Estimation of the Wavelet Variance Using Reflection Boundary Conditions

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

- definition and basic properties of wavelet variance
- application to fractionally differenced processes
- estimation of wavelet variance via discrete wavelet transform
  - unbiased estimator
  - biased estimator
- estimation based upon reflection boundary conditions
- example: ocean shear measurements
- conclusions and future research

Definition of Wavelet Variance: I

- let $\{X_t : t \in \mathbb{Z}\}$ be a zero mean stochastic process, where $\mathbb{Z}$ is the set of all integers
- assume that $\{X_t\}$ has stationary backward differences; i.e.,

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

  forms a stationary process, where
    * d is a nonnegative integer
        * $\star$ $d = 0$ implies $\{X_t\}$ is stationary
        * $\star$ if $d = 1$, Y_t is output from first difference filter
    * B is backward shift operator: $BX_t \equiv X_{t-1}$ & $B^k X_t = X_{t-k}$
- let $\{\tilde{h}_{1,l}, l \in \mathbb{Z}\}$ be scale 1 Daubechies wavelet filter of width L
    - 'width L' implies
        * $\tilde{h}_{1,l} = 0$ when $l < 0$ or $l \geq L$,
        * $\tilde{h}_{1,0} \neq 0$ and $\tilde{h}_{1,L-1} \neq 0$
    - $\sum_l \tilde{h}_{1,l} = 0$
    - $\sum_l \tilde{h}_{1,l}^2 = 1/2$
    - $\sum_l \tilde{h}_{1,l}\tilde{h}_{1,l+2k} = 0$ for nonzero integers k
    - L must be an even integer
    - equivalent to using $L/2$ first difference filters & smoothing filter of width $L/2$

Definition of Wavelet Variance: II

- let $\{\tilde{h}_{j,l}\}$ be scale $\tau_j = 2^{j-1}$ wavelet filter, $j = 2, 3, \ldots$
  - 'scale' is effective half-width of $\{\tilde{h}_{j,l}\}$
  - $\{\tilde{h}_{j,l}\}$ 'stretched out' version of scale 1 filter $\{\tilde{h}_{1,l}\}$
  - actual width of $\{\tilde{h}_{j,l}\}$ is $L_j = (2^j - 1)(L - 1) + 1$
- Fig. 1: Haar and $L = 8$ 'least asymmetric' Daubechies filters (henceforth LA(8))
- assume $L/2 \geq d$ & filter $\{X_t\}$ to create new stochastic process

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

  called scale τ_j wavelet coefficients
- $\{\overline{W}_{j,t}\}$ is a zero mean stationary process with variance

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

  known as the scale τ_j wavelet variance (or wavelet spectrum)

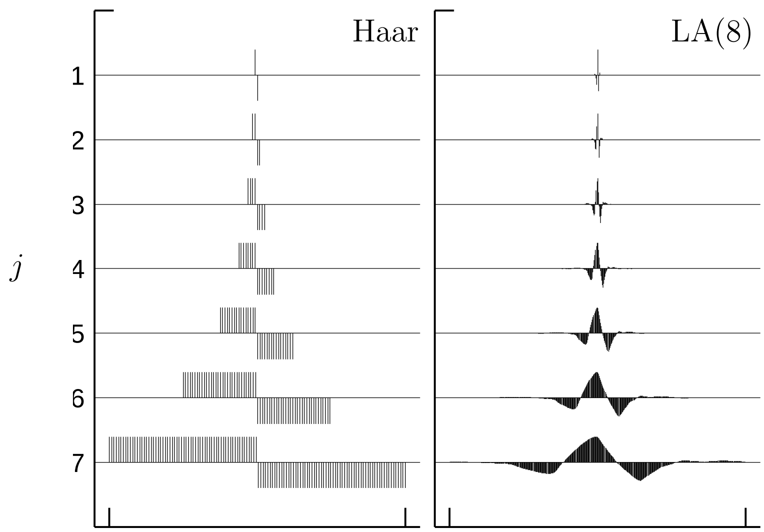

Figure 1: Haar and LA(8) wavelet filters $\{\tilde{h}_{j,l}\}$ for scales indexed by $j = 1, 2, \ldots, 7$.

