A&A manuscript no.
(will be inserted by hand later)

Your thesaurus codes are:
03.13.2; 06.15.1

ASTRONOMY
AND
ASTROPHYSICS
21.10.1996

On the estimation of mode parameters from helioseismic time-series

J. Schou[1,2,⋆]

[1] Institut for Fysik og Astronomi, Aarhus Universitet, DK-8000 Århus C, Denmark
[2] High Altitude Observatory, National Center for Atmospheric Research,[⋆⋆] Box 3000, Boulder, CO 80307, U.S.A.

[the date of receipt and acceptance should be inserted later]

Abstract. Helioseismic time-series have until now been analyzed with a large number of different methods, which take into account the physical properties of the modes to a varying extent. It has not been clear whether the neglect of certain properties is unimportant or has led to systematic errors and increased random errors.

Most methods have until now used a type of least squares fitting of the power spectra, although it is well known that the points in the power spectra are far from normally distributed. In particular Anderson et al. (1991) have shown that the wrong assumption of Gaussian errors leads to both large systematic errors and increased random errors. Also the imperfect isolation of individual modes has been neglected in many cases, and the resulting correlation between different time-series has not been taken into account.

In this paper I present an analysis method that allows for some of these effects and describe some of the details of the implementation I am currently using. While this method does take explicit notice of many aspects of time series data, it is far from complete. Some of the neglected properties will be pointed out later in the paper, while others have undoubtly been overlooked. While this method is substantially more computationally intensive than least squares methods, it is not prohibitively so. I have used it to analyze data taken with the HAO/NSO Fourier Tachometer (see Bachmann, Schou and Brown, 1993). To illustrate the properties of the method, I will present a few results from this analysis and from the analysis of artificial data (see Schou and Brown, 1994).

It is the hope that some of the ideas underlying the method will be generally useful for the analysis of other datasets.

Key words: methods: data analysis – sun: oscillations

1. Introduction

The relationship between the basic mode parameters, the observation technique and the resulting time-series and Fourier transforms thereof has been described in Schou & Brown (1994), and some of the notation from that paper is used here.

Briefly, normal mode oscillations of the Sun are seen as periodic variations in intensity or velocity observed at the solar photosphere, each characterized by a particular temporal frequency and by a variation in the horizontal dimensions proportional to a spherical harmonic. The oscillations are observed by taking intensity or velocity images at a cadence of typically 1 image per minute. In order to isolate the individual modes, these images are multiplied by suitable spatial masks for each (l, m), designed to isolate the target mode without being too sensitive to noise or other modes. The time-series for each target (l, m) is then Fourier transformed in time. As it is generally not possible to observe the Sun continuously, the time-series must be zero-filled or interpolated in some other way to fill the gaps. This leads to problems such as temporal sidelobes and lack of independence of the different frequency points in the Fourier transform; these should ideally be taken into account in the analysis.

Previous methods usually squared the Fourier transforms to turn them into power spectra, thereby ignoring the phase information. As the phase information is used in the presented method, this is not done here.

The solar p-modes are thought to be well modeled by damped oscillators that are excited stochastically. In this case the values of the real and imaginary parts of the Fourier transform of a time-series of an isolated mode at a frequency ν are normally distributed with a mean of 0 and a variance $v(\nu)$. The normal distribution of the points in the Fourier transform follows from the assumption that the perturbations from the excitation are normally distributed or alternatively from the Central Limit Theorem to a very good approximation, as the time-series are generally very long. The variance is given by a Lorenzian in frequency:

$$v(\nu) = \frac{P_{\rm tot}/w}{1 + (\frac{\nu-\nu_0}{w})^2} + r, \qquad (!E1)$$

where ν_0 is the frequency of the mode, w is the Half Width at Half Maximum (HWHM) of the line profile, $P_{\rm tot}$ is a measure of the total power in the mode, and r is the background noise (other sets of parameters can also be used to describe the Lorenzian). Furthermore, under certain assumptions, it may be shown that the values at different frequency points in the discrete Fourier transform are independent. Most importantly, it

⋆ Present address: Center for Space Science and Astrophysics, Stanford University ERL 328, Stanford, CA 94305
⋆⋆ The National Center for Atmospheric Research is supported by the National Science Foundation.

has to be assumed that the mode excitation is random and frequent compared to the lifetime of the mode.

Hence, to estimate the mode parameters (ν_0, w, $P_{\rm tot}$, r, and possibly others, which will be lumped into a vector $\mathbf{a}$), one should maximize, with respect to the mode parameters, the likelihood function L (the joint probability density), given by

$$L(\mathbf{a}) = \prod_i P(\mathbf{a}, \nu_i), \qquad (!E2)$$

where the product is over a suitable number of frequencies in the real and imaginary parts of the discrete Fourier transform and the P's are the individual probability densities. Notice that it has been assumed that the individual points are independent and that the dependence of L and P on the observed values has been supressed. The individual probability densities P are given by

$$P(\mathbf{a}, \nu_i) = \frac{1}{\sqrt{2\pi v(\mathbf{a}, \nu_i)}} e^{-\frac{1}{2}\frac{x(\nu_i)^2}{v(\mathbf{a}, \nu_i)}} \qquad (!E3)$$

for both the real and the imaginary parts, where $x(\nu_i)$ are the observed values in the Fourier transform of the data.

Maximizing L is obviously equivalent to minimizing the negative of the logarithm of L, which except for a constant, is equivalent to minimizing

$$S(\mathbf{a}) = \sum_i \log[v(\mathbf{a}, \nu_i)] + \frac{x(\nu_i)^2}{v(\mathbf{a}, \nu_i)}. \qquad (!E4)$$

Notice that S plays the same role as χ^2 in a least squares fit. This is essentially the method described in Anderson et al. (1991).

Unfortunately, since only one half of the Sun is observed, it is essentially impossible to isolate individual modes based on their angular dependence alone. For velocity observations, where one is only able to observe the line of sight component of the total velocity, the sensitivity decreases strongly close to the solar limb (where there is already a significant foreshortening), leading to an effectively reduced observed area and increased problems with isolating individual modes. While it would seem possible to reduce this problem by increasing the weight of observations near the limb, the noise typically increases rapidly towards the limb, making this approach undesirable.

As an illustration of the degree of correlation between different modes, see Fig. !F1. It is obvious that, at least for the modes shown, the correlation is very strong. The correlation depends on the type of observations (intensity or velocity), on the spatial masks used, and on the values of l and m. Typically intensity observations have lower correlations than velocity observations (again due to the velocity projection effect), and lower m's have higher correlations than higher m's at the same l. In any case, the assumption that the individual Fourier transforms are independent is evidently incorrect.

Since the modes are not truly isolated by the spatial masks used, the Fourier transform for a particular target (l, m) is not given by a Fourier transform of a single mode, but rather as a sum over several modes. The Fourier transform y of an observed time-series is in this case given by

$$y_n(\nu_i) = \sum_k c_{nk} x_k(\nu_i) \qquad (!E5)$$

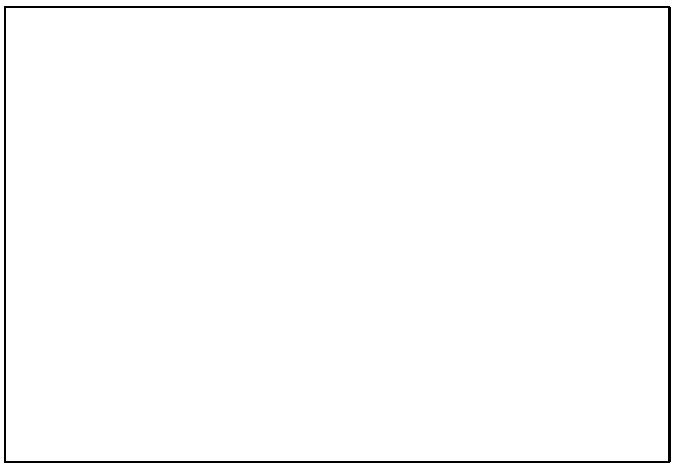

Fig. !F1. A scatter plot showing the values of the time-series for $(l, m) = (30, 2)$ versus the values of $(l, m) = (30, 0)$. The data are from part of an observation run with the Fourier Tachometer from 1989.

where the sum is over the modes contributing to the given time-series, c describes the contribution from each mode and x_k is the Fourier transform of the k'th mode. The 'leakage matrix' c can be calculated as described in Schou and Brown (1994) given the physics of the modes, various instrumental and atmospheric effects and the spatial masks used. Notice that c is independent of i and is the same for the real and imaginary parts of the Fourier transform. Ideally the sum should be over all modes, but as described in more detail later, only modes within a certain l range are generally used, and certain approximations have to be made for l's other than the target l. Also note that the noise has been ignored for the moment.

