Wavelet Transforms in Time Series Analysis

Andrew Tangborn
Global Modeling and Assimilation Office, Goddard Space Flight Center
Andrew.V.Tangborn@nasa.gov
301-614-6178

Outline

1. Fourier Transforms.

2. What is a Wavelet?

3. Continuous and Discrete Wavelet Transforms

4. Construction of Wavelets through dilation equations.

5. Example - Haar wavelets

6. Daubechies Compactly Supported wavelets.

7. Data compression, efficient representation.

8. Soft Thresholding.

9. Continuous Transform - Morlet Wavelet

10. Applications to approximating error correlations

Fourier Transforms

- A good way to understand how wavelets work and why they are useful is by comparing them with Fourier Transforms.

- The Fourier Transform converts a time series into the frequency domain:

Continuous Transform of a function f(x):

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$$

where $\hat{f}(\omega)$ represents the *strength* of the function at frequency ω, where ω is continuous.

Discrete Transform of a function f(x):

$$\hat{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-ikx} dx$$

where k is a discrete number.

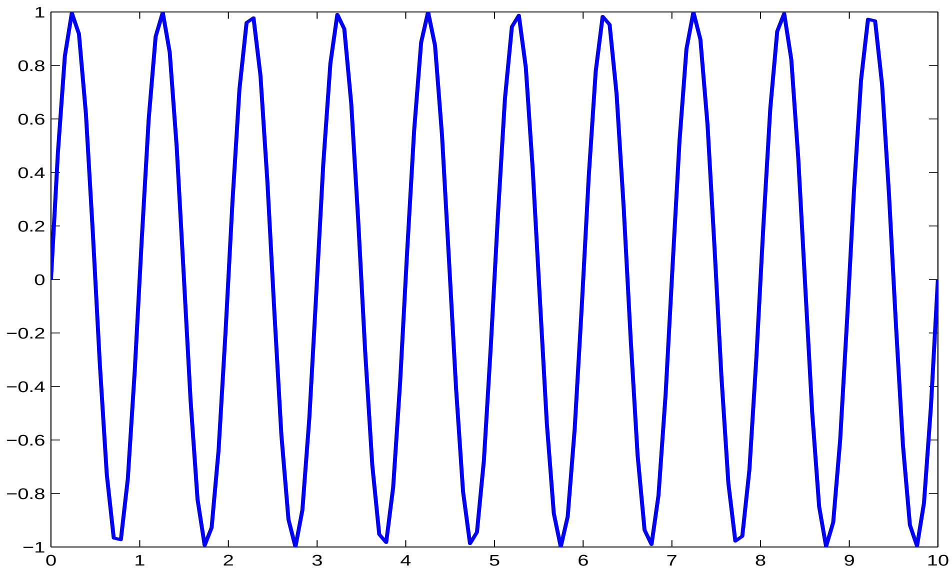
for discrete data $f(x_j)$, $j = 1, ..., N$

$$\hat{f}_k = \sum_{j=1}^{N} f_j e^{(-i2\pi(k-1)(j-1)/N)}$$

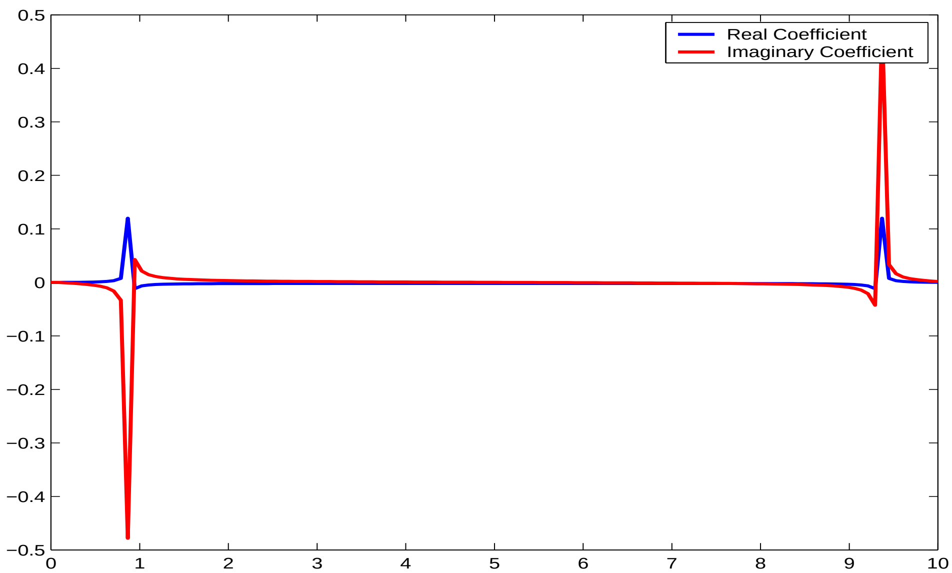
- The Fast Fourier Transform (FFT) is $o(NlogN)$ operations.

Example: Single Frequency Signal

$$f(t) = \sin(2\pi t)$$



Signal

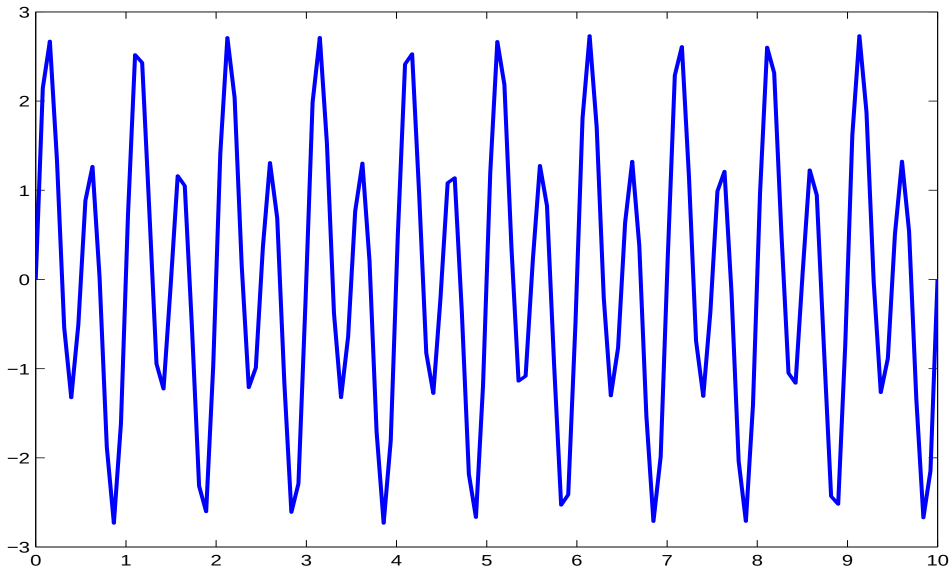


Fourier Transform

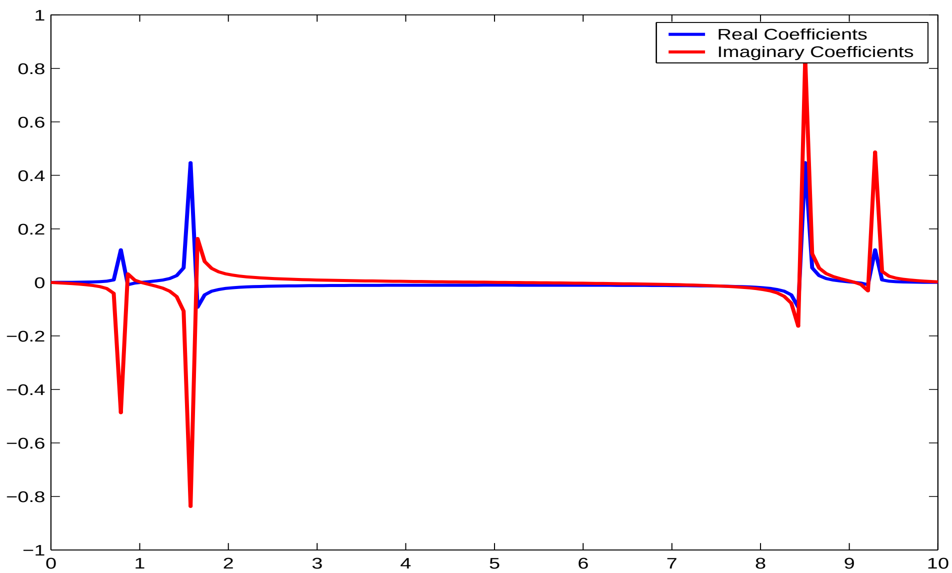
- Discrete Fourier Transform (DFT) locates the single frequency and reflection.
- Complex expansion is not exactly sine series - causes some spread.

