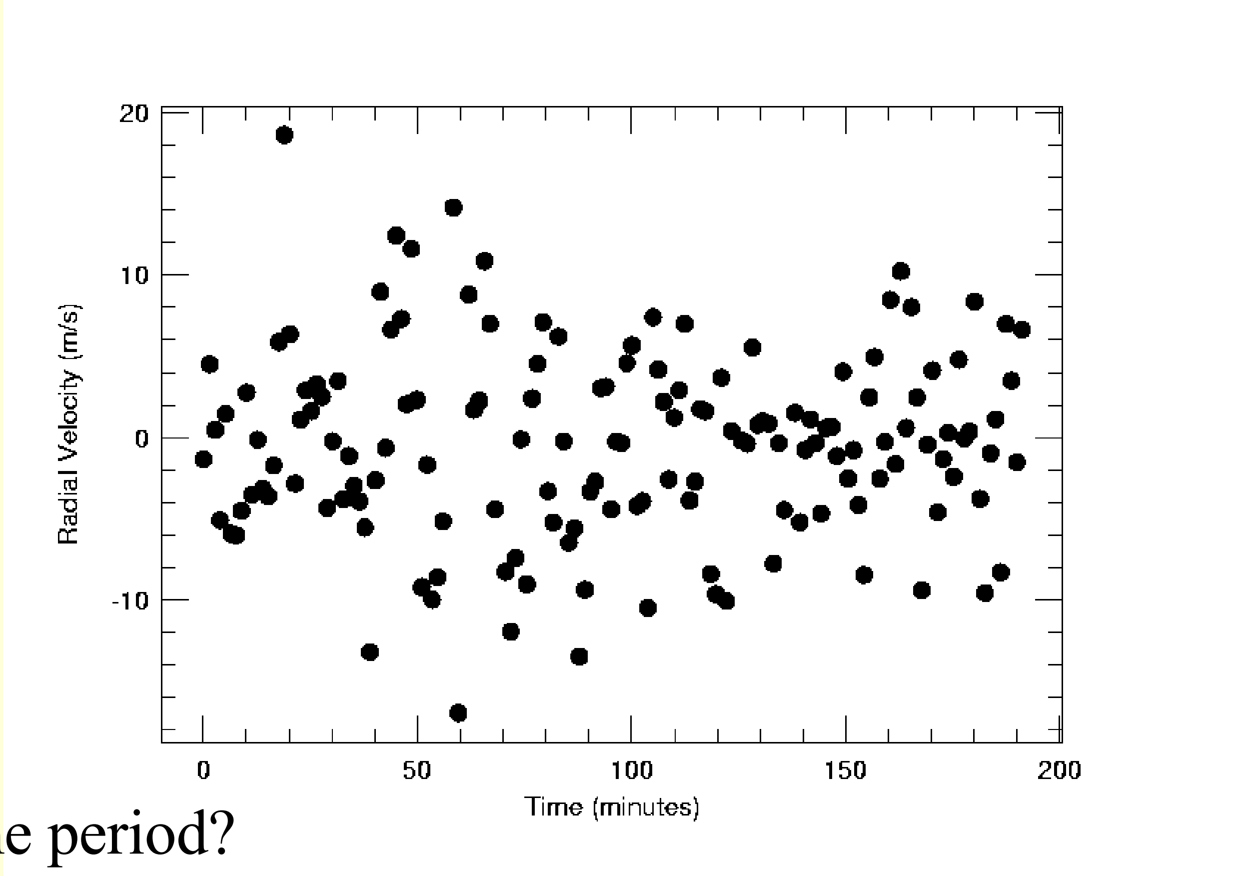
Searching for Periodic Signals in Time Series Data

1. Least Squares Sine Fitting
2. Discrete Fourier Transform
3. Lomb-Scargle Periodogram
4. Pre-whitening of Data
5. Other techniques
   - Phase Dispersion Minimization
   - String Length
   - Wavelets

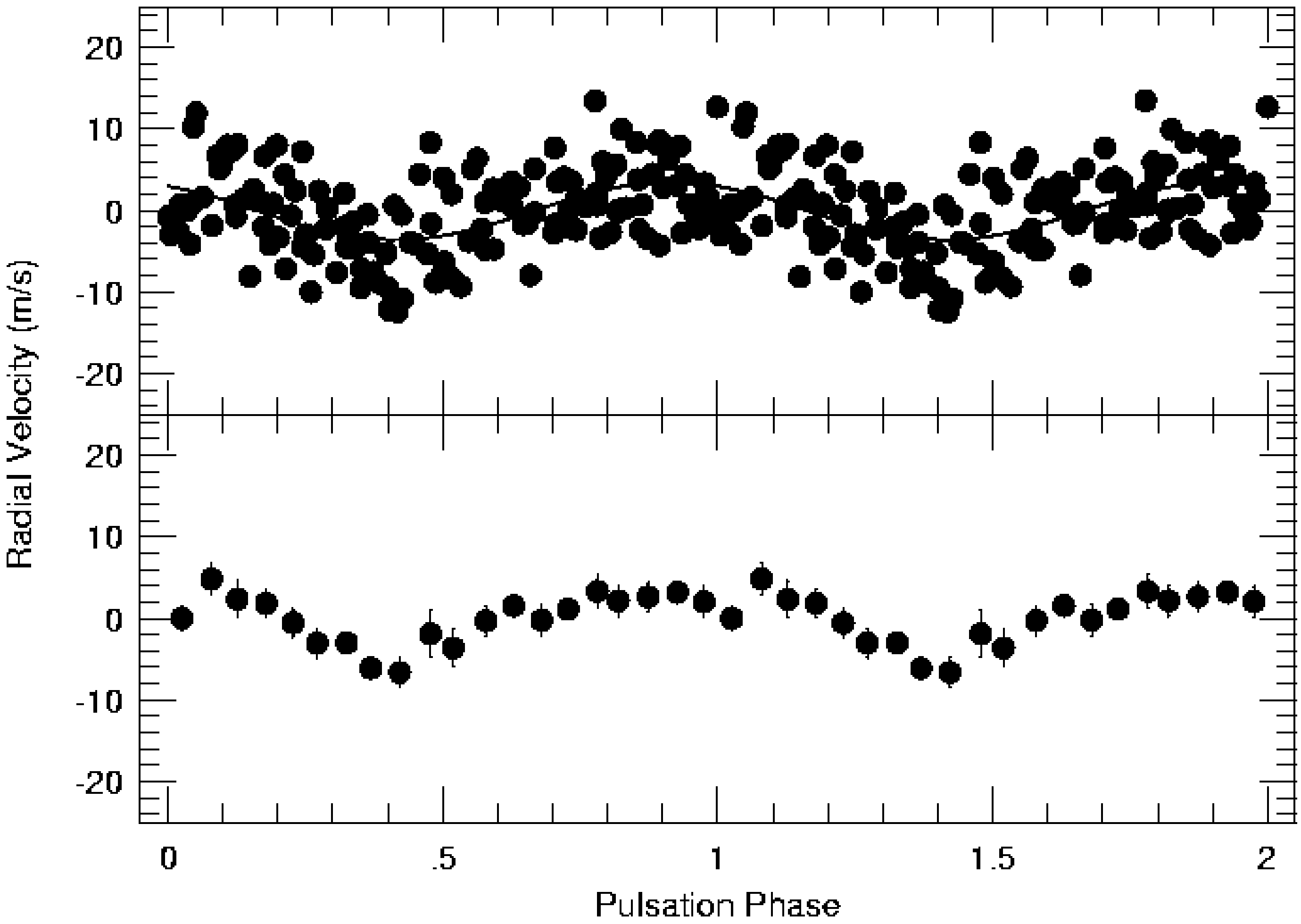
Period Analysis

How do you know if you have a periodic signal in your data?



What is the period?

Try 16.3 minutes:



Least-squares Sine fitting

Fit a sine wave of the form:

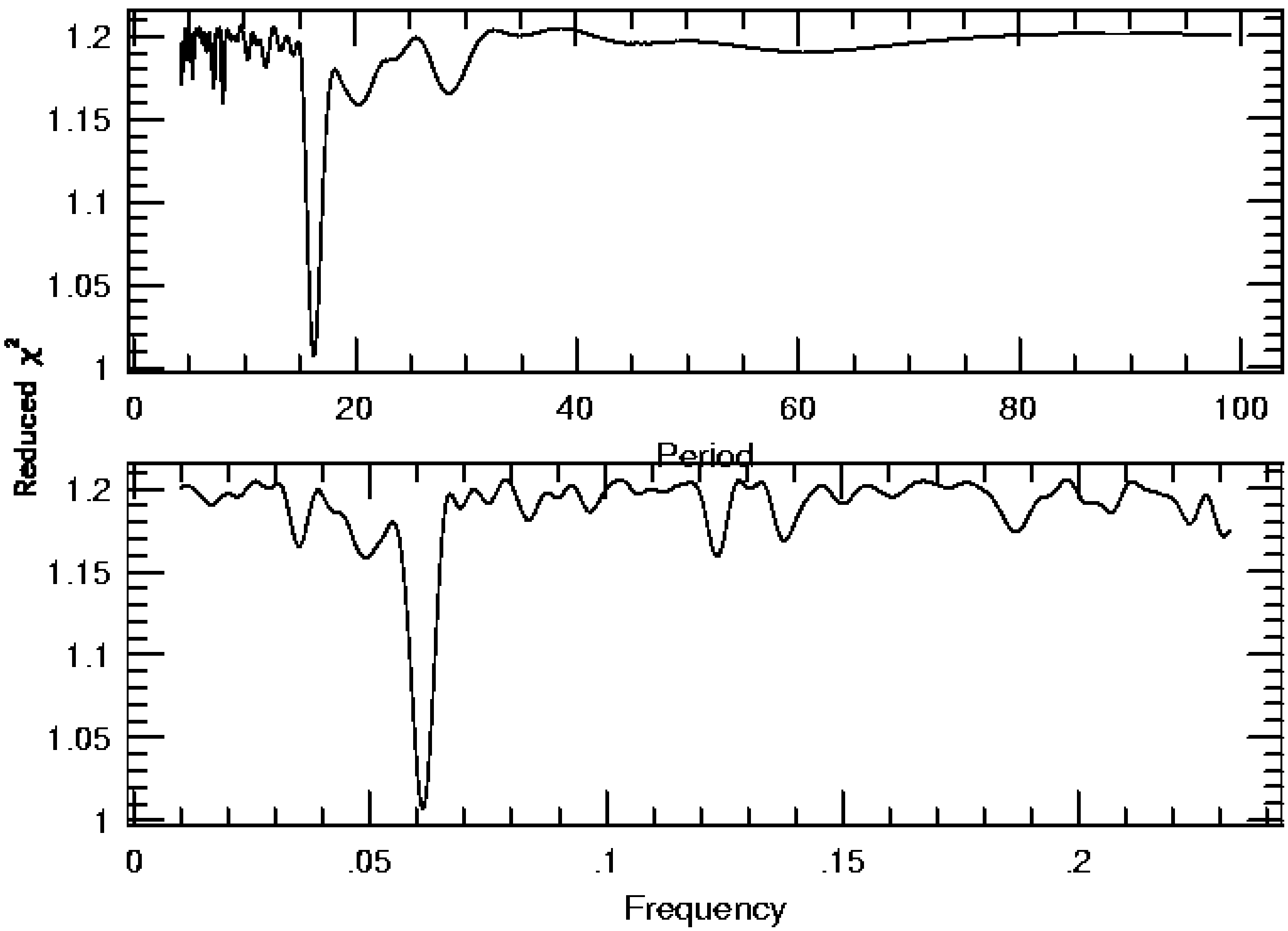
V(t) = A·sin(t +   ) + Constant

Where   = 2 /P,   = phase shift

Best fit minimizes the ^{2}:

^{2} = d_i – $g_i)^2$/N

d_i = data, g_i = fit



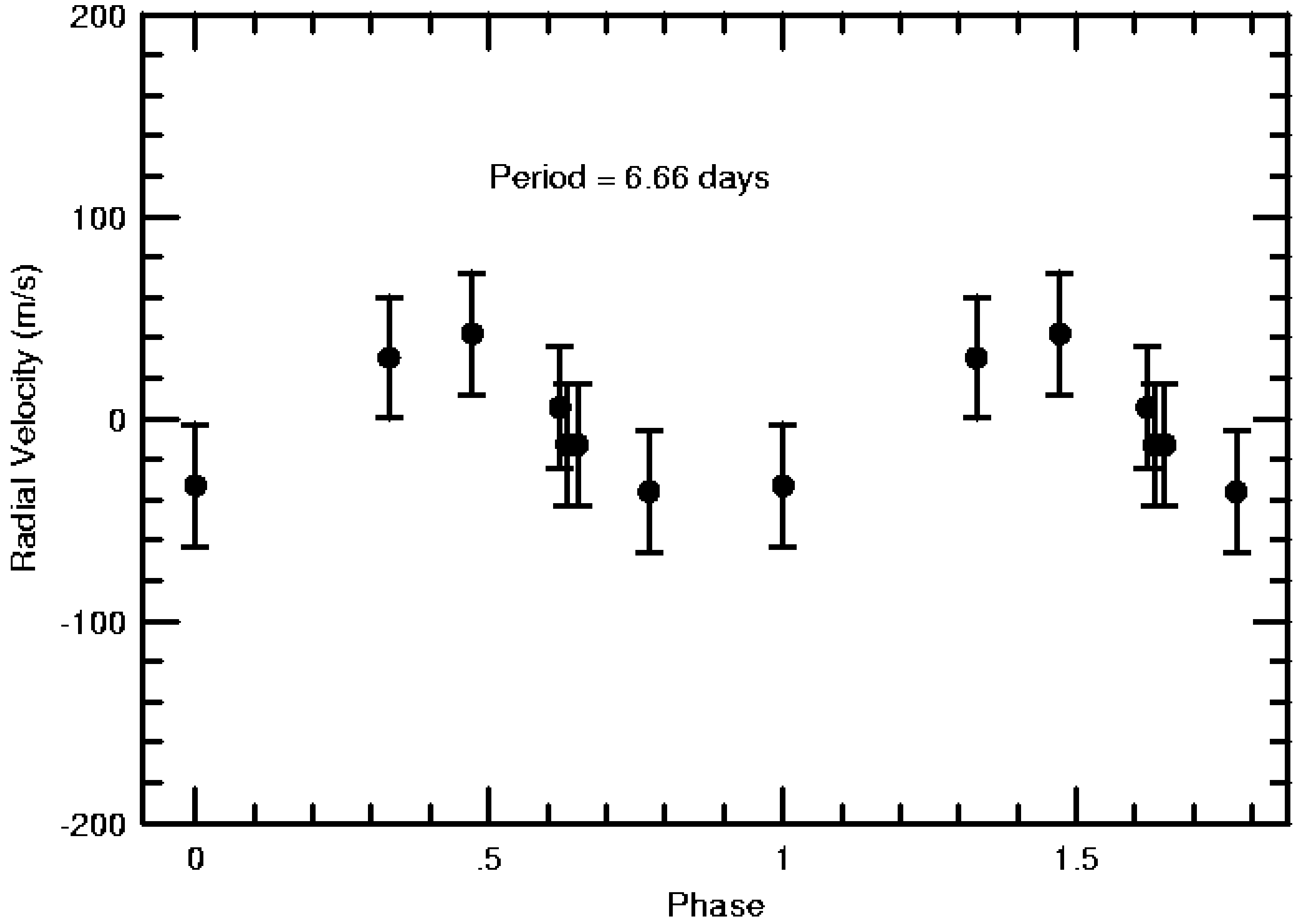
Reduced χ^2
1.2
1.15
1.1
1.05
1
0
20
40
60
80
100
Period
1.2
1.15
1.1
1.05
1
0
.05
.1
.15
.2
Frequency

Advantages of Least Sqaures sine fitting:

- Good for finding periods in relatively sparse data

Disadvantages of Least Sqaures sine fitting:

- Signal may not always be a sine wave (e.g. eccentric orbits)
- No assessement of false alarm probability (more later)
- Don‘t always trust your results



This is fake data of pure random noise with a = 30 m/s. Lesson: poorly sampled noise almost always can give you a period, but it it not signficant

More generally speaking

Fit general model for time series:

$$RSS = \sum_{i=1}^{N} \left[f(t_i) - M(t_i, c_1, c_2, \ldots, c_K) \right]^2$$

Some parameters are linear, some not:

$$M(t, c_1, \ldots, c_K) = \sum_{r=1}^{R} c_r M_r(t, c_{R+1}, \ldots, c_K)$$

Using general linear least-squares fit package reduce task to the nonlinear optimization problem.

