Frequency Analysis Mapping On Unusual Samplings

F. Mignard

OCA/ Cassiopée

Summary

- Statement of the problem
- Objectives and principles of Famous
- Performances
- Application to variable stars
- Conclusions

Times Series

- Times series are ubiquitous in observational science
  - astronomy, geophysics, meteorology, oceanography
  - sociology, demography
  - economy and finance
- They are analysed to find synthetic description
  - trends, periodic pattern, quasi-periodic signatures
- Fourier analysis has been a standard tool for many years
  - well adapted to regularly sampled signal
  - but plagued with aliasing effect

Regular Sampling

- Problems with regular samplings
  - periodic structure in the frequency space
    - aliasing
    - infinitely many replica of a spectral line
    - assumption needed to lift the degeneracy
- Advantages of regular sampling
  - no spurious lines outside the true lines
  - $<\exp^{i2\pi\nu t}, \exp^{i2\pi\nu' t}> \sim 0$ if $\nu - \nu' \neq k/\tau$ :: orthogonality condition
  - with one spectrum one can have all the spectral information

Irregular Sampling

- No definition of what 'irregular' means
  - continuous pattern from fully regular to fully irregular
  - random sampling is much better than 'structured irregular'
- Problems with irregular samplings
  - many ghost lines linked to the true lines
  - $\langle \exp^{i2\pi\nu t}, \exp^{i2\pi\nu' t} \rangle \neq 0$ for many pairs (ν, ν')
    - lack of orthogonality condition
  - with one spectrum one cannot extract the full spectral information
- Advantages of irregular samplings
  - no periodic structure in the frequency space
    - each spectral line appears once over a large frequency range
    - in principle no assumption needed to find the correct line

FAMOUS : Background and overview

- FAMOUS makes the decomposition of a time series as :

$$\phi(t) = c_0 + \sum c_k \cos(2\pi\nu_k t) + s_k \sin(2\pi\nu_k t), \quad k = 1, \cdots, n$$

- c_k and s_k are constant or time polynomials
- The frequencies ν_k are also solved for
- The spectral lines are orthogonal on the sampling (as much as possible)
- FAMOUS never uses a FFT
- It can be used for any kind of time sampling
- It has a built-in system to determine the best sampling in frequency
- It detects uniform sampling and goes into dedicated procedures
- It can search for periodic functions with $\nu_k = k\nu_1$
- It estimates the level of significance of the periods and amplitudes
- It generates a detailed output + all the power spectrums and residuals

Application to Gaia on-board time

	period d	amplitude μs	phase °		
1	365.26401	1664.74	267.373	sidereal year	n_3
2	177.56628	121.74	268.988	lissajous period	σ
3	398.88244	22.63	212.608	synodic jupiter	n_3-n_5
4	182.62961	13.83	264.895	six months	2n_3
5	4333.41190	4.76	238.922	sideral jupiter	n_5
6	378.09968	4.63	18.412	synodic saturn	n_3-n_6
7	10751.37900	2.28	349.510	sideral saturn	n_6
8	345.55283	1.33	272.311	synodic lissajous	σ-n_3
9	291.95491	1.28	76.969	2*synodic venus	2*(n_2-n_3)
10	583.94321	1.13	82.919	synodic venus	n_2-n_3
11	439.32954	1.01	250.119		n_3-2*n_5
12	199.44473	0.80	157.509	2 synodic jupiter	2*(n_3-n_5)
13	119.48292	0.70	266.316	sun+lissajous	σ + n_3
14	1454.84510	0.62	246.329		2*n_2-3*n_3
15	369.65100	0.49	192.786	synodic Uranus	n_3-n_7
16	367.47181	0.46	224.343	synodic Neptune	n_3-n_8

Standard model for FAMOUS

When k frequencies have been identified one has the model :

$$\psi(t) = c_0 + \sum_{1}^{k} c_i \cos(2\pi \nu_i^0 t) + s_i \sin(2\pi \nu_i^0 t)$$

where

$$c_0, c_1, \dots, c_k \quad \text{and} \quad s_0, s_1, \dots, s_k \quad \text{are:}$$

constants or polynomial of time :

$$C_i = a_i^0 + a_i^1 t + a_i^2 t^2 + \cdots + a_i^p t^p$$

$$S_i = b_i^0 + b_i^1 t + b_i^2 t^2 + \cdots + b_i^p t^p$$

where p = p(i) : degree selected for each frequency

Solution with k frequencies

When k frequencies have been identified one has the model :

$$(a_i^r, b_i^r, \nu_i) :: \text{best least-squares fit} \Rightarrow \min \left|S(t) - \psi(t)\right|^2$$

This is a non-linear least-squares very sensitive to the starting values

Solved in two steps :

- SVD with $\nu_i = \nu^0_i$ and

- Levenberg-Marquardt minimisation with all the unknowns

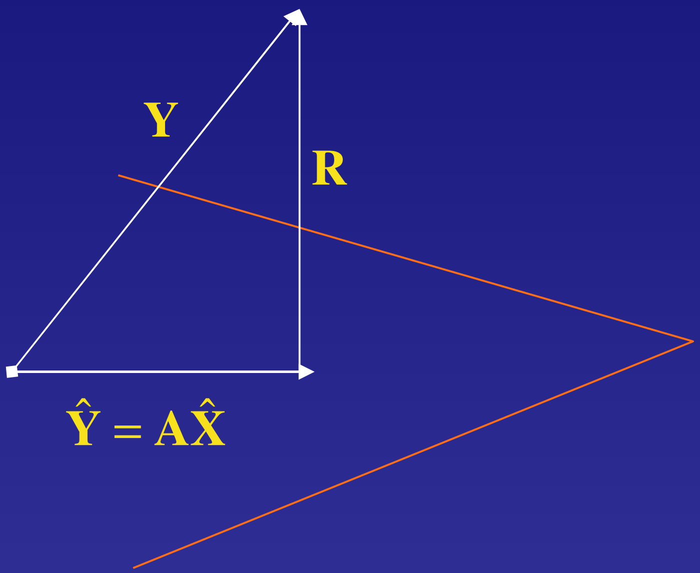
Result : best decomposition of S(t) on the model with *k* frequencies

Orthogonality for the (k+1)th frequency

Least squares solution

$$\mathbf{Y} \quad \mathbf{AX} :: \text{least squares fit } \hat{\mathbf{X}} \Rightarrow \min \left|\mathbf{Y} - \mathbf{AX}\right|^2$$

Characteristic property : $\mathbf{R} = \mathbf{Y} - \mathbf{A}\hat{\mathbf{X}} \perp \mathbf{A}\hat{\mathbf{X}}$



Any new line found in the residual signal in orthogonal to the previous lines

Main steps of FAMOUS

Input data, Settings

Sampling properties
frequency step and range

$\nu = 0$, trend
First residual R_0

$k = 1, n$

Periodogram on R_{k-1}
identify approximate ν_k
high resolution of the line : ν_k, a_k

$\varepsilon^{1/2}$

Non linear LSQ on k frequencies
Residual R_k

ε

Validation, statistics

- 8000 lines of code
- F90

~ 60 functions, subroutines

- cos and sin with recurrences

Settings of FAMOUS

- file_in — Input filename with the data y(x) as xx, yy on each record
- icolx — index of the column with the time data in file_in
- icoly — index of the column with the observations in file_in
- file_out — output filename
- **numfreq — search of at most numfreq lines**
- **flmulti — multiperiodic (true) or periodic (false) search in the signal.**
- **flauto — automatic search (true) or preset value (false) of the max and min frequencies**
- **frbeg — preset min frequency in preset mode**
- **frend — preset max frequency in preset mode**
- fltime — automatic determination (true) or preset value (false) of the time offset
- tzero — preset value of the origin of time if fltime = .false. e
- threshold — threshold in S/N to reject non significant lines (< threshold)
- **flplot — flag for the auxiliary files (power spectrum and remaining signal after k lines)**
- isprint — control of printouts (0 : limited to results, 1 : short report, 2 : detailed report)
- iresid — control the output of the residuals
- **fldunif — flag for the degree of the mixed terms (true : uniform degree for all terms)**
- idunif — degree if fldunif = .true.
- idegf(k) — degree of each line if fldunif = .false. , k=0,numfreq

Two key parameters

- Sampling step in the frequency domain
  - how to determine the optimum value
    - to find every significant line spectral resolution
    - to limit the amount of computation
  - uniform sampling in ν phases in arithmetic progression

