Statistical searches for low signal-to-noise helioseismic oscillations

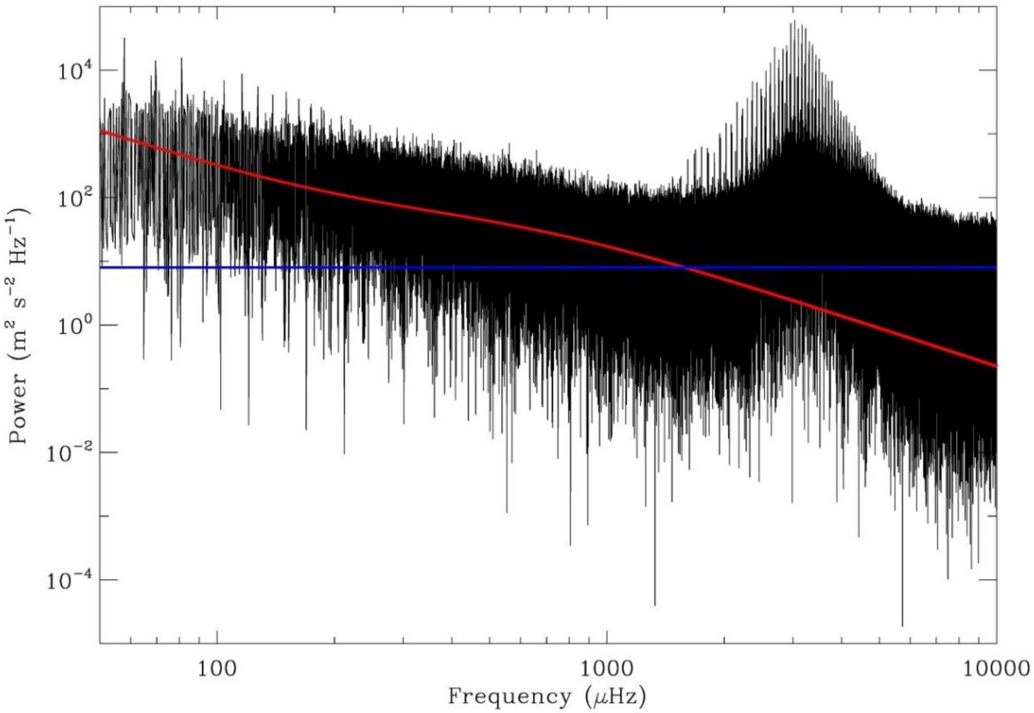
Anne-Marie Broomhall
Centre for Fusion, Space, and Astrophysics,
University of Warwick

Outline

- Contemporaneous data
- How contemporaneous does it need to be?
- Frequentist approach
- Bayesian approach
- Techniques from other areas

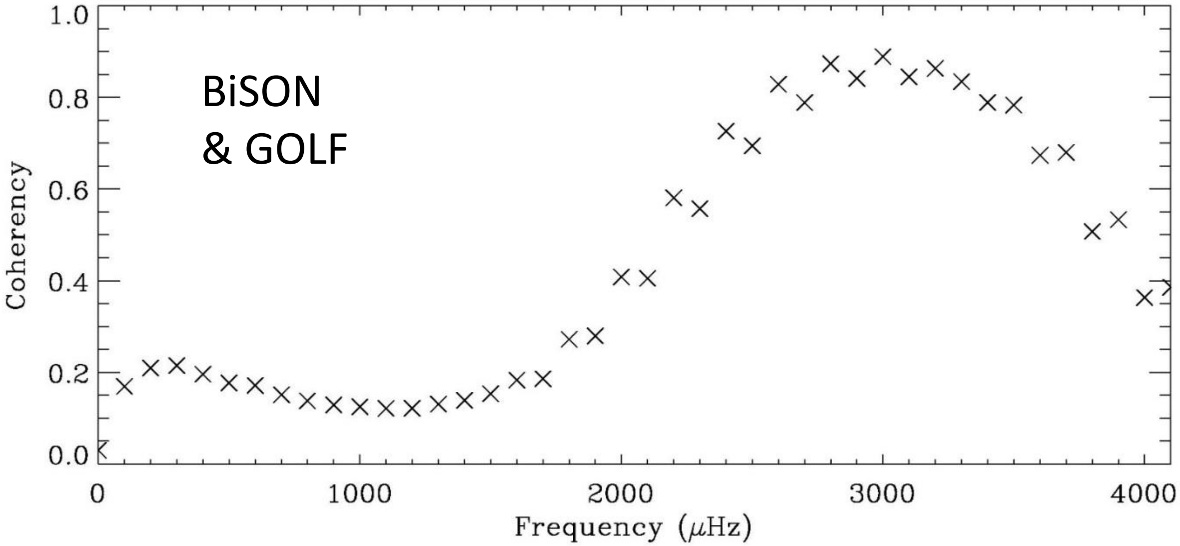
Contemporaneous data

- Main Aim: Emphasise coherent signal.
  - Coherent signal: oscillations.
  - Coherent noise: solar noise.
  - Incoherent noise: instrument, atmospheric.



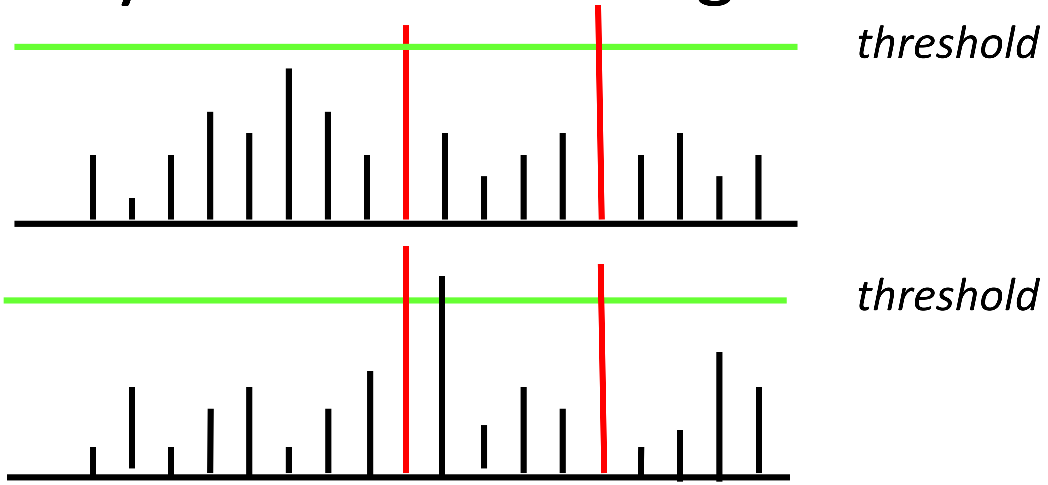
Common noise

- Solar noise will be common to data from two instruments.
- Proper allowance must be made for the level of noise common to the two sets of data.



Joint probability

- Allows searches for coincidences in contemporaneous data.
- Calculate probability of observing these coincidences in noise.
- Search for concentrations of power that lie significantly above the background noise level.



