Wavelet Analysis of Discrete Time Series

Andrew T. Walden

Abstract. We give a brief review of some of the wavelet-based techniques currently available for the analysis of arbitrary-length discrete time series. We discuss the maximal overlap discrete wavelet packet transform (MODWPT), a non-decimated version of the usual discrete wavelet packet transform, and a special case, the maximal overlap discrete wavelet transform (MODWT). Using least-asymmetric or coiflet filters of the Daubechies class, the coefficients resulting from the MODWPT can be readily shifted to be aligned with events in the time series. We look at several aspects of denoising, and compare MODWT and cycle-spinning denoising. While the ordinary DWT basis provides a perfect decomposition of the autocovariance of a time series on a scale-by-scale basis, and is well-suited to decorrelating a stationary time series with 'long-memory' covariance structure, a time series with very different covariance structure can be decorrelated using a wavelet packet 'best-basis' determined by a series of white noise tests.

1. Introduction

One scientific area where wavelet methods have been finding many applications is that of the analysis of discrete time series. A time series is defined to be a sequence of observations associated with an ordered independent variable t. Here we only consider the case of discrete values of t, but note that t could represent time, depth or distance along a line.

In this paper we seek to give a brief review of a few discrete wavelet techniques which have proven useful for analysing time series with a view to answering scientific questions. The fundamental ideas are introduced, but many more details and examples can be found in the references and in the comprehensive book [15] which includes many other useful approaches and aspects not touched on here.

In §2 we particularly concentrate attention on the 'maximal overlap' versions of the discrete wavelet transform and discrete wavelet packet transform which are applicable for arbitrary sample sizes. The ability to 'time align' the transform coefficients for certain choices of filters is also stressed. In §3 we look at the progress of wavelet denoising from its 'universal threshold' roots, while in §4 we discuss the scale-by-scale decomposition of the autocovariance sequence of a stationary process via the discrete wavelet transform. As discussed in §5 the correlations of the discrete wavelet transform coefficients of a time series from a 'long-memory' process are mostly negligible, because the transform is well-matched to the covariance structure of the process; recent research suggests that a time series with very different covariance structure can be successfully decorrelated using a wavelet packet 'best-basis' determined by a series of white noise tests.

2. The Pyramid Algorithms

2.1. The discrete wavelet transform

For discrete compactly supported filters of the Daubechies class, ([4, Chapter 6]), we denote the even-length scaling (low-pass) filter by $\{g_l : l = 0, \ldots, L-1\}$ and the wavelet (high-pass) filter $\{h_l : l = 0, \ldots, L-1\}$. The low-pass filter satisfies

$$\sum_{l=0}^{L-1} g_l^2 = 1, \quad \sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0\,, \tag{1}$$

for all nonzero integers n, so that the filter has unit energy and is orthogonal to its even shifts. The high-pass filter is also required to satisfy equation (1) but additionally the high and low-pass filters are chosen to be quadrature mirror filters (QMFs) satisfying:

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

Denote the series to be transformed by $\{X_t : t = 0, \ldots, N-1\}$. With $V_{0,t}^{(D)} \equiv X_t$, the jth stage input to the pyramid algorithm is $\{V_{j-1,t}^{(D)} : t = 0, \ldots, N_{j-1} - 1\}$, where $N_j \equiv N/2^j$. For the discrete wavelet transform (DWT) pyramid algorithm the jth stage outputs are the jth level wavelet and scaling coefficients given by, respectively,

$$W_{j,t}^{(D)} = \sum_{t=0}^{L-1} h_l V_{j-1,(2t+1-l) \bmod N_{j-1}}^{(D)}, \quad V_{j,t}^{(D)} = \sum_{t=0}^{L-1} g_l V_{j-1,(2t+1-l) \bmod N_{j-1}}^{(D)}\,,$$

$t = 0, \ldots, N_j - 1$. If we write $\{W_{j,t}^{(D)} : t = 0, \ldots, N_j - 1\}$ as $\mathbf{W}_j^{(D)}$ and $\{V_{j,t}^{(D)} : t = 0, \ldots, N_j - 1\}$ as $\mathbf{V}_j^{(D)}$, then if $N = 2^J$ the pyramid algorithm is complete after J repetitions yielding $\mathbf{W}_1^{(D)}, \ldots, \mathbf{W}_J^{(D)}, \mathbf{V}_J^{(D)}$, where the latter two vectors contain only one coefficient each. This defines the *full DWT*. If however N is an integer multiple of 2^{J_0} say, then we can carry out a *partial DWT* to level J_0. The DWT is an orthonormal transform of $\{X_t : t = 0, \ldots, N-1\}$.

The jth level wavelet and scaling coefficients may be linked directly to the series $\{X_t\}$. We define the jth level wavelet filter $\{h_{j,l}\}$ formed by convolving together

$$\begin{array}{ll} \text{filter } 1 & g_0, g_1, \cdots, g_{L-2}, g_{L-1}\,; \\ \quad\vdots & \qquad\vdots \\ \text{filter } j-1 & g_0,\ \underbrace{0,\dots,0,}_{2^{j-2}-1 \text{ zeros}}\ g_1,\ \underbrace{0,\dots,0,}_{2^{j-2}-1 \text{ zeros}}\ \dots, g_{L-2},\ \underbrace{0,\dots,0,}_{2^{j-2}-1 \text{ zeros}}\ g_{L-1}; \text{ and} \\ \text{filter } j & h_0,\ \underbrace{0,\dots,0,}_{2^{j-1}-1 \text{ zeros}}\ h_1,\ \underbrace{0,\dots,0,}_{2^{j-1}-1 \text{ zeros}}\ \dots, h_{L-2},\ \underbrace{0,\dots,0,}_{2^{j-1}-1 \text{ zeros}}\ h_{L-1}\,. \end{array} \tag{2}$$

Filter $\{h_{j,l}\}$ has $L_j = (2^j - 1)(L-1) + 1$ terms. To obtain the jth level scaling filter $\{g_{j,l}\}$ the only difference is that filter j in the list above is replaced by

$$\text{filter } j \quad g_0,\ \underbrace{0,\dots,0,}_{2^{j-1}-1 \text{ zeros}}\ g_1,\ \underbrace{0,\dots,0,}_{2^{j-1}-1 \text{ zeros}}\ \dots, g_{L-2},\ \underbrace{0,\dots,0,}_{2^{j-1}-1 \text{ zeros}}\ g_{L-1}\,. \tag{3}$$

