Automatic search for optimal models using Levenberg-Marquardt algorithm

Sébastien Deheuvels[1] & Yveline Lebreton[2]

[1] IRAP, Observatoire de Midi-Pyrénées, France ; [2] GEPI, Paris Observatory, France

23 May 2016

Numerical codes

Internal structure and evolution code

- Cesam2k *(Morel & Lebreton, 2008)*

  if rotation : cestam *(Marques et al., 2013)*, not used in H&H

- options for input physics and parameters, *e.g. Lebreton & Goupil (2014)*

  - solar mixture : GN93 (Grevesse & Noels 93), AGSS09 (Asplund et al. 09)
  - EoS : OPAL 2005 (update of Rogers & Nayfonov 02)
  - opacities : OPAL96 (Iglesias & Rogers 96), Wichita (Ferguson et al. 05)
  - nuclear rates : NACRE (Angulo et al. 99), LUNA (Formicola et al. 04, for e.g. $^{14}N(p,\gamma)^{15}O$)
  - convection : MLT (Böhm-Vitense, 65), CGM (Canuto et al. 96)
  - atomic diffusion : Michaud & Profitt 93, Burgers 69
  - overshooting $d_{\rm ov} = min(\alpha_{\rm ov} \times H_p, \alpha_{\rm ov} \times R_{cc})$
  - atmosphere : T(τ) laws ; Eddington (grey), Hopf, Kurucz, MARCS, etc.

- compared to ASTEC, CLES, GARSTEC, STARROX by the ESTA CoRoT team (Monteiro et al. 06, Lebreton et al. 08a,b).

Numerical codes

Oscillation codes (adiabatic)

- LOSC (*Scuflaire et al., 2008*)
- GYRE (*Townsend & Teitler, 2014*)

The Levenberg-Marquardt method

- Newton algorithm
  - Based on a 2nd-order Taylor expansion of χ^2 around current location
  - Exact for quadratic functions
  - Efficient to find a close minimum, inefficient to approach a distant one

$$\boldsymbol{\delta}\mathbf{a} = -\mathbf{H}^{-1} \cdot \nabla\chi^2$$

Hessian matrix

$$H_{kl} \equiv \frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = 2\sum_{i=1}^{N}\left[\frac{1}{\sigma_i^2}\frac{\partial y_i(\mathbf{a})}{\partial a_k}\frac{\partial y_i(\mathbf{a})}{\partial a_l} - \frac{y_i - y_i(\mathbf{a})}{\sigma_i^2}\frac{\partial^2 y_i(\mathbf{a})}{\partial a_k \partial a_l}\right]$$

Gradient

$$\nabla_k \chi^2(\mathbf{a}) \equiv \frac{\partial \chi^2}{\partial a_k} = -2\sum_{i=1}^{N}\frac{y_i - y_i(\mathbf{a})}{\sigma_i^2}\frac{\partial y_i(\mathbf{a})}{\partial a_k}$$

$\mathbf{a}$ = set of free parameters

- Gradient-descent method
  - Steepest descent opposite to the gradient: $\delta a_k = -\gamma_k \nabla_k \chi^2$
  - Optimal step: $\gamma_k = 1/H_{kk}$
  - Efficient to approach the minimum, but often slow to converge

The Levenberg-Marquardt method

$$\boldsymbol{\delta}\mathbf{a} = -\mathbf{M}^{-1} \cdot \nabla\chi^2$$

Newton algorithm

$$M_{kl} = H_{kl}$$

Gradient-search method

$$M_{kl} = H_{kl} \times \delta_{k,l}$$

- **Levenberg-Marquardt method**
  - Combine both methods:

$$\boldsymbol{\delta}\mathbf{a} = -\mathbf{M}^{-1} \cdot \nabla\chi^2 \quad \text{with} \quad \begin{cases} M_{kk} & = H_{kk}(1+\lambda) \\ M_{kl} & = H_{kl} \ \text{ if } \ k \neq l \end{cases}$$

  - λ >> 1: Gradient-search method (approach)
  - λ << 1: Newton method (convergence)
  - Initialize λ to a "large" value
  - Decrease λ as the function to be minimized decreases
  - Typically **only ~ 10 iterations** needed to converge

Model optimization

Levenberg-Marquardt method, *Miglio & Montalbán, (2005)*

- Choose a set of observational constraints $x_{\mathrm{i,obs}}$
  non seismic : T_{eff}, L, $[\mathrm{Fe/H}]_{\mathrm{surf}}$, $\log \mathrm{g}$, radius, etc.
  seismic : frequencies $\nu_{n,l}$, frequency separations or ratios
- Adjust unknown star properties
  age, mass, initial helium abundance Y_0, initial metallicity $[\mathrm{Fe/H}]_0$
  mixing-length parameter α_{MLT}, overshooting parameter α_{ov}, etc.
- Other properties are outputs of the optimized model (R, internal properties)
- Need at least as many constraints as unknown parameters

