Astron. Astrophys. 333, 362–368 (1998)

ASTRONOMY AND ASTROPHYSICS

Power spectrum modelisation of helioseismic data: an application to the measurement of solar p-mode uncertainties

D. Fierry Fraillon[1], **B. Gelly**[1], **F.X. Schmider**[1], **F. Hill**[2], **E. Fossat**[1], **and A. Pantel**[1]

[1] Département d'Astrophysique, C.N.R.S. UMR 6525, Université des Sciences, F-06108 Nice Cedex 2, France
[2] National Solar Observatory / N.O.A.O., P.O. Box 26732, Tucson, AZ 85726, USA

Received 5 November 1997 / Accepted 19 December 1997

Abstract. We estimate the statistical uncertainties of low-l solar p-modes parameters based on a Monte Carlo approach. Random perturbations of ideal Lorentz profiles $L(\mathbf{a}, \nu_i)$ can provide many estimations of the set of p-modes parameters $\mathbf{a}$ and allow one to estimate statistical error-bars $\sigma_{\mathbf{a}}$ by modelling the parameters' distribution function. Unlike frequencies, which show symmetric distributions, amplitudes and linewidths have asymmetric probability density function similar to the distribution function for time-averaged energies of stochastically excited solar p-modes (Kumar, 1988). A comparison between σ_ν and uncertainties based on Hessian's computation (Libbrecht 1992, Toutain and Appourchaux 1994) shows a nice agreement. However, our error-bars take into account more statistical effects, and rely less on the initial parameters' estimation. Such a technique has been used on the IRIS power spectra computed from gapped data, and on one GONG power spectrum computed from almost continuous data. We also present IRIS linewidths and error bars averaged over the years 1989-92 and computed with a fitting strategy using imposed frequency which improves the value of both the parameter and its uncertainty.

Key words: Sun: oscillations – methods: statistical

1. Introduction

The solar eigenmodes of oscillation appear in the solar background velocity spectrum as Lorentz shaped lines. The position of the lines on the frequency axis is obviously one fundamental datum, and some solar properties can be inferred from the individual or collective study of the other parameters of these lines. Mode frequencies tables are used in the determination of solar internal structure, sound speed velocity or chemical composition, the combination of width and amplitude of the peaks can produce an estimation of the power fed to a mode from the excitation source, the measurement of the separation between split components gives an estimation of the solar internal rotation and the line asymmetry gives the location of the excitation source. As 'collective' properties, the pseudo-equidistance of the successive harmonics' 'big separation', or the 'small separation' of even-ℓ or odd-ℓ groups, are well known to be excellent tests of solar models. All these studies include an important amount of modelling of the data and optimisation of some cost function (fitting) to extract the desired model parameters from the data (Anderson et al. 1990, Gelly et al. 1997, Appourchaux et al. 1995, Elsworth et al . 1994). We have used such a fitting procedure on the IRIS network data to produce 4 tables of p-modes frequencies with error-bars (Gelly et al., 1997) and we wish to review here the modelling of the power spectrum which enables us to quote statistical error-bars on all the parameters of our fits. A study of the solar p-modes linewidths illustrates the discussion.

Send offprint requests to: fierry@irisalfa.unice.fr

2. The Monte Carlo approach to p-modes parameters uncertainties

IRIS is a non-imaging atomic resonance spectrophotometer making a full-disk measurement of the integrated solar velocity, and producing a discrete one-dimensional time-series $v(t_i)$. The existence of gaps in the IRIS network data, as well as in any other interrupted solar observation, alters the spectral density $s(\nu_i)$ of the time series mainly by the existence of sidelobes of the peaks which represent the solar eigenmodes. Our team has investigated two options: gap filling techniques allow to a certain extent to get rid of these spurious peaks (Pantel 1996). However interesting, this technique works best with 75% to 80% duty cycles, which we did not yet achieve. The other possibility is to take into account those sidelobes in the analysis of the power spectra. To the first order, they are Lorentz shaped peaks located at multiples of 11.57 μHz around each p-mode peak, and they impact significantly on the fitting strategy leading to the extraction of the mode parameters from the power spectrum. They also influence the precision of the result to an amount which, so far, has been poorly known. To compute the statistical uncertainty on p-mode parameters, we have used a Monte-Carlo (MC) approach, which can be described as a 3 stage process,

- We perform a maximum likelihood (ML) fit in a frequency range including a $l = 0 - 2$ or $l = 1 - 3$ group in order to extract a set $\mathbf{a}$ of initial parameters.
- we generate many realizations using random perturbations of the real and imaginary part of the Fourier transform, given by the initial values $\mathbf{a}$.

