# Assessing Arctic Sea Ice Thickness using Wavelets

Don Percival

Applied Physics Laboratory
University of Washington, Seattle

overheads for talk available at

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

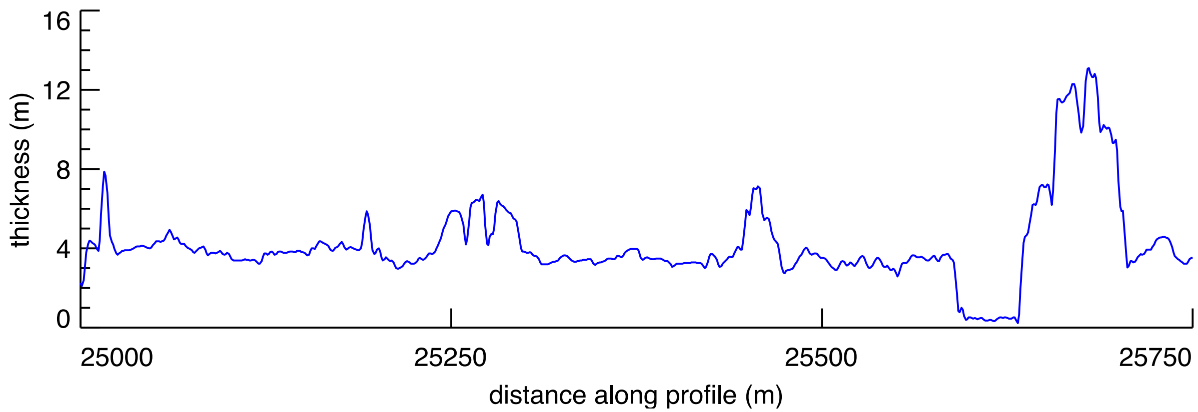
# Overview

- thickness of Arctic sea ice is subject of ongoing scientific interest
- seminar will discuss
  - role of thickness distributions
  - data available to evaluate changes in distributions
  - simplification based upon notion of ice types
  - Haar wavelet variance and its potential role in
    - evaluating changes
    - providing additional insights into Arctic sea ice variability
- work in progress with Yanling Yu, Polar Science Center, APL (future work will involve Mike Keim, QERM)

# Ice Thickness Distributions

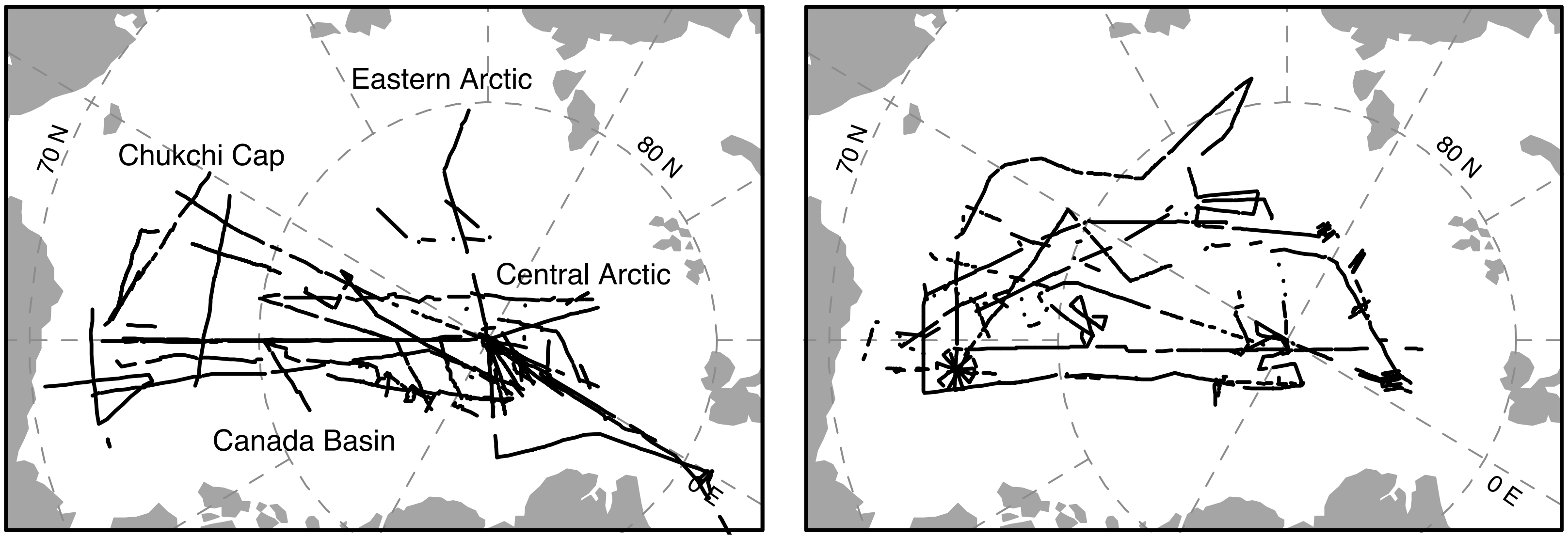
- basic description of Arctic sea ice is distribution of its thickness
- one distribution for entire Arctic basin of less interest than ones
  - localized to specific regions
  - limited to certain time spans
- good physical reasons to hypothesize that distributions are
  - not uniform in space
  - not invariant with time
- inhomogeneities in space and time due to many factors
  - thermodynamic processes control ice growth and melt
  - winds and currents drive ice motion, creating pressure ridges or areas of open water and thin ice

