataset we obtain an estimate of 0.445 for $f \leq 1.5$, indicating evidence of long memory.

Using an LA(8) wavelet filter (which can handle a cubic polynomial trend) and analyzing to level $J = 7$, we obtain the DWT decomposition of the deseasonalized deviations shown in Figure 6. The thick gray vertical lines denote the partition between the boundary (outside) and nonboundary wavelet coefficients (inside) on each wavelet level. The nonboundary wavelet coefficients on lower scales ($j = 1, 2, 3$) are more variable in earlier years, which violates an assumption behind our proposed method for estimating $d$. We can also look at normal Q-Q plots, ACFs, PACFs and periodogram for the nonboundary wavelet coefficients on each scale (not shown). The Gaussian assumption for the data seems reasonable, although the non-constant variance is evident in the lower wavelet levels by the over-dispersion in the Q-Q plots. Lag 1 auto-correlation on levels 4 and 5 imply that the AR(1) wavelet model is more appropriate than the white noise model. If we ignore the non-constant variance problem (as has also been done in the earlier analyses cited above), we obtain an estimate of $\widehat{d}_M = 0.361$ (with a 95% CI of [0.317,0.408]) and $\widehat{\sigma}^2(\widehat{d}_M) = 0.045$ using the AR(1) model.

To assess the affect of non-constant variance, we repeated our analysis using just the last 96 years of data (removing the peak around $f = 1$ again). In this case the heteroscedacity in the boundary-independent wavelet coefficients reduces, and we obtain using the AR(1) wavelet model $\widehat{d}_M = 0.368$ (with a 95% CI of $[0.323, 0.415]$) and $\widehat{\sigma}^2_M(\widehat{d}_M) = 0.032$. The increased variability at the start of series has little effect on the estimate of $d$, but $\widehat{\sigma}^2_M(\widehat{d}_M)$ is reduced somewhat.

An alternative procedure to handle the periodicity at around one year is to analyze the yearly averages. A periodogram similar to Figure 5 show evidence of long memory in this case. We perform a DWT on these data using a D(6) filter to level $J = 4$ (the lower values of $L$ and $J$ are dictated by the decrease in sample size). The equivalent diagnostic plots show few problems in the distribution of the boundary-independent wavelet coefficients (probably due to the small sample sizes – note that we can only handle a quadratic trend now). When we use the AR(1) wavelet model, we obtain $\widehat{d}_M = 0.343$ (with a 95% CI of $[0.101, 0.648]$) and $\widehat{\sigma}^2_M(\widehat{d}_M) = 0.020$, comparable with the previous results. The smaller value of the innovations variance is due to the averaging involved.

Thus, independently of the possible presence of a low order polynomial trend of order $K$ (as long as $L/2 \geq K+1$), there is evidence of significant long memory. For the deseasonalized deviations the long memory process is stationary (since the CI for $d$ does not contain values greater than or equal to 0.5), but we cannot conclude stationarity for the yearly averaged series (due to the reduction in sample size). These deductions support the ideas of Smith (1993) and Beran and Feng (2002), that we should be cautious in testing for a significant trend in this series, unless we can adequately account for the long memory dependence (we investigate the question of the significance of a trend in Craigmile, Percival, and Guttorp (2003)).

## 10 Discussion and summary

The key property of the DWT that we have exploited in our work is that it approximately decorrelates FD processes. The degree to which this approximation holds must be assessed by considering the correlations between wavelet coefficients on the same scale and on different scales. As the wavelet filter width $L$ increases, the correlation between coefficients on different scales necessarily decreases to zero (Craigmile and Percival 2003); however, the same cannot be said for within-scale correlations. Since coefficients within scale are correlated, we consider an AR model to capture this dependency structure. The combination of a moderate filter width ($L = 8$) and the AR model is sufficient to give a very good description of FD processes in the wavelet domain. We have demonstrated through our Monte Carlo experiments that the large sample theory that is based upon this wavelet-based description is reasonably accurate, even for modest sample sizes ($N = 256$). While it should be possible to derive a large sample theory that would take into account the correlations between the wavelet coefficients not accounted for by our approximations, the justification for this nontrivial extension would have to be as an interesting mathematical exercise: the theory that we have developed here is sufficient for all practical purposes.

