Astronomy
&
Astrophysics

Noise reduction in helioseismic power spectra by non-orthogonal wavelets

S. K. Solanki[1], C. Régulo[2,3], M. Fligge[4], and A. G. Kosovichev[5]

[1] Max-Planck-Institut für Aeronomie, 37191 Katlenburg-Lindau, Germany
[2] Instituto de Astrofisica de Canarias, 38205 La Laguna, Tenerife, Spain
[3] Departamento de Astrofisica de la Universidad de La Laguna, 38206 Tenerife, Spain
[4] Institute of Astronomy, ETH-Zentrum, 8092 Zürich, Switzerland
[5] W. W. Hansen Experimental Physics Laboratory, Stanford University, USA

Abstract. We present a method to reduce noise in helioseismic power spectra using a non-orthogonal wavelet transform based on quadratic spline functions. The quality of our method is tested by applying it to artificially generated time-series approximating solar p-modes. The mode frequencies and line widths obtained from least-squares fits to the smoothed spectra are compared with the corresponding parameters deduced from maximum likelihood fits to the original spectra. The results from both approaches are very similar and suggest that there is no major bias in either of these rather independent approaches. As a practical example we denoise parts of the power spectrum obtained from the two first years of operation of the GOLF instrument onboard SOHO.

Key words. methods: data analysis – methods: statistical – Sun: oscillations

1. Introduction

The frequencies of solar p-modes are the main observational data constraining the internal structure of the Sun. The higher the accuracy of these frequencies the more stringent the constraints. In addition to the absolute values of the frequencies the modes of a given degree l are split into their $(2l + 1)$ subfrequencies, each with a different m (azimuthal degree). These splittings provide information on departures from spherical symmetry, primarily due to solar rotation. The reasons for the finite accuracy of the measured p-mode frequencies are manifold. They include 1) incomplete temporal sampling (for Earth-based observations), producing sidebands that can overlap with the primary peaks, 2) a noise background due mainly to convection (and seeing in the case of ground-based data) that varies slowly in amplitude with frequency (Harvey 1985; Pallé et al. 1995), 3) the finite lifetimes of the modes, which broaden the modes and produce noise (stochastic excitation noise) that depends strongly on frequency.

The first source of uncertainty has been significantly reduced by the introduction of ground-based networks, e.g. GONG (Leibacher et al. 1995), IRIS (Fossat 1995), BISON (Chaplin et al. 1996), and to an even larger extent by almost uninterrupted series of data from space, in

particular with the SOHO satellite (Gabriel et al. 1995, 1997; Fröhlich et al. 1997; Scherrer et al. 1997). Going into space has also significantly reduced the background noise, by removing the noise introduced by turbulence in the Earth's atmosphere.

A variety of numerical techniques have been applied to solar oscillation data to reduce the noise from source (2) and in particular (3), which is χ^2 distributed (Anderson et al. 1990; Duvall & Harvey 1986; Toutain & Appourchaux 1994).

Since it is not possible to Fourier smooth a power spectrum obtained by Fourier transformation Elsworth et al. (1994) have reduced noise by specially weighted running means over a few frequency bins. Anguera Gubau et al. (1992) and Jiménez et al. (1994) have taken the approach of breaking the signal into pieces of 2 month each, transforming them and averaging the power spectra. Other techniques include homomorphic deconvolution (applied by Baudin et al. 1993), multitaper analysis and smoothing using orthogonal wavelets (Komm et al. 1998a,b).

