An Introduction to Wavelets with Applications in Climatology

Don Percival

Applied Physics Lab, University of Washington

overheads for talk available at

http://staff.washington.edu/dbp/talks.html

joint work with:

• P. Craigmile (Ohio State)
• P. Guttorm (UW)
• S. Sardy (EPFL, Lausanne)
• B. Whitcher (NCAR)
Overview of Talk

- overview of discrete wavelet transform (DWT)
  - wavelet coefficients and their interpretation
  - DWT as a time series decorrelator

- three uses for wavelets (many more!)
  1. testing for variance changes
  2. bootstrapping auto/cross-correlation estimates
  3. estimating \( \delta \) for stationary/nonstationary fractional difference processes with trend
Overview of DWT

- Let \( X = [X_0, X_1, \ldots, X_{N-1}]^T \) be observed time series (for convenience, assume \( N \) integer multiple of \( 2^{J_0} \))

- Let \( \mathcal{W} \) be \( N \times N \) orthonormal DWT matrix (more precisely: partial DWT of level \( J_0 \))

- \( \mathbf{W} = \mathbf{W} \mathbf{X} \) is vector of DWT coefficients

- Can partition \( \mathbf{W} \) as follows:

\[
\mathbf{W} = \begin{bmatrix}
\mathbf{W}_1 \\
\vdots \\
\mathbf{W}_{J_0} \\
\mathbf{V}_{J_0}
\end{bmatrix}
\]

- \( \mathbf{W}_j \) contains \( N_j = N/2^j \) wavelet coefficients
  - Related to changes of averages at scale \( \tau_j = 2^{j-1} \) (\( \tau_j \) is \( j \)th ‘dyadic’ scale)
  - Related to times spaced \( 2^j \) units apart

- \( \mathbf{V}_{J_0} \) contains \( N_{J_0} = N/2^{J_0} \) scaling coefficients
  - Related to averages at scale \( \lambda_{J_0} = 2^{J_0} \)
  - Related to times spaced \( 2^{J_0} \) units apart
Example: DWT of FD Process

• $X_t$ called fractional difference (FD) process if it has a spectral density function (SDF) given by

$$S_X(f) = \frac{\sigma^2}{|2 \sin(\pi f)|^{2\delta}},$$

where $\sigma^2 > 0$ and $-\frac{1}{2} \leq \delta < \frac{1}{2}$

• note: for small $f$, have $S_X(f) \approx C/|f|^{2\delta}$; i.e., ‘$1/f$ type process’

• if $\delta = 0$, FD process is white noise

• if $0 < \delta < \frac{1}{2}$, process stationary with ‘long memory’

• can extend definition to $\delta \geq \frac{1}{2}$
  – nonstationary $1/f$ type process
  – also called ARFIMA(0,$\delta$,0) process

• Fig. 1: DWT of FD time series with $\delta = 0.4$
Two Consequences of Orthonormality

• multiresolution analysis (MRA)

\[ X = \mathcal{W}^T W = \sum_{j=1}^{J_0} \mathcal{W}_j^T W_j + \mathcal{V}_0^T V_0 \equiv \sum_{j=1}^{J_0} D_j + S_{J_0} \]

(\mathcal{W}_j \text{ partitions } \mathcal{W} \text{ commensurate with } \mathcal{W}_j)

- scale-based additive decomposition
- \(D_j\)'s & \(S_{J_0}\) called details & smooth
- Fig. 2: Nile River minimum flood levels

• analysis of variance:

\[ \hat{\sigma}^2_X \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \overline{X})^2 = \frac{1}{N} \left( \sum_{j=1}^{J_0} \| W_j \|^2 + \| V_0 \|^2 \right) - \overline{X}^2 \]

- scale-based decomposition (cf. frequency-based)
- can define wavelet variance \(\nu^2_X(\tau_j)\)
- for FD process, can deduce \(\delta\) from log/log plots since

\[ \nu^2_X(\tau_j) \approx C \tau_j^{2\delta-1} \]

