DFT vs Lomb-Scargle

A critical comparison of the Lomb-Scargle and the classical periodograms

arXiv:1807.01595

Roberto Vio,[1*] P. Andreani,[2]
[1]*Chip Computers Consulting s.r.l., Viale Don L. Sturzo 82, S.Liberale di Marcon, 30020 Venice, Italy*
[2]*ESO, Karl Schwarzschild strasse 2, 85748 Garching, Germany*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
The detection of signals hidden in noise is one of the oldest and common problems in astronomy. Various solutions have been proposed in the past such as the parametric approaches based on the least-squares fit of theoretical templates or the non-parametric techniques as the phase-folding method. Most of them, however, are suited only for signals with specific time evolution. For generic signals the spectral approach based on the periodogram is potentially the most effective. In astronomy the main problem in working with the periodogram is that often the sampling of the signals is irregular. This complicates its efficient computation (the fast Fourier transform cannot be directly used) but overall the determination of its statistical characteristics. The Lomb-Scargle periodogram (LSP) provides a solution to this last important issue, but its main drawback is the assumption of a very specific model of the data which is not correct for most of the practical applications. These issues are not always considered in literature with theoretical and practical consequences of no easy solution. Moreover, apart from pathological samplings, it is common believe that the LSP and the classical periodogram (CP) usually provide almost identical results. In general, this is true but here it is shown that there are situations where the LSP is less effective than the CP in the detection of signals in noise. There are no compelling reasons, therefore, to use the LSP instead of the CP which is directly connected to the correlation function of the observed signal with the sinusoidal functions at the various frequencies of interest.

Key words: Methods: Statistical – Methods: Data Analysis – Methods: Numerical

DFT vs Lomb-Scargle

A critical comparison of the Lomb-Scargle and the classical periodograms by Vio & Andreani

arXiv:1807.01595

"… apart from pathological samplings, it is common believe that the LSP and the classical periodogram (CP) usually provide almost identical results. **In general, this is true** but here it is shown that there are situations where the LSP is less effective than the CP in the detection of signals in noise. There are no compelling reasons, therefore, to use the LSP instead of the CP …"

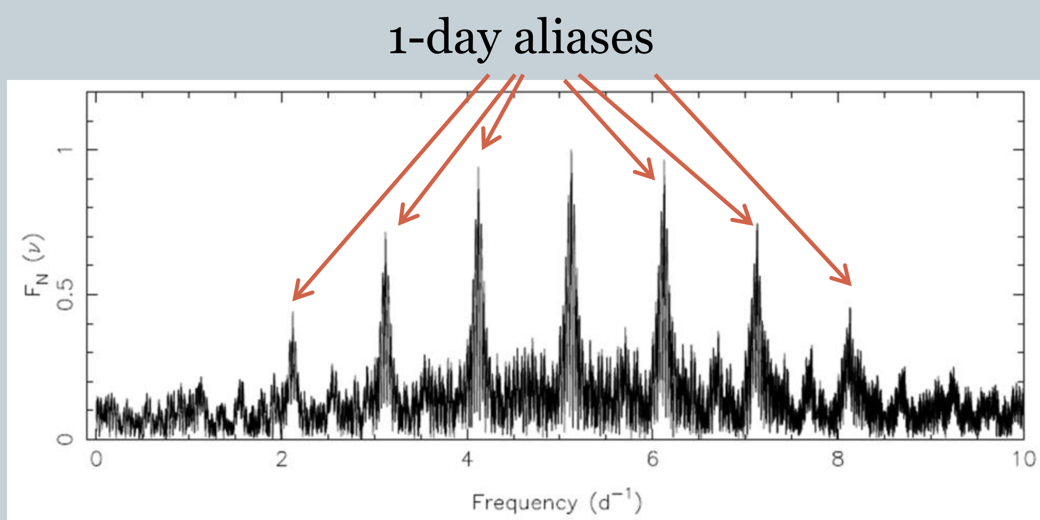
However (it is my remark):

There are situations where the LSP is more effective than the CP.
Thus, there are no compelling reasons to use the DFT instead of the LSP.

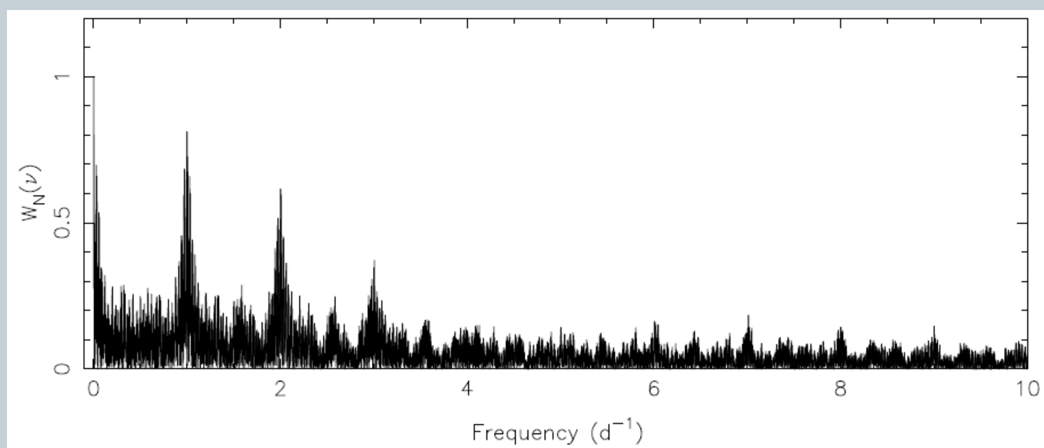
Conclusion: both methods provide almost identical results.
However, the LSP is statistically robust.

Dealing with Aliases. CLEAN algorithm.

Power Spectrum:



Spectral Window:

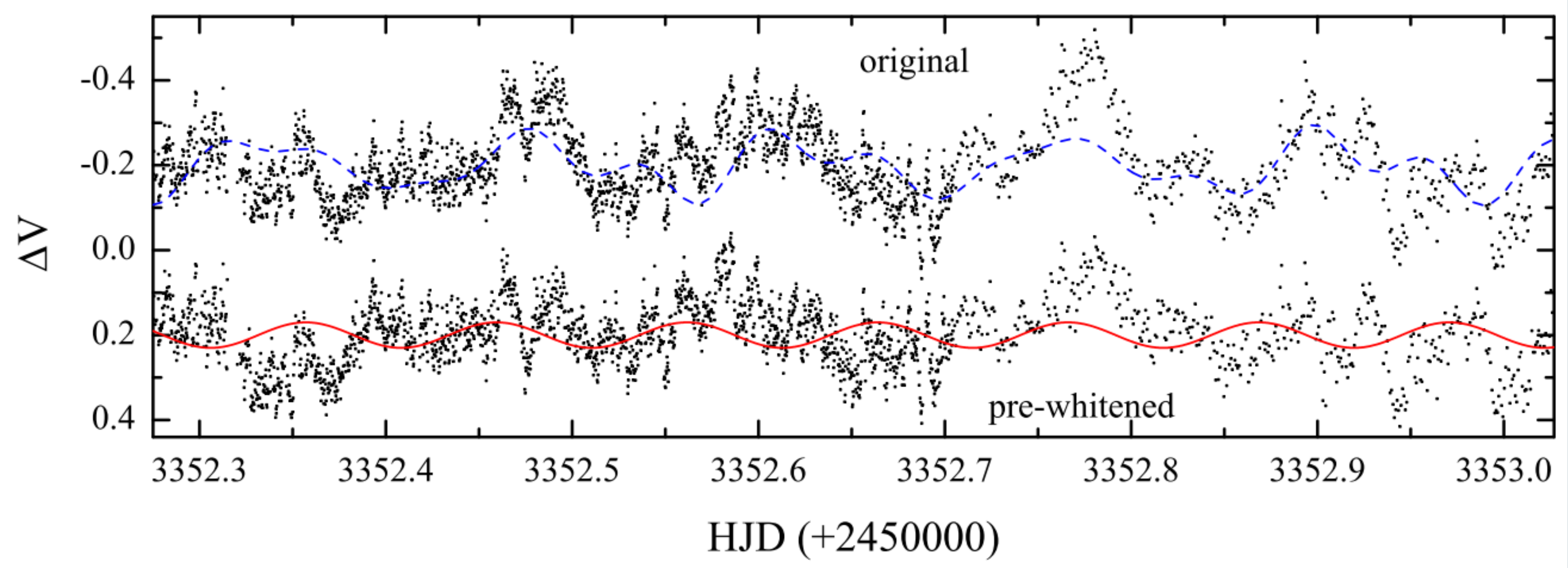


Dealing with Aliases

- How to deal with aliases?
  - Pre-whitening
  - Cleaning (Clean algorithm)

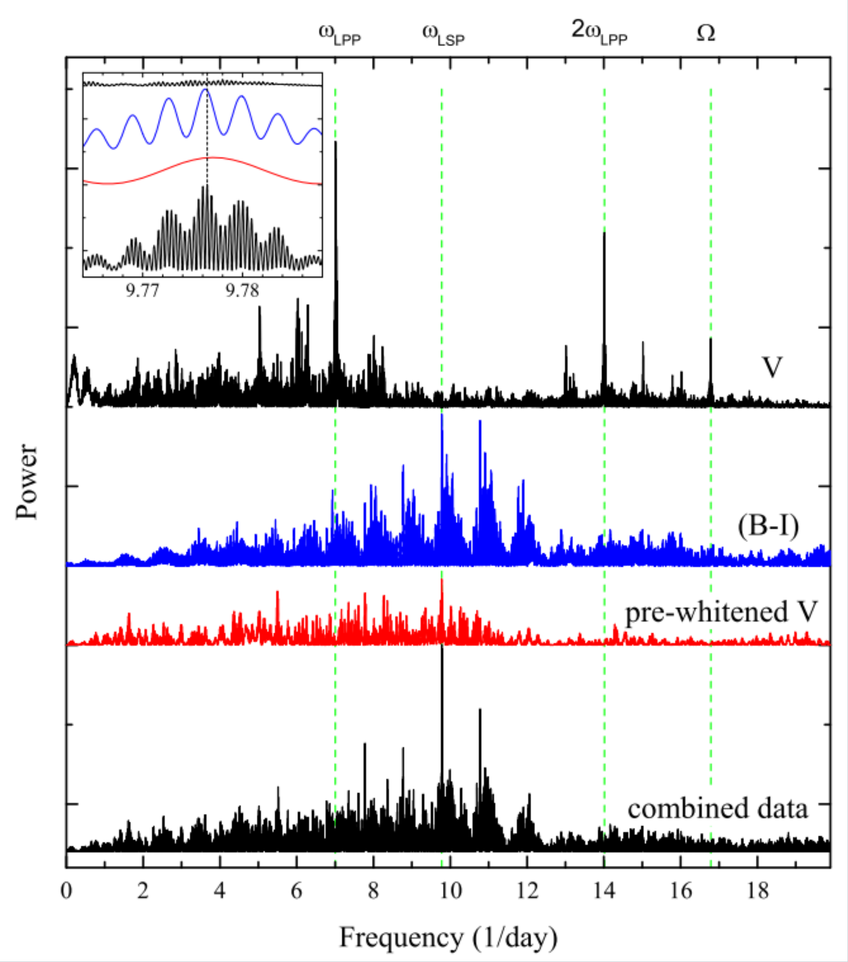
Pre-whitening

- If the light curve contains more than one periodic modulations, and the signal of interest with an unknown frequency is weak and hidden in noise, then one can try to remove the strongest signal of known frequency from the light curve:
  fit the light curve with a sine-wave (and its harmonics) and subtract it.



Pre-whitening

- If the light curve contains more than one periodic modulations, and the signal of interest with an unknown frequency is weak and hidden in noise, then one can try to remove the strongest signal of known frequency from the light curve.