Given the expression (!E5) the expected covariance matrix $E(\mathbf{a}, \nu_i)$ between the different y's is

$$E_{nm}(\mathbf{a}, \nu_i) = \mathrm{Cov}[y_n(\nu_i), y_m(\nu_i)] = \sum_k c_{nk} c_{mk} v_k(\mathbf{a}, \nu_i). \quad (!E6)$$

In this case the probability density at a frequency point ν_i is given by

$$P(\mathbf{a}, \nu_i) = \frac{1}{\sqrt{|2\pi E(\mathbf{a}, \nu_i)|}} e^{-\frac{1}{2} y_i^T E(\mathbf{a}, \nu_i)^{-1} y_i}, \qquad (!E7)$$

where $|\ |$ denotes the determinant, and the likelihood function is thus maximized by minimizing

$$S(\mathbf{a}) = \sum_i \log |E(\mathbf{a}, \nu_i)| + y_i^T E(\mathbf{a}, \nu_i)^{-1} y_i, \qquad (!E8)$$

where it has again been assumed that observed points with different frequencies are independent.

To illustrate the idea behind this approach, consider the (hypothetical) case in which there is no leakage from other l's and no noise. Let $x(\nu_i)$ be the Fourier transforms of the original time-series for all the m's at a given l and frequency ν_i and $y(\nu_i)$ be the observed Fourier transforms. In this case $y(\nu_i) = Cx(\nu_i)$, where C is an $N \times N$ matrix and N is the number of m's, typically $N = 2l+1$. Let $v(\mathbf{a}, \nu_i)$ be the model variances for the original Fourier transforms and $D(\mathbf{a}, \nu_i) = \mathrm{diag}(\sqrt{v(\mathbf{a}, \nu_i)})$. The probability density P_x for the original Fourier transform at

the frequency ν_i is now given by the product of the individual probabilities

$$P_x(\mathbf{a},\nu_i,x) = \prod_m \frac{1}{\sqrt{2\pi}v_k(\mathbf{a},\nu_i)} e^{-\frac{1}{2}\frac{x_k^2}{v_k(\mathbf{a},\nu_i)}} = \frac{1}{|\sqrt{2\pi}D(\mathbf{a},\nu_i)|} e^{-\frac{1}{2}|D(\mathbf{a},\nu_i)^{-1}x(\nu_i)|^2}, \quad (!E8n)$$

where the argument x to P_x has been made explicit.

If it is assumed that C is not singular then $x(\nu_i) = C^{-1}x(\nu_i)$ and the probability density P for the observed Fourier transforms can be found from a change of variables:

$$\begin{aligned} P(\mathbf{a},\nu_i,y) &= |C^{-1}|P_x(\mathbf{a},\nu_i,C^{-1}y) \\ &= \frac{1}{|C|}\frac{1}{|\sqrt{2\pi}D|}e^{-\frac{1}{2}|D^{-1}C^{-1}y_i|^2} \\ &= \frac{1}{|\sqrt{2\pi}CD|}e^{-\frac{1}{2}y_i^T C^{-1}D^{-1}D^{-1}C^{-1}y_i} \\ &= \frac{1}{\sqrt{|2\pi CDDC|}}e^{-\frac{1}{2}y_i^T (CDDC)^{-1}y_i} \\ &= \frac{1}{\sqrt{|2\pi E|}}e^{-\frac{1}{2}y_i^T E^{-1}y_i} \end{aligned}, \quad (!E8p)$$

which is equivalent to (!E7). In the derivation the arguments to D were dropped and $E = CDDC$, which is just the matrix version of (!E6), was used.

In other words, in the absense of noise and crosstalks, the procedure described is equivalent to solving for the original Fourier transforms and maximizing the product of the probabilites for the individual transforms, that is assuming that the individual m's are independent.

2. Implementation

Clearly it is not possible to fit for all modes simultaneously, which would be the ideal way to do the fitting. Doing so would lead to both purely numerical problems and be computationally impossible due to the size of the covariance matrix. Among the numerical problems encountered would be that the covariance matrix would be very close to singular due to the large crosstalks between modes.

Fortunately the crosstalks are generally very low between modes with significantly different l or m. Also only a few modes within the range where the crosstalks are significant have frequencies close to each other. Given these facts, a natural way to proceed is to fit one (n,l) at a time, trying to take into account the leakage from other modes, at least partially.

Given that the m's are closely spaced in frequency, compared to their linewidths, it is clear that the leakage among the m's must be taken into account. Unfortunately, it turns out that the covariance matrix is close to singular, for moderately high l's, if all m's are considered at a given frequency. A partial solution to this problem is, as described in more detail later, to only consider a fairly small range in m at each frequency, covering those m's with a significant expected power at that frequency.

The close spacing of the m's also means that it is likely to be difficult to fit the frequencies for the individual m's. Fortunately it appears that the variation of the mode frequency with m is fairly smooth. Hence, the individual mode frequencies are not fitted for, but are instead represented by an 'unperturbed' mode frequency ν_{nl} and number of so-called a-coefficients in a polynomial

$$\nu_{nlm} = \nu_{nl} + \sum_{i=1}^{s_{max}} a_i(n,l)P_i^l(m). \quad (!E9)$$

The fit can then be done for ν_{nl} and the a-coefficients. Previous workers (eg. Libbrecht 1989) have generally taken the polynomials P_i^l to be Legendre polynomials $P_i^l(m) = lP_i(m/l)$ or $P_i^l(m) = LP_i(m/L)$, where P_i is a Legendre polynomial of degree i, but I have chosen to use polynomials that are orthogonal with respect to summation, as suggested by Ritzwoller and Lavely (1991) (see also Schou, Christensen-Dalsgaard, and Thompson, (1994) for an exact definition of the polynomials used here). The use of a-coefficients rather than individual mode frequencies has been common to all published observations of medium l modes until now. Notice, however, that the current method should be able to fit individual frequencies, although there are, as mentioned, likely to be stability problems, at least for the observations available at the moment. Also note that the algorithm presented here can use the traditional types of a-coefficients.

In the following some of the details of the fitting method and what has been taken into account and what has been neglected will be presented. Some possible improvements will also be discussed later.

(i) General description of minimization procedure

Examples of contour plots of S as a function of selected pairs of variables are shown in Fig. !F2. The values of the other parameters were held fixed at a value close to the minimum.

As can be seen from the contour plots in Fig. !F2, most of the parameter combinations are fairly well behaved, that is reasonably well approximated by a quadratic form. The contours are somewhat tilted for some combinations, indicating that the determinations of the parameters are correlated; in those cases the indicated error bars are the projections of the confidence ellipsoid onto the parameter axis. Generally the highest correlations occur between the parameters determining the magnitude of the spectrum (amplitude, linewidth, noise level, and, if used, the power level of modes other than the target mode, as described later). Also the group of parameters determining the average (over m) mode frequency (ν, a_2, a_4, a_6 etc.) are correlated. Note that although the polynomials used for the a-coefficients are orthogonal or very nearly so, the effective weights assigned to the m's are not equal, leading to some correlations. There also tend to be small correlations among the odd a-coefficients.

As Fig. !F2 shows, S is not always perfectly represented by a quadratic form far from the minimum parameter values, but this should not lead to major problems. Notice that the mode used as an example has a fairly high S/N (signal to noise), for modes with very low S/N, the contours can be less well behaved.

