Data Analysis in Asteroseismology

Tiago Campante

University of Birmingham

campante@bison.ph.bham.ac.uk

19 July 2016

Introduction
Digital signal processing and spectral analysis
- Nyquist sampling theorem and aliasing
- Time-domain filtering
- Power spectral density estimation
- Power spectrum statistics and hypothesis testing
- Non-Fourier periodograms

Extracting global asteroseismic parameters
- Detectability of oscillations
- Background signal
- Large frequency separation ($\Delta\nu$)
- Frequency of maximum amplitude ($\nu_{\max}$)

Peak-bagging
- Power spectrum of a solar-like oscillator
- Modeling the power spectrum
- Bayesian parameter estimation using MCMC

This lecture is intended as a crash course on some of the main data analysis concepts and techniques employed contemporarily in the asteroseismic study of stars exhibiting **solar-like oscillations**.

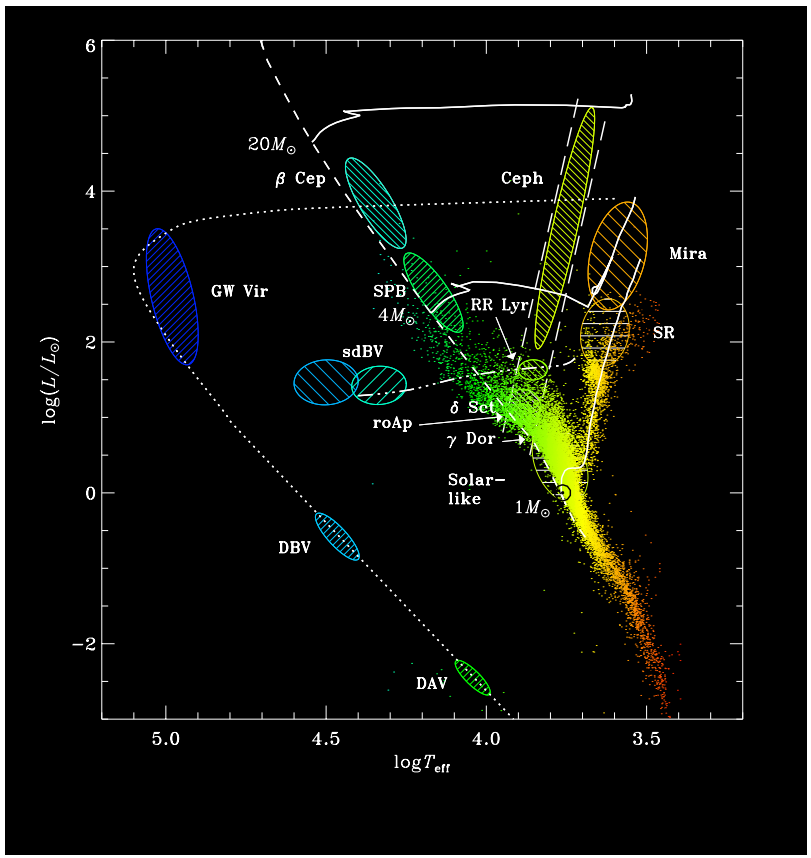
The several concepts and techniques will be presented as we follow the typical **workflow** of the data analysis process.

The contents of this lecture strongly reflect my own experience as a data analyst. For that reason, I have been careful enough to provide references to the work conducted by others, so that you can easily expand on the material presented here.

Solar-like oscillations in the HR diagram

Solar-like oscillations are excited by **turbulent convection** in the outer layers of stars. Consequently, all stars cool enough to harbor an outer convective envelope may be expected to exhibit solar-like oscillations.

Among several other classes of pulsating stars, solar-like oscillations are detectable in main-sequence core, and post-main-sequence shell, hydrogen-burning stars residing on the cool side of the Cepheid instability strip.

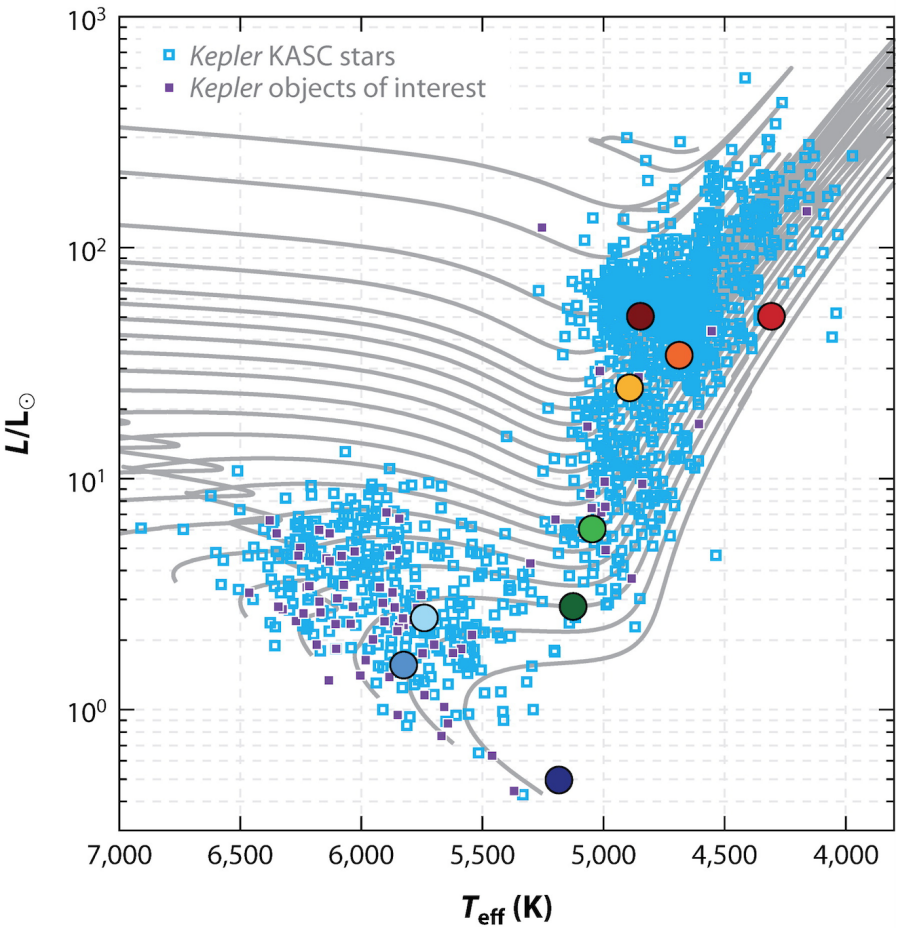


http://astro.phys.au.dk/~jcd/HELAS/puls_HR/

The *Kepler* legacy

The NASA *Kepler* mission has led to a **revolution** in the field of cool-star asteroseismology by detecting solar-like oscillations in several hundred solar-type stars and in over ten thousand red giants.

Of all these stars about 100 are also *Kepler* Objects of Interest (KOIs), i.e., candidate exoplanet-host stars.

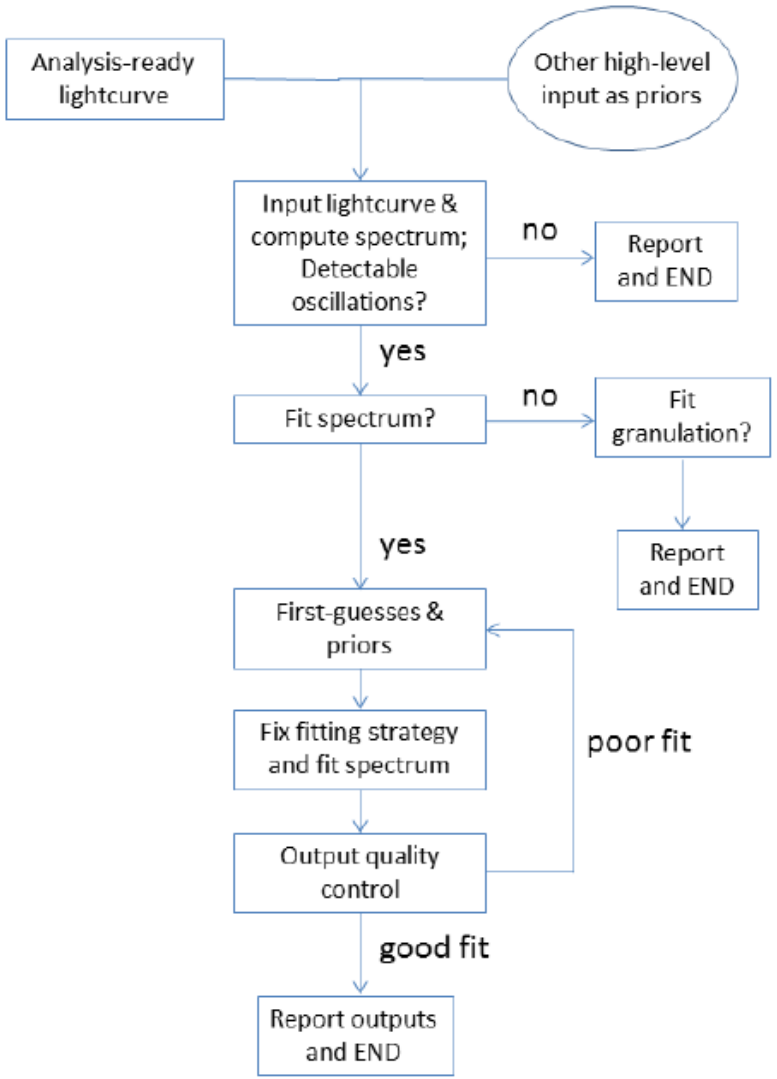


Chaplin & Miglio (2013, ARA&A, 51, 353)

Data analysis workflow

One first establishes whether signatures of solar-like oscillations are detectable in the power spectrum of the light curve. If they are, an attempt is made at extracting **global asteroseismic parameters** from the data.