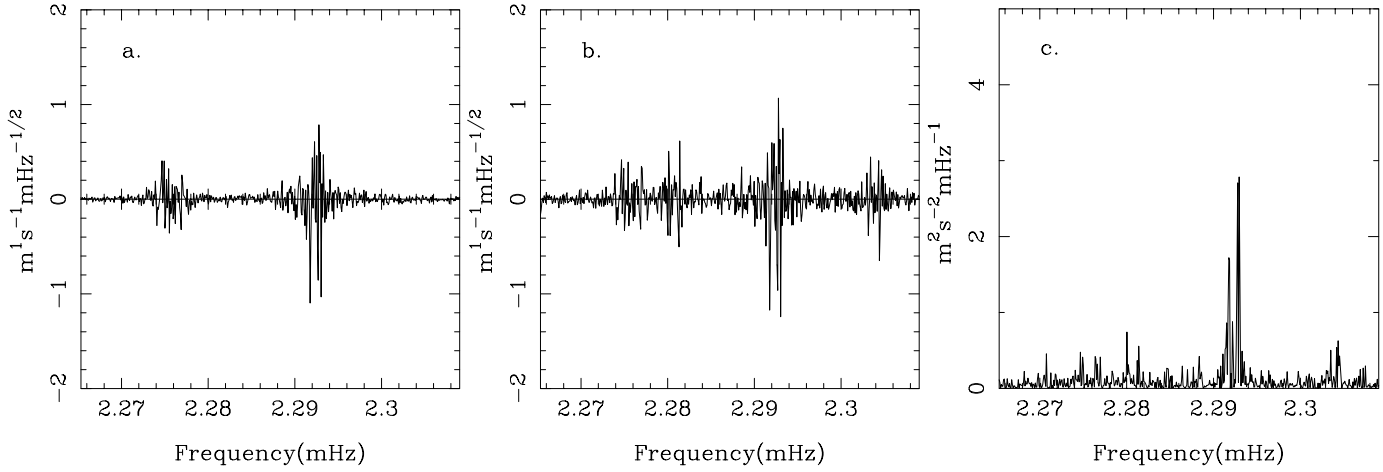

Fig. 1. a One realisation of the real part of a $\ell = 1 - 3$ group without background noise. **b** Same after convolution by the true complex window, showing the sidelobes. **c** Corresponding spectral density after addition of the background noise.

– we perform a ML fit on each simulation in order to determine the noisy parameters. Then we model the probability density function of each p-mode parameter (frequency, width and amplitude) and evaluate its statistical uncertainty.

2.1. Initial estimation of the parameters

A typical $\ell = 0$ mode is modeled by a Lorentz function depending on the line frequency ν_0, the linewidth Γ_0, the mode amplitude A_0 and the background noise B like:

$$L(A_0, \nu_0, \Gamma_0, B, \nu_i) = \frac{A_0}{1 + \left(\frac{\nu_i - \nu_0}{\Gamma_0}\right)^2} + B \qquad (1)$$

To handle cases with more than a singlet mode, we introduced a composite profile $L(\mathbf{a}, \nu_i)$ where $\mathbf{a}$ is a variable size vector of parameters which can describe just as well an $\ell = 0$ mode, $(a_1, a_2, a_3, a_4) = (A_0, \nu_0, \Gamma_0, B)$, or a group of 2 peaks like $\ell = 1$ and $\ell = 3$, $(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8) = (A_1, \nu_1, \Gamma_1, A_3, \nu_3, \Gamma_3, B, Sdls - ratio)$, where the parameter Sdls-ratio adjust the ratio between the amplitude of a peak and its nearest sidelobe. This is just an example and we refer the reader to our previous paper (Gelly et al. 1997), as to the detailed fitting strategy. A Gauss-Newton algorithm is then used to minimize (within bounds) the expression of the maximum likelihood:

$$F(\mathbf{a}) = \sum_i \frac{s(\nu_i)}{L(\mathbf{a}, \nu_i)} + log(s(\nu_i)) \qquad (2)$$

where $s(\nu_i)$ is the spectral density of the solar signal at the discrete frequencies ν_i. This gives the first estimate of the vector of parameters $\mathbf{a}$. As to the initial "guess parameters" needed by the optimization to work, they have no particular value refering to an "a-priori" knowledge of the result. The minimisation converges toward a global minimum in most cases, and two different maximum likelihood fits (with very different guess parameters) will give parameters $\mathbf{a}$ matching $\pm$ 0.2 σ.

