Some History of Multitaper

David J. Thomson, Queen's Univ., Kingston, Ontario
djt@queensu.ca

BIRS East, June 24,25, 2022

Thanks to: My Students, CRC, Queen's, NSERC,
(the old) Bell Labs, Murray Hill, NJ, &
Maja–Lisa Thomson

— — A series of random events — —

"I have the pleasure of (not) seeing my audience nod approbation while they sleep." — Sidney Smith

40^{th} Anniversary of 1982 paper!

Outline

- Background and History
- Introduction to spectrum estimation: Kingston temperature.
- CO_2 and Temperature, 1958–2011. Problem: Annual Cycle.
- Central England Temperature, 1659–2010. *Not* 1 cycle/year.
- What's New: Later!
- Summary

Some Personal History

(1961) Started at Acadia. Met Maja–Lisa. Read: *"Prolate Spheroidal Wave Functions (PSWF), Fourier Analysis and Uncertainty"*.

(1964) ". . . there is a lot of work being done on time series, but I have a feeling it's on the wrong track." — K. D. C. Haley

(1965) Started work at Bell Labs, Murray Hill, NJ.
Fortunate coincidence: Tukey & Cooley published the FFT

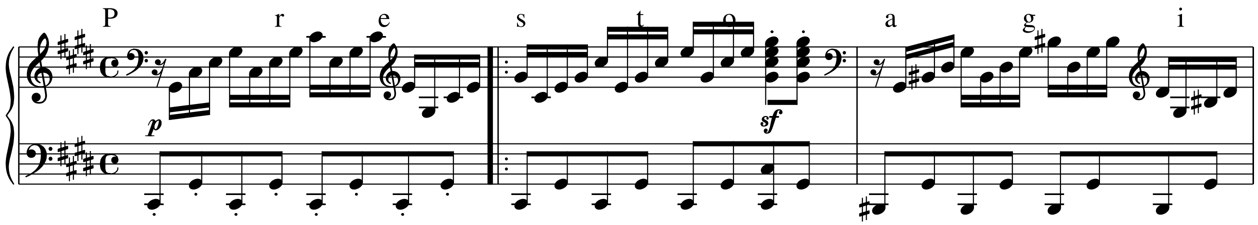
(66–77) Millimeter waveguide project. "Classical" spectrum estimation, robustness, quality control, special functions.
First multitaper, c. 1974

(77–84) Cellular phones: highly non–Gaussian data, hardware spectrum estimation, hypothesis testing, robustness, circuits.
(1982) "Spectrum estimation and harmonic analysis"
(1983) Scripps Institution of Oceanography

(84–01) "Mathematics of Communications" research department.
climate, laser reliability, induced voltages on ocean cables and power systems, satellite anomalies, Ulysses spacecraft project,
Touch–tone detectors, test ban treaty, *etc*.

(02–??) Math and Statistics, Queens University

What is a Spectrum?



Time: Left to Right Frequency: Bottom to Top

Notation: Volume, duration, waveform

Spectral Representation, discrete time, $t = \cdots, -1, 0, 1, \cdots$

$$x(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi ft}\, dX(f)$$

Statistics of **Stationary Processes** do not depend on time

$$\mathbf{E}\{dX(f)dX^*(\xi)\} = S(f)\,\delta(f-\xi)df\,d\xi\,; S(f) \geq 0$$

$S(f)$ is the *spectrum* and defines variance as a function of frequency, f. Its Fourier transform is the **Autocovariance**

$$R(\tau) = \mathbf{E}\{x(t)x(t+\tau)\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} S(f)e^{i2\pi f\tau}df$$

– Einstein (1914)

The Spectrum — some history

(1671) Newton named the spectrum from the Greek "to see"
(1822) Fourier, "Théorie analytique de la chaleur"
(1854–) Sir William Thomson & Sir George G. Stokes, Letters
"On the nature and possibilities of spectrum analysis"
(1870–1890) various, "Empirical Fourier Series"
(1889) Stokes suggested the form of the periodogram
for studying common perodicities in the Sun and climate
(1898) Schuster named "the Periodogram" in a paper on statistical deficiencies of earlier work and gave its distribution

$$P(f) = \frac{1}{T}\left[\left| \int_0^T x(t) \cos 2\pi f t \, dt \right|^2 + \left| \int_0^T x(t) \sin 2\pi f t \, dt \right|^2 \right]$$

(1903) Rayleigh: Periodogram doesn't converge (inconsistent)
(1912) Needs "some way to take the edge off"
– Fermi's comment!

The Spectrum — some practical problems

A Major Problem: Effective Sample Size

Music (CD): 44,000 samples/sec, each channel

Climate: *Longest* daily series (Uppsala, 1722 to present)

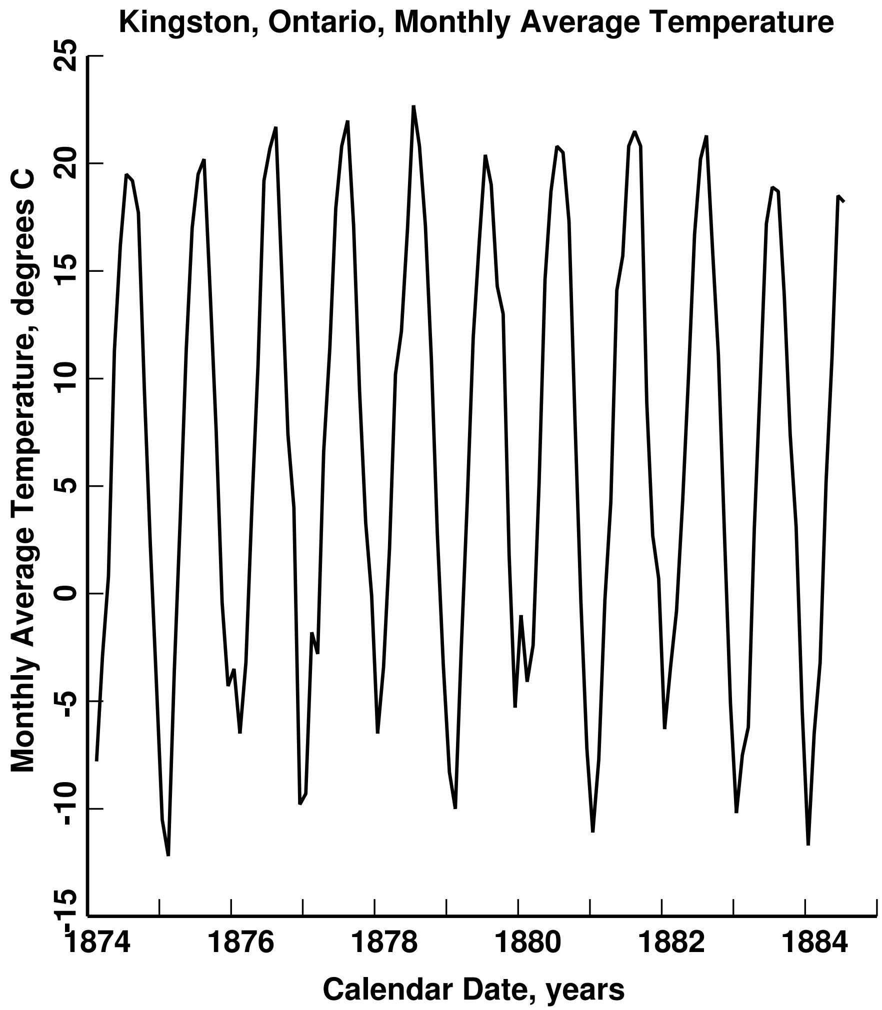
$\sim 112,000$ samples

Climate spectra are analogous to the last few notes...

