An Introduction to Wavelet Analysis

Don Percival

Applied Physics Laboratory
Box 355640
University of Washington
Seattle, WA 98195–5640

206–543–1368

dbp@apl.washington.edu

http://faculty.washington.edu/dbp

Overview: I

- as a subject, wavelets are
  - relatively new (1983 to present)
  - a synthesis of old/new ideas
  - keyword in 29,826+ articles and books since 1989 (4032 more since 2005: an inundation of material!!!)
- wavelets can help us understand
  - time series (i.e., observations collected over time)
  - images

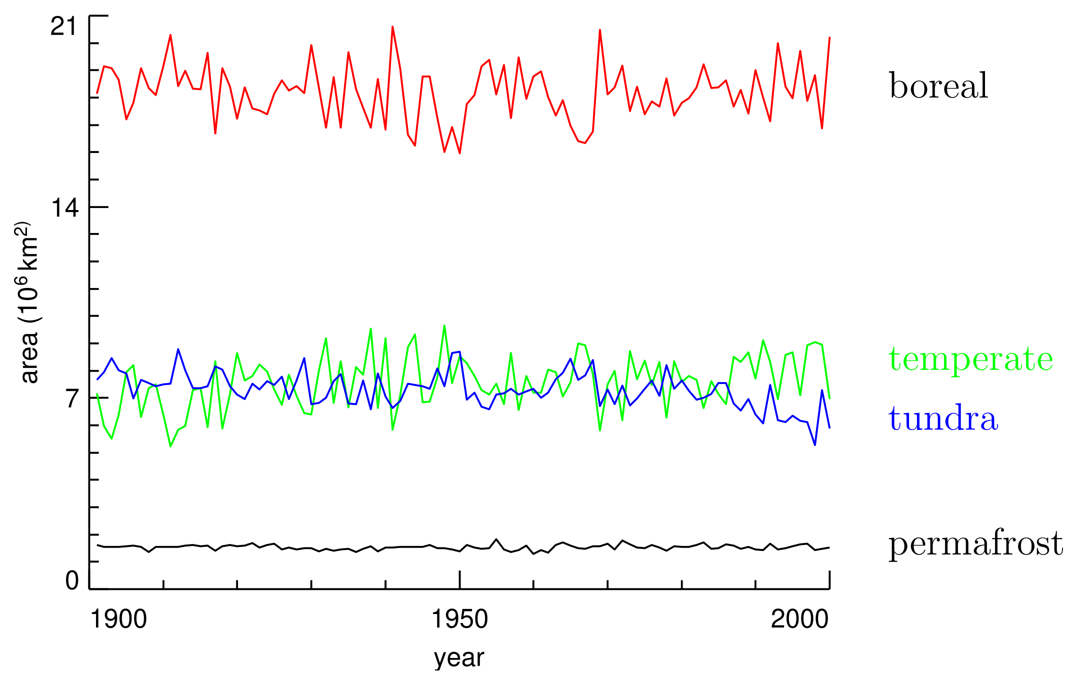
Overview: II

- 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: III

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



Overview: IV

- 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. What are the pairwise relationships between these series on a scale by scale basis (e.g., year to year, decade to decade)?

Outline of Remainder of Talk: I

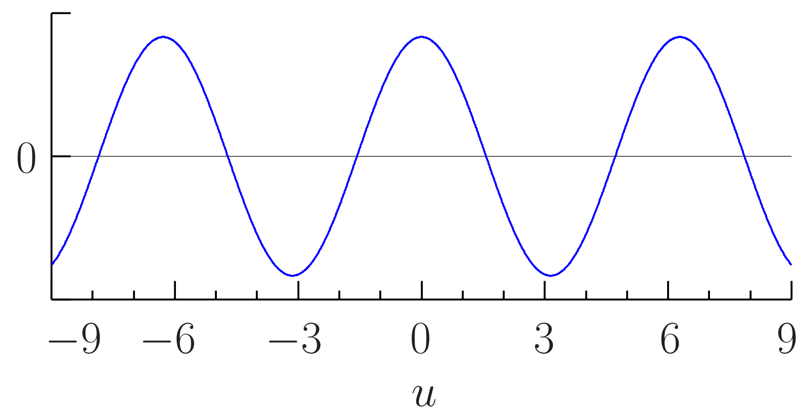
- discuss what exactly a wavelet is
- discuss wavelet analysis (emphasis on physical interpretation)
- point out two fundamental properties of the continuous wavelet transform (CWT):
  1. CWT is fully equivalent to the transformed time series
  2. CWT tells how ‘energy’ in time series is distributed across different scales and different times
- describe the discrete wavelet transform (DWT)
- point out two analogous fundamental properties of DWT

Outline of Remainder of Talk: II

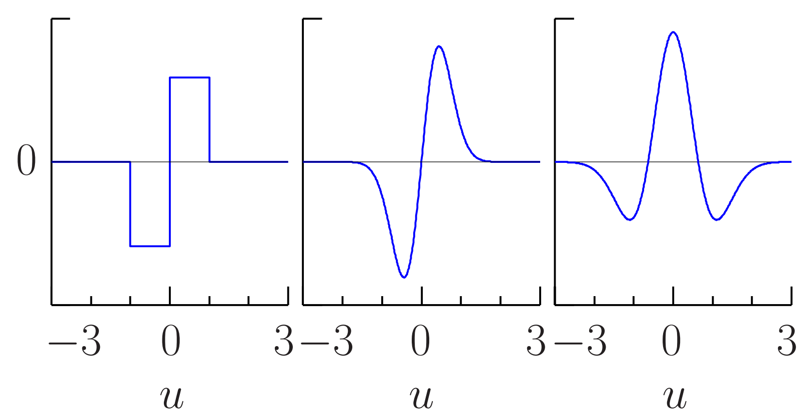
- look at DWT of one of the vegetation area time series (boreal)
  - addresses questions 1 (homogeneity across time) and 2 (prominence of yearly/decadal variations)
- describe wavelet variance
  - addresses questions 2 and 3 (how statistical properties of permafrost & boreal compare)
- look at wavelet 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'

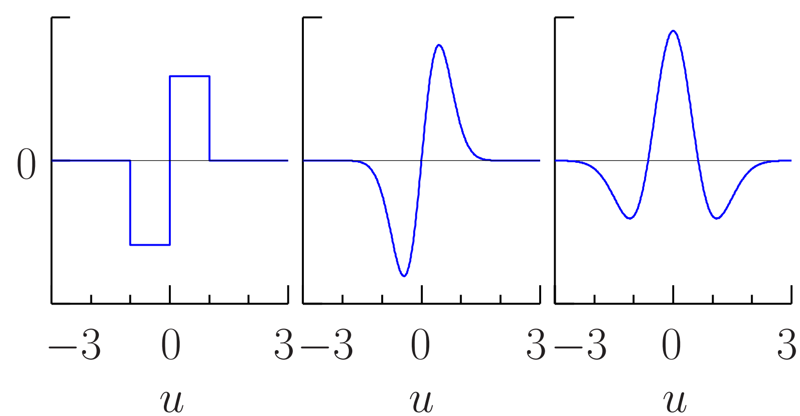


- wavelets are 'small waves' (left-hand is Haar wavelet $\psi^{(\mathrm{H})}(\cdot)$)



