

Editorial Manager(tm) for Solar Physics
Manuscript Draft

Manuscript Number:

Title: Fourier analysis of gapped time series: Improved estimates of solar and stellar oscillation parameters

Article Type: Topical Issue: HELAS

Keywords: "Helioseismology, Observations"; "Oscillations, Solar"; "Oscillations, Stellar"

Corresponding Author: Laurent Gizon,

Corresponding Author's Institution:

First Author: Thorsten Stahn

Order of Authors: Thorsten Stahn; Laurent Gizon

Abstract: Quantitative helio- and asteroseismology require very precise measurements of the frequencies, amplitudes, and lifetimes of the global modes of stellar oscillation. It is common knowledge that the precision of these measurements depends on the total length (\$T\$), quality, and completeness of the observations. Except in a few simple cases, the effect of gaps in the data on measurement precision is poorly understood, in particular in Fourier space where the convolution of the observable with the observation window introduces frequency correlations. Here we describe and implement a rather general method to retrieve maximum likelihood estimates of the oscillation parameters, taking into account the proper statistics of the observations. Our fitting method applies in complex Fourier space and exploits the phase information. We consider both solar-like stochastic oscillations and long-lived harmonic oscillations, plus random noise. Using numerical simulations, we demonstrate the existence of cases for which our improved fitting method is less biased and has a greater precision than when the frequency correlations are ignored. This is especially true of low signal-to-noise solar-like oscillations. For example, we discuss a case where the precision on the mode frequency estimate is increased by a factor of five, for a duty cycle of 15%. In the case of long-lived sinusoidal oscillations, a proper treatment of the frequency correlations does not provide any significant improvement; nevertheless we show that the mode frequency can be measured at a much better precision than the $1/T$ Rayleigh resolution.

one of the best studied objects in this class is the pre-white dwarf PG1159–035 or GW Vir (Winget *et al.*, 1991). In practice, observations of solar-like or deterministic pulsations always have an additional stochastic component due to instrumental, atmospheric, stellar, or photon noise.

An important step in helio- and asteroseismology is the determination of the parameters of the global modes of oscillation, especially the mode frequencies. In the case of the Sun, it is known (Woodard, 1984) that the measurement precision is limited by the stochastic nature of the oscillations (realization noise). Libbrecht (1992) and Toutain and Appourchaux (1994) have shown that realization noise is expected to scale like $1/\sqrt{T}$, where T is the total length of the observations (no gaps). A common practice is to extract the solar mode parameters from the power spectrum using maximum likelihood estimation (MLE, see e.g. Anderson, Duvall, and Jefferies, 1990; Schou, 1992; Toutain and Appourchaux, 1994; Appourchaux, Gizon, and Rabello-Soares, 1998; Appourchaux et al.2000). In its current form, however, this method of analysis is only valid for uninterrupted time-series. This is a significant limitation because gaps in the data are not uncommon (daily cycle, bad weather, technical problems). The gaps complicate the analysis in Fourier space: the convolution of the data with the observation window leads to correlations between the different Fourier components. The goal of this paper is to extend the Fourier analysis of solar and stellar oscillations to time-series with gaps, using appropriate maximum likelihood estimators based on the correct statistics of the data.

Section 2 poses the problem of the analysis of gapped time series in Fourier space. In Section 3 we derive an expression for the joint probability density function (PDF) of the observations, taking into account the frequency correlations. Our answer is consistent with an earlier (independent) derivation by Gabriel (1994). Based on this PDF, we derive maximum likelihood estimators in Section 4. In Section 5 we recall the “old method” of maximum likelihood estimation based on the unjustified assumption that frequency bins are statistically independent. Section 6 explains the set up of the Monte-Carlo simulations, used to test the fitting methods on artificial data sets. In Section 7 we present the results of the Monte-Carlo simulations and compare the new and old fitting methods. For the sake of simplicity, we consider only one mode of oscillation at a time (solar-like or sinusoidal). We exhibit several cases for which our new fitting method leads to a significant improvement in the determination of oscillation parameters, and in particular the mode frequency. As a side result, we also show that the frequency of a sinusoidal function can be inferred in Fourier space with a precision that is much better than $1/T$, as had already been shown in the time domain by e.g. Cuypers (1987).

2. Statement of the problem

2.1. The observed signal in Fourier space

Let us denote by $\tilde{y} = \{\tilde{y}_i\}$ the time series that we wish to analyse. It is sampled at times $t_i = i\Delta t$, where i is an integer in the range $0 \leq i \leq N - 1$, and $\Delta t = 1$ min is the sampling time. All quantities with a tilde are defined in the time domain. The total duration of the time series is $T = N\Delta t$. By choice, all the missing data points were assigned the value zero: this enables us to work on a regular time grid. Formally, we write

$$\tilde{y}_i = \tilde{w}_i \tilde{x}_i, \quad i = 0, 1, \dots, N - 1. \quad (1)$$

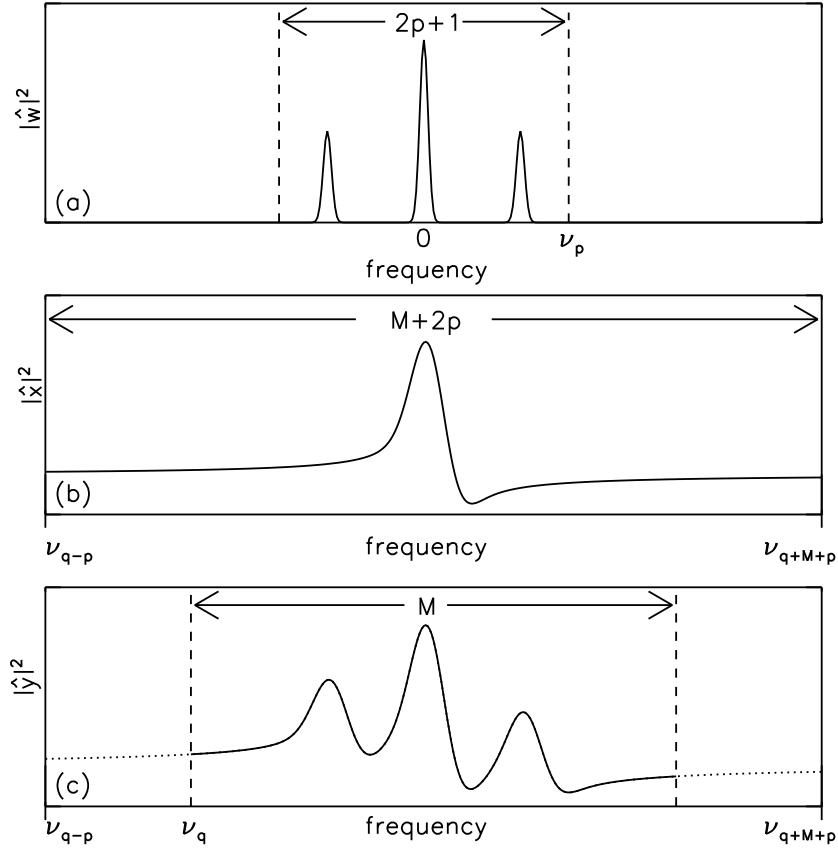


Figure 1. Schematic representation in Fourier space of the convolution of the signal \hat{x} with the window function \hat{w} . For the sake of simplicity, only the power spectra of the different quantities are shown here. Panel (a) shows the window function \hat{w} and its cutoff frequency ν_p , panel (b) the unconvolved signal \hat{x} , and panel (c) the observed signal \hat{y} . Note that the selected section of the observed signal, starting at frequency ν_p , is of length M , while the unconvolved signal is of length $M + 2p$.

2.2. Statistics of the unconvolved signal

Here we describe the basic assumptions that we make about the statistics of the data in the Fourier domain. The unconvolved signal, x , consists of a deterministic component, d , and a zero-mean stochastic component, e , such that

$$x = d + e. \quad (8)$$

The deterministic component, d , may include deterministic stellar oscillations that are long-lived compared to the total length of the observation. The stochastic component e may include various sources of noise (e.g. stellar convection, photon noise, atmospheric noise. . .) and stochastically excited pulsations as observed on the Sun.

