[handwritten: → Helioseismic (?)]

Improved Analysis of Medium-l Data from the Michelson Doppler Imager

tim larson[1] · Jesper Schou[2]

Abstract We present detailed information on the analysis of intermediate spherical harmonic degree data from the Michelson Doppler Imager (MDI). Both current and historical usage are explained, and the various differences between the two are investigated to determine their effects on global mode parameters and systematic errors in the analysis. These differences include a number of geometric corrections made during spherical harmonic decomposition; updated routines for generating window functions, detrending timeseries, and filling gaps; and consideration of physical effects such as line asymmetry, horizontal displacement at the solar surface, and distortion of eigenfunctions by differential rotation. We apply these changes one by one to three years of data, and then reanalyze the entire MDI mission applying all of them, using both the original 72 day timeries and 360 day timeseries. We find significant changes in mode parameters, both as a result of the various changes to the processing, as well as between the 72 day and 360 day analyses. We find reduced residuals of inversions for internal rotation, but artifacts such as the high-latitude rotation rate near the surface remain. An annual periodicity in the f-mode frequencies is also investigated.

Keywords: Helioseismology, Observations; Oscillations, Solar

1. Introduction

The Michelson Doppler Imager (MDI) onboard the *Solar and Heliospheric Observatory (SOHO)* took data from December 1995 to April 2011. Equipped with a 1024×1024 CCD, it was capable in full disk mode of sending down dopplergrams with a 2 arcsec per pixel spatial resolution at a cadence of 60 seconds using the Ni 6768 spectral line. However, due to telemetry constraints, MDI was operated in full disk mode for only a few months total each year. For the rest of the time, we have only data that were convolved with a gaussian vector in each direction, subsampled by a factor of five, and severely cropped in order to fit into the available bandwidth. It was these vector-weighted dopplergrams that comprised the Medium-l Program and were the input to the analysis pipeline described here.

[1] Stanford University, email: tplarson@sun.stanford.edu
[2] Max-Planck-Institut für Sonnensystemforschung

[handwritten margin notes: "See link"; "Say MDI Doppler → helioseismology! + a little context"; "Very Technical. 'Vector' confusing in this context"; "?"]
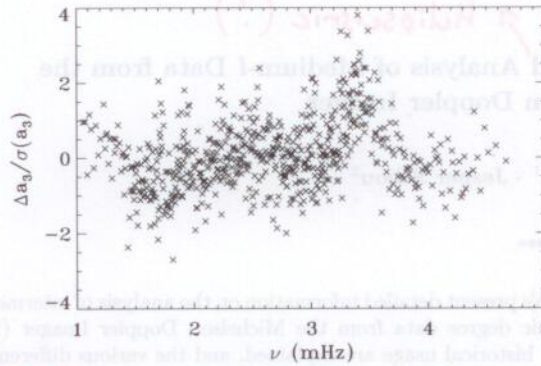
Figure 1. Normalized residuals of a_3 as a function of frequency for the 72 day period beginning on day number 4024 (see Table 1). Results have been binned by a factor of three for clarity.

Dopplergrams are decomposed into their spherical harmonic components described by their degree l and azimuthal order m, which are formed into timeseries and Fourier-transformed. In the medium-l regime, peaks in the power spectrum, corresponding to the oscillation modes, are well-separated from those of different degrees. Sets of modes with the same radial order n form ridges; modes with $n = 0$ are labelled f-modes, those with $n > 0$ are p-modes. The Fourier transforms are fit to yield the mode frequencies (among other parameters) for multiplets described by l and n, as well as a-coefficients, which parameterize the variation of the frequency with m (see Section 3.3). The frequencies and a-coefficients can be inverted to infer the sound speed or angular velocity in the solar interior as a function of latitude and radius. In this work we have used the odd a-coefficients to perform regularized least squares (RLS) inversions for angular velocity, which attempt to balance fitting the data with the smoothness of the solution. With an internal rotation profile in hand, one can compare the fit to the data to investigate systematic errors in the mode parameters. Although the χ^2 values have been reduced in the new analysis, the fact that they are still quite large indicates that the errors are still dominated by systematics.