Since our theory agrees with our simulations best for smaller values of $d$, a useful strategy in practice is to difference the process when there is evidence that $d \geq 0.5$ (i.e., the FD process is nonstationary). The theory developed in this paper will then apply to estimation of the long parameter for the differenced series (since our theory applies to the estimation of FD processes for $d < 0$).

The results presented here for FD processes can be extended naturally to, for example, ARFIMA processes by modeling the autoregressive and moving average component in the spectrum of the process. Estimation via likelihood follows naturally, with equivalent limit theorems. The limit variance of the difference parameter will then depend on the other parameters in the model. When extending the results to other error processes we need to assess the extent to which we can decorrelate the error process, and thus whether the white noise or AR(1) wavelet model fits to the nb wavelet coefficients in this case. Plots allow us to investigate this question theoretically (see, e.g., Figure 1) and in practice (by looking at normal Q-Q plots, ACFs, PACFs and periodograms for the nb wavelet coefficients).

In summary, we have investigated estimation of the parameters of trend contaminated FD processes using the DWT. Our proposed method is valuable in the case of low order polynomial trend (relative to the wavelet order), since it provides for an elegant partitioning of the noise and trend components. This leads to an computationally efficient estimator of $d$ (the wavelet transform is $O(N)$, and the solution of the profile likelihood equation is fast if we use division schemes such as the bisection method, or a Newton-Raphson algorithm). We can also improve estimation by modeling the within wavelet scale correlations using an AR(1) model, and using exact wavelet variance calculations rather than the bandpass approximation.

**Acknowledgments** The authors gratefully acknowledge support for this research from an STTR grant from AFOSR (Insightful, Inc., and the University of Washington), an NSF grant (University of Washington) and an EPA grant (National Research Center for Statistics and the Environment).

## References

Abry, P., P. Gonçalvès, and P. Flandrin (1993). Wavelet-based spectral analysis of $1/f$ processes. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Volume 3, pp. 237–240. Minneapolis, MN, USA.

Abry, P., P. Gonçalvès, and P. Flandrin (1995). Wavelets, spectrum analysis and $1/f$ processes. Volume 103 of *Lecture Notes in Statistics*, pp. 15–29. New York: Springer-Verlag.

Bardet, J., G. Lang, E. Moulines, and P. Soulier (2000). Wavelet estimator of long range dependence processes. *Statistical Inference for Stochastic Processes 3*(1-2), 85–99.

Beran, J. (1994). *Statistics for Long Memory Processes*, Volume 61 of *Monographs on Statistics and Applied Probability*. New York: Chapman & Hall.

Beran, J. and Y. Feng (2002). SEMIFAR models – A semiparametric approach to modelling trends, long-range dependence. *Computational Statistics and Data Analysis 40*(2), 393–419.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994). *Time series analysis, forecasting and control (third edition)*. Englewood Cliffs, New Jersey: Prentice-Hall.

Craigmile, P. F. (2003). Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis 24*(5), 505–511.

Craigmile, P. F. and D. B. Percival (2003). Asymptotic decorrelation of between-scale wavelet coefficients. Technical Report 711, Department of Statistics, The Ohio State University.

Craigmile, P. F., D. B. Percival, and P. Guttorp (2003). Assessing nonlinear trends using the discrete wavelet transform. *Environmetrics*. forthcoming.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. Philadelphia: SIAM.

Davies, R. B. and D. S. Harte (1987). Tests for Hurst effect. *Biometrika 74*, 95–101.

Deo, R. S. and C. M. Hurvich (1998). Linear trend with fractionally integrated errors. *Journal of Time Series Analysis 19*, 379–397.

Giraitis, L., P. Kokoszka, and R. Leipus (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability 38*(4), 1033–1054.

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

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika 68*(1), 165–176.

Hurvich, C. and W. W. Chen (2000). An efficient taper for potentially overdifferenced long memory time series. *Journal of Time Series Analysis 21*(2), 155–180.

Jensen, M. J. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control 24*(3), 361–387.

Künsch, H. (1987). Statistical aspects of self-similar processes. In Y. Prohorov and V. Sazanov (Eds.), *Proceedings of the First World Congress of the Bernoulli Society*, Volume 1, Utrecht, pp. 67–74. VNU Science Press.

Lai, M.-J. (1995). On the digital filter associated with Daubechies' wavelets. *IEEE Transarctions on Signal Processing 43*, 2203–2205.

Lehmann, E. L. (1998). *Theory of Point Estimation* (Second ed.). New York: Springer-Verlag.