Basic Properties of Wavelet Variance: I

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

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

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

- if $\{X_t\}$ nonstationary, then

$$\sum_{j=1}^{\infty} \nu_X^2(\tau_j) = \infty$$

- in either case, $\nu_X^2(\tau_j)$ is contribution to $\mathrm{var}\,\{X_t\}$ due to scale τ_j

Basic Properties of Wavelet Variance: II

- let $S_X(\cdot)$ be spectral density function (SDF) for $\{X_t\}$ (well-defined for processes with stationary increments)
- if $\{X_t\}$ stationary process, then
$$\int_{-1/2}^{1/2} S_X(f)\,df = \mathrm{var}\,\{X_t\}$$
i.e., decomposes $\mathrm{var}\,\{X_t\}$ across frequencies f
- if $\{X_t\}$ nonstationary, then
$$\int_{-1/2}^{1/2} S_X(f)\,df = \infty$$
- $\{\tilde{h}_{j,l}\} \approx$ bandpass over $|f| \in [1/2^{j+1}, 1/2^j]$ and hence
$$\nu_X^2(\tau_j) \approx 2\int_{1/2^{j+1}}^{1/2^j} S_X(f)\,df$$
- as $L \to \infty$,
  - $\{\tilde{h}_{j,l}\} \to$ ideal bandpass filter
  - $\mathrm{cov}\,\{\overline{W}_{j,t}, \overline{W}_{j',t'}\} \to 0$ for all $j \neq j'$
    (i.e., asymptotic 'between-scale' decorrelation)

Fractionally Differenced (FD) Processes: I

- will consider wavelet variance for FD processes as examples
- if $\{X_t\}$ has SDF $S_X(\cdot)$, then $Y_t = X_t - X_{t-1}$ has SDF
$$S_Y(f) = 4\sin^2(\pi f) S_X(f)$$
since $4\sin^2(\pi f)$ is squared gain function for first difference filter
- $\{X_t\}$ called FD(δ) process if it possesses SDF given by
$$S_X(f) = \sigma_\varepsilon^2 [4\sin^2(\pi f)]^{-\delta}, \quad |f| \le 1/2$$
where $\sigma_\varepsilon^2 > 0$ and $-\infty < \delta < \infty$
- if $\delta < 1/2$, $\{X_t\}$ is stationary with autocovariance sequence $s_{X,\tau} = \operatorname{cov}\{X_t, X_{t+|\tau|}\}$ given by
$$s_{X,0} = \frac{\sigma_\varepsilon^2 \Gamma(1-2\delta)}{\Gamma^2(1-\delta)} \text{ and } s_{X,\tau} = s_{X,\tau-1} \frac{\tau+\delta-1}{\tau-\delta}, \;\; \tau = 1, 2, \ldots$$
- if $\delta \ge 1/2$, $\{X_t\}$ is nonstationary process with dth order stationary backward differences $\{Y_t\}$
  - $d = \lfloor \delta + 1/2 \rfloor$, where $\lfloor x \rfloor$ is integer part of x
  - $\{Y_t\}$ is stationary FD($\delta - d$) process
- if $\delta < 0$, FD process is antipersistent
- if $\delta = 0$, FD process becomes white noise
- if $\delta > 0$, FD process has 'long memory'
- if $\delta = 1$, FD process is random walk (sampled Brownian motion)

Fractionally Differenced (FD) Processes: II

- at low (small) frequencies f,

$$S_X(f) = \sigma_\varepsilon^2[4\sin^2(\pi f)]^{-\delta} \approx \sigma_\varepsilon^2[2\pi f]^{-2\delta},$$

  i.e., an approximate power-law

- at large scales, thus have

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

- since

$$\log\left(\nu_X^2(\tau_j)\right) \approx \log\left(C\right) + (2\delta-1)\log\left(\tau_j\right),$$

  log/log plot of $\nu_X^2(\tau_j)$ vs. τ_j looks approximately linear with slope $2\delta - 1$ for τ_j large enough

