

---

# Analysis of Geophysical Time Series Using Discrete Wavelet Transforms: An Overview

Donald B. Percival

Applied Physics Laboratory, University of Washington, Box 355640, Seattle, WA 98195-5640, USA, [dbp@apl.washington.edu](mailto:dbp@apl.washington.edu)

**Summary.** Discrete wavelet transforms (DWTs) are mathematical tools that are useful for analyzing geophysical time series. The basic idea is to transform a time series into coefficients describing how the series varies over particular scales. One version of the DWT is the maximal overlap DWT (MODWT). The MODWT leads to two basic decompositions. The first is a scale-based analysis of variance known as the wavelet variance, and the second is a multiresolution analysis that reexpresses a time series as the sum of several new series, each of which is associated with a particular scale. Both decompositions are illustrated through examples involving Arctic sea ice and an Antarctic ice core. A second version of the DWT is the orthonormal DWT (ODWT), which can be extracted from the MODWT by subsampling. The relative strengths and weaknesses of the MODWT, the ODWT and the continuous wavelet transform are discussed.

## 1 Introduction

The wide-spread use of wavelets to analyze data in the geosciences can be traced back to work by Morlet and coworkers [1, 2] in the early 1980s. Their efforts were motivated by signal analysis in oil and gas exploration and resulted in the continuous wavelet transform (CWT). Work in the late 1980s by Daubechies, Mallat and others [3, 4, 5, 6] led to various discrete wavelet transforms (DWTs), which are the focus of this article. While CWTs and DWTs are closely related, DWTs are more amenable to certain types of statistical analysis, making them the transform of choice for tackling certain – but not all – problems of interest in geophysical data analysis. The intent of this article is to give an overview of how DWTs can be used in the analysis of geophysical time series, i.e., a sequence of observations recorded over time (usually at regularly spaced intervals such as once per second).

The remainder of this article is structured as follows. In Sect. 2 we review the important notion of scale and the basic ideas behind the maximal overlap DWT (MODWT). The MODWT leads to two basic decompositions. The first (the subject of Sect. 3) is a scale-based analysis of variance known as the

wavelet variance (or wavelet spectrum). The second (Sect. 4) is an additive decomposition known as a multiresolution analysis, in which a time series is reexpressed as the sum of several new series, each associated with a particular physical scale. In Sect. 5 we discuss another form of the DWT known as the orthonormal DWT (ODWT) that can be extracted from the MODWT and that has certain strengths and weaknesses in comparison to the MODWT. Our overview concentrates on the so-called Haar wavelet, but we note the existence of other wavelets in Sect. 6 and discuss why they might be preferred over the Haar wavelet for certain types of analyses. Finally we make some concluding comments in Sect. 7, including a comparison of the strengths and weaknesses of DWTs and CWTs.

## 2 Maximal Overlap Discrete Wavelet Transform

Let  $X_n$ ,  $n = 0, 1, \dots, N - 1$ , represent the  $n$ th value of a time series that has  $N$  values in all. We assume that, for all  $n$ , the time at which  $X_n$  was observed can be expressed as  $t_0 + n\Delta$ , where  $t_0$  is the time associated with  $X_0$ , and  $\Delta$  is the sampling interval between any two adjacently recorded values  $X_n$  and  $X_{n+1}$ . Given  $\tau_j = 2^{j-1}$  for some positive integer  $j$ , consider

$$A_{j,n} = \frac{1}{\tau_j} \sum_{l=0}^{\tau_j-1} X_{n-l}, \quad (1)$$

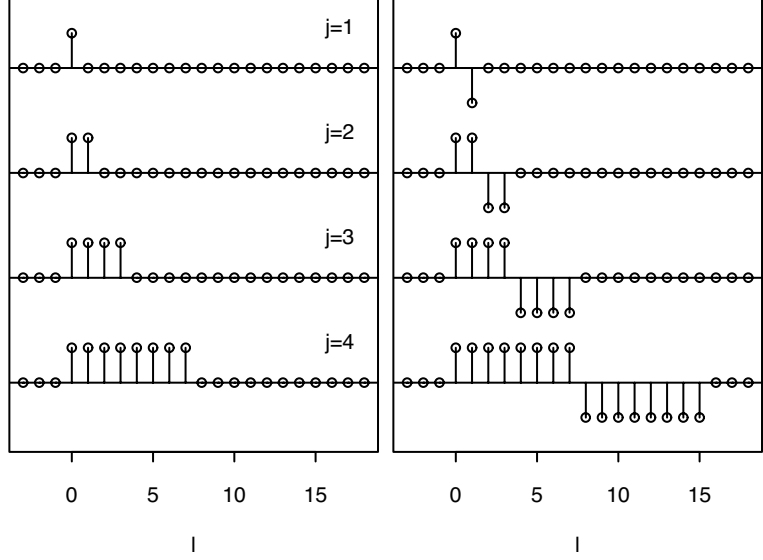
which is the average of  $\tau_j$  adjacent values of the series starting with  $X_{n-\tau_j+1}$  and ending at  $X_n$ . We refer to the above as a scale  $\tau_j$  average. The variable  $\tau_j$  is sometimes called a dyadic scale since its values are restricted to be powers of two. It is a dimensionless scale that is associated with a physical scale of  $\tau_j \Delta$ . Since  $\tau_1 = 1$  and hence  $A_{1,n} = X_n$ , we can think of the original series as being unit scale ‘averages’.

The definition for  $A_{j,n}$  makes sense as long as  $\tau_j - 1 \leq n \leq N - 1$ ; however,  $A_{j,n}$  is ill-defined when  $\tau_j \geq 2$  and  $0 \leq n \leq \tau_j - 2$  because (1) would then involve  $X_{-1}$  and possibly other values of the time series we don’t have access to. To force  $A_{j,n}$  to be well defined for the full range  $0 \leq n \leq N - 1$ , we assume that the time series is periodic with a period of  $N$ ; i.e.,  $X_n = X_{n+N}$  for all integers  $n$ . With this definition,  $X_{-1} = X_{N-1}$ ,  $X_{-2} = X_{N-2}$  and so forth. This assumption introduces some ‘boundary’ averages such as  $A_{2,0} = (X_0 + X_{N-1})/2$ , which combine nonadjacent values from the original series when  $N > 2$ . For  $\tau_2 = 2$ , the only possible boundary average is  $A_{2,0}$ , while the other  $N - 1$  averages  $A_{2,1}, \dots, A_{2,N-1}$  involve adjacent values from the time series.

If we let  $a_{j,l} = 1/\tau_j$  for  $0 \leq l \leq \tau_j - 1$ , we can reexpress (1) in filtering notation as

$$A_{j,n} = \sum_{l=0}^{\tau_j-1} a_{j,l} X_{n-l}, \quad n = 0, 1, \dots, N - 1.$$

The left-hand portion of Fig. 1 shows the filters  $a_{j,l}$  for the dyadic scales indexed by  $j = 1, 2, 3$  and 4 (we define  $a_{j,l}$  to be zero when  $l < 0$  or  $l \geq \tau_j$ ).