Technical Definition of a Wavelet

- real-valued function $\psi(\cdot)$ defined over real axis is a wavelet if
  1. integral of $\psi^2(\cdot)$ is unity: $\int_{-\infty}^{\infty} \psi^2(u)\,du = 1$
     (called 'unit energy' property, with apologies to physicists)
  2. integral of $\psi(\cdot)$ is zero: $\int_{-\infty}^{\infty} \psi(u)\,du = 0$
     (technically, need an 'admissibility condition,' but this is almost equivalent to integration to zero)



What is Wavelet Analysis?

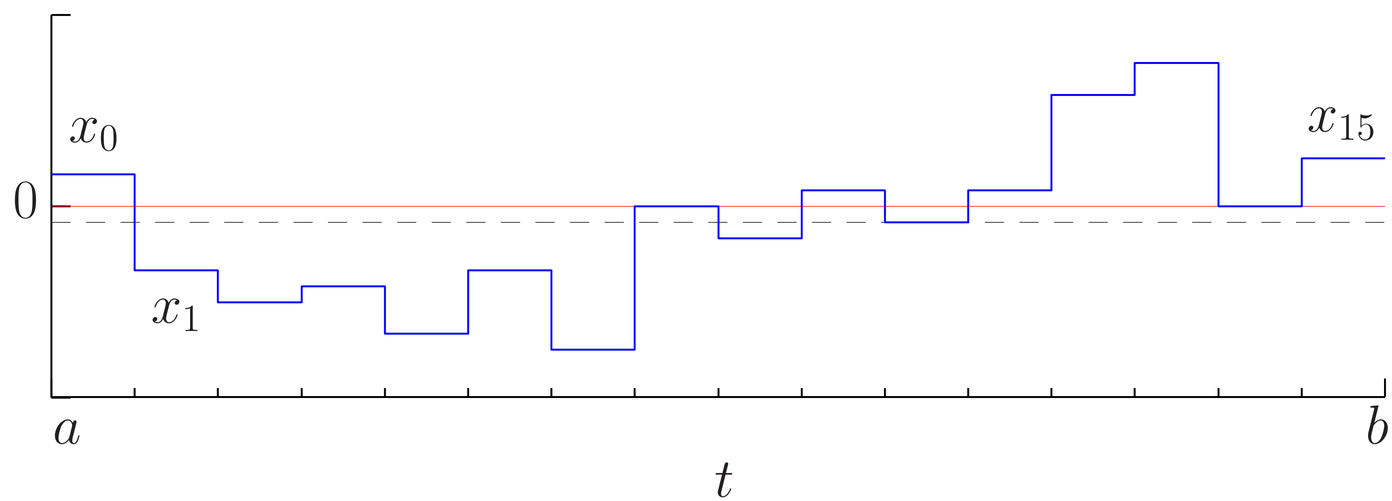
- wavelets tell us about variations in local averages
- to quantify this description, let $x(\cdot)$ be a time series
  - real-valued function of t defined over real axis
  - will refer to t as time (but it need not be such)
- consider 'average value' of $x(\cdot)$ over $[a, b]$:

$$\frac{1}{b-a}\int_a^b x(t)\,dt$$

(above notion discussed in elementary calculus books)

Example of Average Value of a Time Series

- let $x(\cdot)$ be step function taking on values $x_0, x_1, \ldots, x_{15}$ over 16 equal subintervals of $[a, b]$:



- here we have

$$\frac{1}{b-a}\int_a^b x(t)\,dt = \frac{1}{16}\sum_{j=0}^{15} x_j = \text{ height of dashed line}$$

Average Values at Different Scales and Times

- define the following function of λ and t

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

  – λ is width of interval – refered to as 'scale'

  – t is midpoint of interval

- $A(\lambda, t)$ is average value of $x(\cdot)$ over scale λ centered at t
- average values of time series have wide-spread interest

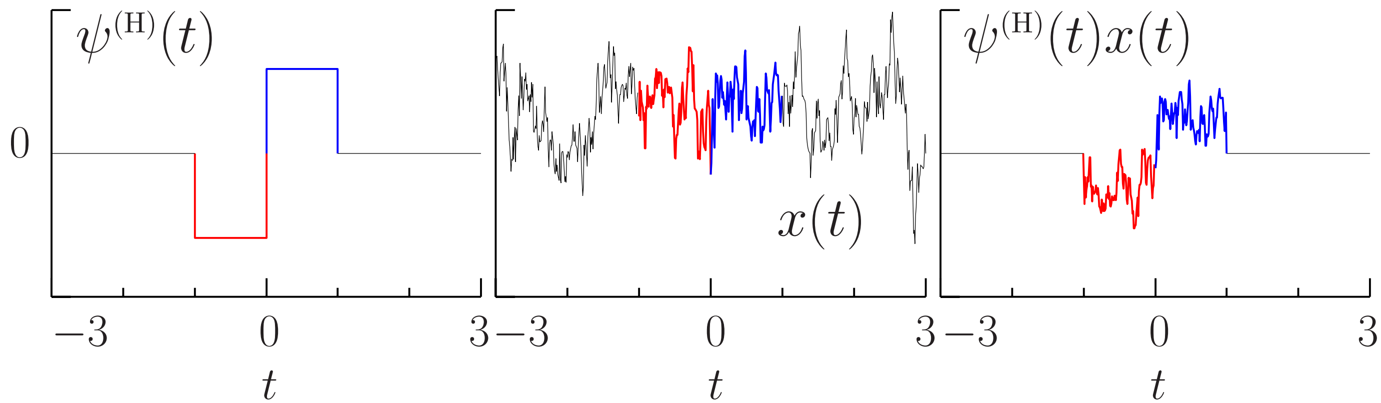
  – one second average temperatures over forest

  – ten minute rainfall rate during severe storm

  – yearly average temperatures over central England

Defining a Wavelet Coefficient W

- multiply Haar wavelet & time series $x(\cdot)$ together:



- integrate resulting function to get 'wavelet coefficient' $W(1,0)$:

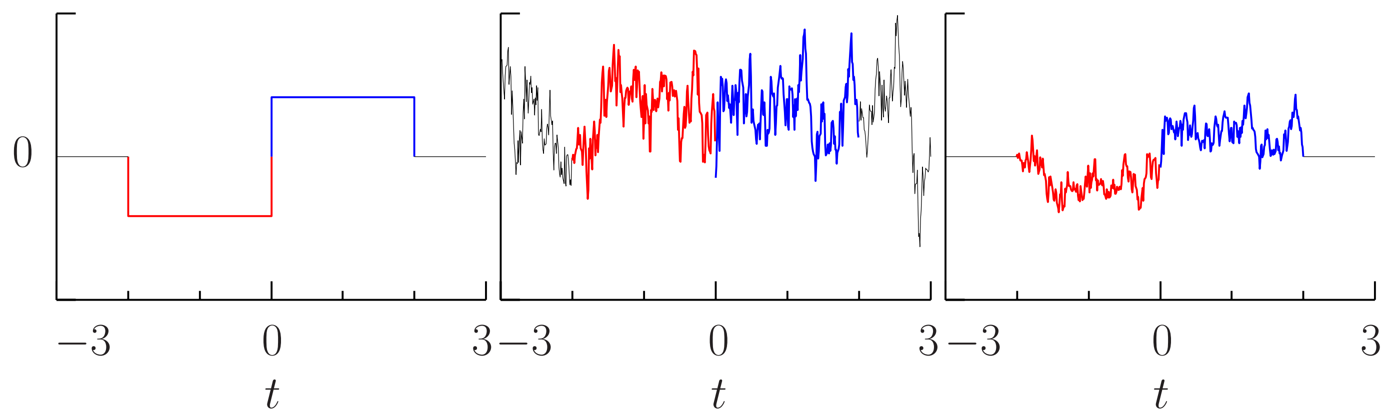
$$\int_{-\infty}^{\infty} \psi^{(\mathrm{H})}(t)x(t)\,dt = W(1,0)$$

- to see what $W(1,0)$ is telling us about $x(\cdot)$, note that

$$W(1,0) \propto \frac{1}{1}\int_0^1 x(t)\,dt - \frac{1}{1}\int_{-1}^0 x(t)\,dt = A(1,\tfrac{1}{2}) - A(1,-\tfrac{1}{2})$$

Defining Wavelet Coefficients for Other Scales

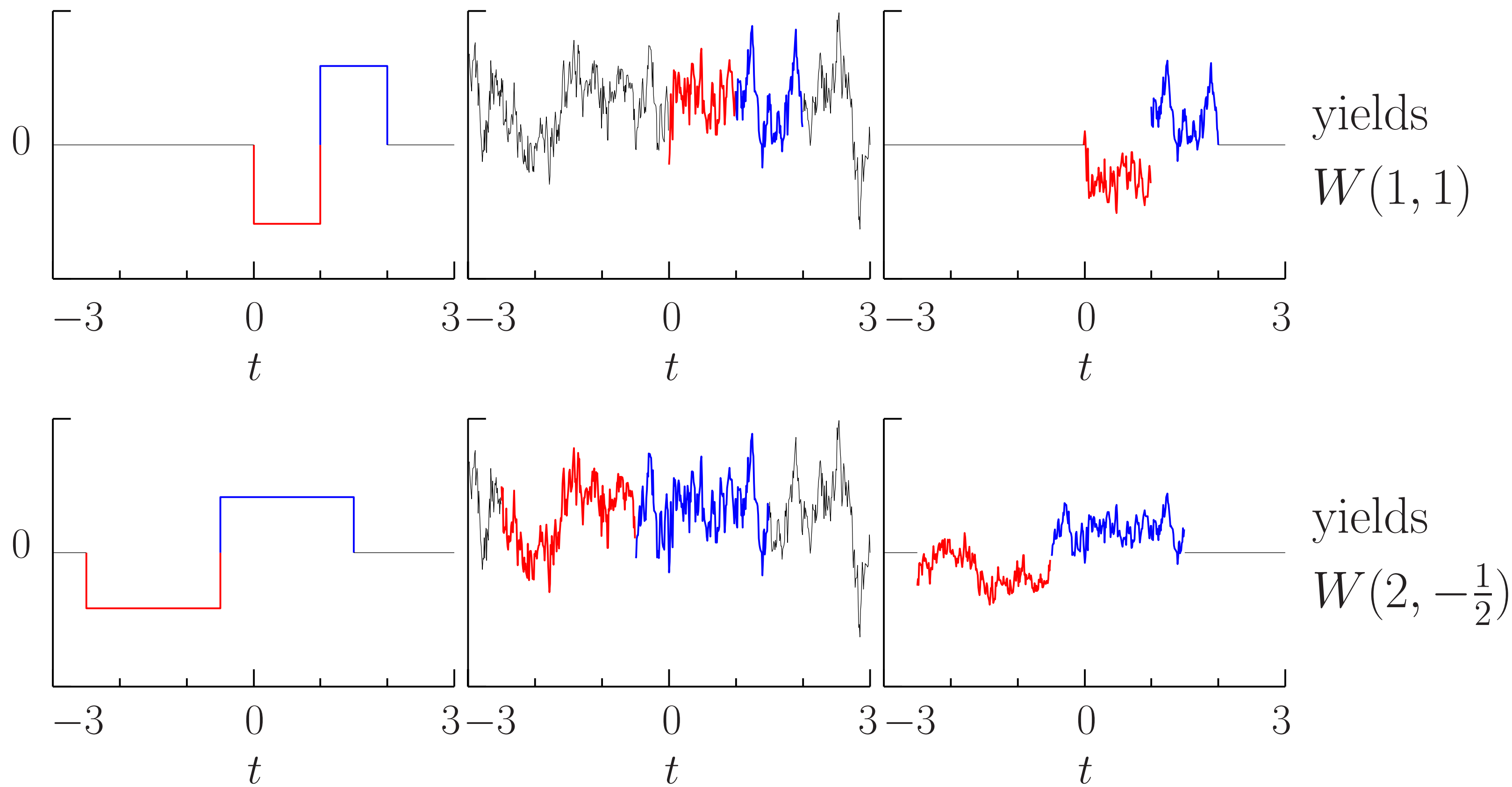
- $W(1,0)$ proportional to difference between averages of $x(\cdot)$ over $[-1,0]$ & $[0,1]$, i.e., two unit scale averages before/after $t=0$
  - '1' in $W(1,0)$ denotes scale 1 (width of each interval)
  - '0' in $W(1,0)$ denotes time 0 (center of combined intervals)
- stretch or shrink wavelet to define $W(\tau,0)$ for other scales τ:



yields $W(2,0)$

Defining Wavelet Coefficients for Other Locations

- relocate to define $W(\tau, t)$ for other times t:



Haar Continuous Wavelet Transform (CWT)