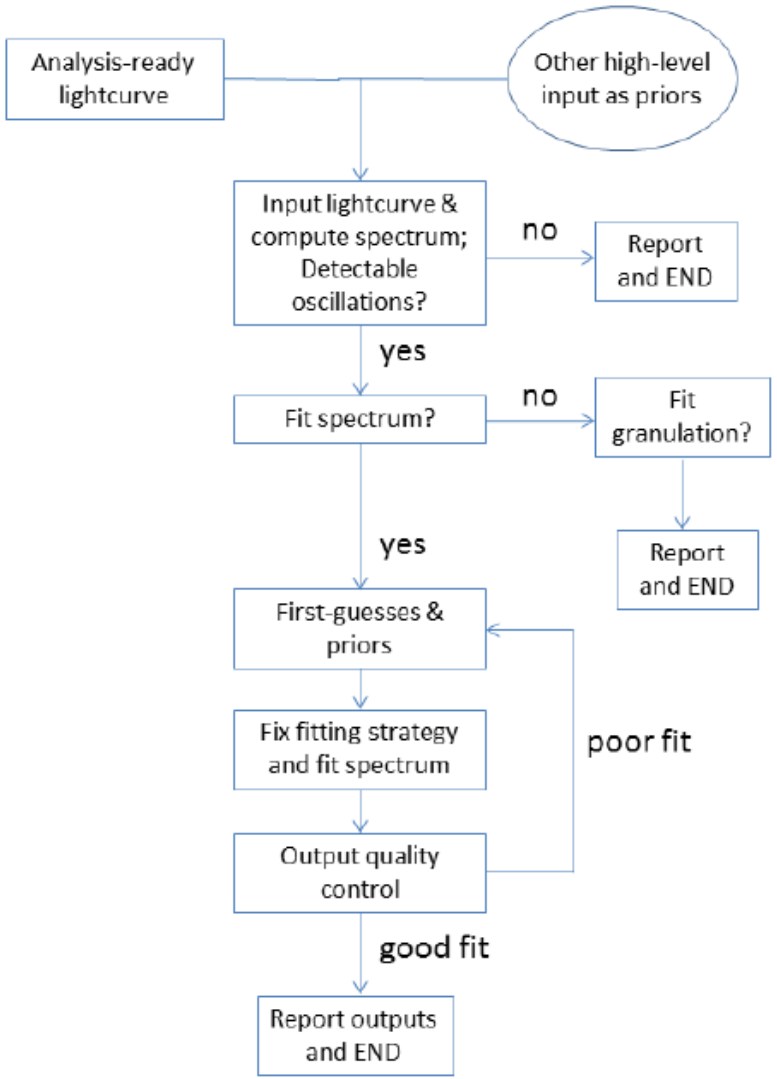
One then establishes whether the oscillation spectrum is of sufficient quality to allow extraction of individual frequencies. If the answer is yes, **individual mode parameters** are then extracted by fitting a multiparameter model to the oscillation spectrum, i.e., by **peak-bagging** the spectrum.



Data analysis workflow (cont.)

The pre-processing of the light curves, although an integrant part of the data analysis process, is beyond the scope of this lecture.

In anticipation of the flood of observations from future space missions such as *PLATO*, a major challenge will be the delivery of **full automation** of the front-to-end analysis. Decisions must then be made concerning the complexity of the fitting model, which modes are to be fitted, and a robust set of first-guess parameters and priors must also be defined.



Whereas some temporal phenomena can be understood through models in the time domain involving deterministic trends or stochastic autoregressive behavior, others are dominated by periodic behavior that is most effectively modeled in the frequency domain.

The functional form of solar-like oscillations is that of a **stochastically-excited harmonic oscillator**. This being a periodic functional form, the **Fourier transform** is the obvious choice for performing data analysis.

Let us consider the idealized case of a continuous signal $x(t)$ sampled by a set of impulse functions regularly spaced by Δt.

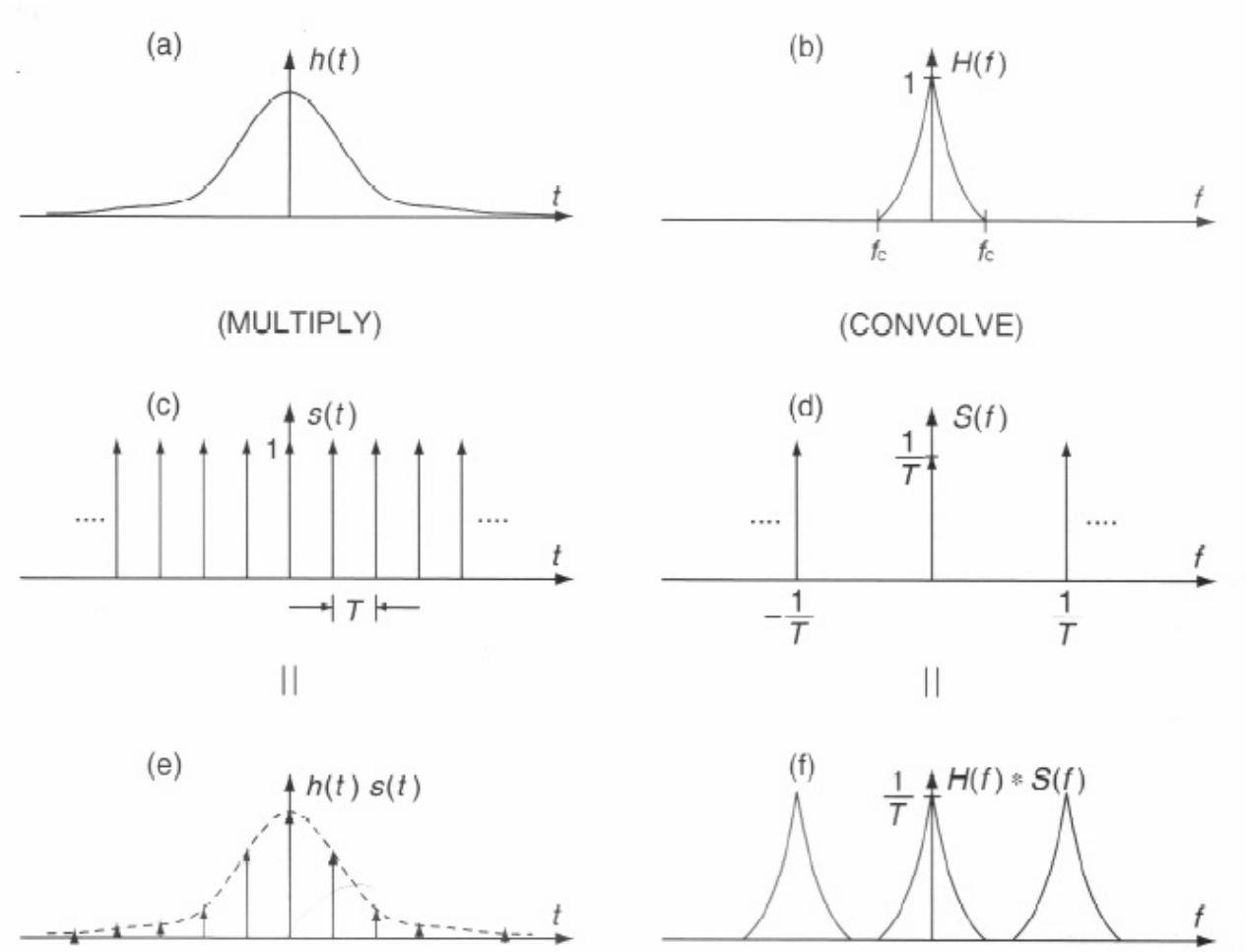
Since the Fourier transform of such a set of impulse functions is another set of impulse functions with separation $1/\Delta t$ in the frequency domain, one can use the convolution theorem to show that the transform of the sampled signal is **periodic**:

$$x(t) \sum_{n=-\infty}^{+\infty} \delta\left(t - n\,\Delta t\right) \Longleftrightarrow X(\nu) * \frac{1}{\Delta t} \sum_{n=-\infty}^{+\infty} \delta\left(\nu - \frac{n}{\Delta t}\right), \quad (1)$$

where $X(\nu)$ is the Fourier transform of $x(t)$.

The **Nyquist sampling theorem** states that if the Fourier transform of a continuous signal is band-limited, i.e., is zero for all $|\nu| \geq \nu_{\rm lim}$, then $x(t)$ can be uniquely reconstructed from a knowledge of its sampled values at uniform intervals of $\Delta t \leq 1/(2\,\nu_{\rm lim})$.