**Fig. 1.** Averaging filters  $a_{j,l}$  (left-hand panel) for dyadic scales  $\tau_j = 2^{j-1}$ ,  $j = 1, 2, 3$  and 4, and related differencing filters  $d_{j,l}$  (right). The averaging filters are proportional to Haar scaling filters, and the differencing filters, to Haar wavelet filters.

While averages of time series over various scales are of interest in their own right, what is often of more interest is how these averages change over time. For example, a key question about various indicators of climate is whether their average values over certain time scales have changed significantly with time. The wavelet transform is a mechanism that allows us to quantify how averages of a time series over particular scales change from one interval of time to the next. These changes are quantified in wavelet coefficients, which form the bulk of any DWT.

Wavelet coefficients in a DWT are organized into sets. There is one set for each dyadic scale  $\tau_j$ , and each coefficient in this set is proportional to the difference between two adjacent nonoverlapping averages. Mathematically, these differences are given by

$$D_{j,n} = A_{j,n} - A_{j,n-\tau_j} = \sum_{l=0}^{2\tau_j-1} d_{j,l} X_{n-l}, \quad n = 0, 1, \dots, N-1, \quad (2)$$

where

$$d_{j,l} = \begin{cases} 1/\tau_j, & l = 0, \dots, \tau_j - 1; \\ -1/\tau_j, & l = \tau_j, \dots, 2\tau_j - 1; \\ 0, & \text{otherwise.} \end{cases}$$

The right-hand portion of Fig. 1 shows the differencing filters  $d_{j,l}$  associated with the averaging filters  $a_{j,l}$ . If  $D_{j,n}$  is close to zero, then  $A_{j,n-\tau_j}$  and  $A_{j,n}$  are close to each other, indicating that there is not much change in these adjacent nonoverlapping averages of scale  $\tau_j$ ; on the other hand, if  $D_{j,n}$  has a large magnitude, then the two scale  $\tau_j$  averages differ considerably.

We can now define the Haar maximal overlap discrete wavelet transform (MODWT) of maximum level  $J_0$ , where  $J_0$  is a positive integer that we are free to select. This transform consists of  $J_0 + 1$  sets of  $N$  coefficients, for a total of  $(J_0 + 1) \times N$  coefficients in all. There are  $J_0$  sets of wavelet coefficients, and the remaining set consists of the so-called scaling coefficients. For  $j = 1, \dots, J_0$ , the wavelet coefficients are given by  $\widetilde{W}_{j,n} = D_{j,n}/2$ , while the single set of scaling coefficients is given by  $\widetilde{V}_{J_0,n} = A_{J_0+1,n}$ , where  $n = 0, 1, \dots, N - 1$  in both cases. Let  $\mathbf{X}$  be an  $N$  dimensional column vector containing the time series  $X_n$ , and let  $\widetilde{\mathbf{W}}_j$  be a similar vector containing the level  $j$  MODWT wavelet coefficients  $\widetilde{W}_{j,n}$ . We can then write

$$\widetilde{\mathbf{W}}_j = \widetilde{\mathcal{W}}_j \mathbf{X}, \quad (3)$$

where  $\widetilde{\mathcal{W}}_j$  is an  $N \times N$  matrix whose rows can be deduced by studying (2). For example, if  $N = 7$  and  $j = 2$  so that  $\tau_2 = 2$ , we find that

$$\widetilde{\mathcal{W}}_2 = \begin{bmatrix} 1/4 & 0 & 0 & 0 & -1/4 & -1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 0 & 0 & -1/4 & -1/4 \\ -1/4 & 1/4 & 1/4 & 0 & 0 & 0 & -1/4 \\ -1/4 & -1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & -1/4 & -1/4 & 1/4 & 1/4 & 0 & 0 \\ 0 & 0 & -1/4 & -1/4 & 1/4 & 1/4 & 0 \\ 0 & 0 & 0 & -1/4 & -1/4 & 1/4 & 1/4 \end{bmatrix}.$$

Note that any of the bottom six rows in  $\widetilde{\mathcal{W}}_2$  can be obtained by circularly shifting the row above it to the right by one, a pattern that holds for all  $\widetilde{\mathcal{W}}_j$ . Note also that the first three rows yield boundary wavelet coefficients since they combine together values of the time series that are not contiguous in time (in general, there are  $\min\{2\tau_j - 1, N\}$  boundary coefficients). In a similar manner, if  $\widetilde{\mathbf{V}}_{J_0}$  is an  $N$  dimensional column vector containing the scaling coefficients  $\widetilde{V}_{J_0,n}$ , then we can write

$$\widetilde{\mathbf{V}}_{J_0} = \widetilde{\mathcal{V}}_{J_0} \mathbf{X}, \quad (4)$$

where  $\widetilde{\mathcal{V}}_{J_0}$  is an  $N \times N$  matrix whose rows are dictated by (1).

In practice the MODWT wavelet and scaling coefficients are not computed directly via (3) and (4), but rather via an efficient recursive procedure known as the pyramid algorithm (for pseudo-code describing this algorithm, see pp. 177–178 of [7]).

### 3 Analysis of Variance via the Wavelet Variance

The MODWT leads to two basic decompositions for a time series  $X_n$ . The first is an analysis of variance (ANOVA) that is based on a decomposition of the ‘energy’ in  $X_n$  (the second is discussed in Sect. 4). By definition the energy in a time series is just the sum of its squared values:

$$\sum_{n=0}^{N-1} X_n^2 = \mathbf{X}^T \mathbf{X} = \|\mathbf{X}\|^2,$$

where ‘ $T$ ’ denotes the transpose operation, and  $\|\mathbf{X}\|$  is the Euclidian norm of  $\mathbf{X}$ . This decomposition states that

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2, \quad (5)$$

so the energy in the series is preserved in its MODWT wavelet and scaling coefficients.

Let  $\sigma_X^2$  be the sample variance for the time series:

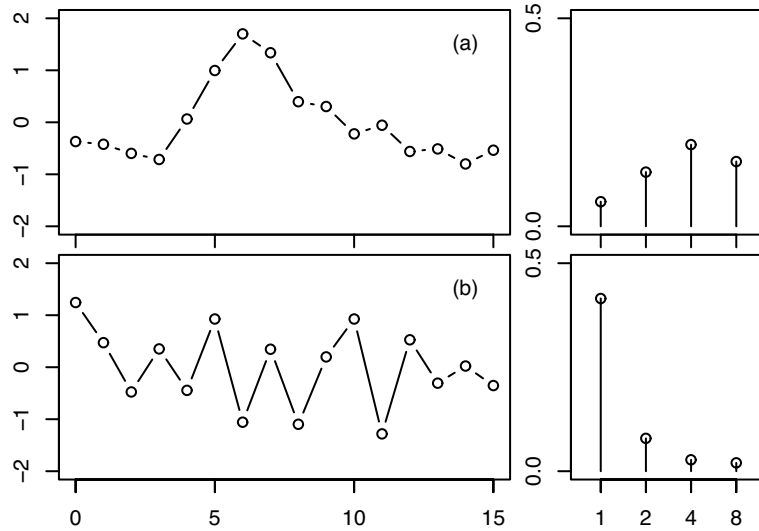
$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{n=0}^{N-1} (X_n - \bar{X})^2 = \frac{1}{N} \sum_{n=0}^{N-1} X_n^2 - \bar{X}^2, \quad \text{where } \bar{X} = \frac{1}{N} \sum_{n=0}^{N-1} X_n.$$