One problem with the original analysis can be seen in a plot of the normalized residuals of a_3, shown in Figure 1. If the model were a good fit to the data, one would expect these to be normally distributed around zero with unit variance. A significant deviation from this expectation is the bump at around 3.4 mHz, which can be seen in all the odd a-coefficients and their residuals. Furthermore, the shape of the bump depends on the width of the fitting interval used, which by itself indicates a problem with the fits. Also visible in this plot are deviations from a continuous function at the ends of ridges. This feature, known as "horns", is visible in several of the mode parameters and is not reproducible by any reasonable internal rotation profile (see Section 4.2 and Figure 17).
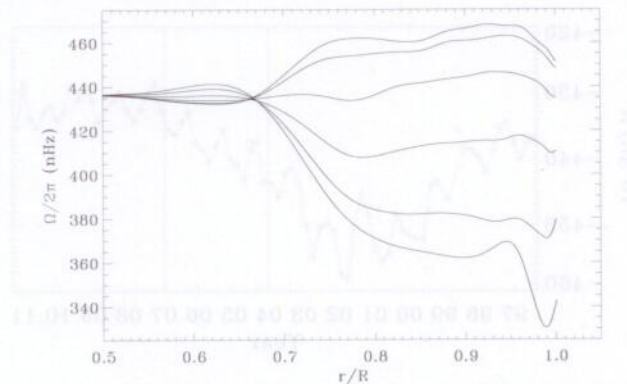
Figure 2. Rotation profile for the same period shown in Figure 1. From top to bottom, the curves correspond to latitudes of 0°, 15°, 30°, 45°, 60°, and 75°.

In parallel to the MDI analysis, the Global Oscillation Network Group (GONG) has done an independent medium-*l* analysis of dopplergrams taken from six ground-based observatories (the GONG network), using the same spectral line and cadence as MDI. Although the two analyses are generally in good agreement, in certain areas the inferences drawn by the two projects differ by more than their errors. In particular, the above mentioned bump is absent in the GONG analysis. Likewise, the MDI analysis indicates a polar jet at a latitude of about 75°, shown in Figure 2 which is not seen in the GONG analysis. Even exluding the modes which contribute to the bump does not remove this high latitude jet. Although the jet may be a real feature, the fact that it is also not seen in the full disk analysis of MDI data makes this unlikely. Until such discrepancies can be resolved, the analysis results must remain in doubt, and the issue has been studied at length by several investigators with little success (Schou *et al.*, 2002).

Another seeming systematic error seen in the original MDI analysis is a one-year periodicity in the fractional change in the seismic radius of the Sun (see Figure 3), which is proportional to the fractional change in *f*-mode frequency (Antia *et al.*, 2001). This cannot be studied with the GONG results because they do not fit enough *f*-modes, and the full disk data does not help either since it is only taken for approximately one time interval (long enough for global analysis) per year. Although it was presumed that this effect had to do with an annual variation in leakage between the modes, early investigations revealed that using a corrected P-angle, B0, and solar radius did not make a substantial difference (Schou and Bogart, 2002).

It was to address all these issues that a reanalysis of the medium-*l* data was undertaken. The original analysis was in general very successful, but it is based on certain approximations. Physical effects such as line asymmetry, horizontal displacement at the solar surface, distortion of eigenfunctions by the differential rotation, and a potential error in orientation of the Sun's rotation axis as given
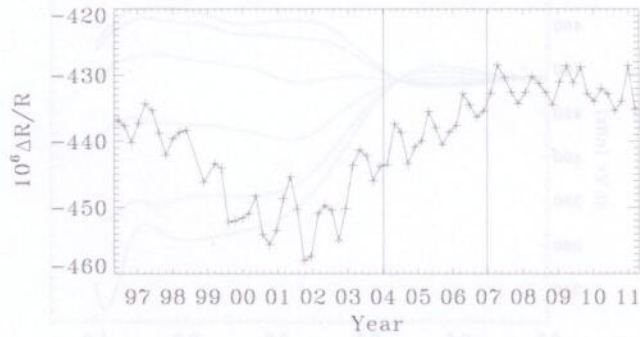
Figure 3. Fractional difference in seismic radius between observations and a model as a function of time. Vertical lines show the three years we reanalyzed for each correction (see Section 2).

by the Carrington elements, were not taken into account. Likewise, instrumental effects such as cubic distortion, nonzero P-angle, an alleged CCD tilt, and plate scale errors were ignored. Furthermore, new algorithms for generating the window functions, detrending the timeseries, and filling the gaps had become available. We updated the processing to include each of these considerations in turn to see what effect, if any, they had on the mode parameters and systematic errors.