Cleaning

THE ASTRONOMICAL JOURNAL VOLUME 93, NUMBER 4 APRIL 1987

TIME SERIES ANALYSIS WITH CLEAN. I. DERIVATION OF A SPECTRUM

DAVID H. ROBERTS
Physics Department, Brandeis University, Waltham, Massachusetts 02254

JOSEPH LEHÁR AND JOHN W. DREHER
Physics Department, 26-315, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
Received 7 October 1986; revised 25 November 1986

ABSTRACT

We present a method of time-series spectral analysis which is especially useful for unequally spaced data. Based on a complex, one-dimensional version of the CLEAN deconvolution algorithm widely used in two-dimensional image reconstruction, this technique provides a simple way to understand and remove the artifacts introduced by missing data. We describe the method, give several examples, and point out various analogies with the conventional use of CLEAN.

Clean Algorithm

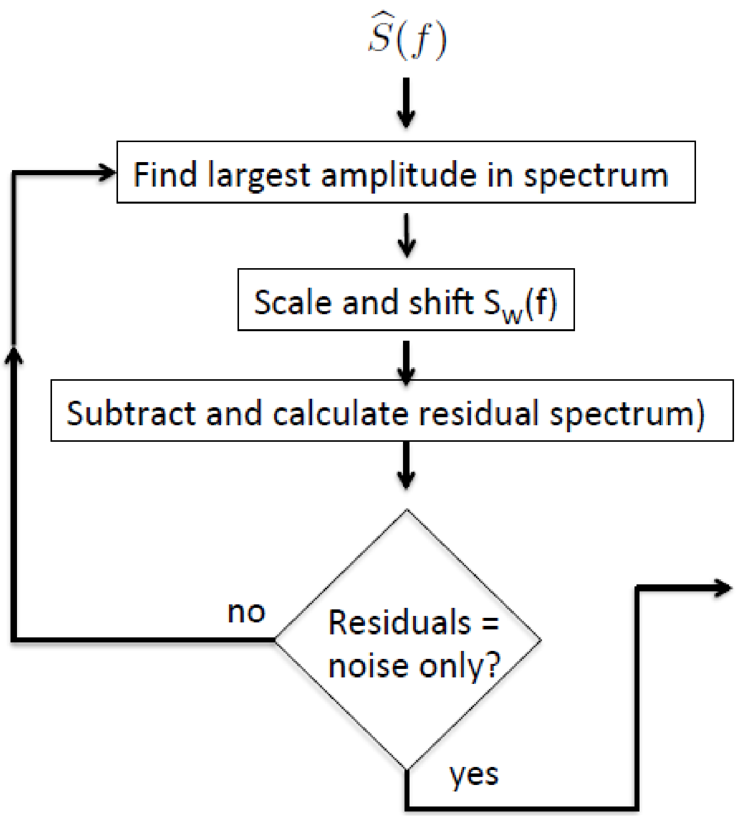
- The premise of CLEAN is that our data consist not only of the data amplitudes but also the detailed sampling in time.
- We therefore know that the true spectrum is convolved with a known window function.
- The actual algorithm is based on the fact that any function can be represented as a sum or integral over delta functions.

Spectral Analysis
$S(f)$ = spectral estimate
$S_W(f) = \lvert FT\{W(t)\}\rvert^2$ = spectral window (calculated as the FT of sample times)
$S_C(f) = \Sigma\, C_j\, \Delta(f - f_j)$ CLEANed Spectrum Sum over CLEAN components $\Delta(f)$ = restoration function that represents the inherent frequency resolution

The restoring function is needed to fairly represent the resolution imposed by Fourier transform properties (uncertainty principle)

Clean Algorithm

Data yield $\widehat{S}(f)$ and $S_W(f) = |\tilde{W}(f)|^2$



$\widehat{S}_i(f_i)_{\max}$ @ $f_i \quad i = \text{iteration count}$

$\gamma_i S_W(f - f_i)$

$\widehat{S}_i(f) \rightarrow \widehat{S}_{i+1}(f) = \widehat{S}_i(f) - \gamma_i S_W(f - f_i)$

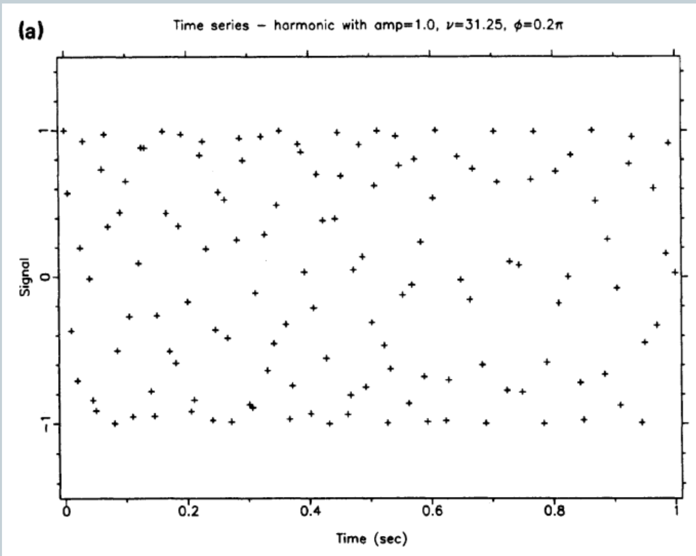
Collect CLEAN components

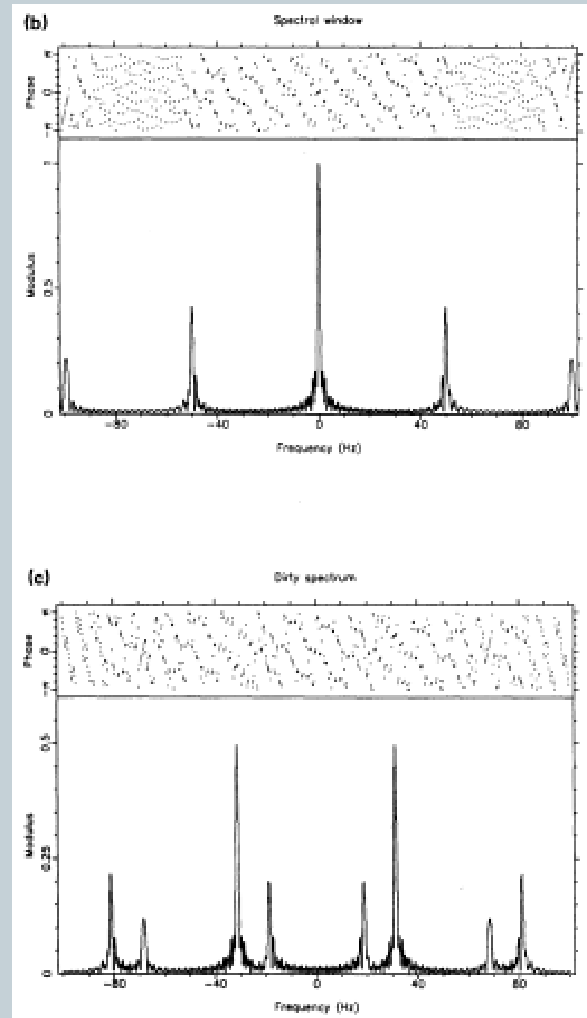
$\{\gamma_i, f_i, \; i = 1, N_c\}$

Form CLEAN spectrum

$$S_c(f) = \sum_{i=1}^{N_c} \gamma_i \Delta(f - f_i)$$

Clean Algorithm





a) Time-Series
b) The window function
c) The dirty spectrum

Clean Algorithm

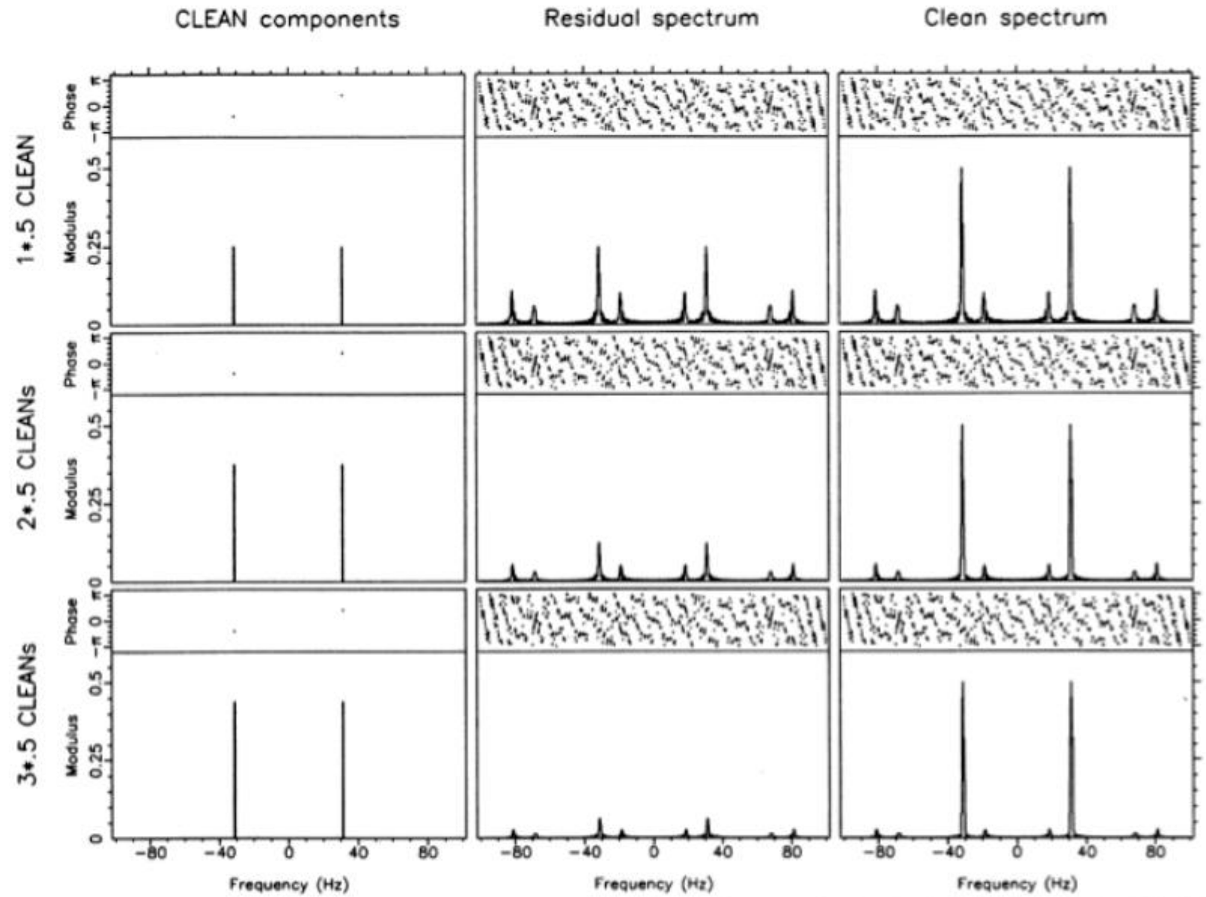


FIG. 2. Analysis of the time series in Fig. 1. (a)–(e). The clean components, residual spectra, and clean spectra after one, two, three, five and one hundred iterations with gain $g = 0.5$, and (f) after one iteration with $g = 1$. Note the change to a logarithmic scale for (d)–(f).

