An Introduction to Wavelet Analysis with Applications to Vegetation Monitoring

Don Percival

Applied Physics Laboratory, University of Washington

Seattle, Washington, USA

overheads for talk available at

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

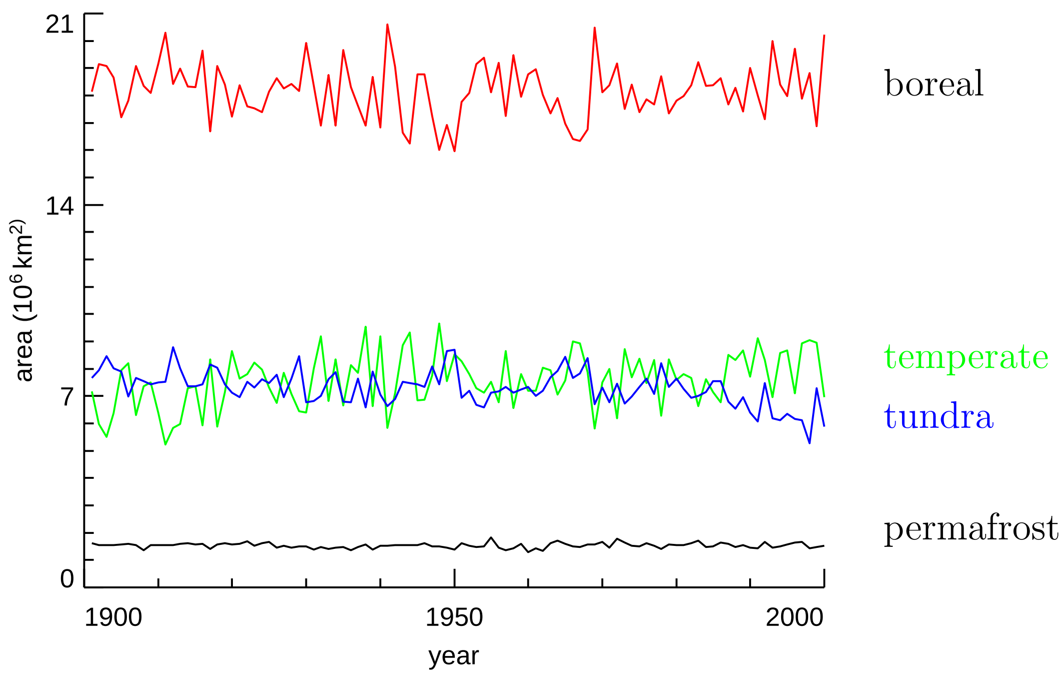
Overview: I

- as a subject, wavelets are
  - relatively new (1983 to present)
  - a synthesis of old/new ideas
  - keyword in $18,465+$ articles & books since 1989 (2700+ since 2002: an inundation of material!!!)
- wavelets can help us understand
  - time series (i.e., observations collected over time)
  - images
- wavelets capable of describing how
  - time series evolve over time on a given scale
  - images change from one place to the next on a given scale,

  where here 'scale' is either
  - an interval (span) of time (hour, year, ...) or
  - a spatial area (square kilometer, acre, ...)

Overview: II

- example: time series of vegetation areas over land (50°–90° N) (based on monthly SAT data from Climate Research Unit, UK)



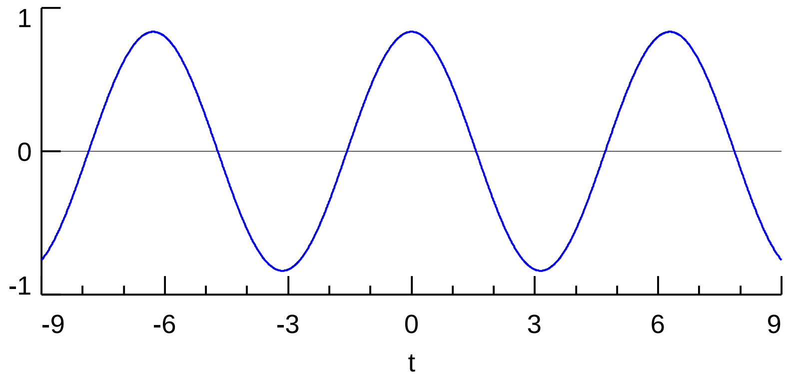
- some questions that wavelets can help up address:
  1. are variations homogeneous across time?
  2. are variations from one year to the next more prominent than variations from one decade to the next?
  3. permafrost is less variable than boreal, but do they have other statistical properties that are significantly different?
  4. how are any two of these series related on a scale by scale basis (e.g., year to year, decade to decade)?