Frequentist approach

$$p = \sum_{n=0}^{\infty} \frac{1}{(n!)^2} \frac{\{\alpha[1 + k(1-\alpha) + \alpha]\}^{2n}}{2^{2n}\sigma_a^{4n+4}[\alpha^2 + \alpha(k-1) - k]^{4n+2}}$$

$$\times \frac{1}{4A_1A_2}\left[\sum_{m=0}^{n} \frac{n!r_1^{2(n-m)}}{(n-m)!A_1^m}\right]\left[\sum_{q=0}^{n} \frac{n!r_2^{2(n-q)}}{(n-q)!A_2^q}\right]$$

$$\times \exp\{-A_1r_1^2 - A_2r_2^2\},$$

where

$$A_1 = \frac{\alpha^2 + [k(1-\alpha)+\alpha]^2}{2\sigma_a^2[\alpha^2 + \alpha(k-1) - k]^2}$$

and

$$A_2 = \frac{(1+\alpha^2)}{2\sigma_a^2[\alpha^2 + \alpha(k-1) - k]^2}.$$

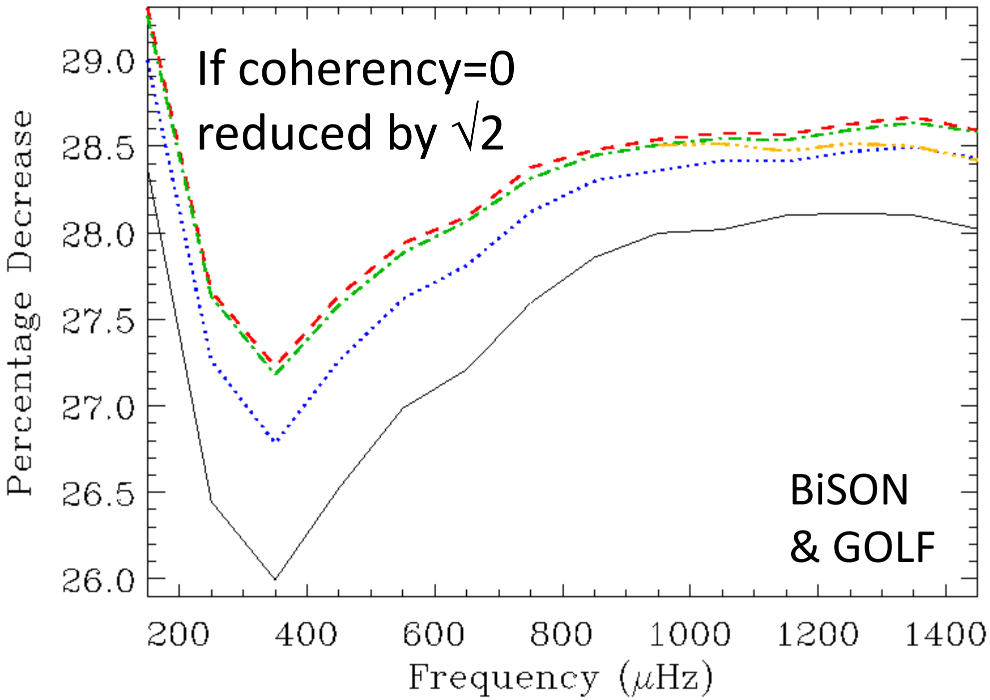
p=joint probability of a prominent data point occurring at same frequency in 2 sets of not-independent data.

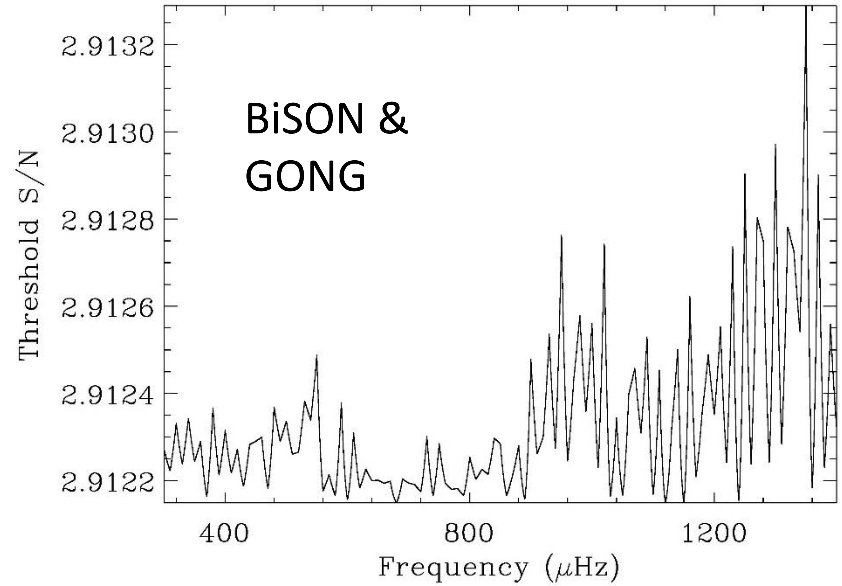
- r is AMPLITUDE of data point
- From data find
  - α: measure of proportion of common noise.
  - k and σ_a describe the variance of the data

Reduction in amplitude thresholds

% decrease in amplitude detection threshold levels

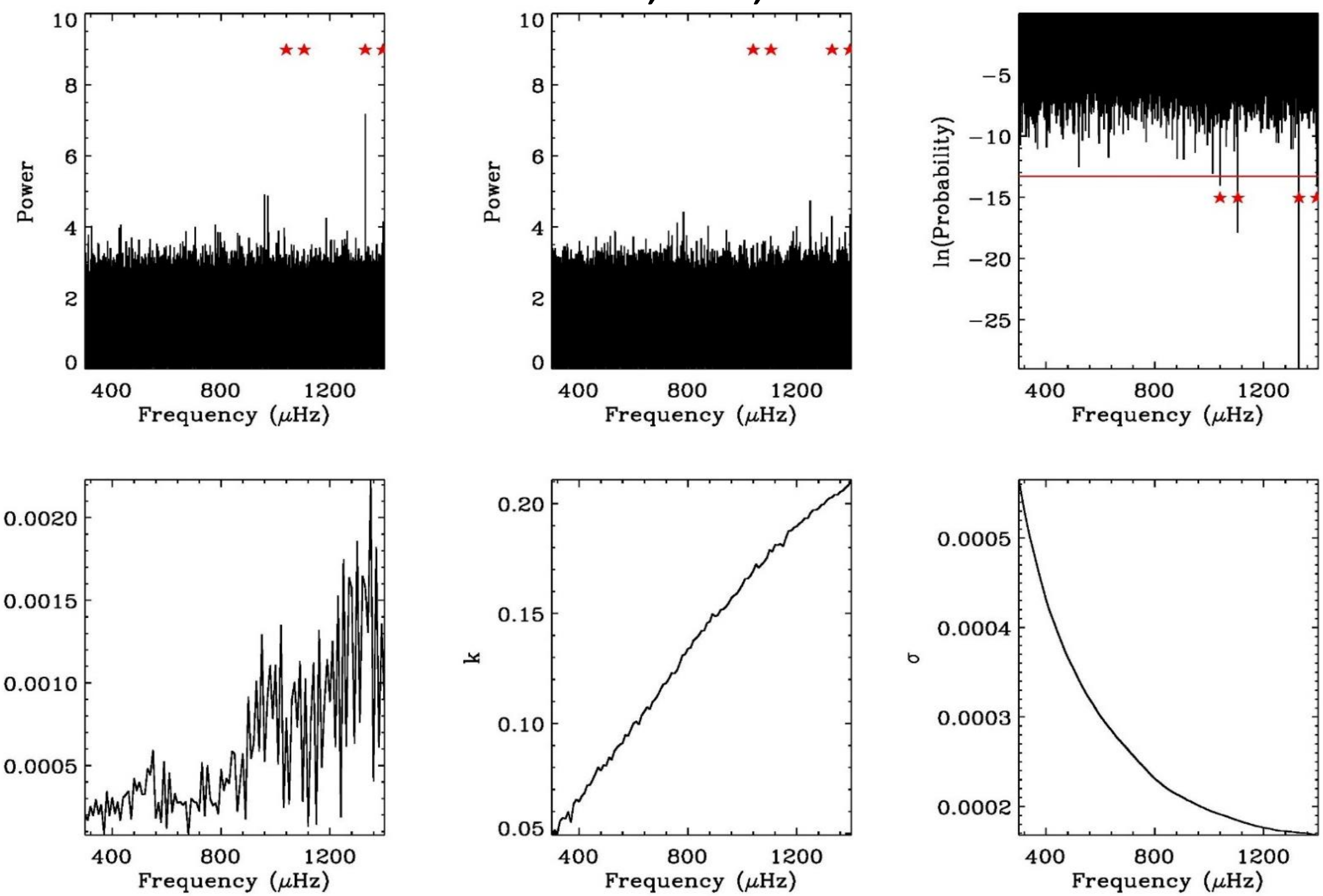
Single spike
2 consecutive spikes
2 or more spikes
2 rotationally split spikes
3 rotationally split spikes