# Ice Thickness Profiles



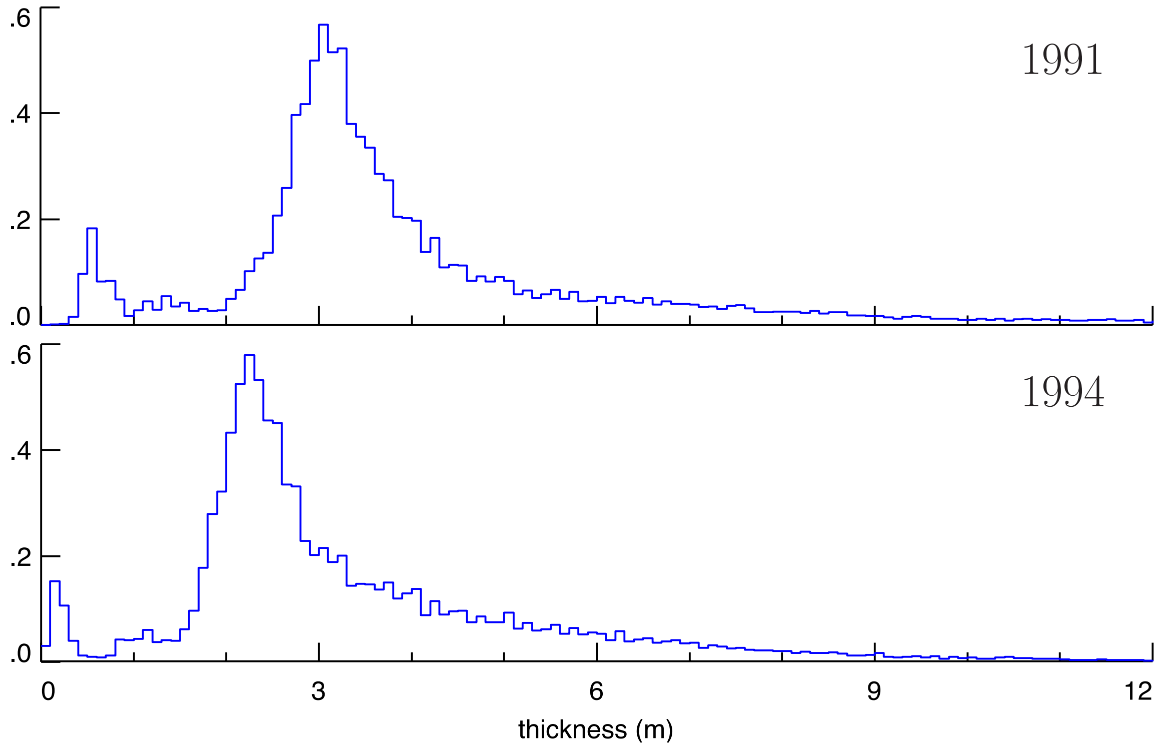

- 750 meter portion of ice thickness profile $h(l)$ measured by a submarine near North Pole in April, 1991
- $l$ is distance in meters along submarine's transect
- measurements are ice drafts, converted to thickness ($\times 1.12$)
- entire profile extends over 50 km (one data point each meter)

# Submarine Tracks 1976–97



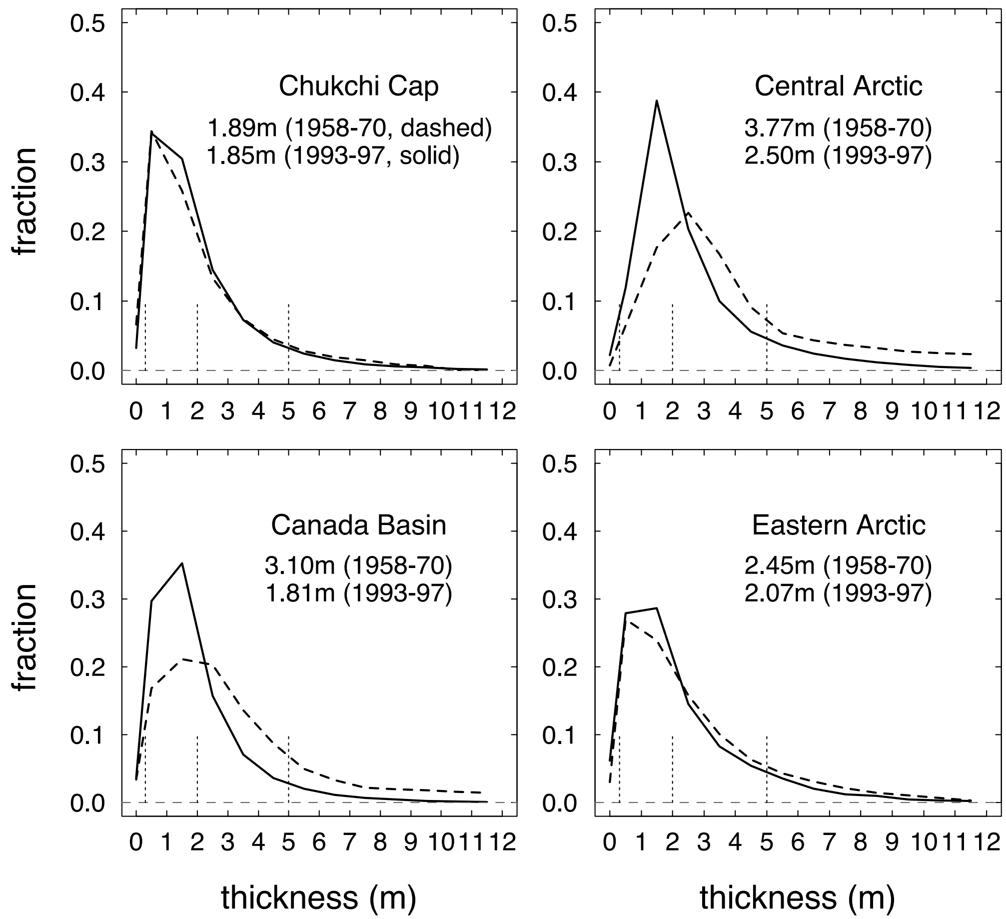

- left-hand plot shows spring (April–May) cruises
- right-hand plot shows fall (Sept–Oct) cruises
- possibility of more data eventually becoming available

# Estimated Ice Thickness Distributions: I




- above give estimates of ice thickness distributions $g(h)$ based on transects near the North Pole, suggesting a thinning of ice from 1991 to 1994, but are differences statistically significant?

