Wavelet-Based Estimation Procedures for Seasonal Long-Memory Models

Brandon Whitcher
EURANDOM
P.O. Box 513
5600 MB Eindhoven
The Netherlands
whitcher@eurandom.tue.nl

May 16, 2000

Abstract

We introduce the multiscale analysis of seasonal persistent processes; i.e., time series models with a singularity in their spectral density function at a or multiple frequencies in $[0, 1/2]$. The discrete wavelet packet transform (DWPT) and an non-decimated version of it known as the maximal overlap DWPT (MODWPT) are introduced as an alternative method to spectral techniques for analyzing time series that exhibit seasonal long-memory. Approximate maximum likelihood estimation is performed by replacing the variance/covariance matrix with diagonalized matrix based on the ability of the DWPT to approximately decorrelate a seasonal persistent process. Simulations are performed using this wavelet-based maximum likelihood technique on a variety of time series models. An application of this methodology to atmospheric CO_2 measurements is also presented.

Keywords. Discrete wavelet packet transform, Gegenbauer process, long memory, multitaper spectral estimation, periodogram, wavelet variance.

1 Introduction

A simple generalization of the fractional difference process (or fractional ARIMA model) was mentioned, in passing, by Hosking (1981) and allows the singularity in the spectral density function (SDF) of the process to be located at any frequency $0 < f < 1/2$. Such a process has been referred to as a Gegenbauer process (Gray *et al.* 1989) and also a seasonal persistent process (SPP) (Anděl 1986). We prefer the latter term because it more accurately and concisely describes the content of the time series. That is, a sinusoid of particular frequency is associated with the singularity present in the spectrum thus causing a persistent oscillation in the process. The autocovariance sequence is a hyperbolically decaying oscillation. While Gray *et al.* (1989) fit an SPP to the Wolfer sunspot data, where short-range dependence was allowed through fitting ARMA components, recent attention has also appeared in the economics literature; e.g., Ooms (1995), Lobato (1997) and Arteche and Robinson (1999).

The discrete wavelet transform (DWT) has been widely used in the analysis of time series which exhibit long-range dependence; usually characterized by a slowly decaying autocovariance sequence or SDF with a singularity at the frequency $f = 0$. The DWT is efficiently implemented through a series of filtering and downsampling operations via the pyramid algorithm (Mallat 1989). Let $\{h_l\}$ and $\{g_l\}$ denote the unit scale wavelet (high-pass) and scaling (low-pass) filters, respectively. Let $H(f)$ and $G(f)$ denote the transfer functions (Fourier transforms) for the filters $\{h_l\}$ and $\{g_l\}$, respectively. Wavelet filters for higher scales are obtained through the inverse Fourier transform of $H_j(f) = H(2^{j-1}f)\prod_{l=0}^{j-2} G(2^l f)$ for $0 \le f \le 1/2$. The key feature of the wavelet filters is that $H_j(0) = 0$ for all j, and hence, the effect of the singularity in the spectrum is essentially eliminated for each scale of wavelet coefficients. The downsampling operation corresponds to stretching and folding the SDF of the filtered coefficients, thereby 'flattening' their spectrum even further. The result of these operations is that the wavelet coefficients of a long memory process are approximately uncorrelated.

Although the DWT works very well for long memory processes, it fails to approximately decorrelate processes with more general SDFs. An example of such a process is the MA(1) time series model $X_t = \epsilon_t - \theta\epsilon_{t-1}$ which is dominated by higher frequencies in its spectrum as $\theta \to -1$. The SDF of this process exhibits a peak in its spectrum at $f = 1/2$. The orthonormal basis associated with the DWT corresponds to a partitioning of the frequency axis, which is all wrong for this MA(1) process. A generalization of the DWT – the discrete wavelet packet transform (DWPT) – involves

a redundant partitioning of the frequency axis. An orthonormal transform may be thought of as a subset of basis functions (disjoint frequency partitions) of the DWPT. The DWT is one example of such a subset.

The model of interest here, the seasonal persistent process, is introduced in Section 2. Its spectrum and asymptotic expression of the autocorrelation function are provided. Section 3 gives a brief introduction to the DWPT and a non-decimated version of it – the maximal overlap DWPT. Parameter estimation for seasonal persistent processes, through approximate maximum likelihood in the wavelet domain, is provided in Section 4. Results from simulations are given in Section 5 and we anlayze monthly CO_2 measurements from Mauna Loa, Hawaii, in Section 6. Conclusions are provided in Section 7.

2 Seasonal Persistent Processes

2.1 Single Factor Model

Let $\{Y_t\}$ be a stochastic process such that

$$Y_t = \left(1 - 2\phi B + B^2\right)^{-\delta} \epsilon_t \tag{1}$$

is a stationary process, then $\{Y_t\}$ is a seasonal persistent process (SPP); a simple example of a seasonal long-memory process. Gray *et al.* (1989) showed that $\{Y_t\}$ is stationary and invertible for $|\phi| = 1$ and $-1/4 < \delta < 1/4$ or $|\phi| < 1$ and $-1/2 < \delta < 1/2$. Clearly, the definition of an SPP also includes a fractional difference process; a simple example of a long-memory process. When $\phi = 1$ we have that $\{Y_t\}$ is a fractional difference process with parameter $d = 2\delta$ (Hosking 1981). If $\{\epsilon_t\}$ is a Gaussian white noise process, then $\{Y_t\}$ is also called a Gegenbauer process since Equation (1) may be written as an infinite moving-average process via

$$Y_t = \sum_{k=0}^{\infty} C_{k,\phi}^{(\delta)} \epsilon_{t-k},$$

where $C_{k,\phi}^{(\delta)}$ is a Gegenbauer polynomial (Rainville 1960, Ch. 17). The SDF of $\{Y_t\}$ is given by

$$S_Y(f) = \sigma_\epsilon^2 \{2|\cos(2\pi f) - \phi|\}^{-2\delta}, \quad \text{for} \quad -\frac{1}{2} < f < \frac{1}{2}, \tag{2}$$

so that $S_Y(f)$ becomes unbounded at frequency $f_G \equiv (\cos^{-1}\phi)/(2\pi)$, sometimes called the Gegenbauer frequency. Example spectra are given in Figure 1, these are taken from Anděl (1986).