χ^2-minimization, accounting for correlations

$$\chi^2 = \sum_{i=1}^{N_{\mathrm{obs}}} \left(x_{\mathrm{i,mod}} - x_{\mathrm{i,obs}}\right)^{\mathrm{T}} . C^{-1} . \left(x_{\mathrm{i,mod}} - x_{\mathrm{i,obs}}\right), \tag{1}$$

Model optimization

Sets of seismic parameters

Individual frequencies : $\nu_{n,l}$

- whole observed range
- truncated range : keep modes with orders close to ν_{max} ; drop high order modes (large freq. uncertainties)

Frequency separations : $\langle \Delta\nu \rangle$, $\Delta\nu_l$, $\langle d_{02/01/10} \rangle$, $d_{02/01/10}(n)$

Both cases need to be corrected from surface effects

Treatment of surface effects : several options

Kjeldsen et al. 2008, Brandão et al. 2011

Ball & Gizon, 2014

Sonoi et al. 2015

Model optimization

Sets of seismic parameters

Frequency separation ratios : $r_{02/01/10}(n)$

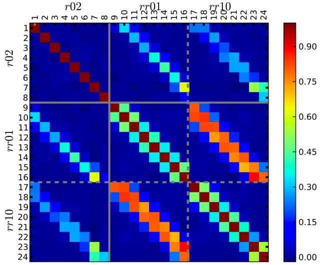
- définitions : *Roxburgh & Vorontsov (2003)*
- model ratios interpolated to observed freq. (Roxburgh & Vorontsov, 13)
- data are correlated : need to evaluate the covariance matrix

Case of correlated observables

- Correlated observables (e.g. large, small separations, ratios…):

$$\chi^2 = [\mathbf{y} - \mathbf{y}(\mathbf{a})]^T \cdot \mathbf{C}^{-1} \cdot [\mathbf{y} - \mathbf{y}(\mathbf{a})]$$

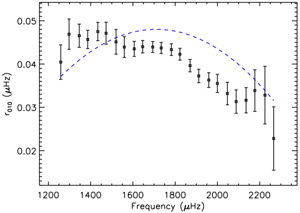
  – where **C** is the covariance matrix between the observables
  – can be calculated analytically for linear combinations of frequencies, or else numerically through Monte Carlo simulations



HD 52265

Conditioning of covariance matrix

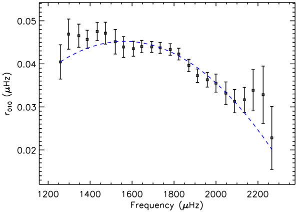
Conditioning of covariance matrix can be very poor!



- Example: fit of a 2nd-order polynomial to $\mathbf{r}_{010}$ ratios

$$\det(\mathbf{C}) \sim 10^{-100}$$
$$\mathrm{cond}(\mathbf{C}) \sim 10^{6}$$

- Solution: project onto the subspace corresponding to the N highest eigenvalues to improve the conditioning of $\mathbf{C}$



$$\chi^2 = [\mathbf{P}(\nu) - \mathbf{r}_{010}]^T \cdot (\mathbf{M}^T\mathbf{W}\mathbf{M}) \cdot [\mathbf{P}(\nu) - \mathbf{r}_{010}]$$

Model initialization

Levenberg-Marquardt method

$\rightarrow$ occurence of secondary minima ; helium-mass degeneracy

Optimization in 3 steps

- Optimization ($O1$), no seismic constraints
  - 3 observational constraints : $T_{\rm eff}$, L, $[\rm Fe/H]_{surf}$
  - take Y_0 from $(\Delta Y/\Delta Z)_\odot$, $\alpha_{\rm MLT} = \alpha_{\rm MLT,\odot}$
  - fit age A, mass M, $[\rm Fe/H]_0$
- Optimization ($O2$), seismic constraints
  - 5 constraints : $T_{\rm eff}$, L, $[\rm Fe/H]_{surf} + \langle\Delta\nu\rangle$, $\nu_{\rm max}$
  - use scaling relations $(M, R) \Longrightarrow (\langle\Delta\nu\rangle, \nu_{\rm max})$
  - fit A, M, $[\rm Fe/H]_0$, Y_0, $\alpha_{\rm MLT}$
- Further optimizations : approach from $O1$ or/and $O2$ parameters

Model initialization

Id	constraints	optimized	fixed
$O1$	$T_{\rm eff}$, L, $[\mathrm{Fe/H}]_{\rm surf}$	A, M, $[\mathrm{Fe/H}]_0$	$(\Delta Y/\Delta Z)_\odot, \alpha_{\mathrm{MLT},\odot}$
$O2$	$O1 + \langle\Delta\nu\rangle$, $\nu_{\max}$	$O1 + Y_0, \alpha_{\rm MLT}$	-
$O3$	$O1 + r_{02,n} + r_{01,n} + \nu_{n_{min},l}$	$O2$	-

