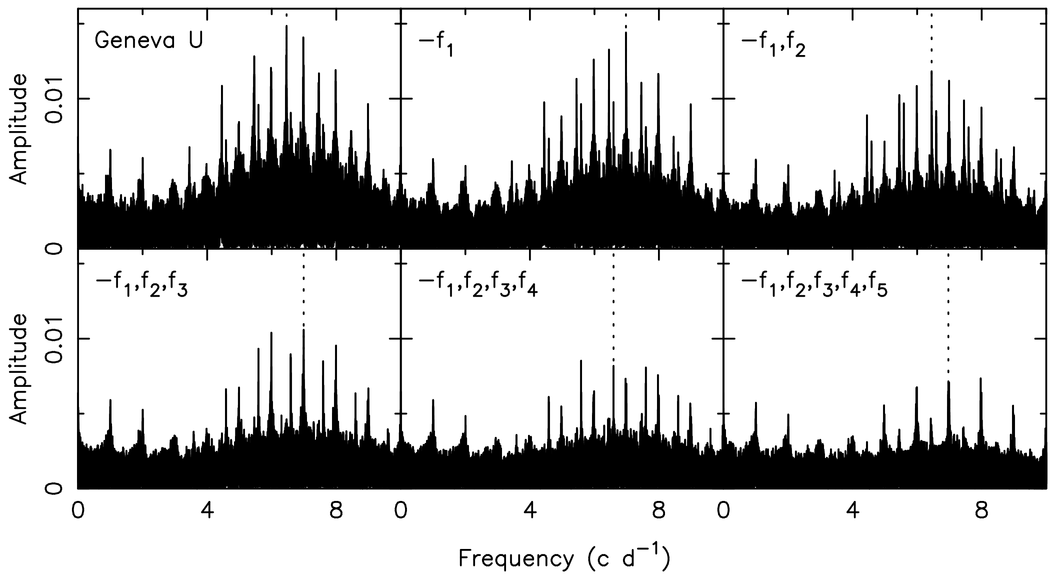


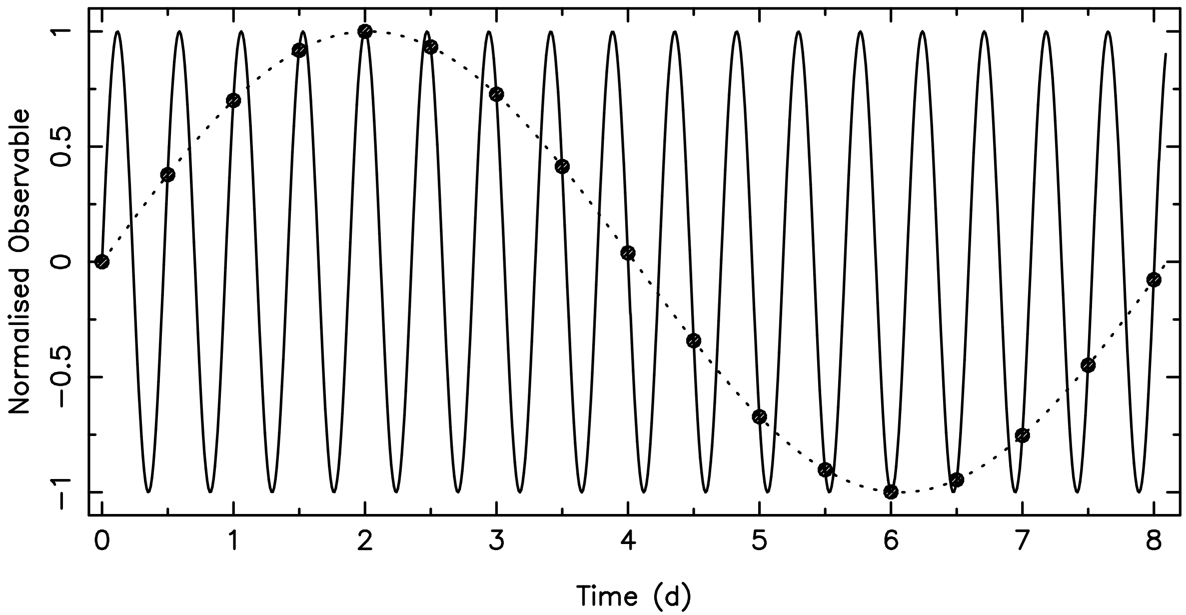
Introduction into

Time Series Analysis

of Unequally-spaced and Gapped

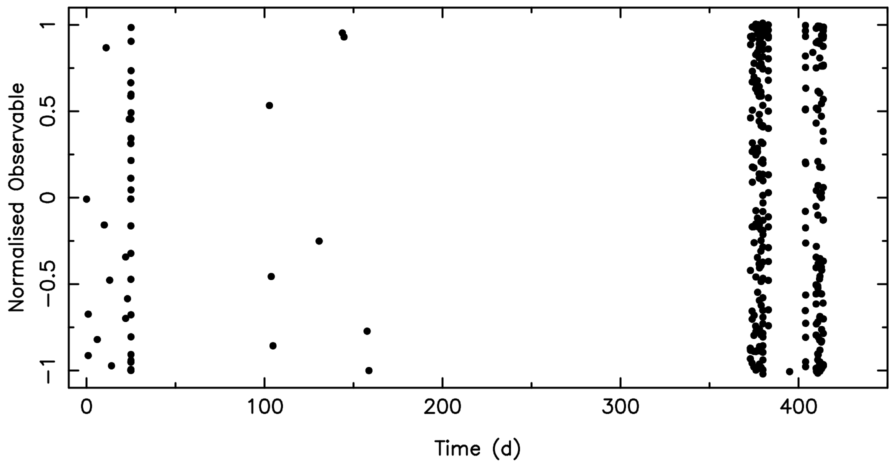
Astronomical Data

1. Some preliminaries



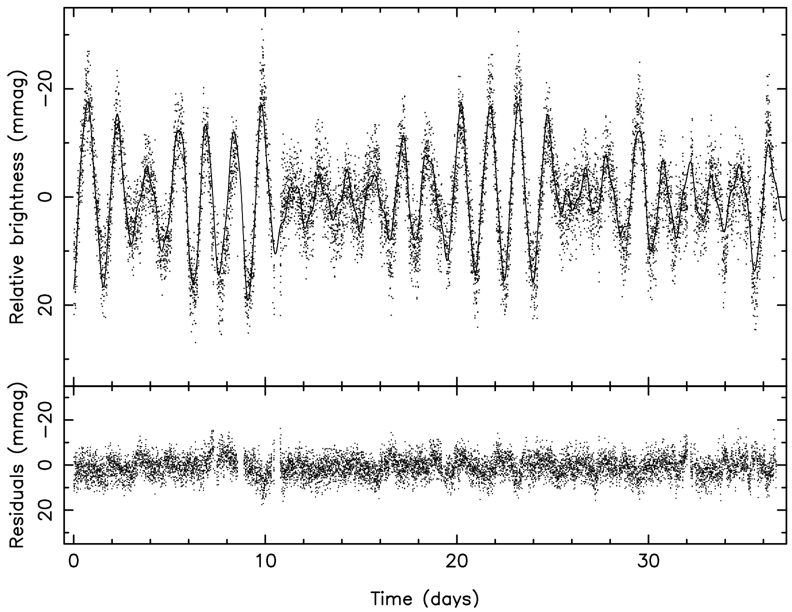
Simulated data (dots) representing a periodic signal with frequency $\nu = 0.123456789\,\mathrm{d}^{-1}$. The dotted line is a harmonic fit for this frequency. The full line represents a fit with the frequency $2.123456789\,\mathrm{d}^{-1}$.

1. Some preliminaries



Simulated gapped data representing a typical time series for a single-site campaign of a pulsating star.

1. Some preliminaries



Real data representing a typical time series for a space campaign of a pulsating star.

2. Methods based on Phase Dispersion Minimisation

Consider quantity x observed at t_i, $x_i(t_i)$ with $i = 1, \ldots, N$.

The *phase* $\varphi(t_i)$ for *cyclic frequency* f or the *period* $P = 1/f$ with respect to the *reference epoch* t_0 is:

$$\varphi(t_i) = [f(t_i - t_0)] = \left[\frac{t_i - t_0}{P}\right],$$

