Wavelet-Based Parameter Estimation for Polynomial Contaminated Fractionally Differenced Processes

Peter F. Craigmile, Peter Guttorp, and Donald B. Percival.

***Abstract*— We consider the problem of estimating the parameters for a stochastic process using a time series containing a trend component. Trend, i.e., large scale variations in the series that are best modeled outside of a stochastic framework, is often confounded with low frequency stochastic fluctuations. This problem is particularly evident in models such as fractionally differenced (FD) processes, which exhibit slowly decaying autocorrelations and can be extended to encompass nonstationary processes with substantial low frequency components. We use the discrete wavelet transform (DWT) to estimate parameters for stationary and nonstationary FD processes in a model of polynomial trend plus FD noise. Using Daubechies wavelet filters allows for automatic elimination of polynomial trends due to embedded differencing operations. Parameter estimation is based on an approximate maximum likelihood approach made possible by the fact that the DWT decorrelates FD processes approximately. We consider this decorrelation in detail, examining the between and within scale wavelet correlations separately. Better between scale decorrelation can be achieved by increasing the length of the wavelet filter, while the within scale correlations can be handled via explicit modeling by a low order autoregressive process. We demonstrate our methodology by applying it to a popular climate dataset.**

***Index Terms*— fractionally differenced processes; discrete wavelet transform; trend; approximate Gaussian likelihood; confidence intervals.**

I. 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 [1]. One common example of a long memory process, the fractionally differenced (FD) process [2], [3], extends existing (integer) integrated processes. The mathematical tractability of FD processes allows for a varied range of estimation methods.

In the absence of a trend component, a common method of FD parameter estimation involves calculating the exact likelihood and maximizing with respect to the parameters. Beran [4] gives a review and evaluation of this method. 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 [4]. Some of these methods exploit fast transforms of the data such as the fast Fourier transform (Robinson [5] – see Moulines and Soulier [6] for an in depth analysis of this estimator) or wavelet transforms (Wornell [7] and McCoy and Walden [8] consider likelihood approaches; Abry et al. [9]–[12], Bardet et al. [13] study least square methods). Vannucci and Corradi [14] consider Bayesian estimation schemes for long memory processes, and Jensen [15] 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 [16]–[18]. Teverovsky and Taqqu [19] consider tests for long memory dependence in the presence of two types of trend (shifting means and slowly decaying trend). Percival and Bruce [20] extend the wavelet-based approximate likelihood estimates of McCoy and Walden [8] to work in the presence of polynomial trends. Deo and Hurvich [21] consider linear trends with fractionally integrated errors. Hurvich and Chen [22] provide a spectral estimation method that can handle some nonstationary ARFIMA processes with a low order polynomial trend component. Giraitis et. al. [23] consider families of tests for long memory, observed in the presence of deterministic trends. Leipus and Vinao [24] extend the work of the previous paper to the case of stochastic trends. Beran and Feng [25] use variable bandwidth smoothing to estimate such processes with additive trend.

In this paper we consider estimation of the long memory parameters of a polynomial contaminated FD process using the discrete wavelet transform (DWT) (see Craigmile et al. [26] for details on the estimation of the trend component). 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 cleanly separate polynomial trends from noise, thus allowing us to analyze time series with a trend well approximated by a polynomial.
3) Wavelets have excellent time and frequency localization, which can be useful for investigating local deviations

The authors gratefully acknowledge support for this research from an STTR grant from AFOSR (Insightful, Inc., and the University of Washington), NSF grant No. 0222115 (University of Washington) and an EPA grant (National Research Center for Statistics and the Environment). We also extend our thanks to the Associate Editor, three anonymous referees, and Agnieszka Jach who gave suggestions that improved this paper.

Peter F. Craigmile is with the Department of Statistics, The Ohio State University, Columbus, OH.

Peter Guttorp is with the Department of Statistics, University of Washington, Seattle, WA.

Donald B. Percival is with the Applied Physics Laboratory, University of Washington, Seattle, WA.

from a statistical model.

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 long memory 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) First order autoregressive (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 specific AR(1) models within each scale. While the AR models are different for each scale, they are in fact coupled together and are fully determined by just two FD parameters.

In Section II we define the DWT. We define the FD process in Section III and demonstrate the statistical properties of the DWT of these processes (with and without trend) in Section IV. We outline the approximate maximum likelihood scheme for the white noise wavelet model in Section V, and for the AR(1) wavelet model in Section VI. We provide theory for these estimators, under the assumption that the approximating models are true, in Section VII, and obtain approximate confidence intervals for the model parameters. In Section VIII, Monte Carlo simulations are used to assess these methods in practice. We also compare our methods to the spectral-based method of Hurvich and Chen [22] and the wavelet-based method of Veitch and Abry [12]. In Section IX 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 X (proofs of the results presented in this paper can be downloaded from `http://www.stat.ohio-state.edu/~pfc/`).

II. 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

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

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 [27] distinguishes between two (of the possible) choices: the extremal phase, D(L), filters are the ones that exhibit the smallest delay (have maximum cumulative energy) over other choices of scaling filter, and the least asymmetric, LA(L), filters (which differ 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 [28], [29]). Define $L_j = (2^j - 1)(L-1) + 1$. 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,

$$\begin{aligned} H_{j,L}(f) &= e^{-i2\pi(L_{j-1}-1)f} H_{1,L}(2^{j-1}f) \\ &\quad \times \prod_{k=0}^{j-2} H_{1,L}(1/2 - 2^k f) \end{aligned} \qquad (2)$$

which in turn defines the jth 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, for $k = 0, \ldots, N_j - 1$,

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

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, $\{\widetilde{W}_{j,k} \equiv W_{j,B_j+k} : j = 1, \ldots, J;\ k = 0, \ldots, M_j - 1\}$. The statistical properties of the boundary coefficients can be quite different from those of the nonboundary coefficients.