We assume that the e_i are $M + 2p$ independent random variables in the Fourier domain. This is equivalent to saying that the stochastic component of signal in the

time domain is stationary. We further assume that e is a Gaussian random vector with independent real and imaginary parts and covariance matrix

$$E[e_i^* e_j] = \sigma_i^2 \delta_{ij}, \quad i, j = 0, 1, \dots, M + 2p - 1, \quad (9)$$

where E denotes the expectation value and σ_i is the standard deviation of e_i at frequency ν_i . One may invoke the central limit theorem to justify the choice of Gaussian distributions. In terms of a complex Gaussian random vector g with unit covariance matrix, $E[g^* g^T] = I_{M+2p}$, we can rewrite e as

$$e = Sg, \quad (10)$$

where S is the $(M + 2p) \times (M + 2p)$ diagonal matrix

$$S = \text{diag}(\sigma_0, \sigma_1, \dots, \sigma_{M+2p-1}). \quad (11)$$

We emphasise that, although the e_i are uncorrelated random variables, the y_i are correlated because of the multiplication of x by the window matrix (eq. [5]).

3. Joint PDF of the complex Fourier spectrum

In this section we derive an expression for the joint probability density function of the observed signal y . This problem had already been solved by Gabriel (1994). We reach the same conclusion, independently and with more compact notations. We start by rewriting the master equation (5) as:

$$y = Wd + Cg, \quad (12)$$

where

$$C = WS \quad (13)$$

is a $M \times (M + 2p)$ matrix with rank M and singular value decomposition (Horn and Johnson, 1985, chapter 7.3)

$$C = U\Sigma V^H. \quad (14)$$

Here the superscript H denotes the hermitian conjugate and U and V are unitary matrices of dimensions $M \times M$ and $(M + 2p) \times (M + 2p)$ respectively, i.e. $U^H U = I_M$ and $V^H V = I_{M+2p}$. The $M \times (M + 2p)$ matrix Σ can be written as

$$\Sigma = [\Lambda \mid 0], \quad \Lambda = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{M-1}), \quad (15)$$

where $\lambda_0, \lambda_1, \dots, \lambda_{M-1}$ are the M (positive) singular values of the matrix C . Thus, there exists a vector $\xi = V^H g$ such that

$$y = Wd + U[\Lambda \mid 0]\xi. \quad (16)$$

Since g has unit covariance matrix and V is unitary, the vector ξ is a complex Gaussian random vector of size $M + 2p$ with unit covariance matrix. It is obvious

from equation (16) that there exists a lower-rank complex Gaussian random vector of length M , $\eta = [\xi_0, \xi_1, \dots, \xi_{M-1}]^T$, such that

$$y = Wd + U\Lambda\eta. \quad (17)$$

The variables $\xi_M, \xi_{M+1}, \dots, \xi_{M+2p-1}$ are dummy variables which do not enter in the description of y . Equation (17) is an important step, as the vector y of length M is now expressed in terms of M independent complex Gaussian variables. This enables us to write the PDF of y as

$$p_y(y) = \frac{1}{J} p_\eta((U\Lambda)^{-1}(y - Wd)), \quad (18)$$

where $p_\eta(\eta)$ denotes the PDF of η and J is the Jacobian of the linear transformation $\eta \rightarrow y$. Since η is a complex Gaussian random vector with unit covariance, i.e. $E[\eta^* \eta^T] = I_M$, we have

$$p_\eta(\eta) = \frac{\exp(-\|\eta\|^2)}{\pi^M}, \quad (19)$$

where we used the notation $\|\eta\|^2 = \eta^H \eta$. Since U is unitary and Λ is diagonal and real, the Jacobian of the transformation is given by

$$J = |\det(U\Lambda)|^2 = (\det \Lambda)^2 = \prod_{i=0}^{M-1} \lambda_i^2. \quad (20)$$

Combining equations (18), (19), and (20), we get the joint PDF of the observed vector y :

$$p_y(y) = \frac{\exp(-\|\Lambda^{-1}U^H y - Wd\|^2)}{\pi^M (\det \Lambda)^2} \quad (21)$$

The above expression is, perhaps, more elegantly written as

$$p_y(y) = \frac{\exp(-\|C^\dagger(y - Wd)\|^2)}{\pi^M (\det \Lambda)^2}, \quad (22)$$

in terms of C^\dagger , the $(M + 2p) \times M$ Moore-Penrose generalized inverse of C (Horn and Johnson, 1985, chapter 7.3),

$$C^\dagger = V\Sigma^\dagger U^H = C^H(CC^H)^{-1}, \quad (23)$$

where Σ^\dagger is the transpose of Σ in which the singular values are replaced by their inverse. One may ask, after the fact, if the quantity $(U\Lambda)^{-1}$ in equation (18) is always defined. The answer would appear to be yes since the Moore-Penrose generalised inverse of C is perfectly well defined. It is not excluded, however, that some singular values λ_i could be infinitesimally small. We have not encountered any such difficulty with the test cases given in section 7. Should C be ill-conditioned in other cases, a simple truncated SVD would help avoid a numerical problem.

4. Maximum likelihood estimation of stellar oscillation parameters

Let us assume that the stellar oscillation model that we are trying to fit to the data depends on a set of parameters $\mu = (\mu_0, \mu_1, \dots, \mu_k)$. These parameters may be the amplitude, the phase, the frequency, the line asymmetry, the noise level, etc. The basic idea of maximum likelihood estimation is to pick the estimates μ_* that maximise the likelihood function. The likelihood function is another name for the joint PDF (eq. [22]) evaluated at the sample data. In practice, one minimises

$$\ell(\mu) = -\ln p_y = \|C^\dagger(y - Wd)\|^2 + 2 \sum_{i=0}^{M-1} \ln \lambda_i + Const., \quad (24)$$

rather than maximise the likelihood function itself. In the above expression, the quantities C^\dagger and λ_i all depend implicitly on the model parameters μ through the covariance matrix S . The vector d also depends on the model parameters in the case deterministic oscillations. The probability of observing the sample data is greatest if the unknown parameters are equal to their maximum likelihood estimates μ_*

$$\mu_* = \arg \min_{\mu} \ell(\mu). \quad (25)$$

The method of maximum likelihood has many good properties (Brandt, 1970). In particular, in the limit of a large sample size (M large), the maximum likelihood estimator is unbiased and has minimum variance.

What is particularly new about our work is the implementation and validation of the minimization problem, equation (25). We use the direction set method, or Powell's algorithm, to solve the minimization problem with a computer. In practice, the result of the fit depends on the initial guess and the fractional tolerance of the minimisation procedure (the relative decrease of ℓ in one iteration). The dependence of the fitted parameters on the initial guess is due to the fact that the function ℓ may have local minima in addition to the global minimum. We will address this issue in more detail in section 7.

5. The old way: fitting the power spectrum ignoring the correlations

Maximum likelihood estimation has been used in the past to infer solar and stellar oscillation parameters, even in the case of gapped time series. The joint PDF of the observations was assumed to be the product of the PDFs of the individual y_i , as if the frequency bins were uncorrelated. For comparison purposes, we briefly review this (unjustified) approach.

According to equation (12), the PDF of y_i is a normal distribution

$$p_{y_i}(y_i) = \frac{\exp(-|y_i - \bar{y}_i|^2/v_i)}{\pi v_i} \quad (26)$$

with mean

$$\bar{y}_i = \sum_{j=0}^{M+2p-1} W_{ij} d_j \quad (27)$$

and variance

$$v_i = \sum_{j=0}^{M+2p-1} |W_{ij}|^2 \sigma_j^2. \quad (28)$$

Under the (wrong) assumption that the y_i are independent random variables, the joint PDF of y becomes

$$p_y^{\text{n.c.}}(y) = \prod_{i=0}^{M-1} p_{y_i}(y_i), \quad (29)$$

where the superscript “n.c.” stands for “no correlation”. This joint PDF uses the correct mean (\bar{y}_i) and variance (v_i) of the data, but it ignores all the non-vanishing cross-terms $E[y_i y_j]$. In other words, the spread of power implied by the convolution by the window is taken care of, but not the proper statistics.