- Fig. 2: $\nu_X^2(\tau_j)$ & sample realizations for four FD processes

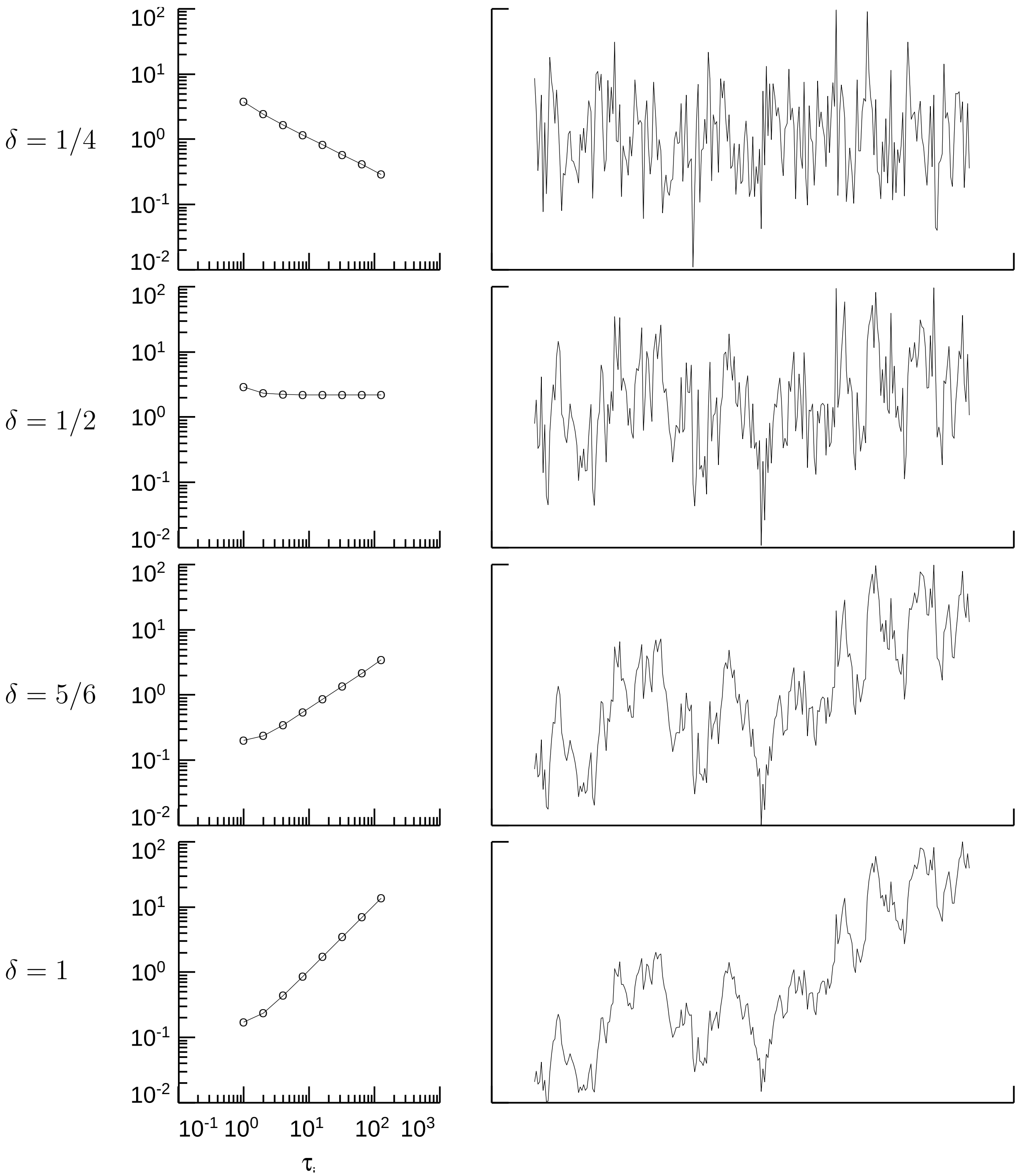

Figure 2: LA(8) wavelet variances $\nu_X^2(\tau_j)$, $j = 1, \ldots, 8$, for four FD(δ) processes (left-hand column), along with one realization of length $N = 256$ from each processs generated by the circulant embedding method using the same set of $2N = 512$ standard Gaussian random deviates (right-hand).

Maximal Overlap Discrete Wavelet Transform

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ be a time series (i.e., part of $\{X_t\}$)
- for $j = 1, \ldots, J_0$, form MODWT wavelet coefficients

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

  where $X_{-1 \bmod N} = X_{N-1}$, $X_{-2 \bmod N} = X_{N-2}$, etc
  (note: actually computed via an efficient 'pyramid' algorithm)
- let $\widetilde{\mathbf{W}}_j = [\widetilde{W}_{j,0}, \widetilde{W}_{j,1}, \ldots, \widetilde{W}_{j,N-1}]^T$
- also form vector $\widetilde{\mathbf{V}}_{J_0}$ of MODWT 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. 3: Haar & LA(8) scaling filters $\{\tilde{g}_{J_0,l}\}$
  - $\widetilde{V}_{J_0,t}$ is weighted average over scale $2\tau_j$
- obtain 'scale by scale' 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 $N = 2^{J_0}$, then $\|\widetilde{\mathbf{V}}_{J_0}\|^2 / N = \overline{X}^2$