Id	A (Myr)	$M/M_\odot$	Y_0	$\alpha_{\rm MLT}$	$R/R_\odot$	$\chi^2/N_{\rm obs}$
Exact	3216	1.182	0.250	0.50	1.335	-
$O1$	5436 ± 1505	1.145 ± 0.038	0.266	0.60	1.386	$3.\ 10^{-6}$
$O2$	4398 ± 1343	1.201 ± 0.020	0.242 ± 0.022	0.60 ± 0.10	1.338	1.5
$O3_2$	3685 ± 195	1.153 ± 0.008	0.266 ± 0.006	0.49 ± 0.02	1.329	1.6
$O3_{12}$	3299 ± 131	1.182 ± 0.008	0.260 ± 0.005	0.50 ± 0.01	1.338	1.2
$O3_1$	3502 ± 152	1.216 ± 0.009	0.247 ± 0.005	0.53 ± 0.02	1.355	1.3

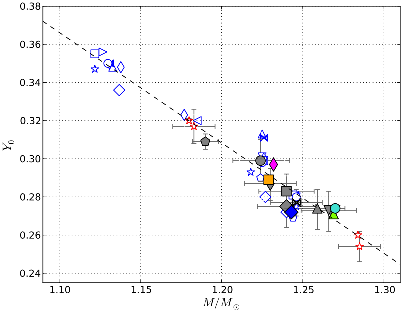
Models $O3_{1,12,2}$

- different starting parameters
- similar fit to seismic constraints
- different (Y, M) combinations → degeneracy

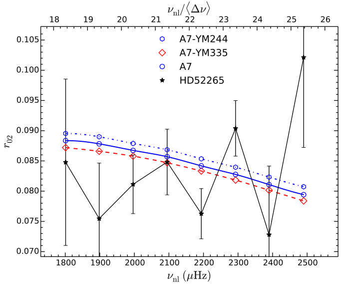
Model optimization : helium-mass degeneracy

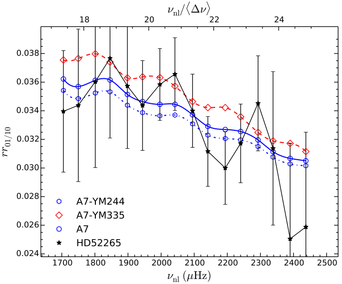
A range of solutions for different (Y_0, M)

The better the global parameters, the narrower the range.



$$Y_0 \simeq -0.58 \times M/M_\odot + 1.00$$

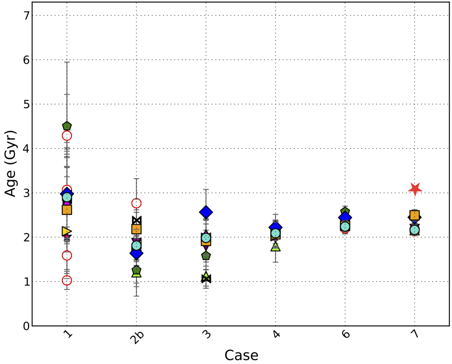




Sensitivity to model input physics & constraints

Case of HD 52265

seismic constraints	adjusted parameters	fixed
no	A, M, $(Z/X)_0$	Y, α_{conv}
$<\Delta\nu>$	A, M, $(Z/X)_0$ α_{conv}	Y
$<\Delta\nu>$, ν_{max}	A, M, $(Z/X)_0$, Y, α_{conv}	-
$<\Delta\nu>$, $<d_{02}>$	A, M, $(Z/X)_0$, Y, α_{conv}	-
$r_{02}(n)$ $rr_{01/10}(n)$	A, M, $(Z/X)_0$, Y, α_{conv}	-
$\nu_{n,l}$	A, M, $(Z/X)_0$, Y, α_{conv}	-



Fit with individual frequencies :

surface effect corrections : $A = 2.17 \pm 0.02$ Gyr ; $M = 1.27 \pm 0.02 M_\odot$

no surface effects : $A = 3.15 \pm 0.03$ Gyr ; $M = 1.30 \pm 0.02 M_\odot$

Estimation of errors

- Estimating errors for the fitted parameters:
  - Errors given by the **diagonal coefficients of the inverse of the Hessian matrix**
  - Hessian matrix obtained numerically by setting $\lambda = 0$ after convergence
  - Results may depend on the choice of steps that used to estimate the derivatives $\partial y_i(\mathbf{a})/\partial a_k$ for the Hessian