Outline of Remainder of Talk

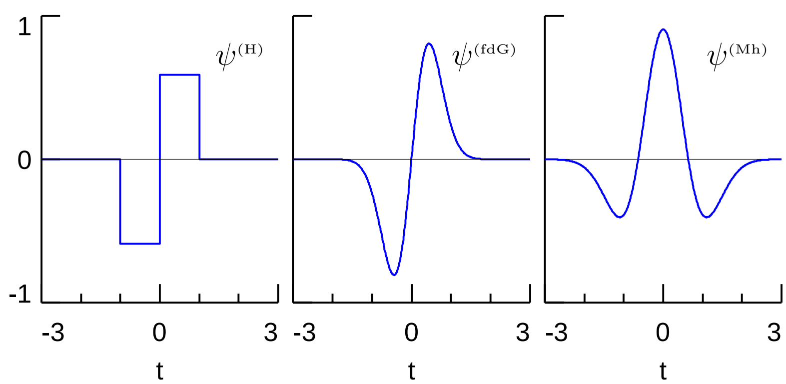
- discuss what exactly a wavelet is
- discuss wavelet analysis (emphasis on physical interpretation)
- point out two fundamental properties of the continuous wavelet transform (CWT):
  1. CWT is fully equivalent to the transformed time series
  2. CWT tells how 'energy' in time series is distributed across different scales and different times
- describe the discrete wavelet transform (DWT)
- point out two analogous fundamental properties of DWT
- look at DWT of one of the vegetation area time series (boreal)
  - addresses questions 1 (homogeneity across time) and 2 (prominence of yearly/decadal variations)
- describe wavelet variance
  - addresses questions 2 and 3 (how statistical properties of permafrost & boreal compare)
- look at wavelet covariance between boreal & temperate series
  - addresses question 4 (scale by scale relationship of two series)
- concluding remarks

What is a Wavelet?

- sines & cosines are 'big waves'



- wavelets are 'small waves'
- three wavelets, including the Haar wavelet $\psi^{(H)}(\cdot)$:



- conditions for a function $\psi(\cdot)$ to be a wavelet:
  - $\psi(\cdot)$ integrates to 0; i.e., $\psi(\cdot)$ balances itself above/below 0
  - $\psi^2(\cdot)$ integrates to 1
  - 'admissibility' (mild technical condition)

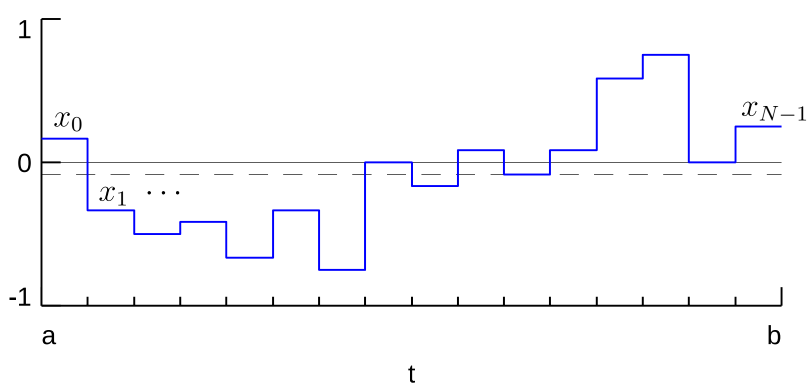
Basics of Wavelet Analysis: I

- wavelets tell us about variations in local averages
- to quantify, let $x(\cdot)$ be a time series
  - real-valued function of t defined over real axis
  - will refer to t as time (but it need not be such)
- consider average value of $x(\cdot)$ over $[a, b]$:

$$\frac{1}{b-a}\int_a^b x(u)\,du \equiv \alpha(a,b)$$

(above notion discussed in elementary calculus books)

- related to idea of sample mean
  - suppose $x(\cdot)$ is a step function with N steps in $[a, b]$:



  - then

$$\frac{1}{b-a}\int_a^b x(u)\,du = \frac{1}{b-a}\sum_{j=0}^{N-1} x_j \frac{b-a}{N} = \frac{1}{N}\sum_{j=0}^{N-1} x_j$$

Basics of Wavelet Analysis: II

- reparameterize using width λ and time t of center of interval:

$$A(\lambda, t) \equiv \alpha(t - \tfrac{\lambda}{2}, t + \tfrac{\lambda}{2}) = \frac{1}{\lambda}\int_{t-\frac{\lambda}{2}}^{t+\frac{\lambda}{2}} x(u)\,du$$

  - $\lambda \equiv b - a$ is called scale
  - $t = (a + b)/2$ is time associated with center of interval

- $A(\lambda, t)$ is average value of $x(\cdot)$ over scale λ at time t
- average values of time series are of considerable interest
  - vegetation area time series are spatial/temporal averages
  - proportion of gaps in transect over a forest canopy
  - etc.
- Q: how much do averages change from one interval to the next?
- can quantify changes in $A(\lambda, t + \frac{1}{2})$ by considering

$$\begin{aligned} D(\lambda, t) &\equiv A(\lambda, t + \tfrac{\lambda}{2}) - A(\lambda, t - \tfrac{\lambda}{2}) \\ &= \frac{1}{\lambda}\int_{t}^{t+\lambda} x(u)\,du - \frac{1}{\lambda}\int_{t-\lambda}^{t} x(u)\,du \end{aligned}$$

- $D(\lambda, t)$ often of more interest than $A(\lambda, t)$

Basics of Wavelet Analysis: III

- can connect $D(1,0)$ to Haar wavelet: note that

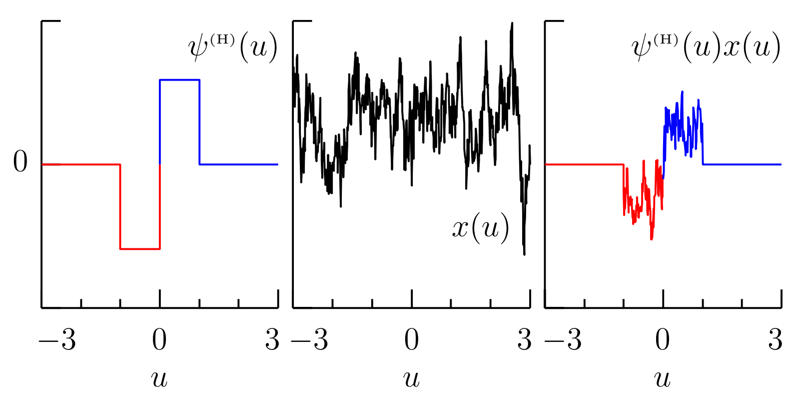
$$D(1,0) = \int_0^1 x(u)\,du - \int_{-1}^0 x(u)\,du = \int_{-\infty}^{\infty} \tilde{\psi}_{1,0}(u) x(u)\,du$$

if we define

$$\tilde{\psi}_{1,0}(u) = \begin{cases} -1, & -1 < u \le 0; \\ 1, & 0 < u \le 1; \\ 0, & \text{otherwise} \end{cases}$$

- comparing $\tilde{\psi}_{1,0}(u)$ with $\psi^{(H)}(\cdot)$ shows that $\tilde{\psi}_{1,0}(u) = \sqrt{2}\psi^{(H)}(u)$
- Haar wavelet extracts information about difference between unit scale averages at $t = 0$ via

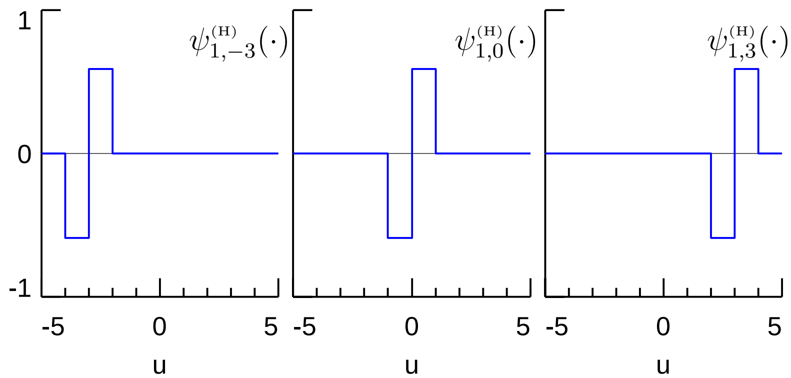
$$W(1,0) \equiv \int_{-\infty}^{\infty} \psi^{(H)}(u) x(u)\,du \propto D(1,0)$$



