Chapter 9
Wavelets

9.1 Introduction

A wavelet is a wave-like oscillation that is localized in the sense that it grows from zero, reaches a maximum amplitude, and then decreases back to zero amplitude again. It thus has a location where it maximizes, a characteristic oscillation period, and also a scale over which it amplifies and declines. Wavelet analysis developed in the largely mathematical literature in the 1980's and began to be used commonly in geophysics in the 1990's. Wavelets can be used in signal analysis, image processing and data compression. They are useful for sorting out scale information, while still maintaining some degree of time or space locality. Wavelets have been used to compress and store fingerprint information. Because the wavelet and scaling functions are obtained by scaling and translating one or two "mother functions", time-scale wavelets are particularly appropriate for analyzing fields that are fractal. Wavelets can be appropriate for analyzing non-stationary time series, whereas Fourier analysis generally is not. They can be applied to time series as a sort of fusion (or compromise) between filtering and Fourier analysis. Wavelets can be used to compress the information in two-dimensional images from satellites or ground based remote sensing techniques such as radars. Wavelets are useful because as you remove the highest frequencies, local information is retained and the image looks like a low resolution version of the full pictures. With Fourier analysis, or other global functional fits, the image may lose all resemblance to the picture, after a few harmonics are removed. This is because wavelets are a hierarchy of local fits, and retain some time localization information, and Fourier or polynomial fits are global fits, usually.

In general, you can think of wavelets as a compromise between looking at digital data at the sampled times, in which case you maximize the information about how things are located in time, and looking at data through a Fourier analysis in frequency space, in which you maximize your information about how things are localized in frequency and give up all information about how things are located in time. In wavelet analysis we retain some frequency localization and some time localization, so it is a compromise.

9.2 Wavelet Types

According to Meyer(1993), two fundamental types of wavelets can be considered, the Grossmann-Morlet time-scale wavelets and the Gabor-Malvar time-frequency wavelets. The more commonly used type in geophysics is probably the time-scale wavelet. These wavelets form bases in which a signal can be decomposed into a wide range of scales, in what is called a "multiresolution analysis". From this comes the obvious application in image compression, as one can call up additional detail as required until the exact image at the original resolution is reconstructed. The intervening coarse resolution images will look like the full resolution one, just fuzzier. This is not true in general of Fourier analysis, where throwing out the last few harmonics can cause the picture to change dramatically.

Time-scale wavelets are defined in reference to a "mother function" $\psi(t)$ of some real variable t. The mother function is required to have several characteristics: it must oscillate, and it must be localized in the sense that it decreases rapidly to zero as $|t|$ tends to infinity. It is also very helpfult to require that the mother function have a certain number of zero moments, according to,

$$\int_{-\infty}^{\infty} t^{m-1}\,\psi(t)\,dt = 0 \quad m = 1, 2, 3, \,.\,. \tag{9.1}$$

Here m is the *approximation condition order* of the wavelet. If the order is one, the mean of the wavelet is zero; if the order is two, the trend of the wavelet is zero, and so forth.

The mother function can be used to generate a whole family of wavelets by translating and scaling the mother wavelet.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\,\psi\Big(\frac{t-b}{a}\Big), \quad a > 0,\ b \in \Re \tag{9.2}$$

Here b is the translation parameter and a is the scaling parameter. Provided that $\psi(t)$ is real-valued, this collection of wavelets can be used as an orthonormal basis to describe any function $f(t)$. The coefficients of this expansion can be obtained through the usual projection.

$$\Psi_{a,b} = \int_{-\infty}^{\infty} \psi_{a,b}(t)\,f(t)\,dt \tag{9.3}$$

These coefficients measure the variations of the field $f(t)$ about the point b, with the scale given by a. A set of parameters a_k and b_j , representing different scales and locations, can be chosen to form an orthonormal basis set. In that case we can reconstruct the original data from the wavelets and their coefficients.

$$f(t) = \sum_j \sum_k \Psi_{a_k,b_j}\,\psi_{a_k,b_j}(t) \tag{9.4}$$

Wavelet analysis of this type can be performed on discrete data using quadrature mirror filters and pyramid algorithms. It is also possible to compute the transform using a Fourier transform technique. Sometimes a_k and b_j are varied more continuously to make useful diagrams with continuous variations of scale and location. This gives up the orthogonality, but has the advantage of making pictures with more resolution, as scale and location generally vary by factors of 2 (dyadic wavelets) in orthogonal wavelets.