2. Data

The velocity data were initially (Schou, 1999) analyzed in 74 timeseries of length 72 days, beginning 1996.05.01_00:00:00 [1]. The last data point used was at 2011.04.12_23:20:00. In late June 1998, however, control of *SOHO* was lost, resulting in a gap of more than 108 days. This was followed by a period of about two months of usable data at the end of 1998, and then another gap of more than 36 days. Therefore the 12th timeseries is offset from the others by 36 days and begins $12*72+36 = 900$ days after the first, while the 13th timeseries begins $14*72 = 1008$ days after the first, as shown in Table 1 (note the low duty cycles around day 2116). We have reanalyzed these same 74 time intervals, as well as used them to make one year (360 day) timeseries. Therefore only three of the 72 day timeseries were used to make the third year, and the last 72 day timeseries was unused in the one year analysis. All timeseries used are available for download [2].

---

[1] All dates and times have been converted from international atomic time (TAI).

[2] All final data products described here can be dowloaded from http://jsoc.stanford.edu/. See the appendix for details. Although not analyzed by us, a 5472 day timeseries spanning the entire mission is also available.

To see the effect of the various corrections, we apply them one by one to the analysis of 15 timeseries covering a period of three years beginning on 2004.01.08. This is long enough to see an annual component in the f-mode frequencies, but short enough to approximate the solar cycle variation as linear during its declining phase. Beginning with the plate scale correction, we apply, in order, corrections for cubic distortion from the instrument optics, the P-angle error, the Carrington inclincation error, and CCD tilt. These are all the corrections we made during the spherical harmonic decomposition, and we regenerate timeseries for the entire mission with all of them applied. The next two improvements applied are to the detrending and then the gapfilling. Again, detrended and gapfilled timeseries have been regenerated for the entire mission. For the one-year analysis, the 360 day timeseries were created by concatenating the detrended and gapfilled 72 day timeseries. The remaining changes to the processing all take place in the fitting. We first take into account horizontal displacement, and then distortion of eigenfunctions following the prescription given by Woodard (1989). Mode parameters for the entire mission have been recomputed with these applied, and again using asymmetric line profiles in addition.

3. Method

Analysis proceeds as follows. An observed oscillation mode is taken as proportional to the real part of a spherical harmonic given by $Y_l^m(\phi, \theta) = P_l^m(\cos\theta)e^{im\phi}$, where the P_l^m are associated Legendre functions with the property that

$$\int_{-1}^{1} \left[P_l^m(x) \right]^2 dx = 1 \qquad (1)$$

and that $P_l^{-m} = P_l^m = P_l^{|m|}$. As used here, l and m are integers with $l \geq 0$ and $-l \leq m \leq l$. However, since modes with positive m and negative m look identical at any instant in time, we only compute spherical harmonic coefficients for $m \geq 0$. For medium-l analysis, we use degrees up to $l = 300$. Beyond this, peaks along the f-mode ridge begin to blend into each other. For the p-modes, this is already happening around $l = 200$.

To find the spherical harmonic coefficients, each image is remapped to a uniform grid in longitude and sin(latitude) using a cubic convolution interpolation, and apodized with a cosine curve in fractional image radius from 0.83 to 0.87. The resulting map is Fourier-transformed in longitude and for each m a dot product is taken with a set of associated Legendre functions of sin(latitude), which yields the complex amplitudes of the spherical harmonics as a function of l and m in the ranges given above. These amplitudes are arranged into timeseries 72 days long, and the timeseries for each l and m is detrended, gapfilled, and Fourier-transformed, at which point the negative frequency part of the transform is identified with negative m and the conjugate of the positive frequency part is identified with positive m. The peaks in (the magnitude of) the Fourier transform are fit (a process that has become known as peakbagging), resulting in a peak frequency, amplitude, linewidth, and background for each l and radial order

Table 1. Timeseries used. Day numbers are given relative to the MDI epoch of 1993.01.01_00:00:00_TAI. Both these and the dates refer to the first day of the timeseries. Duty cycles are given for the original timeseries (DC0), the new timeseries (DC1) and the new timeseries after gapfilling (DC2). The difference DC0-DC1 tends to be positive at the beginning of the mission (at most 0.031) and negative at the end (not less than -0.02).