Examples from real data

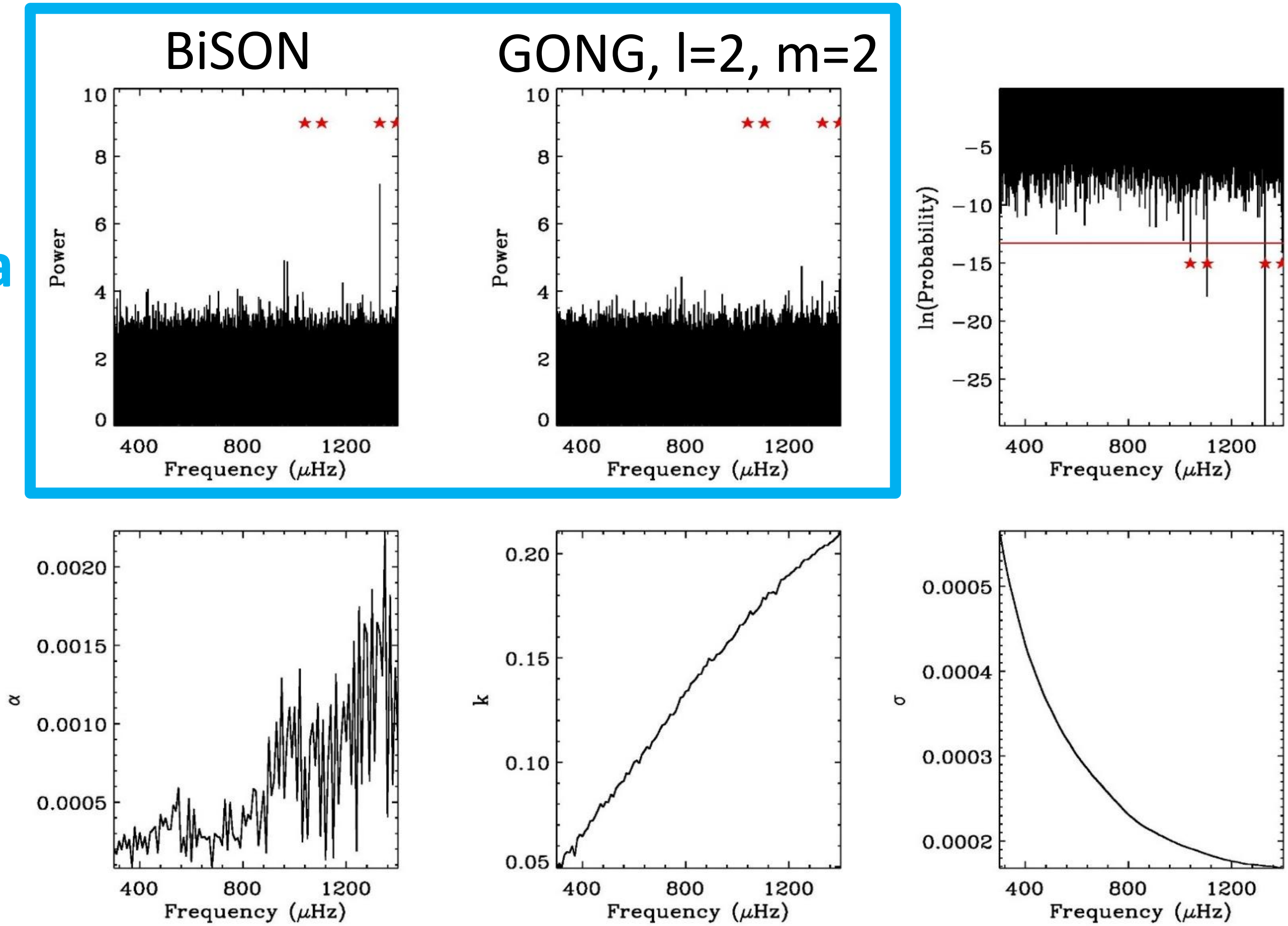
BiSON GONG, l=2, m=2



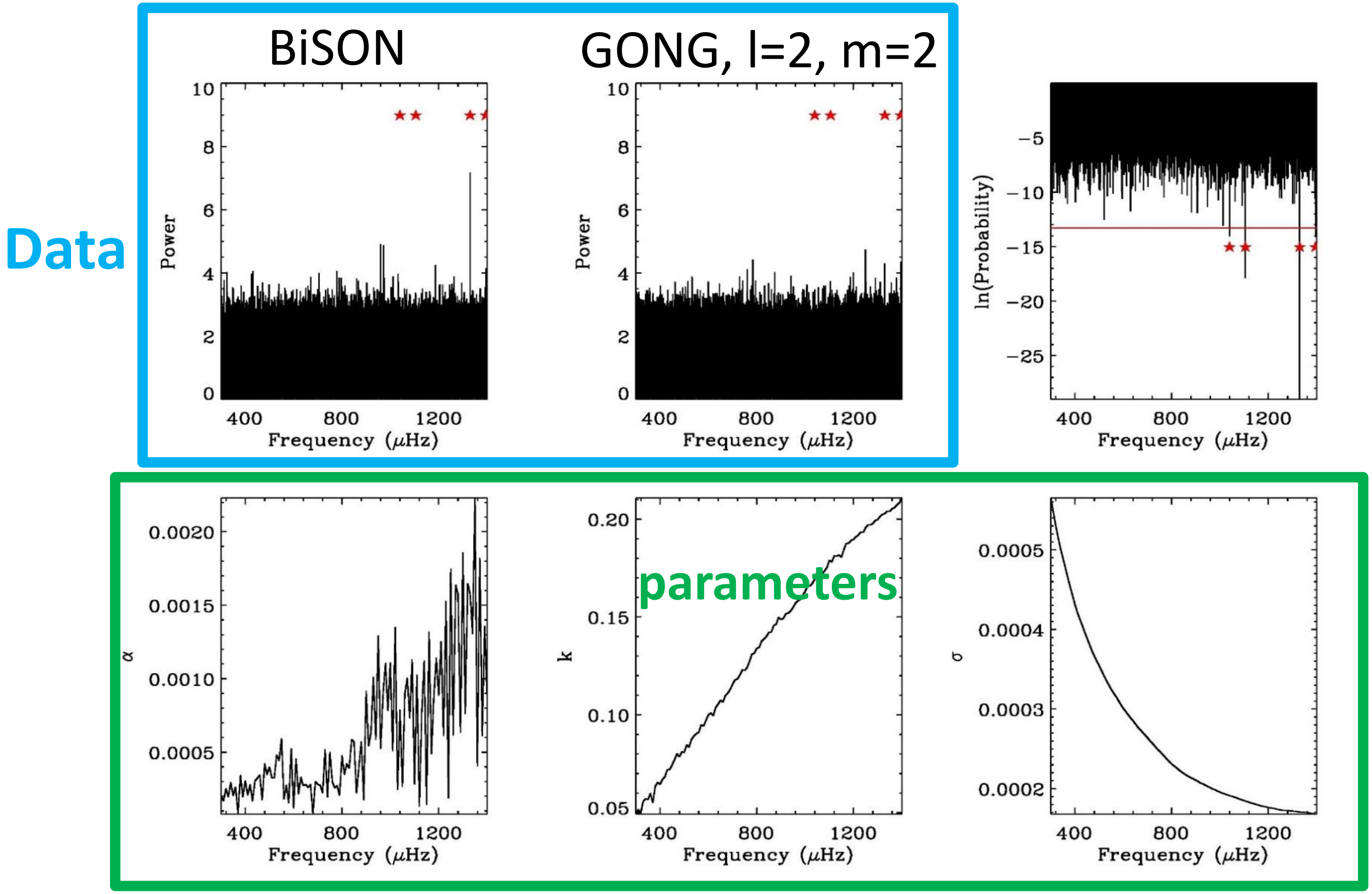
WARWICK
THE UNIVERSITY OF WARWICK

Examples from real data

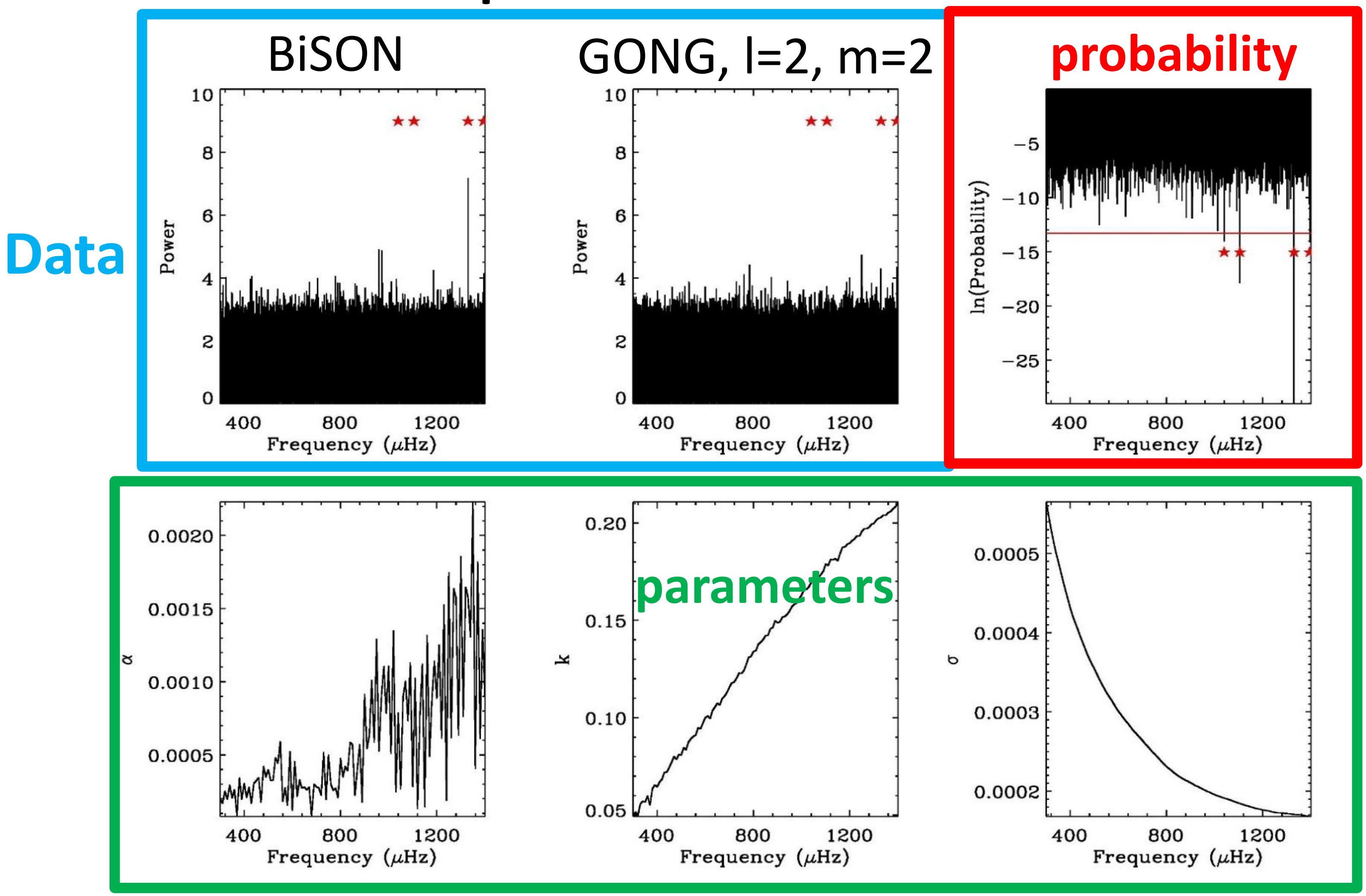
Data



Examples from real data

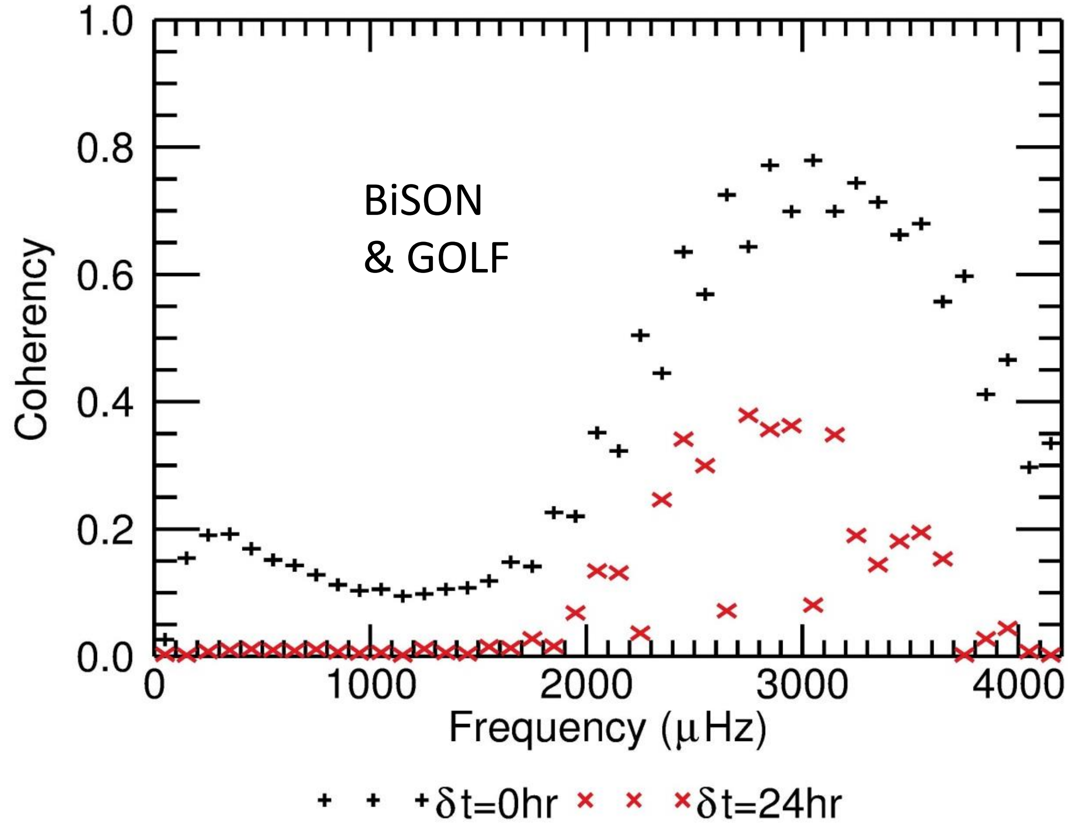


Examples from real data



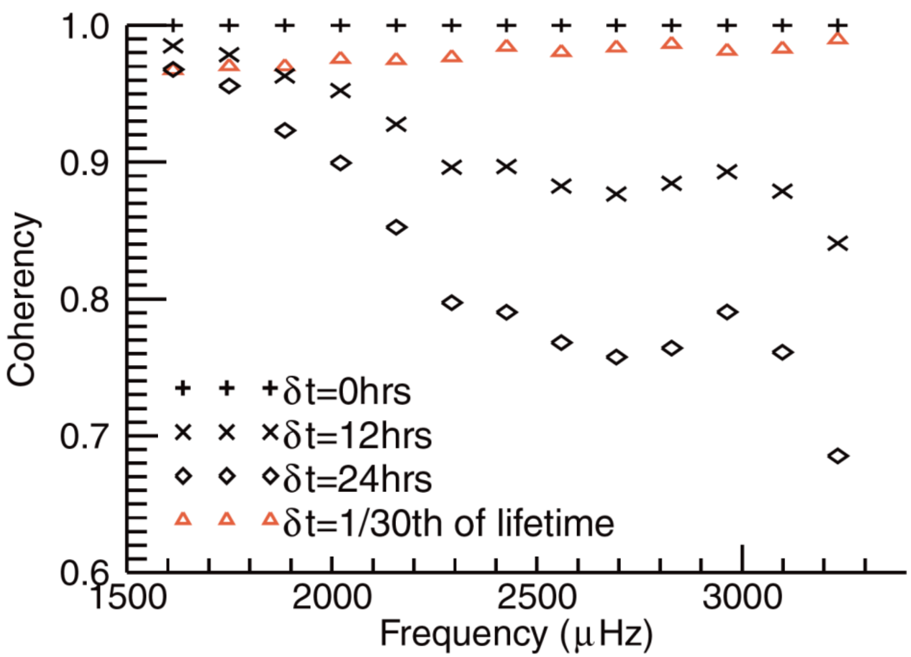
NEAR-contemporaneous data

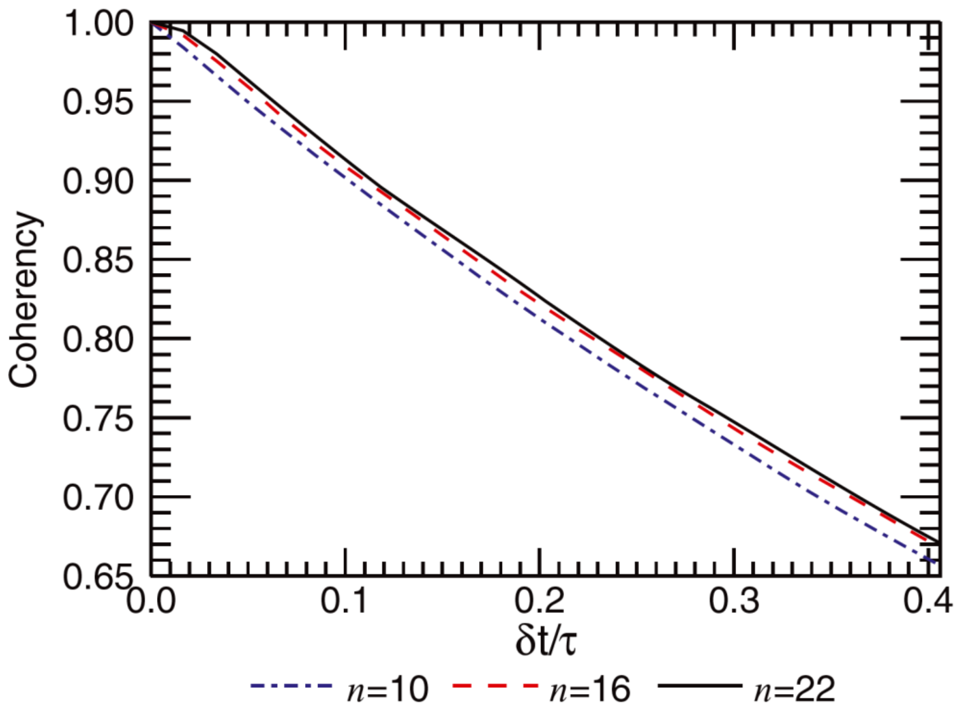