In using wavelets for data analysis, it is important to find a set of them that provides a data description that is best-suited to the problem at hand. If wavelet analysis in general, or the particular set of wavelets chosen, are not well-suited to the problem at hand, they may not lead to any useful insight. For the non-expert, who just wants to get a useful representation, one is probably restricted to choosing from among a library of established wavelet bases, and most probably from among those for which software is already available. This library is very well developed, and techniques are available for determining whether an appropriate representation has been chosen. Python and Matlab both have highly developed wavelet tool kits.

We focus here in these notes on discrete wavelets and the discrete wavelet transform (DWT) and their applications. Wavelets are basis sets for expansion which, unlike Fourier series, have not only a characteristic frequency or scale, but also a location. They can be orthogonal, biorthogonal, or nonorthogonal.

9.3 The Haar Wavelet

Haar (1910) and others were seeking functional expansions that were alternatives to the sine and cosine series of Fourier (1822). He sought an orthonormal system $h_n(t)$ of functions on the interval [0,1] such that for any function $f(t)$, the series,

$$f(t) = \sum_n \langle f, h_n\rangle\, h_n(t) \tag{9.5}$$

would converge uniformly. The angle brackets indicate a suitably defined inner product on the interval [0,1]. Haar began with the initial function,

$$h(t) = \begin{cases} 1.0 & [0.0,\ 0.5] \\ -1.0 & [0.5,\ 1.0] \\ 0.0 & \text{elsewhere} \end{cases} \tag{9.6}$$

Building on this basic mother wavelet, Haar defines his sequence of expansion functions according to,

$$\begin{aligned} n &= 2^j + k \quad j \geqslant 0,\ \ 0 \leqslant k < 2^j \\ h_n(t) &= 2^{j/2}\, h(2^j t - k) \end{aligned} \tag{9.7}$$

Each of these functions is supported (has non-zero values) on the dyadic interval,

$$I_n = \left[k\, 2^{-j},\ (k+1)\, 2^{-j}\right] \tag{9.8}$$

which is included in the interval [0,1] if $0 \leqslant k < 2^j$. Here j is the level, from the mother wavelet level ($j = 0$), to the smallest baby wavelets $j = j_{max}$, k is the spatial index for each level, and n is a mode index, starting with mother ($n = 1$). To complete the set, one must add the function $H_0(t) = 1$ on the interval [0,1], which we can refer to as father, the smoothest level of detail, in this case a constant.

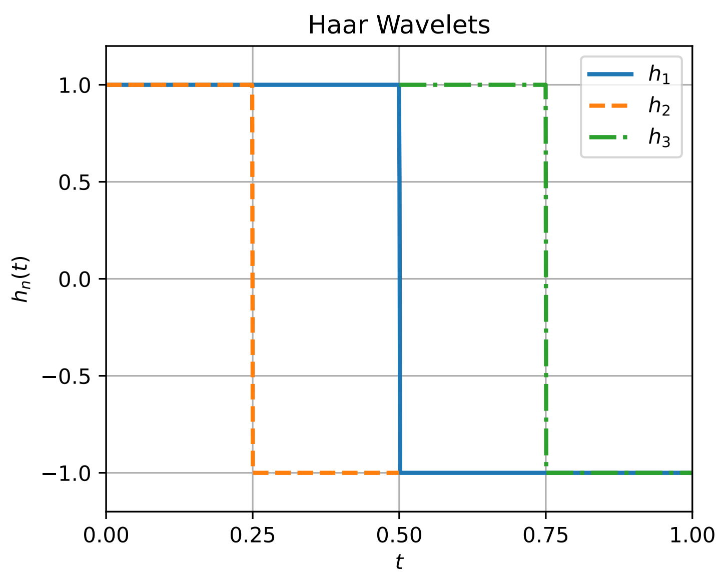

Figure 9.1 Continuous mother Haar wavelet $h_1(t)$ and her first two children.