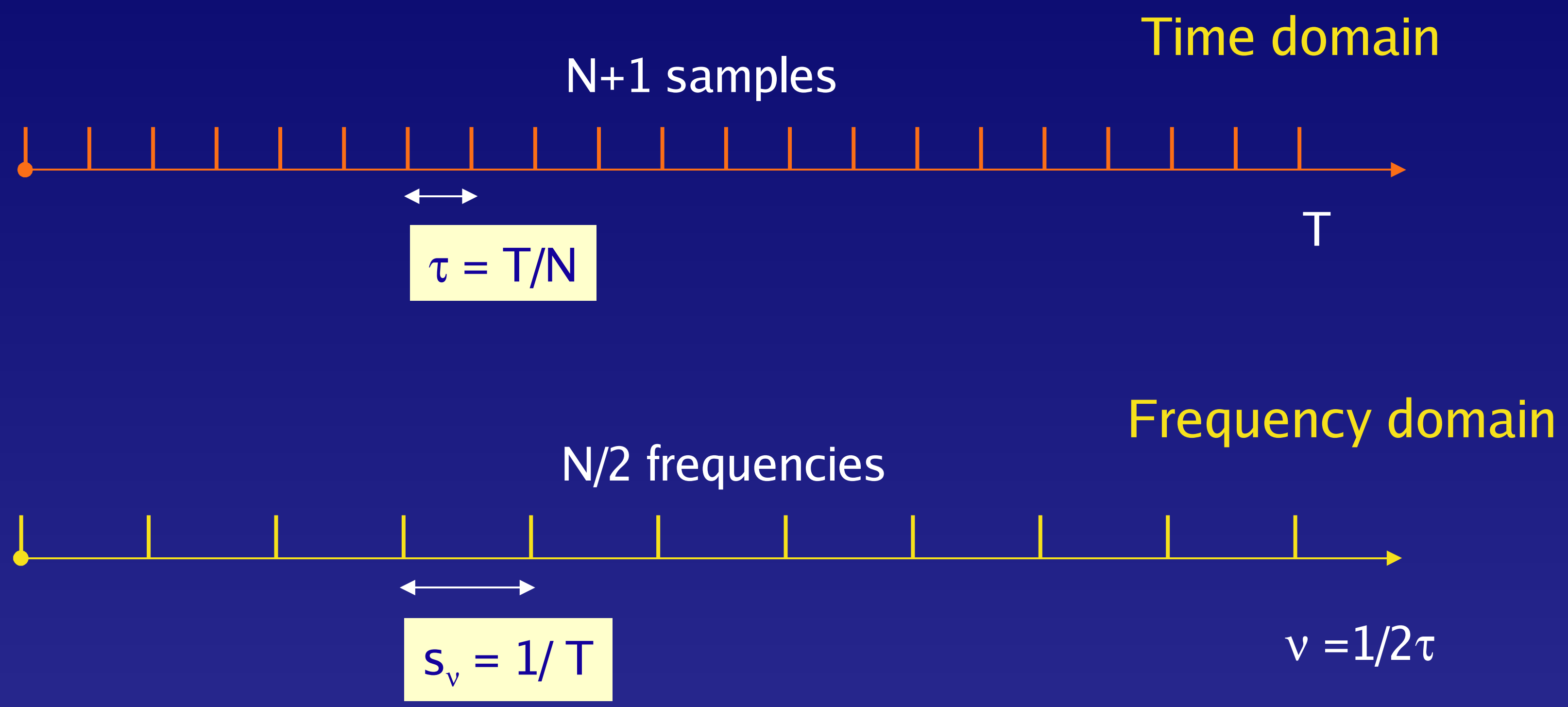
- Range of exploration in the frequency domain
  - big running penalty in searching in the high frequency range
  - easy rule for regular sampling
  - nothing obvious for irregular sampling
  - practical rules have been applied based on :
    - the average step in time domain
    - the smallest step in time domain

Step in the frequency domain

- FAMOUS needs a built-in system optimum for every sampling
- The power spectrum is a continuous function in ν
- The sampling must allow the reconstruction of $P(\nu)$
- There is no obvious and optimum choice
- The choice has important implications :
  - small steps increase the running time
  - large steps : not every line can be discovered

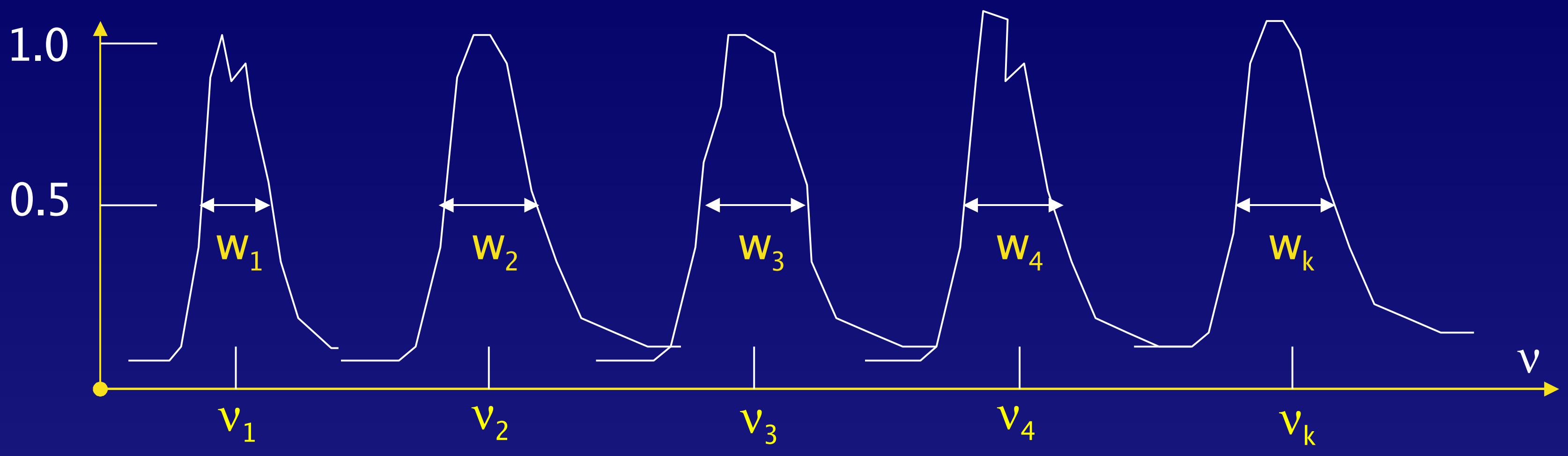
Step in the frequency domain

- In FFT with regular sampling : N+1 data points over T $s_\nu = 1/T$



- In DFT with regular sampling : more freedom on s_ν , but less efficiency

Sampling in frequencies



- High resolution of well chosen spectral lines
  - $s(t) = \cos(2\pi \nu t)$, $\nu = \nu_1, \nu_2, ,\ldots \nu_k$ in $\nu_{min} .. \nu_{max}$
- Statistics of the *k* widths at half-maximum
  - k = 1 for uniform time sampling
  - k ~ 15 for irregular sampling
- Several protections against peculiar line shapes
- Then $s_\nu \sim \langle w \rangle / 6$
- Resolution good enough to go through all the lines

Largest solvable frequency

- The trickiest problem met during development
- Related to the generalisation of the Nyquist frequency
- Relatively well founded solution for uniform sampling
  - ν_{max} = Nyquist frequency or multiple
- No natural maximum for irregular sampling
  - inverse of the smallest, average, median … interval ?
- Practical solution adopted for FAMOUS :
  - Either :
    - ν_{max} user provided recommended solution
  - Otherwise : search of a representative timestep
    - statistics of the 2-point intervals in the time domain
    - then $\tau \sim$ 2nd decile and $\nu_{max} = 1/2\tau$

Performances

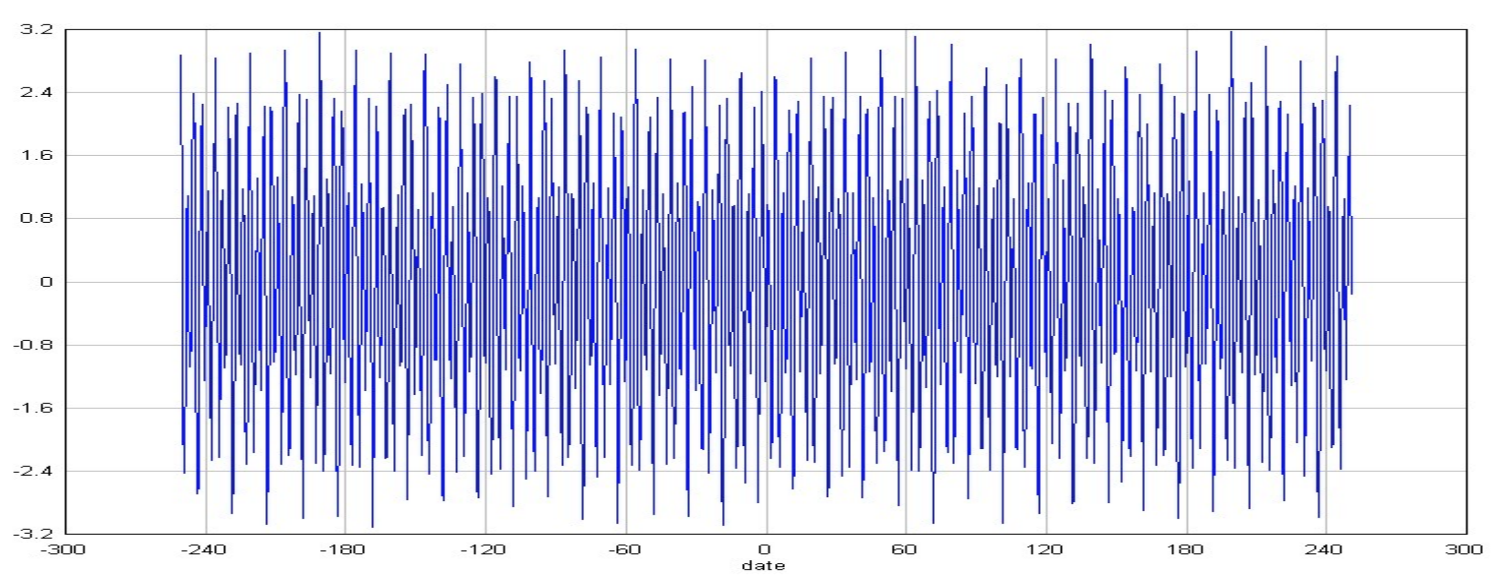
Simulation

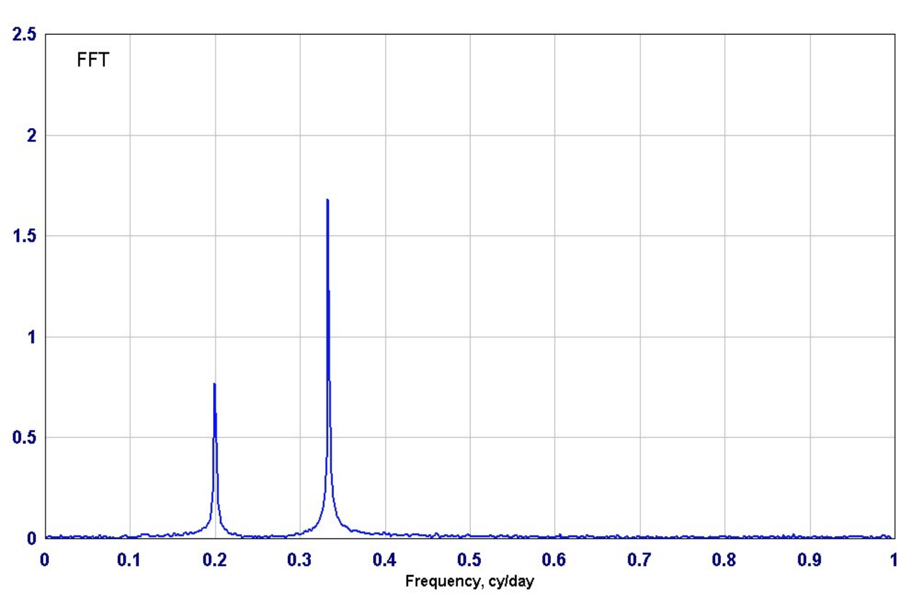
- The simulation generates a periodic or multi-periodic signal
- Sampling can be regular or with some randomness
- $s(t_k) = 2\cos(2\pi/p1 * t_k) + \cos(2\pi/p2 * t_k)$
- Gaussian random noise with $\sigma = 0.1$
  - n = 1000 samples
  - P = 3, 5, … days
  - T = 500 days
  - $\tau = 0.5$ day (for regular sampling)
  - $1/\tau = 2$ cy/day
  - $N_y. = 1/2\tau = 1$ cy/day

