An Introduction to the Wavelet Variance and Its Statistical Properties

Don Percival

Applied Physics Laboratory, University of Washington

Seattle, Washington, USA

overheads for talk available at

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

Overview of Talk

- review of wavelets
  - wavelet filters
  - wavelet coefficients and their interpretation
- wavelet variance decomposition of sample variance
- theoretical wavelet variance for stochastic processes
  - stationary processes
  - nonstationary processes with stationary differences
- sampling theory for Gaussian processes with an example
- sampling theory for non-Gaussian processes with an example
- use on time series with time-varying statistical properties
- summary

Wavelet Filters & Coefficients: I

- let $\{X_t : t = 0, \ldots, N-1\}$ be a time series (e.g., X_t is temperature at noon on tth day of year)
- filter $\{X_t\}$ to obtain wavelet coefficients:

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t = 0, 1, \ldots, N-1;$$

  here $X_t = X_{t \bmod N}$ if $t < 0$ (assumes 'circularity')
- $\{\tilde{h}_{j,l}\}$ is wavelet filter for scale $\tau_j = 2^{j-1}$, $j = 1, 2, \ldots$
  - 'scale' refers to effective half-width of $\{\tilde{h}_{j,l}\}$; i.e., $\widetilde{W}_{j,t}$ effectively determined by $2\tau_j$ values in $\{X_t\}$
  - $\{\tilde{h}_{j,l}\}$ 'stretched out' version of $j = 1$ filter $\{\tilde{h}_{1,l}\}$
  - actual width of $\{\tilde{h}_{1,l}\}$ is $L = L_1$
  - actual width of $\{\tilde{h}_{j,l}\}$ is $L_j = (2^j - 1)(L - 1) + 1$
  - Daubechies filters $\{\tilde{h}_{1,l}\}$ have special properties
    * $\sum_{l=0}^{L-1} \tilde{h}_{1,l} = 0$
    * $\sum_{l=0}^{L-1} \tilde{h}_{1,l}^2 = 1/2$
    * $\sum_{l=0}^{L-1} \tilde{h}_{1,l} \tilde{h}_{1,l+2k} = 0$ for nonzero integers k

Wavelet Filters & Coefficients: II

- Fig. 1: $\{\tilde{h}_{j,l}\}$ for Haar wavelet filter ($L = 2$)
- note form of Haar wavelet coefficients for scale τ_j:

$$\widetilde{W}_{j,t} \propto \overline{X}_t(\tau_j) - \overline{X}_{t-\tau_j}(\tau_j),$$

  where

$$\overline{X}_t(\tau_j) \equiv \frac{1}{\tau_j} \sum_{l=0}^{\tau_j - 1} X_{t-l}$$

- $\widetilde{W}_{j,t} \propto$ change in adjacent averages of τ_j values
  - change measured by simple first difference
  - average is localized sample mean
  - if $\widetilde{W}_{j,t}^2$ small, not much variation over scale τ_j
  - if $\widetilde{W}_{j,t}^2$ large, lot of variation over scale τ_j

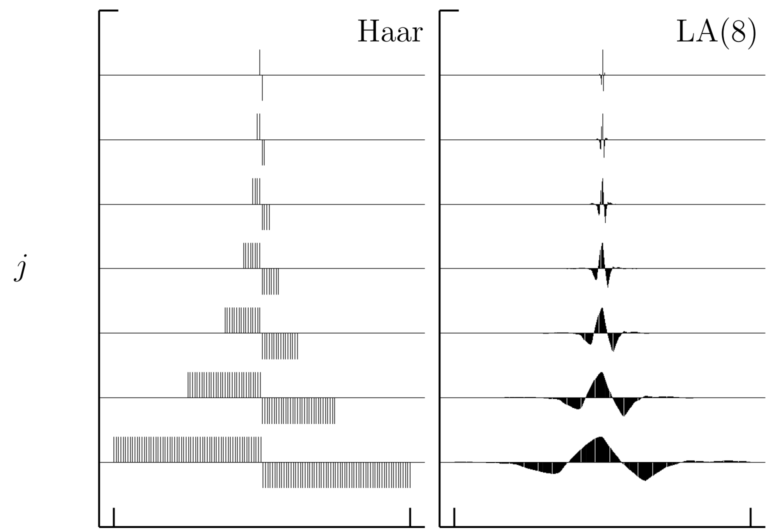

Figure 1: Haar and LA(8) wavelet filters $\{\tilde{h}_{j,l}\}$ for scales indexed by $j = 1, 2, \ldots, 7$.

Wavelet Filters & Coefficients: III

- Haar is $L = 2$ member of Daubechies wavelet filters (L even & typically ranges from 2 up to 20)
- Fig. 1: $\{\tilde{h}_{j,l}\}$ for LA(8) wavelet filter ($L = 8$); here 'LA' stands for 'least asymmetric'
- filtering $\{X_t\}$ with $\{\tilde{h}_{j,l}\}$ yields LA(8) wavelet coefficients $\widetilde{W}_{j,t}$
- $\widetilde{W}_{j,t} \propto$ change between average over scale τ_j and its surroundings
  - change measured by $L/2 = 4$ first differences
  - average is localized weighted average
- pattern holds for all Daubechies wavelet filters: $\widetilde{W}_{j,t} \propto$ difference between localized weighted average and its surroundings

$$\{X_t\} \longrightarrow \boxed{\{a_{j,l}\}} \longrightarrow \underbrace{\boxed{\{1,-1\}} \longrightarrow \cdots \longrightarrow \boxed{\{1,-1\}}}_{L/2 \text{ of these}} \longrightarrow \{\widetilde{W}_{j,t}\},$$

where $\{a_{j,l}\}$ produces localized weighted averages

Empirical Wavelet Variance

- collect $\widetilde{W}_{j,t}$ into $\widetilde{\mathbf{W}}_j$ for levels $j = 1, 2, \ldots, J_0$
- also compute vector $\widetilde{\mathbf{V}}_{J_0}$ of scaling coefficients:

$$\widetilde{V}_{J_0,t} \equiv \sum_{l=0}^{L_{J_0}-1} \tilde{g}_{J_0,l} X_{t-l}, \quad t = 0, 1, \ldots, N-1;$$

  $\{\tilde{g}_{J_0,l}\}$ called scaling filter (depends just on $\{\tilde{h}_{1,l}\}$)

- Fig. 2: Haar & LA(8) scaling filters $\{\tilde{g}_{J_0,l}\}$
  - $\widetilde{V}_{J_0,t}$ is weighted average over scale $2\tau_j$
- obtain analysis of sample variance:

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

  (if $N = 2^{J_0}$, can argue that $\|\widetilde{\mathbf{V}}_{J_0}\|^2/N = \overline{X}^2$).