"Oversampling"



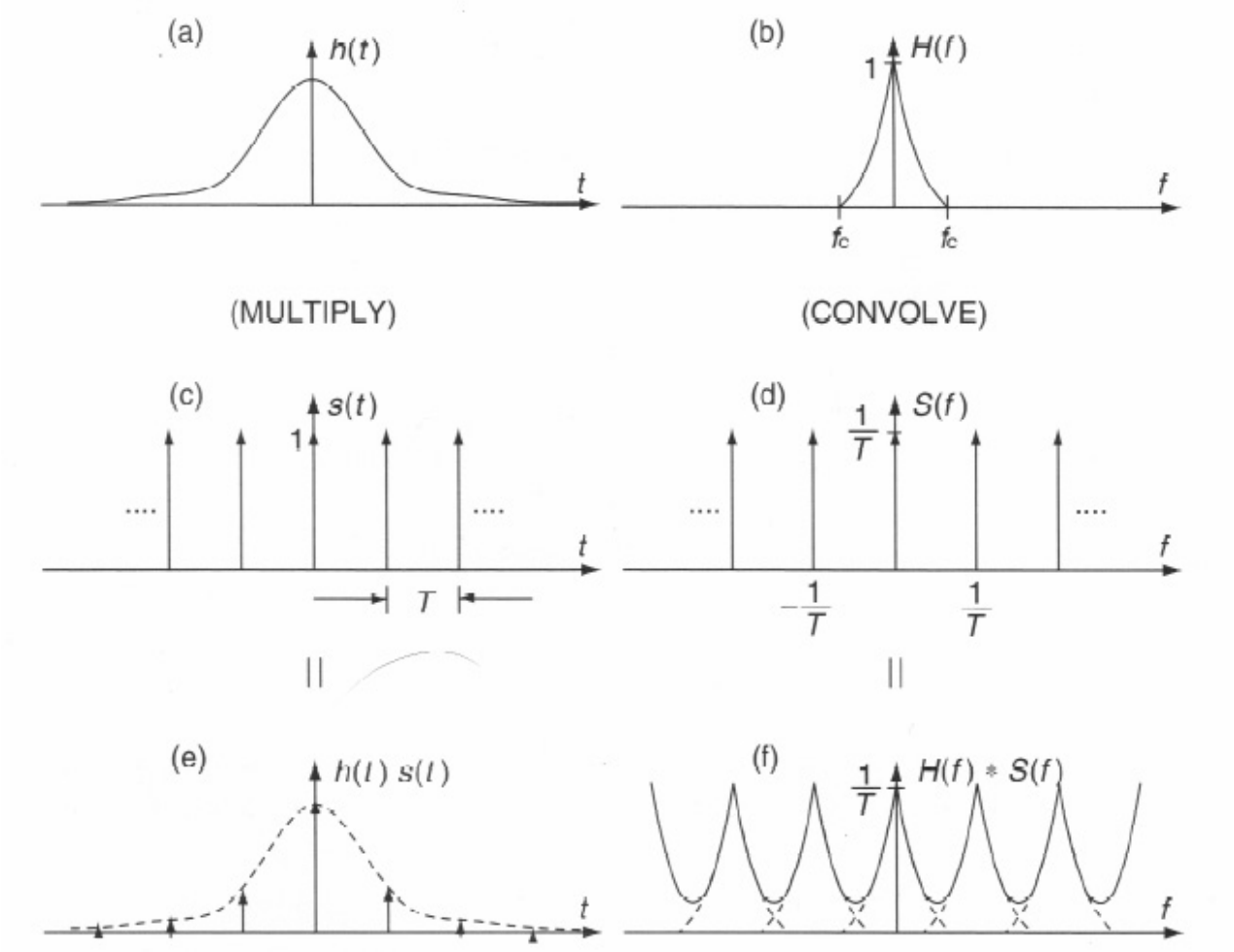
Gregory (2005, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press)

For a given uniform sampling interval Δt, the **Nyquist frequency** is defined as $\nu_{\mathrm{Nyq}} = 1/(2\Delta t)$.

In case the continuous signal being sampled contains frequency components above the Nyquist frequency, these will give rise to an effect known as **aliasing**, whereby the transform of the continuous signal is distorted due to spectral leakage.

The signal is then said to be **undersampled** and can no longer be uniquely recovered.

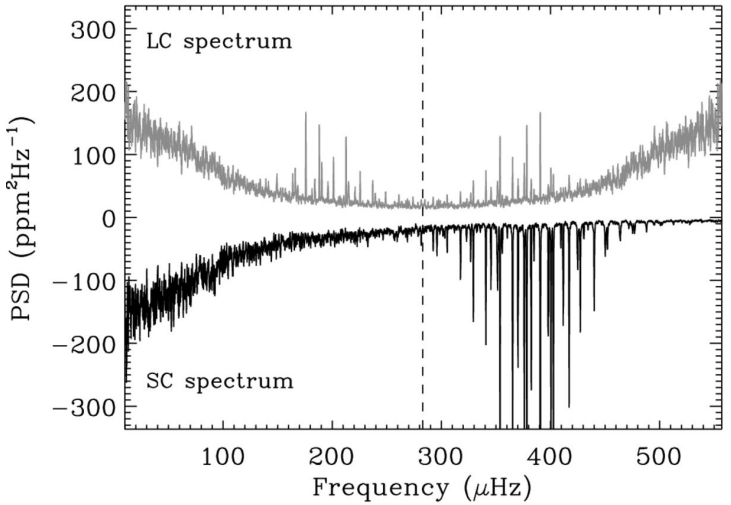
"Undersampling"

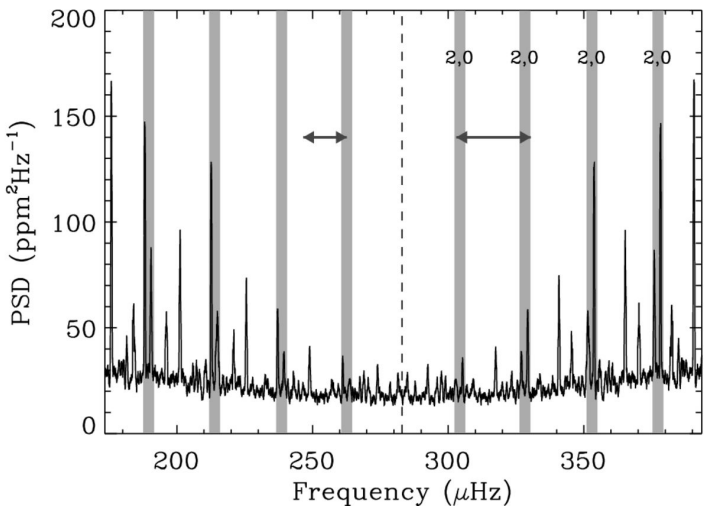


Gregory (2005, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press)

The Nyquist frequency can be thought of as the highest useful frequency to search for in the power spectrum. However, based on astrophysical arguments, one can also accept frequencies above $\nu_{\rm Nyq}$.

Prospects for detecting solar-like oscillations in the **super-Nyquist** regime of *Kepler* long-cadence data, i.e., above the associated Nyquist frequency of $\sim 283\,\mu$Hz, are now being explored. Targets of interest are cool subgiants and stars lying at the base of the red-giant branch.

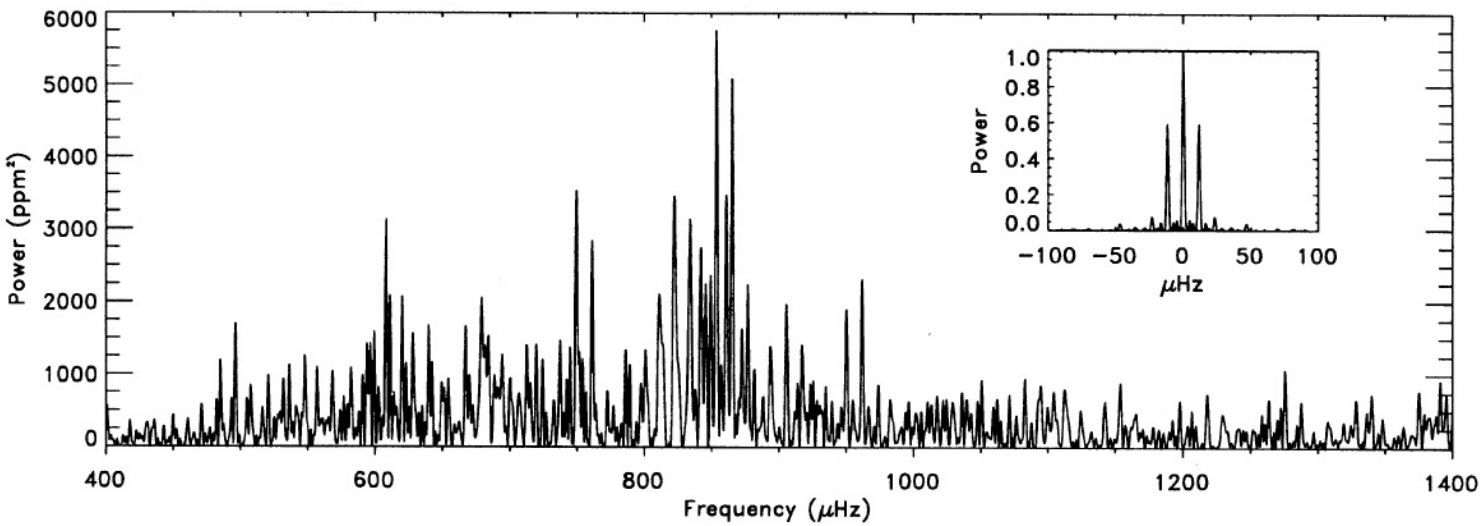




Chaplin et al. (2014, MNRAS, 445, 946)

Regular daily gaps in the light curve are usually present in observations carried out from the ground and also give rise to frequency aliasing. **Daily aliases**, appearing at splittings of $\pm 1\,\mathrm{cycle/day}$ (or, equivalently, $\pm 11.57\,\mu\mathrm{Hz}$), are particularly problematic when observing solar-like oscillations, since frequency separations of that same magnitude are common.