The autocovariance sequence of an SPP may be expressed via

$$s_\tau = \int_{-1/2}^{1/2} S_Y(f) \cos(2\pi f \tau)\, df.$$

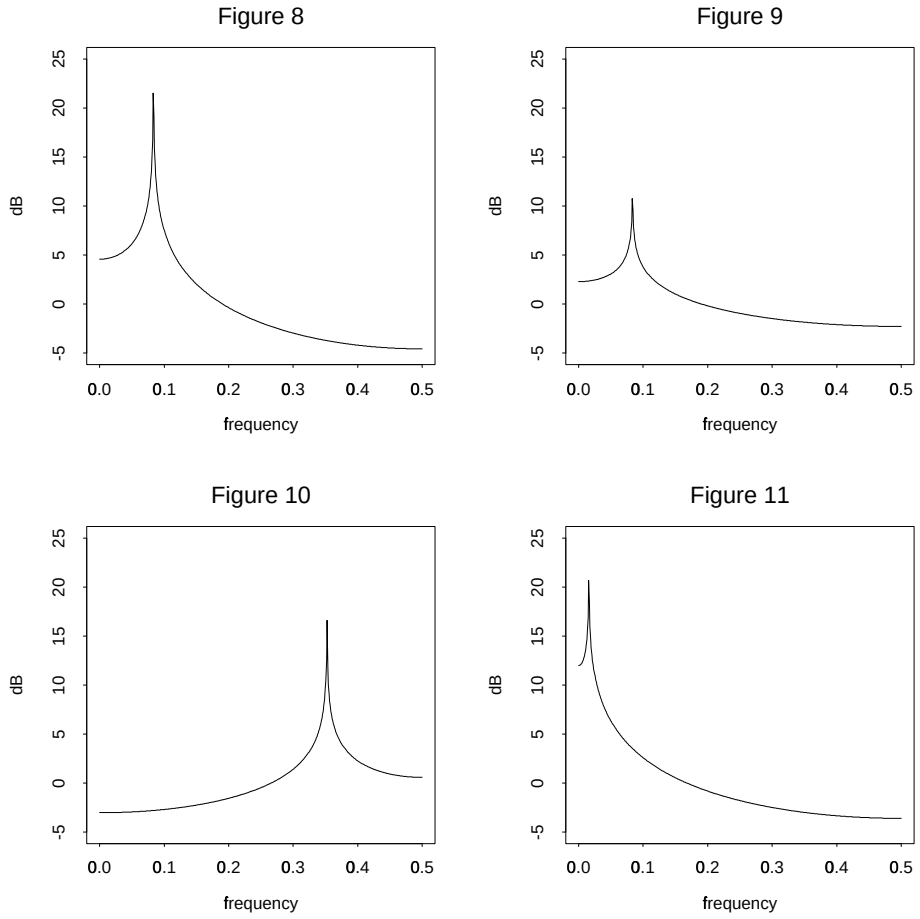


Figure 1: Spectral density functions (in decibels) for SPPs with (8) $\delta = 0.4$, $f_G = 1/12$, (9) $\delta = 0.2$, $f_G = 1/12$, (10) $\delta = 0.3$, $f_G = 0.352$ and (11) $\delta = 0.3$, $f_G = 0.016$. These are Figures 8-11 in Anděl (1986).

An explicit solution is known only for special cases. Gray *et al.* (1994) showed that the autocorrelation sequence of an SPP is given by

$$r_\tau \sim \tau^{2\delta-1} \cos(2\pi f_G \tau) \quad \text{as} \quad \tau \to \infty. \tag{3}$$

Two sequences are related via $a_\tau \sim b_\tau$ as $\tau \to \infty$ if $\lim_{\tau\to\infty}\{a_\tau/b_\tau\} = c$ where c is a finite nonzero constant.

2.2 k-factor Model

An obvious extension of the single factor SPP would be to allow multiple singularities to appear in the SDF of the process. Consider the k-factor Gegenbauer process $\{Y_t\}$ (Woodward *et al.* 1998)

with zero mean given by

$$Y_t = \prod_{i=1}^{k} \left(1 - 2\phi_i B + B^2\right)^{-\delta_i} \epsilon_t,$$

exhibiting k asymptotes located at the frequencies $f_i = (\cos^{-1}\phi_i)/(2\pi)$, $i = 1, \ldots, k$, in its spectrum

$$S_Y^{(k)}(f) \equiv \sigma_\epsilon^2 \prod_{i=1}^{k} \{2|\cos(2\pi f) - \phi_i|\}^{-2\delta_i},$$

$|f| < 1/2$. Using this model allows one to incorporate several observed oscillations, such as the fundamental frequency and its harmonics.

3 Discrete Wavelet Packet Transforms

The orthonormal discrete wavelet transform (DWT) has a very specific band-pass structure which partitions the spectrum of a long memory process finer and finer as $f \to 0$, where the spectrum is unbounded. This is done through a succession of low-pass and high-pass filtering operations; see, for example, Percival and Walden (2000, Ch. 4) for an in depth introduction to the DWT. In order to exploit the approximate decorrelation property for seasonally persistent processes we need to generalize the partitioning scheme of the DWT. This is easily obtained by performing the discrete wavelet packet transform (DWPT) on the process; see, e.g., Wickerhauser (1994, Ch. 7) and Percival and Walden (2000, Ch. 6). Instead of one particular filtering sequence, the DWPT executes all possible filtering combinations to obtain a wavelet packet tree, denoted by $\mathcal{T} = \{(j,n) \mid j = 0, \ldots, J; n = 0, \ldots, 2^j - 1\}$. An orthonormal basis $\mathcal{B} \subset \mathcal{T}$ is obtained when a collection of DWPT coefficients is chosen, whose ideal band-pass frequencies are disjoint and cover $[0, 1/2]$.

3.1 The Discrete Wavelet Packet Transform

We start of with a vector of observations $\mathbf{X}$, and let $h_0, \ldots, h_{L-1}$ be the unit scale wavelet (high-pass) filter coefficients from a Daubechies compactly supported wavelet family (Daubechies 1992) of even length L. In the future, we will denote the Daubechies family of extremal phase compactly supported wavelets with D(L) and the Daubechies family of least asymmetric compactly supported wavelets with LA(L). Another interesting family of wavelets that better approximate ideal band-pass filters, for comparable lengths, is the minimum-bandwidth discrete-time (MBDT) wavelets of Morris and Peravali (1999). We denote them by MB(L) and will utilize them in the future.