- Using data with start times that differ by 24hr can remove correlated noise.



Impact on mode coherency

- Depends on mode lifetime – what are g mode lifetimes?





Δt/τ=time shift/lifetime

Bayesian statistics

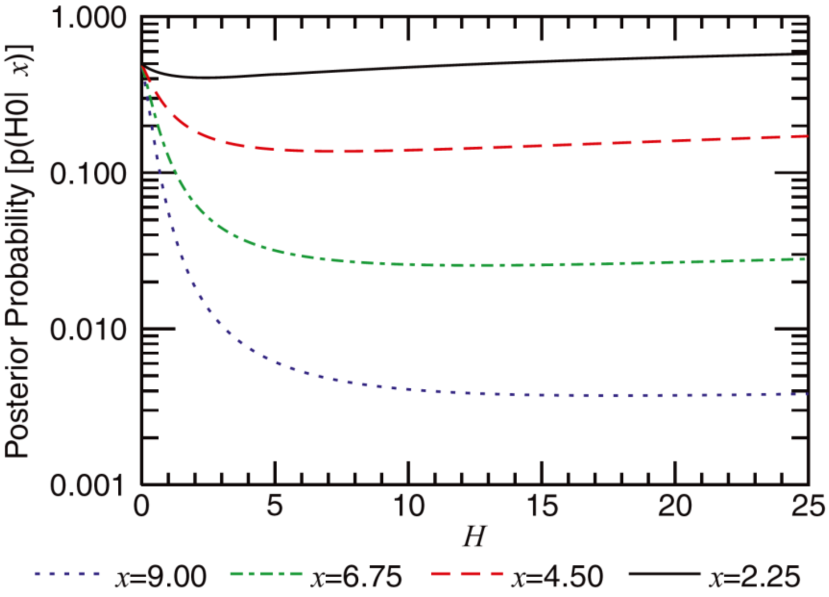
- Posterior probability given by (e.g. Appourchaux et al., 2009):

$$p(\mathrm{H0}|x) = \left[1 + \frac{(1 - p_0)}{p_0} \frac{p(x|\mathrm{H1})}{p(x|\mathrm{H0})}\right]^{-1}$$

- p_0 = prior (simplest approach it $p_0 = 0.5$).
- $p(x|\mathrm{H0})$: probability that data point with height x observed in both spectra if H0 true.
- $p(x|\mathrm{H1})$: probability that data point with height x is observed in both spectra if H1 true.

What amplitudes to g modes have?

- H0: data contain only noise.
- H1: data contain signal with max. height, *H.*
- What value should we choose for *H*?