Single-site observations of the G0IV star η Boo



Kjeldsen et al. (1995, AJ, 109, 1313)

Asteroseismic time series are often affected by low-frequency drifts, which can be either of instrumental origin or else intrinsic to the star. These low-frequency drifts introduce a background in the Fourier domain that ultimately leads to a decrease of the SNR of the oscillation modes. High-pass filters are widely used to reduce this effect while preserving the relevant signals.

Let us start by shedding some light on the process of smoothing of a time series. **Smoothing** consists in convolving a signal $x(t)$ with a weighting function $w(t)$:

$$x_{\rm low}(t) = x(t) * w(t) \Longleftrightarrow X_{\rm low}(\nu) = X(\nu)\, W(\nu)\,, \qquad (2)$$

where $X(\nu)$ and $W(\nu)$ are the transforms of $x(t)$ and $w(t)$, respectively.

Conversely, a **high-pass filter** can be implemented by simply computing $x_{\rm high}(t) = x(t) - x_{\rm low}(t)$:

$$x_{\rm high}(t) \Longleftrightarrow X_{\rm high}(\nu) = X(\nu)\,[1 - W(\nu)]\ . \tag{3}$$

Typical examples of the weighting function $w(t)$ are a boxcar function, a triangular function (equivalent to the convolution of two boxcar functions), and a bell-shaped function (equivalent to the convolution of four boxcar functions or two triangular functions). The transform of the boxcar function is the sinc function and thus leads to an excessive ringing (or Gibbs-like) effect in the Fourier domain. Multiple-boxcar smoothing is therefore advisable.

We begin by estimating the Fourier transform of $x(t)$ based on a finite number of samples. Suppose there are N evenly spaced samples $x(t_n)=x(n\Delta t)$, with $n=0,1,\ldots,N-1$. The **Discrete Fourier Transform**[1] (DFT) is defined as:

$$X_{\rm DFT}(\nu_p) = \sum_{n=0}^{N-1} x(t_n)\,{\rm e}^{{\rm i}\,2\pi\nu_p t_n} \quad \text{for } \nu_p = p/(N\Delta t)\,,\ p=0,1,\ldots,N-1\,. \tag{4}$$

$X_{\rm DFT}(\nu_p)$ is the truncated transform of the sampled signal, which has periodicity $1/\Delta t$ or twice the Nyquist frequency. Then $p=0$ corresponds to the transform at zero frequency and $p=N/2$ to the value at $\pm\nu_{\rm Nyq}$. Values of p between $N/2+1$ and $N-1$ correspond to the transform for negative frequencies.

[1] Cooley & Tukey (1965, Math. Comp., 19, 297) introduced the Fast Fourier Transform (FFT), an efficient method of implementing the DFT.

Finally, I introduce the one-sided **power density spectrum** or **power spectrum**, $P(\nu_q)$, defined only for nonnegative frequencies (with $q = 0, 1, \ldots, N/2$):

$$\begin{aligned} P(\nu_0) &= \frac{\Delta t}{N} \left| X_{\rm DFT}(\nu_0) \right|^2 , \\ P(\nu_q) &= \frac{\Delta t}{N} \left[\left| X_{\rm DFT}(\nu_p) \right|^2 + \left| X_{\rm DFT}(\nu_{N-p}) \right|^2 \right] , \qquad (5) \\ P(\nu_{N/2}) &= \frac{\Delta t}{N} \left| X_{\rm DFT}(\nu_{N/2}) \right|^2 , \end{aligned}$$

where $\nu_{N/2} = 1/(2\Delta t)$ (i.e., the Nyquist frequency). Based on **Parseval's theorem**, we may then normalize $P(\nu_q)$ according to

$$\sum_{q=0}^{N/2} P(\nu_q)\, \Delta\nu = \frac{1}{N} \sum_{n=0}^{N-1} x^2(t_n) \, . \qquad (6)$$

According to the **Wiener-Khintchine theorem**, the power spectrum and the autocorrelation function, $\phi(\tau)$, are a Fourier pair:

$$\phi(\tau) = \int_{-\infty}^{+\infty} P(\nu)\, \mathrm{e}^{-\mathrm{i}\, 2\pi\nu\tau}\, \mathrm{d}\nu \iff P(\nu) = \int_{-\infty}^{+\infty} \phi(\tau)\, \mathrm{e}^{\mathrm{i}\, 2\pi\nu\tau}\, \mathrm{d}\tau\,, \tag{7}$$

where

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\tau)\, \mathrm{d}t\,. \tag{8}$$

The Wiener-Khintchine theorem is absolutely crucial to understanding the spectral analysis of random processes. It straightforwardly explains, for instance, why white noise, whose autocorrelation function is the Dirac delta function, has constant power spectral density.

What is the **statistics** of the power spectrum of a pure noise signal?

Let $x(t)$ represent a random process from which a finite number of samples $x(t_n)$ are drawn. The samples are assumed to be independent and identically distributed (i.i.d.), and the process is further assumed to be stationary, with $\mathrm{E}\left[x(t_n)\right] = 0$ and $\mathrm{E}\left[x^2(t_n)\right] = \sigma_0^2$ for all n. The DFT of the set $x(t_n)$ may be decomposed into its real and imaginary parts as:

$$\begin{aligned} X_{\mathrm{DFT}}(\nu_p) &= X^{\mathrm{Re}}_{\mathrm{DFT}}(\nu_p) + \mathrm{i}\, X^{\mathrm{Im}}_{\mathrm{DFT}}(\nu_p) \\ &= \sum_{n=0}^{N-1} x(t_n) \cos(2\pi\nu_p t_n) + \mathrm{i} \sum_{n=0}^{N-1} x(t_n) \sin(2\pi\nu_p t_n)\,. \end{aligned} \tag{9}$$

It follows from the Central Limit theorem that, for large N, both $X_{\rm DFT}^{\rm Re}$ and $X_{\rm DFT}^{\rm Im}$ are normally distributed with

$$\mathrm{E}\left[X_{\rm DFT}^{\rm Re}(\nu_p)\right] = \mathrm{E}\left[X_{\rm DFT}^{\rm Im}(\nu_p)\right] = 0\,, \tag{10}$$

$$\mathrm{E}\left[\left(X_{\rm DFT}^{\rm Re}(\nu_p)\right)^2\right] = \mathrm{E}\left[\left(X_{\rm DFT}^{\rm Im}(\nu_p)\right)^2\right] = \frac{N}{2}\sigma_0^2\,. \tag{11}$$

Finally, since $X_{\rm DFT}^{\rm Re}$ and $X_{\rm DFT}^{\rm Im}$ are independent and have the same normal distribution, the power spectrum, $|X_{\rm DFT}|^2$, then has by definition a **chi-squared distribution with 2 degrees of freedom** (i.e., χ_2^2).

Adopting $|X_{\mathrm{DFT}}|^2 \Delta t/N$ as our normalization of the power spectrum yields a constant power spectral density for the noise of $\sigma_0^2 \Delta t$ and variance $(\sigma_0^2 \Delta t)^2$. Consequently, as N tends to infinity by sampling a longer stretch of data, the variance in the power spectrum remains unchanged.

Furthermore, the probability density, $p(z)$, that the observed power spectrum takes a particular value z at a fixed frequency bin is given by

$$p(z) = \frac{1}{\langle z \rangle} \exp\left(-\frac{z}{\langle z \rangle}\right) , \tag{12}$$

where $\langle z \rangle = \sigma_0^2 \Delta t$.

Equation (12) enables one to derive the probability that the power in one bin is greater than m times the mean level of the continuum, $\langle z \rangle$:

$$F(m) = \mathrm{e}^{-m}\,. \tag{13}$$

For instance, a **confidence level** of 99% or, equivalently, a **false alarm probability** of 1%, leads to $m \approx 4.6$.