Clean Algorithm

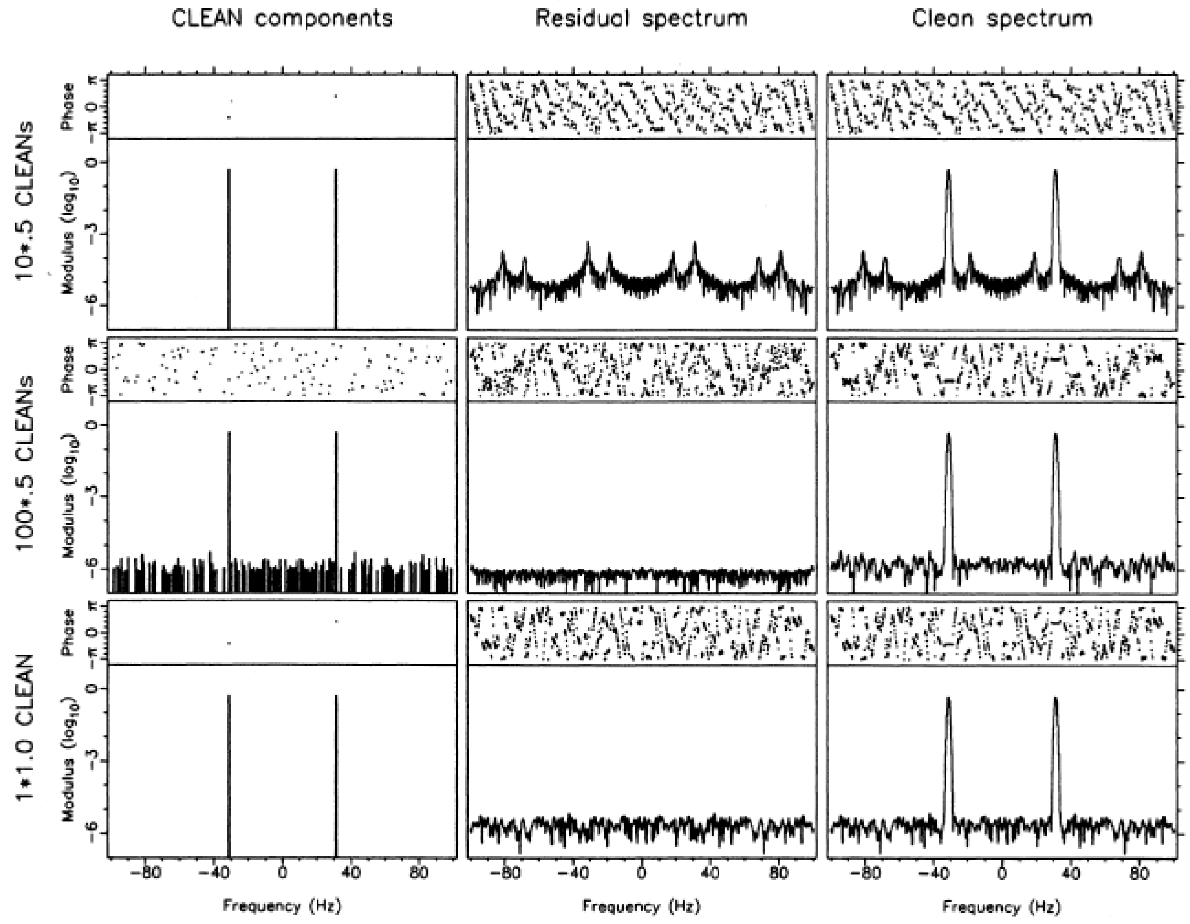


FIG. 2. (continued)

Non-parametric Frequency Analysis Methods

STRING LENGTH METHODS
PHASE DISPERSION MINIMIZATION
OTHER NON-PARAMETRIC METHODS

Non-parametric Methods

- Non-parametric methods imply that one does not *a priori* assume a chosen model function to describe the data.

- This is in contrast to any method based on Fourier transforms, where harmonic model functions are assumed from the start.

String Length Methods

- The method was originally introduced by Lafler & Kinman (1965) → Lafler - Kinman method.
- Clarke (2002) presented a clear recent evaluation of these methods and proposed their generalization to the application for multivariate data, the so-called Rope Length Method.
- This methodology is very suitable to analyse time series of multicolour photometric observations or of radial velocity variations from different spectral lines.

String Length Methods

- For each trial frequency ν, taken from a grid of test frequencies, the original data $x(t_i)$ are first assigned phases $\varphi(t_i)$, which are then ordered in ascending value $0 \le \varphi_1, \ldots, \varphi_N < 1$.
- For each trial frequency, the original Lafler-Kinman statistic performs a "string length" summation of the squares of the differences between the consecutive phase-ordered values.

String Length Methods

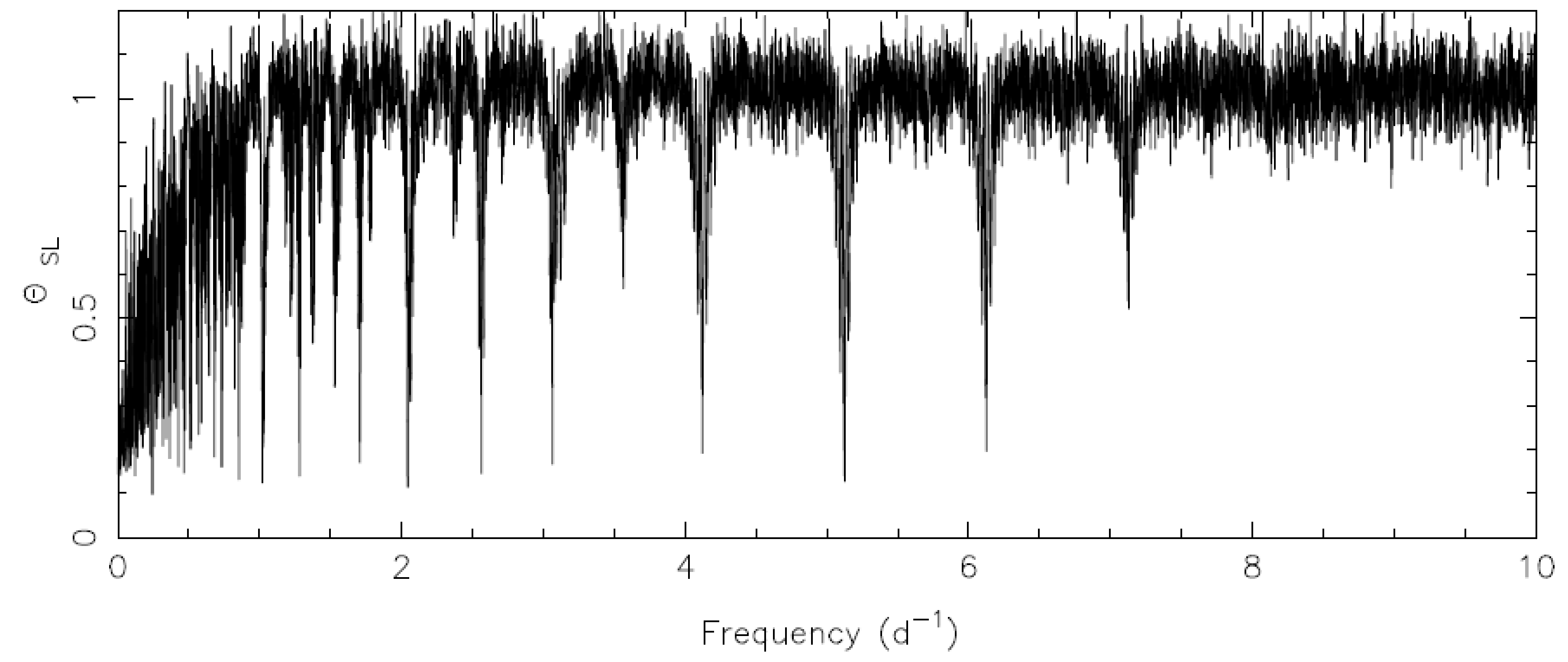
- Clarke (2002) advises the use of the following modified string length statistic:

$$\Theta_{\mathrm{SL}}(\nu) \equiv \frac{\sum_{i=1}^{N} \left[x(\phi_{i+1}) - x(\phi_i)\right]^2}{\sum_{i=1}^{N} \left[x(\phi_i) - \overline{x}\right]^2} \times \frac{N-1}{2N},$$

- where $\bar{x}$ is the mean value of the measurements and $x(\varphi_N+1)$ is taken to be equal to $x(\varphi_1)$.
- If the time series contains periodicity with frequency ν, then Θ_{SL} will reach a minimum at ν while fluctuations in Θ_{SL} due to the noise will result in a level $\Theta_{\mathrm{SL}} \approx 1.0$.

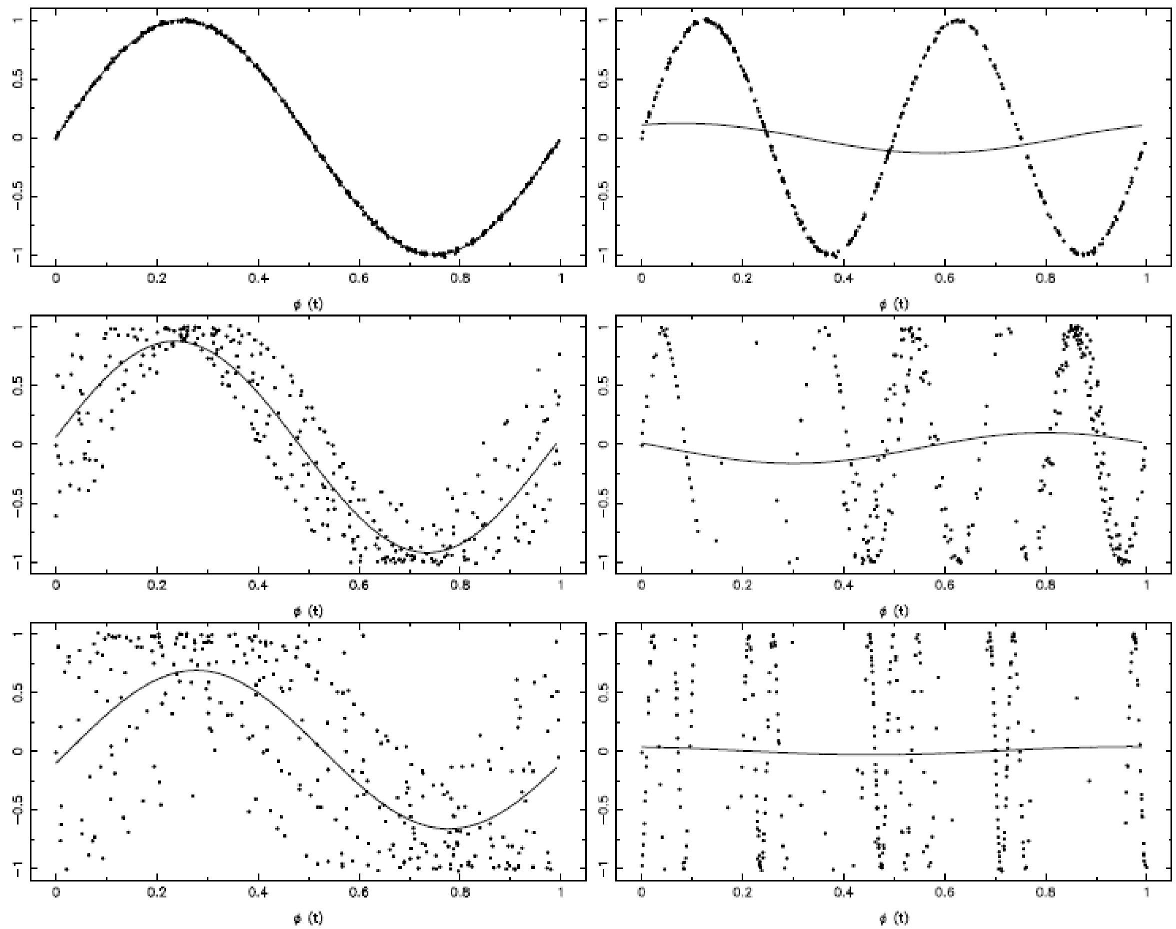
String Length Methods

- The prime disadvantage - the multitude of false peaks compared with Fourier methods.



String Length Methods

Phase diagrams for six minima in Θ_{SL} found from the previous slide.



String Length Methods

- A major advantage of the string length methods that they allow straightforward generalisation to a multivariate treatment.
- The brightness variations in different photometric bands due to oscillations are strongly correlated. Depending on whether or not there are phase differences between the colour curves of the pulsating star, the measurements plotted in a brightness-brightness diagram for two different filters lie on a straight line or an ellipse-like structure.

String Length Methods

- Clarke (2002) proposes the following statistic for multivariate time series:

$$\Theta_{\mathrm{RL}}(\nu) \equiv \frac{1}{Z}\sum_{k=1}^{Z}\left(\frac{\sum_{i=1}^{N[k]}\left[x_k(\phi_{i+1}) - x_k(\phi_i)\right]^2}{\sum_{i=1}^{N[k]}\left[x_k(\phi_i) - \overline{x_k}\right]^2} \times \frac{N[k]-1}{2N[k]}\right),$$

- where $x_k(\varphi_i)$ is the magnitude in filter k or radial velocity from line profile k for each of the measurements taken at times $t_1, \ldots, t_N$ after re-arranging the data such that $\varphi_1, \ldots, \varphi_N$ increases from 0 to 1 for each of the test frequencies ν.
- It is rather cumbersome, however, to interpret the outcome of this statistic for extensive multicolour time series due to the numerous false frequency peaks.