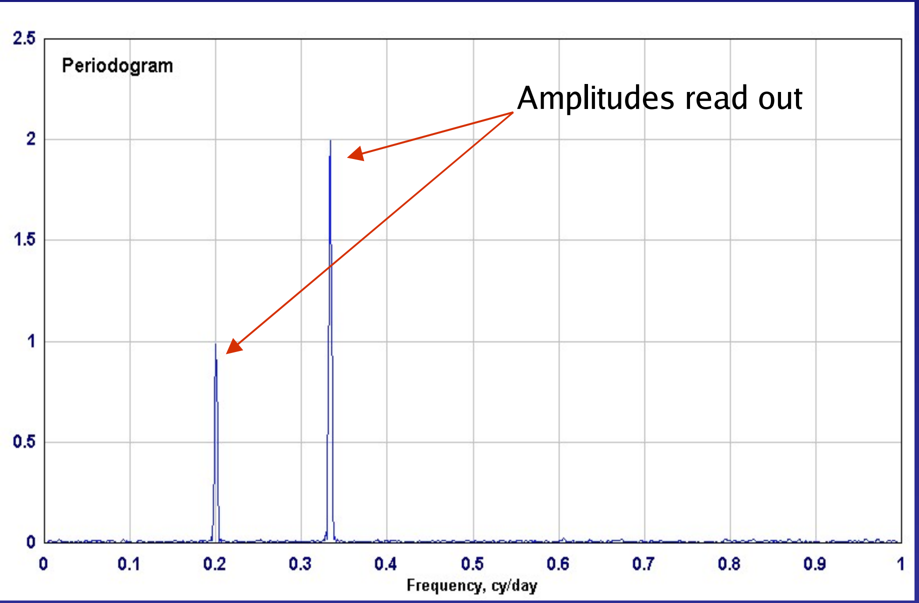
Examples

p_1 : 3 d
p_2 : 5 d

τ = 0.5 d



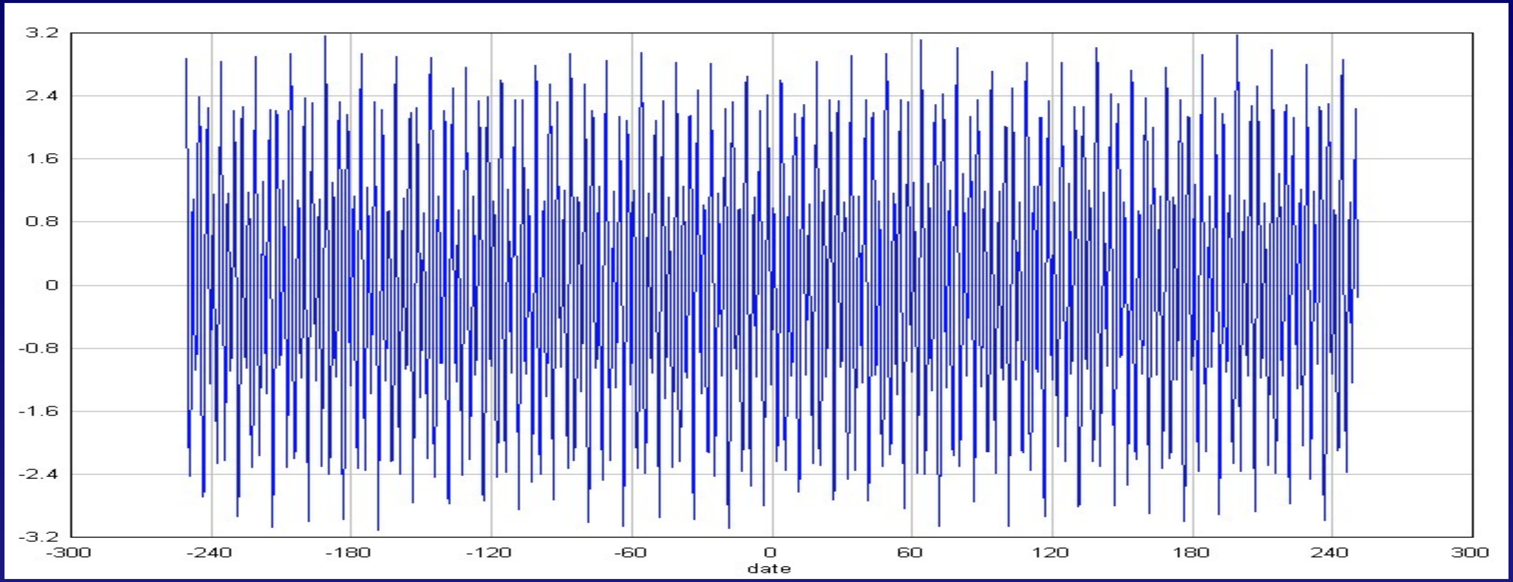


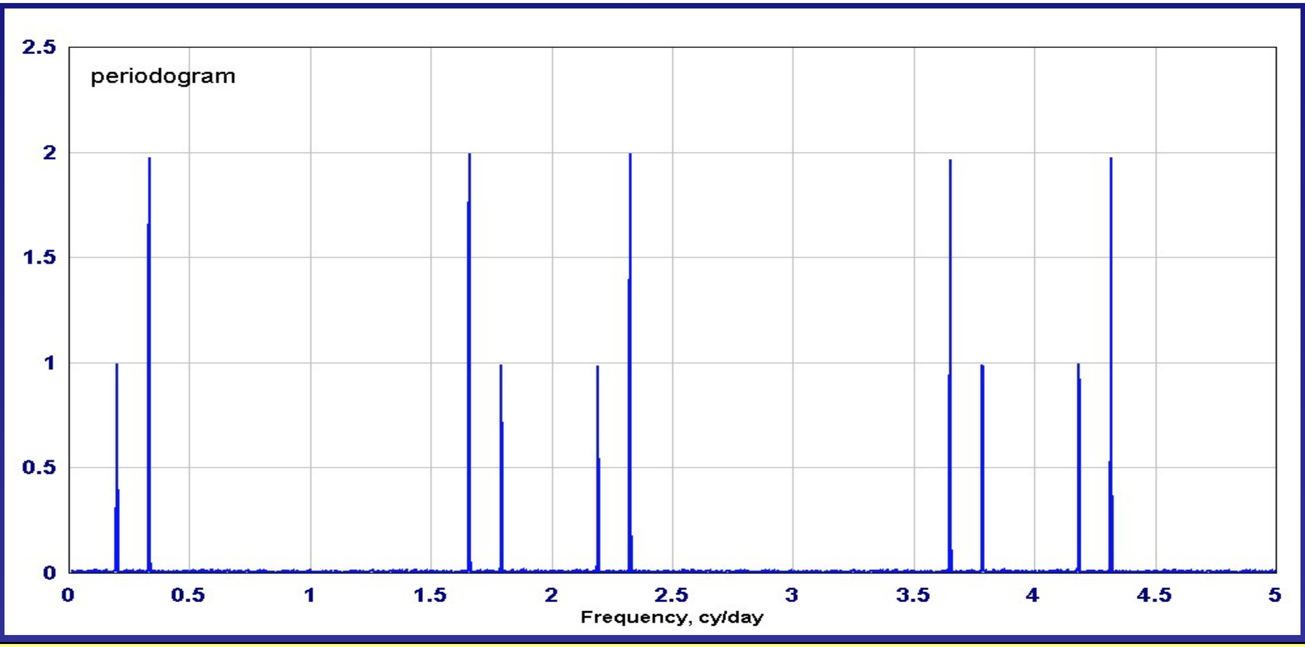


Examples

p_1 : 3 d
p_2 : 5 d

τ = 0.5 d



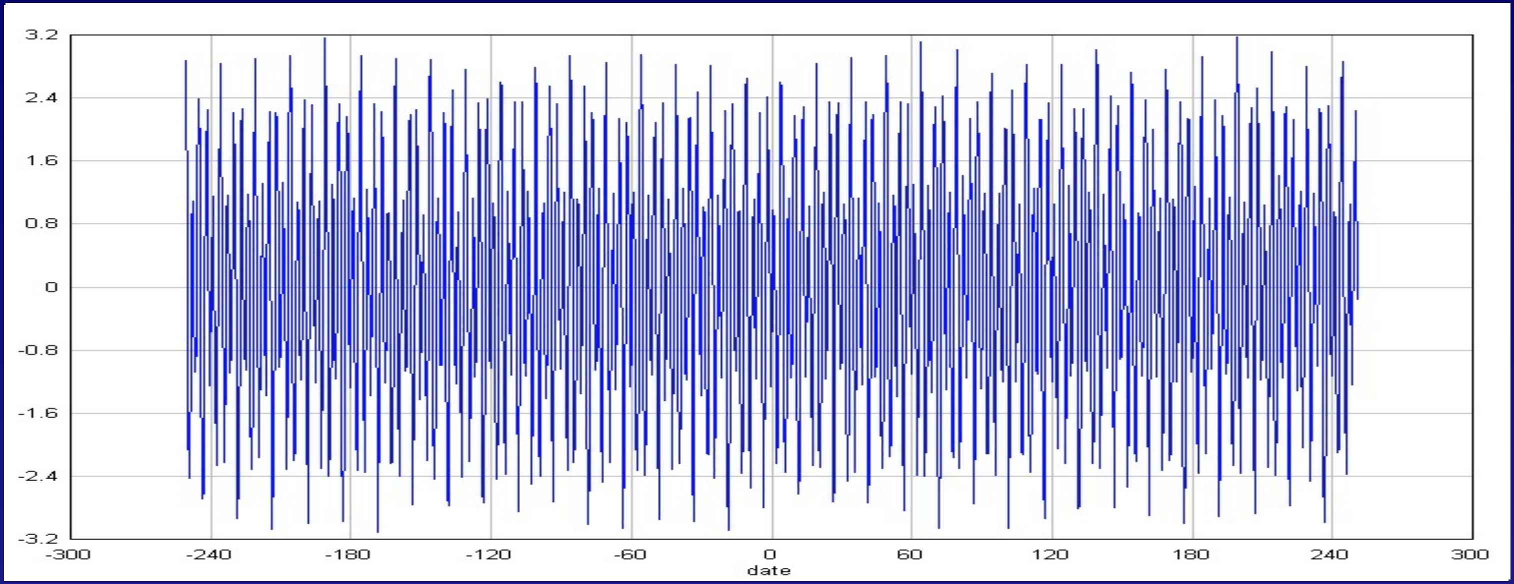


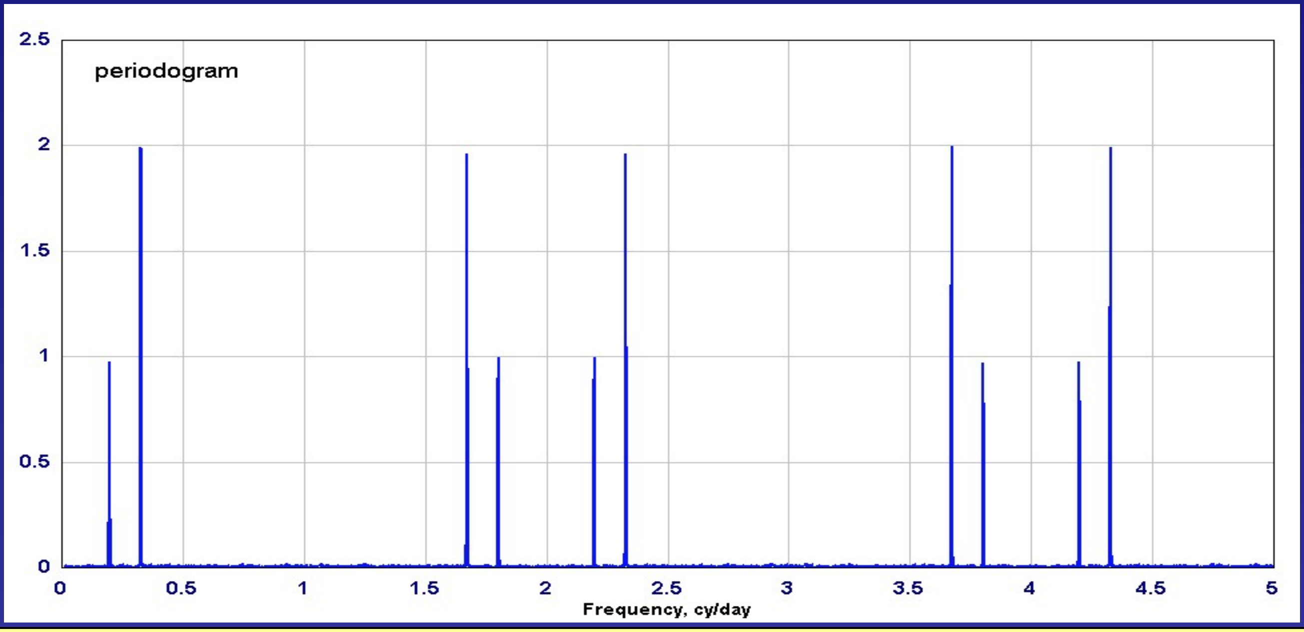
Examples

p_1 : 0.43 d
p_2 : 0.45 d

