

**Welcome!**

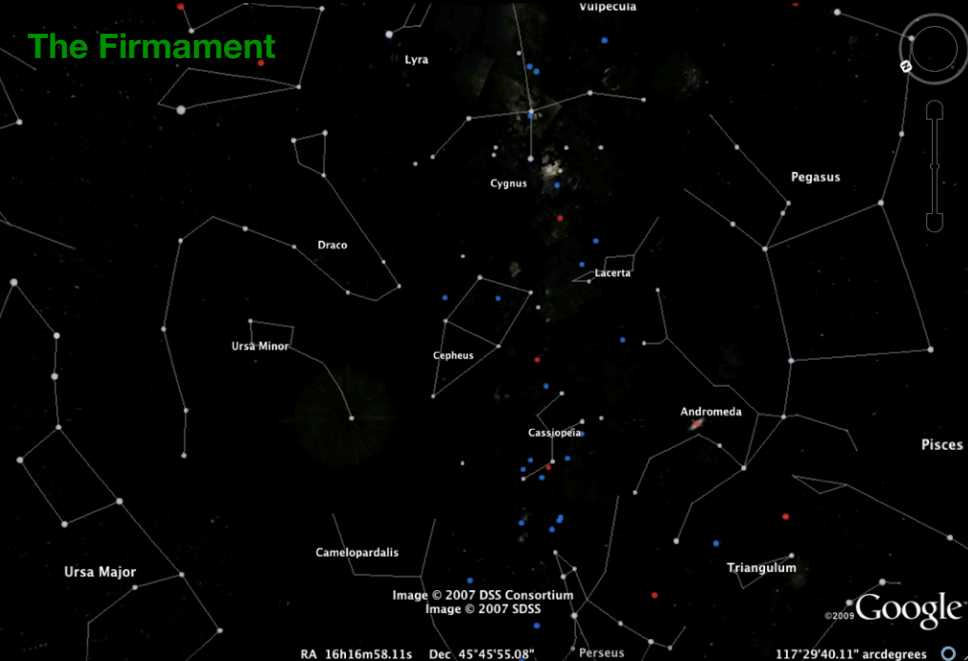
**17th HEAD Meeting Special Session**

**Exploring time series data  
in high energy astrophysics**

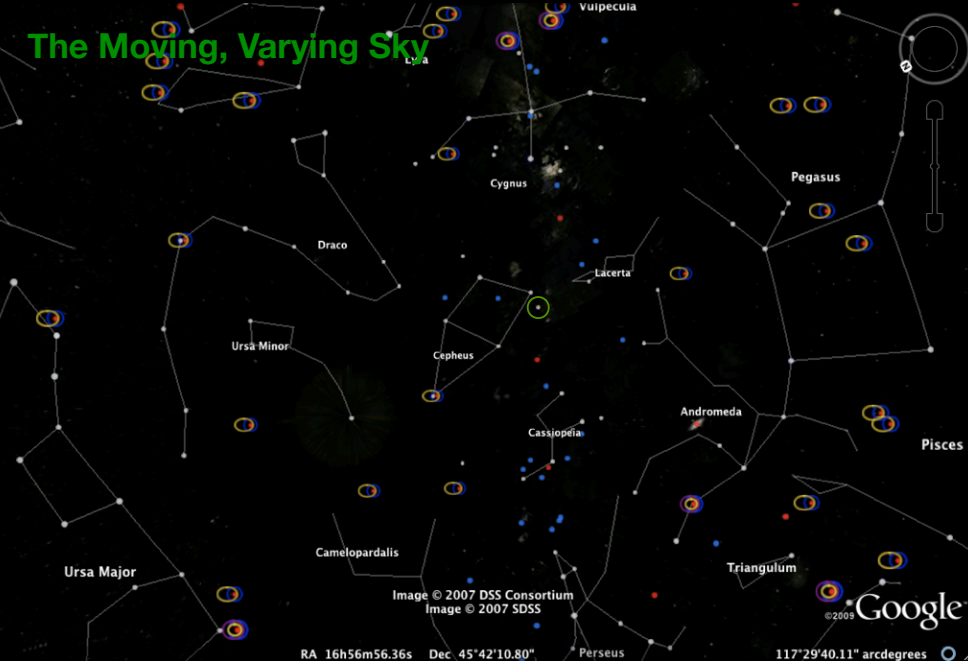
**Session material public GitHub repo:**