Phase Dispersion Minimization

- The "Phase Dispersion Minimization" (**PDM**) technique was originally presented by Stellingwerf (1978).
- This is basically a folding of the data, together with a binning analysis of the variance at each candidate frequency.
- It is also based on the principle of least squares fit, as in the periodogram, but to a mean curve that is determined by the **data**, rather than a **sine** wave.
- Very widely used in variable star research.

Phase Dispersion Minimization

- The PDM analysis computes the sum of the bin variances divided by the total variance of the data.
- For uncorrelated data this ratio is close to unity.
- At a possible period the bin variances are less than the overall variance, and the statistic (called Theta - Θ) drops to some value greater than zero.

Phase Dispersion Minimization

- For each test frequency *v* one divides the phase interval [0, 1] into *B* equal sub-intervals, called *bins*.
- Suppose that the *j*-th bin contains N_j measurements. The average value of the data, the sum of the quadratic deviations and the variance for this bin are

$$\overline{x_j} = \sum_{i=1}^{N_j} \frac{x_{ij}}{N_j},$$

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

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

where x_{ij} *is* the observation $x(t_i)$ with bin index *j*.

Phase Dispersion Minimization

- The analogous quantities for all data, $\bar{x}$, V^2 and s^2, are defined as

$$\overline{x} = \sum_{i=1}^{N} \frac{x_i}{N},$$

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

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

Phase Dispersion Minimization

- The partition of the phase diagram into B equal bins can have disadvantages. It may very well happen that some bins are almost empty if B is chosen to be large or if we have only few data points with a particular time spread.
- For this reason one makes use of a more complicated *bin/cover structure* (B,C). The phase diagram is divided into B bins, each of length $1/B$. This partition is then applied C times, such that each partition is shifted over $1/(B\times C)$ with respect to the previous one. The incomplete bin near phase 1 is completed with the data of the corresponding phase interval near $\varphi = 0$.
- In this way one covers the phase diagram C times, and each partition contains B bins. Such a bin structure allows one to make sure that each observation belongs to at least one bin. Further on we denote the total number of bins as $B_C \equiv B \times C$.

Phase Dispersion Minimization

- We introduce the statistic Θ_{PDM}:

$$\Theta_{\mathrm{PDM}} \equiv \frac{\left(\sum_{j=1}^{B_C} (N_j - 1)\, s_j^2\right) / \left(\sum_{j=1}^{B_C} N_j - B_C\right)}{\left(\sum_{i=1}^{N} (x_i - \overline{x})^2\right) / (N - 1)},$$

where s_j^2 is defined as:

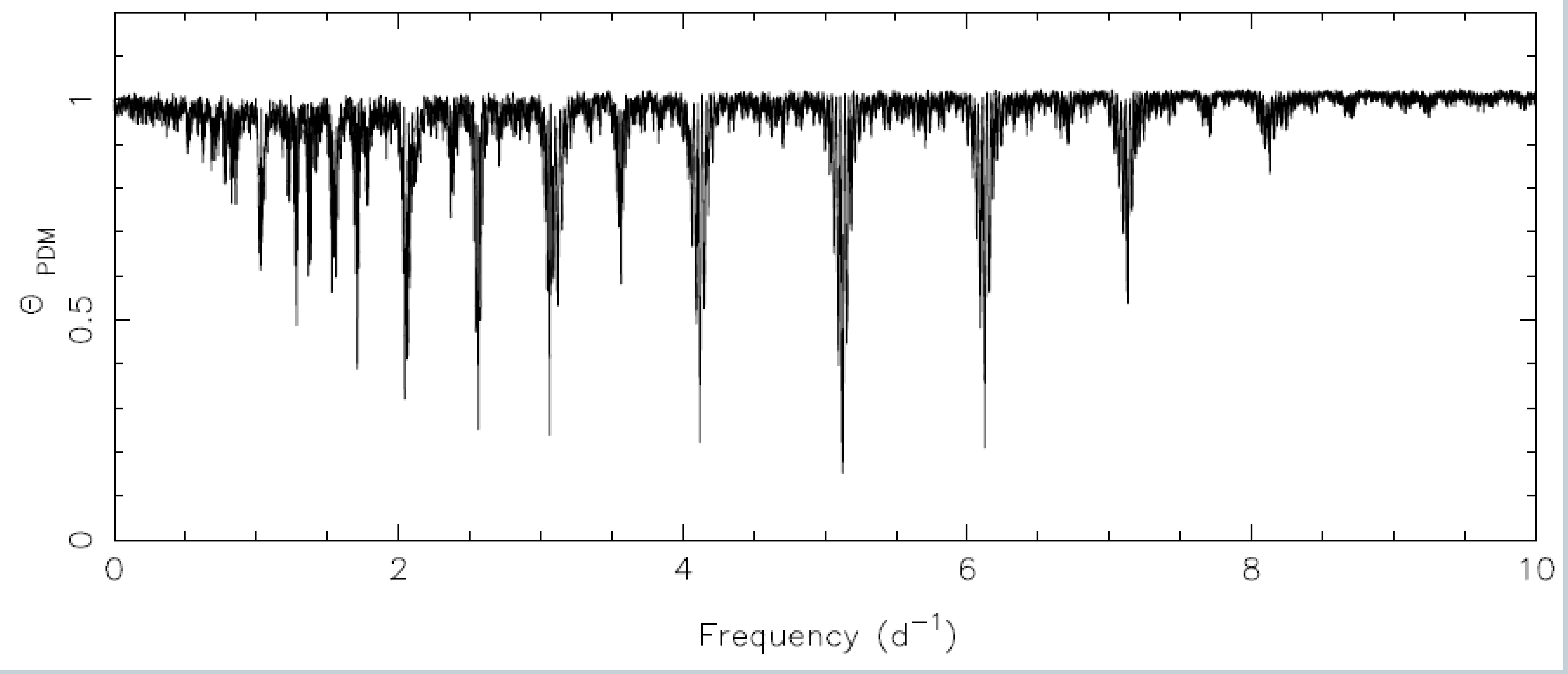
$$s_j^2 \equiv \frac{\sum_{i=1}^{N_j} (x_{ij} - \overline{x_j})^2}{N_j - 1}.$$

Phase Dispersion Minimization

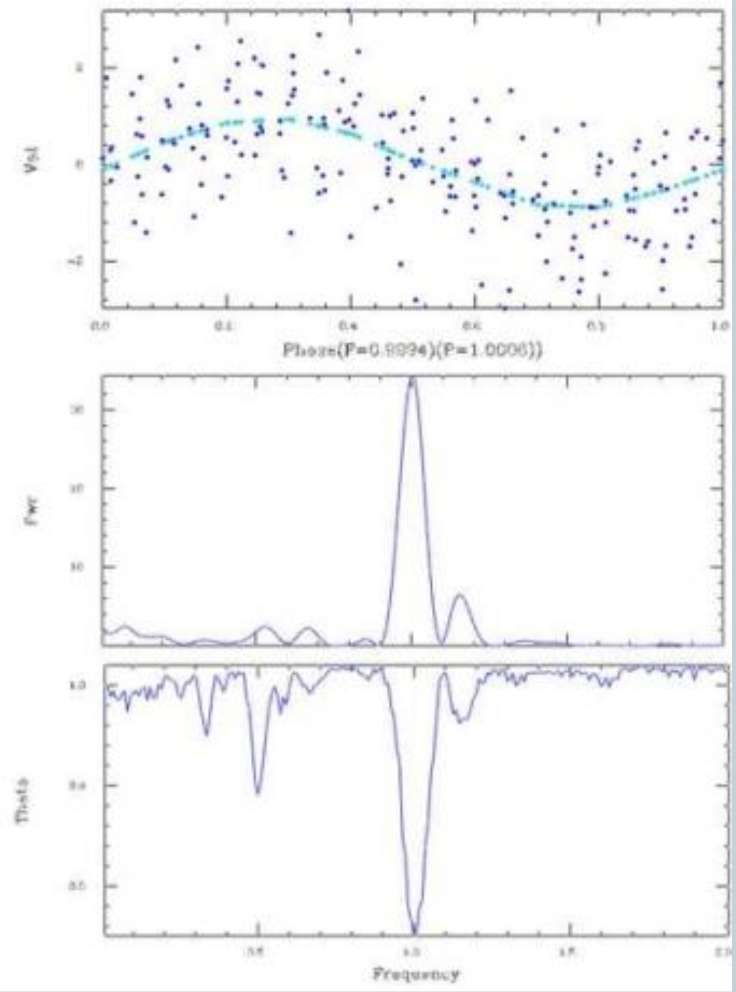
- The search for the most likely frequency in the data comes down to the search for the minimum of Θ_{PDM}.
- For each test frequency that is not present in the data we will find $\Theta_{\mathrm{PDM}} \approx 1$.
- The Θ_{PDM}–statistic was introduced by Stellingwerf (1978) and is a generalisation of the Θ statistic used by Jurkevich (1971) which is only based on bins (C=1).

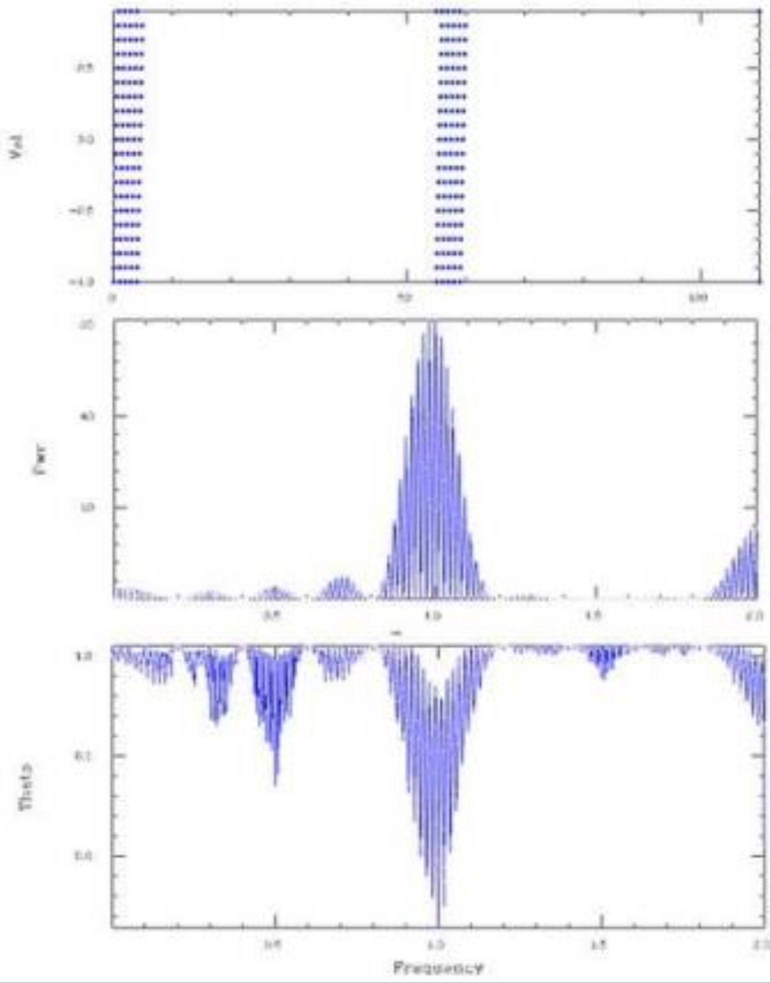
- Experience has shown that 10 bins are adequate (and usually optimum) for variable star data sets. For data sets with more than about 100 points, these bins are non-overlapping. For data sets with less than about 100 points, better results are obtained if the bins are double-wide, and overlap.