The series $h_n(t)$ then forms an orthonormal basis on [0,1]. By looking carefully at (9.6)-(9.8) one can see that the series is the basic step function repeated on intervals that decrease in scale and increase in number by the factor of two at each level, where j is the level index and k is the number of functions at that level of detail necessary to span the interval [0,1]. Note that the mean of the Haar wavelet is zero, but that it's trend is non-zero, so that its approximation condition order is one. The 2^j in front of the Haar function in 9.7 is to normalize the functions on the interval [0, 1], but we will ignore this factor when plotting them in Figure 9.1.

9.4 Discrete Wavelet Transforms

In working with data, we have values at discrete times not at continuous times. In transforming data from time space to wavelet space. We can do this as a matrix operation, and rather that starting with the mother wavelet, we start from the finest detail that can be resolved and work our way up to the mother and father wavelet coefficients. Since the Haar wavelet is dyadic, the whole time series length must be a power of two for this to converge on the mother and father wavelets.

Since the Haar functions are orthogonal, we can derive their coefficients α_i using the relation,

$$\alpha_i = \langle \phi_i,\ x(t) \rangle \tag{9.9}$$

where the angle bracket indicates a suitably defined inner product. It may be easier to see how this is all working by considering how (9.9) looks when expressed in matrix notation, and using the abbreviation $a = 1/\sqrt{2}$.

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ .. \end{bmatrix} = \begin{bmatrix} a & a & & & & & \\ a & -a & & & & & \\ & & a & a & & & \\ & & a & -a & & & \\ & & & & a & a & \\ & & & & a & -a & \\ & & & & & & \dots \end{bmatrix} \begin{bmatrix} x(1) \\ x(2) \\ x(3) \\ x(4) \\ x(5) \\ x(6) \\ \dots \end{bmatrix} = \begin{bmatrix} y_1(1) \\ y_2(1) \\ y_1(2) \\ y_2(2) \\ y_1(3) \\ y_2(3) \\ \dots \end{bmatrix} \tag{9.10}$$

Note that the wavelet transform is divided into a smoothing part $[a, a]$ and a wavelet part $[a, -a]$. In the final column we have divided the coefficients into the smoothed coefficients $y_1(t)$ and the wavelet coefficients $y_2(t)$, each with a value at every other time step. We then continue the wavelet transform by reserving the wavelet coefficients as the highest level of detail, then perform the wavelet transform on the smoothed coefficients.

We can think of y_1 and y_2 as the time series of the coefficients of the even and odd Haar wavelets, respectively. These have only half the time resolution of the original series. You can think of y_1 as a low-frequency representation of $x(t)$ and y_2 as the high frequency details. Often in wavelet analysis literature, the smooth function $[a, a]$ would be called the scaling function , and the wavy one $[a, -a]$ would be called the wavelet . The projection into the coefficient space of the two Haar functions is equivalent to filtering followed by "down sampling", by taking only every other point of the filtered time series. The Haar transform is an example of a two-channel filter bank. It sorts the original series into two filtered data sets. The Haar filter functions are members of a special class of filter function pairs called a quadrature mirror filter pair. After the filtering is done the sum of the energies (or variances) in the two filtered time series is equal to the variance in the original time series.

$$|y_1|^2 + |y_2|^2 = |x|^2 \tag{9.11}$$

Since we are thinking of a wavelet transform as a filtering operation, now is a good time to think about the scaling achieved by this filtering process. Remember from chapter 8 on filtering of time series how we determine the frequency response of the filter from its coefficients. The scaling function $[a, a]$ is a filtering operation that does this,

$$y(t) = a\, x(t + \frac{\Delta t}{2}) + a\, x(t + \frac{-\Delta t}{2}) \tag{9.12}$$

The Fourier Tranform of this is,

$$Y(\omega) = X(\omega)\left(a e^{i\omega\Delta t/2} + a e^{-i\omega\Delta t/2}\right) = X(\omega)\, 2a\cos(\omega\Delta t/2) \tag{9.13}$$

So the response function is $R(\omega) = 2a\cos(\omega\Delta t/2)$. If you wanted a unit response at zero frequency then You would choose $a = 1/2$, but because the wavelets are normalized to have unit length $a = 1/\sqrt{(2)}$, and the response function at zero frequency is $\sqrt{(2)}$. The frequency response goes from $2a\cos(0)$ to $2a\cos(\pi/2)$ while the frequency goes from zero to $\pi/\Delta t$. Just one slow transit from maximum to zero across the Nyquist interval.