τ = 0.5 d

τ = 0.05 d





Examples : random sampling

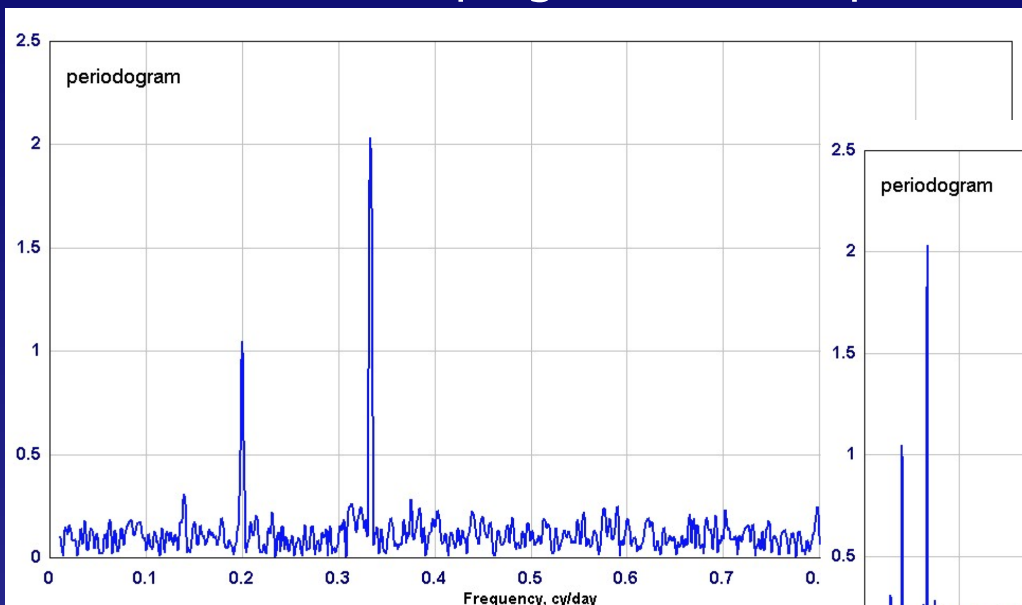
p_1 : 3 d

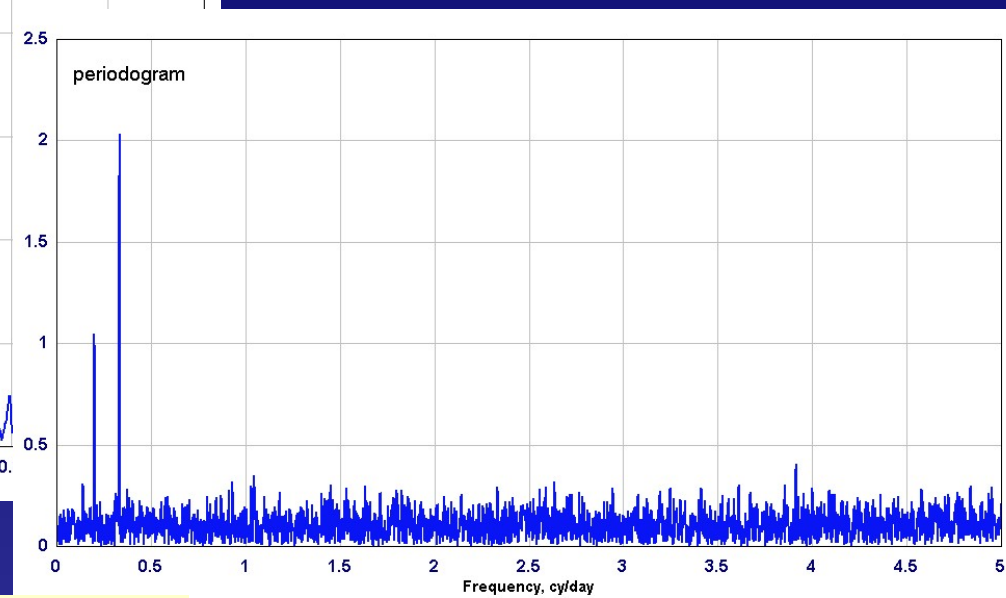
p_2 : 5 d

$$S(t_k) = 2\cos(2\pi t_k / P_1) + \cos(2\pi t_k / P_2)$$

$\langle\tau\rangle = 0.5$ d $\sigma = 0.1$

Uniform random sampling of 1000 data points over 500 days.