Leipus, R. and M.-C. Viano (2003). Long memory and stochastic trend. *Statistics & Probability*

*Letters 61*(2), 177–190.

Li, T.-H. and H.-S. Oh (2002). Wavelet spectrum and its characterization property for random processes. *IEEE Transactions on Information Theory 48*, 2922–2937.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*(7), 674–693.

McCoy, E. and A. Walden (1996). Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics 5(1)*, 26–56.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models (Second Edition)*. New York: Chapman & Hall.

Moulines, E. and P. Soulier (2000). Broadband log-periodogram regression of time series with long range dependence. *The Annals of Statistics 27*(4), 1415–1439.

Percival, D. and A. Bruce (1998). Wavelet-based approximate maximum likelihood estimates for trend-contaminated fractional difference processes. Technical Report 67, Insightful.

Percival, D. and A. Walden (2000). *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press.

R Development Core Team (2003). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-00-3.

Robinson, P. M. (1994). Rates of convergence and optimal spectral bandwidth for long range dependence. *Probability Theory and Related Fields 99*, 443–473.

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. New York: Chapman & Hall.

Smith, R. L. (1989a). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science 4*, 367–377.

Smith, R. L. (1989b). Reply to comments on "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone". *Statistical Science 4*, 389–393.

Smith, R. L. (1993). Long-range dependence and global warming. In *Statistics for the Environment*, pp. 141–161. Wiley (New York).

Teverovsky, V. and M. Taqqu (1997). Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *Journal of Time Series Analysis 18*, 279–304.

Vannucci, M. and F. Corradi (1999). Modeling dependence in the wavelet domain. In *Bayesian Inference In Wavelet-Based Models*, Lecture notes in statistics, Chapter 12, pp. 173–186. New York: Springer-Verlag.

Veitch, D. and P. Abry (1999). A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory 45*(3), 878–897.

Wood, A. T. A. and G. Chan (1994). Simulation of stationary Gaussian processes in $[0,1]^d$. *Journal of Computational and Graphical Statistics 3*, 409–432.

Wornell, G. W. (1995). *Signal Processing with Fractals: A Wavelet-Based Approach*. Upper Saddle River, New Jersey: Prentice Hall.

Yaglom, A. (1958). Correlation theory of processes with random stationary $n$th increments. *American Mathematical Society Translations (Series 2) 8*, 87–141.

Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *The Annals of Statistics 16*, 791–807.

Table 1: Calculation of $\psi_k^2(d)$ for various filter widths $L$, under either the white noise ($k = 0$) or the AR(1) model ($k = 1$) within each wavelet scale. We set $J = 6$ in each case.

| | | $d$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $L$ | Model | 0 | 0.25 | 0.50 | 0.75 | 1.00 | 1.25 | 1.50 |
| 2 | White Noise | 1.260 | 1.036 | 0.896 | 0.781 | 0.664 | 0.541 | – |
| 2 | AR(1) | 1.260 | 1.020 | 0.886 | 0.795 | 0.664 | 0.433 | – |
| 4 | White Noise | 1.060 | 0.982 | 0.921 | 0.867 | 0.816 | 0.764 | 0.712 |
| 4 | AR(1) | 1.060 | 0.961 | 0.884 | 0.828 | 0.793 | 0.778 | 0.761 |
| 8 | White Noise | 0.991 | 0.956 | 0.923 | 0.893 | 0.864 | 0.836 | 0.809 |
| 8 | AR(1) | 0.991 | 0.936 | 0.884 | 0.838 | 0.800 | 0.771 | 0.755 |
| 16 | White Noise | 0.966 | 0.943 | 0.921 | 0.900 | 0.880 | 0.862 | 0.844 |
| 16 | AR(1) | 0.966 | 0.925 | 0.886 | 0.850 | 0.817 | 0.788 | 0.764 |

Table 2: Monte Carlo comparison of four methods to estimate $d$. In each case we simulated $N = 513$ FD($d$) series, 500 times in each case, and estimated $d$ using the basic Gaussian semiparametric method (GSE), the tapered version (GSET), the AR(1) wavelet model (with 6 wavelet levels), and the AR(1) wavelet model (using levels $j = 3, \ldots, 6$, which we call the subset AR(1) model). Both wavelet estimators use an LA(8) wavelet filter decomposition of the first 512 time points.