For the wavelet we have

$$y(t) = a\,x(t + \frac{\Delta t}{2}) - a\,x(t + \frac{-\Delta t}{2}) \tag{9.14}$$

and the Fourier transform is

$$Y(\omega) = X(\omega)\left(ae^{i\omega\Delta t/2} - ae^{-i\omega\Delta t/2}\right) = X(\omega)\,2a\,\sin(\omega\Delta t/2) \tag{9.15}$$

So the response functions for the Haar scaling and wavelet are,

$$R_{scaling}(\omega) = 2a\cos(\omega\Delta t/2) \quad R_{wavelet}(\omega) = 2a\sin(\omega\Delta t/2) \tag{9.16}$$

From these formulas one can see that the response functions are complements of each other, so that the amplitude that is rejected by one is the amplitude that is passed by the other. This is the required characteristic of quadrature mirror filters, and will result in the preservation of power as the expansion in these wavelets continues. The Haar wavelet representation has the advantage of very good time localization, but the frequency resolution is minimal. Discrete wavelets with more weights will be able to provide better frequency resolution at the expense of less precise time localization.

9.5 The Pyramid Scheme of Discrete Wavelet Transforms.

Applying the Haar transform reduces the original N data point time series $x(t)$ into two time series of length $N/2$, which are y_1 and y_2 , as defined in (9.10). One of these contains the smoothed information and the other contains the detail information. The smoothed one could be transformed again with the Haar wavelets again, producing two time series of length $N/4$, with smoothed and detail information, and so on, keeping the details and doing an additional transform of the smoothed time series each time. If the original time series was some power of 2, $N = 2^n$, then this process, called a pyramid algorithm, would terminate when the last two time series were the coefficients of the time mean and the difference between the mean of the first half of the time series and the last half of the time series. The number of coefficients at the end would total N, and would contain all of the information in the original time series, organized according to scale and location, as defined by the Haar wavelet family. The original fine wavelet weights of (a,a) and (a,-a) on an interval of two time points are stretched, or dilated in factors of 2 to create a sequence of wavelets with increasingly large scale, culminating in the mother and father wavelets that span the entire time series.

Let's suppose we start with a time series of 8 data $x_n \;\; n = 1, 8$, and perform successive Haar transforms on this time series. The resulting Haar transformed vector, y_{jk} represents the k time steps corresponding to each of j levels of detail. The diagram below is intended to give some idea of how the original data vector would be transformed into a representation vector in Haar function amplitudes using the pyramid scheme. In this representation, the first 4 values are the amplitudes of the first level of detail, defined at 4 time locations. The next two values represent the wavelet transform of the smoothed data set, which has 4 smoothed values and results in two wavelet coefficients, y_{21} and y_{22} The last two values in the wavelet vector are the mother wavelet y_{31} and the father wavelet y_{32}. Because the Haar wavelet transform is orthogonal, the original time series can be reconstructed from the wavelet coefficient vector y_{jk}.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{bmatrix} \Rightarrow \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} \tag{9.17}$$

Let's consider the specific example of a sine wave with wavelength of 8 time steps, of which we have a total of $2^6 = 64$ data points. Figure 9.2 shows the time series and its Haar transform. The Haar transform is organized with the father and mother wavelet amplitudes on the left and the greatest level of detail in the

32 positions on the right of the transform vector. Note how the Haar coefficient is constant and large for the third level of detail, which corresponds to a period of 8 time units. It is fortuitous that the Haar wavelet of period 8 projects exactly onto the period and phase of the sine wave. If a cosine wave had been chosen, then this would not be the case and the amplitude would be spread over more Haar wavelets. The Haar wavelet, or any orthogonal wavelet, has very poor frequency resolution, as the frequency changes by a factor of two with each change in level. The coefficients for levels higher than the third level of detail are zero, since they have periods of 16, 32 and 64 time steps, and so do not project onto the wave of period 8. Similarly the mean is zero, so the father wavelet (a constant value) is zero.