Phase Dispersion Minimization

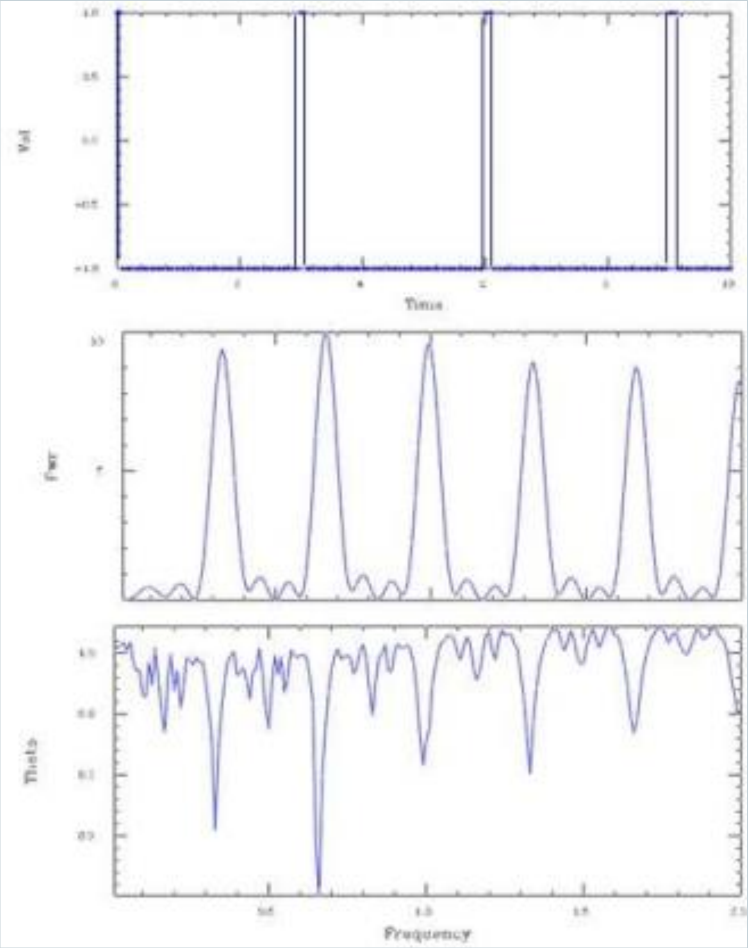


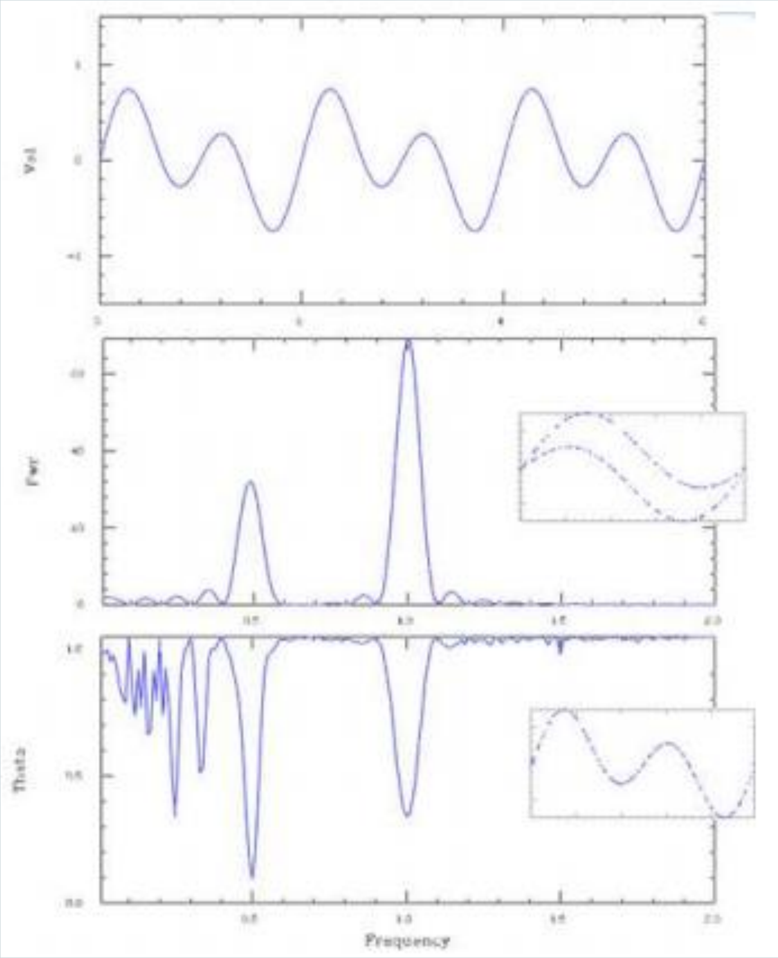
Comparison of PDM and the LS method





Comparison of PDM and the LS method

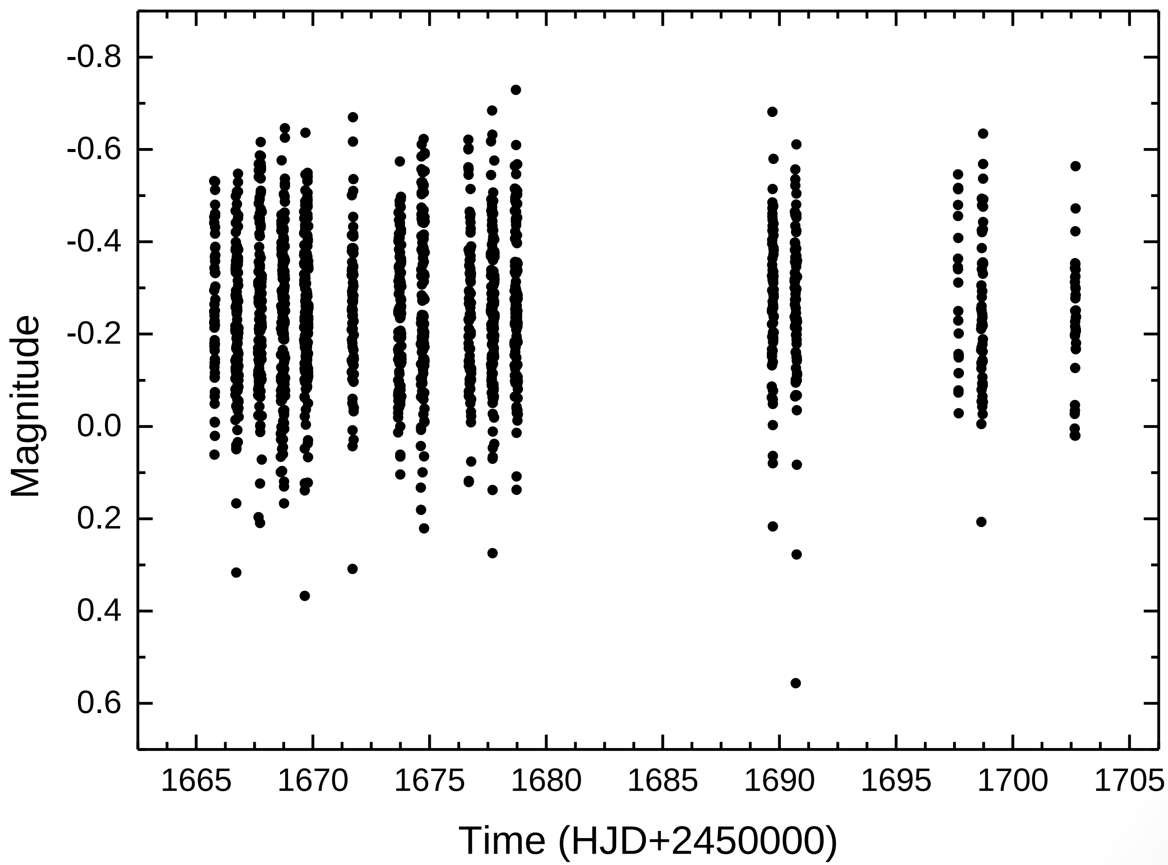




EX Hydrae

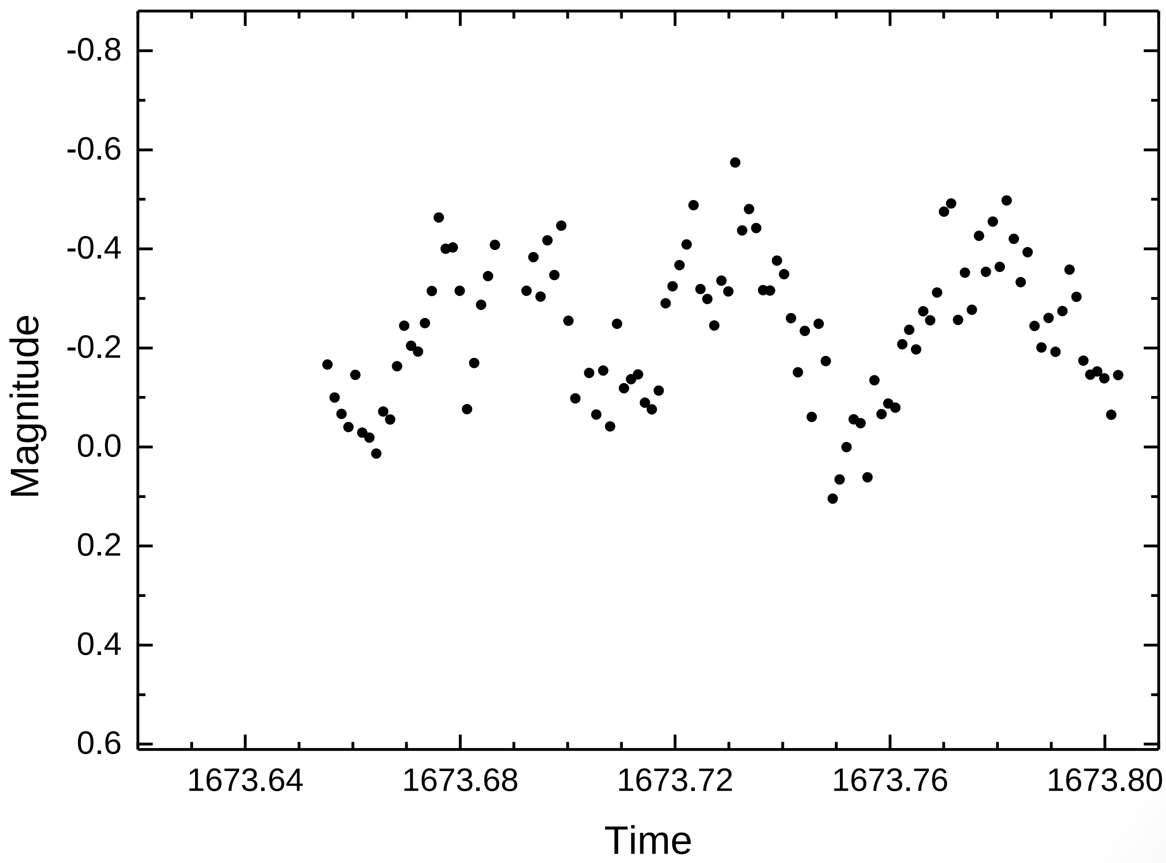
- Intermediate Polar
- Periods:
  - Orbital: 0.068233846 d = 5895.4s, seen in optical and X-rays
  - Spin: 0.046546504d = 4021.617946s, seen in optical photometry, optical spectroscopy and X-rays.
  - Others: Sideband(s) seen only during outbursts.