Day	Date	DC0	DC1	DC2	Day	Date	DC0	DC1	DC2
1216	1996.05.01	0.895	0.888	0.907	4024	2004.01.08	0.986	0.991	1.000
1288	1996.07.12	0.964	0.949	0.966	4096	2004.03.20	0.782	0.770	0.858
1360	1996.09.22	0.964	0.954	0.969	4168	2004.05.31	0.897	0.898	0.989
1432	1996.12.03	0.976	0.962	0.982	4240	2004.08.11	0.853	0.852	0.941
1504	1997.02.13	0.952	0.950	0.964	4312	2004.10.22	0.969	0.968	0.981
1576	1997.04.26	0.981	0.981	1.000	4384	2005.01.02	0.991	0.991	1.000
1648	1997.07.07	0.970	0.976	0.986	4456	2005.03.15	0.991	0.992	0.996
1720	1997.09.17	0.973	0.965	0.976	4528	2005.05.26	0.983	0.989	1.000
1792	1997.11.28	0.979	0.982	1.000	4600	2005.08.06	0.989	0.988	0.996
1864	1998.02.08	0.969	0.968	0.976	4672	2005.10.17	0.985	0.985	0.996
1936	1998.04.21	0.884	0.883	0.896	4744	2005.12.28	0.988	0.992	1.000
2116	1998.10.18	0.731	0.726	0.737	4816	2006.03.10	0.990	0.992	1.000
2224	1999.02.03	0.894	0.885	0.894	4888	2006.05.21	0.962	0.971	0.978
2296	1999.04.16	0.982	0.974	0.986	4960	2006.08.01	0.988	0.992	1.000
2368	1999.06.27	0.986	0.987	1.000	5032	2006.10.12	0.990	0.991	1.000
2440	1999.09.07	0.930	0.917	0.941	5104	2006.12.23	0.895	0.900	0.907
2512	1999.11.18	0.870	0.839	0.852	5176	2007.03.05	0.976	0.977	0.986
2584	2000.01.29	0.986	0.983	0.989	5248	2007.05.16	0.985	0.984	0.994
2656	2000.04.10	0.994	0.994	1.000	5320	2007.07.27	0.988	0.991	1.000
2728	2000.06.21	0.988	0.988	0.996	5392	2007.10.07	0.965	0.968	0.980
2800	2000.09.01	0.986	0.984	0.995	5464	2007.12.18	0.985	0.987	1.000
2872	2000.11.12	0.947	0.937	0.945	5536	2008.02.28	0.996	0.996	1.000
2944	2001.01.23	0.985	0.986	1.000	5608	2008.05.10	0.989	0.993	1.000
3016	2001.04.05	0.990	0.990	1.000	5680	2008.07.21	0.988	0.991	1.000
3088	2001.06.16	0.964	0.961	0.975	5752	2008.10.01	0.983	0.986	0.994
3160	2001.08.27	0.991	0.991	1.000	5824	2008.12.12	0.983	0.989	1.000
3232	2001.11.07	0.971	0.970	0.979	5896	2009.02.22	0.996	0.996	1.000
3304	2002.01.18	0.859	0.862	0.870	5968	2009.05.05	0.951	0.954	0.960
3376	2002.03.31	0.987	0.985	1.000	6040	2009.07.16	0.709	0.729	0.736
3448	2002.06.11	0.978	0.984	0.996	6112	2009.09.26	0.985	0.989	0.996
3520	2002.08.22	0.991	0.990	1.000	6184	2009.12.07	0.989	0.993	1.000
3592	2002.11.02	0.994	0.994	1.000	6256	2010.02.17	0.992	0.993	1.000
3664	2003.01.13	0.992	0.989	1.000	6328	2010.04.30	0.988	0.995	1.000
3736	2003.03.26	0.982	0.982	0.996	6400	2010.07.11	0.952	0.961	0.971
3808	2003.06.06	0.822	0.826	0.852	6472	2010.09.21	0.879	0.881	0.929
3880	2003.08.17	0.981	0.981	0.996	6544	2010.12.02	0.732	0.744	0.753
3952	2003.10.28	0.878	0.878	0.952	6616	2011.02.12	0.812	0.812	0.822

n. The m-dependence of the frequencies is parameterized by the a-coefficients, which are fit for directly in the peakbagging, with the other mode parameters assumed to be the same for all m.

Because of leakage between the modes, predominantly caused by our inability to see most of the Sun, the Fourier transform of the target l and m contain peaks from neighboring modes as well, which have to be accounted for in the peakbagging. This is done through the so-called leakage matrix, which quantifies the amplitude of each mode as it appears in spectra of neighboring modes. The leakage matrix is calculated by generating artificial images containing spherical harmonics and decomposing them in the same way as the the actual data. The same leakage matrix has been used for all times (see Section 3.3).