Then

$$W_{j,t}^{(D)} = \sum_{l=0}^{L_j-1} h_{j,l} X_{(2^j(t+1)-1-l) \bmod N} \quad \text{and} \quad V_{j,t}^{(D)} = \sum_{l=0}^{L_j-1} g_{j,l} X_{(2^j(t+1)-1-l) \bmod N}\,.$$

The jth level filters have the following properties:

$$\sum_{l=0}^{L_j-1} g_{j,l} = 2^{j/2}; \ \sum_{l=0}^{L_j-1} g_{j,l}^2 = 1; \ \sum_{l=0}^{L_j-1} g_{j,l} h_{j,l} = 0; \ \sum_{l=0}^{L_j-1} h_{j,l} = 0; \ \sum_{l=0}^{L_j-1} h_{j,l}^2 = 1\,. \tag{4}$$

At level j the nominal frequency band to which the corresponding wavelet coefficients $\{W_{j,t}^{(D)}\}$ are associated is given by $|f| \in (1/2^{j+1}, 1/2^j]$; an 'octave' band. For example $\{W_{1,t}^{(D)}\}$, $\{W_{2,t}^{(D)}\}$ and $\{W_{3,t}^{(D)}\}$ have nominal pass-bands of $(1/4, 1/2], (1/8, 1/4]$ and $(1/16, 1/8]$ respectively.

2.2. The maximal overlap discrete wavelet transform

The DWT has several limitations:

- It requires the sample size to be an integer multiple of 2^{J_0} for a partial DWT, or to be exactly a power of 2 for the full transform.
- It is sensitive to where we 'break into' a time series. The wavelet and scaling coefficients are not circularly shift equivariant, i.e., circularly shifting the time series by some amount will not circularly shift the DWT wavelet and scaling coefficients by the same amount.
- The number of wavelet and scaling coefficients, N_j, decreases by a factor of 2 for each increasing level of the transform, limiting the ability to carry out statistical analyses on the coefficients.

These deficiencies can be overcome if the downsampling in the DWT can be avoided. This can be achieved by using the maximal overlap discrete wavelet transform (MODWT) ([12, 13]); see also the undecimated discrete wavelet transform ([19]), and stationary discrete wavelet transform ([10]). A rescaling of the defining filters is required to conserve energy: $\tilde{g}_l = g_l/\sqrt{2}$ and $\tilde{h}_l = h_l/\sqrt{2}$ so that now for example, $\sum_{l=0}^{L-1} \tilde{g}_l^2 = 1/2$, while the filters are still QMFs. Now the DWT pyramid algorithm consists of filter-subsample steps repeated for each level. If we define $V_{0,t}^{(M)} = X_t$, then the MODWT pyramid algorithm generates the MODWT wavelet coefficients $\{W_{j,t}^{(M)}\}$ and the MODWT scaling coefficients $\{V_{j,t}^{(M)}\}$ from $\{V_{j-1,t}^{(M)}\}$ using the 'new filters' (2) and (3) with non-zero coefficients divided by $\sqrt{2}$; the circular filterings (convolutions) can be written as

$$W_{j,t}^{(M)} = \sum_{l=0}^{L-1} \tilde{h}_l V_{j-1,(t-2^{j-1}l) \bmod N}^{(M)}, \quad V_{j,t}^{(M)} = \sum_{l=0}^{L-1} \tilde{g}_l V_{j-1,(t-2^{j-1}l) \bmod N}^{(M)},$$

$t = 0, \dots, N-1$.

These coefficients can also be formulated in terms of a filtering of $\{X_t\}$, using the filters $\{\tilde{h}_{j,l} = h_{j,l}/2^{j/2}\}$ and $\{\tilde{g}_{j,l} = g_{j,l}/2^{j/2}\}$:

$$W_{j,t}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{(t-l) \bmod N} \quad \text{and} \quad V_{j,t}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{(t-l) \bmod N}.$$

Notice that the DWT of $\{X_t\}$ can be extracted from the MODWT via a rescaling and subsampling:

$$W_{j,t}^{(D)} = 2^{j/2} W_{j,2^j(t+1)-1}^{(M)} \quad \text{and} \quad V_{j,t}^{(D)} = 2^{j/2} V_{j,2^j(t+1)-1}^{(M)}. \tag{5}$$

The MODWT coefficients at level j are associated to the same nominal frequency band $|f| \in (1/2^{j+1}, 1/2^j]$ as for the DWT, but there are always N of them at each level. The overdetermined MODWT is not an orthonormal transform of $\{X_t : t = 0, \dots, N-1\}$.

2.3. The discrete wavelet packet transform

The discrete wavelet packet transform (DWPT) is a generalization of the DWT which at level j of the transform partitions the frequency axis into 2^j equal width frequency bands, often labelled $n = 0, \dots, 2^j - 1$. Increasing the transform level increases frequency resolution, but, starting with a series of length N, at level j there are only $N/2^j$ DWPT coefficients for each frequency band n.

We denote the time series to be transformed, $\{X_t : t = 0, \dots, N-1\}$ by $\mathbf{X}$, thought of as a column vector. Initially we set $\mathbf{W}_{0,0}^{(D)} \equiv \mathbf{X}$. At the first step of the algorithm $\mathbf{X}$ is circularly filtered by the low-pass filter $\{g_l\}$, with corresponding transfer function $G(f) = \sum_{l=0}^{L-1} g_l e^{-i2\pi fl}$, and then downsampled by 2, to give first level coefficients $\mathbf{W}_{1,0}^{(D)} = \{W_{1,0,t}^{(D)}, t = 0, \dots, (N/2)-1\}$ and $\mathbf{X}$ is also circularly filtered by the high-pass filter $\{h_l\}$, with corresponding transfer function $H(f)$, and then downsampled by 2, to give $\mathbf{W}_{1,1}^{(D)} = \{W_{1,1,t}^{(D)}, t = 0, \dots, (N/2)-1\}$. Both $\mathbf{W}_{1,0}^{(D)}$ and $\mathbf{W}_{1,1}^{(D)}$ are of length $N_1 = N/2$ due to the downsampling.

Subsequent levels j of the transform repeat these steps: entries for the DWPT table up to level $j = 3$ are processed as shown in figure 1; the figure emphasizes that circular filtering is used by showing that the jth level coefficients are obtained by filtering the $(j-1)$th level coefficients with the circular filter having discrete Fourier transform $\{H(\frac{k}{N_{j-1}})\}$ or $\{G(\frac{k}{N_{j-1}})\}$, followed by downsampling by 2.