Basics of Wavelet Analysis: IV

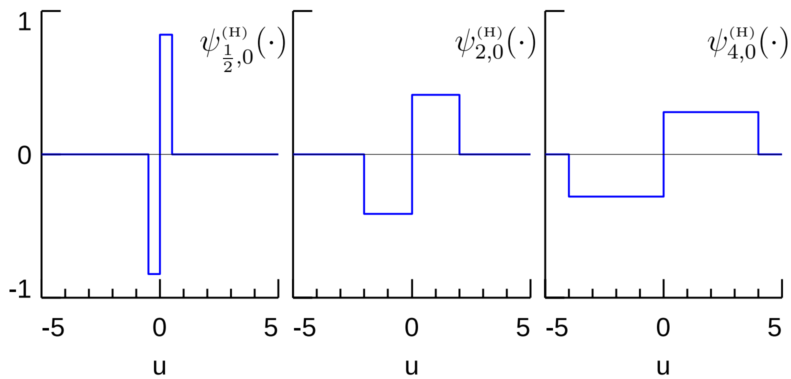
- to extract information at other t's, just shift $\psi^{(H)}(u)$ to form

$$\psi^{(H)}_{1,t}(u) \equiv \psi^{(H)}(u - t)$$



- to extract information about other λ's, form

$$\psi^{(H)}_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}}\psi^{(H)}\left(\frac{u-t}{\lambda}\right)$$



- can check that $\psi^{(H)}_{\lambda,t}(\cdot)$ are indeed wavelets

Basics of Wavelet Analysis: V

- use $\psi^{(H)}_{\lambda,t}(\cdot)$ to obtain

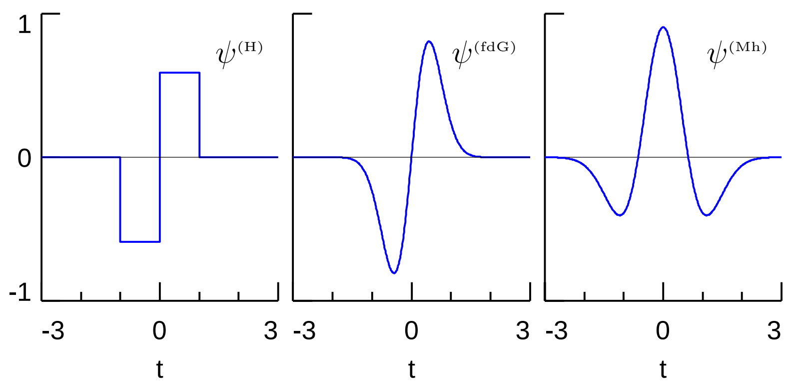
$$W(\lambda, t) \equiv \int_{-\infty}^{\infty} \psi^{(H)}_{\lambda,t}(u) x(u)\, du \propto D(\lambda, t)$$

  left-hand side is Haar continuous wavelet transform (CWT)

- can do the same with other wavelets:

$$W(\lambda, t) \equiv \int_{-\infty}^{\infty} \psi_{\lambda,t}(u) x(u)\, du, \text{ where } \psi_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}} \psi\left(\frac{u-t}{\lambda}\right)$$

  left-hand side is CWT based on $\psi(\cdot)$



- $\psi^{(fdG)}(\cdot)$ & $\psi^{(Mh)}(\cdot)$ yield differences of adjacent *weighted* averages

Two Fundamental Properties of CWT

1. can recover $x(\cdot)$ from its CWT:

$$x(t) = \frac{1}{C_\psi} \int_0^\infty \left[\int_{-\infty}^\infty W(\lambda, u) \frac{1}{\sqrt{\lambda}} \psi \left(\frac{t-u}{\lambda} \right) du \right] \frac{d\lambda}{\lambda^2}$$

   – here C_ψ is a constant depending on just $\psi(\cdot)$
   – conclusion: $W(\lambda, t)$ equivalent to $x(t)$