"$\cdots$ time series is the worst subject to teach. First, you have to teach the standard theory. Then, if you are being honest, you have to tell the students 'None of this stuff works, and this is what people really do'."

Doug Martin (U of Washington), 1979 RSS

An example: Ten Years of Kingson Temperature Data



Kingston, Ontario

Temperature

Degrees C

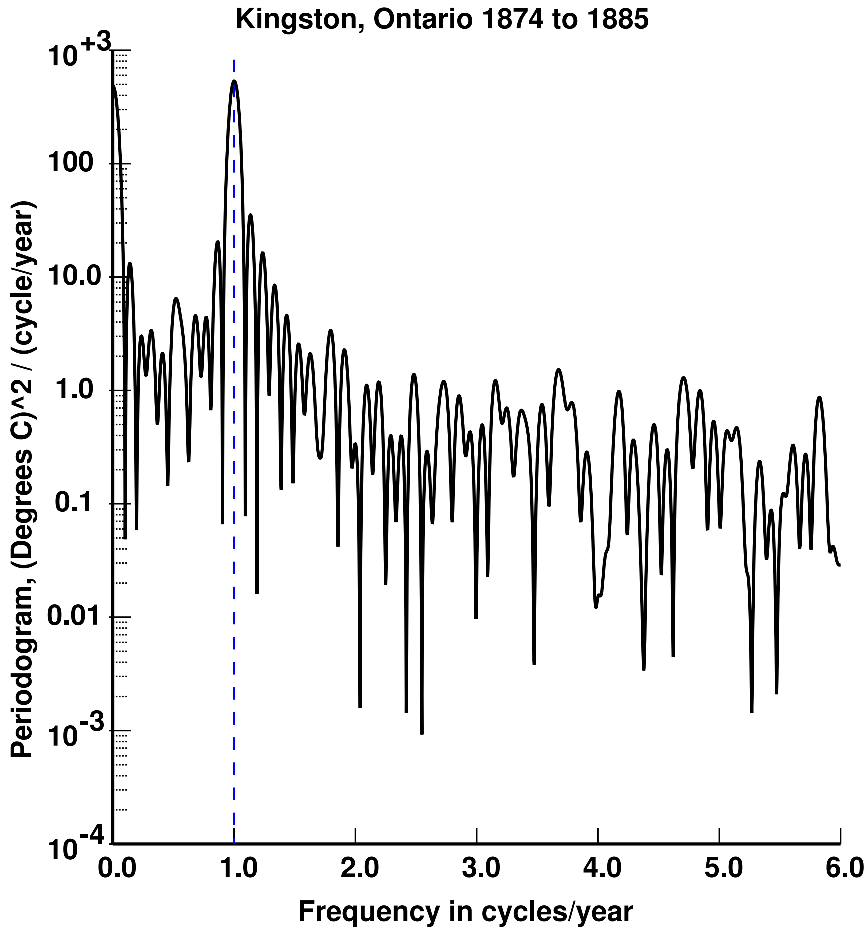
Monthly Ave

1874 – 1885

(120 Samples)

From ncdc.noaa.gov

First Method for Estimating the Spectrum, The Periodogram = |Fourier Transform{Data}|^2



Invented: Stokes (1889)

Gets annual cycle,

but not much else

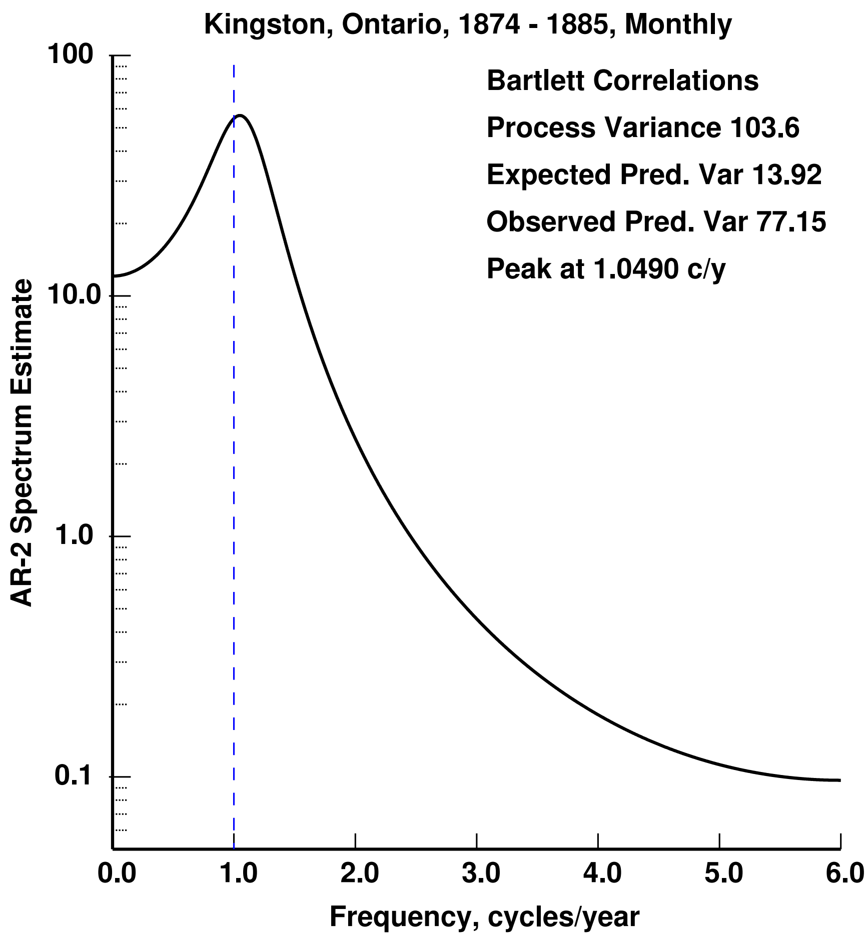
Problems:

– Schuster (1898)

– Rayleigh (1903)

– Tukey (several)

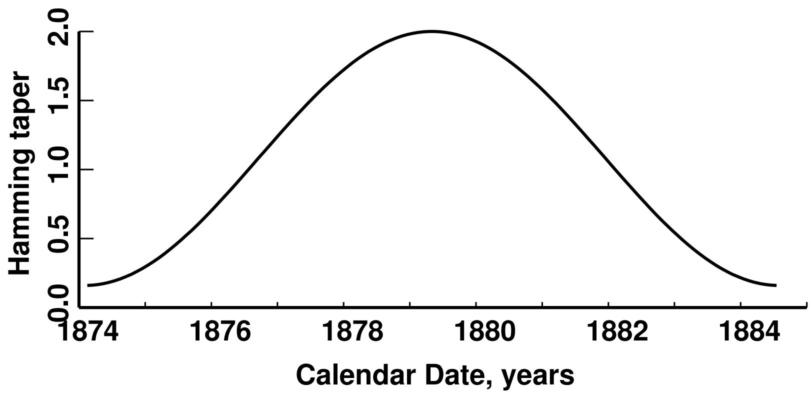
Second (Bad) Method for Estimating the Spectrum,
A Parametric (AR, MA, ARMA, *etc.*) Model