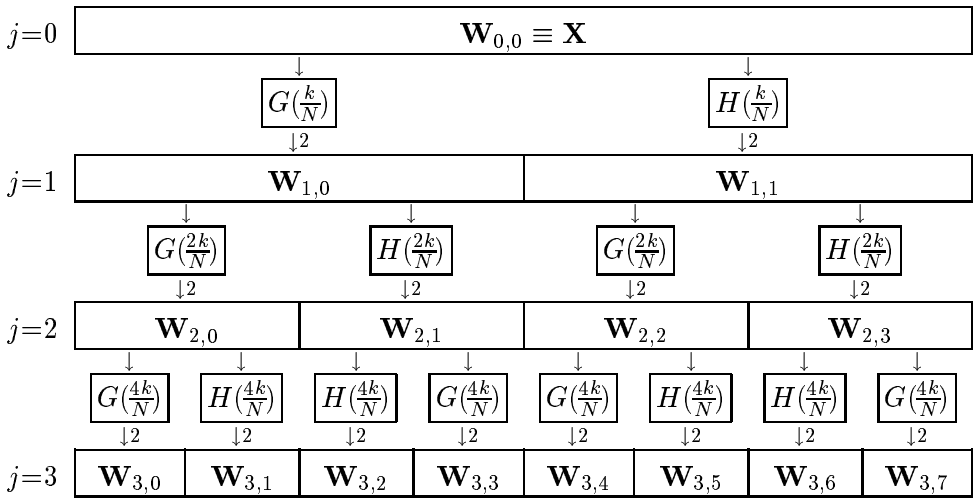


Figure 2: Graphical representation of the DWPT applied to a dyadic length vector $\mathbf{X}$.

The scaling (low-pass) coefficients may be computed via the quadrature mirror relationship

$$g_l = (-1)^{l+1} h_{L-l-1}, \quad l = 0, \ldots, L-1.$$

Now define

$$u_{n,l} \equiv \begin{cases} g_l, & \text{if } n \bmod 4 = 0 \text{ or } 3; \\ h_l, & \text{if } n \bmod 4 = 1 \text{ or } 2, \end{cases}$$

to be the appropriate filter at a given node of the wavelet packet tree. This ordering is necessary in order to force the frequency intervals to be monotonically increasing; it is also called a *sequency* ordering by Wickerhauser (1994).

Let $W_{j,n,t}$ denote the tth element of the length $N_j \equiv N/2^j$ wavelet coefficient vector $\mathbf{W}_{j,n}$, $(j,n) \in \mathcal{T}$. Given the DWPT coefficients $\mathbf{W}_{j-1,\lfloor \frac{n}{2} \rfloor,t}$, of length N_{j-1}, we can directly compute $W_{j,n,t}$ via

$$W_{j,n,t} \equiv \sum_{l=0}^{L-1} u_{n,l} W_{j-1,\lfloor \frac{n}{2} \rfloor, 2t+1-l \bmod N_{j-1}}, \quad t = 0, 1, \ldots, N_j - 1.$$

Note, the recursion is started off with the data such that $\mathbf{W}_{0,0} \equiv \mathbf{X}$. This is one possible formulation of the DWPT, we may also directly filter the observations by generating unique filter coefficients at each level or apply a series of matrix operations; see Percival and Walden (2000, Ch. 6) for more details. As with the DWT, the DWPT is most efficiently computed using a pyramid algorithm (Mallat 1989) of filtering and downsampling steps. Figure 2 provides a graphical representation of the filtering operations required to construct the vectors of a wavelet packet tree down to level 3.

The notation $\mathbf{X} \rightarrow \boxed{H(f)} \stackrel{\downarrow 2}{\rightarrow} \mathbf{W}$ means that the length N vector $\mathbf{X}$ has been convolved with the filter $\{h_l\}$, whose Fourier transform is $H(f)$, and downsampled by two in order to produce a new vector $\mathbf{W}$ of length $N/2$. The algorithm has $O(N \log N)$ operations, like the discrete fast Fourier transform.

An analysis (decomposition) of variance of the original time series may be performed via the DWPT by selecting an orthonormal basis $\mathcal{B}$; i.e.,

$$\|\mathbf{X}\|^2 = \sum_{(j,n) \in \mathcal{B}} \|\mathbf{W}_{j,n}\|^2.$$

Let us define the wavelet packet variance $\nu^2(\lambda_{j,n})$ associated with frequencies in the interval $\lambda_{j,n} \equiv (n/2^{j+1}, (n+1)/2^{j+1}]$ to be the variance of the DWPT coefficients $\mathbf{W}_{j,n}$. The unbiased DWPT-based estimator is given by

$$\hat{\nu}^2(\lambda_{j,n}) \equiv \frac{1}{N'_j 2^j} \sum_{t=L'_j}^{N/2^{j+1}-1} W_{j,n,t}^2,$$

where $N'_j \equiv N/2^{j+1} - L'_j$ and $L'_j \equiv \lceil (L-2)(1-1/2^j) \rceil$. The estimator is unbiased because all coefficients that are affected by the boundary have been removed.

3.2 The Maximal Overlap Discrete Wavelet Packet Transform

Definition of the maximal overlap DWPT (MODWPT) is straightforward, given the DWPT. Simply define the new filter $\tilde{u}_{n,l} \equiv u_{n,l}/2^{1/2}$, replace it with the filter for computing DWPT coefficients and do *not* downsample the filtered output. Hence, the vector of MODWPT coefficients $\widetilde{\mathbf{W}}_{j,n}$ is computed recursively via

$$\widetilde{W}_{j,n,t} \equiv \sum_{l=0}^{L-1} \tilde{u}_{n,l} \widetilde{W}_{j-1,\lfloor \frac{n}{2} \rfloor, t-2^{j-1} l \bmod N}, \quad t = 0, 1, \ldots, N-1.$$

Thus, each vector of MODWPT coefficients has length N (to begin the recursion define $\widetilde{\mathbf{W}}_{0,0} \equiv \mathbf{X}$). This formulation leads to efficient computation using a pyramid-type algorithm (Percival and Walden 2000, Ch. 6). As with the DWPT, the MODWPT is an energy preserving transform and we may define an unbiased MODWPT-based estimator of the wavelet packet variance to be

$$\tilde{\nu}^2(\lambda_{j,n}) \equiv \frac{1}{N_j} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,n,t}^2,$$

where $N_j \equiv N - L_j + 1$ and $L_j = (2^j - 1)(L-1) + 1$. As with the DWPT-based estimator, all coefficients affected by the boundary have been removed for the calculation.