3.1. Spherical Harmonic Transform

Spherical harmonic decomposition begins with a remap, which gives us an opportunity to apply certain corrections to the data. The most significant of these is for the plate scale. Although assumed to be a constant in the original analysis, changes in the instrument with temperature and over time actually caused it to vary. The angular size of the Sun is given by $\arcsin(D/R_{ref})$, where D is the observer distance and R_{ref} is defined as exactly 696 Mm. This value is divided by the plate scale to give the solar radius in pixels. Hence the original value used for the solar radius in pixels was in error. In the current analysis the plate scale is given by a multiplicative factor times the original constant plate scale of 1.97784 arcsec per pixel. The inverse of this factor (hence the radius correction) is given as a function of time t by $a_0 + [a_1 + a_2 * (t - t_0) + a_3 * (t - t_0)^2] * D$ where a_0, a_1, a_2, a_3, and t_0 result from a fit to $(MA + MI)/(2R_0)$, where MA and MI are the lengths of the major and minor axes of the solar image returned by the routine used to fit the solar limb and R_0 is the original value used for the solar radius in pixels (Keh-Cheng Chu, private communication). The parameters of the fit change throughout the mission, typically at a focus change. Hence the radius correction is a piecewise continuous function.

To account for optical distortion in the instrument, we apply a correction given by a cubic distortion model. The fractional change in coordinates is given by $C_{dist}(r^2 - R^2)$, where r is the distance from the center of the CCD, R is the (updated) radius of the solar image, and all quantities are given in terms of full disk pixels. For C_{dist} we have used 7.06×10^{-9}, which was derived from a ray-trace of the MDI instrument [3].

For the P-angle and B0 we apply a simple sinusoidal correction. Since the error from the node position is not significant (Beck and Giles, 2005), if δI is the error from the Carrington inclination, δP is the error on the P-angle resulting from misalignment of the CCD, $B0$ is the original B-angle and P is the original P-angle, then the new values are given by

$$B' = B0 + \delta I * \sin((t_{obs} - t_{ref}) * 2\pi) \tag{2}$$

---

[3] This differs from the value used in Korzennik, Rabello-Soares, and Schou (2004), which resulted from a different model. It is unclear how to resolve the discrepancy. It is hoped that ongoing investigation of the MDI distortion will help.

and

$$P' = P + \delta P + \delta I * \cos\left[(t_{obs} - t_{ref}) * 2\pi\right] \qquad (3)$$

where t_{obs} is the observation time and t_{ref} is a time when B0 is zero, both measured in years [4]. For the value of δP we have used $-0.2°$, which agrees with values obtained both by cross-correlations with GONG and from the Mercury transit in November 1999 (Cliff Toner, private communication). For the value of δI we have used $-0.1°$, a value derived by Beck and Giles (2005).

The ellipticity of the observed solar image is much greater than the actual ellipticity of the Sun. A possible explanation for this is that the CCD is tilted with respect to the optical axis of the instrument. To correct for this, we follow the prescription given in the appendix of Korzennik, Rabello-Soares, and Schou (2004). The required parameters are β, the amount to rotate the x-axis to give the direction around which the CCD is tilted; α, the amount of the tilt; and f_{eff}, the effective focal length. We have adopted the values $\beta = 56.0°$, $\alpha = 2.59°$, $f_{eff} = 12972.629$ pixels, which are consistent with the values found by Korzennik, Rabello-Soares, and Schou. Although there is some doubt as to whether the CCD is actually tilted, the model still reproduces the observed ellipticity reasonably well (Korzennik, Rabello-Soares, and Schou, 2004).

3.2. Detrending and Gapfilling

Once the 72 day timeseries have been assembled, the next step in the processing is the generation of the window function. In the original analysis, the $l = 0$ timeseries was examined to ensure that gaps resulting from known spacecraft and instrument events were accurately reflected in the timeseries generated. These events included such things as station keeping, momentum management, problems with the ground antennas, emergency Sun reacquisitions (ESR's), and tuning changes due to instrumental drifts. Additionally, any day whose duty cycle was less than 95% was investigated to ensure that all potentially available data had gone into the spherical harmonic decomposition. Unfortunately, the original analysis employed a simple algorithm that performed detrending of the timeseries on full mission days only, thus requiring any day that contained a discontinuity in the data (such as was caused by tuning changes) to have its window function zeroed to the nearest day boundary. Also, the instrument occasionally stopped taking images, which caused thermal transients in the instrument until equilibrium was reestablished. These turn-on transients, and other data deemed unusable, were also manually identified in the timeseries and set to zero in the window function. After this, 10 timeseries were examined and thresholds on acceptable values in them were set by hand in order to reject outliers. These 10 timeseries are the real parts of $l = 0, m = 0$; $l = 1, m = 0$; $l = 1, m = 1$; the imaginary part of $l = 1, m = 1$; and the sum over m of the real part squared plus the imaginary part squared for $l = 1, 2, 5, 10, 20, 50$.