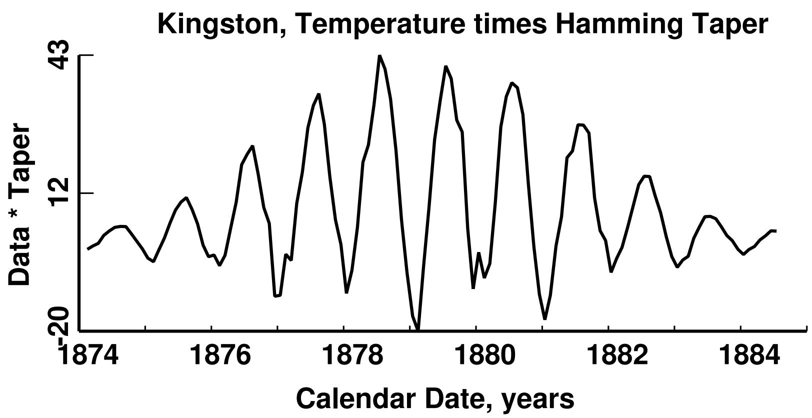
Autoregressive, AR-2
(Bartlett Correlations)
Invented:
– Yule (1927)
– Walker (1931)

Problems:
Quenouille: depends on δt
Arato: not sufficient
Tukey: (many)
Here:
Wrong Length of Year
Wrong line shape
Numerically unstable

A Better Method, Part 1, Taper the Data

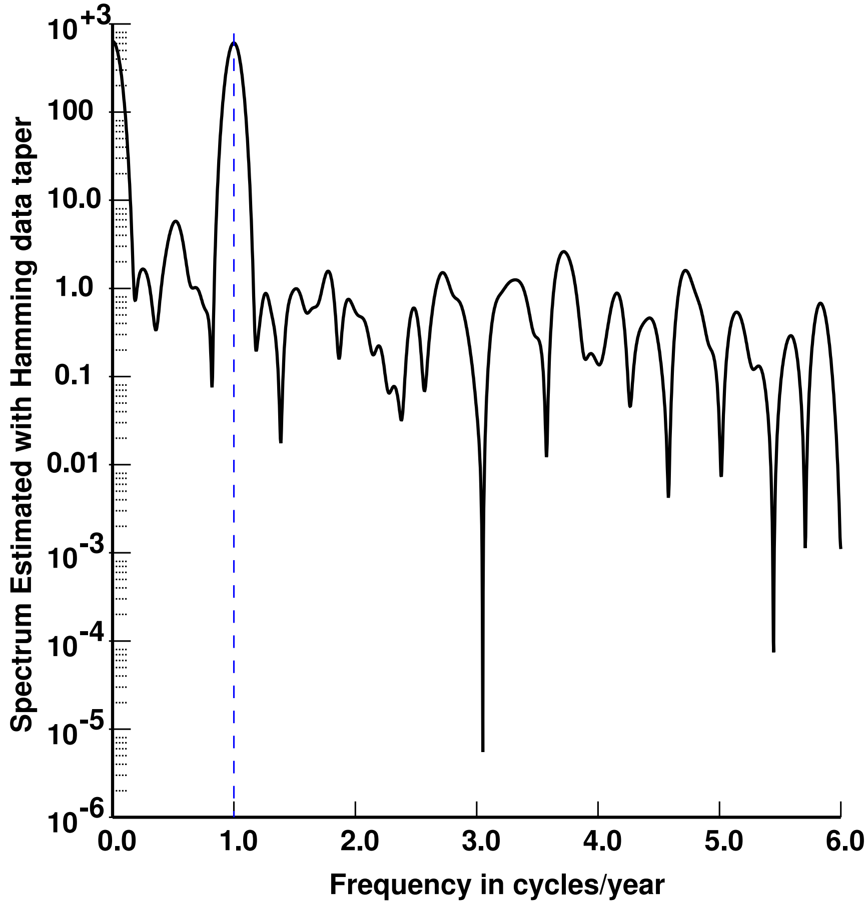


A Hamming Taper



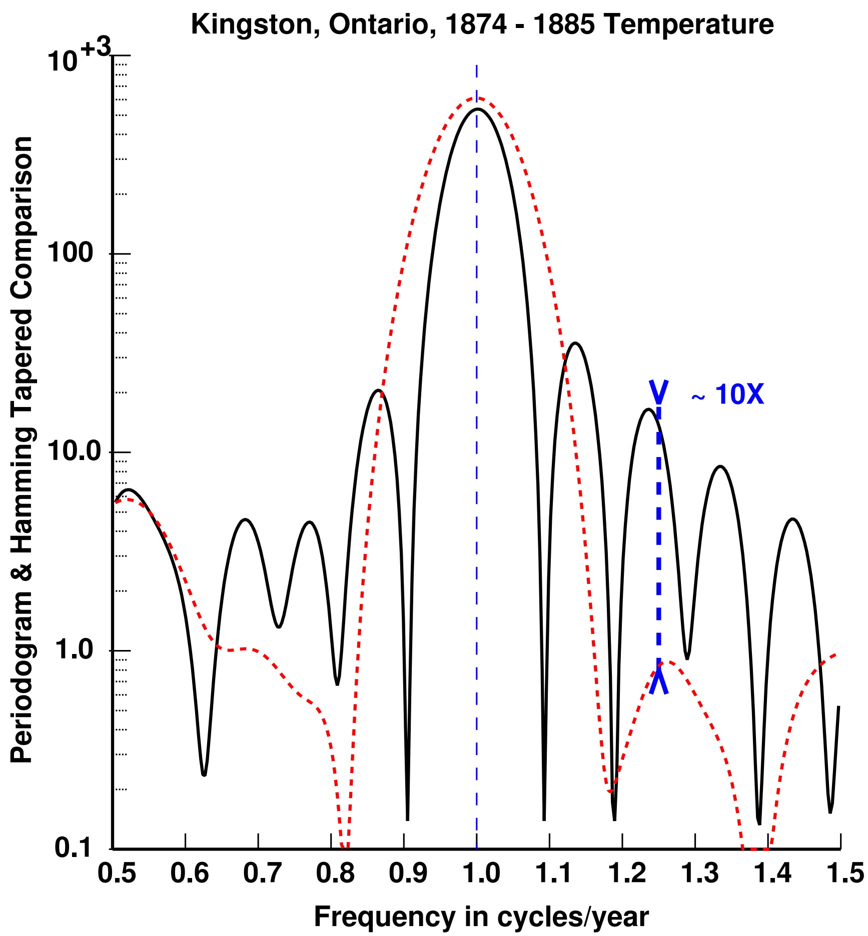
Kingston Temperature
× Hamming Taper

A Better Method, Part 2, Hamming Tapered Spectrum



Hamming Taper
"Direct" Estimate
Hamming (1950)
Tukey (1959,1967)
Lower bias, but
still inconsistent

Hamming Tapered Spectrum and Periodogram Compared



Two spectrum estimates

Periodogram

Hamming Taper

Hamming: wider peaks
Large differences (factor of 10 or more) in power.
State–of–the–art, 1950.

Windowed Estimates, Tukey (1959, 1966)

Tapered estimates have much lower bias than the periodogram.
Variance = (Expected Value $)^2$, implies lower variance.

A problem: all the data is equally valid, however;
different tapers give significantly different estimates,
typical tapers weight the center of data higher than at the ends.
–Best polynomial approximation (Chebyshev) gives high weights at the ends of the interval.
–Shannon theory gives higher weights to less predictable (so more informative) parts of data, the ends.

Only theory (c1970):

John Tukey and Norbert Wiener said it's a good idea!