# Estimated Ice Thickness Distributions: II

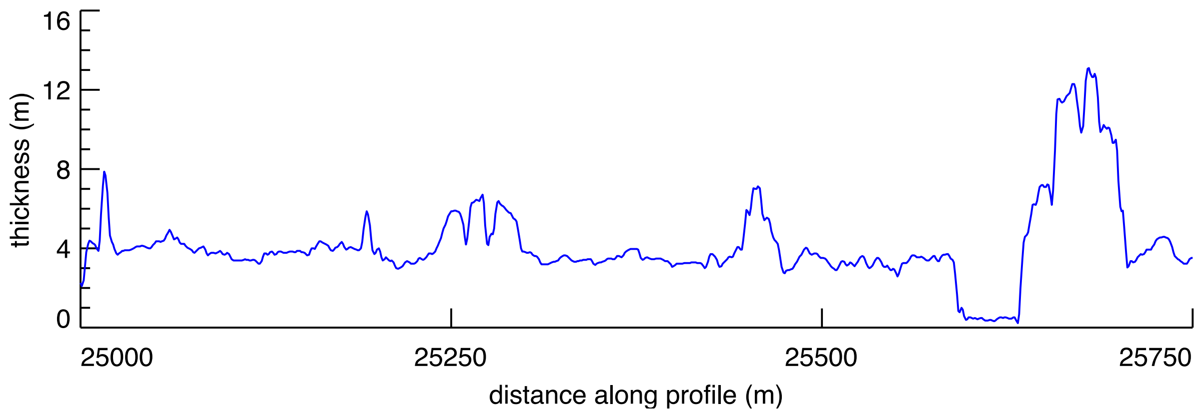



- 8 more estimated distributions (4 regions & 2 time periods)
- note: not well-modeled by normal (Gaussian) distribution

## Statistical Assessment of Thickness Distributions: I

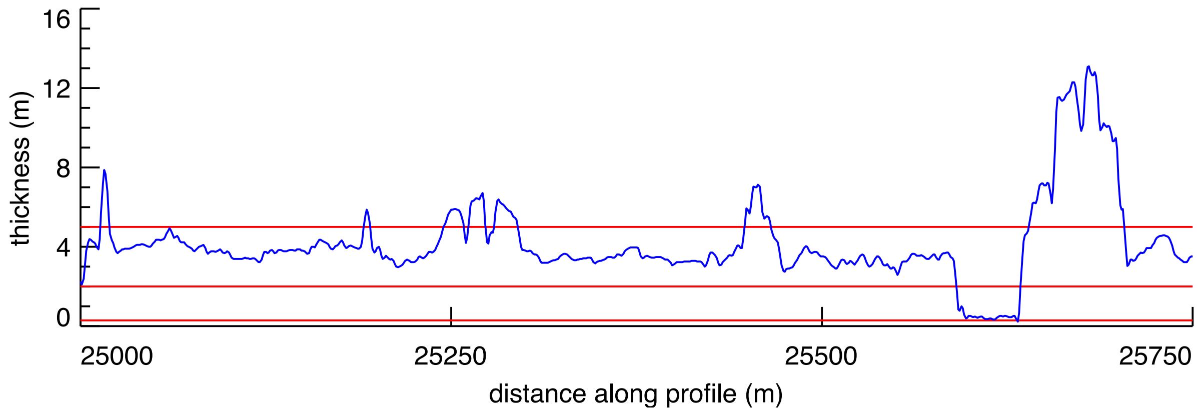
- while visual differences are suggestive, little definitive is known about spatial-temporal properties of ice thickness distributions
- need to understand if observed differences are due primarily to
  1. spatial variability that is consistent from year to year
  2. spatial variability that is changing significantly over time and hence represents important climate changes or
  3. sampling variations that are consistent with the amount of available data
- standard statistical methodology for evaluating null hypothesis that two samples of data come from same distribution assumes independent samples

# Statistical Assessment of Thickness Distributions: II




- independence assumption unrealistic – profiles $h(l)$ are highly autocorrelated
- since histograms are markedly non-Gaussian, have difficult statistical problem of assessing changes in estimated distributions for correlated data from unknown null distribution $g(h)$