III. 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 [2] and Hosking [3] 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

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

Here d is known as the fractional difference parameter and σ^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 $d \ge 1/2$ in (4), and obtain a class of non-stationary processes that become stationary after differencing $\lfloor d + 1/2 \rfloor$ times [30]. Taking differences of the process, we can let $d \le -1/2$ to obtain a stationary, but noninvertible, process. For $d \in [-1/2, 1/2)$ the autocovariance sequence can be shown to be (Beran [4] and Hosking [3] for the $d = -1/2$ case)

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

Fast simulation of FD processes is possible using the Davies–Harte algorithm [31]–[33]. Further properties and an extensive history of the FD process can be found in, e.g., Beran [4]; Samorodnitsky and Taqqu [34, Sections 7.13 and 14.7].

IV. 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 process are Gaussian. By definition of the level j wavelet filter, $\{h_{j,l}\}$, $\sum_l h_{j,l} = 0$ [29, 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 [29, Equation (348a)]

$$\begin{aligned}\operatorname{cov}(\widetilde{W}_{j,k}, \widetilde{W}_{j',k'}) &= \int_{-1/2}^{1/2} e^{i2\pi[2^{j'}(k'+1)-2^j(k+1)]f} \\ &\quad \times H_{j,L}(f)\, H^*_{j',L}(f)\, \sigma^2 (2\sin(\pi f))^{-2d}\, df,\end{aligned}$$

where * denotes the complex conjugation operator. Between scales the DWT acts as a whitening transform for an FD process; that is, for $j \neq j'$, $\operatorname{cov}(\widetilde{W}_{j,k}, \widetilde{W}_{j',k'}) \approx 0$. This approximation improves with increasing L. In fact as $L \to \infty$ the covariance tends to zero, as the next result due to Craigmile and Percival [35] states.

Theorem 4.1: Let $\widetilde{W}_{j,k}$ and $\widetilde{W}_{j',k'}$ 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 $\operatorname{cov}(\widetilde{W}_{j,k}, \widetilde{W}_{j',k'}) \to 0$ 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 ϵ. 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 nb wavelet coefficients. This result 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 nb wavelet coefficients within a given level j are a portion of a zero mean stationary process with autocovariance sequence given by $\operatorname{cov}(\widetilde{W}_{j,k}, \widetilde{W}_{j',k+\tau}) = \sigma^2\, s_{j,\tau}(d)$ where

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

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

Hence within a particular wavelet scale the nb 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 first panel shows the spectrum of the process along with the approximate passbands that correspond to the first five wavelet levels. The second panel shows $S_j(\cdot)$ for $j = 1, \ldots, 5$. The rightmost two 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 third panel. Clearly the fourth 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 and σ^2 alone.

Now let $Y_t = T_t + X_t$, where T_t is a deterministic polynomial trend of order K, and perform a DWT on these data [26]. 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 nb wavelet coefficients if $K \leq L/2-1$; that is, only the boundary wavelet coefficients will be influenced by the trend component. The above results apply and the nb wavelet coefficients can be regarded approximately as either uncorrelated or following an AR(1) model on each level.

V. 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 nb wavelet coefficients $\{\widetilde{W}_{j,k} : j = 1, \ldots, J,\ k = 0, \ldots, M_j - 1\}$ form an independent sample with $\widetilde{W}_{j,k} \sim N(0, s_{j,0}(d)\sigma^2)$. The likelihood function for this model is

$$\begin{aligned}L_M(d, \sigma^2 | \widetilde{W}_{j,k}) &= \prod_{j=1}^{J} \prod_{k=0}^{M_j-1} (2\pi s_{j,0}(d)\sigma^2)^{-1/2} \exp\left(-\frac{\widetilde{W}_{j,k}^2}{2 s_{j,0}(d)\sigma^2}\right).\end{aligned}$$

If we let $R_j = \sum_{k=0}^{M_j-1} \widetilde{W}_{j,k}^2$ denote the sum of squares of the level j nb 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,

$$\begin{aligned}-2\, l_M(d, \sigma^2 | \widetilde{W}_{j,k}) &= M \log(2\pi\sigma^2) + \sum_{j=1}^{J} \left[M_j \log(s_{j,0}(d)) + \frac{R_j}{s_{j,0}(d)\sigma^2} \right].\end{aligned} \tag{7}$$

For a given d, the above is a function of σ^2 that is minimized when

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

Substituting this estimator into (7), we obtain a function of d alone, known as the profile log likelihood [36]:

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

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

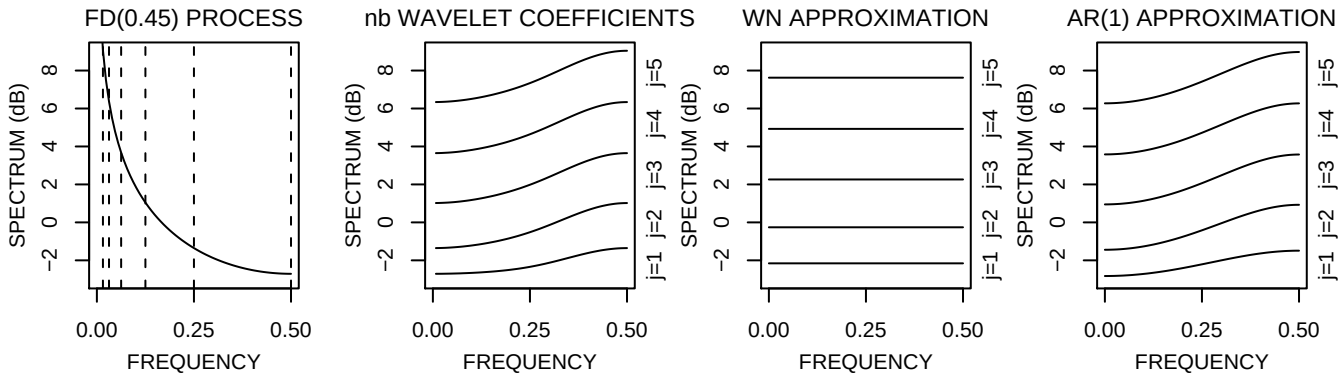

Fig. 1. Going from left to right plots show the spectrum of an FD(0.45) process (dotted vertical lines indicate the approximate passbands for the first five wavelet levels), the spectra of the LA(8) nb wavelet coefficients, and the spectra assumed in the white noise (WN) and AR(1) models.

VI. The first order autoregressive model

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

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

where $\{(Z_{nb})_{j,k} \sim N(0, \eta_j(d)\sigma^2) : j = 1 \dots J, \ k = 0 \dots M_j - 1\}$ are a set of independent random variables. The parameters of the AR(1) process on each wavelet scale j are potentially different, but across scales they are coupled together through the dependence on the FD process parameters. For any given level j, the Yule–Walker equations (e.g., Box, Jenkins, and Reinsel [1]) yield

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

Assuming again independence between coefficients on different scales, it follows from [1] that minus two times the profile likelihood is

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

where the estimate of σ^2 is given by

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

Minimizing (12) with respect to d we obtain the maximum likelihood estimator, $\widehat{d}_M$.

VII. Properties of the wavelet-based estimators

In this section we provide theory for the estimators under the models discussed in Sections V and VI. 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" [37]. We examine further properties of these estimators by simulation in Section VIII.

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 $\widehat{\boldsymbol{\theta}}_M = (\widehat{d}_M, \widehat{\sigma}_M^2(\widehat{d}_M))^T$ 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 $\Delta_1(g(x)) = \left(\frac{d}{dx}g(x)\right)/g(x)$. 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 nb 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

$$\Sigma_0(\boldsymbol{\theta}) \quad = \quad \frac{1}{2}\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & \sigma_\epsilon^{-4} \end{bmatrix},$$

with $a_{11} = \sum_{j=1}^{J} m_j \Delta_1^2(s_{j,0}(d))$, and $a_{12} = \sigma_\epsilon^{-2}\sum_{j=1}^{J} m_j \Delta_1(s_{j,0}(d))$.

(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

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

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

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

$$\frac{d}{dd}s_{j,\tau}(d) = -4\int_0^{1/2}[\log\sin(\pi f)]\cos(2^{j+1}\pi f\tau) \times\mathcal{H}_{j,L}(f)(2\sin(\pi f))^{-2d}df, \quad (13)$$

and 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 nb 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) $\sqrt{M}(\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0) \to_d N(0, \mathfrak{Z}_1^{-1}(\boldsymbol{\theta}_0))$, as $M \to \infty$, where

$$\mathfrak{Z}_1(\boldsymbol{\theta}) = \frac{1}{2}\begin{bmatrix} b_{11} & b_{12} \\ b_{12} & \sigma_\epsilon^{-4} \end{bmatrix},$$

with $b_{11} = \sum_{j=1}^J m_j\Delta_1^2(\eta_j(d))$ and $b_{12} = \sigma_\epsilon^{-2}\sum_{j=1}^J m_j\Delta_1(\eta_j(d))$.

(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

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

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

$$\frac{d}{dd}\phi_j(d) = \frac{\frac{d}{dd}s_{j,1}(d)}{s_{j,0}(d)} - \phi_j(d)\Delta_1(s_{j,0}(d)),$$
$$\frac{d}{dd}\eta_j(d) = \left(\frac{d}{dd}s_{j,0}(d)\right)\left(1-\phi_j^2(d)\right) + 2s_{j,0}(d)\phi_j(d)\left(\frac{d}{dd}\phi_j(d)\right).$$

Table I 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 confidence intervals for d based upon the above models and their profile likelihoods via the log likelihood ratio statistic $2\log\lambda(d, \widehat{\sigma}^2_M(d)) = 2\left[l_M(\widehat{d}, \widehat{\sigma}^2_M(\widehat{d})) - l_M(d, \widehat{\sigma}^2_M(d))\right]$ (the Wald or Rao test statistics could also be used to provide a confidence interval). Standard statistical theory (e.g., Lehmann [38]) suggests that an approximate $100(1-\alpha)\%$ confidence interval is then given by $\{d : 2\log\lambda(d, \widehat{\sigma}^2_M(d)) \leq q_1(1-\alpha)\}$, where here $q_1(1-\alpha)$ denotes the $(1-\alpha)$th quantile of a chi squared random variable with 1 degree of freedom.