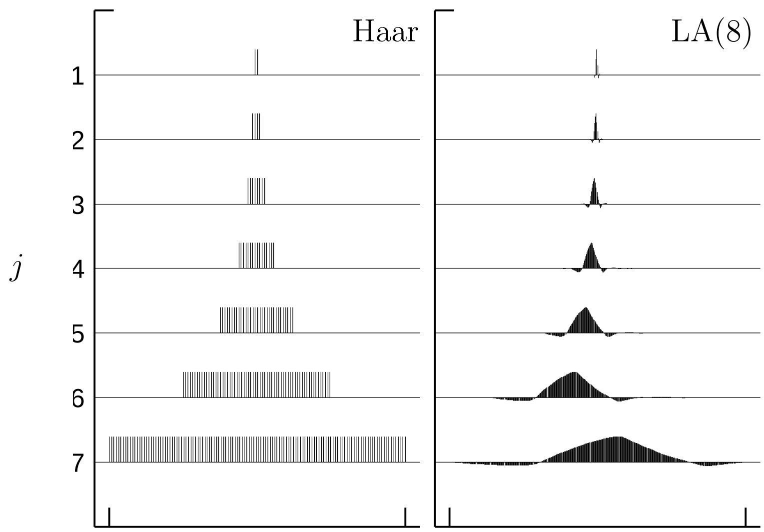

Figure 3: Haar and LA(8) scaling filters $\{\tilde{g}_{J_0,l}\}$ for scales indexed by $J_0 = 1, 2, \ldots, 7$.

Unbiased Estimator of Wavelet Variance

- recall that $\nu_X^2(\tau_j) = \text{var}\,\{\overline{W}_{j,t}\} = E\{\overline{W}_{j,t}^2\}$
- compare MODWT coefficients

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

to

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

- $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ 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 $\hat{\nu}_X^2(\tau_j)$ tractable, but ...
  - for $L \geq 4$ and large j, filter width $L_j = (2^j - 1)(L - 1) + 1$ approximate $L - 1$ times longer than for $L = 2$ (i.e., Haar)
  - Fig. 4: effective width of $\{\tilde{h}_{j,l}\}$ is $2\tau_j$ for all L
- Q: can we use profitably use $\widetilde{W}_{j,t}^2$, $j = 0, \ldots, L_j - 2$?

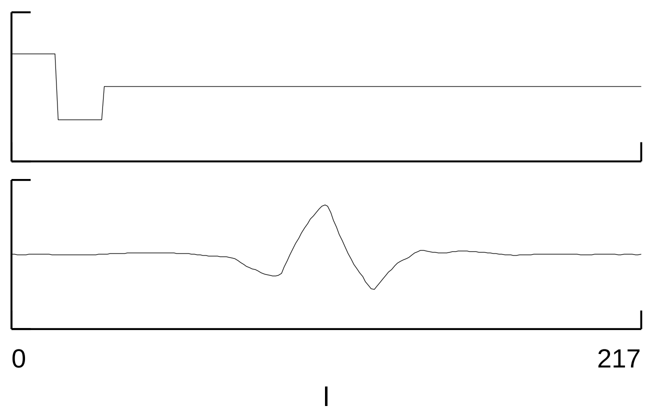

Figure 4: Haar wavelet filter $\{h_{5,l}\}$ for scale $\tau_5 = 16$ (top plot) and corresponding LA(8) wavelet filter (bottom). The actual widths of the Haar and LA(8) filters are $L_5 = 32$ and $L_5 = 218$. Their effective widths are both 32.