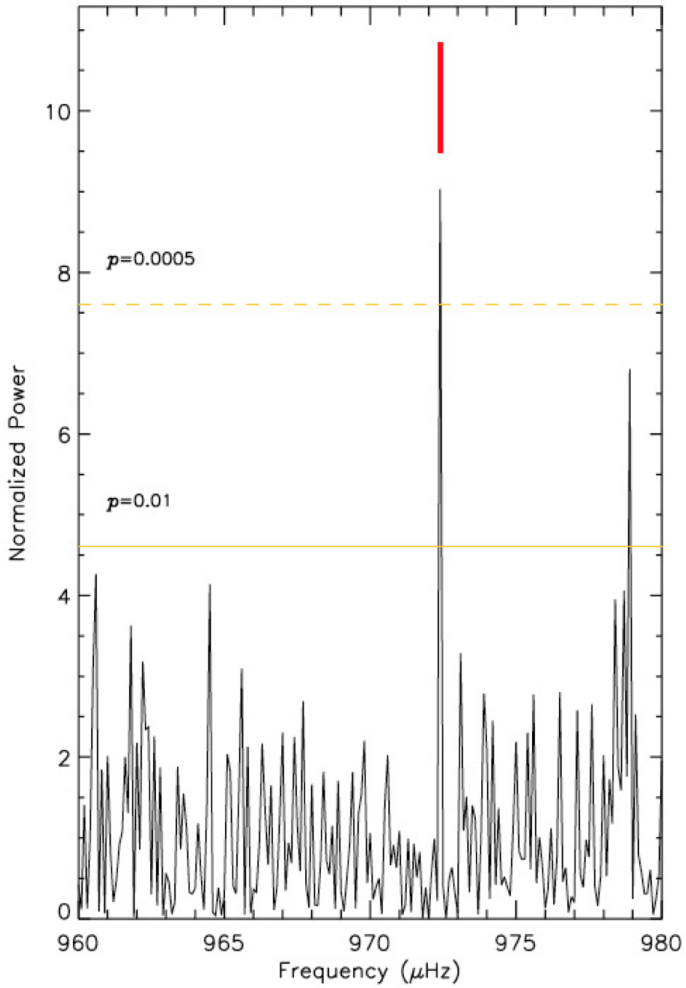
For a frequency band containing M bins, the probability that at least one bin has a normalized power greater than m is then:

$$F_M(m) = 1 - (1 - \mathrm{e}^{-m})^M\,, \tag{14}$$

which approximates to $F_M(m) = M\mathrm{e}^{-m}$ for $\mathrm{e}^{-m} \ll 1$.

Here we apply a test based on the **null hypothesis** to the detection of an unresolved p mode. The solid horizontal line corresponds to a false alarm probability of 1 % in Eq. (13). According to Eq. (14), the chance of finding at least one noise spike within the displayed window (200 bins wide) above this detection threshhold is of about 87%.

A more conservative approach is to set to, say, 10%, the probability of finding at least one spike within this window, resulting in a false alarm probability of 0.05 % (dashed horizontal line).



Campante (2012, PhD thesis)

In astrophysics it is very common to deal with unevenly sampled time series. In that event, an existing frequentist statistic known as the **Lomb–Scargle periodogram**[2] is widely used as an estimator of the power spectral density.

The Lomb–Scargle periodogram can be formulated either as a modified Fourier analysis or as a least-squares regression of the data set to sine waves with a range of frequencies. It has the attractive property of retaining the χ^2_2 statistics.

[2] Fast computation of the periodogram is achieved using the algorithm presented in Press & Rybicki (1989, ApJ, 338, 277), whose trick is to carry out extirpolation of the data onto a regular mesh and subsequently employ the FFT.

Astronomers have developed and extensively used a variety of **non-Fourier periodograms** for period searches in unevenly spaced data sets (e.g., Clarke 2002, A&A, 386, 763). The most common strategy involves folding the data modulo a trial period, computing a statistic on the folded time series (now a function of phase rather than time), and plotting the statistic for all independent periods.

These methods measure the strength of signals that are strictly periodic, but not necessarily sinusoidal in shape. They are also relatively insensitive to the duration and uneven spacing of the data set, and some methods readily permit heteroscedastic weighting from measurement errors.

Recommended reading

- Aerts, C., Christensen-Dalsgaard, J., & Kurtz, D. W. 2010, *Asteroseismology*, 1st ed., Springer
- Appourchaux, T. 2014, in *Asteroseismology*, 22nd Canary Islands Winter School of Astrophysics, Cambridge University Press, eds. P. L. Pallé & C. Esteban, 123
- Campante, T. L. 2012, PhD thesis, Universidade do Porto
- Shumway, R. H. & Stoffer, D. S. 2006, *Time Series Analysis and Its Applications with R Examples*, 3rd ed., Springer

In order to fully characterize a star using asteroseismology, it is desirable to obtain precise estimates of individual mode parameters (e.g., frequencies, amplitudes and linewidths). However, this is only possible for data above a certain SNR.

Global asteroseismic parameters, indicative of the overall stellar structure, are on the other hand readily extractable using **automated pipelines** that are able to incorporate data with a lower SNR and for which a full peak-bagging analysis is not always possible. Furthermore, the automated nature of these pipelines is required if we are to efficiently exploit the plenitude of data made available by space-based missions.

In this section I introduce an automated pipeline[3] designed to measure global asteroseismic parameters of main-sequence and subgiant stars from the power spectrum (Campante 2012, PhD thesis).

The pipeline allows extracting the following information from the power spectrum (points 1–4 are covered here):

1. Frequency range of the oscillations;
2. Parameterization of the stellar background signal;
3. Average large frequency separation, $\Delta\nu$;
4. Frequency of maximum amplitude, $\nu_{\rm max}$;
5. Maximum mode amplitude, $A_{\rm max}$.

[3] A comparison of different pipelines used to extract global asteroseismic parameters is presented in Verner et al. (2011, MNRAS, 415, 3539).

We want to look for a frequency range in the power spectrum in which peaks appear at nearly regular intervals, one of the main signatures of the presence of solar-like oscillations. I note that the assumption of **quasi-regularity** may, however, be too strong in the case of evolved stars due to the presence of mixed modes.

We thus start by partitioning the power spectrum into overlapping windows of variable width, w. The width w depends on the central frequency of the window, $\nu_{\rm central}$, used as a proxy for $\nu_{\rm max}$. We make use of the fact that the width of the p-mode bump roughly scales with $\nu_{\rm max}$, and so w is defined as $w = (\nu_{\rm central}/\nu_{\rm max,\odot})w_\odot$.

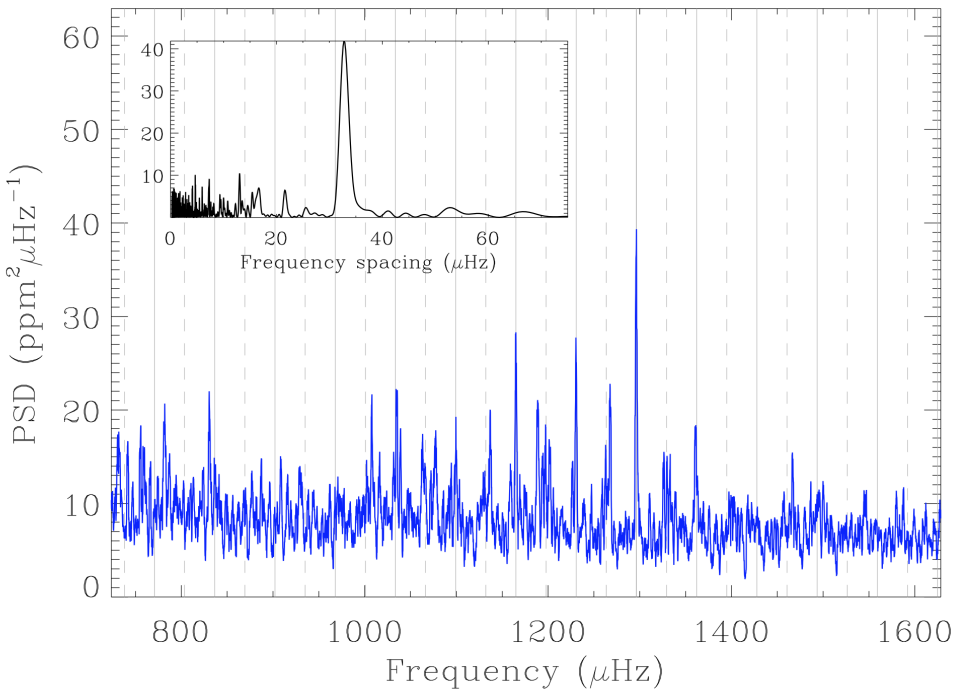
The next step consists in computing the **power spectrum of the power spectrum**, PS⊗PS, for each of these frequency windows. The presence of prominent features in the PS⊗PS around the predicted[4] values of $\Delta\nu/2$, $\Delta\nu/4$, and $\Delta\nu/6$ (the first, second, and third harmonics, respectively) is then examined.

An hypothesis test is subsequently applied, whereby the presence of oscillations in a given window is established if the probability of the three above features being due to noise is less than 1%. Finally, the **frequency range of the oscillations** is determined based on the overall span of the windows with detected oscillations.

[4] The predicted value of $\Delta\nu$ is computed according to the relation $\Delta\nu \propto \nu_{\rm central}^{0.77}$ (Stello et al. 2009, MNRAS, 400, L80).

Here I show the detection of oscillations in the *K2* power spectrum of a solar-type star. Sets of vertical gray solid and dashed lines are separated by the estimated $\Delta\nu$, and mark the spacing on which we would expect to see modes.

The inset shows the PS⊗PS, computed from the region around $\nu_{\rm max}$. The significant peak in the PS⊗PS lies at $\Delta\nu/2$ and is a signature of the near-regular spacing of oscillation peaks.



Chaplin et al. (2015, PASP, 127, 1038)

The model of the **stellar background signal** is kept simple, merely containing a granulation component and photon shot noise. We fit this model to a smoothed version of the power spectrum employing a nonlinear least-squares fitting algorithm.

The frequency range of the oscillations (if detected) is excluded from the fitting window. The fitting window starts at 100 μHz to allow for the decay of any possible activity component, characterized by considerably longer timescales, and extends all the way up to the Nyquist frequency of *Kepler* short-cadence data ($\sim$8300 μHz).