with $0 \leq \varphi < 1$.

A plot of $x_i(t_i)$ as a function of $\varphi(t_i)$ is called a *phase diagram*.

For test period P: divide phase interval $[0, 1]$ in M equal *bins* with bin index $J_i = INT(M\varphi_i) + 1$, $INT(x) \equiv x - [x]$.

Suppose that the $j-$th bin contains N_j measurements.

Per bin:

$$\overline{x_j} = \sum_{i=1}^{N_j} \frac{x_{ij}}{N_j}, \;\; V_j^2 = \sum_{i=1}^{N_j} (x_{ij} - \overline{x_j})^2 = \sum_{i=1}^{N_j} x_{ij}^2 - N_j \overline{x_j}^2,$$

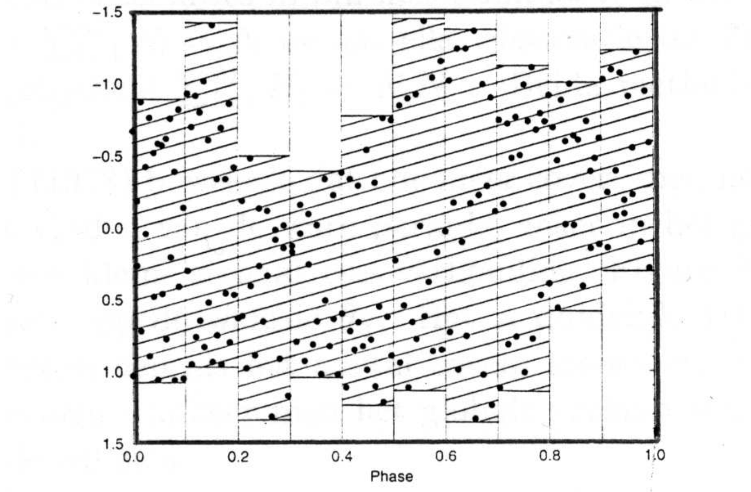
$$s_j^2 = \frac{V_j^2}{N_j - 1},$$

with x_{ij} the observation x_i with bin index $J_i = j$.

For whole dataset, $\overline{x}, V^2$ and s^2, are defined as

$$\overline{x} = \sum_{i=1}^{N} \frac{x_i}{N}, \;\; V^2 = \sum_{i=1}^{N} (x_i - \overline{x})^2 = \sum_{i=1}^{N} x_i^2 - N\overline{x}^2,$$

$$s^2 = \frac{V^2}{N - 1}.$$



For the M bins:

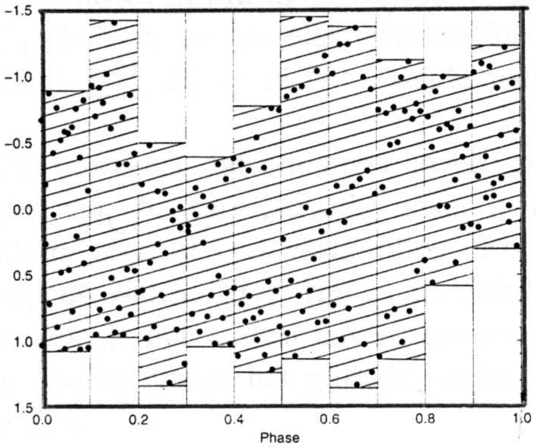
$$V_M^2 = \sum_{j=1}^{M} V_j^2, \ V_G^2 = \sum_{j=1}^{M} N_j \left(\overline{x_j} - \overline{x}\right)^2 .$$

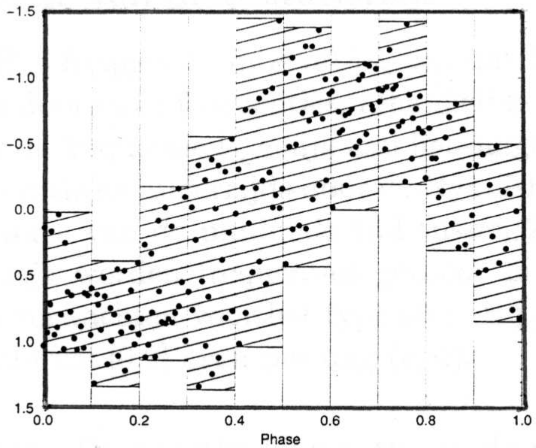
Thus $V^2 = V_M^2 + V_G^2$.