TABLE I

CALCULATION OF $\psi_k^2(d)$ FOR VARIOUS FILTER WIDTHS L, UNDER EITHER THE WHITE NOISE (WN) ($k = 0$) 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	WN	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	WN	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	WN	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	WN	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

VIII. MONTE CARLO STUDIES

Our aim in 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 the estimators of Hurvich and Chen [22] and Veitch and Abry [12]. 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).

A. 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 1.5 in steps of 0.25. In each case we added a linear trend of the form $T_t = 0.5t/N$, $t = 0, \ldots, N-1$, and the resulting 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 [39], was performed by minimizing the negative log-profile likelihood for values of d ranging in [-1,$\min\{(L+1)/2, 3\}$] (an arbitrary choice). We used two methods for calculating $s_{j,\tau}(d)$ (defined in (6)):

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

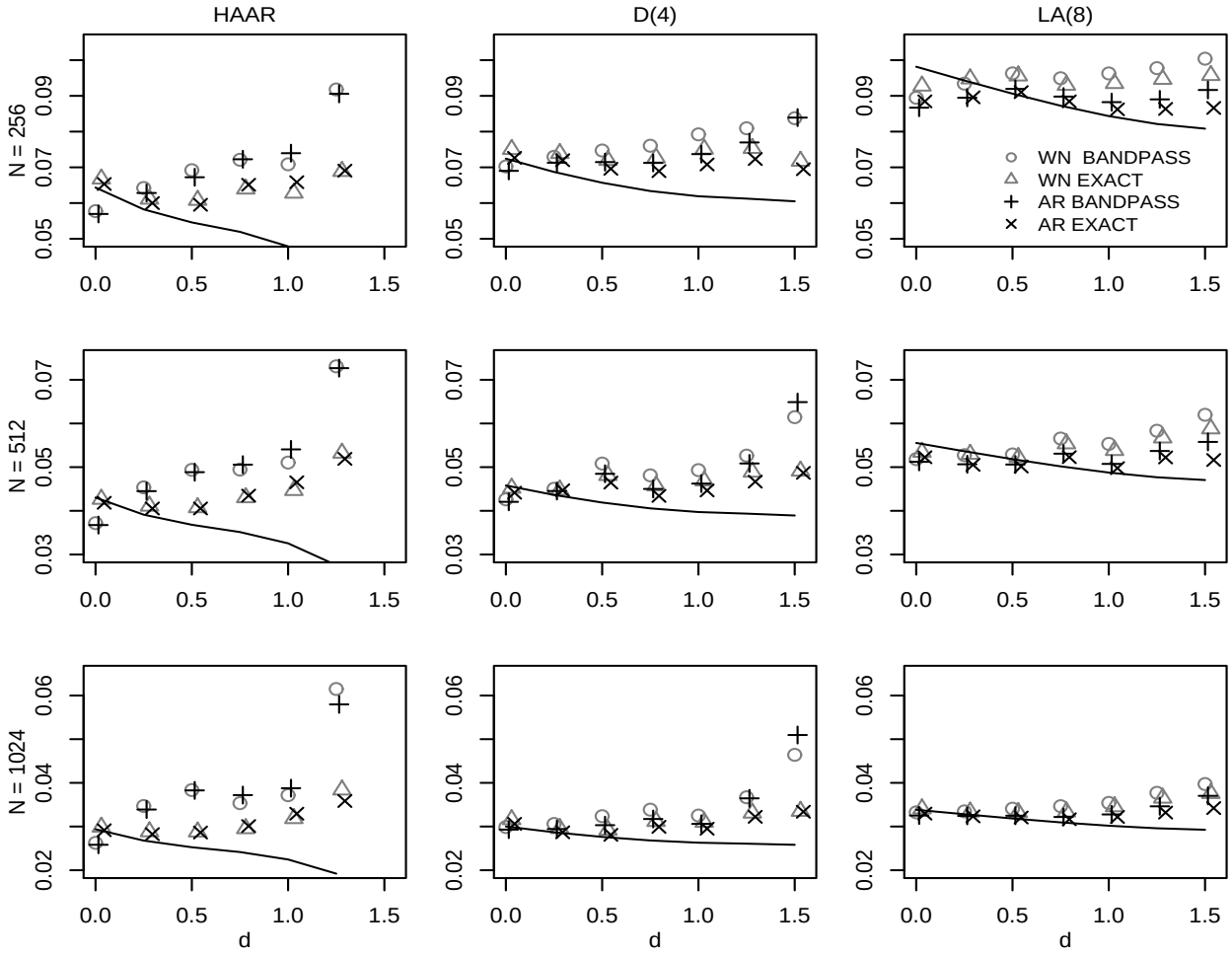

Fig. 2. RMSE in estimating the difference parameter using either the white noise or autoregressive wavelet models, for various wavelet filters, sample lengths, N, and difference parameters. The different symbols denote the different wavelet models and whether an exact or bandpass variance was used. The standard deviations of the estimated RMSEs are bounded by 0.0025. The lines on each panel denotes the theoretical RMSE, $\psi_1(d)/\sqrt{M}$, for the AR(1) method with exact variance calculations.

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 $L \to \infty$ [40]. 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.

Figure 2 also enables us to evaluate the theory of Section VII (where we assume an approximating model for the wavelet coefficients) on the basis of the simulation results. The lines on Figure 2 displays the theoretical RMSE, $\psi_1(d)/\sqrt{M}$, for the AR(1) model, calculated using the equation in Theorem 7.2 with exact variance calculations for $s_{j,\tau}(d)$. The theoretical and simulated values for the AR(1) model 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 is better approximated by longer filter widths L), using an approximating model within scale, which fits better for values of d closest to zero.