It follows from (5) that

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \frac{1}{N} \|\widetilde{\mathbf{V}}_{J_0}\|^2 - \bar{X}^2. \quad (6)$$

In the above, we refer to  $\|\widetilde{\mathbf{W}}_j\|^2/N = \hat{\nu}_j^2$  as the empirical wavelet variance. We can regard  $\hat{\nu}_j^2$  as an appropriate definition for the sample variance of the level  $j$  wavelet coefficients. This assumes that the mean value of  $\widetilde{\mathbf{W}}_j$  can be taken to be zero, which is reasonable for certain  $X_n$  because of the differencing operation inherent in the filters used in (2). On the other hand,  $\frac{1}{N} \|\widetilde{\mathbf{V}}_{J_0}\|^2 - \bar{X}^2$  is the sample variance of the scaling coefficients because  $\widetilde{\mathbf{V}}_{J_0}$  is a running average of  $\mathbf{X}$  and hence has a sample mean of  $\bar{X}$ . Equation (6) thus gives us a scale-based ANOVA, in that we are breaking  $\hat{\sigma}_X^2$  up into  $J_0 + 1$  pieces, each of which can be interpreted in terms of sample variances of either differences in averages over the dyadic scales  $\tau_1, \dots, \tau_{J_0}$  or averages over a scale of  $2\tau_{J_0} = \tau_{J_0+1}$ . If  $N = 2^J$  for some positive integer  $J$  and if we set  $J_0 = J$ , the contribution to  $\hat{\sigma}_X^2$  due to the scaling coefficients drops out because  $\widetilde{\mathbf{V}}_{J_0}$  becomes a vector whose elements are all equal to  $\bar{X}$ , and we then have

$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 = \sum_{j=1}^{J_0} \hat{\nu}_j^2. \quad (7)$$

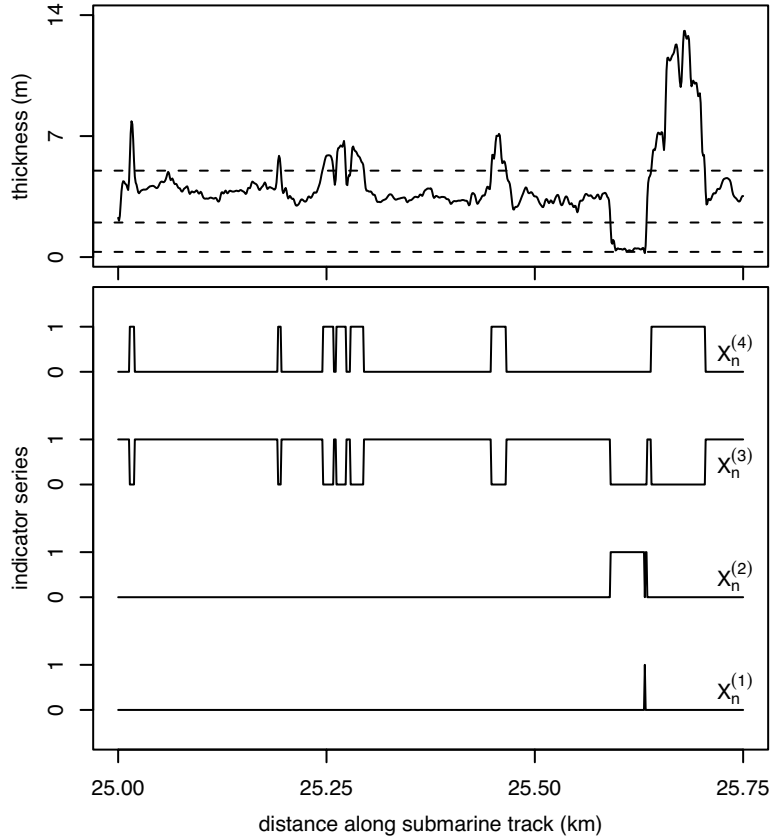


**Fig. 2.** Two time series (left-hand plots), each with  $N = 16$  values. Both series have the same sample means and variances. The right-hand plots show the corresponding the Haar MODWT wavelet variances over the dyadic scales 1, 2, 4 and 8.

Even if these stipulations on  $N$  and  $J_0$  are dropped, the above is still a good approximation as long as  $J_0$  is large enough so that  $\tau_{J_0+1}$  is close to  $N$ .

To see how the wavelet variance can help characterize time series, consider the two artificial series shown in the left-hand column of plots in Fig. 2. By construction both series have exactly the same sample mean and variance, but their appearances are quite different. Series (a) varies more slowly than series (b), which tends to fluctuate back and forth from one time point to the next. The right-hand plots show the corresponding empirical wavelet variances versus the dyadic scales  $\tau_1, \dots, \tau_4$ . The wavelet variances for the two series have their largest values at different scales, namely, scale  $\tau_3 = 4$  for (a) and  $\tau_1 = 1$  for (b). Small-scale fluctuations are thus an important part of the overall variability of series (b), but less so for (a), where larger scale fluctuations are more prominent. Although the sample mean and variance are here incapable of distinguishing between the two series, the scale-based ANOVA given by the wavelet variance can in a manner that is intuitively reasonable.

The next two subsections consider ‘real world’ examples, both involving Arctic sea ice. Other examples of the use of the wavelet variance in geophysics include the study of the El Niño–Southern Oscillation [8], surface albedo and temperature in desert grasslands [9], soil variations [10], the relationship between rainfall and runoff [11], ocean surface waves [12], solar coronal activity [13], North Atlantic sea levels [14], atmospheric turbulence [15] and the impact of large multi-purpose dams on water temperature variability [16].