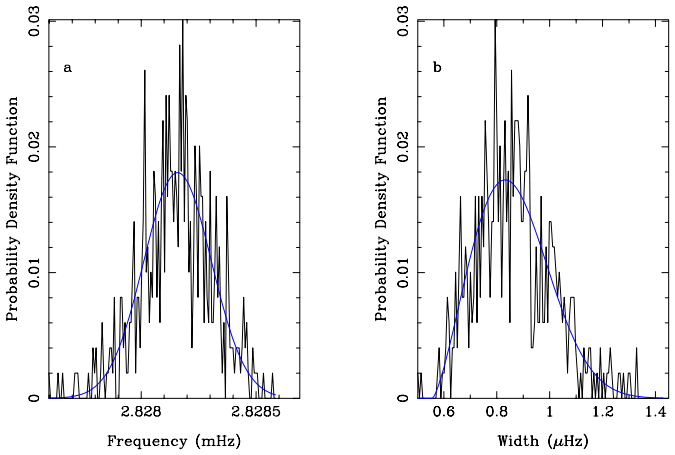

Fig. 2a and b. Probability density function of frequency (**a**) and width (**b**) of the n=19, l=1 mode for the year 1990 dataset in the MC simulation. The superimposed fits give m_ν=2828.25 $\pm$ 0.15 mHz and $m_{\Gamma/2}$=0.864 $\pm$ 0.138 μHz.

2.2. Spectrum modelisation

The distribution function of helioseismic spectral density is a χ^2_2 deriving from the gaussian distribution of the imaginary and real parts of the Fourier transform of $v(t_i)$. Consequently we have used two independent random gaussian variables $g(\nu_i)$ and $g'(\nu_i)$, having a zero mean value and an unit variance to simulate a given frequency range of the Fourier transform $h(\nu_i)$:

$$h(\nu_i) = g(\nu_i)\sqrt{\frac{L(\mathbf{a}, \nu_i)}{2}} + ig'(\nu_i)\sqrt{\frac{L(\mathbf{a}, \nu_i)}{2}} \qquad (3)$$

To be realistic this $h(\nu_i)$ must be convolved with the complex window in order to show the daily sidelobes. This last operation introduces additional difficulties:

– it increases the widths of the peaks. Therefore the initial 'width' parameter, which was estimated on the true spectrum, must be modified in order to restitute the true width

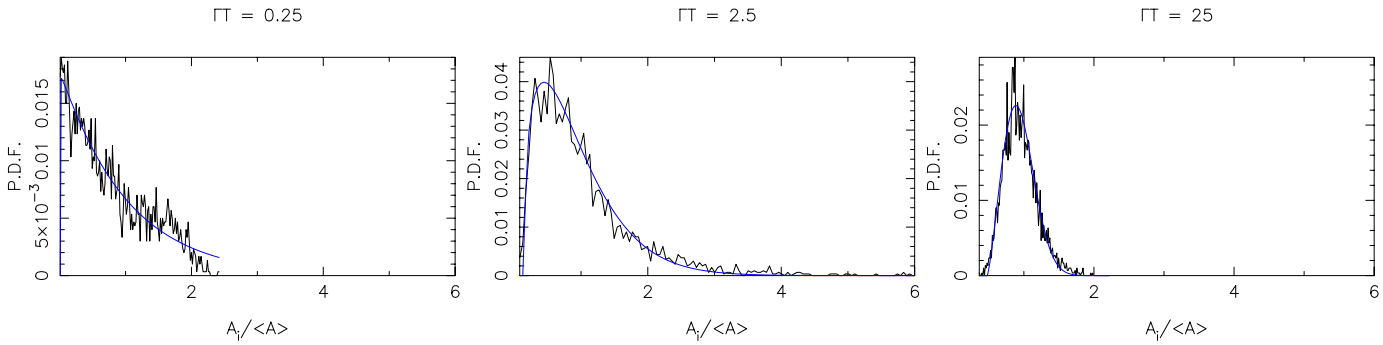

Fig. 3a–c. Variations of the amplitude probability density function as a function of ΓT, in agreement with the stochastic excitation of solar p-modes (Kumar, 1988).

of the peaks after this convolution. We perform two Monte Carlo computations: the first for evaluating the increase of the width and the second to give the error bars computed with the modified initial parameter.

– to take care of the power redistribution of the peaks in the sidelobes we used an amplitude normalized window function
– finally the background noise is added in the form of a complex gaussian random perturbation only after the convolution.

One realisation of the spectral density is obtained by taking the squared modulus of the Fourier transform. Fig. 1 shows the different stages of the spectrum modelling of a $\ell = 1 - 3$ group for the 1990 dataset ($\ell = 1$, $n = 15$ and $\ell = 3$, $n = 14$).

2.3. Error-bars estimation

We compute 400 Monte Carlo perturbed spectral densities in order to ensure the convergence of the mean toward the initial parameters. The error bars are then given by modeling the probability density function (e.g. the histogram) of the values taken from the many realizations of the parameters.

