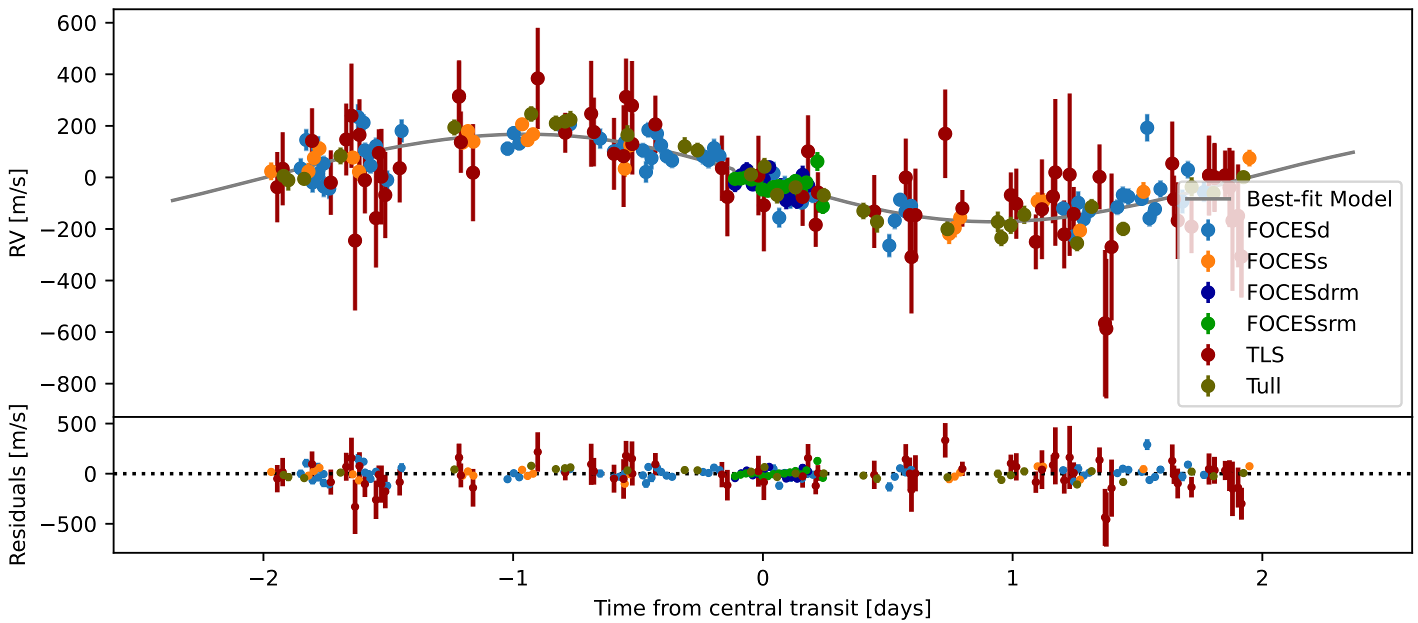
The analysis of time Series
Ondrejov, Monday August 26, 2024

Eike. W. Guenther

Thüringer Landessternwarte Tautenburg, Germany



What is a time series?

A time series is a set of data-points obtained of at discrete times
certain amount of time.
t1, RV-1
t2, RV-2

The difference between astronomical data, and other data is that astronomical data has discrete values that usually unequally spaces in time.

We can classify time-series in several categories:

1.) A linear (or quadratic, or exponential, or…) trend.
2.) A periodic signal.
3.) A multi-period signal.
4.) A quasi-periodic signal.
5.) Random noise.
6.) A discontinuity.

Examples:

1.) Linear trend: Radial-velocity measurements of star with a planet that has an orbital period which is much longer than the time it was observed.
2.) Radial-velocity measurements of a star with a planet.
3.) Radial-velocity measurements of an active star with a planet.
4.) The solar activity cycle.
5.) The photon statistics.
6.) The change of the bias-level in a detector after being hit by an energetic particle.

The periodogram analysis

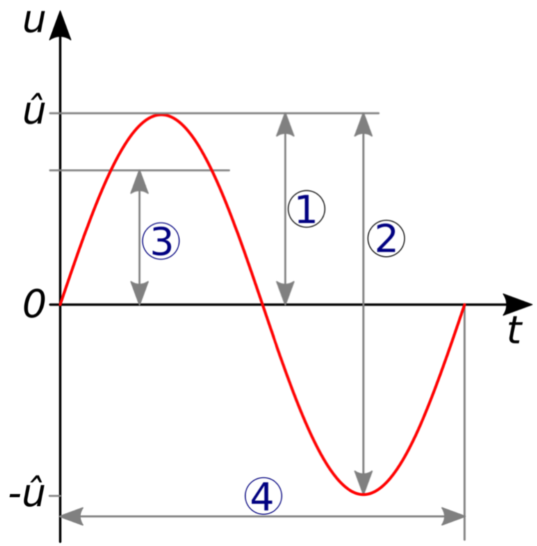
Any time series can be expressed as a combination of cosine and sine waves with differing periods.

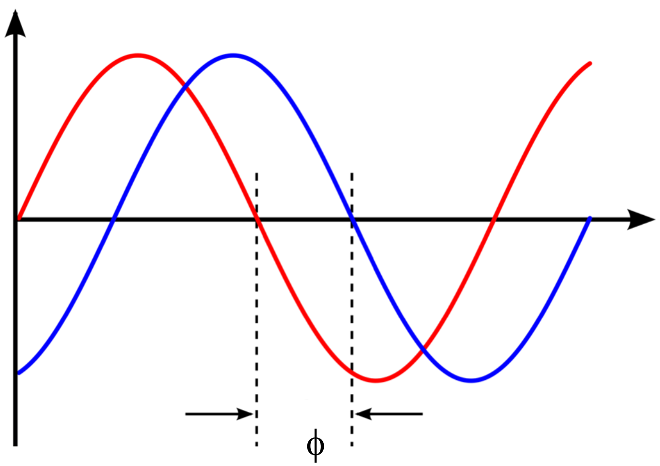
A periodogram is used to identify the dominant periods (or frequencies) of a time series. This can be a helpful tool for identifying the dominant cyclical behaviour in a series, particularly when the cycles are not related to the commonly encountered monthly or quarterly seasonality.

The frequency ω= 1/T. It is the fraction of the complete cycle that's completed in a single time-period T (4).

The (K-) amplitude A (1) determines the maximum absolute height of the curve.
The phase determines the starting point, in angle degrees, for the cosine wave. Peak-to-peak amplitude is (2), and the Root-Mean-square (RMS) is show in (3).

$$x_t = A \cos (2\pi\omega t + \phi)$$





Spectroscopic binaries:

If both components are visible in the spectra, we obtain the mass radio of the two stars M1/M2 from the RV-amplitude. For obtaining the absolute masses, we need to know the inclination, because the RV-amplitude is m_1*sin i.

If only the spectral-lines of one star is seen, we obtain the mass-function:

$$(M1 + M2) * \left(\frac{M2}{M_1+M_2}\right)^3 \sin^3 i = \frac{{M_2}^3 \sin^3 i}{(M_1+M_2)^2} = \frac{4\pi^2}{G}\frac{(a_1 \sin i)^3}{P^2}$$

Statistically $\sin^3 i=0.59$.
Exoplanet case: The mass of the planet is much smaller than that of the star. The K-amplitude is the RV-amplitude in m/s.:

$$K = \frac{28.4\,\mathrm{m\,s^{-1}}}{\sqrt{1-\epsilon^2}}\, m_p[\mathrm{M}_J]\, \sin i \left(\frac{1}{m_S[\mathrm{M}_\odot]}\right)^{2/3} \left(\frac{1}{P[\mathrm{yr}]}\right)^{1/3}$$

The Doppler effect is given by:

with β=v/c and z=Δλ/λ

$$\beta = v/c \qquad z = \frac{\lambda_{\rm ob} - \lambda_{\rm em}}{\lambda_{\rm em}}.$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1$$

Small velocities:

$$Z = \frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

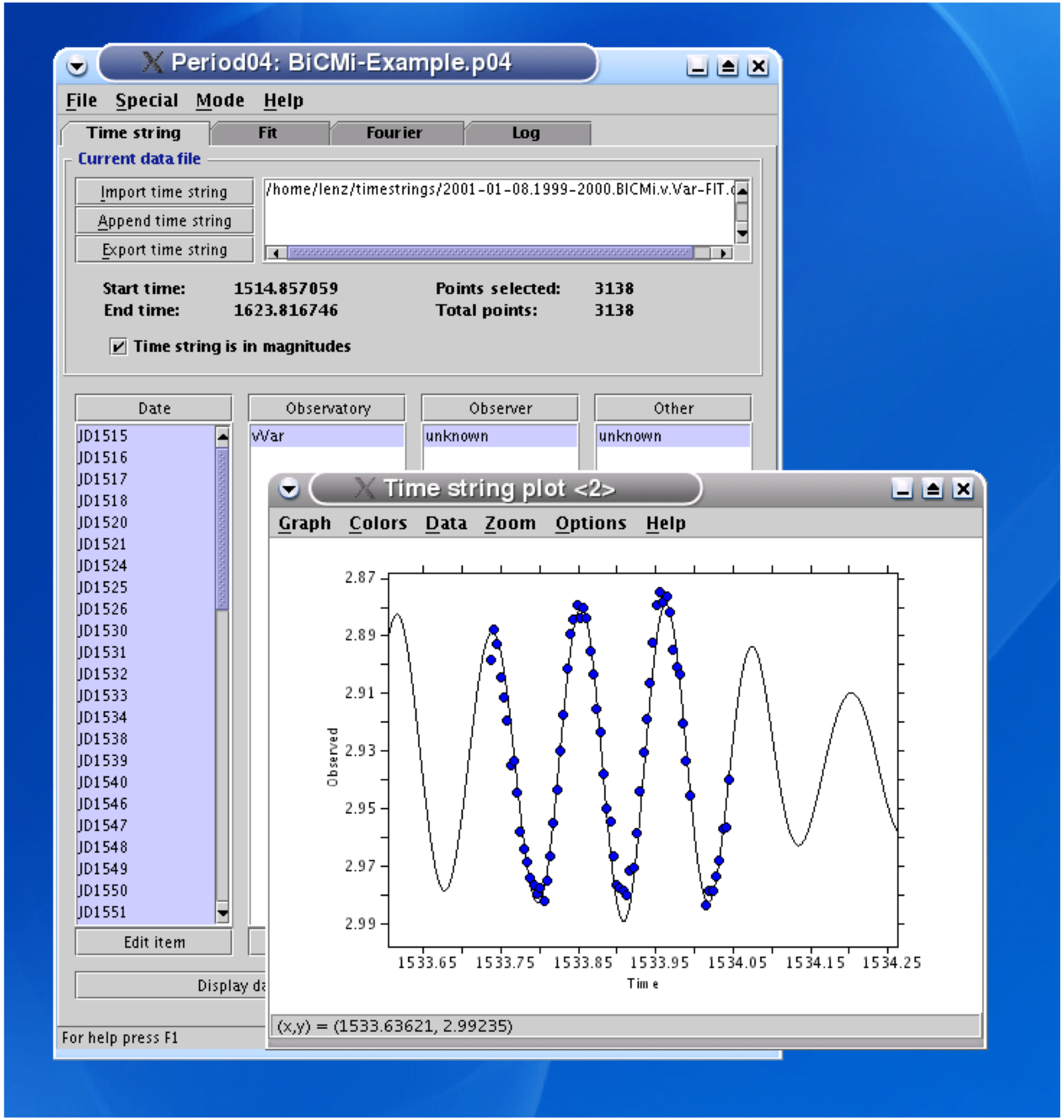
Period04

Period04 is a computer program especially dedicated to the statistical analysis of large astronomical time series containing gaps. The program offers tools to extract the individual frequencies from the multiperiodic content of time series and provides a flexible interface to perform multiple-frequency fits.