Under the same simplifying “no correlation” assumption, the log-likelihood function is

$$\ell^{\text{n.c.}}(\mu) = \sum_{i=0}^{M-1} \frac{|y_i - \bar{y}_i|^2}{v_i} + \sum_{i=0}^{M-1} \ln v_i + \text{Const.}, \quad (30)$$

where the \bar{y}_i and v_i are implicit functions of the model parameters μ .

If the signal has no deterministic component ($d = 0$), then the power spectrum, $P_i(\mu) = |y_i|^2$, has the expectation value $\bar{P}_i = E[P_i] = v_i$. Thus, in the case of purely solar-like oscillations, we recover the standard expression (Toutain and Appourchaux, 1994):

$$\ell^{\text{n.c.}}(\mu) = \sum_{i=0}^{M-1} \left(\frac{P_i}{\bar{P}_i} + \ln \bar{P}_i \right) + \text{Const.} \quad \text{when } d = 0. \quad (31)$$

While the above expression is perfectly valid for uninterrupted data, it is not justified when gaps are present. The parameters $\mu_\star^{\text{n.c.}}$ that minimize $\ell^{\text{n.c.}}(\mu)$ are not optimal, as will be shown later using Monte-Carlo simulations.

6. Simulation of artificial time series

So far we have considered a general signal which includes a deterministic component and a stochastic component. The parametrisation of each component depends on prior knowledge about the physics of the stellar oscillations. Solar-like pulsations are stochastic in nature and no deterministic component is needed in this case. On the other hand, long-lived stellar pulsations are treated as deterministic. Some stars may support both deterministic and stochastic oscillations. In this section, we model the two cases separately.

We want to test the fitting method (eqs. [24] and [25]) by applying it to simulated time series with gaps. For comparison, we also want to apply the old fitting method (section 5) to the same time series. We need to generate many realizations of the same random process in order to test the estimators for bias and precision: this is called Monte-Carlo simulation. In section 6.1 we discuss the generation of the synthetic window functions. We then discuss the parametrisation of the solar-like oscillations

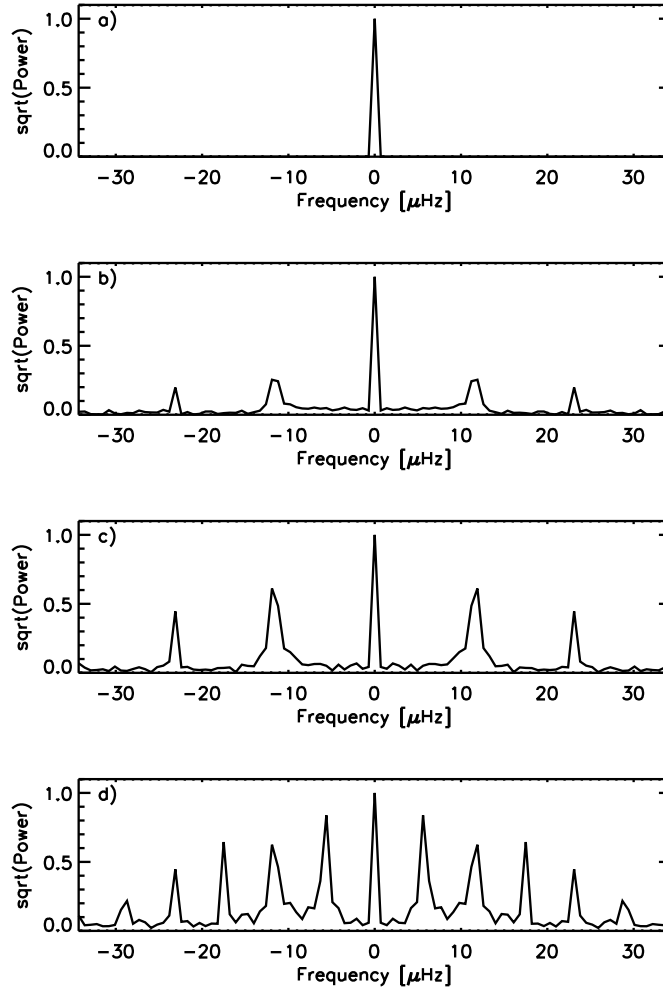


Figure 2. Synthetic window functions \hat{w} (square root of power spectrum). From top to bottom, the duty cycle is (a) 100%, (b) 66%, (c) 30%, and (d) 15%. The main periodicity of the window is 24 hr for cases (a) and (b), and 48 hr for window (d). All windows are truncated at frequency $\nu_p = 34.3 \mu\text{Hz}$.

(section 6.2) and the deterministic oscillations (section 6.3) used to simulate the unconvolved signal.

6.1. Synthetic window functions

We generate three different observation windows, corresponding to different duty cycles. The observation windows are first constructed in the time domain. By definition, \tilde{w}_i is set to one if an observation is available and zero otherwise. The total length of all time series is fixed at $T = 16.5$ days (frequency resolution $\Delta\nu = 0.7 \mu\text{Hz}$). A

window function is characterized by two main properties: the duty cycle (fraction of ones) and the average periodicity. A typical window function for a single ground-based site has a 24 hr periodicity. In order to deviate slightly from purely periodic window functions we introduce some randomness for the end time of each observation block.

Figure 2b-d shows the power spectra of the three window functions. The first and second window functions have a main periodicity of 24 hr and duty cycles of 66% and 30% respectively. Two side lobes occur at frequencies 11.6 μHz and 23.1 μHz . The non-vanishing power between the side lobes is due to the deviation from purely periodic window. The third window function has a main periodicity of 48 hr and a duty cycle of only 15%. All these window functions are not unrealistic.

We apply a sharp low pass filter at frequency $\nu_p = 34.3 \mu\text{Hz}$ ($p = 49$) to all window functions. The power at higher frequencies corresponds to about 5% of the total power in the windows. This truncation is needed to apply the fitting algorithm, which assumes that there exists a frequency ν_p beyond which the power in the window vanishes.

6.2. Modeling solar-like oscillations

We generate the realizations of the unconvolved solar-like oscillation signal directly in the Fourier domain. We consider a purely stochastic signal ($d = 0$) and a single mode of oscillation. Since we assumed stationarity in the time domain, the Fourier spectrum of the unconvolved signal for one single mode can be written as

$$x_i = e_i = [SL(\nu_i) + N]^{1/2} \eta_i, \quad i = 0, 1, \dots, M + 2p - 1, \quad (32)$$

where L describes the line profile of the mode in the power spectrum, S is the mode's maximum power, N is the variance of the background noise, and η_i is a centered complex Gaussian random variable with unit variance and independent real and imaginary parts. Solar-like oscillations are stochastically excited and intrinsically damped by turbulent convection (Stein *et al.*, 2004). The expectation value of the power spectrum is nearly Lorentzian, except for some line asymmetry (e.g. Duvall *et al.* 1993). Here we use a simple asymmetric line profile:

$$L(\nu) = \frac{(1 + bX)^2 + b^2}{1 + X^2} \quad \text{with} \quad X = \frac{\nu - \nu_0}{\Gamma/2}, \quad (33)$$

where ν_0 is the resonant frequency, b is the asymmetry parameter of the line profile ($|b| \ll 1$), and Γ is a measure of the width of the line profile. Since $L(\nu_0) = 1 + b^2 \simeq 1$, we refer to S/N as the signal-to-noise ratio in the power spectrum. As b tends to zero, Γ becomes the full width at half maximum (FWHM) of the power spectrum and $1/(\pi\Gamma)$ the mode lifetime. There are five model parameters, $\mu = (\nu_0, \Gamma, b, S, N)$.