- $\frac{1}{N}\|\widetilde{\mathbf{W}}_j\|^2$ portion of $\hat{\sigma}_X^2$ due to changes in averages over scale τ_j; i.e., 'scale by scale' analysis of variance
- cf. 'frequency by frequency' analysis of variance:

$$\hat{\sigma}_X^2 = \frac{1}{N}\|\mathbf{F}\|^2 - \overline{X}^2 \text{ with } F_k \equiv \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi tk/N}$$

- scale τ_j related to frequency interval $[1/2^{j+1}, 1/2^j]$
- Fig. 3: example of empirical wavelet variance

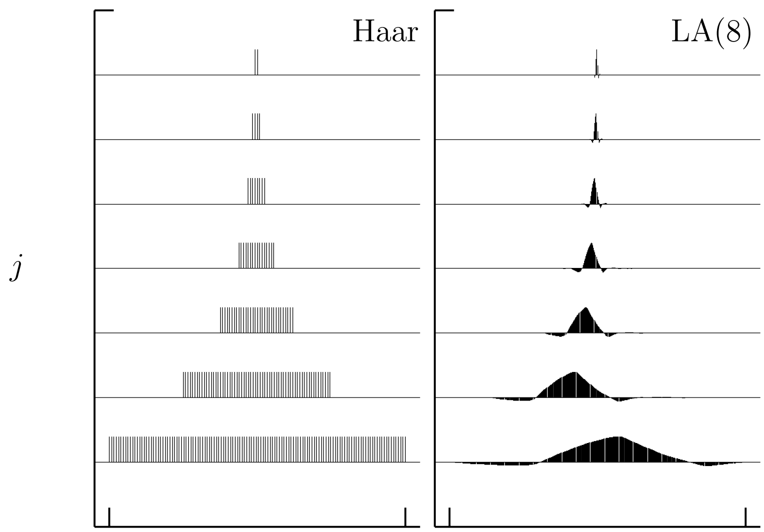

Figure 2: Haar and LA(8) scaling filters $\{\tilde{g}_{J_0,l}\}$ for scales indexed by $J_0 = 1, 2, \ldots, 7$.

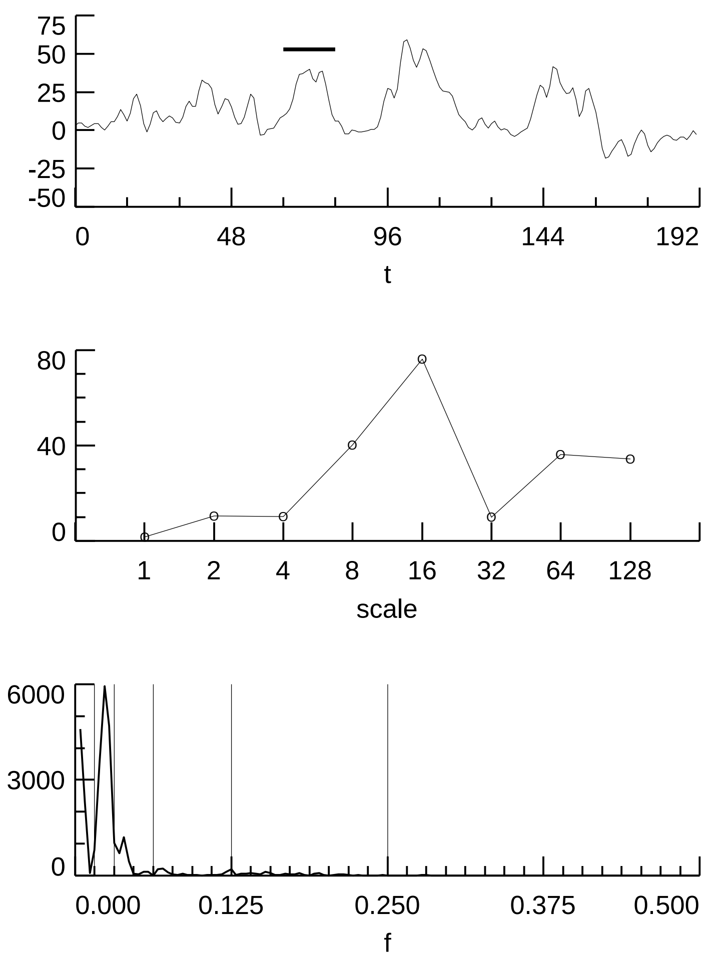

Figure 3: Time series of subtidal sea levels (top plot), along with associated empirical wavelet variances $\|\widetilde{\mathbf{W}}_j\|^2/N$ versus scales $\tau_j = 2^{j-1}$ for $j = 1, \ldots, 8$ (middle) and periodogram versus frequency (bottom).

Theoretical Wavelet Variance

- now assume X_t is real-valued random variable
- $\{X_t : t \in \mathbb{Z}\}$ is stochastic process ($\mathbb{Z}$ is set of all integers)
- filter $\{X_t\}$ to create new stochastic process:

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t \in \mathbb{Z},$$

  which should be contrasted with

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t = 0, 1, \ldots, N-1$$

  where $X_t = X_{t \bmod N}$ when $t < 0$

- definition of time dependent wavelet variance
  (also called wavelet spectrum):

$$\nu^2_{X,t}(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\},$$

  with conditions on $\{X_t\}$ so that $\operatorname{var}\{\overline{W}_{j,t}\}$ exists and is finite

- $\nu^2_{X,t}(\tau_j)$ depends on τ_j and t
- will focus time independent wavelet variance

$$\nu^2_X(\tau_j) \equiv \operatorname{var}\{\overline{W}_{j,t}\}$$

  (can adapt theory to handle time varying situation)

Rationale for Wavelet Variance

- decomposes variance on scale by scale basis
- useful substitute/complement for spectral density function (SDF)
- useful substitute for process/sample variance
- well-defined for certain nonstationary processes

Variance Decomposition

- if $\{X_t\}$ stationary process with SDF, then

$$\int_{-1/2}^{1/2} S_X(f)\, df = \mathrm{var}\,\{X_t\};$$

  i.e., SDF decomposes $\mathrm{var}\,\{X_t\}$ across frequencies f

  - have analogous result for sample variance
  - involves uncountably infinite number of f's
  - $S_X(f)\,\Delta f \approx$ contribution to $\mathrm{var}\,\{X_t\}$ due to f's in interval of length Δf centered at f

- wavelet variance yields analogous decomposition:

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \mathrm{var}\,\{X_t\}$$

  i.e., decomposes $\mathrm{var}\,\{X_t\}$ across scales τ_j

  - have analogous result for sample variance
  - involves countably infinite number of τ_j's
  - $\nu_X^2(\tau_j)$ contribution to $\mathrm{var}\,\{X_t\}$ due to scale τ_j
  - $\nu_X(\tau_j)$ has same units as X_t

SDF Substitute/Complement: I

- because $\{\tilde{h}_{j,l}\} \approx$ bandpass over $[1/2^{j+1}, 1/2^j]$,

$$\nu_X^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_X(f)\, df \tag{1}$$

- if $S_X(\cdot)$ 'featureless', $\{\nu_X^2(\tau_j)\}$ as informative as $S_X(\cdot)$
- $\{\nu_X^2(\tau_j)\}$ more succinct: one value per 'octave band'
- example: $S_X(f) \propto |f|^\alpha$, i.e., pure power law process
  - can deduce α from slope of $\log S_X(f)$ vs. $\log f$
  - (1) implies $\nu_X^2(\tau_j) \propto \tau_j^{-\alpha-1}$ approximately
  - can deduce α from slope of $\log \nu_X^2(\tau_j)$ vs. $\log \tau_j$
  - no real loss in using $\nu_X^2(\tau_j)$ in place of $S_X(\cdot)$

SDF Substitute/Complement: II

- $\nu_X^2(\tau_j)$ easier to estimate than SDF
- basic estimator of SDF is periodogram: given $X_0, \ldots, X_{N-1}$,

$$\hat{S}_X^{(p)}(f_k) \equiv \frac{1}{N} \left| \sum_{t=0}^{N-1} (X_t - \overline{X}) e^{-i2\pi f_k t} \right|^2, \quad f_k \equiv \frac{k}{N}$$

  - inconsistent because $\text{var}\{\hat{S}_X^{(p)}(f_k)\} \approx S_X^2(f_k)$ (i.e., does not decrease to 0 as $N \to \infty$)
  - need smoothers etc. to get consistency
  - can be badly biased
- basic estimator of $\nu_X^2(\tau_j)$ is

$$\tilde{\nu}_X^2(\tau_j) \equiv \frac{1}{N} \sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2$$

  - biased since $\widetilde{W}_{j,t}, t = 0, \ldots, L_j - 1$, influenced by circularity
  - unbiased if these L_j terms are dropped
  - estimator so constructed is consistent

Substitute for Variance: I

- can be difficult to estimate process variance for stationary $\{X_t\}$
- argument: $\nu_X^2(\tau_j)$ easier to estimate
- to understand why, suppose $\{X_t\}$ has
  - known mean $\mu_X = E\{X_t\}$
  - unknown variance σ_X^2
- can estimate σ_X^2 using

$$\tilde{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \mu_X)^2$$

- estimator above is unbiased: $E\{\tilde{\sigma}_X^2\} = \sigma_X^2$
- now suppose μ_X is unknown
- can estimate σ_X^2 using

$$\hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \overline{X})^2, \text{ where } \overline{X} \equiv \frac{1}{N} \sum_{t=0}^{N-1} X_t$$

Substitute for Variance: II

- can argue that $E\{\hat{\sigma}_X^2\} = \sigma_X^2 - \text{var}\,\{\overline{X}\}$
- implies $0 \leq E\{\hat{\sigma}_X^2\} \leq \sigma_X^2$ because $\text{var}\,\{\overline{X}\} \geq 0$
- $E\{\hat{\sigma}_X^2\} \to \sigma_X^2$ as $N \to \infty$ if SDF exists ... **but**
  - for any small $\epsilon > 0$ (say, $0.00\cdots01$) and
  - for any sample size N (say, $N = 10^{10^{10}}$)

  there exists a (nonpathological!) $\{X_t\}$ such that

  $$E\{\hat{\sigma}_X^2\} < \epsilon\sigma_X^2$$

  for chosen N; i.e., $\hat{\sigma}_X^2$ badly biased even for very large N

Substitute for Variance: III

- consider fractional Gaussian noise (FGN) with parameter H (called Hurst coefficient)
- for $H = 1/2$, FGN is white noise (i.e., uncorrelated)
- $1/2 < H < 1$ is stationary 'long memory' process (i.e., has slowly decaying autocovariance sequence)
- can argue that $\mathrm{var}\,\{\overline{X}\} = \sigma_X^2/N^{2-2H}$
  - $H = 1/2$: $\mathrm{var}\,\{\overline{X}\} = \sigma_X^2/N$ ('classic' rate of decay)
  - $H = 1 - \delta/2$, $0 < \delta < 1$: $\mathrm{var}\,\{\overline{X}\} = \sigma_X^2/N^{\delta}$; i.e., slower rate of decay than classic
- for given $0 < \epsilon < 1$ and $N > 1$, have

$$E\{\hat{\sigma}_X^2\} < \epsilon\sigma_X^2 \text{ if we pick } H > 1 - \frac{\log(1-\epsilon)}{2\log(N)}$$

- Fig. 4: realization of FGN, $\sigma_X^2 = 1$, $H = 0.9$ & $N = 1000$
  - using $\mu_X = 0$, obtain $\hat{s}'_0 \doteq 0.99$
  - using $\overline{X} \doteq 0.53$, obtain $\hat{\sigma}_X^2 \doteq 0.71$; note that $E\{\hat{\sigma}_X^2\} \doteq 0.75$
  - need $N \geq 10^{10}$ so that $s_{X,0} - E\{\hat{\sigma}_X^2\} \leq 0.01$; i.e., for the bias to be 1% or less of true σ_X^2
- conclusion: $\hat{\sigma}_X^2$ can be badly biased if μ_X unknown (can patch up by estimating H, but need model)

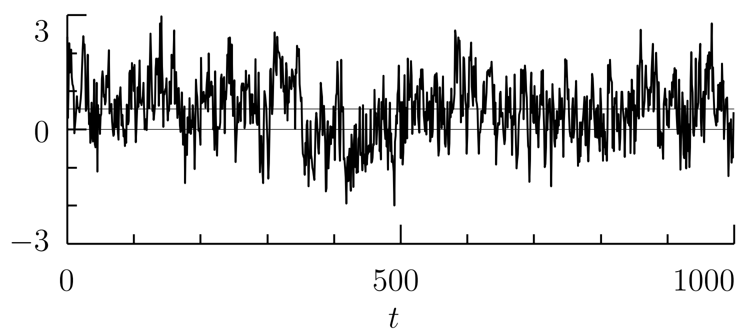

Figure 4: Realization of a fractional Gaussian noise (FGN) process with Hurst coefficient $H = 0.9$. The sample mean of approximately 0.53 and the true mean of zero are indicated by the thin horizontal lines (taken from Figure 300, Percival and Walden, 2000, copyright Cambridge University Press).

