

HELIOSEISMIC DATA REDUCTION

F. HILL

National Optical Astronomy Observatories

National Solar Observatory

PO Box 26732

Tucson, AZ 85726-6732, USA

1. Introduction

Helioseismology seeks to infer the properties of the solar interior using measurements of the global normal mode oscillations as a function of spherical harmonic degree ℓ , azimuthal order m , and radial order n as observed on the solar surface. The frequencies, $\nu_0(\ell, m, n)$, of the modes are influenced by the physical conditions of the solar plasma through which the p-mode (pressure) waves propagate, while the power, $P(\ell, m, n)$, and line widths, $\Gamma(\ell, m, n)$, provide clues about the excitation and damping of the oscillations. These mode parameters are extracted from observations of the solar surface using a long and complex data reduction procedure. It is thus important to precisely describe the steps in the reduction and to assess the influence of the choices on the resulting inferred physical conditions.

A comprehensive and detailed discussion of the reduction process is outside the scope (and page limit) of this review, but can be found in Hill *et al.* 1998. Thus, this paper will provide only a top-level overview of the processing chain, and will then focus on two topics which significantly affect the inferred internal solar structure and dynamics. The first topic is the determination of the geometry of the images, in particular the estimation of the radii and the position of solar north of the images that provide the basic data. The second topic is the estimation of the mode parameters from the multi-dimensional spectra that emerge from the basic reduction processing.

2. Processing Overview

A conceptual outline of helioseismic data processing is shown in Figure 1. The core step of the reduction is the decomposition of a time series of images of the solar surface into a set of spherical harmonics. The observations are typically obtained in either Doppler velocity or intensity; here Doppler velocity images from the GONG project are used to illustrate the process. The decomposition can be expressed mathematically as a matrix multiplication:

$$[V_t] \times [Y_{\ell,m}] = [A_{t,\ell,m}]$$

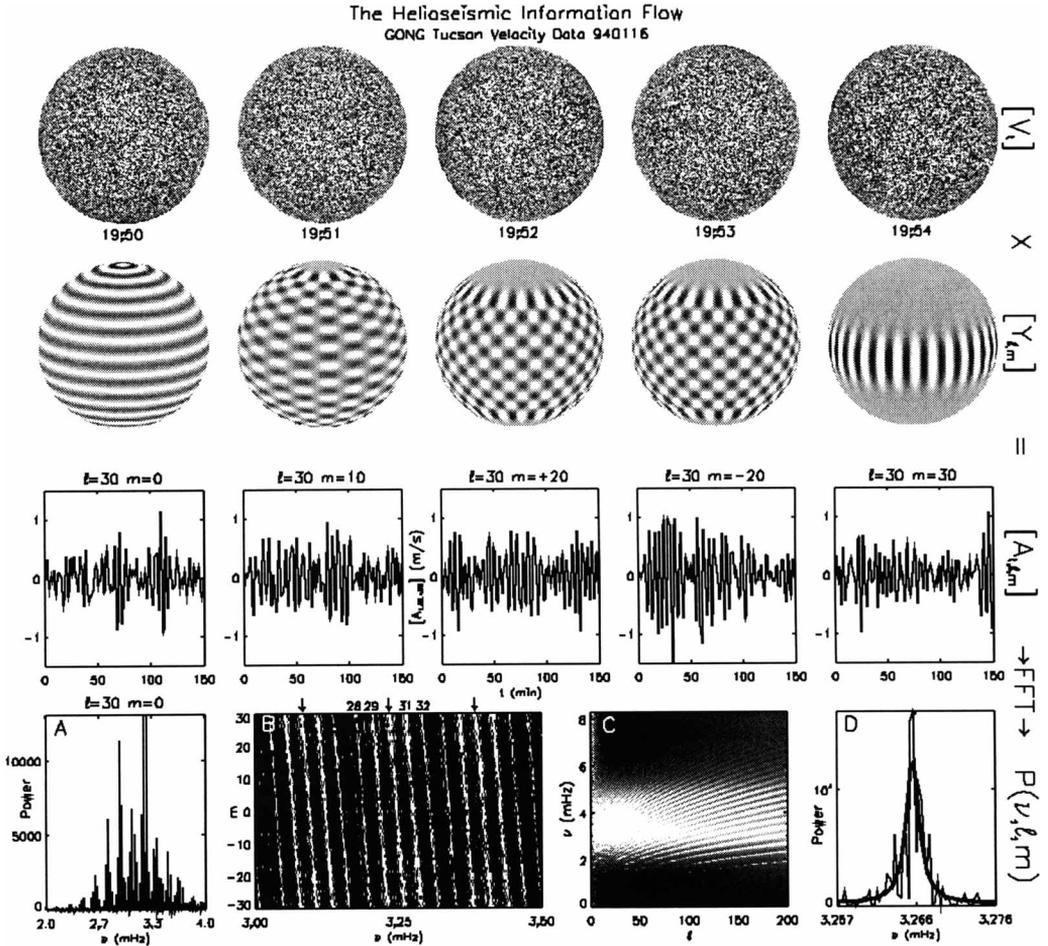


Figure 1. A conceptual outline of helioseismic data processing

where $[V_t]$ is a vector representing a series of velocity images V obtained at a set of times t ; $[Y_{\ell,m}]$ is a matrix of spherical harmonic functions for a set of ℓ and m values with

$$Y_{\ell,m} = P_{\ell,m}(\cos \theta) e^{im\phi},$$

$P_{\ell,m}(\cos \theta)$ are the Associated Legendre functions, θ is the heliographic colatitude, ϕ is the heliographic longitude; and $[A_{t,\ell,m}]$ is a matrix of spherical harmonic coefficient time series.

In Figure 1, the top row shows a small five-minute portion of GONG Doppler velocity images to illustrate $[V_t]$. These images have been high-pass temporally filtered

to remove non-p-mode velocities, and show the typical center-to-limb variation of the primarily radial oscillatory velocity field. The second row provides images of sample spherical harmonic patterns $[Y_{\ell,m}]$ for $\ell = 30$, with a zonal mode ($m = 0$), a sectoral mode ($m = 30 = \ell$), and three tesseral modes ($0 < |m| < \ell$). Two of the tesseral modes differ in the sign of m , and thus in the spatial phase of the harmonic. The third row show small portions of the five time series $[A_{t,\ell,m}]$ corresponding to the spherical harmonics in row 2.

In practice, the decomposition is not performed as a matrix multiplication. Since the individual image geometry varies temporally, this would require the generation of a complete set of spherical harmonic images for each image. It is much more computationally efficient to remap the input images onto a heliographic θ, ϕ grid, perform a fast Fourier transform (FFT) in the ϕ direction to obtain the m decomposition, and then apply a one-dimensional $P_{\ell,m}$ transform in the θ direction.

After the images have been projected onto the set of $Y_{\ell,m}$ functions, a temporal FFT is applied to each time series to produce a power spectrum $P(\nu, \ell, m)$. The bottom row of Figure 1 shows four representations of these spectra. Panel A is the one-dimensional slice $P(\nu, 30, 0)$. Panel B shows the two-dimensional image $P(\nu, 30, m)$ constructed from a set of one-dimensional slices. This panel shows three sets of slanted ridges with a central arrow indicating the ridge corresponding to the target spherical harmonic $\ell = 30$. Each ridge set corresponds to a specific value of n . The shape of the ridges is a depth-averaged measurement of the rotation of the solar interior. Most of the individual ridges within each set are spatial sidelobes that arise from the spherical harmonic decomposition of the data over a restricted area on the sphere. This results in the leakage of power from ℓ values adjacent to the target ℓ . In the figure, each n set contains five ridges, corresponding to $\ell = 28, 29$, the target 30, 31, and 32; these are labeled in the central set.

