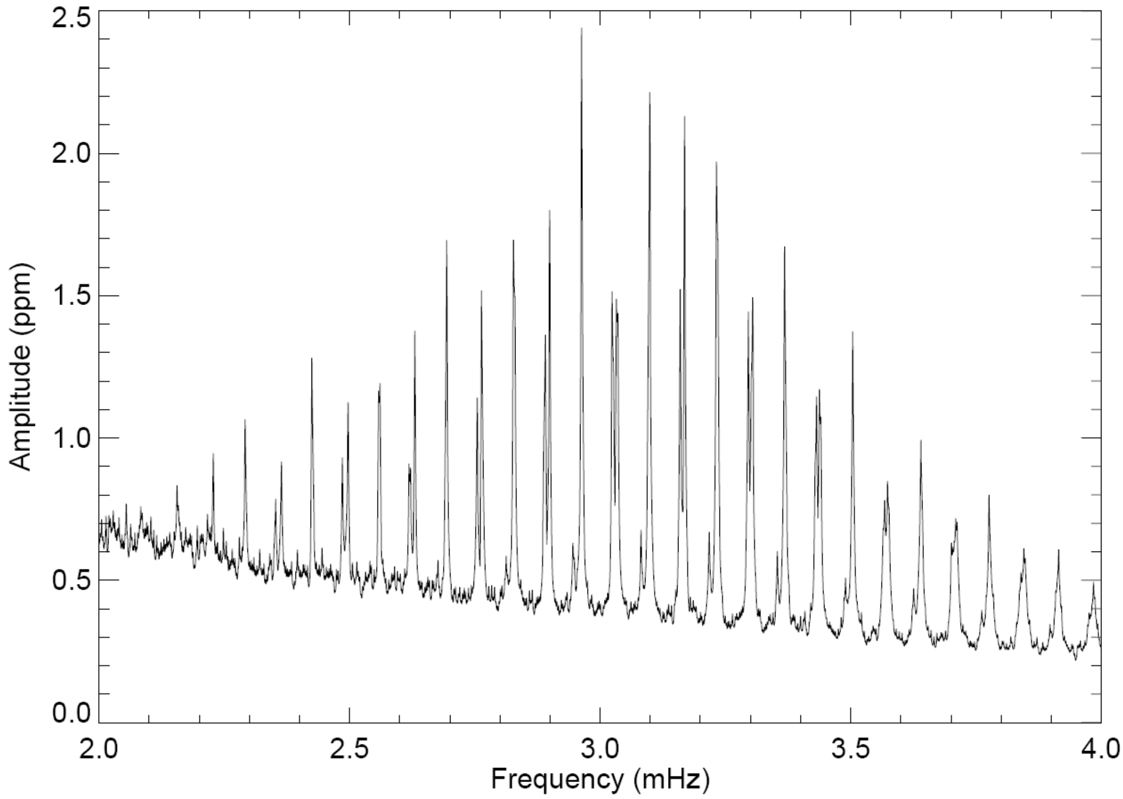
The KASOC Pipeline

– The Kepler Asteroseismic analysis pipeline –



Danish AsteroSeismology Centre - DASC
Department of Physics and Astronomy
University of Aarhus, Denmark

October 2007

The Kepler Asteroseismic analysis pipeline

Danish AsteroSeismology Centre - DASC
Department of Physics and Astronomy
University of Aarhus, Denmark

Document no.: **DASC/KASOC/0002(1)**
Date: 23.10.2007
Prepared by: Hans Kjeldsen, Jørgen Christensen-Dalsgaard, Ronald Gilliland and Torben Arentoft
Classification: **Open:** The document is unclassified and there are no restrictions in circulation.

1. Required activities for Kepler Asteroseismic Science Consortium

The required activities for the asteroseismic investigation within the Kepler project are described in the "Letter of Direction". The purpose of the Kepler Asteroseismic Investigation (KAI) is to ensure that full use is made of the Kepler time series data and that the full benefit of asteroseismology is provided for the Kepler investigations of extra-solar planetary systems. In order to structure the work that is needed for KAI we have established the Kepler Asteroseismic Science Consortium (KASC), a group of collaborating scientists and/or institutions.

The following acronyms will be used in the present document:

KAI	Kepler Asteroseismic Investigation
KASC	Kepler Asteroseismic Science Consortium
KASOC	Kepler Asteroseismic Science Operations Center
DASC	Danish AsteroSeismology Centre

1.1. Asteroseismology Products to be Provided to the Kepler Project

The activities required by KAI/KASOC is described in the "Letter of Direction" that can be downloadad from the following internet-address: http://astro.phys.au.dk/KASC/Letter_of_direction.pdf

KAI/KASOC will need to provide:

- Asteroseismic characterization of planet-hosting stars, including mass and particularly radius.

- Ability to distinguish cool giants from main-sequence stars through asteroseismic measures, such as estimates of the stellar radius or the shape of the background power spectrum characterizing the granulation time scale.

- Understanding of general stellar properties, including stellar structure modelling, contributing to stellar characterization.

1.2. KAI Activities

The KAI activities can be characterized as

- Selection of asteroseismic targets
- Characterization of asteroseismic targets
- Development and verification of data analysis pipeline
- Development and verification of data interpretation pipeline
- Developing stellar modelling techniques

The selection of asteroseismic targets shall aim at selecting enough targets such that 512 short-cadence (1-min) targets can be chosen for asteroseismic observations during the initial roll segment and at least 240 short-cadence (1-min) targets are available for change each quarter. Of the 240 targets available for asteroseismology, at least 140 may be selected by the KAI, and up to 100 may be selected by the Project from transit candidate stars meeting brightness and spectral type criteria expected to allow asteroseismic results. On top of this we need to select 100 or more (as resources permit) long-cadence targets specially for comparison with MS dwarfs. These can include cool giants to be observed for 3.5 years (and possibly longer if the mission is extended beyond the 3.5 years of initial operation), as well as beta Cephei stars, slowly pulsating B stars, and long-period delta Scuti stars.

As it is described in the "Letter of Direction" the KAI activities before launch are

- Develop and test "high-pass" filter, or other mechanism, to remove planet information from data to be used for asteroseismic investigation. This will take place in collaboration with the SOC and will be subject to the approval of the PI.

- Develop a pipeline to extract frequencies or frequency properties from observed time-series.

- Develop a pipeline to derive stellar properties from frequencies or frequency properties.
- Make available to the Project either a separate pipeline or other system to estimate stellar radius and/or identify giant stars, from power spectra based on long-cadence timeseries for cool giants. (Alternatively, at the discretion of the Project, use the power spectra provided by the SOC for these analyses.)
- Organize KASC to support these activities.
- Provide the selection of initial targets for asteroseismic program, as well as asteroseismic targets to be observed throughout the mission, for the Kepler Mission Planning.

After launch the KAI activities can be described as

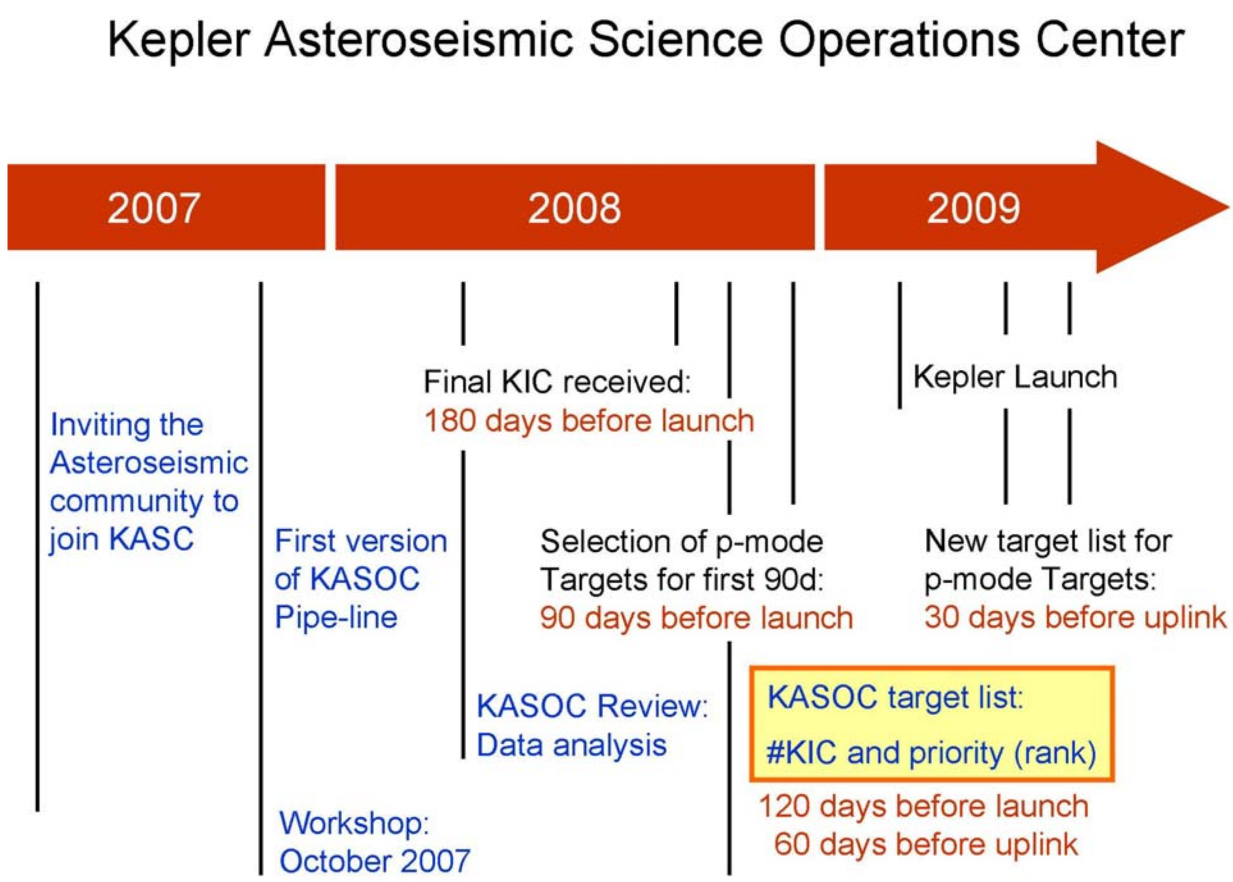
- Provide stellar parameters (particularly size) in a timely fashion to the Project; the goal is to provide the results in three months or less after receiving timeseries data.
- Perform asteroseismic analyses on any additional short-cadence targets upon request from the Project.
- Revise the asteroseismic target list as appropriate. In particular, identify cases suitable for asteroseismic analysis from those targets selected for short-cadence planet-transit observations.
- Comparison of the general stellar properties of cool MS stars with those of evolved stars.
- Ensure timely publication of the asteroseismic results.

2. Schedule for the KASC work: 2007- 2014

The schedule for the work within KAI, KASC and KASOC shall be coordinated such that it follows the overall Kepler mission development schedule. The selection of asteroseismic targets shall happen on a time scale that allows upload of targets to the onboard computer on Kepler and in order to ensure enough margin within the schedule both in relation to targets selection but also in relation to development and verification of the KASOC pipeline the schedule for the KAI work will be (note that the dates stated below are preliminary dates and they will be subject to revision):

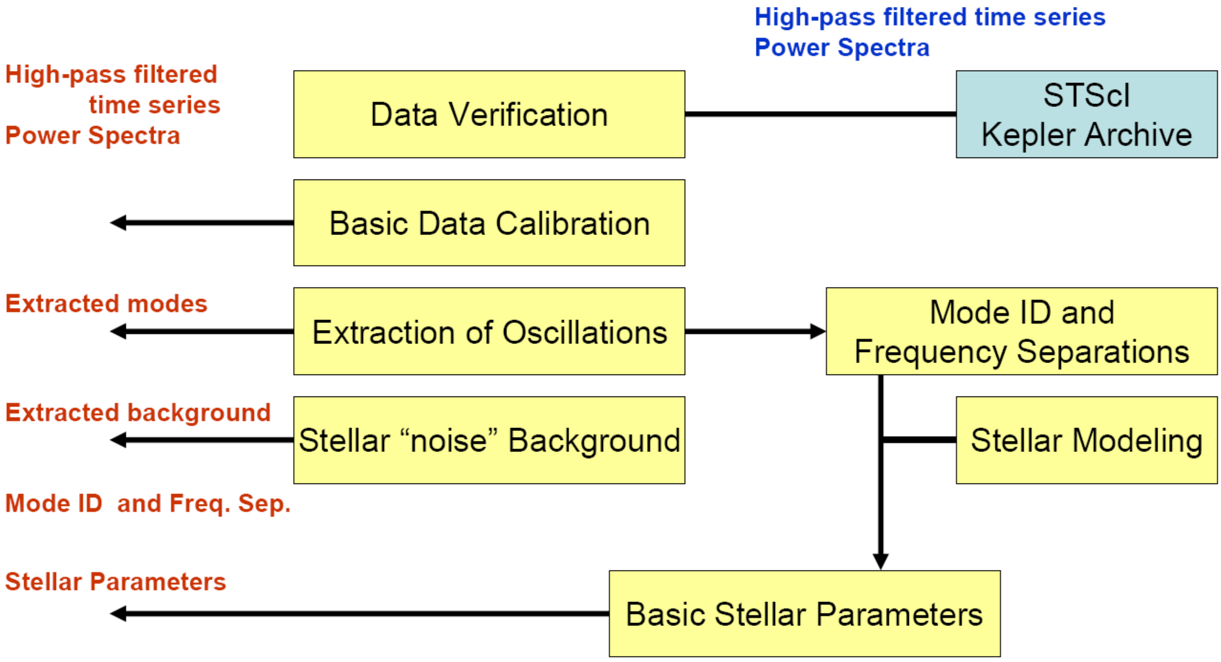
July-December 2007	KASC is set up, Work Packages (WP) will be defined and a first version of the KASOC pipeline will exist.
October 2007	First KASC workshop (October 29-31)
March 2008	KASOC Review #1: Data analysis review (internal)
June 2008	Second KASC workshop in Aarhus: Target Selection Procedure *The preliminary target selection will be based on the preliminary version of the KIC.* KASOC Steering Committee meeting
August 2008	Final Kepler Input Catalogue (KIC) received
September 2008	KASOC Steering Committee meeting: Target Selection
October 1, 2008	**First set of asteroseismology targets transferred to Kepler Science Office**
November 2008	KASOC Review #2: Final review (internal)