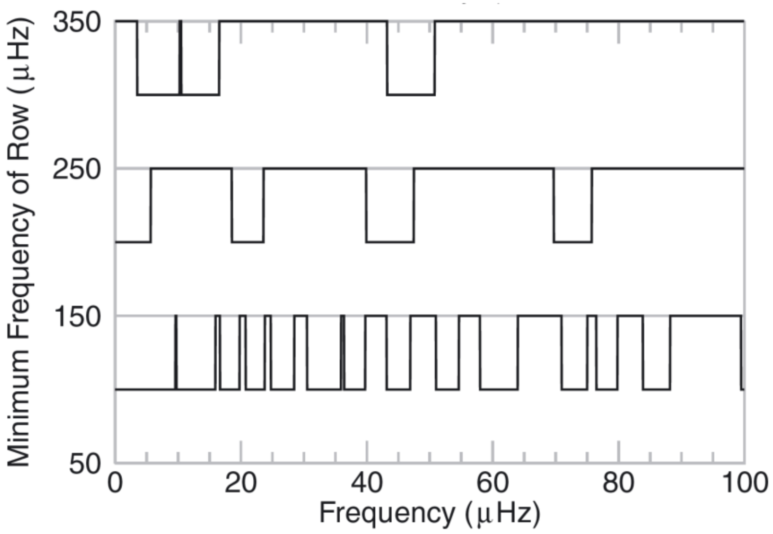


x=S/N

$$p(x|\mathrm{H1}) = \frac{1}{H}\int_0^H \frac{1}{1+H'} \mathrm{e}^{-x/(1+H')}\,\mathrm{d}H'$$

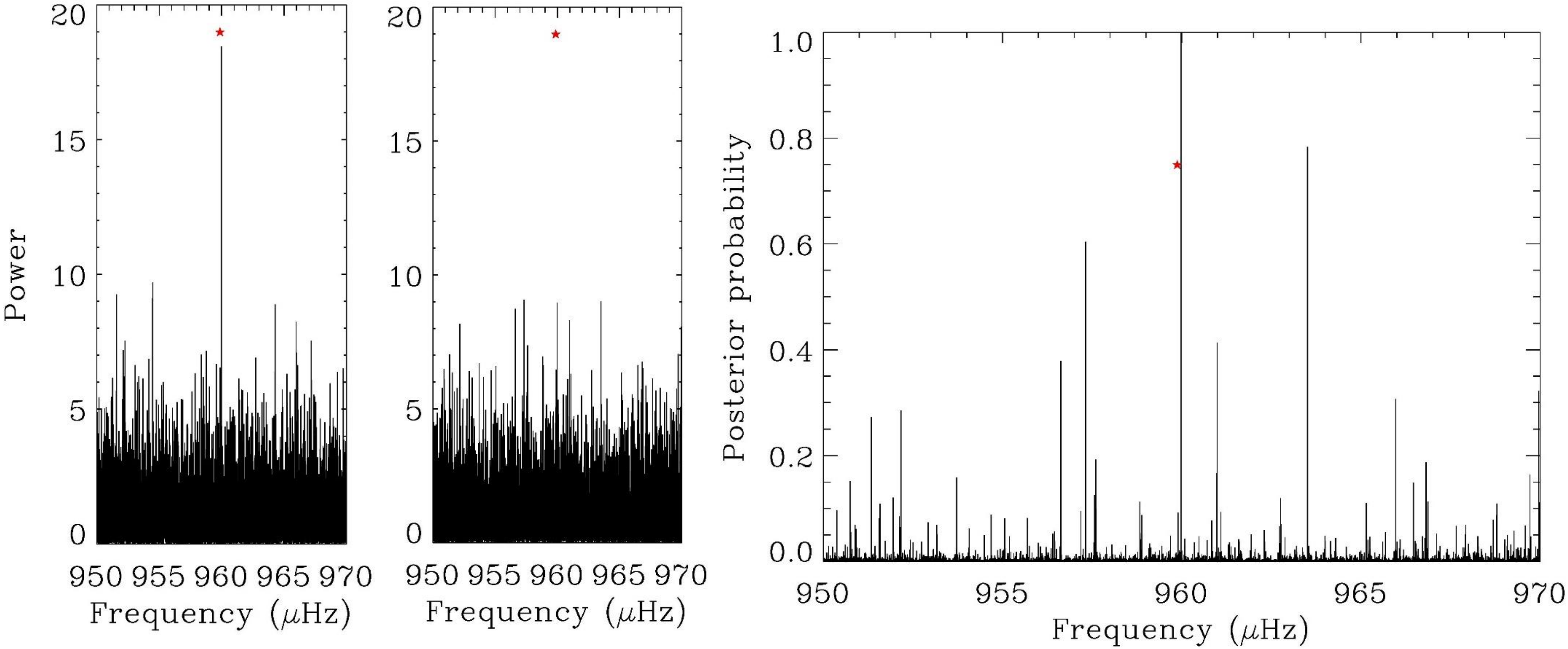
What are the frequencies of g modes?

- Mode frequencies can be used to constrain prior probability.
  - How reliable are these mode frequencies
- Are peaks expected to have width?
  - Search rebinned spectra BUT..
  - Need prior information on likely widths.