Once the unconvolved signal x has been generated in the Fourier domain, the observed signal y is obtained by multiplication by the window matrix W , as explained above.

6.3. Modeling deterministic sinusoidal oscillations

In the time domain, we consider a purely sinusoidal function on top of white background noise:

$$\tilde{x}_i = A \sin(2\pi\nu_0 t_i + \varphi) + \sigma_t \eta_i \quad i = 0, 1, \dots, N - 1. \quad (34)$$

The first term describes the deterministic component of the signal, where A is the amplitude, ν_0 the frequency, and φ the phase of the mode. The second term is stochastic noise with standard deviation σ_t . The η_i are N normally distributed independent real random variables with zero mean and unit variance. The observed signal is obtained by multiplying \tilde{x}_i by the window \tilde{w}_i in the time domain. The model parameters are $\mu = (\nu_0, \varphi, A, \sigma_t)$.

We have defined the signal and the noise in the time domain, but a definition of signal-to-noise ratio in the Fourier domain is desirable. On the one hand, the variance of the noise in the Fourier domain is

$$\sigma_\nu^2 = \frac{\sigma_t^2}{N} \sum_{i=-p}^p |\hat{w}_i|^2, \quad (35)$$

where $\sum_i |\hat{w}_i|^2$ is the total power in the window. On the other hand, the maximum power of the signal in Fourier space is $P_{\max} = A^2 |\hat{w}_0|^2 / 4$, where $|\hat{w}_0|^2$ is the power of the window at zero frequency. Thus, by analogy with the solar-like case, it makes sense to define the signal to noise ratio in the Fourier domain as

$$S/N = \left(\frac{N |\hat{w}_0|^2}{4 \sum_i |\hat{w}_i|^2} \right) \frac{A^2}{\sigma_t^2}. \quad (36)$$

In practice we fix A and S/N and deduce the corresponding noise level σ_t .

7. Testing the fitting methods

Several hundreds of realizations are needed in order to assess the quality of a fitting method. We do not intend to test all possible combinations of mode parameters but we want to show a few cases for which the new fitting method provides a significant improvement compared to the old fitting method.

7.1. Solar-like oscillations: Window function with 30% duty cycle

Figure 3 shows one realization of a simulated mode of solar-like oscillation with input parameters $\nu_0 = 3000 \mu\text{Hz}$, $\Gamma = 3.2 \mu\text{Hz}$, $S = 0.9$, $N = 0.15$, and $b = 0.1$. The signal to noise ratio is $S/N = 6$ and the window function is 30% full (see Fig. 2c). The mode lifetime is $1/(\pi\Gamma) = 27.6$ hr. Figure 3a displays the real and imaginary parts of the Fourier transform, y , together with the standard deviation of the data (n.c. fit in blue, new fit in red, expectation value in green). Figure 3b shows the power spectrum and the fits. Notice the sidelobes introduced by the convolution of the signal with the window functions. The “no-correlation” fit is done on the power spectrum (eq. [31]), while the new fit is performed in complex Fourier space (eq. [24]).

Each fit shown in Figure 3 corresponds in fact to the best fit out of 5 fits with different initial guesses. For each realization, we use the frequency guesses $3000 + (0, \pm 5.5, \pm 11.9) \mu\text{Hz}$ for ν_0 . The last two frequency guesses correspond to the frequencies of the two main sidelobes of the window function (Fig. 2c). For the other parameters, we choose random guesses within $\pm 20\%$ of the input values. The reason for using several guesses is to ensure that the fit converges to the global maximum of the likelihood not to a nearby local maximum, i.e. that the estimates

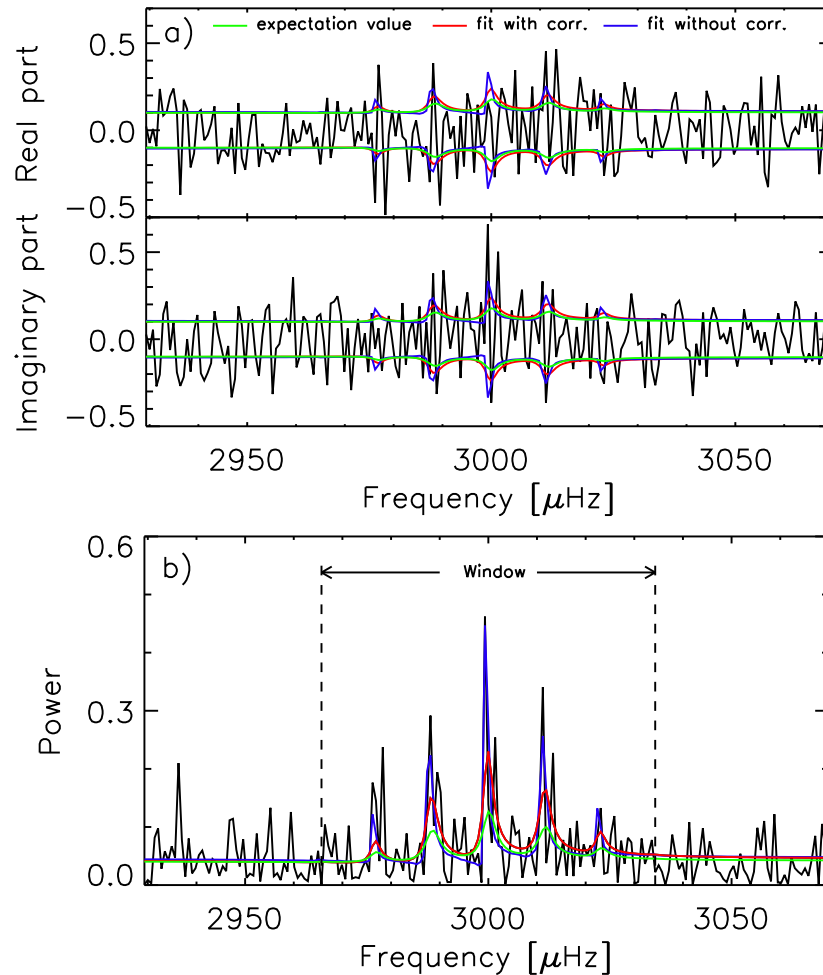


Figure 3. Example of a realization of one mode of a solar-like oscillation with input frequency $\nu_0 = 3000 \mu\text{Hz}$, linewidth $\Gamma = 3.2 \mu\text{Hz}$, and $S/N = 6$. The window function is 30% full. Panel (a) shows the real and imaginary parts of the Fourier spectrum. Panel (b) shows the power spectrum. The vertical dashed lines represent the width of the window function. Also shown are the new fit (red), the old fit (blue), and the expectation value (green).

returned by the code are the MLE estimates defined by equation (25). In some cases, the global maximum coincides with a sidelobe at $\pm 11.9 \mu\text{Hz}$ from the main peak. We note that the new fitting method requires a much longer computing time than the old n.c. method: typically, 3 hr on a single CPU core for a single realization (5 guesses, 5 fits).

For the particular realization of Figure 3, the new fit is closer to the expectation value (i.e. is closer to the answer) than the old n.c. fit. No conclusion should be drawn, however, from looking at a single realization.

In order to test the reliability of each fitting method, we computed a total of 750 realizations with the same input parameters as in Figure 3 and the same window function (30% full). The quality (bias and precision) of the estimators can be studied from the distributions of the inferred parameters. As shown by the distributions of Figure 4 the new fitting method is superior to the old n.c. method. This is true for all the parameters, in particular the mode frequency, ν_0 . The distributions for the mode frequency (Fig. 4a) are quite symmetric and Gaussian-like, although the old fitting method leads to a significant excess of values beyond the 2-sigma mark. We note that, in general, the old fitting method is more sensitive to the initial frequency guess. Also the estimates of the linewidth Γ and the mode power S are significantly more biased with the old fitting method than with the new one (Figs. 4b, 4c). It is worth noting that the fits return a number of small Γ /large S estimates away from the main peaks of the distributions, less so for the new fits. These values correspond to instances when the signal barely comes out of the noise background. The new fit returns the noise level, N , with a higher precision and a lesser number of underestimated outliers than the old method (the outliers are represented by the vertical bars in Fig. 4d). Although the estimation of the asymmetry parameter is unbiased with the new fitting method (Fig. 4e), the uncertainty on b is so large that it probably could have been ignored in the model.