In the present paper we smooth using highly redundant non-orthogonal wavelets. These turn out to be well adapted to the signals being studied. One aim of the paper is to see whether it is possible to improve the accuracy of frequencies and other mode parameters by smoothing. Another aim is to compare the results from fits to smoothed and unsmoothed data. Since both types of data have been treated very differently and fits are based on
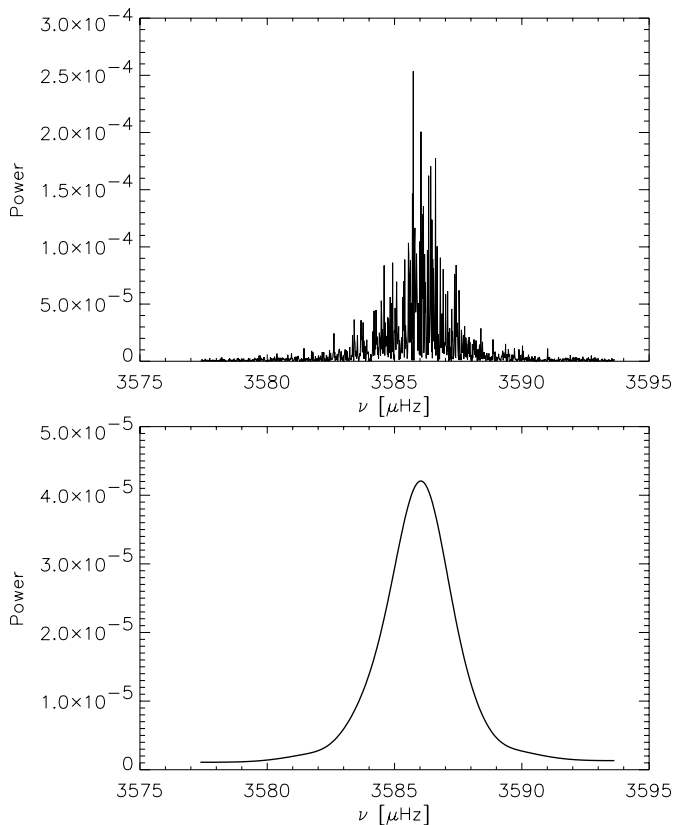

Fig. 1. Noisy model power spectrum (upper panel) and the same power spectrum after wavelet noise reduction (lower panel). The input frequency of the peak is $\nu_1 = 3585.937$ μHz.

different statistics the two approaches are in many ways independent. In this way it is possible to test whether the fits introduce any bias into the results.

2. Technique

The wavelets we use are quadratic spline functions proposed by Mallat & Zhong (1992). The corresponding wavelet transform is a highly redundant representation of the original signal since the analyzing wavelet is translated by the signal's sampling rate only – independent of the frequency scale considered. Fligge & Solanki (1997) used these wavelets in the context of a wavelet packets analysis. They showed that for typical signals resulting from spectroscopy this approach is superior to a whole range of other noise-reduction techniques, including methods employing orthogonal wavelets. We follow that paper closely here and apply the code developed in the course of that investigation to power spectra.

First, the power spectrum to be denoised is wavelet transformed. The used non-orthogonal wavelet transform has the property of transforming Gaussian white noise into a spectrum with increasing power towards higher frequencies thus allowing it to be better separated from the signal. This is a particularly useful property since stochastic excitation noise is actually rather blue, i.e. it corresponds to high-frequency noise in the power spectrum. We expect this property to improve the separation between signal
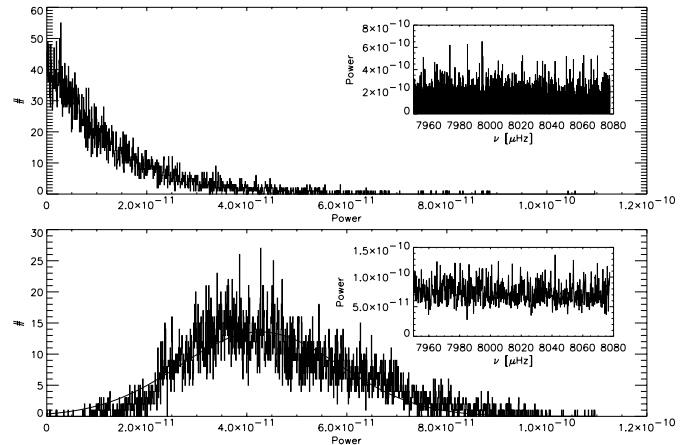

Fig. 2. Upper panel: Histogram of a peak-free region between 7960 μHz and 8080 μHz of the artificial signal. The signal (stochastic excitation noise) follows a χ^2 distribution. Below: "De-noised" signal of the same spectral region. Now, the noise follows a Gaussian distribution much more closely (thick curve).

and noise due to the non-orthogonal wavelet transform. In principle it is possible to apply the powerful wavelet packets-based technique proposed by Fligge & Solanki (1997), but it turns out that the noise is so benign for the p-mode power spectra that simple thresholding after a single wavelet transform gives results that are already very similar to the application of the wavelet packets technique. Thresholding implies that all points in the wavelet transformed signal with an amplitude below the threshold are set to zero.