January 2009	Third KASC workshop: Data Analysis KASOC Steering Committee meeting
February 2009	**Kepler Launch**
March 2009	Commissioning (Kepler in space)
April 2009	Flight Operations begins (science phase, phase E)
May 1, 2009	**Second set of asteroseismology targets transferred to Kepler Science Office (uplink: July 2009).** Transfer of targets to Science Office for next three years: February 1, May 1, August 1 and November 1.
November 2009	**First set of data is ready for KASC analysis through KASOC.** Data release during next three years will be: February, May, August and November
November 2009	KASOC Steering Committee meeting: Data release
March 2010	Fourth KASC workshop: First Results of the KAI KASOC Steering Committee meeting
March 2011	Fifth KASC workshop KASOC Steering Committee meeting
October 2012	End of Flight Operations, Phase E (extended mission?)
April 2014	End of Data Analysis, Phase E



3. KASOC data analysis

The KASOC pipeline will contain a number of key elements that will extract and modify the time series data. At KASOC at DASC, University of Aarhus, a set of basic analysis tools already exists and those tools will be extended with input from KASC. A number of Work Packages will be defined to allow different KASC groups to provide analysis tools and analysis packages. There will be two kinds of software within KASOC and KASC. One will be the main pipeline software that will run on all data in a coherent and structured way. The basic elements in this pipeline will be:



The data flow from the Kepler satellite to KASOC will be:

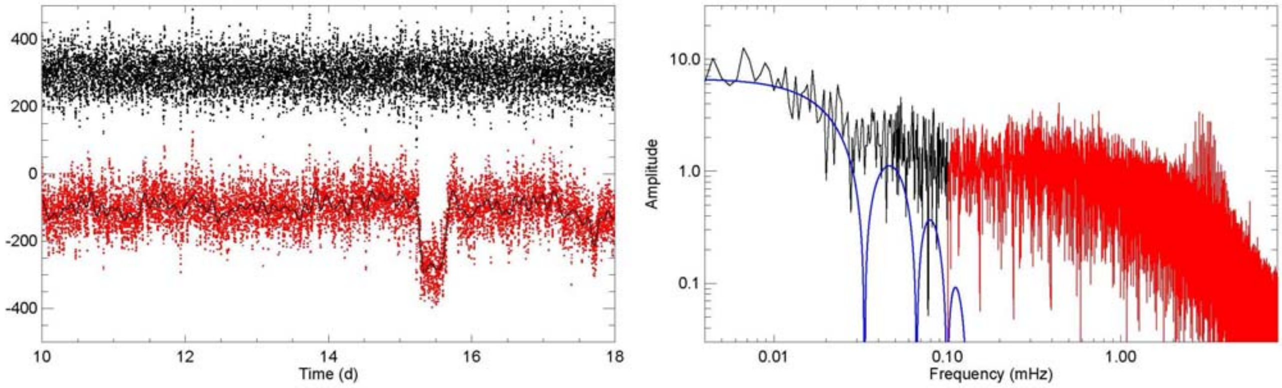
- Light curves (time series data) created at the Kepler Science Operations Center (SOC) are transferred to the STScI Kepler Archive and ingested.
- Kepler Co-I Ronald Gilliland retrieves short-cadence data for which transfer to the KASOC has been approved and applies the high-pass filter / transit removal operation to these data
- The resulting filtered light curves are transferred from Gilliland to the KASOC.

3.1. Data Verification

When data are received from Kepler Co-I Ronald Gilliland (after filtering) a basic verification will take place in order to ensure that the data quality requirements are fulfilled. Any anomaly will be flagged and in case of major anomalies or other unexpected problems with the data the STScI/Kepler Archive as well as Kepler Science Office will be contacted. Output from the data verification will be two main data sets:

1. High-pass filtered time series (with a cut-off frequency that does not allow any transit signal to leak through.
2. Power Spectra (un-filtered but with no phase information) with a sampling that is 10 times higher than critical sampling. In special cases one may get even higher over-sampling. It is important to note that the means for generating the power spectra and sharing those with KASOC has not yet been developed. This development is expected to take place during 2008.

Basic analysis of the time series data may also be performed at the Kepler Archive or at the Kepler Science Office and this may result in extracted oscillation mode parameters that can perhaps be made available for analysis within KASC.



An example of a high-pass filtered time series with a cut-off frequency of 100 µHz (corresponding to an oscillation period of 2 hr. 47 min.). The left figure shows at the top the time series after high-pass filtering and below the original time series with a simulated planet transit. At right the amplitude spectrum is shown. The red part of the spectrum shows the part of the spectrum that is unaffected by the high-pass filter while the black part shows the amplitude of the spectrum that is removed from the time series. The blue curve shows the amplitude spectrum of a time series with a single planet transit. As one can see most of the power for a transit is concentrated at frequencies below the cut-off frequency for the high-pass filter.

3.2. Basic Data Calibration

At KASOC we will have access to calibration data that will allow a series of calibrations (amplitude, frequency, time, false signals etc.) to be carried out. The raw data that will be available from KASOC will contain this data calibration. Examples of basic calibration data that may be available are Satellite Attitude Control (ACS) information (roll, pitch and yaw time series data), telescope focus information, CCD temperatures and particle radiation levels.

3.3. Extraction of Oscillations

This will be the main analysis of the data. Fourier analysis of high-pass filtered data will be used to extract frequencies, amplitudes and phases (and in some cases mode energy, mode lifetime etc.) for the individual oscillation modes. Power spectra – if available - will be used to extract frequencies, amplitudes, mode energy and mode lifetime also for low-frequency modes where no information will be available from the high-pass filtered time series.

3.3.1. Fourier Analysis

Fourier Time Series Analysis is a classical discipline and within the Kepler Asteroseismic Science Consortium several research groups and individuals have a long and world-leading experience on time series analysis. The KASOC pipeline will therefore contain several parallel analysis lines and through comparison we will get information on the quality of the extracted information as well as information on the robustness of used routines and algorithms.

Although the Kepler Time Series data are expected to be uniform both in quality per data point and in duty cycle throughout nominal operation, we need to be able to correct for individual bad data points and periods of non-uniform sampling. One way of doing this is by calculating a weighted power spectrum, which is done by use of the following equations. α and β are orthogonal components containing the amplitude of a weighted fit to sine (α) and cosine (β). We will use the cyclic frequency ($\nu = \omega / 2\pi$) for all KASOC pipeline analysis.

$$P(\nu) = \alpha(\nu)^2 + \beta(\nu)^2$$

$$\alpha(\nu) = \frac{s \cdot cc - c \cdot sc}{ss \cdot cc - sc^2}$$

$$\beta(\nu) = \frac{c \cdot ss - s \cdot sc}{ss \cdot cc - sc^2}$$

$$s = \sum w(t) \cdot f(t) \cdot \sin(2\pi \cdot \nu \cdot t)$$

$$c = \sum w(t) \cdot f(t) \cdot \cos(2\pi \cdot \nu \cdot t)$$

$$ss = \sum w(t) \cdot \sin^2(2\pi \cdot \nu \cdot t)$$

$$cc = \sum w(t) \cdot \cos^2(2\pi \cdot \nu \cdot t)$$

$$sc = \sum w(t) \cdot \sin(2\pi \cdot \nu \cdot t) \cdot \cos(2\pi \cdot \nu \cdot t)$$

where *w(t)* is the statistical weight, *v* is the frequency and *f(t)* is the Kepler time series data. *α(v)* and *β(v)* contain the phase information for a given frequency and the square root of the power *P(v)* is the amplitude. The power spectrum will be calculated for high-pass filtered data.

Using *α(v)*, *β(v)* and *P(v)* we are able to locate and extract the stellar oscillations. The classical inversion problem is to minimize the following term, given *N* data points and *n* frequencies:

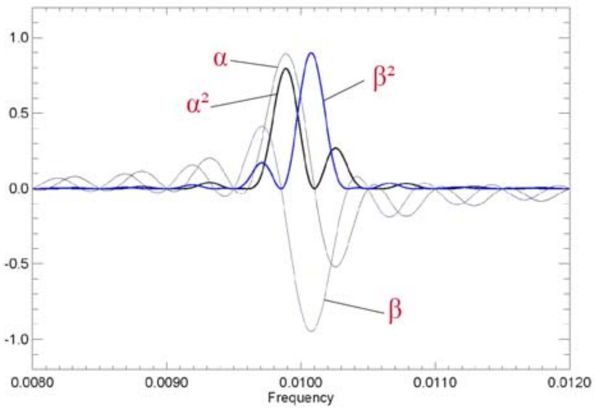
$$\left(\sum_{j=1}^{N} \left[f(t_j) - \sum_{i=1}^{n} \left(\alpha(\nu_i) \cdot \sin(2\pi \cdot \nu_i \cdot t_j) + \beta(\nu_i) \cdot \cos(2\pi \cdot \nu_i \cdot t_j) \right) \right] \right)^2$$

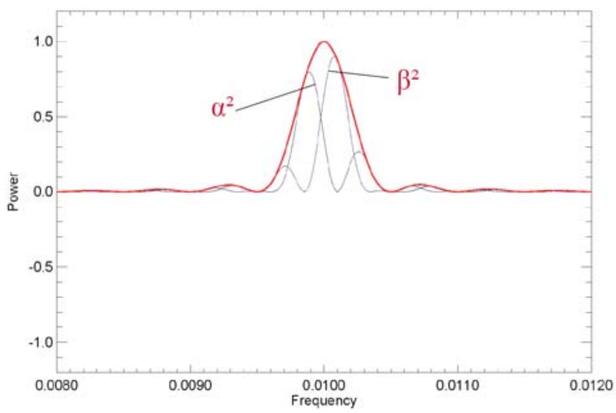
Several routines for solving this inversion problem exist in the literature and within KASC. In principle this inversion could take place before KASOC get the data, however this issue is still to be decided on.

One of the routines that will be used to analyze the Kepler data will be the so-called iterative sine-wave fitting which is an iterative CLEANing process. The process is just one of several routines that exist for solving the inversion problem. Iterative sine-wave fitting contains the following steps:

1. The power spectrum *P(v)* is calculated for all frequencies for *f(t).*

2. The frequency v_{max} for the highest peak in the power spectrum is located and *α* and *β* is calculated for this frequency. Below we show an example of α, β, α^2, β^2, $\alpha^2+\beta^2$ as a function of frequency. The frequency v_{max} corresponds to the frequency for the max. value of the power (= $\alpha^2 + \beta^2$).

