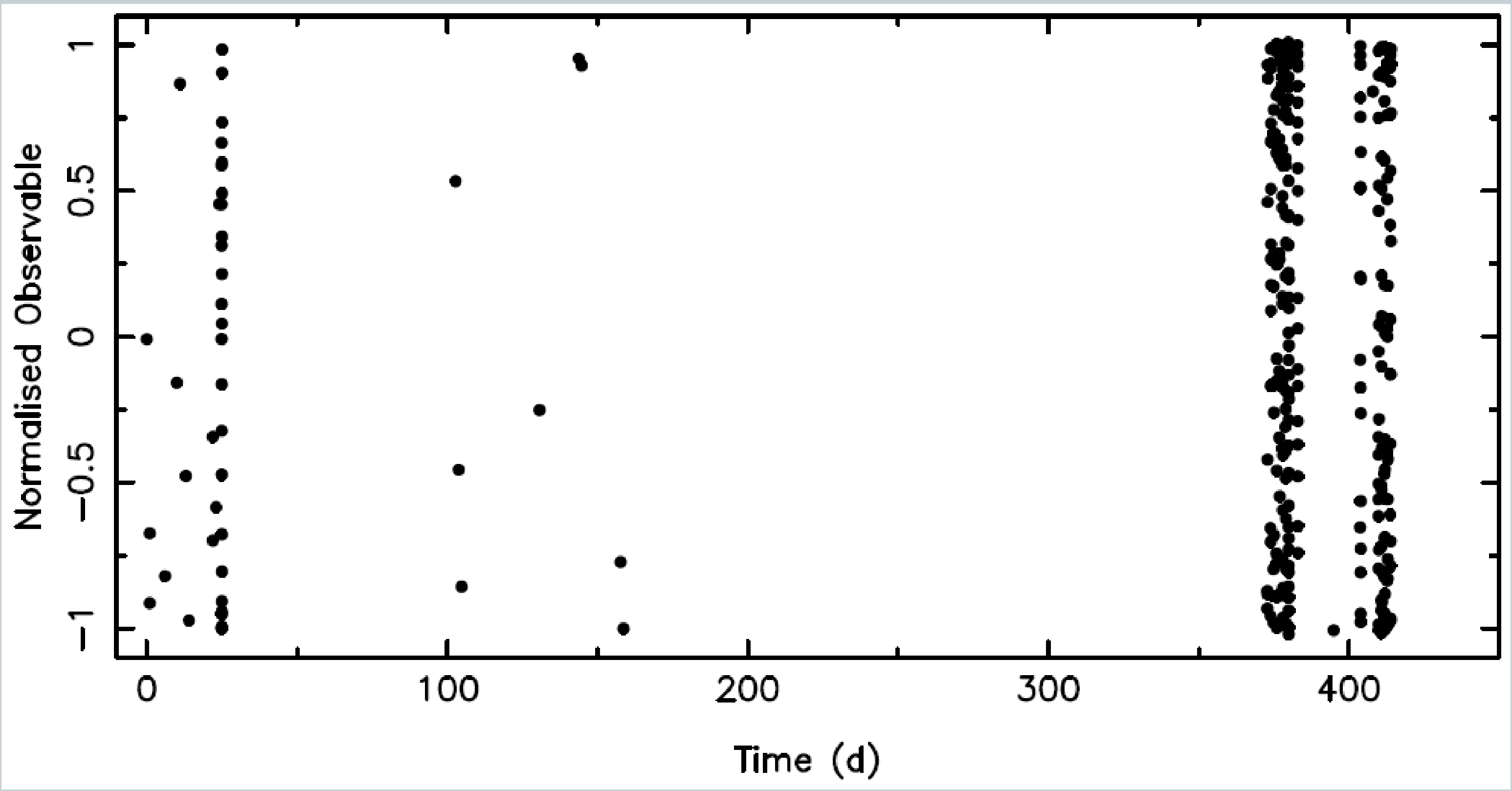
Spectral Analysis with Unevenly-Spaced Data

DISCRETE FOURIER TRANSFORM
LOMB-SCARGLE PERIODOGRAM

Spectral Analysis with Unevenly-Spaced Data

- Gapped data representing a typical time series for a ground-based single-site observational campaign:



Spectral Analysis with Unevenly-Spaced Data

FOURIER ANALYSIS WITH UNEQUALLY-SPACED DATA*

T. J. DEEMING

Dept. of Astronomy, The University of Texas at Austin, Tex., U.S.A.