Despite the highly nonlinear terms in Eq. (!E8), it thus appears that the function to be minimized is fairly well approximated by a quadratic form and can thus be minimized using standard methods used for nonlinear least squares minimizations. Also note that it is possible to calculate analytical

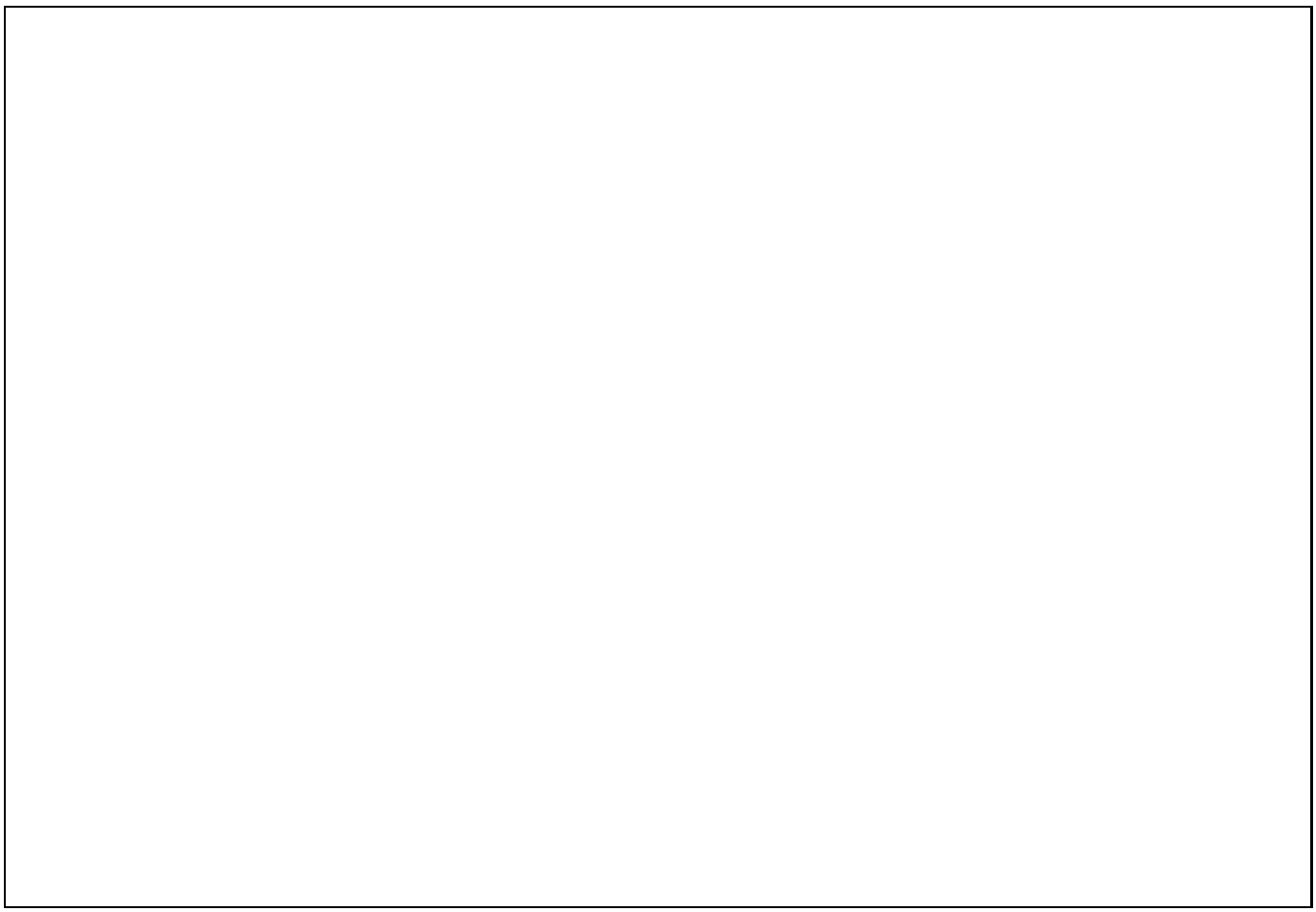

Fig. !F2. Contour plots of the function S. The independent variables on each of the sub plots are indicated on those. The values of the other independent variables have been held at a value close to the minimum. On each plot the minimum has been indicated with a cross of size 1σ. On the upper left plot, the contour spacing is 5, on the upper right hand plot the spacing is 10 and on the lower two plot the spacing is 2. The zero point of the frequency scale was chose to be at the frequency bin closest to the minimum. The units for the amplitudes, linewidths and background power are arbitrary.

derivatives of S. Estimates of the random errors on the determined parameters can also be found using methods similar to those in the least squares case, that is an estimate of the covariance matrix can be obtained from the inverse of the second derivative matrix of S. The iteration method used here is a modified version of the Levenberg-Marquardt method, an inefficient implementation of which may be found in Press et al. (1986) (among the problems with their implementation are that functions and derivatives are calculated several times for the same parameter values, second derivatives are recalculated even for even the tiniest parameter changes, the control parameter may underflow, and various similar problems).

(ii) Calculation of derivatives

As mentioned, it is possible to calculate analytic derivatives of S. Dropping the summation over the frequency points (and the corresponding subscripts) and defining $z \equiv E^{-1}y$ one has:

$$\begin{aligned}\frac{\partial S}{\partial v_i} &= \frac{1}{|E|}\frac{\partial |E|}{\partial v_j} + y^T\left(-E^{-1}\frac{\partial E}{\partial v_i}E^{-1}\right)y \\ &= \left(\frac{\partial E}{\partial v_i}\cdot E^{-1}\right) - z^T\frac{\partial E}{\partial v_i}z\end{aligned} \qquad (!E10)$$

and

$$\begin{aligned}\frac{\partial^2 S}{\partial v_j \partial v_i} &= \frac{\partial^2 E}{\partial v_j \partial v_i}\cdot E^{-1} + \frac{\partial E}{\partial v_i}\cdot\left(-E^{-1}\frac{\partial E}{\partial v_j}E^{-1}\right) \\ &\quad - y^T\left(-2E^{-1}\frac{\partial E}{\partial v_j}E^{-1}\frac{\partial E}{\partial v_i}E^{-1} + E^{-1}\frac{\partial^2 E}{\partial v_j \partial v_i}E^{-1}\right)y.\end{aligned} \qquad (!E11)$$

Rearranging a bit using the fact that $\frac{\partial^2 E}{\partial v_j \partial v_i} = 0$ for all i, j (as follows from Eq. (!E6)) yields

$$\begin{aligned}\frac{\partial^2 S}{\partial v_j \partial v_i} &= -\left(E^{-1}\frac{\partial E}{\partial v_j}\right)\cdot\left(E^{-1}\frac{\partial E}{\partial v_i}\right)^T \\ &\quad + 2z^T\frac{\partial E}{\partial v_j}E^{-1}\frac{\partial E}{\partial v_i}z,\end{aligned} \qquad (!E12)$$

where the expressions

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(A^{-1}\right) = -A^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t),$$

$$\frac{\mathrm{d}|A|}{\mathrm{d}t} = |A|\sum_{i,j}\frac{\mathrm{d}a_{ij}}{\mathrm{d}t}\left(A^{-1}\right)_{ji} = |A|\frac{\mathrm{d}A}{\mathrm{d}t}\cdot\left(A^{-1}\right)^T,$$

$$A \cdot B \equiv \sum_{i,j} A_{ij} B_{ij} = \sum_i (AB^T)_{ii}$$

and

$$A \cdot (BC) = (B^T A) \cdot C \qquad (!E13)$$

have been used.

Finally we have

$$\frac{\partial S}{\partial a_k} = \sum_i \frac{\partial S}{\partial v_i} \frac{\partial v_i}{\partial a_k} \qquad (!E14)$$

and

$$\begin{aligned} \frac{\partial^2 S}{\partial a_k \partial a_l} &= \sum_{i,j} \frac{\partial^2 S}{\partial v_j \partial v_i} \frac{\partial v_j}{\partial a_l} \frac{\partial v_i}{\partial a_k} + \sum_i \frac{\partial S}{\partial v_i} \frac{\partial^2 v_i}{\partial a_k \partial a_l} \\ &\approx \sum_{i,j} \frac{\partial^2 S}{\partial v_j \partial v_i} \frac{\partial v_j}{\partial a_l} \frac{\partial v_i}{\partial a_k} \end{aligned} \qquad (!E15)$$

where the second derivatives of v with respect to $\mathbf{a}$ have been dropped. Note that since only the second derivative is modified in this way and not the function be minimized, the fitted value is (ideally) not affected, only the iteration path taken to reach it. It is possible to use other approximations, but this is the most straightforward to derive.

The derivatives may also be calculated by using E instead of the variances as intermediate variable. This way of expressing the derivatives is useful when calculating the derivatives with respect to the background noise. The noise is added as a term $rE_{\rm noise}$ to Eq. (!E8), where r is the noise level and $E_{\rm noise}$ is an estimate of the noise covariance matrix. (Actually the logarithm of the noise is used in the fitting.) In this case $\frac{\partial E}{\partial r} = E_{\rm noise}$.