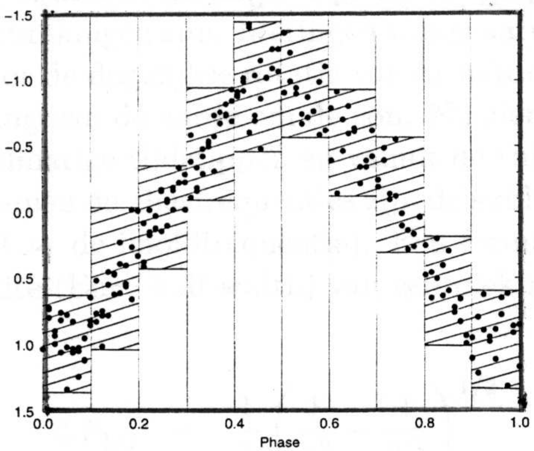
$\overline{x_j} \neq \overline{x}$ for good period: $V_G^2 \simeq V^2$

$\overline{x_j} \simeq \overline{x}$ for bad period: $V_G^2 \ll V^2$

search maximum of $V_G^2 \Leftrightarrow$ search minimum of V_M^2


-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
Phase


-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
Phase


-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
Phase

Better: bin-cover structure (N_b, N_c): divide phase diagram in N_b bins of length $1/N_b$ and apply this partition N_c times, with each partition shifted over $1/N_b N_c$

Define

$$\Theta \equiv \frac{\sum_{j=1}^{M} (N_j - 1)\, s_j^2 / \sum_{j=1}^{M} N_j - M}{\sum_{i=1}^{N} (x_i - \overline{x})^2 / N - 1} \in [0, 1]$$

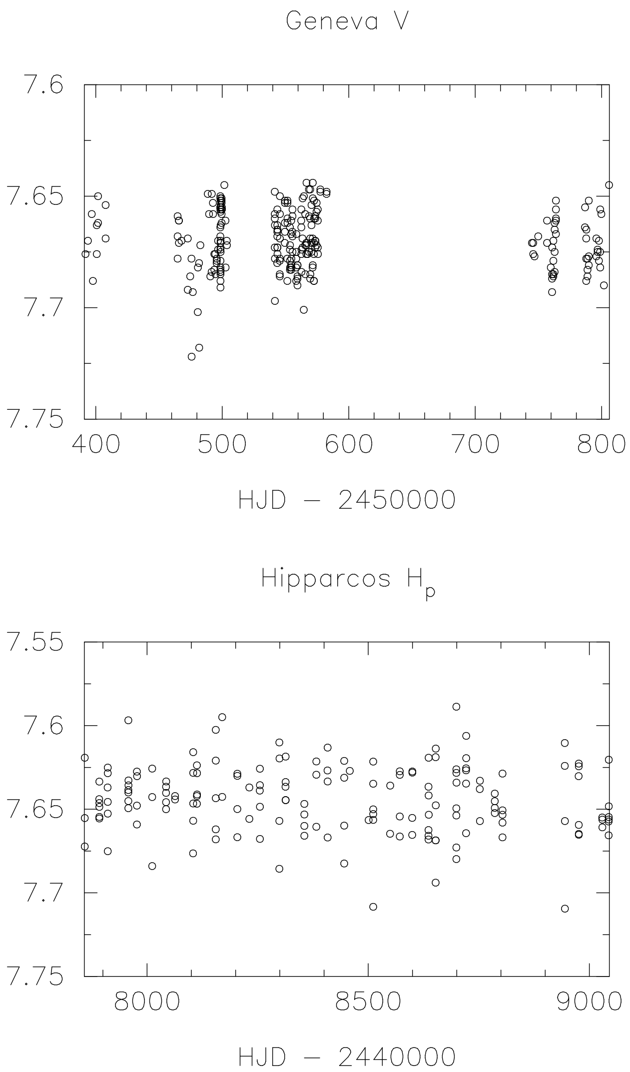
where s_j^2 is defined as:

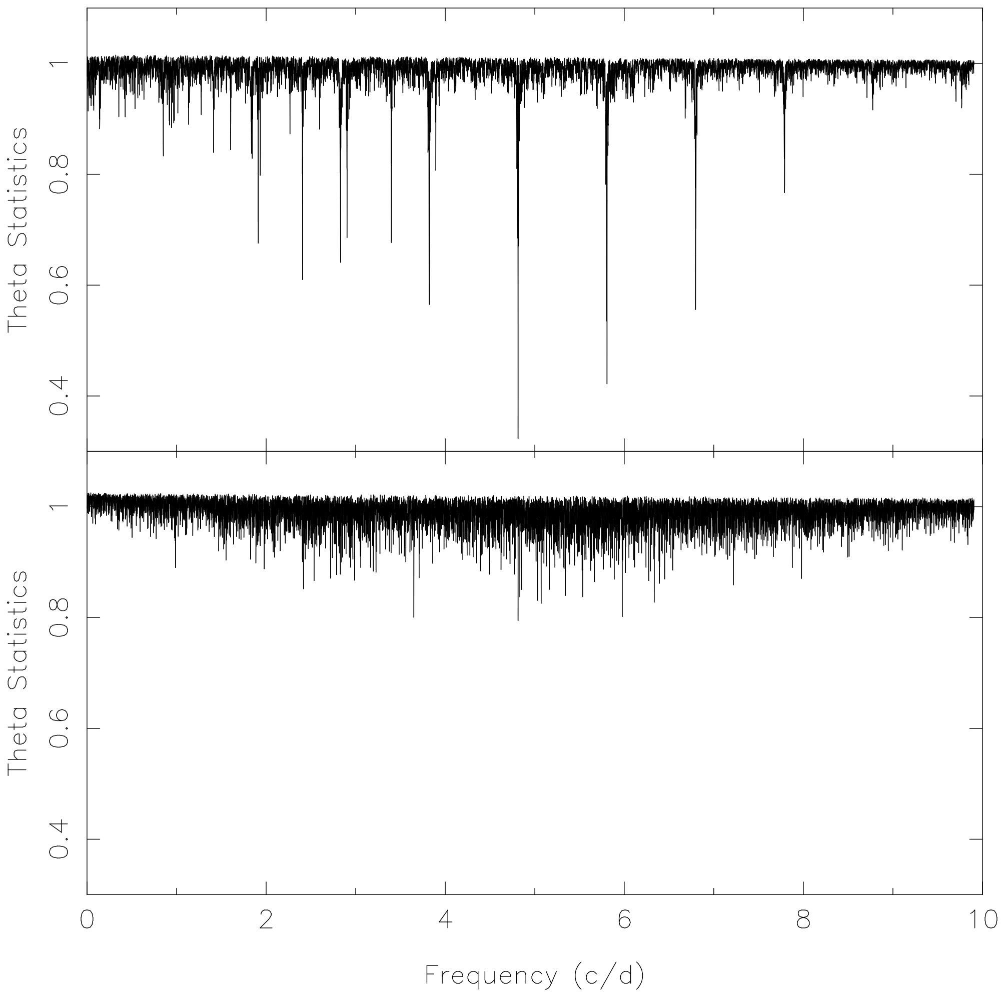
$$s_j^2 \equiv \frac{\sum_{i=1}^{N_j} (x_{ij} - \overline{x_j})^2}{N_j - 1}.$$