Periods and amplitudes found

2.99998 +/- 0.00002	1.995 +/- 0.005
4.99998 +/- 0.0001	1.005 +/- 0.005

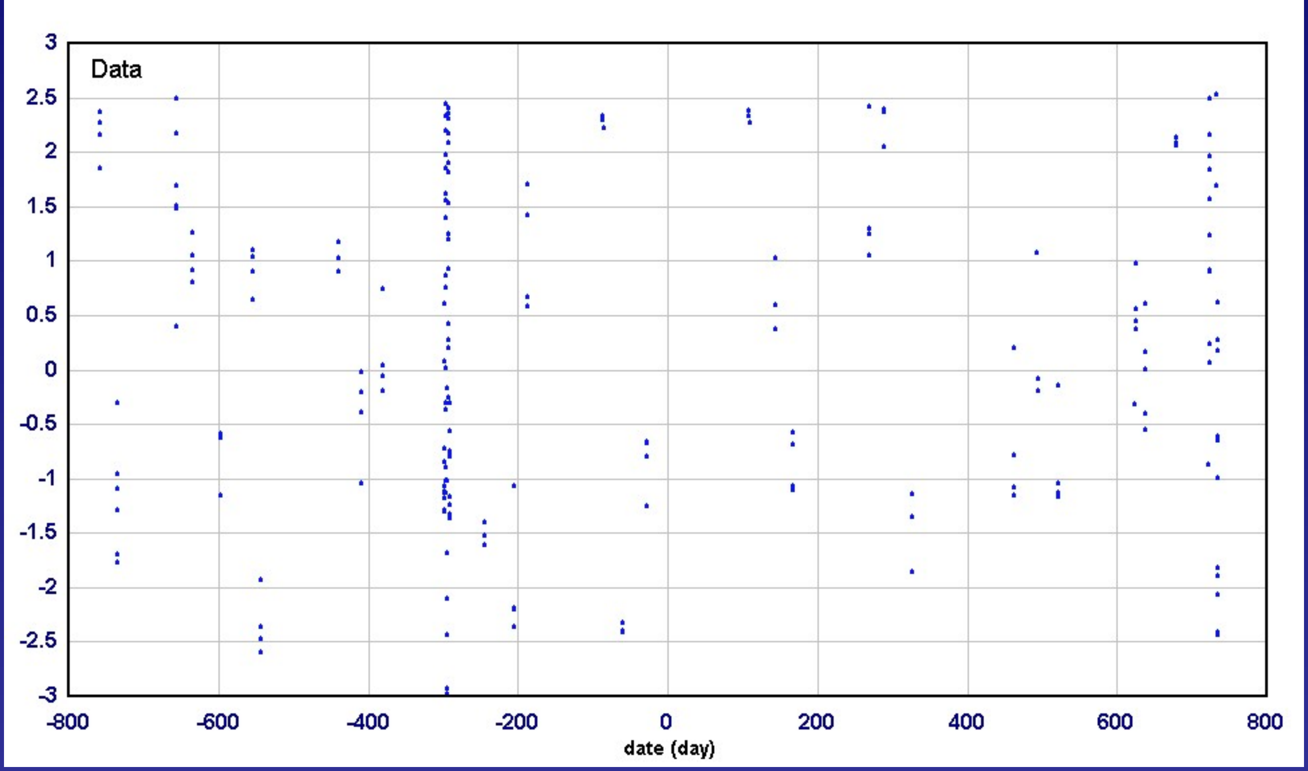
Examples : Gaia-like sampling

$$S(t) = \sum a_i \cos\left(2\pi \frac{t}{P_i}\right)$$

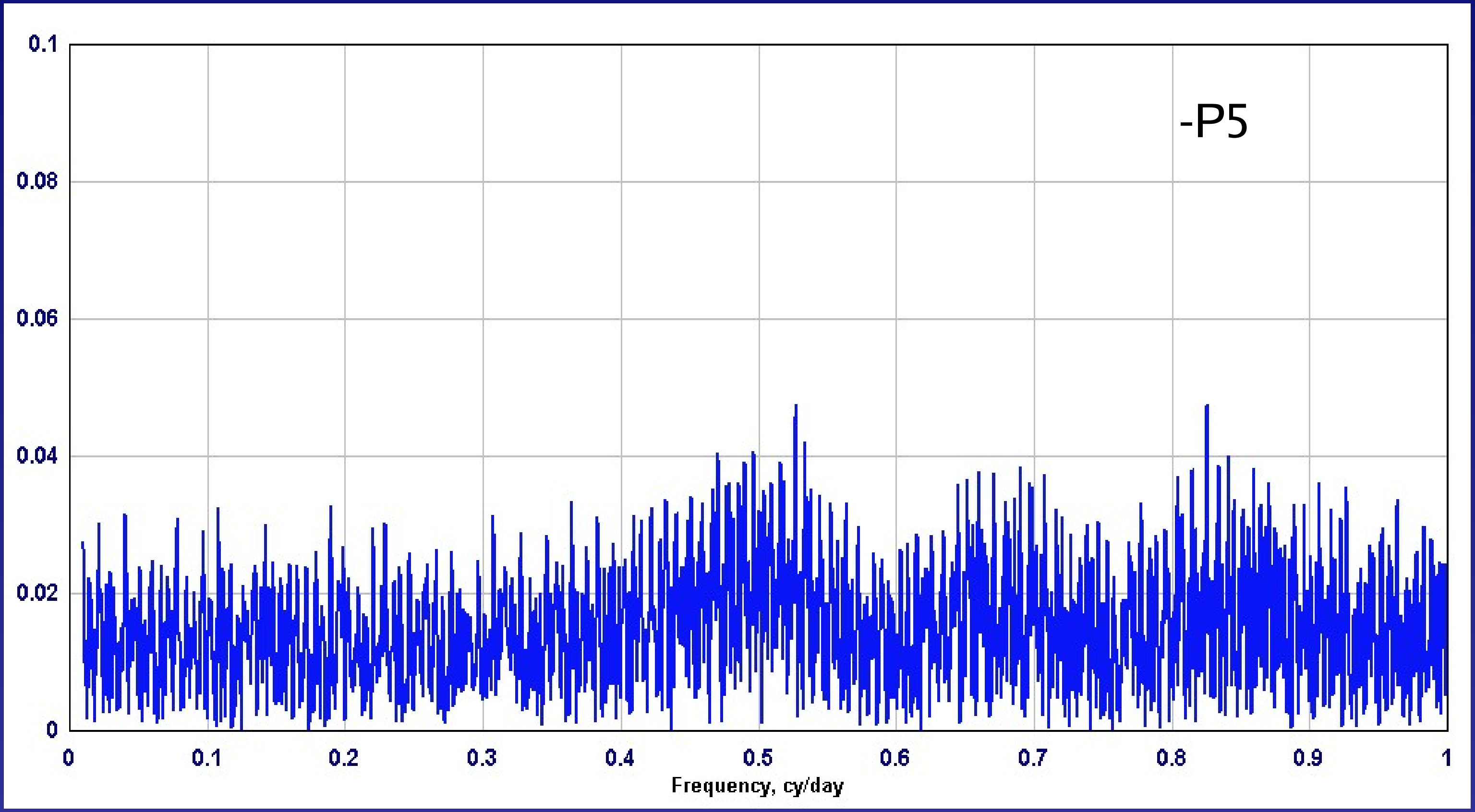
i	P	a
1	3	2
2	5	1.5
3	1	1
4	7	0.5
5	20	0.2

$\sigma = 0.1$

220 samples over 1600 days.



i	P	a
1	3	2
2	5	1.5
3	1	1
4	7	0.5
5	20	0.2



Examples : Gaia-like sampling

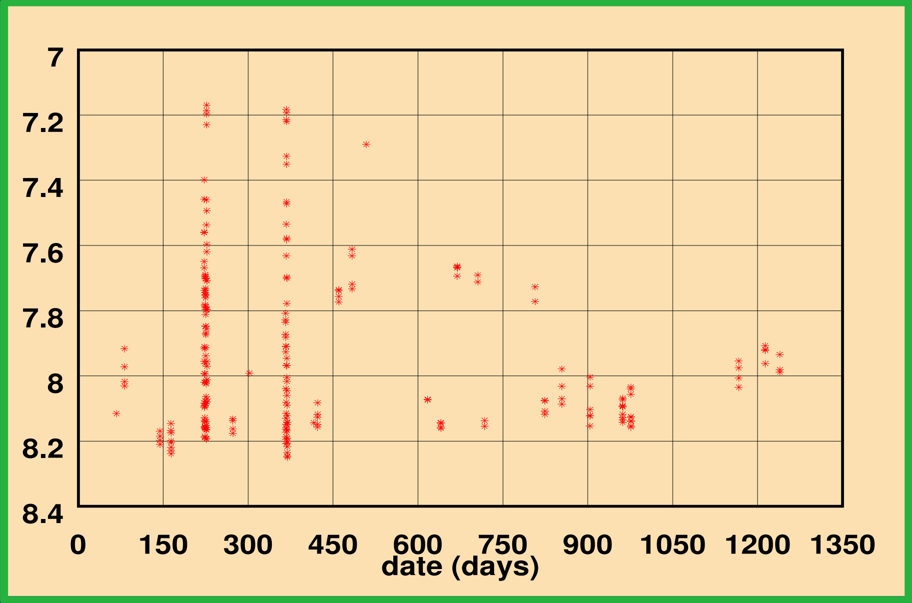
i	P	a
1	3	2
2	5	1.5
3	1	1
4	7	0.5
5	20	0.2

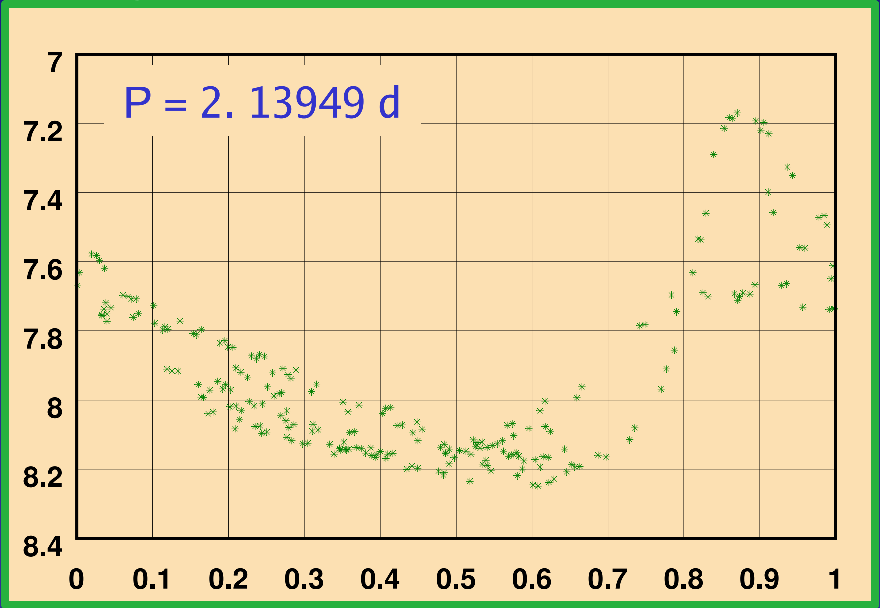
- Results from FAMOUS

Periods		amplitudes	
3	2.99998 +/- 0.00002	2	1.993 +/- 0.02
3	4.99995 +/- 0.00006	1.5	1.523 +/- 0.02
1	0.99999 +/- 0.00002	1.0	0.990 +/- 0.02
7	7.00008 +/- 0.0003	0.5	0.483 +/- 0.02
20	20.0083 +/- 0.008	0.2	0.187 +/- 0.02

2-mode Cepheids

- Hip 2085 = TU Cas
  - problem known for many years
  - large residuals in Hipparcos data with a single period
  - well visible in the folded light-curve
- FAMOUS can solve for several unrelated periods
  - $p_1 = 2.1395$, $p_2 = 1.5186$, $p_3 = 1.1753$ days
  - $a_1 = 0.316$, $a_2 = 0.086$, $a_3 = 0.074$ mag