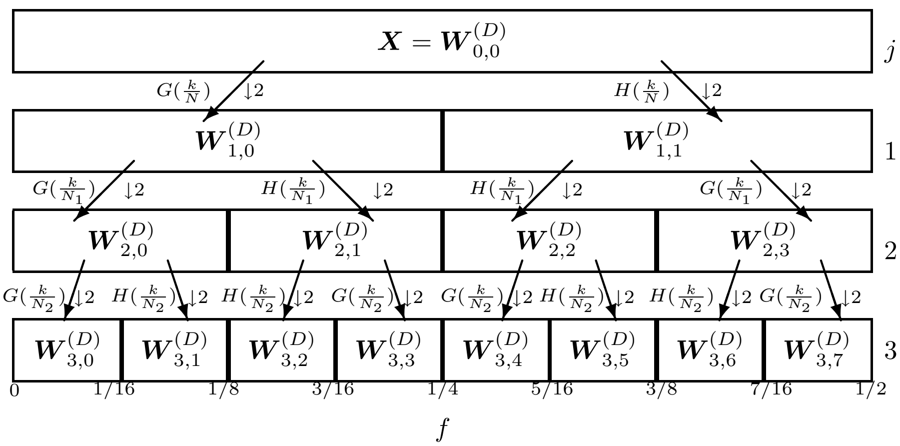

FIGURE 1. Wavelet packet table showing the DWPT of $\mathbf{X}$.

Figure 1 can be summarised as follows. To compute the DWPT coefficients for levels $j = 1, \ldots, J_0$, we circularly filter the wavelet packet coefficients at the previous stage and downsample by 2. Given the series $\{W^{(D)}_{j-1,\lfloor n/2\rfloor,t}\}$ of length N_{j-1} we calculate $\{W^{(D)}_{j,n,t}\}$ using

$$W^{(D)}_{j,n,t} \equiv \sum_{l=0}^{L-1} r_{n,l} W^{(D)}_{j-1,\lfloor n/2\rfloor,(2t+1-l) \bmod N_{j-1}}, \quad t = 0, \ldots, N_j - 1\,,$$

where

$$r_{n,l} = \begin{cases} g_l, & \text{if } n \bmod 4 = 0 \text{ or } 3\,; \\ h_l, & \text{if } n \bmod 4 = 1 \text{ or } 2\,. \end{cases} \tag{6}$$

Here $\lfloor\cdot\rfloor$ denotes 'the integer part' operator. The ordering of the filters means that at level j band n is nominally associated with frequencies in the interval $(\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$. Such a DWPT would be said to be *sequency ordered* [22].

The transform that takes $\mathbf{X}$ to $\mathbf{W}^{(D)}_{j,n}$, $n = 0, \ldots, 2^j - 1$, for any j between 0 and J_0 is called a DWPT, and is orthonormal.

We can also write $\{W^{(D)}_{j,n,t}\}$ in terms of filtering $\{X_t\colon t = 0, \ldots, N-1\}$. Suppose we let

$$\{u_{1,0,l}\} = \{g_l\colon l = 0, \ldots, L-1\} \quad \text{and} \quad \{u_{1,1,l}\} = \{h_l\colon l = 0, \ldots, L-1\}\,,$$

and for general (j,n) define

$$u_{j,n,l} = \sum_{k=0}^{L-1} r_{n,k}\, u_{j-1,\lfloor n/2\rfloor,l-2^{j-1}k}, \quad l = 0,\dots,L_j - 1\,, \tag{7}$$

where $L_j = (2^j - 1)(L-1) + 1$ is the length of $\{u_{j,n,l}\}$. For example, for $j = 2$,

$$u_{2,1,l} = \sum_{k=0}^{L-1} r_{1,k}\, u_{1,0,l-2k} = \sum_{k=0}^{L-1} h_k g_{l-2k}$$

which is the convolution of the upsampled version of $\{h_l\}$ with $\{g_l\}$; if $\{X_t : t = 0,\dots,N-1\}$ is circularly convolved with this filter and the result downsampled by 4, then $\{W^{(D)}_{2,1,t}\}$ results. In general, for $j = 1,\dots,J_0$, we can write $\{W^{(D)}_{j,n,t}\}$ in terms of a filtering of $\{X_t : t = 0,\dots,N-1\}$, via

$$W^{(D)}_{j,n,t} = \sum_{l=0}^{L_j-1} u_{j,n,l} X_{(2^j[t+1]-1-l) \bmod N}, \quad t = 0,1,\dots,N_j - 1\,.$$

2.4. The maximal overlap discrete wavelet packet transform

The downsampling step in the DWPT can also be removed. The maximal overlap (undecimated, stationary) discrete wavelet packet transform (MODWPT) as developed in [20] can be briefly summarized as follows.

Let $\widetilde{G}(f) = \sum_{l=0}^{L-1} \tilde{g}_l e^{-i2\pi f l}$, be the transfer function of $\{\tilde{g}_l\}$, and let the transfer function corresponding to $\{\tilde{h}_l\}$ be defined similarly. Initially we set $\mathbf{W}^{(M)}_{0,0} \equiv \mathbf{X}$. At the first step of the algorithm $\mathbf{X}$ is circularly filtered by the low-pass filter $\{\tilde{g}_l\}$, with corresponding transfer function $\widetilde{G}(f)$, to give first level coefficients $\mathbf{W}^{(M)}_{1,0} = \{W^{(M)}_{1,0,t},\, t = 0,\dots,N-1\}$ and $\mathbf{X}$ is also circularly filtered by the high-pass filter $\{\tilde{h}_l\}$, with corresponding transfer function $\widetilde{H}(f)$, to give $\mathbf{W}^{(M)}_{1,1} = \{W^{(M)}_{1,1,t},\, t = 0,\dots,N-1\}$. Both $\mathbf{W}^{(M)}_{1,0}$ and $\mathbf{W}^{(M)}_{1,1}$ are of length N since there is no downsampling.