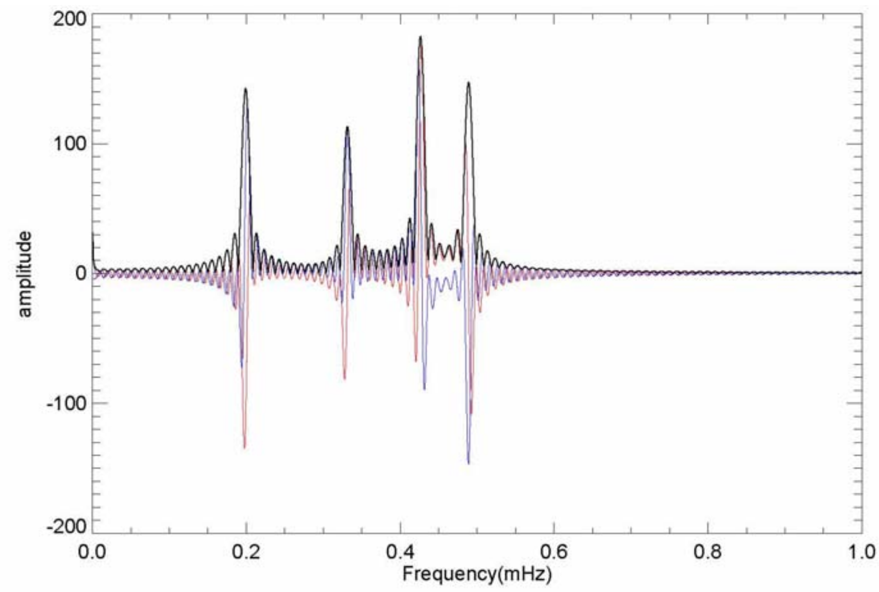




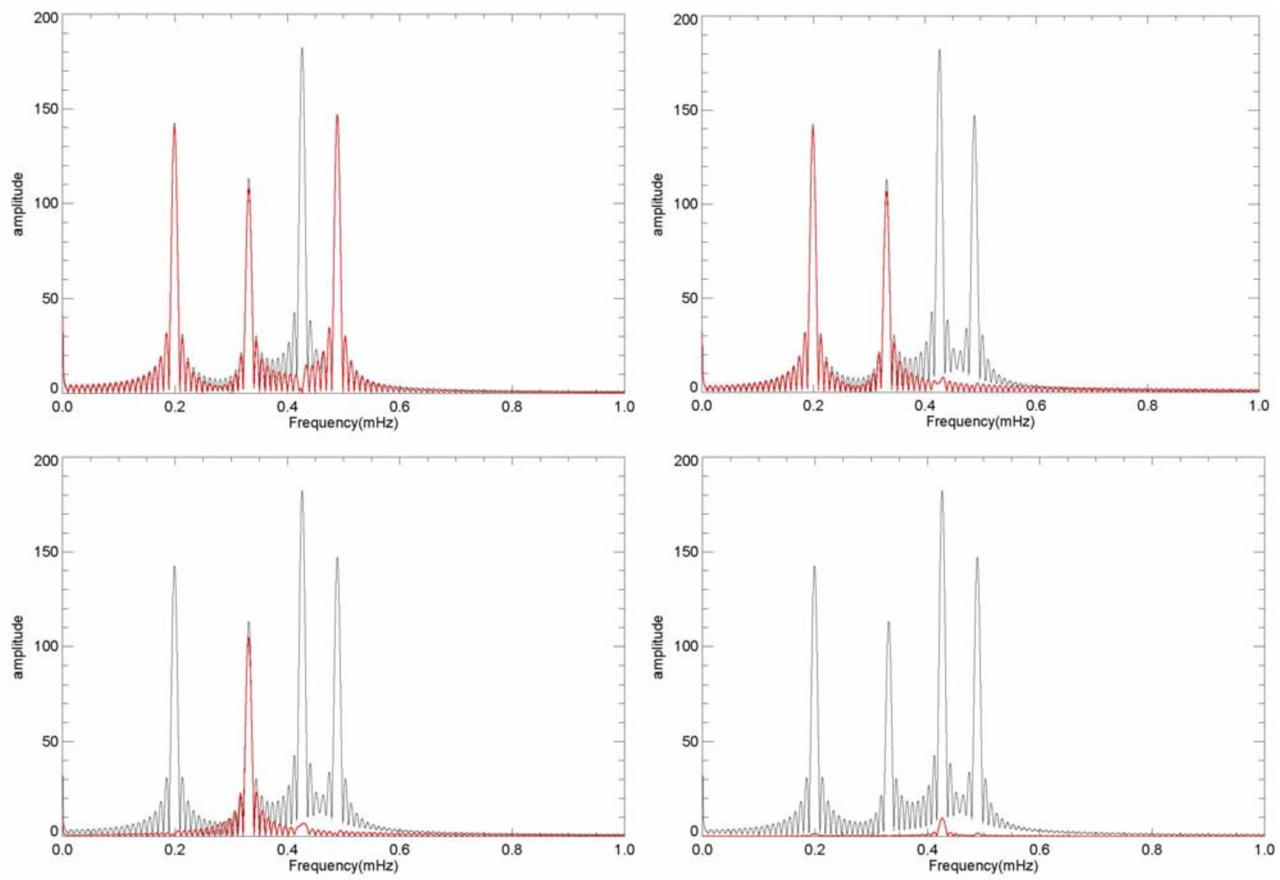
3. A new time series is constructed as (for all *t*):

$$f(t) - \alpha(\nu_{\max}) \cdot \sin(2\pi \cdot \nu_{\max} \cdot t) - \beta(\nu_{\max}) \cdot \cos(2\pi \cdot \nu_{\max} \cdot t)$$

4. Step 1-3 is repeated until the highest peaks in the power spectrum are insignificant compared to the noise level.



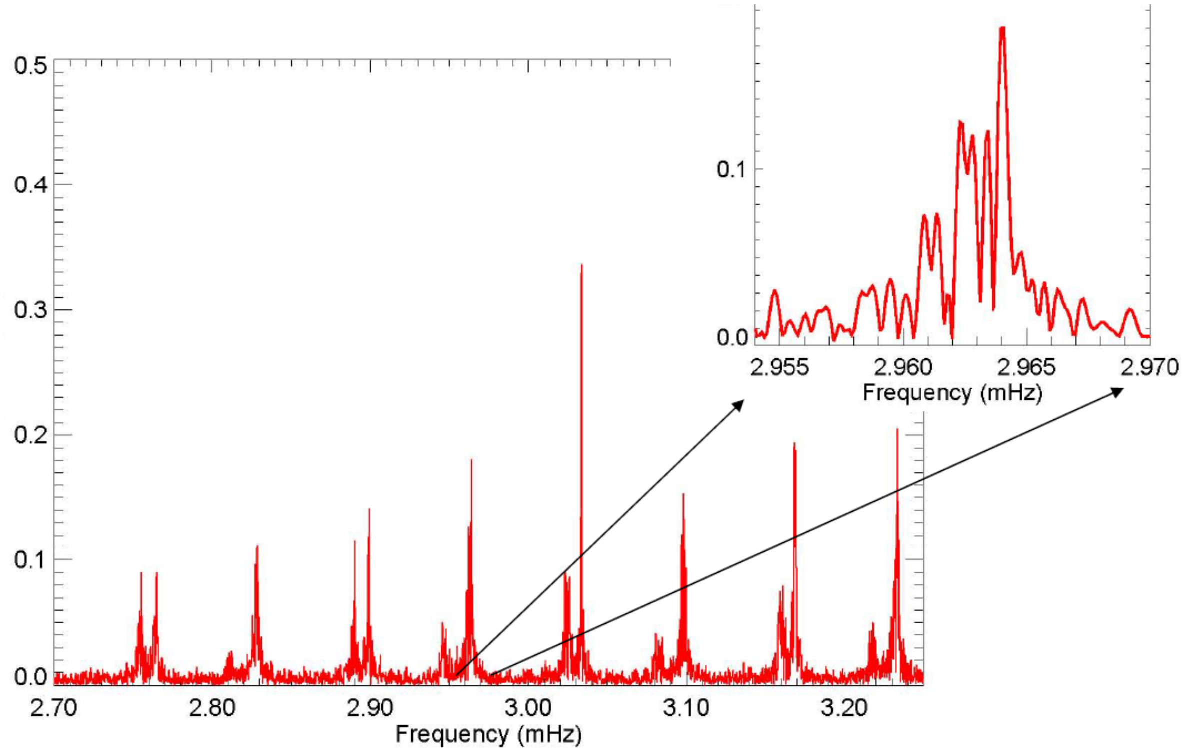
At left an example of four oscillation peaks is seen. Amplitude, α and β are shown (red and blue curves) as a function of frequency. The iterative sine-wave fitting for those data is shown in the four figures below. The black curve is the original spectrum and the red curves show the amplitude spectrum after removing one, two, three and four oscillations via iterative sine-wave fitting.



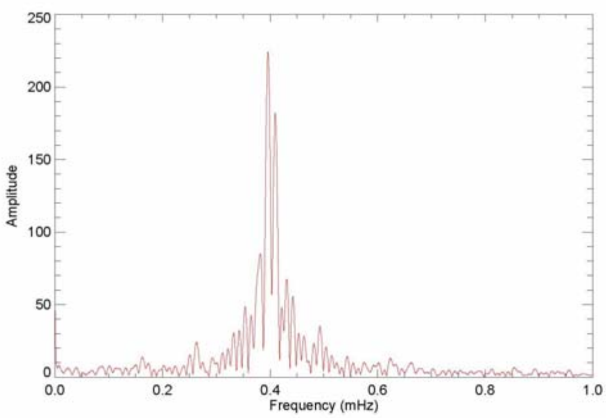
The result of the above described process is a series of extracted frequencies, power for those frequencies and phases (give via α and β). The process has only physical meaning if the oscillations can be described as a set of coherent oscillations with constant amplitude. For classical pulsating stars this is a good description of the oscillations; however, for solar-like oscillations one will need to deal with the fact that any oscillation mode has a finite lifetime and that a mode is re-excited throughout the data series. The iterative sine-wave fitting will not be an optimum process for those types of oscillations and we therefore need other techniques which are known and used in time series analysis for helioseismology. Analysis of the power spectrum is in this case the preferred technique.

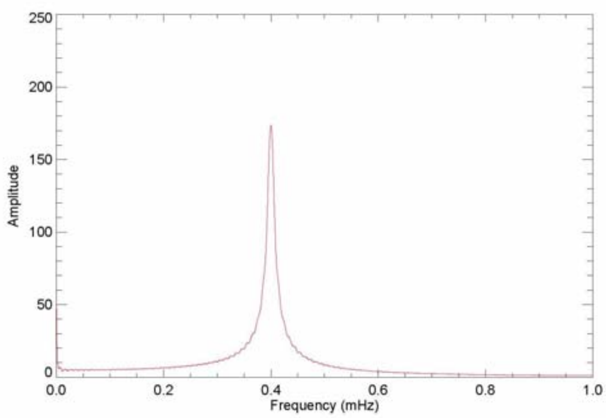
3.3.2. Power Spectrum analysis

For evolved stars (red giants) and long period oscillators (beta Cepheid stars and delta Scuti stars) information on low frequency oscillations will in the beginning of the mission only be available through analysis of the power spectrum (with no phase information). For damped and re-excited oscillators we also only need the power spectrum to extract the basic oscillation properties since we in this case fit the power spectrum directly. At present the details of the transit removing filter has not been finally decided, however a simple power spectrum with no phase information will in general not contain transit information, since a transit can only be regenerated from the power spectrum if phase information is available. However, the power spectrum will contain information on the existence of inner orbit planets in modulation of reflected light, thus placing limits on the general availability of power spectra for wide dissemination.



As shown above (where a part of the solar p-mode spectrum is seen) energy (or power) for an oscillation mode is not concentrated in a single frequency. The reason for this is the finite mode lifetime which will result in mode energy being spread over a frequency range. The shorter the mode lifetime the more the power will be spread.





The power for a damped and re-excited oscillator will be distributed as a Lorentzian profile for a given observing period (window) which can be parameterized by mode lifetime τ, size of peak H, and frequency ν. The amplitude spectrum for the Lorentzian profile is seen at right and the amplitude spectrum for a damped and re-excited oscillator is seen at left.

The total power from the stellar oscillations are therefore described as

$$P(\nu) = \sum_{i=1}^{n} L(\tau_i, H_i, \nu_i)$$

where $L(\tau,H,\nu)$ is the Lorentzian profile for the observing window for a mode lifetime τ, size of peak H, and frequency ν. The Lorentzian profile is defined as

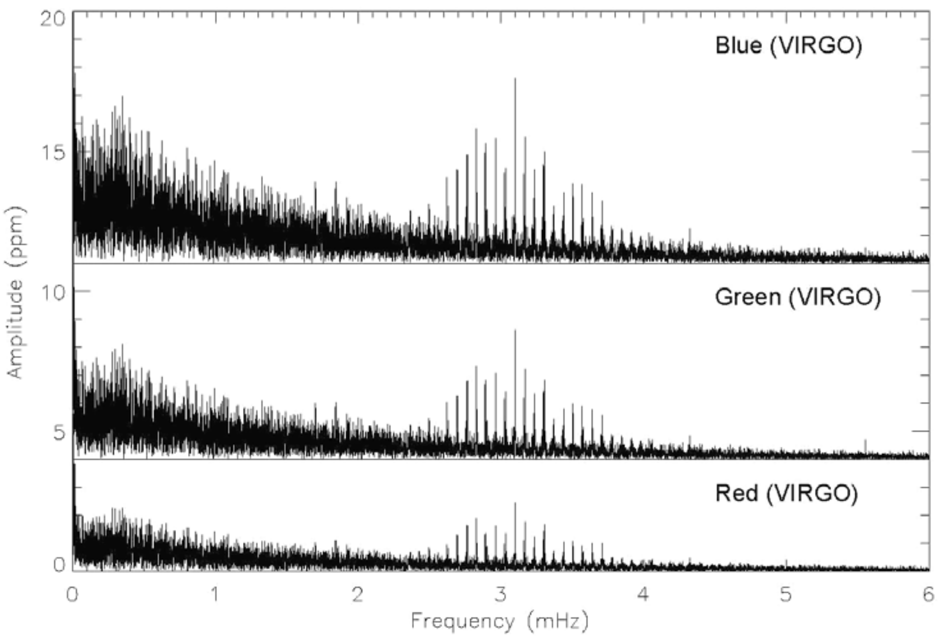
$$P_{Lorentzian}(\nu) = L(\tau_i, H_i, \nu_i) = \frac{H_i}{1 + \left(2\pi \cdot (\nu - \nu_i) \cdot \tau_i\right)^2}$$

