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 $0 \leq \varphi_1, \ldots, \varphi_N < 1$.
- 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

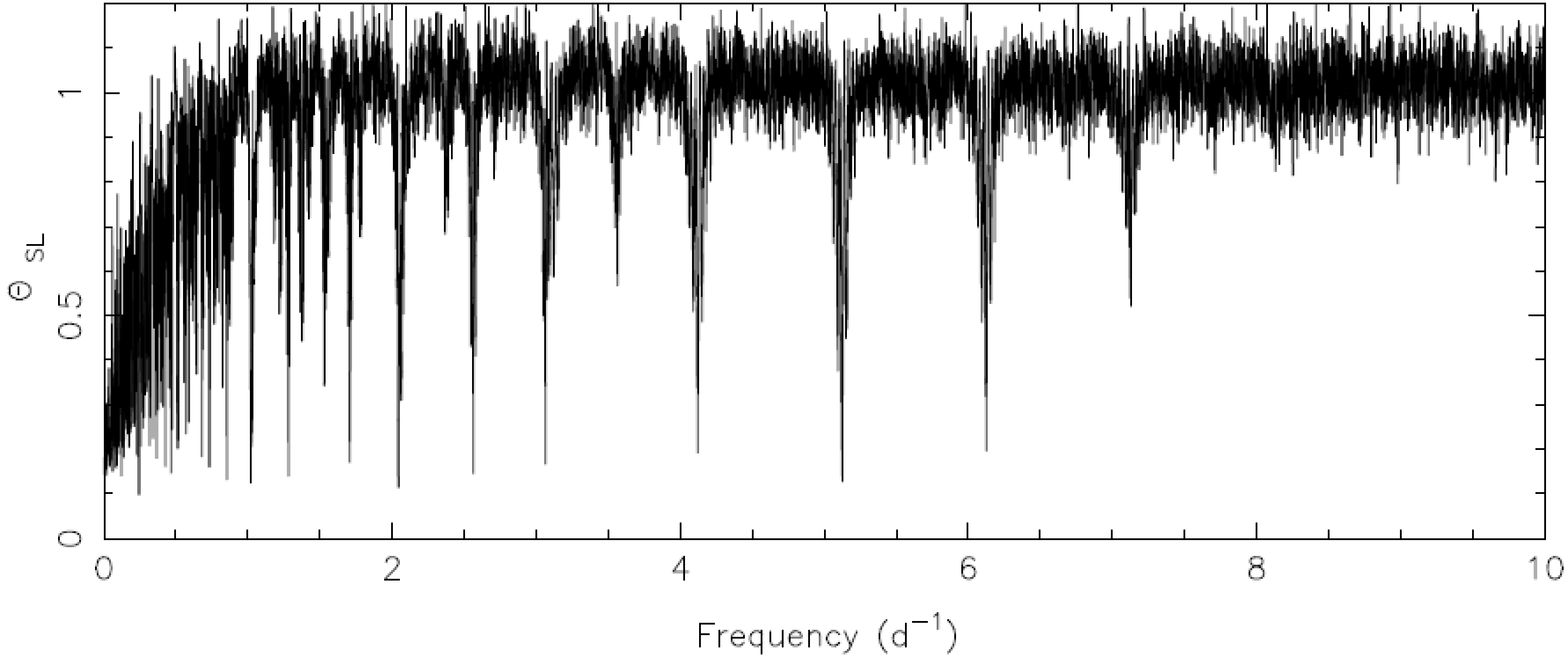
- Clarke (2002) advises the use of the following modified string length statistic:

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

- 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 $\Theta_{\mathrm{SL}} \approx 1.0$.

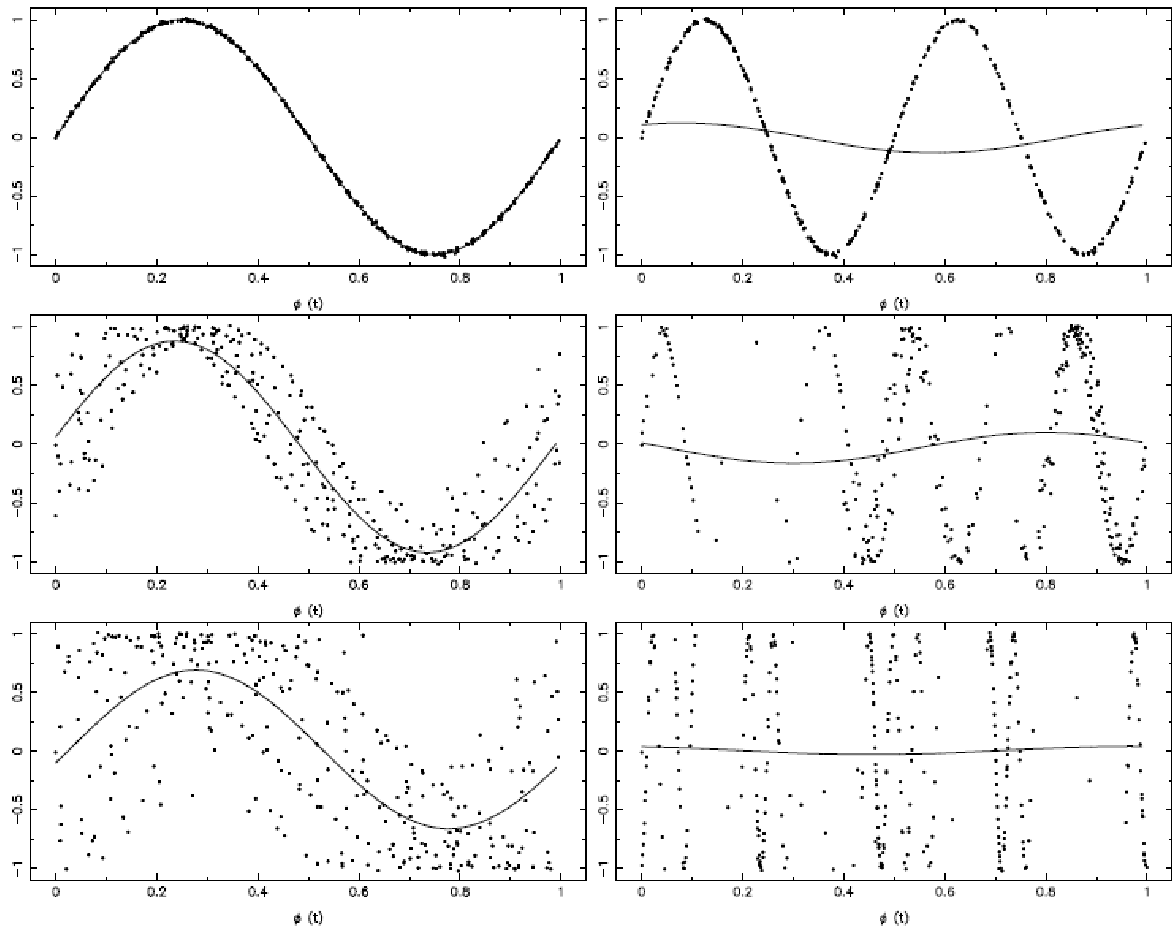
String Length Methods

- 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:

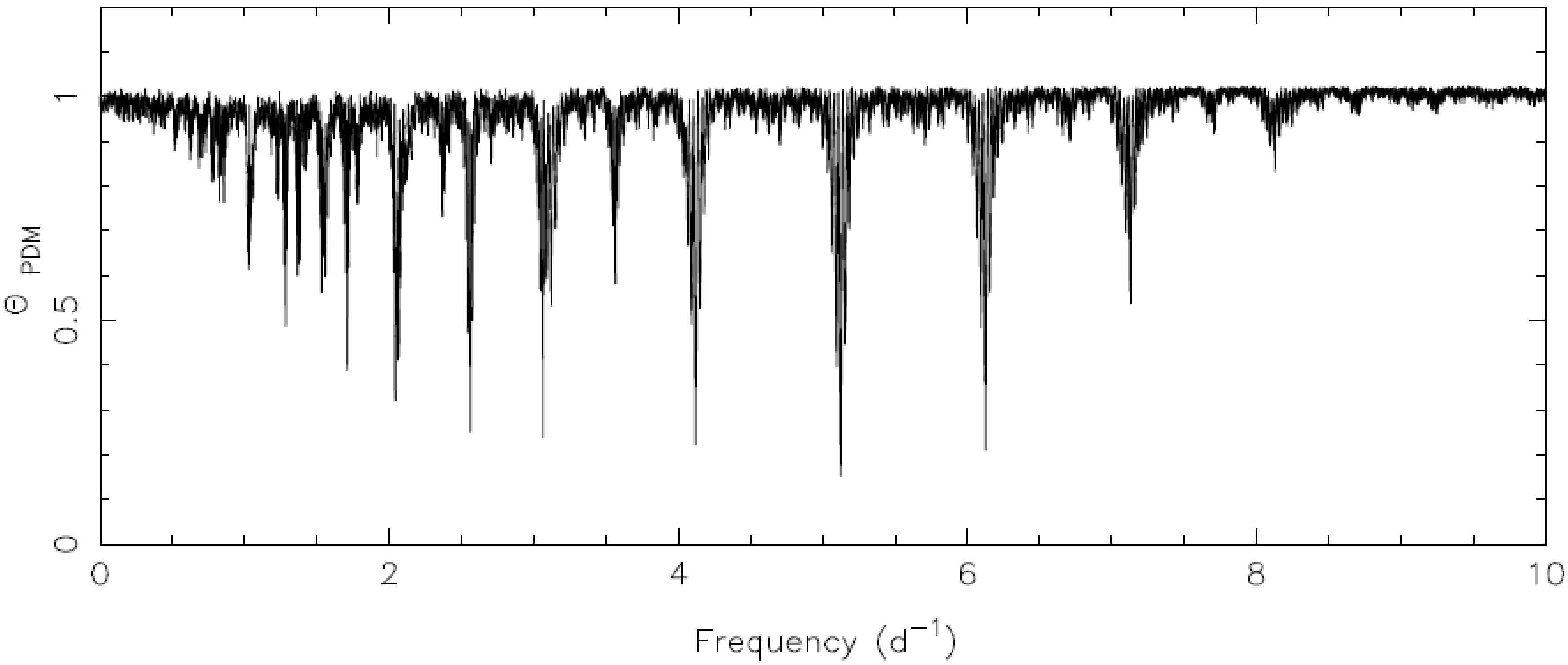
$$s_j^2 \equiv \frac{\sum_{i=1}^{N_j} (x_{ij} - \overline{x_j})^2}{N_j - 1}.$$

Phase Dispersion Minimization

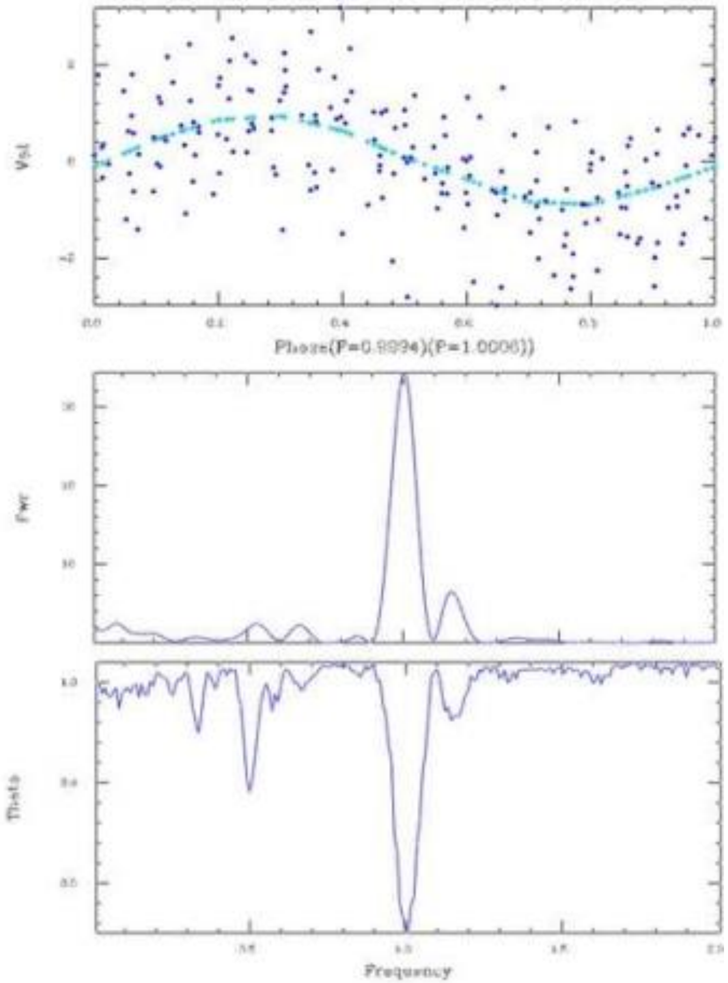
- The search for the most likely frequency in the data comes down to the search for the minimum of Θ_{PDM}.
- For each test frequency that is not present in the data we will find $\Theta_{\mathrm{PDM}} \approx 1$.
- The Θ_{PDM}–statistic was introduced by Stellingwerf (1978) and is a generalisation of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