"If one were not blinded by the mathematical elegance of the conventional approach, making unfounded assumptions as to the values of unmeasured data and changing the data values that one knows would be totally unacceptable"

— Burg *Thesis* (Stanford, 1975)

Comments from my Bell Labs Department Heads

Consider if a problem is worth doing. Do a back–of–the–envelope calculation. If it won't save the company at least a million dollars, don't waste your time on it. — M. C. Biskeborn, 1965

"I can find at least 100 different ways to estimate the spectrum in the literature. What I want to know from you is which is the right one, and why. And, if you are wrong, we are talking about a measurable fraction of the US GNP" — D. A. Alsberg, c1974

Multitapers — I: Fundamental Integral Equation

Fourier Transform of available data: Equivalent to data

$$y(f) = \sum_{t=0}^{N-1} x(t) e^{-i2\pi f t} \tag{1}$$

Spectral Representation *Definition of the Spectrum*

$$x(t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi\xi t} dX(\xi) \qquad \mathbf{E}\{|dX(f)|^2\} = S(f)df \tag{2}$$

(1) & (2) give a convolution. Think of it as an integral equation.

$$y(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\sin N\pi(f-\xi)}{\sin \pi(f-\xi)} dX(\xi) \tag{3}$$

Do a least–squares solution in a series of Slepian functions, $V_k(f)$. $k = 0, 1, \cdots K-1$. Typically $K = 4$ to 20 terms.

$$\lambda_k V_k(f) = \int_{-W}^{W} \frac{\sin N\pi(f-\xi)}{\sin \pi(f-\xi)} V_k(\xi) d\xi \tag{4}$$

Multitaper Solution (1982)

For a center frequency f, do an eigenfunction solution on $(f-W, f+W)$ and compute the $K = 2NW$ *eigencoefficients*

$$y_k(f) = \frac{1}{\lambda_k}\int_{-W}^{W} y(f-\xi)\, V_k(\xi) d\xi \qquad k = 0, 1, \cdots, K \lesssim \lfloor 2NW \rfloor \quad (5)$$

$\cdots$ some magic $\cdots$

$$= \sum_{t=0}^{N-1} x(t)\ v_t^{(k)}\ e^{-i2\pi ft} \quad (6)$$

$\Rightarrow$ Fourier Transform of $\{$ data $\times$ taper $\}$

$\Rightarrow$ Orthonormal expansion of $\{$ data $\times\, e^{-i2\pi ft}$ $\}$

The simplest multitaper spectrum estimate is

$$\hat{S}(f) \sim \frac{1}{K}\sum_{k=0}^{K-1} |\, y_k(f)\,|^2 \quad (7)$$

VERY IMPORTANT: $K > 1$, 8 or 10 in examples here

Best Method: Multitaper Spectrum Estimates

Derived from "First Principles", not *ad–hoc*

Two "free" parameters: Block Length & Time–Bandwidth

Optimum Energy Concentration, Efficient, Minimal bias

Numerically Stable, Uses FFT for speed

Separates Periodic Components and Background "Noise"

Jackknife over *tapers*

Approximately Maximum–Likelihood

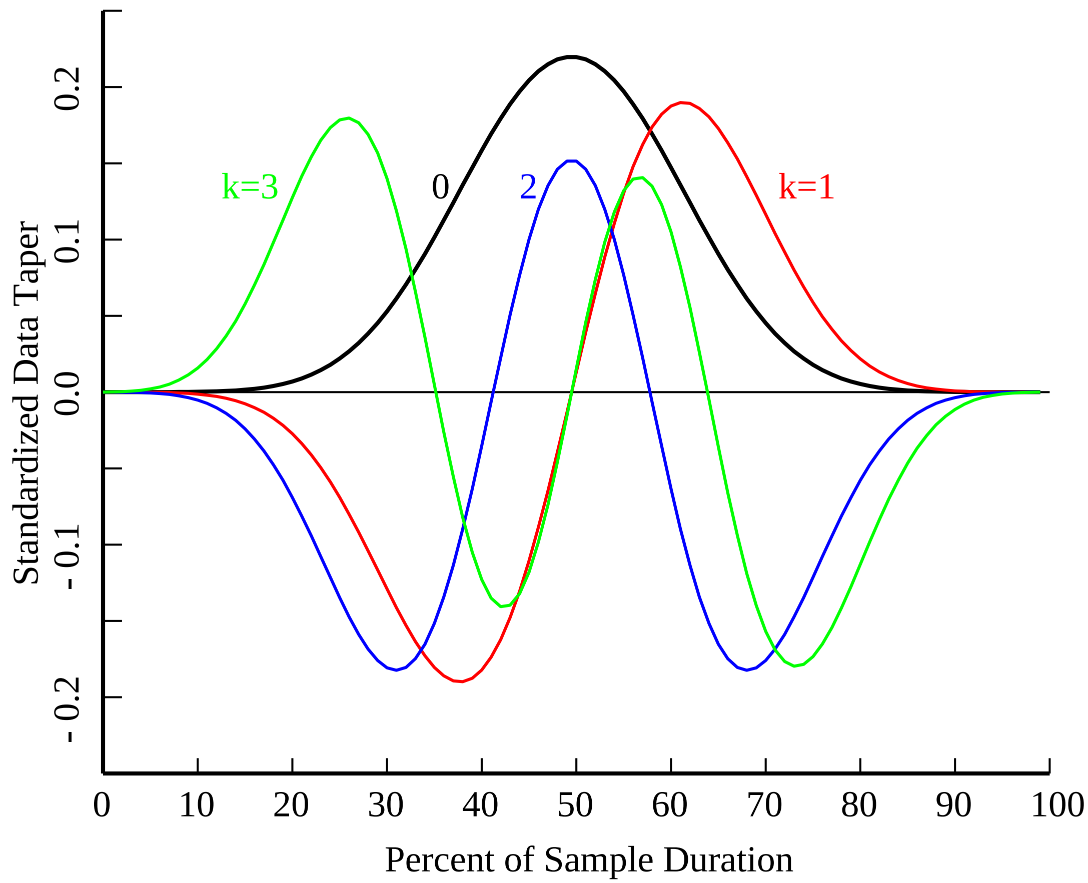
Fixes problems found by Schuster, Rayleigh, Kendall, Bartlett

Series Solution: Tukey's Direct Estimate $\approx$ first term

Maximum–likelihood: Stoica & Sundin, 1999

"People spend all their time on convergency, and never learn how to use the series." - J.J.Thomson