(Received 22 March; in revised form 11 November, 1974)

Abstract. The general problems of Fourier and spectral analysis are discussed. A discrete Fourier transform $F_N(\nu)$ of a function $f(t)$ is presented which (i) is defined for arbitrary data spacing; (ii) is equal to the convolution of the true Fourier transform of $f(t)$ with a spectral window. It is shown that the 'pathology' of the data spacing, including aliasing and related effects, is all contained in the spectral window, and the properties of the spectral windows are examined for various kinds of data spacing. The results are applicable to power spectrum analysis of stochastic functions as well as to ordinary Fourier analysis of periodic or quasiperiodic functions.

Discrete Fourier transform

- Time series, x_k, $k=0, \dots, N\text{-}1$
- **Evenly** spaced data:
  - The discrete Fourier transform decomposes the signal into N sine waves, a_j, $j= -N/2+1, \dots, N/2$

$$a_j = \sum_{k=0}^{N-1} x_k \, e^{i2\pi jk/N} \qquad j = -\frac{N}{2}+1, \dots, \frac{N}{2}$$

- **Unevenly** spaced data:
  - The discrete Fourier transform decomposes the signal into N sine waves, a_j, $j= -N/2+1, \dots, N/2$

$$a_j = \sum_{k=0}^{N-1} x_k \, e^{i2\pi \nu_j t_k} \qquad j = 1, \dots, M$$

M and ν_j are now arbitrary.