EX Hydrae

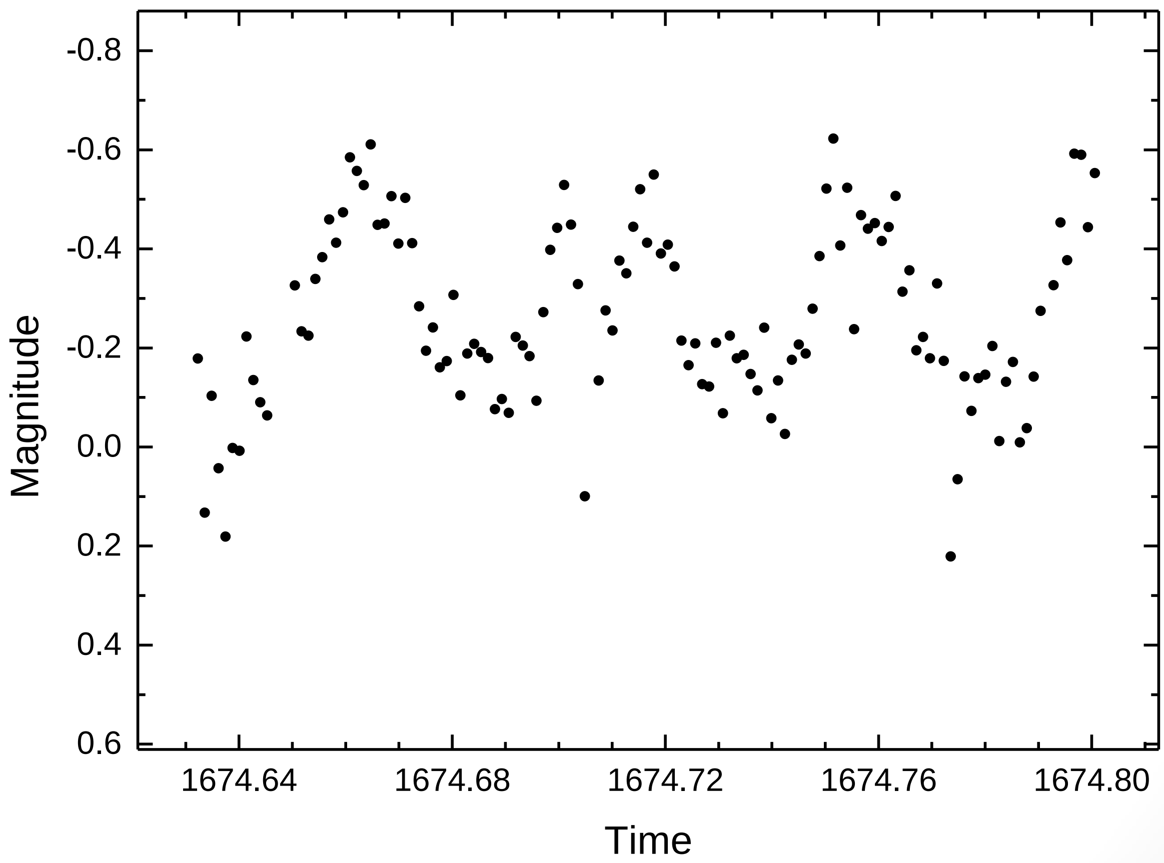


-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Magnitude
1665
1670
1675
1680
1685
1690
1695
1700
1705
Time (HJD+2450000)

EX Hydrae

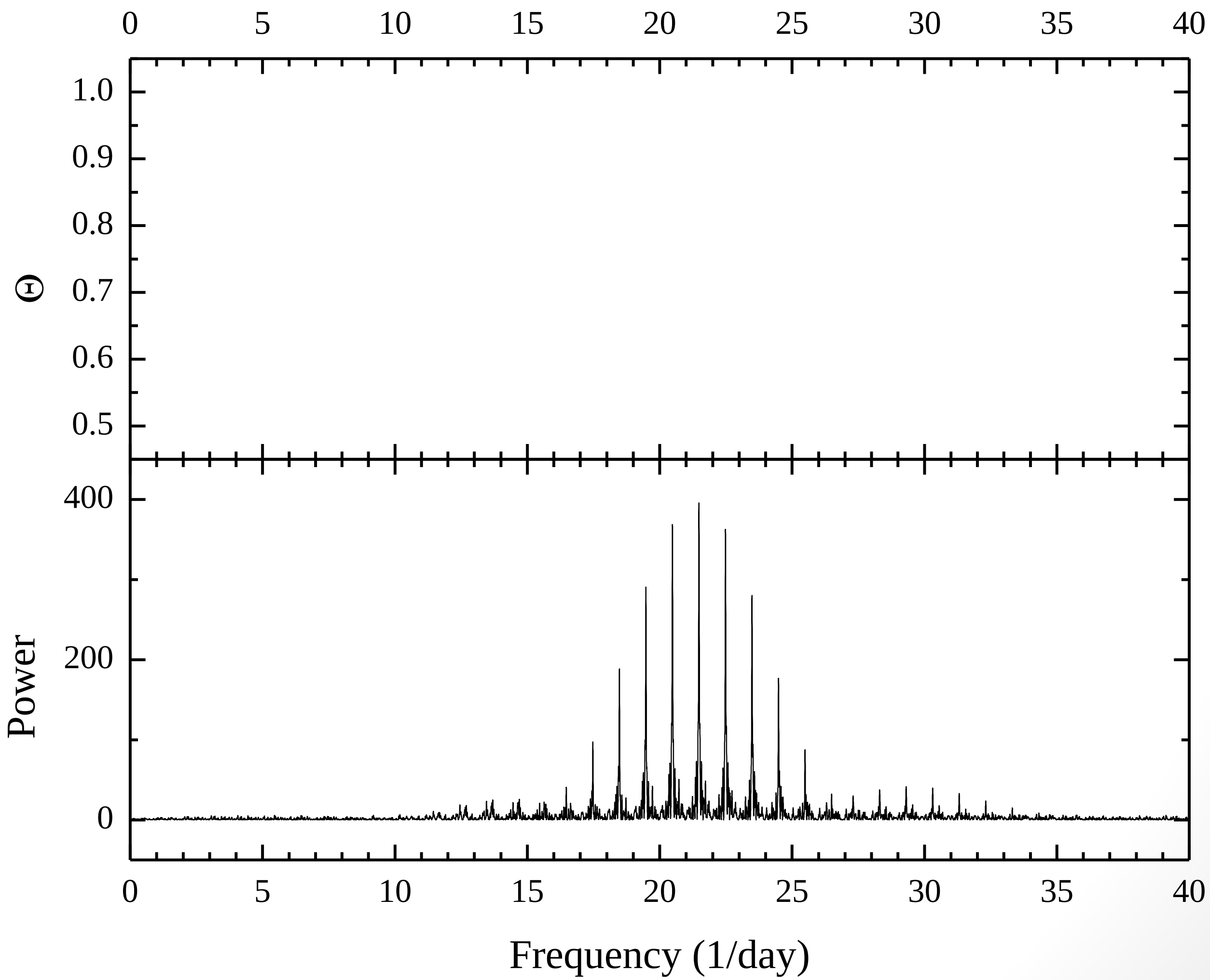

-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Magnitude
1673.64
1673.68
1673.72
1673.76
1673.80
Time

EX Hydrae



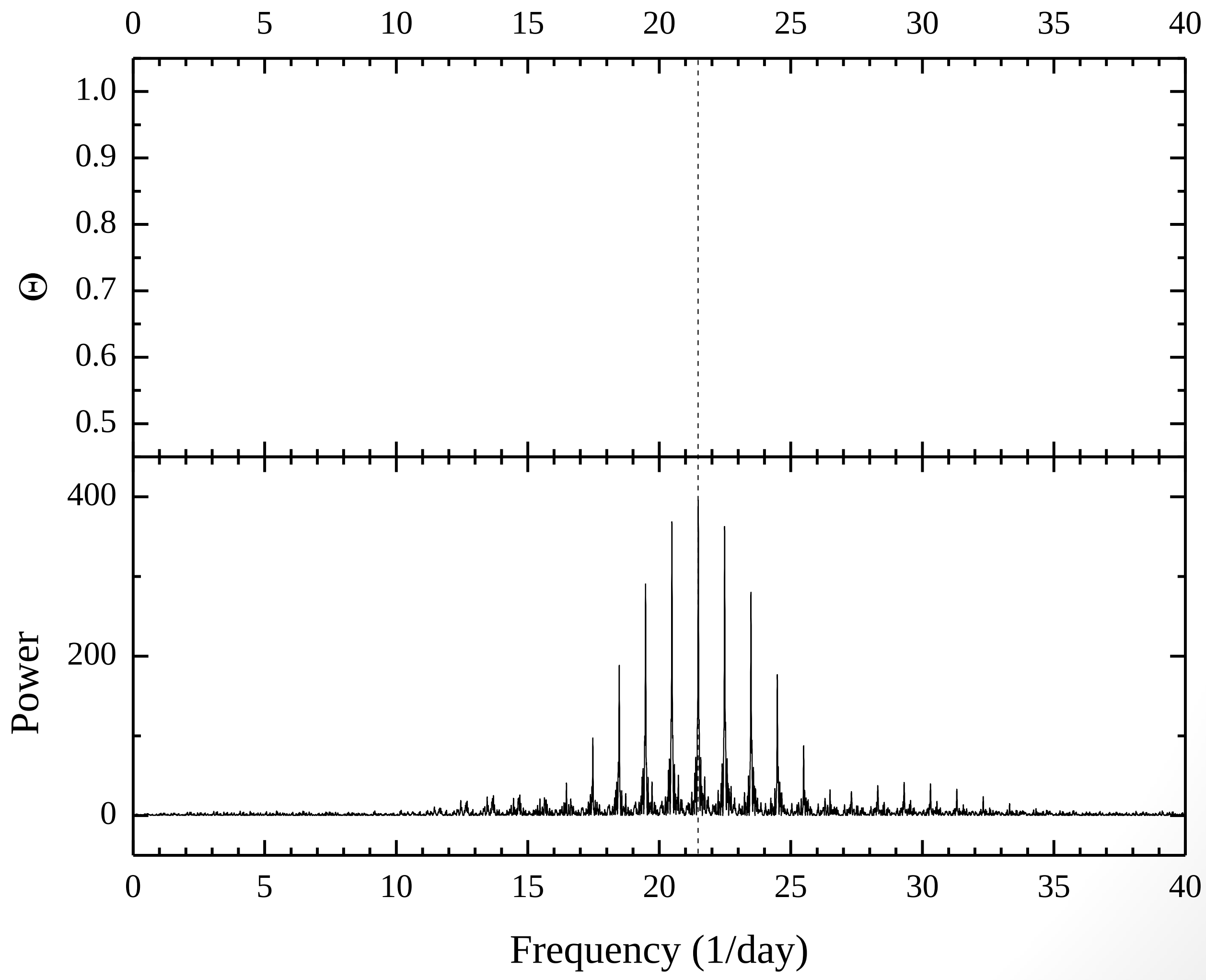
Magnitude
Time
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
1674.64
1674.68
1674.72
1674.76
1674.80