| | GSE | | | GSET | | | AR(1) | | | Subset AR(1) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d$ | bias | var. | MSE | bias | var. | MSE | bias | var. | MSE | bias | var. | MSE |
| 0.0 | 0.287 | 0.039 | 0.122 | 0.075 | 0.014 | 0.020 | 0.018 | 0.001 | 0.001 | 0.048 | 0.005 | 0.007 |
| 0.2 | 0.119 | 0.024 | 0.038 | 0.052 | 0.014 | 0.017 | -0.003 | 0.002 | 0.002 | -0.011 | 0.015 | 0.015 |
| 0.4 | 0.040 | 0.015 | 0.016 | 0.034 | 0.014 | 0.015 | -0.003 | 0.002 | 0.002 | -0.017 | 0.017 | 0.016 |
| 0.6 | 0.008 | 0.011 | 0.011 | 0.017 | 0.014 | 0.014 | -0.003 | 0.002 | 0.003 | -0.009 | 0.016 | 0.016 |
| 0.8 | -0.006 | 0.011 | 0.011 | 0.001 | 0.014 | 0.014 | -0.003 | 0.003 | 0.003 | -0.009 | 0.016 | 0.016 |
| 1.2 | -0.009 | 0.011 | 0.011 | -0.025 | 0.014 | 0.015 | -0.003 | 0.003 | 0.003 | -0.012 | 0.017 | 0.017 |
| 1.4 | -0.012 | 0.009 | 0.009 | -0.043 | 0.011 | 0.013 | -0.003 | 0.003 | 0.003 | -0.013 | 0.018 | 0.018 |