Then we have:

$$\frac{\partial S}{\partial E_{ij}} = (E^{-1})_{ji} - z_i z_j, \qquad (!E16)$$

and

$$\begin{aligned} \frac{\partial^2 S}{\partial E_{ij} \partial E_{mn}} &= -(E^{-1})_{im} (E^{-1})_{nj} \\ &\quad + z_j (E^{-1})_{im} z_n + z_m (E^{-1})_{nj} z_i, \end{aligned} \qquad (!E17)$$

and thereby, using the chain rule

$$\frac{\partial S}{\partial a_k} = \sum_{ij} \frac{\partial S}{\partial E_{ij}} \frac{\partial E_{ij}}{\partial a_k}, \qquad (!E18)$$

and

$$\begin{aligned} \frac{\partial^2 S}{\partial a_l \partial a_k} &= \sum_{ij} \frac{\partial S}{\partial E_{ij}} \frac{\partial^2 E_{ij}}{\partial a_k \partial a_l} + \sum_{ijmn} \frac{\partial^2 S}{\partial E_{ij} \partial E_{mn}} \frac{\partial E_{ij}}{\partial a_k} \frac{\partial E_{mn}}{\partial a_l} \\ &\approx \sum_{ijmn} \frac{\partial^2 S}{\partial E_{ij} \partial E_{mn}} \frac{\partial E_{ij}}{\partial a_k} \frac{\partial E_{mn}}{\partial a_l}, \end{aligned} \qquad (!E19)$$

where the second derivatives of E with respect to $\mathbf{a}$ have been neglected.

A way to find the noise covariance matrix $E_{\rm noise}$ is to estimate it from the noise outside the p-mode band. Unfortunately the different noise sources contribute different fractions of the noise at different frequencies, and it is thus not quite clear which part of the frequency spectrum to use. Different methods for estimating the noise covariance matrix will be discussed later.

(iii) Effects taken into account

The presence of gaps in the time-series due to day/night cycles, bad weather and instrument failures also needs to be taken into account. The periodic nature of the gaps typically leads to so called temporal sidelobes in the power spectra/Fourier transforms. This is modeled by convolving the model spectra (the v's) and their derivatives with the sidelobe structure (which is assumed known a priori). This only solves half of the problem, however. It takes care of properly modelling the sidelobes, but the correlations between different frequency points in the Fourier transform introduced by the sidelobes is not taken into account. The possible effects of this omission will be discussed later.

Another effect to be taken into account is the signal contributed from the modes not explicitly fitted for (eg. those with (n,l) different from the target mode). These modes cause additional peaks to appear in the power spectra, both through their main lobes and their temporal sidelobes. Generally the main lobes are of little concern as they for most l's have frequencies far from the target mode. For the very lowest l's, they can however present some problems. The temporal sidelobes of the neighbouring modes on the other hand often cause problems as they can have frequencies close to those of the target mode.

One fairly simple way of taking these other modes into account is to treat their contribution to the covariance matrix for the target mode as an additional source of (highly frequency dependent) noise. To be specific, the contribution $E_{\rm other}$ to the covariance matrix from the other l's and n's is calculated from the initial estimates of the parameters for the other modes, using Eq. (!E6). Equation (!E6) is then modified by adding a term $P_{\rm other} E_{\rm other}$. The parameter $P_{\rm other}$ is either kept fixed $P_{\rm other} = 1$, fitted for or adjusted according to the correction to the power of the target mode. The 'justification' for the latter approach is that modes with almost the same frequency but slightly different l (those most likely to interfere) are likely to have their initial estimates in error by roughly the same amount. This method also has the advantage that it stabilizes the fit, as there is one less parameter to fit for. This is clearly not the ideal way to do things, but most other ways have undesirable properties, such as increasing the computing time by an order of magnitude. I will discuss this point further later.

Since the power in a mode falls off rapidly away from the center frequency, the mode parameter fitting needs to use only a small range of frequencies around the expected frequency. Indeed it may be shown (Schou, 1994) that for frequency ranges of only a few times the mode linewidth, the loss in potential accuracy is fairly small. This fact, combined with the rapid variation of the frequency with m due to the average solar rotation rate, means that at each frequency only a certain range in m needs to be considered. This leads to a very significant reduction in the computing time, and also has the desirable side-effect of removing the near singularity of the covariance matrix, which can otherwise lead to numerical problems. If the linewidth is large (which is the case for high frequency modes) it is necessary to restrict the frequency range to the point of reducing the accuracy of the parameter determinations, or to get around these numerical problems in some other way. Some possible ways of doing this will be discussed later.

Another convenient result is that, at least for the types of masks generally used, the even and odd m's can be treated separately; there is no crosstalk between even and odd m's.

(iv) Diagnostics

Given all the aspects taken into account and the associated potential problems, a natural question to ask is whether it is possible to find diagnostics telling if the real data have the expected statistical properties. In particular, it would be useful to know if the spatial leakages have been incorrectly estimated. It turns out that it is indeed possible to find such fairly general diagnostics. At a given frequency point the observed values y should have a covariance E given by expression (!E6). If a matrix D exists such that $E = DD^T$, it may be shown that the covariance matrix of $y' \equiv D^{-1}y$ is the identity matrix. Since E is a covariance matrix, and is thus positive semidefinite, such a matrix D always exists. Moreover, the standard way to solve a linear system involving a positive definite matrix is to use a Cholesky decomposition, which as a byproduct produces such a matrix D. Notice that while the Cholesky decomposition produces one such matrix, several other matrices generally exist, and it is not clear that this is the most desirable D to use. If E is indeed a good estimate of the true covariance matrix, as it should be if the model is correct and the iteration converged, y' should be a vector of (hopefully) independent unit variance numbers at each frequency point. If on the other hand the model is incorrect and the crosstalks have been incorrectly estimated, the elements in the vector y' are likely to be correlated. Checking this vector for correlations should thus be a good test for problems with the crosstalks. Unfortunately it is not easy to test a single realisation of a random vector for correlations, especially not to get numbers describing the degree and characteristics of the correlations. In any case a few numbers characterizing all the frequency points are likely to be more useful than a series of numbers at each frequency point. Since the crosstalks most likely to be significantly incorrect (that is likely to lead to problems) are those between close m's, a way proceed is to calculate the covariance between all pairs of points at all the individual frequencies with Δm of 0, 2, 4, ... That is

$$\frac{\sum_i \sum_{|m-m'|=\Delta m} y'(\nu_i)_m y'(\nu_i)_{m'}}{\sum_i \sum_{|m-m'|=\Delta m} 1} = \frac{\sum_i \sum_{|m-m'|=\Delta m} y'(\nu_i)_m y'(\nu_i)_{m'}}{\text{Number of points used}}, \qquad (!E20)$$

where $y'(\nu_i) = D(\mathbf{a},\nu_i)^{-1} y(\nu_i)$ with $D(\mathbf{a},\nu_i)D(\mathbf{a},\nu_i)^T = E(\mathbf{a},\nu_i)$. As there is no crosstalk between the even and odd m's and we have only used the even m's in the analysis of real data, only even Δm's have been considered.

Another diagnostic that has been useful is to simply check that the power as a function of m is the same in the real data and the best fit model spectrum. Since the number of m's is large for high l's the ratio of the observed to the fitted power has been expanded in polynomials in m/l. To be specific define

$$Q_m = \frac{\sum_{\nu_i \text{ used}} \frac{y_m(\nu_i)^2}{E_{mm}(\nu_i)}}{\sum_{\nu_i \text{ used}} 1} = \frac{\sum_{\nu_i \text{ used}} \frac{y_m(\nu_i)^2}{E_{mm}(\nu_i)}}{\text{number of modes used}} \qquad (!E21)$$

and expand using the polynomials used for the a-coefficients

$$Q_m = \sum_i b_i P_i^l(m)/l. \qquad (!E22)$$

If the power is correct as a function of m, $b_0 = 1$ and $b_i = 0$ for $i > 1$. This diagnostic has been particularly useful in showing the effect of a wrong assumed asymmetry of the point spread function (see eg. Schou and Brown, 1994) used in the calculation of the crosstalks (which affects different m's at the same l differently).

Obviously it is possible to devise a whole series of other diagnostics, but these two are examples we have found particularly useful. Unfortunately the fact that the diagnostics all have the correct values does not prove that there are no systematic errors, just as incorrect values do not mean that the fitted parameters are all wrong.

3. Results

Given the increased computational burden using this method relative to least squares fitting of power spectra, it is obviously useful to see how well the method does on artificial data. For this and other purposes we have developed a program to generate realistic artificial time-series (see Schou and Brown, 1994). Figures !F3 through !F8 illustrate various aspects of the performance of the analysis scheme both on fake data generated by this program and on real data.

As the number of things one can check is very large, only selected results have been shown here. We are currently in the process of systematically analysing artificial data with different analysis programs, including the one presented here, to compare the different algorithms, and plan to publish our results shortly.

Figure !F3 shows how the fitted parameters from a run with artificial data compare to the 'true' input parameters to the fake data generation program.