Two Frequency Signal

$$f(t) = \sin(2\pi t) + 2\sin(4\pi t)$$



Signal

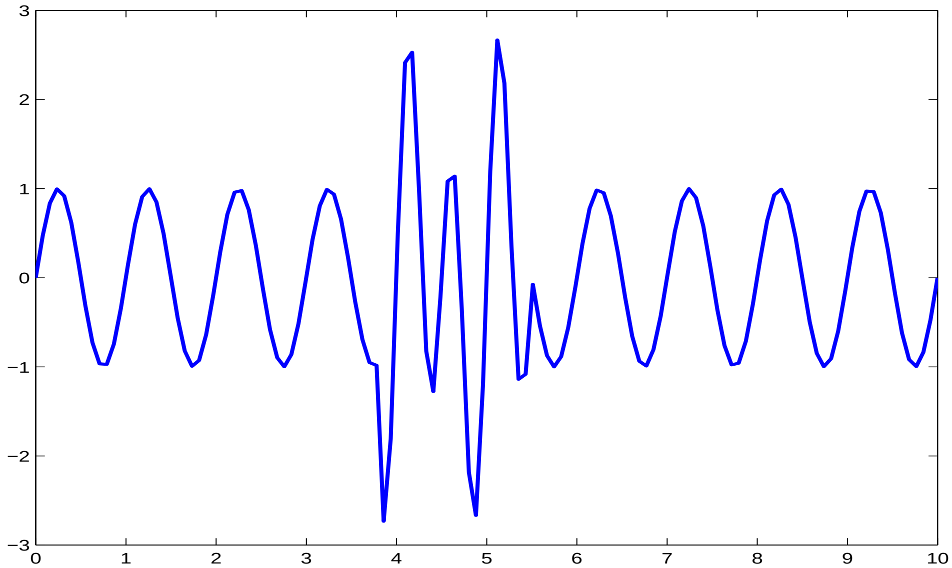


Fourier Transform

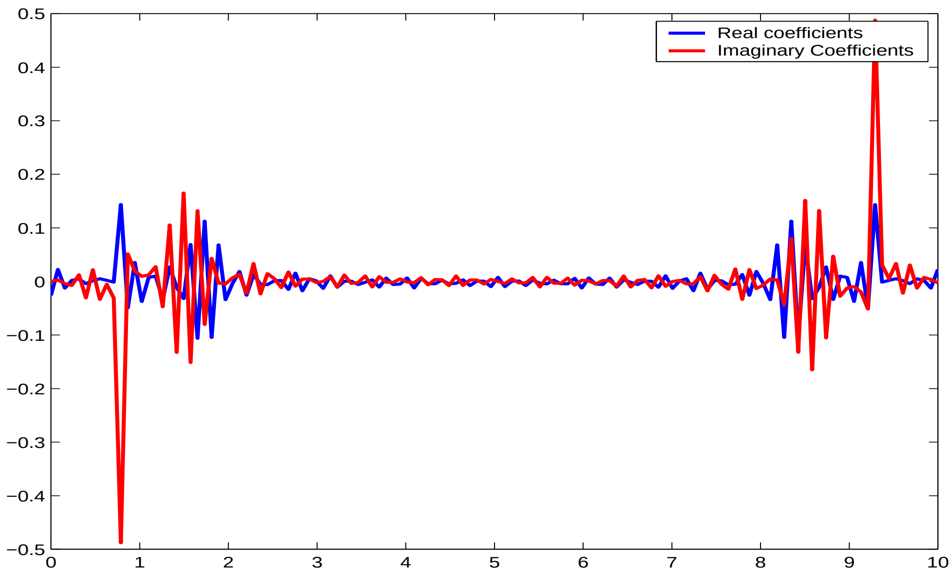
- Both signal frequencies are represented by the Fourier Coefficients.

Intermittent Signal

$f(t) = \sin(2\pi t) + 2\sin(4\pi t)$ when $.37 < t < .55$
$f(t) = \sin(2\pi t)$ otherwise



Signal



Fourier Transform

• First frequency is found, but higher intermittent frequency appears as many frequencies, and is not clearly identified.

What is a Wavelet?

• A function that is localized in time and frequency, generally with a zero mean.

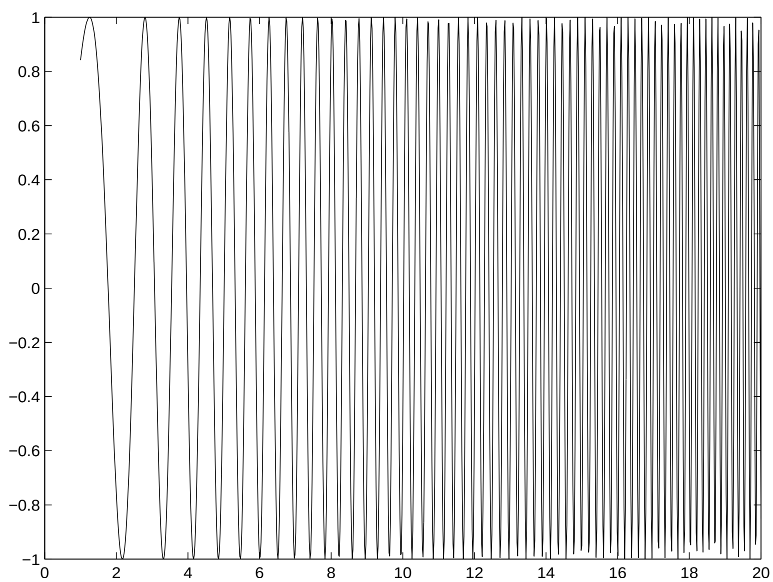
• It is also a tool for decomposing a signal by location *and* frequency.

Consider the Fourier transform:

A signal is only decomposed into its frequency components.

No information is extracted about location and time.

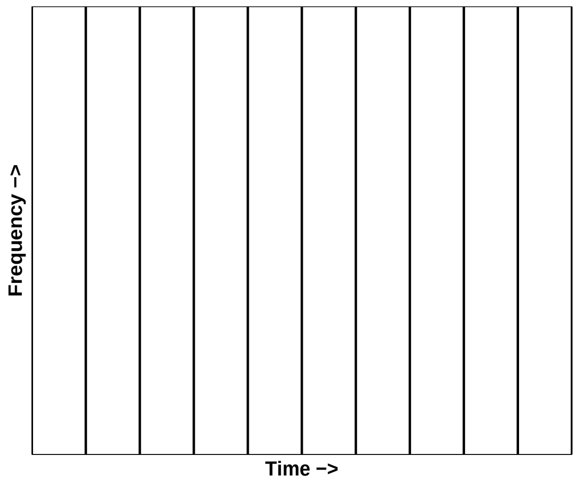
• What happens when applying a Fourier transform to a signal that has a time varying frequency?



The Fourier transform will only give some information on which frequencies are present, but will give no information on when they occur.

Schematic Representation of Decomposition

Original Signal

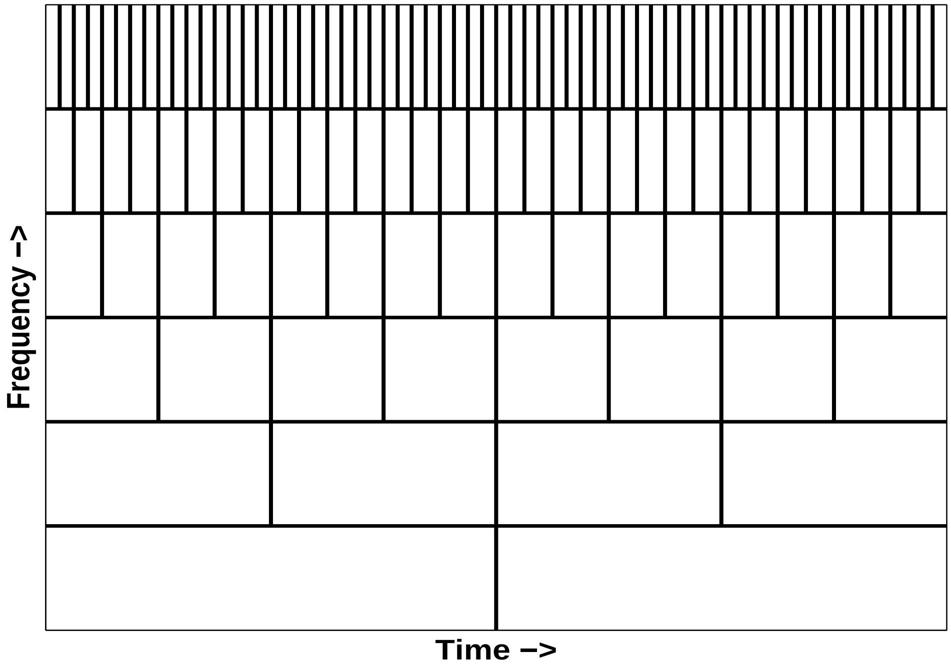


- The signal is represented by an amplitude that is changing in time.
- There is no explicit information on frequency.

Fourier Transform

- The Fourier transform results in a representation that depends only on frequency.
- It gives no information on time.

Wavelet Transform



• The wavelet transform contains information on both the time location and frequency of a signal.

Some typical (but not required) properties of wavelets

• Orthogonality - Both wavelet transform matrix and wavelet functions can be orthogonal.

Useful for creating basis functions for computation.

• Zero Mean (admissibility condition) - Forces wavelet functions to *wiggle* (oscillate between positive and negative).

• Compact Support - Efficient at representing localized data and functions.

How are Wavelets Defined?

- **Families of basis functions that are based on**

  (1) dilations:
  $$\psi(x) \to \psi(2x)$$
  (2) translations:
  $$\psi(x) \to \psi(x+1)$$
  of a given general "mother wavelet" $\psi(x)$.