The FWHM (cyclic frequency) of the power of the Lorentzian profile is

$$FWHM = \frac{1}{\pi \cdot \tau_i}$$

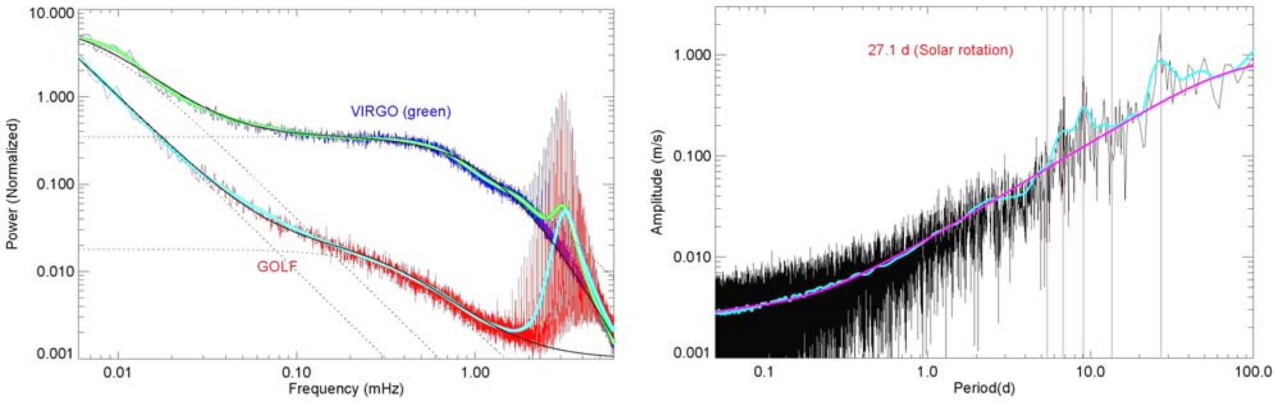
In the case of the Sun it has been observed that the individual oscillation peaks do indeed approximately follow a Lorentzian profile although data with very high SNR has shown the profiles to be slightly asymmetric indicating that the damping and re-excitation of the p modes are not fully described by a simple damped and re-excited oscillator.

3.4. Extraction of stellar background "noise"



The stellar background "noise" contains a lot of information on stellar surface activity, stellar rotation and stellar granulation. At left we show a 20 day series from the VIRGO instrument onboard the SoHO spacecraft (in three colours). The rise in the background at low frequencies is due to the granulation noise and a fit to the power at low frequencies will therefore provide information on the level of granulation.

The two figures below show solar power and amplitude as a function of frequency and period. One can clearly extract power components from granulation, activity and solar rotation (right panel). For other stars than the sun similar analysis will lead to extraction of general properties for granulation, rotation and stellar activity. In the figures below one can at the left panel see the background for intensity (VIRGO) and velocity (GOLF). The p-mode signal can be seen at 3 mHz and the background is described by a two component fit (granulation and at low frequencies solar activity). The right panel shows the velocity amplitude as a function of period ($1/\nu$). The main signal is coming from solar spot activity. The vertical lines indicate signal from the solar rotation, including the high order components of the rotation (P, P/2, P/3, P/4 and P/5). It demonstrates the capability to measure stellar surface rotation from the Kepler time series.



The fit to the power spectrum for a given Kepler time series will therefore also contain a background component that will be modelled as (in agreement with the theoretical models for stellar "noise" background), where the last component is the rotational signal caused by stellar spot activity:

$$P(\nu) = P_{white} + \sum_{j=1}^{M} \frac{4 \cdot \sigma_j^2 \cdot \tau_j}{1 + (2\pi \cdot \nu \cdot \tau_j)^2} + \sum_{i=1}^{N} L(\tau_{spots}, H_i, i \cdot \nu_{ROT})$$

where *M* is the number of "Harvey"-like background noise components included in the fit (typical 4-5) and *N* is the number of rotational related peaks (main rotation and high order rotational components). Analysis of solar background "noise" has demonstrated that a "Harvey background" (see Harvey J. 1984, In *Probing the depths of a star: the study of the solar oscillations from space,* ed. R. W. Noyes & E. J. Rhodes Jr., (Pasadena, JPL/NASA), 327) provides a good fit to the observed power distribution. In the left figure, the dotted lines indicate such components from (in order of decreasing frequency) granulation, supergranulation and activity. In the equation above we describe the rotational signal as a series of Lorentzian profiles. This may not turn out to be the correct description in all cases, however if the activity is characterized as spots that are appearing and disappearing with a typical time scale of τ_{spots} one should expect the profile to be represented by a Lorentzian profile.

Based on the above considerations the fit to the total power spectrum is characterized as:

$$P(\nu) = \sum_{i=1}^{K} L(\tau_i, H_i, \nu_i) + P_{white} + \sum_{j=1}^{M} \frac{4 \cdot \sigma_j^2 \cdot \tau_j}{1 + (2\pi \cdot \nu \cdot \tau_j)^2} + \sum_{i=1}^{N} L(\tau_{spots}, H_i, i \cdot \nu_{ROT})$$

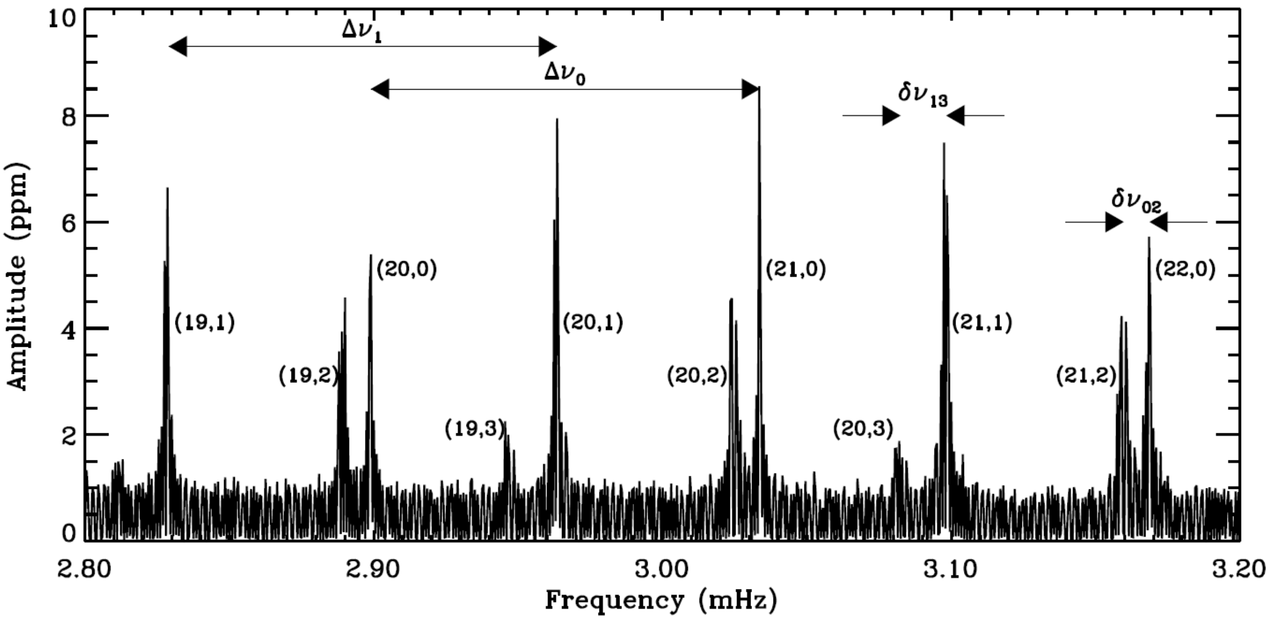
where *K* is the number of oscillation modes above the noise.

3.5. Mode ID and extraction of frequency separations

From the power spectrum it is not possible to identify the order and degree of a given oscillation mode. This is, however, possible for solar-like oscillations through the p-mode structure. A solar-like star is expected to oscillate with frequencies that to first order can be described according to the asymptotic relation:

$$\nu_{n,l} \approx \Delta\nu \cdot (n + l/2 + \varepsilon) - (l+1) \cdot l \cdot D_0$$

where *l* is the mode degree and *n* is the mode order. Based on the asymptotic relation one may define a characteristic set of frequency separations that describes the structure of the oscillation frequencies for a solar-like star.



The classical separations are the so-called large separations (that contain information on the mean density of the star):

$$\Delta\nu_0(n) = \nu_{n+1,0} - \nu_{n,0}$$
$$\Delta\nu_1(n) = \nu_{n+1,1} - \nu_{n,1}$$
$$\Delta\nu_2(n) = \nu_{n+1,2} - \nu_{n,2}$$
$$\Delta\nu_3(n) = \nu_{n+1,3} - \nu_{n,3}$$

$$\Delta\nu_0 \approx \Delta\nu_1 \approx \Delta\nu_2 \approx \Delta\nu_3$$

and the small separations (that are sensitive to the core helium content):

$$\delta\nu_{01}(n)_0 = \frac{1}{2}\cdot(\nu_{n,0} + \nu_{n+1,0}) - \nu_{n,1} \approx 2D_0$$
$$\delta\nu_{01}(n)_1 = \nu_{n+1,0} - \frac{1}{2}\cdot(\nu_{n,1} + \nu_{n+1,1}) \approx 2D_0$$
$$\delta\nu_{02}(n) = \nu_{n+1,0} - \nu_{n,2} \approx 6D_0$$
$$\delta\nu_{13}(n) = \nu_{n+1,1} - \nu_{n,3} \approx 10D_0$$

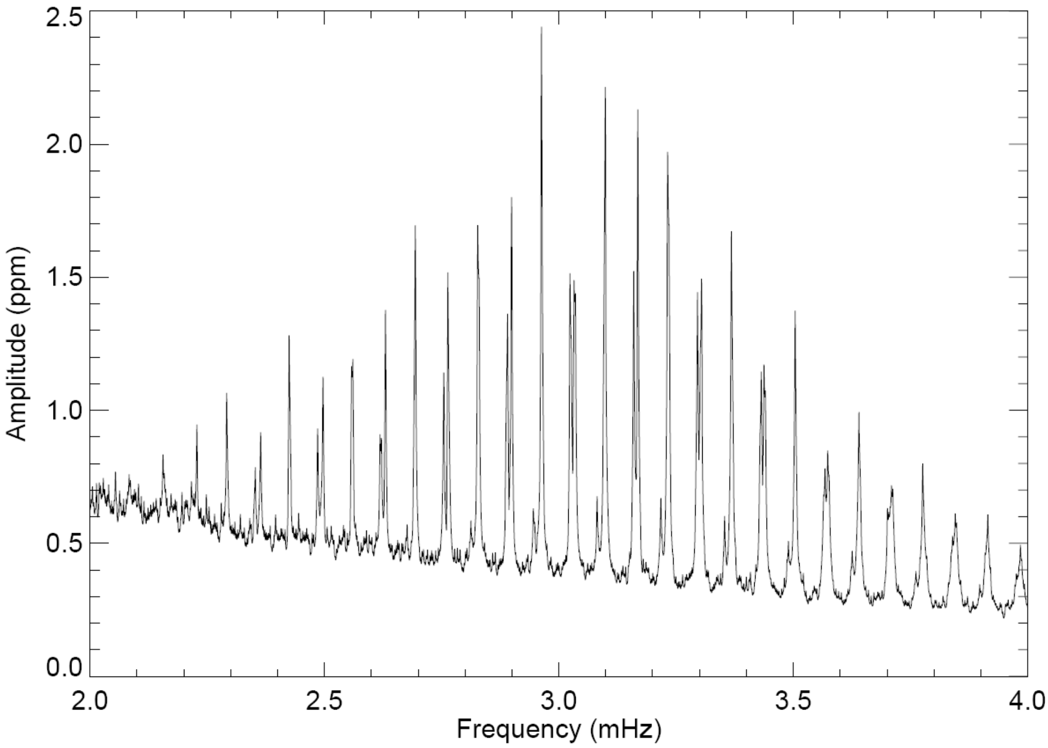
$$5\cdot\delta\nu_{01}(n) \approx \frac{5}{3}\delta\nu_{02}(n) \approx \delta\nu_{13}(n)$$

Extracting the small and large separations for a solar-like star is therefore one of the main goals of the KASOC pipeline analysis. Extracting this information will allow a mode identification which is crucial for any comparison between a stellar model and the observed oscillation frequencies.

In order to extract the separations one will need to estimate the regularity of the power spectrum. There are several methods used for this process and we expect to include all those methods in the KASOC pipeline. The major methods are:

- Power spectrum auto-correlation (with or without a weight envelope)
- Power spectrum cross-correlation (between the observed spectrum and a model spectrum)
- Power of power spectrum (with or without a weight envelope)
- Matched filter (between the observed spectrum and a model spectrum)

The tests performed at the KASOC preliminary pipeline show that the Matched filter algorithm seems to work for quite faint stars and this technique is therefore implemented in the first version of the KASOC pipeline. The figures below demonstrate how the Matched filter routine works. The first part of the process contains the construction of a smoothed power spectrum. An example of such a smoothed amplitude spectrum is seen below (the figure shows the mean amplitude of 52 one-week power spectra (1 year) for a solar "twin". The time series contain only stellar noise).