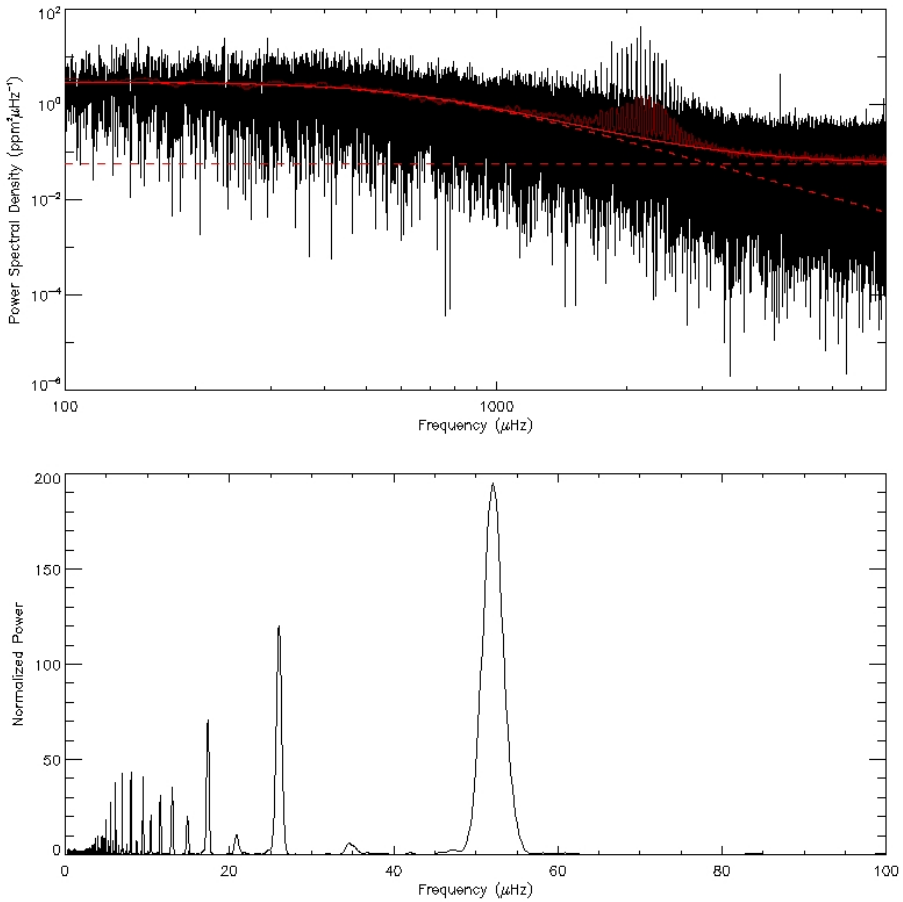
The granulation component is represented by a **Harvey-like profile** (e.g., Kallinger et al. 2014, A&A, 570, A41) to which an offset is added representing the **shot noise** component:

$$B(\nu) = B_0 + \eta^2(\nu) \left[\frac{B_{\rm gran}}{1 + (2\pi\nu\,\tau_{\rm gran})^a} \right] , \qquad (15)$$

where $B_{\rm gran}$ is the height at $\nu = 0$ of the granulation component, $\tau_{\rm gran}$ is the characteristic turnover timescale and a calibrates the amount of memory in the process. Such a functional form is representative of a random non-harmonic field whose autocorrelation decays exponentially with time. The attenuation factor $\eta^2(\nu)$ takes into account the apodization of the oscillation signal due to the finite integration time.

The top panel displays the smoothed power spectrum of 16 Cyg A (in dark red) superimposed on the original spectrum (in black). The fit to the background signal (red solid line) and both its components (red dashed lines) are also shown.

The bottom panel displays the PS⊗PS over the frequency range of the oscillations. The features at $\Delta\nu/2$ ($\sim$52 μHz), $\Delta\nu/4$ ($\sim$26 μHz) and $\Delta\nu/6$ ($\sim$17 μHz) are conspicuous.



Campante (2012, PhD thesis)

In order to estimate the **average large frequency separation**, $\Delta\nu$, we compute the PS$\otimes$PS over the frequency range of the oscillations. The feature at $\Delta\nu/2$ (first harmonic) in the PS$\otimes$PS is then located and its power-weighted centroid computed to provide an estimate of $\Delta\nu$.

The standard deviation of grouped data, given by $\sqrt{\left[\sum hx^2 - \left(\sum hx\right)^2 / \sum h\right] / \left(\sum h - 1\right)}$, is adopted as the error on $\Delta\nu$, meaning that the feature in the PS$\otimes$PS is interpreted as an assembly of spectral heights (h) over a number of bins (with midpoint x).

In order to estimate the **frequency of maximum amplitude**, $\nu_{\rm max}$, we average the p-mode power (after subtraction of the background fit) over contiguous rectangular windows of width $2\Delta\nu$ and convert to power per radial mode by multiplying by $\Delta\nu/c$, where c measures the effective number of modes per order (see Kjeldsen et al. 2008, ApJ, 682, 1370).

An estimate of $\nu_{\rm max}$ is then given by the power-weighted centroid, with the associated uncertainty derived from the standard deviation of grouped data.

Recommended reading

- Campante, T. L., et al. 2010, MNRAS, 408, 542
- Campante, T. L. 2012, PhD thesis, Universidade do Porto
- Hekker, S., et al. 2010, MNRAS, 402, 2049
- Huber, D., et al. 2009, Communications in Asteroseismology, 160, 74
- Mosser, B. & Appourchaux, T. 2009, A&A, 508, 877
- Verner, G. A., et al. 2011, MNRAS, 415, 3539

In this section I introduce a **Bayesian** peak-bagging tool that employs **Markov chain Monte Carlo** (MCMC) techniques. Besides making it possible to incorporate relevant **prior information** through Bayes' theorem, this tool also allows obtaining the marginal **probability density function** (pdf) for each of the model parameters.

These techniques are in many ways an extension of the **Maximum Likelihood Estimation** (MLE) methods traditionally used in helioseismology.

Understanding the characteristics of the power spectrum of a solar-like oscillator is fundamental in order to extract information on the physics of the modes.

The **stochastic driving of a damped oscillator** can be described by

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} y(t) + 2\eta \frac{\mathrm{d}}{\mathrm{d}t} y(t) + \omega_0^2 \, y(t) = f(t)\,, \tag{16}$$

where $y(t)$ is the amplitude of the oscillator, η is the linear damping rate, ω_0 is the frequency of the undamped oscillator and $f(t)$ is a random forcing function. The Fourier transform of Eq. (16) is then expressed as

$$-\omega^2 \, Y(\omega) - \mathrm{i}\, 2\eta\omega \, Y(\omega) + \omega_0^2 \, Y(\omega) = F(\omega). \tag{17}$$

When a realization of $y(t)$ is observed for a finite amount of time, an estimate of the power spectrum is then given by

$$P(\omega) = |Y(\omega)|^2 = \frac{|F(\omega)|^2}{(\omega_0^2 - \omega^2)^2 + 4\,\eta^2\omega^2}\,. \tag{18}$$

In the limit of taking the average of an infinite number of realizations, and assuming the damping rate to be very small compared to the frequency of oscillation, one obtains near the resonance the following expression for the **limit spectrum**:

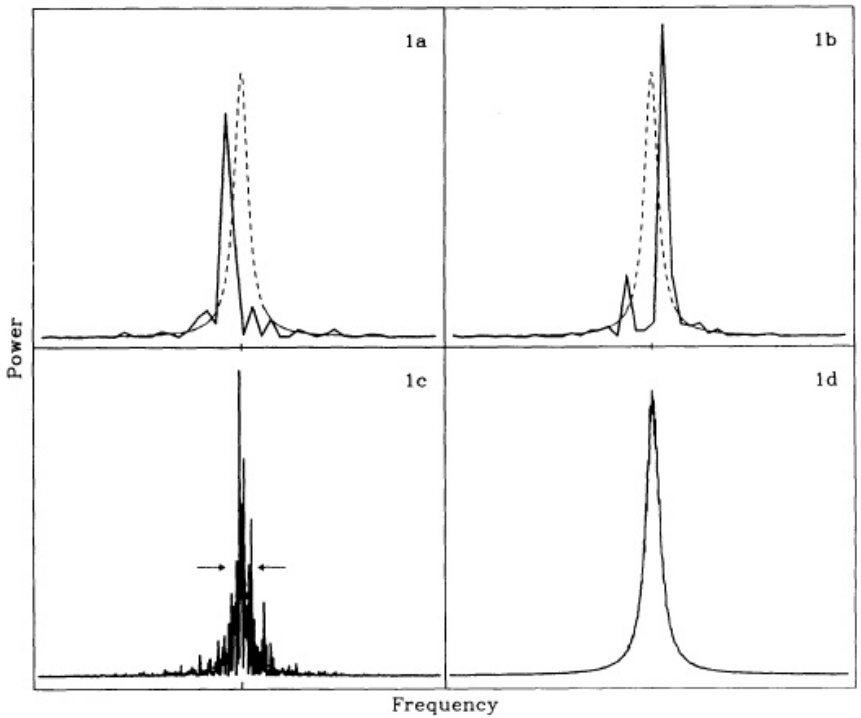
$$\langle P(\omega)\rangle \simeq \frac{1}{4\,\omega_0^2}\,\frac{\langle P_f(\omega)\rangle}{(\omega - \omega_0)^2 + \eta^2}\,. \tag{19}$$

The average power spectrum of the random forcing function, $\langle P_f(\omega)\rangle$, is a slowly-varying function of frequency. The result is thus a **Lorentzian profile**, characterized by the central frequency ω_0 and a width determined by the linear damping rate η.