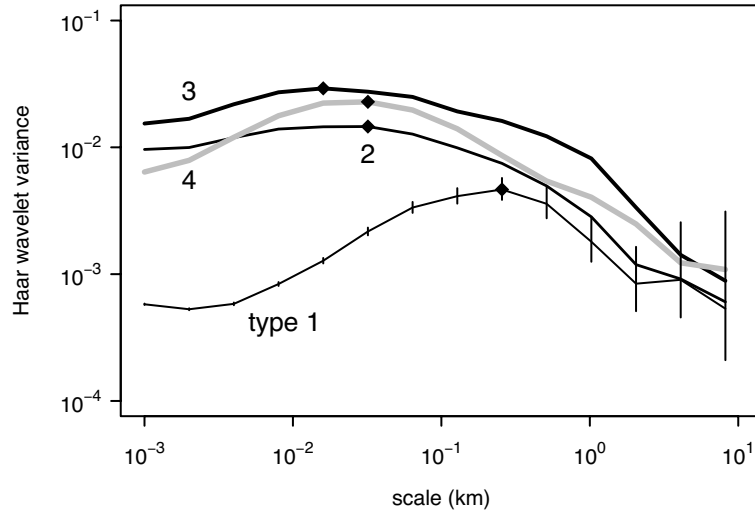


**Fig. 3.** Portion of a series of Arctic ice thickness measurements  $X_n$  versus distance along a submarine track (upper plot), along with four binary-valued series (lower) indicating the absence/presence (0/1) of four ice types: leads and new ice (defined as  $X_n < 0.3$  m and denoted as  $X_n^{(1)}$ ), first year ice ( $0.3 \leq X_n < 2$ ,  $X_n^{(2)}$ ), medium multiyear ice ( $2 \leq X_n < 5$ ,  $X_n^{(3)}$ ) and ridged ice ( $X_n \geq 5$ ,  $X_n^{(4)}$ ). The sampling interval is  $\Delta = 0.001$  km. The horizontal dashed lines in the upper plot depict the defining boundaries for the ice types. These data were taken near the North Pole in April of 1991 and are archived at the National Snow and Ice Data Center (<http://nsidc.org/>).

### 3.1 Wavelet Variance Analysis of Arctic Ice Types

Naval submarines with upward-looking sonars have collected data on sea-ice thickness in the Arctic Ocean since 1958. Currently data from 34 cruises conducted by the U.S. Navy between 1975 and 2000 have been publically archived. These data provide a unique direct look at the climatology of Arctic ice thickness as a function of space and time. The upper plot of Fig. 3 shows a 0.75 km portion of one such series of ice thickness measurements  $X_n$  (in meters) taken

near the North Pole in April of 1991 (the entire set of measurements extends over 50 km). We can regard  $X_n$  as a time series with  $\Delta = 0.001$  km, where here ‘time’ is considered as a surrogate for distance along the submarine track under the ice (the observations were recorded at regular intervals of time, but the submarine was moving at a constant speed).



**Fig. 4.** Empirical Haar wavelet variances  $\hat{\nu}_j^2$  versus physical scales  $\tau_j \Delta$ ,  $j = 1, 2, \dots, 14$ , for the four binary-valued ice type series  $X_n^{(i)}$  shown in Fig. 3. The vertical lines emanating from each  $\hat{\nu}_j^2$  for  $X_n^{(1)}$  represent 95% confidence intervals for a hypothesized theoretical wavelet variance. The largest wavelet variances for each ice type are indicated by a solid diamond – their locations define wavelet-based characteristic scales.

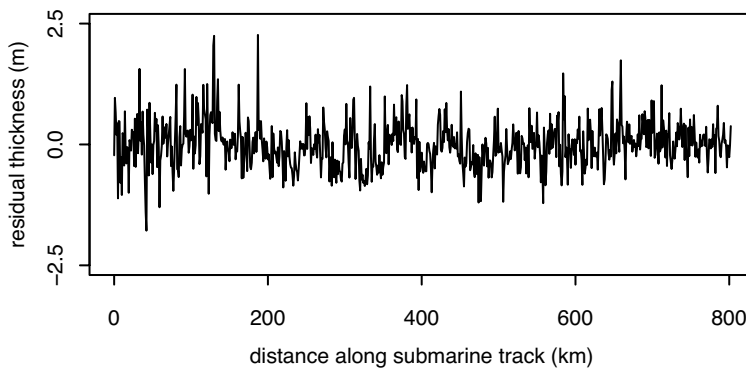
Ice thickness can be classified into four types, which are driven by different physical processes [17, 18]. The first type consists of leads and new ice and has a thickness below 0.3 m; the second is first year ice and ranges from 0.3 to 2 m; the third is medium multiyear ice, from 2 to 5m; and the fourth is ridged ice, anything above 5 m. The divisions between the four types are marked on Fig. 3 by horizontal dashed lines. Let  $X_n^{(i)}$  be a binary-valued series indicating the absence or presence (using 0 or 1) of ice type  $i$  at measurement  $X_n$ . These four indicator series are plotted in the bottom of Fig. 3.

Figure 4 shows empirical Haar wavelet variances for the four indicator series  $X_n^{(i)}$  plotted versus  $\tau_j \Delta$  for  $j$  ranging from 1 to 14 (i.e, physical scales from 0.001 up to 8.192 km). If we regard  $\hat{\nu}_j^2$  as an estimate of a hypothesized theoretical wavelet variance, we can determine how far our estimates are likely



to be off from the true wavelet variances (for details, see Chap. 8 of [7]). The vertical lines in Fig. 4 indicate 95% confidence intervals (CIs) for the true wavelet variances for ice type 1 (the three other ice types would have CIs with similar widths). Note that the widths of the CIs increase as  $\tau_j$  increases.

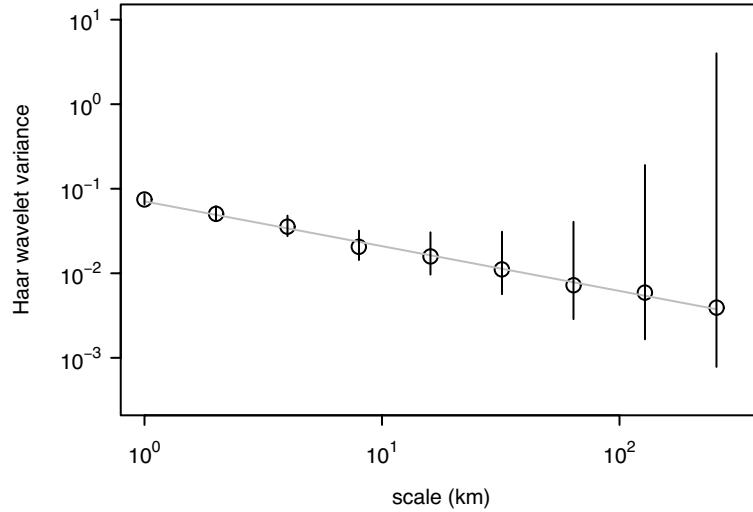
All four wavelet variance curves in Fig. 4 have a single broad peak. The largest  $\hat{\nu}_j^2$  for each ice type is marked with a solid diamond. While the scale at which the largest value occurs is similar for types 2, 3 and 4 (either 16 or 32 m), the one for type 1 is an order of magnitude larger (256 m). We can consider the location of these peak values as defining a characteristic scale for each ice type. A question of geophysical interest is how stable these characteristic scales are both spatially and temporally. This question can be addressed by using  $\hat{\nu}_j^2$  to determine these scales from data taken at other locations and times across the Arctic basin. For this application, the wavelet variance thus extracts a summary statistic that picks out the largest scale-based contributor to the sample variance of an ice-type indicator series, and this statistic can be studied across space and time to deduce possible changes in the climatology of Arctic ice thickness.