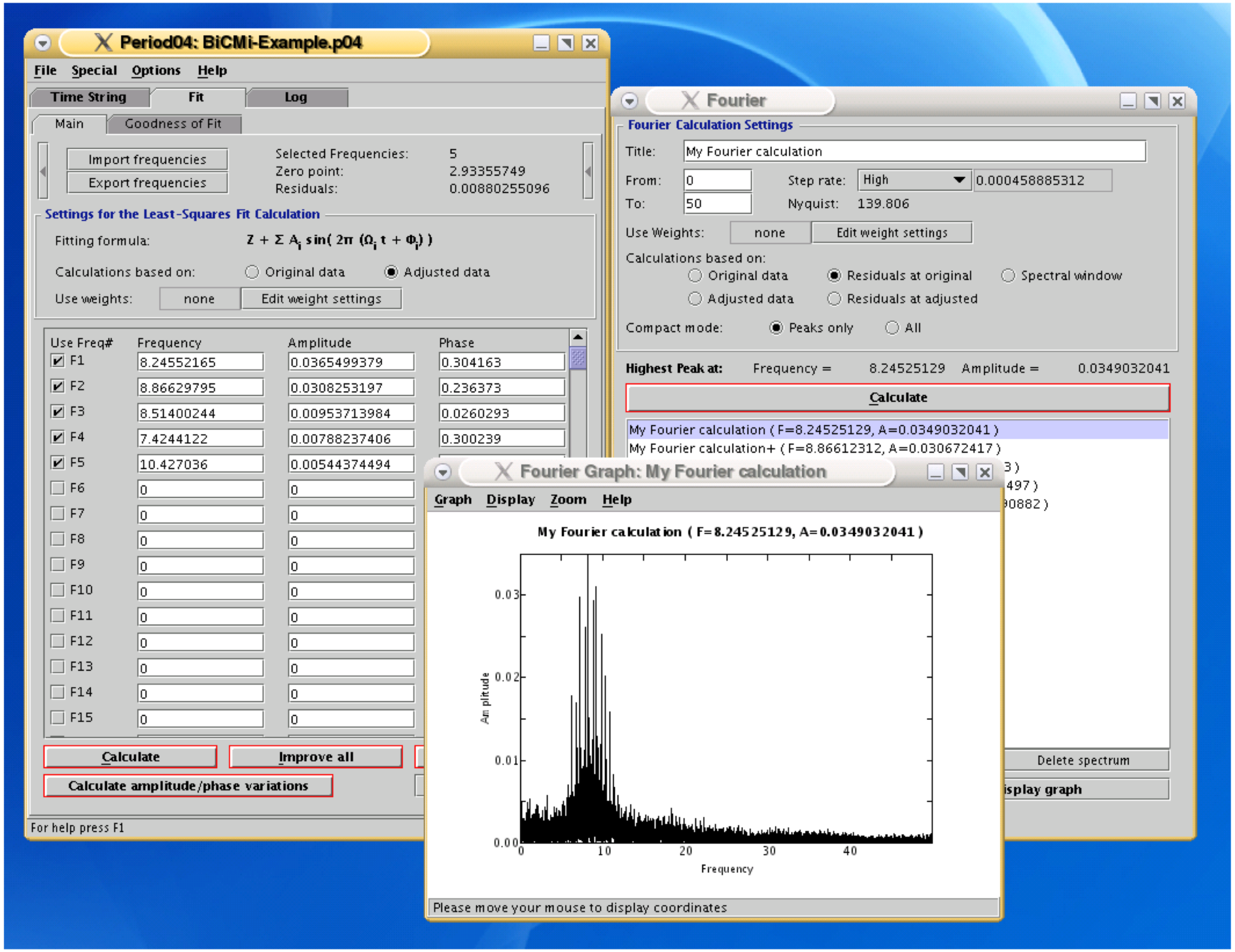
Reference:
The package has been published in 'Communications in Asteroseismology':
Lenz P., Breger M. 2005, CoAst, 146, 53

General features:
- discrete Fourier transformation
- least-squares fitting of multiple frequencies to the data
- tools for the calculation of formal uncertainties
- possibility to calculate amplitude and/or phase variations

Time String Tab

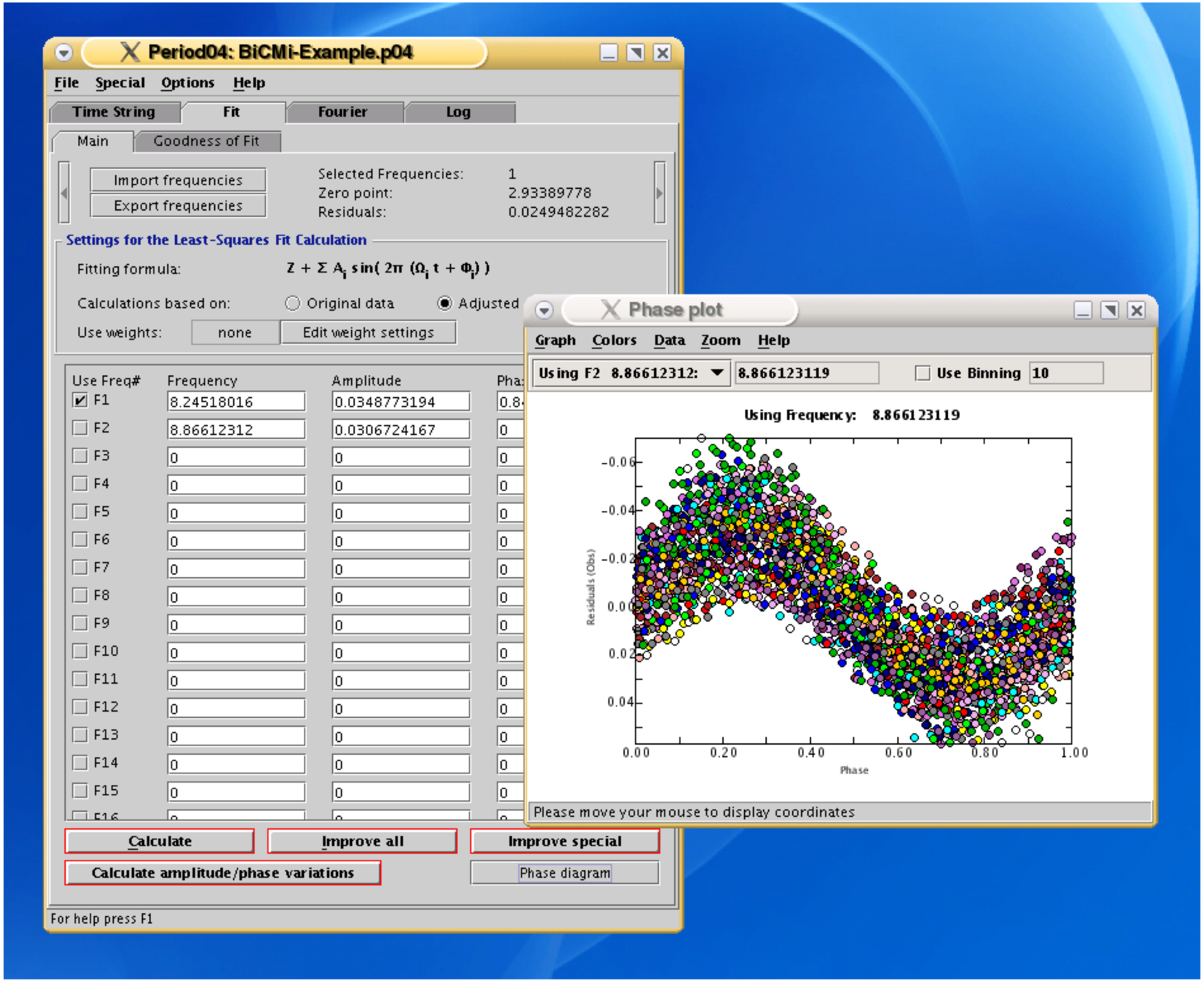


Fourier Tab



Period04: BiCMi-Example.p04
File Special Options Help
Time String Fit Log
Main Goodness of Fit
Import frequencies
Export frequencies
Selected Frequencies: 5
Zero point: 2.93355749
Residuals: 0.00880255096
Settings for the Least-Squares Fit Calculation
Fitting formula:
Calculations based on: Original data Adjusted data
Use weights: none Edit weight settings
Use Freq# Frequency Amplitude Phase
F1 8.24552165 0.0365499379 0.304163
F2 8.86629795 0.0308253197 0.236373
F3 8.51400244 0.00953713984 0.0260293
F4 7.4244122 0.00788237406 0.300239
F5 10.427036 0.00544374494
Calculate Improve all
Calculate amplitude/phase variations
For help press F1
Fourier
Fourier Calculation Settings
Title: My Fourier calculation
From: 0 Step rate: High 0.000458885312
To: 50 Nyquist: 139.806
Use Weights: none Edit weight settings
Calculations based on:
Original data Residuals at original Spectral window
Adjusted data Residuals at adjusted
Compact mode: Peaks only All
Highest Peak at: Frequency = 8.24525129 Amplitude = 0.0349032041
Calculate
My Fourier calculation (F=8.24525129, A=0.0349032041)
My Fourier calculation+ (F=8.86612312, A=0.030672417)
Delete spectrum
Fourier Graph: My Fourier calculation
Graph Display Zoom Help
My Fourier calculation (F=8.24525129, A=0.0349032041)
Amplitude
Frequency
Please move your mouse to display coordinates

Fit Tab


Period04: BiCMi-Example.p04
File Special Options Help
Time String Fit Fourier Log
Main Goodness of Fit
Import frequencies
Export frequencies
Selected Frequencies: 1
Zero point: 2.93389778
Residuals: 0.0249482282
Settings for the Least-Squares Fit Calculation
Fitting formula:
Calculations based on: Original data Adjusted
Use weights: none Edit weight settings
Use Freq# Frequency Amplitude
F1 8.24518016 0.0348773194
F2 8.86612312 0.0306724167
Calculate
Improve all
Improve special
Calculate amplitude/phase variations
Phase diagram
For help press F1
Phase plot
Graph Colors Data Zoom Help
Using F2 8.86612312: 8.866123119
Use Binning 10
Using Frequency: 8.866123119
Phase
Please move your mouse to display coordinates

The Fourier transform

The Fourier transform is an *analysis* process, decomposing a function f(x) into its constituent frequencies and their amplitudes.

Gilt für eine Funktion f(x) mit der Periode 2π eine Darstellung als Fourierreihe

$$f(x) = \frac{a_o}{2} + \sum_{n=1}^{\infty} a_n \cos nx + \sum_{n=1}^{\infty} b_n \sin nx,$$

dann sind die Koeffizienten bestimmt durch die Beziehungen

$$a_o = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, dx \qquad (22\text{-}3)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\, dx \qquad (22\text{-}4)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\, dx \qquad (22\text{-}5)$$

$$\underbrace{Ae^{i\theta}}_{\text{polar coordinate form}} = \underbrace{A\cos(\theta) + iA\sin(\theta)}_{\text{rectangular coordinate form}}.$$

The Discrete Fourier transform

Reference: The Doppler Method for the Detection of Exoplanets by A. P. Hatzes.

The Fourier transform is the classic method for finding periodic signals in your time series data. Since with experimental data we are always dealing with discrete time series, we use the Discrete Fourier Transform (DFT) defined by:

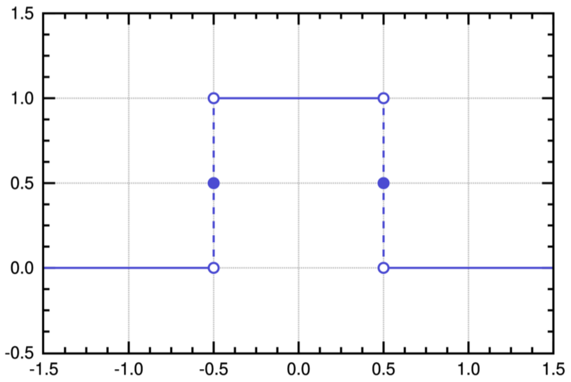
$$\mathrm{DFT_X}(\omega) = \sum_{i=1}^{N} X(t_j) e^{-i\omega t_j}$$

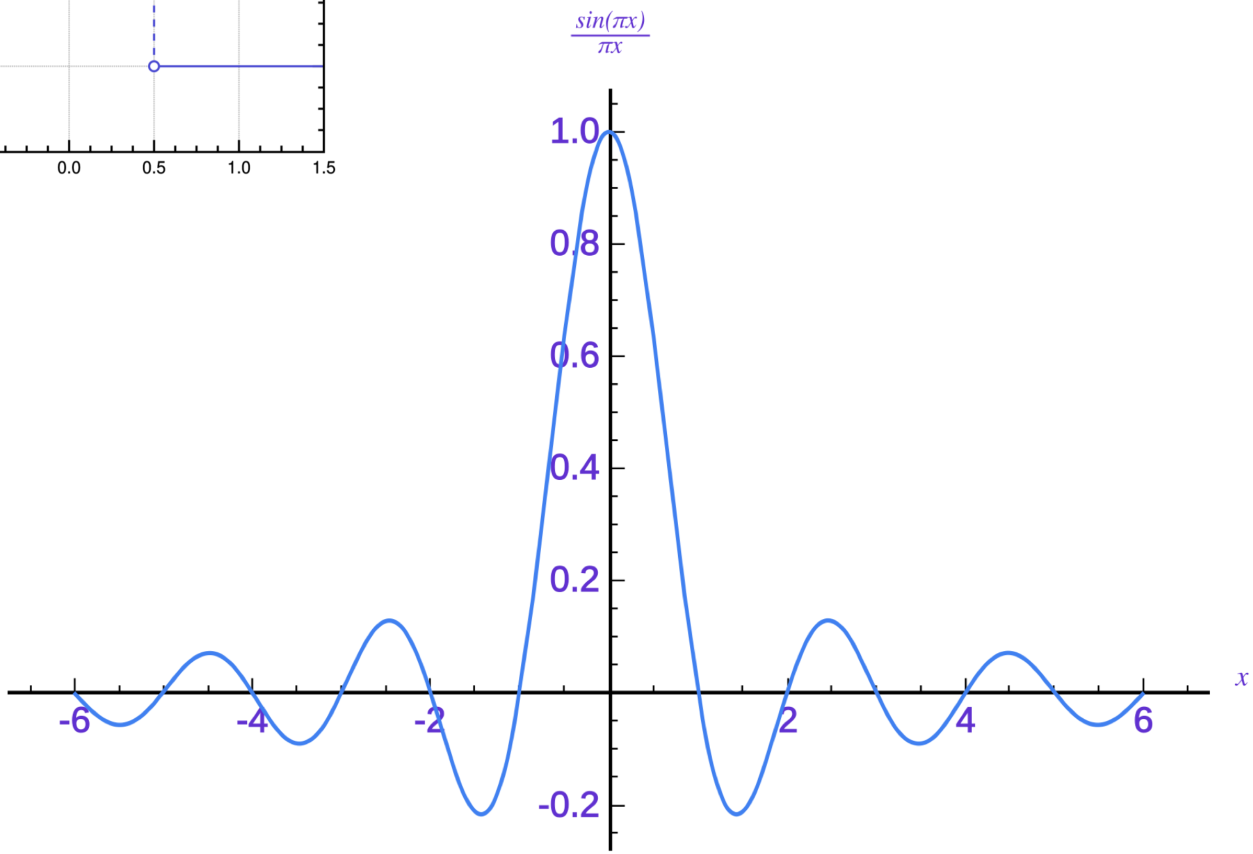
where $e^{i\omega t}$ is the complex trigonometric function $\cos(\omega t) + i\sin(\omega t)$, N is the number of data points sampled at times t_j, and ω the frequency. The DFT is often called the classic periodogram.