Given a particular level j of the transform, we may also reconstruct $\mathbf{X}$ by projecting the MODWPT coefficients back onto the filter coefficients via

$$X_t = \sum_{n=0}^{2^j-1} \sum_{l=0}^{L_j-1} \tilde{u}_{j,n,l} \widetilde{W}_{j,n,t+l \bmod N}, \quad t = 0, \ldots, N-1. \tag{4}$$

Let $\widetilde{\mathcal{D}}_{j,n}$ be the wavelet packet detail associated with the frequency interval $\lambda_{j,n}$. Then

$$\widetilde{\mathcal{D}}_{j,n,t} \equiv \sum_{l=0}^{L_j-1} \tilde{u}_{j,n,l} \widetilde{W}_{j,n,t+l \bmod N}, \quad t = 0, \ldots, N-1,$$

and an additive decomposition in Equation (4) may be rewritten as $X_t = \sum_{(j,n)\in\mathcal{B}} \widetilde{\mathcal{D}}_{j,n,t}$ for any orthonormal basis $\mathcal{B}$. These details, when using Daubechies LA(L) wavelets, are associated with zero-phase filters and there features line up perfectly with those in the original time series $\mathbf{X}$ at the same time (Percival and Walden 2000, Sec. 6.6).

4 Parameter Estimation for Seasonal Persistence

The common techniques for estimating the long-memory parameter for a fractional ARIMA model have recently been extended to SPPs, including log-periodogram and semiparametric analysis (Arteche and Robinson 2000). As an alternative to the periodogram, the wavelet variance has proved quite effective in estimating the long-memory parameter in fractional ARIMA models (McCoy and Walden 1996; Jensen 1999a, 1999b). We introduce methodology for estimating single and multiple factor seasonal time series models using approximate likelihood.

4.1 Initial Parameter Estimates

To facilitate the rapid convergence to a solution of the likelihood, we introduce initial estimates $(\hat{\delta}_0, \hat{f}_{G,0})$ for the SPP of interest. The Gegenbauer frequency, for a single-factor model, is straightforward to estimate by simply taking the Fourier frequency associated with the maximum periodogram coordinate. Using this we may estimate the fractional difference parameter through least squares regression across the wavelet packet variances $\nu^2(\lambda_{j,n})$ for $(j,n) \in \mathcal{B}$. This follows from the fact that the wavelet variance is an estimator of the SDF on $\lambda_{j,n}$ and the particular form of the spectrum of an SPP to yield

$$\log \nu^2(\lambda_{j,n}) = \alpha + \beta \log 2 |\cos(2\pi\mu_{j,n}) - \cos(2\pi f_G)|, \tag{5}$$

where $\mu_{j,n}$ is the midpoint of the frequency interval $\lambda_{j,n}$. To be precise, the (j,n)th MODWPT variance covers the entire interval of frequencies $\lambda_{j,n}$ but it suffices to represent this interval by its midpoint here. The slope from a simple linear regression of $\log \tilde{\nu}^2(\lambda_{j,n})$ on $\log 2|\cos(2\pi\mu_{j,n}) - \cos(2\pi f_G)|$, appropriately normalized, provides an estimate of the fractional difference parameter via $\hat{\delta}_0 = -\beta/2$. Simplifying Equation (5) to just the frequencies, and not the full SDF, yields

$$\log \nu^2(\lambda_{j,n}) \approx \alpha + \beta \log 2|\mu_{j,n} - f_G|. \tag{6}$$

We make use of Equation 6 to determine $\hat{\delta}_0$ in practice. The least-squares estimator is also worthy of further investigation given its simplicity to compute.

4.1.1 Approximate Maximum Likelihood Estimation

McCoy and Walden (1996) and Jensen (1999a) have both provided an approximate maximum likelihood estimator (MLE) to the fractional difference parameter for long-memory time series models. The DWT provides a simple and effective method for approximately diagonalizing the variance/covariance of the original process. We extend their results to the case of SPPs, where two parameters δ and f_G define the SDF. As before, we utilize the DWPT under a particular basis function $\mathcal{B}$ to approximately diagonalize the variance/covariance matrix of an SPP.

Let $\mathbf{X}$ be a realization of a zero mean stationary SPP with unknown parameters δ, f_G and $\sigma_\epsilon^2 > 0$. The likelihood function for $\mathbf{X}$, under the assumption of multivariate Gaussianity, is given by

$$L(\delta, f_G, \sigma_\epsilon^2 \mid \mathbf{X}) \equiv \frac{1}{(2\pi)^{N/2}|\Sigma_{\mathbf{X}}|^{1/2}} e^{-\mathbf{X}^T\Sigma_{\mathbf{X}}^{-1}\mathbf{X}/2},$$

where $\Sigma_{\mathbf{X}}$ is the variance/covariance matrix of $\mathbf{X}$ and $|\Sigma_{\mathbf{X}}|$ is the determinant of $\Sigma_{\mathbf{X}}$. As previously alluded to, we avoid computing the exact MLEs of the parameters of interest and instead use the DWPT to approximately diagonalize $\Sigma_{\mathbf{X}}$; i.e., $\Sigma_{\mathbf{X}} \approx \widehat{\Sigma}_{\mathbf{X}} \equiv \mathcal{W}_{\mathcal{B}}^T \Omega_N \mathcal{W}_{\mathcal{B}}$, where $\mathcal{W}_{\mathcal{B}}$ is an $N \times N$ orthonormal matrix defining the DWPT through the basis $\mathcal{B}$ and Ω_N is a diagonal matrix containing the band-pass variances of an SPP. It is convenient to work with the rescaled band-pass variance $\omega_{j,n}^2 \equiv \sigma_\epsilon^2 \bar{\omega}^2(\lambda_{j,n})$ such that

$$\bar{\omega}_{j,n}^2 \equiv 2^{j+1} \int_{\frac{n}{2^{j+1}}}^{\frac{n+1}{2^{j+1}}} \frac{1}{\{4[\cos(2\pi f) - \cos(2\pi f_G)]^2\}^\delta}\, df,$$

for all $(j,n) \in \mathcal{B}$. The approximate log-likelihood function is now

$$\begin{aligned}\widehat{\mathcal{L}}(\delta, f_G, \sigma_\epsilon^2 \,|\, \mathbf{X}) &\equiv -2\log\left(\widehat{L}(\delta, f_G, \sigma_\epsilon^2 \,|\, \mathbf{X})\right) - N\log(2\pi) \\ &= N\log(\sigma_\epsilon^2) + \sum_{(j,n)\in\mathcal{B}} N_{j,n}\log(\bar{\omega}_{j,n}^2) + \frac{1}{\sigma_\epsilon^2}\sum_{(j,n)\in\mathcal{B}} \frac{\mathbf{W}_{j,n}^T\mathbf{W}_{j,n}}{\bar{\omega}_{j,n}^2}. \qquad (7)\end{aligned}$$