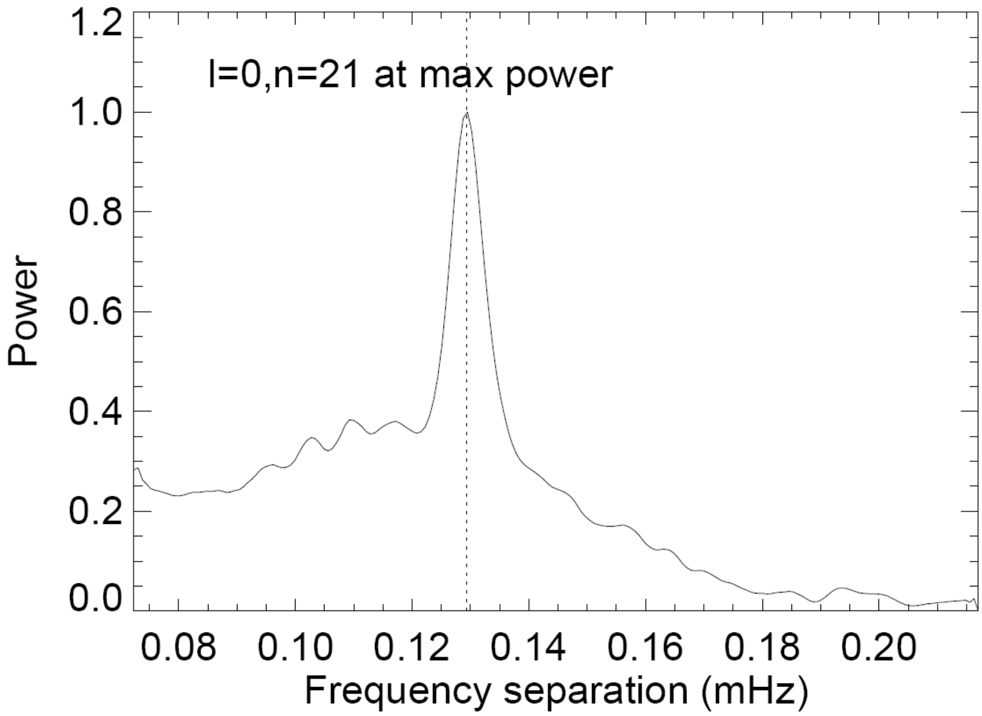


Based on this spectrum a model spectrum is constructed that basically contains power at the frequencies defined by the asymptotic relation. The observed smoothed power spectrum and the model spectrum are then matched and the power of the matched filter is calculated for each model spectrum. The best match is then assumed to be the correct p-mode spectrum. An example of such a match is shown below for a simulated high SNR smoothed power spectrum. A very significant peak is detected at the right value for the large separation.

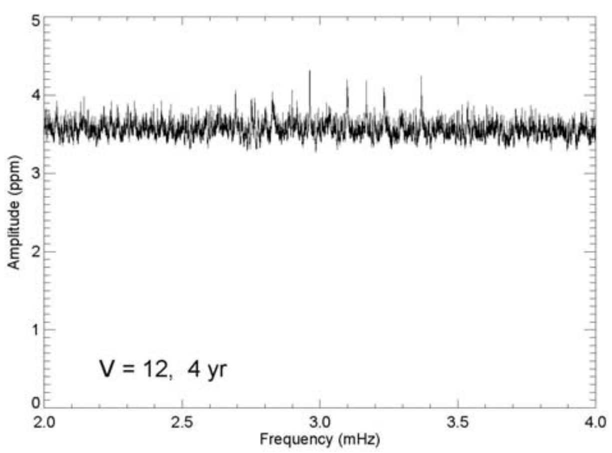
The values for the radial orders, *n* (and the angular degree, *l*) will be extracted by fitting to the asymptotic relation

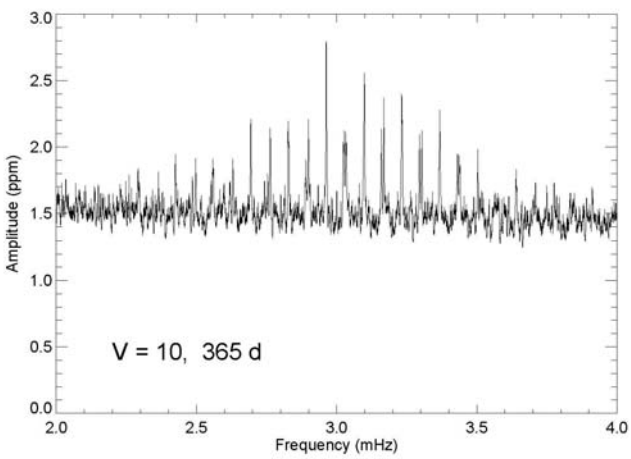
$$\nu_{n,l} \approx \Delta\nu \cdot (n + l/2 + \varepsilon) - (l+1) \cdot l \cdot D_0$$

And assuming the value for ε is between 1 and 2 (which is true for all stellar models of solar-type stars)



The simulation shows that for most cases where the large separation can be extracted one is also able to extract the small separation as well as the *n* and *l* value for maximum power. Below we show two examples of Kepler amplitude spectra for solar "twins" at magnitude V=12 and V=10.





The parameters that will be extracted from the pipeline for solar-like stars will be (the parameters at the top will be detected in the faintest stars and/or shorter observation series, while only for the brightest targets will we be able to detect all the individual frequencies):

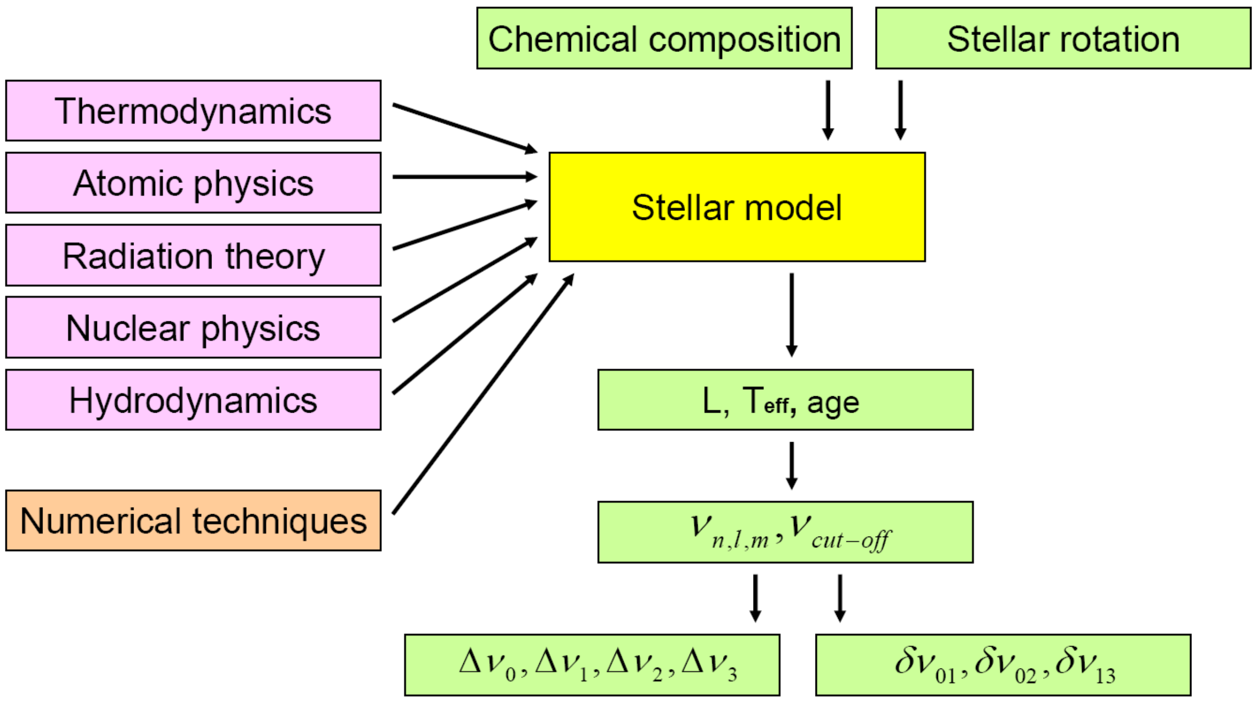
- Large separation: $\Delta\nu$
- Small separation: $\delta\nu_{01}, \delta\nu_{02}$
- Frequency at peak power and identification of radial orders for main peaks: $\nu_{n(\max),l(\max)}$
- Summed power from *l* = 3 modes via the small separation: $\delta\nu_{13}$
- Individual modes near peak power: $\nu_{n,l}$

- Echelle diagram and individual values for large separations: $\Delta\nu_0, \Delta\nu_1, \Delta\nu_2, \Delta\nu_3$
- Individual modes far from peak power, rotational splitting (*m*), mode lifetime: $\nu_{n,l}$

3.6. Stellar Modelling

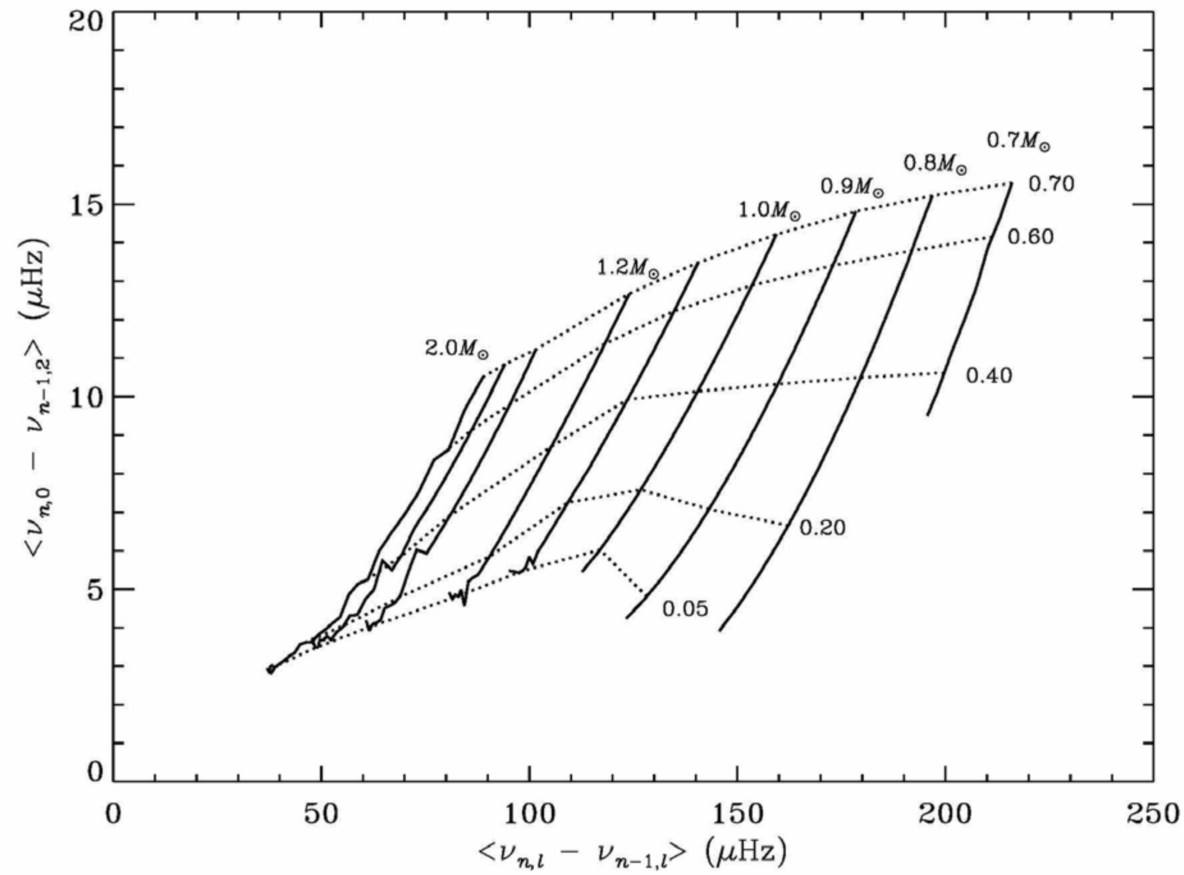
The KASOC pipeline is not only aimed at producing time series data and extracting oscillation frequencies. A very important part of the pipeline is the extraction of basic stellar parameters for any star that is observed under KAI.

In order to be able to perform pipeline extraction of basic stellar parameters KASOC and KASC need a large program to calculate stellar models and investigate the error propagation due to limits in the accuracy of stellar modelling techniques and the limits in the description of the stellar physics.