For subsequent levels j of the transform we again insert $2^{j-1}-1$ zeros, $j \geq 1$, between the elements of $\{\tilde{g}_l\}$ as in equation (3); the resulting filter has a transfer function given by $\widetilde{G}(2^{j-1}f)$. We can do likewise for $\{\tilde{h}_l\}$ to obtain $\widetilde{H}(2^{j-1}f)$. The set of entries for the MODWPT table up to level $j = 3$ are processed as shown in figure 2. The jth level coefficients are obtained by filtering the level $j-1$ coefficients with the circular filter having discrete Fourier transform $\{\widetilde{H}(2^{j-1}\frac{k}{N})\}$ or $\{\widetilde{G}(2^{j-1}\frac{k}{N})\}$, as appropriate. Figure 2 can be summarised as follows. To compute the MODWPT coefficients for levels $j = 1,\dots,J_0$, we circularly filter the wavelet packet coefficients at the previous stage. Given the series $\{W^{(M)}_{j-1,\lfloor n/2\rfloor,t}\}$ of length N we calculate $\{W^{(M)}_{j,n,t}\}$ using

$$W^{(M)}_{j,n,t} \equiv \sum_{l=0}^{L-1} r_{n,l} W^{(M)}_{j-1,\lfloor n/2\rfloor,(t-2^{j-1}l) \bmod N}, \quad t = 0,\dots,N-1\,,$$

where now

$$r_{n,l} = \begin{cases} \tilde{g}_l, & \text{if } n \bmod 4 = 0 \text{ or } 3\,; \\ \tilde{h}_l, & \text{if } n \bmod 4 = 1 \text{ or } 2\,. \end{cases} \tag{8}$$

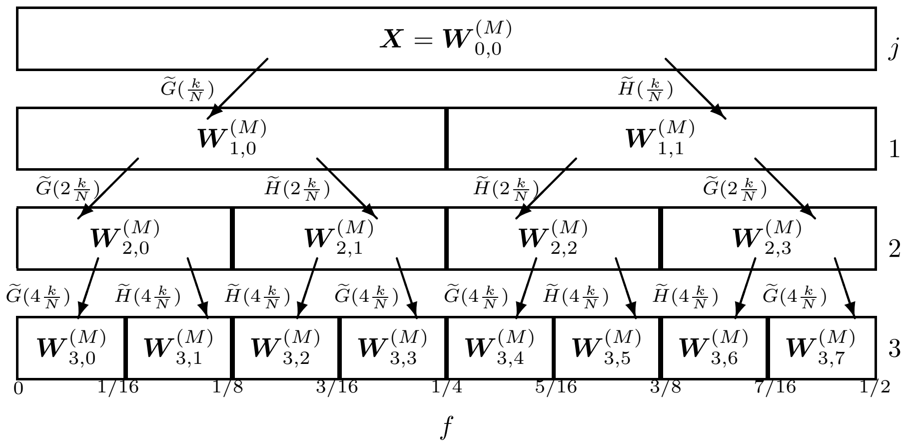

FIGURE 2. Undecimated wavelet packet table showing the MODWPT of **X**.

We can also write $\{W^{(M)}_{j,n,t}\}$ in terms of filtering $\{X_t\colon t = 0, \ldots, N-1\}$. Suppose we now let

$$\{u_{1,0,l}\} = \{\tilde{g}_l\colon l = 0, \ldots, L-1\} \quad \text{and} \quad \{u_{1,1,l}\} = \{\tilde{h}_l\colon l = 0, \ldots, L-1\}\,,$$

and for general (j, n) define $\{u_{j,n,l}\}$ as in (7) so for example, for $j = 2$,

$$u_{2,1,l} = \sum_{k=0}^{L-1} r_{1,k}\, u_{1,0,l-2k} = \sum_{k=0}^{L-1} \tilde{h}_k \tilde{g}_{l-2k}$$

which is the convolution of the upsampled version of $\{\tilde{h}_l\}$ with $\{\tilde{g}_l\}$; if $\{X_t\colon t = 0, \ldots, N-1\}$ is circularly convolved with this filter then $\{W^{(M)}_{2,1,t}\}$ results. In general, for $j = 1, \ldots, J_0$, we can write $\{W^{(M)}_{j,n,t}\}$ in terms of a filtering of $\{X_t\colon t = 0, \ldots, N-1\}$, via

$$W^{(M)}_{j,n,t} = \sum_{l=0}^{L_j-1} u_{j,n,l} X_{(t-l) \bmod N}, \quad t = 0, \ldots, N-1\,.$$

The transfer function $U_{j,n}(f) = \sum_{l=0}^{L_j-1} u_{j,n,l} e^{-i2\pi f l}$, of $\{u_{j,n,l}\colon l = 0, \ldots, L_j - 1\}$ is defined entirely by the two filters $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$. Suppose we let $\widetilde{M}_0 = \widetilde{G}(f)$, say, and $\widetilde{M}_1 = \widetilde{H}(f)$, say, as in [3]. Further, suppose we attach to each pair (j, n) in the wavelet packet tree a sequence of ones and zeros according to the following rule. Let $\{\mathbf{c}_{1,0}\} = \{0\}$ and $\{\mathbf{c}_{1,1}\} = \{1\}$ and let $\{\mathbf{c}_{j,n}\} =$

$\{c_{j,n,0},\ldots,c_{j,n,j-1}\}$ where $c_{j,n,k} \in \{0,1\}$ for $k = 0,\ldots,j-1$. Define, for $n = 0,\ldots,2^j-1$,

$$\mathbf{c}_{j,n} = \begin{cases} \{\mathbf{c}_{j-1,\lfloor n/2 \rfloor}, 0\}, & \text{if } n \bmod 4 = 0 \text{ or } 3\,; \\ \{\mathbf{c}_{j-1,\lfloor n/2 \rfloor}, 1\}, & \text{if } n \bmod 4 = 1 \text{ or } 2\,. \end{cases}$$

Hence to obtain $\{\mathbf{c}_{j,n}\}$ we append a zero to the sequence $\{\mathbf{c}_{j-1,\lfloor n/2 \rfloor}\}$ if $n \bmod 4 = 0$ or 3 and append a one if $n \bmod 4 = 1$ or 2. Then,

$$U_{j,n}(f) = \prod_{m=0}^{j-1} \widetilde{M}_{c_{j,n,m}}(2^m f)\,. \tag{9}$$

As an example consider $U_{3,3}(f)$, to which is associated the binary sequence $\{\mathbf{c}_{3,3}\} = \{c_{3,3,0}, c_{3,3,1}, c_{3,3,2}\} = \{0,1,0\}$. We then have $U_{3,3}(f) = \widetilde{M}_0(f)\widetilde{M}_1(2f)\widetilde{M}_0(4f) = \widetilde{G}(f)\widetilde{H}(2f)\widetilde{G}(4f)$.