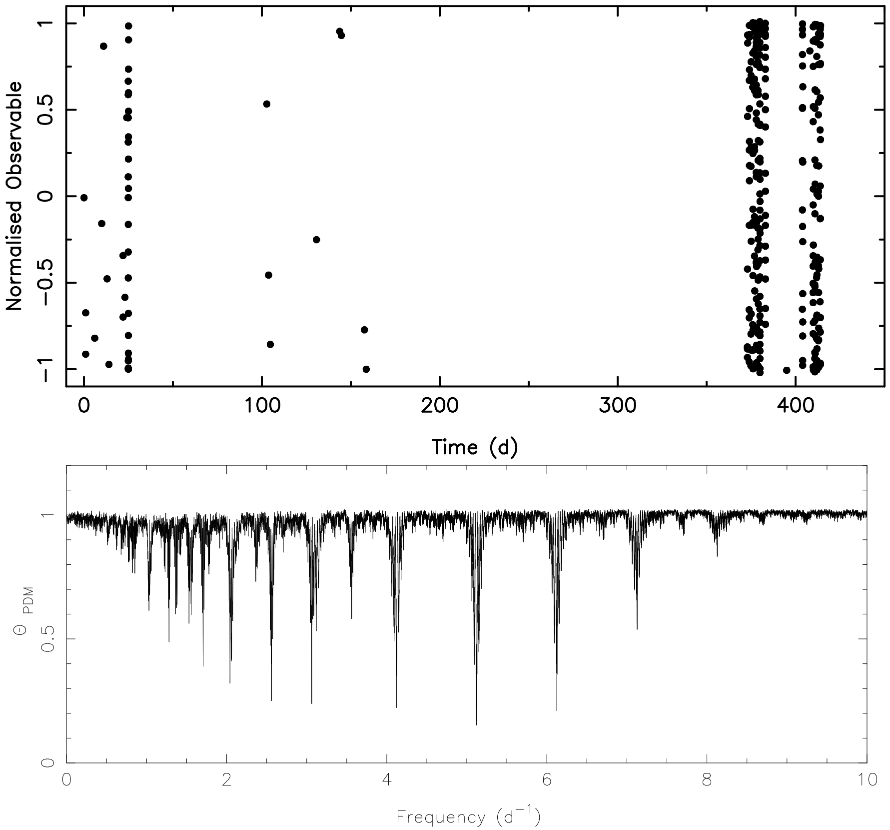
With the introduced notation:

$$\Theta = \frac{V_M^2 / \left(\sum_{j=1}^{M} N_j - M\right)}{V^2 / (N - 1)} = \frac{V_M^2 / N_c (N - N_b)}{V^2 / (N - 1)}.$$

A minimum in the $\Theta-$statistic corresponds to a minimum of V_M

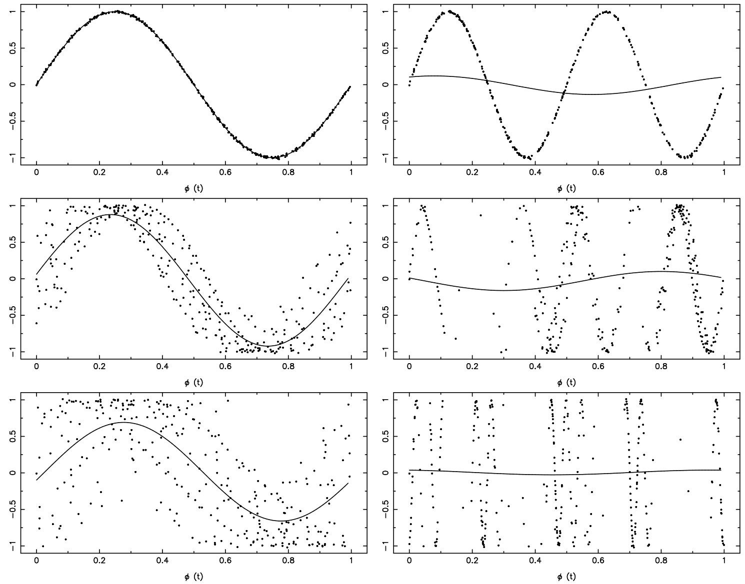

Geneva V
7.6
7.65
7.7
7.75
400
500
600
700
800
HJD − 2450000
Hipparcos H_p
7.55
7.6
7.65
7.7
7.75
8000
8500
9000
HJD − 2440000


Theta Statistics
0.4
0.6
0.8
1
Theta Statistics
0.4
0.6
0.8
1
0
2
4
6
8
10
Frequency (c/d)



Statistic Θ_{PDM} of the data using 10 bins and 2 covers.

PDM and any other string-length method are sensitive to harmonics and subharmonics . . . Be careful !



Phase diagrams for six minima in the $\Theta_{\rm PDM}$: 5.123 d^{-1} (upper left), 4.121 d^{-1} (middle left), 7.129 d^{-1} (lower left), 2.562 d^{-1} (upper right), 1.021 d^{-1} (middle right), 0.244 d^{-1} (lower right).