2.3.1. Frequency

Due to the symmetrical Lorentzian modelling of the p-modes, the frequency PDF shows a symmetric distribution around the mean value, which is fitted by a gaussian distribution Eq. (4) (Gelly et al., 1997). Fig. 2a shows the frequency PDF of $n = 19$, $l = 1$ mode for the year 1990.

$$f(\nu) = A \ \exp \frac{(\nu - m_\nu)^2}{2\sigma_\nu^2} \tag{4}$$

The frequency uncertainty is given by σ_ν.

2.3.2. Amplitude and width

The amplitude and width PDF's are not symmetric because of the χ_2^2 power spectrum distribution and are best modeled by :

$$f(x) = K(\frac{x}{c})^\alpha e^{(\left(\frac{x}{c}\right)^{\alpha+1})} \tag{5}$$

K being a normalization constant and c being an adjustable shape parameter. The statistical uncertainties σ_A and σ_Γ are then derived numerically from the modeled PDF.

Fig. 2b shows the width probability density function of $n = 19$, $\ell = 1$ for the year 1990. It is worth noticing that the width and amplitude PDF have the same characteristics as the distribution function for time averaged energies of stochastically excited solar p-modes (Kumar, 1988). This is just the effect of gaussian random perturbation in the MC simulation of the real and imaginary part of the Fourier transform, that corresponds to the hypothesis of the stochastic nature of the excitation. Fig. 3a–c shows three normalized PDF of the amplitudes, computed by MC simulations, as a function of ΓT (where T is the duration of observation and Γ is proportional to the inverse of the lifetime of the mode), and the corresponding modeled distribution given by Eq. (5). For low values of ΓT, the PDF are well modeled by a Boltzmann law (α=0) and by a gaussian law (high values of α) for large values of ΓT.

3. Frequencies error bars results

The theoretical frequencies error-bars are based on Eq. (6) (Toutain and Appourchaux, 1994):

$$\sigma_{th} = \sqrt{\frac{\Gamma}{4\pi T d}} f(\beta) \tag{6}$$

Γ being the linewidth of the mode, T the duration of the observation, d the duty cycle and $f(\beta)$ a function of the noise to signal ratio. This formula only applies to a singlet ($\ell = 0$) mode. For multiplet modes ($\ell > 0$), if the $\ell + 1$ known splitting components visible by the full disk observations were completely uncorrelated, they would, in principle, allow $\ell + 1$ independent determinations of the same central frequency and subsequently decrease the error-bar of Eq. (6) like:

$$\sigma_\nu^{'} = \frac{\sigma_{th}}{\sqrt{\ell + 1}} \tag{7}$$

We took this quantity as reference for the comparisons.

3.1. IRIS data, sidelobes

Figs. 4 and 5 show a comparison between the former theoretical error-bars and the computed ones, computed on the 1990 dataset

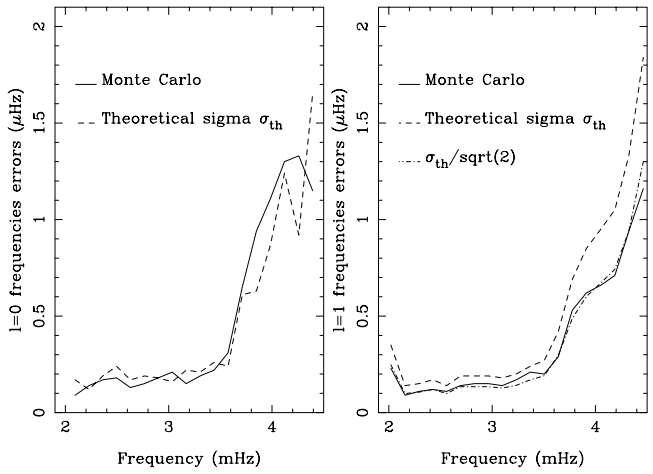

Fig. 4. $\ell = 0$ and $\ell = 1$ frequencies error bars for the year 1990, showing a good agreement between the Monte Carlo results and theory.

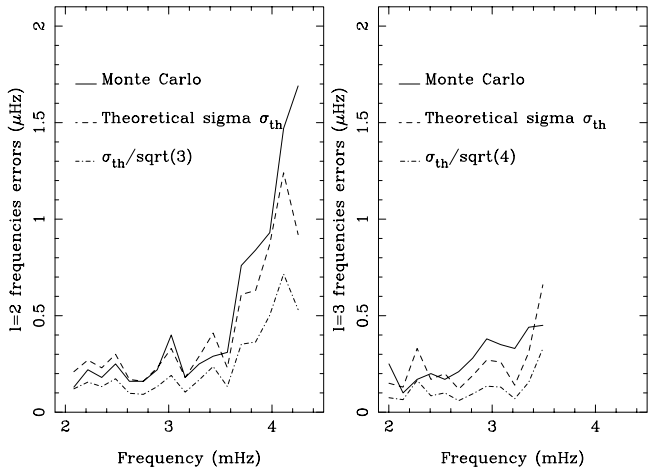

Fig. 5. $\ell = 2$ and $\ell = 3$ frequencies error bars in year 1990. The MC results give higher numbers than the theory, showing the difficulty to observe the $\ell = 2$ triplet and the $\ell = 3$ quadruplet.

of the IRIS data with a bin in the spectral density of 0.095 μHz. In all the cases, all the errors range from $.1$ μHz to 1.5 μHz. All the curves show a frequency dependence, related to the presence of the linewidth in the numerator of Eq. 6. Nevertheless, the error-bar curves are *not* identical to the width curves because they are also weighted by the SNR of the peaks which behaves differently.