# Simplification via Ice Type Distributions: I



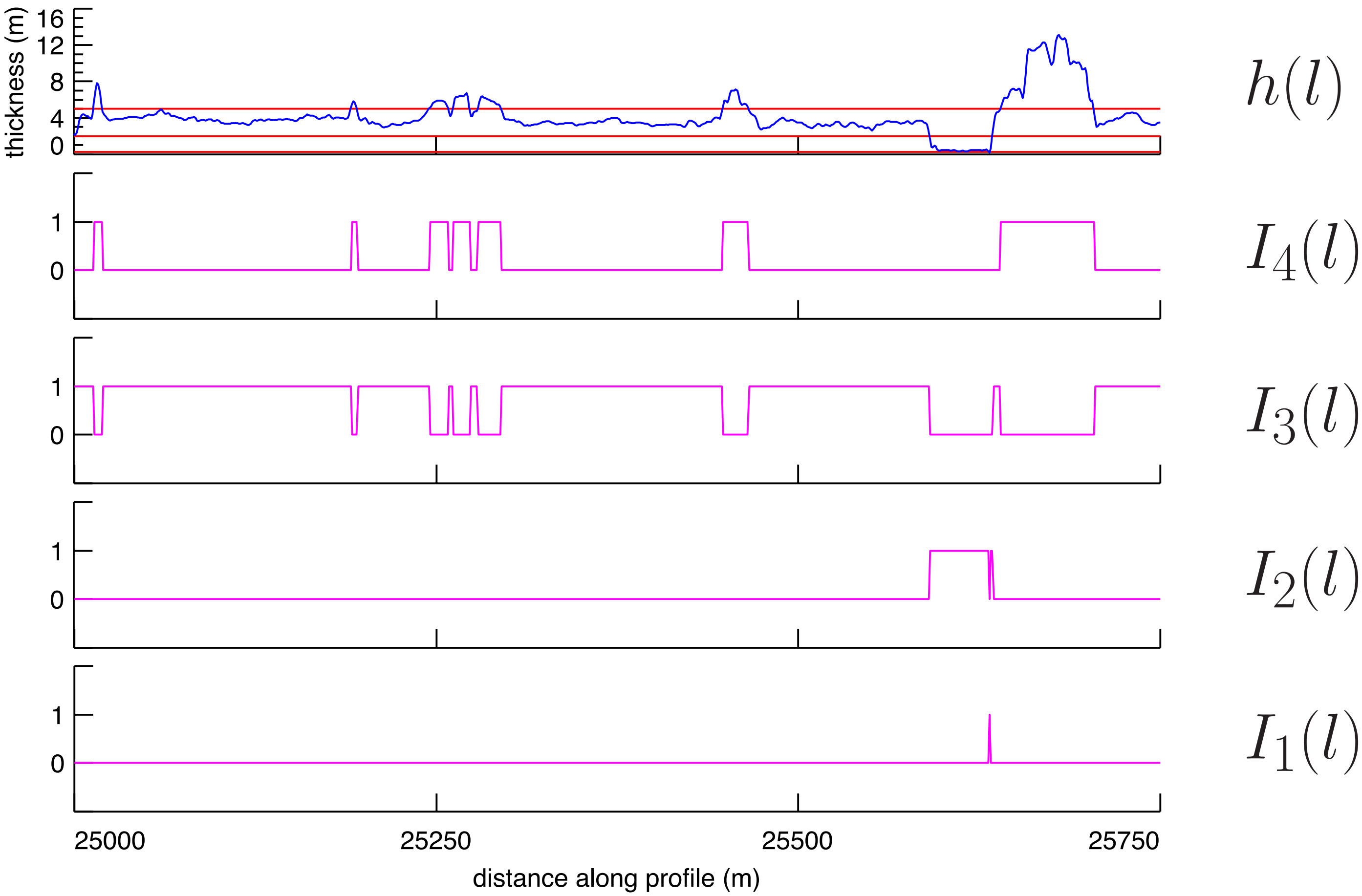

- can simplify problem considerably by binning data according to physically meaningful ice types:
  1. leads and new ice, for which $h(l) < 0.3$ m;
  2. first year ice, for which $0.3 \text{ m} \leq h(l) < 2$ m;
  3. medium multiyear ice, for which $2 \text{ m} \leq h(l) < 5$ m; and
  4. ridged ice, for which $h(l) \geq 5$ m
- horizontal red lines indicate binning into ice types

## Simplification via Ice Type Distributions: II

- let $h_i^{(l)}$ and $h_i^{(u)}$ be the lower and upper limits for the $i$th ice type (with $h_1^{(l)} = 0$ and $h_4^{(u)} = \infty$)
- define an indicator series for $i$th ice type as

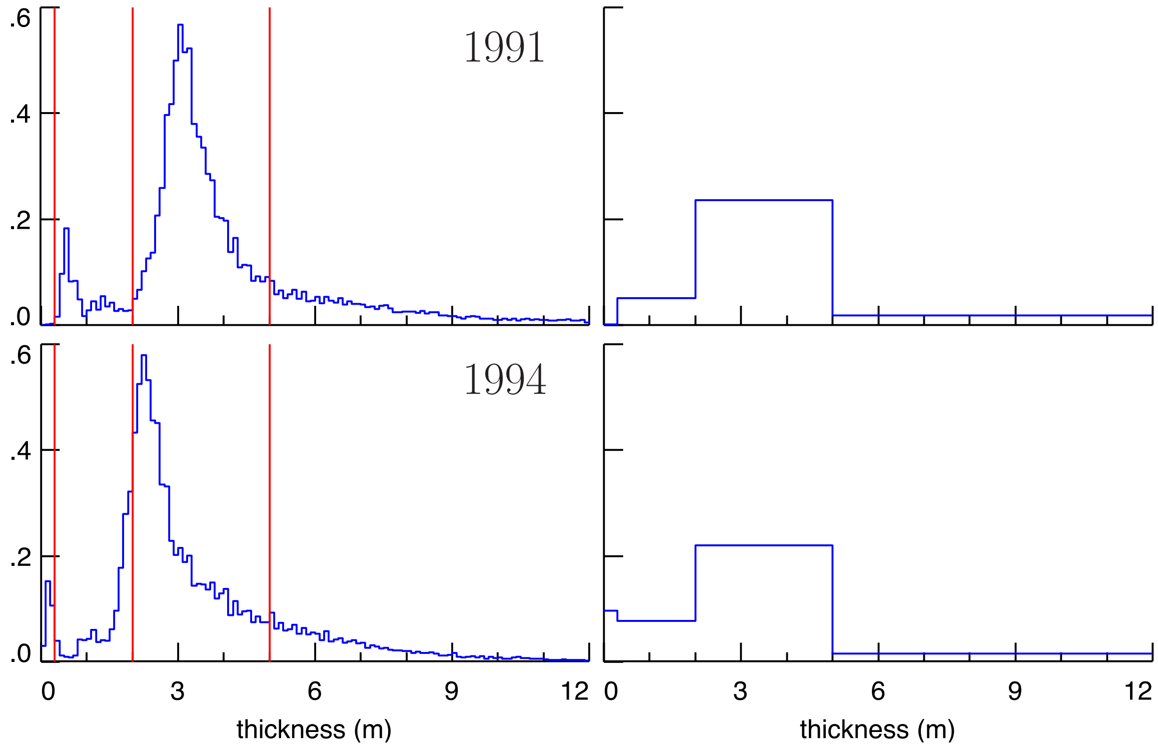
$$I_i(l) = \begin{cases} 1, & \text{if } h_i^{(l)} \le h(l) < h_i^{(u)}\text{; and} \\ 0, & \text{otherwise.} \end{cases}$$

# Simplification via Ice Type Distributions: III




- bottom 4 plots show indicator series (these are binary valued)

# Simplification via Ice Type Distributions: IV




- binned data described by 4 values (constrained to sum to 1)
- ice type distributions coarser than original ice thickness distributions, but former captures essence of latter

## Assessing Changes in Ice Type Distributions: I

- after binning, question of significance of changes still remains, but can be tackled using the following ideas
- by construction, area under an estimated ice type distribution between $h_i^{(l)}$ and $h_i^{(u)}$ is observed proportion $\hat{p}_i$ of occurrences of the $i$th ice type in $h(l)$
- proportion $\hat{p}_i$ is estimator of corresponding theoretical quantity

$$p_i = \int_{h_i^{(l)}}^{h_i^{(u)}} g(h)\, dh$$

## Assessing Changes in Ice Type Distributions: II

- observed proportion $\hat{p}_i$ is just sample mean of $i$th indicator series $I_i(l)$:

$$\hat{p}_i = \frac{1}{N} \sum_{l=0}^{N-1} I_i(l),$$

  where $N$ is # of measurements in profile (typically $> 50,000$)

- corresponding sample variance for $I_i(l)$ is given by

$$\hat{\sigma}_i^2 = \frac{1}{N} \sum_{l=0}^{N-1} \left(I_i(l) - \hat{p}_i\right)^2 = \hat{p}_i(1 - \hat{p}_i).$$