- Fig. 3: Nile River minimum flood levels
DWT in Terms of Filters

• filter $X_0, X_1, \ldots, X_{N-1}$ to obtain

$$2^{j/2} \tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \mod N}, \quad t = 0, 1, \ldots, N - 1$$

where $h_{j,l}$ is $j$th level wavelet filter

– note: circular filtering

• subsample to obtain wavelet coefficients:

$$W_{j,t} = 2^{j/2} \tilde{W}_{j,2^j(t+1)-1}, \quad t = 0, 1, \ldots, N_j - 1,$$

where $W_{j,t}$ is $t$th element of $W_j$

• Figs. 4 & 5: Haar, D(4), C(6) & LA(8) wavelet filters

• $j$th wavelet filter is band-pass with pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$

• note: $j$th scale related to interval of frequencies

• similarly, scaling filters yield $V_{J_0}$

• Figs. 6 & 7: Haar, D(4), C(6) & LA(8) scaling filters

• $J_0$th scaling filter is low-pass with pass-band $[0, \frac{1}{2^{J_0+1}}]$
Wavelets as Whitening Filters

• recall Fig. 1: DWT of FD time series with $\delta = 0.4$

• since FD process is stationary, $W_j$ is also
  (ignoring terms influenced by circularity)

• Fig. 8: SDFs for each $W_j$

• DWT acts as whitening filter
  – requires SDF of $X$ to be $\approx$ flat over pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$
  – if not true, can use ‘wavelet packet’ transform (DWPT)
    – used by Flandrin, Tewfik & Kim, Wornell, McCoy & Walden

• three examples built on whitening property
  1. testing for variance changes
  2. bootstrapping auto/cross-correlation estimates
  3. estimating $\delta$ for stationary/nonstationary fractional difference processes with trend

• whitening property should help with other problems
Homogeneity of Variance: I

• claim: DWT approximately ‘decorrelates’ FD processes

• implication: $W_j$ should resemble white noise (ignoring coefficients influenced by circularity)
  
  \[ \text{cov} \{W_{j,t}, W_{j,t'}\} \approx 0 \text{ when } t \neq t' \]

  \[ \text{var} \{W_{j,t}\} \text{ should not vary with } t \]
  (homogeneity of variance)

• can test for homogeneity of variance using $W_j$

• suppose $Y_0, \ldots, Y_{N-1}$ independent normal RVs with $E\{Y_t\} = 0$ and var \{Y_t\} = $\sigma_t^2$

• want to test null hypothesis

  \[ H_0 : \sigma_0^2 = \sigma_1^2 = \cdots = \sigma_{N-1}^2 \]

• can test $H_0$ versus a variety of alternatives, e.g.,

  \[ H_1 : \sigma_0^2 = \cdots = \sigma_k^2 \neq \sigma_{k+1}^2 = \cdots = \sigma_{N-1}^2 \]

  using normalized cumulative sum of squares
Homogeneity of Variance: II

- to define test statistic $D$, start with
  \[ P_k \equiv \frac{\sum_{j=0}^{k} Y_j^2}{\sum_{j=0}^{N-1} Y_j^2}, \quad k = 0, \ldots, N - 2 \]
  and then compute
  \[ D^+ \equiv \max_{0 \leq k \leq N-2} \left( \frac{k + 1}{N - 1} - P_k \right) \quad \& \quad D^- \equiv \max_{0 \leq k \leq N-2} \left( P_k - \frac{k}{N - 1} \right) \]
  from which we form $D \equiv \max (D^+, D^-)$

- can reject $H_0$ if observed $D$ is ‘too large’

- can quantify ‘too large’ by considering distribution of $D$ under $H_0$

- need to find critical value $x_\alpha$ such that
  \[ P[D \geq x_\alpha] = \alpha \]
  for, e.g., $\alpha = 0.01, 0.05$ or 0.1