**Fig. 5.** Arctic ice thickness residuals  $X_n$  versus distance along a submarine track. The residuals are the deviations from a least squares fit of a line to a series of 1 km average thicknesses (the sampling interval  $\Delta$  is also 1 km). There are  $N = 803$  thickness measurements, and these were collected from a SCientific ICe EXpedition (SCICEX) cruise within the Arctic Ocean in September of 1997 and are archived at the National Snow and Ice Data Center (<http://nsidc.org/>).

### 3.2 Wavelet Variance Analysis of Averaged Ice Thickness

As a second example, let us consider another series of ice thickness measurements, but now consisting of one kilometer averages that have been detrended by subtracting off a line fit via least squares (the sampling interval is  $\Delta = 1$  km). The residuals from this fit are plotted in Fig. 5.



**Fig. 6.** Empirical Haar wavelet variances  $\hat{\nu}_j^2$  versus physical scales  $\tau_j \Delta$ ,  $j = 1, 2, \dots, 9$ , for the residual thickness series shown in Fig. 5. The vertical lines emanating from each  $\hat{\nu}_j^2$  represent 95% confidence intervals for a hypothesized theoretical wavelet variance. The line through the variances is a least squares fit of  $\log_{10}(\hat{\nu}_j^2)$  versus  $\log_{10}(\tau_j \Delta)$  and has a slope of  $-0.53$ .

The empirical wavelet variances  $\hat{\nu}_j^2$  for  $j = 1, \dots, 9$  for the residual thicknesses are shown in Fig. 6, along with 95% confidence intervals for a hypothesized theoretical wavelet variance (the vertical lines). A linear least squares fit of  $\log_{10}(\hat{\nu}_j^2)$  versus  $\log_{10}(\tau_j \Delta)$  is also shown (the line with a slope of  $-0.53$ ). With 1 km averaging, the largest wavelet variance occurs at the smallest scale, but what is of more interest is the rate of decay of  $\hat{\nu}_j^2$  with increasing scale. This decay is very close to linear on a log/log scale. This form of decay is indicative of a stationary process whose spectral density function (SDF)  $S(f)$  is approximately proportional to a power law  $|f|^\alpha$ , where  $f$  is a Fourier frequency. For such a process, it can be argued that the theoretical wavelet variance should be approximately proportional to  $\tau_j^{-\alpha-1}$ , which implies that a log/log plot of  $\hat{\nu}_j^2$  versus  $\tau_j \Delta$  should be approximately linear, with a slope given by  $-\alpha - 1$ . The observed slope of  $-0.53$  thus maps into a power-law exponent of  $\alpha = -0.47$ . A process whose SDF is proportional to  $|f|^{-0.47}$  exhibits long-range dependence, which is characterized by an autocovariance function that decays at a slower rate than standard models such as autoregressive and/or moving average processes. This slower rate of decay has implications in assessing the sampling variability in various statistics derived from ice thickness measurements (for details, see [19, 20]).

## 4 Multiresolution Analysis

We now turn to the second basic decomposition afforded by the MODWT, which is an additive decomposition known in the wavelet literature as a multiresolution analysis (MRA). This decomposition says that we can reexpress  $\mathbf{X}$  as the sum of  $J_0 + 1$  new time series, each of which has a scale-based interpretation. In particular, define

$$\tilde{\mathbf{D}}_j = \tilde{\mathbf{W}}_j^T \tilde{\mathbf{W}}_j \quad \text{and} \quad \tilde{\mathbf{S}}_{J_0} = \tilde{\mathbf{V}}_{J_0}^T \tilde{\mathbf{V}}_{J_0}, \quad (8)$$

where  $\tilde{\mathbf{D}}_j$  and  $\tilde{\mathbf{S}}_{J_0}$  are  $N$  dimensional vectors known as, respectively, the  $j$ th level detail and the  $J_0$ th level smooth. We can now write

$$\mathbf{X} = \sum_{j=1}^{J_0} \tilde{\mathbf{D}}_j + \tilde{\mathbf{S}}_{J_0}, \quad (9)$$

where  $\tilde{\mathbf{D}}_j$  is a time series reflecting variations in averages over a scale of  $\tau_j$  in  $\mathbf{X}$ , whereas  $\tilde{\mathbf{S}}_{J_0}$  is a series reflecting averages over a scale of  $\tau_{J_0+1}$ . Note that we can recover our original time series  $\mathbf{X}$  from its MODWT, which tells us that no information about the series has been lost in transforming it and that (8) constitutes the pieces of an inverse MODWT. Thus, if we know how a time series varies at the dyadic scales  $\tau_1, \dots, \tau_{J_0}$  and if we know its averages over a scale of  $\tau_{J_0+1}$ , then we can reconstruct the series perfectly. If we compare (9) to a level  $J_0 + 1$  decomposition, namely,

$$\mathbf{X} = \sum_{j=1}^{J_0+1} \tilde{\mathbf{D}}_j + \tilde{\mathbf{S}}_{J_0+1},$$

we can deduce that, for all  $j$ ,

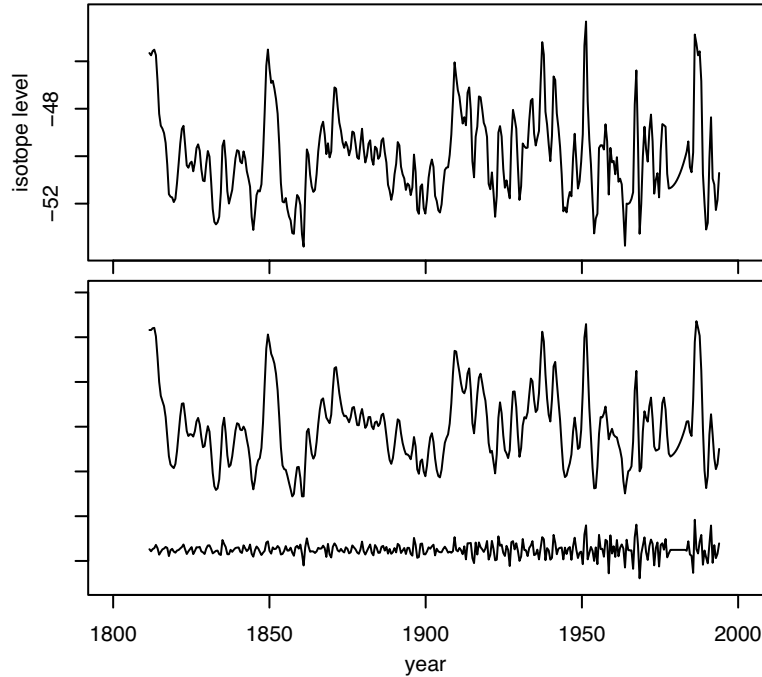
$$\tilde{\mathbf{S}}_j = \tilde{\mathbf{S}}_{j+1} + \tilde{\mathbf{D}}_{j+1}, \quad (10)$$

and hence the details can be interpreted as the differences between successive smooths. If  $N = 2^J$  and if we again set  $J_0 = J$ , then (9) becomes

$$\mathbf{X} = \sum_{j=1}^{J_0} \tilde{\mathbf{D}}_j + \bar{X}\mathbf{1}, \quad (11)$$

where  $\mathbf{1}$  is an  $N$  dimensional vector, all of whose elements are ones.