- **How do we use $\psi(x)$?**

  The general form is:
  $$\psi_k^j(x) = 2^{j/2}\psi(2^j x - k)$$
  where

  j: dilation index

  k: translation index

  $2^{k/2}$ needed for normalization

- **How do we get $\psi(x)$?**

  Dilation Equations

Construction of Wavelets

- **We consider here only orthogonally/compactly supported wavelets**

  - Orthogonality means:

$$\int_{-\infty}^{\infty} \psi_k^j(x)\psi_{k'}^{j'}(x)dx = \delta_{kk'}\delta_{jj'}$$

- **Wavelets are constructed from scaling functions,** $\phi(x)$ **:** $\phi(x)$ come from the dilation equation:

$$\phi(x) = \sum_k c_k \phi(2x - k)$$

  c_k: **Finite** set of filter coefficients

- **General features:**

  - Fewer non-zero c_k's mean **more** compact and **less** smooth functions
  - More non-zero c_k's mean **less** compact and **more** smooth functions

Restrictions on the Filter Coefficients

- **Normality:**

$$\int_{-\infty}^{\infty} \phi(x)dx = 1$$

$$\int_{-\infty}^{\infty} \sum_k c_k \phi(2x - k)dx = 1$$

$$\sum_k c_k \int_{-\infty}^{\infty} \phi(2x - k)dx = 1$$

$$\sum_k c_k \int_{-\infty}^{\infty} \frac{1}{2}\phi(2x - k)d(2x - k) = 1$$

$$\sum_k c_k \int_{-\infty}^{\infty} \frac{1}{2}\phi(\xi)d(\xi) = 1$$

$$\sum_k c_k (\frac{1}{2})(1) = 1$$

or

$$\sum_k c_k = 2$$

- **Simple Examples**
    - Smallest number of c_k is 1: just get $\phi(x) = \delta$, $\rightarrow$ zero support.
    - Haar Scaling Function: $c_0 = 1$, $c_1 = 1$.

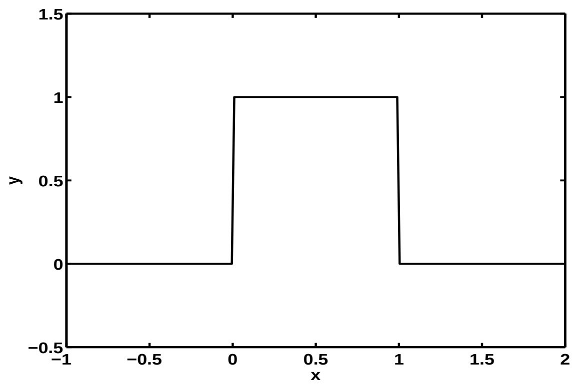
Haar Scaling Function

The scaling function equation is:

$$\phi(x) = \phi(2x) + \phi(2x - 1)$$

The only function that satisfies this is:

$\phi(x) = 1$ if $0 \leq x \leq 1$
$\phi(x) = 0$ otherwise

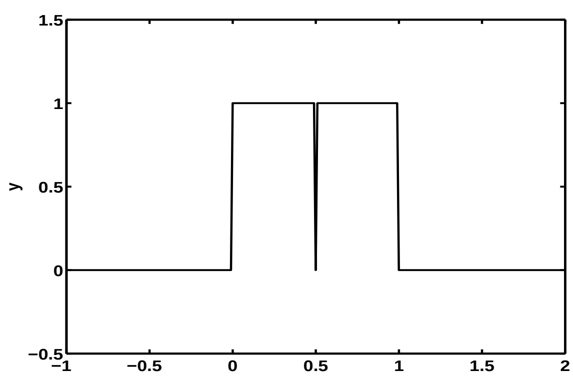


- **Translation and Dilation of** $\phi(x)$:

  $\phi(2x) = 1$ if $0 \leq x \leq 1/2$
  $\phi(2x) = 0$ elsewhere

  $\phi(2x - 1) = 1$ if $1/2 \leq x \leq 1$
  $\phi(2x - 1) = 0$ elsewhere

  so the sum of the two functions is then:



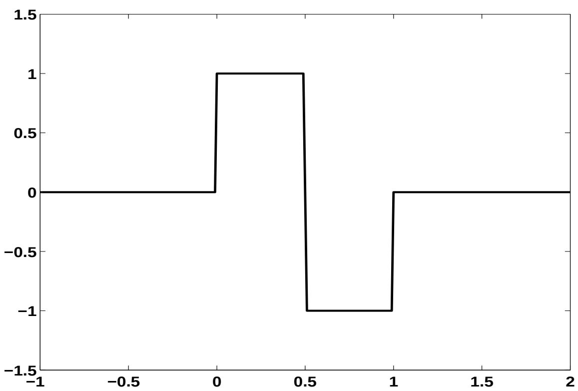
Haar Wavelet

• Wavelets are constructed by taking differences of scaling functions

$$\psi(x) = \sum_k (-1)^k c_{1-k} \phi(2x - k)$$

differencing is caused by the $(-1)^k$:

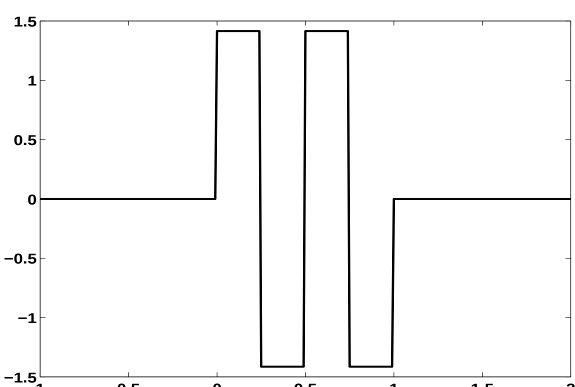
so the basic Haar wavelet is:



and the family comes from dilating and translating:

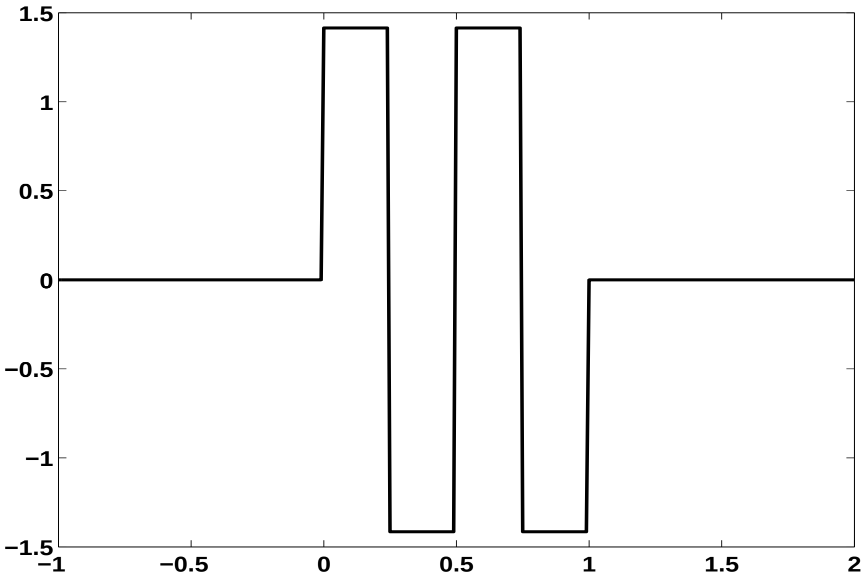
$$\psi_k^j(x) = 2^{j/2} \psi(2^j x - k)$$

so that the $j = 1$ wavelets are:

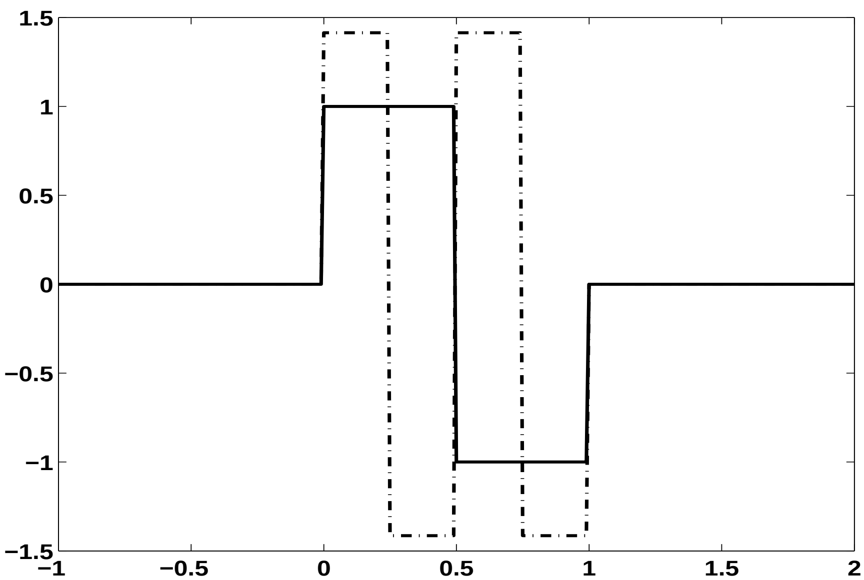


Orthogonality of Haar wavelets

- **Translation ⇒ no overlap**



- **Dilation ⇒ cancellation**

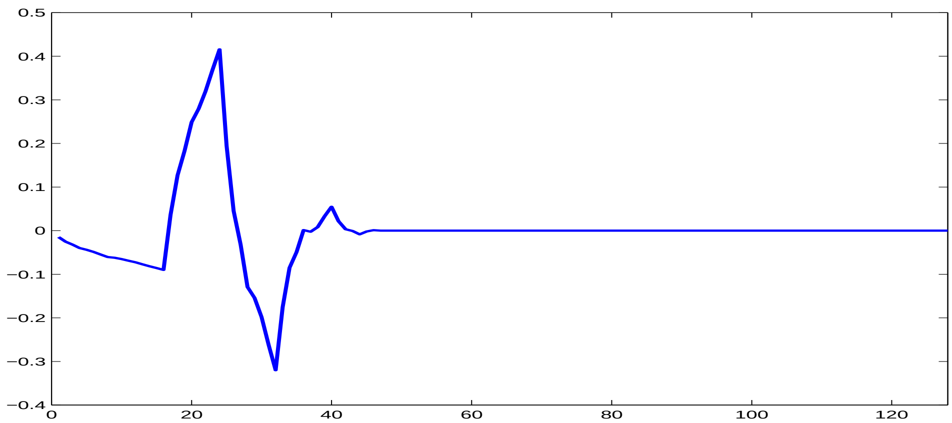


Daubechies' Compactly Supported Wavelets

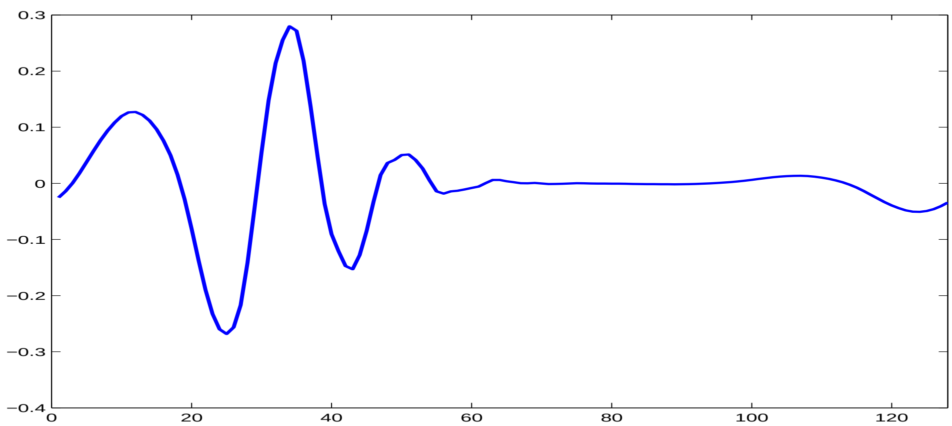
The following plots are wavelets created using larger numbers of filter coefficients, but having all the properties of orthogonal wavelets. For example, D_4 has coefficients:

$$c_k = \frac{1}{4}(1+\sqrt{3}), \frac{1}{4}(3+\sqrt{3}), \frac{1}{4}(3-\sqrt{3}), \frac{1}{4}(1-\sqrt{3})$$

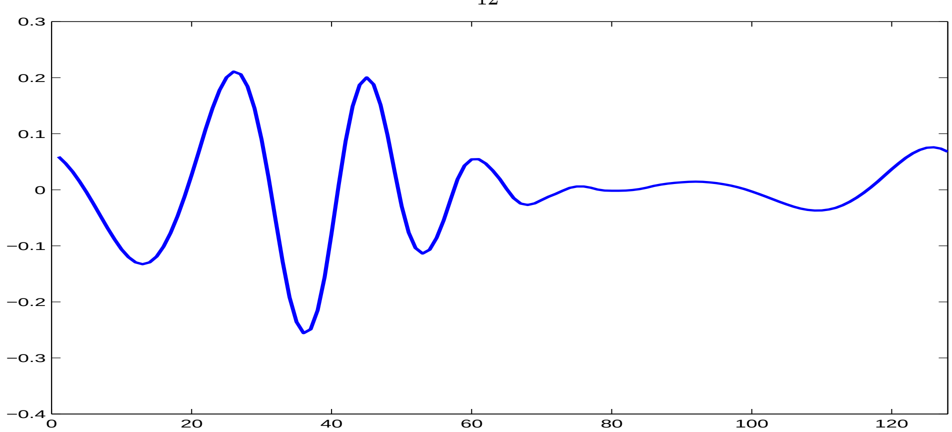
(Reference: Daubechies, 1988)



D_4



D_{12}



D_{20}

Continuous and Discrete Wavelet Transforms

Continuous

$$(T^{wav}f)(a,b) = |a|^{-1/2} \int dt f(t) \psi(\frac{t-1}{b})$$

where

a: translation parameter

b: dilation parameter

Discrete

$$T^{wav}_{m,n}(f) = a_o^{-m/2} \int dt f(t) \psi(a_o^{-m} t - n b_o)$$

m: dilation parameter

n: translation parameter

a_o,b_o depend on the wavelet used

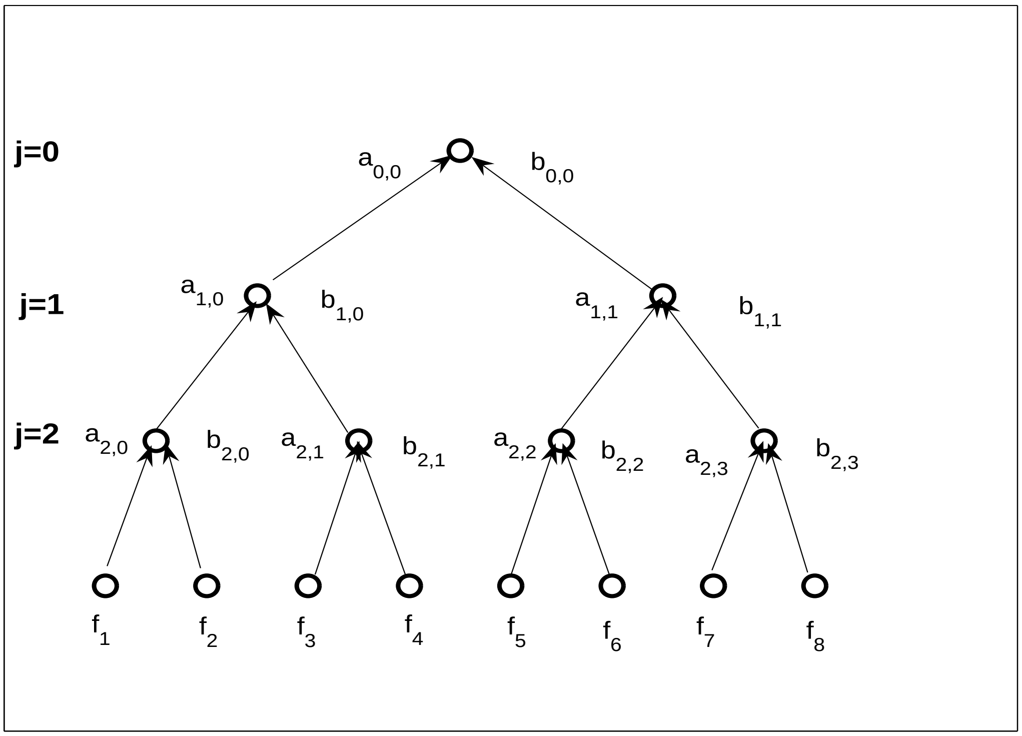
Fast Wavelet Transform

(Reference: S. Mallat, 1989)

- **Uses the discrete data:**

$$\left[\begin{array}{cccccccc} f_0 & f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & f_7 \end{array}\right]$$

- **Pyramid Algorithm $\Rightarrow$ o(N) !!** **- Start at finest scale and calculate differences**



and averages

- Use Averages at next coarser scale to get new set of differences ($b_{j,k}$) and averages ($a_{j,k}$)

And the coefficients are simply the differences ($b_{j,k}$) and the average for the coarsest scale ($a_{0,0}$):

$$\left[\begin{array}{cccccccc} a_{0,0} & b_{0,0} & b_{1,0} & b_{1,1} & b_{2,0} & b_{2,1} & b_{2,2} & b_{2,3} \end{array}\right]$$

For Haar Wavelet:

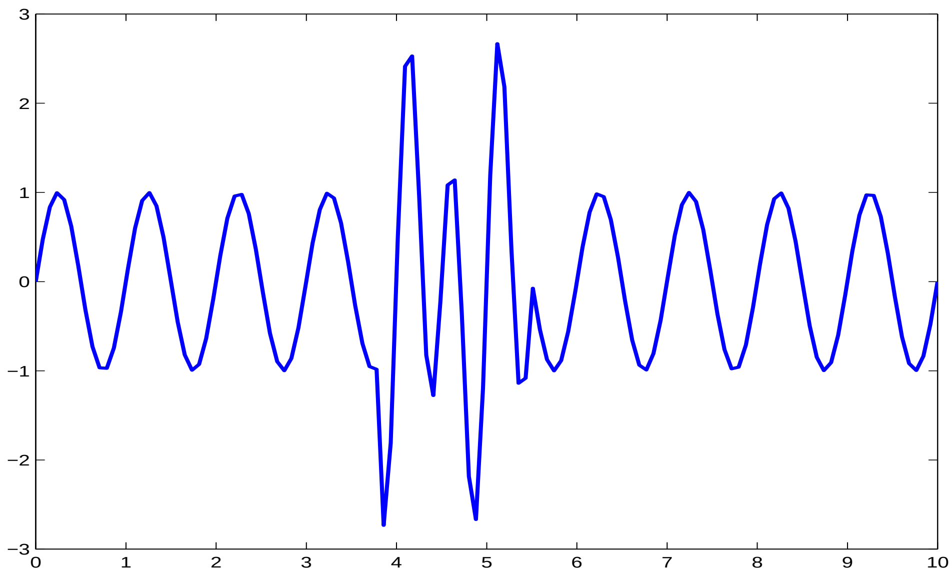
$$a_{j-1,k} = c_0 a_{j,k} + c_1 a_{j,k+1}$$