Biased Estimator of Wavelet Variance

- can construct 'biased' estimator of $\nu_X^2(\tau_j)$:

$$\tilde{\nu}_X^2(\tau_j) \equiv \frac{1}{N}\|\widetilde{\mathbf{W}}_j\|^2 = \frac{1}{N}\left(\sum_{t=0}^{L_j-2} \widetilde{W}_{j,t}^2 + \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2\right)$$

- biased estimator offers exact analysis of $\hat{\sigma}_X^2$
- if $\{X_t\}$ stationary, bias goes to 0 as $N \rightarrow \infty$; not true in general if $\{X_t\}$ nonstationary
- Fig. 5: $E\{\tilde{\nu}_X^2(\tau_j)\}$ for LA(8) wavelet and $N = 256$
- problem: possible large mismatch between X_0 & X_{N-1} (cf. Fig. 2)

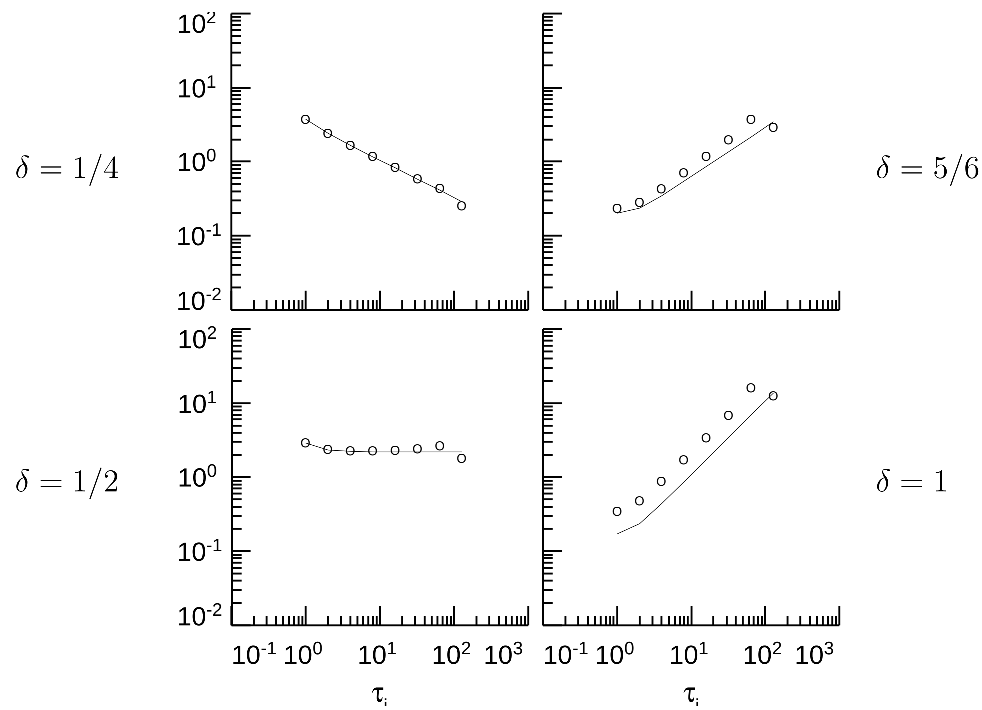

Figure 5: LA(8) wavelet variances $\nu_X^2(\tau_j)$ for four FD(δ) processes (thin curves) and corresponding $E\{\tilde{\nu}_X^2(\tau_j)\}$ for $N = 256$ (circles) versus τ_j for $j = 1, \ldots, 8$.

Reflection Boundary Conditions

- construct 2nd biased estimator using idea from Fourier analysis
- extend $X_0, \ldots, X_{N-1}$ to length $2N$ by 'reflection':

$$X'_t = \begin{cases} X_t, & t = 0, \ldots, N-1; \\ X_{2N-1-t}, & t = N, \ldots, 2N-1 \end{cases}$$

- Fig. 6: examples of reflected series
- $2N$ series has same sample mean & variance
- let $\widetilde{\mathbf{W}}'_j$ denote wavelet coefficients of $X'_0, \ldots, X'_{2N-1}$
- second biased estimator of $\nu^2_X(\tau_j)$ is thus

$$\tilde{\nu}^2_{X'}(\tau_j) \equiv \frac{1}{2N} \|\widetilde{\mathbf{W}}'_j\|^2$$

- Fig. 7: $E\{\tilde{\nu}^2_{X'}(\tau_j)\}$ for LA(8) wavelet and $N = 256$
- Fig. 8: root mean square errors for $\hat{\nu}^2_X(\tau_j)$ and $\tilde{\nu}^2_{X'}(\tau_j)$