- Estimating errors for other parameters (e.g. radius):
  - No direct estimate provided through the fit
  - For a parameter π which is not a free parameter of the fit, $\pi(a_j)$

$$\sigma_\pi = \sqrt{\sum_{i,j=1}^{N} C_{ij} \left(\frac{\partial \pi}{\partial a_i}\right)\left(\frac{\partial \pi}{\partial a_j}\right)}$$

Covariance matrix =
inverse of Hessian

Estimation of errors

- Testing the errors produced by LM algorithm
  - **Input model** close to HH2a
  - Use of **classical** (T_{eff}, Z/X_{surf}, $L/L_{\odot}$) and **seismic constraints**
  - **Add random normally-distributed noise** to all observables according to the statistical errors of their measurements
  - **Series of ~ 50 optimizations** performed using LM algorithm (low precision)

Using individual frequencies ($\ell = 0, 1, 2$)

	True value	Average of fitted values	RMS of fitted values	LM errors
Mass ($M_{\odot}$)	1.22	1.223	0.038	0.029
Age (Myr)	3000	3007	149	75
Y_0	0.28	0.278	0.019	0.012
$(Z/X)_0$	0.0215	0.0214	0.0020	0.0012
α_{conv}	0.60	0.612	0.023	0.022
R ($R_{\odot}$)	1.487	1.482	0.042	0.028

Estimation of errors

- Testing the errors produced by LM algorithm
  - **Input model** close to HH2a
  - Use of **classical** (T_{eff}, Z/X_{surf}, $L/L_\odot$) and **seismic constraints**
  - **Add random normally-distributed noise** to all observables according to the statistical errors of their measurements
  - **Series of ~ 50 optimizations** performed using LM algorithm (low precision)

Using combinations of frequencies ($\delta\nu_{01}$, $\delta\nu_{10}$, $\delta\nu_{02}$, ν_0)

	True value	Average of fitted values	RMS of fitted values	LM errors
Mass ($M_\odot$)	1.22	1.227	0.026	0.019
Age (Myr)	3000	3068	194	44
Y_0	0.28	0.275	0.014	0.010
$(Z/X)_0$	0.0215	0.0220	0.002	0.001
α_{conv}	0.60	0.592	0.04	0.02
R ($R_\odot$)	1.487	1.491	0.011	0.009

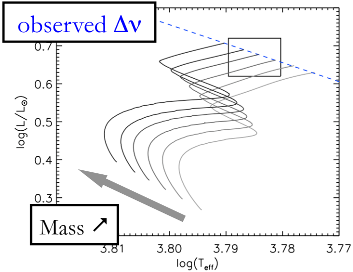
Examples of the use of LM algorithm

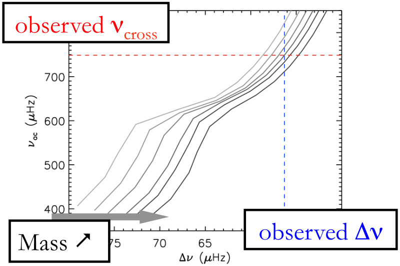
- **Main sequence stars**:
  - Detailed seismic modeling of specific targets (e.g. HD49933 Goupil et al. 2011, HD52265 Lebreton et al. 2014)
  - Test case of HD52265 to estimate the influence of seismic constraints in modeling (Lebreton et al. 2014)
  - Measuring the size of convective cores (Deheuvels et al. 2010, 2016)
  - …

- **Subgiants** with mixed modes:
  - Classical Levenberg-Marquardt procedure fails because of the fast evolution of g-mode frequencies (core contraction)
  - Bijection between (Mass, age) and ($\Delta\nu$, ν_{cross}) for a given set of other parameters

Examples of the use of LM algorithm

- **Subgiants** with mixed modes:





Free parameters of the fit: $(\cancel{M},\ \cancel{\tau},\ Y_0,\ \alpha_{\rm ov},\ (Z/X)\ \ldots)$

fixed to $(\tilde{M}, \tilde{\tau})$ by fitting $(\Delta\nu, \nu_g)$

Successfully applied to CoRoT subgiant HD49385 (Deheuvels et al. 2011)

Levenberg-Marquardt optimization

Advantages

- No restrictions on the number of unknown parameters to be estimated provided there are enough observational constraints
  $\rightarrow$ no necessary a priori assumption on Y, $(\Delta Y/\Delta Z)$, $\alpha_{\mathrm{MLT/ov}}$
  $\rightarrow$ easy to modify/add physical processes and refit parameters
- Errors on the fitted parameters easily obtained via the Hessian matrix
- the procedure is adaptable to use other stellar evolution codes (MESA), oscillation codes (ADIPLS, GYRE, LOSC)

Caveats

- Secondary minima
- Error estimates for parameters that are not fitted (ex. radius, convection zone boundaries, etc.) have to be evaluated a posteriori.
- The order of magnitude of the error bars is correct but rather imprecise (work to be done)