We have applied this technique to two sets of signals: firstly to the power spectra of model time-series, secondly to the power spectrum obtained from the two first years of the GOLF instrument on SOHO (Gabriel et al. 1995, 1997). The data we used have been calibrated by Ulrich et al. (1998) using the S method. In addition to the full time series we also considered the first and second years separately in order to check whether our technique gives similar results for the two.

3. Tests

To test the efficiency of the denoising technique we carried out a "hare-and-hounds" exercise. One of us (AGK) generated artificial time series and provided them for testing without revealing the oscillation characteristics. These characteristics were determined by denoising the power spectra and fitting line profiles. Then the results were compared with the actual parameters of the artificial data. A standard fitting procedure without denoising was also applied to the artificial data.

The artificial time series were generated using a stochastic model of harmonic oscillations excited by a random forcing function which was represented in the frequency domain by a complex function of frequency with random normally distributed real and imaginary parts.

The amplitude variance was approximated by a smooth Gaussian to mimic the power distribution of solar oscillations. The oscillation frequencies were taken from a standard solar model, and only 29 modes of $l = 0$ in the frequency range from 1 to 5 mHz were included. The line profile of each oscillation mode was assumed to be Lorentzian with the line width approximated by a fourth-degree polynomial of frequency. The frequency resolution was 0.016 μHz. This corresponds to a 728-day time series of 1-min cadence data.

In Fig. 1 we show portions of a power spectrum constructed as described above (upper panel). Also shown is the same spectrum after denoising (lower panel). The reduction in the noise is obvious. Instead of a peak heavily marked by stochastic excitation noise a much cleaner profile is left after the wavelet de-noising. Not just the noise is reduced, however, but the noise statistics are also changed. In Fig. 2 we plot histograms of the noise, i.e. of the power at the frequencies between the modes, at a sufficient distance for the Lorentzian wings to have become very small compared to the noise. In the top panel histograms for the original power spectrum (purely stochastic excitation noise) are shown, in the lower panel for the denoised spectra. Besides the reduction in amplitude the wavelet denoising also makes the noise distribution almost Gaussian (Régulo & Roca Cortés 2000).

Next, Lorentzian fits are made to both the raw and the denoised power spectra. Due to the different noise statistics it is not appropriate to employ the same fitting technique for both spectra. Following standard procedure we fit the original spectrum employing the maximum likelihood technique. For the de-noised spectra, due to the nearly Gaussian statistics of the remaining noise, least-squares fits were deemed more appropriate.

The central frequencies of these Lorentzians, ν_i and their width, Γ_i are finally compared with the corresponding parameters underlying the original signal. In Fig. 3 we plot versus frequency the difference between the synthetic original frequencies and the frequencies derived from both fits, maximum-likelihood fits to the original spectrum (asterisks) and least-squares fits to the smoothed spectra (squares). On the whole both techniques give results that are approximately equally accurate. The error bars (1 σ errors), however, are very different in magnitude. They appear appropriate for the maximum-likelihood fits, but are too small in the case of the least-squares fits. As can be seen from Fig. 4 the linewidth values obtained from the de-noised spectra (squares in the figure) are sightly higher than those obtained from the original spectra (asterisks in the figure), due to the broadening of peaks associated with practically any sort of noise-reduction procedure.