Differentiating Equation (7) with respect to σ^2 and setting the result equal to zero, the MLE of σ^2 is equal to

$$\hat{\sigma}_\epsilon^2(\delta, f_G) \equiv \frac{1}{N}\sum_{(j,n)\in\mathcal{B}} \frac{\mathbf{W}_{j,n}^T\mathbf{W}_{j,n}}{\bar{\omega}_{j,n}^2}.$$

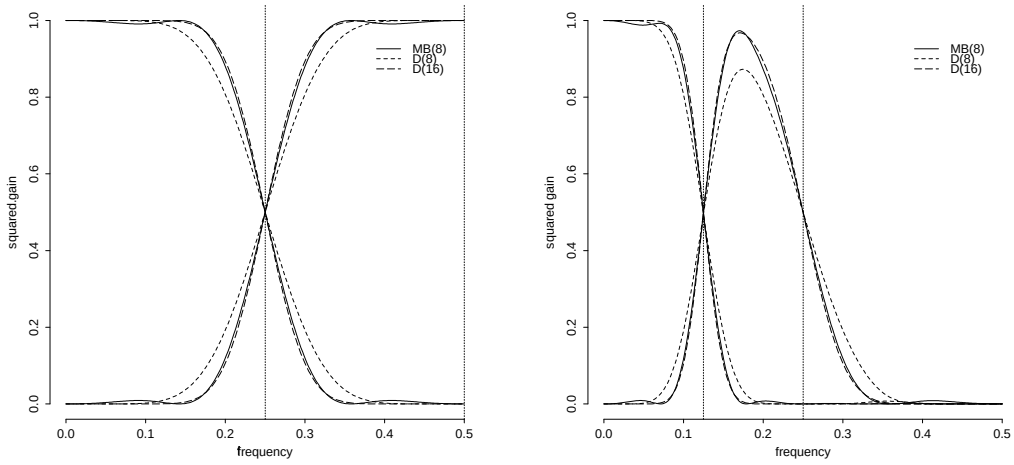
Replacing σ_ϵ^2 with its MLE, we reduce the complexity of Equation (7) to obtain

$$\widehat{\mathcal{L}}(\delta, f_G \,|\, \mathbf{X}) = N\log(\hat{\sigma}_\epsilon^2(\delta, f_G)) + \sum_{(j,n)\in\mathcal{B}} N_{j,n}\log(\bar{\omega}_{j,n}^2). \qquad (8)$$

The reduced log-likelihood in Equation (8) is now a function of only two parameters δ and f_G, whose space of possible solutions lives on $(-1/2, 1/2) \times (0, 1/2)$. In most practical situations, we will be interested in fractional difference parameters which are strictly positive (thus reducing the solution space even more). This estimation procedure differs from the frequency-based semiparametric estimator of Arteche and Robinson (2000) by simultaneously determining MLEs for both the fractional difference parameters and Gegenbauer frequency, whereas their procedure requires user-specified frequencies for the asymptotes in the SDF of the model.

4.2 Basis Selection Procedure

Given that we are working with time series that exhibit a wide range of characteristics, through rather loose assumptions on their SDFs, selecting the orthonormal basis for the wavelet transform is important. We want to adapt as best as possible to the underlying SDF, but only have the observations to help us. For long-memory processes, the DWT works extremely well at approximately decorrelating the process (McCoy and Walden 1996). Whitcher (1998) related this ability to the fact that the SDFs of the wavelet coefficient vectors are essentially flat; e.g., only varying by 3 dB for the unit scale DWT coefficients when the fractional difference (long-memory) parameter is associated with stationary and invertible fractional ARIMA models ($-1/2 < d < 1/2$). Figure 3 shows the squared gain functions of the wavelet and scaling filters for the first two levels of the DWT; these correspond to the frequency intervals $\lambda_{1,0}, \lambda_{1,1}, \lambda_{2,0}, \lambda_{2,1}$. As noted in Whitcher (1999), the approximation to an ideal band-pass filter by the wavelet packet filter is crucial to successfully



(a) $\lambda_{1,0}$ and $\lambda_{1,1}$ (b) $\lambda_{2,0}$ and $\lambda_{2,1}$

Figure 3: Squared gain functions for the MB(8), D(8) and D(16) wavelet filters. The ideal pass-band is found between the dotted lines. The frequency intervals displayed correspond to the orthonormal basis associated with the DWT ($J = 2$).

producing approximately uncorrelated wavelet coefficient vectors. Hence, the ability of the MB(8) wavelet filter to achieve the same approximation as a Daubechies wavelet filter twice its length is highly desirable.

A constant SDF is associated with a white noise process, where $\int S(f)\, df = \sigma^2$. Several methods have been proposed in order to test for white noise in time series, such as the cumulative periodogram and portmanteau tests; see, e.g., Brockwell and Davis (1991). There are a few variations of the portmanteau test, two applied to the sample autocorrelations of the raw data and one applied to the sample autocorrelations of the squared data. A cumulative sum of squares (CSS) test statistic was proposed by Brown *et al.* (1975) for testing the constancy of regression relationships over time and successfully applied to test for nonstationary features in the output from the DWT by Whitcher *et al.* (1998).