The presence of sidelobes of one day in the spectral density biases the measurement of the linewidths above 3.7 mHz, affecting both the computed and theoretical curves in the same way, so that they continue to match, up to 4.2 mHz. The presence of the daily sidelobes, taken into account by the fit, also affects the quality of the result in degrading by a small amount the frequency of a given peak, but in degrading its immediate neighbours at about 11.57 μHz, split or unsplit.

$\ell = 0$ and $\ell = 2$ degrade themselves reciprocally, although $\ell = 2$ is more affected, being of poorer SNR, and $\ell = 1$ degrades $\ell = 3$. Subsequently, we expect as a general trend to have smaller error bars on $\ell = 0$ and $\ell = 1$ than on the two others, and this is actually the case.

As to the hypothesis of the gain in precision afforded by the multiplets, there is a difference between Fig. 4 and 5. For $\ell = 1$, the relation seems verified and the MC numbers are closer to the σ'_ν of Eq. 7 than to σ_{th}. It is not the case for $\ell = 2$ or 3 where the MC values are closer to σ_{th}. We think that the precision in the central frequency of a multiplet can benefit from independent determinations on each component only when those components are unambiguously identified by the fit for most of our 400 statistical realisations. The situation is then easier for $\ell = 1$ because of its high relative amplitude and the cleanliness of its surroundings, than for $l = 2$, affected by the left sidelobe of $l = 0$. Also we shall see that the split components do interfere each other in the complex space, leading to interferences in the spectral density that can wipe out completely one or several components of the multiplet. The higher the number of components, the higher the probability of interferences leading to a misidentification of the individual components by the fit.

3.2. GONG data, no sidelobes

From Figs. 4 and 5, it is not clear which effect plays the major role, although we suspect that the sidelobe pollution has a stronger effect than the 'blurring' of the splitting.

For this reason, we also performed MC simulations using 1 month of GONG data with a 98% duty cycle and a better SNR than the IRIS data (but a lower resolution). Fig. 6a–c shows the difference between MC and theoretical uncertainties for singlet, doublet and triplet modes. In this case, free from sidelobes, we can really size the effect of the splitting on the error bars: we move from an almost perfect match for $\ell = 0$ to a significant discrepancy for the doublet and triplet modes. We conclude from this that no gain in precision is to be expected from the split nature of a peak and that error-bars computed from an $\ell = 0$ formula are a good guideline for all the modes. Fig. 7 shows one example of two possibles realizations of a doublet generated with the same set of initial parameters, and one can see that a maximum likelihood fit will give two incompatible values of splitting, and that the frequency determination will be greatly affected.

4. P-modes widths and widths error bars

4.1. Strategy

We studied the p-modes linewidths (FWMH) for the 4 years of IRIS data from 1898 to 1992, with duty ranging from 40 to 57 %. We used the same technique as in the frequency determination (Gelly et al., 1997). However, we forced the central frequency of any given peak to our tabulated value, and the splitting value to $451nHz$ (Lazrek et al., 1996). The widths were free parameters, and were different within even$-\ell$ and odd$-\ell$ groups, even though they were determined together. The linewidth is related to the lifetime of the mode by $\tau = 1/4\pi\Gamma$. When the linewidth becomes of the same magnitude as the distance to the daily