**`https://github.com/tloredo/HEAD2019-TSE`**

# The Firmament



# The Moving, Varying Sky



# Delta Cephei — Variability!

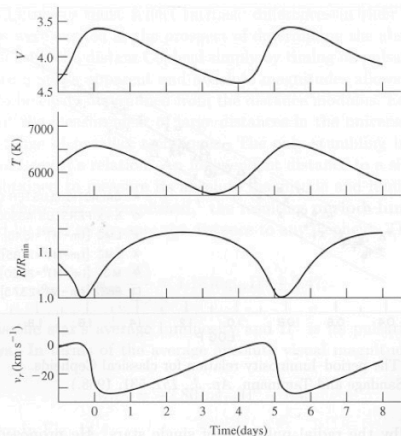
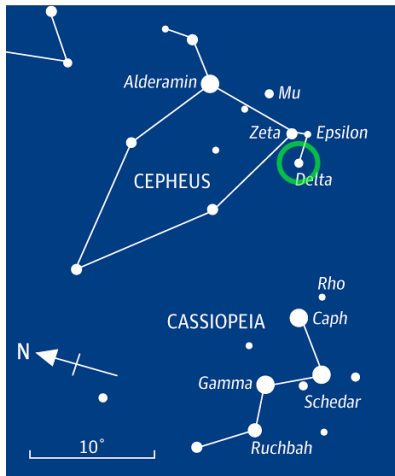


Figure 14.5 Observed pulsation properties of  $\delta$  Cephei.

Discovered in 1700s; 5.4 d period, 0.9 mag ampl  
(Mira & Algol periodic variables discovered in 1600s; “Mira” = “wonderful,”  
“astonishing”)

## Leavitt law for Cepheids

*An early time-domain astronomy triumph*



A straight line can readily be drawn among each of the two series of points corresponding to maxima and minima, thus showing that there is a simple relation between the brightness of the variables and their periods.

— *Henrietta Swan Leavitt* —

AZ QUOTES

# Varied variability in high energy astro

- Periodic (e.g., pulsars)
- Stochastic: (AGN, XRBs; QPOs), transient (XRBs, GRBs, SGRs)
- Spectro-temporal — add energy dimension

*Talks in this session cover each type*

# Context: Time series software for astronomy

Packages with generality/breadth/depth (recent/maintained):

- VARTOOLS: Command-line light curve analysis (C)
- SITAR: S-lang/ISIS Timing Analysis Routines
- CULSP: Lomb-Scargle periodograms on GPUs
- LightcurveMC: LC simulation, testing tools in C++, R
- BGLS: Bayesian generalized Lomb-Scargle periodogram
- FATS: Feature analysis for time series
- gatspy: General tools for astro time series (AstroML)
- Spectra: Power spectra for unequally-spaced data
- agatha: Period finding in correlated noise (R)
- Gaussian process packages: George, celerite
- Mission/project-specific tools: *Fermi* tools, *Kepler*/*TESS* lightkurve, Starlink...
- Julia: JuliaAstro/LombScargle, cerite, CARMA.jl...
- *carma\_pack*: Bayesian CARMA modeling via MCMC (C++, Python)
- *Stingray*: Next-generation spectral-timing software (Python)
- TSE Project: Python and MATLAB packages, e.g., inference, stanfitter, batse5bp...