Panel C in the bottom row is a two-dimensional spectrum constructed by averaging for each ℓ an image as seen in Panel B along the ridges (so-called m averaging). This results in a set of one-dimensional spectra $P(\nu, \ell)$ which are then combined to make the image known as an $\ell - \nu$ diagram in Panel C. Each of the curved ridges in this diagram corresponds to a single value of n .

Panel D isolates a single peak in the p-mode spectrum at $\ell = 30, m = -30$. The spectral line profile shows considerable jagged structure due to the stochastic driving of the solar oscillations. Superimposed on the data is an estimate of the smooth limit spectrum obtained from a maximum-likelihood fit to the data. The parameters of this estimate, in particular the central frequency ν_0 , are then used to infer the internal solar structure and dynamics.

It is evident that there is a long and complex sequence of data reduction steps between the observations and the inferred internal conditions. It is thus not surprising that the data reduction can substantially influence the final results. A few of these effects are discussed in the next sections.

3. Image Geometry

The geometry of the full-disk images used for helioseismology varies continuously. Terrestrial atmospheric refraction imposes an essentially elliptical shape with a diurnal variation in major axis, minor axis, and angular orientation. The image generally moves and rotates on the detector. Finally, the annual orbital motion of the earth creates temporal variations in the solar B_0 angle (the inclination of the solar rotation

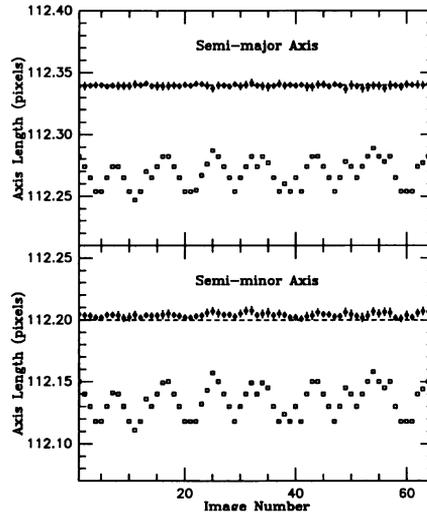


Figure 2. The image dimensions from the Hankel zero routine (filled circles with $1 - \sigma$ error bars) and the Laplacian limb fit (open squares) compared to the true dimensions (dashed lines) for a time series of artificial images with varying MTF.

axis with respect to the ecliptic), P_0 angle (the angle between the solar rotation axis and terrestrial north), and apparent radius. These changes are accounted for in the remapping stage of the spherical harmonic decomposition, but accurate measurements of the geometry of each individual image are essential to reduce subsequent errors in the analysis. Two of the most critical measurements are the apparent radii of the elliptical image, and P_0 , the position angle of the solar rotation axis.

3.1. RADII ESTIMATION

It is essential that the scale of ℓ of the oscillation spectrum be accurately calibrated so that the inversion techniques that infer the internal structure correctly associate the kernels with the data. In turn, the ℓ scale is calibrated by measuring the radii of the images. Then, since ℓ is proportional to the solar radius, an error in the radius is linearly related to an error in ℓ .

The measurement of the apparent semi-major and semi-minor axes of the observed elliptical image requires both a definition of the solar limb, and a determination of its position. Usually, the limb is defined as the zero crossing of the second derivative of the solar intensity profile and is frequently determined by computing the Laplacian of the image. Tests with simulated images show that this method systematically underestimates the radii by a part in 10^3 . A more sophisticated method based on the Hankel transform of the image has been developed by Toner and Jefferies (1993).

As seen in Figure 2, this method reduces the systematic error by an order of magnitude. In addition, the method also provides an estimated modulation transfer function (MTF), and a measurement of the limb-darkening function of the images. The MTF provides valuable information on the scattering, seeing, and focus quality of each image, and is an essential ingredient of the GONG merging scheme.