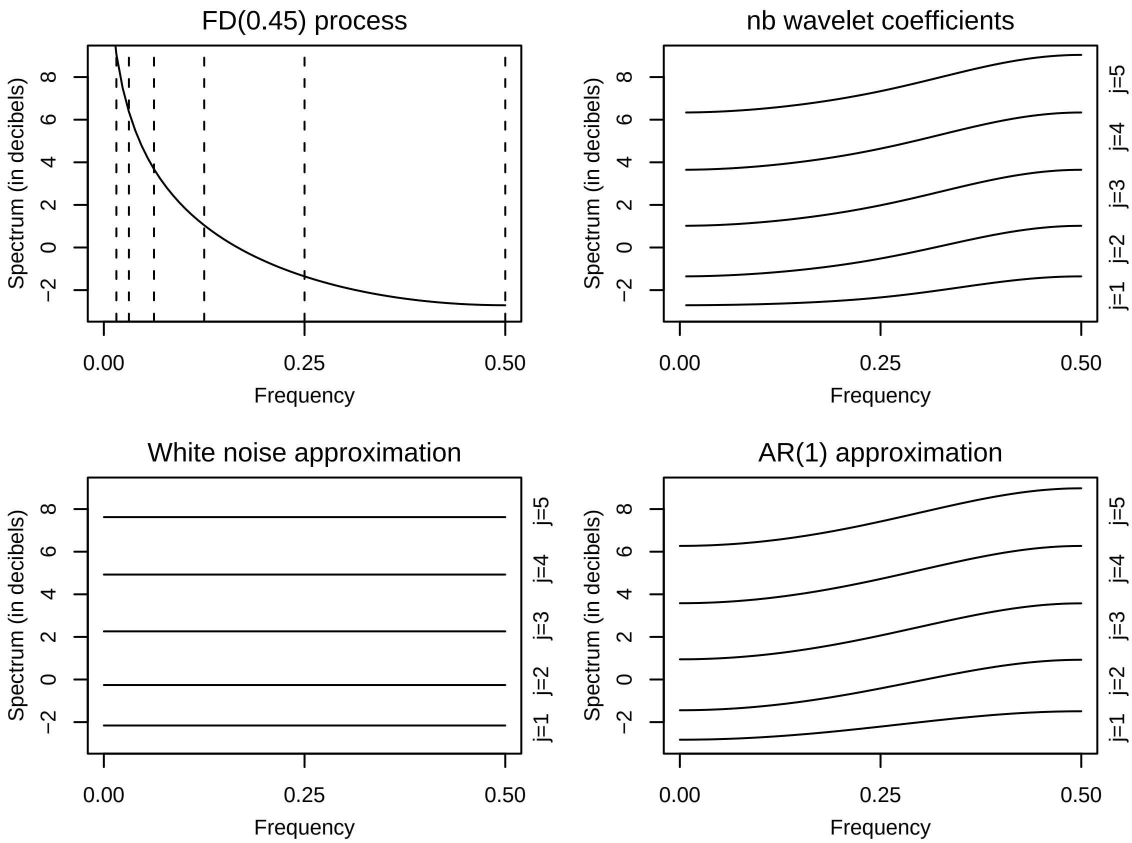


Figure 1: Going from left to right, top to bottom, plots show the spectrum of an FD(0.45) process (dotted vertical lines indicate the approximate passbands for the first five wavelet levels), the spectra of the LA(8) nonboundary (nb) wavelet coefficients, and the spectra assumed in the white noise and AR(1) models.

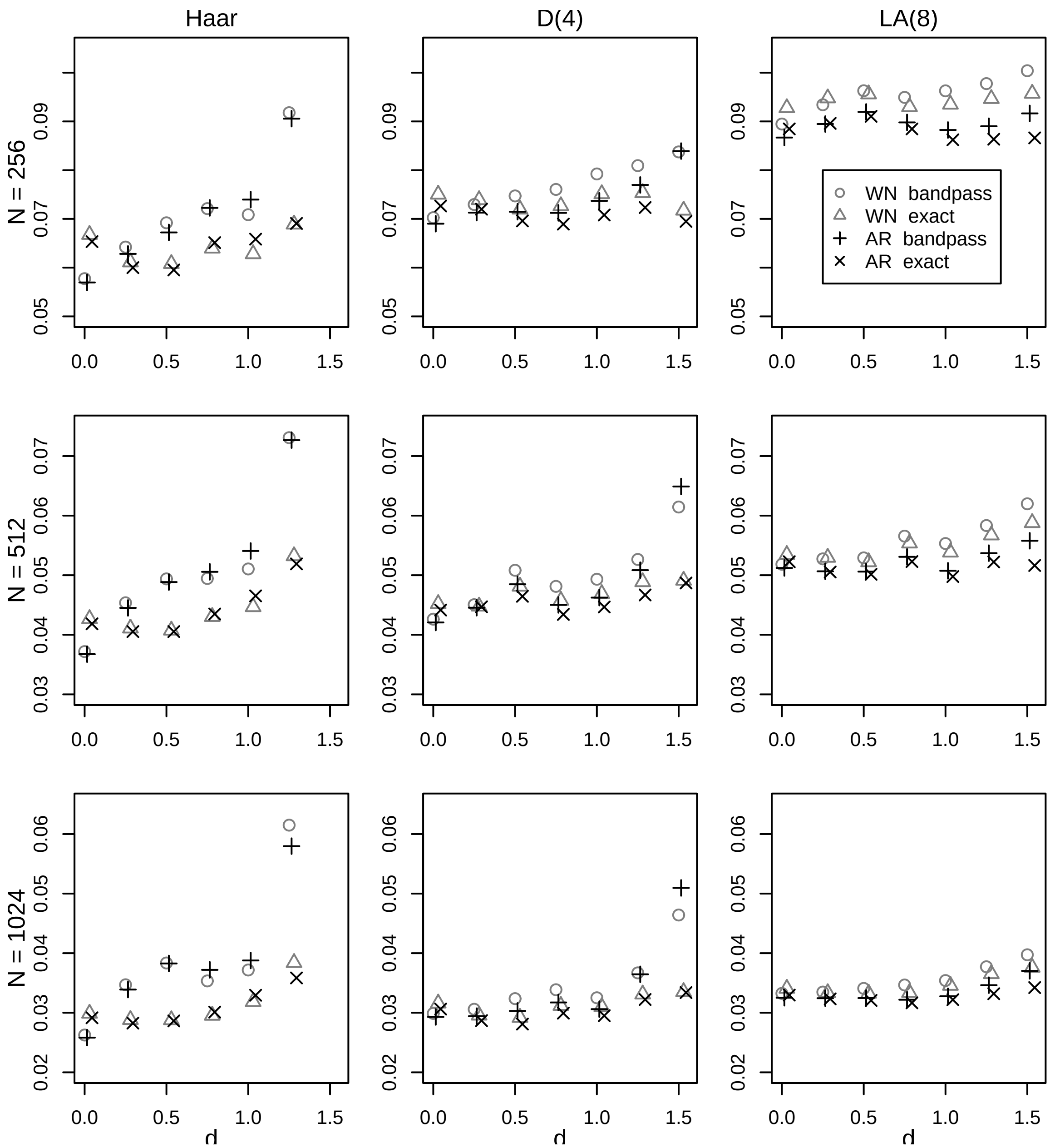


Figure 2: Root mean squared error in estimating the difference parameter using either the white noise or autoregressive wavelet models, for various wavelet filters, sample lengths, $N$, and difference parameters. The different symbols denote the different wavelet models and whether an exact or bandpass variance was used. The standard deviations of the RMSEs are bounded by 0.0025.