Stellar modelling requires input from several physical disciplines including atomic and nuclear physics, radiation physics, hydrodynamics and thermodynamics. The main issues related to the physical description concerns the detailed equation of state (EOS), pressure (internal energy) from other sources than the gas (e.g. radiation pressure, turbulent pressure from convection and other motion), convection (including the so-called overshoot), rotation (including differential rotation), mixing of the stellar matter, diffusion (including diffusion of individual chemical elements), interaction between radiation and matter (i.e., opacity), nuclear energy generation and surface convection and interaction with stellar activity and stellar rotation. Also the interaction between stellar pulsations and the internal properties of the star need to be included in the calculation for high-amplitude pulsations. In order to generate realistic models one also needs to include 3D hydrodynamical modelling of parts of the stellar interior (e.g., the surface and outer part of the convection zones for cooler stars).

The output from the stellar modelling is, apart from the global physical properties such as luminosity, effective temperature, age and radius, also frequencies for the oscillation modes and the acoustic cut-off frequency in the stellar atmosphere. We are therefore able to produce a number of diagrams that can be used to extract the basic stellar parameters from the oscillation modes. Below we show one example of calculated stellar evolution tracks in a (Δν, δν) – diagram. Observing those main separations one will be able to extract information on stellar mass and core hydrogen content (indicated as dashed curves in the figure below).

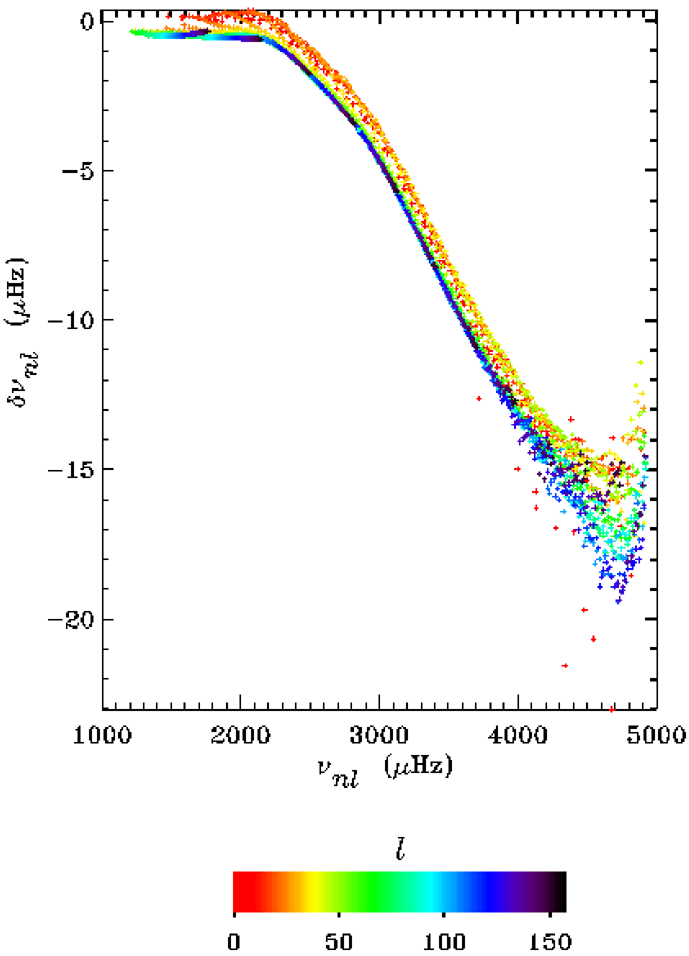


3.7. Extraction of Basic Stellar Parameters

In principle extraction of the basic stellar parameters is an inversion problem where the observed oscillation frequencies are used to estimate the stellar parameters. The inversion shall be understood such that the best estimated parameters for a given star correspond to the model input parameters for the model that shows frequencies most similar to the observed ones. However, this process is not as simple as it may sound. There are basically two problems in the inversion process. First we know from helioseismology that the calculated frequencies suffer from errors caused by near-surface errors in the stellar modelling. On top of this we find that for some stellar models two different sets of input parameters may result in nearly similar frequencies which will create problems concerning the robustness of a given set of extracted stellar parameters.

3.7.1. Surface offset

One of the major problems for asteroseismology of solar-like stars is the fact that the observed oscillation frequencies and the frequencies calculated in a stellar model will show a frequency-dependent offset that is caused by errors in the stellar modeling very close to the surface of the star. The problem may be solved if one can improve on the modeling of convection near the surface, but this has so far not been done. At present the problem is solved in the Sun by using the fact that the offset term is a function of frequency and is almost independent of the mode degree. Using the high-degree modes one therefore - for the Sun - may correct the frequencies for the low-degree modes by applying the offset measured for all oscillation frequencies in the Sun.



The figure at left shows the difference between solar model frequencies and observed solar oscillation frequencies as a function of frequency and mode degree (from Christensen-Dalsgaard et al.).

The fact that the mode frequencies for different degrees show a very similar frequency dependence is used to correct all model frequencies for the surface error term such that model frequencies can be directly compared with observed frequencies.

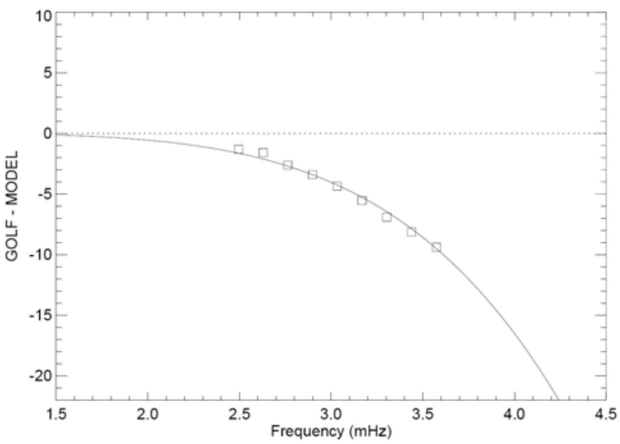
This process is, however, not possible for other stars than the Sun, simply because we have only measured frequencies for modes of low degree.

When we look at the offset structure in the figure it is clear that the size of the offset increases rapidly with frequency and that it is essentially zero for the low-frequency modes.

If we assume this property to be a general feature for the offset term, we may have a chance to estimate the size of this term in other stars than the Sun. One should of course be aware that we do not know the exact mass or the exact radius of other stars and it will therefore be difficult to know if we really estimate the correct offset when we find a given solution for another star.

It is expected that this surface term will be modeled as part of the KASC work and the Kepler time series data will in fact provide new and unique information on this modelling problem. In order to extract precise stellar properties we cannot ignore this problem which is illustrated by the fact that the observed values for the large separation in the Sun 0.8 % lower than the large separation found from the raw frequencies in the solar model. Also the observed frequencies are lower than the raw frequencies in the solar model (by 0.16 % near the peak of the p-mode power).

The synergy between asteroseismology and detailed modeling of transits for Hot Jupiters and high mass ratio eclipsing binaries will allow direct calibration of this offset correction factor. A primary measurement from ultra-high signal-to-noise transit observations, i.e. as will certainly be available for the V = 11.4 solar twin TrES-2 is the mean density of the host star. The latter determination depends only on simple geometry, the measured light curve and application of Kepler's third law. In estimating stellar radii from transit modeling and asteroseismology mass enters in a common way as the limiting factor.



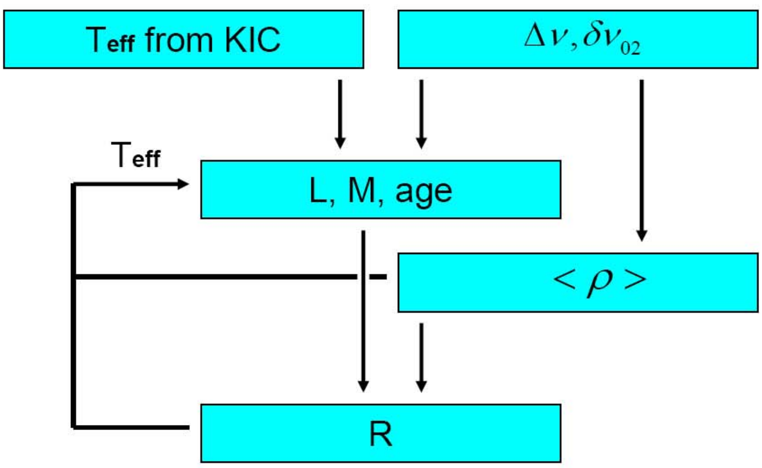
One technique that we expect to implement in the KASOC pipeline is a scaled solar surface offset. Scaling frequencies with peak power frequency or acoustic cutoff frequentcy in the stellar atmosphere will allow us to implement a surface term that only contains one or two free parameters.

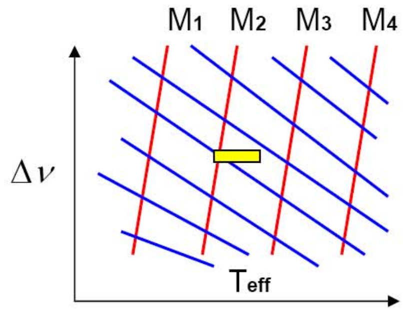
The figure at left shows the difference between the observed GOLF frequencies (radial modes with n=17-25) and the frequencies calculated for a model of the Sun. The solid curve shows a simple approximation for the surface term. This simple approximation may then be used (and rescaled) to correct for the surface term in other stars.

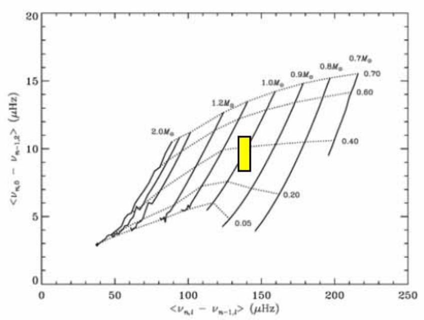
In the Sun we know that at low frequencies the surface term is small and we will therefore also use the lowest detected individual frequency as a check of the offset term by fitting the stellar model directly to this frequency.

3.7.2. Extracting the radius and age of a given star

If we focus the stellar modelling process on the determination of a few crucial parameters we may get a more robust fitting for the pipeline. For the planet transit programme of Kepler the most interesting parameters to extract are the stellar radius and the stellar age. The pipeline process for this extraction will follow a procedure where radius is estimated based on the frequency separations and the parameters from the Kepler Input Catalogue.







The procedure within KASOC for the determination of stellar radius will be:

1. The effective temperature (and raw values for luminosity and radius) will be extracted from the Kepler Input Catalogue (KIC).

2. The small and large frequency separation will be used to estimate independent values of luminosity, mass and age. Combining seismic measurements and the KIC effective temperature will give improved values of luminosity, mass and age.

3. The large (and to some extent the small) separation is used to calculate an accurate value of the stellar mean density.

4. Combining the estimate of mass and measurement of mean density will be used to estimate the stellar radius

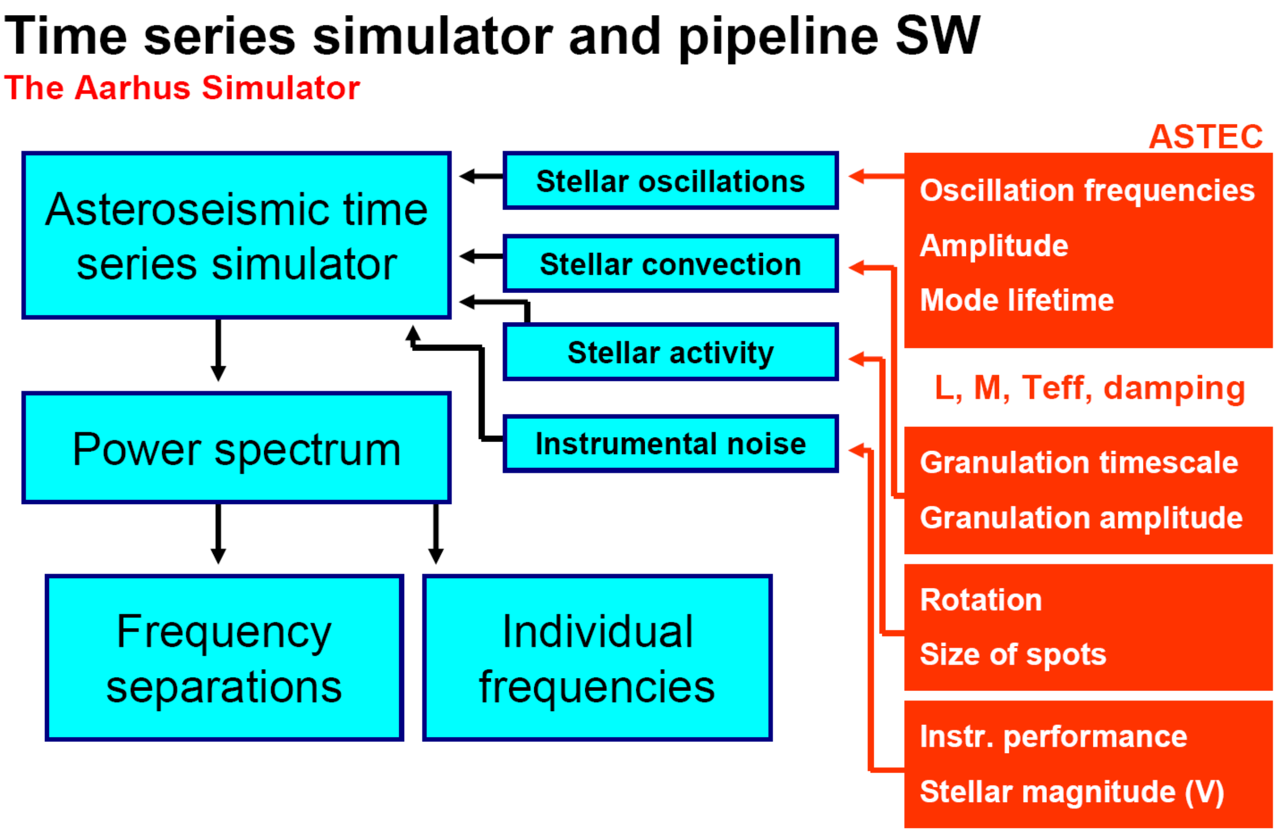
.