Substitute for Variance: IV

- Q: why is wavelet variance useful when $\hat{\sigma}_X^2$ is not?
- replaces 'global' variability with variability over scales
- if $\{X_t\}$ stationary with mean μ_X, then

$$E\{\overline{W}_{j,t}\} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} E\{X_{t-l}\} = \mu_X \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0$$

  because always have $\sum_l \tilde{h}_{j,l} = 0$
- $E\{\overline{W}_{j,t}\}$ known, so estimator of var $\{\overline{W}_{j,t}\} = \nu_X^2(\tau_j)$ unbiased

Generalization to Nonstationary Processes

- if L is properly chosen, $\nu_X^2(\tau_j)$ well-defined for processes with stationary backward differences
- let B be such that $BX_t \equiv X_{t-1} \Rightarrow B^k X_t = X_{t-k}$
- X_t has dth order stationary backward differences if

$$Y_t \equiv (1-B)^d X_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}$$

forms a stationary process (d nonnegative integer)

$$\{X_t\} \longrightarrow \underbrace{\boxed{\{1,-1\}} \longrightarrow \cdots \longrightarrow \boxed{\{1,-1\}}}_{d \text{ of these}} \longrightarrow \{Y_t\}$$

- if $\{X_t\}$ stationary, $\{Y_t\}$ is also with

$$S_Y(f) = [4\sin^2(\pi f)]^d S_X(f) \equiv \mathcal{D}^d(f) S_X(f)$$

- if $\{X_t\}$ nonstationary but dth order differences are, can *define* SDF for $\{X_t\}$ via

$$S_X(f) \equiv \frac{S_Y(f)}{[4\sin^2(\pi f)]^d} = \frac{S_Y(f)}{\mathcal{D}^d(f)}$$

(Yaglom, 1958)

- attaches meaning to, e.g., $S_X(f) \propto |f|^{-5/3}$
- Fig. 5: examples

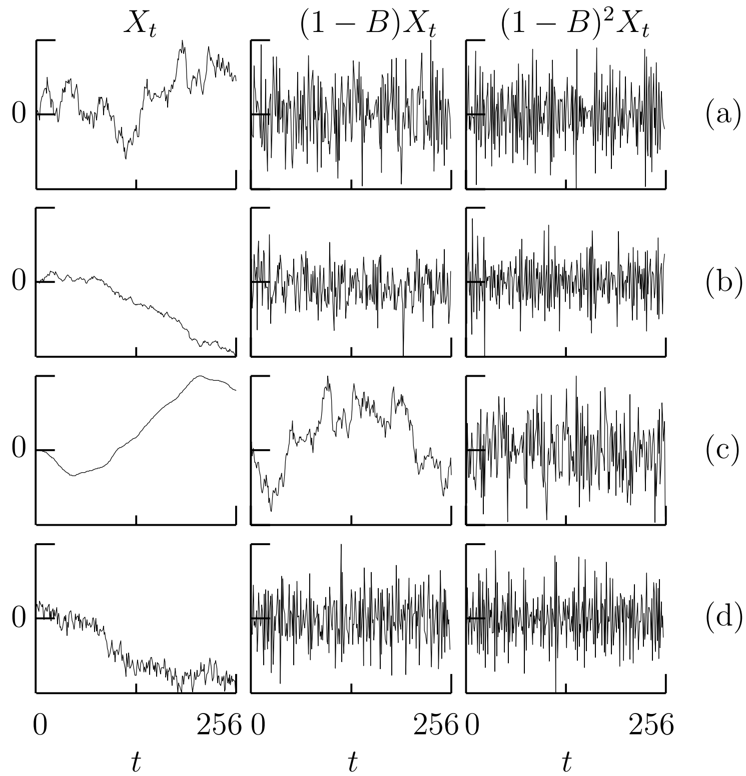

Figure 5: Simulated realizations of nonstationary processes $\{X_t\}$ with stationary backward differences of various orders (first column) along with their first backward differences $\{(1 - B)X_t\}$ (second column) and second backward differences $\{(1-B)^2X_t\}$ (final column). From top to bottom, the processes are (a) a random walk; (b) a modified random walk, formed using a white noise sequence with mean $\mu_\varepsilon = -0.2$; (c) a 'random run' (i.e., cumulative sums of a random walk); and (d) a process formed by summing the line given by $-0.05t$ and a simulation of a stationary FD process with $\delta = 0.45$ (taken from Figure 289, Percival and Walden, 2000, copyright Cambridge University Press).

Wavelet Variance for Processes with Stationary Backward Differences: I

- suppose $\{X_t\}$ has dth order stationary differences
- recall that

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \quad t \in \mathbb{Z}$$

- claim: if $L \geq 2d$, $\{\overline{W}_{j,t}\}$ stationary with SDF

$$S_j(f) = \widetilde{\mathcal{H}}_j^{(D)}(f) S_X(f)$$

where $\widetilde{\mathcal{H}}_j^{(D)}(\cdot)$ is squared gain function for $\{\tilde{h}_{j,l}\}$

- proof: $\{\tilde{h}_{j,l}\} \Leftrightarrow \frac{L}{2}$ first differences & then $\{a_{j,l}\}$ so

$$\{X_t\} \longrightarrow \underbrace{\boxed{\{1,-1\}} \longrightarrow \cdots \longrightarrow \boxed{\{1,-1\}}}_{L/2 \text{ of these}} \longrightarrow \{Y_t\}$$

is stationary with SDF $S_Y(f) = \mathcal{D}^d(f) S_X(f)$;

$$\{Y_t\} \longrightarrow \underbrace{\boxed{\{1,-1\}} \longrightarrow \cdots \longrightarrow \boxed{\{1,-1\}}}_{L/2-d \text{ of these}} \longrightarrow \boxed{\{a_{j,l}\}} \longrightarrow \{\overline{W}_{j,t}\}$$

Wavelet Variance for Processes with Stationary Backward Differences: II

- with $\mu_Y \equiv E\{Y_t\}$, have
  - $E\{\overline{W}_{j,t}\} = 0$ if either
    - $L > 2d$ or
    - $L = 2d$ & $\mu_Y = 0$
  - $E\{\overline{W}_{j,t}\} \neq 0$ if $L = 2d$ & $\mu_Y \neq 0$
- conclusions: $\nu_X^2(\tau_j)$ well-defined for $\{X_t\}$ that is
  - stationary: any L will do & $E\{\overline{W}_{j,t}\} = 0$
  - nonstationary with dth order stationary backward differences: need $L \geq 2d$, but might need $L > 2d$ to get $E\{\overline{W}_{j,t}\} = 0$
- have