*Documentation, VCS, appealing API are essential for buy-in*

# R packages (by statisticians)

*“Best of” list c/o Eric Feigelson*

## *Base-R functions*

- acf-pacf-ccf: correlation functions with significance levels
- arima-prewhiten: autoregressive modeling
- Box.test: test for autocorrelation
- density-spline-loess: kernel & local polynomial interpolations
- fft & convolve : Fast Fourier Transform, convolutions
- fitdistr: maximum likelihood fitting of statistical distributions
- plot: display time series
- runmed-smooth-supsmu: running median-like smoothers
- spec.pgram: Fourier periodogram with tapering & smoothing



## CRAN packages

- bspec: Bayesian autocorrelation & spectral analysis
- cobs: cobs quantile spline interpolation
- changepoint-Rseg-segmented-strucchange: changepoint detection & segmented regression
- dlm: Bayesian dynamic modeling
- dtw-dtwclust: dynamic time warping & clustering
- dyn-dyn.lm: regression for irregular time series
- *forecast: auto.arima-arfima ARIMA modeling with model selection & 1/f-noise*
- imputeTS: na.Kalman ARIMA interpolation
- its, xts & zoo: infrastructure for irregular time series
- locfit: locfit local interpolation with bootstrap, weighting & censoring

## CRAN packages

- *lomb: lsp Lomb-Scargle periodogram*
- meboot: bootstrap for nonstationary time series
- MSBVar: dynamic multivariate autoregressive modeling
- msl.trend: linear, spline, SSA interpolation of gaps for irregular time series
- mvtsplot: visualization of multivariate time series
- nortest: ad.test test for normality
- robfilter: robust treatments of outliers
- RobPer: RobPer robust periodograms: PDM, LSP, etc
- sde: stochastic differential equations
- tseries-TSA: extensive time series analysis & testing
- TSDist-TSClust: distance and clustering ensembles of times series
- wavelets-wavethresh-wmtsa-adlift: wavelet transform, denoising & analysis
- WeightPortTest: tests for autocorrelation with heteroscedastic weights

# This session



- Periodic, power spectra: TSE, inference (Tom & Jeff)
- Stochastic: carma\_pack (Malgosia)
- Spectro-temporal (spectral timing): Stingray (Abbie)

*Underlying theme:*

*Interplay of time-domain and frequency-domain perspectives*

# Time series exploration in Python and MATLAB: Bayesian blocks, periodograms, and all that

Tom Lored<sup>\*</sup> & Jeff Scargle<sup>†</sup>

<sup>\*</sup>Cornell Center for Astrophysics and Planetary Science

<sup>†</sup>NASA Ames Research Center

*Funded by NASA ADAP;  
includes some NSF AAG-funded contributions*

17th HEAD Meeting — 18 March 2019

# TSE: Time Series Explorer project

- MATLAB (Jeff Scargle, NASA Ames)
  - ▶ Exploratory data analysis (graphical)
  - ▶ Irregularly-spaced data: Global and local power spectra via discrete correlation function (DCF)
  - ▶ Event data: New developments with Bayesian Blocks
- Python (Tom Lored, CCAPS @ Cornell U.)
  - ▶ Parametric & semiparametric time-domain modeling
  - ▶ Spectro-temporal pulse decomposition using Lévy processes
  - ▶ Python translation of MATLAB algorithms

Documentation: *Handbook of Practical Time Series Analysis*

*Mostly pre-release “alpha” status. . .*

# Bayesian blocks developments

## Bayesian blocks

= locally adaptive histogramming of point/event data

= change point modeling

- Algorithm: Faster  $O(N)$  algorithm, usable up to  $N \sim 10^8$
- Diverse regularizers to control binning criteria
- Generalized block shapes (flat, linear...)
- 3-D time-energy-space decomposition

## Online demos (R Shiny translation of MATLAB):

- BB BATSE GRB pulse decomposition
- BB with piecewise-linear blocks

# Probabilistic perspectives on periodograms



Wikipedia

“More lives have been lost looking at the raw periodogram than by any other action involving time series!”