Figure 4 shows results from a small simulation study comparing the three proposed methods for selecting an orthonormal basis. Realizations from an SPP ($f_G = 1/12, \delta = 0.4$), of length $N = 1024$, were generated using an exact time-domain method (Hosking 1984). A partial DWPT ($J = 6$) was applied using the MB(8) wavelet filter. As seen from the figure, all methods capture the general shape of the SDF; see Figure 1. However, the portmanteau and cumulative periodogram tests appear to select a basis which closely matches the true SDF more often than the CSS test (this

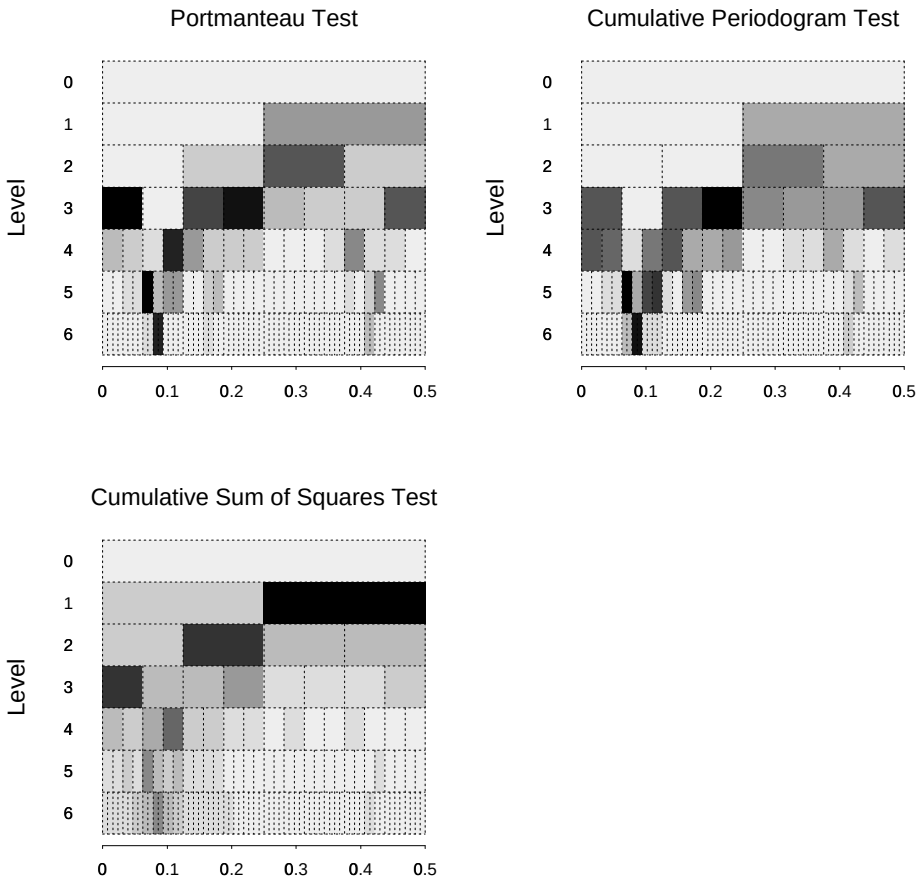


Figure 4: Wavelet packet table ($J = 6$) for three white noise tests applied to an SPP ($N = 1024$) with $\delta = 0.4$ and $f_G = 1/12$, summed over 500 simulations. The frequency a particular node of the wavelet packet tree $\mathcal{T}$ was chosen in the orthonormal basis $\mathcal{B}$ is given by the shade of the rectangle – darker shades correspond to higher frequencies. All hypothesis tests were performed at the $\alpha = 0.05$ level of significance.

is seen by the higher number of 'dark' rectangles in the portmanteau and cumulative periodogram wavelet packet tables). This is not surprising, since the CSS test is not specifically designed to test for white noise, only for constant variance.

One disadvantage of the cumulative periodogram test is that, by applying the DFT to each vector of wavelet packet coefficients, the number of values used in the test is halved. Given the inherent downsampling of the DWPT, each level j of the DWPT has only $N/2^j$ coefficients, and hence, the cumulative periodogram test will only contain $N/2^{j+1}$ periodogram ordinates. This is quite restrictive on the depth of the DWPT for a given sample size.

5 Simulations

To assess the performance of this approximate ML methodology, we simulate the four time series in Figure 1 using numeric integration to compute their autocovariance sequences. Table 1 summarizes the results of this simulation study for 500 iterations. The average MLEs $\hat{\delta}$ and $\hat{f}_G$ are given along with their empirical bias, standard deviation and empirical mean squared error (MSE). These time series models provide an adequate representation of seasonal long-memory processes. The first two provide an annual periodicity with two levels of persistence. The orthonormal basis was chosen by applying a portmanteau test to the squared wavelet coefficients for all vectors in the wavelet packet table.

When $\delta = 0.4$, the MLEs for all three wavelet filters exhibit a slight negative bias but the estimated Gegenbauer frequency is right on. Reduce the level of persistence to $\delta = 0.2$ does not diminish the ability of the method to accurately estimate the parameters, in fact both the fractional difference parameter and Gegenbauer frequency show reduced bias and only slightly increased standard deviation versus $\delta = 0.4$. The third model provides a high-frequency oscillation $f_G = 0.3524$ and its MLEs show a slight negative bias in the case of $\hat{\delta}$ and negligible positive bias in the Gegenbauer frequency. The empirical MSEs are greater than those observed. This pattern of empirical bias and MSE is similar when the Gegenbauer frequency is reduced to $f_G = 0.159$, producing a very low-frequency (large period) oscillation. There is a slight improvement in using a wavelet filter that better approximates an ideal band-pass filter, the MB(16), but not overwhelming evidence with respect to these specific time series models.

6 Application to Atmospheric CO_2 Data

Woodward *et al.* (1998) analyzed monthly atmospheric CO_2 measurements from the Mauna Loa Observatory, Hawaii. We analyze an extended version of these data obtained from the Carbon Dioxide Information Analysis Center (CDIAC) website[1]. The current record of CO_2 measurements is from 1958 through 1998, but contains several missing values in the early years. The longest continuous record begins in June 1964 with $N = 416$; see Figure 5. We observe an obvious periodicity and time-dependent mean structure in the series.

Exploratory analysis was performed via a multiresolution analysis of the data; see Figure 6. For simplicity, the standard orthonormal basis was used to apply a partial MODWT ($J = 6$) to these