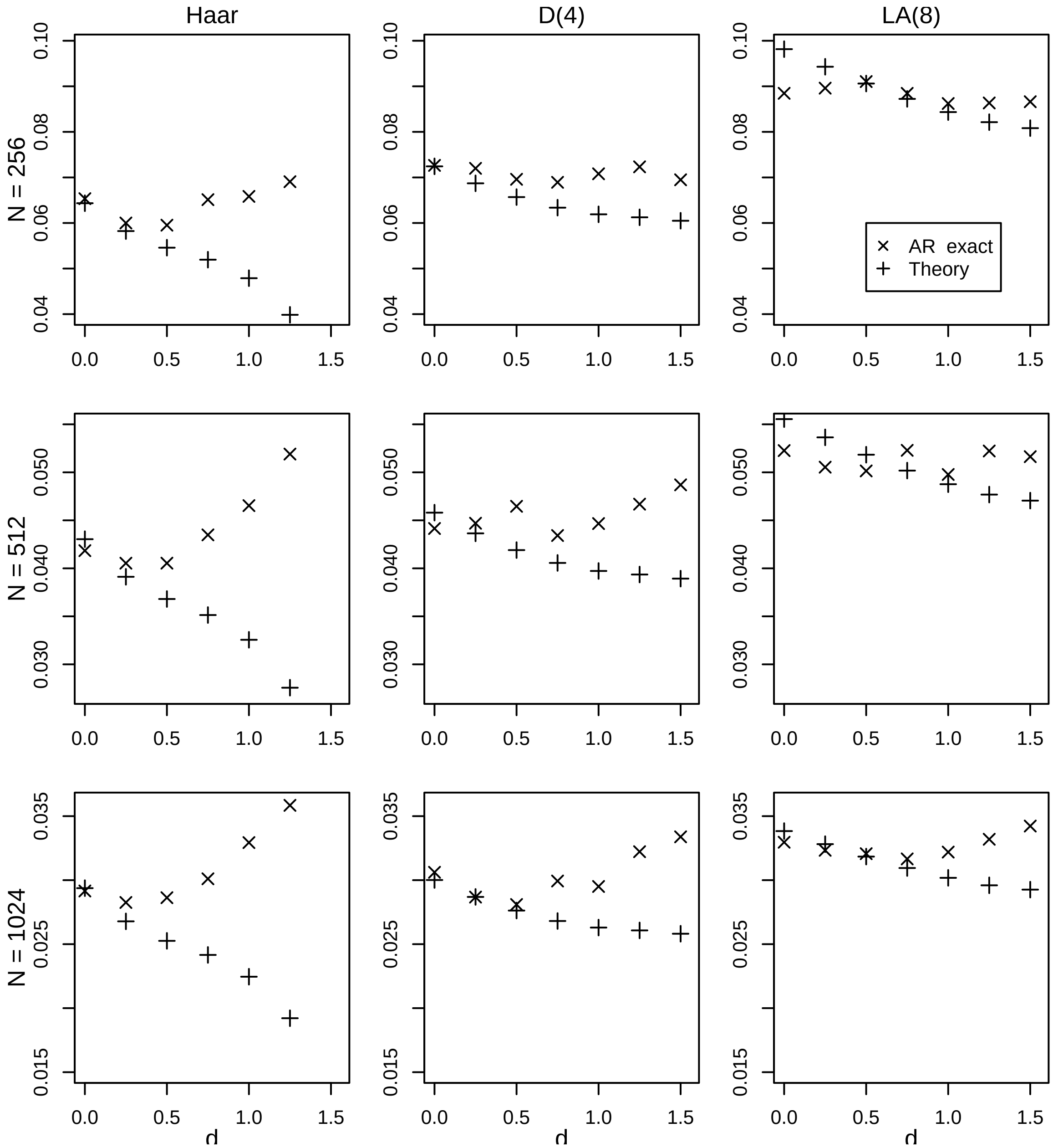


Figure 3: Comparisons of the theoretical RMSE, $\sqrt{\psi_1^2(d)/M}$ with the root mean squared error in estimating the difference parameter using the AR(1) method with exact variance calculations. The standard deviations of the RMSEs are bounded by 0.0025.