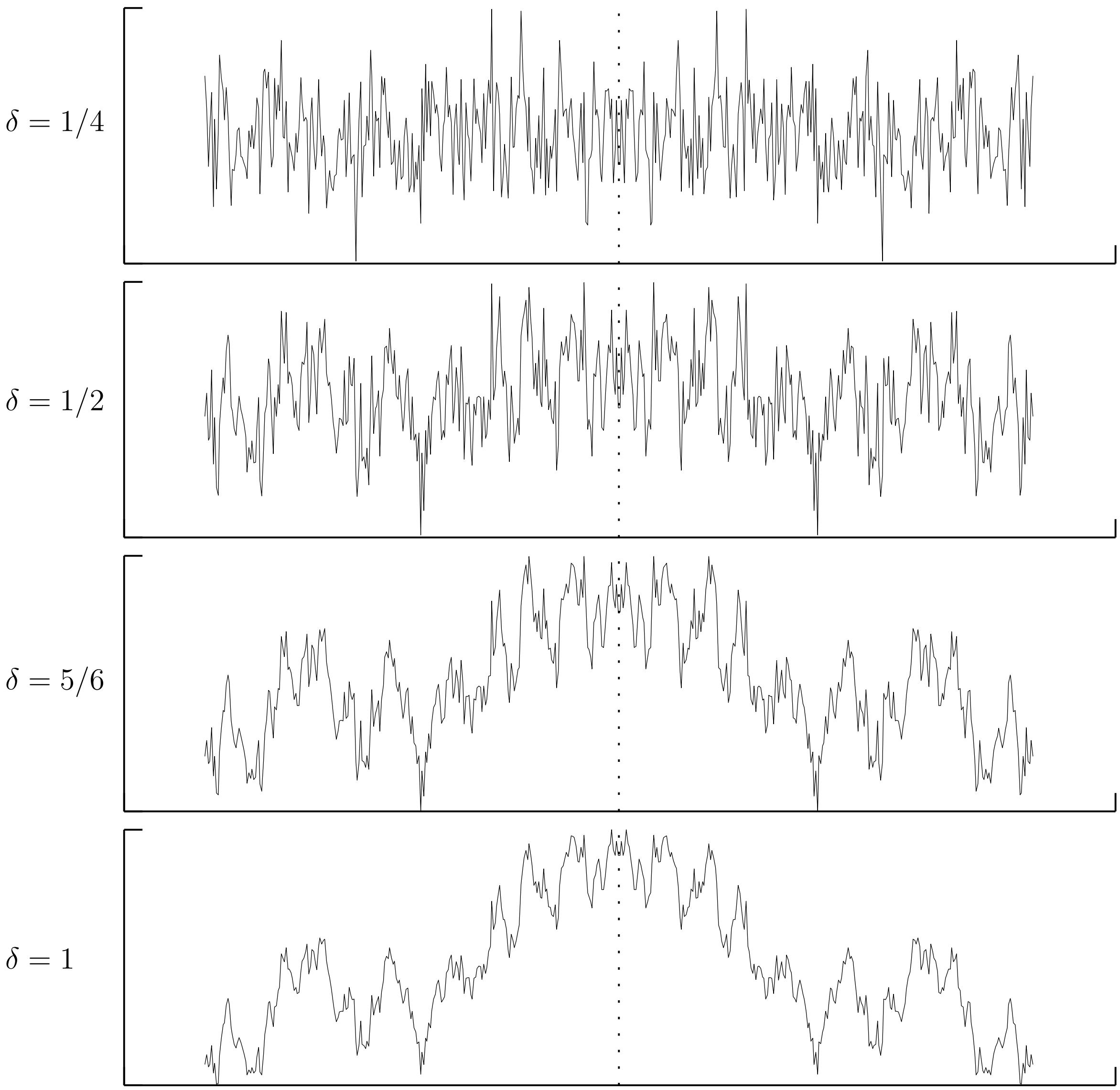

Figure 6: Realizations of length $N = 256$ from four FD(δ) processes extended to length $N = 512$ by tacking a time-reversed version of the original series onto the end of the series.