2. can decompose 'energy' in time series using CWT:

$$\int_{-\infty}^\infty x^2(t)\, dt = \int_0^\infty \int_{-\infty}^\infty \frac{W^2(\lambda, t)}{C_\psi \lambda^2}\, dt\, d\lambda$$

   – left-hand side called energy in $x(\cdot)$
   – plot of $x^2(t)$ versus t gives energy distribution across time
   – right-hand integrand gives energy distribution across time & scale

The Discrete Wavelet Transform (DWT)

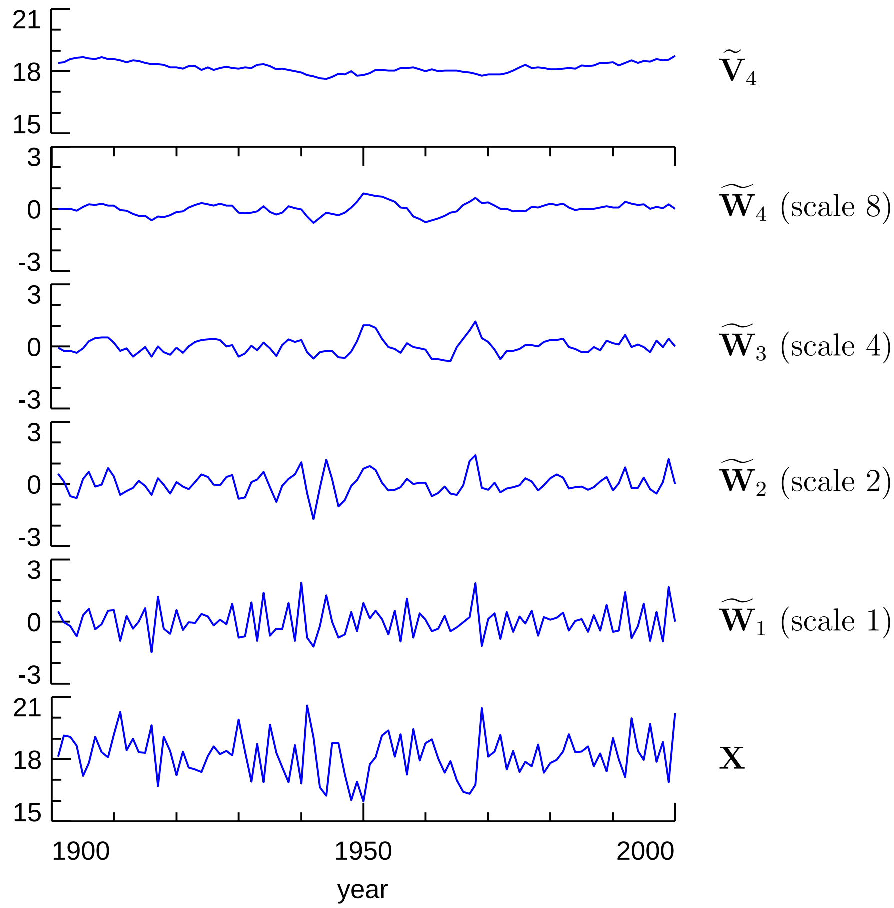
- when dealing with samples x_0, x_1, ... x_{N-1} from $x(\cdot)$, more convenient to deal with DWT than CWT
- can regard DWT as 'slices' through CWT
  - restrict λ to 'dyadic' scales $\tau_j \equiv 2^{j-1}$, $j = 1, 2, \ldots, J_0$
  - restrict times to integers $t = 0, 1, \ldots, N-1$
  - note: considering 'maximal overlap' DWT (MODWT) (can restrict times further to get orthonormal DWT)
- yields wavelet coefficients $\widetilde{W}_{j,t} \propto W(\tau_j, t)$
- also get scaling coefficients $\widetilde{V}_{J_0,t}$
  - related to averages over a scale of $\lambda = 2\tau_{J_0}$
  - summary of information in $W(\lambda, t)$ at $\lambda \geq 2\tau_{J_0} = 2^{J_0}$
- collect $\widetilde{W}_{j,t}$ into vector $\widetilde{\mathbf{W}}_j$ for levels $j = 1, 2, \ldots, J_0$
- also collect $\widetilde{V}_{J_0,t}$ into vector $\widetilde{\mathbf{V}}_{J_0}$
- $\widetilde{\mathbf{W}}_1, \ldots, \widetilde{\mathbf{W}}_{J_0}$ and $\widetilde{\mathbf{V}}_{J_0}$ form the DWT of $\mathbf{X} \equiv [x_0, \ldots, x_{N-1}]^T$