5. The estimated stellar radius, luminosity, mass, and mean density and age from seismology will be used to improve the estimated value for the effective temperature.

6. Steps 2 - 4 are then used to refine the estimate of the stellar radius. Simulations show that we are able to reach a high precision in the determinations of stellar radii (relative accuracy below 2-3 per cent) and ages (better than 5-10 per cent of the total main sequence lifetime).

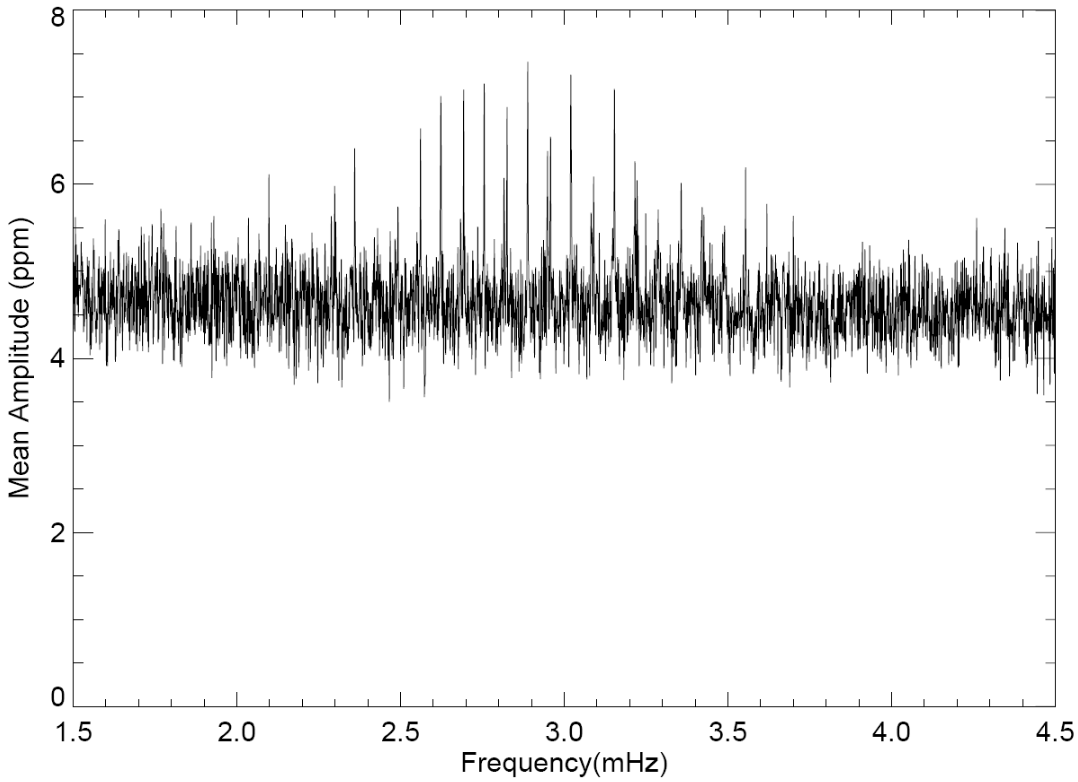
4. KASOC pipeline verification

A crucial part of the KASOC pipeline verification procedure, including verification of the extraction of the accurate basic stellar parameters, is to run realistic simulations of the whole analysis pipeline. In order to do this individual KASC team members are already performing extensive Hare and Hound exercises where one group is generating artificial data that another group is analyzing in order to extract the input information. This exercise is done under the so-called AsteroFLAG collaboration (PI: Bill Chaplin). At KASOC we have been running an extended data generator for several years and this has been used to verify central components in the present version of the KASOC data analysis pipeline. The basic structure of the Aarhus time series simulator can be seen in the figure below.

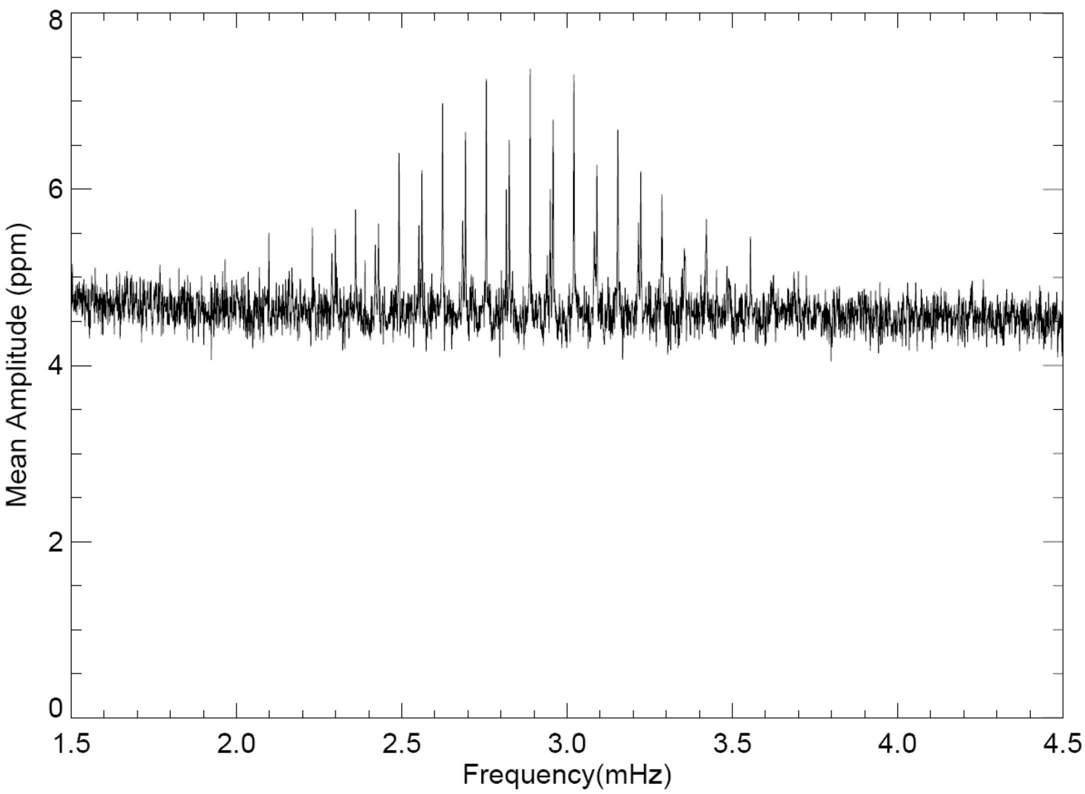


Oscillation frequencies were calculated by use of the ASTEC code (Developed and maintained by Jørgen Christensen-Dalsgaard at University of Aarhus). Amplitudes were estimated by a scaled solar power excess (scaled using simple scaling relations for L, M and T_{eff}). For the mode lifetime we use at present solar damping rates (with a frequency dependence scaled relative to the acoustic cut-off frequency). Granulation timescale and amplitude is estimated by scaling from the Sun and instrumental noise properties are estimated by use of the requirement specifications for Kepler (assuming that the instrument will perform according to the specifications).

Below we show two examples of output from the data simulator for a solar “twin” at magnitude V=11.5 observed with the Kepler photometer. A major task for KASC will be to verify all the data analysis techniques by running a series of Hare and Hound exercises.



The figure above shows the mean amplitude of 52 one-week power spectra (1 year) for the Kepler simulation, based on the Aarhus data generator and time series analysis simulator.

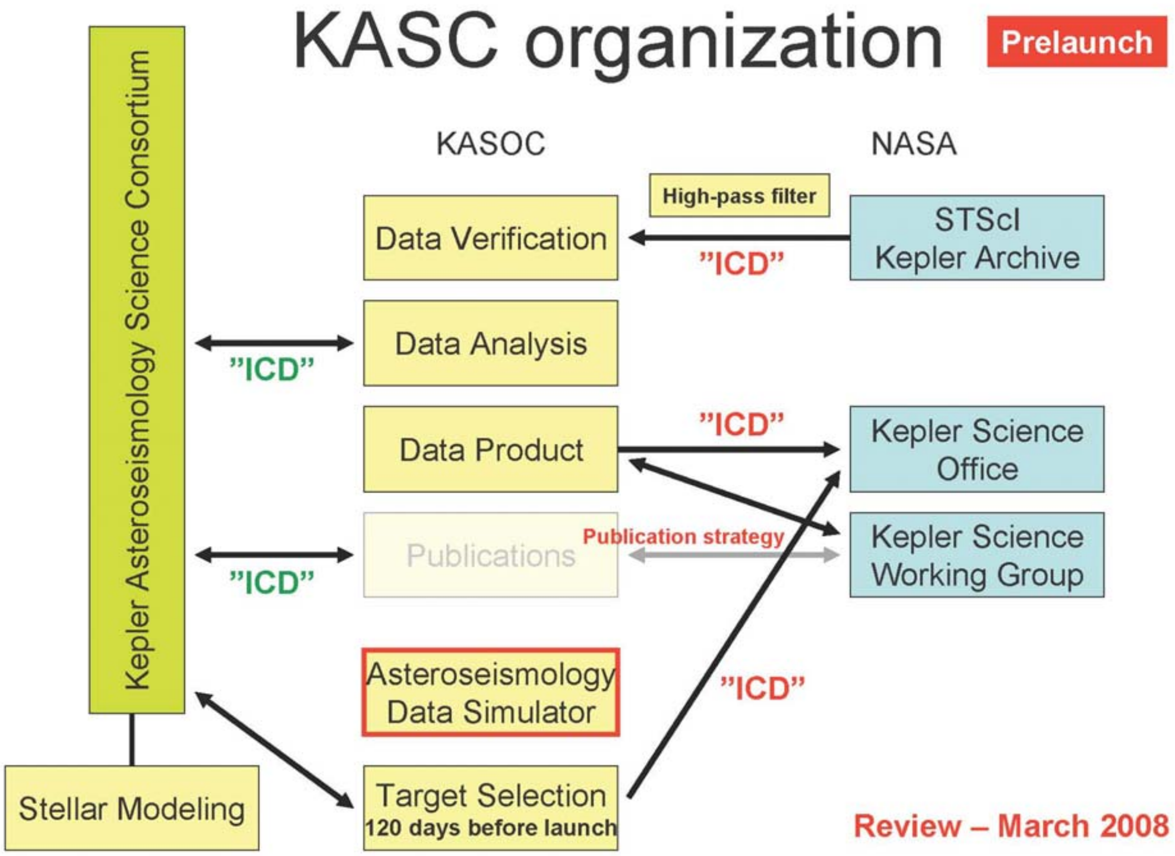


In the figure above we show the mean amplitude of 209 one week power spectra (4 years) for the Kepler simulation.

5. KASOC / KASC organization prelaunch activities

The KASOC pipeline is described in detail in the previous sections. The work is structured such that KASOC is run and managed by the Danish AsteroSeismology Centre (DASC) at University of Aarhus in Denmark (Project Scientist for KASOC is Hans Kjeldsen). Interaction between the Kepler Project and KASOC is mainly through the Kepler Co-I Ronald Gilliland at the Kepler Archive at STScI in Baltimore and directly to the Kepler Science Office at Ames Space Center in California. Details of this are described in the "Letter of Direction". After data are received from the Kepler satellite they will be processed at the NASA Ames Science Operations Center, transferred to the STScI Kepler Archive, retrieved by Co-I Ronald Gilliland, high-pass filtered, then sent to the KASOC. When these data have been analyzed the results will be sent to the Kepler Science Office.

All interfaces between the NASA side of Kepler and KASOC in Denmark will be described via a series of Interface Control Documents (ICD's). Draft version of the ICD's will be ready during 2007 or beginning of 2008 for approval by the Kepler team. During the prelaunch phase of Kepler all interfaces will need to be tested and verified. This also concerns approval of the publication strategy for KAI, which shall be reviewed for approval by the Kepler Science Team and the Kepler PI in November 2007. The ICD for the High-pass filter application and data transfer that is shown in the figure below is in fact an ICD between KASOC and Kepler Co-I Ronald Gilliland as described in detail in section 3.



In order to ensure that full use is made of the Kepler time series data and that the full benefit of asteroseismology is provided for the Kepler investigations of extra-solar planetary systems the Kepler Asteroseismic Science Consortium (KASC) has been established. KASC is a group of collaborating scientists and/or institutions that are connected to the Kepler project via KASOC. The KASC work is lead by the KASC Steering Committee (head: Jørgen Christensen-Dalsgaard) and the KASC Steering Committee will make recommendations to the KAI Steering Committee (head: Ron Gilliland) which is the direct link between KAI and the Kepler Science Team.