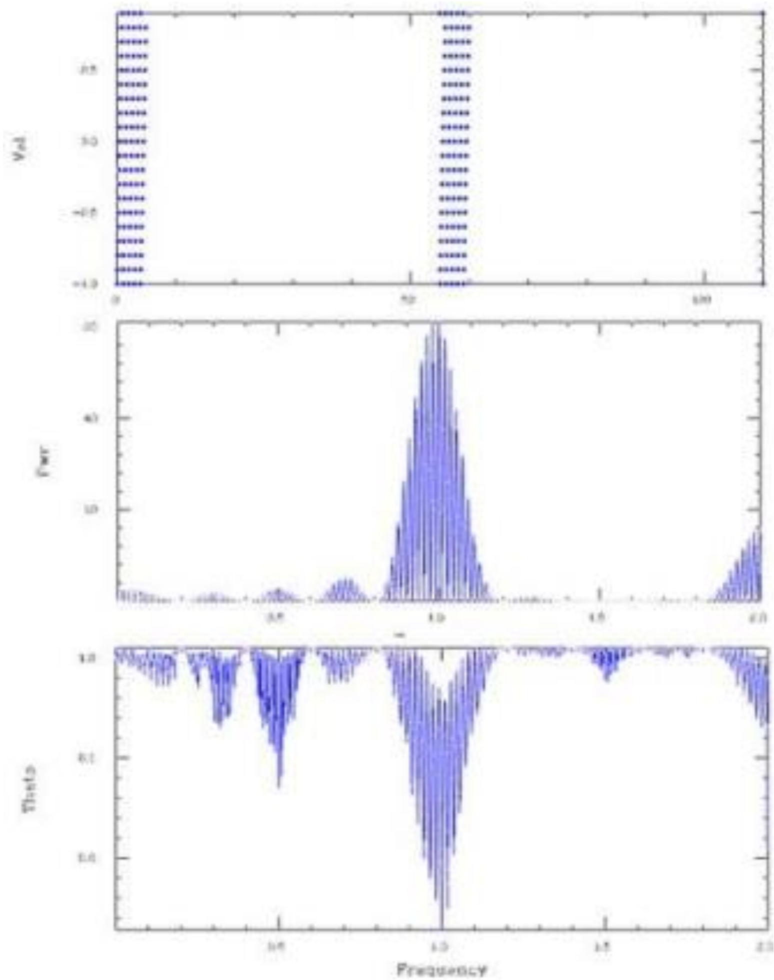
- Experience has shown that 10 bins are adequate (and usually optimum) for variable star data sets. For data sets with more than about 100 points, these bins are non-overlapping. For data sets with less than about 100 points, better results are obtained if the bins are double-wide, and overlap.