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \begin{cases} \text{var}\,\{X_t\} < \infty & \text{if } \{X_t\} \text{ stationary;} \\ \infty & \text{if } \{X_t\} \text{ nonstationary} \end{cases}$$

Unbiased Estimator of Wavelet Variance

- suppose have realization of $X_0, X_1, \ldots, X_{N-1}$, where $\{X_t\}$ has dth order stationary differences
- want to estimate $\nu_X^2(\tau_j)$ for wavelet filter such that $L \geq 2d$ & $E\{\overline{W}_{j,t}\} = 0$:

$$\nu_X^2(\tau_j) = \operatorname{var}\{\overline{W}_{j,t}\} = E\{\overline{W}_{j,t}^2\}$$

- can base estimator on squares of

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N}, \qquad t = 0, 1, \ldots, N-1$$

- recall that

$$\overline{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l}, \qquad t \in \mathbb{Z},$$

so $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ if 'mod N' not needed; i.e., $L_j - 1 \leq t < N$

- if $N - L_j \geq 0$, unbiased estimator of $\nu_X^2(\tau_j)$ is

$$\hat{\nu}_X^2(\tau_j) \equiv \frac{1}{N - L_j + 1} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2$$

where $M_j \equiv N - L_j + 1$

Statistical Properties of $\hat{\nu}_X^2(\tau_j)$ (Gaussian)

- suppose $\{\overline{W}_{j,t}\}$ Gaussian with mean zero & SDF $S_j(\cdot)$ (note: filtering tends to yield normality)
- suppose square integrability condition holds:

$$A_j \equiv \int_{-1/2}^{1/2} S_j^2(f)\, df < \infty \ \ \& \ \ S_j^2(f) > 0 \ \text{ almost everywhere}$$

- can show $\hat{\nu}_X^2(\tau_j)$ asymptotically normal with mean $\nu_X^2(\tau_j)$ & large sample variance $2A_j/M_j$
- meaning of square integrability condition:
  - let $s_{j,\tau} = \operatorname{cov}\{\overline{W}_{j,t}, \overline{W}_{j,t+\tau}\}$
  - if $\sum_\tau s_{j,\tau}^2 < \infty$, then $\{s_{j,\tau}\} \longleftrightarrow S_j(\cdot)$, so

$$\sum_{\tau=-\infty}^{\infty} s_{j,\tau}^2 = \int_{-1/2}^{1/2} S_j^2(f)\, df = A_j$$

  - A_j finite if autocovariance damps quickly to 0
  - if A_j infinite, usually because $S_j(f) \to \infty$ as $f \to 0$: can correct by increasing L
- conclusion: square integrability easy to satisfy
- Monte Carlo studies: large sample theory good if $M_j \geq 128$

Estimation of A_j

- in practical applications, need to estimate

$$A_j \equiv \int_{-1/2}^{1/2} S_j^2(f)\, df$$

- $S_j(\cdot)$ is SDF of $\{\overline{W}_{j,t}\}$, so estimate via periodogram:

$$\hat{S}_j^{(p)}(f) \equiv \frac{1}{M_j} \left| \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t} e^{-i2\pi f t} \right|^2$$

- statistical theory says: for $0 < |f| < 1/2$ & large N

$$\frac{2\hat{S}_j^{(p)}(f)}{S_j(f)} \stackrel{\mathrm{d}}{=} \chi_2^2,$$

  yielding (for large M_j) $\approx$ unbiased estimator:

$$\hat{A}_j \equiv \frac{1}{2} \int_{-1/2}^{1/2} [\hat{S}_j^{(p)}(f)]^2\, df = \frac{\left(\hat{s}_{j,0}^{(p)}\right)^2}{2} + \sum_{\tau=1}^{M_j-1} \left(\hat{s}_{j,\tau}^{(p)}\right)^2,$$

  where $\{\hat{s}_{j,\tau}^{(p)}\} \longleftrightarrow \hat{S}_j^{(p)}(\cdot)$:

$$\hat{s}_{j,\tau}^{(p)} \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1-|\tau|} \widetilde{W}_{j,t} \widetilde{W}_{j,t+|\tau|}, \quad 0 \le |\tau| \le M_j - 1$$

- Monte Carlo results: $\hat{A}_j$ reasonably good for $M_j \ge 128$

Confidence Intervals for $\nu_X^2(\tau_j)$: I

- for finite M_j, Gaussian-based CI problematic: lower limit of CI can very well be negative
- can avoid by basing CIs on assumption

$$\hat{\nu}_X^2(\tau_j) \equiv \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 \stackrel{\mathrm{d}}{=} a\chi_\eta^2$$

where η is equivalent degrees of freedom (EDOF)

- moment matching yields

$$\eta = \frac{2\left(E\{\hat{\nu}_X^2(\tau_j)\}\right)^2}{\operatorname{var}\{\hat{\nu}_X^2(\tau_j)\}}$$

Three Ways to Set η

- use large sample theory with appropriate estimates:

$$\eta_1 \equiv \frac{M_j \hat{\nu}_X^4(\tau_j)}{\hat{A}_j}$$

- assume nominal SDF for $\{X_t\}$: $S_X(f) = hC(f)$
  - function $C(\cdot)$ known, but level h unknown
  - in practice, $C(\cdot)$ often deduced from data (!?)
  - though questionable, get acceptable CIs using

$$\eta_2 = \frac{2\left(\sum_{k=1}^{\lfloor (M_j-1)/2 \rfloor} C_j(f_k)\right)^2}{\sum_{k=1}^{\lfloor (M_j-1)/2 \rfloor} C_j^2(f_k)}$$