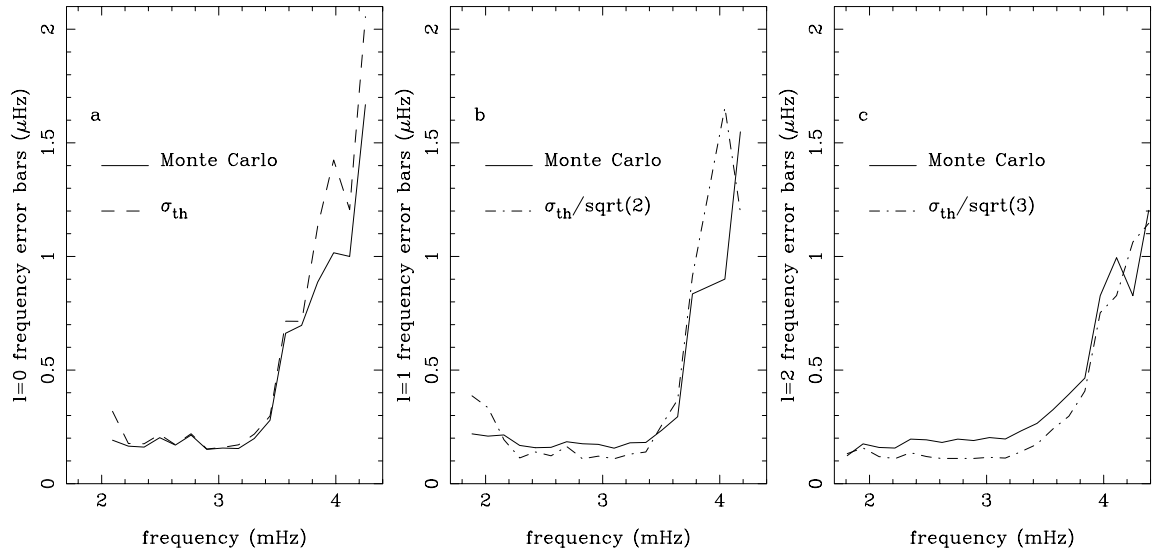

Fig. 6a–c. Frequency error bars on GONG 1-month data for a singlet (**a**), a doublet (**b**) and a triplet (**c**). The frequency bin of the data is 0.34 μHz.

Table 1. IRIS p-mode widths for 4 years of dataset (1989-92)

Order	l=0	l=1	l=2
12			0.26 ± 0.10
13	0.44 ± 0.19	0.39 ± 0.13	0.69 ± 0.31
14	0.69 ± 0.20	0.68 ± 0.26	0.65 ± 0.27
15	0.50 ± 0.18	0.73 ± 0.16	0.91 ± 0.29
16	0.91 ± 0.23	0.73 ± 0.18	1.35 ± 0.28
17	1.16 ± 0.24	0.84 ± 0.16	1.28 ± 0.27
18	1.09 ± 0.20	1.36 ± 0.19	0.85 ± 0.21
19	1.30 ± 0.21	1.04 ± 0.16	1.05 ± 0.24
20	1.11 ± 0.24	1.20 ± 0.16	1.40 ± 0.24
21	1.44 ± 0.29	1.24 ± 0.17	1.56 ± 0.25
22	1.64 ± 0.24	1.47 ± 0.17	2.30 ± 0.39
23	1.72 ± 0.23	2.85 ± 0.30	2.46 ± 0.37
24	2.27 ± 0.35	3.35 ± 0.55	4.06 ± 0.49
25	3.47 ± 0.49	5.32 ± 0.92	5.78 ± 0.57
26	4.97 ± 1.35	6.79 ± 0.87	8.03 ± 1.30

sidelobes of the peak, our measurement is biased and the former relation no longer applies. This happens above $n = 26$.

As mentionned in 2, the linewidth in the spectral density is increased because of the convolution by the window function. The given values are corrected to first order, i.e. by substracting the amount by which a given linewidth is increased after the convolution, although it has been checked that this correction is well inside the error bars. Moreover, those small corrections do not exhibit a clear relationship with the linewidth in the frequency range that we considered, justifying this linear correction.

We have checked for linewidths variation among the years 1989 to 1992, and we did not find any, at our current level of precision. No dependency with the duty-cycle has been noticed. We present in Table 1 the mean widths for l=0, 1 and 2 for those 4 years.

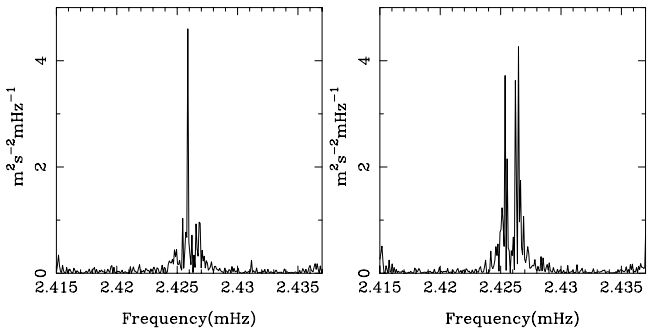

Fig. 7. Splitting 'blurring'. Two Monte Carlo simulations of (l=1, n=16) for the year 1990 showing two differents realizations of the doublet.

Fig. 8 shows the IRIS results and a comparison with the IPHIR (Toutain et al. , 1992) and LOI measurements (Appourchaux et al., 1997). Our result also shows a flat level of width, which was first described by Libbrecht (1988), showing that the mode damping may be due to another processes than turbulent or radiative viscosity, but we do not observe such a dip as presented for LOI results, which is interpreted as a resonance effect with convection (Frohlich et al., 1997). There is only a hint of such a feature for the $\ell = 2$ mode.