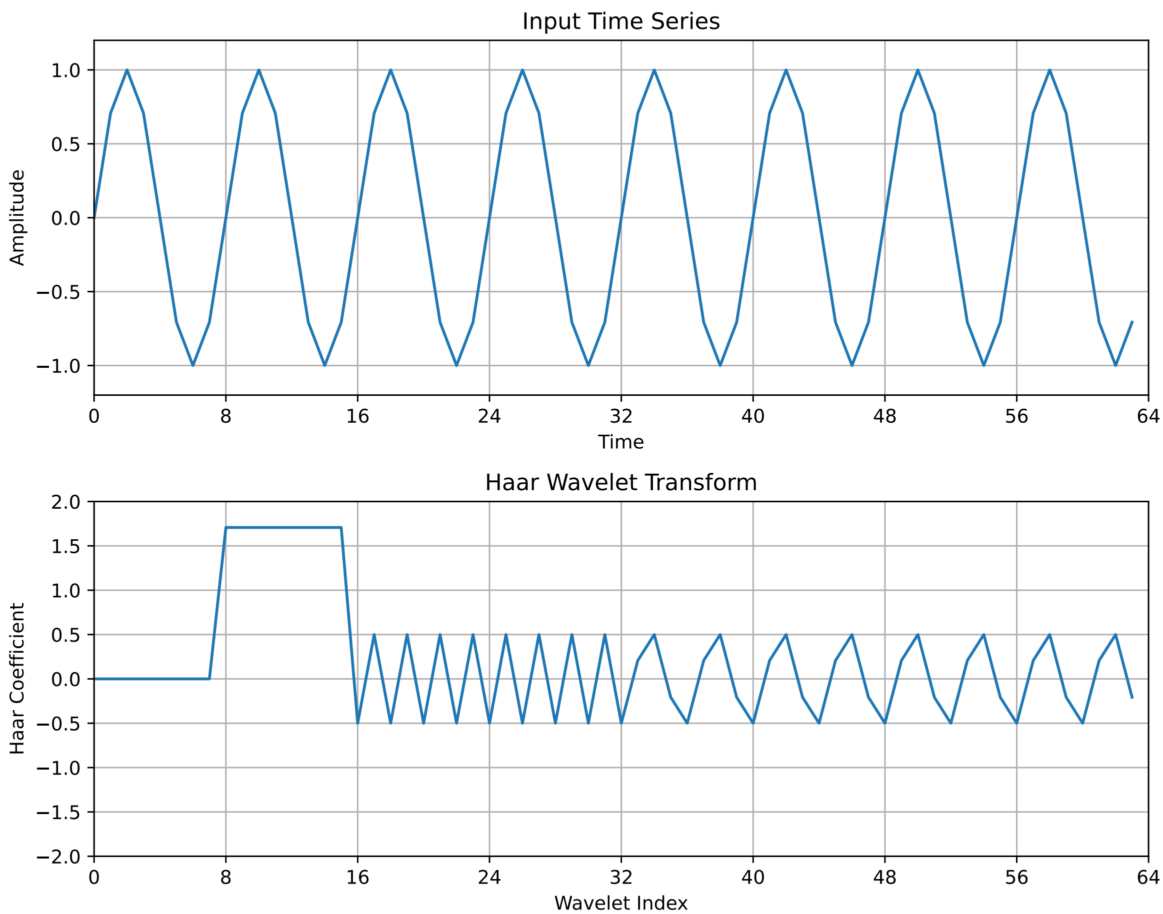

Figure 9.2 Time series of a sine wave with a period of 8 time units (top) and the Haar wavelet transform of the time series (bottom).