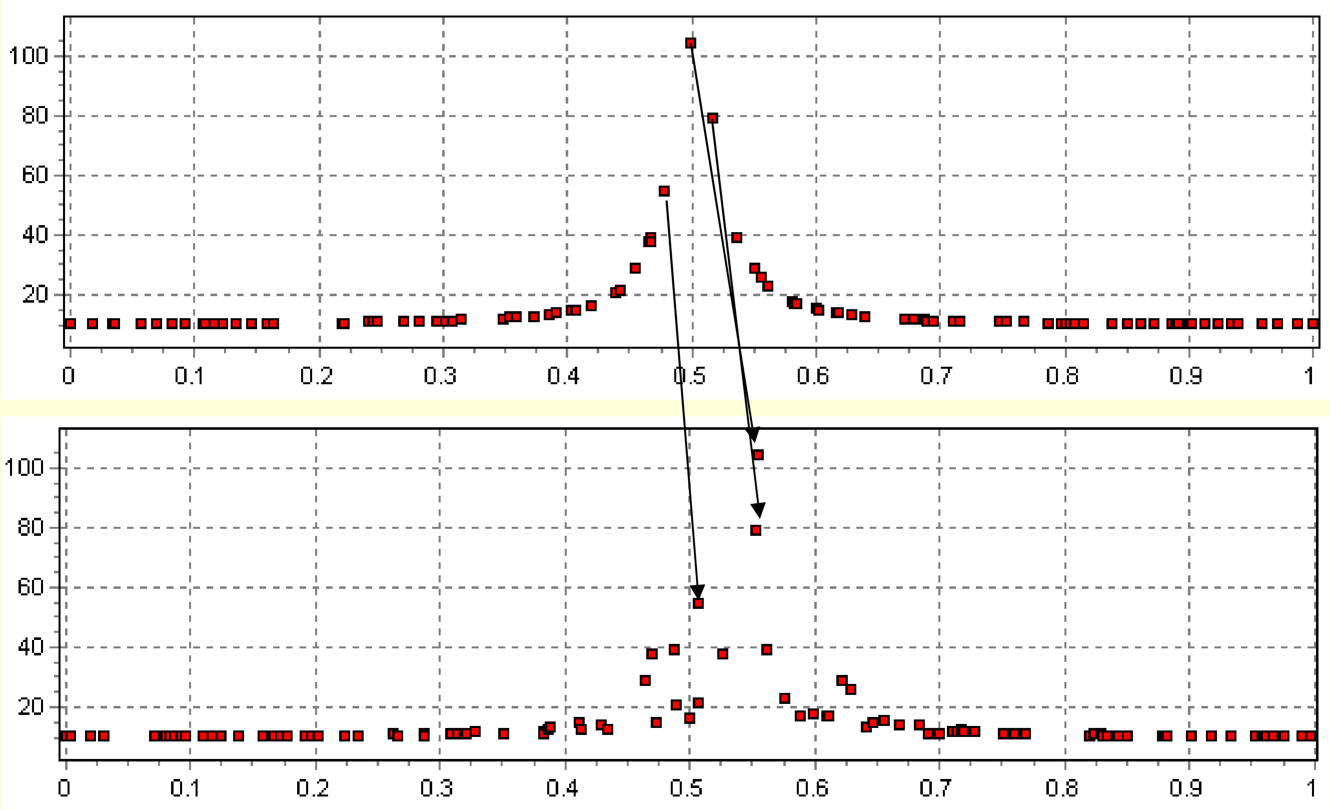
$$LSS(c_{R+1}, \ldots, c_K) = \sum_{i=1}^{N} \left(f(t_i) - M(t_i, \hat{c}_1, \hat{c}_2, \ldots, \hat{c}_R, c_{R+1}, \ldots, c_K) \right)^2$$

Typical model consists of multiple of sinusoids:

$$LSS(\nu) = \sum_{i=1}^{N} \left\{ f(t_i) - \hat{a}_0(\nu) - \sum_{r=1}^{R} \left[\hat{a}_r(\nu)\cos(2\pi r \nu t_i) + \hat{b}_r(\nu)\sin(2\pi r \nu t_i) \right] \right\}^2$$

Dispersions must be computed for a full grid of frequencies:

$$\nu_l = (l-1)\Delta\nu + \nu_{\min}, l = 1, 2, \ldots, L, \nu_L = \nu_{\max}$$



$$\Delta\nu = \frac{\Delta\varphi}{t_N - t_1}$$

Speed of points on phase-process diagram depends on distance

Total number of the trial frequencies can be quite large

$$L = \frac{(\nu_{\max} - \nu_{\min})(t_N - t_1)}{\Delta\varphi}$$

Here is the strategic plan for a simple multistage search:

- Divide the data range into k subranges, with maximal length Δt

$$L_{\text{crude}} = \frac{(\nu_{\max} - \nu_{\min})\Delta t}{\Delta\varphi}.$$

- Compute LSS for every subrange.
- Compute the average LSS.
- Search the averaged LSS to obtain crude estimates of probable frequencies.
- Compute fine spectra in the small ranges (with width) $\Delta\nu$ around the crude estimates

$$L_{\text{fine}} = \frac{M\Delta\nu(t_N - t_1)}{\Delta\varphi}$$

  where M is the number of crude estimates.
- Refine iteratively computed frequencies by standard Levenberg-Marquardt iteration procedure [10] to get estimates with maximum attainable precision.

The Discrete Fourier Transform

Any function can be fit as a sum of sine and cosines

$$FT(\ \ ) = {}_{j=1}^{N_0} X_j\ (t)\ e^{-i\ \ t}$$

Recall $e^{i\ \ t} = \cos\ \ \ t + i \sin\ \ \ t$

X(t) is the time series

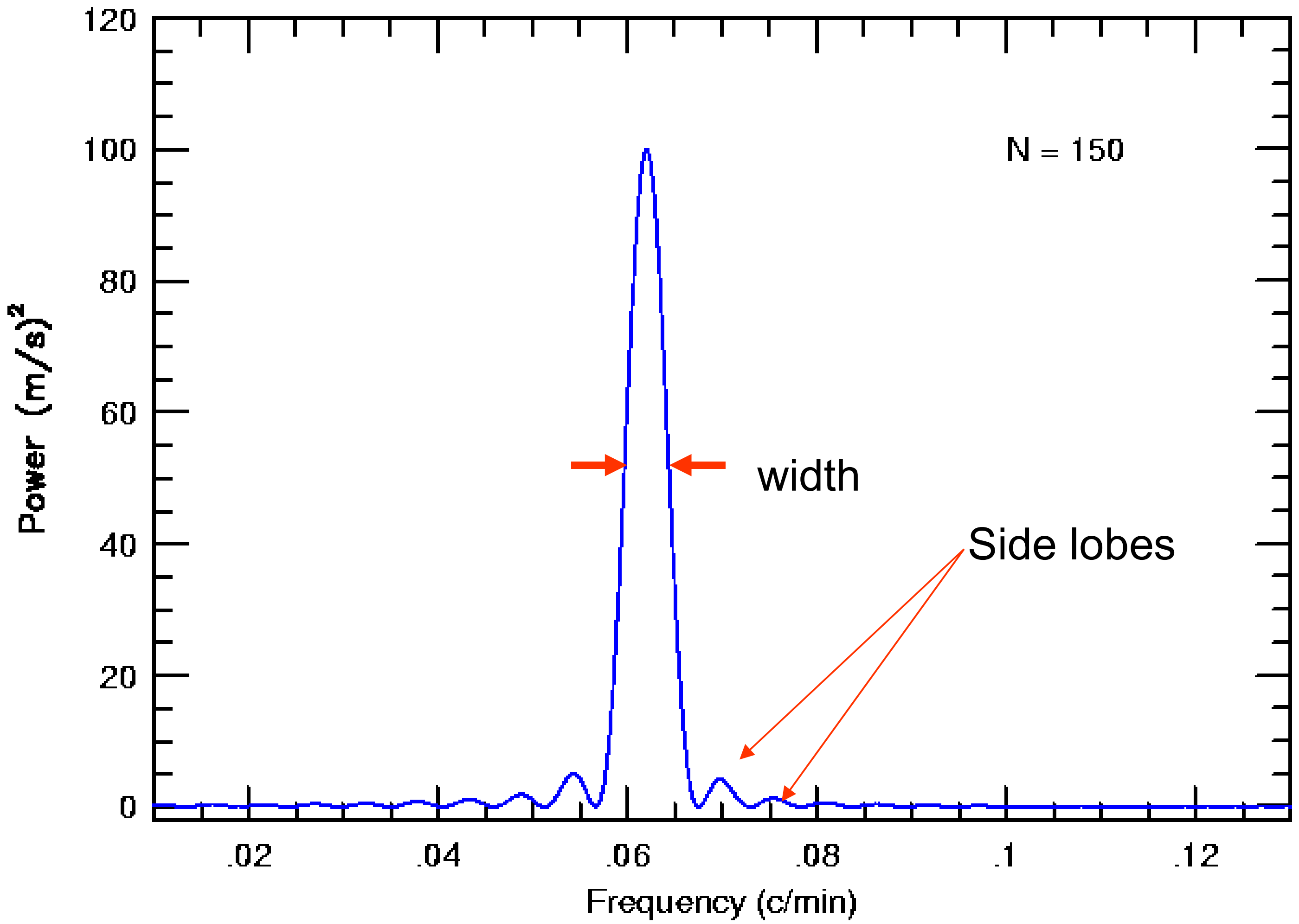
Power: $P_x(\ \ ) = \frac{1}{N_0} |\ FT_X(\ \ )|^2$

N_0 = number of points

$$P_x(\ \ ) = \frac{1}{N_0} \Big[\Big(\ \ \ X_j \cos\ \ \ t_j \Big)^2 + \Big(\ \ \ X_j \sin\ \ \ t_j \Big)^2 \Big]$$

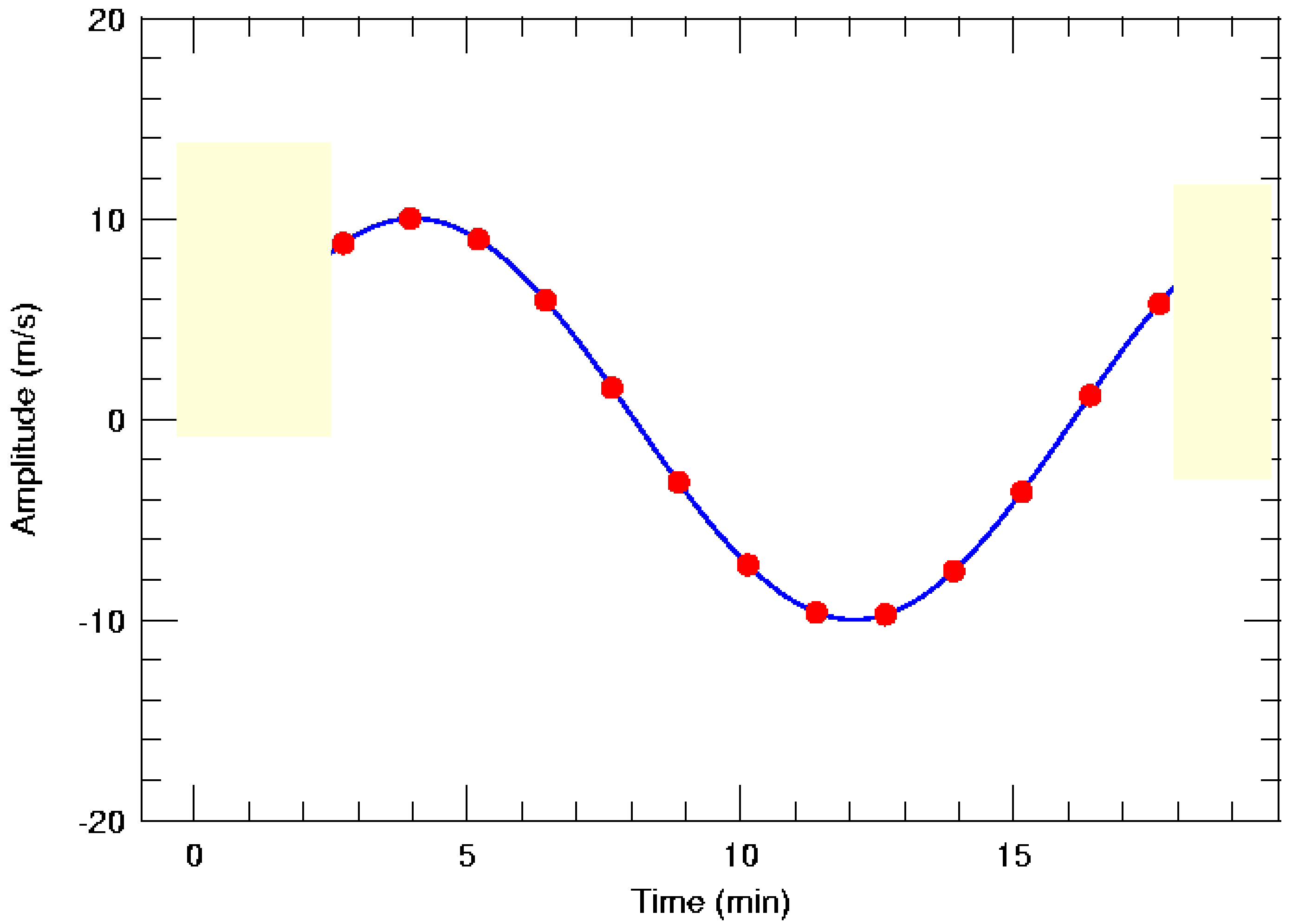
A DFT gives you as a function of frequency the amplitude (power) of each sine wave that is in the data

DFT of a pure sine wave:

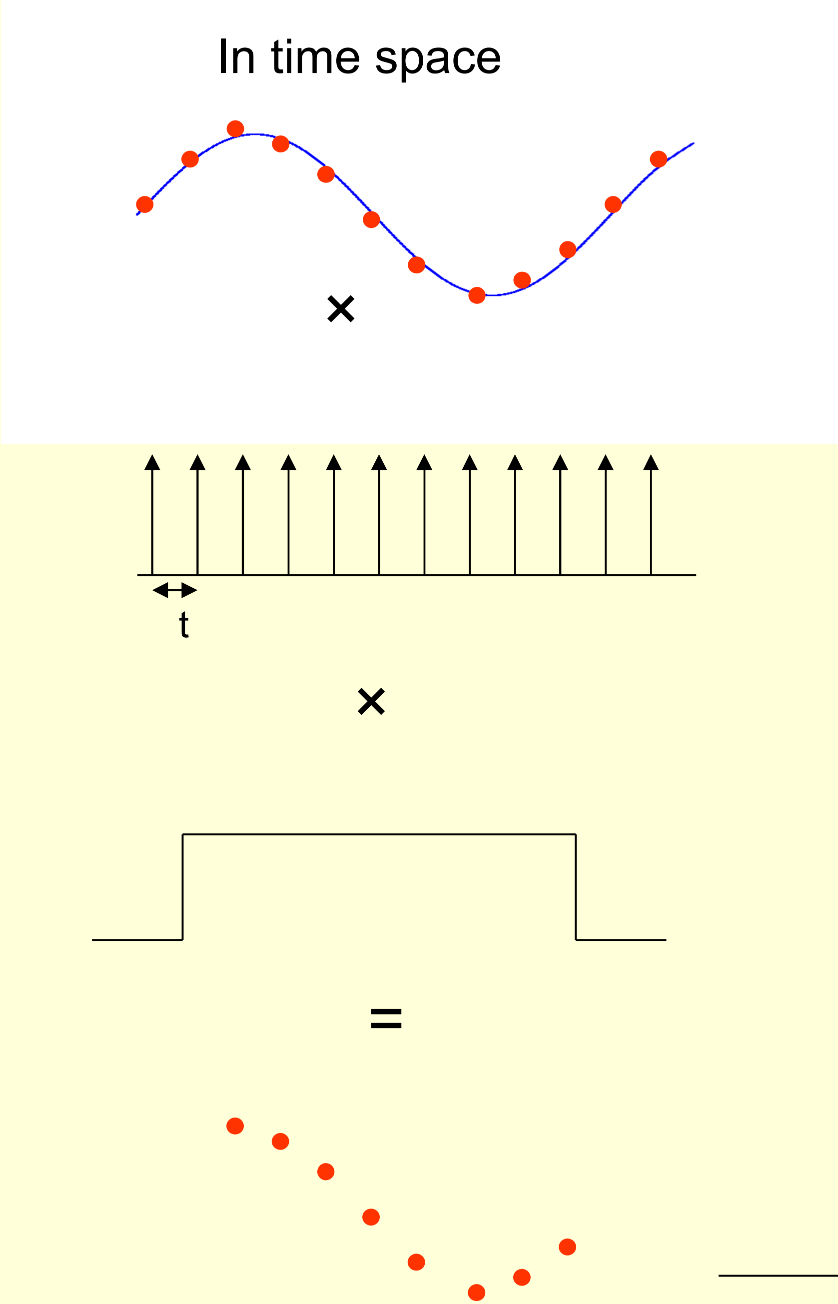


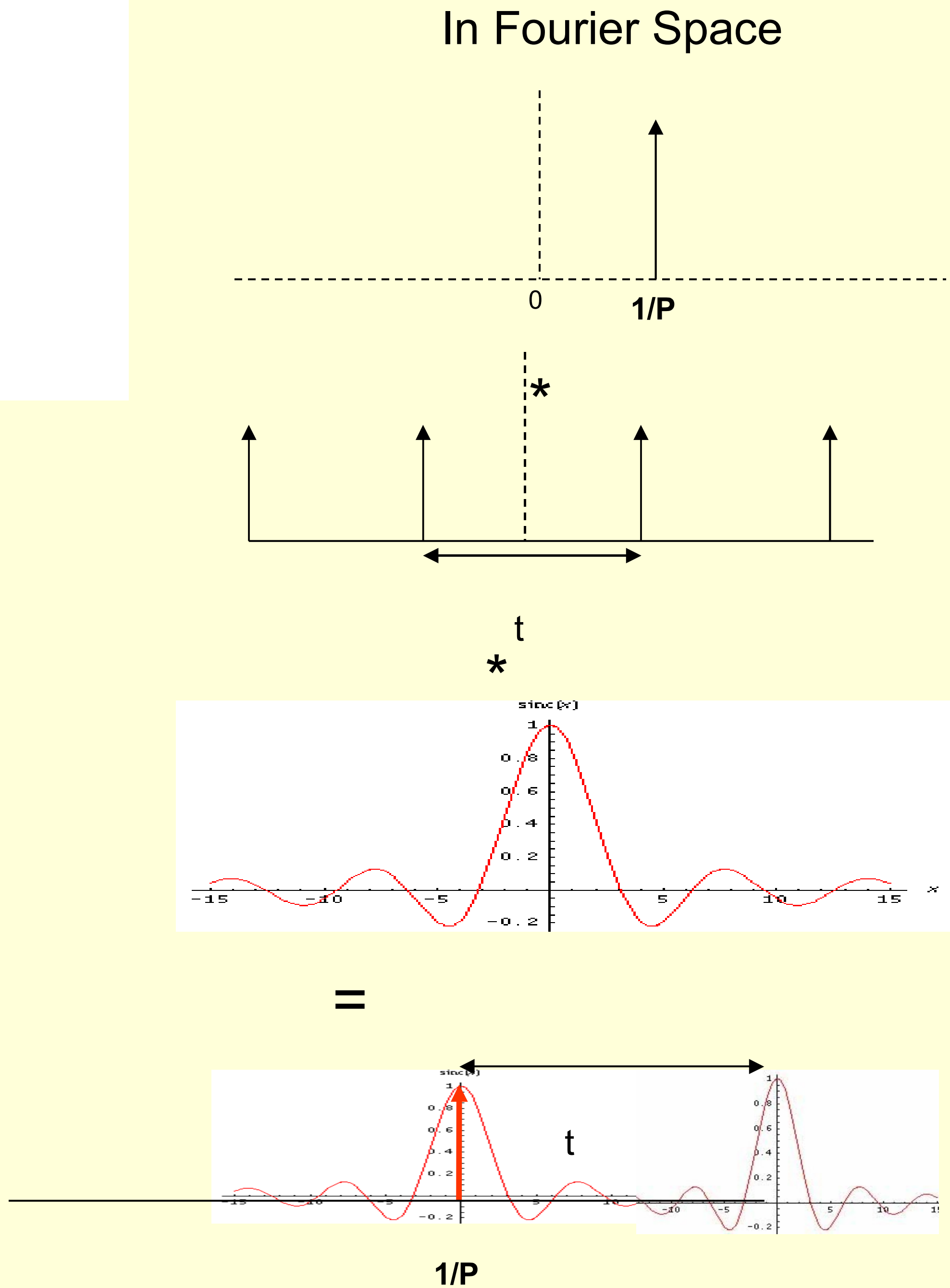
So why isn‘t it a -function?

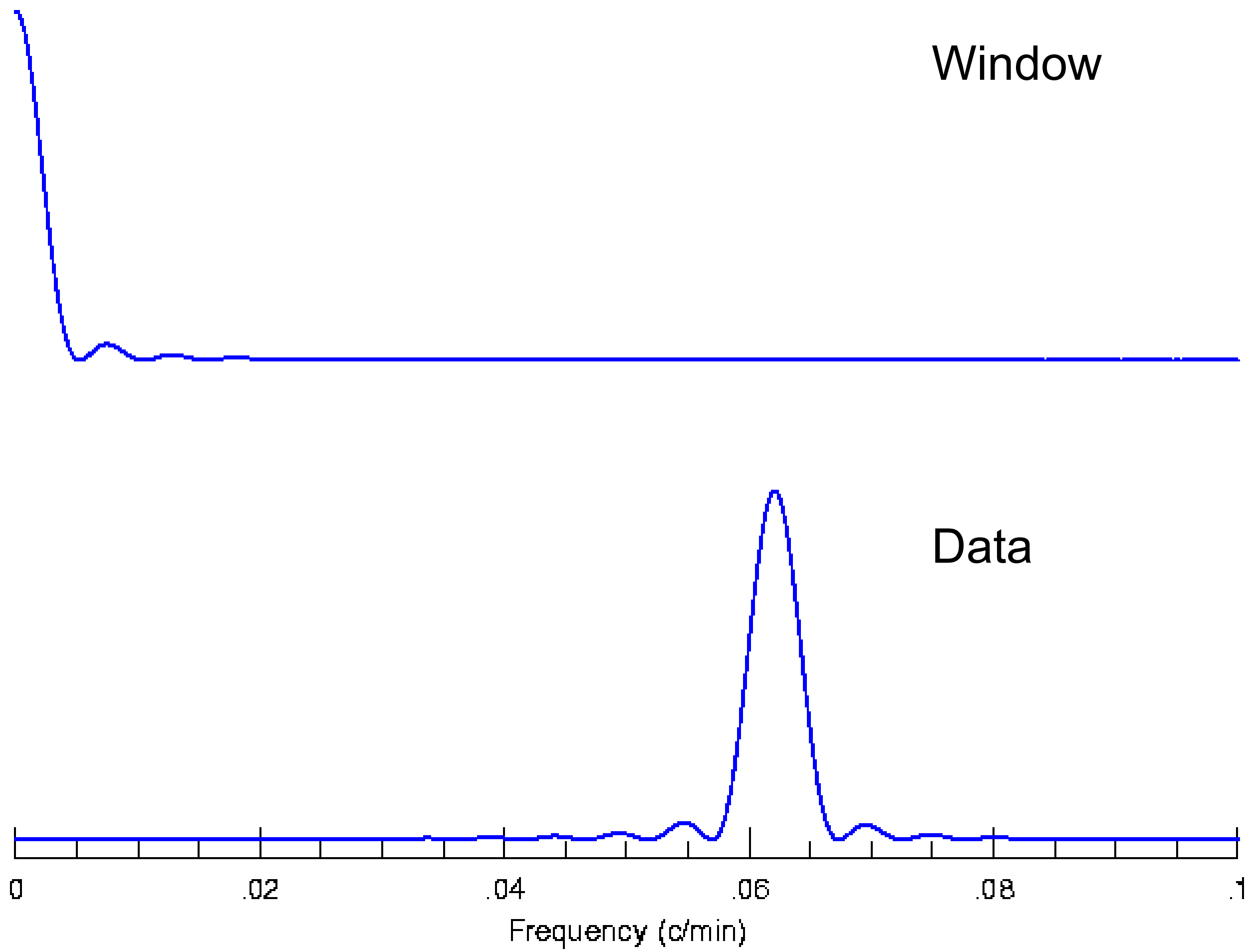
It would be if we measured the blue line out to infinity:



But we measure the red points. Our sampling degrades the delta function and introduces sidelobes

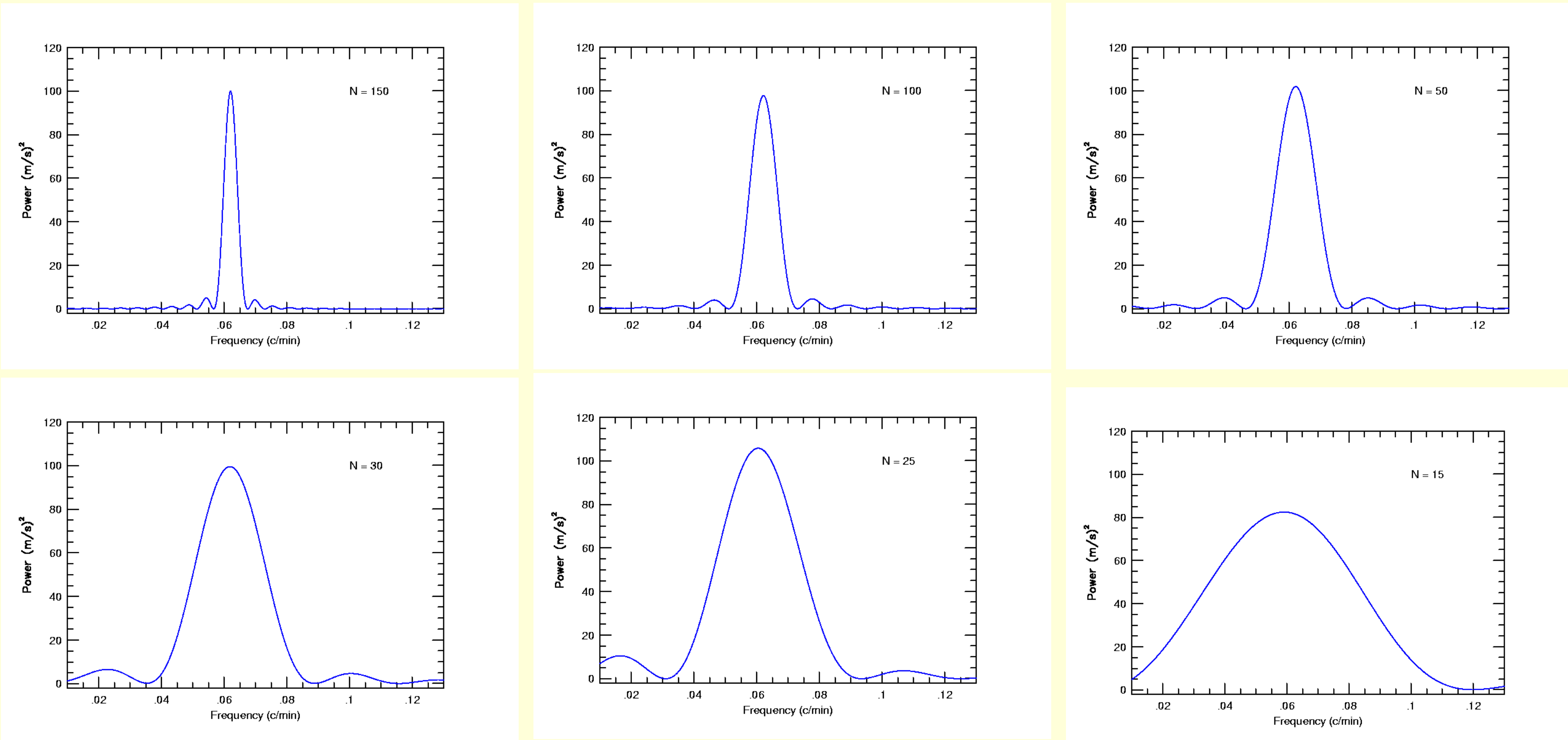

In time space
×
t
×
=


In Fourier Space
0
1/P
*
t
*
sinc(x)
=
t
1/P



16 min period sampled regularly for 3 hours

The longer the data window, the narrower is the width of the sinc function window:



Resolution of the spectra

Error in the period (frequency) of a peak in DFT:

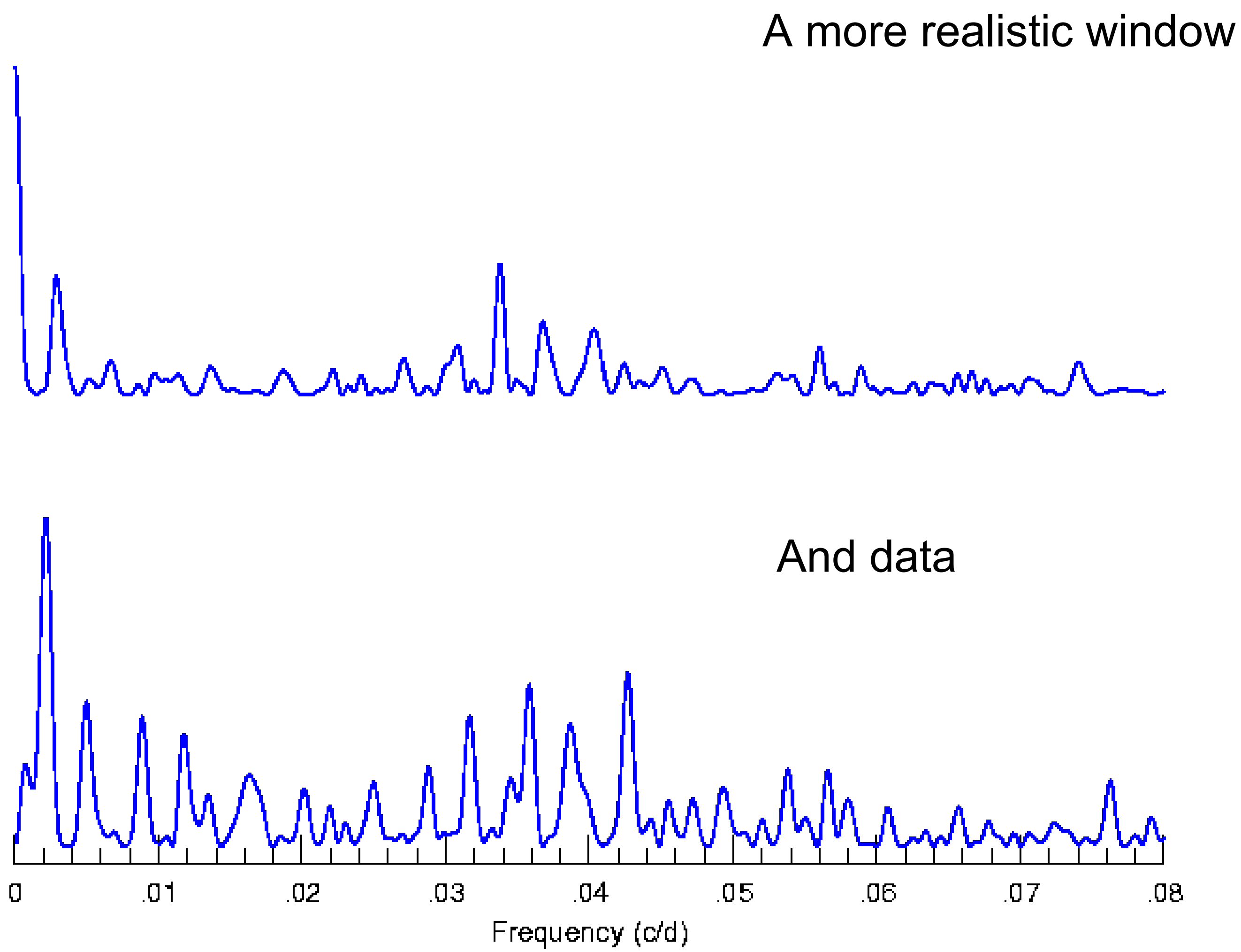
$$\frac{3}{2 N^{1/2} T A}$$

= error of measurement

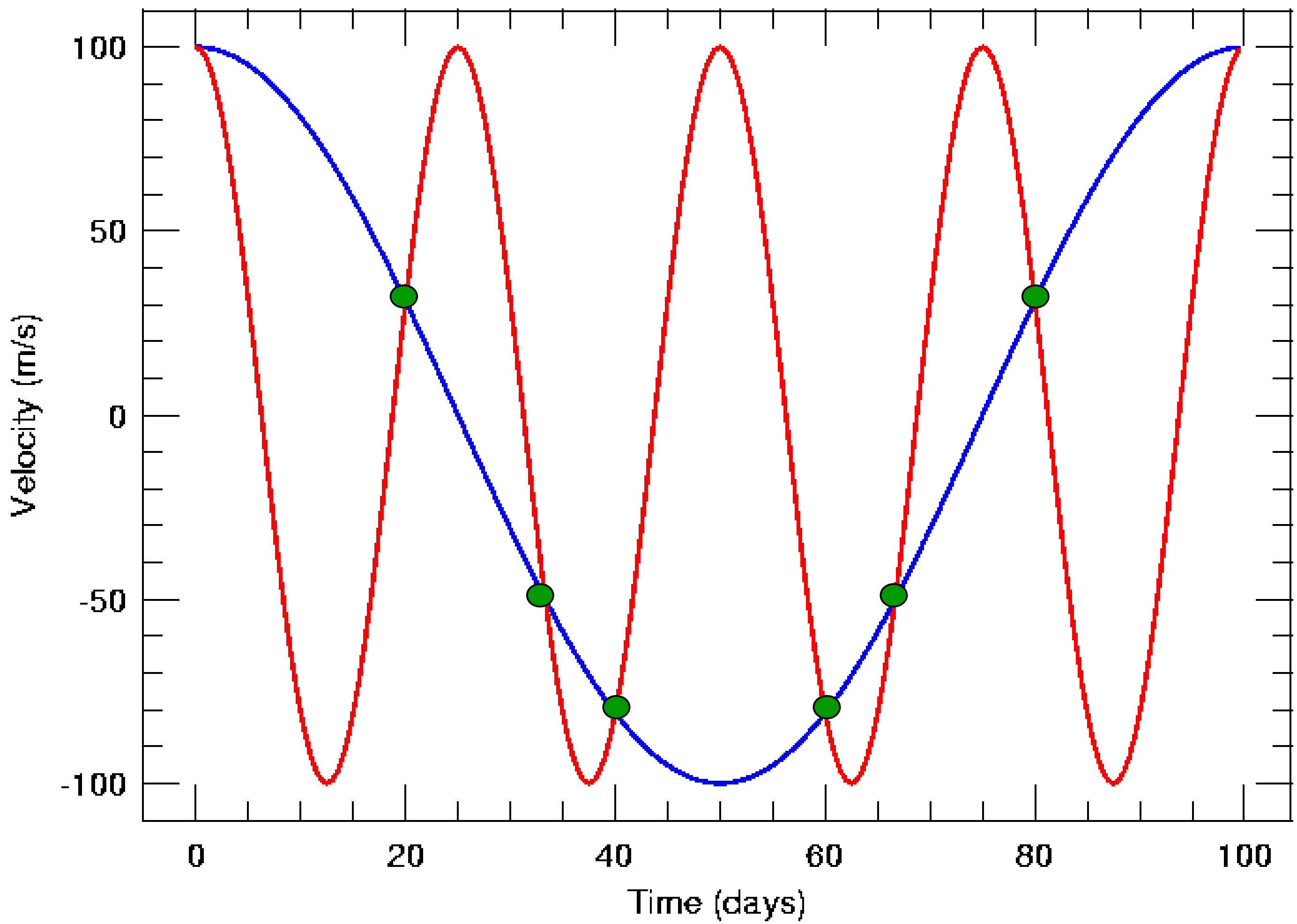
T = time span of your observations

A = amplitude of your signal

N = number of data points


A more realistic window
And data
0
.01
.02
.03
.04
.05
.06
.07
.08
Frequency (c/d)