3. Methods based on Fourier Analysis

Fourier transform of $x(t)$:

$$F(f) \equiv \int_{-\infty}^{+\infty} x(t) \exp(2\pi \mathrm{i}\, ft) dt$$

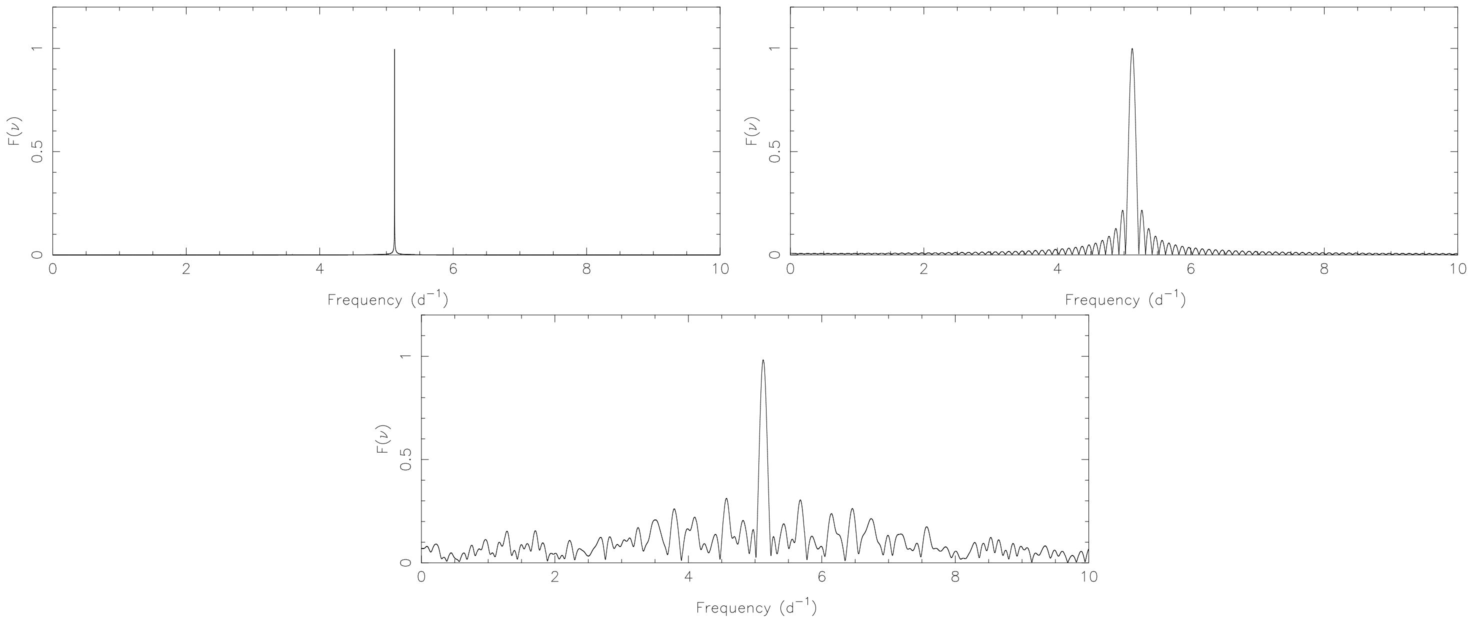
Fourier transform $F(f)$ of sum of harmonic functions with frequencies $f_1, \ldots, f_n$ and amplitudes $A_1, \ldots, A_n$:

$$x(t) = \sum_{k=1}^{n} A_k \exp(2\pi \mathrm{i}\, f_k t) : \quad F(f) = \sum_{k=1}^{n} A_k \delta(f - f_k)$$

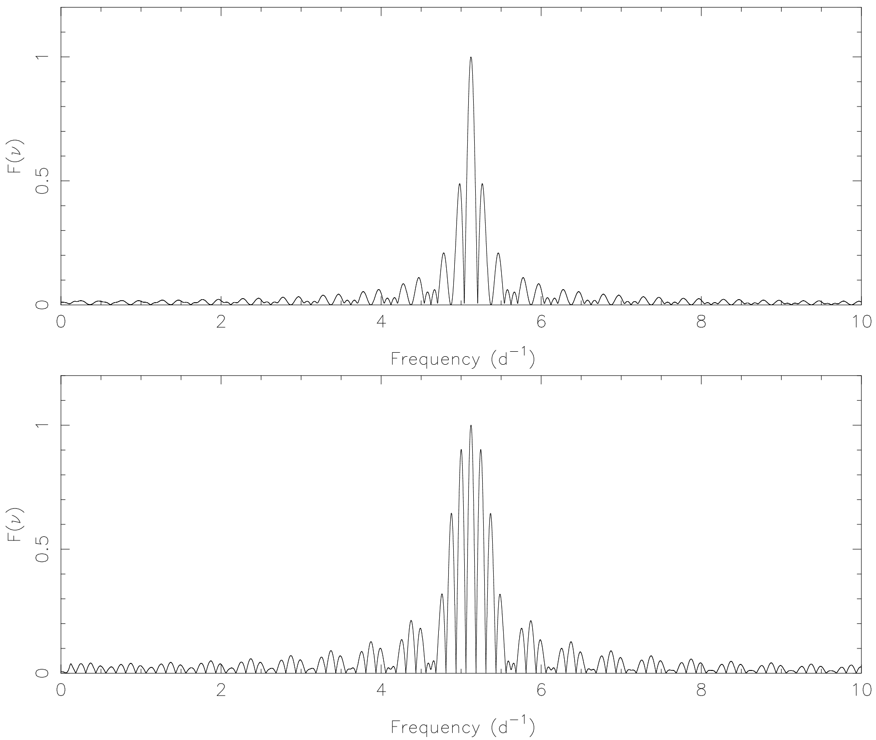
For $x(t) =$ sine with frequency f_1, $F(f) \neq 0$ for $f = \pm f_1$

For $x(t) =$ sum of n harmonic functions with frequencies $f_1, \ldots, f_n$, $F(f) =$ sum of $\delta-$functions $\neq 0$ for $\pm f_1, \ldots, \pm f_n$

Real data set: $x(t)$ known for a discrete number of times $t_j, j = 1, \ldots, N$



Left: 10^6 points over 1000 d, right: 10^4 points over 10 d, bottom: 4472 points over 10 d.



Fourier transforms of a noiseles time series of a sine function with frequency 5.123456789 d^{-1} generated for a finite time span of 10 days and containing one large gap from day 4 until day 6 (top) and from day 2 until day 8 (bottom).