## Assessing Changes in Ice Type Distributions: III

- if $I_i(l)$, $l = 0, \ldots, N-1$, were a realization of $N$ independent binary-valued random variables, could compute an approximate 95% confidence interval for $p_i$ based upon an estimate of the variance of $\hat{p}_i$:

$$\operatorname{var}\{\hat{p}_i\} \approx \frac{\hat{\sigma}_i^2}{N} = \frac{\hat{p}_i(1-\hat{p}_i)}{N},$$

- independence does not hold due to autocorrelation, so need to determine correlation structure of indicator series $I_i(l)$
- can quantify correlation structure using wavelets

## Decomposing Sample Variance of 'Time' Series

- let $X_0, X_1, \ldots, X_{N-1}$ represent a 'time' series with $N$ values (for sea ice, 'time' is actually distance along a transect)
- let $\overline{X}$ denote sample mean of $X_t$'s: $\overline{X} \equiv \frac{1}{N}\sum_{t=0}^{N-1} X_t$
- let $\hat{\sigma}_X^2$ denote sample variance of $X_t$'s:

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

- can quantify certain properties of $X_t$ by an analysis of variance, i.e., by decomposing (breaking up) $\hat{\sigma}_X^2$ into pieces
- analysis of variance using wavelets based upon differences between adjacent averages over certain 'scales'

## Notion of Scale

- scale $\tau$ refers to width of an interval (e.g., chunch of a transect)
- scale-based analysis looks at averages over intervals of width $\tau$:

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

- localized average of $X_t$ and its $\tau - 1$ prior values
- $\overline{X}_t(1) = X_t$ is scale 1 'average'
- $\overline{X}_{N-1}(N) = \overline{X}$ is scale $N$ average

## Wavelet Coefficients and Filters

- wavelet coefficients tell us about variations in adjacent averages
- use wavelet filter to create wavelet coefficients
- given $X_0, X_1, \ldots, X_{N-1}$, define wavelet coefficients via

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

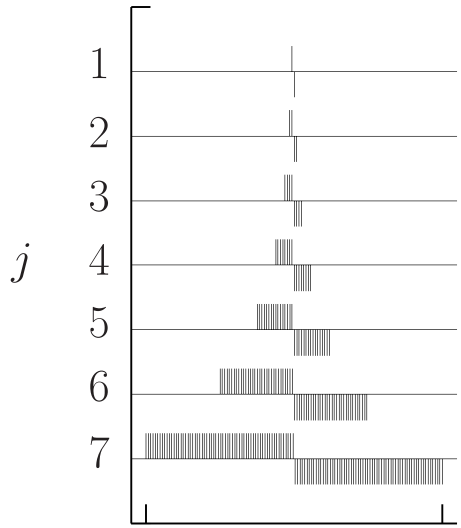
  where $\tilde{h}_{j,l}$ is a wavelet filter with $L_j$ coefficients, and

$$\begin{aligned} X_{t-l \bmod N} &= X_t \text{ if } 0 \le t-l \le N-1 \\ X_{-1 \bmod N} &= X_{N-1} \\ X_{-2 \bmod N} &= X_{N-2} \text{ etc (`circularity')} \end{aligned}$$

- index $j$ specifies associated scale as $\tau_j \equiv 2^{j-1}$, $j = 1, 2, \ldots$; i.e., scales are powers of two (1, 2, 4, 8, ...)

# Haar Wavelet Filters

- Haar wavelet filters $\tilde{h}_{j,l}$ for scales indexed by $j = 1, \ldots, 7$




1 positive & 1 negative coefficient

2 positive & 2 negative coefficients

4 & 4

8 & 8

16 & 16

32 & 32

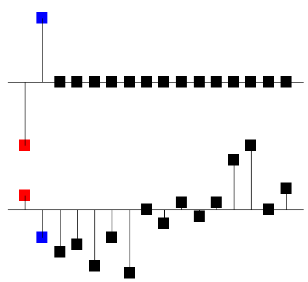
64 & 64

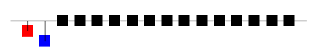
# Haar Wavelet Coefficients: I

- consider how $\widetilde{W}_{1,1} = \sum_l \tilde{h}_{1,l} X_{1-l \bmod N}$ is formed ($N = 16$):

$\tilde{h}_{1,l}$

$X_t$

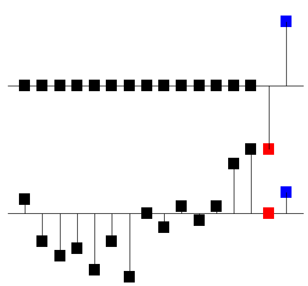


product  sum $\propto \overline{X}_1(1) - \overline{X}_0(1)$

- similar interpretation for $\widetilde{W}_{1,15} = \sum_l \tilde{h}_{1,l} X_{15-l \bmod N}$:

$\tilde{h}_{1,l}$

$X_t$



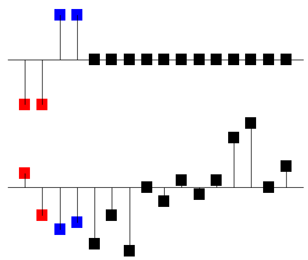
product  sum $\propto \overline{X}_{15}(1) - \overline{X}_{14}(1)$

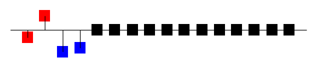
# Haar Wavelet Coefficients: II

- now consider form of $\widetilde{W}_{2,3} = \sum_l \tilde{h}_{2,l} X_{3-l \bmod N}$:

$h_{2,l}$

$X_t$



product  sum $\propto \overline{X}_3(2) - \overline{X}_1(2)$

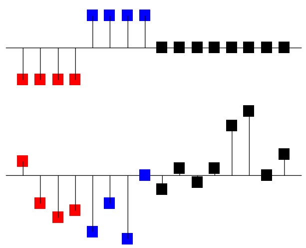
- similar interpretation for $\widetilde{W}_{2,4}, \widetilde{W}_{2,5}, \ldots, \widetilde{W}_{2,15}$
- note: $\widetilde{W}_{2,0}, \widetilde{W}_{2,1}$ and $\widetilde{W}_{2,2}$ aren't proportional to differences of adjacent averages (called 'boundary' coefficients)

# Haar Wavelet Coefficients: III

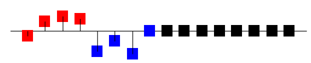
- $\widetilde{W}_{3,7} = \sum_l \tilde{h}_{3,l} X_{7-l \bmod N}$ takes the following form:

$\tilde{h}_{3,l}$

$X_t$



product



sum $\propto \overline{X}_7(4) - \overline{X}_3(4)$

- Haar wavelet coefficients $\widetilde{W}_{j,t}$ for scale $\tau_j = 2^{j-1}$ proportional to $\overline{X}_t(\tau_j) - \overline{X}_{t-\tau_j}(\tau_j)$. i.e., to change in adjacent $\tau_j$ averages
  - change measured by simple first difference
  - average is localized sample mean
  - if $\widetilde{W}^2_{j,t}$ small, not much variation over scale $\tau_j$
  - if $\widetilde{W}^2_{j,t}$ large, lot of variation over scale $\tau_j$

# Empirical Wavelet Variance

- define empirical wavelet variance for scale $\tau_j$ as

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

- if $N = 2^J$, obtain analysis (decomposition) of sample variance:

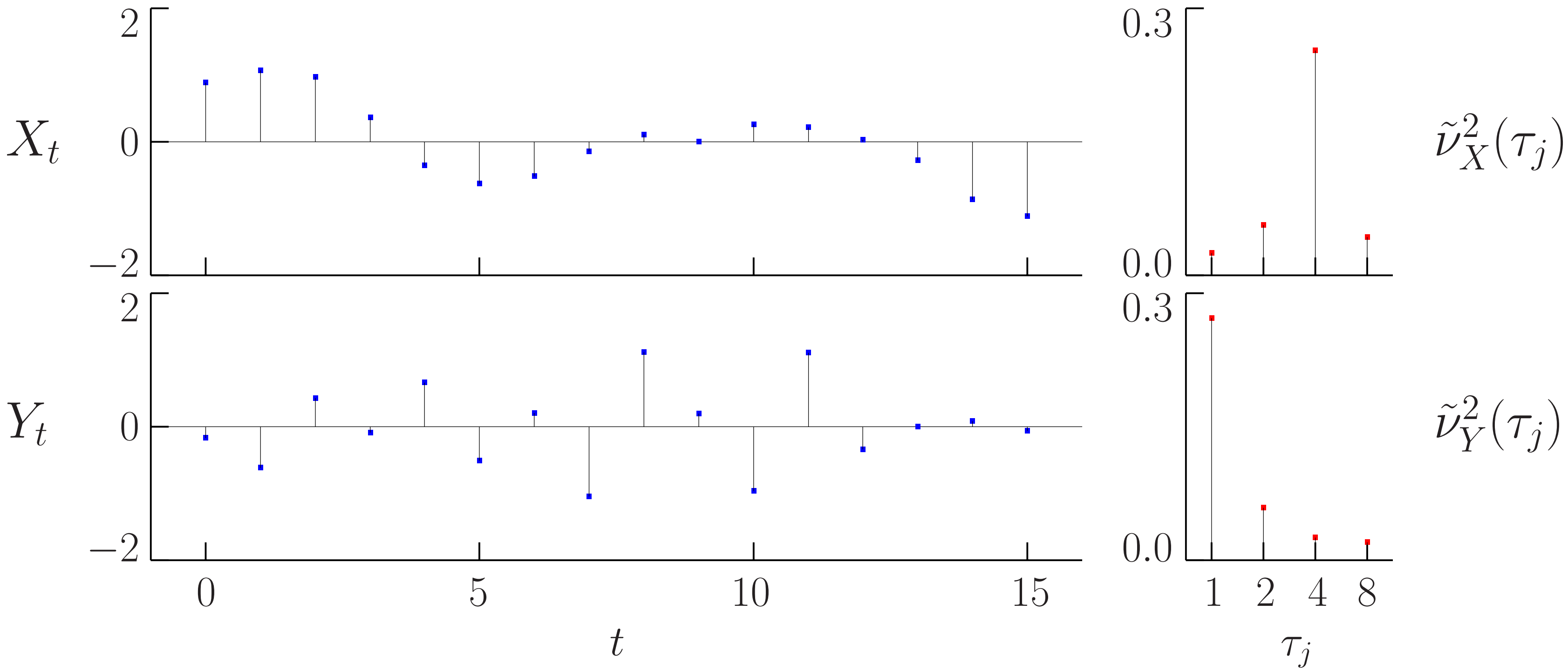
$$\hat{\sigma}_X^2 = \frac{1}{N} \sum_{t=0}^{N-1} \left(X_t - \overline{X}\right)^2 = \sum_{j=1}^{J} \tilde{\nu}_X^2(\tau_j)$$

  (if $N$ not a power of 2, can still obtain an analysis of variance to a given level $J_0$, but have component due to 'scaling' filter)

- interpretation: $\tilde{\nu}_X^2(\tau_j)$ is portion of $\hat{\sigma}_X^2$ due to changes in averages over scale $\tau_j$; i.e., 'scale by scale' analysis of variance

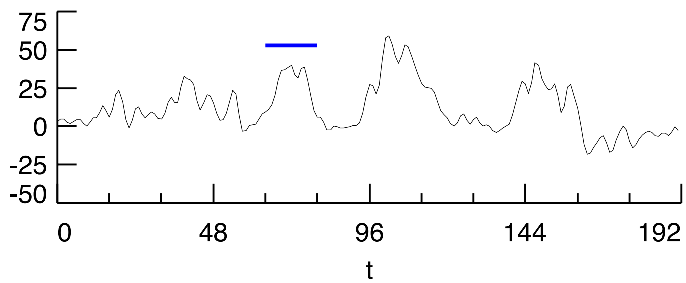
## Example of Empirical Wavelet Variance

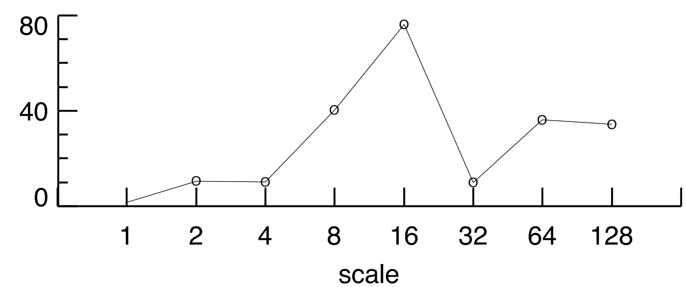
- wavelet variances for time series $X_t$ and $Y_t$ of length $N = 16$, each with zero sample mean and same sample variance