Each factor in the product in equation (9) can be written

$$\widetilde{M}_{c_{j,n,m}}(2^m f) = |\widetilde{M}_{c_{j,n,m}}(2^m f)| \exp\{i\theta_{c_{j,n,m}}(2^m f)\}\,,$$

where

$$\theta_{c_{j,n,m}}(2^m f) = \begin{cases} \theta^{(\tilde{G})}(2^m f), & \text{if } c_{j,n,m} = 0\,; \\ \theta^{(\tilde{H})}(2^m f), & \text{if } c_{j,n,m} = 1\,, \end{cases}$$

where $\theta^{(\tilde{G})}(f)$ and $\theta^{(\tilde{H})}(f)$ are the phase functions of the filters $\{\tilde{g}_l\}$ and $\{\tilde{h}_l\}$, respectively.

For the Daubechies least asymmetric filters of lengths $L = 8(2)20$, denoted here LA(L), and the Daubechies coiflet filters of lengths $L = 6(6)30$, denoted here C(L), it may be shown that $\theta^{(\tilde{G})}(f) \approx 2\pi f \nu$ and $\theta^{(\tilde{H})}(f) \approx -2\pi f(L-1+\nu)$, where, for the LA($L$) filters

$$\nu = \begin{cases} -(L/2)+1, & \text{if } L = 8, 12, 16 \text{ or } 20\,; \\ -(L/2), & \text{if } L = 10 \text{ or } 18\,; \\ -(L/2)+2, & \text{if } L = 14\,; \end{cases}$$

and for the C(L) filters, $\nu = -(2L/3)+1$, $L = 6, 12, 18, 24$ and 30. The overall phase function of $U_{j,n}(f)$ is thus

$$\begin{aligned} \sum_{m=0}^{j-1} \theta_{c_{j,n,m}}(2^m f) &\approx 2\pi f \left(\nu \sum_{\substack{m=0 \\ \{m: c_{j,n,m}=0\}}}^{j-1} 2^m - (L-1+\nu) \sum_{\substack{m=0 \\ \{m: c_{j,n,m}=1\}}}^{j-1} 2^m \right) \\ &= 2\pi f \nu_{j,n}, \quad \text{say}\,. \end{aligned}$$

Since $\nu_{j,n} < 0$ for the filters discussed above, we obtain the closest approximation to zero phase filtering if we associate the coefficient $W^{(M)}_{j,n,(t+|\nu_{j,n}|) \bmod N}$ with the input X_t. (For further details, including the necessity of reversing some of the filters listed in [4] for phase correction purposes, see [20] and [15].)

3. Denoising

3.1. The DWT and universal thresholding

Suppose that our observed time series consists of a signal plus noise, i.e.,

$$\mathbf{X} = \mathbf{D} + \boldsymbol{\epsilon}\,, \tag{10}$$

where $\mathbf{D}$ is a deterministic signal and $\boldsymbol{\epsilon}$ represents an N dimensional vector of independent and identically distributed (IID) Gaussian noise, each random variable having variance σ_ϵ^2.

For the purpose of denoising via thresholding Donoho and Johnstone ([5]) recommend firstly computing a level J_0 partial DWT giving coefficient vectors $\mathbf{W}_1^{(D)}, \ldots, \mathbf{W}_{J_0}^{(D)}$ and $\mathbf{V}_{J_0}^{(D)}$. Component-wise, we have, say,

$$W_{j,t}^{(D)} = d_{j,t} + e_{j,t} \quad j = 1, \ldots, J_0;\;\; t = 0, \ldots, N_j - 1\,.$$

(J_0 must be specified by the user.) Then, only the coefficients in the $\mathbf{W}_k^{(D)}$ vectors are subjected to thresholding; i.e., the elements of $\mathbf{V}_{J_0}^{(D)}$ are untouched, so that portion of $\mathbf{X}$ is automatically assigned to the signal $\mathbf{D}$.

Next a threshold level must be chosen. A key property about an orthonormal transform, (such as the partial DWT), of IID Gaussian noise is that the transformed noise has the same statistical properties as the untransformed noise so that the $\{e_{j,t}\}$ are also IID Gaussian with mean zero and variance σ_ϵ^2. The *universal threshold* was defined in [5] as

$$\delta_U \equiv \sqrt{[2\sigma_\epsilon^2 \log(N)]}\,.$$

To understand the rationale for this threshold suppose that the signal $\mathbf{D}$ is in fact a vector of zeros so that the transform coefficients $\{W_{j,t}^{(D)}\}$ are a portion of an IID Gaussian sequence $\{e_{j,t}\}$ with zero mean and variance σ_ϵ^2. Then, as $N \to \infty$, we have

$$\mathbf{P}\big[\max\{|W_{j,t}^{(D)}|\} \le \delta_U\big] \equiv \mathbf{P}\big[\max\{|e_{j,t}|\} \le \delta_U\big] \to 1\,,$$

so that asymptotically we will correctly estimate the signal vector. Universal thresholding typically removes all the noise, but, in doing so, it can mistakenly set some small signal transform coefficients to zero. Universal thresholding thus ensures, with high probability, that the reconstruction is at least as smooth as the true deterministic signal. If σ_ϵ^2 is unknown as is frequently the case in applications, a practical procedure is to estimate it based upon the *median absolute deviation* (MAD) standard deviation estimate using just the $N/2$ level $j = 1$ coefficients in $\mathbf{W}_1^{(D)}$. By definition, this standard deviation estimator is

$$\hat{\sigma}_{\mathrm{MAD}} \equiv \frac{\mathrm{median}\{|W_{1,0}^{(D)}|, |W_{1,1}^{(D)}|, \ldots, |W_{1,\frac{N}{2}-1}^{(D)}|\}}{0.6745}\,.$$

The factor 0.6745 rescales so that $\hat{\sigma}_{\mathrm{MAD}}$ is also a suitable estimator for the standard deviation for Gaussian white noise. $\hat{\sigma}_{\mathrm{MAD}}$ is calculated from the elements of $\mathbf{W}_1^{(D)}$ because the smallest scale wavelet coefficients should be noise dominated, with the possible exception of the largest values. The MAD standard deviation estimate is designed to be robust against large deviations and hence should reflect the noise variance rather than the signal variance.