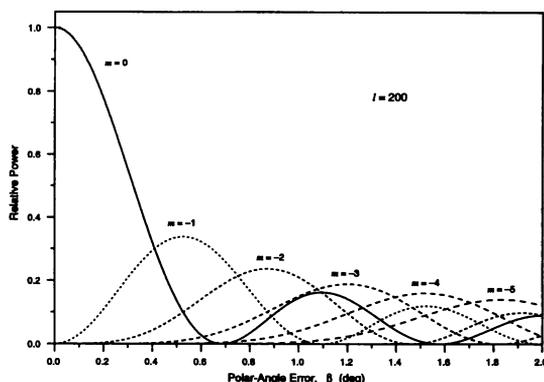


Figure 3. The relative redistribution of power from the $\ell = 200$, $m = 0$ mode into other values of m as a function of the error β in the position of the solar rotation axis.

3.2. P_0 ESTIMATION

The amplitude of the spherical harmonic transform $A_{t,\ell,m}$ for a given image at time t depends on the the coordinate system θ, ϕ . A rotation of this coordinate system changes the relative amplitudes as a function of ℓ and m . If the sun did not rotate, then the oscillation frequencies would be degenerate in m , the well-known transformation of spherical harmonic amplitudes between rotated coordinate systems could be applied, and there would be no need to determine P_0 .

The sun does indeed rotate, and the dependence of the rotation as a function of depth and latitude is of fundamental importance for astrophysical fluid dynamics. It is thus crucial to determine P_0 . Figure 3 illustrates the redistribution of power in the $\ell = 200$, $m = 0$ mode as a function of the error β in P_0 (Kennedy 1997). This figure shows a β of only 0.2° can reduce the power in the mode by 30%. The effect increases with ℓ .

The effect can be detected by reducing a set of images assuming different values of P_0 (Kennedy and Williams 1997), and it may be possible to eventually refine Carrington's solar rotation elements using helioseismic data. For a helioseismic network, the relative angular orientation of the images can be determined from cross-correlations of simultaneous images, and an optimal set of rotations found around the network. The absolute orientation can then be set by a drift scan at any station in the network (Toner and Harvey 1997).

4. Peak Fitting and Mode Parameter Estimation

The extraction of the mode parameters from the multi-dimensional spectra is the last step in the reduction of helioseismic data. It is also the least well-defined, and currently an area of vigorous research. There are several factors that complicate this step:

1. The stochastic excitation creates a spectrum with noise that does not decrease with increasing time series length
2. The spatial leakage produces a large number of extra features in the spectrum
3. The mode parameters have a large temporal variation over the solar cycle

4. The line shape is not symmetric and the underlying physical model is not yet understood
5. The spectrum is qualitatively different when observed in velocity or intensity

The stochastic excitation of the modes adds noise to the observed spectrum as seen in Panel D of Figure 1. This random process dominates the spectrum, and is proportional to the power in a particular frequency interval. The underlying statistical distribution of the process is not Gaussian, but is instead approximates a χ^2 distribution with two degrees of freedom that must be explicitly incorporated into a fitting procedure that seeks the limit spectrum of the oscillation also seen in Panel D of Figure 1. A maximum likelihood minimization method must be used rather than the more common least-squares approach (Anderson, Duvall, and Jefferies 1990).

Spatial leakage was mentioned in Section 2 and can be seen in Panel B of Figure 1. It arises because the spherical harmonics are not orthogonal over a portion of a sphere, that is

$$\int_{\phi_1}^{\phi_2} \int_{\theta_1}^{\theta_2} Y_{\ell,m}(\theta, \phi) Y_{\ell',m'}^*(\theta, \phi) d\theta d\phi \neq 0$$

when $|\phi_2 - \phi_1| < 2\pi$, or $|\theta_2 - \theta_1| < 2\pi$. In addition to leakage in ℓ , there is leakage in m which introduces a spurious correlation between modes with nearly equal m . Finally, there can be n leaks, where the peak from a mode with a different n (and hence a very different ℓ) actually overlaps the target peak. This complex leak structure must be incorporated into the model of the spectrum that is fitted to the data, and the calculation of the relative amplitudes of the spatial leaks for a specific observational set is a large computational task in itself (Schou 1992, Howe and Thompson 1997). The leakage problem was even more complicated prior to the space and network observing strategies that eliminated additional temporal sidelobes. There is some evidence that the leakage matrix is most effectively used in fitting the complex Fourier amplitudes of the oscillations rather than the power spectrum (e.g. Schou 1992; Appourchaux, Gizon and Rabello-Soares 1997). However, these methods currently do not produce frequencies for individual values of m but instead project the m dependence onto a set of orthogonal polynomials.