- for all $\tau > 0$ and all $-\infty < t < \infty$, can write
$$W(\tau, t) = \frac{1}{\sqrt{\tau}} \int_{-\infty}^{\infty} x(u) \psi^{(\mathrm{H})} \left(\frac{u-t}{\tau} \right) du$$
  - $\frac{u-t}{\tau}$ does the stretching/shrinking and relocating
  - $\frac{1}{\sqrt{\tau}}$ needed so $\psi^{(\mathrm{H})}_{\tau,t}(u) \equiv \frac{1}{\sqrt{\tau}} \psi^{(\mathrm{H})}\left(\frac{u-t}{\tau}\right)$ has unit energy
  - since it also integrates to zero, $\psi^{(\mathrm{H})}_{\tau,t}(\cdot)$ is a wavelet
- $W(\tau, t)$ over all $\tau > 0$ and all t is Haar CWT for $x(\cdot)$
- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of averages

Other Continuous Wavelet Transforms: I

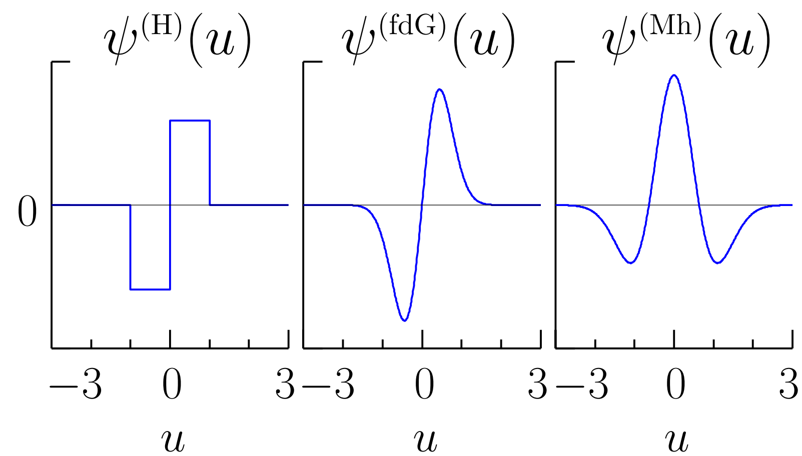
- can do the same for wavelets other than the Haar
- start with basic wavelet $\psi(\cdot)$
- use $\psi_{\tau,t}(u) = \frac{1}{\sqrt{\tau}}\psi\left(\frac{u-t}{\tau}\right)$ to stretch/shrink & relocate
- define CWT via

$$W(\tau, t) = \int_{-\infty}^{\infty} x(u)\psi_{\tau,t}(u)\,du = \frac{1}{\sqrt{\tau}}\int_{-\infty}^{\infty} x(u)\psi\left(\frac{u-t}{\tau}\right)du$$

- analyzes/breaks up/decomposes $x(\cdot)$ into components
  - associated with a scale and a time
  - physically related to a difference of *weighted* averages

Other Continuous Wavelet Transforms: II

- consider two buddies of Haar wavelet



- $\psi^{(\mathrm{fdG})}(\cdot)$ proportional to 1st derivative of Gaussian PDF
- 'Mexican hat' wavelet $\psi^{(\mathrm{Mh})}(\cdot)$ proportional to 2nd derivative
- $\psi^{(\mathrm{fdG})}(\cdot)$ looks at difference of adjacent weighted averages
- $\psi^{(\mathrm{Mh})}(\cdot)$ looks at difference between weighted average and sum of weighted averages occurring before & after

First Hairy-Looking Equation

- CWT equivalent to $x(\cdot)$ because we can write

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

  where C is a constant depending on specific wavelet $\psi(\cdot)$
- can synthesize (put back together) $x(\cdot)$ given its CWT; i.e., nothing is lost in reexpressing time series $x(\cdot)$ via its CWT
- regard stuff in brackets as defining 'scale τ' time series at t
- says we can reexpress $x(\cdot)$ as integral (sum) of new time series, each associated with a particular scale
- similar additive decompositions are a central theme of wavelet analysis

Second Hairy-Looking Equation

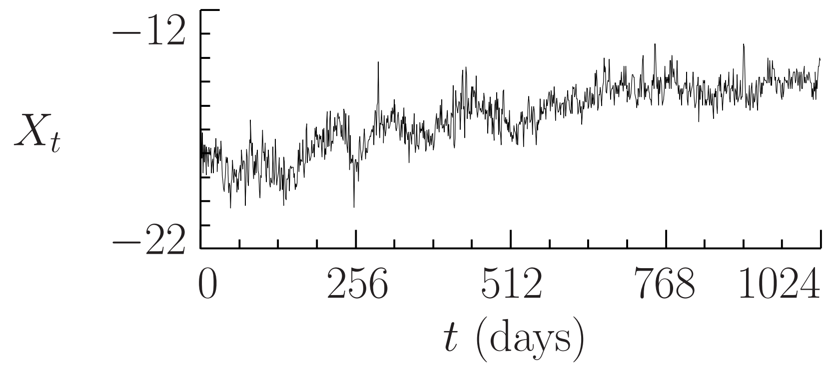
- energy in $x(\cdot)$ is reexpressed in CWT because

$$\text{energy} = \int_{-\infty}^{\infty} x^2(t)\,dt = \int_0^{\infty} \left[\frac{1}{C\tau^2}\int_{-\infty}^{\infty} W^2(\tau,t)\,dt\right] d\tau$$

- can regard $x^2(t)$ versus t as breaking up the energy across time (i.e., an ‘energy density’ function)
- regard stuff in brackets as breaking up the energy across scales
- says we can reexpress energy as integral (sum) of components, each associated with a particular scale
- function defined by $W^2(\tau,t)/C\tau^2$ is an energy density across both time and scale
- similar energy decompositions are a second central theme