- once determined, can perform $\alpha$ level test of $H_0$:
  - compute $D$ statistic from data $Y_0, \ldots, Y_{N-1}$
  - reject $H_0$ at level $\alpha$ if $D \geq x_\alpha$
Homogeneity of Variance: III

• can determine critical values \( x_\alpha \) in two ways
  – Monte Carlo simulations
  – large sample approximation to distribution of \( D_\cdot \):

\[
P[(N/2)^{1/2}D \geq x] \approx 1 + 2 \sum_{l=1}^{\infty} (-1)^l e^{-2l^2x^2}
\]

(reasonable approximation for \( N \geq 128 \))

• idea: given time series \( X \), compute \( D \) using

\[
W_{j,t} = 2^{j/2}\hat{W}_{j,2^{j(t+1)}-1}, \quad \left( (L - 2) \left( 1 - \frac{1}{2^j} \right) \right) \leq t \leq \left\lfloor \frac{N}{2^j} - 1 \right\rfloor,
\]

where \( L \) is length of \( j = 1 \) level wavelet filter and

\[
2^{j/2}\hat{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l}X_{t-l \mod N}
\]

– results in ‘level by level’ tests

– above formulation allows for general \( N \)
  (i.e., \( N \) need not be multiple of \( 2^{J_0} \))

• Q: is DWT decorrelation of FD processes good enough?

• Fig. 9: yes!
Example: Nile River Minima

• recall MRA & wavelet variance plots

• application of homogeneity of variance test:

<table>
<thead>
<tr>
<th>scale</th>
<th>$D$</th>
<th>$x_{0.1}$</th>
<th>$x_{0.05}$</th>
<th>$x_{0.01}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 year</td>
<td>0.1559</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 years</td>
<td>0.1754</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 years</td>
<td>0.1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 years</td>
<td>0.2313</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• if $H_0$ rejected, use ‘nondecimated’ DWT to detect change point:
  – compute rotated cumulative variance curve
  – look for time of largest excursion from 0

• Fig. 10: change point detection
  – 720 AD for level $j = 1$
  – 722 AD for level $j = 2$
  – agrees well with mosque construction in 715 AD

• interpretation differs from Beran & Terrin (1996)
Wavelet-Based Bootstrapping: I

• Davison & Hinkley, 1998, Chapter 8, discusses bootstrapping in context of time series analysis

• whitening allows wavelet-based bootstrapping for certain statistics (but not all)

• first example: lag 1 autocorrelation estimate

\[ \hat{r}_1 \equiv \frac{\sum_{t=0}^{N-2} X_t X_{t+1}}{\sum_{t=0}^{N-1} X_t^2} \]

• idea: to get standard error of \( \hat{r}_1 \),
  
  – compute DWT of \( X \)
  
  – sample with replacement from \( W_j \) to form \( W_j^{(b)} \) 
    (do same with \( V_{J_0} \))
  
  – synthesize \( X^{(b)} \) using \( W_j^{(b)} \)'s & \( V_{J_0}^{(b)} \)
  
  – compute \( \hat{r}_1^{(b)} \) for \( X^{(b)} \)
  
  – repeat until computer gets tired
  
  – use standard error of \( \hat{r}_1^{(b)} \)'s for \( X^{(b)} \)'s to assess standard error of \( \hat{r}_1 \) for \( X \)
Wavelet-Based Bootstrapping: II

• to test scheme, did Monte Carlo study involving
  – AR(1) process: \( X_t = 0.9X_{t-1} + \epsilon_t \)
  – MA(1) process: \( X_t = \epsilon_t + \epsilon_{t-1} \)
  – FD process with \( \delta = 0.45 \)
• average \( \hat{r}_1^{(b)} \)'s have negligible bias
• comparison of standard errors, \( N = 128 \):