9.6 Daubechies Wavelet Filter Coefficients

In seeking other possible basis function sets on which we would like to expand we consider the following desirable characteristics:

(1) Good localization in both time and frequency (these conflict so we must compromise) (2) Simplicity, and ease of construction and characterization (3) Invariance under certain elementary operations such as translation (4) Smoothness, continuity and differentiability (5) Good moment properties, zero moments up to some order.

From the example of the Haar wavelet, we can see that a wavelet transform is equivalent to a filtering process with two filters that are quadrature mirror filters and divide the time series into a wavelet part, which represents the detail, and another smoothed part. Daubechies (1988, 1992) discovered an important and useful class of such filter coefficients. The simplest set has only 4 coefficients (DB2), and will serve as a useful illustration. Consider the following transformation matrix acting on a data vector to its right.

$$\begin{bmatrix} c_0 & c_1 & c_2 & c_3 & & & & & & & & \\ c_3 & -c_2 & c_1 & -c_0 & & & & & & & & \\ & & c_0 & c_1 & c_2 & c_3 & \cdots & & & & & \\ & & c_3 & -c_2 & c_1 & -c_0 & \cdots & & & & & \\ & & & & & & & \cdots & c_0 & c_1 & c_2 & c_3 \\ & & & & & & & \cdots & c_3 & -c_2 & c_1 & -c_0 \\ c_2 & c_3 & & & & & & \cdots & & & c_0 & c_1 \\ c_1 & -c_0 & & & & & & \cdots & & & c_3 & -c_2 \end{bmatrix} \tag{9.18}$$

Here we are only showing only the top two rows, the bottom two rows, and a subset of the columns. The blank spaces are occupied by zeros. The matrix is arranged in such a way that cyclic continuity of the data is assumed, much as in Fourier Analysis. Other options are possible. Dots represent where the matrix should be continued. The action of this matrix is to perform two convolutions with different, but related, filters, $[c_0, c_1, c_2, c_3]$ is the scaling filter and smooths the input if all the coefficients are positive and $[c_3, -c_2, c_1, -c_0]$ is the wavelet filter. These coefficients have been chosen such that the inner product of the smoothing and wavelet coefficients is zero, so that the two filters are orthogonal mirror filters. The pyramid algorithm can be applied, as with the Haar filter, so that successive levels of wavelet data are retained. We still have 4 unknown coefficients that we can solve for by using an approximation condition of two, and also requiring that the matrix be orthonormal. This matrix is called Daubechies-2 or DB2 because its approximation condition is 2. To ensure that it has approximation condition 2, we want to choose the coefficients of the wavelet so that their mean and trend are zero.

$$\begin{aligned} c_3 - c_2 + c_1 - c_0 &= 0 \\ 0c_3 - 1c_2 + 2c_1 - 3c_0 &= 0 \end{aligned} \tag{9.19}$$

For the transformation of the data vector to be useful, one must be able to reconstruct the original data from its smooth and detail components. This can be assured by requiring that the matrix (9.18) is orthogonal, so that its inverse is just its transpose. In discrete space, this is the equivalent of the orthogonality condition for continuous functions. The orthogonality condition places two additional constraints on the coefficients, which can be derived by multiplying (9.18) by its transpose and requiring that the product be the unit matrix. This yields two additional conditions on the coefficients, so that we now know them uniquely.

$$\begin{aligned} c_0^2 + c_1^2 + c_2^2 + c_3^2 &= 1 \\ c_3c_1 + c_2c_0 &= 0 \end{aligned} \tag{9.20}$$

These four equations for the coefficients (9.19 and 9.20) have a unique solution up to a left-right reversal. DB2 is only the simplest of a family of wavelet sets with the number of coefficients increasing by two each time (2, 4, 6, 8, 12, . . .). Note that the Haar wavelet would be DB1 in this family of wavelets. Each time we add two more coefficients we add an additional orthogonality constraint and raise the number of zero moments, or the approximation condition order, by one. Daubechies (1988) has tabulated the coefficients for lots of these, and they are available in most wavelet software packages.