Precise measurements of $\nu_0(\ell, m)$ are essential for further advances in helioseismology, and these in turn require very long time series of observations spanning several years. During these long periods the values of ν_0 (as well as P and Γ) change as the sun progresses through its activity cycle, degrading the precision of the measurements. The cycle dependence of the mode parameters is itself an active area of research.

The underlying functional form of the oscillation limit spectrum provides another complication. Virtually all fitting methods have so far assumed that the spectrum has a symmetrical Lorentzian line profile arising from the stochastically excited damped harmonic oscillator model of the oscillations. However, observations clearly show that the peaks in the spectrum are actually asymmetric in shape as a function of ν . The cause of this asymmetry is thought to lie in the physical details of the source mechanism (Jefferies 1997). At the time of this review, the correct physical mechanism is a research subject. Since it has been demonstrated that fitting the wrong line shape to the data results in systematically biasing the measured frequencies, it is feasible that prior results of helioseismology may need revision in the light of more accurate models of the oscillation spectrum.

Another indication that earlier results may need revision lies in the recent discovery of the qualitative difference in oscillation spectra obtained simultaneously in Doppler velocity and total intensity (Harvey *et al.* 1997). GONG spectra in these

two observables show very large (as much as $50 \mu\text{Hz}$) differences in the apparent positions of the oscillation ridges above 5.5 mHz . In the 3-mHz band, the differences are smaller (a few tenths of a μHz) but still significant. Since the modes must have the same central frequency regardless of the physical variable they are observed in, the definition of a mode frequency is brought into question. The explanation of the differences probably lies again in the theory of the mode excitation, which also must explain the observed flip in the sense of the asymmetry between oscillation spectra observed in velocity and intensity.

It is evident that there are several difficult issues that complicate mode parameter estimation, arguably the most critical link in the helioseismic data reduction chain. The development of methods to improve this step is proceeding along a number of paths. More accurate calculations of the leakage matrix, including the contribution from the small but non-negligible horizontal component of the oscillatory velocity field, are underway. Fitting methods that consider more than one spectral dimension and instead simultaneously fit in both ℓ and ν have been successfully used. These methods are being extended to the much more challenging full 3-dimensional fit. The physical model describing the line shapes and background is currently undergoing rapid development. It is likely that simultaneous fitting of both the velocity and intensity spectra, as well as the $(V - I)$ phase difference spectra, will be necessary to get the correct frequencies.

Substantial progress has been made in applying more advanced spectral analysis methods to helioseismic data. These methods seek to reduce the realization noise component of the spectra that arise from the stochastic excitation. Homomorphic deconvolution promises to actually separate the realization spectrum from the limit spectrum (Baudin and Hill 1997). Multi-taper spectral analysis, coupled with wavelet denoising, has proved to be very effective in reducing the noise in the spectrum and improving the results of fitting the peaks (Komm *et al.* 1997). Figure 4 illustrates the smoothing of the spectrum that results from applying these techniques. Tests with GONG data indicate that the application of a multi-taper analysis in tandem with a wavelet denoising increases the number of good fits by 50%, and decreases the formal variances of the estimated frequencies by a factor of 10 without introducing any systematic bias into the results.

5. Conclusion

The processing of helioseismic data is a lengthy and complex chain of steps that can affect the inferred internal solar structure. In particular, the determination of the image geometry and the fitting of the features in the spectra can introduce systematic errors into the estimated frequencies of the modes. A vigorous development activity is currently underway to improve the mode fitting methods, particularly in the physical model of the asymmetric line shape observed in both intensity and velocity. The outcome of this activity will substantially improve our knowledge of the solar interior.