Alias periods:



Undersampled periods appearing as another period

Irregularly spaced data
How to compute transforms?

General forms of sums we need on irregular grids:

$$F^k C(s_n) = \sum_{m=1}^{M} f^k(t_m)\cos(2\pi t_m s_n), \quad F^k S(s_n) = \sum_{m=1}^{M} f^k(t_m)\sin(2\pi t_m s_n)$$

$$\begin{aligned}
C(s) &= \sum_{m=1}^{M} \cos(2\pi t_m s), \\
S(s) &= \sum_{m=1}^{M} \sin(2\pi t_m s), \\
FC(s) &= \sum_{m=1}^{M} f(t_m)\cos(2\pi t_m s), \\
FS(s) &= \sum_{m=1}^{M} f(t_m)\sin(2\pi t_m s), \\
F^{(2)}C(s) &= \sum_{m=1}^{M} f^2(t_m)\cos(2\pi t_m s), \\
F^{(2)}S(s) &= \sum_{m=1}^{M} f^2(t_m)\sin(2\pi t_m s), \mathit{etc.},
\end{aligned}$$

We need to compute Fourier Transforms for formal sums:

$$f(t) = \sum_{m=1}^{M} f(t_m)\delta(t - t_m), \ -\infty < t < \infty$$

$$F(s) = \sum_{m=1}^{M} f(t_m)e^{-2\pi i t_m s}, \ -\infty < s < \infty$$

The further work is based on convolution theorem:

$$f(t) * k(t) = \int_{-\infty}^{\infty} f(u)k(t-u)du \quad \Longleftrightarrow \quad G(s) = F(s)K(s)$$

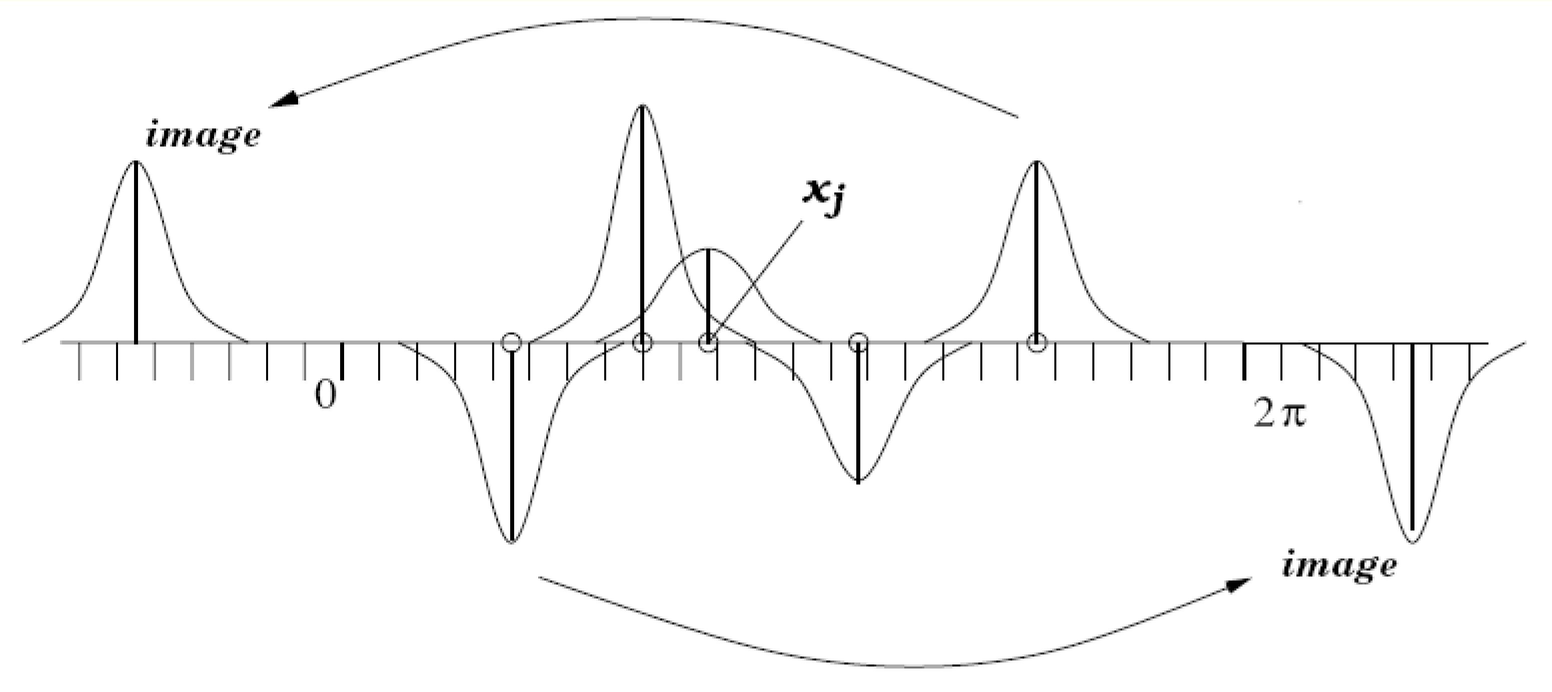
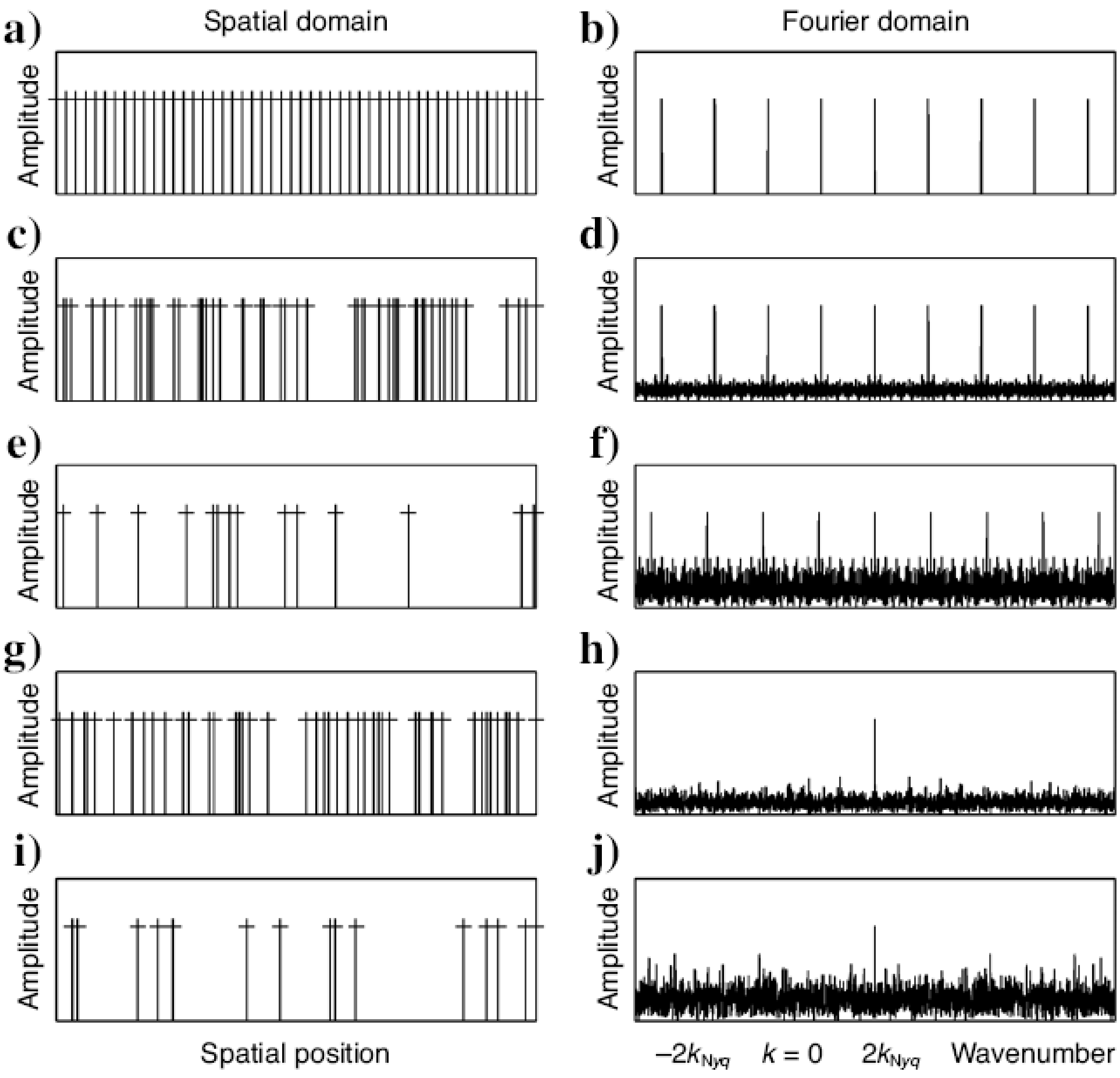
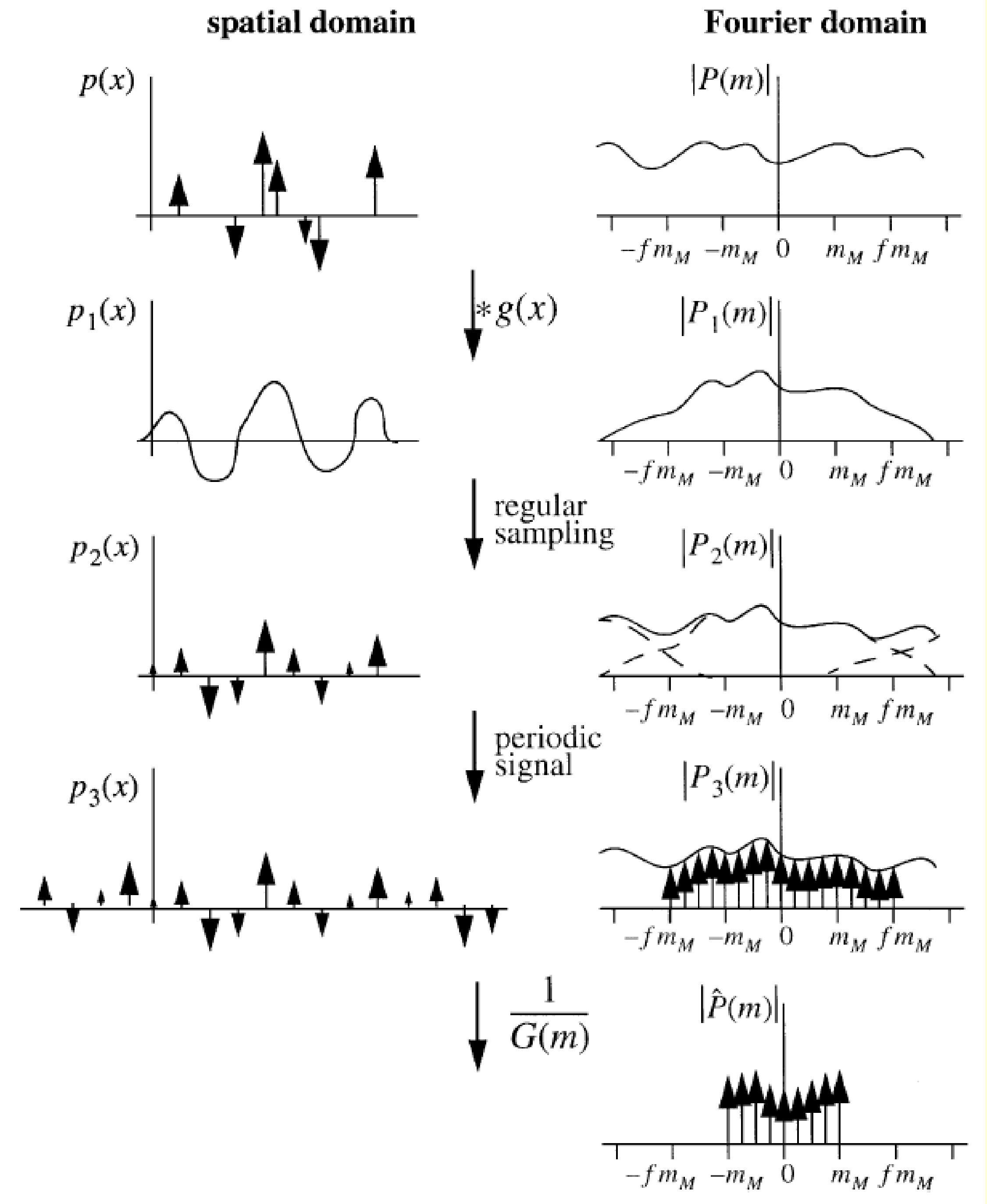


image
x_j
0
2π
image


a)
Spatial domain
b)
Fourier domain
Amplitude
c)
d)
e)
f)
g)
h)
i)
j)
Spatial position
−2k_Nyq
k = 0
2k_Nyq
Wavenumber



spatial domain
Fourier domain
p(x)
|P(m)|
−fm_M −m_M 0 m_M fm_M
p_1(x)
*g(x)
|P_1(m)|
−fm_M −m_M 0 m_M fm_M
regular sampling
p_2(x)
|P_2(m)|
−fm_M −m_M 0 m_M fm_M
periodic signal
p_3(x)
|P_3(m)|
−fm_M −m_M 0 m_M fm_M
1/G(m)
|P̂(m)|
−fm_M −m_M 0 m_M fm_M

1. Smear the sum of the weighted δ-functions with appropriate kernel function to obtain continuous function.
2. Back-transform obtained continuous function by using numerical evaluation of the Fourier Transform (this is the point where well-known Fast Fourier Transform is involved).
3. Compensate result by dividing it with the Fourier transformed smearing kernel.