<table>
<thead>
<tr>
<th></th>
<th>LA(8) DWT</th>
<th>LA(8) DWPT</th>
<th>true</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.057</td>
<td>0.052</td>
<td>0.048</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.071</td>
<td>0.068</td>
<td>0.063</td>
</tr>
<tr>
<td>FD</td>
<td>0.094</td>
<td>0.083</td>
<td>0.107</td>
</tr>
</tbody>
</table>

• comparison of standard errors, \( N = 1024 \):

<table>
<thead>
<tr>
<th></th>
<th>LA(8) DWT</th>
<th>LA(8) DWPT</th>
<th>true</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(1)</td>
<td>0.016</td>
<td>0.015</td>
<td>0.014</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.026</td>
<td>0.024</td>
<td>0.022</td>
</tr>
<tr>
<td>FD</td>
<td>0.044</td>
<td>0.042</td>
<td>0.053</td>
</tr>
</tbody>
</table>

• handles both short & long memory models
Wavelet-Based Bootstrapping: III

• second example: cross-correlation estimate

\[ \hat{r}_0^{(XY)} \equiv \frac{\sum_{t=0}^{N-1} X_t Y_t}{\left( \sum_{t=0}^{N-1} X_t^2 \sum_{t=0}^{N-1} Y_t^2 \right)^{1/2}} \]

• to assess null hypothesis \( r_0^{(XY)} = 0 \),
  
  – separately generate \( X^{(b)} \) & \( Y^{(b)} \)
  
  – bootstrapped \( \hat{r}_0^{(XY)} \) should reflect variability in \( \hat{r}_0^{(XY)} \) under null

• Fig. 11: two time series
  
  – maximum annual snow-pack level at Mt. Rainier
  
  – Pacific decadel oscillation index

• cross-correlation estimate is \( \hat{r}_0^{(XY)} = -0.26 \)

• Q: is this significantly different from 0?

• Fig. 12: yes, can reject null at critical level \( p = 0.01 \)
Estimation for FD Processes: I

• extension of work by Wornell, McCoy & Walden

• problem: estimate $\delta$ from time series $U_t$ such that

$$U_t = T_t + X_t$$

where

- $T_t \equiv \sum_{j=0}^r a_j t^j$ is polynomial trend
- $X_t$ is FD process, but can have $\delta \geq \frac{1}{2}$

• DWT wavelet filter of width $L$ has embedded differencing operation of order $L/2$

• if $\frac{L}{2} \geq r + 1$, reduces polynomial trend to 0

• can partition DWT coefficients as

$$W = W_s + W_b + W_w$$

where

- $W_s$ has scaling coefficients and 0s elsewhere
- $W_s$ has boundary-dependent wavelet coefficients
- $W_w$ has boundary-independent wavelet coefficients
Estimation for FD Processes: II

- since $U = \mathcal{W}^T \mathbf{W}$, can write
  
  $$U = \mathcal{W}^T (\mathbf{W}_s + \mathbf{W}_b) + \mathcal{W}^T \mathbf{W}_w \equiv \hat{T} + \tilde{X}$$

- Fig. 13: Hansen–Lebedeff global temperature index
  - can use values in $\mathbf{W}_w$ to form likelihood:
  
  $$L(\delta, \sigma^2_\epsilon) \equiv \prod_{j=1}^{J_0} \prod_{t=1}^{N_j'} \frac{1}{(2\pi \sigma^2_j)^{1/2}} e^{-W^2_{j,t} + L'_{j-1} - 1/2(2\sigma^2_j)}$$

  where $\sigma^2_\epsilon$ is innovations variance;

  $$\sigma^2_j \equiv \int_{-1/2}^{1/2} \mathcal{H}_j(f) \frac{\sigma^2_\epsilon}{|2\sin(\pi f)|^{2\delta}} df;$$

  and $\mathcal{H}_j(\cdot)$ is squared gain for $\{h_{j,l}\}$

- works well in Monte Carlo simulations

- for Hansen–Lebedeff series, get $\hat{\delta} = 0.40 \pm 0.08$

- can also test for significance of $\hat{T}$