— *John Tukey*

“More lives have been lost looking at the *raw* periodogram  
than by any other action involving time series!”  
— *John Tukey*

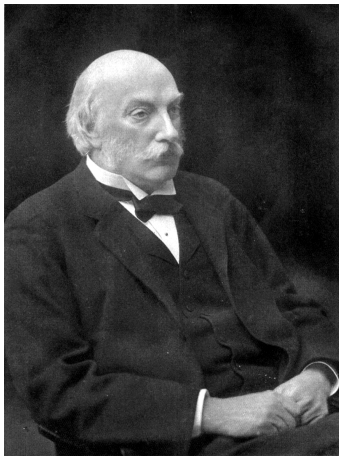
It's how you *manipulate the periodogram* that makes it useful—for  
*multiple purposes*

Literature (vast!) relevant to understanding periodograms:

- Spectral analysis of time series
- Applied harmonic analysis
- Harmonic regression
- Trigonometric regression



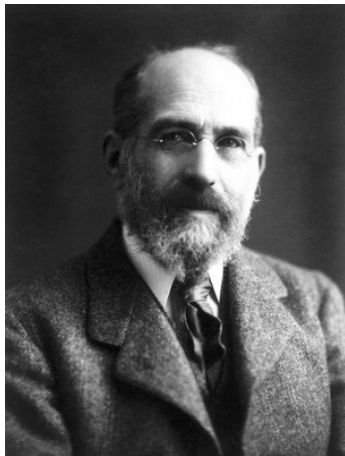
**Lord Rayleigh**  
**(John William Strutt)**  
1842–1919



Wikipedia

Cavendish Lab, University of Cambridge  
(succeeded Maxwell);  
Royal Inst. of Great Britain

**Arthur Schuster**  
1851–1934



Wikipedia

Cavendish Lab (w. Maxwell, Rayleigh);  
Owens College, Victoria University;  
succeeded by Rutherford

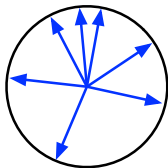
# Schuster (& Rayleigh) inventing periodograms

## *Motivation*

- Find “*hidden periodicities*” (vs. “obvious periodicities” like tides, sunspot maxima)
- *Quantify uncertainty*: “[A]pply the theory of probability in such a way that we may be able to assign a definite number for the probability that the effects found by means of the usual methods are real, and not due to accident.” *[sic! A p-value is NOT a FAP!]*

## *Name & interpretation*

- “*Fourier’s analysis* here serves the same purpose as the prismatic analysis of a luminous disturbance. . . .”
- “It is convenient to have a word for some representation of a variable quantity which shall correspond to the ‘spectrum’ of a luminous radiation. I propose the word *periodogram*. . . .”
- Envisioned application to *periodic* (“lines” from musical instruments), *quasiperiodic* (“bands” from sunspots) and *stochastic* (“broad band” from noise) variability



## Two types of periodograms

- “Occurrence periodogram” for event times (*Rayleigh statistic*): Add wrapped/folded unit vectors
- “Magnitude periodogram” for equally-spaced scalar measurements (*Schuster periodogram*): Add wrapped/folded vectors

## Null distributions (no periodic signal)

- Resultant vector magnitude,  $R$ : Rayleigh distribution,  
 $p(R) \propto R e^{-R^2/\sigma^2}$
- Squared magnitude (power): Exponential distribution,  
 $p(S) \propto e^{-S/\sigma^2}$

(Use  $\sigma^2 \rightarrow N$  for occurrence periodograms; beware normalizations!)

# Three uses for periodograms

- Nonparametric spectrum estimation
- Parametric period detection & estimation (least squares, Bayes)
- Semiparametric stochastic process (Gaussian process) modeling

# Periodogram as a spectrum estimator

Underlying idea:

- A periodic signal may be recognized by looking for a  $\delta$ -function in the *signal's* power spectrum
- $\Rightarrow$  Devise an *estimator* for the signal's power spectrum and look for  $\delta$ -function-like peaks in an estimate

Periodogram is the power spectrum of the *data*,  $d_i = f(t_i) + \epsilon_i$ :

$$P(\omega) = \frac{2}{N} \left[ \left( \sum_i d_i \cos \omega t_i \right)^2 + \left( \sum_i d_i \sin \omega t_i \right)^2 \right]$$

(Note: Normalization conventions differ!)