In addition to the errors in the determined parameters, the observed scatter and the rms errors determined internally in the program from the second derivatives of S have been shown in Fig. !F4. In order to be able to use a significant number of realizations, a low l of 6 was used. This allows the scatter to be determined with a reasonably good accuracy.

Shown in Fig. !F5 are the values of the first diagnostic discussed for a set of artificial data and a set of real data. Figure !F6 shows the change in this diagnostic caused by a change in the assumed apodization.

Values of the second diagnostic discussed are shown in Fig. !F7 for the same sets of artificial and real data as used for Fig. !F5. The change in one of the second type of diagnostics, caused by a change in the assumed PSF, is shown in Fig. !F8. The main change was in the horizontal width of the PSF.

4. Discussion

(i) Performance on real and artificial data

As can be seen from Fig. !F3, the program generally estimates the mode parameters quite accurately, when used to analyze artificial data. There are, however, small but very (statistically) significant errors for the mode amplitudes and linewidths. Note that the length of the timeseries used was such that the each frequency bin had a size of $\approx 0.13\mu Hz$, and that it is therefore not too surprising that it is difficult to estimate the linewidths for the lowest frequency modes where the linewidths are comparable to this bin size. At $1700\mu Hz$, for instance, $w \approx 0.2\mu Hz$, depending somewhat on l. For these modes most of the power

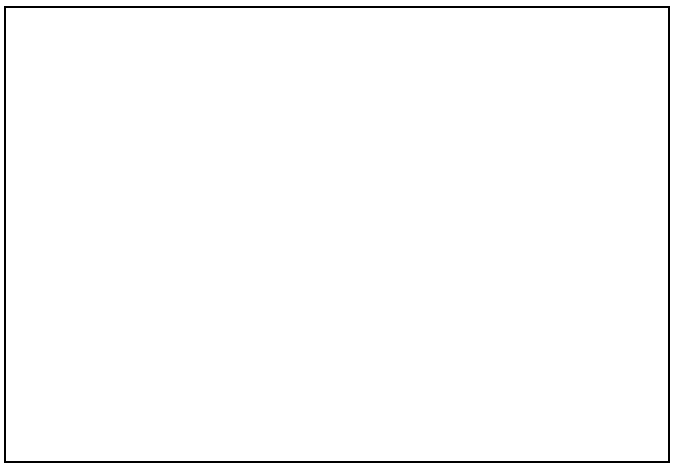

Fig. !F4. The average internally estimated standard errors on the mode frequency compared to the scatter determined from an analysis of 110 time series of artificial data for $l = 6$. Solid lines indicate the average internally estimated standard error for ungapped time-series, + the actual scatter for the ungapped runs, dotted lines the estimates for gapped time series and diamonds the actual scatter for the gapped time-series. The length of the time series and the gaps were as in Fig. !F3. For some modes the analysis program failed for some of the realisations, leading to less than 110 realisations from which to determine the scatter.

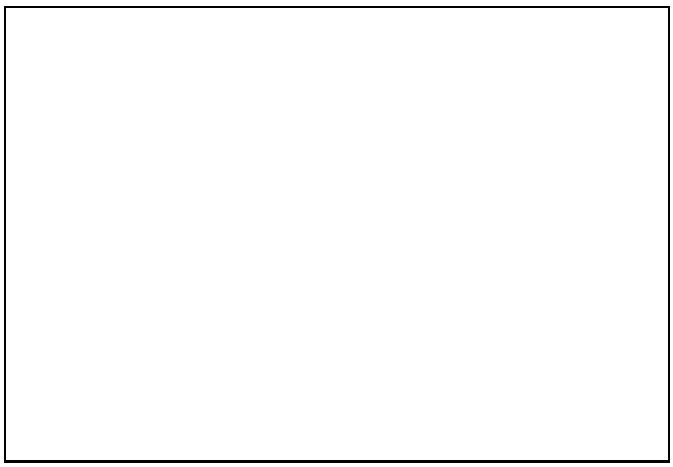

Fig. !F6. The change in the first diagnostic caused by a wrong assumed crosstalk. The original crosstalks assumed (correctly) that the part of the images with a horizontal distance of more than $0.5R$, where R is the radius of the solar image had been cut away. The incorrect crosstalks were calculated assuming that everything outside of $0.6R$ was cut away. Again only the change for $\Delta m = 2$ has been shown. Results are from the data used for Fig. !F5, but using only the first month of data.

is in a few frequency bins close to the mode frequency making it very difficult to estimate the linewidth.

For the high frequency modes, where the linewidths are very large, it was necessary to limit the size of the fitting window to less than several times the linewidths (normally 5 times the linewidth was used). This was done in order to reduce the size of the covariance matrix, and is presumably the main reason for the large errors in parameter values at those frequencies. The most worrisome of these errors is the error in a_1 for the high l, high frequency modes. However, as these modes have been difficult to fit in the real data, presumably due to the increased linewidth and low S/N, a limited effort has until now been put into trying to find the cause of this problem. Clearly it will be necessary to look for the cause of this problem at some point in the future.

The error estimates shown in Fig. !F4 are generally fairly good. As expected (since 65% of the observations have been thrown away) the scatter is generally somewhat larger for the gapped series than for the ungapped series. Even taking into account the fact that only approximately 100 realisations were used, leading to some uncertainties in the estimates of the scatter, it is obvious that the estimates are not perfect, even for the ungapped series. This is not too surprising. One reason is that the second derivatives of the merit function S (the inverse of which is the covariance matrix) is only calculated approximately and normally only in the first iteration (due to the computational expense). Also note that the estimates are generally less accurate for the gapped series than for the ungapped series. A reason for this is probably that the correlations between different frequency points due to the time-gap structure are not taken into account.

Examples of the first type of diagnostic discussed for artificial time-series are shown in Fig. !F5. As can be seen this diagnostic is indeed close to the expected values for the artificial data. Figure !F6 shows the change in this diagnostic caused by a change in the assumed crosstalks (wrong assumed apodization), and as can be seen this diagnostic does indeed indicate that something is wrong. Unfortunately this diagnostic is not too informative as to what is the exact cause of the problems.

Examples of the second type of diagnostic are shown in Fig. !F7. The effect of changing the width of the assumed PSF shown in Fig. !F8 is very obvious at high l. In this case the exact cause of the change is easier to identify. A larger smearing in the horizontal direction preferentially degrades the sensitivity to modes with high m compared to those with low m at a given l, causing a change in the dependence of observed power on m. It is interesting, however, that the changes in some of the parameters, like the mode frequencies and the a-coefficients, are very small, though not totally statistically insignificant.

While it is difficult to look for systematic errors in the real data, where the correct parameter values are unfortunately not known, it is still possible to check whether the diagnostics have the correct values. As can be seen from Fig. !F5 and Fig. !F7, the diagnostics are generally good for the low l's, while there are significant problems for high l's. The likely cause for at least some of these problems is the PSF (compare Fig. !F7 and Fig. !F8). For the analysis shown here, it is known that the PSF was incorrectly estimated; in particular the extra degradation of the PSF caused by an interpolation of the images was ignored.

Unfortunately, as described in Bachmann, Schou and Brown (1993), it still appears that there is some sort of an instrumental/analysis problem with a_1 in the reduction of the real data, the cause of which we have not been able to pinpoint. While the diagnostics shown in Fig. !F5 and Fig. !F7 indicate that some property of the oscillations or the instrument is incorrectly treated, we have not been able to find the cause. Changing the PSF has, as already mentioned, an insignificant effect on the a-coefficients.