The estimation bias is not shown in Figure 2. In general the bias decreases as we increase the wavelet order, and for $d \neq 0$ is bounded by ± 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 AR(1) model using an LA(8) filter with a bandpass variance calculation are displayed in Table II). Because of the relatively low biases the variance largely determines the MSE of these estimators.

B. Comparisons with the Hurvich–Chen estimator

Hurvich and Chen [22] propose a complex valued taper which can be used to estimate d in the presence of a low order polynomial trend when $d \in (-0.5, 1.5)$. This estimator is based on the Gaussian semiparametric estimator due to

Künsch [41]. For $f_j = j/N$ such that $0 < f_j \leq 1/2$, let

$$F^{(p)}(f_j) \quad = \quad \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t \, e^{-i2\pi f_j t} \qquad (14)$$

denote the (orthonormal) discrete Fourier transform of the data [29, Chapter 3], and let $S^{(p)}(f_j) = \left|F^{(p)}(f_j)\right|^2$ denote the periodogram of the data. The Künsch [41] estimator is the value of $d \in (-0.5, 0.5)$ which minimizes

$$\begin{aligned} Q(d, \widehat{\sigma}^2_Q(d)) \quad &= \quad \log(\widehat{\sigma}^2_Q(d)) \\ &\quad -\frac{2d}{m} \sum_{j=1}^{m} \log(2\sin(\pi f_j)), \end{aligned} \qquad (15)$$

for some $m < N/2$, where $\widehat{\sigma}^2_Q(d) = m^{-1} \sum_{j=1}^{m} (2\sin(\pi f_j))^{2d} S^{(p)}(f_j)$. The refinement of Hurvich and Chen [22] is to first 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), which can perform badly for low values of d as we have overdifferenced the time series. To compensate for this, a complex spectral taper is used (basically an extension of Tukey's cosine bell taper to the complex plane). The spectral estimator for this taper is given by $S^{(hc)}(f_j) = \left|F^{(p)}(f_j) - e^{-i\pi/N} F^{(p)}(f_{j+1})\right|/2$, and then we minimize expression (15) with $S^{(p)}(f_j)$ replaced by $S^{(hc)}(f_j)$ in the expression for $\widehat{\sigma}^2_Q(d)$. Under conditions set out in Hurvich and Chen [22], this estimator, denoted GSET ('T' stands for tapered), is consistent and has a Gaussian limiting distribution. 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 [22], 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. However, in our simulations we also consider the same linear trend component as above, i.e., $T_t = 0.5t/N$, for $t = 0, \ldots, 512$. 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 II. We see that the wavelet-based estimator clearly outperforms the other two estimators in terms of RMSE, variance and 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 II 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 at $d = 0.8$), but the variance is larger. 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 [22] and Bardet et al. [13]).

C. Comparisons with the Veitch–Abry estimator

Veitch and Abry [12] consider a weighted least squares wavelet-based estimator of the long memory parameter d. Their method is based on a unbiased estimate of the wavelet variance at level j given by $\omega_j = \sum_{k=0}^{M_j-1} \widetilde{W}^2_{j,k}/M_j$. Letting $y_j = \log_2(\omega_j) - \psi(M_j/2)/\log 2 + \log_2(M_j/2)$ where $\psi(\cdot)$ denotes the digamma function, and under the assumption that the wavelet coefficients are Gaussian and independent both across and within scales, they show that

$$\begin{aligned} E(y_j) \quad &= \quad \log_2(\sigma^2) + d(2j), \\ \mathrm{var}(y_j) \quad &= \quad \zeta(2, M_j/2)/(\log 2)^2, \end{aligned}$$

for each j, where $\zeta(\cdot,\cdot)$ denotes the generalized Zeta function. Estimation of d is obtained using weighted least squares estimation in a simple linear regression model where $\{y_j : j = J_0, \ldots, J_1\}$ is the response variable, and the explanatory variable is $\{2j : j = J_0, \ldots, J_1\}$. Here $J_0, \ldots, J_1$ denotes the range of values over which the power-law relationship, i.e., $S(f) \propto |f^2|^{-d}$, holds.

Table III summarizes the results of a Monte Carlo simulation to compare the performance of the AR(1) wavelet methods with exact variances calculations to the Veitch–Abry estimator. In each case we simulated 500 realizations of an FD(d) process with additive linear trend $T_t = \beta t/512$, for $t = 0, \ldots, 511$, with the slope parameter set at $\beta = 0.5$. We analyzed each process to $J = 6$ levels using an LA(8) wavelet. For both methods we calculated the estimate of d using only the wavelet coefficients on levels $J_0, \ldots, J_1$. We fixed $J_1 = 6$, and let J_0 be either 1 or 2. As J_0 increases the spectrum will be closer to a power-law. We repeated the experiment for $d = 0$ to 1.4 in steps of 0.2, and calculated the bias, variance and MSE for each estimator. As we can see the AR(1) wavelet method has a small bias for each d, whereas, with the Veitch–Abry estimator, the bias tends to increase as d increases. The variances of both estimators are comparable, with the variance of the Veitch–Abry estimator being slightly larger than the variance of the AR(1) estimator when $J_0 = 2$. In terms of the MSE, the AR(1) wavelet method performs better for larger d. Because there are fewer wavelet coefficients included in each of the estimators, the MSEs of both estimators increase as J_0 increases.