As a simple example of an MRA, let us consider a series  $\mathbf{X}$  of  $N = 352$  oxygen isotope measurements from an ice core taken at one location on a spatial array with 3.5 to 7 km spacing in Dronning Maud Land, Antarctica. Here the spacing between observations is taken to be  $\Delta = 0.5$  years (the raw measurements are indexed by distance along the core, but these are then mapped to values at half-year intervals). The series is plotted in the upper panel of



**Fig. 7.** Oxygen isotope measurements from an Antarctic ice core (top panel), along with a Haar MODWT-based multiresolution analysis of level  $J_0 = 1$  consisting of the smooth series  $\tilde{\mathbf{S}}_1$  and a single detail series  $\tilde{\mathbf{D}}_1$  (bottom panel, upper and lower plots, respectively). Due to the large difference between the beginning and end of the series, the MODWT was computed using so-called reflection boundary conditions rather than the periodic conditions described in Sect. 2 (for details, see p. 140 of [7]). Data is courtesy of Lars Karlöf, Norwegian Polar Institute, Polar Environmental Centre, Tromsø, Norway.

Fig. 7 and has a temporal span of 176 years. For each of the cores in the array, an MRA was conducted in order to compare details with similar scales to ascertain which scales are dominated by environmental noise and which might contain a common signal (see [21] for details). Here we demonstrate that the simplest possible MRA for the core shown in Fig. 7 reveals some interesting properties not readily apparent in a plot of the data itself.

The lower panel of Fig. 7 depicts a level  $J_0 = 1$  Haar MODWT-based MRA, consisting of a smooth series  $\tilde{\mathbf{S}}_1$  and a single detail series  $\tilde{\mathbf{D}}_1$ , which, upon being added together, yield  $\mathbf{X}$ . The smooth series is the portion of  $\mathbf{X}$  that can be attributed to averages over a scale of a year, whereas the detail series represents variations over a half-year scale. What is interesting is that the local variability in  $\tilde{\mathbf{D}}_1$  increases gradually with the passage of time. This increase is not readily apparent in the plot of  $\mathbf{X}$  itself, but the MRA pulls it out

clearly. The physical mechanism behind this increase is not fully understood, but is thought to be due to diffusion.

In addition the MRA reveals an artifact in  $\tilde{\mathbf{D}}_1$  centered at 1981, around which the detail series is flat for a stretch of 5.5 years. This is due to a linear interpolation scheme used to fill in a break in the ice core. Differencing a series of the form  $X_n = a + bn$  leads to wavelet coefficients that are proportional to the slope  $b$  and a detail series that is flat. For the questions that this MRA and those for other cores in the spatial array were used to address, filling in a small number of short gaps by linear interpolation is acceptable. Had we been interested in estimating the wavelet variance for a series with many gaps, linear interpolation could bias the estimates unacceptably towards zero. In this case it is advisable to use either a wavelet variance estimator that is specifically designed to work with gappy time series [22] or a stochastic interpolation scheme that preserves the small scale properties of the time series based upon a nominal stochastic model (see [19], Appendix B).

Other examples of the use of MRAs in geophysics include the analysis of subtidal sea level fluctuations [23], magnetic storm activity [24], the Lisbon and Gibraltar North Atlantic Oscillation winter indices [25], spatial variation of microflora abundance in agricultural soil [26], the December 26th 2004 tsunami as recorded along the southeastern coast of Brazil [27] and large-scale coherent structures in turbulent separation bubbles [28].

## 5 Orthonormal Discrete Wavelet Transform

While a level  $J_0$  MODWT of a time series of length  $N$  consists of a total of  $(J_0 + 1) \times N$  values, it is also possible to define a discrete wavelet transform that consists of just  $N$  values. This transform is orthonormal, which means that the transpose of  $N \times N$  matrix  $\mathcal{W}$  relating the time series to the transform coefficients is the inverse of  $\mathcal{W}$ . We hence use the acronym ‘ODWT’ to denote this transform. We can readily define the ODWT in terms of the MODWT if  $N$  happens to be an integer multiple of  $2^{J_0}$  (if  $N$  is not of this form, an ODWT can still be defined, but not as easily – see pp. 141–145 of [7] for details). The ODWT wavelet coefficients are given by

$$W_{j,n} = 2^{j/2} \tilde{W}_{j,2^j(n+1)-1}, \quad n = 0, 1, \dots, \frac{N}{2^j} - 1, \quad j = 1, 2, \dots, J_0;$$

i.e., the ODWT coefficients are obtained by subsampling and rescaling the MODWT coefficients. For example, at level  $j = 1$ , the ODWT coefficients are formed by taking the MODWT coefficients with odd indices and multiplying them by  $\sqrt{2}$ , whereas, at level  $j = 2$ , we subsample every fourth coefficient and multiply them by 2:

$$\sqrt{2} \begin{bmatrix} \widetilde{W}_{1,1} \\ \widetilde{W}_{1,3} \\ \widetilde{W}_{1,5} \\ \vdots \\ \widetilde{W}_{1,N-3} \\ \widetilde{W}_{1,N-1} \end{bmatrix} = \begin{bmatrix} W_{1,0} \\ W_{1,1} \\ W_{1,2} \\ \vdots \\ W_{1,\frac{N}{2}-2} \\ W_{1,\frac{N}{2}-1} \end{bmatrix} = \mathbf{W}_1 \quad \& \quad 2 \begin{bmatrix} \widetilde{W}_{2,3} \\ \widetilde{W}_{2,7} \\ \widetilde{W}_{2,11} \\ \vdots \\ \widetilde{W}_{2,N-5} \\ \widetilde{W}_{2,N-1} \end{bmatrix} = \begin{bmatrix} W_{2,0} \\ W_{2,1} \\ W_{2,2} \\ \vdots \\ W_{2,\frac{N}{4}-2} \\ W_{2,\frac{N}{4}-1} \end{bmatrix} = \mathbf{W}_2 .$$

As the level  $j$  increases, we need to subsample fewer and fewer MODWT wavelet coefficients in order to create the corresponding ODWT coefficients in  $\mathbf{W}_j$ . In a similar manner the ODWT scaling coefficients are defined by

$$V_{J_0,n} = 2^{J_0/2} \widetilde{V}_{J_0,2^{J_0}(n+1)-1}, \quad n = 0, 1, \dots, N_{J_0} - 1,$$

and can be placed in an  $N_{J_0}$  dimensional vector denoted as  $\mathbf{V}_{J_0}$ . The ODWT of level  $J_0$  consists of the collection of vectors  $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{J_0}$  and  $\mathbf{V}_{J_0}$ , whose dimensions are, respectively,  $N/2, N/4, \dots, N/2^{J_0}$  and  $N/2^{J_0}$ , which collectively sum to  $N$ .