# Second Example of Empirical Wavelet Variance

- top: subtidal sea level series $X_t$ (blue line shows scale of 16)




- bottom: empirical wavelet variances $\tilde{\nu}_X^2(\tau_j)$

# Theoretical Wavelet Variance: I

- now assume $X_t$ is a real-valued random variable (RV)
- let $X_t$, $t \in \mathbb{Z}$ denote a stochastic process, i.e., collection of RVs indexed by 'time' $t$ (here $\mathbb{Z}$ denotes the 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

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

# Theoretical Wavelet Variance: II

- if $Y$ is any RV, let $E\{Y\}$ denote its expectation
- let $\text{var}\,\{Y\}$ denote its variance: $\text{var}\,\{Y\} \equiv E\{(Y - E\{Y\})^2\}$
- definition of wavelet variance:

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

  with conditions on $X_t$ so that $\text{var}\,\{\overline{W}_{j,t}\}$ exists, is finite and does not depend on $t$
- $\nu_X^2(\tau_j)$ well-defined for stationary processes, so let's review concept of stationarity

# Definition of a Stationary Process

- if $U$ and $V$ are two RVs, denote their covariance by

$$\operatorname{cov}\{U,V\} = E\{(U - E\{U\})(V - E\{V\})\}$$

- stochastic process $X_t$ called stationary if
  - $E\{X_t\} = \mu_X$ for all $t$, i.e., constant independent of $t$
  - $\operatorname{cov}\{X_t, X_{t+\tau}\} = s_{X,\tau}$, i.e., depends on lag $\tau$, but not $t$
- $s_{X,\tau}$, $\tau \in \mathbb{Z}$, is autocovariance sequence (ACVS)
- $s_{X,0} = \operatorname{cov}\{X_t, X_t\} = \operatorname{var}\{X_t\}$; i.e., variance same for all $t$

## Example of a Stationary Process: White Noise

- simplest example of a stationary process is 'white noise'
- process $X_t$ said to be white noise if
  - it has a constant mean $E\{X_t\} = \mu_X$
  - it has a constant variance $\text{var}\,\{X_t\} = \sigma_X^2$
  - $\text{cov}\,\{X_t, X_{t+\tau}\} = 0$ for all $t$ and nonzero $\tau$; i.e., distinct RVs in the process are uncorrelated
- ACVS for white noise takes a very simple form:

$$s_{X,\tau} = \text{cov}\,\{X_t, X_{t+\tau}\} = \begin{cases} \sigma_X^2, & \tau = 0; \\ 0, & \text{otherwise.} \end{cases}$$

# Wavelet Variance for Stationary Processes

- for stationary processes, wavelet variance decomposes var $\{X_t\}$:

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

(above result similar to one for sample variance)

- $\nu_X^2(\tau_j)$ is thus contribution to var $\{X_t\}$ due to scale $\tau_j$
- example: for a white noise process, have

$$\nu_X^2(\tau_j) = \frac{\mathrm{var}\,\{X_t\}}{2^j} = \frac{\mathrm{var}\,\{X_t\}}{2\tau_j} \propto \tau_j^{-1},$$

so largest contribution to var $\{X_t\}$ is at smallest scale $\tau_1$

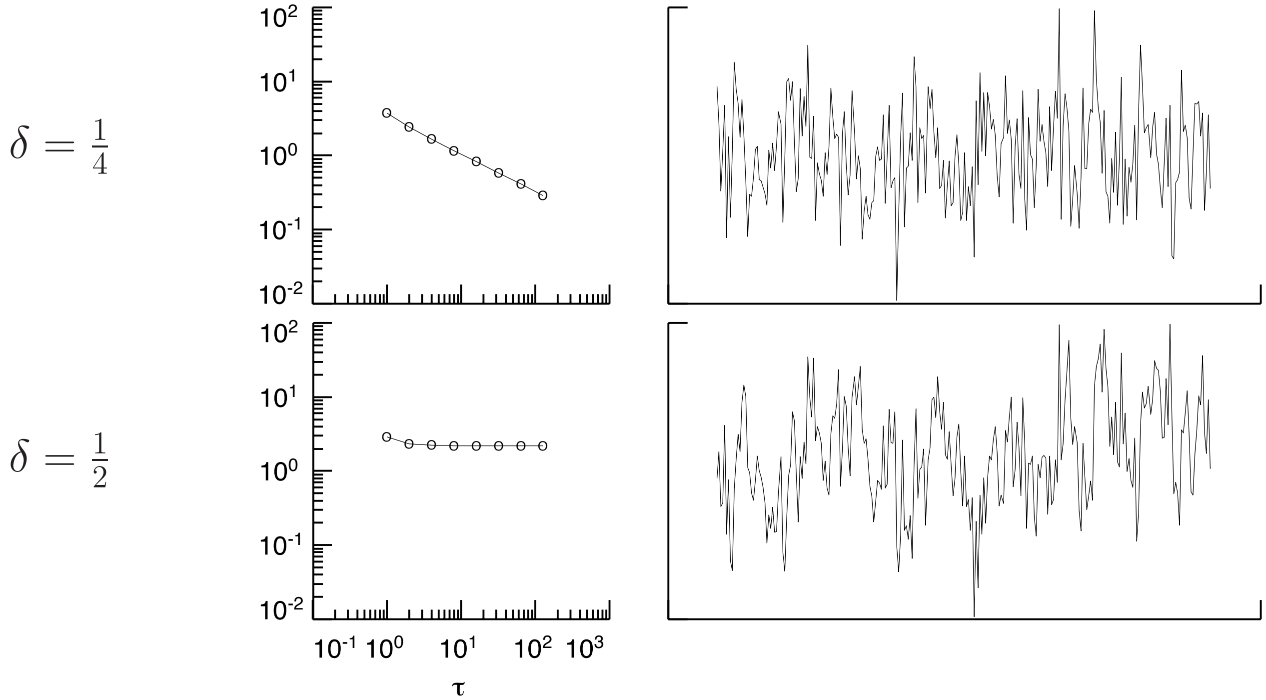
# Fractionally Differenced (FD) Processes: I