Finally, for $W_{j,t}^{(D)}, j = 1, \ldots, J_0$ and $t = 0, \ldots, N_j - 1$, we apply a chosen thresholding rule, such as hard thresholding defined by

$$\widehat{W}_{j,t}^{(D)} = \begin{cases} 0, & \text{if } |W_{j,t}^{(D)}| \leq \delta_U\,; \\ W_{j,t}^{(D)}, & \text{otherwise}\,, \end{cases}$$

to obtain the thresholded coefficients $\{\widehat{W}_{j,t}^{(D)}\}$, which are then used to form $\widehat{\mathbf{W}}_j^{(D)}$, $j = 1, \ldots, J_0$. $\mathbf{D}$ is estimated as $\widehat{\mathbf{D}}$ obtained by inverse transforming $\widehat{\mathbf{W}}_1^{(D)}, \ldots, \widehat{\mathbf{W}}_{J_0}^{(D)}$ and $\mathbf{V}_{J_0}^{(D)}$.

3.2. Level-dependent thresholding

In the event that the noise wavelet coefficients $\{e_{j,t}\}$ are uncorrelated, but not identically distributed, as will be the case if the additive noise is still IID but is non-Gaussian, a level-dependent (universal thresholding) can be used where now

$$\delta_{U,j} \equiv \sqrt{[2\sigma_j^2 \log(N)]}\,,$$

and σ_j^2 is the variance of the jth level coefficients. As pointed out in [7] in the general case of unknown variances, we could use the MAD estimate at each level j, namely,

$$\hat{\sigma}_{\text{MAD},j} \equiv \frac{\text{median}\{|W_{j,0}^{(D)}|, |W_{j,1}^{(D)}|, \ldots, |W_{j,N_j-1}^{(D)}|\}}{0.6745}\,,$$

and then use the universal threshold

$$\hat{\delta}_{U,j} \equiv \sqrt{[2\hat{\sigma}_{\text{MAD},j}^2 \log(N)]}$$

at each level. The estimate $\hat{\sigma}_{\text{MAD},j}$ is only appropriate for small j, where there is a considerable number of coefficients in a given level, and the signal is sparse.

Scale-dependent thresholding was investigated in the context of spectrum estimation in [21]; in this application the variances σ_j^2 at each level were known, and did not need to be estimated.

3.3. Cycle-spinning denoising and the MODWT

It was noted in section (2.2) that one potential problem with the DWT is its sensitivity to where we 'break into' a time series. As a consequence the result of denoising using the DWT will depend somewhat on the starting point of the series. To try to alleviate this problem the idea of *cycle spinning* was introduced in [2]. Consider a partial DWT of level J_0 calculated from a time series of length N an integer multiple of 2^{J_0}. The idea of denoising via cycle spinning is to apply denoising not only to $\mathbf{X}$, but also to all possible unique circularly shifted versions of $\mathbf{X}$, and to average the results. Consider the model in (10). Let $\mathcal{T}\mathbf{X} = [X_{N-1}, X_0, X_1, \ldots, X_{N-2}]$, i.e., $\mathcal{T}$ (circularly) delays $\mathbf{X}$ by one time unit, and let $\mathcal{T}^2\mathbf{X} = \mathcal{T}\mathcal{T}\mathbf{X}$ etc. Also we take $\mathcal{T}^{-1}$ to be a circular advance operator. Suppose that $\widehat{\mathbf{D}}_n$ is the estimate of $\mathbf{D}$

resulting from applying a denoising procedure to $\mathcal{T}^n\mathbf{X}$. Then the cycle spinning denoising estimate of $\mathbf{D}$ is given by

$$\overline{\mathbf{D}} = \frac{1}{2^{J_0}} \sum_{n=0}^{2^{J_0}-1} \mathcal{T}^{-n}\widehat{\mathbf{D}}_n \,.$$

(Shifts greater than or equal to 2^{J_0} are redundant.)

However, we know that the DWT of each $\mathcal{T}^n\mathbf{X}$ can be extracted readily from the MODWT of $\mathbf{X}$: see equation (5) where we note that the variance of the DWT coefficients is 2^j times that of the MODWT coefficients. Hence cycle-spinning can be implemented efficiently in terms of the MODWT as follows ([15]):

- Compute a level J_0 partial MODWT giving coefficient vectors $\mathbf{W}_1^{(M)}, \ldots, \mathbf{W}_{J_0}^{(M)}$ and $\mathbf{V}_{J_0}^{(M)}$.
- For each $j = 1, \ldots, J_0$ apply a chosen thresholding rule to each element of $\mathbf{W}_j^{(M)}$ using the level-dependent universal threshold $\delta_{U,j} \equiv \sqrt{[2\sigma_j^2 \log(N)]}$, with $\sigma_j^2 = \sigma_\epsilon^2/2^j$.
- The thresholded coefficients $\{\widehat{W}_{j,t}^{(M)}\}$ are then used to form $\widehat{\mathbf{W}}_j^{(M)}$, $j = 1, \ldots, J_0$. $\mathbf{D}$ is estimated as $\widehat{\mathbf{D}}$ obtained by applying the inverse MODWT to $\widehat{\mathbf{W}}_1^{(M)}, \ldots, \widehat{\mathbf{W}}_{J_0}^{(M)}$ and $\mathbf{V}_{J_0}^{(M)}$.

The advantages of this approach to cycle spinning are two-fold. Cycle spinning originated assuming that the sample size N is an integer multiple of 2^{J_0}, whereas the MODWT-based approach above is valid for general N. Secondly, if σ_ϵ^2 is unknown, the MAD scale estimate can be adapted by taking

$$\tilde{\sigma}_{\text{MAD}} \equiv \frac{2^{1/2}\,\text{median}\{|W_{1,0}^{(M)}|, |W_{1,1}^{(M)}|, \ldots, |W_{1,N-1}^{(M)}|\}}{0.6745} \,.$$

The scaling factor of '$2^{1/2}$' in the numerator above, reflects the fact that $W_{1,t}^{(D)} = 2^{1/2} W_{1,2t+1}^{(M)}$.

4. Scale-Based Decomposition

Let $\{X_t, t \in \mathbb{Z}\}$ denote a real-valued second-order stationary process for which $\{X_t \colon t = 0, \ldots, N-1\}$ would represent a realization. The autocovariance sequence $\{s_{X,m}\}$ is defined as

$$s_{X,m} \equiv \text{cov}\{X_t, X_{t+m}\} = E\left\{[X_t - \mu_X][X_{t+m} - \mu_X]\right\} = s_{X,-m}\,,$$

where μ_X is the mean of $\{X_t, t \in \mathbb{Z}\}$.