As was true for the MODWT, the ODWT leads to a scale-based ANOVA and an MRA. We start by considering the analogs of (3) and (4):

$$\mathbf{W}_j = \mathcal{W}_j \mathbf{X} \quad \text{and} \quad \mathbf{V}_{J_0} = \mathcal{V}_{J_0} \mathbf{X}, \quad (12)$$

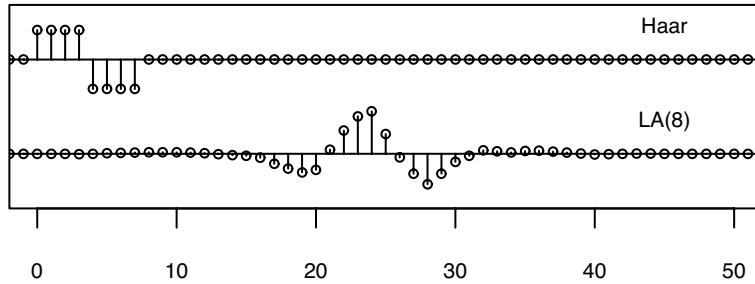
where  $\mathcal{W}_j$  is an  $\frac{N}{2^j} \times N$  matrix whose rows are selected rescaled rows from  $\widetilde{\mathcal{W}}_j$ , while  $\mathcal{V}_{J_0}$  is an  $\frac{N}{2^{J_0}} \times N$  matrix whose rows are selected rescaled rows from  $\widetilde{\mathcal{V}}_{J_0}$ . The ODWT-based ANOVAs and MRAs are easy to state: just remove all the tildes from Equations (5) through (11)! In practice the ODWT wavelet and scaling coefficients are not computed by subsampling the corresponding MODWT coefficients, but rather via an efficient pyramid algorithm (pseudocode for this algorithm is given on pp. 100–101 of [7]).

In general, MODWT-based ANOVAs and MRAs tend to perform better than their ODWT equivalents because of the deleterious effect that subsampling can have on the ODWT (for details, see Sects. 5.1, 5.6 and 8.3 of [7]); however, the ODWT is the transform of choice for certain other types of analyses. For example, if a time series can be modeled as a signal plus Gaussian white noise, then its ODWT consists of a transformed signal plus Gaussian white noise. Certain types of signals are more easily recognized in the ODWT domain than in their original time domain representation, which makes it possible to design effective data-adaptive procedures for extracting signals buried in white noise. This fact is exploited in the large body of literature devoted to wavelet shrinkage; see [29] for a recent review article that emphasize this use of the ODWT. (There is a device called ‘cycle spinning’ in which ODWT-based signal extraction is applied to a time series and all its possible circular shifts, followed by an averaging of the  $N$  extracted signals. This procedure is equivalent to a signal estimation procedure based upon the MODWT; for details, see pp. 429–431 of [7]).

As a second example, the ODWT transforms certain – but not all – time series into a collection of wavelet coefficients that are approximately uncorrelated within and between levels, but that have possibly level-dependent variances. Time series with long-range dependence are examples of ones that are effectively decorrelated by the ODWT. This decorrelating property can be put to good use in formulating wavelet-based approximate maximum likelihood estimators of parameters associated with processes with long-range dependence, in simulating series with long-range dependence and in formulating bootstrap procedures for assessing the sampling variability in certain statistics (for details, see [7, 30]).

### 6 Beyond the Haar Wavelet

Our discussion so far has focused on the Haar MODWT and corresponding ODWT, but there are other versions of both transforms. For a selected maximum level  $J_0$ , these transforms can be formulated in terms of wavelet filters of levels  $j = 1, \dots, J_0$  and a scaling filter of level  $J_0$ . Figure 8 shows the level  $j = 3$  wavelet filters for the Haar transform and LA(8) transform, where ‘LA(8)’ stands for the member of the Daubechies ‘least asymmetric’ family whose level  $j = 1$  filter has width  $L = 8$  [31]. The shape of the Haar filter tells us that the corresponding wavelet coefficients are proportional to differences of adjacent simple averages of scale 4. The shape of the LA(8) filter says that the wavelet coefficients can be interpreted as the difference between a centrally located weighted average and weighted averages occurring before and after it. Once the wavelet and scaling filters have been used to properly formulate the matrices  $\tilde{\mathcal{W}}_j$ ,  $\tilde{\mathcal{V}}_{J_0}$ ,  $\mathcal{W}_j$ , and  $\mathcal{V}_{J_0}$  of (3), (4) and (12), all of the equations involving the Haar MODWT and ODWT presented in Sects. 2 to 5 also hold for the corresponding LA(8) transforms.



**Fig. 8.** Filters used to produce scale  $\tau_3$  wavelet coefficients based upon the Haar wavelet filter (top) and the Daubechies least asymmetric wavelet filter of width 8 (bottom).

The LA(8) transform can yield a more informative ANOVA and MRA than the Haar for certain time series because the latter can suffer from ‘leakage’ effects in which the wavelet coefficients for a particular scale are locked into patterns driven by a nearby dominant scale. The fact that the wavelet coefficients are highly correlated between different levels is undesirable because the transform is then not successfully partitioning out different aspects of a time series into different coefficients. In many geophysical applications, including the ones used as examples in Sects. 3 and 4, an analysis based upon the Haar wavelet is entirely adequate, and there is no need to consider other wavelets. An effective procedure for deciding if the Haar wavelet is adequate or not is to compare analyses based upon the Haar wavelet with those based upon other wavelets. If the analyses are basically the same, there is no need to use anything other than the Haar; if not, an analysis using something other than the Haar wavelet might be called for. Use of non-Haar MODWTs and ODWTs produces more boundary coefficients, so there is a price to pay in abandoning a Haar-based analysis.

## 7 Concluding Comments

Hopefully the overview presented here has given the reader some idea of the potential uses of DWTs in analyzing geophysical time series. There are many aspects of wavelet analysis that we have not touched upon, including the fact that all of the procedures we have discussed can be applied to time series whose statistical properties are evolving over time. The ability of DWTs to handle this case, which is eluded to briefly in the MRA for the oxygen series presented in Fig. 7, is tied up with the fact that the wavelet coefficients extract information not only across different scales, but also across time. For example, a wavelet variance estimator in which the squared wavelet coefficients are averaged locally rather than globally (as in the construction of  $\hat{\nu}_j^2$ ) is an effective way of studying time-varying properties in a time series. The reader should consult [7] for details on this and other aspects of wavelet analysis not covered in this brief overview.

In Sect. 5 we discussed some of the relative strengths and weaknesses of the MODWT and the ODWT. These DWTs are closely related to corresponding continuous wavelet transforms (CWTs), which also are quite commonly used to analyze geophysical time series. A CWT might be called an ‘anti-statistic’ in the sense that, rather than summarizing the information in a time series, it converts it into a two-dimensional field. As a result, there is a considerable amount of redundant information in a CWT, which is both a strength and a weakness. One example where this redundancy is a strength is in the analysis of certain types of singularities (‘cusps’), where the nature of the singularity can be deduced by tracing the wavelet transform modulus maxima across a fine grid of scales (see, e.g., [32], Fig. 6.5). The dyadic scales used in DWTs are typically too coarsely spaced to make this type of singularity analysis feasible.