4.2. Width error bars

The theoretical sigma (Toutain and Appourchaux, 1994) is given by Eq. (8) and corresponds to a fit with an unconstrained frequency on one Lorentz profile.

$$\sigma_\Gamma = \sqrt{\frac{1}{\Gamma \pi T d}} g(\beta) \tag{8}$$

The difference of free parameters in the maximum likelihood fitting procedure modifies the uncertainties values. An imposed frequency fit decreases the error bars of the width (Fig. 9) and of the amplitude, and we believe that this is a real im-

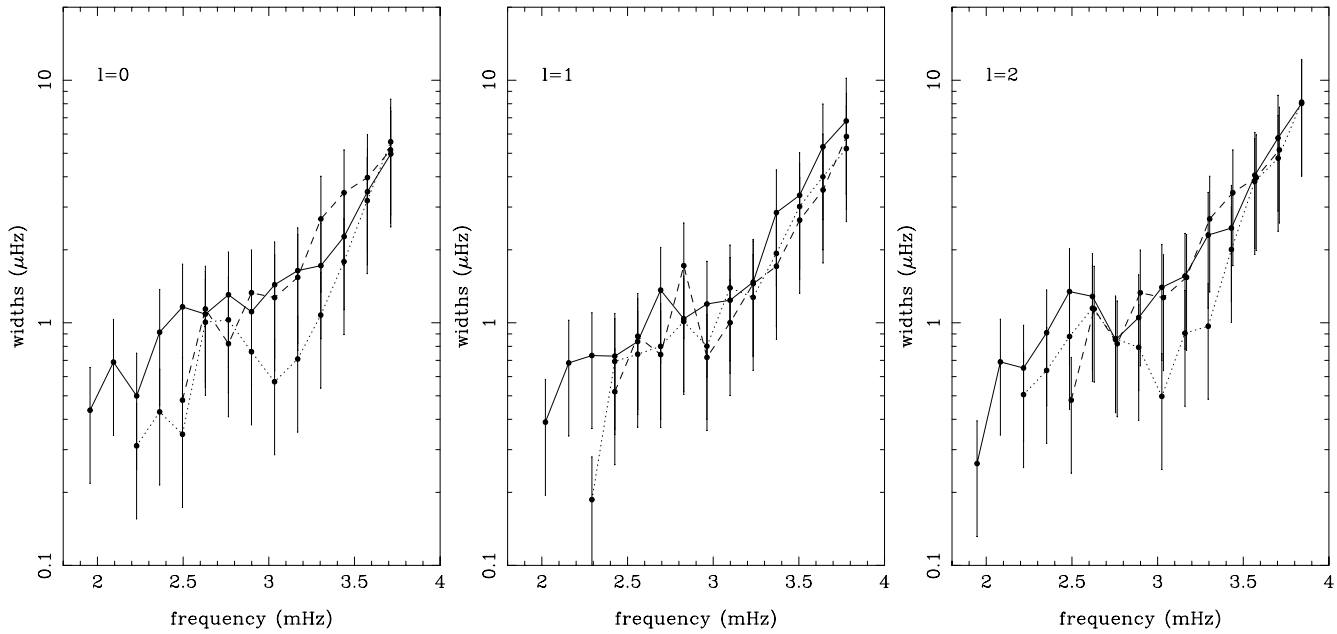

Fig. 8. IRIS width (solid lines) for 1989-92 and a comparison with IPHIR (dashed lines) and LOI (dotted lines).

provement in the FWHM determination of complicated cases involving a multiplet mode *and* a sidelobe.

5. Conclusion

Our power spectrum model is based on normal noise added to some ideal function in the Fourier space, leading to a solar-like p-mode power spectrum. This process is intended to picture a stochastically driven oscillating system. We think that our simulation is realistic enough, and we have been able through it to verify numerically the corresponding theoretical predictions from Kumar. To use Monte Carlo simulation for the computation of error bars brings a few evidences:

- The presence of daily sidelobes of the eigenmode peaks combined with a bad SNR and a blurred multiplet shape do decrease by a factor 1.5 to 2 the precision of the frequency measurements for ℓ =2 and 3. $\ell = 0$ and $\ell = 1$ are better because of the smaller impact of the closest sidelobes and of the better SNR. The theoretical formula valid for $\ell = 0$ seems a good starting point in any circumstances.
- Destructive interference between the real and imaginary parts of the noise in the Fourier transform are important for the p-modes parameters uncertainties. So are the interferences between different peaks with comparable values of frequencies, such as sidelobes of the $\ell = 2$ and $\ell = 0$ peak or between split components. On a single realization they can generate a dramatic variation of the spectral density shape, leading to absurd values for the p-modes parameters and/or error bars. Monte Carlo simulations cannot correct the initially biased estimation, but, as they do not rely on the values of the parameters to derive the error-bars, they will provide error bars that potentially allow the correct result to fall in the $\pm 1\sigma$ interval.