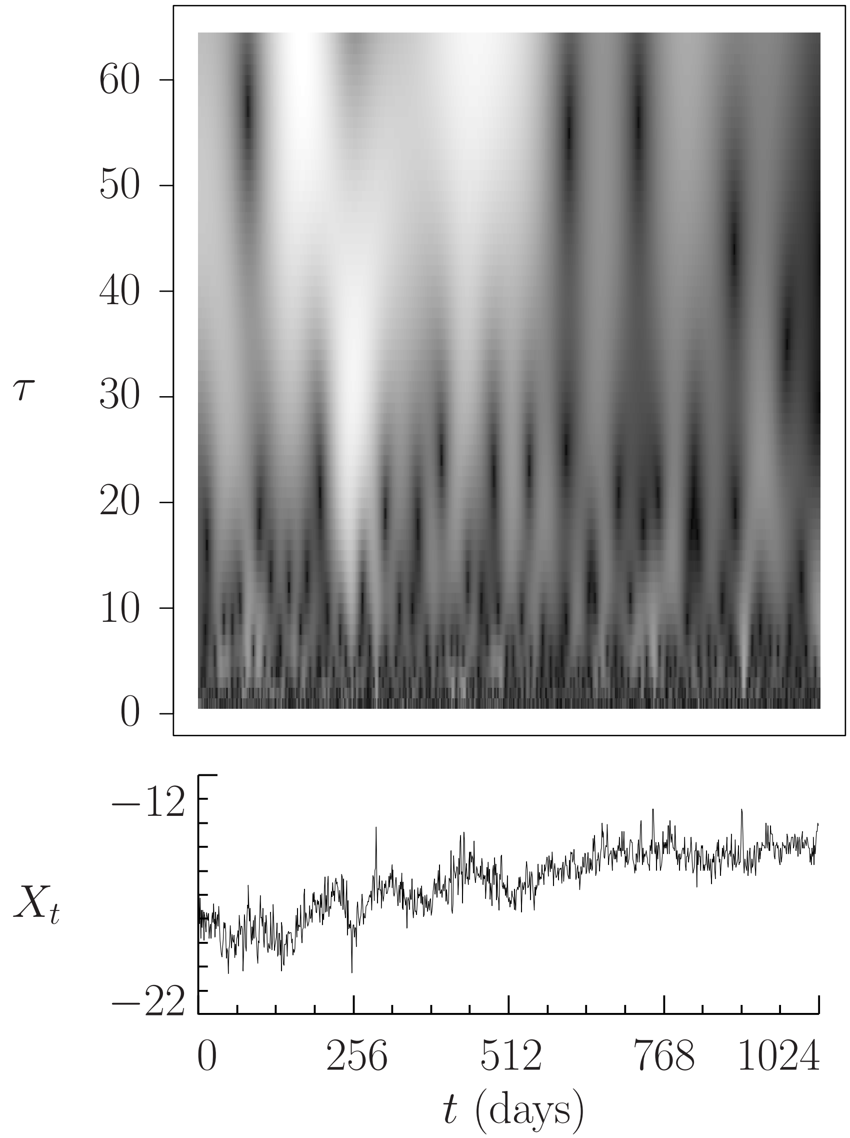
Example: Atomic Clock Data

- example: average daily frequency variations in clock 571

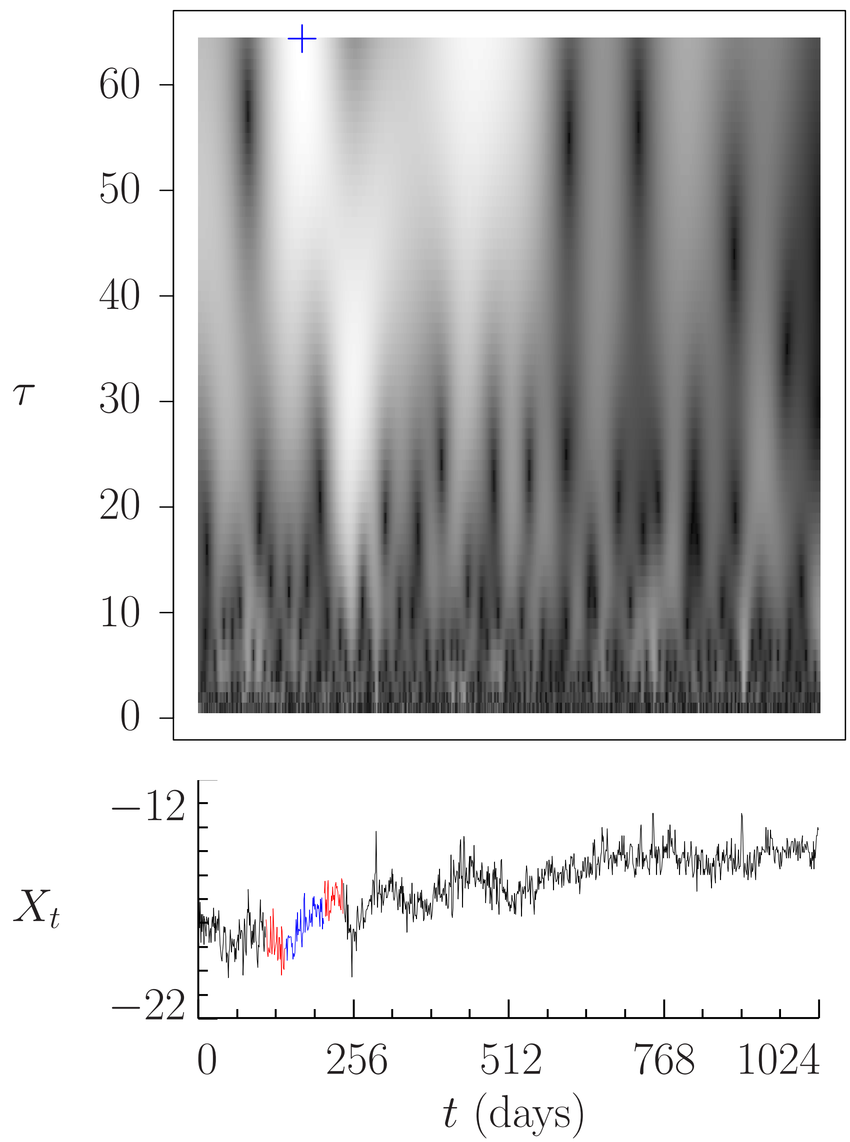


- t is measured in days (one measurment per day)
- plot shows X_t versus integer t
- $X_t = 0$ would mean that clock 571 could keep time perfectly
- $X_t < 0$ implies that clock is losing time systematically
- can easily adjust clock if X_t were constant
- inherent quality of clock related to changes in averages of X_t