(ii) Problems

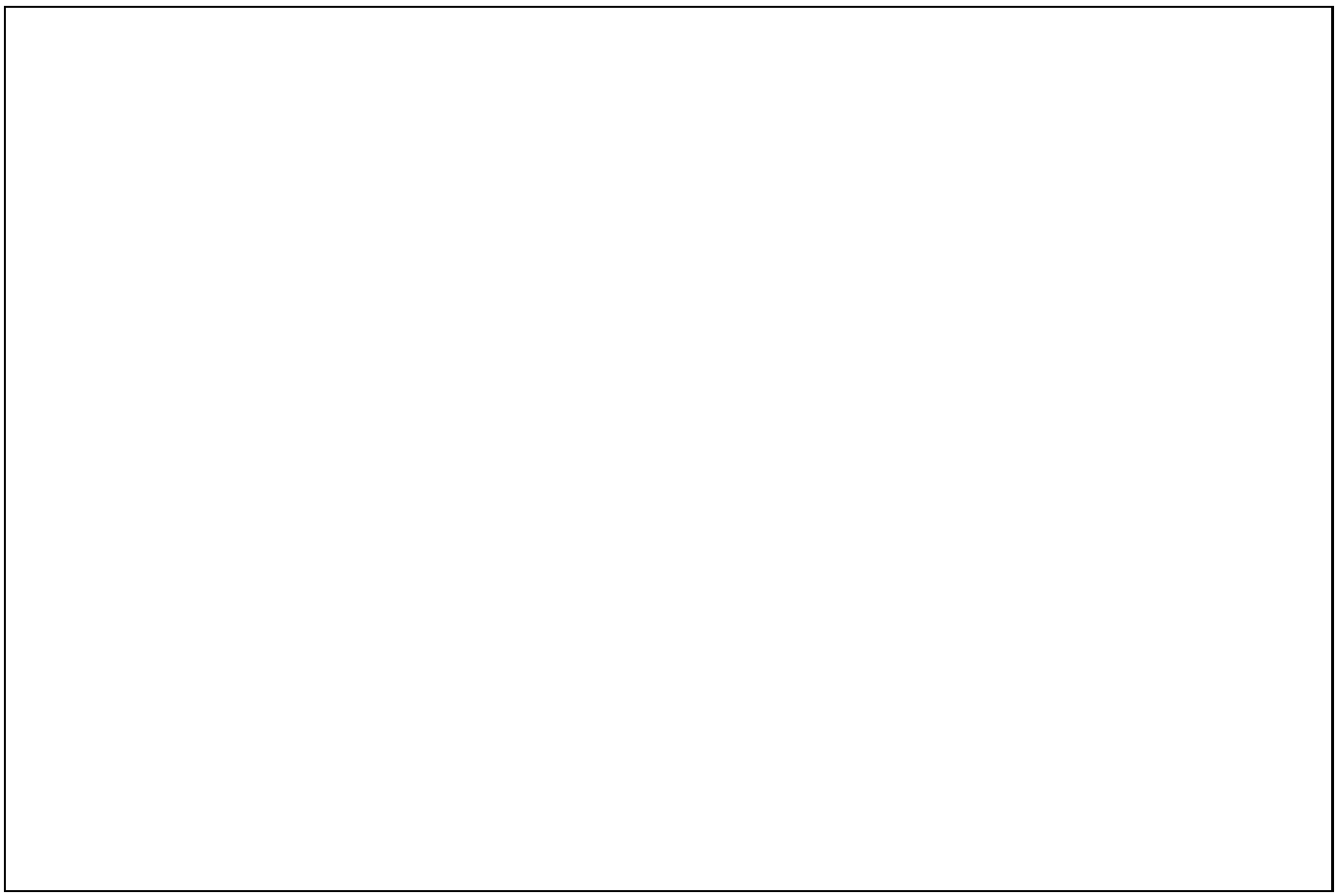

Fig. !F3. Results of runs with artificial data. All results are from a run with $10 \leq l \leq 89$. On the plot of a_1, modes with $10 \leq l \leq 49$ are indicated with $+$, while modes with $50 \leq l \leq 89$ have been indicated with diamonds. The length of each time-series was 100800 timesteps of 75s each. The time gap structure is from a Fourier Tachometer run from 1989 with a duty cycle of 34.94%, and should thus represent realistic single site time gaps.

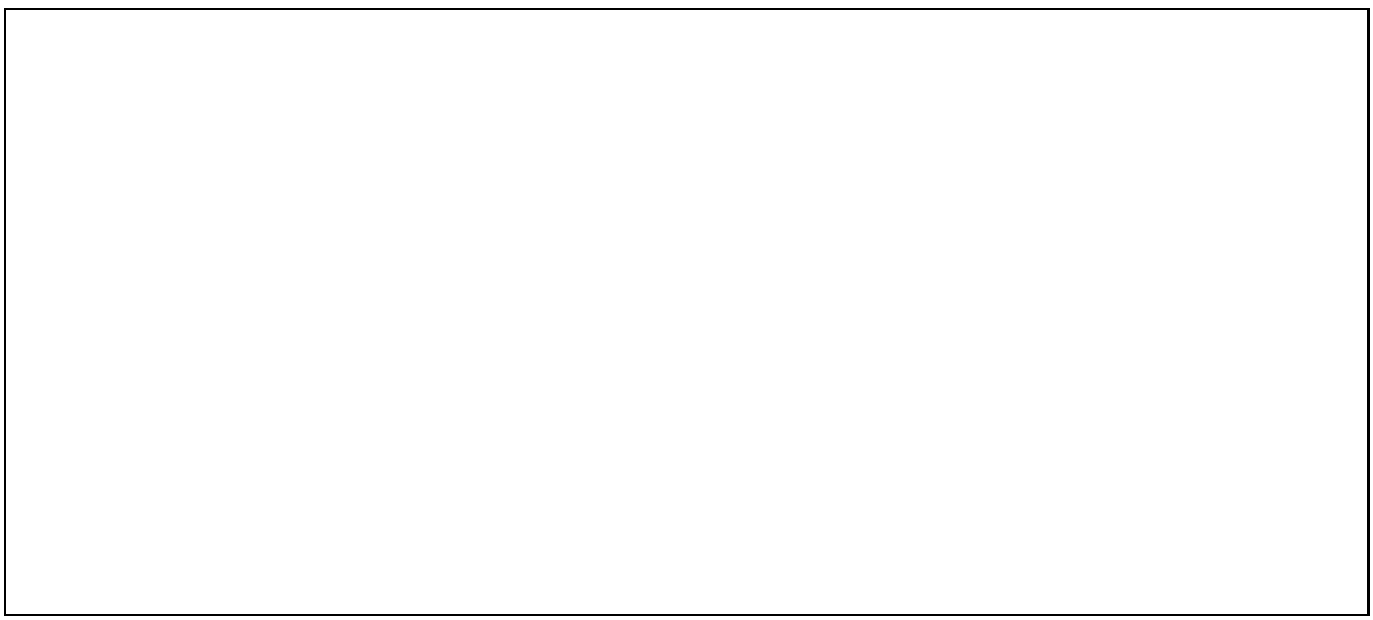

Fig. !F5. Values of the first type of diagnostics discussed for the artificial data from Fig. 3 and for approximately 3 months worth of data from the Fourier Tachometer. Only the values for $\Delta m = 2$ have been shown.

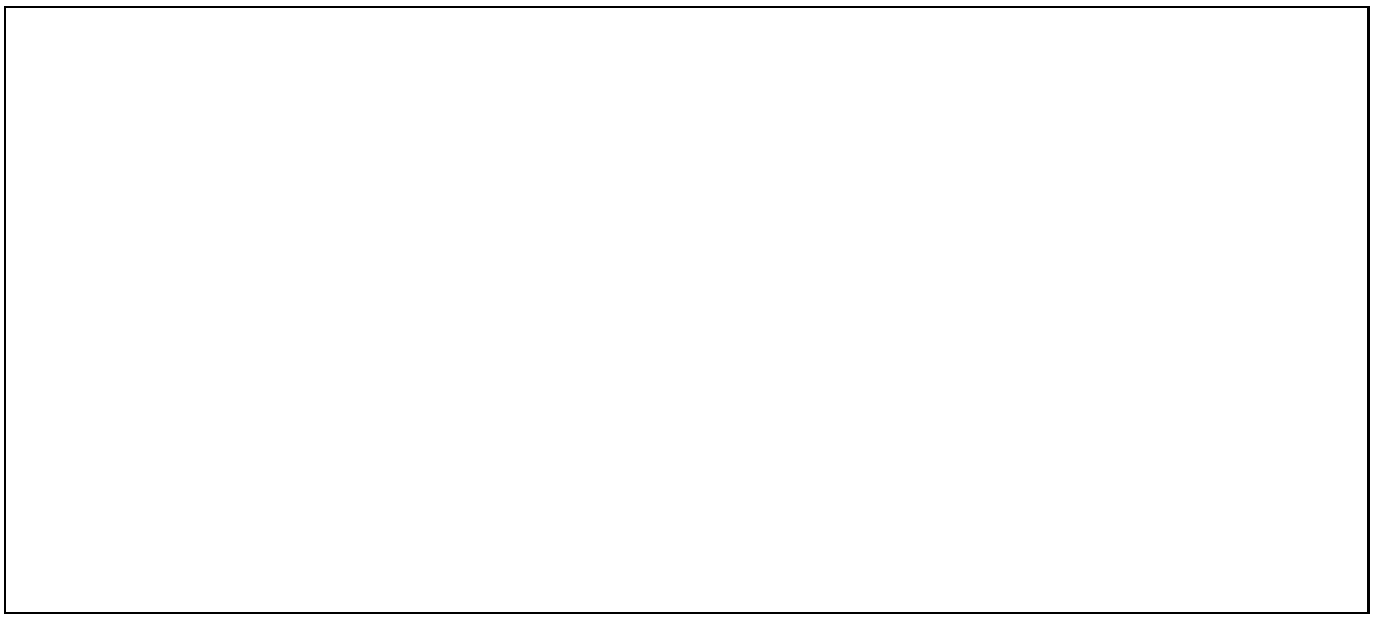

Fig. !F8. The change in the mode amplitude (left plot) and the second diagnostic (right hand plot) caused by a change in the assumed PSF. The main change in the PSF was the horizontal size changed. The data used was that used for Fig. !F6.