In order to structure the work within KASC a set of Work Packages (WP) and Interface Control Documents (ICD) will be developed during end of 2007. The aim of those WP's and ICD's will be to allow the KASOC management to incorporate the KASC contribution in the KASOC pipeline and data test and verification process. Templates for KASOC WP's and KASOC ICD's will be distributed to KASC members in November 2007.

5.1. KASOC Reviews

In order to verify and focus the work that is done in KASOC and inside KASC two internal reviews will take place at KASOC. The details of those reviews as well as the interaction between the KASOC and the Kepler project in relation to the reviews will be discussed during 2007. The first KASOC review will take place in March 2008 and will be focused on the data analysis pipeline. The second and final internal review will take place in November 2008 and this review will focus on the KASOC core activities including interfaces to KASC, the Kepler Archive, Kepler SWG, Co_I's, the Kepler Team and the Kepler Science Office. After launch of Kepler the KASOC pipeline will be commissioned by use of space time series data.

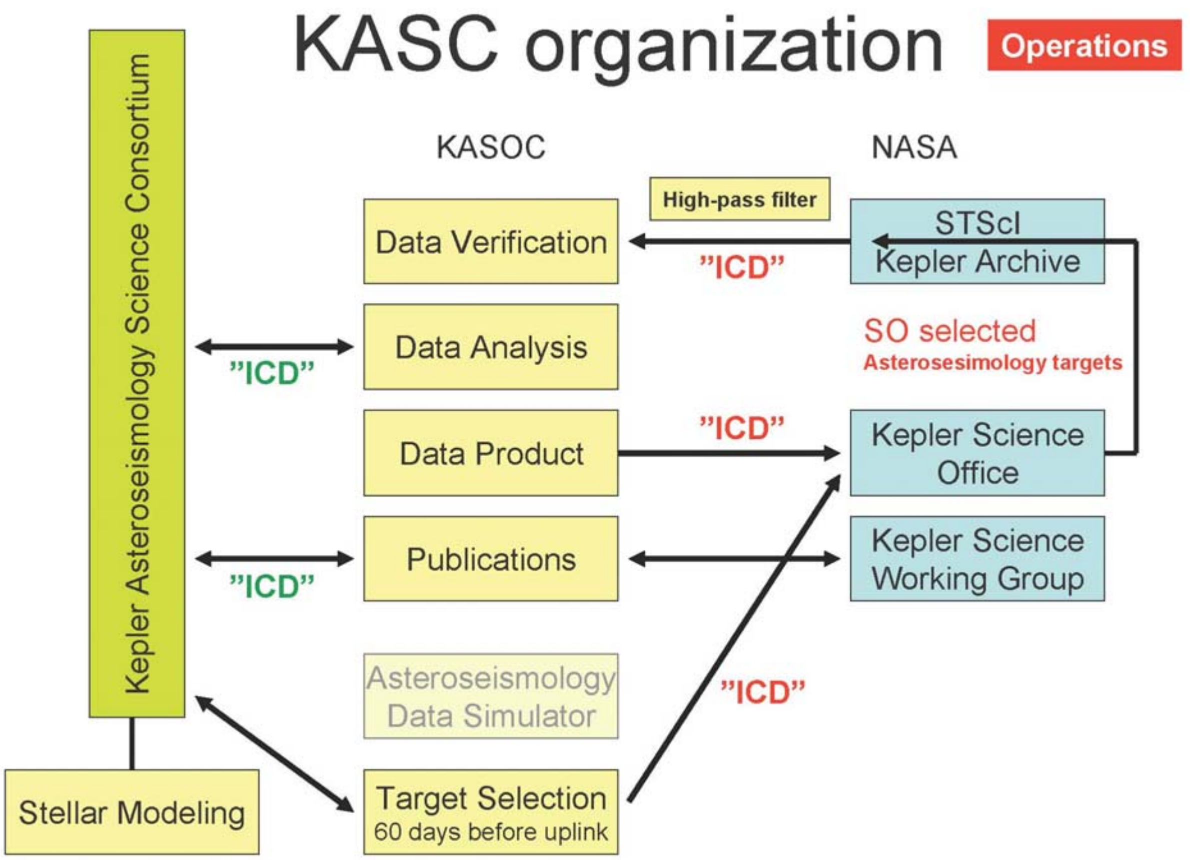
Review activities:

March 2008	KASOC Review #1: Data analysis review (internal)
November 2008	KASOC Review #2: Final review (internal)
March 2009	Commissioning of KASOC (Kepler in space).

6. KASOC / KASC organization of activities after launch

After launch and commissioning of Kepler and KASOC the nominal operations are expected to start in April 2009. The first set of data will be ready for analysis at KASOC in November 2009 and at this time the data will also be available for KASC members through the KASOC. Until October 2012 when the Flight Operations, Phase E will end if the mission is not extended, we expect data release every three months (February, May, August and November).

During operations KASOC will pipeline analyse all KAI and SWG/SO selected asteroseismology targets and KAI/KASOC will every three months (60 days before up-load) select targets for the coming Kepler observing. The KASC/KASOC organization during this phase is illustrated below.

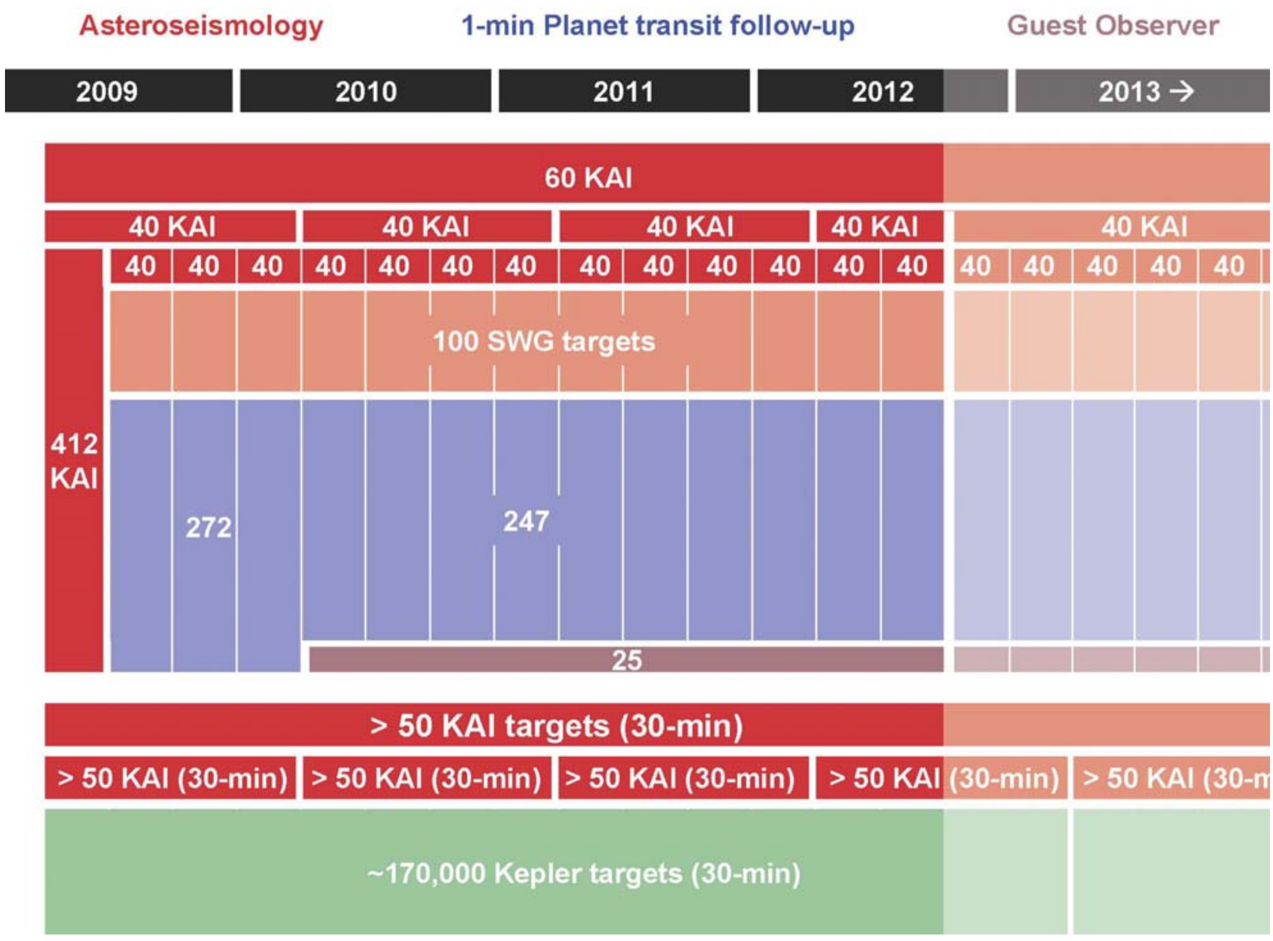


7. Target Selection

The target selection procedure for asteroseismology is based on the "Letter of Direction". Targets are formally selected by the KAI Steering Committee based on recommendations by the KASC Steering Committee and by individual members of the KASC.

The deadlines for target selection depend on the exact launch date. With a scheduled launch of Kepler in February 2009 the dates for transfer of selected asteroseismic targets are:

October 1, 2008	First set of asteroseismology targets transferred to Kepler Science Office.
May 1, 2009	Second set of asteroseismology targets transferred to Kepler Science Office. Transfer of targets to Science Office for next three years: February 1, May 1, August 1 and November 1.



An example of asteroseismic observations with Kepler is shown above; short-cadence targets are at the top and long-cadence targets at the bottom. Dark red denotes targets selected for asteroseismology by KAI. Throughout the mission 140 short-cadence targets are selected by KAI; during the first roll segment all 512 short-cadence targets are available. In the remainder of the mission up to 100 short-cadence targets (called SWG targets in the figure above) may be selected, owing to their interest as planet hosts, by the Kepler Science Team (PI, Co-I's and SWG) for asteroseismic investigation (light red), whereas the remainder of the short-cadence targets are used for detailed investigations of planet transits (blue), or, after the first year, for the 25 Guest Observer targets (magenta). In addition, at least 100 long-cadence targets may be selected by KAI for asteroseismology, Finally, the figure indicates the Kepler main targets for the search for planetary transits (pale green). The nominal mission life time is 3.5 years, but with a possible extension of two years.

The detailed target selection procedure will be discussed at the KASC Steering Committee meeting in October 2007, at the Kepler Science Team (SWG) meeting in November 2007 and at the regular KAI Steering Committee meetings.

The number of targets to be selected depends on the length of observing for the different targets. Based on the example above – and as an illustration – we expect to select the following number of targets:

- 60 targets selected for 3.5 years time series observing with short-cadence.
- 120 targets selected for 1.0 year time series observing with short-cadence.
- 40 targets selected for 180 days time series observing with short-cadence.
- 932 targets selected for 90 days time series observing with short-cadence.

- At least 50 targets selected for 3.5 years time series observing with long-cadence.
- At least 150 targets selected for 1.0 year time series observing with long-cadence.
- At least 50 targets selected for 180 days time series observing with long-cadence.

We will therefore need to select at least 1400 targets for KAI and the first 612 targets will need to be selected before October 1, 2008.

The distribution of targets slots between the different types of pulsating stars will depend on specific target requirements related to stellar magnitude, oscillation periods and density of modes in the frequency domain. However, based on the major objectives for the Kepler Asteroseismic Investigation we expect the target list to be dominated by solar-like stars (including red giants). As stated above the detailed target selection procedure will be discussed at a series of meetings at end of 2007; however, based on the present discussions one can foresee a distribution of targets among the different types of pulsating stars that will allow selection of approximately

1. **Solar-like main sequence stars (780 targets)**

Short-cadence:	40 targets for 3.5 years
	80 targets for 1.0 year
	40 targets for 180 days
	580 targets for 90 days
Long-cadence:	5 targets for 3.5 years
	15 targets for 1.0 years
	20 targets for 180 days

2. **Red giant stars (140 targets)**

Short-cadence:	20 targets for 90 days
Long-cadence:	30 targets for 3.5 years
	90 targets for 1.0 years

   Note that astrometry needs may call for approximately **400 red giant stars** (in long-cadence mode) to be observed throughout the mission, these would be available for asteroseismic analyses.

3. **Delta Scuti stars, roAp stars and gamma Dor stars (240 targets)**

Short-cadence:	10 targets for 3.5 years
	20 targets for 1.0 year
	180 targets for 90 days
Long-cadence:	5 targets for 3.5 years
	15 targets for 1.0 years
	10 targets for 180 days

4. Slowly Pulsating B-stars (SPB-stars) (120 targets)

Short-cadence:	5 targets for 3.5 years 10 targets for 1.0 year 75 targets for 90 days
Long-cadence:	5 targets for 3.5 years 15 targets for 1.0 years 10 targets for 180 days

5. Beta Cephei stars (120 targets)

Short-cadence:	5 targets for 3.5 years 10 targets for 1.0 year 75 targets for 90 days
Long-cadence:	5 targets for 3.5 years 15 targets for 1.0 years 10 targets for 180 days