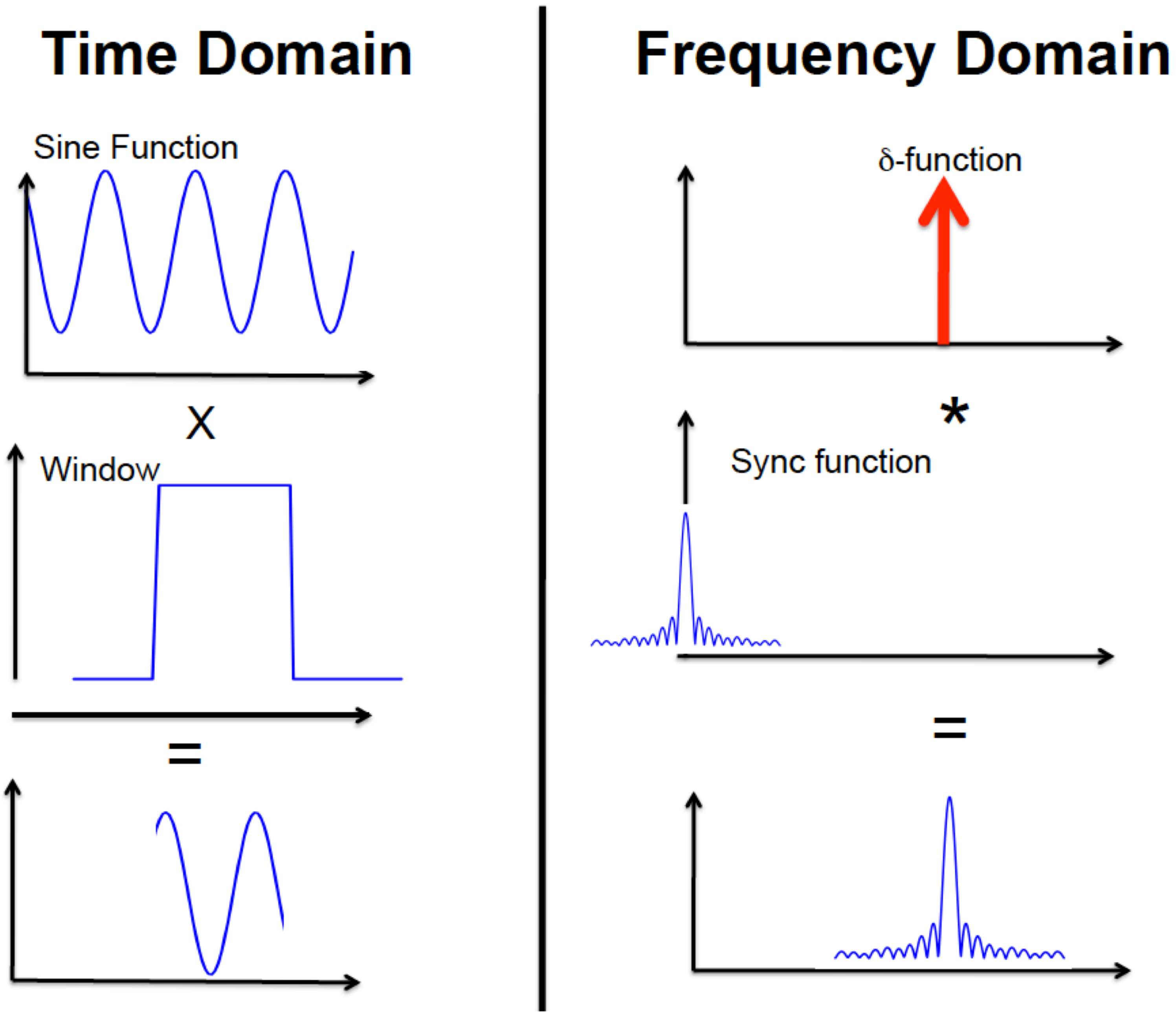
The **power** is defined by

$$P_X(\omega) = \frac{1}{N}|\mathrm{FT_X}(\omega)|^2 = \frac{1}{N}\left[\left(\sum_{j=1}^{N} X_j \cos\omega t_j\right)^2 + \left(\sum_{j=1}^{N} X_j \sin\omega t_j\right)^2\right]$$

The sinc function is the Fourier transform of a rectangular function.







Time Domain
Frequency Domain
Sine Function
δ-function
X
*
Window
Sync function
=
=

The Lomb-Scargle (LS) periodogram

Reference: The Doppler Method for the Detection of Exoplanets by A. P. Hatzes.

The astronomical community often uses the Lomb-Scargle (Lomb 1976; Scalge 1982) periodogram. The Lomb–Scargle periodogram is a well-known algorithm for detecting and characterizing periodicity in unevenly sampled time-series.
The Power in a LS-periodogram is defined as:

$$P_X(\omega) = \frac{1}{2}\left\{ \frac{[\sum_j X_j \cos\omega(t_j - \tau)]^2}{\sum_j \cos^2\omega(t_j - \tau)} + \frac{[\sum_j X_j \sin\omega(t_j - \tau)]^2}{\sum_j \sin^2\omega(t_j - \tau)} \right\}$$

where τ is defined by

$$\tan(2\omega\tau) = \sum_j \sin(2\omega t_j) / \sum_j \cos(2\omega t_j)$$

The generalized Lomb-Scargle (GLS) periodogram

RV-measurements of the same star taken at different times usually different error-bars! The GLS takes the errors of the measurements into account. Furthermore the mean of the data is subtracted, before the GLS is calculated. A constant offset would otherwise prdice a δ–function in the power spectrum.

Estimating the significance using the Lomb-Scargle (LS) periodogram

The false alarm probability (FAP) is given by:

$$FAP = 1 - (1-e^{-z})^{Ni}$$

With **z, the power of the peak in the LS-diagram**, and Ni the number of independent frequencies. The number of independent frequencies is $Ni = -6.362 + 1.193 * N_0 + 0.00098 * N_0^2$.

With N_0 the number of measurements.

For a large number of measurements, we can use **FAP ≈ $N^* e^{-z}$**

Example: 30 measurements, z=8 → FAP=0.01

z = 6 - 10: Probably due to noise.

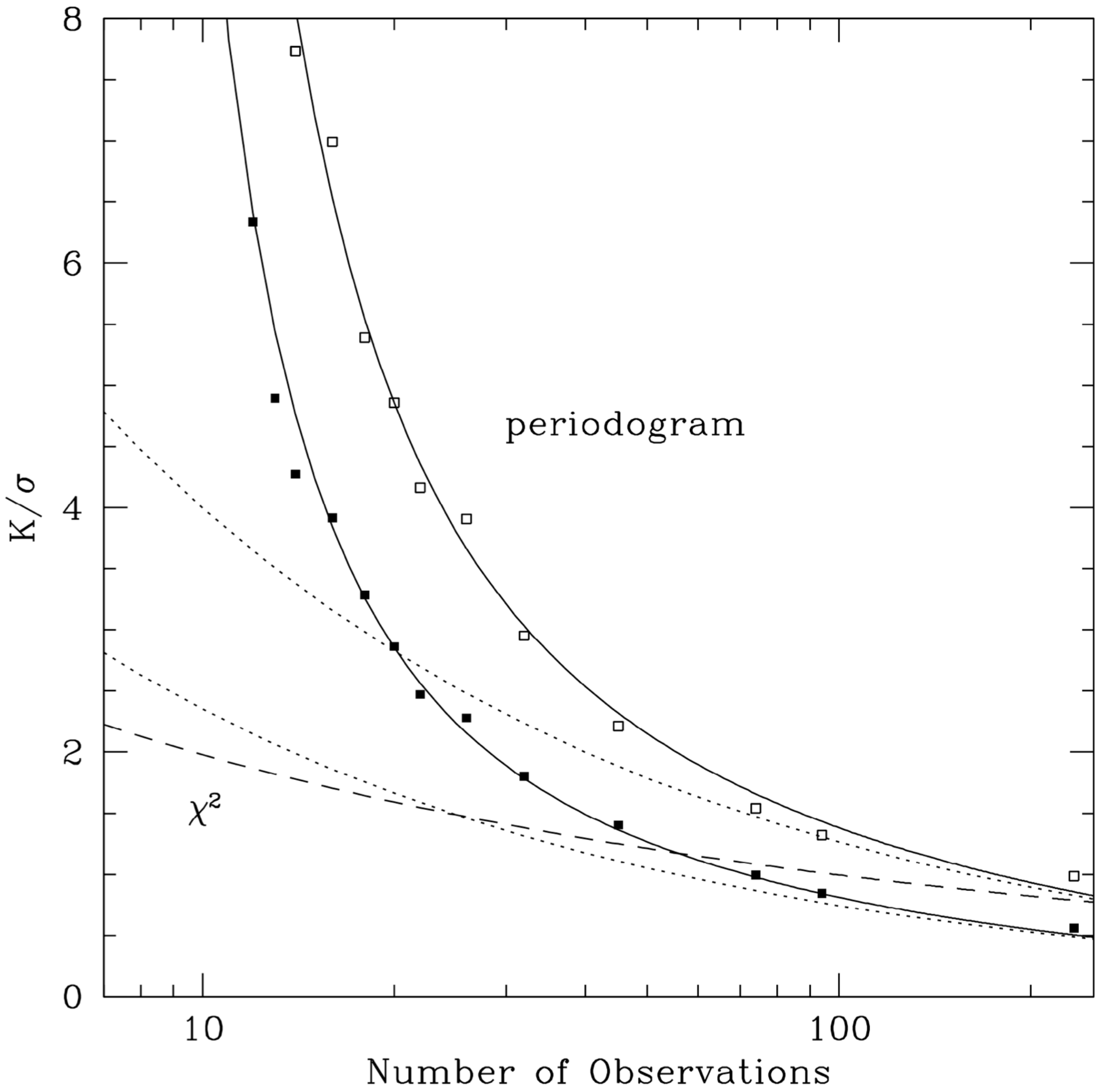
z = 8 - 12: Borderline case, but most likely noise.

z = 12 - 15: Interesting, most likely real, but in some cases can be due to noise.

z = 15-20: With high probability real, but nature can still fool you.

z > 20 Real signal.

Signal-to-noise ratio K/σ that can be detected with N observations, and 99% (upper curve) and 50% detection efficiency (lower curve). Taken from Cumming (2004)



Window function

A window function is a mathematic function that is zero-valued outside a given interval and larger than zero inside the interval. It describes how much of the signal is lost, because of the time-sample.

For example, if we observe a star every night at exactly the same time, we will not be able to detect a planet with an orbital period on one day.

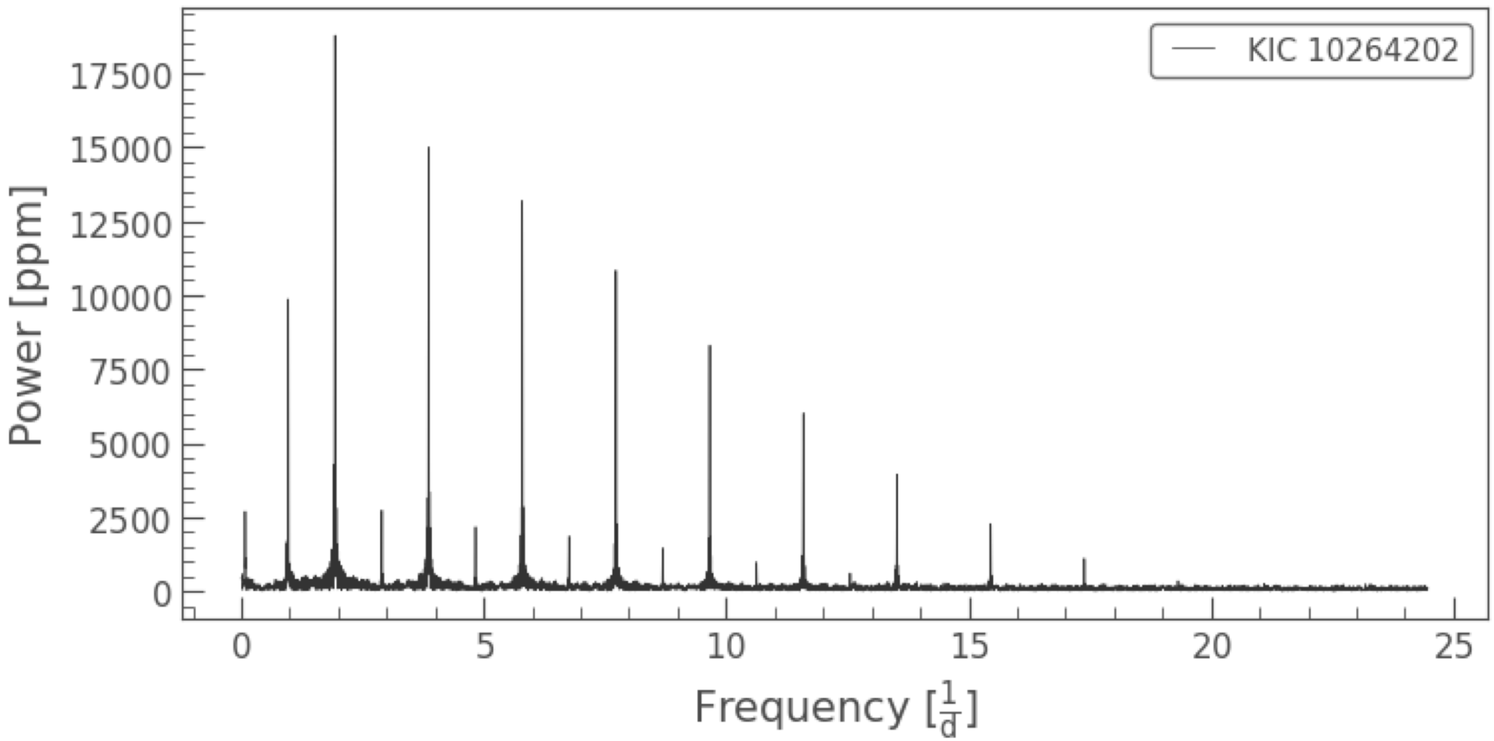
What is even worse, is that if the data contains only noise and we observe a star every night at the same time, we will see a signal that is not real.