4. Sample application to solar data

An application to real data is shown in Fig. 5. The raw power spectra of $n = 13$, $l = 0$ and $n = 12$, $l = 2$ (left) and $n = 12$, $l = 1$ (right) are plotted in the upper row. In the central row the corresponding denoised spectra (thick lines) are exhibited. Overplotted on these are Lorentzian fits (thin lines). Finally, in the lowest row the denoised power spectra for the first (thick lines) and second (thin lines) years of GOLF operations are plotted separately (year 1: 11 April 1996–14 March 1997, year 2: 15 March 1997–15 February 1998). All these fits are symmetric, but as was shown by Régulo & Roca Cortés (2000), although the p-modes profiles present a slightly negative asymmetry, the use of symmetric profiles to fit the power spectra does not introduce any systematic shifts in the obtained frequencies.

The most striking feature of these plots, besides the much cleaner appearance of the denoised power spectra, is the fact that the various denoised power peaks show the correct number of sub-peaks expected from their degree (i.e. $l + 1$ sub-peaks). Thus we expect that the conclusions drawn from de-noising the test spectrum for the mode frequency also apply to the splittings of individual modes. This conclusion is supported by the fact that on the whole the splittings suggested by the two years are similar. At least a part of the differences between the peaks for the two years are probably due to stochastic differences between the excitations and thus reveal typical uncertainties. Changes due to, e.g., the evolving magnetic activity of the Sun may, however, also contribute.

An interesting result related with the linewidth of the modes is shown in Fig. 6, where a comparison between the width obtained after denoising observed data and the one obtained using the standard method (Bertello, private communication) shows that the former seems to work better with observed than with simulated data, at least in the sense that the line widths are closer to those obtained from the untreated data. It gives larger widths only at very low frequencies. There is also a slight dependence on the l value, the maximum likelihood method giving higher widths for $l = 2$ and $l = 3$. Note that the use of randomly distributed real and imaginary parts in the frequency domain for artificial data does not correspond to the phase structure of a real spectrum and may explain why non-orthogonal wavelets may work better with real data than with our artificial data.

We have also first multitapered the signal, before denoising the resulting power spectrum. According to Komm et al. (1998a,b) the combination of multitapering and *orthogonal* wavelet denoising gives better results than applying wavelet denoising to a normally Fourier transformed signal. We found, however, that the results obtained from applying the *non-orthogonal* wavelet transform are practically independent of whether a multitaper is previously applied or not. We actually favour using this technique without prior application of multitapers, since otherwise broader peaks are produced.