The redundancy in the CWT, however, can make proper interpretation of ‘heat’ plots of the CWT problematic, i.e., scale versus time plots in which the magnitudes of the CWT coefficients are color-coded. These plots often have rather striking structures that our eyes are drawn toward, but that can be largely attributed to the fact that CWT coefficients are typically highly correlated both spatially and temporarily. Proper statistical assessment of the significance of these structures involves some subtle issues [33], particularly if they are picked out by eye prior to being assessed. Subsampling to the dyadic scales in the MODWT and ODWT essentially breaks this correlation structure spatially, and subsampling the MODWT to get the ODWT does the same temporarily. The fact that collections of coefficients from these DWTs are approximately uncorrelated makes it easier to devise statistical tests and to implement bootstrapping procedures (the latter are not feasible with CWTs).

Finally we note that the CWT does not formally involve components in a time series that are handled in DWTs by the scaling coefficients. These are often useful for extracting large-scale trends that are an important part of some geophysical time series and that are a key component in wavelet-based signal extraction.

## Acknowledgments

The author gratefully acknowledges partial support for preparation of this article from the U.S. National Science Foundation under award number 0529955.

## References

1. Goupillaud, P., Grossmann, A. and Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, **23**, 85–102.
2. Grossmann, A. and J. Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, **15**, 723–36.
3. Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, **41**, 909–996.
4. Mallat, S.G. (1989). Multiresolution approximations and wavelet orthonormal bases of  $L^2(\mathbb{R})$ . *Transactions of the American Mathematical Society*, **315**, 69–87.
5. Mallat, S.G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 674–693.
6. Mallat, S.G. (1989). Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **37**, 2091–2110.
7. Percival, D.B. and Walden, A.T. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge UK.

8. Torrence, C. and Compo, G.P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, **79**, 61–78.
9. Pelgrum, H., Schmutge, T., Rango, A., Ritchie, J. and Kustas, B. (2000). Length-scale analysis of surface albedo, temperature, and normalized difference vegetation index in desert grassland. *Water Resources Research*, **36**, 1757–1766.
10. Lark, R.M. and Webster, R. (2001). Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet transform. *European Journal of Soil Science*, **52**, 547–562.
11. Labat, D., Ababou, R. and Mangin, A. (2001). Introduction of wavelet analyses to rainfall/runoffs relationship for a karstic basin: The case of Licq–Athérey karstic system. *Ground Water*, **39**, 605–615.
12. Massel, S.R. (2001). Wavelet analysis for processing of ocean surface wave records. *Ocean Engineering*, **28**, 957–987.
13. Rybák, J. and Dorotovič, I. (2002). Temporal variability of the coronal green-line index (1947–1998). *Solar Physics*, **205**, 177–187.
14. Barbosa, S.M., Fernandes, M.J. and Silva, M.E. (2006). Long range dependence in North Atlantic sea level. *Physica A: Statistical Mechanics and its Applications*, **371**, 725–731.
15. Cornish, C.R., Bretherton, C.S. and Percival, D.B. (2006). Maximal overlap wavelet statistical analysis with application to atmospheric turbulence. *Boundary-Layer Meteorology*, **119**, 339–374.
16. Steel, E.A. and Lange, I.A. (2007). Using wavelet analysis to detect changes in water temperature regimes at multiple scales: Effects of multi-purpose dams in the Willamette River basin. *River Research and Applications*, **23**, 351–359.
17. Flato, G.M. (1995). Spatial and temporal variability of Arctic ice thickness. *Annals of Glaciology*, **21**, 323–329.
18. World Meteorological Organization (2007). *Sea-Ice Information Services in the World*. (3rd ed.) Publication 574.
19. Percival, D.B., Rothrock, D.A., Thorndike, A.S. and Gneiting, T. (2008). The variance of mean sea-ice thickness: effect of long-range dependence. *Journal of Geophysical Research – Oceans*, in press.
20. Rothrock, D.A., Percival, D.B. and Wensnahan, M. (2008). The decline in arctic sea-ice thickness: separating the spatial, annual, and interannual variability in a quarter century of submarine data. *Journal of Geophysical Research – Oceans*, in press.
21. Karlöf, L., Winebrenner, D.P. and Percival, D.B. (2006). How representative is a time series derived from a firn core? A study at a low accumulation site on the Antarctic plateau. *Journal of Geophysical Research – Earth Surface*, **111**, F04001, doi:10.1029/2006JF000552.
22. Mondal, D. and Percival, D.B. (2008). Wavelet variance analysis for gappy time series. *Annals of the Institute of Statistical Mathematics*, under review.
23. Percival, D.B. and Mofjeld, H.O. (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association*, **92**, 868–80.
24. Jach, A., Kokoszka, P., Sojka, J. and Zhu, L. (2006). Wavelet-based index of magnetic storm activity. *Journal of Geophysical Research – Space Physics*, **111**, A09215, doi:10.1029/2006JA011635.
25. Barbosa, S., Silva, M.E. and Fernandes, M.J. (2006). Wavelet analysis of the Lisbon and Gibraltar North Atlantic Oscillation winter indices. *International Journal of Climatology*, **26**, 581–593

26. Barnes, R.J., Baxter, S.J. and Lark, R.M. (2007). Spatial covariation of *Azotobacter* abundance and soil properties: A case study using the wavelet transform. *Soil Biology and Biochemistry*, **39**, 295–310.
27. França, C.A.S. and De Mesquita, A.R. (2007). The December 26th 2004 tsunami recorded along the southeastern coast of Brazil. *Natural Hazards*, **40**, 209–222.
28. Chun, S.J., Liu, Y.Z. and Sung, H.J. (2007). Multi-resolution analysis of the large-scale coherent structure in a turbulent separation bubble affected by an unsteady wake. *Journal of Fluids and Structures*, **23**, 85–100.
29. Wang, Y. (2006). Selected review on wavelets in statistics. In *Frontiers in Statistics: Dedicated to Peter John Bickel in Honor of His 65th Birthday* (J. Fan and H.L. Koul, ed.). Imperial College Press, London, pp. 163–182.
30. Percival, D.B., Sardy, S. and Davison, A.C. (2001). Wavestrapping time series: adaptive wavelet-based bootstrapping. In *Nonlinear and Nonstationary Signal Processing* (W.J. Fitzgerald, R.L. Smith, A.T. Walden and P.C. Young, ed.). Cambridge University Press, Cambridge UK, pp. 442–470.
31. Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
32. Mallat, S.G. (1999). *A Wavelet Tour of Signal Processing* (2nd ed.). Academic Press, San Diego.
33. Maraun, D., Kurths, J. and Holschneider, M. (2007). Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing. *Physical Review E* **75**, 016707.