Panels (a) and (b) display two realizations of the same limit spectrum. Both power spectra appear as an erratic function concealing the underlying Lorentzian profile.

Panel (c) displays a realization of the same limit spectrum, although with a resolution twenty times higher. Increasing the total observational span, hence the resolution, did nothing to reduce the erratic behavior.

Panel (d) displays the average of a large number of realizations with the same resolution as in (c).



Anderson et al. (1990, ApJ, 364, 699)

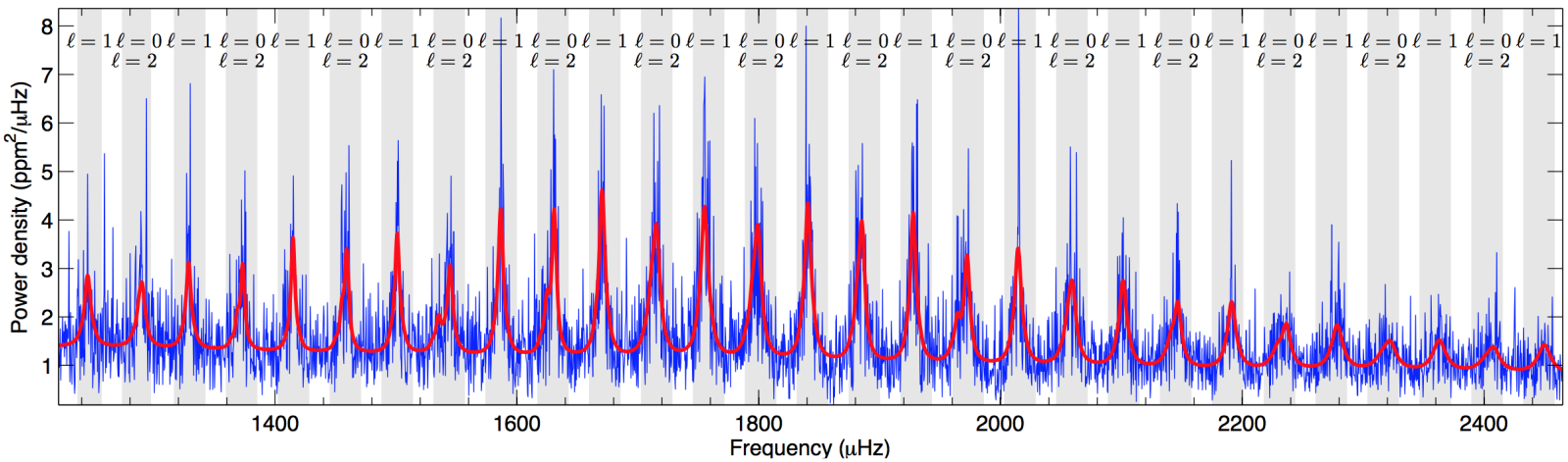
We are primarily interested in performing a **global fit** to the power spectrum, whereby the observed modes are fitted simultaneously over a broad frequency range.

We thus model the limit oscillation spectrum as a sum of standard Lorentzian profiles, $O(\nu)$, which sit atop a background signal described by $B(\nu)$:

$$\begin{aligned} P(\nu;\boldsymbol{\lambda}) &= O(\nu) + B(\nu) \\ &= \sum_{n',l} \sum_{m=-l}^{l} \frac{\mathscr{E}_{lm}(i_{\mathrm{s}}) H_{n'l}}{1 + \left[\frac{2(\nu - \nu_{n'l0} - m\nu_{\mathrm{s}})}{\Gamma_{n'lm}}\right]^2} + B(\nu)\,, \end{aligned} \tag{20}$$

where $\boldsymbol{\lambda}$ represents the set of model parameters.

Shown below is the power spectrum of HD 49933 (blue) based on 180 days of *CoRoT* photometry. The best-fitting model (red) is overlaid, with the shaded areas indicating the ranges of the uniform priors on the mode frequencies.



Handberg & Campante (2011, A&A, 527, A56)

At a given frequency bin j, the probability density, $f(P_j; \boldsymbol{\lambda})$, that the observed power spectrum takes a particular value P_j is related to the limit spectrum, $P(\nu_j; \boldsymbol{\lambda})$, by (cf. Eq. 12):

$$f(P_j; \boldsymbol{\lambda}) = \frac{1}{P(\nu_j; \boldsymbol{\lambda})} \exp\left[-\frac{P_j}{P(\nu_j; \boldsymbol{\lambda})}\right] . \tag{21}$$

We now want to specify the **likelihood function**, i.e., the joint pdf of the data sample $\{P_j\}$. Assuming the frequency bins to be uncorrelated, the joint pdf is simply given by the product of $f(P_j; \boldsymbol{\lambda})$ over some frequency interval of interest spanned by j:

$$L(\boldsymbol{\lambda}) = \prod_j f(P_j; \boldsymbol{\lambda}) . \tag{22}$$

Bayes' theorem

I now describe the formalism of a Bayesian approach to **parameter estimation** and **model comparison** that employs an MCMC algorithm.

Let us consider a set of competing hypotheses, $\{H_i\}$, assumed to be mutually exclusive. One should be able to assign a probability, $p(H_i|D, I)$, to each hypothesis, taking into account the observed data, D, and any available prior information, I. This is done through **Bayes' theorem**:

$$p(H_i|D, I) = \frac{p(H_i|I)p(D|H_i, I)}{p(D|I)} . \tag{23}$$

Bayes' theorem (cont.)

The probability of the hypothesis H_i in the absence of D is called the **prior probability**, $p(H_i|I)$, whereas the probability including D is called the **posterior probability**, $p(H_i|D,I)$. The quantity $p(D|H_i,I)$ is called the **likelihood** of H_i. The denominator $p(D|I)$ is the **global likelihood** for the entire class of hypotheses.

The sum of the posterior probabilities over the hypothesis space of interest is unity, hence one has:

$$p(D|I) = \sum_i p(H_i|I)p(D|H_i,I)\,. \tag{24}$$

Parameter estimation

If a particular hypothesis, i.e., a given model M describing the physical process, is assumed true, then the hypothesis space of interest concerns the values taken by the model parameters, $\boldsymbol{\lambda}$. These parameters are continuous and one will be interested in obtaining their pdf.

The global likelihood of model M is then given by the continuous counterpart of Eq. (24):

$$p(D|I) = \int p(\boldsymbol{\lambda}|I)p(D|\boldsymbol{\lambda}, I)d\boldsymbol{\lambda}\,. \tag{25}$$

Parameter estimation (cont.)

We restate Bayes' theorem to account for this new formalism:

$$p(\boldsymbol{\lambda}|D,I) = \frac{p(\boldsymbol{\lambda}|I)p(D|\boldsymbol{\lambda},I)}{p(D|I)}\,, \tag{26}$$

where $p(D|I)$ plays the role of a normalization constant.

Ultimately, we are interested in using MCMC techniques to map the posterior pdf, $p(\boldsymbol{\lambda}|D,I)$. The procedure of **marginalization** allows computation of the posterior pdf for a subset of parameters $\boldsymbol{\lambda}_A$ by integrating over the remaining parameters (or **nuisance parameters**) $\boldsymbol{\lambda}_B$:

$$p(\boldsymbol{\lambda}_A|D,I) = \int p(\boldsymbol{\lambda}_A,\boldsymbol{\lambda}_B|D,I)d\boldsymbol{\lambda}_B\,. \tag{27}$$

Model comparison

The problem of model comparison is analogous to that of parameter estimation. When facing a situation in which several parameterized models are available for describing the same physical process, one expects Bayes' theorem to allow for a statistical comparison between such models.

Bayesian model comparison has a built-in **Occam's razor** by which a complex model is automatically penalized, unless the available data justify its additional complexity.

Competing models may be either intrinsically different models or else similar but with varying number of parameters (i.e., nested models), or even the same model with different priors affecting its parameters.

Model comparison (cont.)

Given two or more competing models and our prior information, I, being in the present context that one and only one of the models is true, we can assign individual probabilities similarly to what has been done in Eq. (23), after replacing H_i by M_i:

$$p(M_i|D,I) = \frac{p(M_i|I)p(D|M_i,I)}{p(D|I)}\,, \tag{28}$$

where the global likelihood of model M_i, $p(D|M_i,I)$, also called the **evidence** of the model, is given by Eq. (25).

Model comparison (cont.)

We are often interested in computing the ratio of the probabilities of two competing models:

$$O_{ij} \equiv \frac{p(M_i|D,I)}{p(M_j|D,I)} = \frac{p(M_i|I)p(D|M_i,I)}{p(M_j|I)p(D|M_j,I)} = \frac{p(M_i|I)}{p(M_j|I)} B_{ij} \,, \qquad (29)$$

where O_{ij} is the **odds ratio** in favor of model M_i over model M_j, B_{ij} is the so-called **Bayes' factor** and $p(M_i|I)/p(M_j|I)$ is the **prior odds ratio**. The Bayesian odds ratio is the product of the ratio of the prior probabilities of the models and the ratio of their global likelihoods.