Discrete Fourier transform:

$$F_N(f) \equiv \sum_{j=1}^{N} x(t_j) \exp(2\pi \mathrm{i}\, f t_j)$$

$F_N \neq F$! but connected through window function:

$$w_N(t) \equiv \frac{1}{N} \sum_{j=1}^{N} \delta(t - t_j)$$

Hence:

$$\frac{F_N}{N} = \int_{-\infty}^{+\infty} x(t) w_N(t) \exp(2\pi \mathrm{i}\, f t) dt$$

Discrete Fourier transform of window function = spectral window $W_N(f)$:

$$W_N(f) = \frac{1}{N} \sum_{j=1}^{N} \exp(2\pi \mathrm{i}\, f t_j)$$

Discrete Fourier transform = convolution of spectral window and Fourier transform: $F_N(f)/N = F(f) * W_N(f)$.

If $F(f) = \delta$ at f_1, $F_N(f)/N$ same behaviour as $W_N(f)$ at f_1:

$$F_N(f)/N = W_N(f) * \delta(f - f_1) = W_N(f - f_1)$$

$$\Downarrow$$

Comparison of $W_N(f)$ with $F_N(f)/N$ near $f_1 \rightarrow$ is f_1 is real or not?

If $F(f) =$ sum of n $\delta-$functions:

$$\frac{F_N(f)}{N} = W_N(f) * \sum_{k=1}^{n} \delta(f - f_k) = \sum_{k=1}^{n} W_N(f) * \delta(f - f_k)$$
$$= \sum_{k=1}^{n} W_N(f - f_k) = \frac{1}{N} \sum_{k=1}^{n} \sum_{j=1}^{N} \exp(2\pi \mathrm{i}\,(f - f_k)t_j).$$

Hence, $F_N(f)/N =$ sum of n spectral windows centered around f_k

$W_N(f) \neq 0$ at all f (not only at $f_k, k = 1, \ldots, n$) due to interference

$$\Downarrow$$

maxima in periodogram at spurious frequencies due to observing times. . .
ALIAS FREQUENCIES

Most common alias frequencies? Assume $t_j = t_0 + j\triangle t$:

$$W_N(f) = \frac{1}{N}\sum_{j=1}^{N} \exp(2\pi \mathrm{i}\, f t_0) \exp(2\pi \mathrm{i} f j \triangle t)$$
$$= \frac{1}{N} \exp(2\pi \mathrm{i}\, f t_0) \sum_{j=1}^{N} \exp(2\pi \mathrm{i}\, f j \triangle t)$$
$$= \exp(2\pi \mathrm{i}\, f t_0) \exp(\pi \mathrm{i}\, f \triangle t (N+1)) \frac{\sin(\pi f N \triangle t)}{N \sin(\pi f \triangle t)}$$

in which we have made use of

$$\sum_{j=0}^{N-1} z^j = \frac{1 - z^N}{1 - z}$$

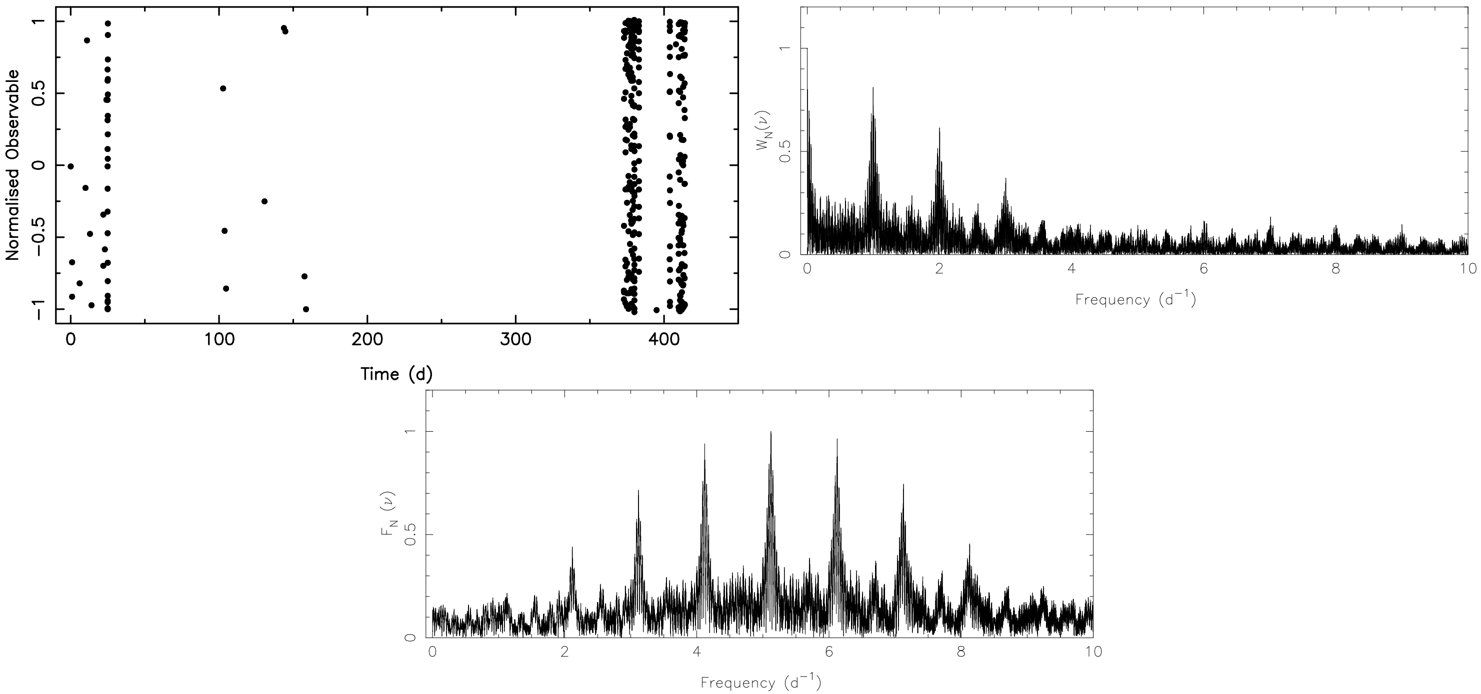
with $z = \exp(2\pi \mathrm{i}\, f \triangle t)$.