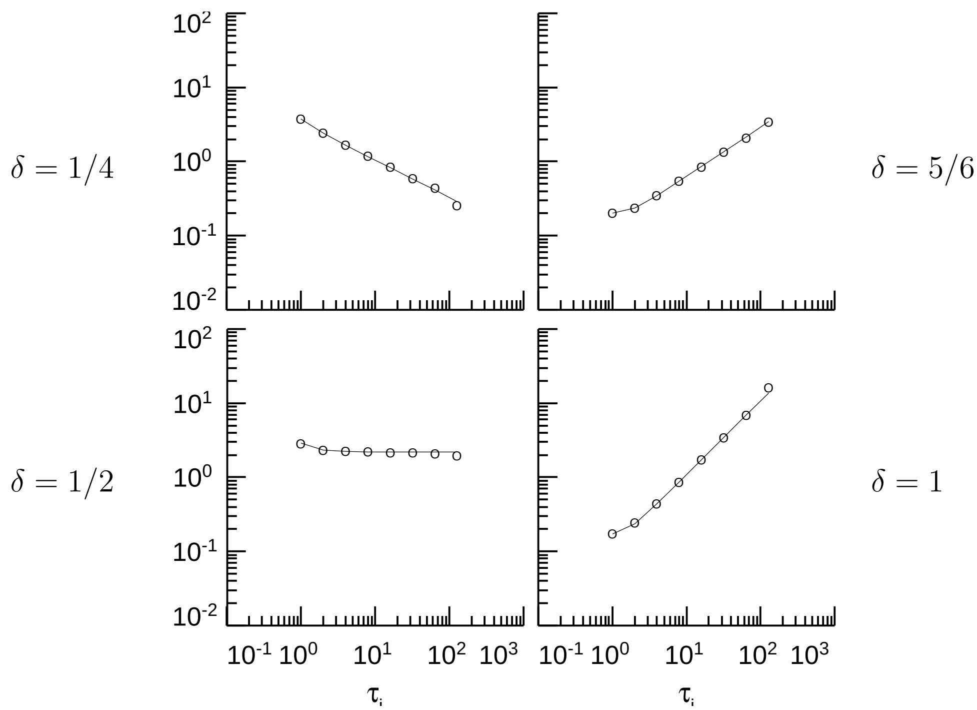

Figure 7: LA(8) wavelet variances $\nu_X^2(\tau_j)$ for four FD(δ) processes (thin curves) along with $E\{\tilde{\nu}_{X'}^2(\tau_j)\}$ for $N = 256$ (circles) versus τ_j for $j = 1, \ldots, 8$.

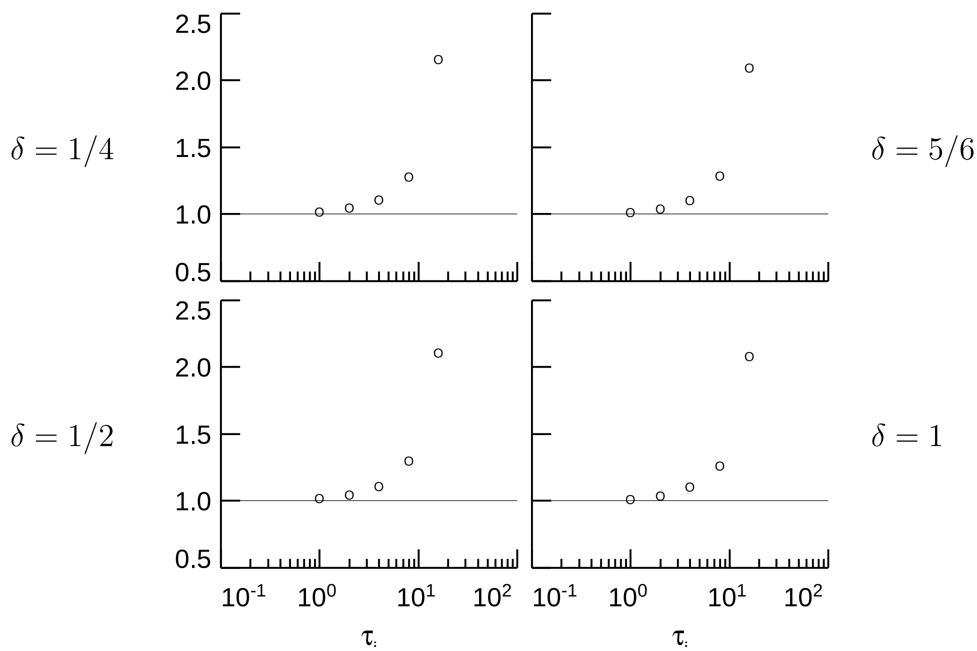

Figure 8: Ratios of root mean square error (rmse) for unbiased estimator $\hat{\nu}^2_X(\tau_j)$ to rmse for biased estimator $\tilde{\nu}^2_{X'}(\tau_j)$ of LA(8) wavelet variance for $N = 256$ and $j = 1, \ldots, 5$ (circles).