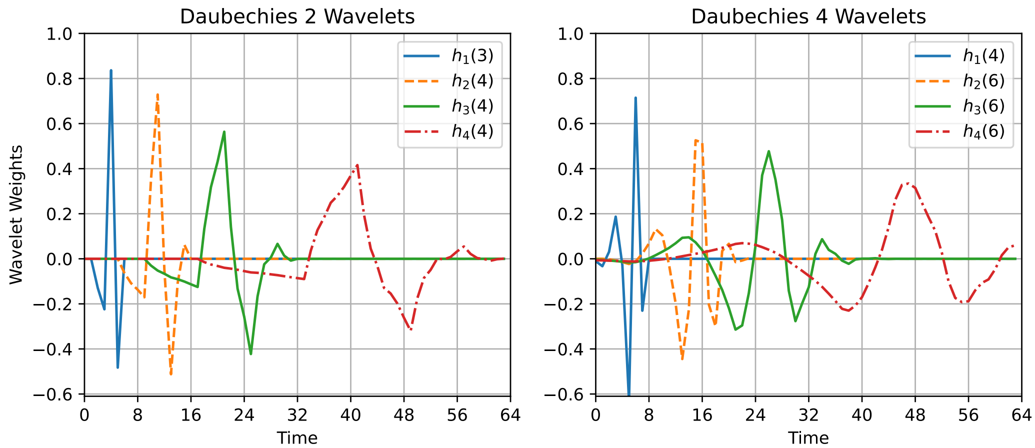

Figure 9.3 Selected examples of the Daubechies-2 and Daubechies-4 wavelets. The subscript indicates the level of approximation and the number in parentheses is the position in time. For example, only the third of the 32 level 1 Daubechies-2 wavelets is shown.

Figure 9.3 shows examples of some of the Daubechies-2 and Daubechies-4 wavelets. Note that as the number of weights is increased, the wavelets become smoother. Edge effects become increasingly important as the number of weights is increased, since the span of the longer wavelets becomes great.

9.7 Continuous, Non-orthogonal Wavelets

The frequency resolution with orthogonal wavelets is constrained to be coarse, so we may wish to use non-orthogonal wavelets in which we vary the wavelength and position of the wavelet more continuously. Some relatively famous wavelets are the Mexican Hat,

$$\psi_\sigma(t) = \frac{2}{\sqrt{3\sigma}\,\pi^{1/4}}\left(1 - \left(\frac{t}{\sigma}\right)^2\right) e^{-\frac{t^2}{2\sigma^2}} \tag{9.21}$$

and the Complex Morlet Wavelet,

$$\begin{aligned} \psi_\sigma(t) &= c_\sigma \pi^{-1/4} e^{-\frac{1}{2}t^2}\left(e^{i\sigma t} - e^{-\frac{1}{2}\sigma^2}\right) \\ c_\sigma &= \left(1 + e^{-\sigma^2} - 2e^{-\frac{3}{4}\sigma^2}\right) \end{aligned} \tag{9.22}$$

Their structures are shown in Figure 9.4. The effect of the scale parameter σ on the Mexican Hat wavelet is shown in th left panel. Since the imaginary part of the Morelet wavelet is phase-shifted relative to its center location, in visual representations of data, only the real part is shown. the scale and location of the wavelet is varied to provide a representation in time-frequency space, as shown in the example below.

Figure 9.5 illustrates the use of a Morlet wavelet representation in frequency and time for the time series of Benthic δ18O constructed by Lisiecki and Raymo (2005). Ocean water gets heavier in $\delta^{18}O$ as the lighter isotope is preferentially stored in ice sheets during ice ages. So the increase shows the growing global ice volume. As the ice volume gets larger it oscillates in time between glacial maxima with high $\delta^{18}O$ and interglacials with lower values. From the time-frequency separation in the plot, one can see that about 2.5 million years ago an oscillation with a period of about 40,000 years (2.5 cycles per 100kyr) begins to