For $t_0 = -(N+1)\triangle t/2$: $W_N(f) = \frac{\sin(\pi N f \triangle t)}{N \sin(\pi f \triangle t)}$:
periodic function with period $1/\triangle t$:

$$\left| W_N \left(f + \frac{n}{\triangle t} \right) \right| = |W_N(f)| \, .$$

$F_N(f)$ has maxima at $f_n = n/\triangle t$

For non-equidistant data: $\triangle t \simeq$1 day, 1 year, specific gaps, length of nightly strings, etc.



Spectal window (upper right) and DFT (bottom) of a noise-free sinusoid with amplitude 1 at 5.123456789 d^{-1} for the sampling shown upper left.

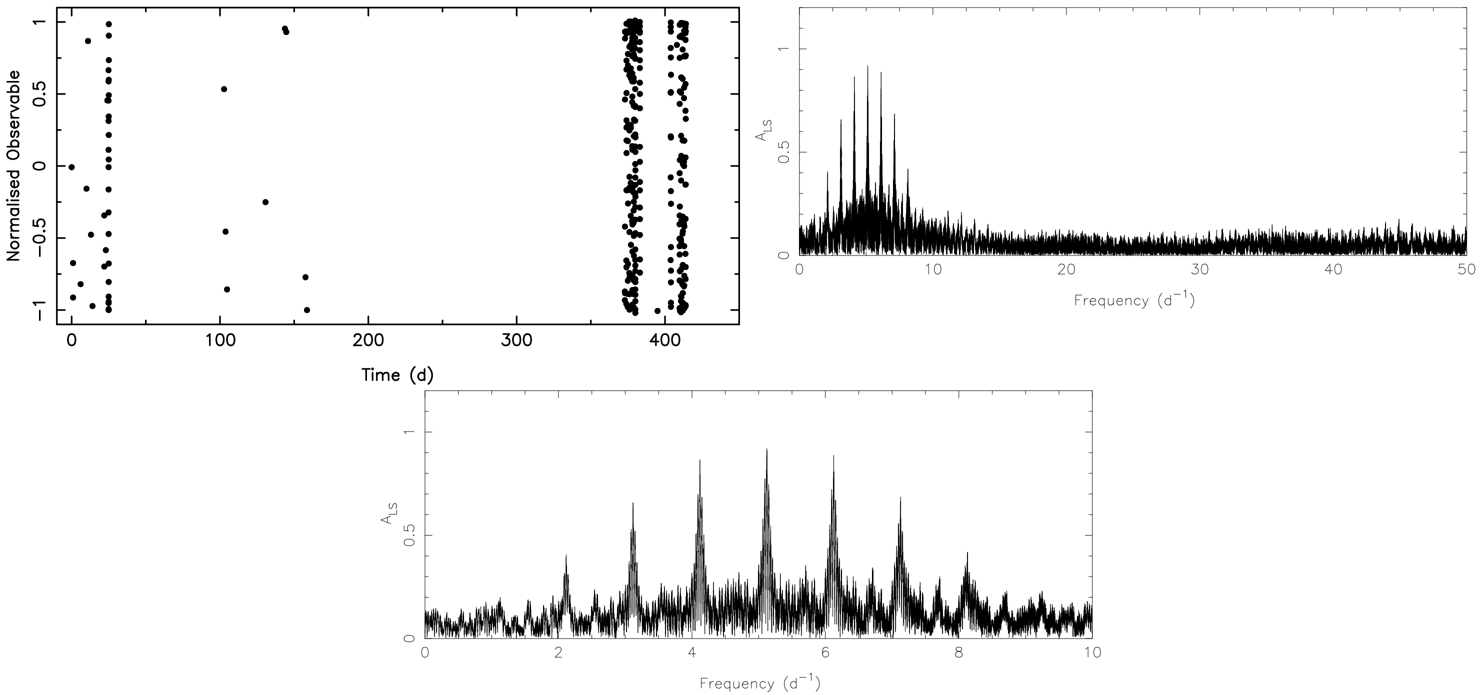
The Lomb-Scargle periodogram:

$$P_N(f) = \frac{1}{2}\left\{ \frac{\left\{\sum_{i=1}^{N} x(t_i)\cos[2\pi f(t_i-\tau)]\right\}^2}{\sum_{i=1}^{N}\cos^2[2\pi f(t_i-\tau)]} + \frac{\left\{\sum_{i=1}^{N} x(t_i)\sin[2\pi f(t_i-\tau)]\right\}^2}{\sum_{i=1}^{N}\sin^2[2\pi f(t_i-\tau)]} \right\}$$

with τ chosen such that

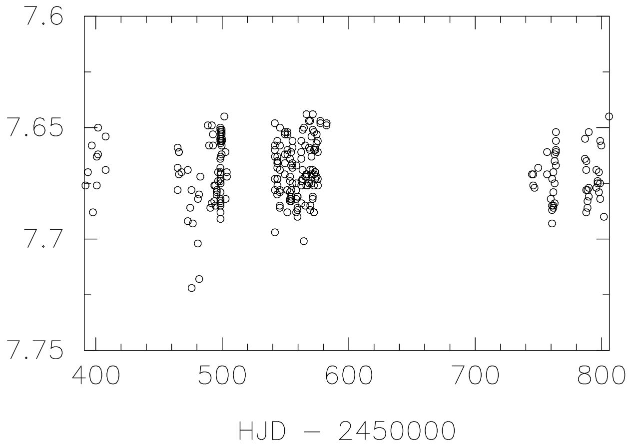
$$\sum_{i=1}^{N}\cos[2\pi f(t_i-\tau)]\sin[2\pi f(t_i-\tau)] = 0.$$