Phase Dispersion Minimization

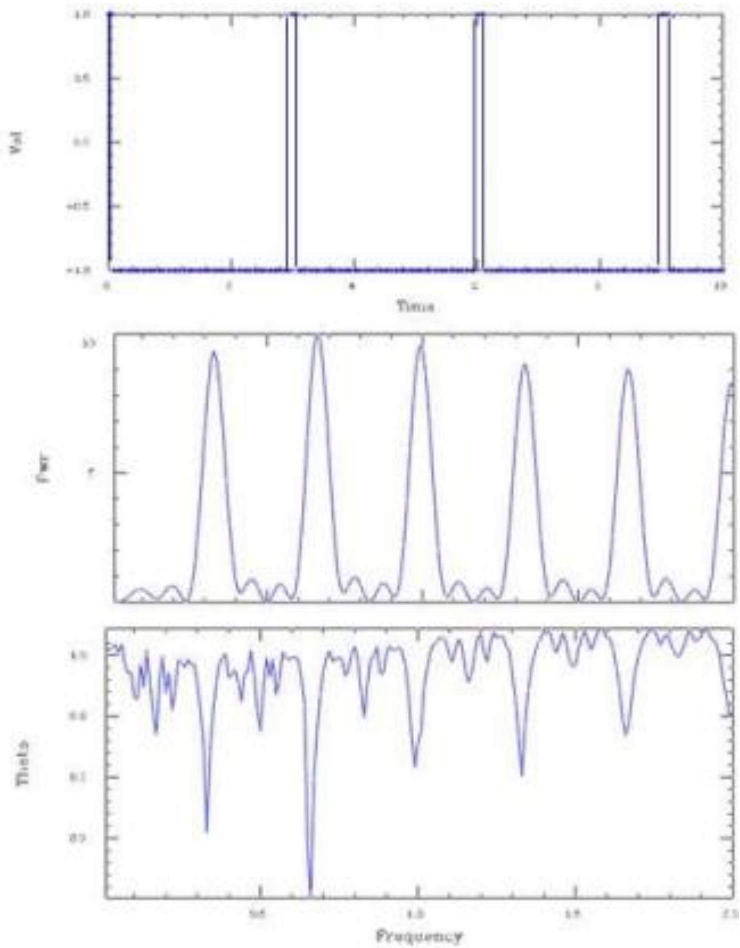


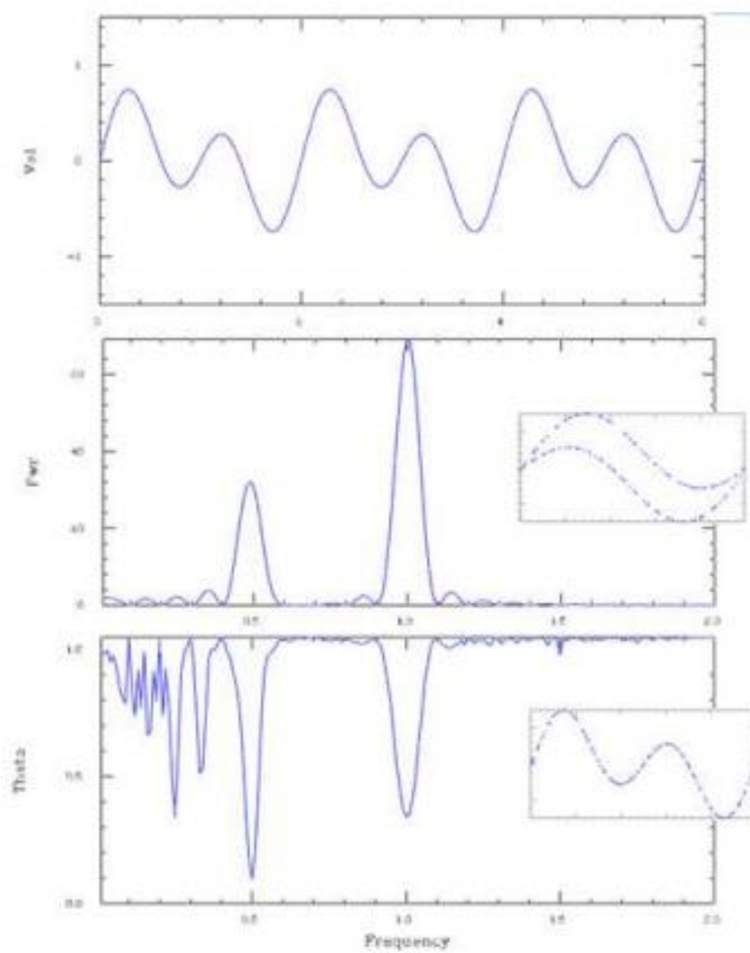
Comparison of PDM and the LS method





Comparison of PDM and the LS method

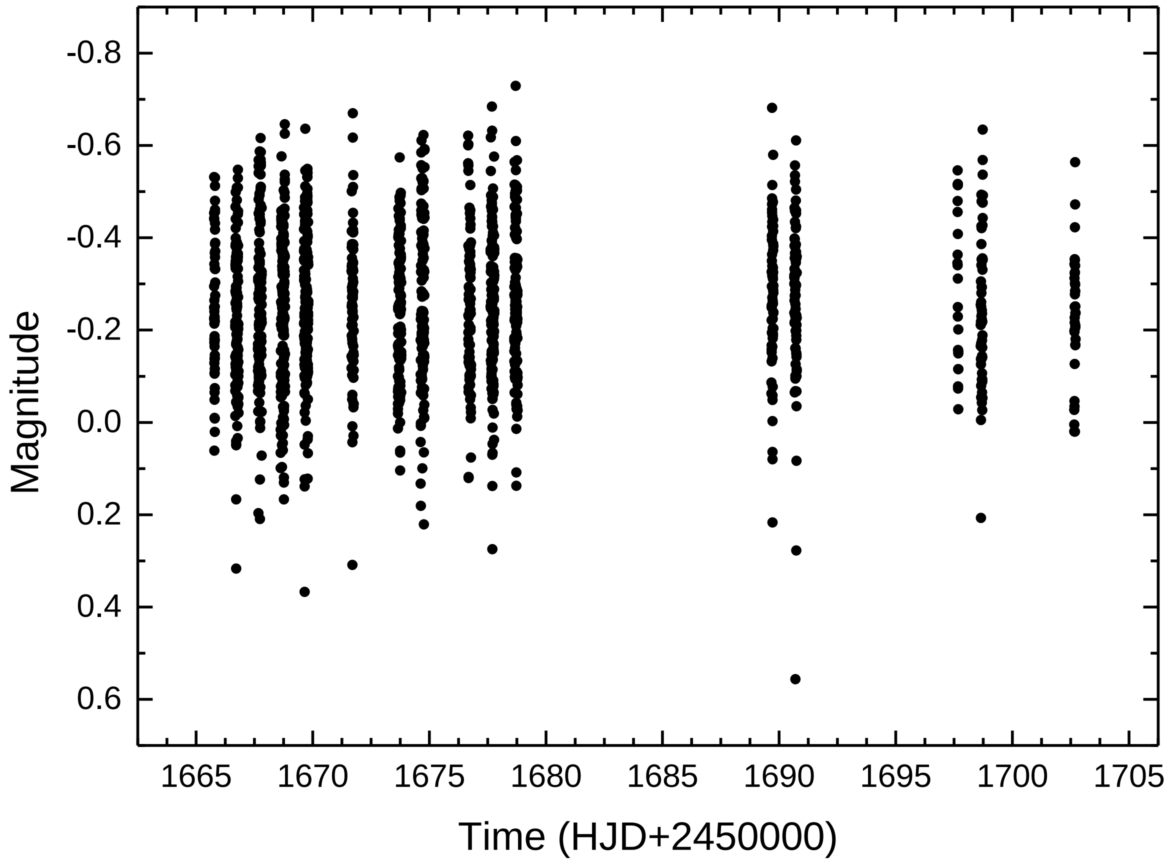




EX Hydrae

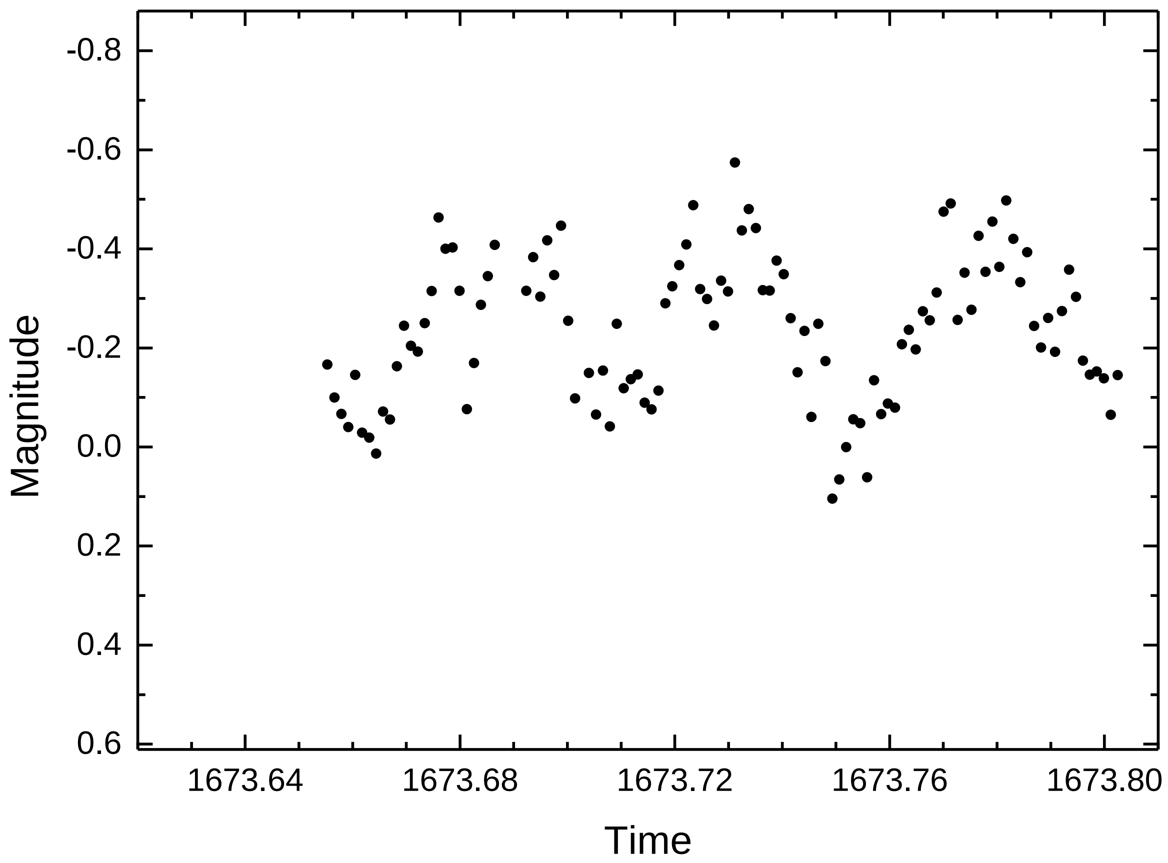
- Intermediate Polar
- Periods:
  - Orbital: 0.068233846 d = 5895.4s, seen in optical and X-rays