IX. A NORTHERN HEMISPHERE TEMPERATURE SERIES

The top panel of Figure 3 shows a time series plot of the deseasonalized 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 original data come from the Climate Research Unit, University of East Anglia, UK. This updated version of the dataset incorporates combinations of grid data (over the sea and land) from 1000 extra sites, new reference periods and an increased resolution. Visually there is an indication of an upward trend and increased variability at the start of the series.

TABLE II

MONTE CARLO COMPARISON OF FOUR METHODS TO ESTIMATE d. IN EACH CASE WE SIMULATED AN $N = 513$ FD(d) SERIES WITH A LINEAR TREND, 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	0.261	0.037	0.105	0.069	0.014	0.019	0.002	0.003	0.003	-0.001	0.016	0.016
0.2	0.127	0.022	0.038	0.049	0.014	0.017	-0.008	0.003	0.003	-0.025	0.018	0.018
0.4	0.062	0.013	0.017	0.042	0.016	0.018	-0.002	0.002	0.002	-0.003	0.017	0.017
0.6	0.010	0.011	0.011	0.020	0.015	0.015	0.000	0.003	0.003	-0.012	0.016	0.017
0.8	-0.007	0.012	0.012	-0.001	0.015	0.015	-0.001	0.003	0.003	-0.009	0.016	0.016
1.2	-0.007	0.010	0.010	-0.026	0.014	0.015	-0.005	0.002	0.002	-0.017	0.017	0.018
1.4	-0.010	0.007	0.007	-0.040	0.011	0.013	-0.001	0.003	0.003	-0.010	0.018	0.018

TABLE III

MONTE CARLO COMPARISON OF THE AR(1) WAVELET AND VEITCH–ABRY METHODS OVER DIFFERENT SUBSETS OF WAVELET LEVELS, $J_0, \ldots, 6$. IN EACH CASE WE SIMULATED AN $N = 512$ FD(d) SERIES WITH A LINEAR TREND, 500 TIMES IN EACH CASE, AND ESTIMATED d USING EACH METHOD.

	AR(1), $J_0 = 1$			Veitch–Abry, $J_0 = 1$			AR(1), $J_0 = 2$			Veitch–Abry, $J_0 = 2$		
d	bias	var.	MSE	bias	var.	MSE	bias	var.	MSE	bias	var.	MSE
0.0	-0.004	0.003	0.003	-0.002	0.003	0.003	-0.007	0.005	0.006	-0.005	0.006	0.006
0.2	-0.007	0.003	0.003	-0.036	0.002	0.003	-0.007	0.006	0.006	-0.015	0.006	0.007
0.4	-0.005	0.003	0.003	-0.061	0.002	0.006	-0.010	0.006	0.006	-0.020	0.007	0.007
0.6	-0.005	0.002	0.003	-0.089	0.002	0.010	-0.004	0.005	0.005	-0.023	0.007	0.007
0.8	-0.001	0.002	0.002	-0.109	0.002	0.014	-0.003	0.005	0.005	-0.031	0.007	0.008
1.0	0.000	0.002	0.002	-0.125	0.002	0.018	-0.002	0.005	0.005	-0.030	0.007	0.008
1.2	-0.006	0.002	0.003	-0.148	0.003	0.024	-0.009	0.006	0.006	-0.042	0.008	0.010
1.4	-0.004	0.003	0.003	-0.160	0.003	0.029	-0.005	0.006	0.006	-0.040	0.009	0.011

There has been much interest in earlier versions of this time series that only went up to 1989 and were averaged over a different reference period (1950–1979). Smith [18] illustrates the problem of trying to fit an autoregressive model to the data. Using a spectral-based estimate of d, he finds significant long memory behavior in the series, with d ranging from 0.29 to 0.40 (depending on the choice of two key parameters in the estimator). Alternatively, using a kernel smoother Beran and Feng [25] obtain an ML estimate for d of 0.33 with a 95% CI of [0.19,0.46]. We now analyze the newer version of 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.

The bottom panel of Figure 3 shows a periodogram of the data. If we take the log of the spectrum given by (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 $\log(S_X(f)) \approx \log(2\sigma^2) - 2d\log(2\pi f \Delta t)$. 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. We obtain an estimate of 0.532 for $f \leq 1$, 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 4. The thick gray vertical lines denote the partition between the boundary (outside) and nb wavelet coefficients (inside) on each wavelet level. The nb 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 periodograms for the nb wavelet coefficients on each scale (not shown). From these plots the Gaussian assumption for the data seems reasonable, although the non-constant variance is evident in the lower wavelet levels by

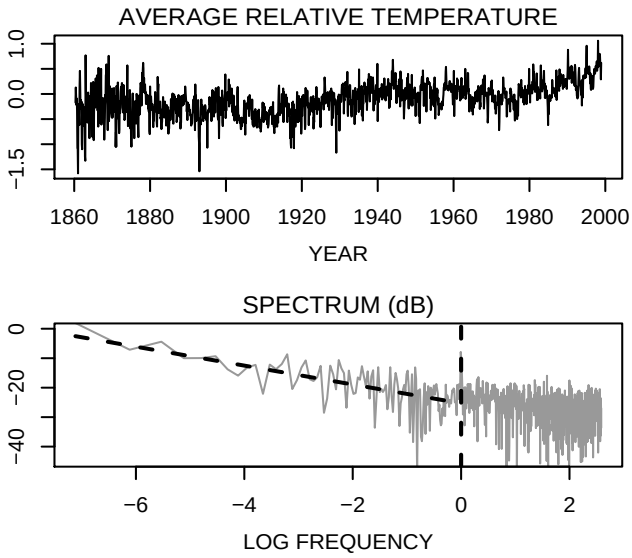

Fig. 3. The top panel denotes a time series plot of the monthly deseasonalized deviations in the northern hemisphere temperatures. The bottom panel is the corresponding periodogram. The spectrum (in decibels) is shown versus the log (base 2) frequency. The lines with the negative slope on the periodogram denotes the least squares fit for $f \leq 1$.