How can we then solve that problem? Answer: The best way is to ask a friend with an observatory at a very different latitude. If just a few data-points are added, the spurious signal will disappear.

→ The detection probability depends on the time-sampling of the data.

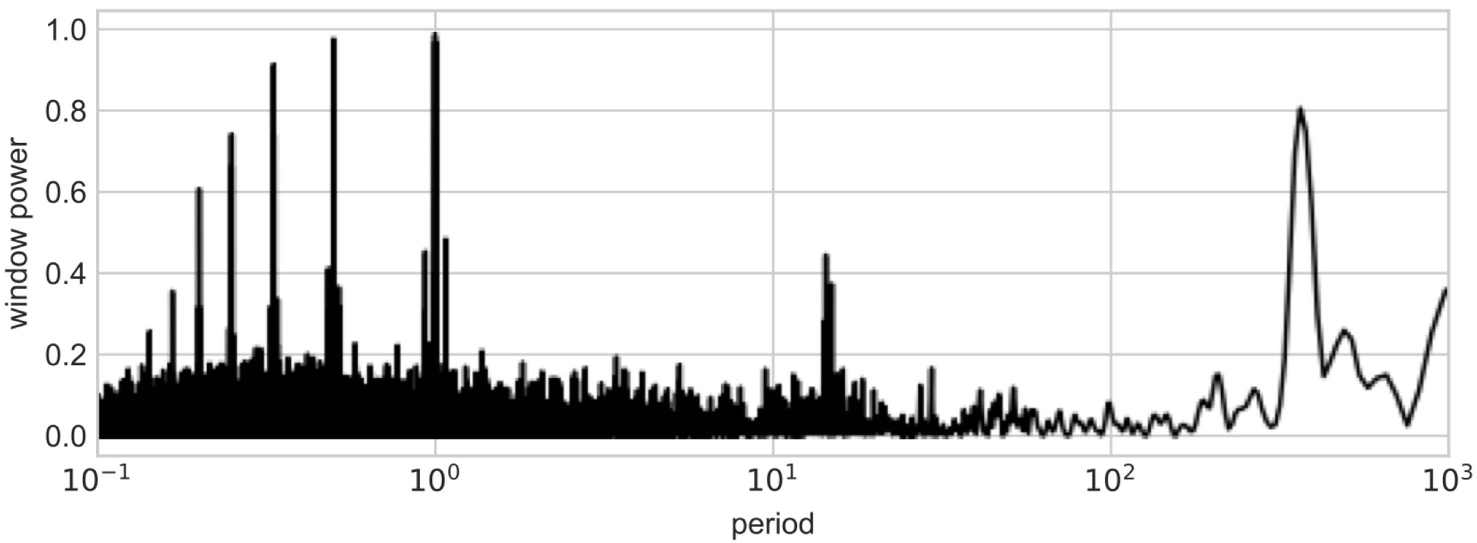
Aliases I

While at first glance it may look like there are lots of different signals, the repeating, evenly spaced peaks in this instance are likely alias harmonics of the true oscillation frequency. Harmonics are integer multiples of the oscillation frequency. They occur when a light, or RV-curve is finite in duration **or contains gaps**, in which case it is common for a DFT to identify integer multiples as additional candidate signals. For example, an eclipse with a period of 10 days can easily be mistaken for an eclipse with a period of 20 days, in particular when the data contains gaps near the transit times.



Aliases II

The power spectrum of the observing window. Notice the strong spike in power at a period of 1 day, and related **aliases at 1/n days** for integer n. There is also a strong spike at 365 days, and a noticeable spike at ∼ 14 days. Each of these indicate time intervals that appear often in the data. A nightly observation pattern—typical of ground-based surveys—leads to a window function with a strong diurnal component that causes each frequency signature f0 to be partially aliased at **f0 +nδf**, for integers n and δf = 1 cycle/day.



Pre-whitening

In pre-whitening, one performs a DFT on the time series in order to find the dominant peak in the data. A sine fit is made to the data using that frequency and this is subtracted from the data. Note that this procedure also removes all the aliases due to this signal. One then performs a DFT on the residual data to find the next dominant peak which you then fit and subtract. The process stops when the final residual peak in the DFT is at the level of the noise. A good level to stop is when the final peak is less than four times the surrounding noise level.

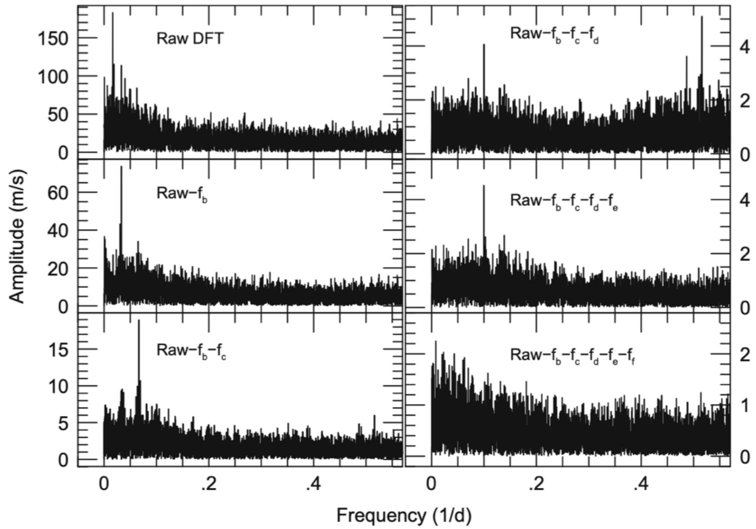


Fig. 1.33 Pre-whitening procedure applied to the RV data of the six planets system GJ 876

Table 1.2 Planets found in the GL 876 system by the pre-whitening process compared to published values (Jenkins et al. 2014)

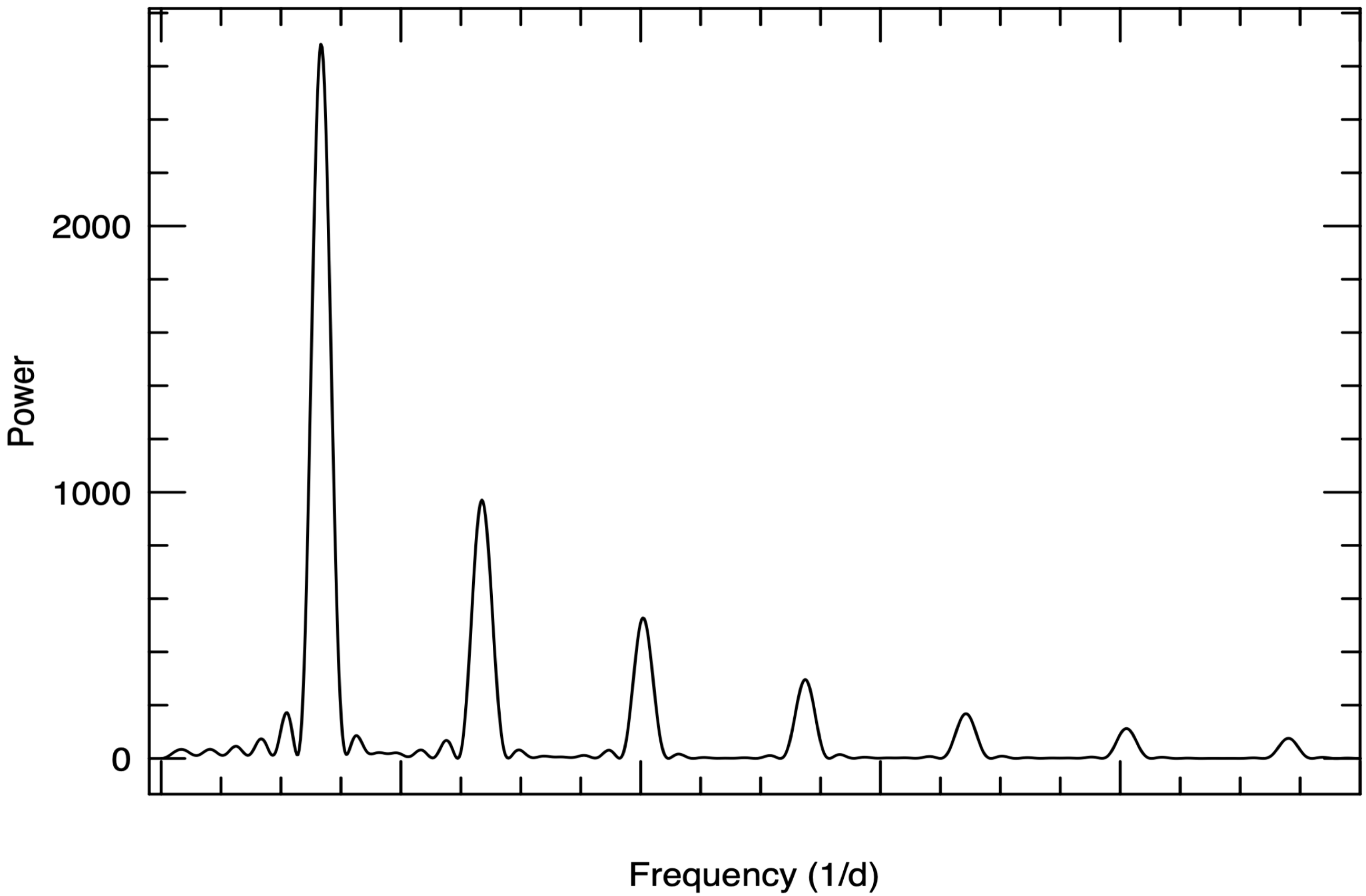
	ν (day^{-1})	P (days)	K (m s^{-1})	$P_{\text{published}}$ (days)	$K_{\text{published}}$ (m s^{-1})
f_b	0.0163	61.03 ± 0.001	211.7 ± 0.4	61.03 ± 0.03	211.6 ± 32.9
f_c	0.0330	30.28 ± 0.004	89.0 ± 0.3	30.23 ± 0.03	88.7 ± 13.2
f_d	0.0664	15.04 ± 0.004	20.76 ± 0.28	15.04 ± 0.004	20.7 ± 3.2
f_f	0.0998	10.01 ± 0.03	5.70 ± 0.27	10.01 ± 0.02	5.00 ± 0.80
f_e	0.5160	1.94 ± 0.001	6.21 ± 0.23	1.94 ± 0.001	5.91 ± 0.98
f_g	0.0080	124.88 ± 0.02	3.19 ± 0.35	124.88 ± 90	3.37 ± 0.53

Two planets, or just one?

Two planets in circular orbits can mimic the radial velocity (RV) signature of a single eccentric giant planet if not enough RV-measurements have been used to determine the true orbital configuration of the system (R. Wittenmeyer 2013).

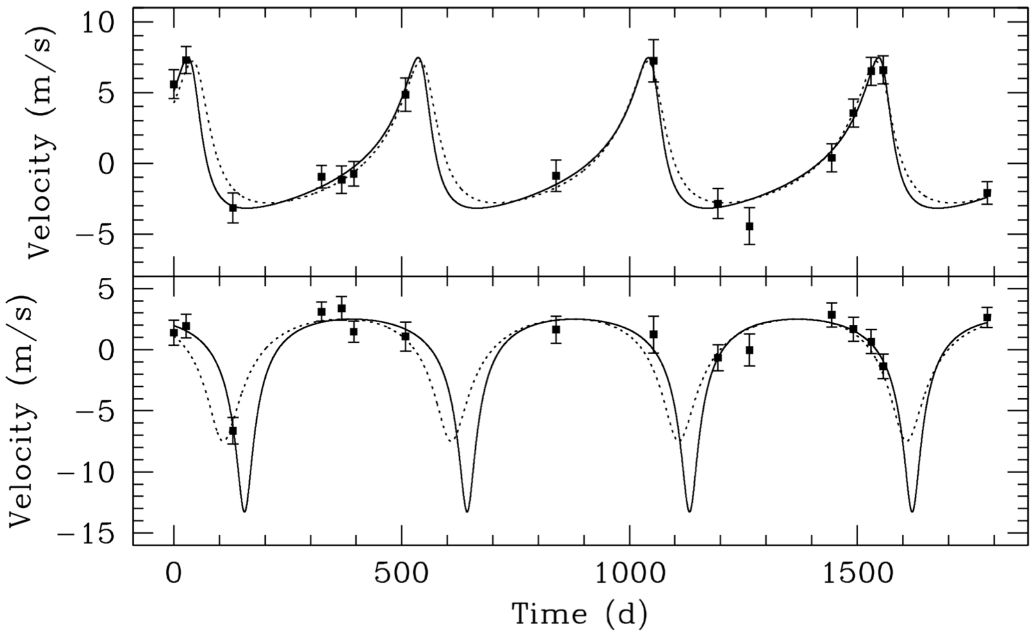
Lomb-Scargle periodogram of RV data from an eccentric orbit (e = 0.75).

Additional peaks to the main frequency occur at harmonics of the orbital period (f_{orb}): $f = n*f_{orb}$ where n is an integer.



Eccentric orbits:

Examples of velocity curves with *e* = 0.5 that are (top panel) and are not (bottom panel) detected. The **dotted line in each case shows the true orbit**; the points are the observed velocities; and the solid curve shows the best-fitting orbit. In both cases, the solid curve gives a lower χ2 than the dotted curve. The lower panel has only a single measurement during the periastron passage, and is not a significant detection. Taken from Cumming (2004).