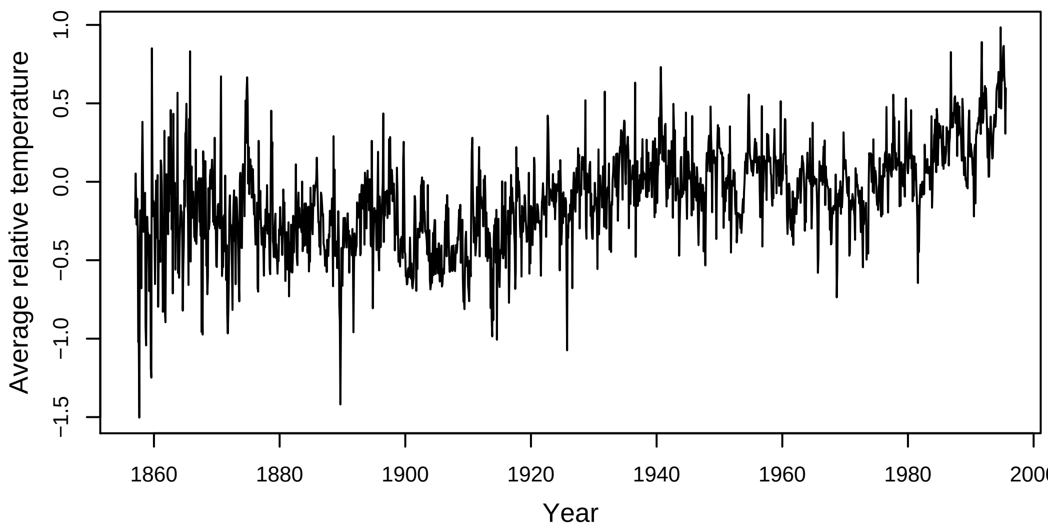


Figure 4: A time series plot of the monthly deseasonalized deviations in the northern hemisphere temperatures.

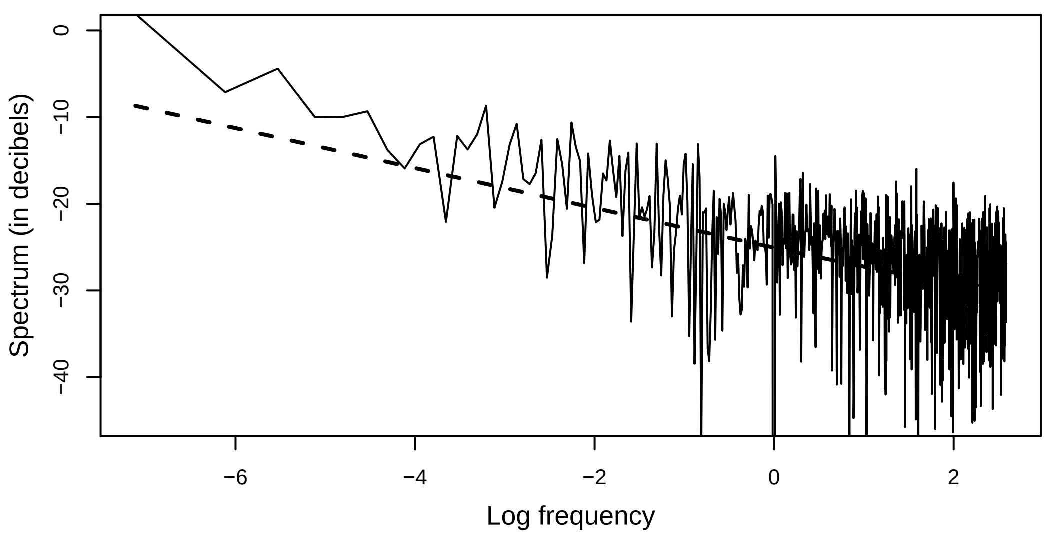


Figure 5: A periodogram of the monthly deseasonalized deviations. The spectrum (in decibels) is shown versus the log (base 2) frequency. The line on the figure denotes the least squares fit for $f \leq 1.5$.