- Evenly sampled data: Schuster periodogram
- Unevenly sampled data: Lomb-Scargle periodogram (LSP) (with an optimal choice of phase)

## *Bias and variability*

Interpret data power spectrum as signal power spectrum corrupted by noise and sampling “window”

Under no-signal “null” hypotheses,  $N/2$  values at the *Fourier frequencies* (as returned by the DFT) are *statistically independent*

(This follows from orthogonality of sines/cosines on a uniform grid; LSP case is more complicated)

⇒ *focus on Fourier frequencies*

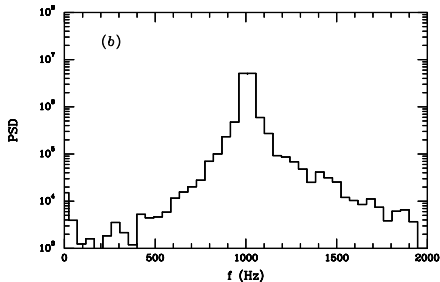
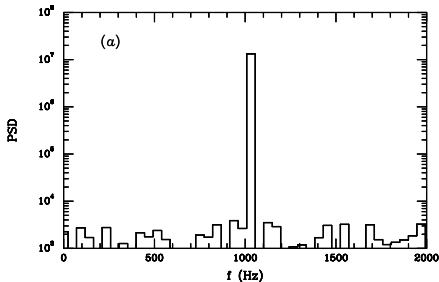
- Variability: The number of spectrum estimates (at Fourier frequencies) grows with  $N \rightarrow$  periodogram value is an *inconsistent* estimate of the signal power spectrum
- Bias: The window function (finite duration, discrete sampling) biases the expectation value of the periodogram away from the true signal power

## Spectral Leakage

1024 samples at 48 kHz sampling rate

$S/N = 5$ , white noise

$f = 1031.25$  Hz (*Fourier frequency*) and  $f = 1008$  Hz



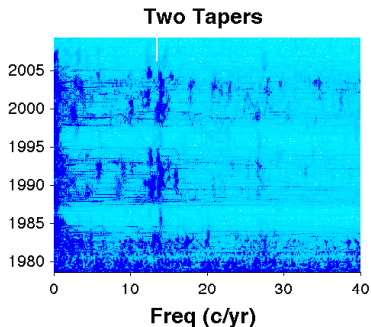
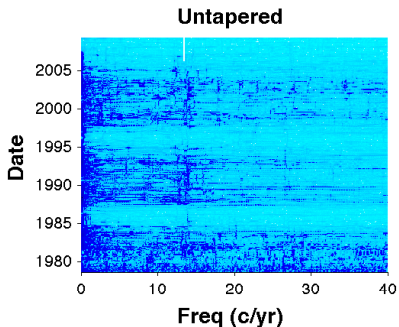
Bias reduction techniques: Periodogram smoothing, data tapering

These are esp. useful for *estimating a continuous spectrum* (vs. period hunting)



## *Time-frequency power spectrum via multitapering*

MATLAB implementation, applied to 33 y of solar Ca II K line data from Sacramento Peak National Solar Observatory (Keil Worden 1984):



## Periodogram as a period finder

Adopt a sinusoid periodic signal model (a  $\delta$ -function spectrum!):

$$\begin{aligned} f(t) &= A \cos(\omega t - \phi) && \text{parameters } \omega, A, \phi \\ &= A_1 \cos \omega t + A_2 \sin \omega t && \text{parameters } \omega, A_1, A_2 \end{aligned}$$

$$d_i = f(t_i) + e_i \quad \text{Gaussian error pdfs; rms} = \sigma$$

Estimate  $\omega$  via profile likelihood, or Bayes (*Jaynes-Bretthorst alg.*):

$$\begin{aligned} p(\omega|D) &\propto \int dA_1 \int dA_2 p(\omega, A_1, A_2) \mathcal{L}(\omega, A_1, A_2) \\ &\propto p(\omega) J(\omega) \exp \left[ \frac{S(\omega)}{\sigma^2} \right] \end{aligned}$$