Quantitative estimates of the mean and the dispersion of the estimators are provided in Table 1. Because the distributions of the estimated parameters are not always Gaussian and may contain several outliers, we compute the median (instead of the mean) and the lower and upper bounds corresponding to $\pm 34\%$ of the points on each side of the median (instead of the $1\text{-}\sigma$ dispersion). This definition has the advantage of being robust with respect to the outliers. The notation $3000.0_{-a}^{+b} \mu\text{Hz}$ in the first row of Table 1 means that the median mode frequency is $3000.0 \mu\text{Hz}$ and that 68% of the fits belong to the interval $[3000.0 - a, 3000.0 + b] \mu\text{Hz}$. We insist that the subscript $-a$ and the exponent $+b$ do not refer to an uncertainty in the determination of the median: the median is known to a much higher precision thanks to the large number of realizations. Later we abuse the language and refer to the “1-sigma uncertainty” to mean the average $\sigma = (a + b)/2$.

The numbers from the last two columns in Table 1 confirm the analysis of Figure 4. The mode frequency can be measured with a precision of $1.4 \mu\text{Hz}$, which is exactly twice better with the new fitting method than with the old one. This gain in precision is very significant and potentially important. Since measurement uncertainty scales like $T^{-1/2}$ (Libbrecht, 1992), one may equate the gain in using the proper fitting procedure to an effective increase in the total length of the time series by a factor of four. As seen in Table 1, the linewidth, the mode power, the background noise, and the line asymmetry parameter are all less biased and more precise with the new fitting method than the old one.

7.2. Solar-like oscillations: Different window functions

Here we study how bias and precision change as the window function changes, in particular as the duty cycle changes. We consider the four window functions defined in section 6.1 with duty cycles, α , equal to 15%, 30%, 66%, and 100%. First we consider input parameters of solar-like oscillations that are exactly the same as in the previous section: $\nu_0 = 3000 \mu\text{Hz}$, $\Gamma = 3.2 \mu\text{Hz}$, $S = 0.9$, $S/N = 6$, and $b = 0.1$. Figure 5 shows the distributions of the inferred mode frequencies and linewidths,

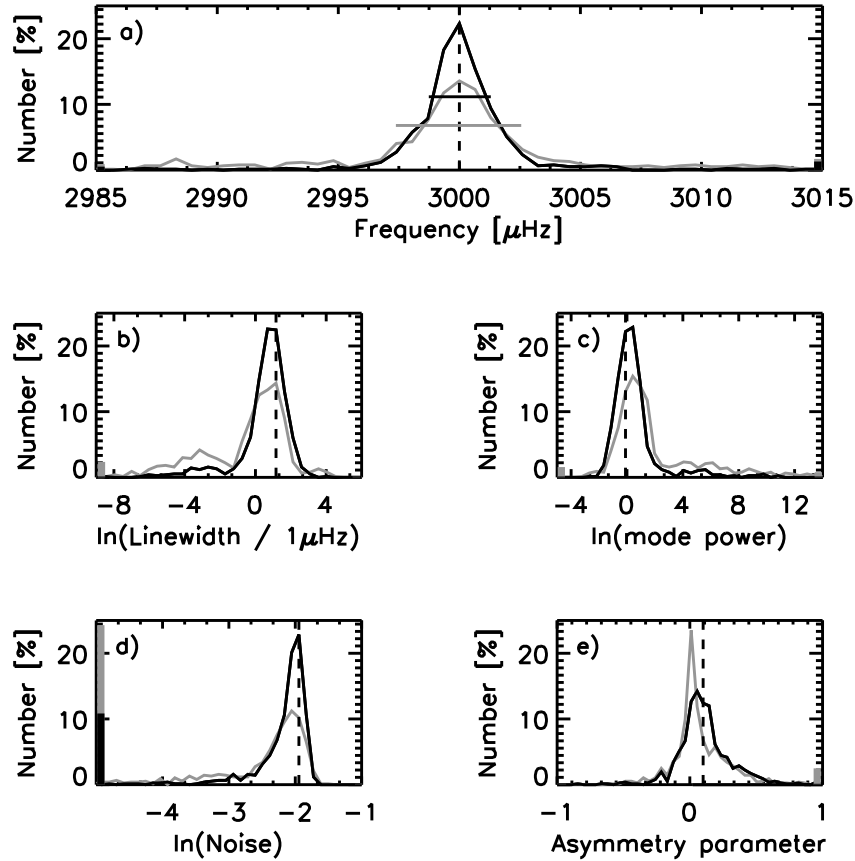


Figure 4. Distributions of the inferred oscillation parameters from fits to 750 realizations of a single mode of solar-like oscillation. The input parameters are given in Table 1 and the window function is 30% full. The five panels show the distributions of the inferred (a) mode frequency ν_0 , (b) linewidth Γ , (c) mode power S , (d) noise level N , and (e) asymmetry parameter b . The black lines show the results obtained with the new fitting method and the grey lines show the old “no-correlation” fits. The vertical dashed line in each plot indicates the input value. The horizontal lines in panel (a) are intervals containing 68% of the fits for the new (black line) and the old (grey line) fitting methods. The thick black and grey vertical lines in panel (d) give the numbers of outliers with $\ln N < -5$.

using the old (Figs. 5a, 5b) and the new (Figs. 5c, 5d) fitting methods. Each fit is the best fit from five different ν_0 guesses (see section 7.1). The distributions for the 100%-window are identical for the two fitting methods; this is expected since the old and new fitting methods are equivalent in the absence of gaps.

The precision on ν_0 using the old “no-correlation” MLE drops fast as the duty cycle decreases (Fig. 5a). This drop is much faster than in the case of the fits that take the frequency correlations into account (Fig. 5b). When the duty cycle is 15%, the frequency estimate is 5 times better with the new than the old method. The difference is perhaps even more obvious for the linewidth. For the 15% window, it

Table 1. Medians and scatters of the distributions of the estimated parameters of solar-like oscillation (see Fig. 4). The window function is 30% full, the input linewidth is $3.2 \mu\text{Hz}$, and the input signal-to-noise ratio is $S/N = 6$. The new and old MLE estimates are given in the last two columns. By definition, 68% of the fits fall within the bounds set by the subscripts/exponents (the notation is explained in detail in the text).

Mode parameter	Input value	New fitting	Old fitting
ν_0 [μHz]	3000.0	$3000.0^{+1.4}_{-1.4}$	$3000.0^{+2.8}_{-2.8}$
$\ln(\Gamma[\mu\text{Hz}])$	1.2	$0.8^{+0.8}_{-1.0}$	$0.2^{+1.1}_{-3.7}$
$\ln(S)$	-0.1	$0.2^{+0.9}_{-0.9}$	$0.9^{+4.3}_{-1.2}$
$\ln(N)$	-1.9	$-2.1^{+0.2}_{-0.9}$	$-2.4^{+0.4}_{-6.8}$
b	0.1	$0.1^{+0.2}_{-0.1}$	$0.0^{+0.2}_{-0.1}$

Table 2. Medians and scatters of the mode frequency estimates (solar-like oscillations) for the window functions defined in Section 6.1. The input mode frequency is $\nu_0 = 3000 \mu\text{Hz}$, the input linewidth is $\Gamma = 3.2 \mu\text{Hz}$, and the signal-to-noise ratio is fixed at $S/N = 6$. The mode lifetime is 27.6 hr.