$$b_{j-1,k} = c_0 a_{j,k} - c_1 a_{j,k+1}$$

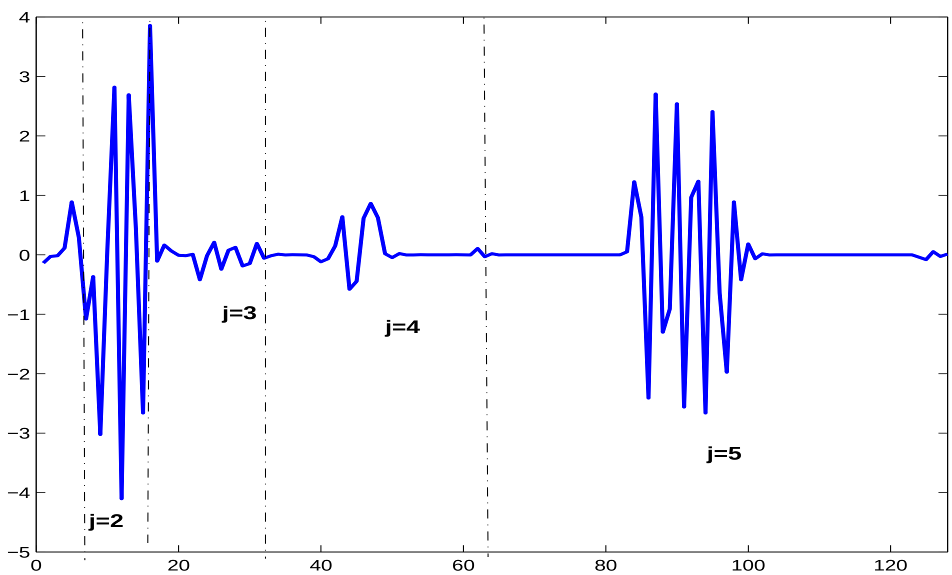
- The differences are the coefficients at each scale. Averages used for next scale.

- **How do we store all this? (eg, Numerical Recipes algorithm).**

Intermittent Signal - Wavelet Transform

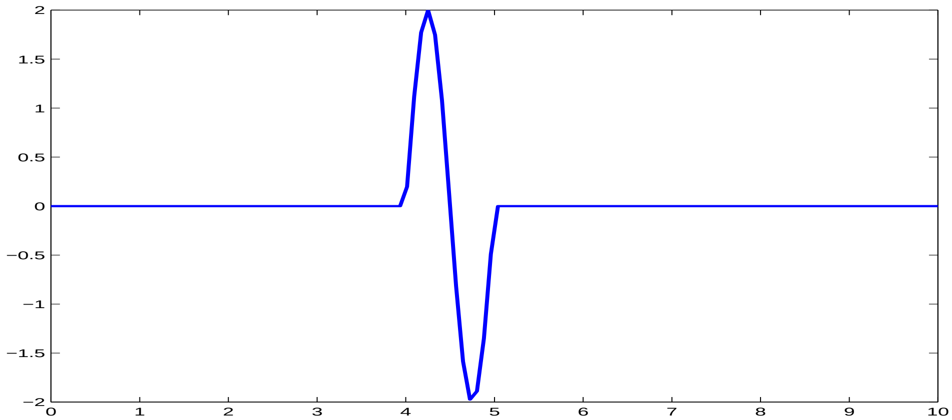


Signal

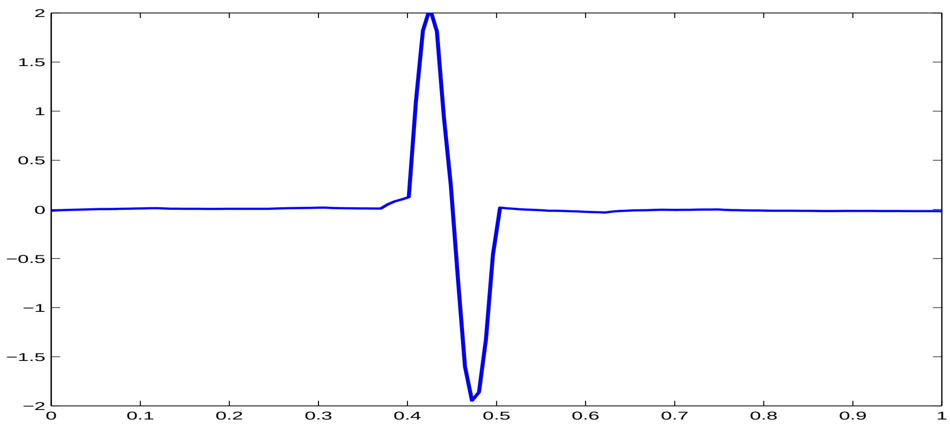


• **Higher Frequency intermittent signal shows up in the $j = 5, 6$ scales, and only in the middle portion of the domain. Localized frequency data is found.**

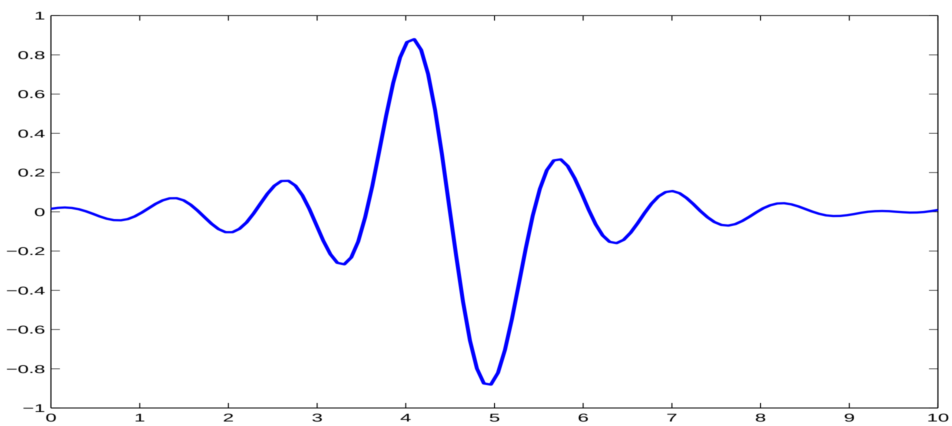
Data Compression - Efficient Representation



Original Localized Data, 128 points



16 term Wavelet Reconstruction



16 term Fourier Reconstruction

• Localized function is represented with more accuracy using 16 wavelet coefficients because only significant coefficients need to be retained. Fourier coefficients are global.

Denoising by Soft Thresholding

Basic Idea:

- **Wavelet Representation highly compresses coherent data into just a few coefficients.**

  Magnitude of each coefficient is relatively large.

- **White noise contains energy at all time scales and time locations.**

  Representation is spread to many (if not all) wavelet coefficients.

  Noise coefficients are relatively small in amplitude.

- **How do we remove the noise contribution in the coefficients?**

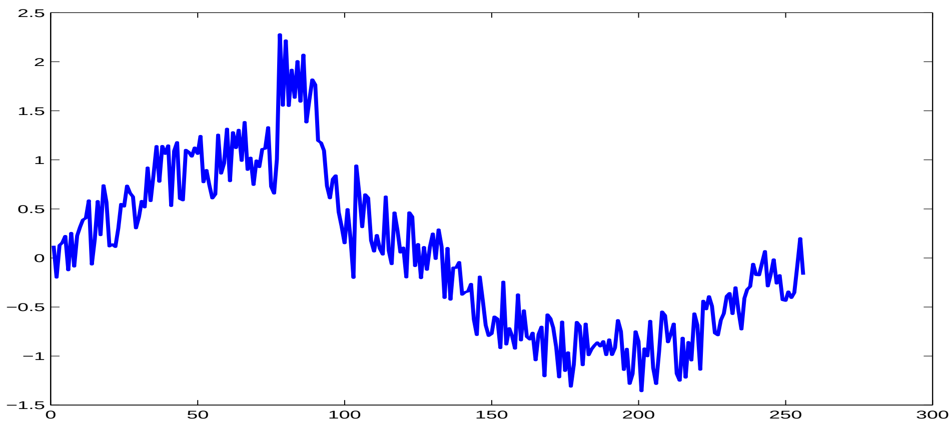
  Ans: Shrink all of them just a little.

- **How much do we shrink each coefficient?**

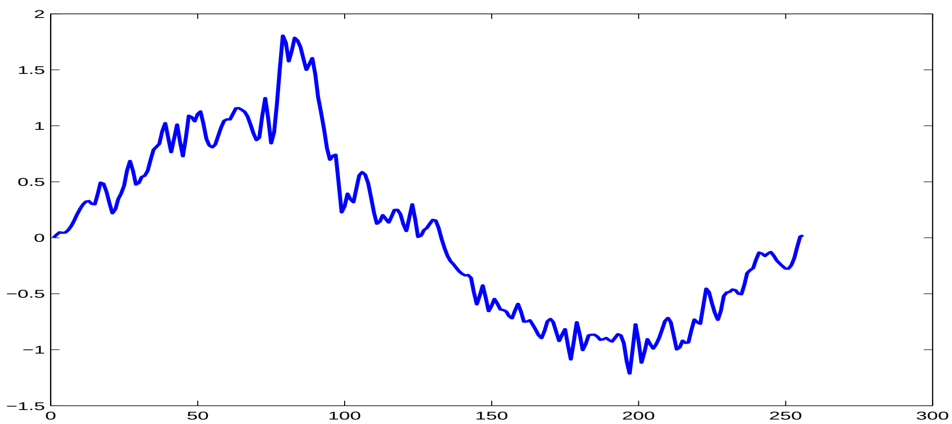
  $t = \sqrt{2log(n)}\sigma/\sqrt{n}$

  where n= number of data points and σ=noise standard deviation.