The bootstrap method for estimating the errors

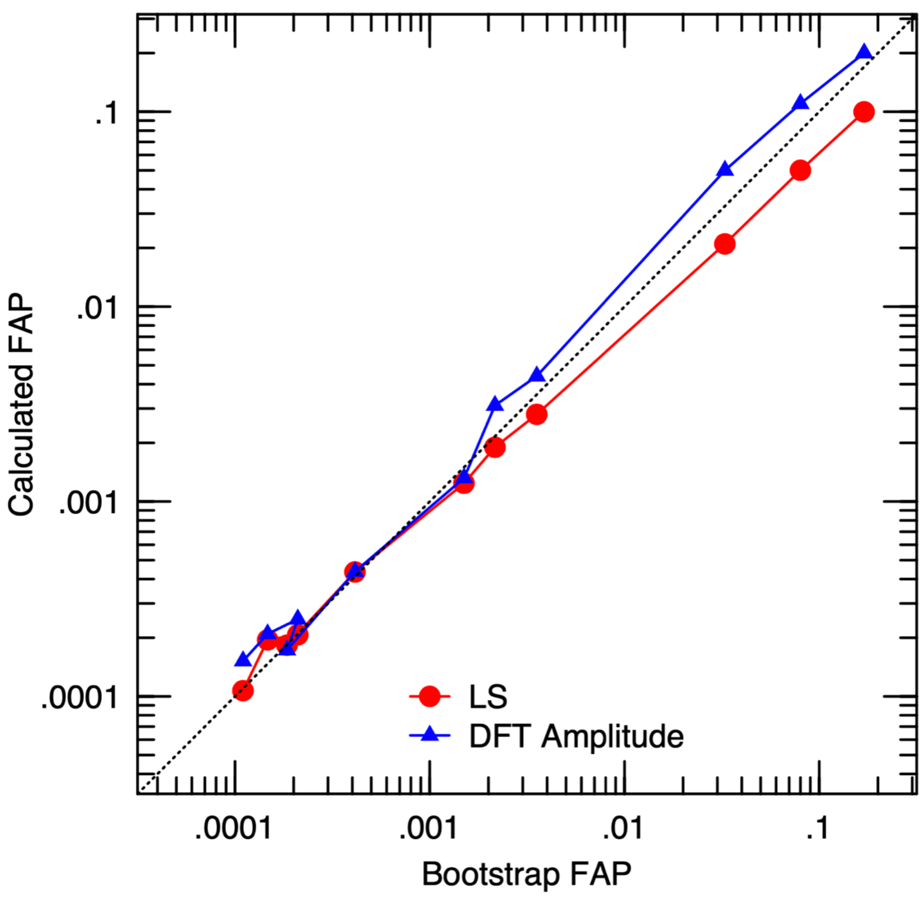
In the absence of a true analytic solution to the FAP, we can turn to computational methods such as the bootstrap. The bootstrap method is a technique in which the statistic in question is computed repeatedly on many random resamplings of the data in order to approximate the distribution of that statistic (see Ivezić et al. 2014, for a useful general discussion of this technique). For the periodogram, in each resampling we keep the temporal coordinates the same, draw observations randomly with replacement from the observed values, and then compute the maximum of the resulting periodogram. For enough resamplings, the distribution of these maxima will approximate the true distribution for the case with no periodic signal present. The bootstrap produces the most robust estimate of the FAP because it makes few assumptions about the form of the periodogram distribution and fully accounts for survey window effects.

Example:

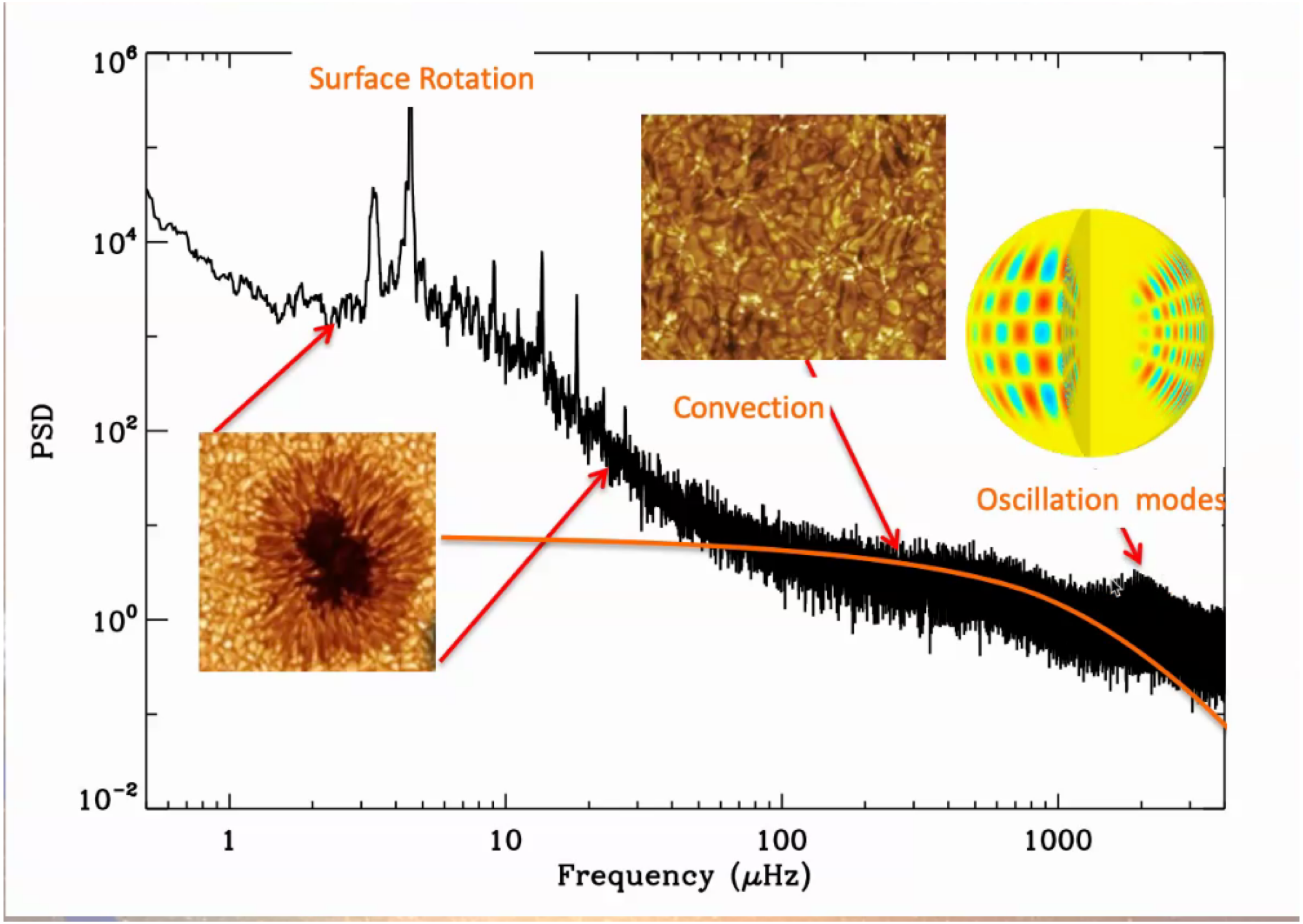
t1-RV1,t2-RV2,t3-RV3 → t1-RV2, t2-RV3,t3-RV1…etc…

Make power spectrum of all the random samples and count how many times the largest peak exceeds the hight of the peak that you think is a signal.

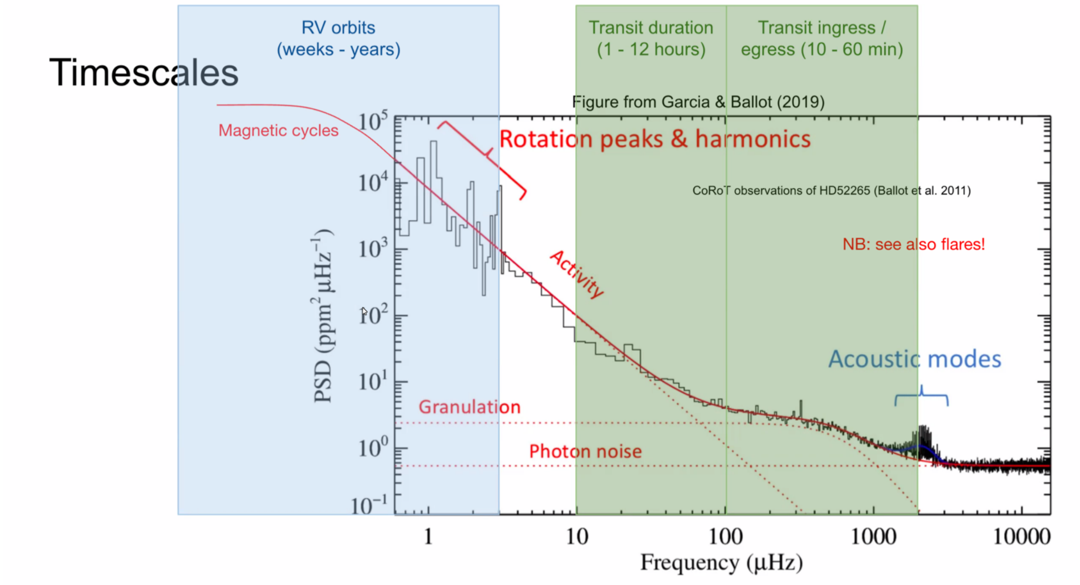
Comparing the significance calculated using the hight of the peak in the Lomb-Scargle (LS) periodogram with the bootstrap method.



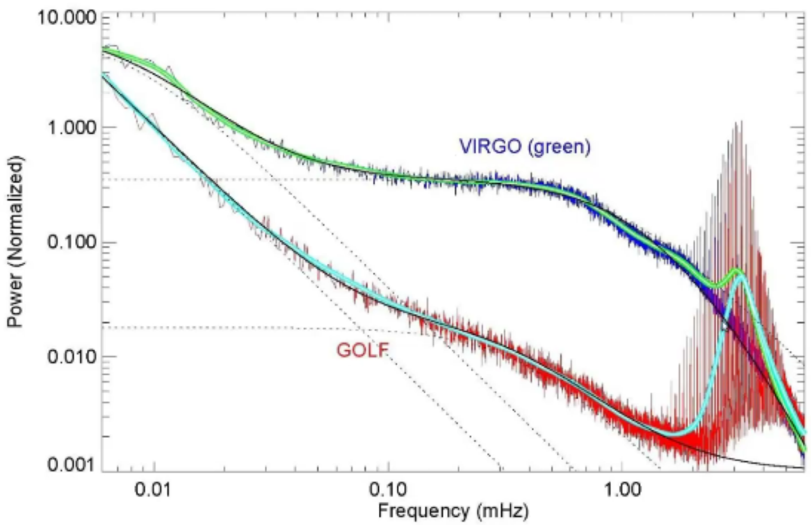
How a periodogram of a solar-like star looks like