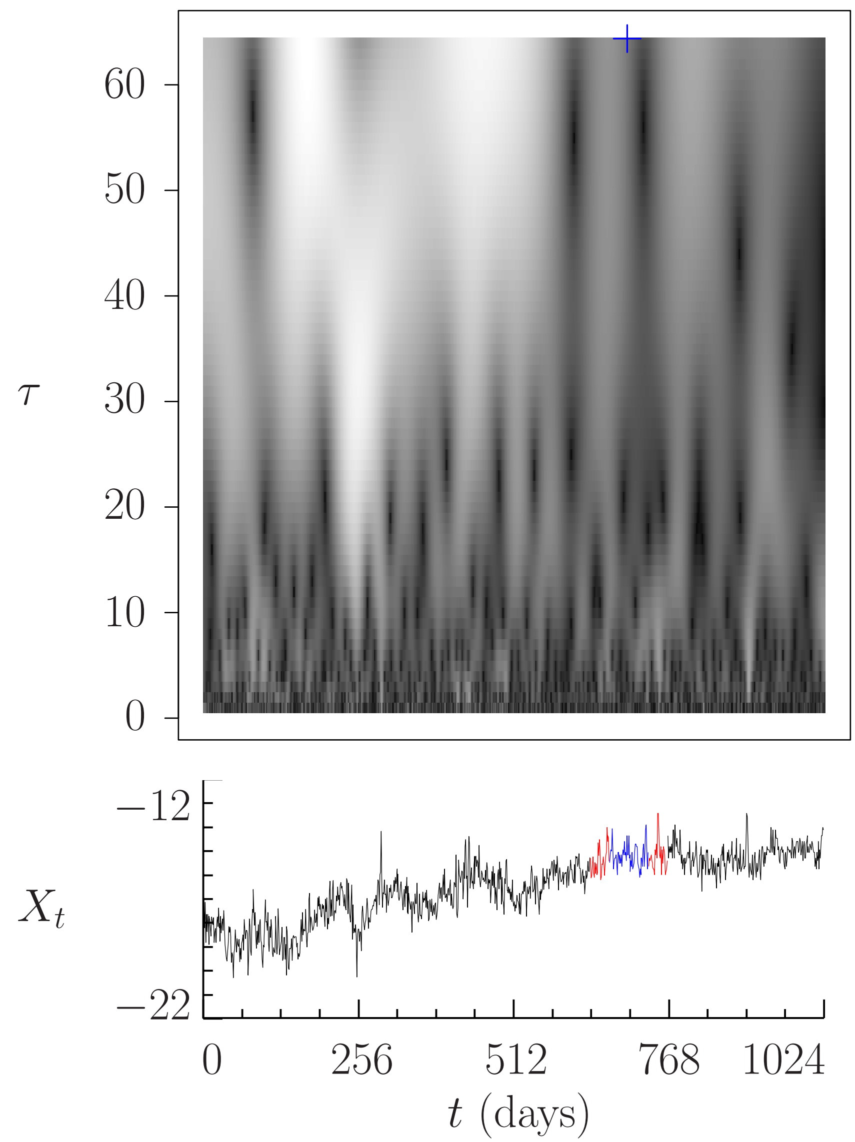
Mexican Hat CWT of Clock Data: I



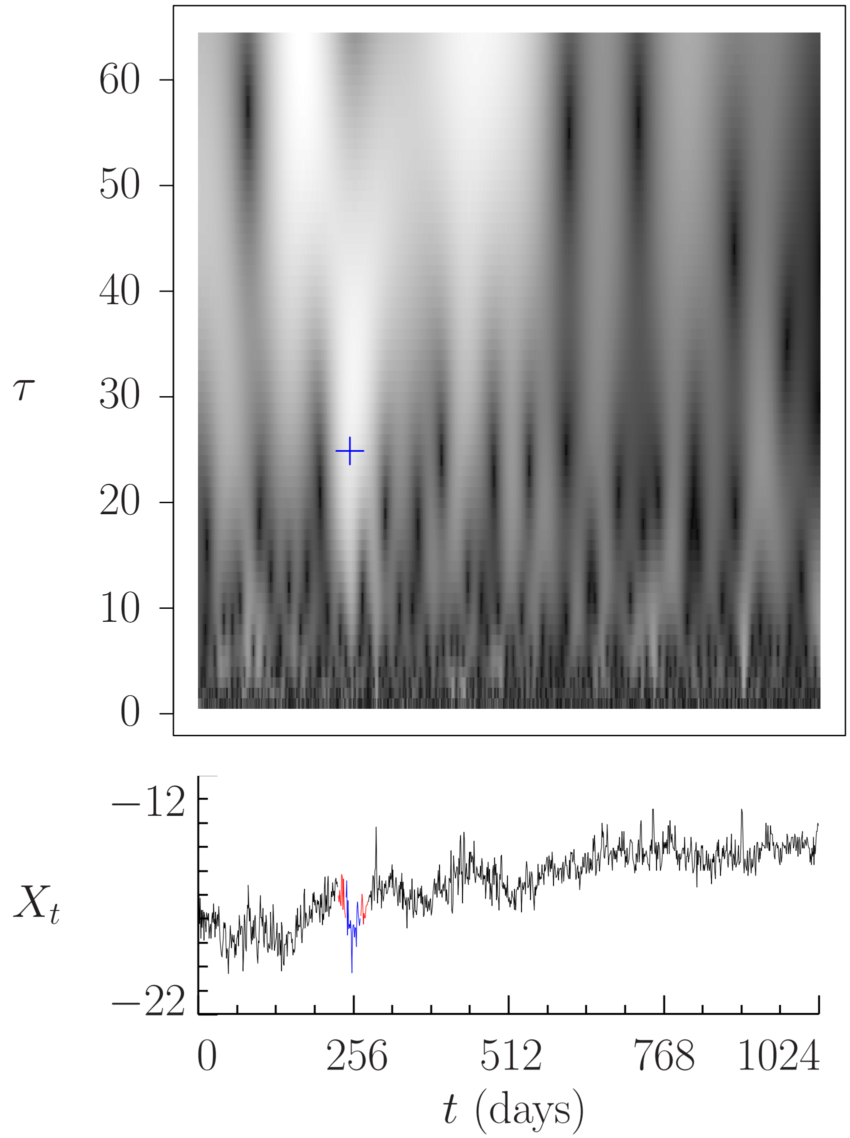
Mexican Hat CWT of Clock Data: II



Mexican Hat CWT of Clock Data: III

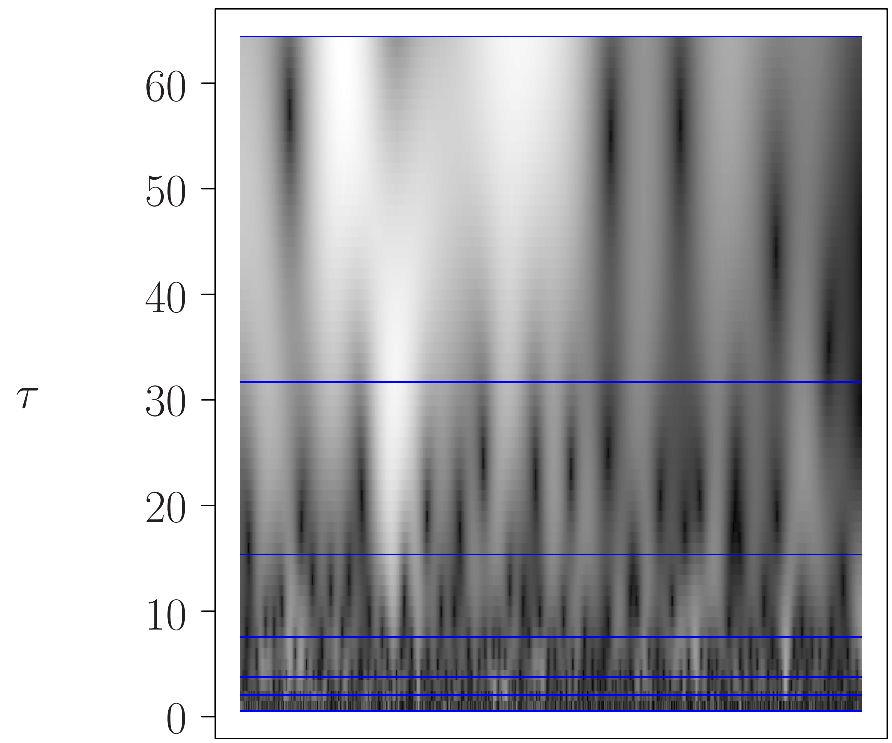


Mexican Hat CWT of Clock Data: IV



Beyond the CWT: the DWT

- can often get by with subsamples of $W(\tau, t)$
- leads to notion of discrete wavelet transform (DWT)



The Discrete Wavelet Transform: I

- when dealing with samples x_0, x_1, $\ldots x_{N-1}$ 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 $\lambda \geq 2\tau_{J_0} = 2^{J_0}$