Example



Noisy Signal with white noise of known variance.



Wavelet coefficients Shrunk toward zero.

• Soft Thresholding takes advantage of the fact that white noise is represented equally by all coefficients.

• Data compression means that coherent signal is represented by just a few coefficients with relatively large values. White noise (or non-white as well) is spread over many coefficients, adding just a small amount to the magnitude of each.

The Continuous Morlet Wavelet Transform

The Morlet mother wavelet is a complex exponential (Fourier) with a Gaussian envelop which ensures localization:

$$\psi(t) = exp(i\omega_0 t)exp(-t^2/2\sigma^2)$$

where ω_0 is the frequency and σ is a measure of the spread or support.

Note that while the footprint is infinite, the exponential decay creates an effective footprint which is relatively compact.

Translations and dilations of the Morlet wavelet:

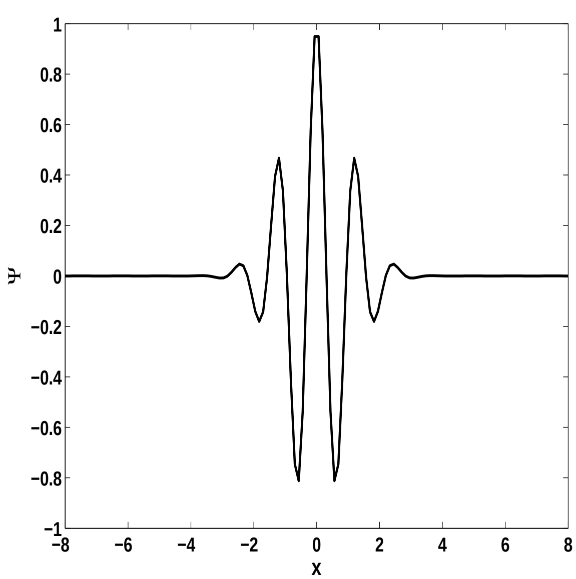
$$\psi_(b,a)(t) = \frac{1}{a}exp\left[i\omega_0\left(\frac{t-b}{a}\right)\right]exp\left[-\left(\frac{t-b}{a}\right)^2/2\sigma^2\right]$$

The Morlet wavelet using matlab

In matlab, the Morlet mother wavelet can be constructed using the command:

```
[psi,x] =  Morlet(-8,8,128);
```

on 128 grid points, and domain of [-8,8].



The Morlet transform with Matlab

Given a time series:

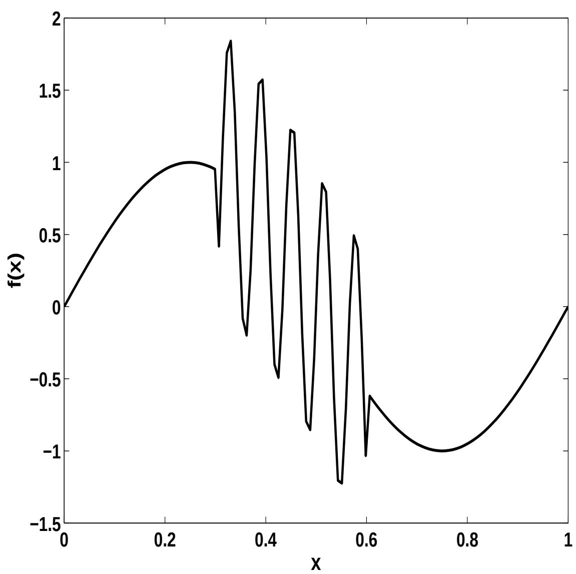
$$y(t) = \begin{cases} sin(2\pi t) + sin(32\pi t) & \textbf{when } .3 \leq t \leq .6 \\ sin(2\pi t) & \textbf{otherwise} \end{cases}$$

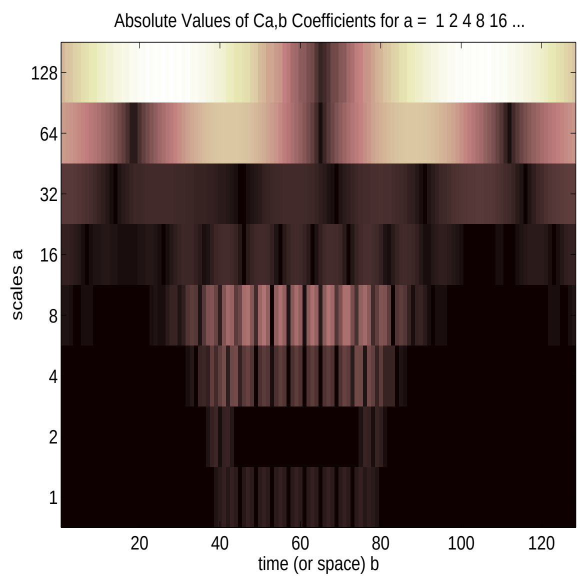
on 128 grid points, the Morlet transform can be calculated and plotted using the command:

```
coef=cwt(y,[1 2 4 8 16 32 64 128],'morl','plot')
```

- the top level of the coefficient plot shows bright (or large) values for scale 128 (largest scale).

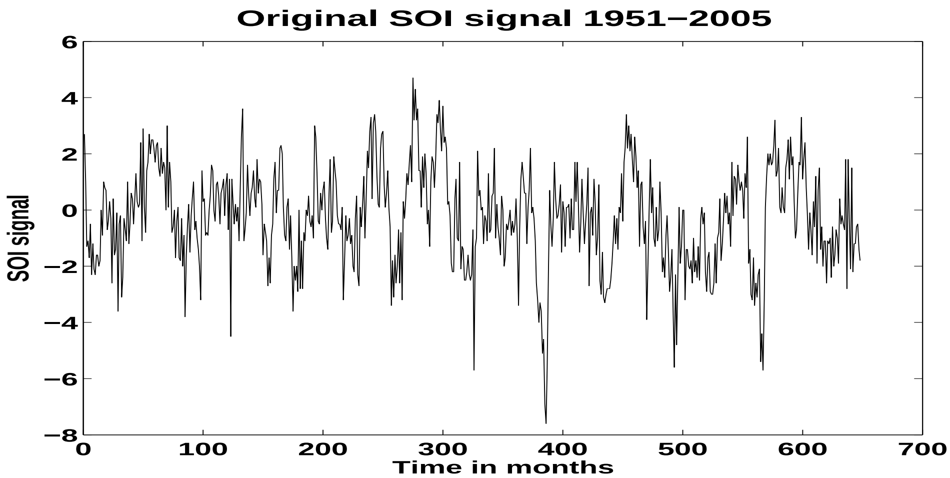
- The next two layers are not zero, indicating *leakage* between different time scales.

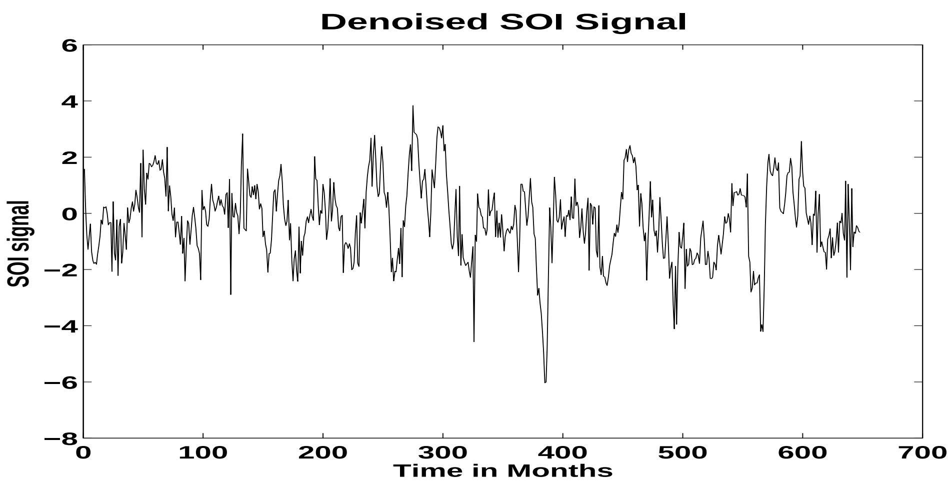




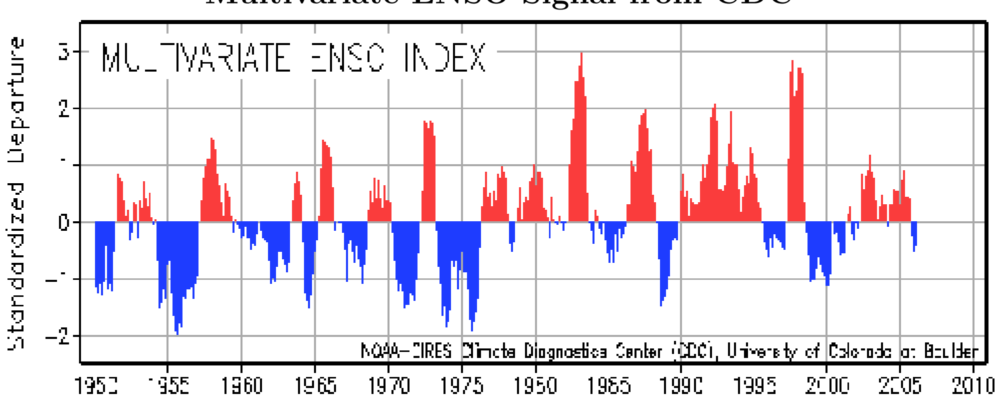
Example: Southern Oscillation Index Time Series 1951-2005

The SOI is the monthly pressure fluctuations in air pressure between Tahiti and Darwin. It is generally a noisy time series and benefits from some smoothing. Soft thresholding is a way to remove the noise in the signal without removing important information. The multivariate ENSO signal comes from sea level pressure, surface wind, sea surface tempemperature, surface air temperature and total amount of cloudiness.