Duty cycle	Window function		Frequency estimate [μHz]	
	Main period	Average gap	New fitting	Old fitting
100%	–	–	$3000.0^{+1.1}_{-1.2}$	$3000.0^{+1.1}_{-1.2}$
66%	24 hr	7.4 hr	$3000.0^{+1.1}_{-1.3}$	$3000.1^{+1.5}_{-1.4}$
30%	24 hr	16.4 hr	$3000.0^{+1.4}_{-1.4}$	$3000.0^{+2.8}_{-2.8}$
15%	48 hr	40.7 hr	$3000.0^{+1.7}_{-1.3}$	$3000.0^{+8.3}_{-6.5}$

is almost impossible to retrieve Γ with the old fitting method (Figure 5b), while the new method gives estimates that are almost as precise as in the no-gap case (Fig. 5d). The estimates of Γ are significantly less biased with the new method. Figure 5 confirms the importance of using the correct expression for the likelihood function.

Table 2 gives the medians and half-widths of the ν_0 distributions. The 1-sigma dispersions are plotted as a function of the duty cycle α in the left panel of Figure 6. The improvement in the fits is quite spectacular. For example, when $\alpha = 15\%$ the dispersion on ν_0 is 5 times less with the new fitting method ($1.5 \mu\text{Hz}$ vs. $7.4 \mu\text{Hz}$).

With the old method, the uncertainty on ν_0 increases much faster than $\alpha^{-1/2}$ as the duty cycle α drops ($\sim \alpha^{-1}$ between the 30% and 15% windows). This

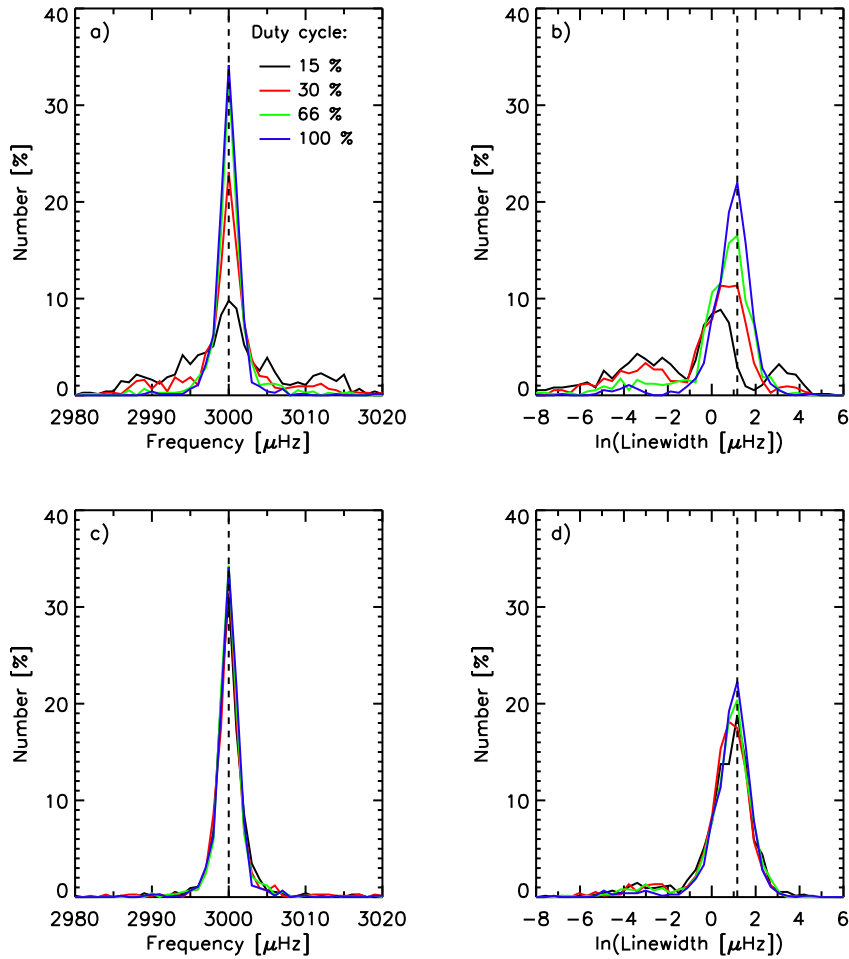


Figure 5. Distributions of the mode frequency and the linewidth for 750 realizations of solar-like oscillations, using the old fitting method (a-b) and the new fitting method (c-d). The observation windows have a duty cycle of 15%, 30%, 66%, and 100%. The vertical dashed lines represent the input values.

step dependence on α is worse than “predicted” by Libbrecht (1992). In his paper, Libbrecht suggested to use the uncertainty $\sigma_{\nu_0} = \sqrt{f\Gamma/(4\pi T)}$ where $f(\beta) = (1 + \beta)^{1/2}[(1 + \beta)^{1/2} + \beta^{1/2}]^3$ and β is an “effective” noise-to-signal ratio. He suggested that the main effect of the gaps is to increase the noise-to-signal ratio N/S , presumably by a factor $\sum_i |\hat{w}_i|^2 / |\hat{w}_0|^2$, itself proportional to $1/\alpha$. This leads however to a dependence of σ_{ν_0} on α which, in our particular case, is closer to $\alpha^{-1/2}$ than α^{-1} . We suspect that the Libbrecht formula underestimates the dispersion because it ignores the frequency correlations.

The new fitting method returns a ν_0 uncertainty that is much less sensitive to the duty cycle, with a variation like $\sim \alpha^{-0.15}$ (left panel of Fig. 6). This is quite

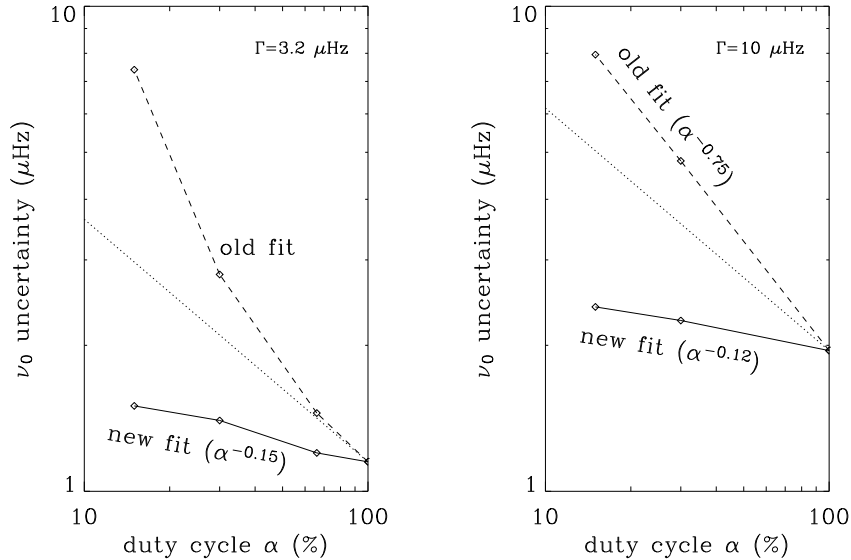


Figure 6. 1-sigma uncertainty of the mode frequency measurement as a function of duty cycle. The window functions are as defined in section 6.1. The left panel corresponds to the input linewidth $\Gamma = 3.2 \mu\text{Hz}$ (see numbers in Table 1). The right panel is for an input linewidth $\Gamma = 10 \mu\text{Hz}$, all other parameters being the same as in the left panel. For reference, the dotted line has slope $\alpha^{-1/2}$.

remarkable. That the frequency uncertainty could remain nearly constant for $\alpha > 30\%$ is not really surprising since the average gap (see numbers in Table 2) is less than the mode lifetime $\tau = 1/(\pi\Gamma) = 27.6$ hr. This regime was studied by Fossat et al. (1999) using a gap filling method: as long as the signal-to-noise ratio is large enough, the signal can be reconstructed. We do not fully understand, however, why the new fit is doing such a good job for the window with duty cycle $\alpha = 15\%$, since the average gap (40.7 hr) is larger than the mode lifetime.