5. Conclusions

We have applied a noise-reduction technique based on the application of a non-orthogonal wavelet transform to helioseismic power spectra. The resulting de-noised spectra
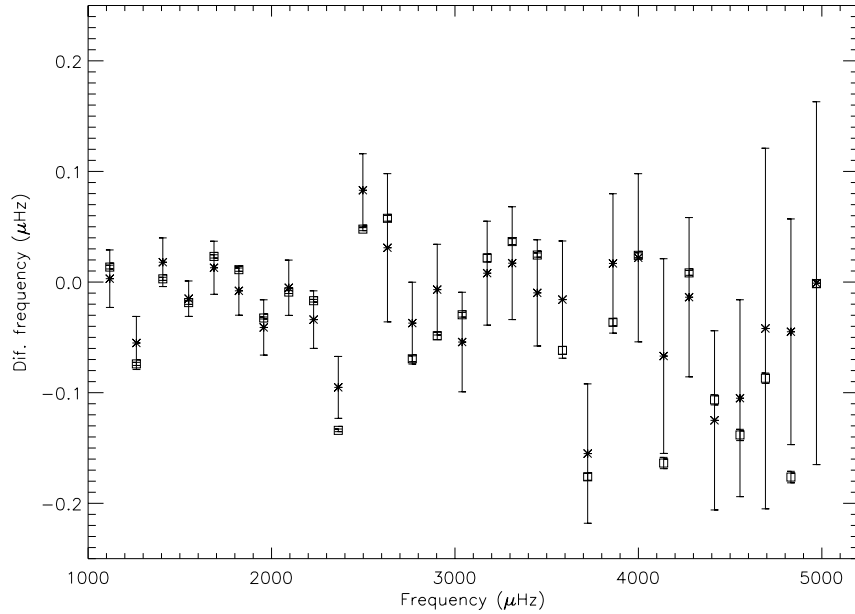
Fig. 3. Difference between the synthetic original frequencies and the frequencies derived from the fits. Asterisk are obtained when the standard fit is used and squares from least-squares fits to the denoised data.
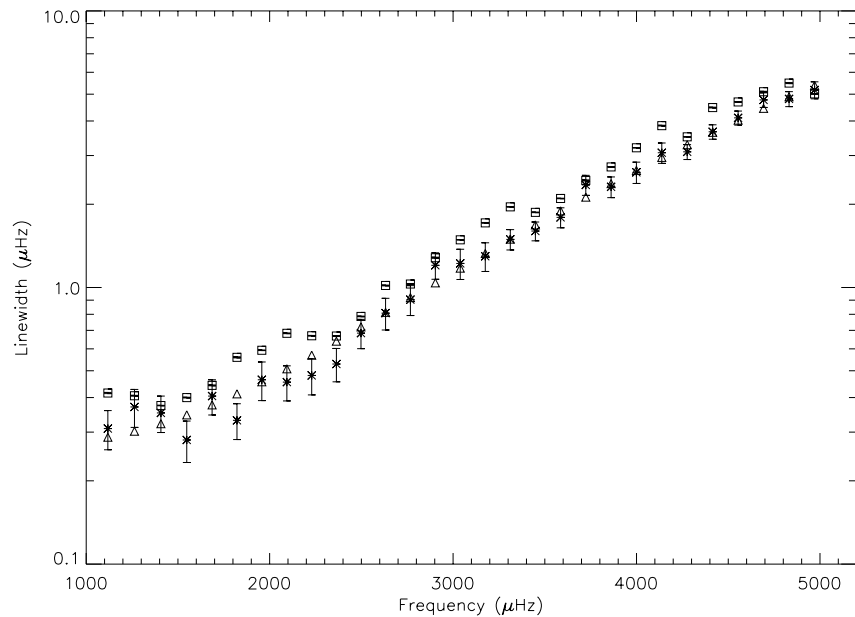

Fig. 4. Linewidth obtained from both fits, maximum-likelihood fits to the original spectrum (asterisks) and least-squares fits to the smoothed spectra (squares). As a reference, the synthetic original linewidths are plotted too (triangles).

are very smooth and individual modes show the number of sub-peaks expected. The noise statistics are changed from χ^2 to nearly Gaussian, so that least-squares fits to the de-noised spectra become possible.

The mode frequencies obtained from the fits to the de-noised and the original power spectra are of similar accuracy. Thus, it does not appear possible to improve significantly on the accuracy achieved with maximum likelihood fits by reducing noise using non-orthogonal wavelets. On the basis of the results of Fligge & Solanki (1998) we expect that orthogonal wavelets do not give more reliable