- assume $S_j(\cdot)$ band-pass white noise:

$$S_j(f) = \begin{cases} h, & 1/2^{j+1} < |f| \leq 1/2^j \\ 0, & \text{otherwise}, \end{cases}$$

yielding simple (but competitive!) approach:

$$\eta_3 = \max\{M_j/2^j, 1\}$$

- Figs. 6 & 7: examples for vertical shear in the ocean

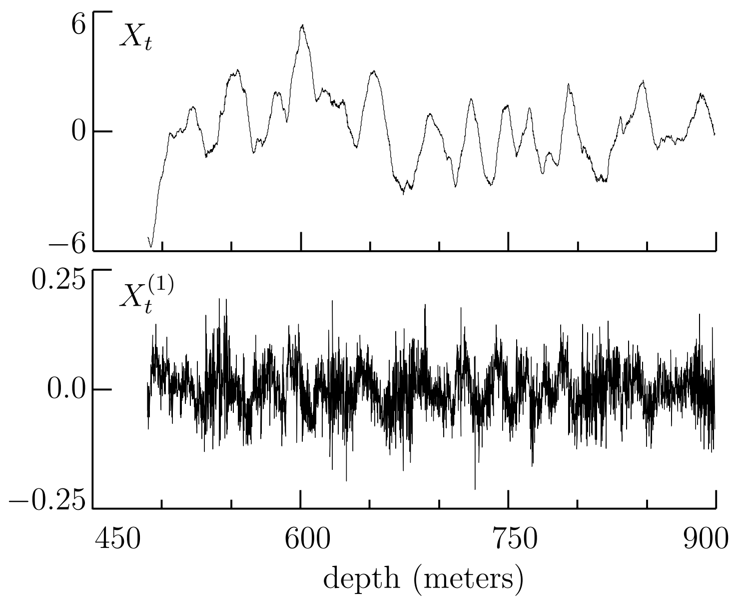

Figure 6: Vertical shear measurements and associated backward differences $\{X_t^{(1)}\}$ (taken from Figure 328, Percival and Walden, 2000, copyright Cambridge University Press).

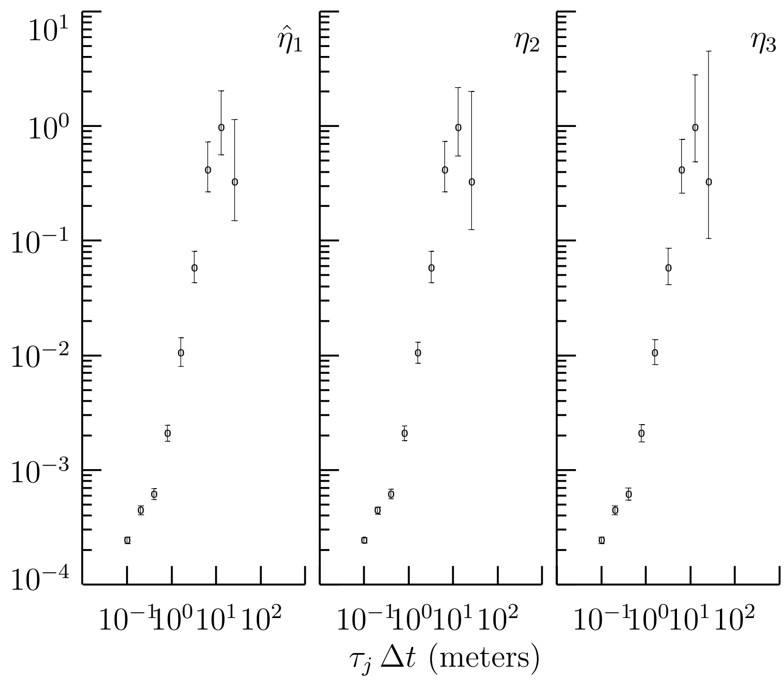

Figure 7: 95% confidence intervals for the D(6) wavelet variance for the vertical ocean shear series. The intervals are based upon χ^2 approximations to the distribution of the unbiased wavelet variance estimator with EDOFs determined by, from left to right, $\hat{\eta}_1$, η_2 using a nominal model for $S_X(\cdot)$ and η_3 (taken from Figure 333, Percival and Walden, 2000, copyright Cambridge University Press).

Statistical Properties of $\hat{\nu}_X^2(\tau_j)$ (Non-Gaussian)

- assume $\{\overline{W}_{j,t}\}$ strictly stationary process satisfying
  - $E\{\overline{W}_{j,t}\} = 0$
  - $E\{|\overline{W}_{j,t}|^{4+2\delta}\} < \infty$ for some $\delta > 0$
  - mixing condition $\alpha_{\overline{W}_j,n} = O(1/n^{\chi})$, where

    $$\alpha_{\overline{W}_j,n} \equiv \sup_{A \in \mathcal{M}_{-\infty}^{0},\, B \in \mathcal{M}_n^{\infty}} |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)|$$

    and
    - $\mathcal{M}_m^n(\overline{W}_j)$ is σ-algebra for $\overline{W}_{j,m}, \ldots, \overline{W}_{j,n}$
    - $\chi > (2+\delta)/\delta$
- let $Z_{j,t} \equiv \overline{W}_{j,t}^2$ have SDF $S_{Z_j}(\cdot)$ such that $0 < S_{Z_j}(0) < \infty$
- $\hat{\nu}_X^2(\tau_j)$ asymptotically normal with mean $\nu_X^2(\tau_j)$ & large sample variance $S_{Z_j}(0)/M_j$
- can estimate $S_{Z_j}(0)$ using standard SDF estimators
  - multitaper SDF estimator with $K = 5$ tapers
  - autoregressive SDF estimator (moderate order p)
- Fig. 8: surface albedo of spring pack ice in Beaufort Sea