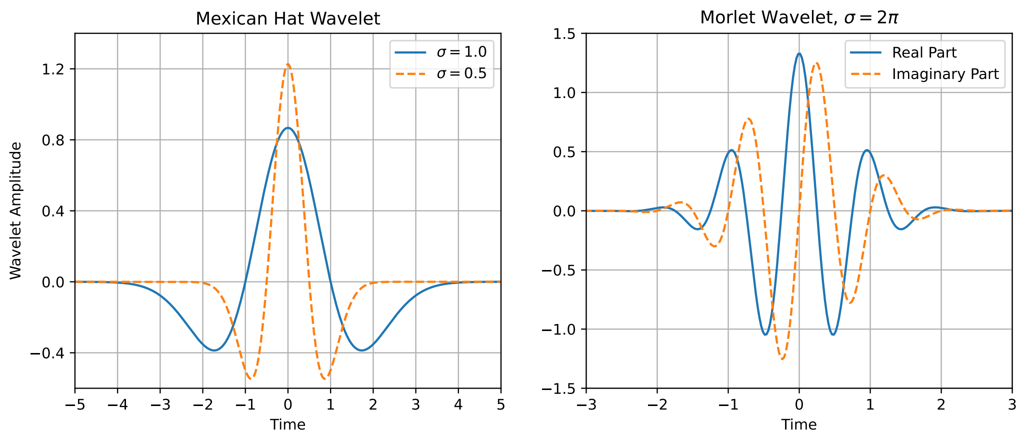

Figure 9.4 Examples of the Mexican Hat and Morlet continuous wavelets.

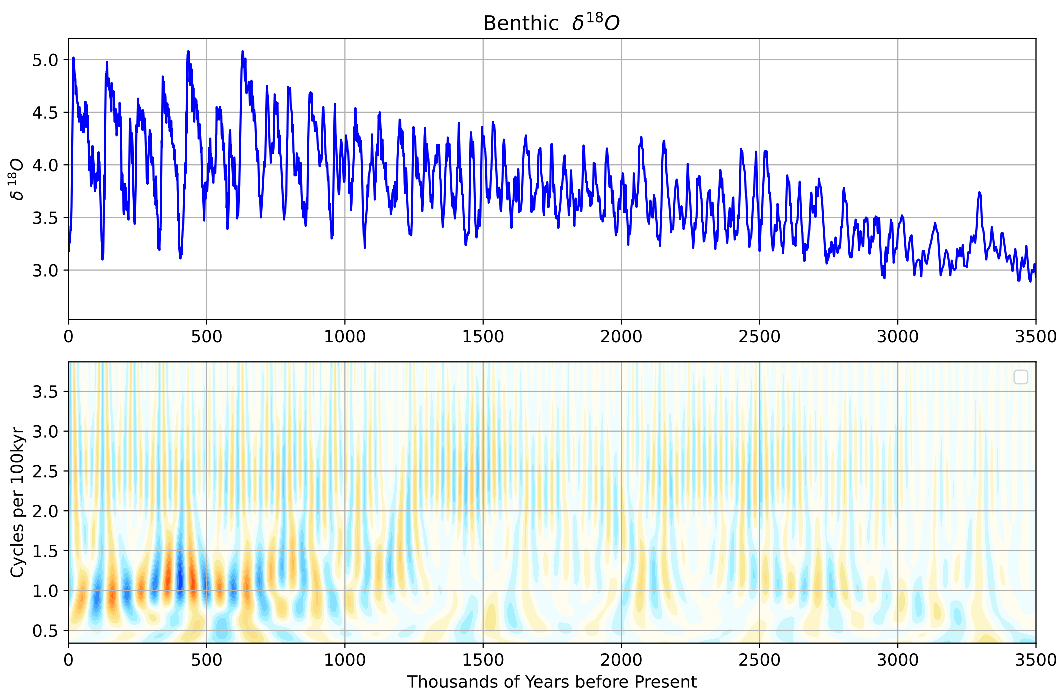

Figure 9.5 Time series of $\delta^{18}O$ from ocean sediment cores for the past 3.5 million years (top) and the Morlet wavelet transform plotted as a function of frequency and time.

occur intermittently. Later at about 1 million years ago a strong oscillation with a period of about 100,000 years begins and continues until the present. Analysis with wavelets reveals the episodic nature of these oscillations is interesting and would not be revealed by power spectral analysis, which discards all time location information in favor of maximal frequency resolution.

Dennis: *A bit more work to do here? What is missing that woud help?*