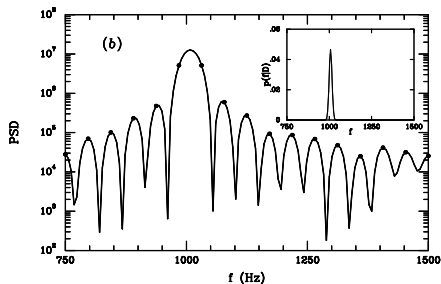
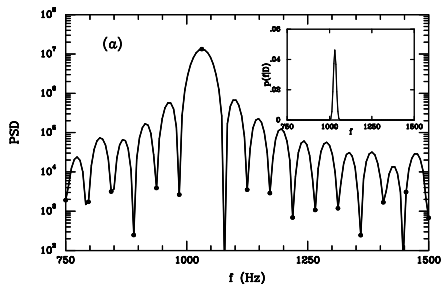
- Equally-spaced samples:  $S(\omega) \rightarrow P(\omega)$  for large  $N$  (when  $\eta$  is nearly diagonal)
- Unequally-spaced samples:  $S(\omega) \approx$  Lomb-Scargle periodogram

The posterior dist'n for  $\omega$  is related to a *continuous* version of the periodogram/power spectrum

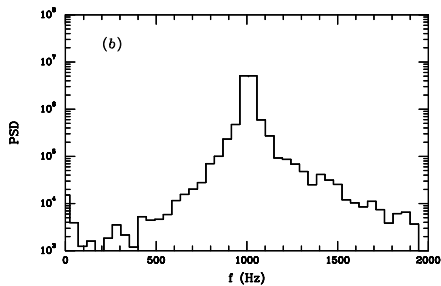
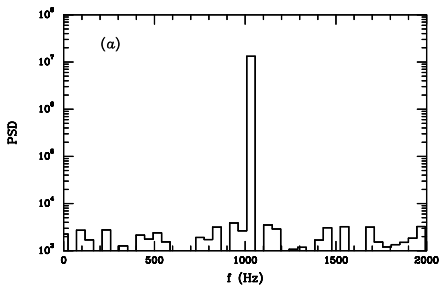
1024 samples at 48 kHz sampling rate

$S/N = 5$

$f = 1031.25$  Hz (Fourier frequency) and  $f = 1008$  Hz



Same function, evaluated only at Fourier frequencies (whence  
“leakage” when true frequency  $\neq$  Fourier frequency):



## Contrast with spectrum estimation

- Periodogram here is not a power spectrum estimator, but the logarithm of the marginal *pdf* for  $\omega$
- No special role for Fourier frequencies
- No “leakage;”  $\log$  *pdf* has similar structure for *all* signal frequencies, but sidelobes get *exponentiately attenuated* (no need for smoothing) — *sidelobes quantify period uncertainty*
- Bayes: Detect signal using signal *marginal likelihood*:

$$\mathcal{L}(\text{signal}) \approx \exp \left[ \frac{P_{\max}}{\sigma^2} \right] \times \frac{\text{peak width}}{\text{prior search range}}$$

Handles multiple testing issues via marginalization

- Conventional periodograms optimal only for *single sinusoids*

## Two Sinusoids

Adopt a model with two sinusoids at distinct frequencies:

$$\begin{aligned}f(t) &= A \cos(\omega_1 t - \phi_1) + B \cos(\omega_2 t - \phi_2) \\&= A_1 \cos \omega_1 t + A_2 \sin \omega_1 t \\&\quad + A_3 \cos \omega_2 t + A_4 \sin \omega_2 t\end{aligned}$$

Use JB algorithm to find  $p(\omega_1, \omega_2 | D)$  or  $p(\omega_1, \delta\omega | D)$