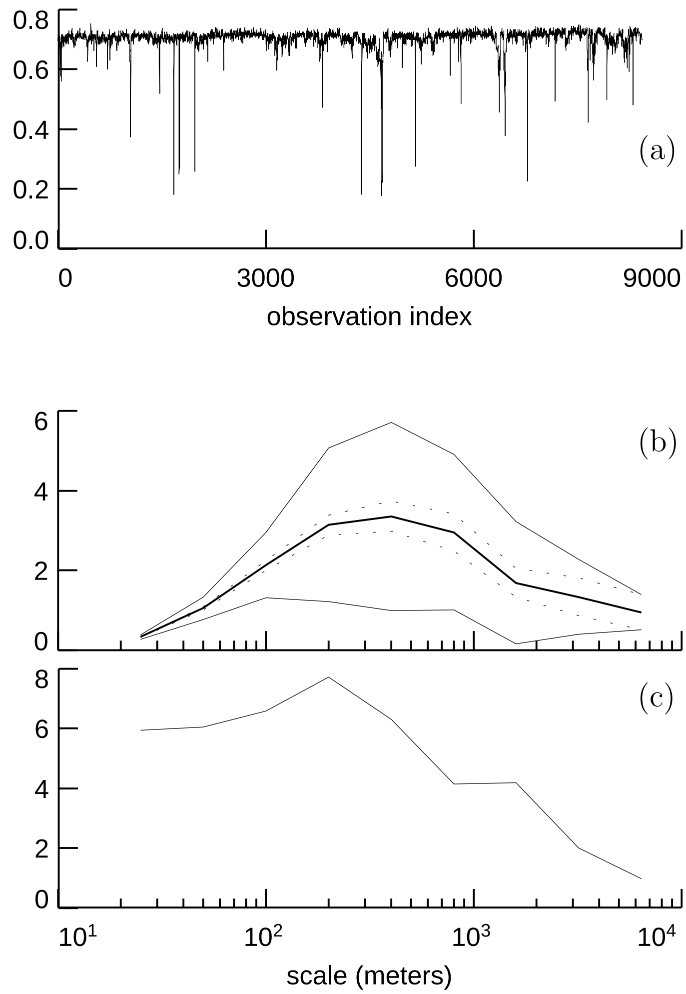

Figure 8: (a) surface albedo of pack ice in the Beaufort Sea (sampling interval between measurements is 25 meters, and there are 8428 measurements in all); (b) estimated LA(8) wavelet variance (thick solid curve), along with upper and lower 90% confidence intervals based upon Gaussian (thin dotted curves) and non-Gaussian theory (tbin solid curve); (c) ratio of estimated non-Gaussian large sample standard deviations to estimated Gaussian large sample standard deviations (adapted from Figure 3, A. Serroukh, A. T. Walden and D. B. Percival, 'Statistical Properties and Uses of the Wavelet Variance Estimator for the Scale Analysis of Time Series,' *Journal of the American Statistical Association*, **95**, pp. 184–96).

Wavelet Variance Analysis of Time Series with Time-Varying Statistical Properties

- each wavelet coefficient $\widetilde{W}_{j,t}$ formed using portion of $\{X_t\}$
- suppose X_t associated with actual time $t_0 + t$
  (t_0 is actual time of first observation X_0)
- suppose $\{\tilde{h}_{j,l}\}$ is Haar or 'least asymmetric' Daubechies wavelet
- can associate $\widetilde{W}_{j,t}$ with actual time interval of form

$$[t_0 + t' - \tau_j, t_0 + t' + \tau_j]$$

- can thus form 'localized' wavelet variance analysis (implicitly assumes stationarity or stationary increments locally)
- Fig. 9: annual minima of Nile River
- Figs. 10, 11 & 12: subtidal sea level fluctuations

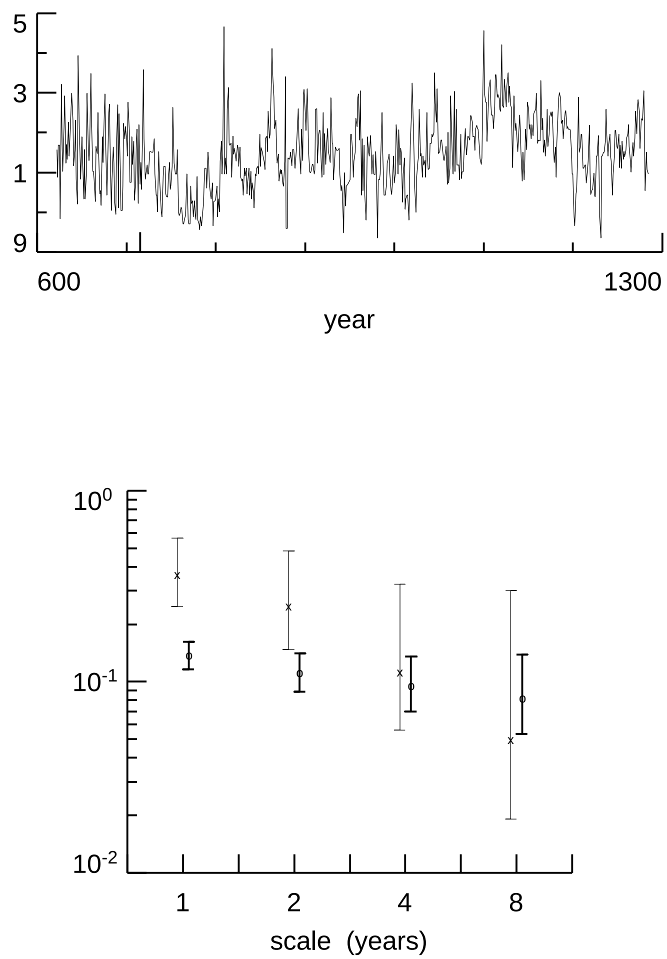

Figure 9: Nile River yearly minima (top plot), along with estimated Haar wavelet variance before and after year 715.5 (`x`'s and `o`'s, respectively) and 95% confidence intervals (thin and thick lines, respectively) based upon a chi-square approximation with EDOFs determined by η_3 (taken from Figures 192 and 327, Percival and Walden, 2000, copyright Cambridge University Press).