[1] http://cdiac.esd.ornl.gov/ndps/ndp001.html

Model (δ, f_G)		MB(8) $\hat{\delta}$	MB(8) $\hat{f}_G$	LA(16) $\hat{\delta}$	LA(16) $\hat{f}_G$	MB(16) $\hat{\delta}$	MB(16) $\hat{f}_G$
				$N = 128$			
(0.4, 0.0833)	mean	0.3733	0.0845	0.3842	0.0857	0.3822	0.0827
	bias	−0.0267	0.0011	−0.0158	0.0024	−0.0178	−0.0006
	SD	0.0509	0.0243	0.0473	0.0222	0.0484	0.0224
	RMSE	0.0575	0.0243	0.0498	0.0223	0.0516	0.0224
(0.2, 0.0833)	mean	0.2006	0.0854	0.1999	0.0861	0.1991	0.0830
	bias	0.0006	0.0021	−0.0001	0.0028	−0.0009	−0.0003
	SD	0.0609	0.0543	0.0577	0.0581	0.0588	0.0596
	RMSE	0.0609	0.0542	0.0576	0.0582	0.0588	0.0595
(0.3, 0.3524)	mean	0.2843	0.3596	0.2813	0.3626	0.2855	0.3606
	bias	−0.0157	0.0072	−0.0187	0.0102	−0.0145	0.0082
	SD	0.0549	0.0289	0.0584	0.0365	0.0559	0.0359
	RMSE	0.0571	0.0297	0.0613	0.0378	0.0577	0.0368
(0.3, 0.0159)	mean	0.2900	0.0245	0.2892	0.0225	0.2876	0.0220
	bias	−0.0100	0.0085	−0.0108	0.0066	−0.0124	0.0061
	SD	0.0442	0.0250	0.0422	0.0243	0.0443	0.0317
	RMSE	0.0453	0.0264	0.0435	0.0252	0.046	0.0322
				$N = 512$			
(0., 0.0833)	mean						
	bias						
	SD						
	RMSE						
(0.2, 0.0833)	mean						
	bias						
	SD						
	RMSE						
(0.3, 0.3524)	mean						
	bias						
	SD						
	RMSE						
(0.3, 0.0159)	mean						
	bias						
	SD						
	RMSE						

Table 1: Simulation results for DWPT-based approximate MLEs $\hat{\delta}$ and $\hat{f}_G$ using the MB(8), LA(16) and MB(16) wavelet filters. An initial parameter estimate of δ was obtained by least-squares estimation where f_G was chosen to be the Fourier frequency with the largest contribution to the periodogram. The portmanteau test ($\alpha = 0.05$) was applied to the squared wavelet coefficients in order to select the orthonormal basis $\mathcal{B} \subset \mathcal{T}$.

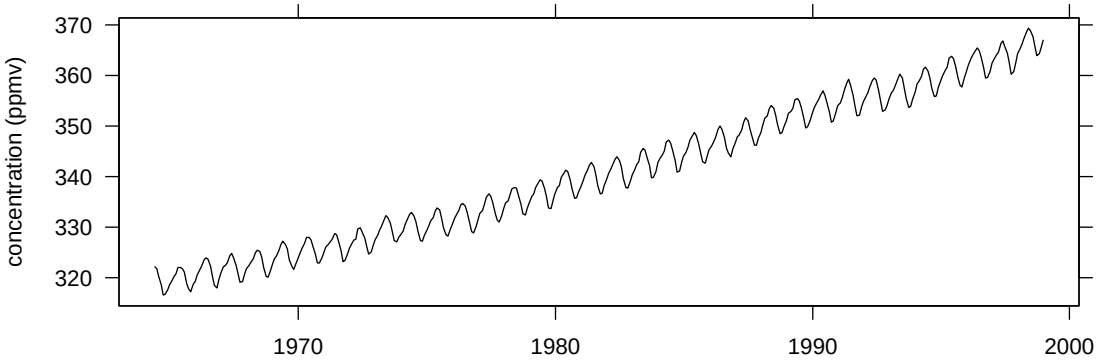


Figure 5: Time series plot of the monthly Mauna Loa CO_2 measurements.

data with an LA(8) wavelet filter. Although this particular orthonormal basis is not adapted to the SDF of this time series, the strongest periodicity appears in $\widetilde{\mathcal{D}}_3 \equiv \widetilde{\mathcal{D}}_{3,1}$ which is associated with the frequency interval $\lambda_{3,1} = (1/16, 1/8]$ or 8–16 month oscillations. The next lowest wavelet detail $\widetilde{\mathcal{D}}_2$, capturing 16–32 month oscillations, also contains a modest periodicity but with an amplitude which may be time dependent.

Before fitting a model to the process, the time-dependent mean should be removed. This was accomplished in Woodward *et al.* (1998) by taking the second difference of the raw series and modelling the residuals. Let $S_{\mathrm{CO}_2}(f)$ denote the true SDF for the CO_2 measurements and let $\mathcal{D}(f) \equiv 4\sin^2(\pi f)$ be the squared gain function for the first order backward difference filter. Hence, the true SDF for the series given is $S_{\mathrm{CO}_2}(f)/\mathcal{D}^2(f)$ and differs from $S_{\mathrm{CO}_2}(f)$ at every frequency!

As an alternative to traditional differencing, consider the multiresolution analysis in Figure 6. By definition of the wavelet transform, all wavelet details are guaranteed to have mean zero as long as the dth order backward difference of the original process is stationary. For Daubechies families of wavelet filters with even length L, they correspond to differences of order $L/2$. Therefore, the LA(8) wavelet filter produces mean zero wavelet details for processes whose 4th order difference is stationary. This appears to be a reasonable assumption for the Mauna Loa CO_2 measurements. The wavelet smooth $\widetilde{\mathcal{S}}_6$, which is associated with the frequency interval $\lambda_{6,0} = [0, 1/128]$, appears to be capturing the time-dependent mean of the time series quite well. Hence, we may produce a low-pass filtered version of the series by summing over the first six wavelet details and ignoring the wavelet smooth – corresponding to the wavelet rough $\widetilde{\mathcal{R}}_6 = \sum_{j=1}^{6} \widetilde{\mathcal{D}}_j$. Filtering in this way only