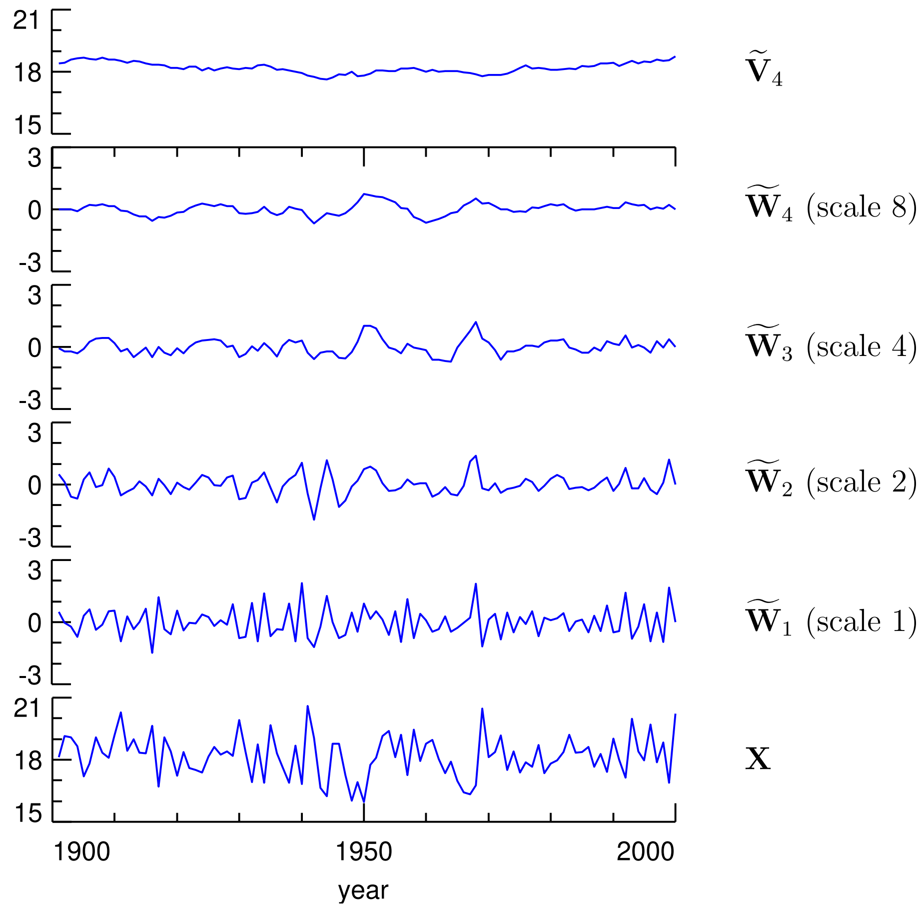
The Discrete Wavelet Transform: II

- 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: I



Example: DWT of Boreal Time Series: I

- 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:

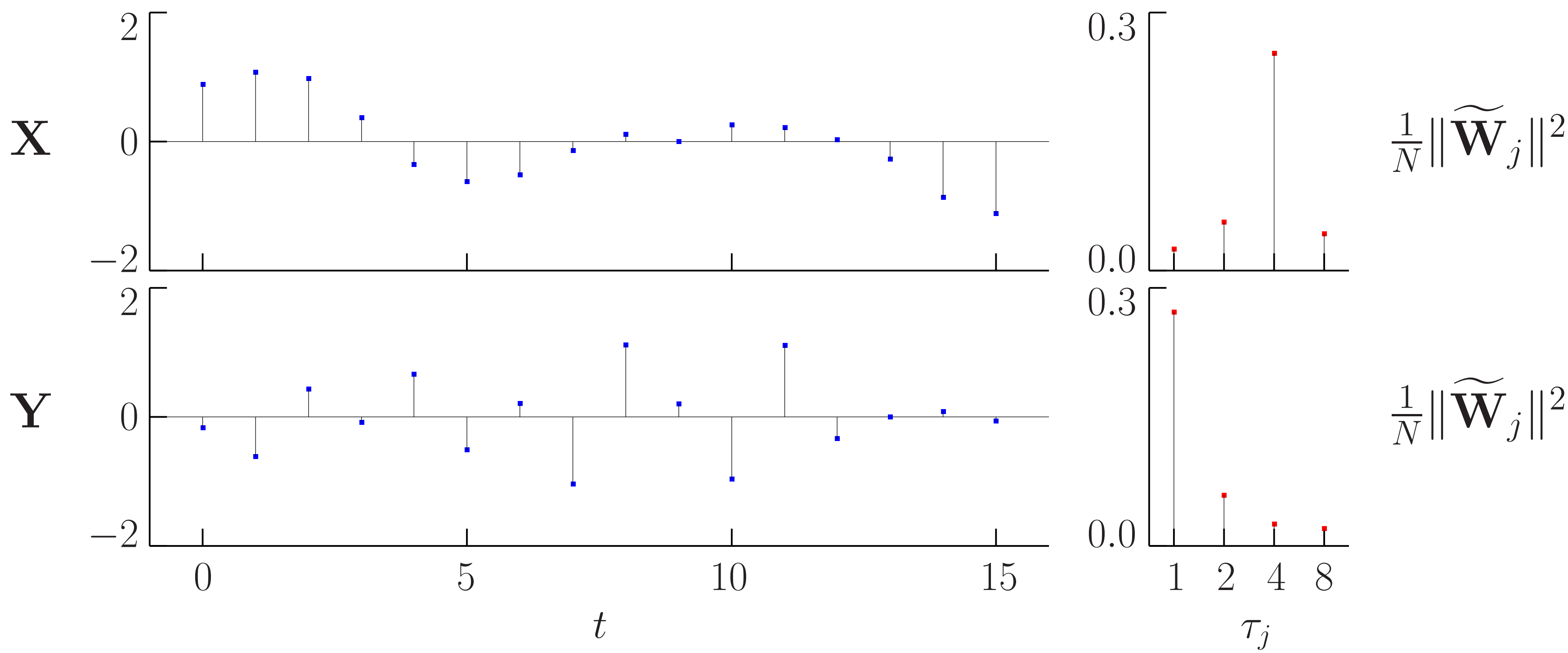
$$\hat{\sigma}_x^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (x_t - \bar{x})^2 = \frac{1}{N} \left(\sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2 \right) - \overline{x}^2,$$