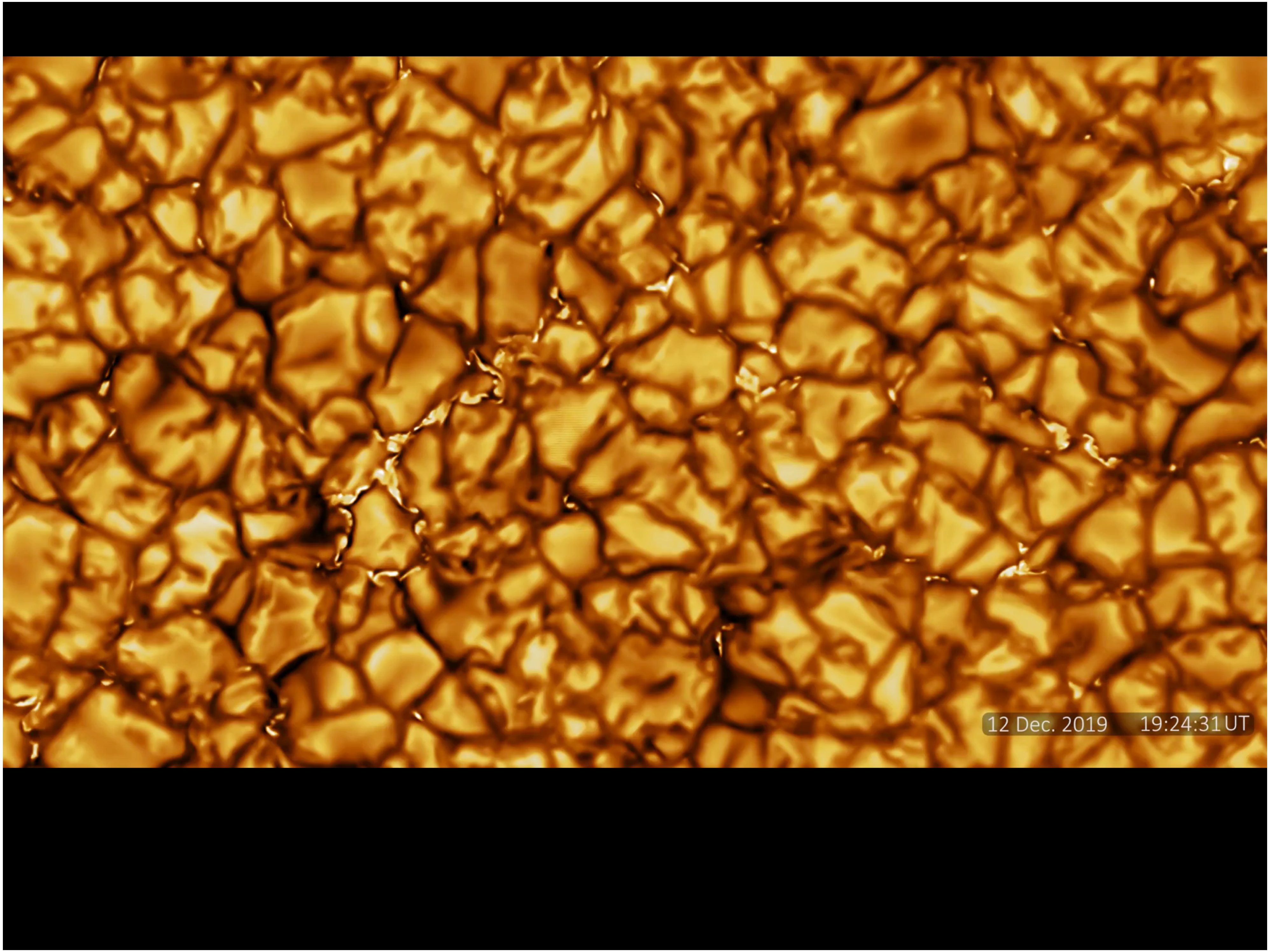


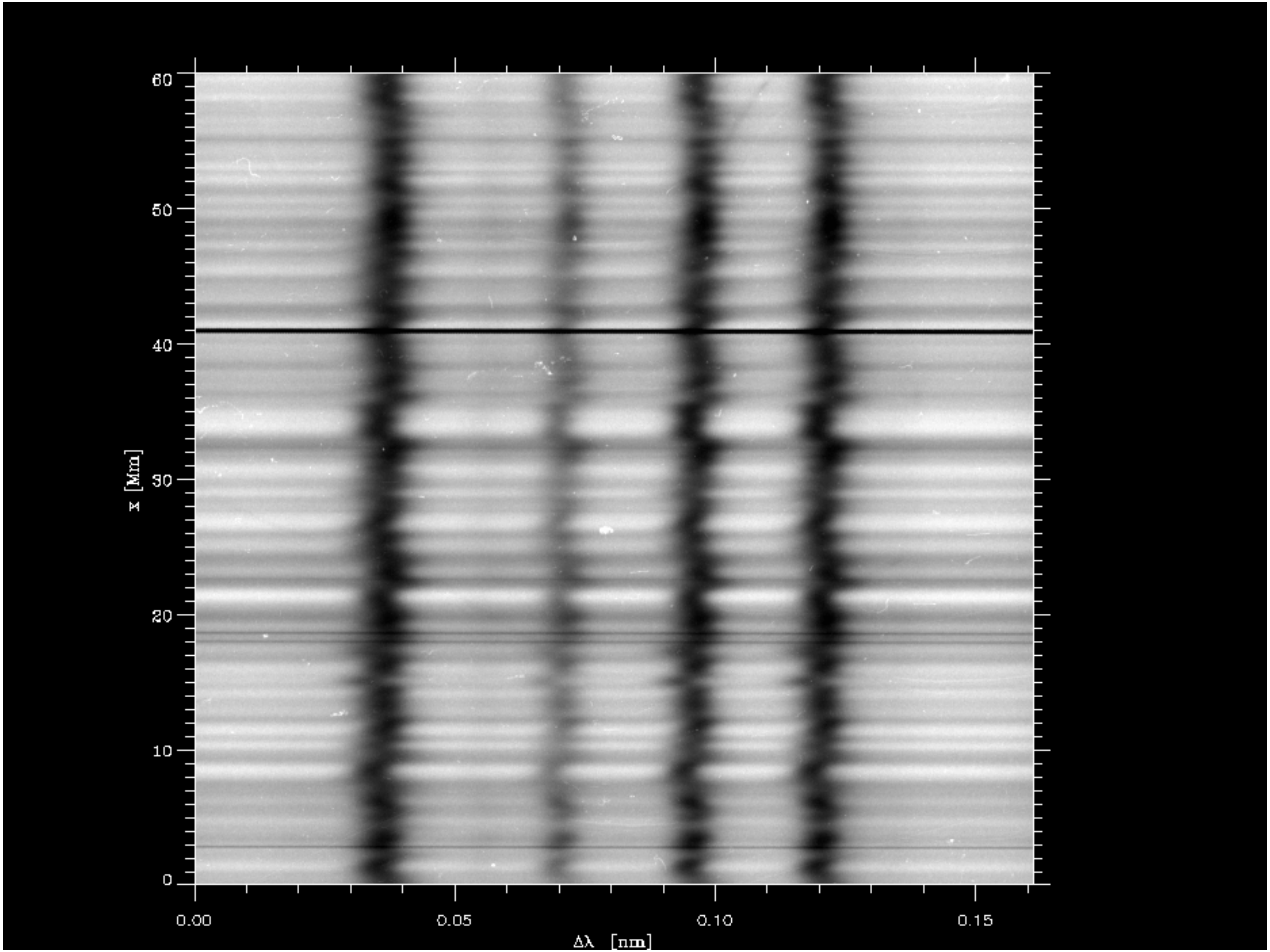
Power spectrum of the brightness and RV-variations of the sun and a star



RV vs. Photometry

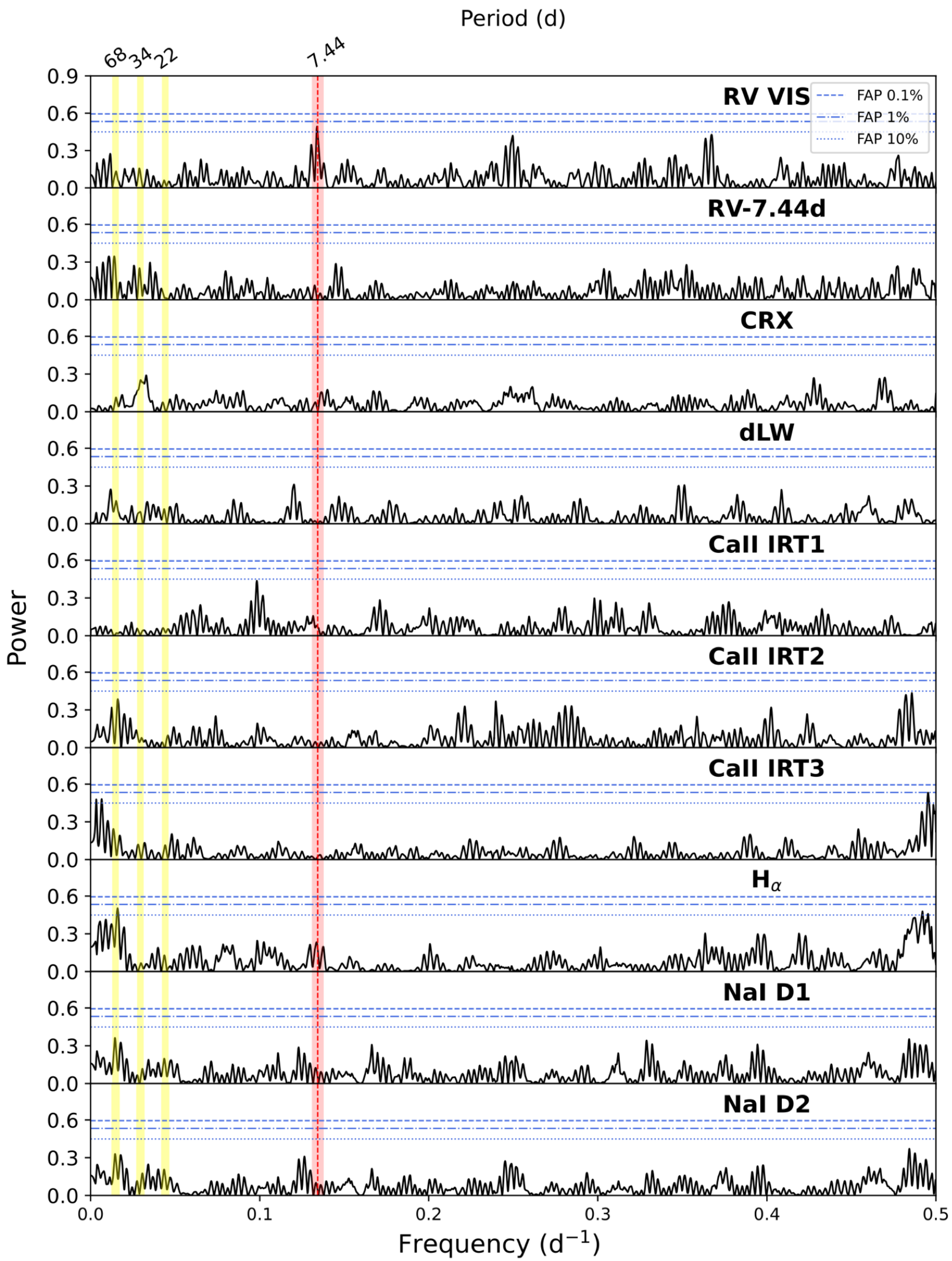



12 Dec. 2019
19:24:31 UT



60
50
40
30
20
10
0
x [Mm]
0.00
0.05
0.10
0.15
Δλ [nm]

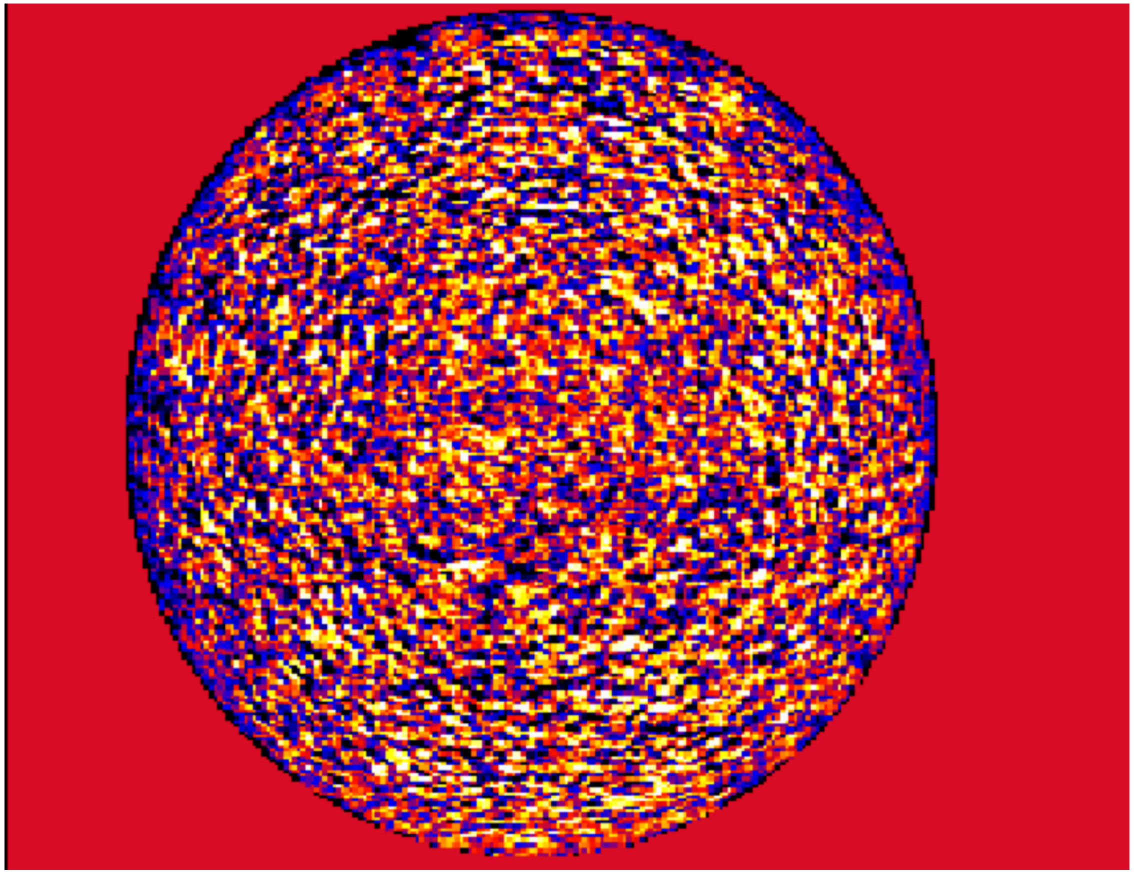
How to distinguish RV-variations caused activity from RV-variations caused by a planet.



An example for a period signal: Solar (stellar) oscillations Discovery

In 1975 solar oscillations were discovered by Franz-Ludwig Deubner and independently by Robert B. Leighton using Doppler-shifts of spectral lines.

The figure shows the Doppler motions in the photosphere.



The oscillations have typical periods of about 5 minutes,

or ω=0.01-0.03 s^{-1}.

Basics about oscillations

The length of the signal T (length of the time that the signal is observed) allows to resolve frequencies of $\Delta\omega=2\pi/T$.

The highest frequency that we can observe is given by the sampling rate (**Nyquist** frequency) Δt: $\omega_{Ny}=\pi/\Delta t$.

In frequency domain: $2\pi/T \leq \omega \leq \pi/\Delta t$

In space domain: $2\pi/L_x \leq k_x \leq \pi/\Delta x$

P-mode oscillations on the sun

- The waves are not transversal, like water-waves but longitudinal, like sound-waves.
- The amplitudes are 0.5 to 1 km/s.
- The life-time of modes are typically hours.
- The waves are presumably excited by turbulence in the outer layers of the sun.

Propagation of the waves

- Let us assume a wave that travels downwards into the sun under a small angle.
- The speed with which the waves propagate is the thermal velocity of the atoms. Thus, the speed is proportional to sqrt(T).
- Because the temperature increases inside the sun, speed of that part of the wave that is closer to the surface is slower.
- The wave bends, and is finally reflected back to the surface.