Call  $\log p$  the *doublet periodogram*

If the frequencies are well-separated, can show that

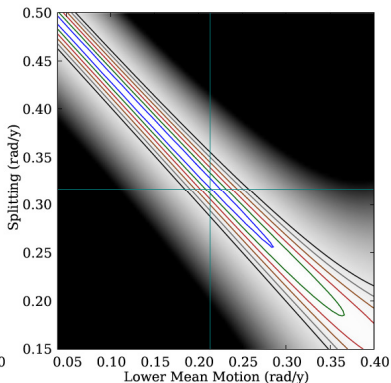
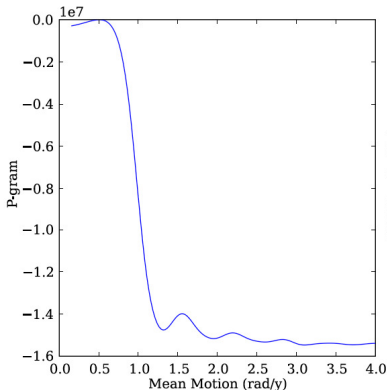
$$p(\omega_1, \omega_2 | D) \propto p(\omega_1, \omega_2) J(\omega_1, \omega_2) \exp \left[ \frac{P(\omega_1)}{\sigma^2} \right] \exp \left[ \frac{P(\omega_2)}{\sigma^2} \right]$$

a product of two *independent* dist'ns, each determined by the periodogram

If the frequencies are close, there is no such simplification; the real and imaginary parts of the DFT at each frequency (projections!) combine in a more complicated way

## Two-planet example

Singlet periodogram (i.e., single-sinusoid, left) and *doublet periodogram* (right) for simulated *SIM* astrometry data from a Jupiter (11.9 y) + Saturn (29.5 y) system at 10 pc with  $\approx 45^\circ$  inclination. The data span 10 y, with  $0.86 \mu\text{as}$  errors. The true mean motions (shown by crosshair) are 0.53 and 0.21 rad/y.



# Periodogram as a Gaussian process log likelihood

Model *stochastic variability* via a *stationary Gaussian process with parametric covariance function*  $C(u; \theta)$  for lag  $u$  and params  $\theta$ :

$$C(u; \theta) = E[f(t)f(t - u)] \quad (\text{zero-mean case})$$

This process has power spectrum

$$P(\omega; \theta) = \mathcal{F}\{C(u; \theta)\}$$

Then (under some conditions on  $C$ ), for equally spaced samples and large  $N$ , the likelihood function for  $\theta$  may be approximately calculated using the periodogram (*Whittle likelihood*,  $O(N \log N)$  i/o  $O(N^2)$ ; see Simon Vaughan's work):

$$\mathcal{L}(\theta) \propto \prod_{j=1}^{N/2} \frac{1}{P(\omega_j; \theta)} \exp \left[ -\frac{S(\omega_j)}{P(\omega_j; \theta)} \right]$$

where  $\{\omega_j\}$  are the Fourier frequencies (can be signif. biased)

*Intuition: Recall that under the white-noise null, periodogram ordinates at Fourier frequencies follow independent  $\chi_2^2$  (i.e., exponential) distributions*



# Using periodograms w/o losing your life

Tailor manipulations to goals:

- Nonparametrically estimate a spectrum (data spectrum  $\neq$  signal spectrum):
  - ▶ Raw power null distribution  $\propto e^{-S(\omega_i)/\sigma^2}$
  - ▶ Raw periodogram is inconsistent and biased (window, leakage)  
→ smooth/taper it
  - ▶ TSE provides multitaper methods, including time-frequency and for irregular sampling (currently MATLAB only)
- Detect a periodic signal (spectrum assumed to have  $\delta$ -function):
  - ▶ Single sinusoid:  $\mathcal{L}_m(\omega) \propto e^{S(\omega)/\sigma^2}$
  - ▶ Multiple sinusoids (harmonics?) → *multiplet periodograms*
  - ▶ TSE provides tools for building generalized periodograms via Jaynes-Brethorst algorithm
- Whittle likelihood for semiparametric stationary GPs