Vertical Ocean Shear Measurements

- Fig. 9: plot of depth series and its first difference $Y_t = X_t - X_{t-1}$
- data sampled vertically every $\Delta t = 0.1$ meters
- Fig. 10: three wavelet variance estimates
  - `x`'s: unbiased Haar estimates $\hat{\nu}_X^2(\tau_j)$ up to $\tau_{12}\,\Delta t$
  - `o`'s: unbiased LA(8) estimates $\hat{\nu}_X^2(\tau_j)$ up to $\tau_9\,\Delta t$
  - `+`'s: biased LA(8) estimates $\tilde{\nu}_{X'}^2(\tau_j)$ up to $\tau_{12}\,\Delta t$
  - `*` shows 'remainder variance' for biased LA(8) estimate: since

$$\hat{\sigma}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}\left(X_t - \overline{X}\right)^2 = \frac{1}{2N}\left(\sum_{j=1}^{12}\|\widetilde{\mathbf{W}}'_j\|^2 + \|\widetilde{\mathbf{V}}'_{12}\|^2\right) - \overline{X}^2$$

remainder variance given by

$$\frac{1}{2N}\|\widetilde{\mathbf{V}}'_{12}\|^2 - \overline{X}^2$$

  - associated with averages over scale $2\tau_{12}\,\Delta = 409.6$ meters (equal to total length of original series)
  - accounts for 1.3% of total variance here

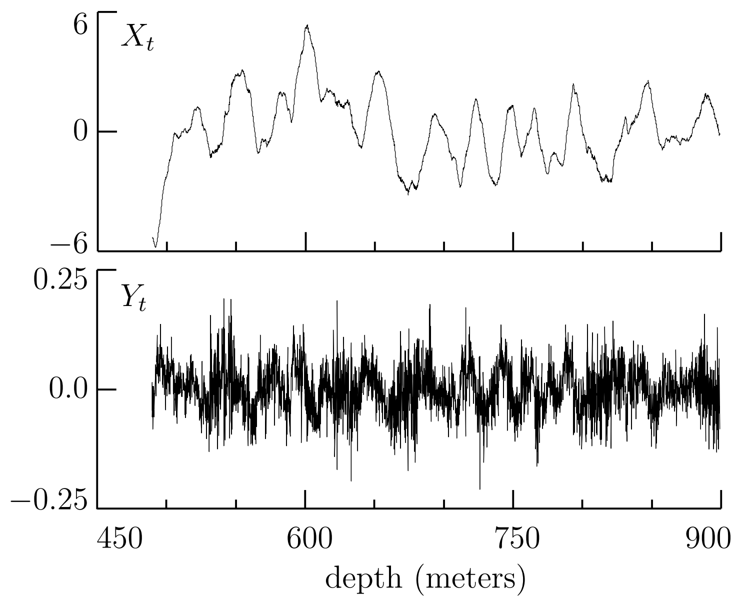

Figure 9: $N = 4096$ vertical shear measurements $\{X_t\}$ (top plot) and associated backward differences $\{Y_t\}$ (bottom).

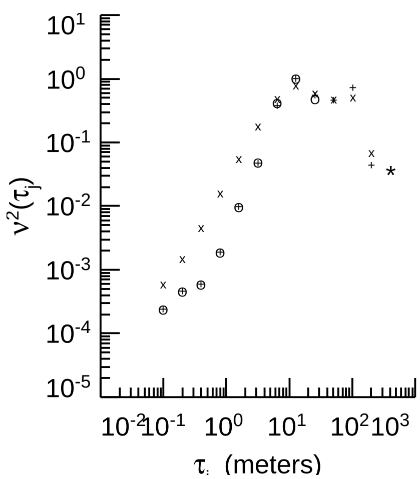

Figure 10: Wavelet variances estimates for vertical shear series. The `x`'s indicate the unbiased Haar estimates $\hat{\nu}_X^2(\tau_j)$; the `o`'s, the unbiased LA(8) estimates $\hat{\nu}_X^2(\tau_j)$; and the `+`'s, the biased LA(8) estimates $\tilde{\nu}_{X'}^2(\tau_j)$. The `*` indicates the 'remainder variance' for the biased LA(8) estimate.

Conclusions and Future Research

- reflection boundary conditions gives viable estimator of $\nu_X^2(\tau_j)$
  - need to assess performance outside of FD processes
  - need to develop large sample theory
  - need to look into question of polynomial drift
- other potential uses for reflection boundary conditions
  - wavelet-based bootstrapping
  - alternative to tapering in spectral analysis?