6. Acknowledgements

This work utilizes data obtained by the Global Oscillation Network Group (GONG) project, managed by the National Solar Observatory, a Division of the National Optical Astronomy Observatories, which is operated by AURA, Inc. under a cooperative agreement with the National Science Foundation. The data were acquired by in-

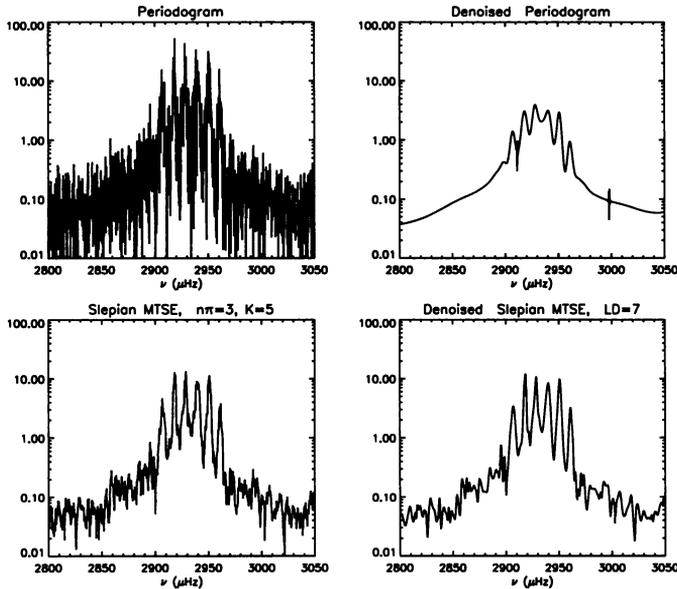


Figure 4. A comparison of a helioseismic spectrum created without any smoothing, with a Slepian multi-taper, a wavelet denoising, and both smoothings.

struments operated by the Big Bear Solar Observatory, High Altitude Observatory, Learmonth Solar Observatory, Udaipur Solar Observatory, Instituto de Astrofísico de Canarias, and Cerro Tololo Interamerican Observatory.

References

- Anderson, E.A., Duvall, T.L. Jr., and Jefferies, S.M.: 1990, *Astrophys. J.*, **364**, 699.
- Appourchaux, T., Gizon, L., and Rabello-Soares, M.-C.: 1997, *Astron. & Astrophys.*, submitted.
- Baudin, F., and Hill, F.: 1997, *Bull. Amer. Astron. Soc.*, **29**, 893.
- Harvey, J.W., Hill, F., Komm, R., Leibacher J.W., Pohl, B., and the GONG Team: 1997, these Proceedings.
- Hill, F., Anderson, E.A., Armet, D., Chavez, E., Erdwurm, W., Goodrich, J., Harvey, J.W., Hubbard, R.P., Jefferies, S.M., Kennedy, J.R., Ladd, G., Leibacher, J.W., Pintar, J.A., Pohl, B., Toner, C.G., Toussaint, R., Williams, W.E., and Wing, T.: 1998, *Solar Phys.*, in preparation.
- Howe, R., and Thompson, M.J.: 1997, *Astron. & Astrophys.*, submitted.
- Jefferies, S.M.: 1997, these Proceedings.
- Kennedy, J.R.: 1997, *Solar Phys.*, in press.
- Kennedy, J.R., and Williams, W.E.: 1997, *Solar Phys.*, submitted.
- Komm, R., Gu, Y., Hill, F., Stark, P., and Fodor, I.: 1997, in *Proc. Cool Stars*, in press.
- Schou, J.: 1992, *On the Analysis of Helioseismic Data*, Ph.D Thesis, Aarhus University, Appendix 4.
- Toner, C.G., and Harvey, J.W.: 1997, in *Sounding Solar And Stellar Interiors*, Proc. IAU Symp. 181, Ed. G. Berthomieu and F.-X. Schmider, in press.
- Toner, C.G., and Jefferies, S.M.: 1993, *Astrophys. J.*, **415**, 852.