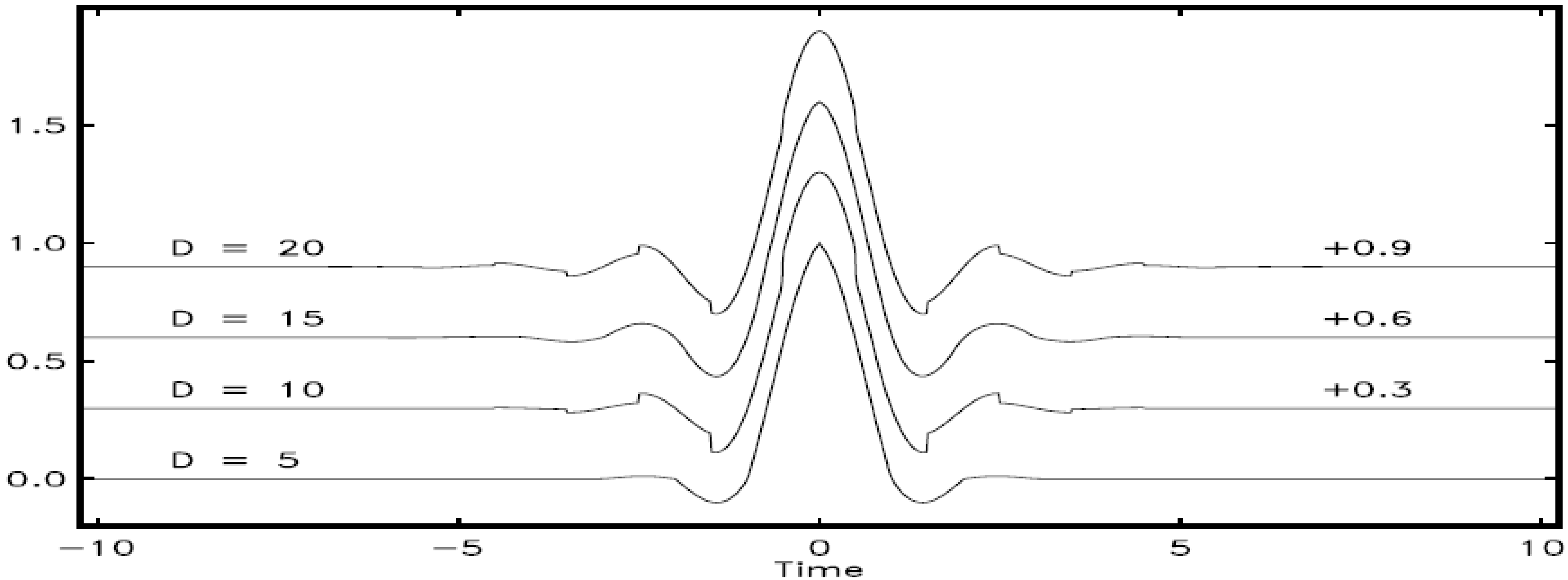


Figure 1. Four Lagrange kernels, with different degrees of interpolating polynomial

1. Truncated Gaussian kernel (see Dutt & Rokhlin (1993))

$$k(t, q, b) = \frac{1}{2\sqrt{\pi b}} \Pi(\frac{t}{q}) e^{-\frac{t^2}{4b}},$$

   where Π is rectangle function.
2. The method of reverse interpolation by Press & Rybicki (1989) can be interpreted as smoothing by Lagrange kernel (see Fig. 1).
3. Combined kernel (see Sramek & Schwab (1988))

$$k(t, q, a, b) = \frac{1}{2\sqrt{\pi b}} \Pi(t/q) \mathrm{sinc}(\pi a t) e^{-\frac{t^2}{4b}},$$

   combines properties of theoretically ideal sinc kernel and Gaussian kernel.

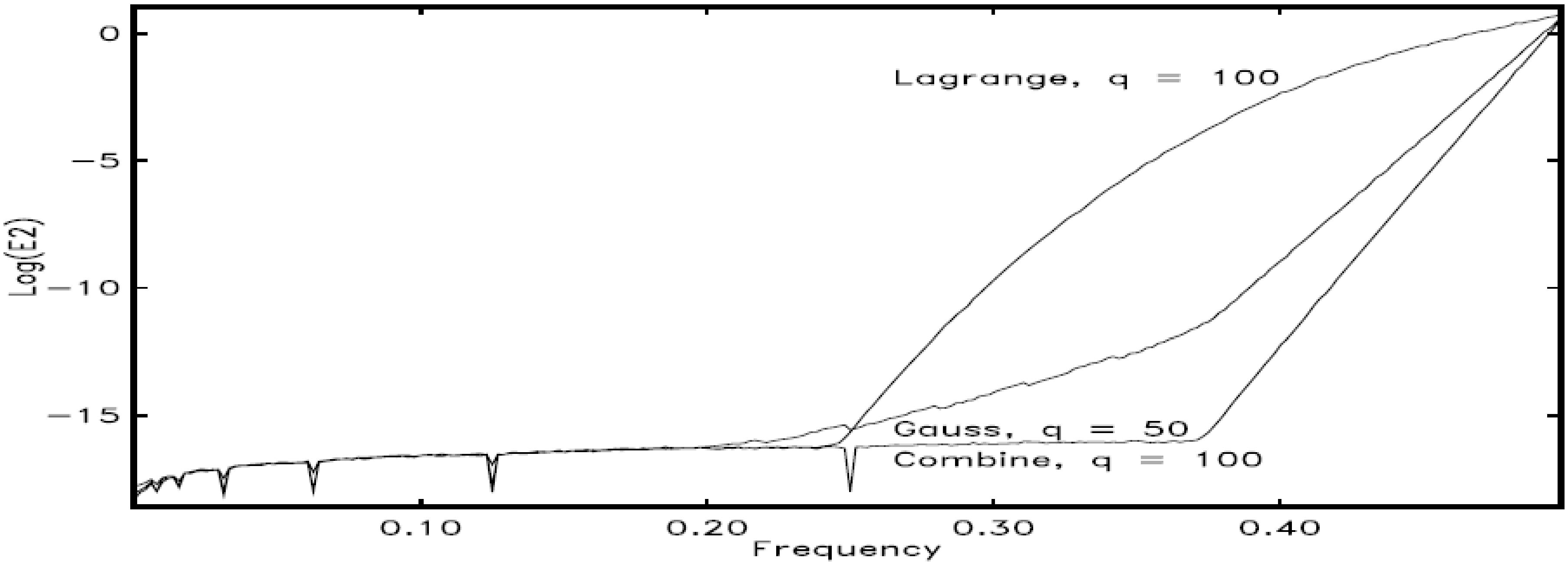


Figure 2. Comparison of the error levels for the three kernels

Alias periods:

$$P_{\text{false}}^{-1} = P_{\text{alias}}^{-1} + P_{\text{true}}^{-1}$$

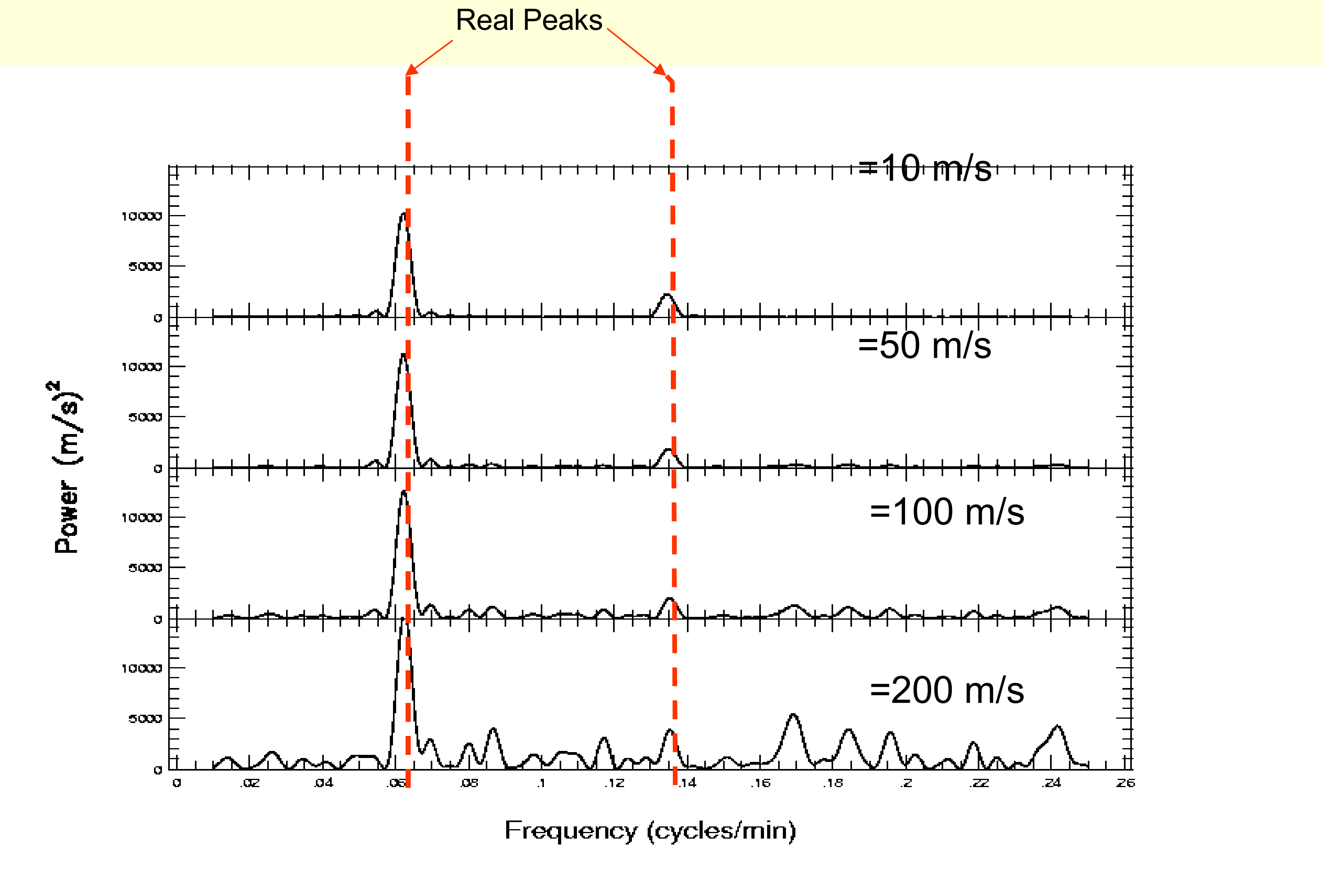
Common Alias Periods:

$$P_{\text{false}}^{-1} = (1\ day)^{-1} + P_{\text{true}}^{-1} \qquad \text{day}$$

$$P_{\text{false}}^{-1} = (29.53\ d)^{-1} + P_{\text{true}}^{-1} \qquad \text{month}$$

$$P_{\text{false}}^{-1} = (365.25\ d)^{-1} + P_{\text{true}}^{-1} \qquad \text{year}$$

The Effects of Noise



2 sine waves ampltudes of 100 and 50 m/s. Noise added at different levels

Lomb-Scargle Periodogram

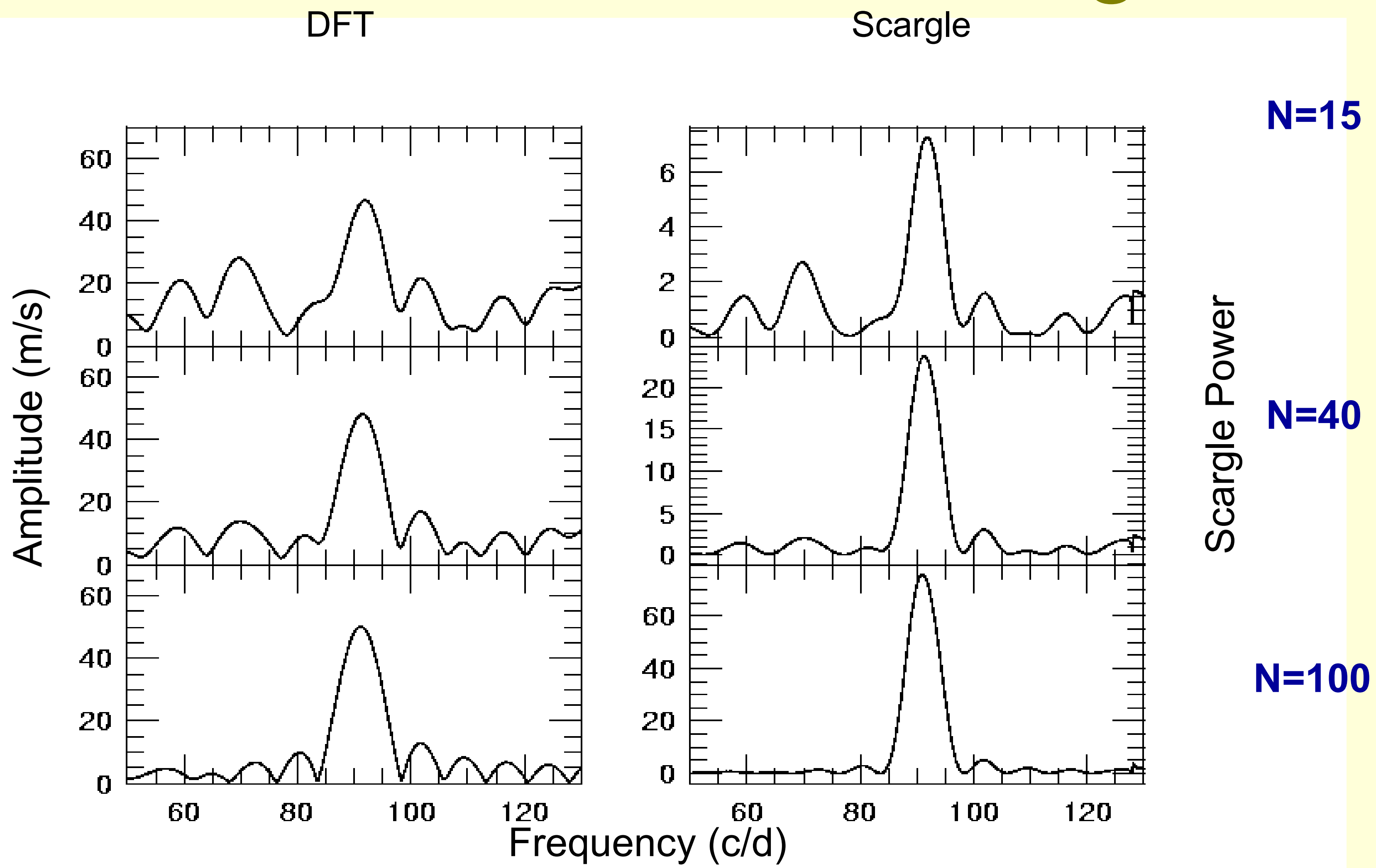
$$P_x(\;\;) = \frac{1}{2}\,\frac{\left[\;_{j}\; X_j \cos \;\; t_j - \;\right]^2}{\;_{j}\; X_j \cos^2 \;\; t_j - \;} + \frac{1}{2}\,\frac{\left[\;_{j}\; X_j \sin \;\; t_j - \;\right]^2}{\; X_j \sin^2 \;\; t_j - \;}$$

$$\tan(2\;\;\;) = \;_{j}\; \sin 2\;\; t_j)/\;_{j}\; \cos 2\;\; t_j)$$

Power is a measure of the statistical significance of that frequency (period):

Scargle, Astrophysical Journal, 263, 835, 1982

DFT versus Lomb-Scargle



DFTs give you the amplitude of a periodic signal in the data. This does not change with more data. The Lomb-Scargle power gives you the statistical significance of a period. The more data you have the more significant the detection is, thus the higher power with more data

False Alarm Probability (FAP)

The FAP is the probability that random noise will produce a peak with Lomb-Scargle Power the same as your observed peak in a certain frequency range

Unknown period:

$$FAP \approx 1 - (1 - e^{-P})^{N}$$

Where P = Scargle Power

N = number of independent frequencies in the frequency range of interest

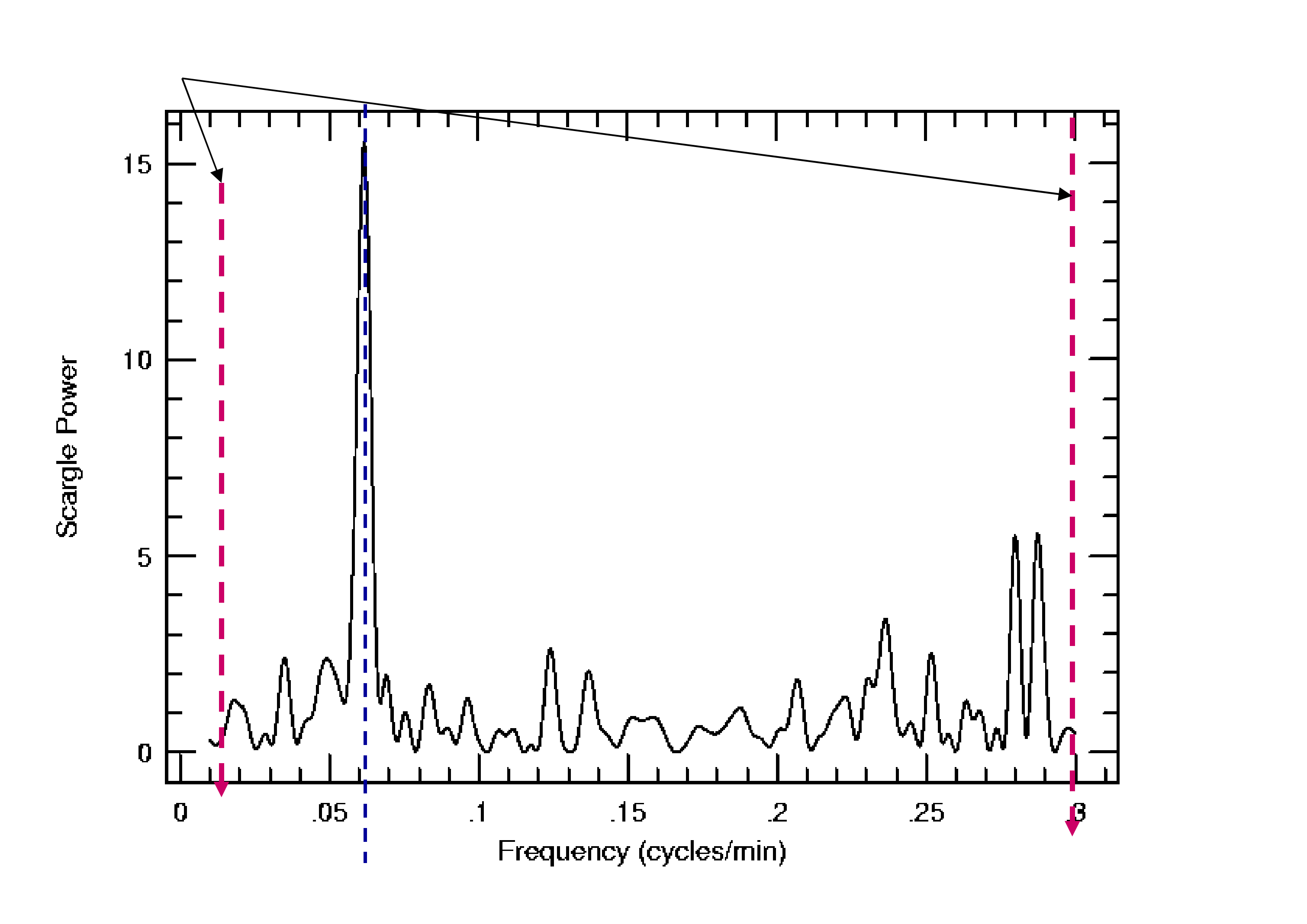
Known period:

$$FAP \approx e^{-P}$$

In this case you have only one independent frequency

Scargle Power (significance) is increased by lower level of noise and/or more data points

The probability that noise can produce the highest peak over a range ≈ $1 - (1 - e^{-P})^N$



The probability that noise can produce the this peak exactly at this frequency = e^{-P}

Determining FAP

To use the Scargle formula you need the number of independent frequencies.