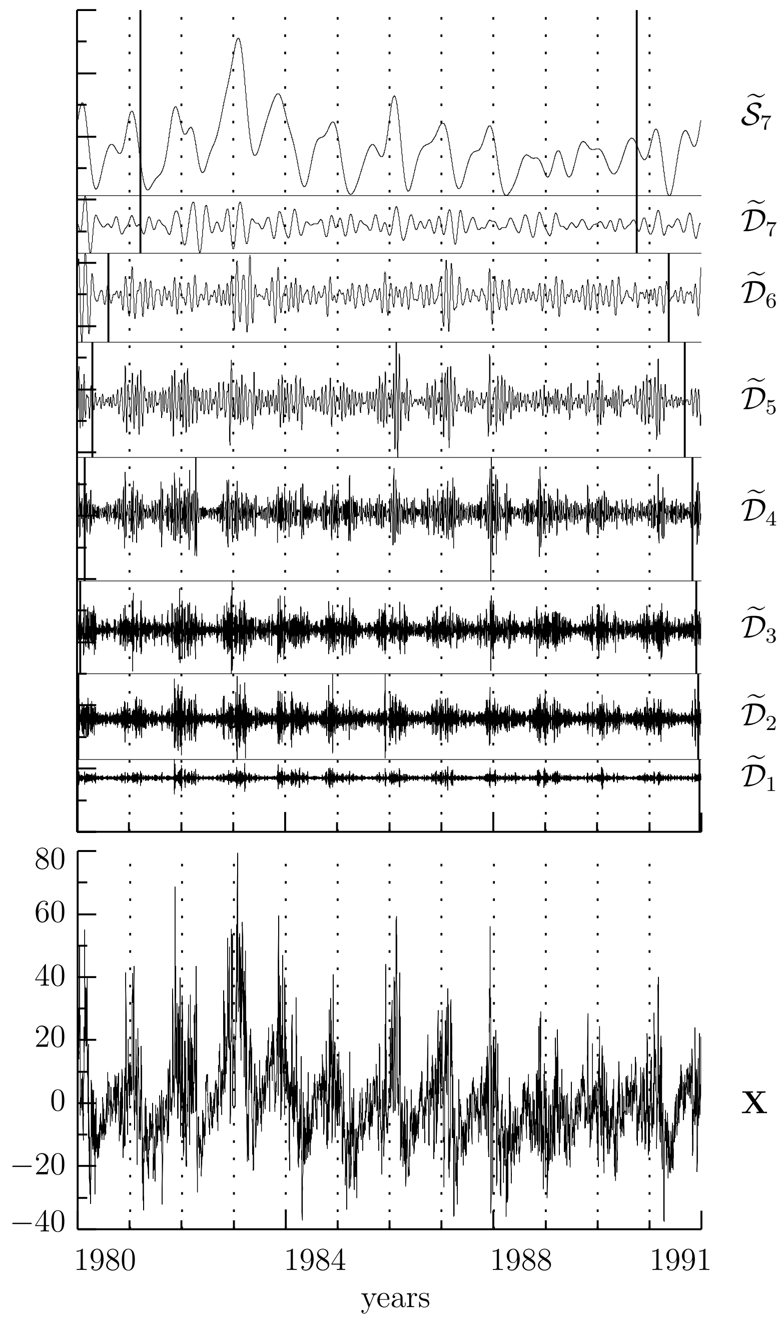

Figure 10: LA(8) MODWT multiresolution analysis for Crescent City subtidal variations measured in centimeters (taken from Figure 186, Percival and Walden, 2000, copyright Cambridge University Press).

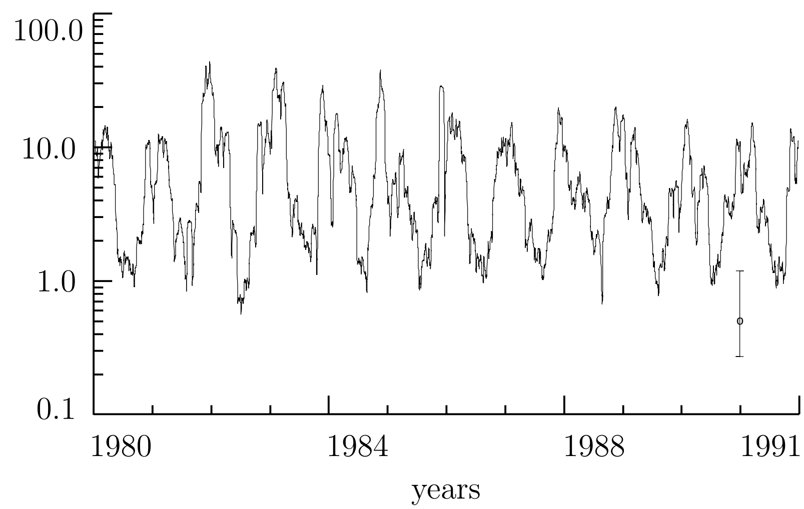

Figure 11: Estimated time-dependent LA(8) wavelet variances for physical scale $\tau_2 \Delta t = 1$ day for the Crescent City subtidal sea level variations, along with a representative 95% confidence interval based upon a hypothetical wavelet variance estimate of 1/2 and a chi-square distribution with $\nu = 15.25$ (taken from Figure 324, Percival and Walden, 2000, copyright Cambridge University Press).

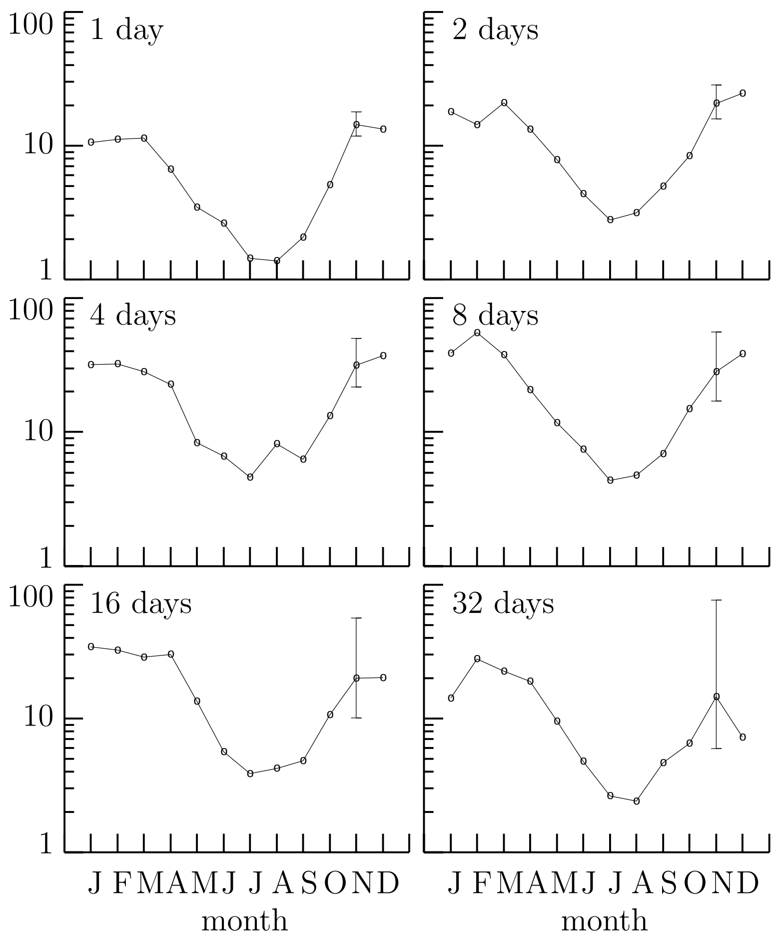

Figure 12: Estimated LA(8) wavelet variances for physical scales $\tau_j \, \Delta t = 2^{j-2}$ days, $j = 2, \ldots, 7$, grouped by calendar month for the subtidal sea level variations (taken from Figure 326, Percival and Walden, 2000, copyright Cambridge University Press).

Summary

- wavelet variance gives scale-based analysis of variance (natural match for many geophysical processes)
- statistical theory worked out for
  - Gaussian processes with stationary backward differences
  - non-Gaussian processes satisfying a mixing condition
- applications include analysis of
  - genome sequences
  - frequency fluctuations in atomic clocks
  - changes in variance of soil properties
  - canopy gaps in forests
  - accumulation of snow fields in polar regions
  - turbulence in atmosphere and ocean
  - regular and semiregular variables stars