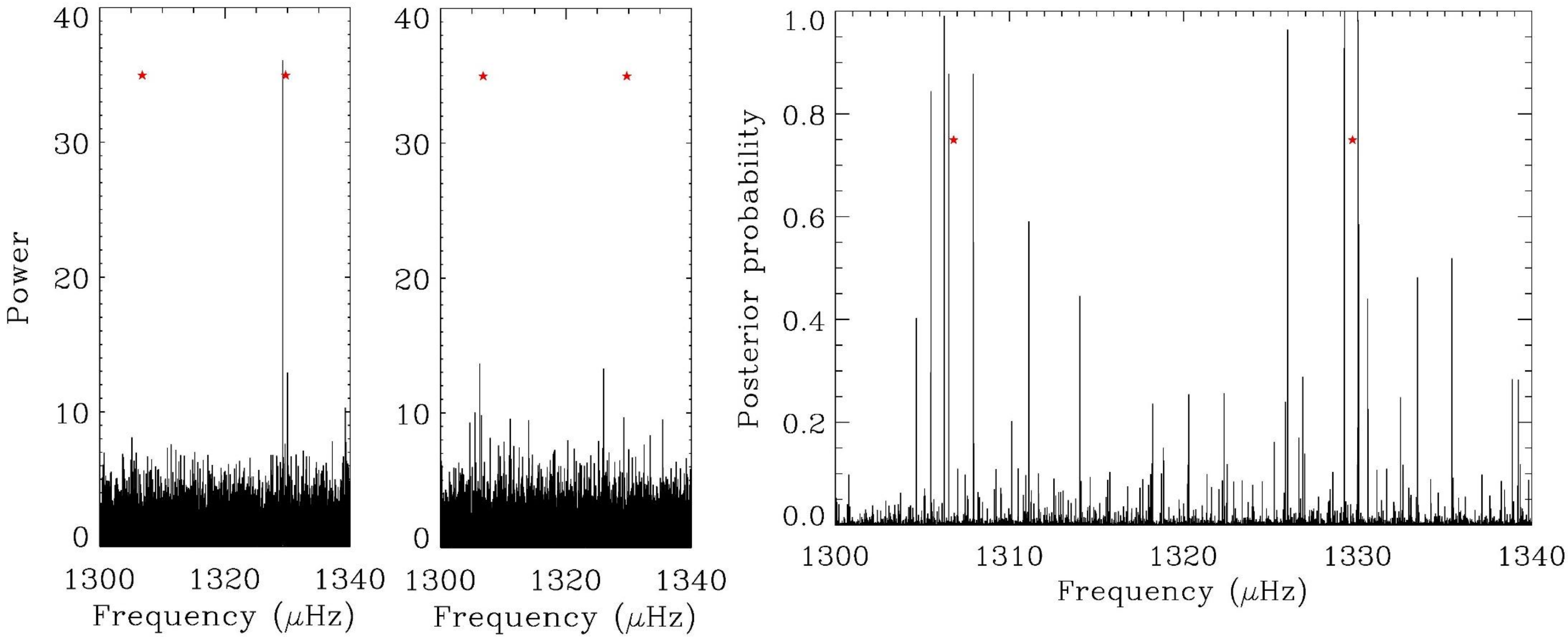
Example result: l=2, n=5

BiSON GONG, l=2, m=0



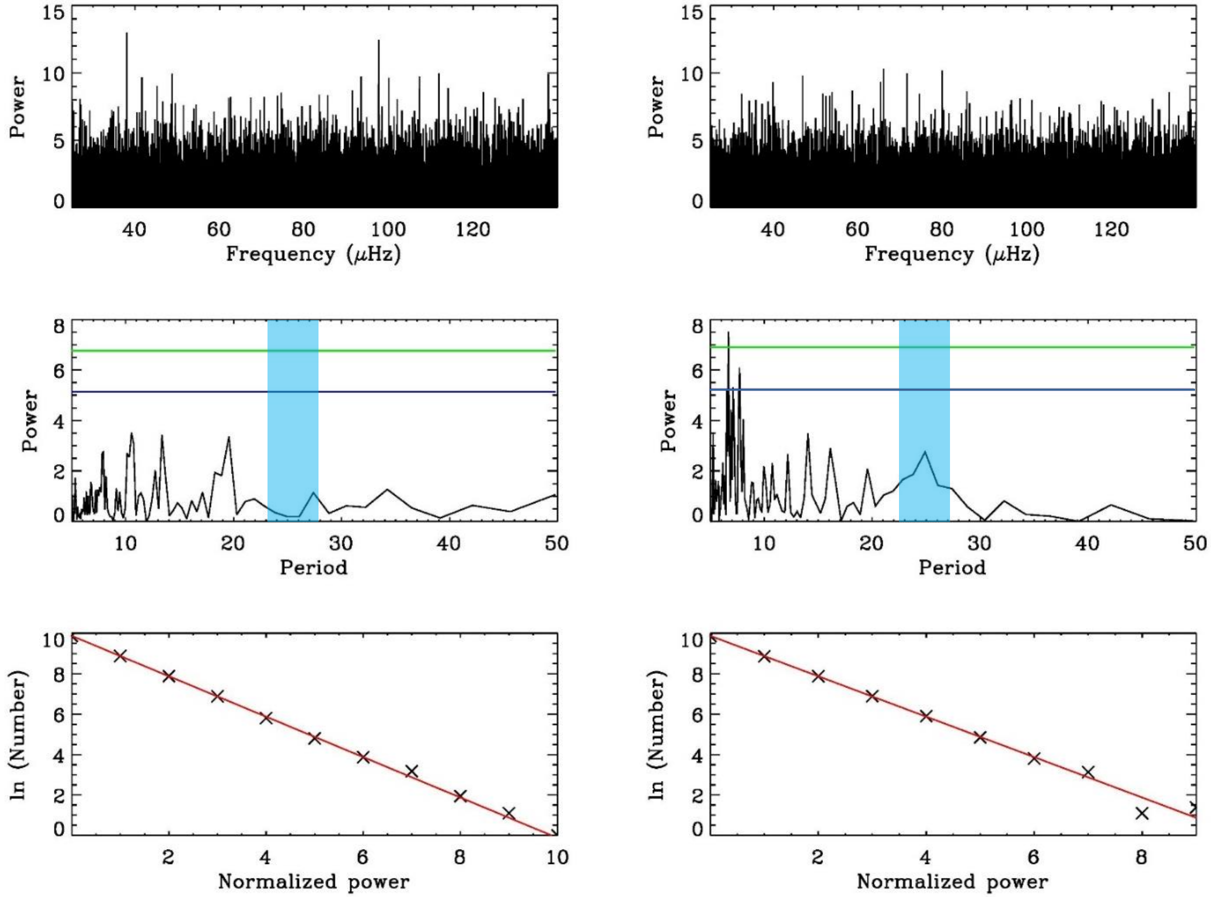
Example result: l=3, n=7

BiSON GONG, l=3, m=3



Joint periodogram

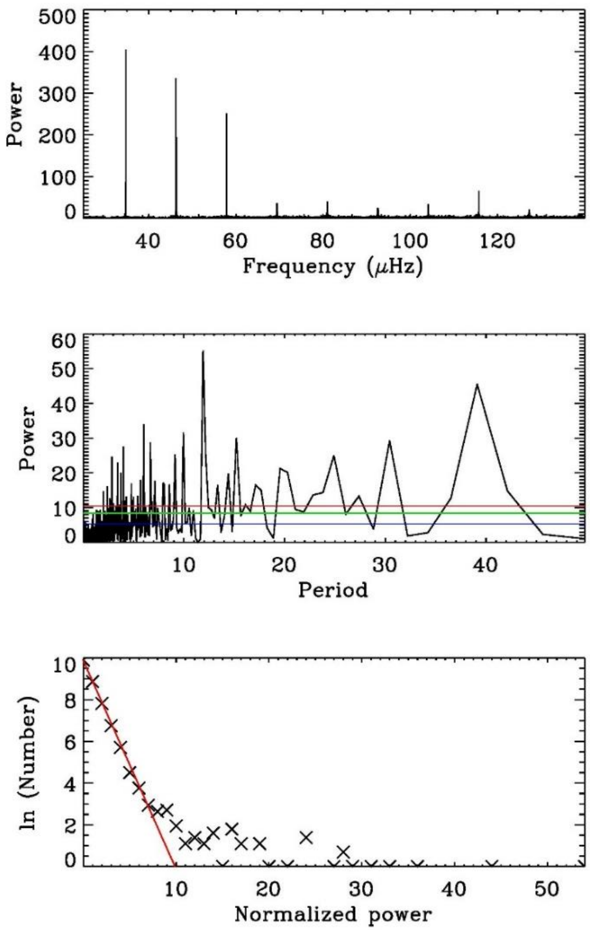
VIRGO GOLF



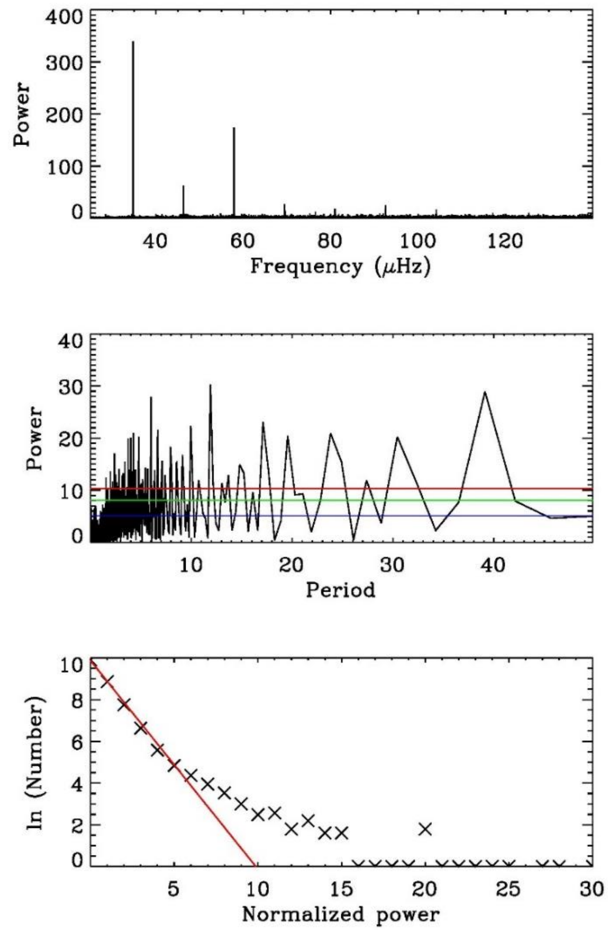
5.81σ
Combined 1 false

Joint periodogram – duty cycle

BiSON



GONG – l=2, m=0



Accounting for coloured noise

- Low-frequency spectra not white… so fit a power law…
- BUT need to account for uncertainties in this when calculating significance levels (Pugh et al, 2017 and references therein).