How do you get the number of independent Frequencies?

First Approximation: Use the number of data points N_0

Horne & Baliunas (1986, Astrophysical Journal, 302, 757):

$N_i = -6.362 + 1.193\, N_0 + 0.00098 N_0^2$ = number of independent frequencies

Use Scargle FAP only as an estimate. A more valid determination of the FAP requires Monte Carlo Simulations:

Method 1:

- Create random noise at the same level as your data
- Sample the random noise in the same manner as your data
- Calculate Scargle periodogram of noise and determine highest peak in frequency range of interest
- Repeat 1.000-100.000 times = N_{total}
- Add the number of noise periodograms with power greater than your data = N_{noise}
- FAP = N_{noise}/N_{total}

Assumes Gaussian noise. What if your noise is not Gaussian, or has some unknown characteristics?

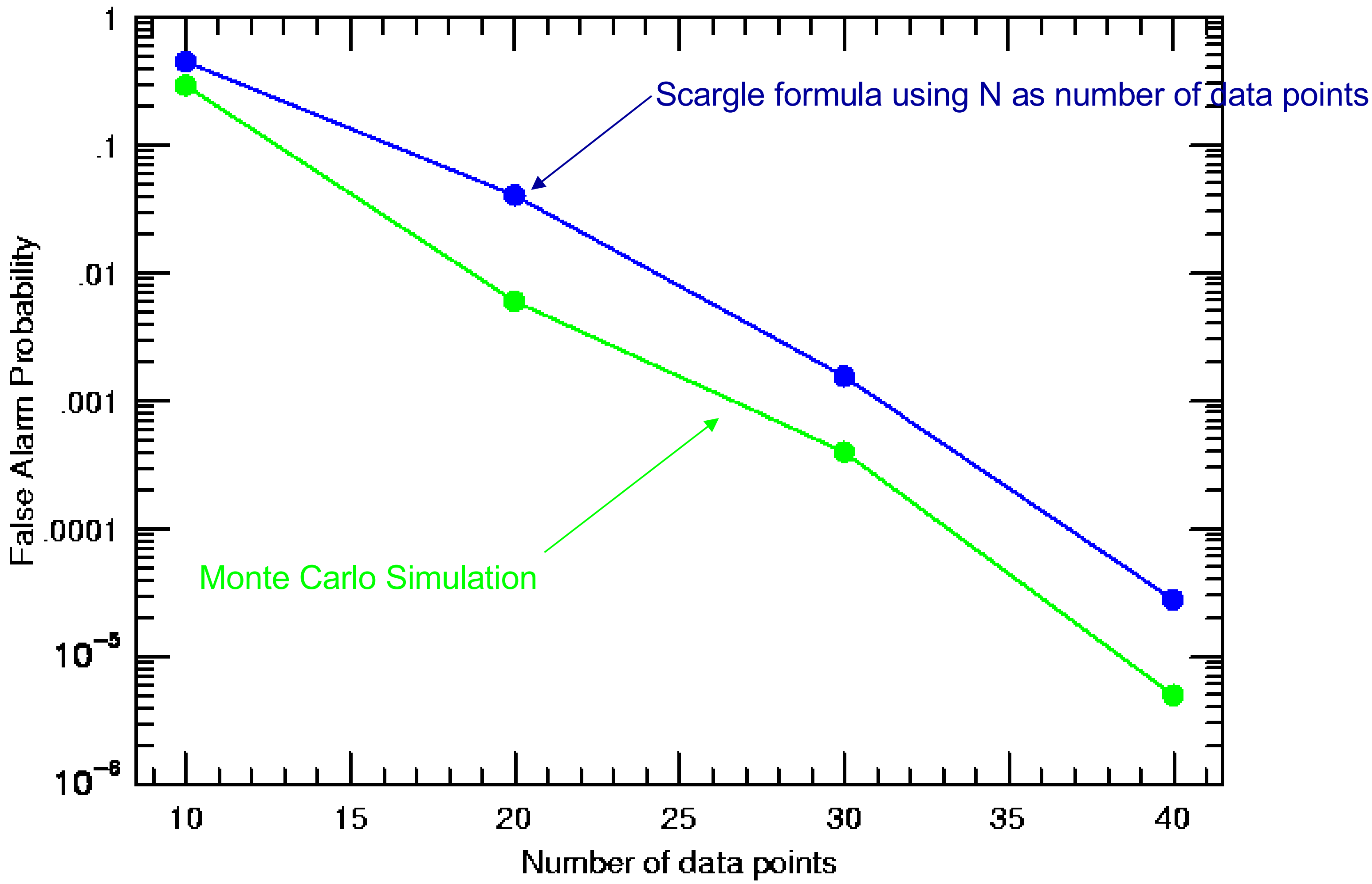
Use Scargle FAP only as an estimate. A more valid determination of the FAP requires Monte Carlo Simulations:

Method 2:

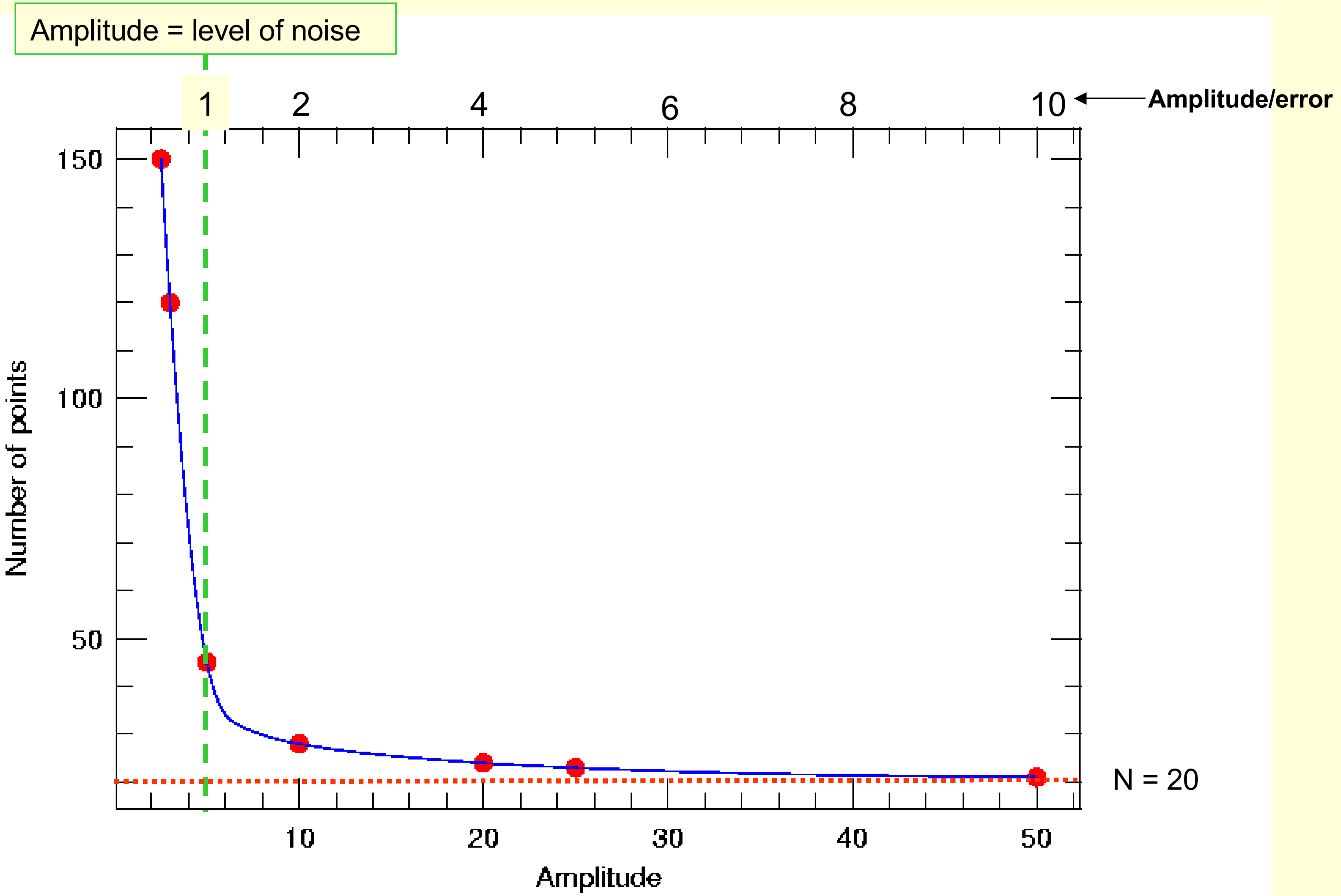
- Randomly shuffle the measured values (velocity, light, etc) keeping the times of your observations fixed
- Calculate Scargle periodogram of random data and determine highest peak in frequency range of interest
- Reshuffle your data 1.000-100.000 times = N_{total}
- Add the number of „random“ periodograms with power greater than your data = N_{noise}
- FAP = N_{noise}/N_{total}

Advantage: Uses the actual noise characteristics of your data

FAP comparisons

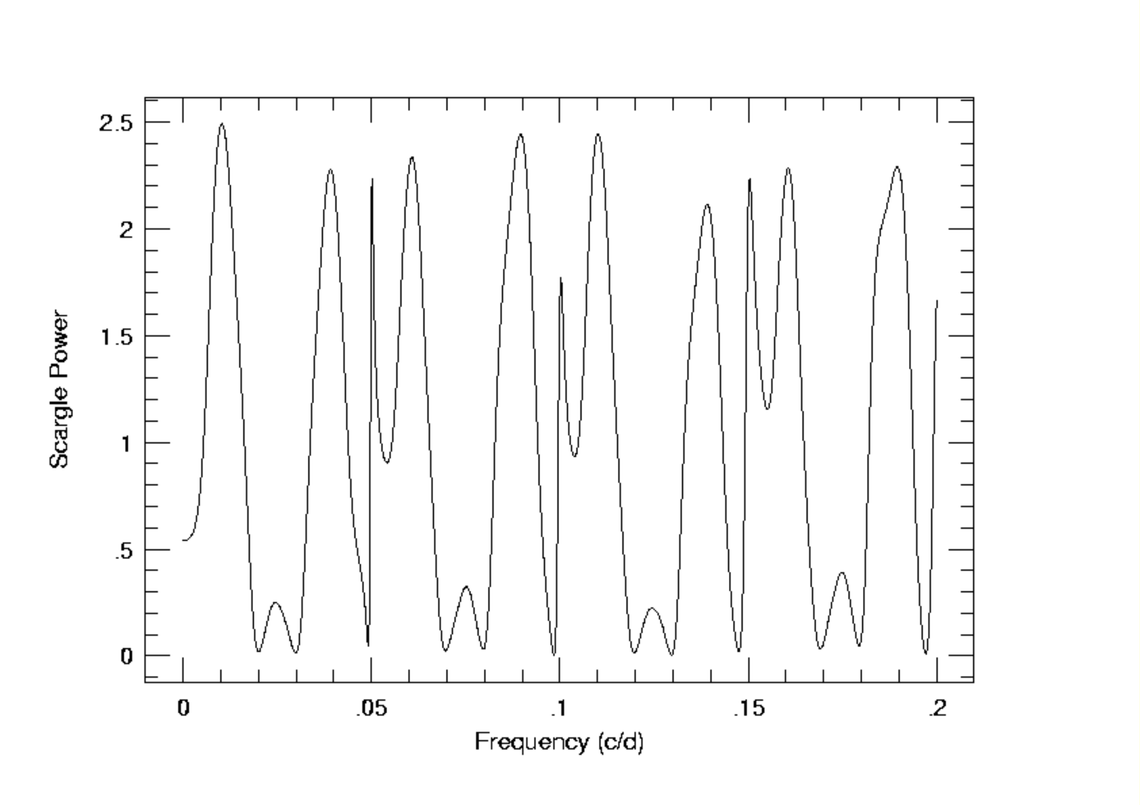


Using formula and number of data points as your independent frequencies may overestimate FAP, but each case is different.



Number of measurements needed to detect a signal of a certain amplitude. The FAP of the detection is 0.001. The noise level is = 5 m/s. Basically, the larger the measurement error the more measurements you need to detect a signal.