Two Fundamental Properties of DWT

1. can recover $\mathbf{X}$ perfectly from its DWT; i.e., $\widetilde{\mathbf{W}}_1, \ldots, \widetilde{\mathbf{W}}_{J_0}$ & $\widetilde{\mathbf{V}}_{J_0}$ are equivalent to $\mathbf{X}$
2. 'energy' in $\mathbf{X}$ preserved in its DWT:

$$\|\mathbf{X}\|^2 \equiv \sum_{t=0}^{N-1} x_t^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2$$

Example: DWT of Boreal Time Series



- large value for a wavelet coefficient indicates large variation at a particular scale & time
- variations fairly homogeneous across scales (i.e., stationarity might be a reasonable assumption here)
- smaller scales seem to be dominant

Wavelet Variance: I

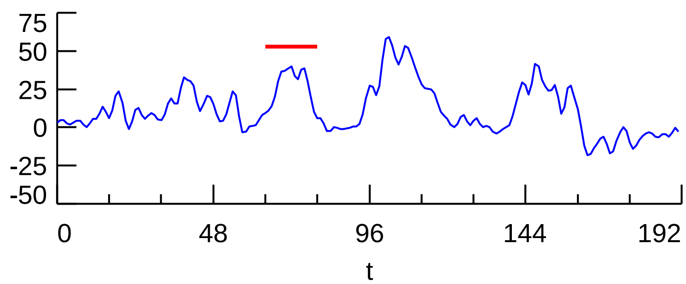
- energy preservation leads to analysis of sample variance:

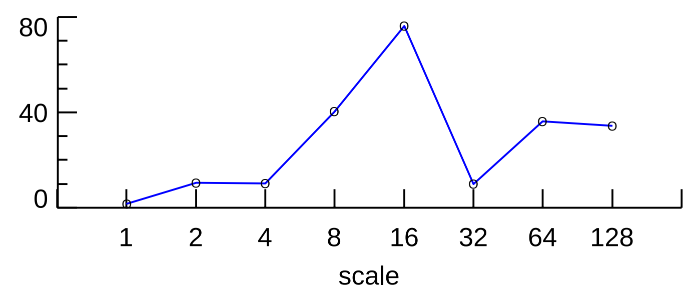
$$\hat{\sigma}_x^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}(x_t - \bar{x})^2 = \frac{1}{N}\left(\sum_{j=1}^{J_0}\|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2\right) - \overline{x}^2,$$

  where $\bar{x} \equiv \sum_t x_t/N$

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

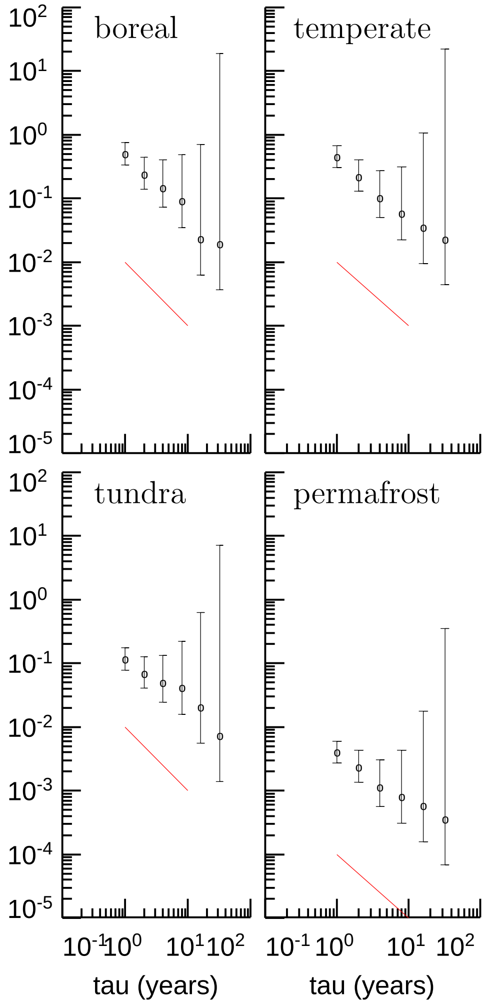
- example: subtidal sea levels and associated wavelet variances $\frac{1}{N}\|\widetilde{\mathbf{W}}_j\|^2$ versus scales $\tau_j = 2^{j-1}$ for $j = 1, \ldots, 8$