$$\Pr\left\{x_j > \gamma_j\right\} = \int_{\gamma_j}^{\infty}\int_0^{\infty} \frac{1}{\sqrt{8\pi}S_j}\exp\left\{\frac{-(\ln w)^2}{2S_j^2} - \frac{wz}{2}\right\} dw\, dz$$

Probability of a value x_j being greater than some threshold γ_j

Log-normal distribution of model uncertainties

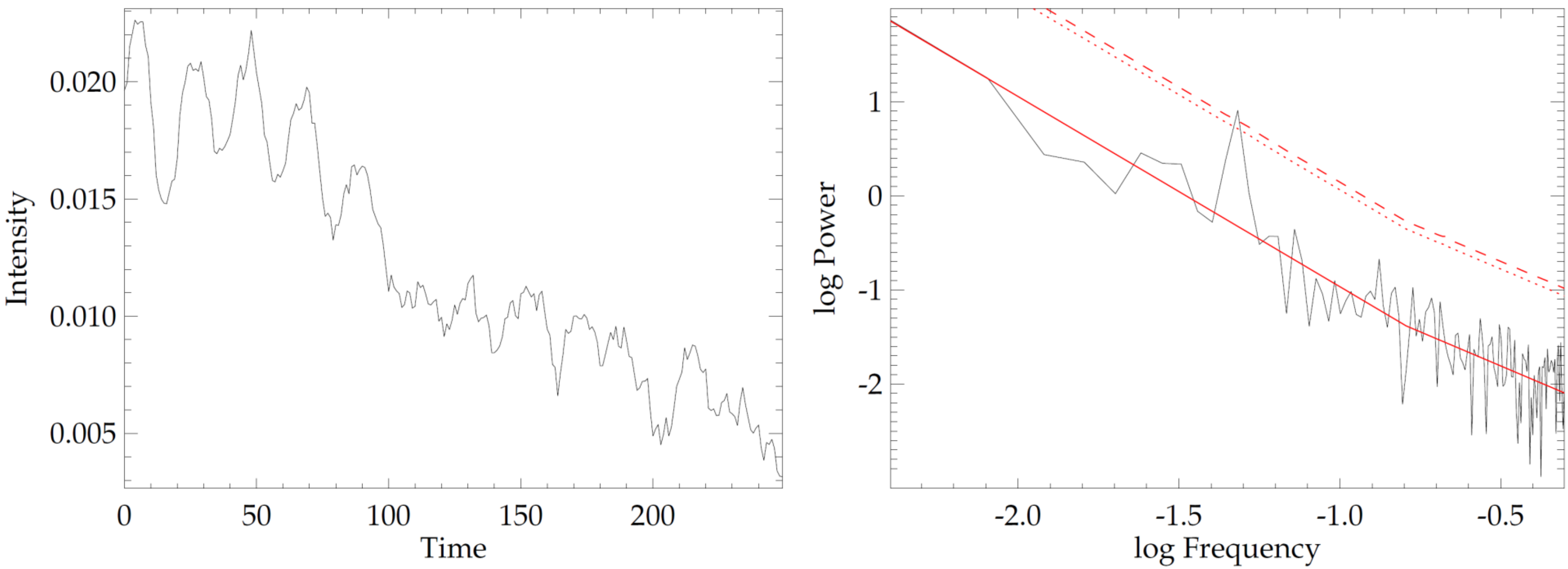
w and z are dummy variables representing power

Uncertainty on the model

Exponential χ^2 2 d.o.f distribution

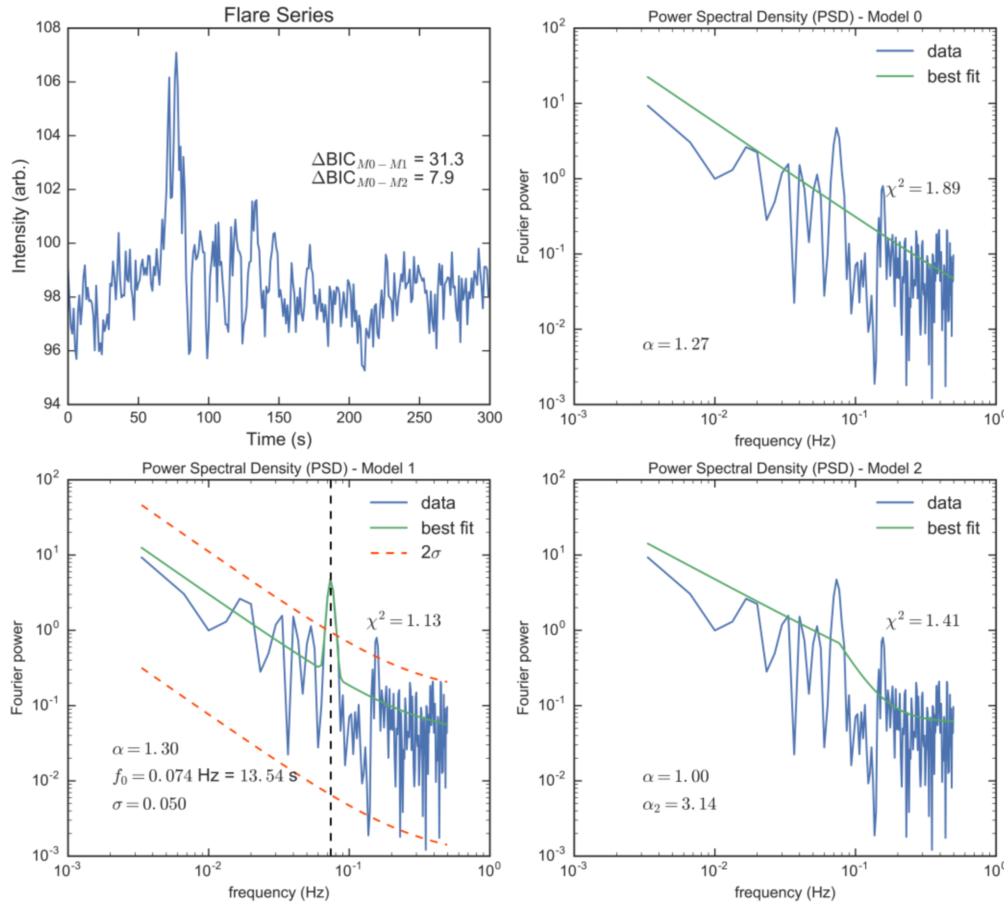
Example from flares

- Pugh et al (2017)
- If peaks have width can use same logic for rebinned power spectra.



AFINO (another example from flares)

- Automated flare inference of oscillations – Inglis et al., 2015, 2016.
- Model comparison: power law with and without Gaussian enhancement i.e. signal.
- Bayesian Information Criterion.



WARWICK
THE UNIVERSITY OF WARWICK

Summary

- Low frequency p modes and g modes are important for inversions of the solar interior.
  - BUT these modes have low-amplitudes.
  - Any detections need to be statistically robust.
  - Various tools have been developed.
- Some new low frequency p modes detected.
- Set thresholds on amplitudes of g modes.
- Future aims:
  - Incorporate SDO data.
  - Comparisons of 3 data sets.
  - Joint Bayesian analysis of periodogram of periodogram.