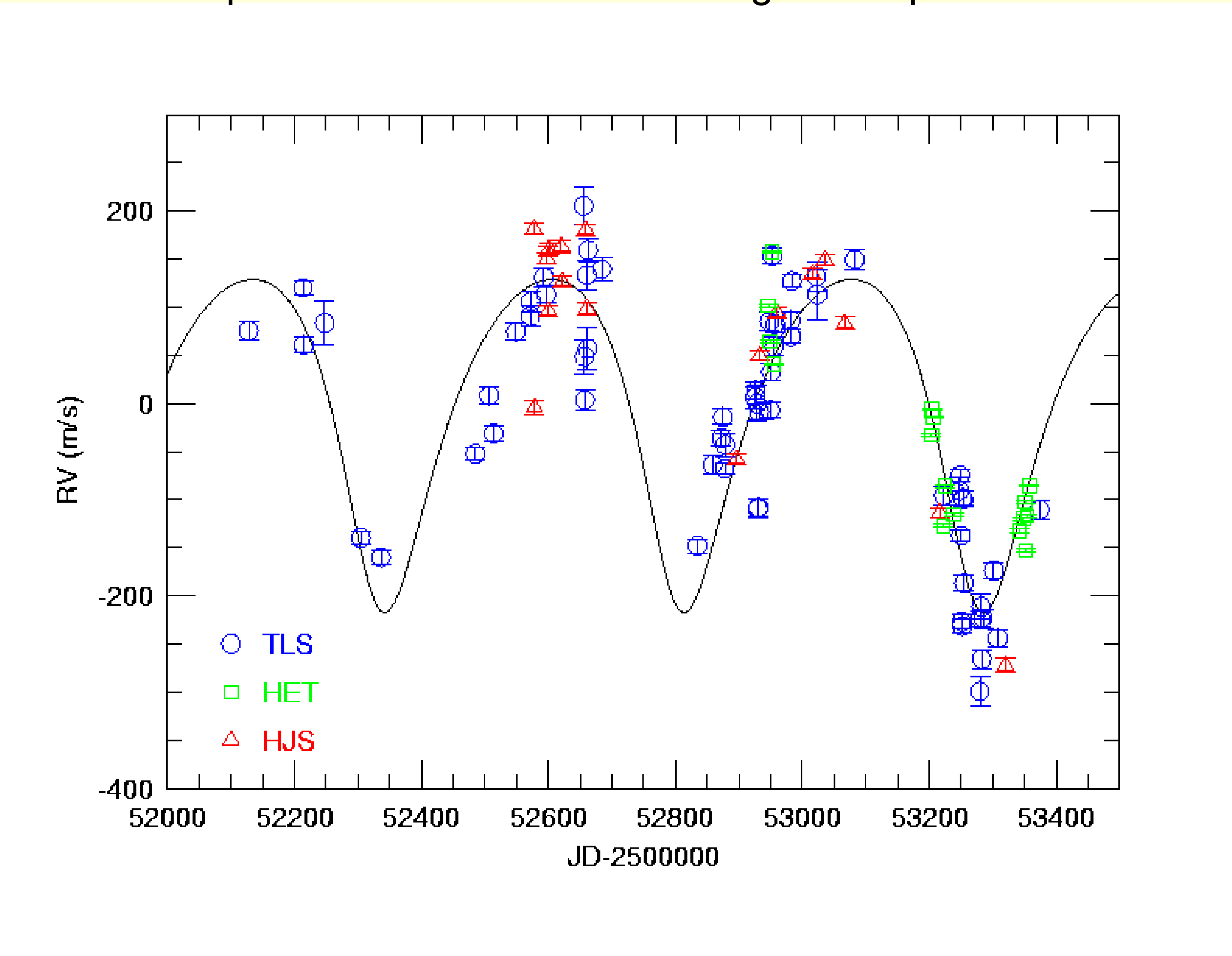
Lomb-Scargle Periodogram of 6 data points of a sine wave:

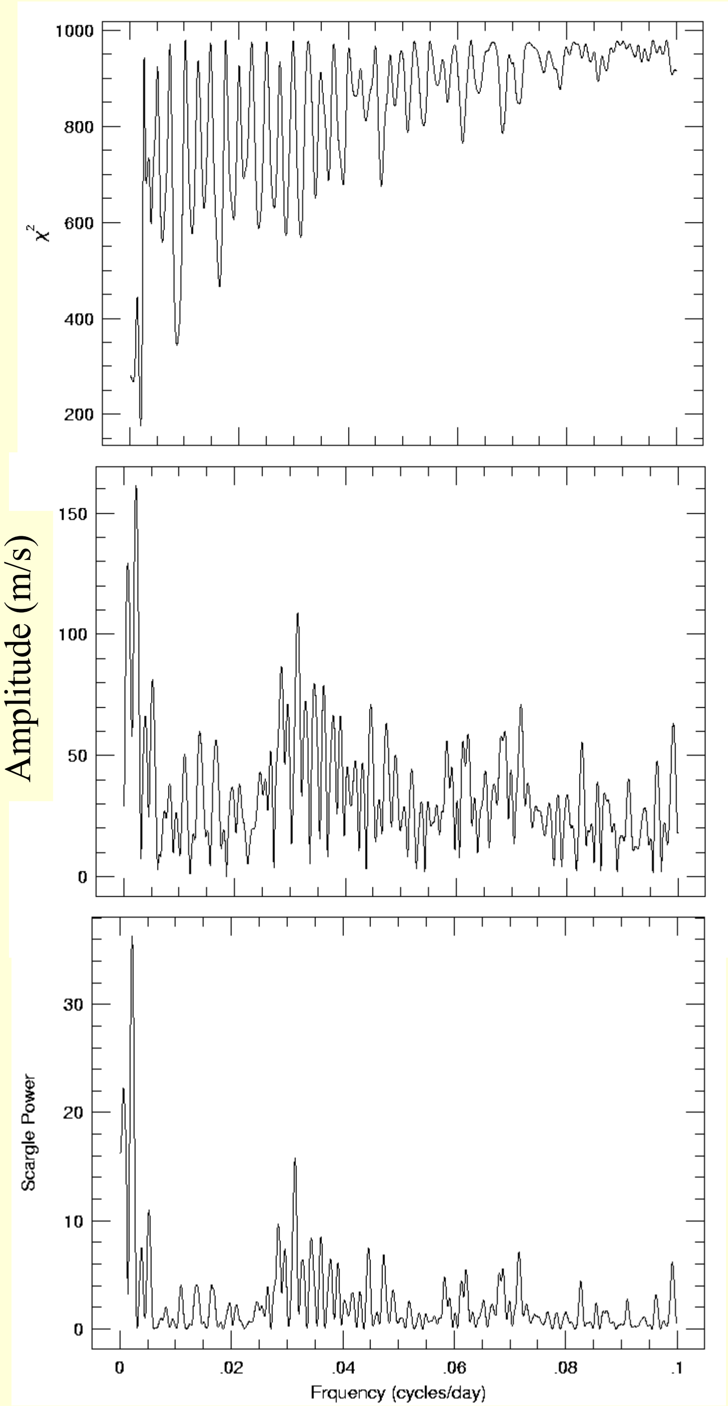


Lots of alias periods and false alarm probability (chance that it is due to noise) is 40%!

For small number of data points do not use Scargle, sine fitting is best. But be cautious!!

Comparison of the 3 Period finding techniques





Least squares sine fitting: The best fit period (frequency) has the lowest ^{2}

Discrete Fourier Transform: Gives the power of each frequency that is present in the data. Power is in $(m/s)^2$ or (m/s) for amplitude

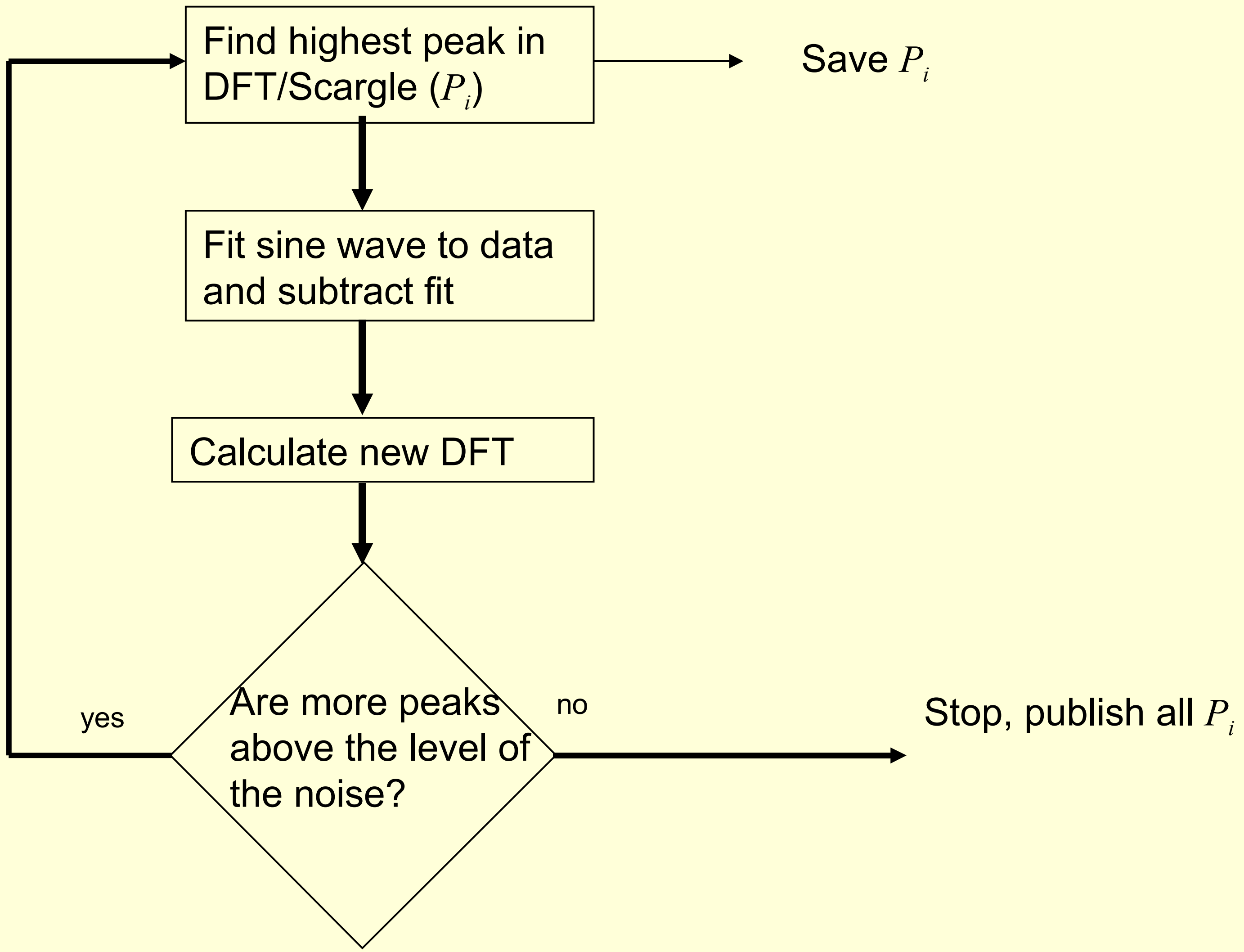
Lomb-Scargle Periodogram: Gives the power of each frequency that is present in the data. Power is a measure of statistical signficance

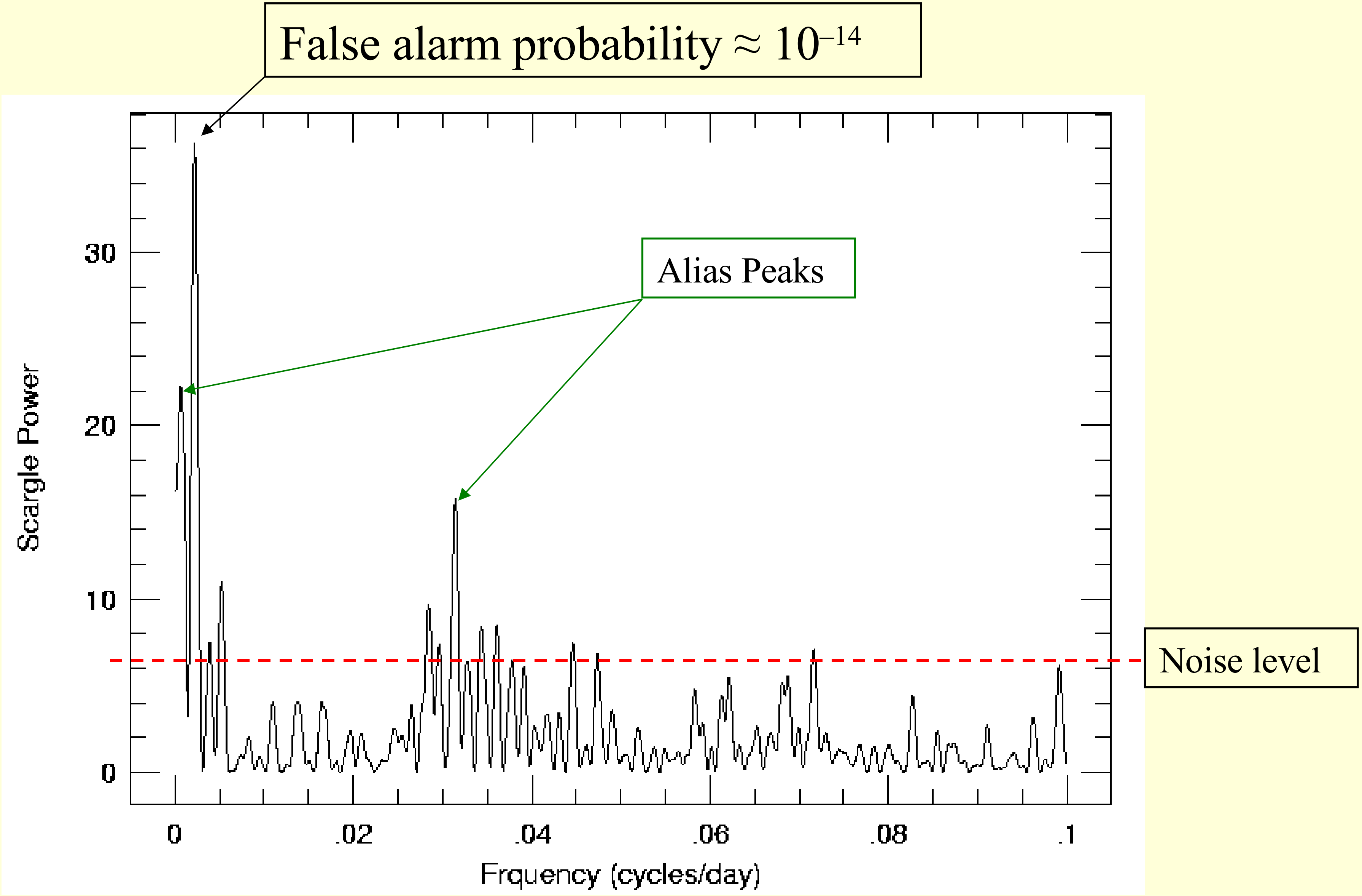
Finding Multiple Periods in Data: Pre-whitening

What if you have multiple periods in your data? How do you find these and make sure that these are not due to alias effects of your sampling window.

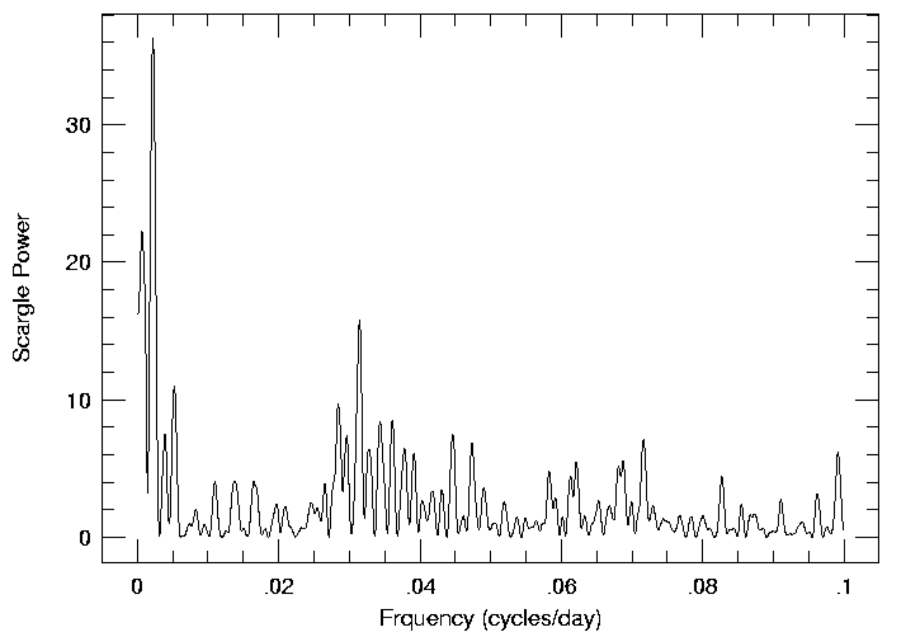
Standard procedure: ***Pre-whitening.*** Sequentially remove periods from the data until you reach the level of the noise

Prewhitening flow diagram:

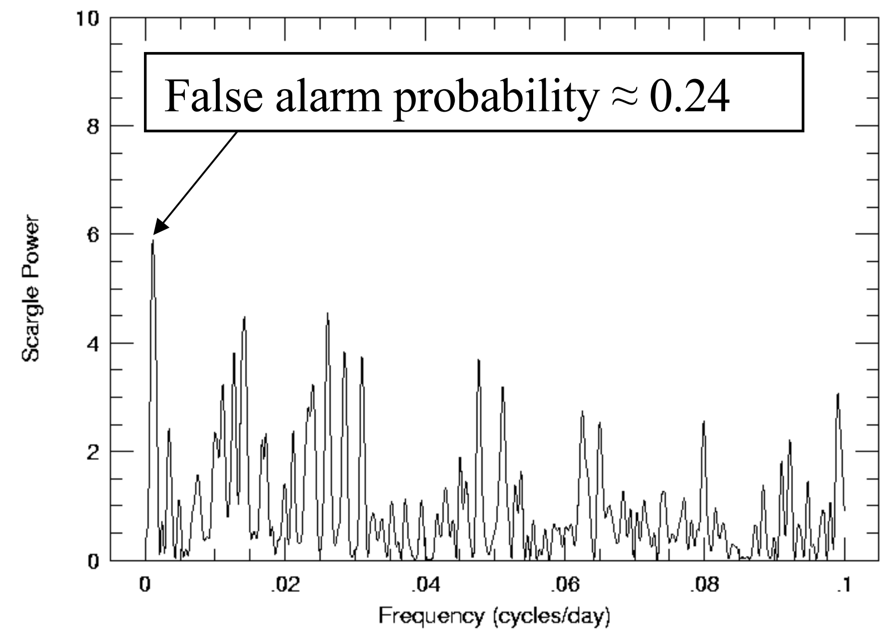




False alarm probability ≈ 10^{-14}
Alias Peaks
Noise level
Scargle Power
30
20
10
0
0
.02
.04
.06
.08
.1
Frquency (cycles/day)