- as another example, consider wavelet variance for FD processes (Granger & Joyeux, 1980; Hosking, 1981)
- FD processes determined by 2 parameters $-\infty < \delta < \infty$ & $\sigma_\epsilon^2 > 0$ (relatively unimportant)
- if $\delta < 1/2$, FD process $X_t$ is stationary, and, in particular,
  - reduces to white noise if $\delta = 0$
  - has 'long range' dependence if $\delta > 0$; i.e., $s_{X,\tau} > 0$ and does not decrease to zero rapidly
  - is 'antipersistent' if $\delta < 0$ (i.e., $\operatorname{cov}\{X_t, X_{t+1}\} < 0$)

# Fractionally Differenced (FD) Processes: II

- at large scales, have
$$\nu_X^2(\tau_j) \approx C\tau_j^{2\delta-1}$$
- thus
$$\log\left(\nu_X^2(\tau_j)\right) \approx \log\left(C\right) + (2\delta - 1)\log\left(\tau_j\right),$$
so a log/log plot of $\nu_X^2(\tau_j)$ vs. $\tau_j$ looks approximately linear with slope $2\delta - 1$ for $\tau_j$ large enough
- for white noise, have $\delta = 0$, $\nu_X^2(\tau_j) = C\tau_j^{-1}$ and hence
$$\log\left(\nu_X^2(\tau_j)\right) = \log\left(C\right) - \log\left(\tau_j\right),$$

# LA(8) Wavelet Variance for 2 FD Processes




- left-hand column: $\nu_X^2(\tau_j)$ versus $\tau_j$
- right-hand: realization of length $N = 256$ from each FD process
- note: slope on log/log plot would be $-1$ for uncorrelated data (white noise)

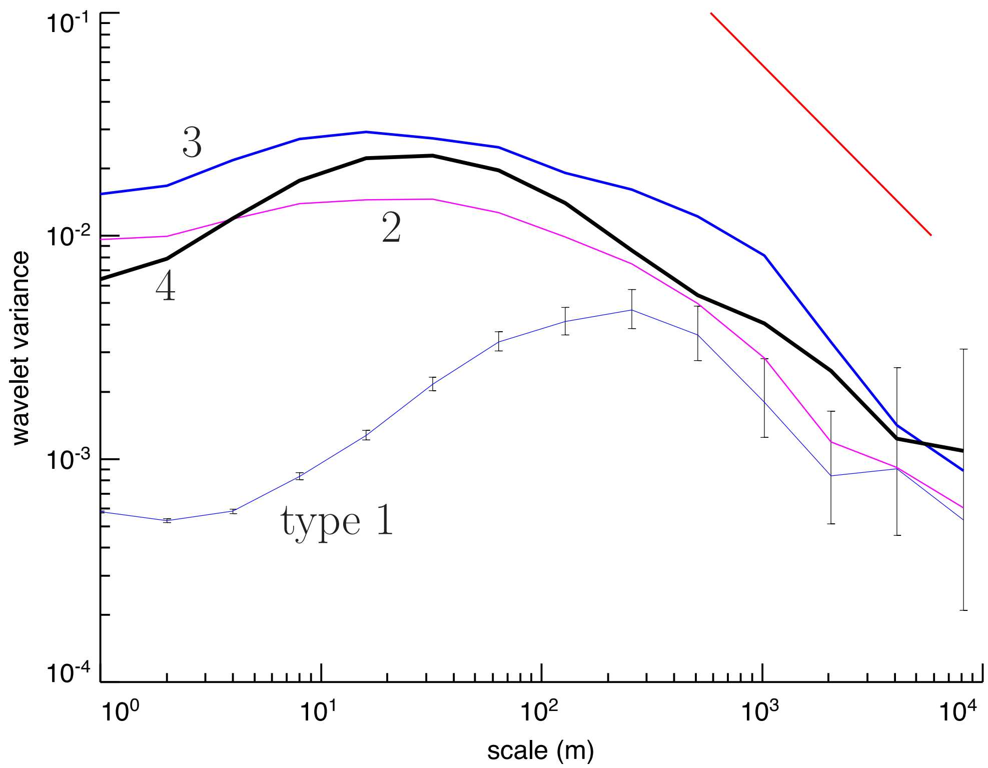
## Haar Wavelet Analysis of Ice Type Series: I

- let $W_{i;j,l}$ denote Haar wavelet coefficient for $i$th ice type and scale $\tau_j = 2^{j-1}$ meters at location along transect indexed by $l$
- coefficient is proportional to difference of adjacent sample means:

$$W_{i;j,l} \propto \frac{1}{\tau_j} \sum_{m=0}^{\tau_j - 1} I_i(l - m) - \frac{1}{\tau_j} \sum_{n=0}^{\tau_j - 1} I_i(l - \tau_j - n),$$

- because $I_i(l)$ is binary-valued, $W_{i;j,l}$ has simple and intuitively appealing interpretation: it is proportional to *differences* in percentage of $i$th ice type between adjacent parts of the ice thickness profile, with each part being of scale (length) $\tau_j$.
- if $W_{i;j,l}^2$ is small, ice type percentage is stable, i.e., not varying much between adjacent parts of scale $\tau_j$

# Haar Wavelet Analysis of Ice Type Series: II




- estimated wavelet variances based upon averaging available $W^2_{i;j,l}$
- red line has slope of $-1$ (appropriate for white noise)

# Haar Wavelet Analysis of Ice Type Series: III

- wavelet variance curves are basically unimodal
- location of peak gives indication of 'characteristic scale' (e.g., $\approx$ 256 m and 32 m for types 1 and 4)
- conjecture: can use characteristic scale and rate of decay of curve as $\tau_j$ gets large to assess variability in $\hat{p}_i$ via

$$\mathrm{var}\{\hat{p}_i\} \approx \frac{\hat{p}_i(1-\hat{p}_i)}{N_e^{-\beta}}$$

- $N_e \propto N$ is 'effective sample size,' with constant of proportionality being related to the characteristic scale
- rate at which $\mathrm{var}\{\hat{p}_i\}$ decreases to zero as $N$ increases is determined by $\beta$, which is the slope of $\log(\hat{\nu}_i^2(\tau_j))$ versus $\log(\tau_j)$ over large $\tau$

# Future Work

- work out statistical theory for testing null hypothesis of constant ice type distribution
- investigate relationship between characteristic scales and physical processes
- look at spatial and temporal variations in distributions and characteristics scales
- thanks for the invitation to speak!