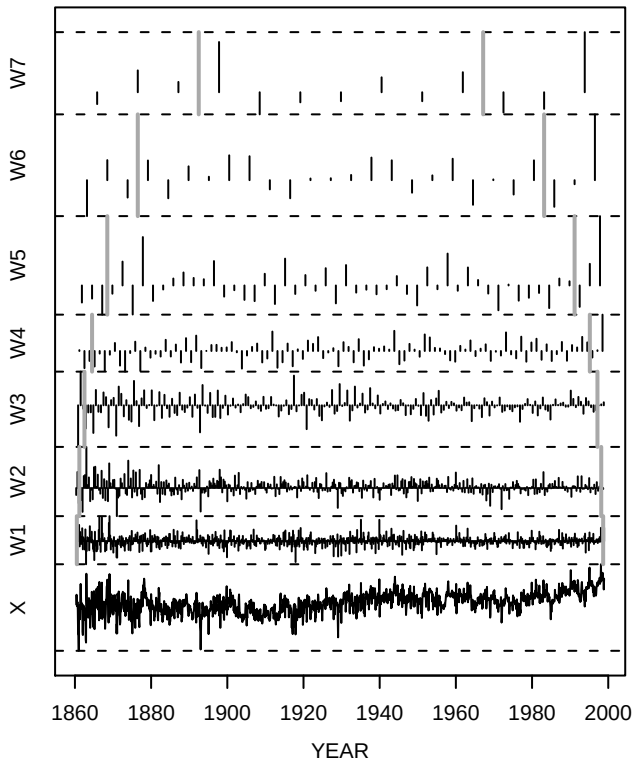

Fig. 4. Wavelet coefficients from DWT decomposition of the northern hemisphere series using an LA(8) wavelet filter analyzing levels $j = 1, \ldots, J = 7$. The thick gray vertical lines denote the partition between the boundary (outside) and non-boundary wavelet coefficients (inside) on each wavelet level.

an over-dispersion in the Q-Q plots. Lag 1 autocorrelation 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 analyzes 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. In this case the heteroscedacity in the boundary-independent wavelet coefficients reduces, and via the AR(1) wavelet model we obtain $\widehat{d}_M = 0.368$ (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 thus has little effect on the estimate of d, but the innovation variance is reduced somewhat.

An alternative to dealing with deseasonalized monthly deviates is to analyze yearly averages of the original monthly deviates. A periodogram similar to the one shown previously shows 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 [18] and Beran and Feng [25], that we should be cautious in testing for a significant trend in this series, unless we can adequately account for the long memory dependence (the question of the significance of trend can be investigated using the methods of Craigmile et al. [26]).

X. 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 [35]; 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.

Our methodology depends upon a sensible choice of the width L for the wavelet filter. Two considerations are important. The first concerns the order K of the polynomial that we are willing to consider in our statistical model. Setting K places a lower bound on L since we must have $L \geq 2(K+1)$. The second consideration is dictated by our simulations, which show that the RMSE in estimating d increases as L increases due in part to the decrease in the number of nb coefficients per scale. There is thus good reason to not let L be too much above $2(K+1)$. (Our estimator of d is designed to be impervious to trends well-modeled by low order polynomials. The question of how to use wavelets to estimate and test for such trends is addressed in Craigmile et al. [26]).

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 memory 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. This allows us to model short and long range dependence in a time series. Estimation via likelihood is supported by 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 are still adequate fits for the nb wavelet coefficients. Plots allow us to investigate this question empirically (see, e.g., Figure 1) and in practice (by looking at normal Q-Q plots, ACFs, PACFs and periodograms for the nb wavelet coefficients). Another way to estimate the magnitude of long range dependence in the presence of short range dependence, is to only consider lower frequencies of the spectrum (but not too low as to be affected by trend). Equivalently using the frequency localization of wavelet transforms, we estimate d using a subset of wavelet scales. This leads to a semiparametric wavelet-based approach for the estimation of long memory processes (e.g., Veitch and Abry [12]; Bardet et al. [13]). Selection of the range of wavelet scales to include in the estimator is critical in obtaining estimators with good statistical properties, especially in the presence of trend (see Veitch and Abry [12], [42] for a discussion of such issues under the assumption that the wavelet coefficients are independent both across and within scale).

In summary, we have investigated estimation of the parameters of polynomial 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.

REFERENCES

[1] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis, Forecasting and Control (Third Edition)*. Englewood Cliffs, New Jersey: Prentice-Hall, 1994.

[2] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, vol. 1, pp. 15–29, 1980.

[3] J. R. M. Hosking, "Fractional differencing," *Biometrika*, vol. 68, no. 1, pp. 165–176, 1981.

[4] J. Beran, *Statistics for Long Memory Processes*, ser. Monographs on Statistics and Applied Probability. New York: Chapman & Hall, 1994, vol. 61.