EX Hydrae



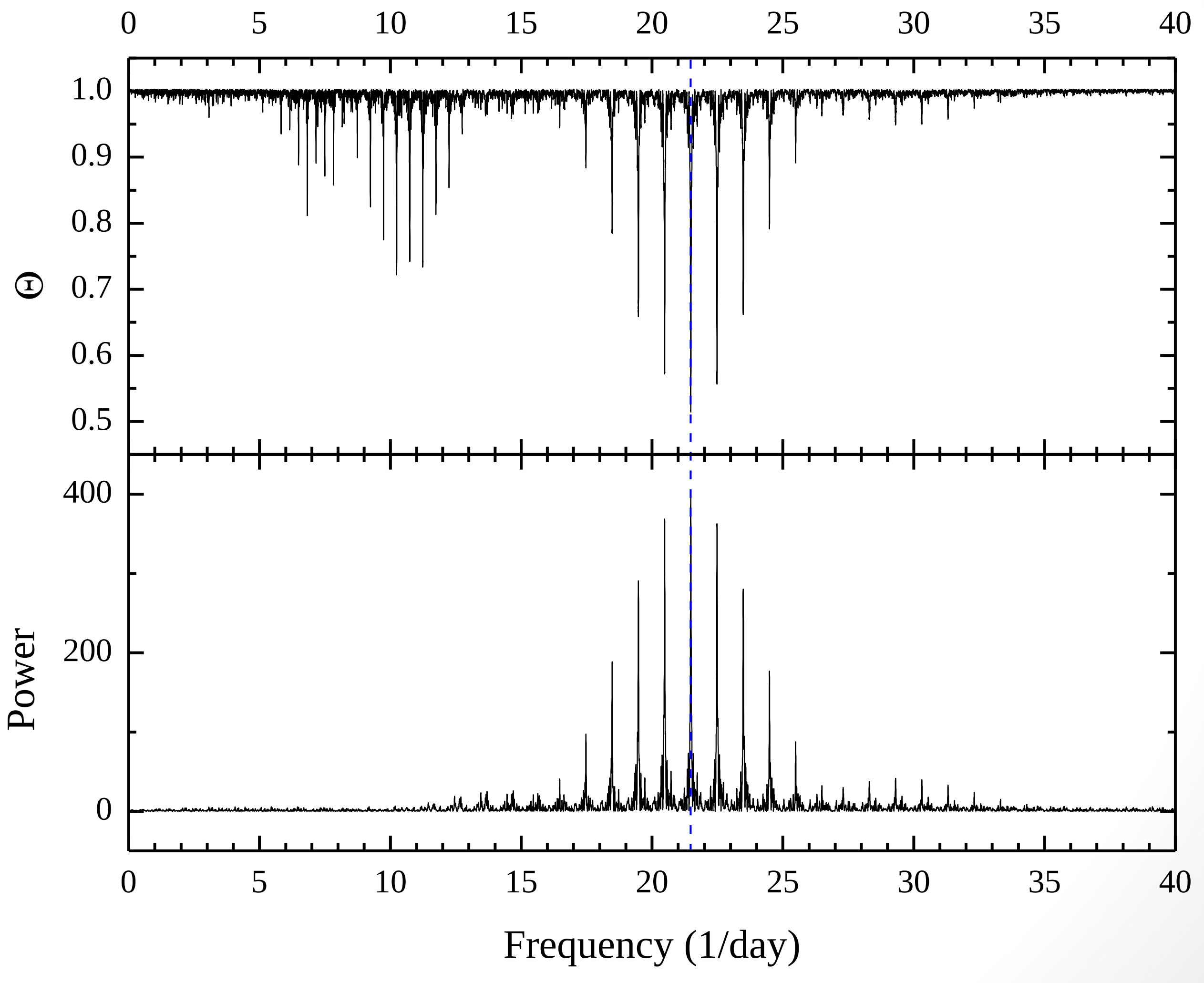
0 5 10 15 20 25 30 35 40
1.0
0.9
0.8
0.7
0.6
0.5
Θ
400
200
0
Power
0 5 10 15 20 25 30 35 40
Frequency (1/day)

EX Hydrae



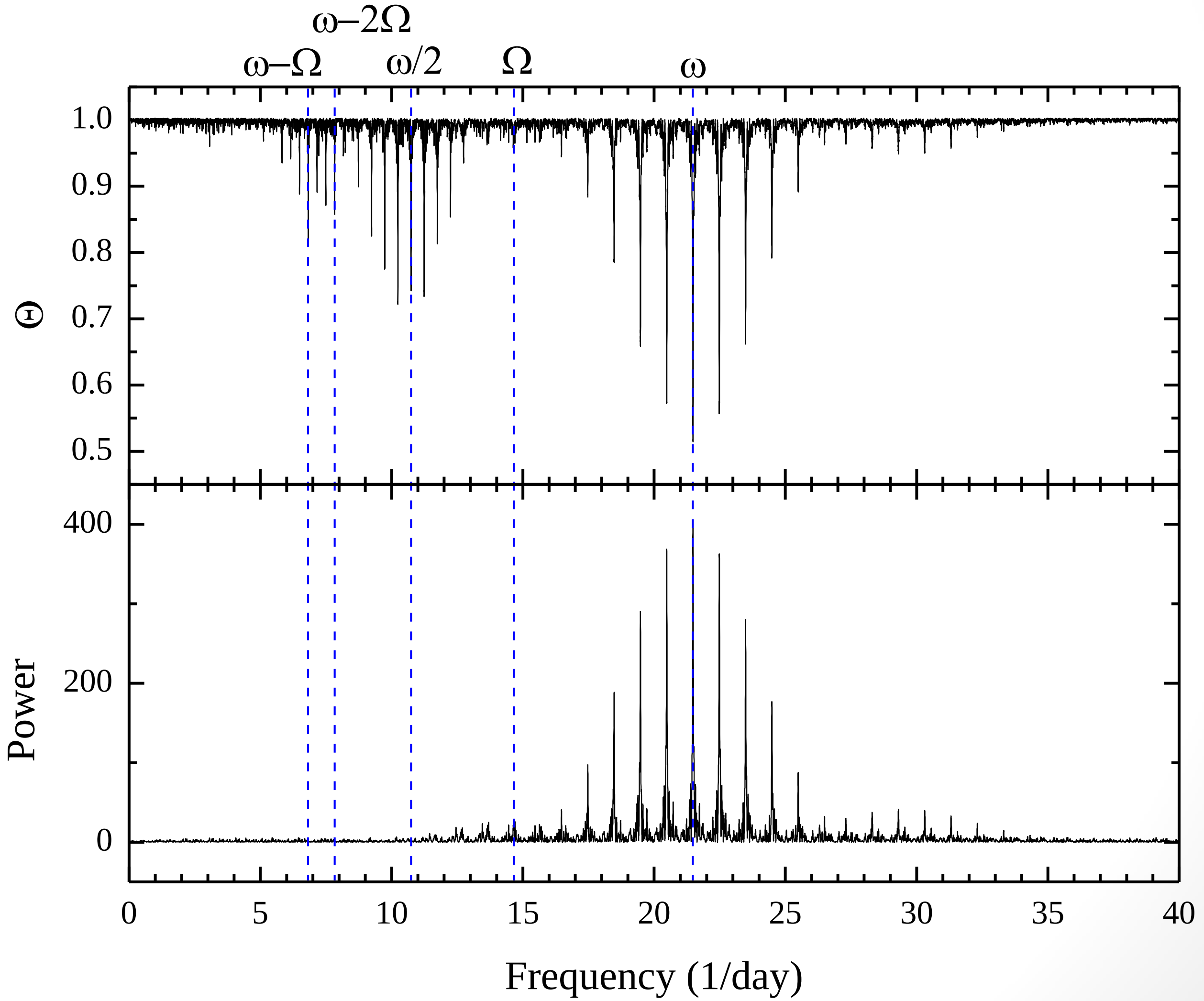
0 5 10 15 20 25 30 35 40
1.0
0.9
0.8
0.7
0.6
0.5
Θ
400
200
0
Power
Frequency (1/day)

EX Hydrae

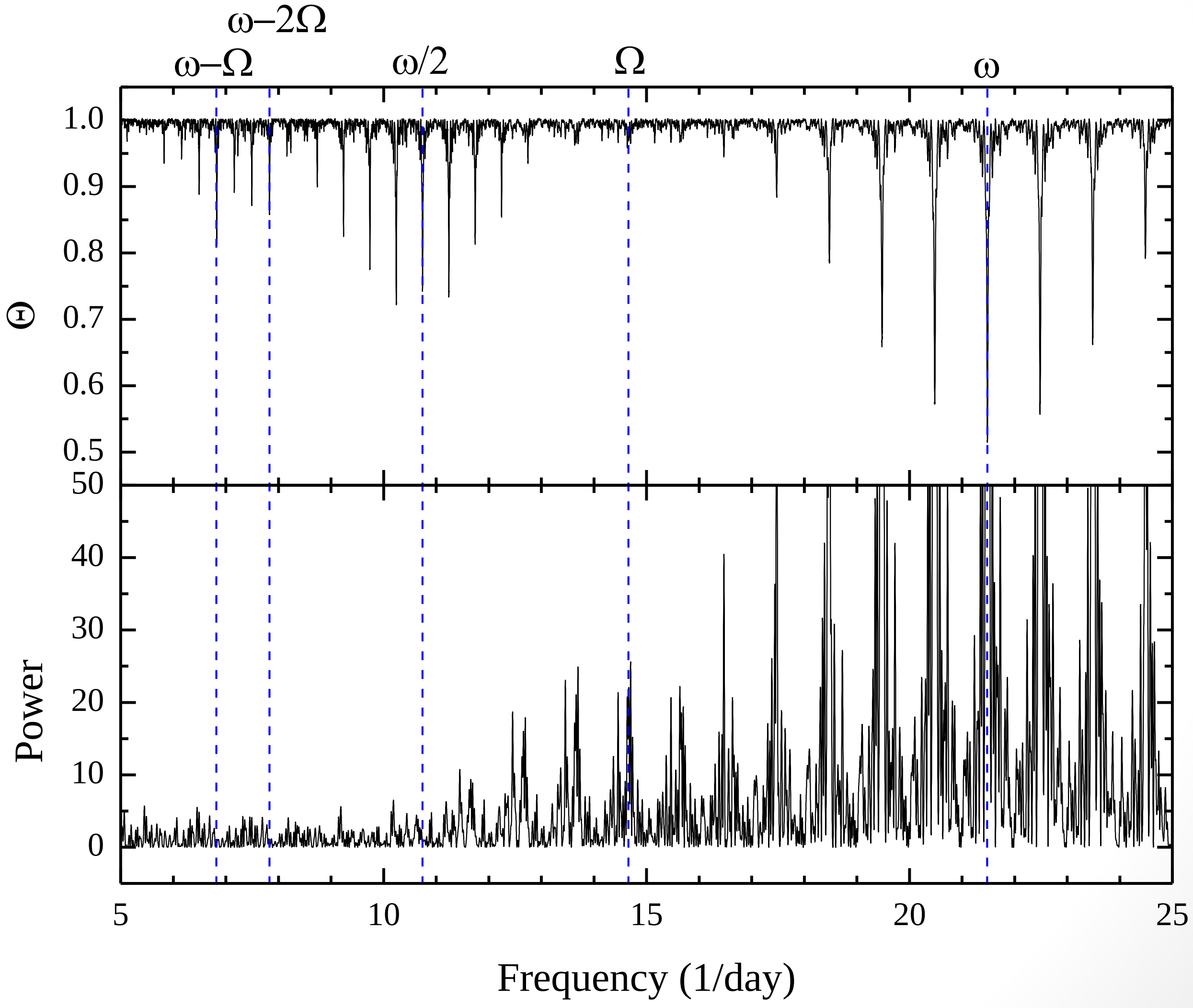


0 5 10 15 20 25 30 35 40
1.0
0.9
0.8
0.7
0.6
0.5
Θ
400
200
0
Power
Frequency (1/day)

EX Hydrae


ω–2Ω
ω–Ω
ω/2
Ω
ω
Θ
Power
Frequency (1/day)
1.0
0.9
0.8
0.7
0.6
0.5
400
200
0
0
5
10
15
20
25
30
35
40

EX Hydrae



ω–2Ω
ω–Ω
ω/2
Ω
ω
1.0
0.9
0.8
0.7
0.6
0.5
Θ
50
40
30
20
10
0
Power
5
10
15
20
25
Frequency (1/day)

Comparison of PDM and the LS method

- Two methods elaborate each other.

Other Non-parametric Methods

- **Analysis of Variations (AoV)** by Schwarzenberg-Czerny (1989).
  Similar to PDM, but uses another statistic.
- **Rayleigh and Z_n^2** tests (Leahy et al. 1983) for periodicity search Poisson distributed photon arrival events. Equivalent to Fourier spectrum at high count rates.
- **Bayesian periodicity search** (Gregory & Loredo 1992). Designed for non-sinusoidal periodic shapes observed with Poisson events. Calculates odds ratio for periodic over constant model and most probable shape.