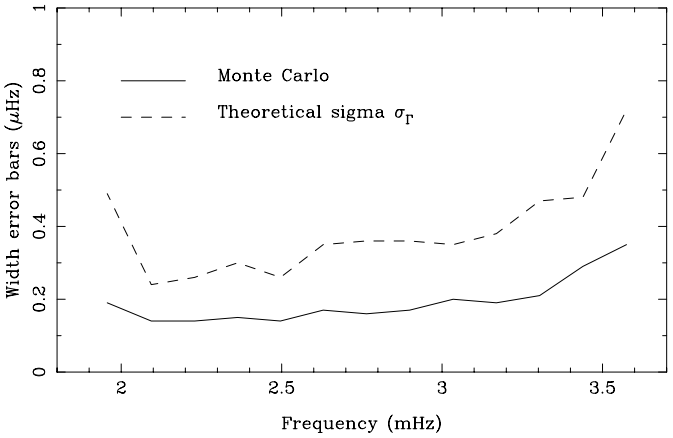

Fig. 9. Comparison between MC error bars computed for an imposed frequency, and a theoritical σ_Γ with no such constraint.

As to the widths, the strategy of reducing the number of free parameters helps very much in reducing the error-bars by a factor of 2 in the best case. This is encouragingly in favor of the "iterative" fitting methods already designed for the collective splitting measurements (Toutain et al. 1992, Lazrek et al. 1996), in which some parameters of the vector $\mathbf{a}$ are in turn locked, then released, then locked again down to the entire convergence of all the set.

The IRIS network was funded and is supported by the French Institut National des Sciences de l'Univers (I.N.S.U.) and the Centre National de la Recherche Scientifique (C.N.R.S.).

Acknowledgements. Data from the IRIS network depends on the coordinated efforts of many peoples from several nations. The authors wish to thank those who have conceived the instrument: E. Fossat and G. Grec; those who have contributed to build and maintain all instruments on site: B. Gelly, J.F. Manigault, G. Rouget, J. Demarcq, G. Galou, A.

Escobar, J.M. Robillot; those who have operated the observing sites: M. Baijumamov, S. Ehgamberdiev, S. Ilyasov, S. Khalikov, I. Khamitov, G. Menshikov, S. Raubaev, T. Hoeksema, Z. Benkhaldoun, M. Lazrek, S. Kadiri, H. Touma, M. Anguera, A. Jimenez, P.L. Palle, A. Pimienta, C. Regulo, T. Roca Cortes, L. Sanchez, F.X. Schmider; R. Luckhurst, those who have developed the analysis software: S. Ehgamberdiev, S. Khalikov, E. Fossat, B. Gelly, M. Lazrek, P.L. Palle, L. Sanchez, E. Gavryuseva, V. Gavryusev; and those who have contributed to the success of the IRIS project in other critical ways: P. Delache, D. Gough, I. Roxburgh, F. Hill, T. Roca-Cortes, G. Zatsepin, T. Yuldashbaev, D. Ben Sari, L. Woltjer, H. Van der Laan, D. Hofstadt, J. Kennewell, D. Cole, P. Scherrer, F. Sanchez., and the Birmingham University Solar Network for sharing the observational facility in Tenerife.

References

Anderson E.R., Duvall T.L. Jr, Jefferies S.M., 1990, ApJ 364, 699-705.
Appourchaux T., Andersen B. N.., Fröhlich C. et al., 1997, Solar Physics 170, 27.
Appourchaux T., Toutain T., Telljohann U., Jimenez A. et al., 1995, A&A 294, L13-L16.
Elsworth, Y., Howe, R., Isaak, G. R. et al., 1994, ApJ 434, 801.
Frohlich C., Andersen B.N., Appourchaux T. et al., 1997, IAU Symposium 181, Schmider Eds, in the press.
Gelly B., Fierry Fraillon D., Fossat E., Palle P. et al., 1997, A&A 323, 235-242.
Kumar P., Franklin J., and Goldreich P., 1988, ApJ 328, 879-887.
Lazrek M., Pantel A., Fossat E. et al., 1996, Solar Physics,
Libbrecht K.G., 1988, ApJ 334, 510-516.
Libbrecht K.G., 1992, ApJ 387, 712-714.
Pantel A., 1996, Ph. D. Thesis, Nice University.
Toutain T. and Appourchaux T., 1994, A&A 289, 649-658.
Toutain T. and Frohlich C., 1992, A&A 257, 287-297.
Woodard M., 1984, Ph. D. Thesis, San Diego University.