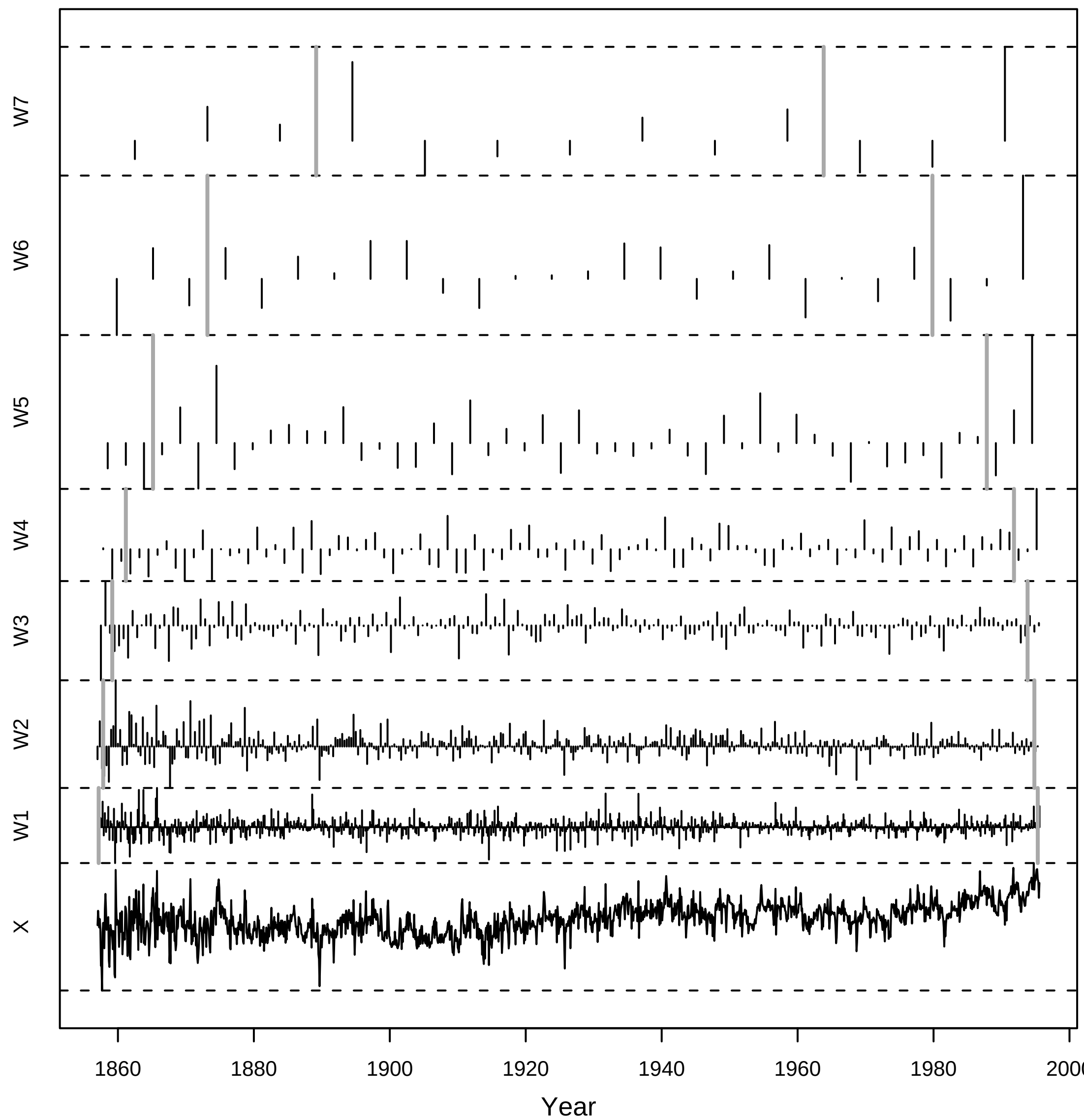


Figure 6: Wavelet coefficients from DWT decomposition of the northern hemisphere series using an LA(8) wavelet filter analyzing to level $J = 7$. The thick gray vertical lines denote the partition between the boundary (outside) and non-boundary wavelet coefficients (inside) on each wavelet level.