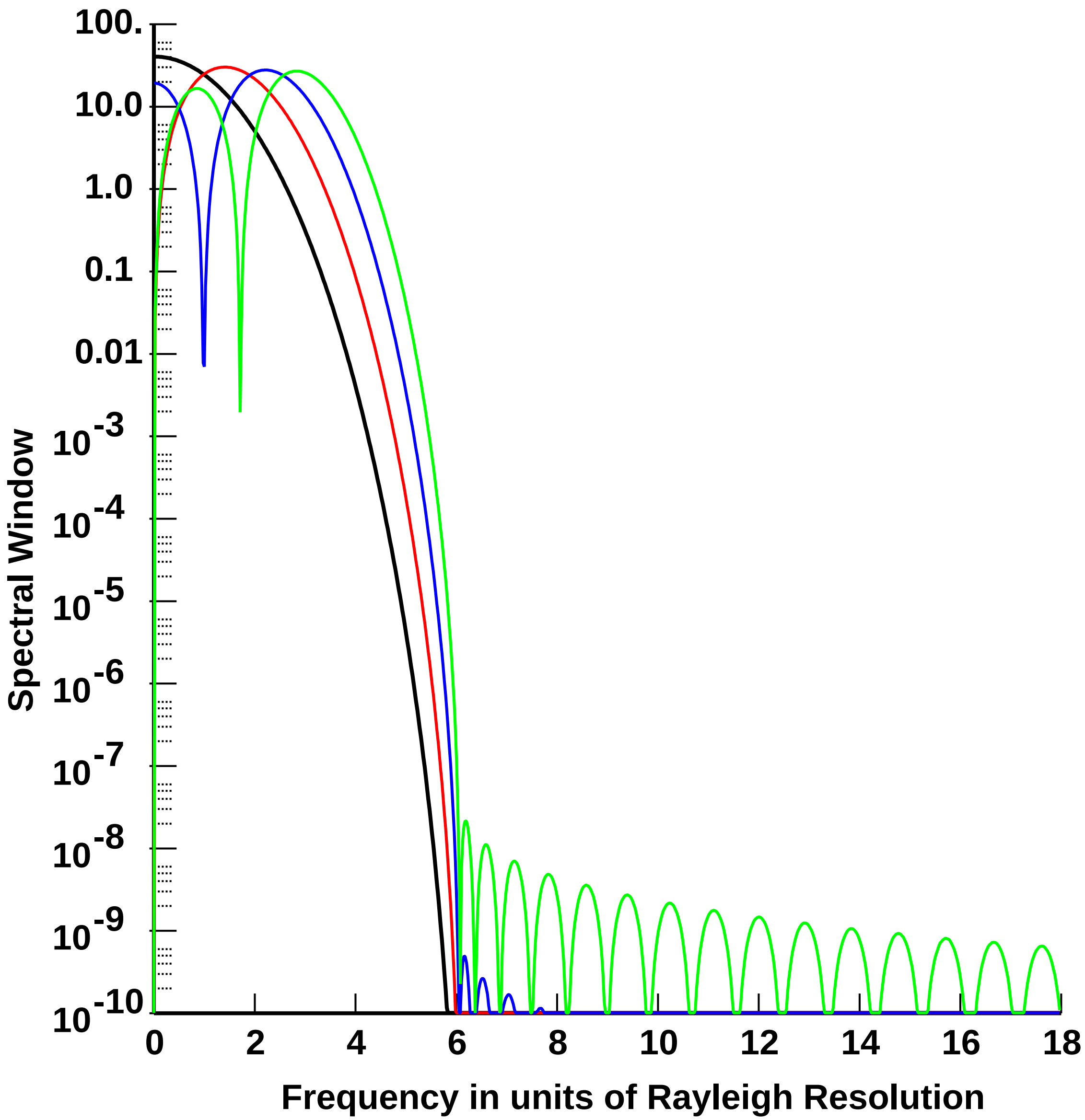
Best Tapers: Slepian Sequences $v_n^{(k)}$



Multiple Tapers
Slepian Sequences
= Discrete Prolate
Spheroidal Sequences

A complete basis.
Orthonormal.
Describes:
signals, noise,
and nonstationarity.

Corresponding spectral windows: Slepian functions



Multiple Tapers

Slepian functions

$$|V_k(f)|^2$$

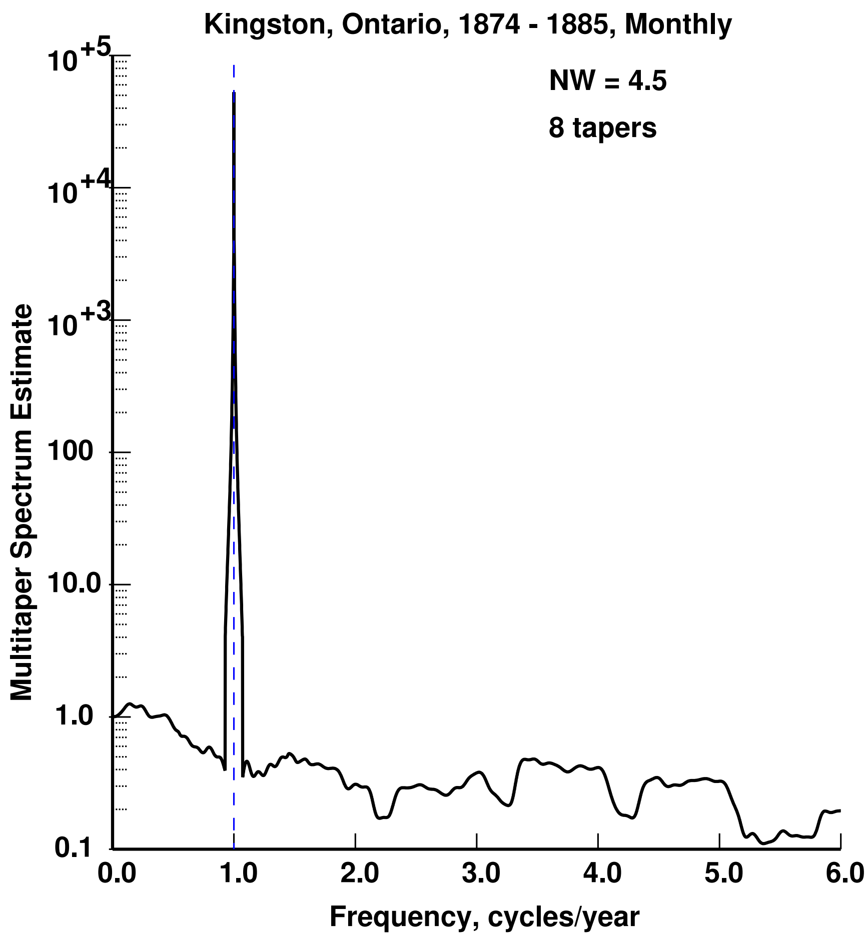
Best Energy

Concentration

for finite sequence

Doubly orthogonal

Best Method: Multitaper Spectrum Estimates



Multitaper Spectrum

Data: Kingston, Ontario

Temperature, 1874–1885

Annual cycle:

– well defined

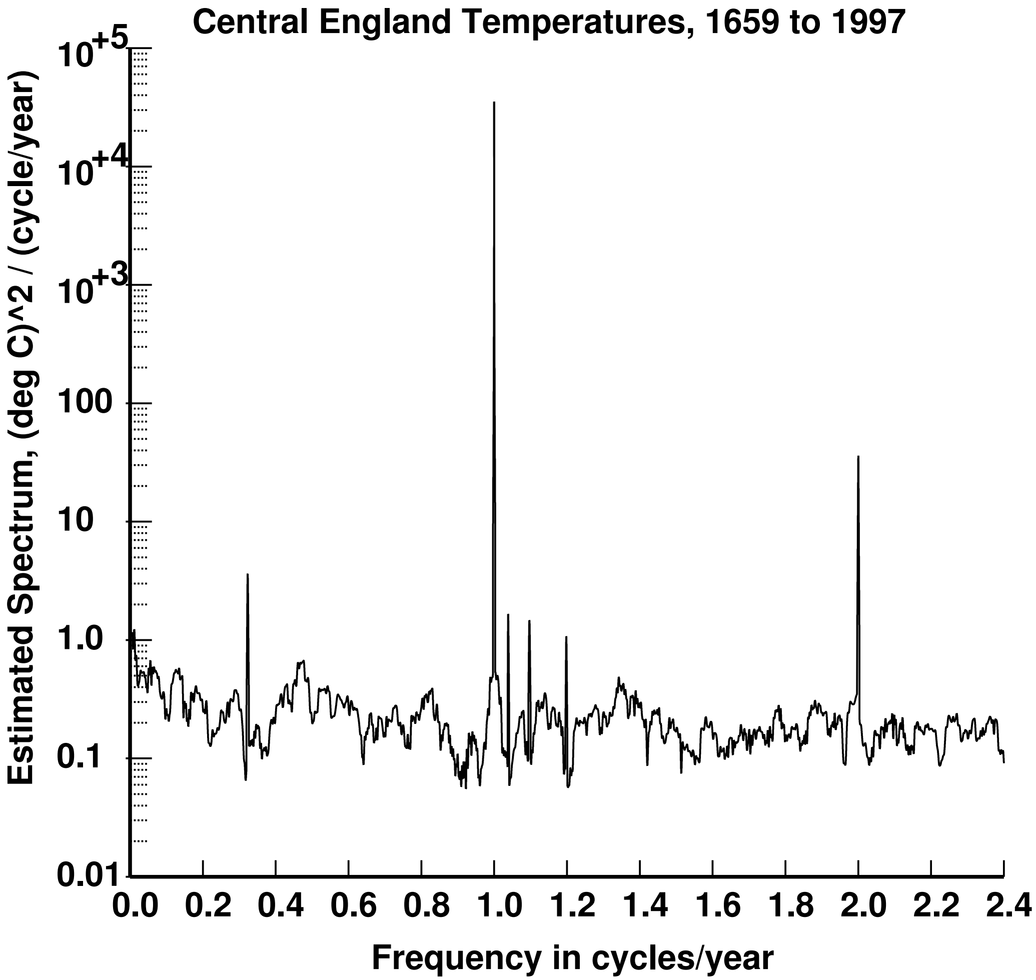
– not zero width

Baseline:

– stable

– more lines?

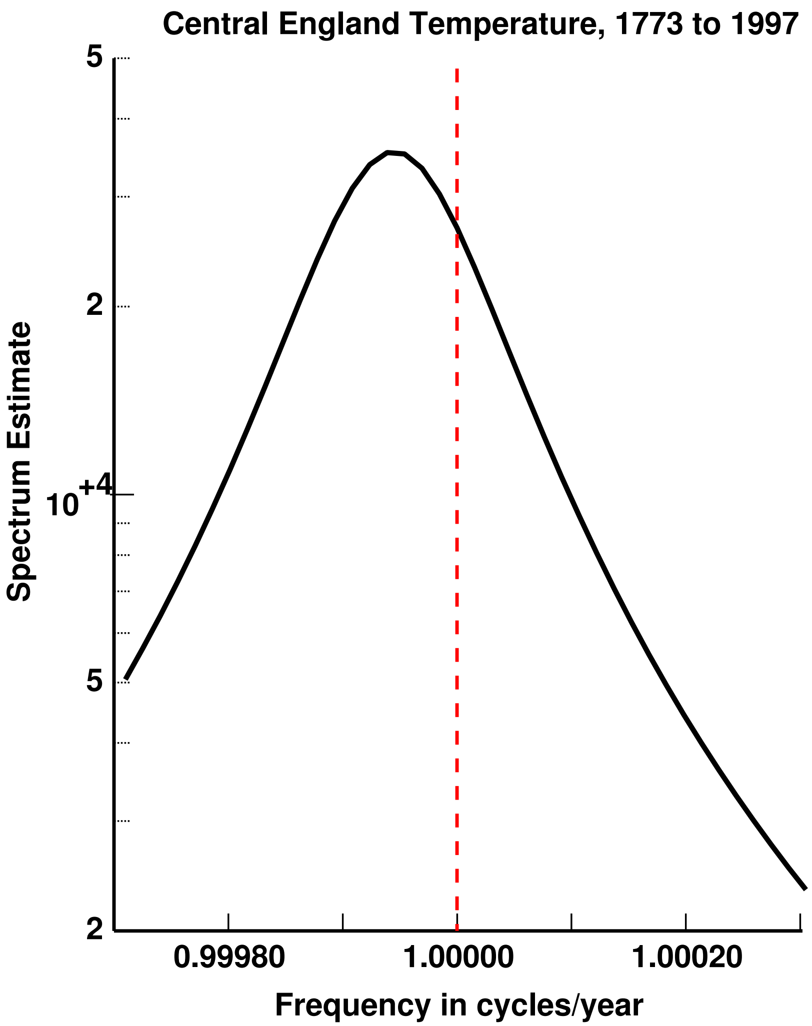
Multitaper Spectrum, Central England Temperature
Why is the annual cycle so odd?



Data:
Central England
Temperature
1659–1997
(Longest Monthly)

Annual cycle:
– Well resolved
– Slightly broadened
Two cycle/year line
Other lines
Lots of “small stuff”

Multitaper Spectrum, Central England Temperature
The annual cycle is odd because it's **NOT** one cycle/year!



(Annual cycle detail)

Data: Central England Temperature, 1659–1997

Here: 1753 on

Harmonic–F test

Frequency:

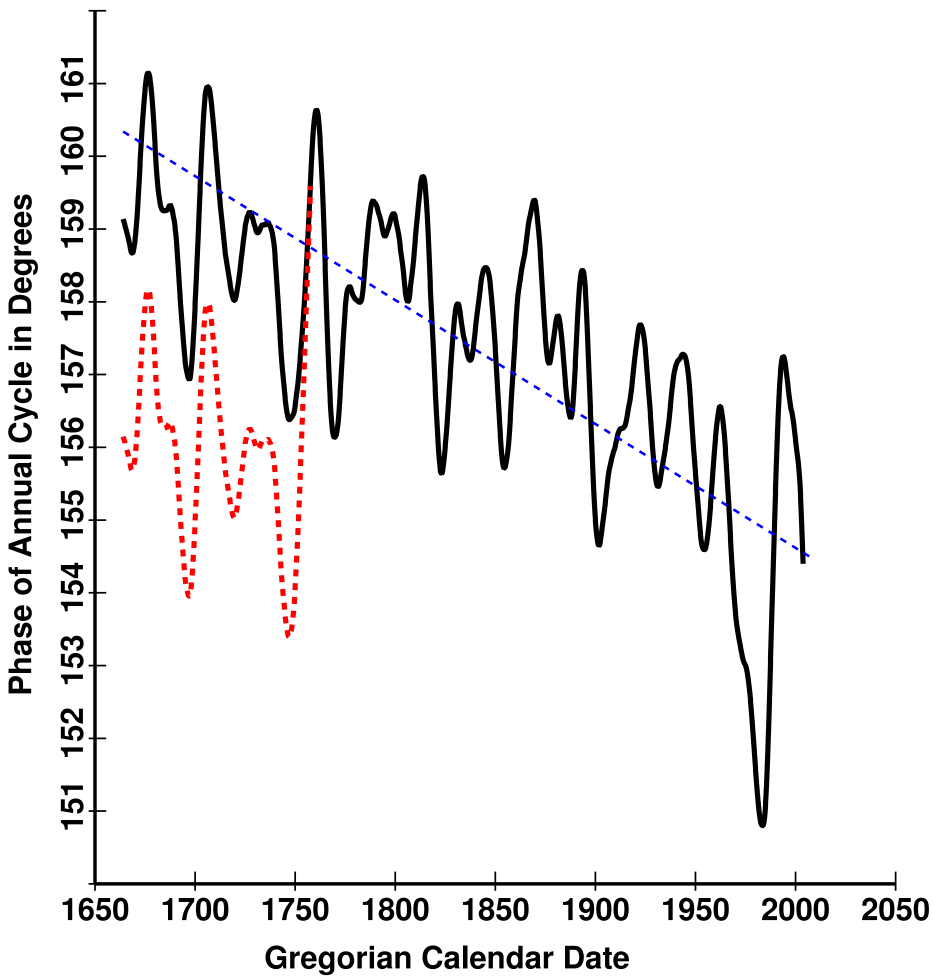
Variance $\propto \frac{1}{T^3}$

($T = N\delta t$)

Peak $> 5\sigma$ from 1.0 c/y

(Next Page)

Annual Cycle Timing, Central England Temperature
(Filtering Program "borrowed" from Cell Phone Development)



Central England Temperature 1659 to 2010 Phase(t)

Blue line represents General precession. (50''.291 per year)

NOT a fit!