In order to investigate this point, we ran another set of simulations using a mode linewidth $\Gamma = 10 \mu\text{Hz}$ corresponding to a mode lifetime $\tau = 8.8$ hr, which is significantly smaller than the average gap lengths of the 30% and the 15% windows. The other input parameters remained the same as above. We computed and fitted 1350 realizations. The results are shown in the right panel of Figure 6. For the new fitting method, the dependence of the frequency uncertainty on the duty cycle is about $\alpha^{-0.12}$, which is comparable to the previous simulations with $\Gamma = 3.2 \mu\text{Hz}$. We conclude that it is really worth solving for the correct minimization problem. We suspect that fitting for the phase information in complex Fourier space is important to get a good match between the model and the data. The fact that we have a perfect knowledge of the model is another important factor, which is rarely the case for real observations.

We should also mention that we ran another set of simulations using a higher signal-to-noise ratio. The difference between the old and the new fitting methods was not so obvious in this case.

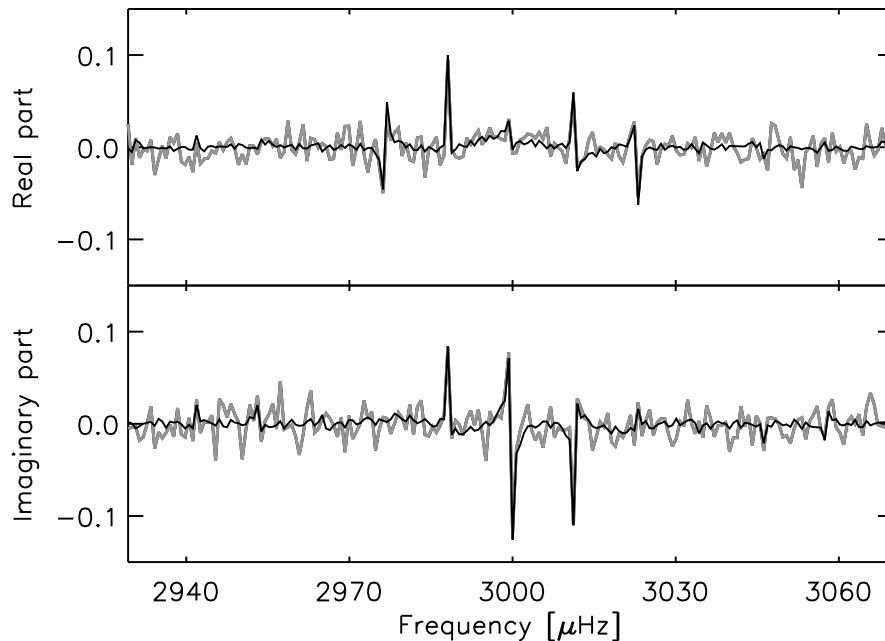


Figure 7. Real and imaginary part of the Fourier transform of a simulated gapped time series containing a sinusoid on top of white background noise. The signal-to-noise ratio is $S/N = 100$. The observation window has a duty cycle of 30%. The simulated data is the thick grey line. The thin black line shows the fit to the data using the new fitting method. The fit with the old method is not shown since it is almost identical.

7.3. Sinusoidal deterministic oscillations

Figure 7 shows the Fourier spectrum of a simulated time series containing a sinusoidal mode of oscillation on top of a white noise background as described in section 6.3. In this particular case the observation window with a duty cycle of 30% is used (see Fig. 2c). The input parameters of the sinusoidal function are the mode frequency $\nu_0 = 3000 \mu\text{Hz}$, the amplitude $A = 1.1$, and the phase $\varphi = 60^\circ$. The signal-to-noise ratio is $S/N = 100$. The fit shown in Figure 7 was obtained with the new fitting method. Since we found no significant difference between the old and the new fitting methods in this case, the old fitting method is not shown. Differences between the data and the fit are essentially due to the noise.

We computed 500 realizations of sinusoidal oscillations with the same mode parameters (frequency, amplitude, phase) as above, the same observation window (30% full), but with a signal-to-noise ratio $S/N = 46$. The resulting distributions of the inferred parameters obtained with the two fitting methods are shown in Figure 8. For this simulation, the known input values were used as an initial guess to speed up the minimization; we checked on several realizations that it is acceptable to do so when the signal to noise ratio is large. The distributions of the inferred parameters (Figure

8) show that for sinusoidal oscillations, the new fitting method does not provide any significant improvement compared to the old fitting method.

We emphasise that the fitting parameters can be determined with a very high precision when the noise level is small. In particular, we confirm that the uncertainty of the frequency estimator can be much smaller than $1/T$ (see Figure 8a). Figure 9 shows the median and the standard deviation of the mode frequency for different signal-to-noise ratios. Each symbol and its error bar in Figure 9 is based on the computation of 500 realizations of sinusoidal oscillations with the same mode parameters as above, the same observation window (30% full), but various signal-to-noise ratios. Since we did not find any significant difference between the two fitting methods, only the results obtained with the new fitting method are shown. Figure 9 illustrates that even for a relatively low signal-to-noise ratio of $S/N = 10$, the standard deviation of the inferred mode frequency is smaller than $1/T$ by a factor of four. For higher signal-to-noise ratios the precision is even more impressive: when $S/N = 100$, the standard deviation of the mode frequency is about 20 times smaller than $1/T$.

We have compared the standard deviation of the mode frequency with the theoretical value obtained Cuypers (1987) in the time domain:

$$\sigma_{\nu_0} = \frac{\sqrt{6} \sigma_t}{\pi A T \sqrt{n}}, \quad (37)$$

where A is the amplitude of the sinusoid in the time domain, σ_t is the rms value of the noise, $n = \alpha N$ is the number of recorded data points, and T is the total observation length. This theoretical uncertainty is overplotted in Figure 9. The match with our Monte-carlo measurements is excellent. This confirms that, in this case, it is equivalent to perform the fits in the temporal and in the Fourier domains. Note that equation (37) is only valid under the assumption that the noise is uncorrelated in the time domain, a condition fulfilled by our simulations. The main reason why the measurement precision is only limited by the noise-to-signal ratio is because perfect knowledge of the model is assumed.

8. Conclusion

In this paper we derived an expression for the joint PDF of solar or stellar oscillations in complex Fourier space, in agreement with the work of Gabriel (1994). This joint PDF explicitly takes into account frequency correlations introduced by the convolution with the window function. We implemented a maximum likelihood estimation method to retrieve the parameters of stellar oscillations. Both stochastic solar-like oscillations and deterministic sinusoidal oscillations were considered.

In the case of solar-like oscillations, we performed Monte-Carlo simulations to show that the improvement provided by our fitting method can be very significant in comparison with a fitting method that ignores the frequency correlations. In one particular example, using an observation window with a duty cycle $\alpha = 30\%$ and a signal-to-noise ratio $S/N = 6$, the new fitting method increased the precision on the mode frequency by a factor of two and the estimates of the linewidth and mode power were less biased and more precise. For a window with a the duty cycle $\alpha = 15\%$, the precision on the mode frequency estimate was increased by a factor of five.