  - Spin: 0.046546504d = 4021.617946s, seen in optical photometry, optical spectroscopy and X-rays.
  - Others: Sideband(s) seen only during outbursts.

EX Hydrae



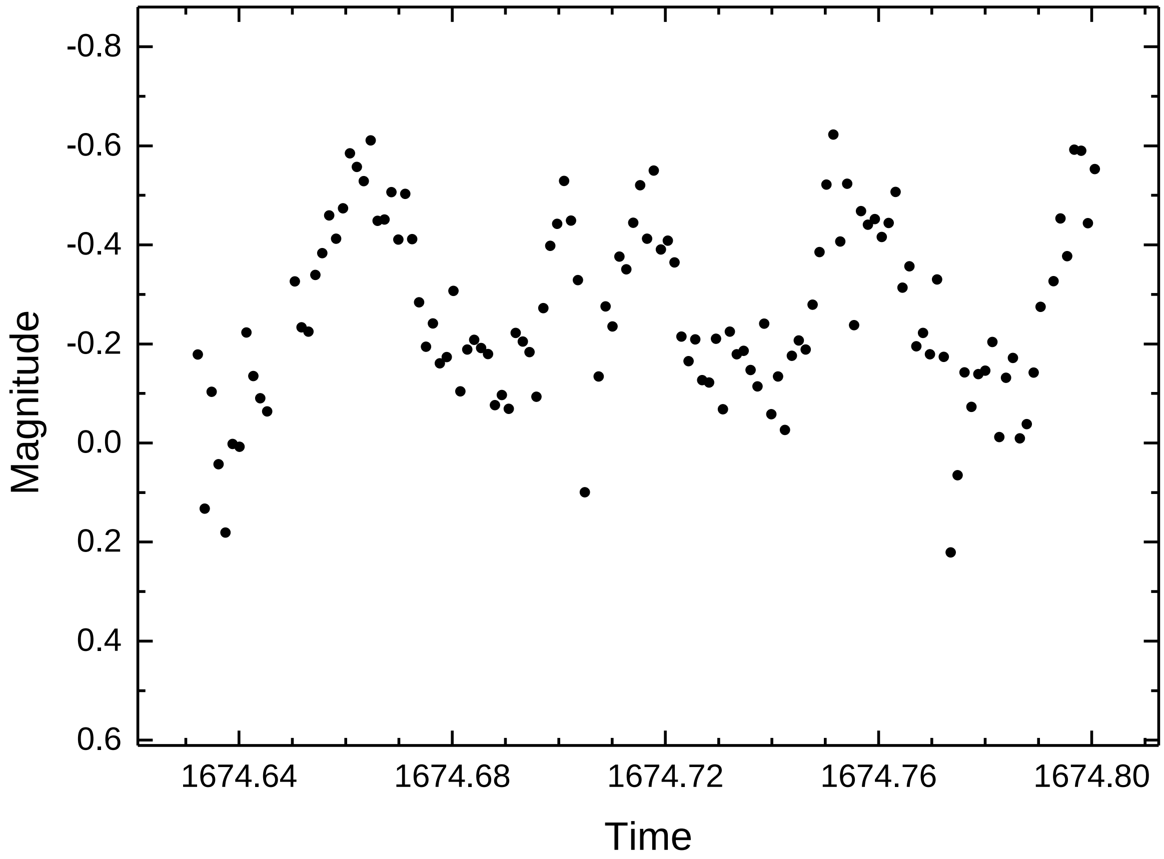
Magnitude
-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
1665
1670
1675
1680
1685
1690
1695
1700
1705
Time (HJD+2450000)