Wavelet Variance: II

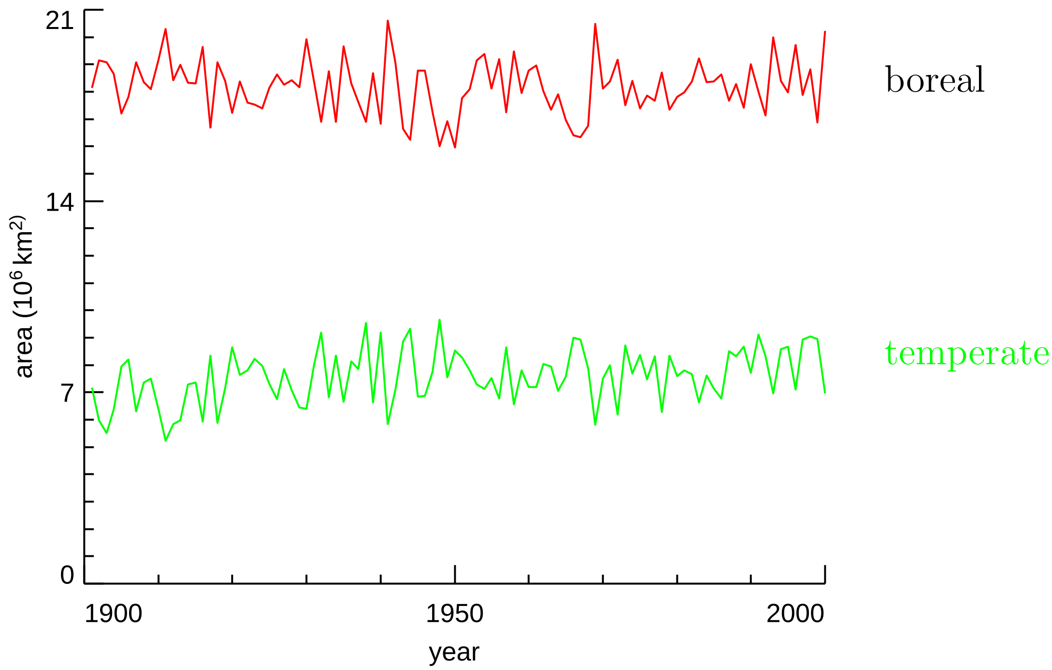
- wavelet variances for vegetation time series



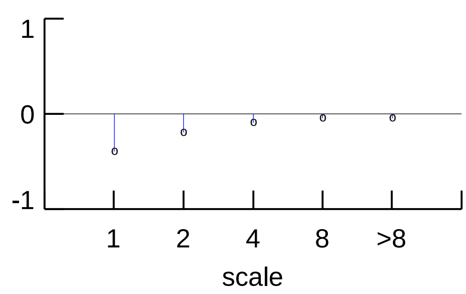
- sum of wavelet variances is equal to sample variance
- 95% confidence intervals based on statistical theory
- confirms that unit scale is dominant
- except for tundra, roll-offs similar & consistent with white noise
- tundra possibly possesses 'long memory' (rolls off more slowly)

Wavelet Covariance

- reconsider boreal and temperate time series:



- sample cross-correlation is -0.79; i.e., series are anticorrelated
- can decompose cross-correlation across different scales:



- two series anticorrelated at all scales, but sample cross-correlation mainly due to two smallest scales (75%)

Concluding Remarks

- wavelets decompose time series with respect to two variables:
  - time (location)
  - scale (extent)
- CWT & DWT have two fundamental properties:
  1. fully equivalent to original time series
  2. energy in time series is preserved
- wavelet variance gives scale-based analysis of variance (natural match for many geophysical processes)
- techniques extends naturally to images
- *many* other uses for wavelets (have barely scratched the surface!)
  - approximately decorrelate certain time series
  - can assess sampling properties of certain statistics
  - signal extraction ('wavelet shrinkage')
  - edge identification in images
  - compression of time series/images
  - fast simulation of time series/images
- come join the fun – article #18,467 is begging to be written!
- big thanks to symposium organizers for invitation to speak!