In the new analysis, we use the old $l = 0$ timeseries, since they had already been examined, to confirm the legitimacy of any data missing in the new $l = 0$

---

[4]For t_{ref} we have used 2001.06.06_06:57:22_TAI.

timeseries. We then automatically set to zero in the window function any point where the Image Stabilization System (ISS) was off, as derived from housekeeping keywords. Next we form 10 timeseries in the same fashion as the original analysis, but we replace squaring the real and imaginary parts in the sum over m with taking the absolute value of the real and imaginary parts. We then run these 10 timeseries through a median filter to reject outliers. We do this by subtracting a 41 point running median, taking the rms of the result excluding the top and bottom 1% of the data, and rejecting any points that differ from zero by more than 6 times the rms.

In the new analysis, the discontinuities, which were typically caused by tuning changes, spacecraft rolls, and any event which powered down the instrument, all had to be identified by hand. This information has to be available for the median filtering, and subsequent detrending can now be done on entire continuous sections of data irrespective of day boundaries. Further, the beginning of every section is automatically checked for the existence of thermal transients in the $l = 0$ timeseries by fitting a sum of two decaying exponentials and a constant[5]. The window function is zeroed wherever the fit differs from the constant by more than the rms of the median-subtracted $l = 0$ timeseries. Also, by defining sections, we were able to manually reject any data lying in between the sections, if such was deemed necessary. In the new analysis, defining the sections of continuous data was the only operation that required human attention, and had to be done only once.

Detrending in the original analysis was performed on whole mission days (1440 time points) by fitting a Legendre polynomial of degree given by $2 + N_{span}/300$ where N_{span} is the number of minutes spanned by the available data points and the division truncates to the next lowest integer. This polynomial was subtracted prior to gapfilling, which was also independently performed on each mission day. The routine used would compute a model from the data and use it to fill gaps up to a maximum size of five. It required six points either before or after each gap to do so, regardless of the size of the gap.

Detrending in the new analysis is done by fitting a Legendre polynomial of degree seven to an interval of data spanning 1600 minutes which is advanced by 1440 minutes for each fit. In other words, the detrending intervals overlap by 160 points. The polynomials are stitched together in the overlap region by apodizing each of them with a $\cos^2$ curve. In the case that the data points in a detrending interval spanned less than 800 minutes, the Legendre polynomial was recomputed for the shorter span and the fit was not apodized. The resulting function is subtracted from the data to give a timeseries with a mean of zero.

In the new analysis, gaps are filled using an autoregressive algorithm based on the work of Fahlman and Ulrych (1982). This method predicts values for the missing data based on the spectral content of the data present. Each point in the known data is expressed as a linear combination of the N preceding and following points, where N is the order of the autoregressive model, the coefficients

[5]We do not fit the decay constants as part of this check. Rather, we fit for them only once and now hold them fixed at values of 15 and 60 minutes. The use of two exponentials comes from a model of the instrument.

of which are found by minimizing the prediction error in the least squares sense. Hence, the order of the model can be no greater than the number of points in the shortest section of data. If a model of a certain order is desired, it imposes a lower limit on the length of data sections that can be used to generate it. In our implementation, we use the maximum order such that at least 90% of the data will be used to generate the model, up to a maximum order of 360. However, the coefficients of a model of order N are determined from a model of order $N-1$, so the model order may actually be less than requested if the prediction error stops decreasing as the model order is increased. Once the model is known, the gaps are filled by minimizing the prediction error in the least squares sense, this time with respect to the unknown data values. The innovation over the method of Falhman and Ulrych is that all gaps shorter than the model order within each filling interval are filled simultaneously. Gaps at the beginning or end of the timeseries are not filled. The model order may possibly then be increased by using the filled values as known data, and the process is repeated, but using the original gap structure. That is, the gaps that were filled on the first iteration will be filled again using the new model. If the model order did not change, or if all the gaps were already filled in the first iteration, the process stops after two iterations. Otherwise a final iteration is run wherein a new model is computed using the newly filled values, and the gaps are filled one last time (Rasmus Larsen, private communication).