EX Hydrae

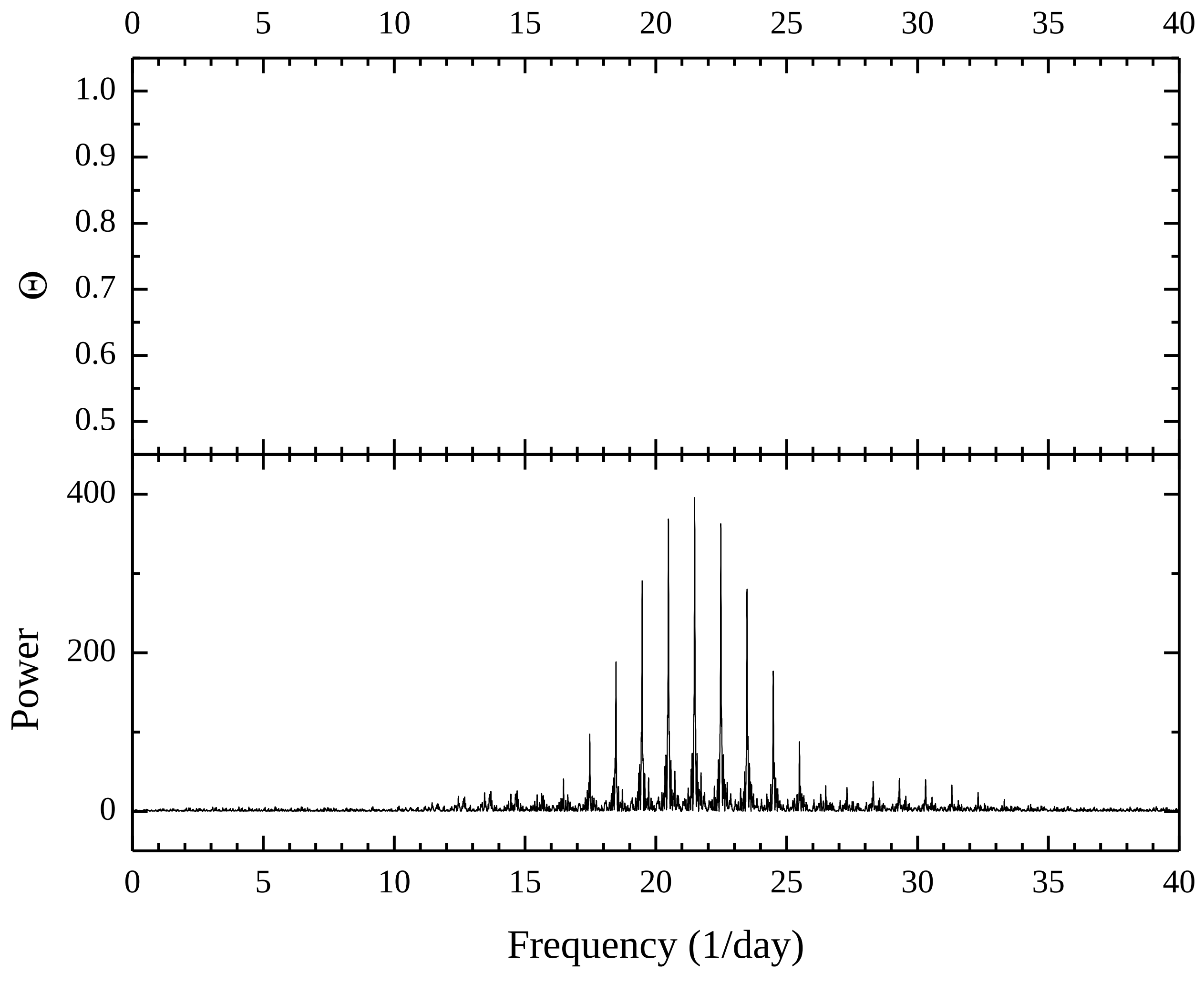


-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Magnitude
1673.64
1673.68
1673.72
1673.76
1673.80
Time

EX Hydrae

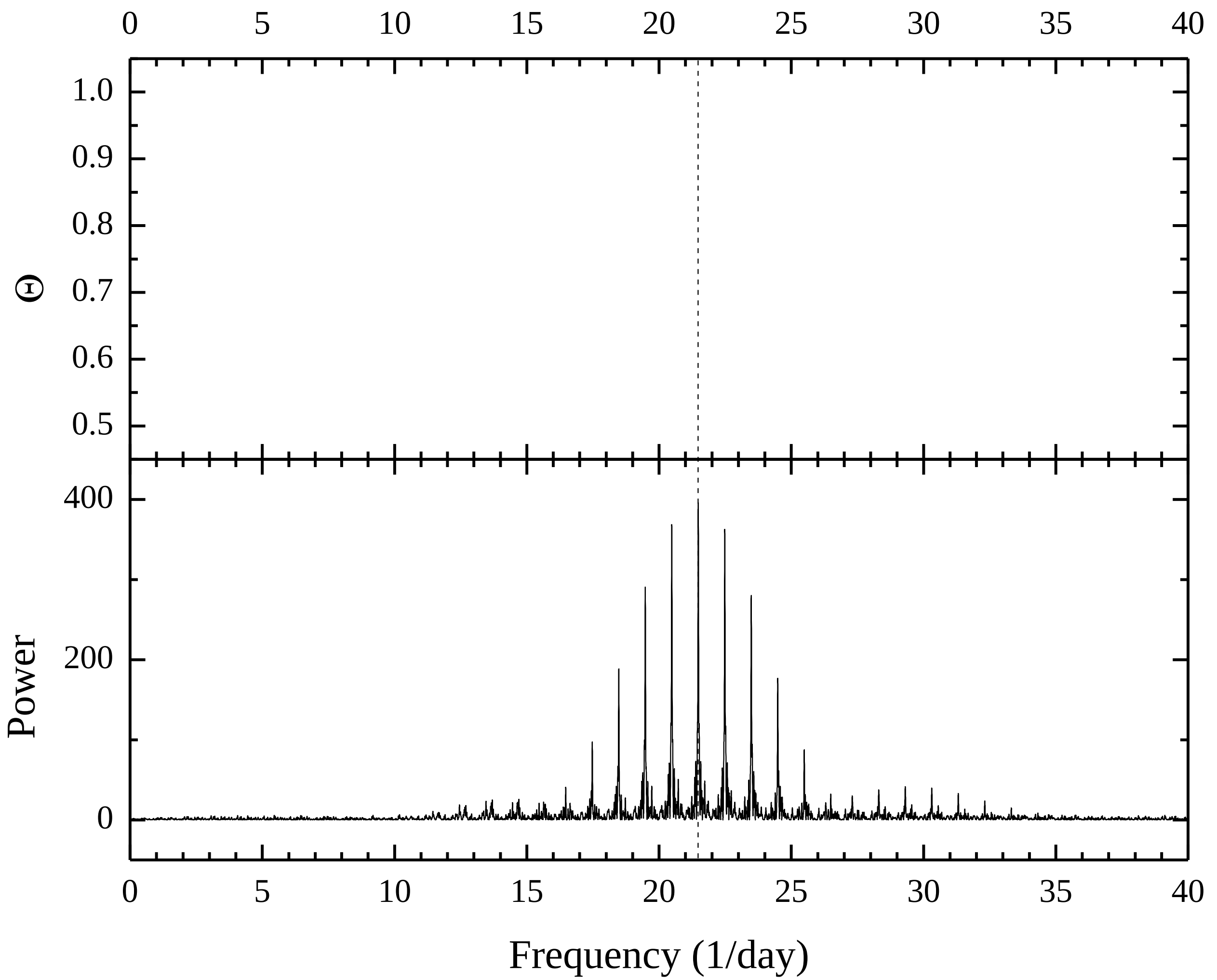

-0.8
-0.6
-0.4
-0.2
0.0
0.2
0.4
0.6
Magnitude
1674.64
1674.68
1674.72
1674.76
1674.80
Time