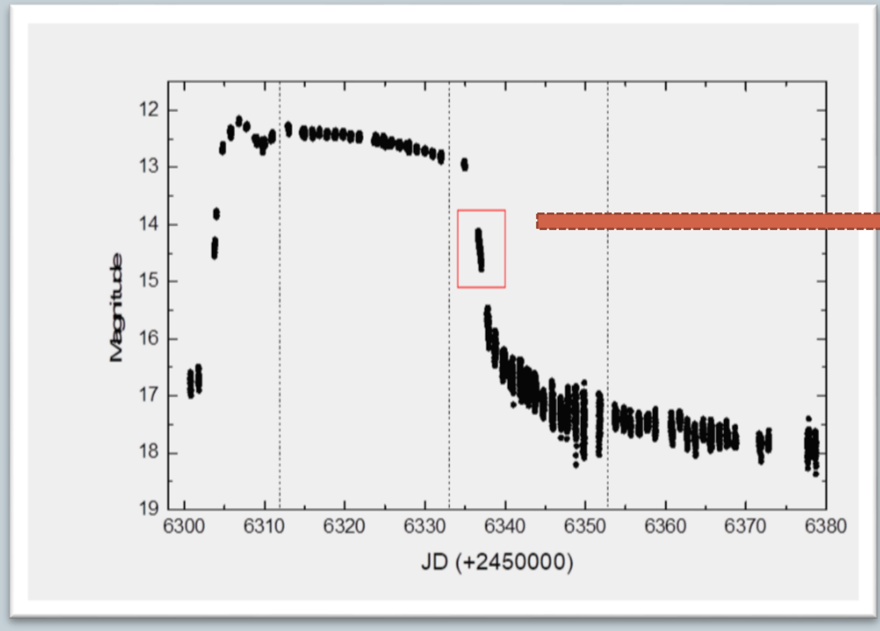
Detrending

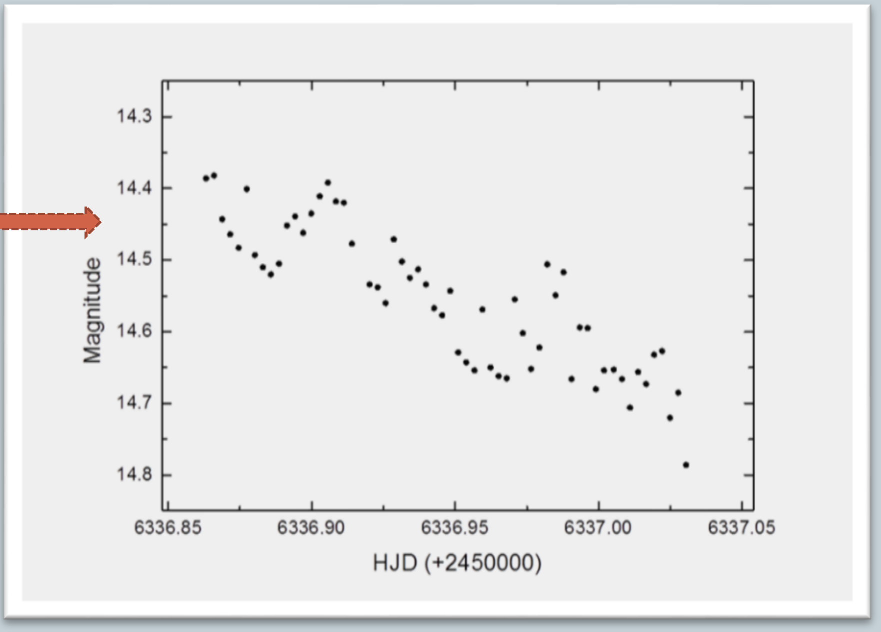
- *Trend* in a time series is a slow, gradual change in some property of the series over the whole interval under investigation.

- Trend is sometimes defined as a long term change in the mean, but can also refer to change in other statistical properties.

- *Detrending* is the statistical or mathematical operation of removing trend from the series.

Detrending


Magnitude
12
13
14
15
16
17
18
19
6300
6310
6320
6330
6340
6350
6360
6370
6380
JD (+2450000)


Magnitude
14.3
14.4
14.5
14.6
14.7
14.8
6336.85
6336.90
6336.95
6337.00
6337.05
HJD (+2450000)

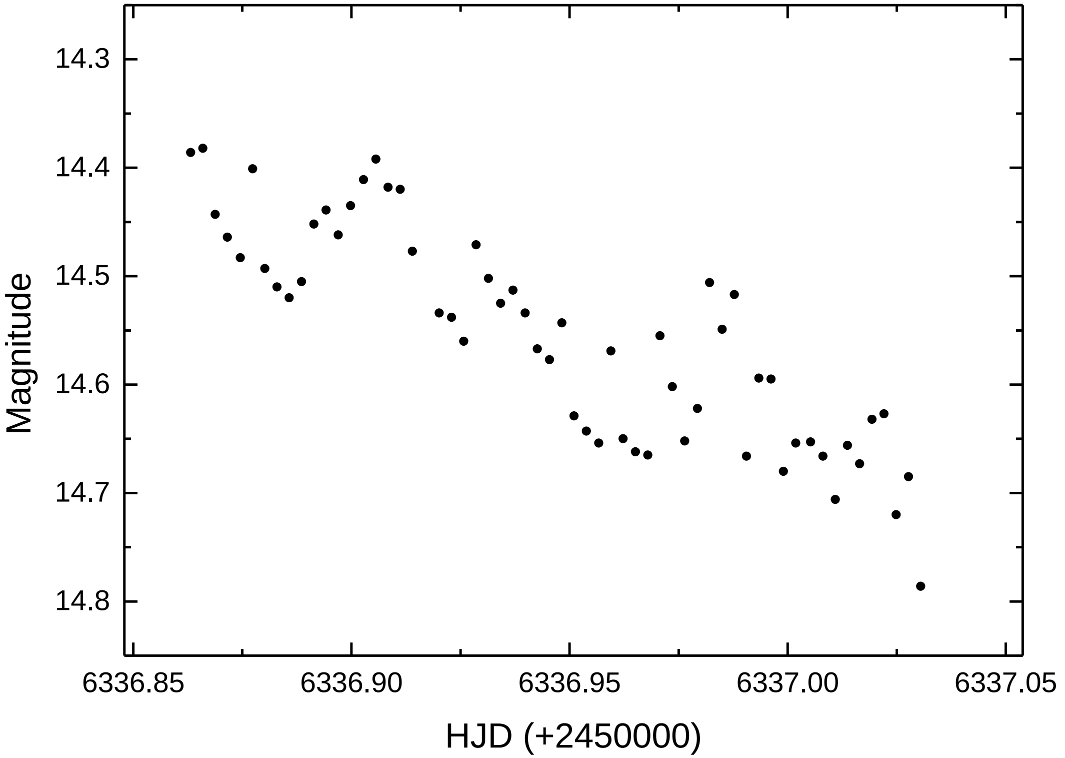
Detrending

- In studying and removing trend, it is important to understand the effect of detrending on the spectral properties of the time series. This effect can be summarized by the *frequency response* of the detrending function.

- Many alternative methods are available for detrending:
  - A simple and widely used function of time is the least-squares-fit straight line, which assumes linear trend.
  - Other functions of time (e.g., quadratic) might be better depending on the type of data.
  - An alternative to fitting a curve to the entire time series (curve fitting) is to fit polynomials of time to different parts of the time series.
  - Sometimes the mathematical form of the trend function has physical basis.

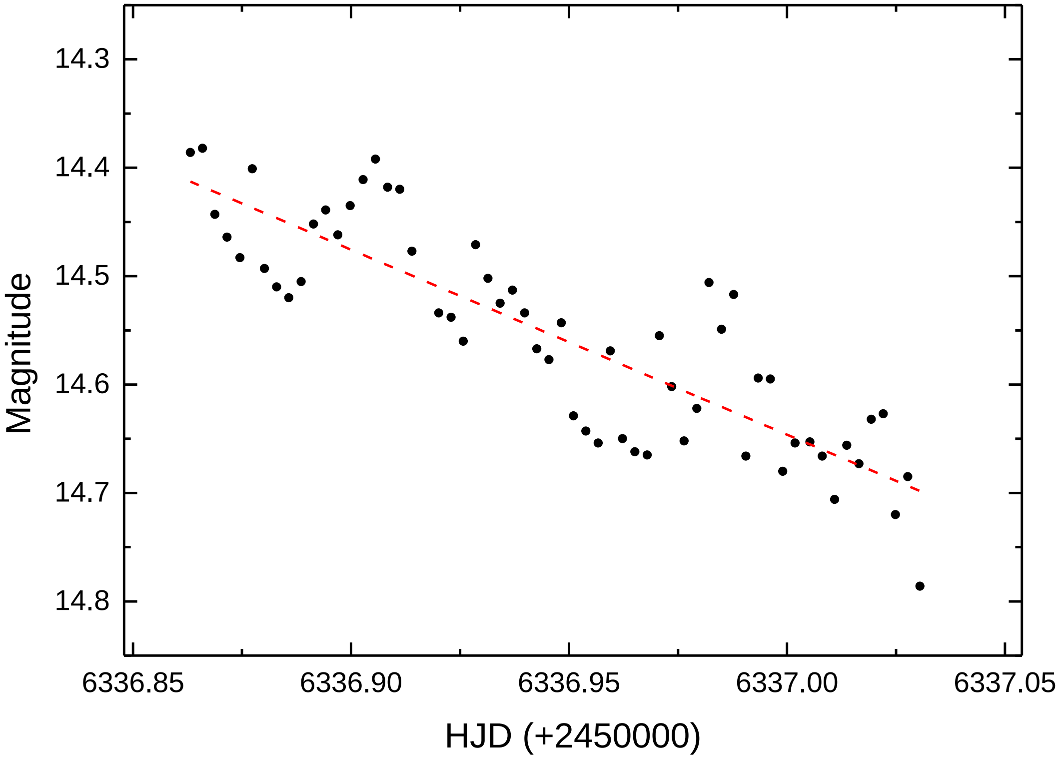
Detrending

Original light curve



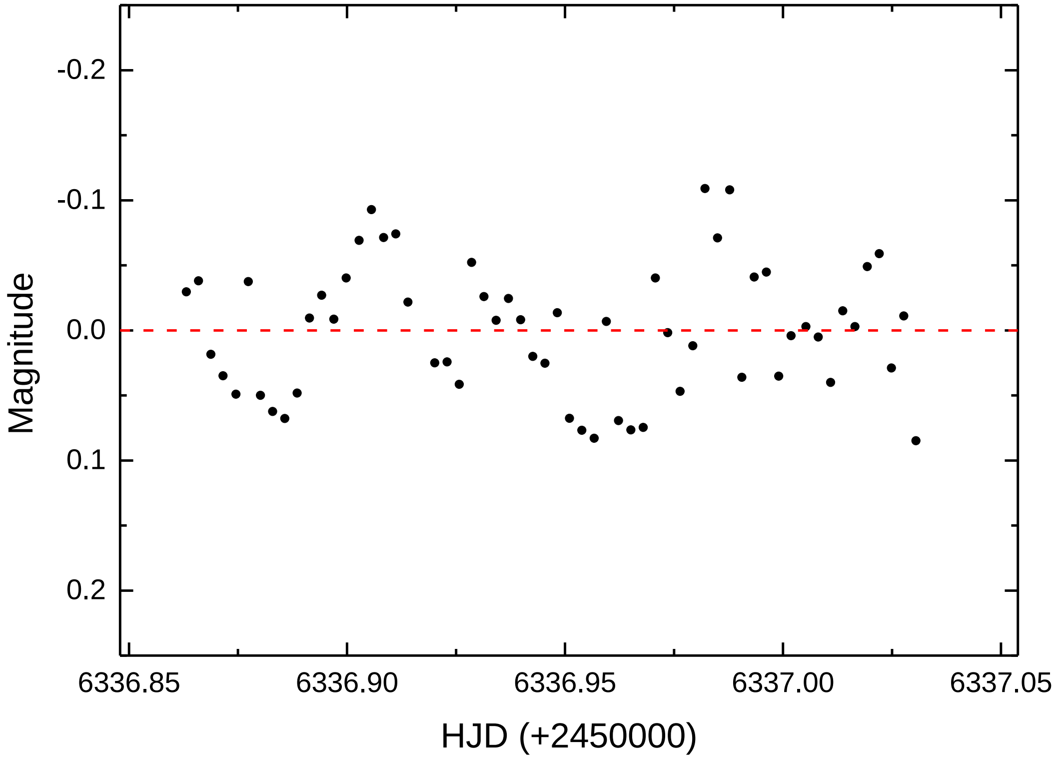
Detrending

The least-squares-fit straight line to the light curve



Detrending

Detrended light curve



Detrending

- Identification of trend in a time series is subjective because trend cannot be unequivocally distinguished from low frequency fluctuations. What looks like trend in a short segment of a time series segment often proves to be a low-frequency fluctuation – perhaps part of a cycle – in the longer series.

- We can view the entire observed time series as a segment of an unknown infinitely long series, and cannot be sure that an observed change in mean over that segment is not part of some low-frequency fluctuation imparted by a stationary process.