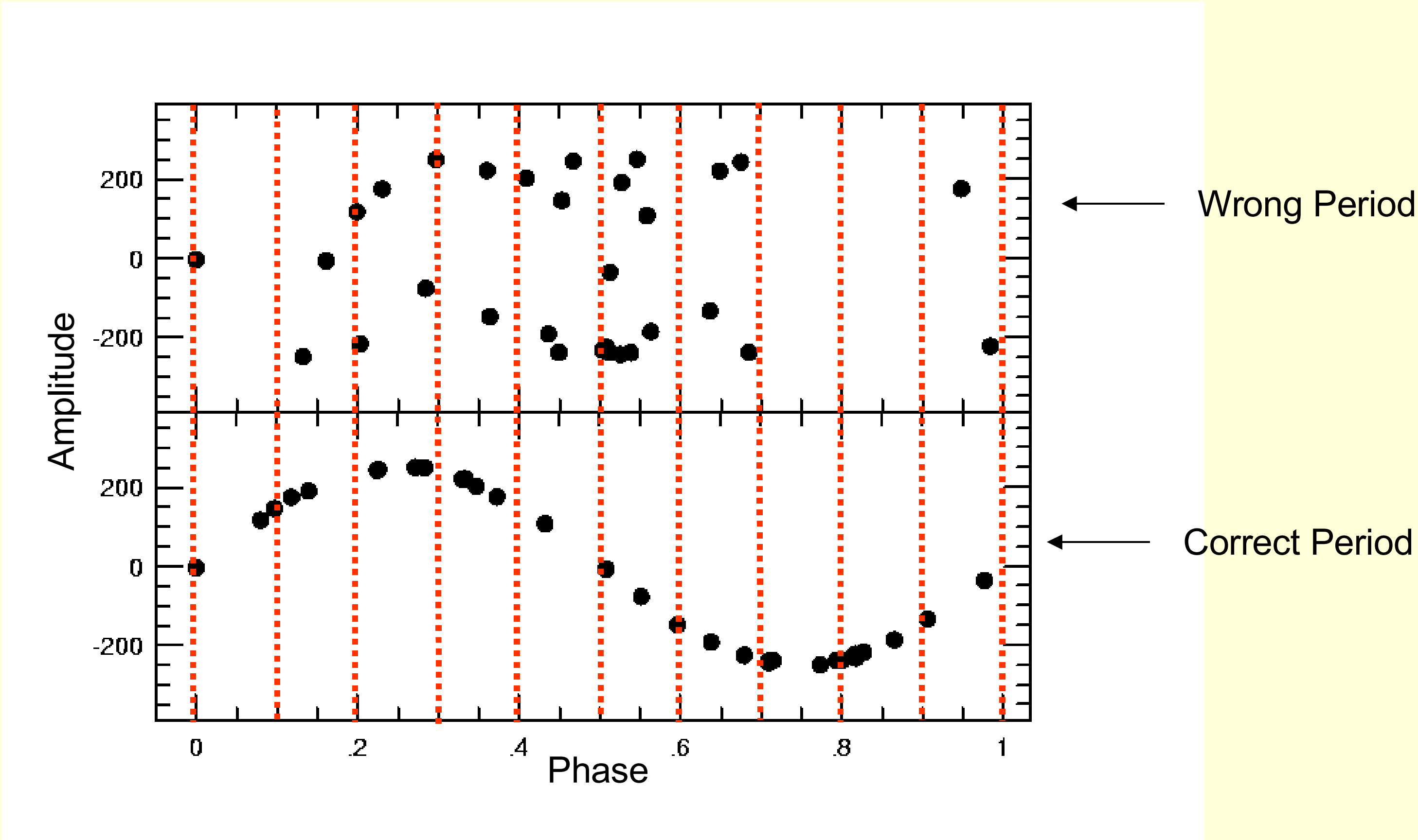


Raw data



After removal of dominant period

Other Techniques: Phase Dispersion Minimization



Choose a period and phase the data. Divide phased data into *M* bins and compute the standard deviation in each bin. If $\ ^2$ is the variance of the time series data and s^2 the total variance of the *M* bin samples, the correct period has a minimum value of

$$= s^2/\ ^2$$

See Stellingwerf, Astromomical Journal, 224, 953, 1978

General phase dispersion

$$\theta(s) = \frac{\sum_{m=1}^{M-1} \sum_{l=m+1}^{M} g(t_m, t_l, s)\left[f(t_m) - f(t_l)\right]^2}{\sum_{m=1}^{M-1} \sum_{l=m+1}^{M} g(t_m, t_l, s)}$$

can be expressed as

$$\theta(s) = \frac{Q_1(s) - Q_2(s)}{Q_3(s)},$$

where truncated partial sums

$$Q_1(s) = \sum_{r=0}^{R} d_r[F^2C(rs) \times F^0C(rs) + F^2S(rs) \times F^0S(rs)],$$

and

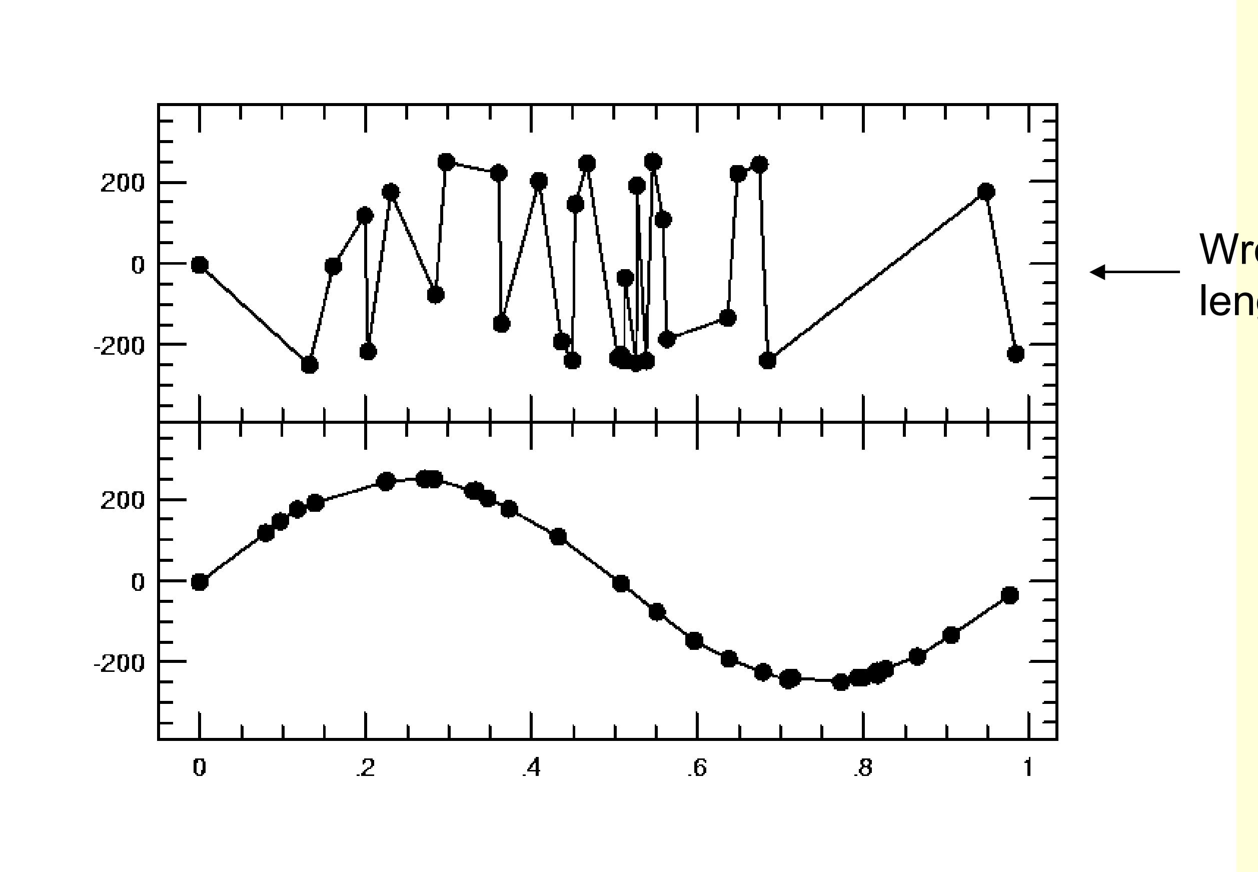
$$Q_2(s) = 2\sum_{r=0}^{R} d_r[FC^2(rs) + FS^2(rs)],$$

and also normalization coefficient

$$Q_3(s) = \sum_{r=0}^{R} d_r[F^0C^2(rs) + F^0S^2(rs)],$$

can be evaluated from trigonometric sums, which can be easily computed using basic algorithm.

Other Techniques: String Length Method



Wrong period longer string length

Phase the data to a test period and minimize the distance between adjacent points

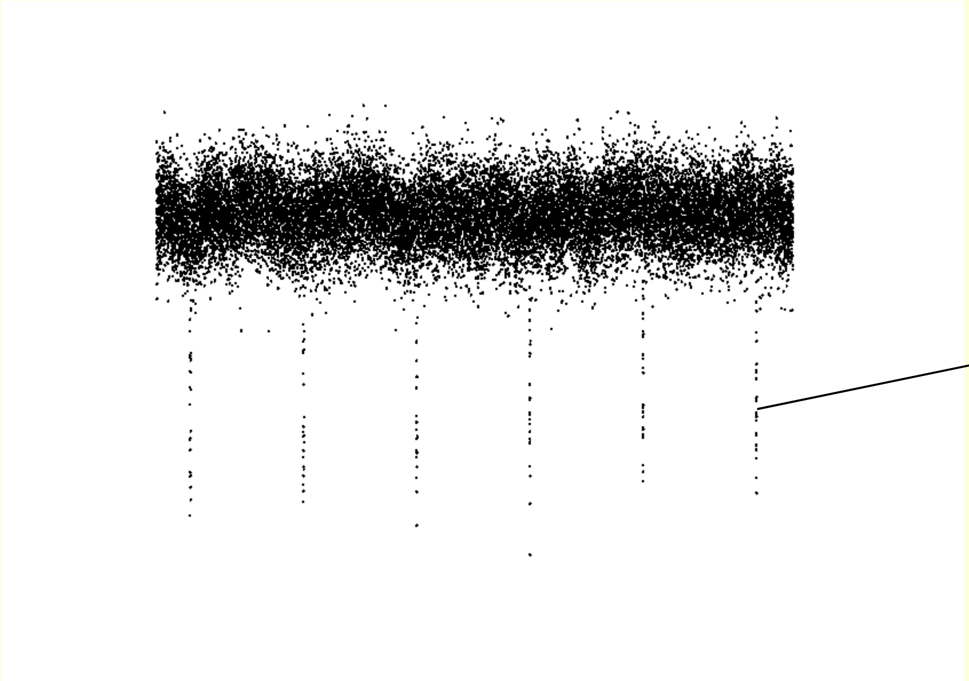
Lafler & Kinman, Astrophysical Journal Supplement, 11, 216, 1965

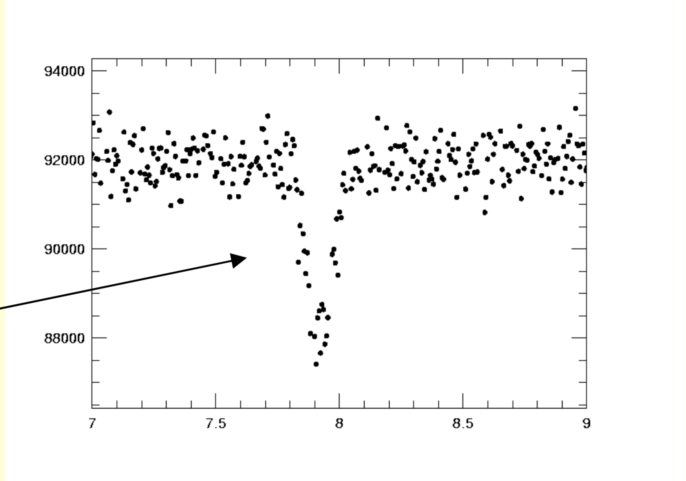
Dworetsky, Monthly Notices Astronomical Society, 203, 917, 1983

5. Other Techniques: Wavelet Analysis

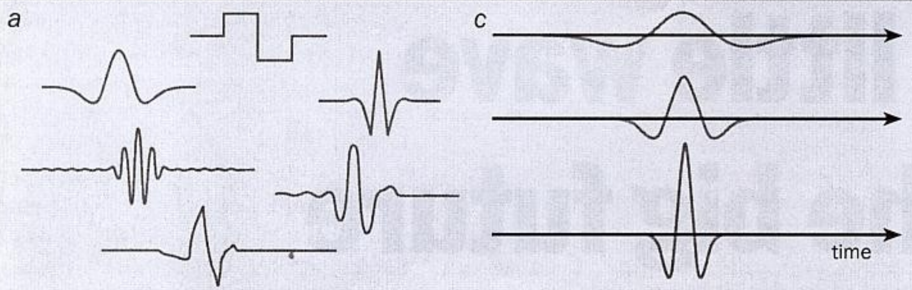
This technique is ideal for finding signals that are aperiodic, noisy, intermittent, or transient.

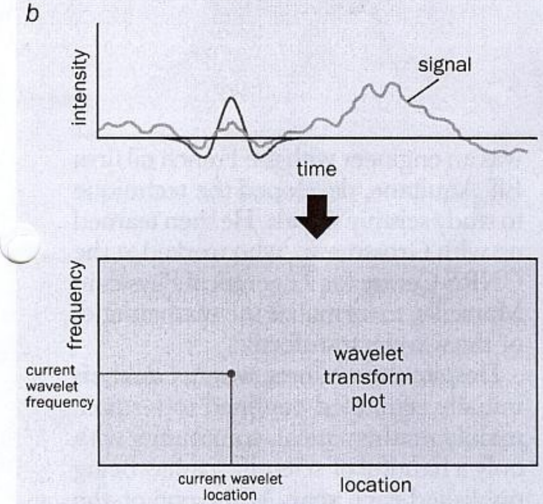
Recent interest has been in transit detection in light curves





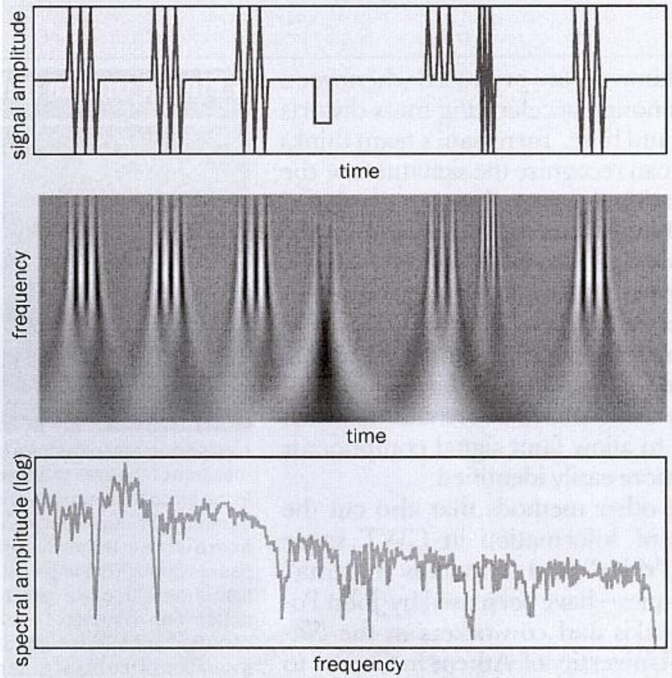
1 The secrets of the little wave





(*a*) Wavelets are used to transform a signal into another representation that presents the information in a more useful form. A few common examples of wavelets are shown here, including the square "Haar" wavelet. (*b*) In the wavelet-analysis technique the wavelet function (red) is "convoluted" with the signal (green). If the wavelet matches the signal well, a large number is obtained. If the wavelet matches the signal poorly, a low number is obtained. The transform is computed at different times in the signal, with wavelets of different frequency used on each occasion. The transform value for each position and frequency of the wavelet function are usually plotted on a 2D plane. (*c*) Stretching and squeezing the wavelet during the transform process changes its frequency make up. The height of the wavelet also changes to conserve its signal energy.

2 Forget Fourier



Wavelet transforms are good at picking out features in a signal that occur only intermittently. This advantage can be seen with a synthetic signal (top), which contains a number of isolated features. It has been transformed using a wavelet that is in the form of a "Mexican hat". The four identical wavegroups all have the same morphology in the wavelet-transform plot (middle), while the remaining features each have their own unique appearance. Although the correlation between features in the signal and those in the transform can be seen visually in this example, statistical techniques have to be used for the messier signals that are more likely to be encountered in real applications. In contrast, the conventional Fourier transform of the original signal (below) provides no useful information about the obvious features in the original signal.