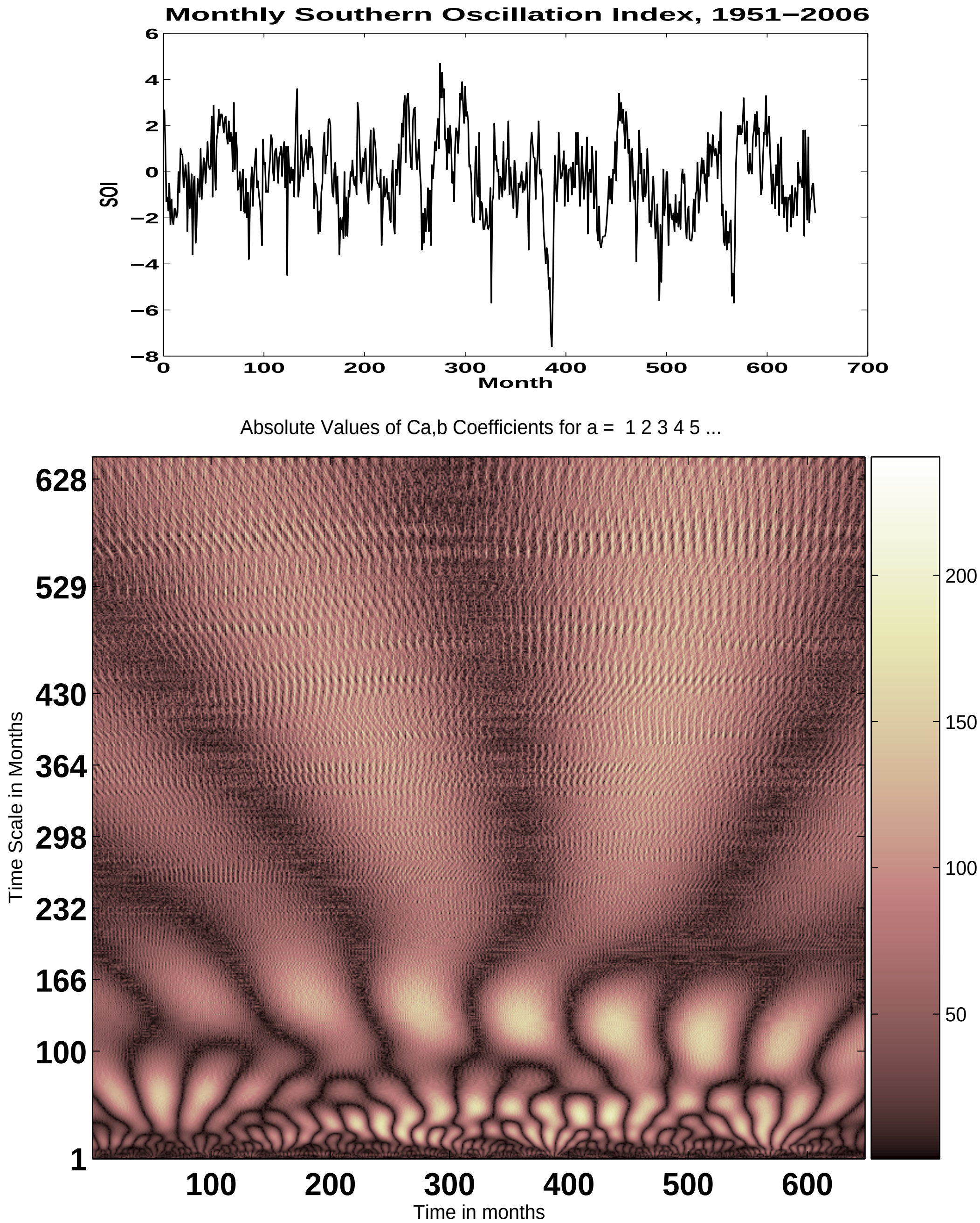




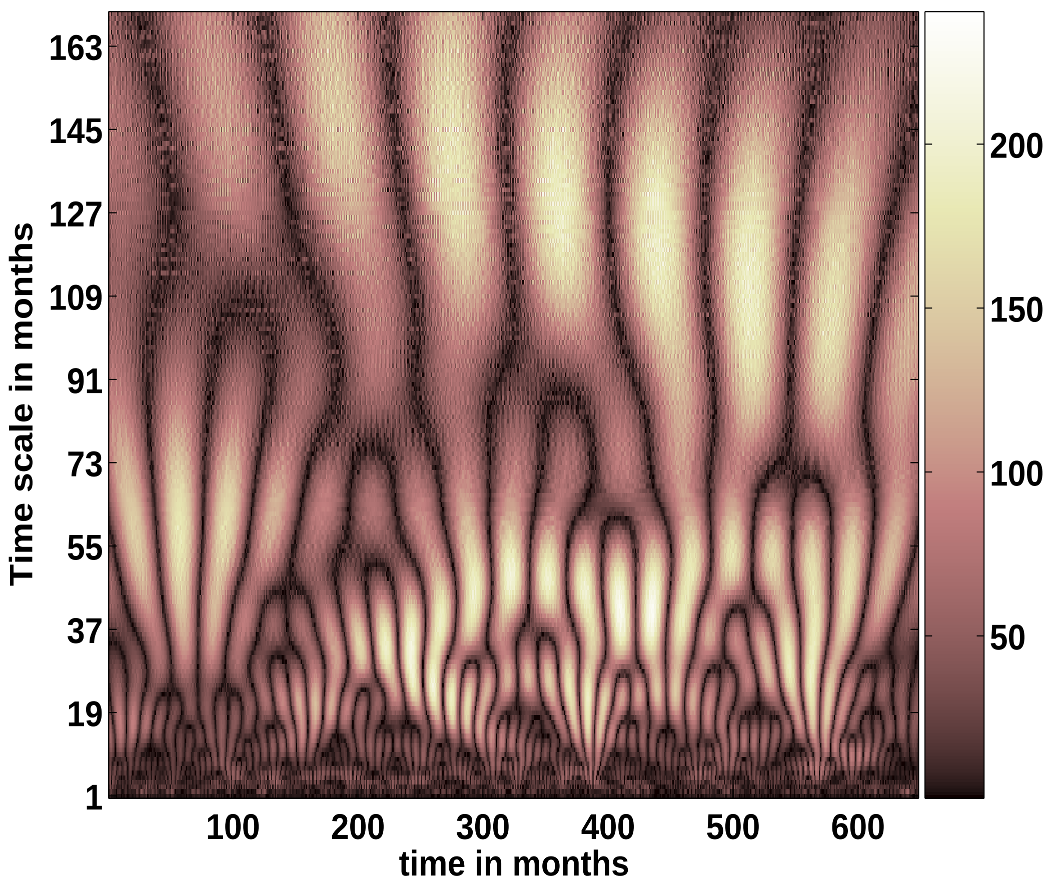
Multivariate ENSO Signal from CDC



Morlet Wavelet Decomposition of SOI time series

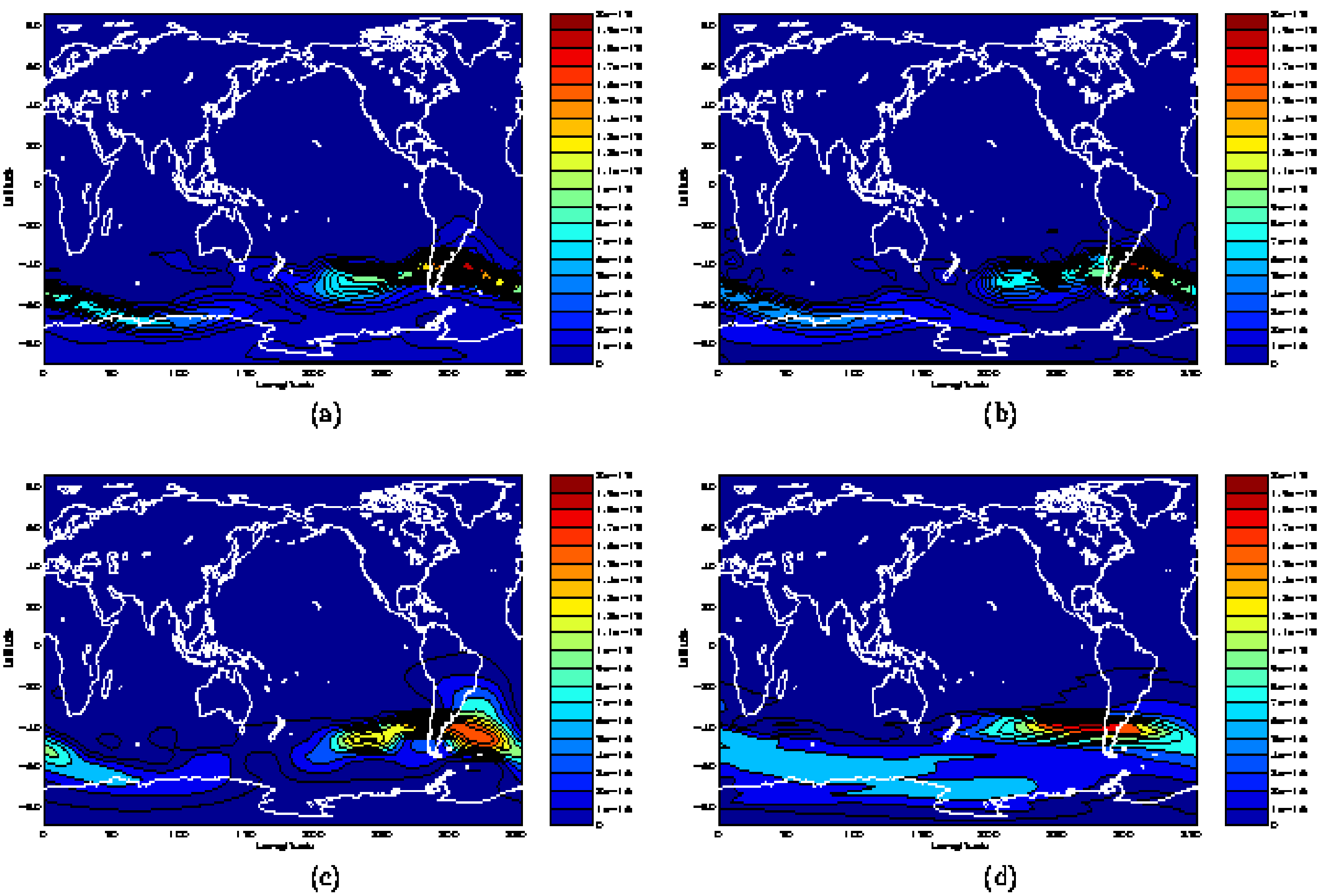


Morlet Coefficients for SOI



Application to Approximation of Error Correlations

- Error correlations are essential for carrying out atmospheric data assimilation.
- They tell us how errors in model outputs are spatially related, and how to spread information from observational data into the model.
- Error correlations are generally the most computationally and memory intensive parts of a data assimilation system.
- Below is the error correlation of a chemical constituent assimilation system (a), and wavelet approximations created by retaining 10 % (b), 10% (c) and 2% (d) of the wavelet coeffients.



(a) (b) (c) (d)

Extracting time scales from climate signals:

Reference: Seasonal-to-Interannual variability of Ethiopia/Horn of Africa Monsoon. Part 1: Associations of Wavelet Filtered large-Scale Atmospheric Circulation and Global Sea Surface Temperature, Segelet et al. J. of Climate, 2009.

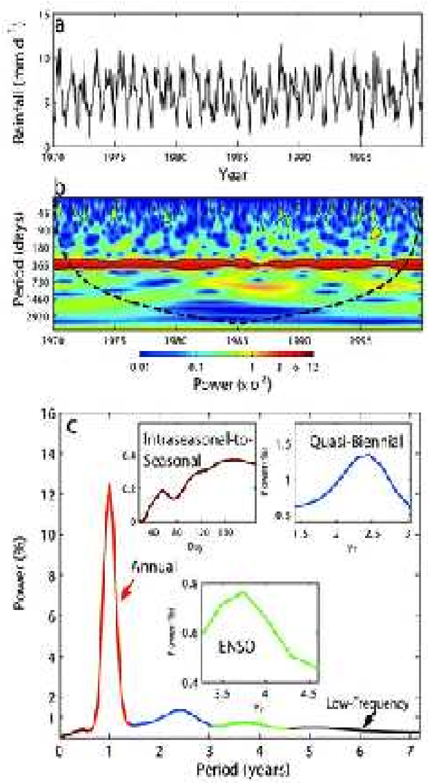


FIG. 2. Spectral characteristics of Ethiopian June–September 1970–99 rainfall averaged across stations in Fig. 1 (dots). (a) Time series of 5-day (pentad) mean station rainfall rates. (b) Local wavelet spectra of time series in (a), normalized by $1/\sigma^2$, where $\sigma^2 = 4.47$ (mm d^{-1})^2 is pentad rainfall variance. Area below thick broken line is the COI. (c) Global wavelet spectra (local power averaged over 1970–99) expressed as percentage of total global power summed over all frequencies. Insets in (c) magnify the major peaks (color coded) at indicated time scales. Here, (a) and (b) use the same abscissa for easy identification of the power at any given time. Thin solid black contours in (b) enclose areas of greater than 95% confidence for a red-noise process estimated from $\alpha = 0.5(\alpha_1 + (\alpha_2)^{1/2}) = 0.74$, where α_1 and α_2 are lag-1 and lag-2 autocorrelations of the time series (Torrence and Compo 1998). Year check marks in (a) and (b) indicate start of first pentad of season.