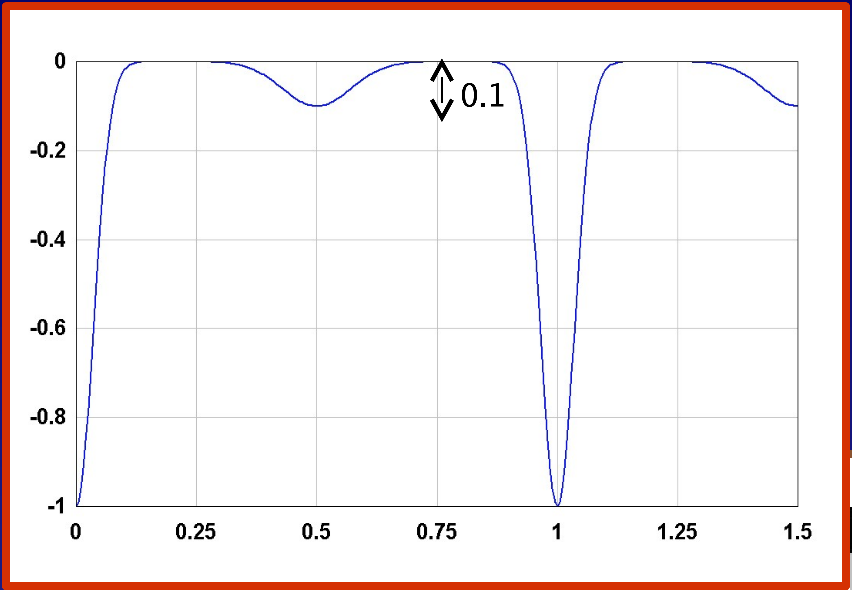




Famous for periodic signals

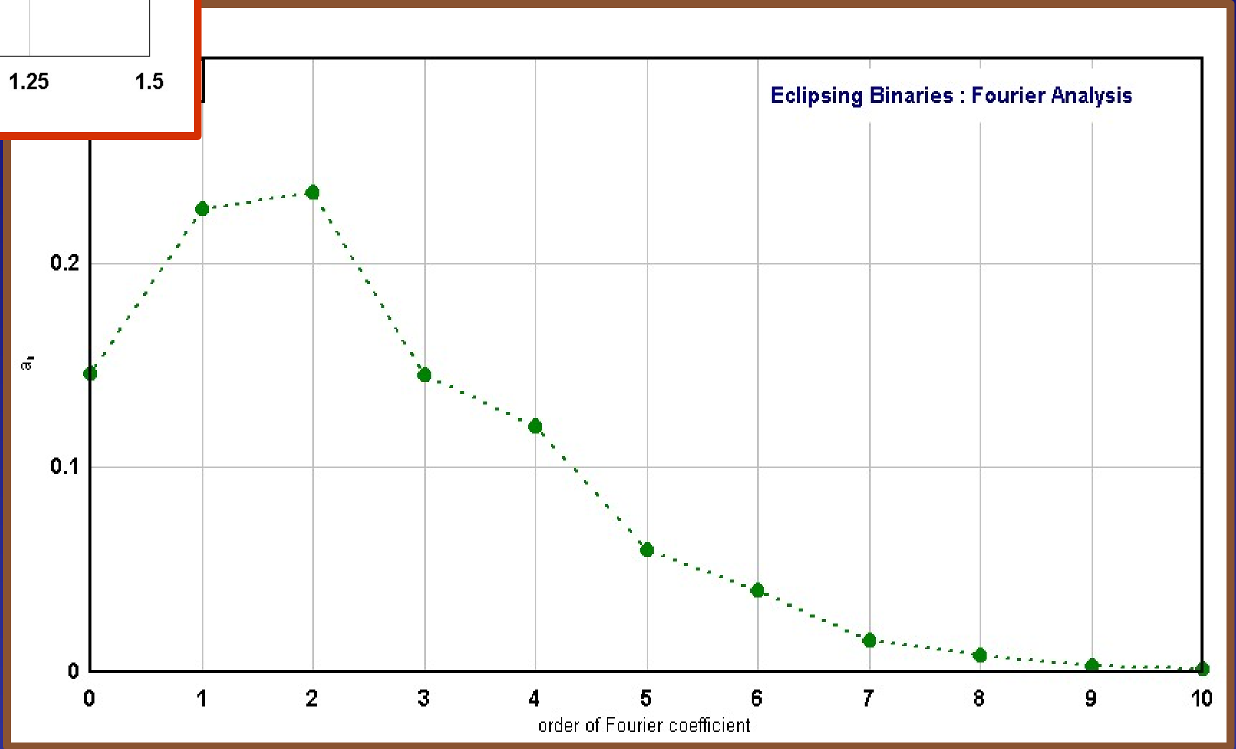
- FAMOUS is not specifically designed for periodic signal
- However one can search only one frequency
  - in most cases of interest this gives the period
  - but : the largest amplitude of a periodic signal can be an harmonic
    - therefore a submutiple of the period is found
- With two frequencies : $\nu_2/\nu_1 = 2$ or 0. 5 tests
- This approach is generalised to locate the fundamental
  - search the first frequency (largest amplitude in the 1st periodogram)
  - search over a narrow bandwidth around $\nu_1/2$, $\nu_1/3$, $\nu_1/4$, $2\nu_1$, $3\nu_1$, ...
  - tests to select the fundamental

Eclipsing Binaries I.

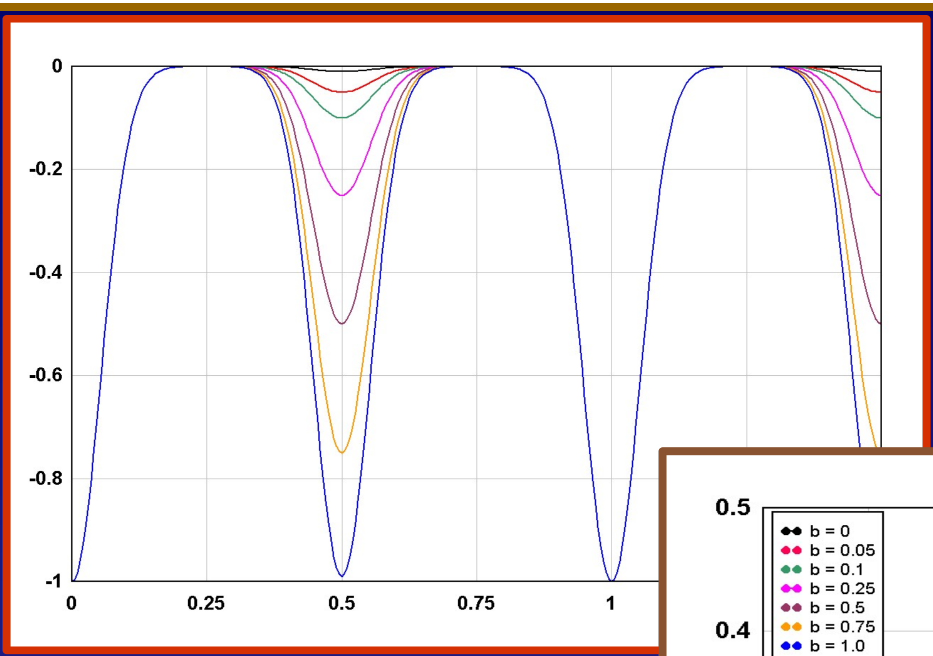


EA star with b/a = 0.1

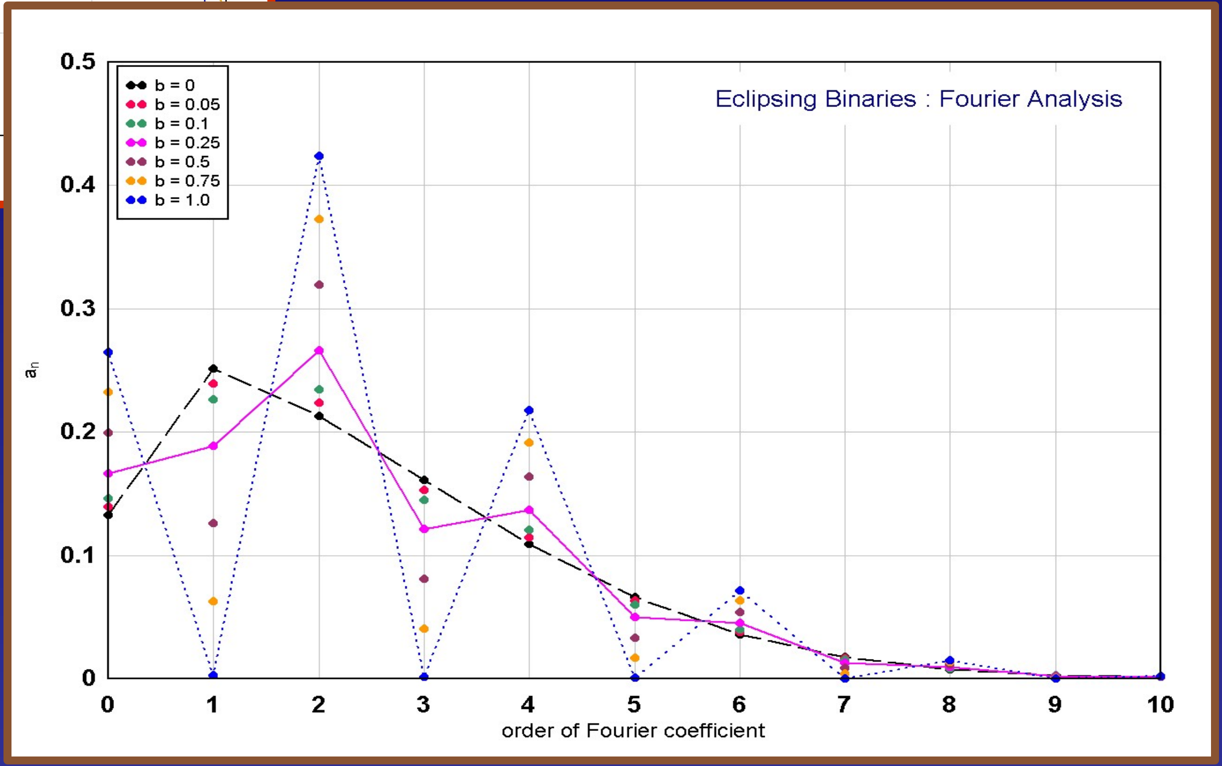
Fourier coefficients



Eclipsing Binaries II.