Fourier Analysis with Unequally-Spaced Data

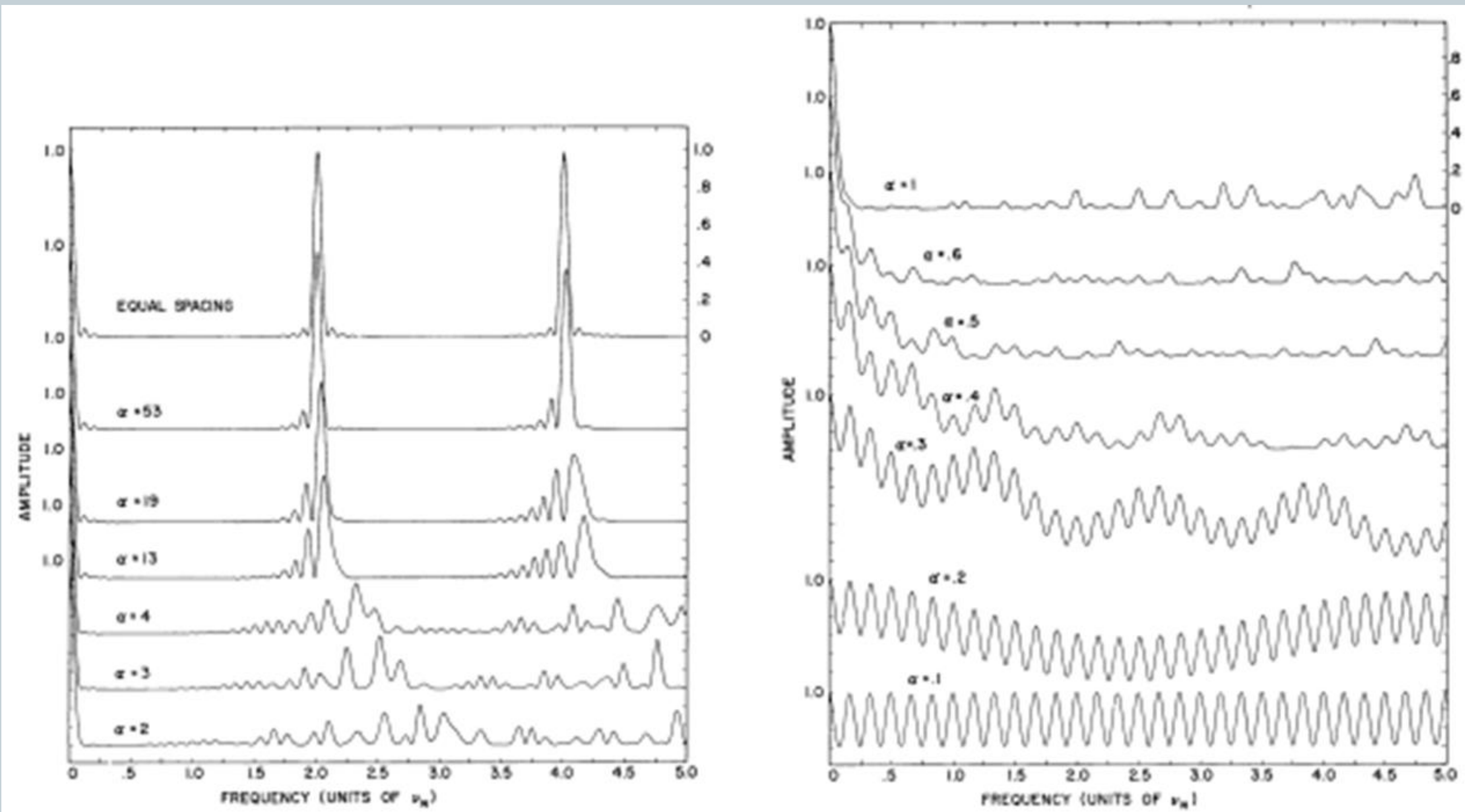
- PSD is computed as the squared Fourier amplitudes:

$$P_j = (Normalization)|a_j|^2$$

- **Deeming:**
  the "pathology" of the data spacing, including aliasing and related effects, is all contained in the spectral window.

Fourier Analysis with Unequally-Spaced Data

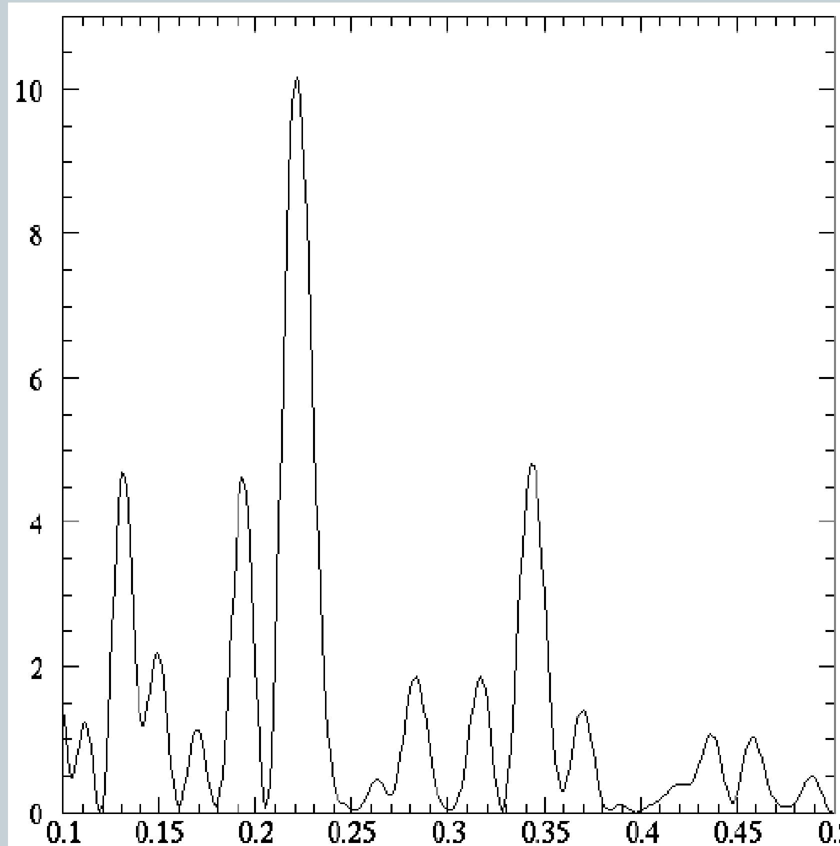
- The variation of spectral window shape:



Fourier Analysis with Unequally-Spaced Data

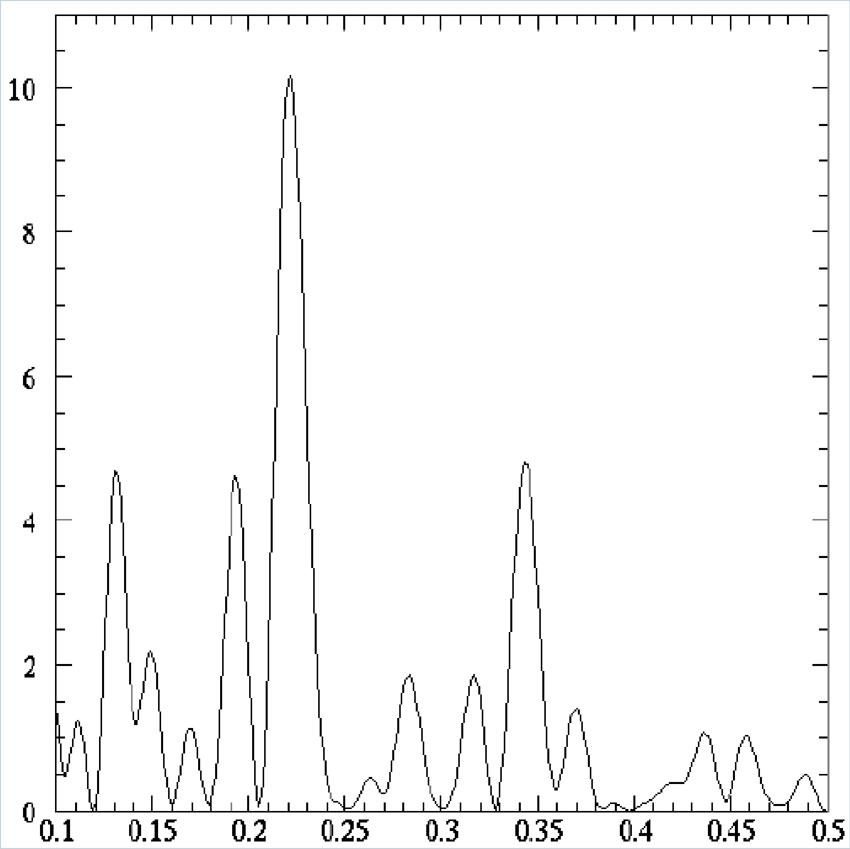
- How to determine the significance of peaks found in power spectra of unevenly-spaced data?

- How many *independent* frequencies do we use? In the plot to the right, 1000 values of v_j are plotted (M=1000), but most frequencies are ***not*** independent!



Fourier Analysis with Unequally-Spaced Data

- Most frequencies are ***not*** independent!
- Best solution: use Monte Carlo data sets to estimate the probability that the largest peak is bigger than P_{det}.
- Second-best rule of thumb: Each peak has a width of $d\omega=2\pi/T$, where T is the length of the entire data set. Here T=365 days, so $d\omega=0.017$. We then estimate that there should be approximately $(0.5-0.1)/0.017 = 23$ independent frequencies over this range.