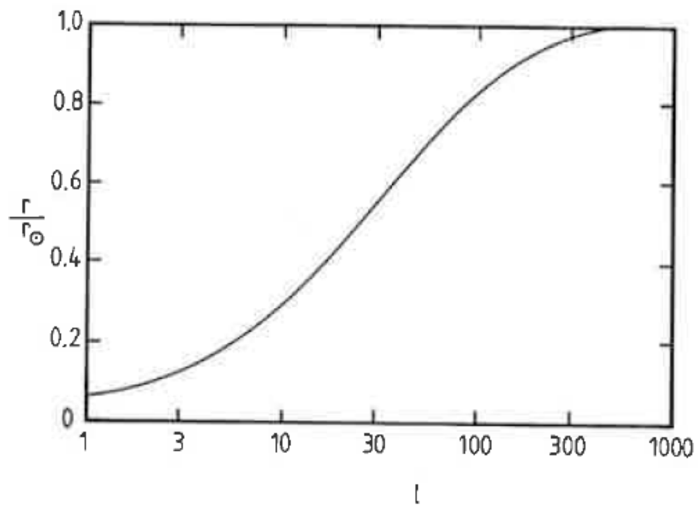


Fig. 5.13. Depth of internal reflection, according to (5.37), as a function of degree l, for modes with an oscillation frequency 3 mHz. Adapted from Noyes and Rhodes (1984)

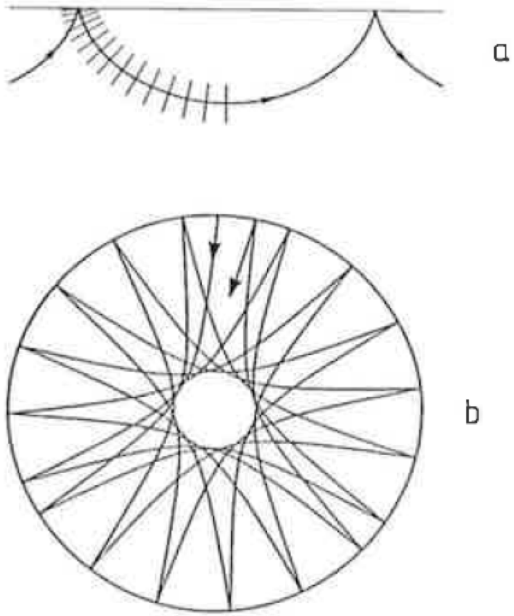


Fig. 5.14. (a) Ray path and surfaces of constant phase for a p mode of high degree l confined in a shallow layer just below the solar surface. **(b)** Ray path around the Sun of a multiply reflected p mode of low degree. From Gough (1983b)

How the oscillations are described

Because the sun is a sphere, we should use spherical coordinates. We thus use the same description which is familiar to us from quantum-mechanics: spherical harmonics which contain the Legendre-polynomes.

$$Y_l^m = P_l^m(cos\theta)e^{im\phi}$$

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$
$$P_3(x) = \frac{1}{2}(5x^3 - 3x)$$
$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$$

$$v(r,\theta,\phi,t) = \sum_{n=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-l}^{+l} V_n(r)Y_l^m(\theta,\phi)e^{-2\pi\nu t}$$

The meaning of the ”quantum-numbers” for a star:

- n : number of radial knots.
- l : total number of knots on the stellar surface.
- m : Number of knots that go over the poles.

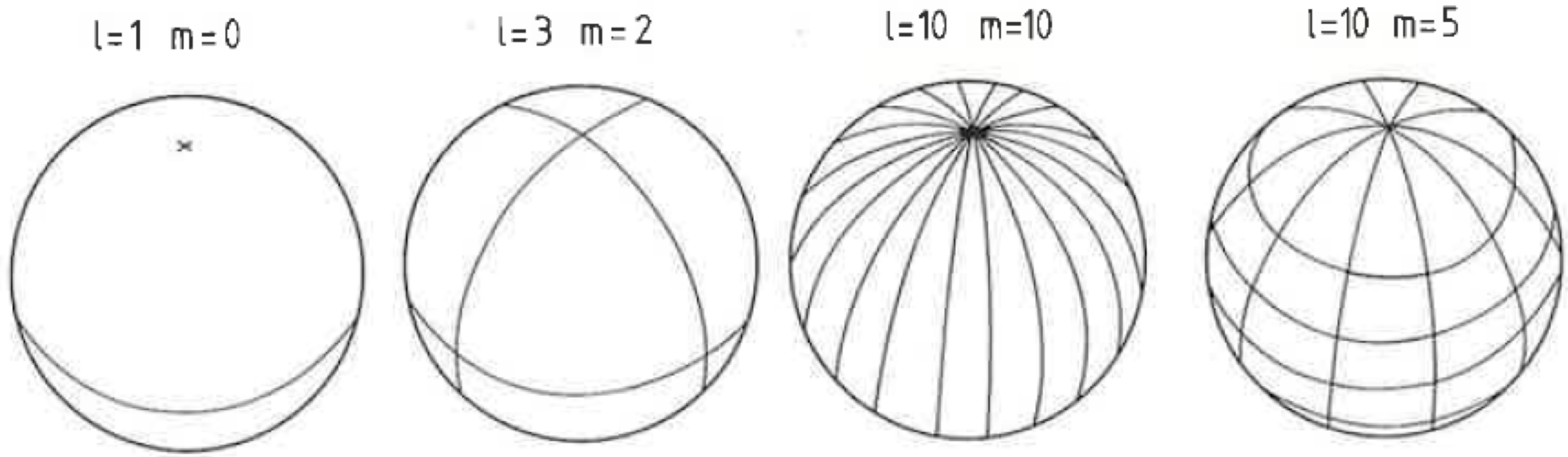


Fig. 5.6. Node circles of spherical surface harmonics. Adapted from Noyes and Rhodes (1984)

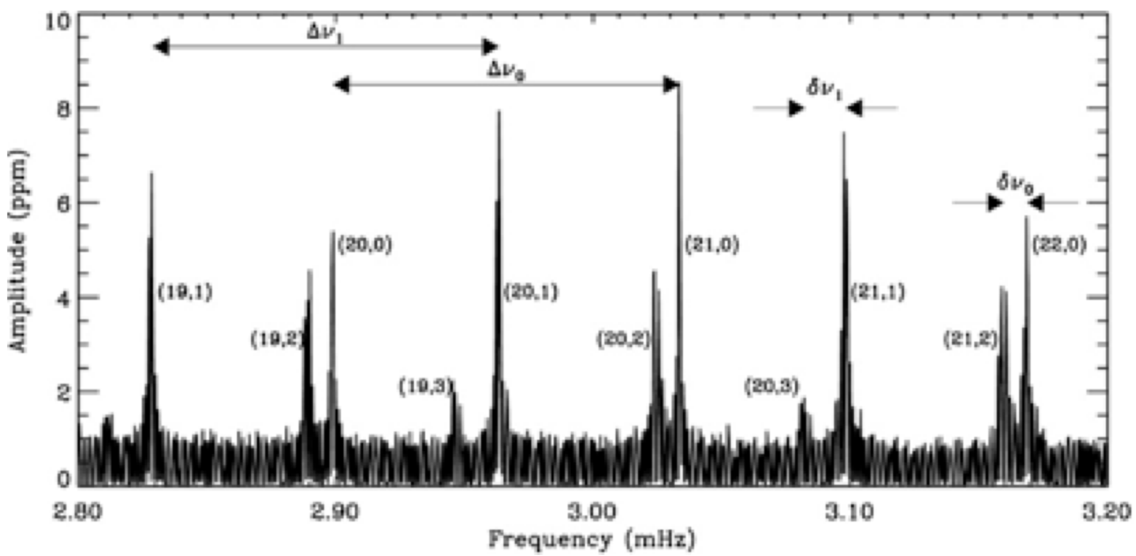
Large and small separations I

Low-degree acoustic modes accessible for distant stars probe their central regions. These are used to determine the overall parameters of the stars.

The oscillation frequencies of a star exhibit some regular patterns which allow diagnostic information on specific characteristics of the stellar structure to be derived.

This regularity can be clearly seen in the solar power spectrum:

The main characteristic of this spectrum is the approximately equal "**large separation**'' of approximately 135 micro-Hz between the larger peaks corresponding to p modes with **(n, l) and (n-1, l),** but important information is also encoded in the "**small separation**'' between peaks with nearly the same frequency corresponding to p modes with (**n, l) and (n-1, l+2).**

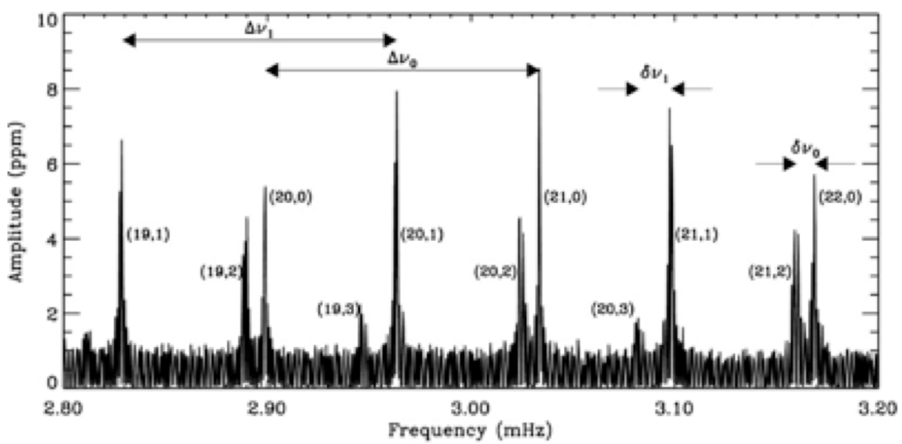


Large and small separations II

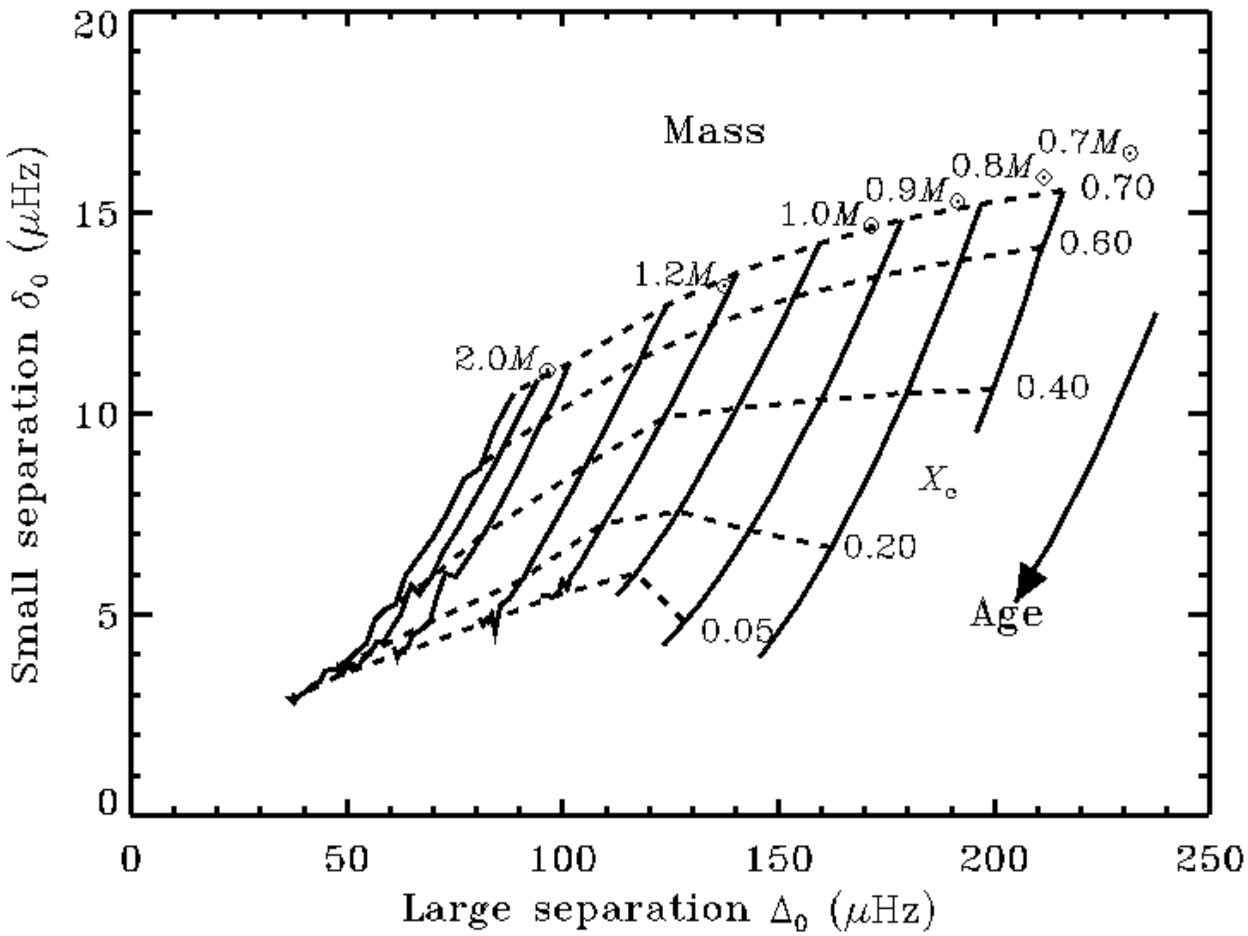
These two numbers alone give significant constraints:

The average **large separation** (n, l to n-1, l) contains information on the mean properties of the star, it is the **mean density** of the star.

The average **small separation** (n, l to n-1, l+2) is sensitive to the chemical composition in the central region, and can therefore measure the central hydrogen content, hence the **age** of the star.