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

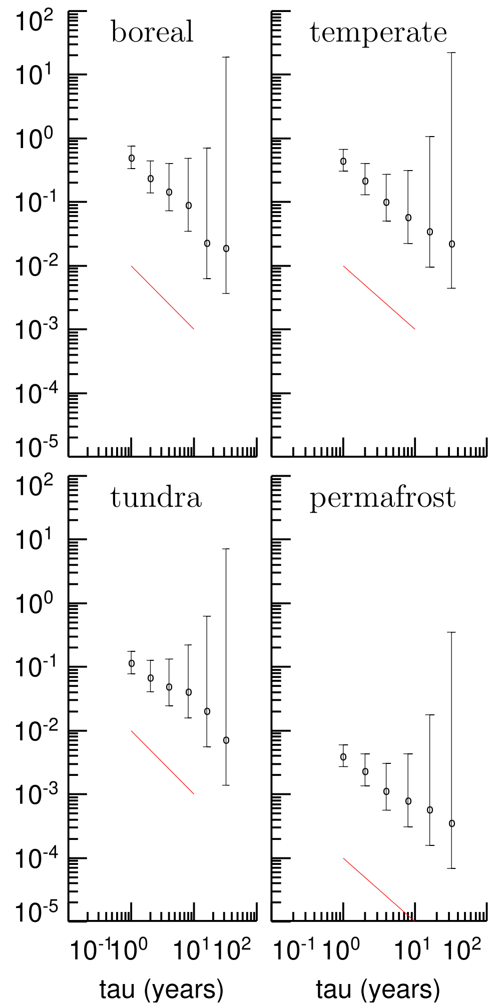
- $\frac{1}{N}\|\widetilde{\mathbf{W}}_j\|^2$ portion of $\hat{\sigma}_X^2$ due to changes in averages over scale τ_j; i.e., 'scale by scale' analysis of variance

Wavelet Variance: II

- wavelet variances for time series $\mathbf{X}$ and $\mathbf{Y}$ of length $N = 16$, each with zero sample mean and same sample variance



Wavelet Variances for Vegetation Time Series: I

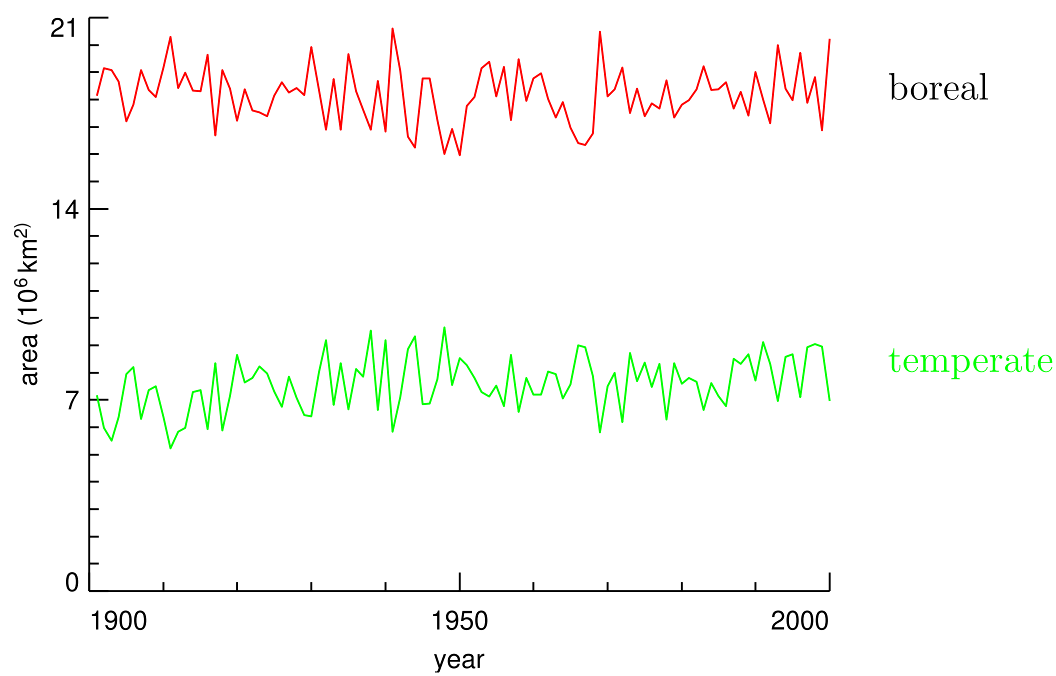


Wavelet Variances for Vegetation Time Series: II

- 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: I

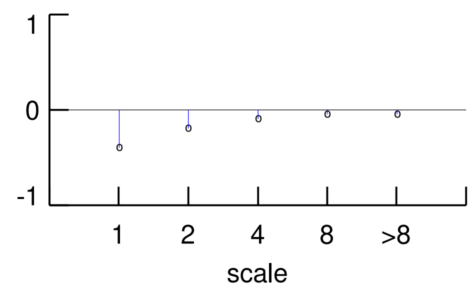
- reconsider boreal and temperate time series:



- sample cross-correlation is -0.79; i.e., series are anticorrelated

Wavelet Covariance: II

- 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: I

- 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

Concluding Remarks: II

- *many* other uses for wavelets (barely scratched the surface!)
  - approximately decorrelate certain time series (DWT needed)
  - assessing sampling properties of statistics (DWT best)
  - signal extraction ('wavelet shrinkage'; DWT best)
  - edge identification in images (CWT best)
  - compression of time series/images (DWT needed)
  - fast simulation of time series/images (DWT needed)

Upcoming Courses

- Stat 530/EE 524 (Spring Quarter)
- short course at APL, Sept. 12, 13 & 15, 2006
  - details will be posted within a month at

    http://www.apl.washington.edu/