IAS
Institut d'Astrophysique Spatiale

Data analysis algorithms: development and verification

T.Appourchaux

KASC, Orsay, 29 October 2007

Contents

- A bit of history
- Toolbox of algorithms
- Algorithms for the pipeline
- Verification of algorithms
- Conclusion

A bit of history

Data analysis in helioseismology:

- Power spectrum statistics
- Synthetic time series
- Mode extraction
- Stellar noise extraction

So please do not reinvent the wheel…

Stellar p-mode statistics

II. THE STOCHASTICALLY EXCITED OSCILLATOR

ı governing the time evolution of the coordinate, q, of a damped harmonic o
om force, $F(t)$, reads

$$\frac{d^2q}{dt^2} + 2\Gamma \frac{dq}{dt} + \omega_0^2 q = \frac{F(t)}{M} ,$$

' is a positive damping constant. We assume that $F(t)$ is a Gaussian random p
the central limit theorem implies that other processes, such as the Poisson pro
N

Kumar and Goldreich, 1988

Synthetic time series

$$\frac{1}{(2\pi\nu_0)^2}\frac{d^2}{dt^2}y(t)+\frac{1}{2\pi\nu_0 Q}\frac{d}{dt}y(t)+y(t)=x(t)\ ,$$

: $y(t)$ is the displacement, ν_0 is the frequency o
mped oscillator, Q is a constant describing the dam
:(t) is a random forcing function. Equation (1) has a
hat can be expressed as a convolution:

$$y(t)=\int_{-\infty}^{+\infty}h(t')x(t-t')dt'\ ,$$

: $h(t)$ is the impulse response of the system that is ind
of $x(t)$ and therefore only depends on the left-hand si
ion (1).
e Fourier transform of equation (2) is:

$$Y(\nu)=H(\nu)X(\nu)\ ,$$

$$p_Y(\nu)=|H(\nu)|^2 p_X(\nu)$$

χ^2 with 2 degrees of freedom

Anderson et al, 1990

Synthetic data

- Full-disk data:
  - In ν: Anderson et al, 1990
    Toutain and Appourchaux, 1994
    Fierry Fraillon et al, 1998
  - In *t*: Chaplin et al, 1997
- Imaged data:
  Schou and Brown, 1994 (in *t*)
  Appourchaux et al., 1998 (in ν)

P-mode extraction

- Maximum Likelihood Estimators:
  - Duvall and Harvey, 1986
  - Anderson et al, 1990
  - Toutain and Fröhlich, 1992
  - Schou, 1992
- Statistics of the spectrum:
  - Woodard, 1984
  - Schou, 1992
  - Gabriel, 1994
- Error bars:
  - Libbrecht, 1992
  - Veitzer et al, 1993
  - Toutain and Appourchaux, 1994
- Monte-Carlo, error bars and Significance
  - Appourchaux et al, 1998

Solar noise extraction

- Harvey et al, 1983 (high frequency slope is -2)
- Rabello-Soares, 1995 (high frequency slope is -2)
- Appourchaux et al, 2002 (high frequency slope slope is -b)

Nota bene: the Harvey *model* is not a model. It is an empirical description of any noisy fluctuations having memory. This *description* is universal. It can be applied to any physical processes.

Side effect: physical processes being universal, there is no way to distinguish between stellar noise and instrumental noise from a single time series.

Toolbox of algorithms

- Spectrum estimators
- Statistical test
- Mode extraction
- Stellar noise extraction
- Error bars

Spectrum estimation

- Spectral decomposition:
  - Fourier spectrum (for uniform sampling)
  - Lomb-Scargle (for non-uniform sampling)
  - Generalized LS (Bretthorst, 2001)
- Mean power spectra estimation:
  - Smooth spectrum
  - Multitapering of time series
- Frequency matching (oversampling and bin shifting)
- Time-frequency spectrum

Statistical tests

- Statistical properties of any spectrum estimation must be *analytically* understood
- Use hypothesis testing (H_0, H_1)
- Use what you know about the spectrum (Bayesian approach)
- Don’t overinterpret the data nor your statistical test

Mode extraction

- Solar-like stars:
  - Use Maximum Likelihood Estimation
  - Follow CoRoT recipe ?
- Classical pulsators
  - Use sine wave fitting (iterative or not)
  - Follow CoRoT recipe?

HH#6: Recipe generation

- COROT frequencies to be used as reference and properly referenced to
- Data reduction group provides generic recipe for:
  - Solar-like stars, heavier stars
  - Classical pulsators (Cepheids, β cepheids, etc…)
- Frequencies are produced using this recipe by one fitter
- Various cases:
  - Generic recipe works: OK!
  - Generic recipe fails:
    - no COROT frequencies
    - needs for more elaborate techniques

HH#6: Recipe for solar-like stars

- General agreement:
  - S1: Normalisation such that the power in [0,Nyquist/2] is half the square of the rms of the time series (σ^2)
  - S2: Inclination angle must be determined beforehand
  - S3: Total power in a multiplet is one (angle compensation)
- Steps of the recipe:
  - Compute power spectra (or Lomb Scargle)
  - Normalize according to S1
  - Perform degree tagging (echelle diagramme) and guess parameters
  - Estimate best possible inclination angle (S2) and use it as a fix parameter
  - Fit a symmetrical profile over a window of $\Delta\nu_0/3$, assuming a white noise, the same linewidth for pair of modes (l=0-2 or l=1-3) and different splitting for each degree (Multiplet as per S3)

HH#6: Recipe for classical pulsators

- General agreement:
  - C1: Detection level set to 1% for the [0 μHz ,5000 μHz] window (taking into account the number of bins in the window)
  - C2: The mean noise level in a 10-μHz window will be determined using the median or other methods
  - C3: After detection, frequencies, amplitudes and phases will be obtained by fitting the time series by the hundred (requiring frequency filtering)
- Steps of the recipe:
  - Compute power spectra (or Lomb Scargle)
  - Normalize according to S1
  - Detection level computation according to C1
  - Mean noise level computation according to C2 in the 10-μHz wide windows over the range [0 μHz ,5000 μHz]
  - Detection by getting all peaks above the product of the mean noise level x detection level
  - Fit of set of 100 sine wave in the time series after filtering in the frequency domain

Stellar noise extraction

Superposition of many short lived components

$$P(\nu) = \frac{A}{1 + (2\pi\nu\tau)^b}$$

- *A* is the amplitude
- τ is the lifetime (granulation, etc…)
- *b* is the slope for high frequency

Kepler extraction ?

- First step:
  - Fit modes with CoRoT recipe
  - Fit stellar noise
- Second step:
  - Use input parameters for global parameters
  - Derive full error matrix
  - Automated…?

Kepler pipeline: a minimal set

- Fourier, LS and multitaper
- Mode parameters:
  - Collapsed Echelle Diagramme (Large separation)
  - Degree tagging (Echelle diagramme)
  - Maximum Likelihood Estimation + error bars
  - Sine wave fitting + error bars
- Stellar noise parameters
- Frequency separations:
  - Large, small,…

Kepler pipeline: a better set?

- Minimal set +
- Asymmetry (noise/mode correlation)
- Bayesian inference
- Global fitting per star
- Global fitting on the HR diagram

Verification

- Minimal set: can be done in the framework of AsteroFlag thru HH
- End-to-end test: from model to model (thru data)
- Better set: some parts being developed, and to be tested, some other to be implemented

Conclusion

- Algorithms exist
- Need to pick up elements from tool box
- Modular approach (evolution)
- Automation needs robust algorithms