b/a = 0.0 to 1.0

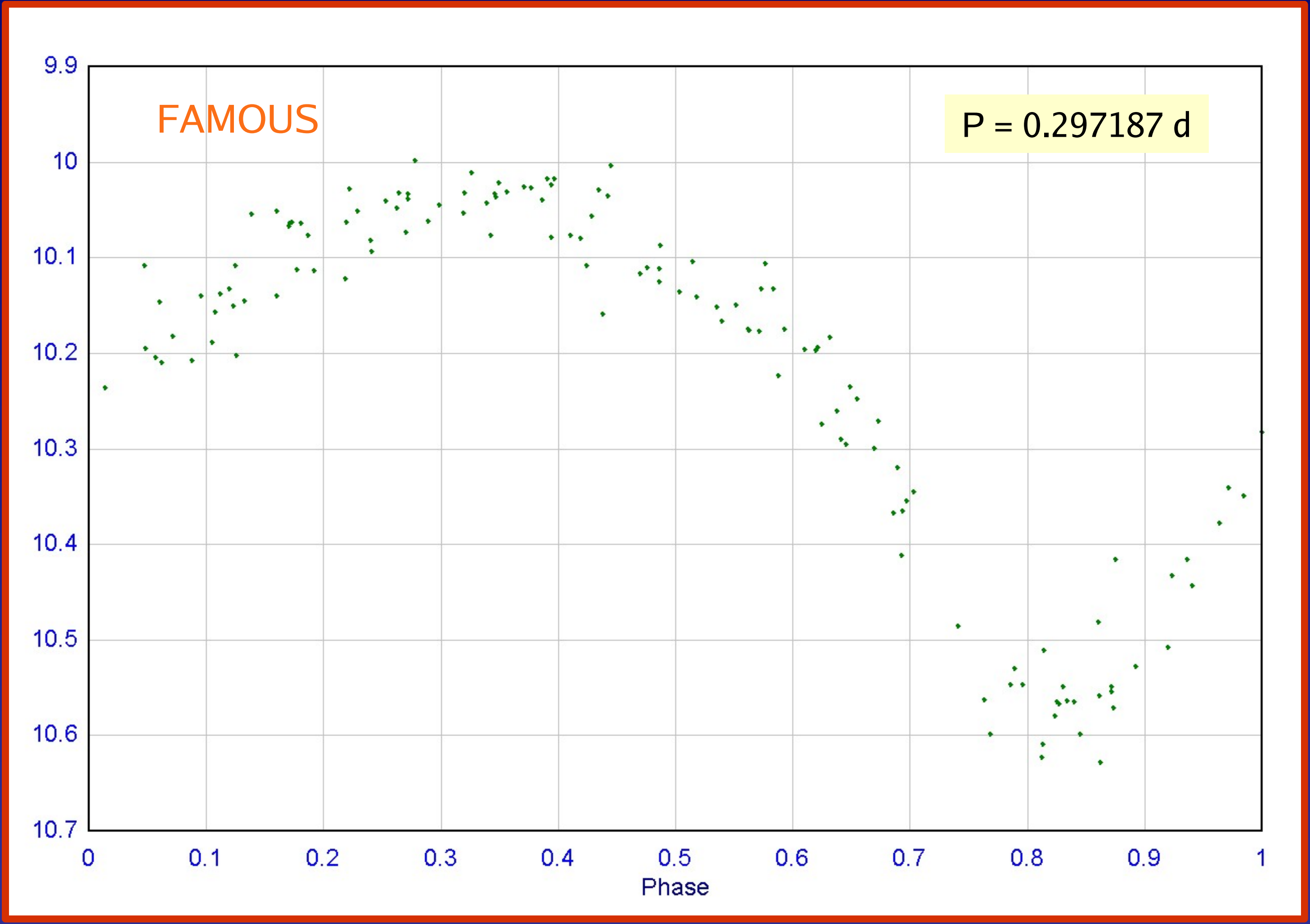


-The first harmonics gets larger from Ea to Ec

- the period found may be half the true period

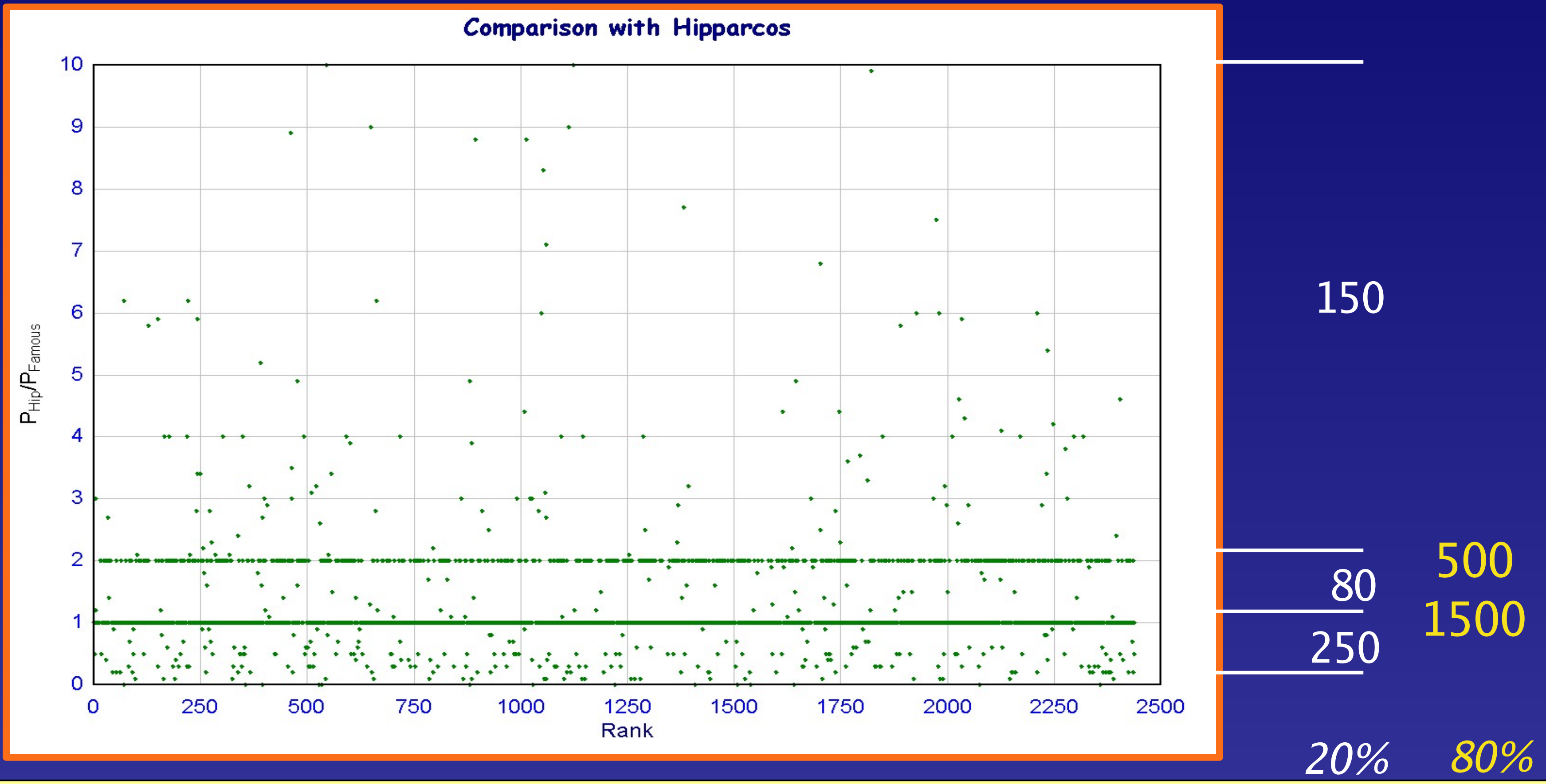
Eclipsing binary : period ?

- Hip 1387 = AQ Tuc



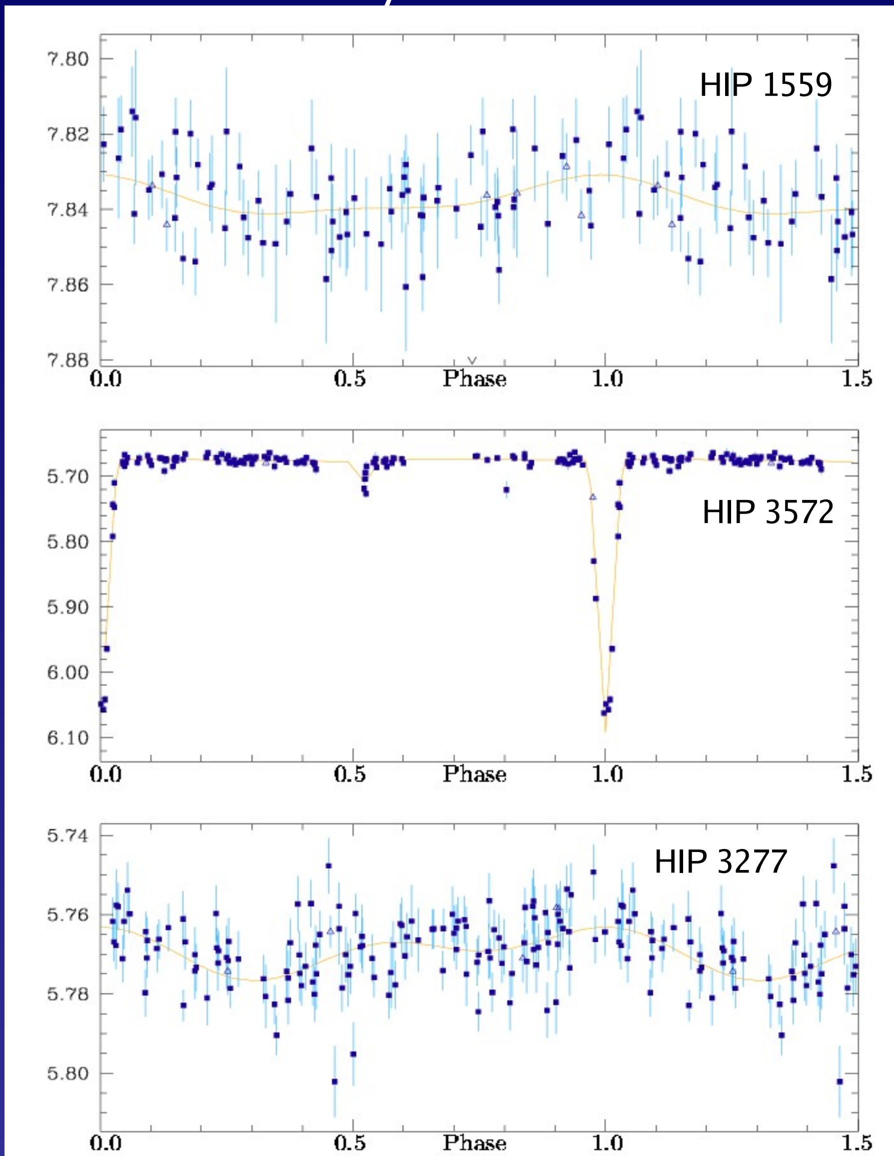
Global exploitation on Hipparcos data

- Run over the photometric data of the ~ 2500 periodic variables
  - periods larger than 2h searched (0 to 12 cy/day) key parameter
  - totally blind search
  - production of folded light-curves
  - running time on laptop : 450 s ~ 0.18 s/star