As with most fitting procedures, the method described above will fail on occasion. The failures can manifest themselves in a number of ways and have different causes. Perhaps the most common failure is when trying to fit a mode drowned in the noise; these 'failures' tend to be easy to identify since the iteration typically will not converge or will give totally unrealistic mode parameters. A more serious failure mode results when the initial guess is sufficiently wrong that the procedure is not able to find the global minimum in S. An initial guess that is sufficiently bad can also lead to the subsequent guesses being bad enough to give numerical problems. Also, seriously erroneous assumed crosstalks can destabilize the fit. A particularly interesting failure mode results when the uncertainty in the determined frequency is comparable to 1 day^{-1} and/or the spacing to the neighbouring l, and the procedure locks onto one of the temporal sidelobes or the neighbouring l. This failure is generally fairly easy to detect, as the frequencies are usually obviously in error and several of the other parameters have atypical values.

A problem related to the failures is that of numerical problems. As mentioned earlier, the individual m's are so strongly correlated that the covariance matrix is almost singular, at least numerically. Given that only a subset of the m's are used at any given frequency, this is only a problem at high frequencies where the linewidths are large. The present solution, which is to limit the frequency interval to the point of losing information, is not satisfactory. A possible solution is to exclude certain m's; since they are redundant, this procedure should lead to little loss of information and make the covariance less singular. This approach has not yet been tested.

Another way to fix the problem with the near singular covariance matrices would be to 'regularize' them by adding, say, a diagonal matrix with very small elements, this is essentially the same as to add the contribution from roundoff to the matrix, as this should behave as uncorrelated noise. In addition it should also be possible to bin the frequency points if the linewidths are very large, which would reduce the computational burden of taking a large frequency interval into account. These approaches have not been tested either.

(iii) Some possible improvements

As mentioned earlier several features of the real solar time-series have been neglected here. Although some of the neglected features are likely to be unimportant, others may not be. Using artificial time-series (created as described in Schou and Brown, 1994) it has been possible to address, at least partially, some of the concerns, while others are very difficult to test. Also some features are likely to have been overlooked.

One of the main neglected features of the time-series is the correlations introduced by the time-gaps. It is not surprising that this leads to problems with error estimation, as illustrated in Fig. !F4. It may be argued that this is not a very significant problem, however, as it only affects the error estimates. Unfortunately, ignoring the statistical effects can also lead to larger random and systematic errors than could have been obtained with a proper statistical treatment. As taking the correlations properly into account would require that a large number of frequencies and thus m's be considered simultaneously (that is a covariance matrix considering all of them, leading to computational problems), there is probably not much that can be done, unless some computational tricks can be found. For high S/N modes the information in the sidelobes is generally redundant, and one probably does best by only analysing the main lobe. For low S/N modes something can potentially be gained by considering the side-lobes, and one possible shortcut is to analyze the individual peaks separately and form a weighted average, knowing that the error estimates are highly unreliable. This has not been tried.

A problem related to that of correlations introduced by time-gaps is correlations introduced by the possibly infrequent excitation of the modes. This also leads to correlations between different frequency points. Unfortunately the statistics of the excitation are not very well known, although some theoretical

estimates have been made (Brown (1991), see also Brown et al., (1992)). **Fossat ref.** For the present purpose the best assumption is probably that the excitation is very frequent compared to the mode lifetime and that the individual points of the Fourier transform are thus independent. On the other hand the infrequent excitation might actually be used for improving the frequency estimates, as it means that the excitation, and thus the random nature of the Fourier transforms, are not independent for different modes. Also the details of the excitation are of considerable physical interest, as they relate to the physics of the convection.

Clearly the treatment of the crosstalk from modes with other l's is less than ideal. The problem is basically that the initial estimates for some of the other parameters for the other modes may be significantly erroneous. When the main lobe of one of these other modes or one of the temporal sidelobes happens to be close to the frequency of the target mode (which is the case for many l's), this can lead to a significant feedback into the target mode. Some examples of the degree of feedback are shown in Fig. !F9. Note that the changes have been plotted as a function of ν/L, as this determines the frequency spacing between modes with the same n but l differing by 1. Also note that the crossing of the temporal sidelobes (which have a spacing of 1 day^{-1} $\approx 11.57 \mu Hz$) from the neighbouring l's occur at $\nu/L \approx 31 \mu Hz$. As can be seen from Fig. !F9, the feedback is generally fairly small, and the effects can thus, presumably, be eliminated by iterating the process using the parameters determined in one iteration as the initial estimates for the next iteration.

One thing that could be done about the feedback problem, at least for some of the parameters, is to let the parameters for modes other than the target mode follow those of the target mode, in a fashion similar to the one implemented for the mode amplitudes. This is likely to work, as initial parameter estimates for modes with the same n but slightly different l are likely err in the same direction. Unfortunately this means that the contribution from the other modes must be calculated together with the derivatives with respect to the various parameters, which would be very expensive. It may be possible to accurately calculate only the contributions in each iteration, and to do an approximate calculation of the derivatives, but this has not been attempted.

Until now the phase information has only been used to partly compensate for the crosstalk among the m's. It should be potentially useful to also use the phase information to partly compensate for the crosstalk among the l's, in particular for the modes where other l's interfere. Ideally one should fit for all (n, l, m)'s simultaneously, but that is clearly too expensive computationally. Even fitting for a small range in l at a time would be very difficult. There are several problems with trying this. The size of the covariance matrix will increase considerably, leading to a considerably increased computation time and possible numerical problems. Also, it is generally the temporal sidelobes of the neighbouring modes that interfere, rather than the main lobes, which creates some problems. For instance it is necessary to fit over the whole range of all the interfering modes at once, leading to even bigger covariance matrices and the problems associated with them. Finally since the phases and magnitudes are not independent between the main lobe and the temporal sidelobes, the phases of the interfering temporal sidelobes are probably best determined by looking at the main lobe and fitting all the frequencies simultaneously, which (as already mentioned) is practically impossible. It may be possible to take at least some of the contributions from neighbouring modes out by saving some of the statistical information from previous iterations, but this has not been attempted.

A possible way around many of the problems associated with the near singular covariance matrices and the leakage from neighbouring l's is to modify the spatial masks used, so as to reduce the unwanted leakages. This may increase the leakage from non interfering modes or decrease the S/N. This modification of the masks may also be done by combining the time-series or Fourier transforms, although that approach does not give one complete freedom in choosing the effective masks. I have not studied this method in detail, but I am currently looking into it, in particular in connection with low l's, where it is likely to help with occasional interfering modes and where it is more computationally feasible than at higher l's. The higher l's also have the problem that they are generally affected by a fairly large number of other (l, m)'s and their temporal sidelobes, due to the lower frequency spacing between adjacent l's.

Another problem relating to the contribution from modes other than the target mode is that of modes with l far from that of the target mode. These modes have small but, when added up, significant crosstalks into the target modes. As their frequencies and contributions are essentially random, their total contribution has a power envelope like that of the p-modes. The modes therefore lead to what looks like noise with a frequency dependency like that of the envelope of the power distribution in the p-mode band. This may be the cause of the rise observed in the background power in the p-mode band as illustrated in Fig. !F10 (see also Schou and Brown 1994). Modes with m outside the band used for the fitting (and which are currently not used in the calculation of the covariance matrix) may also lead to problems, especially since these modes are typically fairly close in frequency to the modes taken into account. These modes may be the cause of some of the systematic errors in the a-coefficients shown in Fig. !F3. This problem can of course be fixed by taking into account the contributions from more m's, but this is, as the solutions to many of the other problems mentioned, expensive computationally.