The stationary stochastic processes resulting from applying the filters $\{\tilde{h}_{j,l}\}$, $j = 1, \ldots, J_0$, and $\{\tilde{g}_{J_0,l}\}$, to $\{X_t\}$ are given by the jth level wavelet coefficients

$\{\mathcal{W}_{j,t}^{(M)}\}$, $j = 1, \ldots, J_0$, and J_0th scaling coefficients $\{\mathcal{V}_{J_0,t}^{(W)}\}$, calculated by non-circular filtering, where

$$\mathcal{W}_{j,t}^{(M)} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l},\ j = 1, \ldots, J_0, \quad \text{and} \quad \mathcal{V}_{J_0,t}^{(M)} = \sum_{l=0}^{L_{J_0}-1} \tilde{g}_{J_0,l} X_{t-l},\ t \in \mathbb{Z}\,.$$

The autocovariance of the wavelet coefficient process at transform level j, $\{\mathcal{W}_{j,t}^{(M)}\}$, is given by

$$s_{\mathcal{W}_j,m} = E\{\mathcal{W}_{j,t}^{(M)} \mathcal{W}_{j,t+m}^{(M)}\}\,.$$

Note that $\{\mathcal{W}_{j,t}^{(M)}\}$ has a mean of zero since by design $\sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0$.

The autocovariance of the scaling coefficient process at transform level J_0, $\{\mathcal{V}_{J_0,t}^{(M)}\}$, is given by

$$s_{\mathcal{V}_{J_0},m} = E\{\mathcal{V}_{J_0,t}^{(M)} \mathcal{V}_{J_0,t+m}^{(M)}\} - E^2\{\mathcal{V}_{J_0,t}^{(M)}\} = E\{\mathcal{V}_{J_0,t}^{(M)} V_{J_0,t+m}^{(M)}\} - \mu_X^2\,,$$

since the mean of $\{\mathcal{V}_{j,t}^{(M)}\}$ is μ_X, because by design $\sum_{l=0}^{L_j-1} \tilde{g}_{j,l} = 1$. Then ([16]),

$$s_{X,m} = \sum_{j=1}^{J_0} s_{\mathcal{W}_j,m} + s_{\mathcal{V}_{J_0},m} \tag{11}$$

i.e., the autocovariance sequence of the process $\{X_t, t \in \mathbb{Z}\}$ can be decomposed in terms of the autocovariance sequences of the wavelet coefficient processes, and the autocovariance sequence of the single scaling coefficient process. (This type of decomposition extends to cross-covariances (see [16]); it was noted for sequence variance in [11].) Most usefully, we can interpret this result as a scale-by-scale decomposition.

A standard measure of effective width of the jth level scaling filter is the 'autocorrelation width,' ([1]), defined as $\text{width}_\text{a}\{\tilde{g}_{j,l}\} = \left(\sum_l \tilde{g}_{j,l}\right)^2 / \sum_l \tilde{g}_{j,l}^2$. But, from (4),

$$\left(\sum_l \tilde{g}_{j,l}\right)^2 = 1 \quad \text{and} \quad \sum_l \tilde{g}_{j,l}^2 = 2^{-j}\,,$$

so that $\text{width}_\text{a}\{\tilde{g}_{j,l}\} = 2^j$. The averaging effect of $\{\tilde{g}_{j,l}\}$ thus extends over a scale of 2^j. Moreover, because the wavelet filter $\{\tilde{h}_{j,l}\}$ has the same length L_j as $\{\tilde{g}_{j,l}\}$, is orthogonal to $\{\tilde{g}_{j,l}\}$ and sums to zero, it represents the difference between two generalized averages, each occupying half the width of 2^j. Thus the wavelet coefficients at transform level j are associated with a scale of 2^{j-1}. For processes with sample interval Δt the jth level scaling coefficients are associated with a physical scale of $2^j\,\Delta t$ and the jth level wavelet coefficients with a physical scale of $2^{j-1}\,\Delta t$. We are thus able to think of the decomposition in (11) as a scale-by-scale decomposition of the autocovariance sequence of $\{X_t, t \in \mathbb{Z}\}$. For estimation considerations and examples of practical applications of such scale-based decompositions see [17] and [18].

5. Decorrelation

It has become well-recognised in recent years ([6, 8, 23]) that wavelet transforms are particularly well suited to the analysis of long-memory or '$1/f$-type' processes; these have power spectra which plot as straight lines on log-frequency/log-power scales over many octaves of frequency and tend to infinity as the frequency tends to zero. The application of the DWT to a particular discrete-time long-memory model class, namely the stationary fractionally differenced (FD) processes, was investigated in [9]. A long-memory FD(d) process can be written an an infinite order moving average process:

$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \varepsilon_{t-k}\,,$$

where $0 < d < 1/2$, $\Gamma(\cdot)$ is the gamma function, and $\{\varepsilon_t\}$ is a sequence of uncorrelated random variables (white noise) with mean zero and variance σ_ε^2. The spectrum of such a process is of the form $S(f) = \sigma_\varepsilon^2 (4\sin^2 \pi f)^{-d}$, $|f| \leq 1/2$. Using a full DWT for $N = 2^J = 2^5 = 32$ the exact 32×32 correlation matrix, (including boundary effects), of the wavelet coefficients $\{W_{j,t}^{(D)}\}$, $j = 1, \ldots, 5$ and the single scaling coefficient $V_{5,0}^{(D)}$ was computed, (for each of three different Daubechies scaling and wavelet filters). The only large off-diagonal correlations occurred between, rather than within, levels of the transform, the largest being due to boundary effects. The lack of correlation within a level can be explained by computing the covariance of the wavelet coefficients, ignoring boundary effects; from [15],

$$\mathrm{cov}\{W_{j,t}^{(D)}, W_{j,t+\tau}^{(D)}\} = \int_{-1/2}^{1/2} e^{i2^{j+1}\pi f \tau} |H_j(f)|^2 S(f) df\,,$$

where $H_j(f) = \sum_{l=0}^{L_j-1} h_{j,l} e^{-i2\pi f l}$ is the transfer function of $\{h_{j,l}\}$. The jth level wavelet filter has an (ideal) pass band $|f| \in (1/2^{j+1}, 1/2^j]$; hence if in addition to $|H_j(f)|^2$ being approximately constant over this band, $S(f)$ is also, then approximately

$$\mathrm{cov}\{W_{j,t}^{(D)}, W_{j,t+\tau}^{(D)}\} \propto \int_{-1/2^j}^{-1/2^{j+1}} e^{i2^{j+1}\pi f \tau} df + \int_{1/2^{j+1}}^{1/2^j} e^{i2^{j+1}\pi f \tau} df = 0 \quad \text{for } \tau \neq 0\,.$$