3.3. Peakbagging

Fourier transforms of the gapfilled timeseries are fit using a maximum likelihood technique, taking into account leakage between the modes. Modelling an oscillation mode as a stochastically excited damped oscillator, both the real and imaginary parts of the Fourier transform will be normally distributed with a mean of zero. The variance due to the mode will be given by

$$v(\nu_0, w, A, \nu) = \frac{2wA^2}{w^2 + 4(\nu - \nu_0)^2} \tag{4}$$

where ν_0 is the frequency of the mode, w is the full width at half maximum, and A is the amplitude (A^2 is a measure of the total power in the mode). To fit an actual observed spectrum, one must also add a background term; our treatment of the background is described below. Furthermore, to account for the redistribution of power caused by gaps in the timeseries, this model will be convolved with the power spectrum of the window function (Anderson, Duvall, and Jefferies, 1990). If x is the real part of the observed value of the Fourier transform, then the probability density for the ith frequency bin in the real part will by given by

$$P_{real}(\nu_0, w, A, \nu_i) = \frac{1}{\sqrt{2\pi v(\nu_i)}} \exp\left(-\frac{x(\nu_i)^2}{2v(\nu_i)}\right) \tag{5}$$

and likewise for P_{imag} with x replaced by y, the imaginary part. The total probability density for the ith bin is then $P = P_{real}P_{imag}$. In these equations the

mode parameters, and hence v, are functions of n, l, and m; we have suppressed their dependence on these for conciseness.

The idea is to maximize the joint probability density of a given mode, which is given by a product of individual probability densities over a suitable number of frequency bins (assuming that each frequency bin is independent, which is not strictly true in the presence of gaps). This is equivalent to minimizing the negative logarithm of this product, which, except for constants, is given by

$$S(\nu_0, w, A) = \sum_i \left[\ln\left(v(\nu_i)\right) + \frac{x(\nu_i)^2 + y(\nu_i)^2}{v(\nu_i)} \right]. \tag{6}$$

For a given value of l, there will be $2l+1$ values of m. Rather than fitting each m separately, we will maximize the joint probability density of all of them together. To do so, we assume the width and amplitude are independent of m and estimate the variation of the background with m from the spectrum far from the peaks. We redefine ν_0 as the mean multiplet frequency for each n and l, and expand the frequency dependence on m as

$$\nu_{nlm} = \nu_0(n,l) + \sum_{i=1}^{N_a} a_i(n,l)\mathcal{P}_i^l(m) \tag{7}$$

where the polynomials $\mathcal{P}$ are those used in Schou, Christensen-Dalsgaard, and Thompson (1994), and the coefficients a_i are fit for directly. In what follows, we will label the set of parameters upon which S depends using the vector $\boldsymbol{p}$. This will include ν_0, w, A, N_a a-coefficients, a background parameter (described below), and optionally a parameter to describe the asymmetry (also described below), for each n and l.

Due to leakage between the modes, the observed timeseries and Fourier transforms are a superposition of the true underlying oscillations. The observed timeseries for a given l and m will be given by

$$o_{lm}(t) = \sum_{n'l'm'} c_{lm,l'm'}^{RR} \mathrm{Re}[a_{n'l'm'}(t)] + i c_{lm,l'm'}^{II} \mathrm{Im}[a_{n'l'm'}(t)] \tag{8}$$

where $a(t)$ is the complex amplitude of the underlying timeseries, and Re[] and Im[] denote the real and imaginary parts, respectively. The sensitivity coefficients c^{RR} and c^{II} give the real-to-real leaks and imaginary-to-imaginary leaks respectively. Approximate expressions for the radial contribution to these coefficients are given in Schou and Brown (1994). Under the same approximations, it can be shown that the real-to-imaginary and imaginary-to-real leaks are identically zero for geometries that are symmetric around the central meridian. Although these are still assumed to be zero for the current work, c^{RR} and c^{II} are computed more accurately, as described below. It can also be shown that

$$c_{lm,l'm'}^{RR} = c_{l'm',lm}^{RR}$$
$$c_{lm,l'm'}^{II} = c_{l'm',lm}^{II}$$
$$c_{l(-m),l'm'}^{RR} = c_{lm,l'm'}^{RR}$$

$$c^{II}_{l(-m),l'm'} = -c^{II}_{lm,l'm'} \qquad (9)$$

Note that since the spherical harmonic decomposition is not able to separate the different values of n, we have suppressed the n-dependence of the leaks in these equations. Later we will consider effects that cause the leaks to vary with n. In frequency space, the observed Fourier transform can then be expressed as

$$\tilde{o}_{lm}(\nu) = x_{lm}(\nu) + iy_{lm}(\nu) = \sum_{n'l'm'} C_{lm,l'm'} \tilde{a}_{n'l'm'}(\nu) \qquad (10)$$