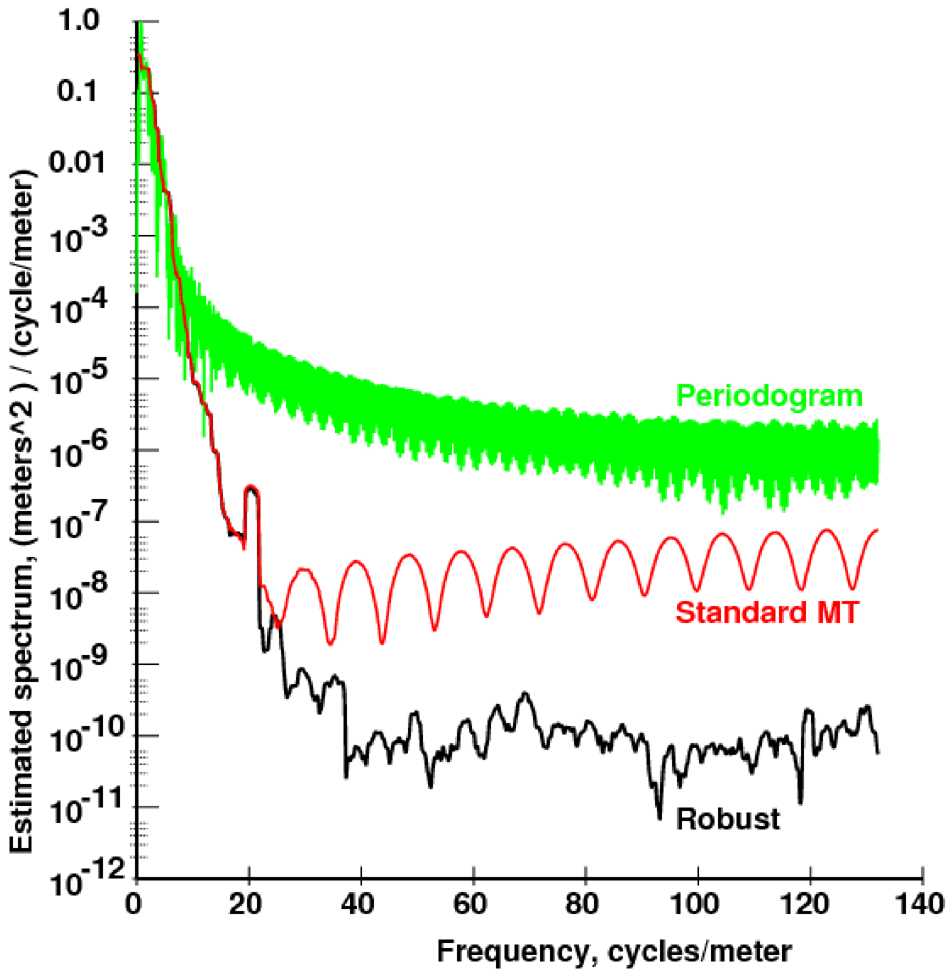
Wiggles are solar.

Dotted red is "raw" data (Julian to Gregorian switch, Sept. 1752).

Science (1995)

Cited: President Clinton's *Supplement to the Budget*

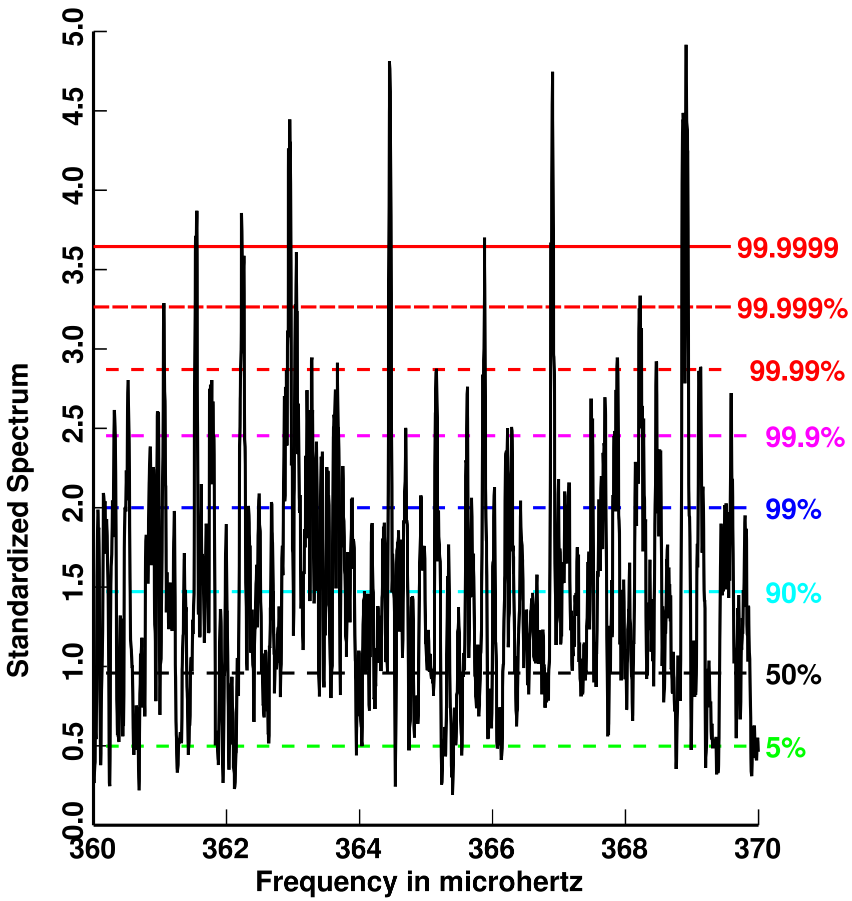
Quality Control Data



Millimeter Waveguide
Diameter (microns)
vs Distance
Raw Data:
"Outlier"
(Repeated Sample)

BSTJ 1977

Processes with MANY lines are surprisingly common.



GOES-10 P4 Protons

July 1998 to Dec. 2004

99.99% level Upcrossings:

Expect: 0.228

Observe: 18

False Peak Rate:

The periodogram is 10 times worse than this multitaper.

DJT, Charlotte Haley

Proc. Royal Soc. A

March, 2014

Some Current topics

Multitaper Autocorrelations — next page!

Processes with *Many* lines — *get more data*!

Expected number of false peaks in a spectrum — ✓

DJT & C. Haley *Proc RSL A* 2014

Others (not discussed):

Canonical coherences — ✓

Canonical bicoherence (third moments)

Seismic “noise”, or ‘Earth’s hum” — ✓

DJT & F.L.Vernon *GJI* 2015

Missing data problems — *neverending*!

see Chave 2019; Springford, Eadie, DJT 2020

Loève spectrum (tomorrow!)

Book !

Laplace distributed data

Solar wind: non–stationarity

Whatever interesting data comes my way!