Fourier Analysis with Unequally-Spaced Data

- **BE VERY SKEPTICAL OF CLAIMS FOR PERIODICITIES THAT COINCIDE WITH NATURAL FREQUENCIES OF DETECTORS OR OBSERVERS (eg. 1-day, 7-day, 1-year).**

Fourier Analysis with Unequally-Spaced Data

- The dependence of the PSD on the data length, T, is different for periodic, non-periodic, and stochastic functions:

$\lvert a_j \rvert \propto T^0$	non-periodic
$\lvert a_j \rvert \propto T^1$	periodic
$\lvert a_j \rvert \propto T^{1/2}$	stochastic

Lomb-Scargle Periodogram

- Lomb (1976) - Scargle (1982) Periodogram:

$$P_{\mathrm{LS}}(\nu) = \frac{1}{2} \frac{\left\{ \sum_{i=1}^{N} x(t_i) \cos[2\pi\nu(t_i - \tau)] \right\}^2}{\sum_{i=1}^{N} \cos^2[2\pi\nu(t_i - \tau)]} + \frac{\left\{ \sum_{i=1}^{N} x(t_i) \sin[2\pi\nu(t_i - \tau)] \right\}^2}{\sum_{i=1}^{N} \sin^2[2\pi\nu(t_i - \tau)]}.$$

  - Good for general uneven sampling
  - Equivalent to linear least-square fit to sin+cos
  - Statistically robust

Lomb-Scargle Periodogram

- In this expression, the reference epoch τ is chosen in such a way that:

$$\sum_{i=1}^{N} \cos[2\pi\nu(t_i - \tau)] \sin[2\pi\nu(t_i - \tau)] = 0,$$

- Or, equivalently

$$\tan(4\pi\nu\tau) = \frac{\sum_{i=1}^{N} \sin(4\pi\nu t_i)}{\sum_{i=1}^{N} \cos(4\pi\nu t_i)}.$$

- It looks complicated, but it's basically the regular periodogram adapted to handle unevenly spaced data. In the limit of equal spacing, it actually reduces to the classical result.

Lomb-Scargle Periodogram

- One of the reasons to have introduced the Lomb-Scargle periodogram is that its value does not change when all time values t_i are replaced by $t_i + T$ because of the definition of τ.

Properties of the Lomb-Scargle periodogram

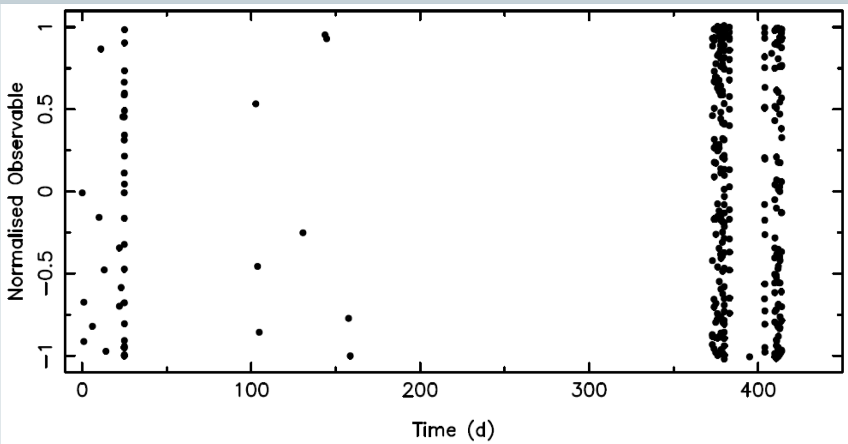
- The most important feature of the Lomb-Scargle periodogram is the significance of the power at an individual frequency:

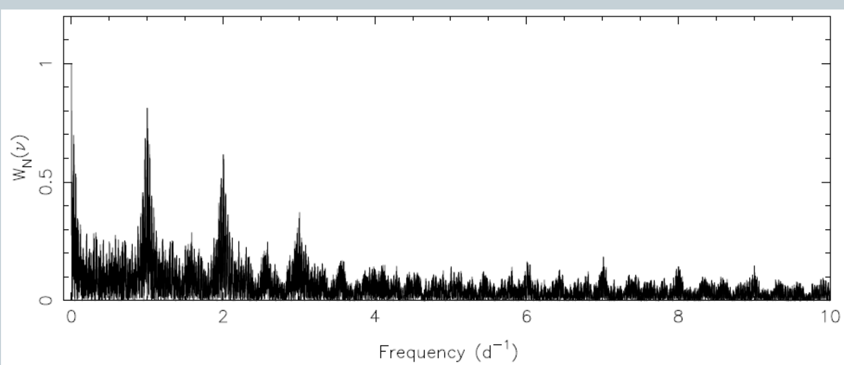
$$\mathrm{Prob}(P(\nu)>P_{det})=\exp(-\,P_{det})$$

- You still have to worry about the number of independent frequencies you test to account for trials factors

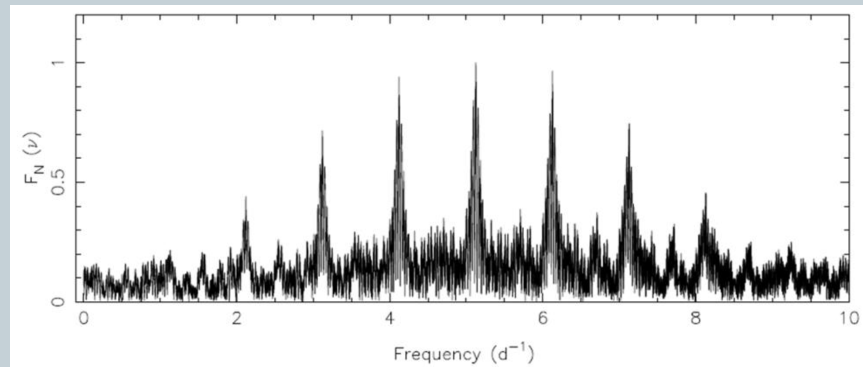
Spectral Analysis with Unevenly-Spaced Data

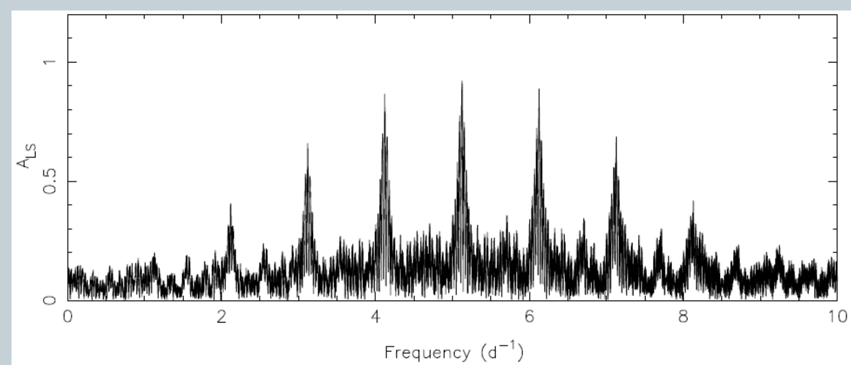
The gapped data and its Spectral window





DFT (top) and LS periodogram (bottom)





Advantages of non-uniform sampling

- Consider the following Lomb-Scargle periodograms of an f=0.8 signal.
  Top: sampling exactly once per day at noon
  Bottom: sampling once per day at a random time within the 24-hour period.
- Regular sampling gives strong alias peak at f=0.2: in fact Nyquist frequency is f<0.5, so you'd conclude there was really a signal at f=0.2
- Random sampling gets rid of alias peak! And it gives sensitivity to higher frequencies - since random times can be close to each other, Nyquist cutoff is not a hard limit anymore!