EX Hydrae



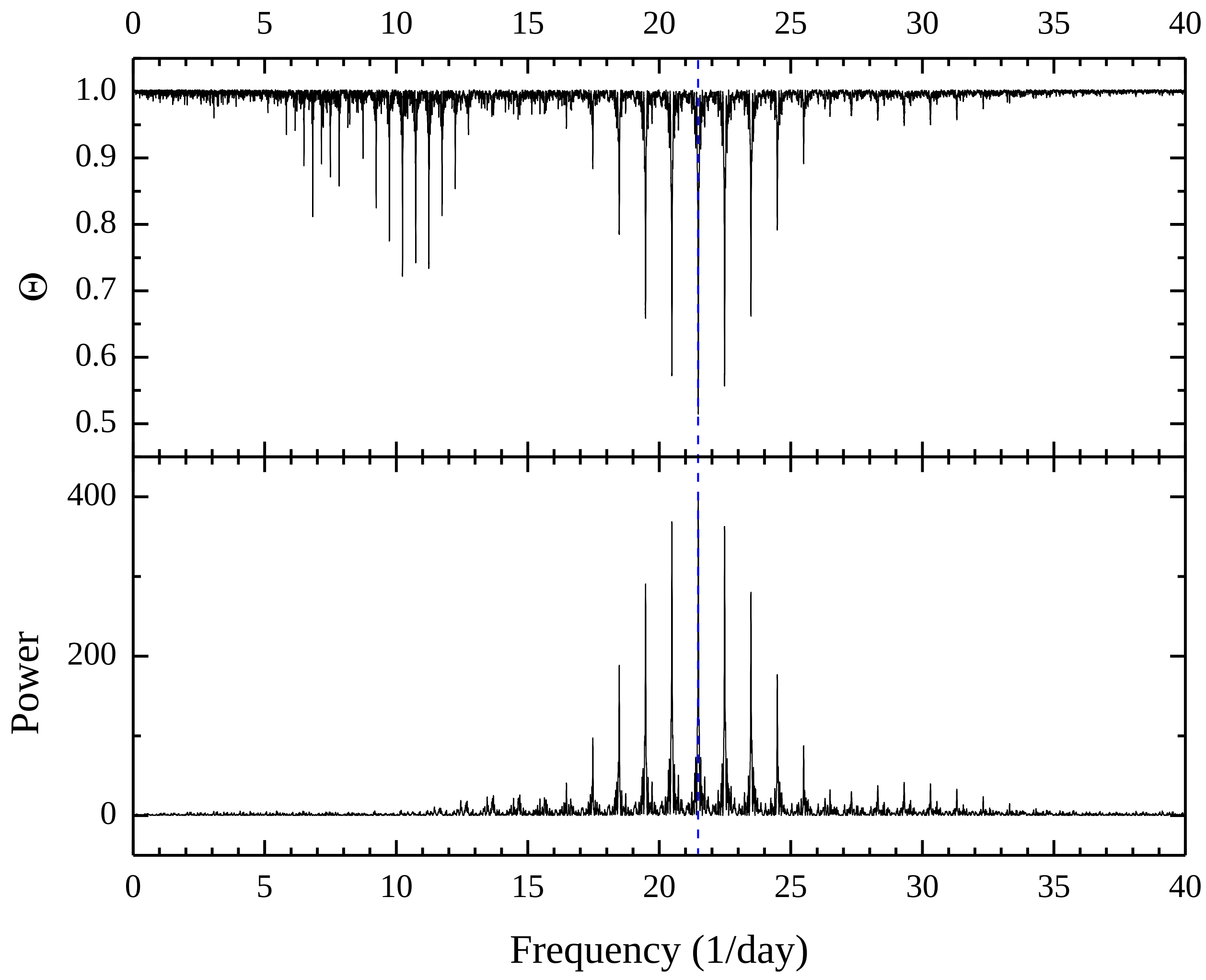
0 5 10 15 20 25 30 35 40
1.0
0.9
0.8
0.7
0.6
0.5
Θ
400
200
0
Power
Frequency (1/day)