Note: $\lim_{N\to\infty} P_N(f) = A^2 N/4$

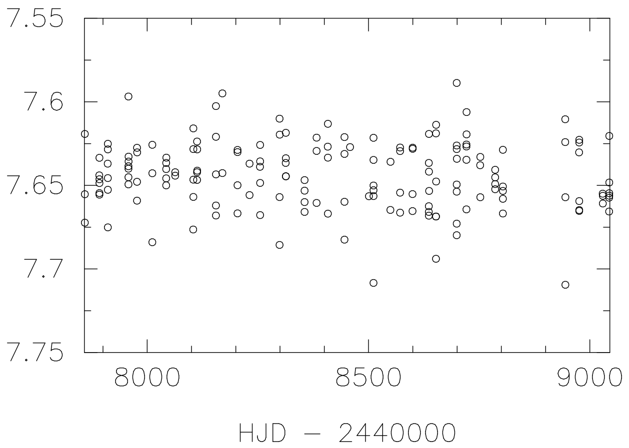


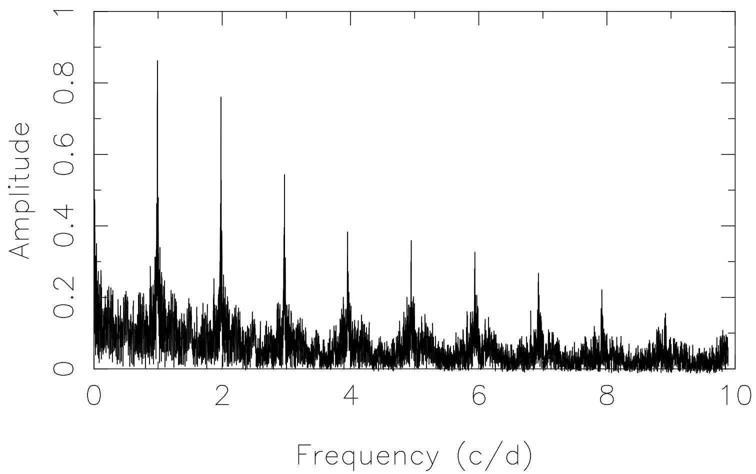
Lomb-Scargle periodograms for the data.

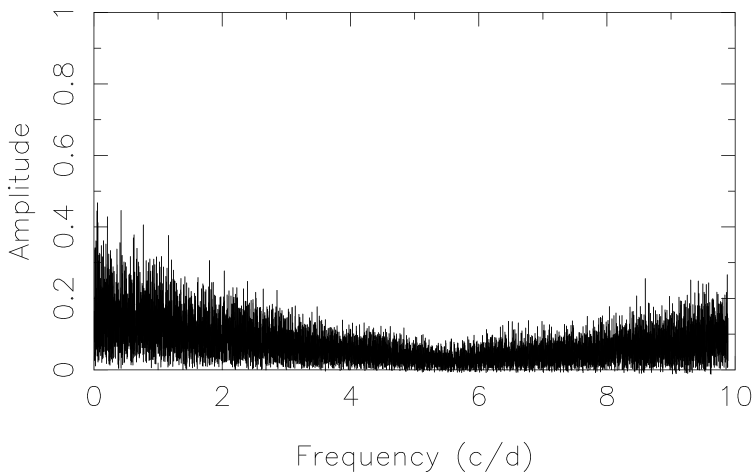
Geneva V

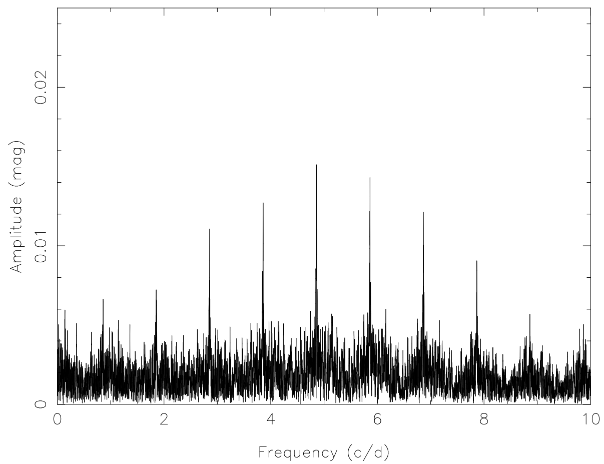


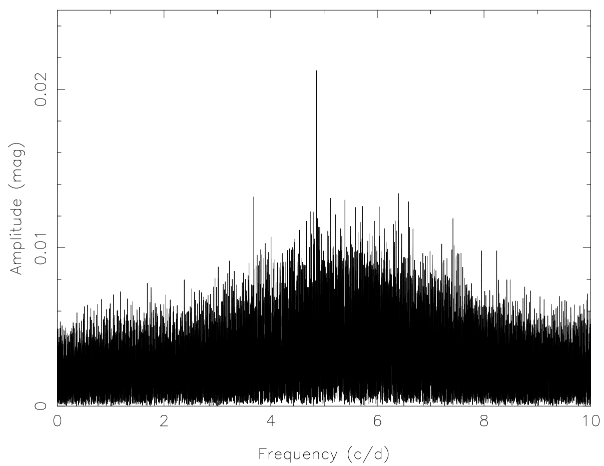
Hipparcos H_p




Amplitude
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10
Frequency (c/d)


Amplitude
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10
Frequency (c/d)


Amplitude (mag)
0
0.01
0.02
0
2
4
6
8
10
Frequency (c/d)


Amplitude (mag)
0
0.01
0.02
0
2
4
6
8
10
Frequency (c/d)

- Error estimate for derived frequency?

  Assuming white noise and no alias problems:

$$\sigma_f = \frac{\sqrt{6}\sigma_R}{\pi\sqrt{N}AT}$$

  with σ_R average error of data ($\sim$ standard deviation of noise), T total time span, N number of data points

- Interval of testfrequencies?

  use interval $[0, Nyq]$ with Nyq the Nyquist frequency.

  For equidistant data: $1/2\triangle T$ with $\triangle T$ the time step.

  For unevenly spaced data: take median of all time intervals between two consecutive measurements.

- Acceptance criterion for a frequency: $A > 4\sigma_{\mathrm{Fourier}}$. Compute $\sigma_{\mathrm{Fourier}}$ in oversampled interval of testfrequencies around candidate frequency.

  Criterion: 99.9% confidence interval that peak is not due to noise.

How good is the solution ?

LS fitting with f fixed:

$$x_i(t_i) = A \sin\left[2\pi(ft_i + \psi)\right] + C$$

Variance reduction $\in [0, 1]$:

$$1 - \frac{\sum_{i=1}^{N} \left\{x_i - \left[A \sin\left(2\pi(ft_i + \psi)\right) + C\right]\right\}^2}{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}$$

Search for new frequencies in residuals $R_i(f) \equiv x_i - x_i^c(f)$ with:

$$x_i^c(f) \equiv A \sin\left[2\pi(ft_i + \psi)\right] + C$$

and so on, until amplitude $A < 4\sigma_{\text{Fourier}}$

At some point, you will reach the noise level . . .