But the basis underlying the DWT ensures that $S(f)$ for a FD process is indeed approximately constant over $|f| \in (1/2^{j+1}, 1/2^j]$; although $S(f)$ rises rapidly towards zero frequency, the interval $(1/2^{j+1}, 1/2^j]$ becomes increasingly narrow towards zero frequency as j increases, 'tracking' the change in the spectrum. Hence the covariance, and correlation, is negligible within a level. Covariances between levels can be considered using similar reasoning —see [15].

The ability to decorrelate a process is a powerful weapon in statistical analysis as it enables the use of statistical tools which will fail in the presence of strong correlations (see for example [14]). For a stochastic process with a spectrum very different from that of a long-memory process, the DWT can be very sub-optimal as a decorrelator. A more flexible basis is of course provided by the DWPT; we can define a disjoint dyadic decomposition of the table in figure 1 by noting that at each potential parent node (j,n) we can either carry out the splitting using both $G(\cdot)$ and $H(\cdot)$ or not split at all. Such a resulting decomposition is also orthonormal, being associated with a nonoverlapping partition of $(0,1/2]$. A way of determining a 'best basis,' suitable for decorrelating, from all the possible orthonormal partitions, is given in [14]. The algorithm looks at each node (j,n) and carries out a statistical white noise test on $W^{(D)}_{j,n,t}$, $t = 0,\ldots,N_j - 1$. If we fail to reject the hypothesis that the DWPT coefficients are from a white noise process, then we retain $W^{(D)}_{j,n,t}$, $t = 0,\ldots,N_j - 1$, i.e., this sequence is processed no further. If we reject the hypothesis, then $W^{(D)}_{j,n,t}$, $t = 0,\ldots,N_j - 1$, is further filtered and downsampled to produce $W^{(D)}_{j+1,2n,t}$, $t = 0,\ldots,N_{j+1} - 1$ and $W^{(D)}_{j+1,2n+1,t}$, $t = 0,\ldots,N_{j+1} - 1$. Further details and application examples can be found in [14].

6. Conclusions

We have shown that a number of very useful wavelet-based tools are currently available for the analysis of discrete time series from both algorithmic and statistical perspectives. We expect many more to be developed in the coming years.

References

[1] R. N. Bracewell (1978). *The Fourier Transform and Its Applications (Second Edition).* New York: McGraw-Hill.

[2] R. R. Coifman and D. L. Donoho (1995). Translation-invariant denoising. In *Wavelets and Statistics (Lecture Notes in Statistics, Volume 103)*, (A. Antoniadis and G. Oppenheim, *eds.*), 125–50, New York: Springer-Verlag.

[3] R. R. Coifman, Y. Meyer and M. V. Wickerhauser (1992). Size properties of wavelet packets. In *Wavelets and Their Applications*, (M. B. Ruskai et al, *eds.*), 453–79, Boston: Jones and Bartlett.

[4] I. Daubechies (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.

[5] D. L. Donoho and I. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425–55.

[6] P. Flandrin (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38, 910–7.

[7] I. M. Johnstone and B. W. Silverman (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B*, 59, 319–51.

[8] E. Masry (1993). The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion. *IEEE Transactions on Information Theory*, 39, 260–4.

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

[10] G. P. Nason and B. W. Silverman (1994). The discrete wavelet transform in *S*. *Journal of Computational and Graphical Statistics*, 3, 163–91.

[11] D. B. Percival (1995). On estimation of the wavelet variance. *Biometrika*, 82, 619–31.

[12] D. B. Percival and P. Guttorp (1994). Long-memory processes, the Allan variance and wavelets. In *Wavelets in Geophysics*, (E. Foufoula-Georgiou and P. Kumar, *eds.*), 325–44, San Diego: Academic Press.

[13] D. B. Percival and H. Mofjeld (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, 92, 868–80.

[14] D. B. Percival, S. Sardy and A. C. Davison (2000). Wavestrapping time series: adaptive wavelet-based bootstrapping. In *Nonlinear and Nonstationary Signal Processing*, (W. Fitzgerald, R. L. Smith, A. T. Walden and P. Young, *eds.*), Cambridge University Press, forthcoming.

[15] D. B. Percival and A. T. Walden (2000). *Wavelet Methods for Time Series Analysis.* Cambridge UK: Cambridge University Press.

[16] A. Serroukh and A. T. Walden (2000). Wavelet scale analysis of bivariate time series I: motivation and estimation. *Journal of Nonparametric Statistics*, to appear.

[17] A. Serroukh and A. T. Walden (2000). Wavelet scale analysis of bivariate time series II: statistical properties for linear processes. *Journal of Nonparametric Statistics*, to appear.

[18] A. Serroukh, A. T. Walden and D. B. Percival (2000). Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series. *Journal of the American Statistical Association*, 95, 184–96.

[19] M. J. Shensa (1992). The discrete wavelet transform: wedding the à trous and Mallat algorithms. *IEEE Transactions on Signal Processing*, 40, 2464–82.

[20] A. T. Walden and A. Contreras Cristan (1998). The phase-corrected undecimated discrete wavelet packet transform and its application to interpreting the timing of events. *Proceedings of the Royal Society of London, Series A.*, 454, 2243–66.

[21] A. T. Walden, D. B. Percival and E. J. McCoy (1998). Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*, 46, 3153–65.

[22] M. V. Wickerhauser (1994). *Adapted Wavelet Analysis from Theory to Software Algorithms.* A K Peters.

[23] G. W. Wornell (1993). Wavelet-based representations for the $1/f$ family of fractal processes. *Proceedings of the IEEE*, 81, 1428–50.

Department of Mathematics,
Imperial College of Science, Technology and Medicine,
Huxley Building, 180 Queen's Gate,
London SW7 2BZ, UK
E-mail address: `a.walden@ic.ac.uk`