EX Hydrae



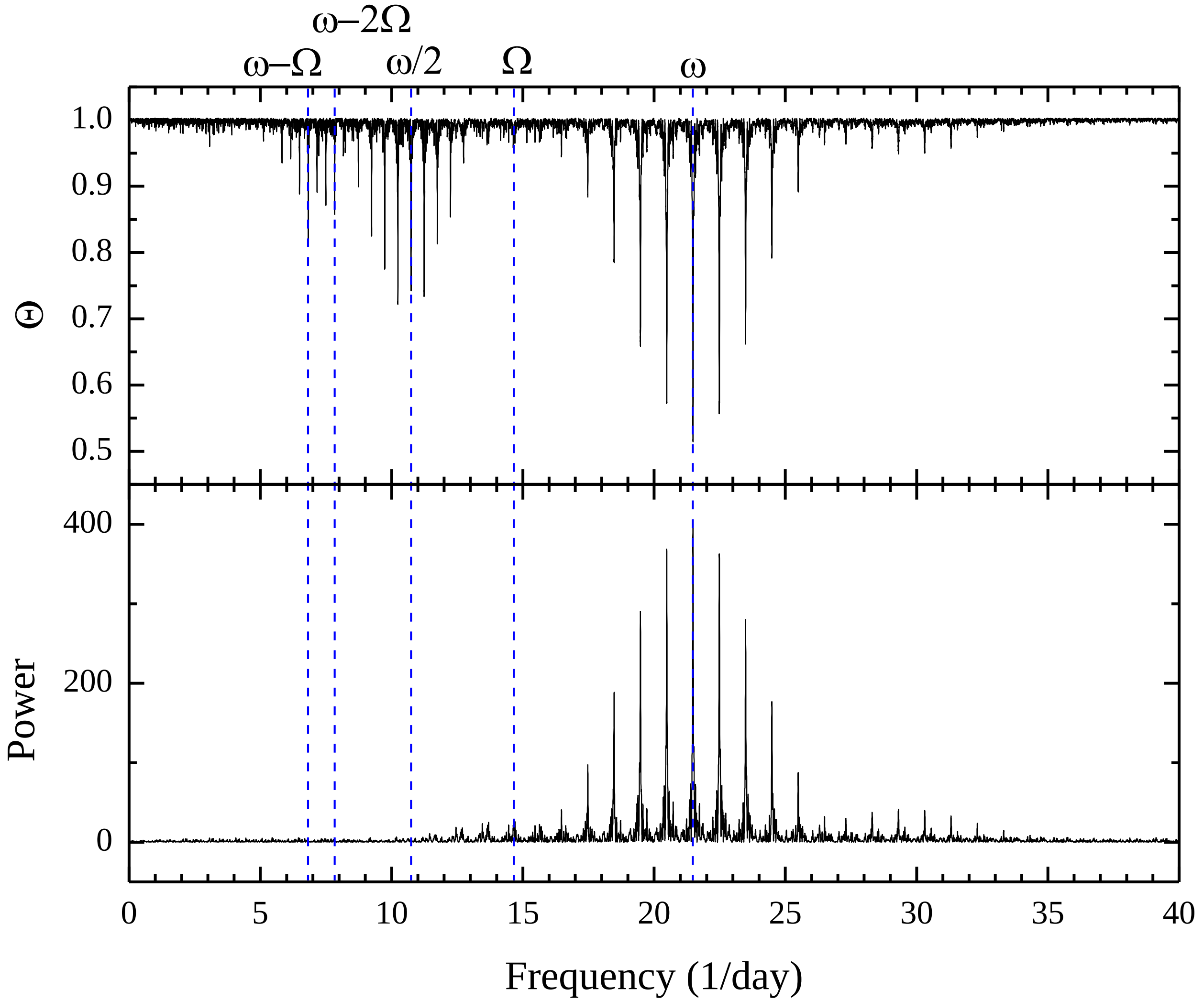
0 5 10 15 20 25 30 35 40
1.0
0.9
0.8
0.7
0.6
0.5
Θ
400
200
0
Power
Frequency (1/day)

EX Hydrae



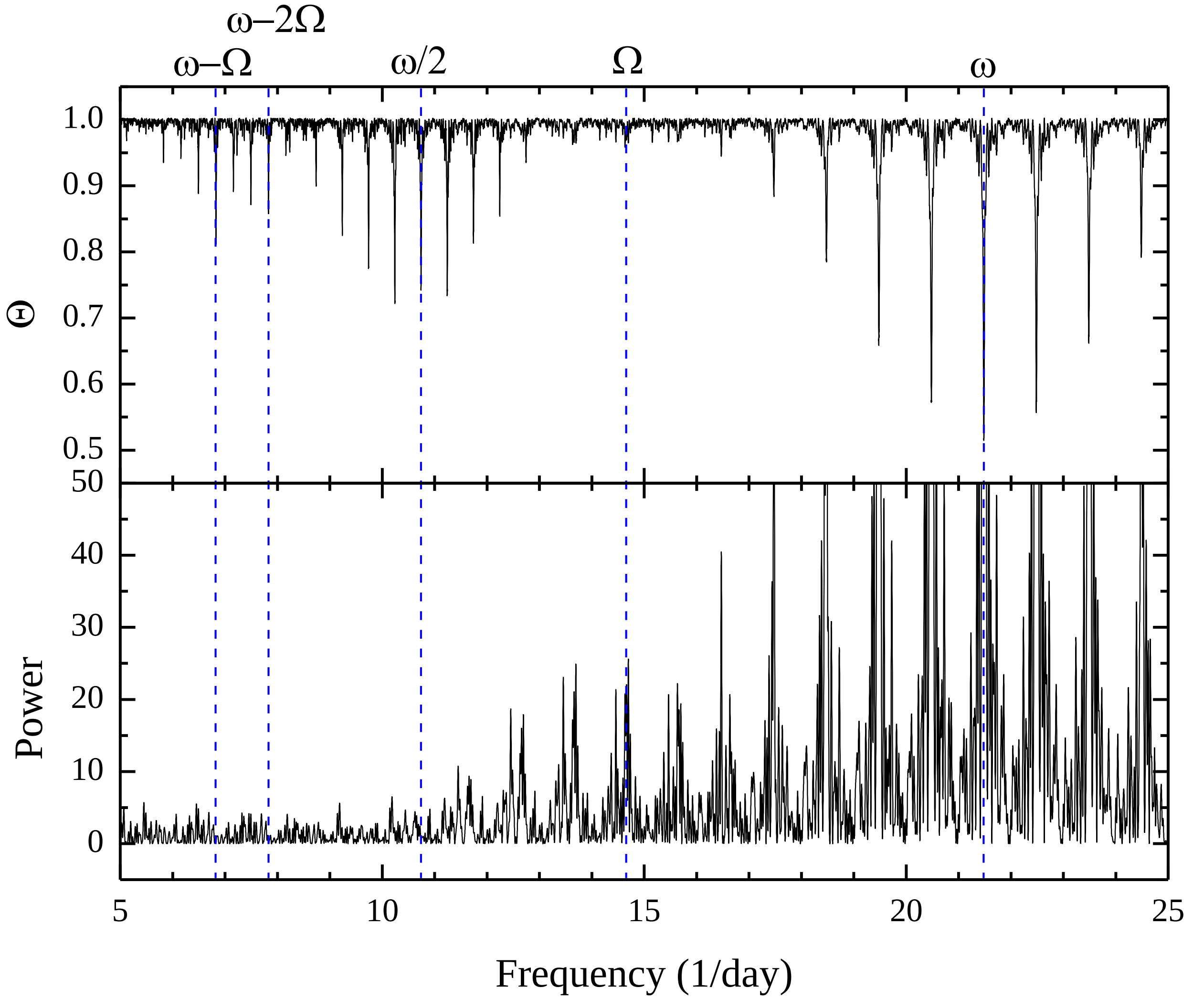
0 5 10 15 20 25 30 35 40
1.0
0.9
0.8
0.7
0.6
0.5
Θ
400
200
0
Power
Frequency (1/day)

EX Hydrae



ω–2Ω
ω–Ω
ω/2
Ω
ω
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



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.