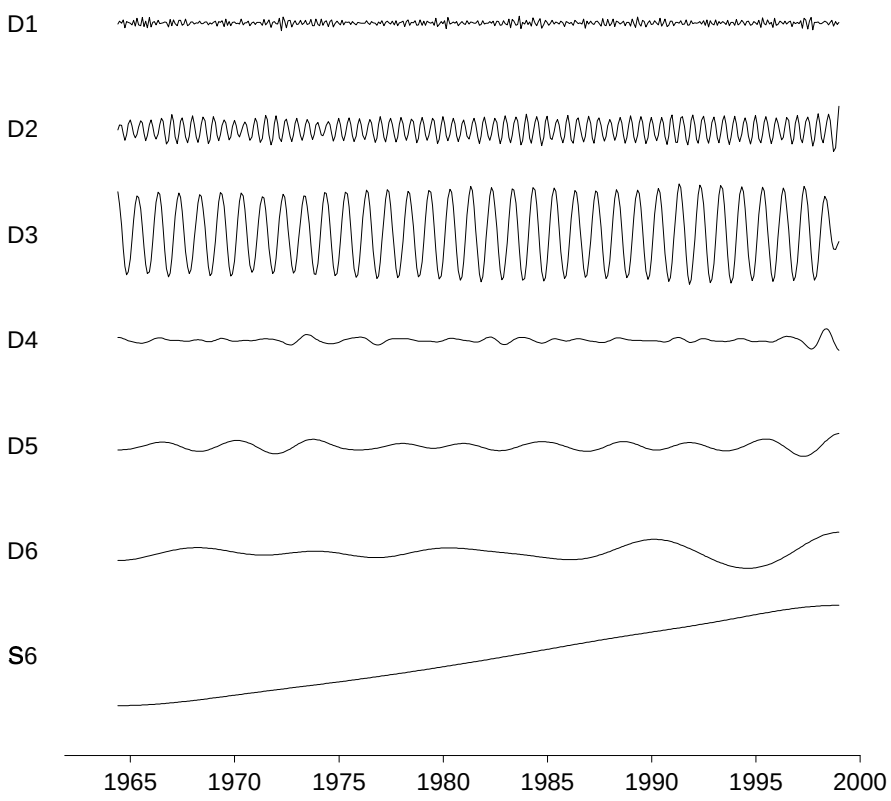


Figure 6: MODWT-based multiresolution analysis ($J = 6$) of the Mauna Loa CO_2 series using the MB(8) wavelet filter and standard orthonormal basis. The same vertical scale is used for the six wavelet details $\widetilde{\mathcal{D}}_1$–$\widetilde{\mathcal{D}}_6$ (which each have mean zero), but not for the wavelet smooth $\widetilde{\mathcal{S}}_6$.

affected frequencies in the range $[0, 1/128]$ (note, this is approximate since all compactly supported wavelet filters are approximations to ideal band-pass filters).

Once an appropriate orthonormal basis $\mathcal{B}$ has been selected, the wavelet variances $\bar{\omega}_{j,n}$ may be computed via numeric integration and optimization of the concentrated likelihood proceeds. Allowing for two asymptotes in the SDF of our process, the wavelet-based seasonal long memory model provides the following MLEs: $\hat{\delta}_1 = 0.47$, $\hat{f}_1 \approx 0.079$, $\hat{\delta}_2 = 0.22$, $\hat{f}_2 \approx 0.165$ and suggests the model

$$(1 - 1.756B + B^2)^{0.47}(1 - 1.013B + B^2)^{0.21}(Y_t - \widetilde{\mathcal{S}}_{6,t}) = \epsilon_t.$$

The first (δ, f_G)-pair corresponds to the strong annual component in the data. This is apparent in

Figure 6 where the majority of energy is contained in the third wavelet detail, corresponding to the frequency interval $\lambda_{3,1} = (1/16, 1/8]$. The second (δ, f_G)-pair is associated with the first harmonic of the annual frequency and contributes less, as indicated by its smaller fractional difference parameter. These estimates tend to agree with the ones obtained in Woodward *et al.* (1998), although they obtained initial parameter estimates and then 'increased' them until the residuals were free from a 12-month cycle.

7 Discussion

Achieving approximate decorrelation in univariate stochastic processes is an appealing feature since it reduces implementing time-consuming methods such as maximum likelihood. Using straightforward tests for white noise, an orthonormal basis may be selected from the wavelet packet table that produces an approximately uncorrelated set of wavelet coefficient vectors. Approximate ML estimation in the wavelet domain produces accurate results for a variety of seasonal long-memory time series models. Although least squares regression of the wavelet packet variance was only used to obtain initial estimates of the fractional difference parameter, this simpler technique deserves further investigation.

References

Anděl, J. (1986). Long memory time series models. *Kybernetika 22*(2), 105–123.

Arteche, J. and P. M. Robinson (1999). Seasonal and cyclical long memory. In S. Ghosh (Ed.), *Asymptotics, Nonparametrics, and Time Series*, Volume 128 of *STATISTICS: Textbooks and Monographs*, pp. 115–148. New York: Marcel Dekker.

Arteche, J. and P. M. Robinson (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis 21*(1), 1–25.

Brockwell, P. J. and R. A. Davis (1991). *Time Series: Theory and Methods* (2 ed.). New York: Springer-Verlag.

Brown, R. L., J. Durbin, and J. M. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B 37*, 149–163.

Daubechies, I. (1992). *Ten Lectures on Wavelets*, Volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics.

Gray, H. L., N.-F. Zhang, and W. A. Woodward (1989). On generalized fractional processes. *Journal of Time Series Analysis 10*(3), 233–257.

Gray, H. L., N.-F. Zhang, and W. A. Woodward (1994). On generalized fractional processes – a correction. *Journal of Time Series Analysis 15*(5), 561–562.

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

Hosking, J. R. M. (1984). Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research 20*(12), 1898–1908.

Jensen, M. J. (1999a). An approximate wavelet MLE of short and long memory parameters. *Studies in Nonlinear Dynamics and Economics 3*(4), 239–253.

Jensen, M. J. (1999b). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting 18*(1), 17–32.

Lobato, I. N. (1997). Semiparametric estimation of seasonal long-memory models: Theory and application to modeling of exchange rates. *Investigaciones Económicas 21*(2), 273–295.

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

Morris, J. M. and R. Peravali (1999). Minimum-bandwidth discrete-time wavelets. *Signal Processing 76*(2), 181–193.

Ooms, M. (1995). Flexible seasonal long memory and economic time series. Technical Report 9515/A, Econometric Institute, Erasmus University.

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

Rainville, E. D. (1960). *Special Functions*. New York: The Macmillan Company.

Whitcher, B. (1998). *Assessing Nonstationary Time Series Using Wavelets*. Ph. D. thesis, University of Washington.

Whitcher, B. (1999). Simulating Gaussian stationary processes with unbounded spectra. Tentatively accepted for publication in the *Journal of Computational and Graphical Statistics*.

Whitcher, B., S. D. Byers, P. Guttorp, and D. B. Percival (1998). Testing for homogeneity of variance in time series: Long memory, wavelets and the Nile River. Submitted for publication.

Wickerhauser, M. V. (1994). *Adapted Wavelet Analysis from Theory to Software*. Wellesley, Massachusetts: A K Peters.

Woodward, W. A., Q. C. Cheng, and H. L. Gray (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis 19*(4), 485–504.