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. Guttorp (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 δ for stationary/nonstationary fractional difference processes with trend

Overview of DWT

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ 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)
- $\mathbf{W} = \mathcal{W}\mathbf{X}$ 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 $\tau_j = 2^{j-1}$ (τ_j is jth '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,δ,0) process
- Fig. 1: DWT of FD time series with $\delta = 0.4$

Two Consequences of Orthonormality

- multiresolution analysis (MRA)

$$\mathbf{X} = \mathcal{W}^T\mathbf{W} = \sum_{j=1}^{J_0} \mathcal{W}_j^T\mathbf{W}_j + \mathcal{V}_{J_0}^T\mathbf{V}_{J_0} \equiv \sum_{j=1}^{J_0} \mathcal{D}_j + \mathcal{S}_{J_0}$$

($\mathcal{W}_j$ partitions $\mathcal{W}$ commensurate with $\mathbf{W}_j$)

  - scale-based additive decomposition
  - $\mathcal{D}_j$'s & $\mathcal{S}_{J_0}$ called details & smooth
  - Fig. 2: Nile River minimum flood levels

- analysis of variance:

$$\hat{\sigma}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}(X_t - \overline{X})^2 = \frac{1}{N}\left(\sum_{j=1}^{J_0}\|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2\right) - \overline{X}^2$$

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

$$\nu_X^2(\tau_j) \approx C\tau_j^{2\delta-1}$$

  - Fig. 3: Nile River minimum flood levels

DWT in Terms of Filters

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

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

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

  – note: circular filtering

- subsample to obtain wavelet coefficients:

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

  where $W_{j,t}$ is tth element of $\mathbf{W}_j$

- Figs. 4 & 5: Haar, D(4), C(6) & LA(8) wavelet filters
- jth wavelet filter is band-pass with pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$
- note: jth scale related to interval of frequencies
- similarly, scaling filters yield $\mathbf{V}_{J_0}$
- Figs. 6 & 7: Haar, D(4), C(6) & LA(8) scaling filters
- J_0th 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, $\mathbf{W}_j$ is also (ignoring terms influenced by circularity)
- Fig. 8: SDFs for each $\mathbf{W}_j$
- DWT acts as whitening filter
  - requires SDF of $\mathbf{X}$ 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 δ 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: $\mathbf{W}_j$ should resemble white noise (ignoring coefficients influenced by circularity)
  - $\operatorname{cov}\{W_{j,t}, W_{j,t'}\} \approx 0$ when $t \neq t'$
  - $\operatorname{var}\{W_{j,t}\}$ should not vary with t (homogeneity of variance)
- can test for homogeneity of variance using $\mathbf{W}_j$
- suppose $Y_0, \ldots, Y_{N-1}$ independent normal RVs with $E\{Y_t\} = 0$ and $\operatorname{var}\{Y_t\} = \sigma_t^2$
- 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

$$\mathcal{P}_k \equiv \frac{\sum_{j=0}^{k} Y_j^2}{\sum_{j=0}^{N-1} Y_j^2}, \quad k = 0, \ldots, N-2$$

  and then compute

$$D^+ \equiv \max_{0 \le k \le N-2} \left(\frac{k+1}{N-1} - \mathcal{P}_k \right) \ \& \ D^- \equiv \max_{0 \le k \le N-2} \left(\mathcal{P}_k - \frac{k}{N-1} \right)$$

  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_α such that

$$\mathbf{P}[D \ge x_\alpha] = \alpha$$

  for, e.g., $\alpha = 0.01$, 0.05 or 0.1

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

Homogeneity of Variance: III

- can determine critical values x_α in two ways
  - Monte Carlo simulations
  - large sample approximation to distribution of D:

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

    (reasonable approximation for $N \geq 128$)

- idea: given time series $\mathbf{X}$, compute D using

$$W_{j,t} = 2^{j/2}\widetilde{W}_{j,2^j(t+1)-1}, \quad \left\lceil (L-2)\left(1-\frac{1}{2^j}\right)\right\rceil \leq t \leq \left\lfloor \frac{N}{2^j} - 1 \right\rfloor,$$

  where L is length of $j = 1$ level wavelet filter and

$$2^{j/2}\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} h_{j,l}X_{t-l \bmod N}$$

  - 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:

scale	D	$x_{0.1}$	$x_{0.05}$	$x_{0.01}$
1 year	0.1559	0.0945	0.1051	0.1262
2 years	0.1754	0.1320	0.1469	0.1765
4 years	0.1000	0.1855	0.2068	0.2474
8 years	0.2313	0.2572	0.2864	0.3436

- 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 $\mathbf{X}$
  - sample with replacement from $\mathbf{W}_j$ to form $\mathbf{W}_j^{(b)}$ (do same with $\mathbf{V}_{J_0}$)
  - synthesize $\mathbf{X}^{(b)}$ using $\mathbf{W}_j^{(b)}$'s & $\mathbf{V}_{J_0}^{(b)}$
  - compute $\hat{r}_1^{(b)}$ for $\mathbf{X}^{(b)}$
  - repeat until computer gets tired
  - use standard error of $\hat{r}_1^{(b)}$'s for $\mathbf{X}^{(b)}$'s to assess standard error of $\hat{r}_1$ for $\mathbf{X}$

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, $N = 128$:

	LA(8) DWT	LA(8) DWPT	true
AR(1)	0.057	0.052	0.048
MA(1)	0.071	0.068	0.063
FD	0.094	0.083	0.107

- comparison of standard errors, $N = 1024$:

	LA(8) DWT	LA(8) DWPT	true
AR(1)	0.016	0.015	0.014
MA(1)	0.026	0.024	0.022
FD	0.044	0.042	0.053

- 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 $\mathbf{X}^{(b)}$ & $\mathbf{Y}^{(b)}$
  - 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 δ from time series U_t such that

$$U_t = T_t + X_t$$

  where

  - $T_t \equiv \Sigma_{j=0}^{r} a_j t^j$ 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

$$\mathbf{W} = \mathbf{W}_s + \mathbf{W}_b + \mathbf{W}_w$$

  where

  - $\mathbf{W}_s$ has scaling coefficients and 0s elsewhere
  - $\mathbf{W}_s$ has boundary-dependent wavelet coefficients
  - $\mathbf{W}_w$ has boundary-independent wavelet coefficients

Estimation for FD Processes: II

- since $\mathbf{U} = \mathcal{W}^T\mathbf{W}$, can write

$$\mathbf{U} = \mathcal{W}^T(\mathbf{W}_s + \mathbf{W}_b) + \mathcal{W}^T\mathbf{W}_w \equiv \widehat{\mathbf{T}} + \widetilde{\mathbf{X}}$$

- Fig. 13: Hansen–Lebedeff global temperature index
- can use values in $\mathbf{W}_w$ to form likelihood:

$$L(\delta, \sigma_\epsilon^2) \equiv \prod_{j=1}^{J_0} \prod_{t=1}^{N_j'} \frac{1}{\left(2\pi\sigma_j^2\right)^{1/2}} e^{-W^2_{j,t+L_j'-1}/(2\sigma_j^2)}$$

where σ_ϵ^2 is innovations variance;

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

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

- works well in Monte Carlo simulations
- for Hansen–Lebedeff series, get $\hat{\delta} \doteq 0.40 \pm 0.08$
- can also test for significance of $\widehat{\mathbf{T}}$