results than the non-orthogonal wavelets. Note that Komm et al. (1998a,b) have not compared frequencies, nor have Baudin et al. (1993), who applied homomorphic deconvolution. Since the two fitting techniques are very different the similarity between the frequencies obtained using them is heartening. Thus a comparison of the frequencies obtained with both techniques applied to solar data, such as the power spectrum recorded by GOLF, would provide an independent test of the uncertainties in these parameters. Such an investigation is planned.
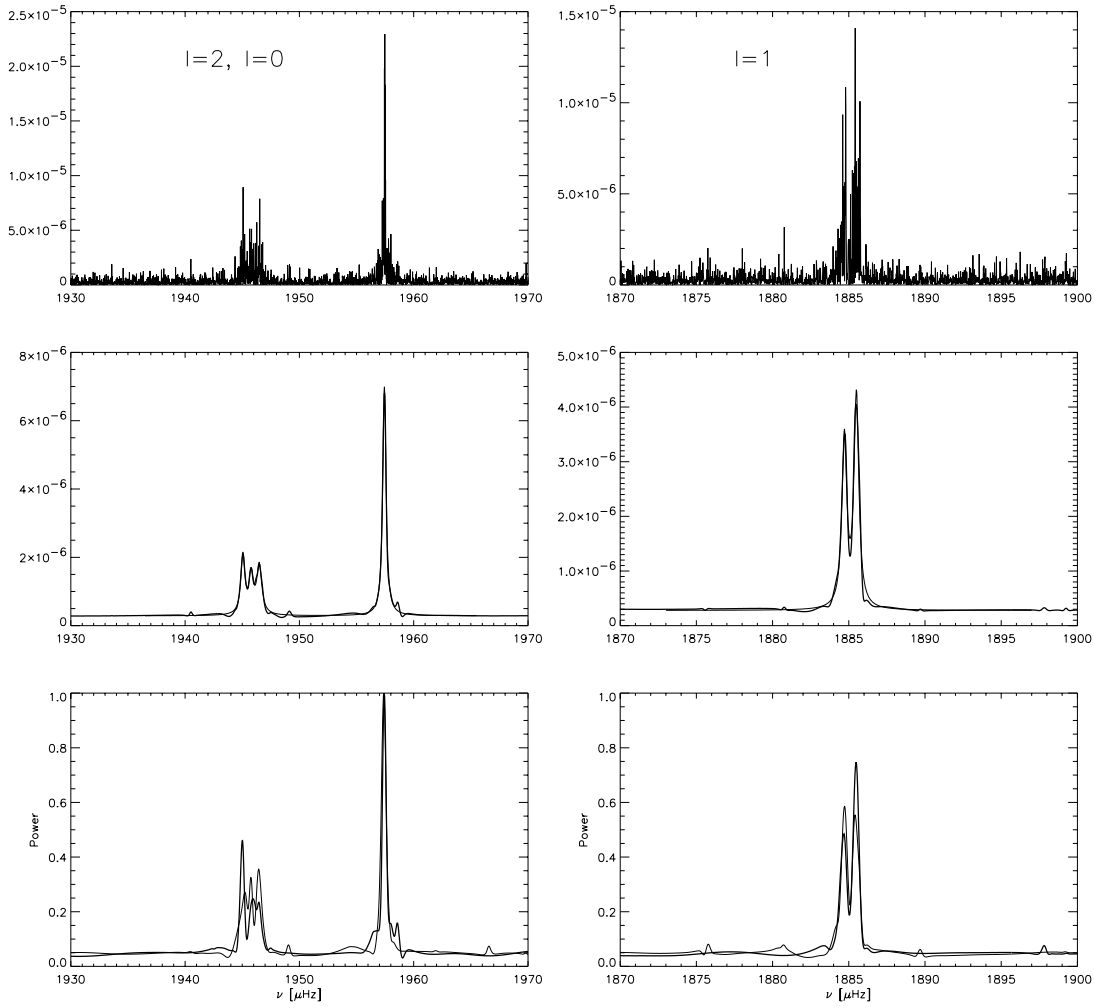
Fig. 5. Power spectra of $n = 13$, $l = 0$ and $n = 12$, $l = 2$ (left frames) and $n = 12$, $l = 1$ (right frames) obtained from two years of GOLF data. Top row: raw data, central row: after wavelet-noise reduction (thick lines), with Lorentzian fits (thin lines), bottom row: separate denoised power spectra of the first (thick lines) and second (thin lines) years of GOLF. These power spectra have been normalized such that the $n = 13$, $l = 0$ mode has unit amplitude.
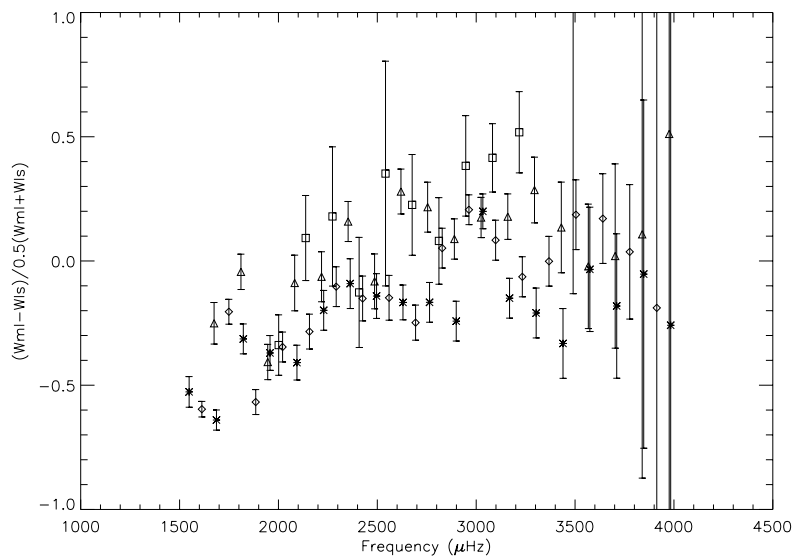