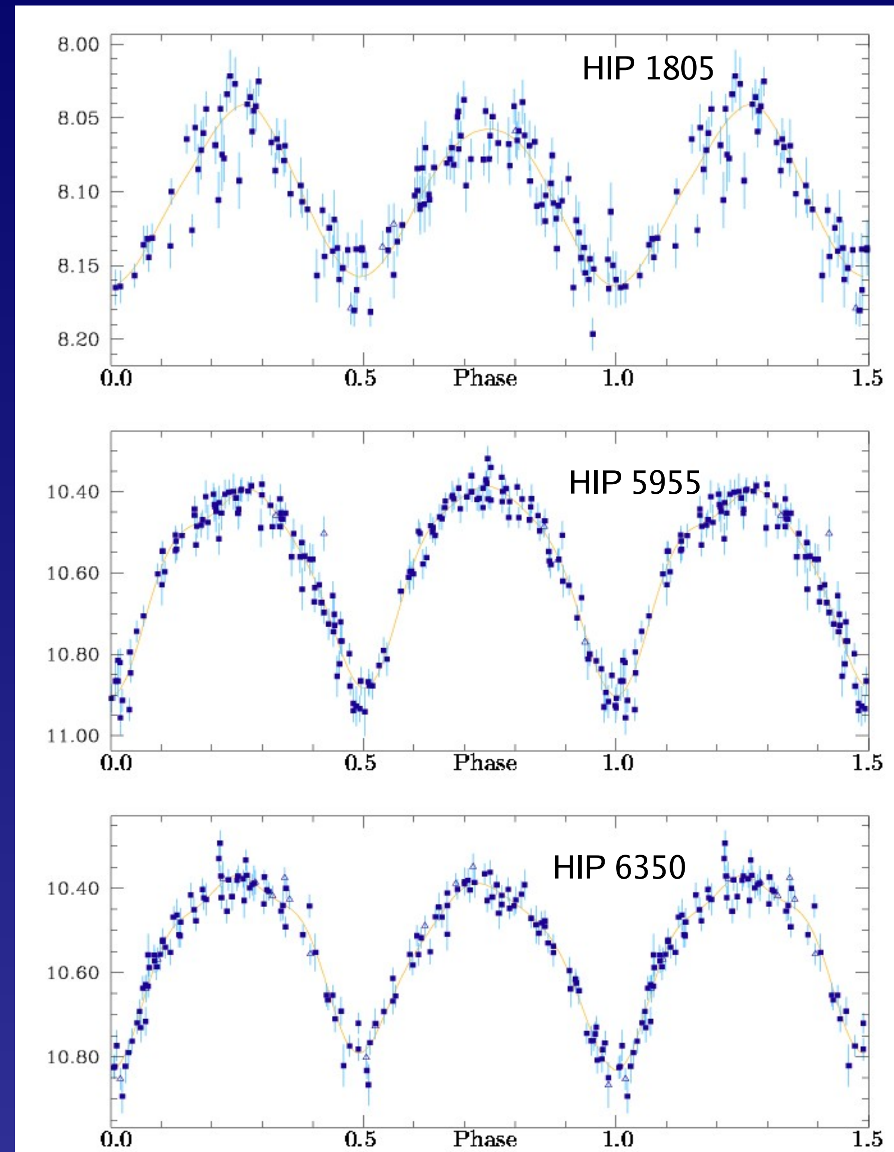


Wrong solutions

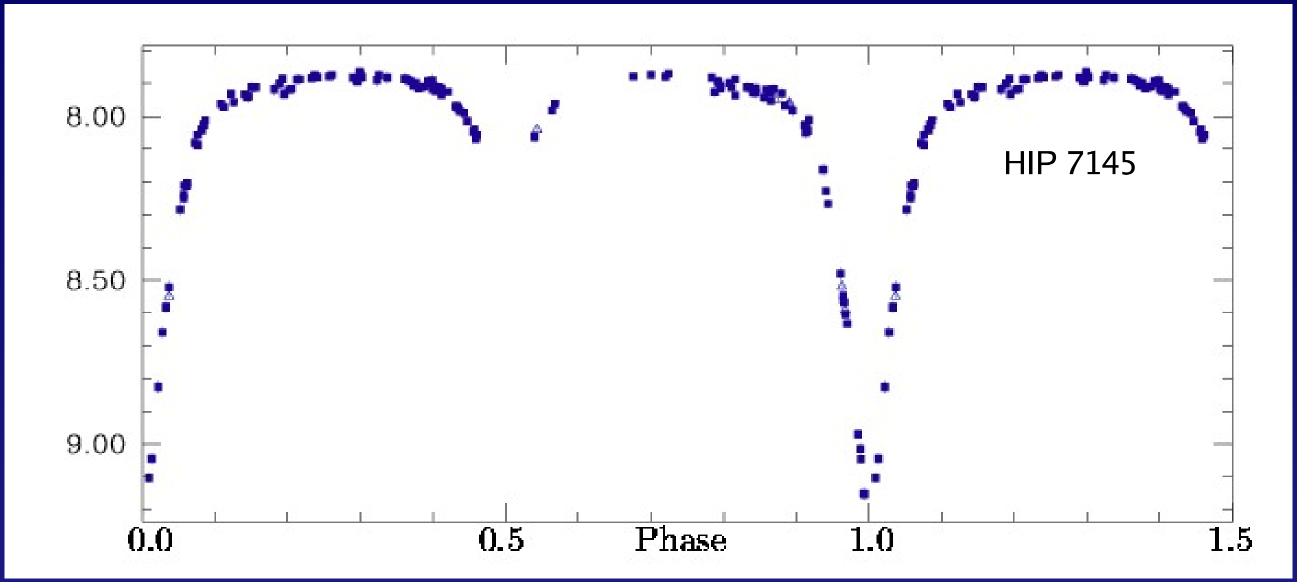
Very different



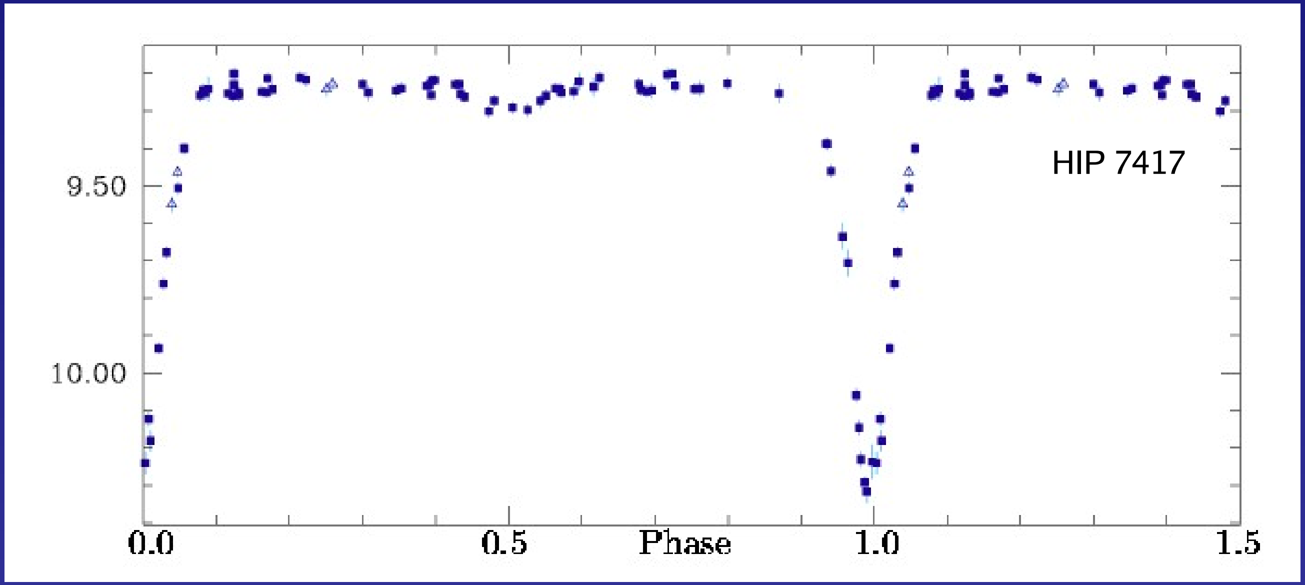
Factor 2



Wrong solutions ?



Factor 2



not found
or
factor 2

Conclusions and Further developments

- FAMOUS performs very well to analyse periodic time series
- It is not optimum for pulse-like signals with numerous harmonics
- But it is very efficient as starting method

- More effort on the theory is needed :
  - significance and error analysis not complete
  - window function for irregular sampling
  - better theoretically validated maximum frequency
  - relation with sufficient statistics not established

- FAMOUS is freely available on line, with Fortran source, test files, and the built-in simulator : website of the VSWG or on

  ftp.obs-nice.fr/pub/mignard/Famous