Markov chain Monte Carlo

The need becomes clear for a mathematical tool that is able to efficiently evaluate the multidimensional integrals required in the computation of the marginal distributions.

The aim is to draw samples from the **target distribution**, $p(\boldsymbol{\lambda}|D, I)$, by constructing a pseudo-random walk in parameter space such that the number of samples drawn from a particular region is proportional to its posterior density. This is achieved by generating a **Markov chain**, whereby a new sample, $\boldsymbol{\lambda}_{t+1}$, depends on the previous sample, $\boldsymbol{\lambda}_t$, according to a time-independent quantity called the **transition kernel**, $p(\boldsymbol{\lambda}_{t+1}|\boldsymbol{\lambda}_t)$. After a burn-in phase, $p(\boldsymbol{\lambda}_{t+1}|\boldsymbol{\lambda}_t)$ should be able to generate samples of $\boldsymbol{\lambda}$ with a probability density converging on the target distribution.

Markov chain Monte Carlo (cont.)

We generate a Markov chain by using the **Metropolis–Hastings algorithm**. Let us denote the current sample by $\boldsymbol{\lambda}_t$. We would like to steer the Markov chain toward the next sampling state, $\boldsymbol{\lambda}_{t+1}$, by first proposing a new sample, $\boldsymbol{\xi}$, to be drawn from a proposal distribution, $q(\boldsymbol{\xi}|\boldsymbol{\lambda}_t)$, that can have almost any form. The proposed sample is then accepted with a probability given by:

$$\alpha(\boldsymbol{\lambda}_t, \boldsymbol{\xi}) = \min(1, r) = \min\left[1, \frac{p(\boldsymbol{\xi}|D, I)}{p(\boldsymbol{\lambda}_t|D, I)} \frac{q(\boldsymbol{\lambda}_t|\boldsymbol{\xi})}{q(\boldsymbol{\xi}|\boldsymbol{\lambda}_t)}\right], \qquad (30)$$

where $\alpha(\boldsymbol{\lambda}_t, \boldsymbol{\xi})$ is the **acceptance probability** and r is called the **Metropolis ratio**. If $\boldsymbol{\xi}$ is not accepted, then the chain will keep the current sampling state, i.e., $\boldsymbol{\lambda}_{t+1} = \boldsymbol{\lambda}_t$.

Markov chain Monte Carlo (cont.)

Once the posterior pdf, $p(\boldsymbol{\lambda}|D, I)$, has been mapped, the procedure of marginalization becomes trivial. The marginal posterior distribution of a given parameter λ, $p(\lambda|D, I)$, is then simply obtained by collecting its samples in a normalized histogram. An estimate of the k-th moment of λ about the origin is then given by

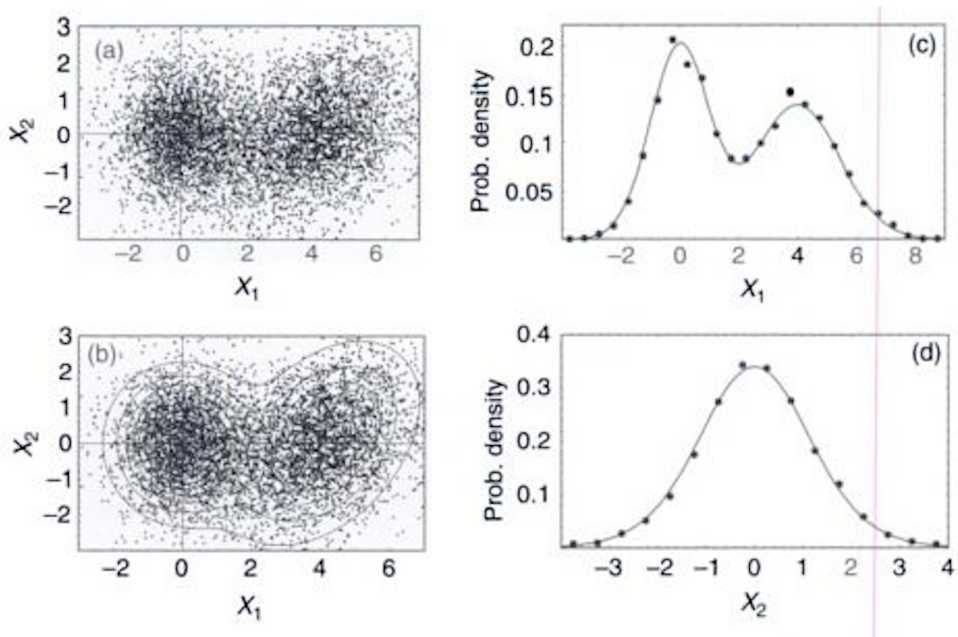
$$\langle \lambda^k \rangle \equiv \int \lambda^k p(\lambda|D, I) d\lambda \approx \frac{1}{N} \sum \lambda_t^k \,, \tag{31}$$

where N is the total number of samples.

Markov chain Monte Carlo (cont.)

Shown here are the results from a two-dimensional MCMC simulation of a double peaked posterior.

Panel (a) shows a sequence of 7950 samples from the MCMC. Panel (b) shows the same points with contours of the posterior overlaid. Panel (c) shows a comparison of the marginal posterior (solid curve) for X_1 and the MCMC marginal (dots). Panel (d) shows a comparison of the marginal posterior (solid curve) for X_2 and the MCMC marginal (dots).



Gregory (2005, *Bayesian Logical Data Analysis for the Physical Sciences*, Cambridge University Press)

Parallel tempering

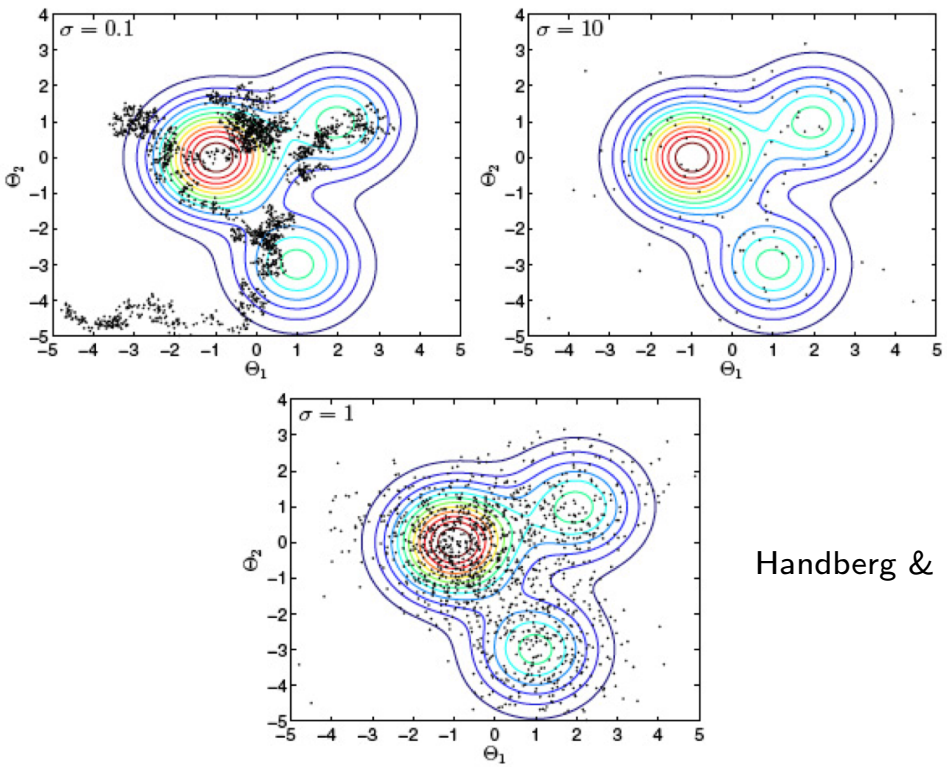
The basic Metropolis–Hastings algorithm runs a risk of becoming stuck in a local mode of the target distribution. A way of overcoming this is to employ **parallel tempering**, whereby a discrete set of progressively flatter versions of the target distribution is created by introducing a **tempering parameter**, γ. We modify Eq. (26) to generate the tempered distributions:

$$p(\boldsymbol{\lambda}|D,\gamma,I) \propto p(\boldsymbol{\lambda}|I)p(D|\boldsymbol{\lambda},I)^{\gamma}\,, \quad 0<\gamma\leq 1\,. \tag{32}$$

For $\gamma=1$, we retrieve the target distribution, while distributions with $\gamma<1$ are effectively flatter versions of the target distribution. By running such a set of chains in parallel and allowing the swap of their parameter states, we increase the mixing properties of the Markov chain.

The need for automation

The Metropolis–Hastings algorithm can also be refined by implementing a statistical control system allowing to automatically fine-tune the proposal distribution during the burn-in phase.



Handberg & Campante (2011, A&A, 527, A56)

Recommended reading

- Appourchaux, T., et al. 2012, A&A, 543, A54
- Campante, T. L. 2012, PhD thesis, Universidade do Porto
- Campante, T. L., et al. 2016, ApJ, 819, 85
- Corsaro, E. & De Ridder, J. 2014, A&A, 571, A71
- Davies, G. R., et al. 2016, MNRAS, 456, 2183
- Gregory, P. C. 2005, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with 'Mathematica' Support*, 1st ed.,Cambridge University Press
- Handberg, R. & Campante, T. L. 2011, A&A, 527, A56