Fig. 6. Relative difference in the linewidth of the solar p-modes obtained from the denoised spectrum followed by a non linear least squares fit (Wls) and the linewidth of the same p-modes obtained from the original spectrum by a maximum likelihood fit (Wml). Asterisks represent $l = 0$ p-modes, diamonds $l = 1$, triangles $l = 2$ and squares $l = 3$ p-modes.

References

Anderson, E. R., Duvall, T. L., & Jefferies, S. M. 1990, ApJ, 364, 699

Anguera Gubau, M., Pallé, P. L., Pérez Hernández, F., Régulo, C., & Roca Cortés, T. 1992, A&A, 255, 363

Baudin, K., Gabriel, A., & Gildert, D. 1993, A&A, 276, L1

Chaplin, W. J., Elsworth, Y., Howe, R., et al. 1996, Sol. Phys., 168, 1

Duvall, T. L., & Harvey, J. W. 1986, in Seismology of the sun and the distant stars, Proc. of the NATO Advanced Research Workshop (Cambridge, England, Dordrecht, D. Reidel Publishing Co.), 105

Elsworth, Y., Howe, R., Isaak, G. R., et al. 1994, ApJ, 434, 801

Fligge, M., & Solanki, S. K. 1997, A&AS, 124, 579

Fossat, E. 1995, in GONG '94: Helio- and Astero-Seismology from the Earth and Space, ed. R. K. Ulrich, E. J. Rhodes Jr., & W. Dappen, PASP Conf. Ser., 76, 387

Gabriel, A. H., Charra, J., Grec, G., et al. 1997, Sol. Phys., 175, 207

Gabriel, A. H., Grec, G., Charra, J., et al. 1995, Sol. Phys., 162, 61

Harvey, J. 1985, in Future Missions in Solar, Heliospheric and Space Plasma Physics, ESA, 199

Jiménez, A., Pérez Hernández, F., Claret, A., et al. 1994, ApJ, 435, 874

Komm, R., Anderson, E., Hill, F., et al. 1998a, in Structure and Dynamics of the Interior of the Sun and Sun-like Stars, SOHO6/GONG98 Workshop Abstract (Boston, Massachusetts)

Komm, R., Gu, Y., Hill, F., Stark, P., & Fodor, I. 1998b, in The Tenth Cambridge Workshop on Cool Stars, Stellar Systems and the Sun, ed. A. Donahue, & J. A. Bookbinder, ASP Conf. Ser., 154, 783

Leibacher, J., & the Gong Project Team, 1995, in GONG '94: Helio- and Astero-Seismology from the Earth and Space, ed. R. K. Ulrich, E. J. Rhodes Jr., & W. Dappen, PASP Conf. Ser., 76, 381

Mallat, St., & Zhong, S. 1992, IEEE Trans. Pattern Anal. and Machine Intell. 14, 710

Pallé, P. L., Jiménez, A., Pérez Hernández, F., et al. 1995, ApJ, 441, 952

Régulo, C., & Roca Cortés, T. 2000, Sol. Phys., in press

Toutain, T., & Appourchaux, T. 1994, A&A, 289, 649

Ulrich, R. K., Garcia, R. A., Robillot, J. M., et al. 1998, in Structure and Dynamics of the Interior of the Sun and Sun-like Stars, SOHO6/GONG98 Workshop Abstract (Boston, Massachusetts), 353