Other Side of Fourier Transform — Autocorrelations

$$R_x(\tau) = \mathbf{E}\{x(t+\tau)x^*(t)\}$$

Implies:

1) Quadratic scaling

2) Non-negativity

3) Modulation covariance:

$$R(\tau\,;\{e^{i\omega t}x(t)\}) = e^{i\omega\tau}R(\tau\,;\,\{x\})$$

Require use of multitaper estimates (McWhorter & Scharf 1998)

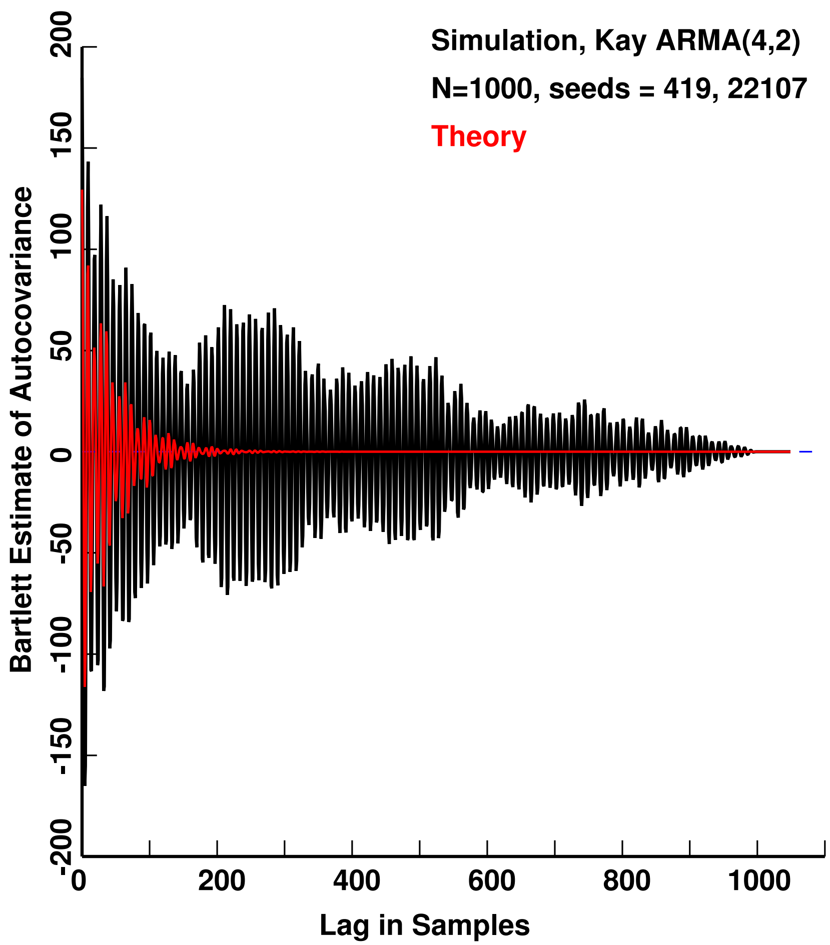
$$\widehat{R}_{mt}(\tau) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \widehat{S}_{mt}(f)\, e^{i2\pi f\tau}\, df$$

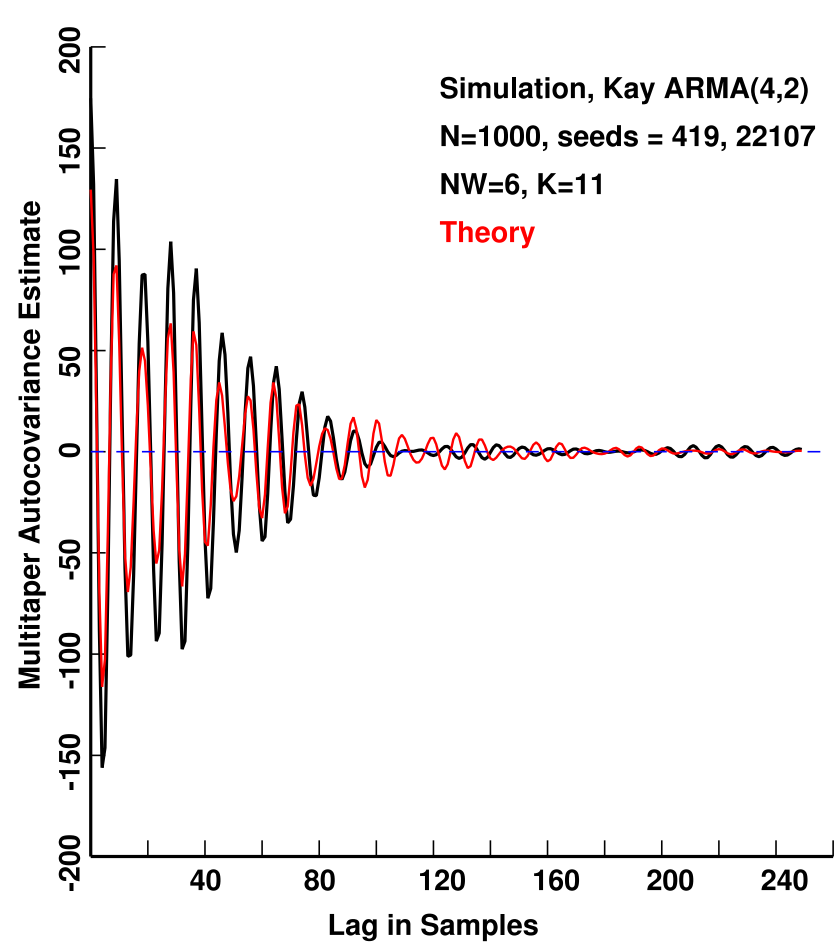
$=$ Average of autocorrelations of $\{\, x(n)\; v_n^{(k)}\,\}$

"an observed autocorrelation always exhibits less damping than the theoretical" – Bartlett

Multitaper autocovariance estimates have *much* lower variance and bias than the "standard" estimate.

Simulation comparing Bartlett & Multitaper Autocovariances





Bartlett $\frac{1}{N}\sum_{n=1}^{N-\tau} x(n)x(n+\tau)$ **Multitaper Theory**

Folk Theorem: Bad Spectra $\Leftrightarrow$ Bad Autocorrelations

Summary - What I'm trying to do

Convert Time–Series from a "Black Art" to a Science

You can teach the multitaper method *without* having to tell students "None of this stuff works $\cdots$

Improve scientific inferences and answer interesting questions

Huge impact so far: **Lots of new science.**

Why were the age scales from obliquity and precession different?

This method saved the first Transatlantic fiber optic cable.

Factor of 10^5 improvement in Touch-Tone detectors.

Lots of interesting math, statistical, and science problems!

Some Lessons more–or–less Learned

Pay attention to maxims. Learn from other people's mistakes,
— you won't live long enough to make them all yourself.

Don't ever use periodograms. Ditto, Bartlett autocorrelations.

Avoid unverifiable assumptions. Test the ones you make.

You are *always* working with small samples.

Keep Frequencies in Hertz. Cycles/year OK for paleoclimate.

Use degrees and avoid Radians. *Really* avoid radian frequency!

Specify frequencies precisely. $\sim$ C–R bound.

Chaoplexology $\Rightarrow$ Run away!

Avoid dogma — it always gets you in trouble!

Analyze data, don't assume dogma is correct

Pay attention to theory, but not too much.

Most *great* scientists have one idea in their working life, except for Lord Rayleigh, who had two. — Fermi.
— Corollary: Expect < 0.01 idea/paper.

Useful wisdom from the past

"It's better to be approximately right than exactly wrong" - Tukey

"If your experiment needs statistics, you ought to have done a better experiment" - Rutherford

"All models are wrong but some are useful" - Box

— Corollary: *Some aren't*

"One can't be too paranoid about spectra" - Dewan

"As simple as possible, but not simpler" - Einstein

KISS: Keep It Simple, Stupid!

"The hallmark of good science is that it uses models and 'theory' but never believes them." - Wilk

All things are made of atoms. - Feynman

"A statistician is someone who is good with numbers, but who lacks the personality to be an accountant" — RSS News

Thank you!