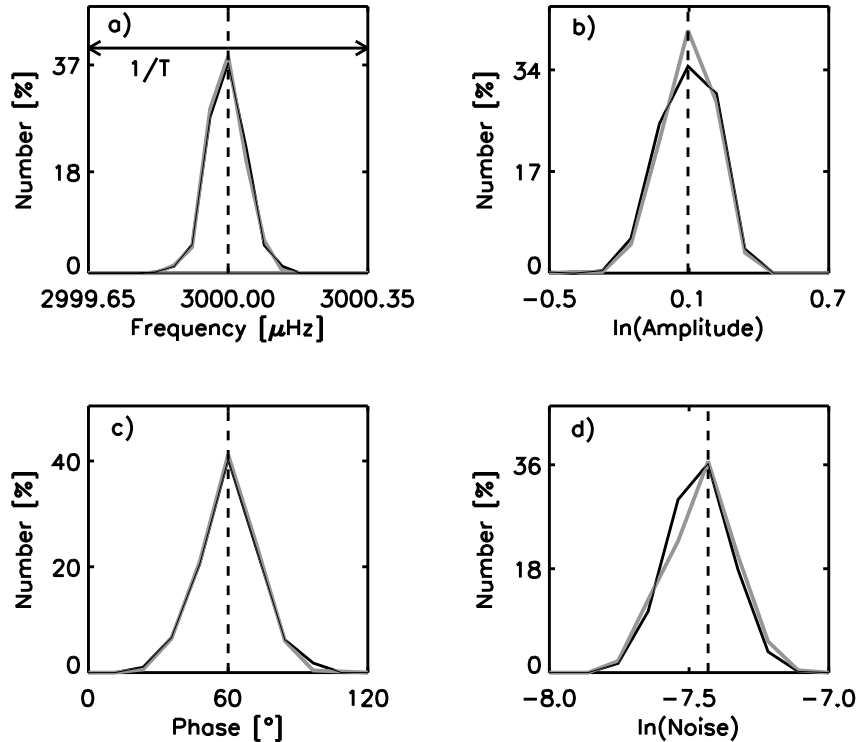


Figure 8. Distributions of the inferred oscillation parameters for a set of 500 realizations of long-lived sinusoidal oscillations with $S/N = 46$. The window function with a duty cycle of 30% is used. The black and the grey lines are for the new and old fitting methods respectively. The vertical dashed line in each plot indicates the input value. The parameters shown are (a) the mode frequency ν_0 , (b) the logarithm of the mode amplitude $\ln A$, (c) the phase of the oscillation ϕ , and (d) the logarithm of the noise level $\ln N$ (see eq. [34]).

In the case of long-lived, purely sinusoidal oscillations, we did not find any significant improvement in using this new fitting method. Yet we confirm that the standard deviation of the mode frequency can be measured in Fourier space with a precision much better than $1/T$ for large signal-to-noise ratios, in accordance with a time-domain calculation by Cuypers (1987).

The analysis of time series containing a lot of gaps can benefit from our work. Applications may include, for example, the re-analysis of solar oscillations from the early days of the BiSON network (Miller et al., 2004) or the solar-like oscillations of α Centauri observed from the ground with two telescopes (Butler et al., 2004).

Acknowledgements We thank T. Appourchaux for useful discussions. T. Stahn is a member of the International Max Planck Research School on Physical Processes in the Solar System and Beyond at the Universities of Göttingen and Braunschweig.

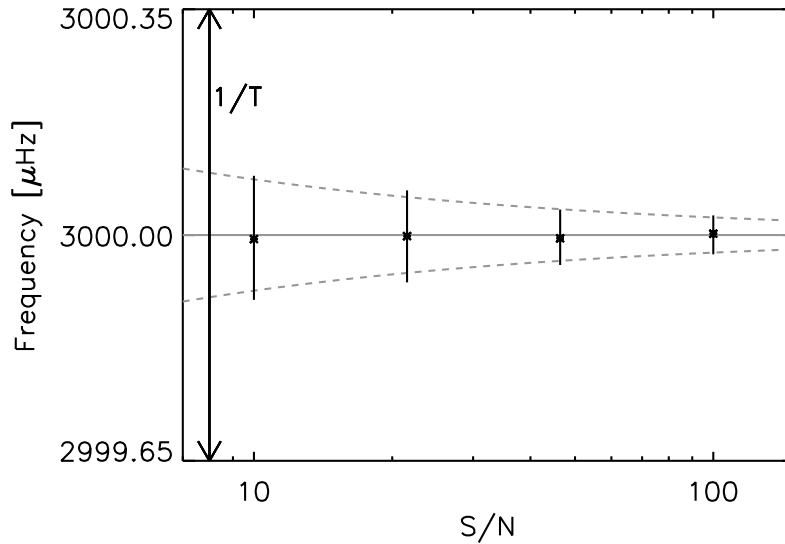


Figure 9. Median (cross) and standard deviation (vertical bar) of the inferred frequency of sinusoidal oscillation, ν_0 , as a function of signal-to-noise ratio S/N . Only the results obtained with the new fitting method are shown. The horizontal grey line shows the input mode frequency. The dashed grey lines is the theoretical value of frequency uncertainty given by eq. (37). The vertical axis of the plot spans the interval $\Delta\nu = 1/T = 0.7 \mu\text{Hz}$.

References

- Anderson, E.R., Duvall, T.L., Jr., and Jefferies, S.M.: 1990, *Astrophys. J.* **364**, 699.
- Appourchaux, T., Gizon, L., and Rabello-Soares, M.-C.: 1998, *Astron. Astrophys. Suppl. Ser.* **132**, 107.
- Appourchaux, T., Chang, H.-Y., Gough, D.O., and Sekii, T.: 2000, *Month. Not. Roy. Astron. Soc.* **319**, 365.
- Bedding, T.R. and Kjeldsen, H.: 2007, *Comm. Asteroseismology* **150**, 106.
- Brandt, S.: 1970, *Statistical and Computational Methods in Data Analysis*, North-Holland, Amsterdam.
- Butler, R.P., Bedding, T.R., Kjeldsen, H., McCarthy, C., O'Toole, S.J., Tinney, C.G., Marcy, G.W., and Wright, J.T.: 2004, *Astrophys. J.* **600**, L75.
- Cuyppers, J.: 1987, *Bulletin de l'académie royale des sciences, des lettres et des beaux arts de Belgique (classe des sciences)* **49**, 21.
- Duvall, T.L., Jr., Jefferies, S.M., Harvey, J.W., Osaki, Y., and Pomerantz, M.A.: 1993, *Astrophys. J.* **410**, 829.
- Fossat, E., Kholikov, S., Gelly, B., Schmider, F.X., Fierry-Fraillon, D., Grec, G., Palle, P., Cacciani, A., Ehgamberdiev, S., Hoeksema, J.T., and Lazrek, M.: 1999, *Astron. Astroph.* **343**, 608.
- Gabriel, M.: 1994, *Astron. Astroph.* **287**, 685.
- Gizon, L. and Solanki, S.K.: 2003, *Astrophys. J.* **589**, 1009.
- Horn, R.A. and Johnson, C.R.: 1985, *Matrix Analysis*, Cambridge Univ. Press, Cambridge.
- Libbrecht, K.G.: 1992, *Astrophys. J.* **387**, 712.
- Miller, B.A., Hale, S.J., Elsworth, Y., Chaplin, W.J., Isaak, G.R., and New, R.: 2004, in D. Danesy (ed.), *Helio- and Asteroseismology: Towards a Golden Future*, ESA, Noordwijk, p. 571.
- Schou, J.: 1992, Ph.D. Dissertation, University of Aarhus.
- Schwarzenberg-Czerny, A.: 1991, *Month. Not. Roy. Astron. Soc.* **253**, 198.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Stein, R., Georgobiani, D., Trampedach, R., Ludwig, H.-G., and Nordlund, Å.: 2004, *Solar Phys.* **220**, 229.

Toutain, T. and Appourchaux, T.: 1994, *Astron. Astroph.* **289**, 649.

Winget, D.E., Nather, R.E., Clemens, J.C., Provencal, J., Kleinman, S.J., Bradley, P.A., Wood, M.A., Claver, C.F., Frueh, M.L., Grauer, A.D., Hine, B.P., Hansen, C.J., Fontaine, G., Achilleos, N., Wickramasinghe, D.T., Marar, T.M.K., Seetha, S., Ashoka, B.N., O'Donoghue, D., Warner, B., Kurtz, D.W., Buckley, D.A., Brickhill, J., Vauclair, G., Dolez, N., Chevreton, M., Barstow, M.A., Solheim, J.E., Kanaan, A., Kepler, S.O., Henry, G.W., and Kawaler, S.D.: 1991, *Astrophys. J.* **378**, 326.

Woodard, M.F.: 1984, Ph.D. Dissertation, University of California, San Diego