[5] P. M. Robinson, "Rates of convergence and optimal spectral bandwidth for long range dependence," *Probability Theory and Related Fields*, vol. 99, pp. 443–473, 1994.

[6] E. Moulines and P. Soulier, "Broadband log-periodogram regression of time series with long range dependence," *The Annals of Statistics*, vol. 27, no. 4, pp. 1415–1439, 2000.

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

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

[9] P. Abry, P. Gonçalvès, and P. Flandrin, "Wavelet-based spectral analysis of $1/f$ processes," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, 1993, pp. 237–240, minneapolis, MN, USA.

[10] ——, "Wavelets, spectrum analysis and $1/f$ processes," ser. Lecture Notes in Statistics, A. Antionadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, vol. 103, pp. 15–29.

[11] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffi c," *IEEE Trans. on Info. Theory*, vol. 44, no. 1, pp. 2–15, 1998.

[12] D. Veitch and P. Abry, "A wavelet based joint estimator of the parameters of long-range dependence," *IEEE Transactions on Information Theory*, vol. 45, no. 3, pp. 878–897, 1999.

[13] J. Bardet, G. Lang, E. Moulines, and P. Soulier, "Wavelet estimator of long range dependence processes," *Statistical Inference for Stochastic Processes*, vol. 3, no. 1-2, pp. 85–99, 2000.

[14] M. Vannucci and F. Corradi, "Modeling dependence in the wavelet domain," in *Bayesian Inference In Wavelet-Based Models*, ser. Lecture notes in statistics. New York: Springer-Verlag, 1999, ch. 12, pp. 173–186.

[15] M. J. Jensen, "An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets," *Journal of Economic Dynamics and Control*, vol. 24, no. 3, pp. 361–387, 2000.

[16] R. L. Smith, "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone," *Statistical Science*, vol. 4, pp. 367–377, 1989.

[17] ——, "Reply to comments on "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone"," *Statistical Science*, vol. 4, pp. 389–393, 1989.

[18] ——, "Long-range dependence and global warming," in *Statistics for the Environment*. Wiley (New York), 1993, pp. 141–161.

[19] V. Teverovsky and M. Taqqu, "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*, vol. 18, pp. 279–304, 1997.

[20] D. Percival and A. Bruce, "Wavelet-based approximate maximum likelihood estimates for trend-contaminated fractional difference processes," Insightful, Tech. Rep. 67, 1998.

[21] R. S. Deo and C. M. Hurvich, "Linear trend with fractionally integrated errors," *Journal of Time Series Analysis*, vol. 19, pp. 379–397, 1998.

[22] C. Hurvich and W. W. Chen, "An effi cient taper for potentially overdifferenced long memory time series," *Journal of Time Series Analysis*, vol. 21, no. 2, pp. 155–180, 2000.

[23] L. Giraitis, P. Kokoszka, and R. Leipus, "Testing for long memory in the presence of a general trend," *Journal of Applied Probability*, vol. 38, no. 4, pp. 1033–1054, 2001.

[24] R. Leipus and M.-C. Viano, "Long memory and stochastic trend," *Statistics & Probability Letters*, vol. 61, no. 2, pp. 177–190, 2003.

[25] J. Beran and Y. Feng, "SEMIFAR models – A semiparametric approach to modelling trends, long-range dependence," *Computational Statistics and Data Analysis*, vol. 40, no. 2, pp. 393–419, 2002.

[26] P. F. Craigmile, P. Guttorp, and D. B. Percival, "Assessing nonlinear trends using the discrete wavelet transform," *Environmetrics*, vol. 15, no. 4, pp. 313–335, 2004.

[27] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.

[28] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[29] D. Percival and A. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.

[30] A. Yaglom, "Correlation theory of processes with random stationary nth increments," *American Mathematical Society Translations (Series 2)*, vol. 8, pp. 87–141, 1958.

[31] R. B. Davies and D. S. Harte, "Tests for Hurst effect," *Biometrika*, vol. 74, pp. 95–101, 1987.

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

[33] P. F. Craigmile, "Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes," *Journal of Time Series Analysis*, vol. 24, no. 5, pp. 505–511, 2003.

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

[35] P. F. Craigmile and D. B. Percival, "Asymptotic decorrelation of between-scale wavelet coeffi cients," *IEEE Transactions on Information Theory*, 2004, to appear.
[36] P. McCullagh and J. A. Nelder, *Generalized Linear Models (Second Edition)*. New York: Chapman & Hall, 1989.
[37] T.-H. Li and H.-S. Oh, "Wavelet spectrum and its characterization property for random processes," *IEEE Transactions on Information Theory*, vol. 48, pp. 2922–2937, 2002.
[38] E. L. Lehmann, *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag, 1998.
[39] R Development Core Team, *R: A Language And Environment For Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2003. [Online]. Available: http://www.R-project.org
[40] M.-J. Lai, "On the digital fi lter associated with Daubechies' wavelets," *IEEE Transarctions on Signal Processing*, vol. 43, pp. 2203–2205, 1995.
[41] H. K¨unsch, "Statistical aspects of self-similar processes," in *Proceedings of the First World Congress of the Bernoulli Society*, Y. Prohorov and V. Sazanov, Eds., vol. 1. Utrecht: VNU Science Press, 1987, pp. 67–74.
[42] D. Veitch and P. Abry, "A statistical test for the time constancy of scaling exponents," *IEEE Trans. of Signal Processing*, vol. 49, no. 10, pp. 2325–2334, 2001.