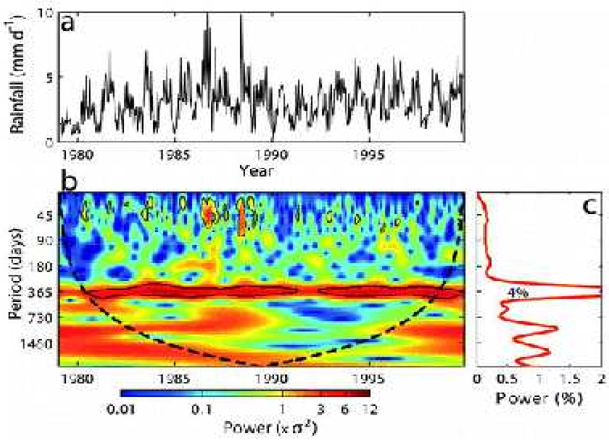


Fig. 3. As in Fig. 2, but for pentad CMAP rainfall estimates for Greater Horn of Africa for 1979–99 averaged over crosses in Fig. 1. Variance of time series in (a) is 2.3 (mm $d^{-1})^2$. Here, (b) and (c) use same ordinate for easy identification of power at any period. Thin solid contours in (b) enclose area of greater than 95% confidence for a red-noise process estimated from $\alpha = 0.5(\alpha_1 + (\alpha_2)^{1/2}) = 0.57$, where α_1 and α_2 are lag-1 and lag-2 autocorrelations of pentad CMAP rainfall.

[previous] [next]

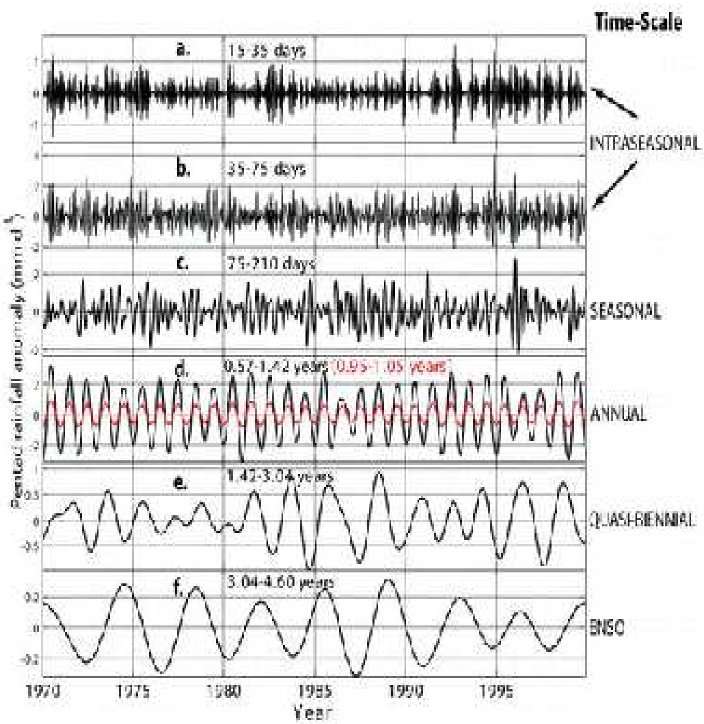


Fig. 4. Time series of pentad June–September Ethiopian wavelet-filtered rainfall anomalies (mm d^{-1}) for 1970–99. The nonoverlapping frequency bands shown contain the major power peaks and were delineated from the global power spectra in Fig. 2c. Year check marks indicate start of first pentad of season.

How can we use information separated out by time-scale to make predictions?

Reference: Webster and Hoyos, BAMS, Vol 85, 2004.

- Prediction of Asian monsoons on 15-35 day timescale.
- Morphology of the Monsoon Intraseasonal Oscillation (MISO)
- Madden-Julian Oscillation (MJO):
  - Eastward propagating convection
  - Largest variance in 20-40 day spectral band

Facts about South Asian Summers

(1) Convection stronger in eastern Indian Ocean.

(2) E. Indian Ocean convection lags western convection.

(3) Northern Indian Ocean convection primarily in the East.

(4) Convection or east equitorial Indian Ocean out of phase with convecton over India.

(5) North Indian Ocean convection coincides with development of monsoons over south asia.