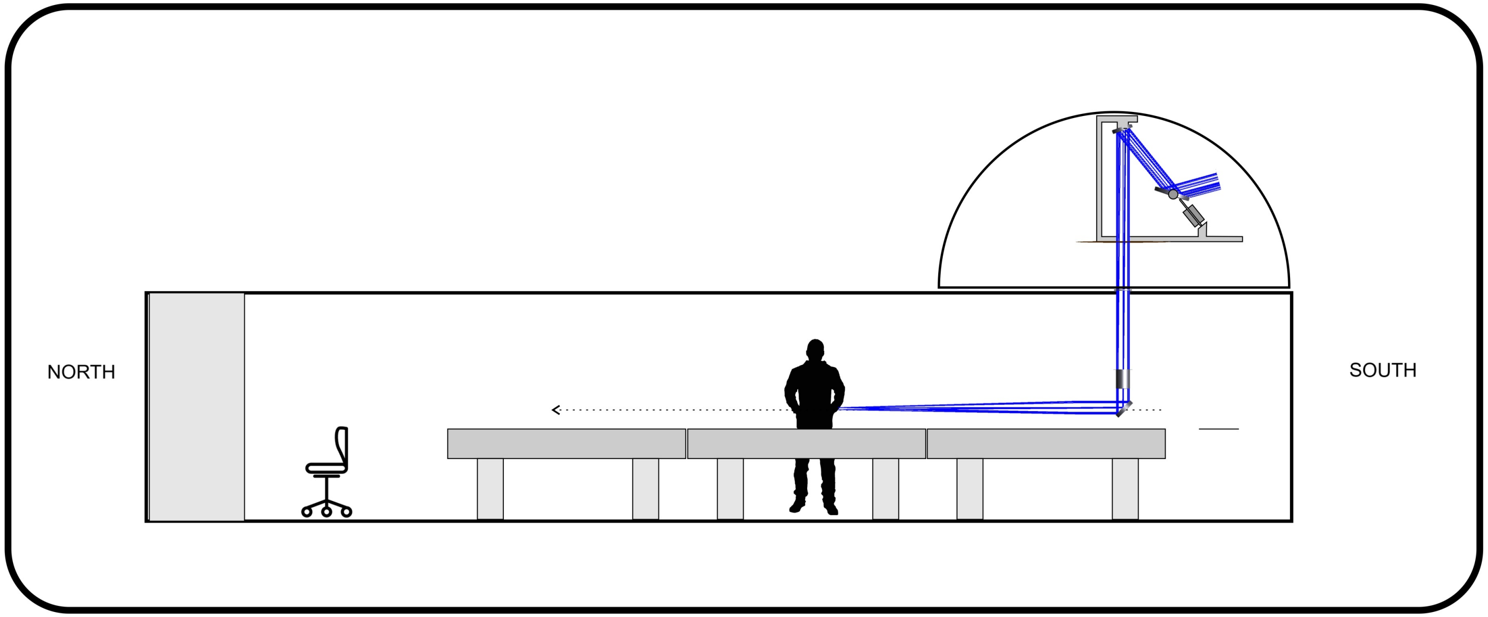


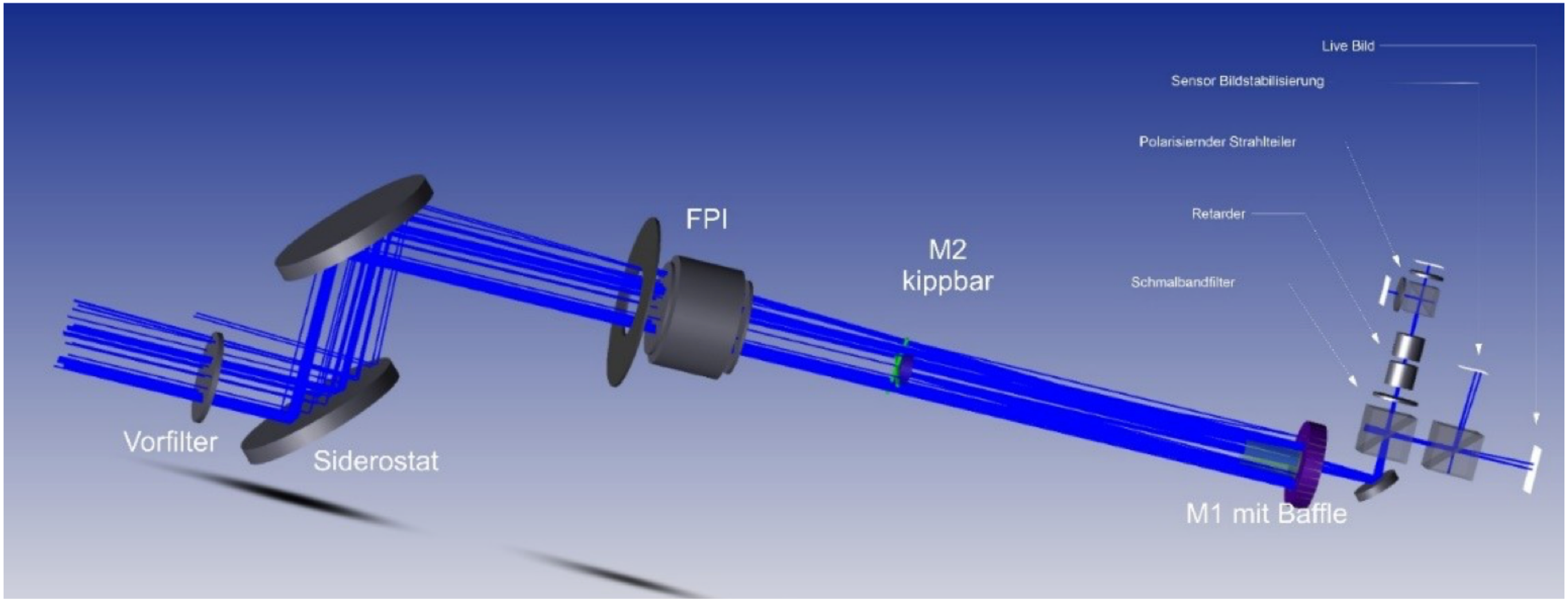
Asteroseismic HR diagram



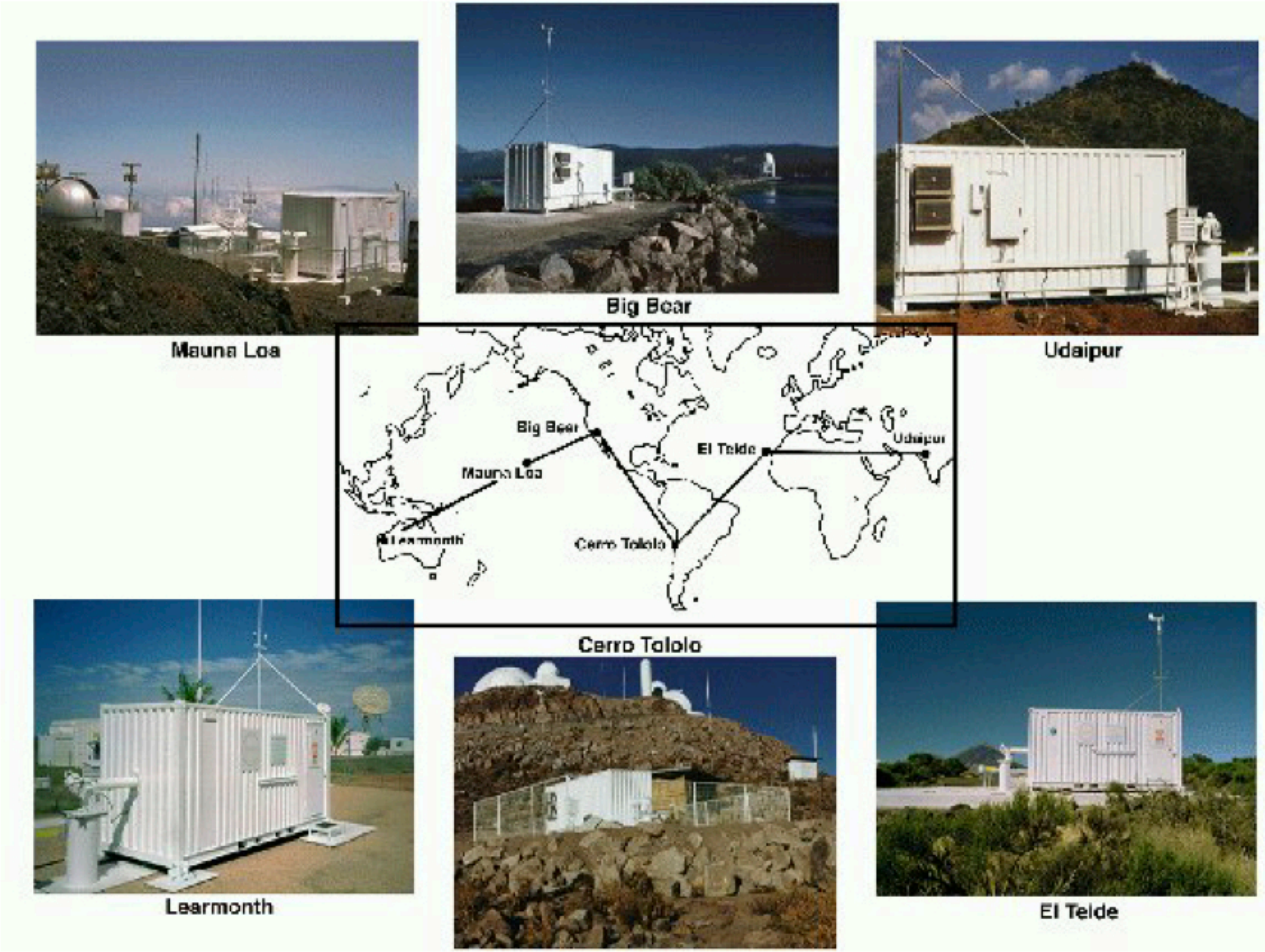
Solar Lab (“next generation GONG)

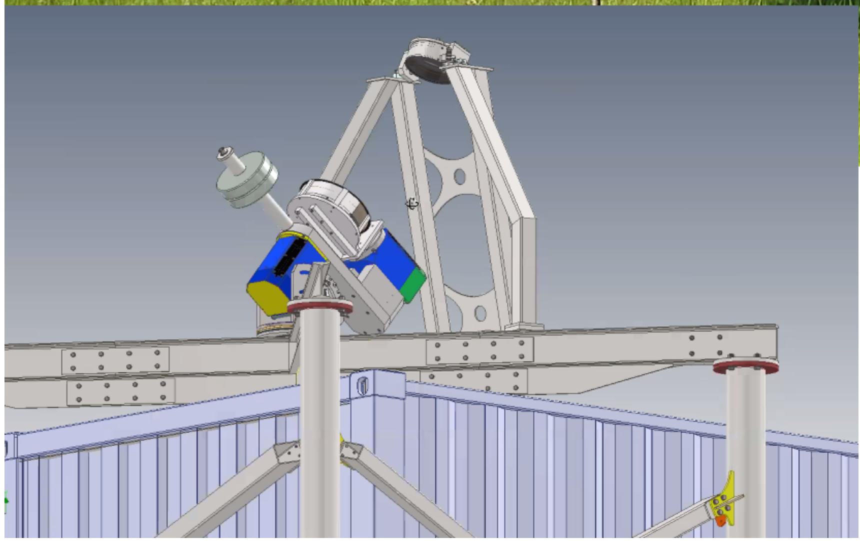
- Multi-line high-resolution magnetic observations of the Sun.
- 3-D magnetic topology of active region magnetic fields.
- First ground based continuous vector magnetometry for real time space weather predictions.
- Flare related changes in magnetic fields and electric currents in the Chormosphere.
- Long-term magnetic field records.

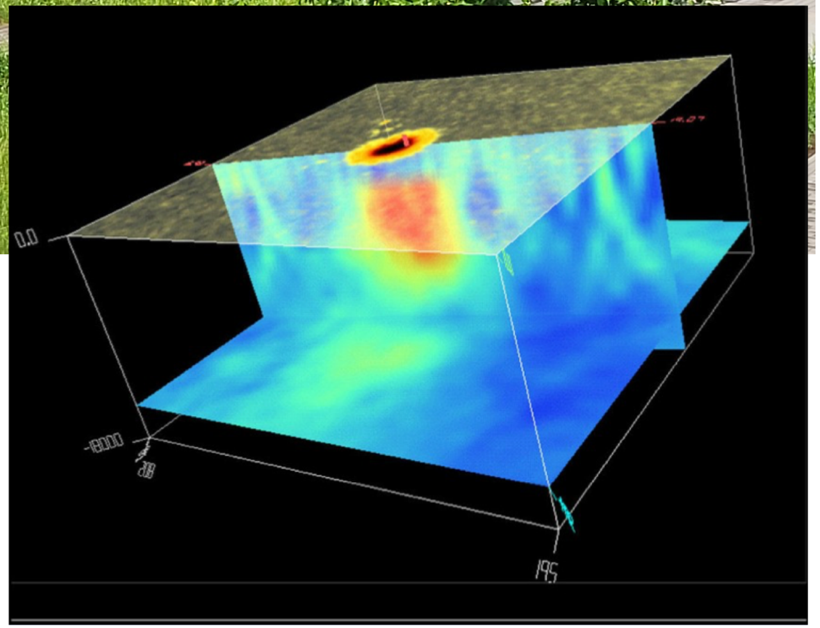

NORTH
SOUTH


Live Bild
Sensor Bildstabilisierung
Polarisiernder Strahlteiler
Retarder
Schmalbandfilter
FPI
M2
kippbar
Vorfilter
Siderostat
M1 mit Baffle

The solar oscillations are studied with a network of telescopes (GONG)






0.0
195