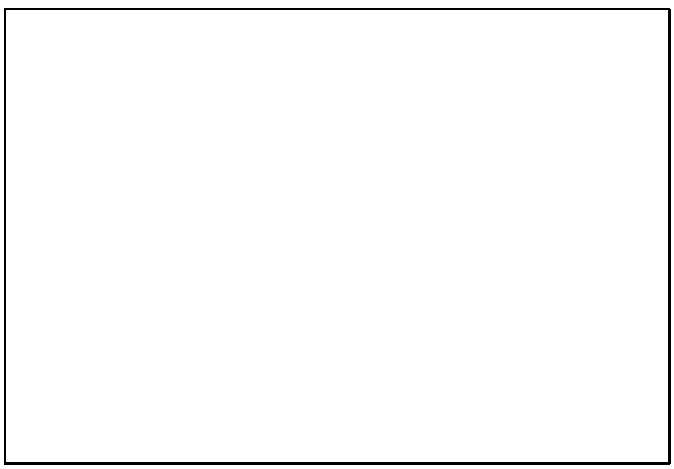

Fig. !F10. The noise as a function of frequency for a dataset from 1989, only $30 \le l \le 50$ has been shown. The two distinct set of points are due to the even and odd l's having different noise characteristics. Only even m's were used in the fitting process.

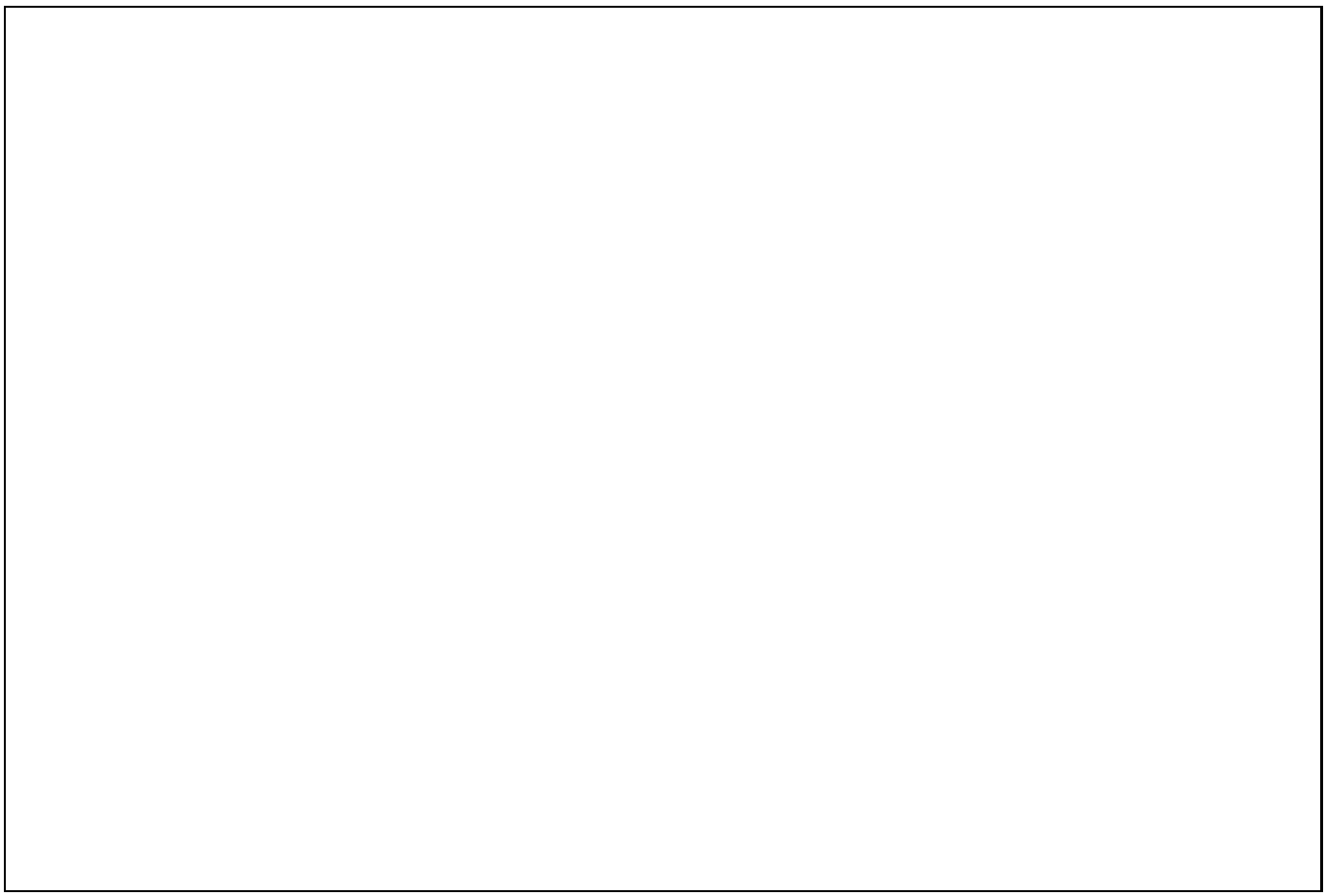

Fig. !F9. Examples of feedback from other modes into the target mode, that is the change induced in the parameter indicated by changing its value for all other modes than the target mode (by an amount shown). The results are for the data used for Fig. !F6. The frequency change has been scaled by $4/w$, where w is the FWHM of the mode, to shown the change caused by a perturbation of $1\mu Hz$.

The problem with the increased noise power in the p-mode band is related to the problem of how to estimate the covariance matrix for the noise. To estimate the properties of the 'true' noise in the p-mode band it is desirable to use a piece of the spectrum without p-modes. This means going to low and/or high frequencies. Given that 'modes' are observed far above the acoustic cut-off frequency and that the Nyquist frequency generally is not very high, it is difficult to use high frequencies. At low frequencies there is presumably only a very low contribution from p-modes, but it is likely that the source of the dominant noise is not the same as in the p-mode band, and that it may therefore have other statistical properties. If, as Fig. !F10 seems to indicate, the noise in the p-mode band is in part due to unresolved modes, then perhaps the noise estimate ought not to be 'mode-free', or even the same for all frequencies. Calculating the noise covariance matrix for the Fourier Tachometer data from a model of the noise on the solar disk has failed, probably due to problems in finding the noise as a function of position on the disk and spatial correlations in the noise. The preferred method is, at the moment, to estimate the noise properties from the high frequency noise.

Obviously the fitting methods can be improved in other directions than the one described here. In particular Duvall et al. (1993) have considered methods in which one fits all frequency points in an interval at a given (l, m) simultaneously, instead of all m's and frequencies in a narrow range, as considered here. One possible variation of this method uses that fact that the spacings between neighbouring modes are well known, and can be held fixed in the iterations or fitted for as a parameter if the frequency interval used is sufficiently large. Although not implemented, the method just described has the advantage that it should be possible to treat the correlations introduced by the time gap structure, particularly if the frequency interval is not too large. On the other hand, while the temporal sidelobes only affect certain modes and will be a minor problem for future observations, such as those from the GONG network, the correlations among m's affect all modes. The magnitude of the leakage between different m's, and l's for that matter, depends on the observation technique and the spatial masks used. In particular Duvall et al. considered intensity observations, which are less affected by these problems than the velocity observations we have considered, and that will be used for future instruments such as the GONG network and the SOI/MDI instrument on the SOHO satellite.

Recent results from Duvall et al. (1993) seem to show that the line profiles are not given by Lorenzians, but are instead somewhat asymmetric. While this obviously changes the function fitted for and the derivatives, it should be possible to incorporate into the procedure presented here.

Acknowledgements. The work reported here was supported in part by the Danish Natural Science Research Council, the Danish Research Academy and the National Aeronautics and Space Administration through grant # W-17, 678.

References

Anderson, E. R., Duvall and T. L. Jr., Jefferies, S. M., 1991, Astrophys. J., **364**, 699.

Bachmann, K. T., Schou, J., and Brown, T. M., 1993, ApJ, submitted.

Brown, T. M., 1991, ApJ, **371**, 396.

Brown, T. M., Bogdan, T. J., Lites, B. W. and Thomas, J. H., 1992, ApJ Letters, **394**, L65.

Duvall, T. L. Jr., Jefferies, S. M., Harvey, J. W., Osaki, Y. and Pomerantz, M. A., 1993, Astrophys. J., submitted.

Libbrecht, K. G., 1989, ApJ, **336**, 1092.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T., 1986, Numerical Recipes, Cambridge University Press.

Ritzwoller, M. H. & Lavely, E. M., 1991, *Astrophys. J.*, **369**, 557.

Schou, J., 1994, in preparation.

Schou, J. and Brown, T. M., 1994, Astronomy and Astrophysics.

Schou, Brown and Bachmann, 1993, in preparation.

Schou, J., Christensen-Dalsgaard, J., and Thompson, M. J., 1994, ApJ.

This article was processed by the author using Springer-Verlag TEX A&A macro package 1991.