where $C = (c^{RR} + c^{II})/2$ (Schou and Brown, 1994). Although in principle the sum above should be over all modes, for a given l and m, only modes in a certain range in l' and m' will have signficant leakage. Therefore the sum above need only be over modes which may have appreciable amplitudes within the fitting window. For this work we have used $\Delta l = l - l'$ in the range ± 6 and $\Delta m = m - m'$ in the range ± 15. Since the modes on the Sun are uncorrelated with each other, the elements of the covariance matrix between the different transforms at each frequency point will be given by

$$E^{modes}_{lm,l'm'}(\nu_i) = Cov(x_{lm}(\nu_i), x_{l'm'}(\nu_i)) = Cov(y_{lm}(\nu_i), y_{l'm'}(\nu_i))$$

$$= \sum_{n''l''m''} C_{lm,l''m''} C_{l'm',l''m''} v_{n''l''m''}(\boldsymbol{p}, \nu_i). \qquad (11)$$

The total covariance will be the sum of the covariance between the modes and the covariance of the noise. Since we fit each l separately and all m for that l simultaneously, the covariance matrix used in the fitting is

$$E_{m,m'}(\nu_i) = E^{modes}_{m,m'}(\nu_i) + \tilde{E}_{m,m'} \frac{\nu_B}{\nu_i} e^b \qquad (12)$$

where $\tilde{E}$ is the measured covariance between m and m' in the frequency range 7638.9 to 8217.6 μHz, ν_B is a constant, and b is a free parameter determined in the fit. Due to our choice of normalization, e^b is proportional to the length of the timeseries. The probability density for a frequency bin then becomes

$$P(\boldsymbol{p}, \nu_i) = \frac{1}{|2\pi \mathbf{E}(\boldsymbol{p}, \nu_i)|} \exp\left(-\frac{1}{2}\Big(\boldsymbol{x}(\nu_i)^T \mathbf{E}(\boldsymbol{p}, \nu_i)\boldsymbol{x}(\nu_i) + \boldsymbol{y}(\nu_i)^T \mathbf{E}(\boldsymbol{p}, \nu_i)\boldsymbol{y}(\nu_i)\Big)\right) \qquad (13)$$

and the function to minimize becomes

$$S(\boldsymbol{p}) = \sum_i \left[\ln|\mathbf{E}(\boldsymbol{p}, \nu_i)| + \boldsymbol{x}(\nu_i)^T \mathbf{E}(\boldsymbol{p}, \nu_i)\boldsymbol{x}(\nu_i) + \boldsymbol{y}(\nu_i)^T \mathbf{E}(\boldsymbol{p}, \nu_i)\boldsymbol{y}(\nu_i)\right] \qquad (14)$$

where $|\ |$ denotes the determinant, $\boldsymbol{x}$ is a vector of the $2l + 1$ real parts of the transforms, and $\boldsymbol{y}$ is a vector of the $2l + 1$ imaginary parts. Note that $\boldsymbol{p}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ are implicit functions of n and l (the dependence of $\boldsymbol{x}$ and $\boldsymbol{y}$ on n come from the frequency range chosen for the fitting window). For the width of the fitting window we have chosen 5 times the estimated width of the peak, with a minimum of 2.9 μHz and a maximum of 81.0 μHz. The peakbagging will yield

the mode parameters specified by p for each multiplet it is able to fit, as well as error estimates on these, generically referred to as σ. The errors are estimated from the inverse of the Hessian matrix at the minimum of S.

The minimization scheme used is a variation of the Levenberg-Marquardt method. For further details, such as approximations made in the calculation of derivatives, the reader is referred to Schou (1992).

Since we fit for one n and l at a time while holding the leaks fixed, the peak-bagging must be iterated to account for the variation of the mode parameters as the fits proceed. For all iterations except the last, we fit six a-coefficients. In the original analysis, the initial guess for the first iteration was taken from the final fits of the previous timeseries. In the new analysis, the same initial guess was used for all time periods, which allows for fitting all of them independently of one another. We found this made no significant difference. Any modes which cannot be fit in the first attempt have the initial guess of their background parameter perturbed by -1 and the fit is reattempted. At this point in the original analysis the resulting set of fitted modes would be weeded by hand to reject outliers. In the new analysis this step is simply skipped; again we found it made no significant difference. In both cases the remaining modes are used to make new initial guesses for the modes that had not converged (or were rejected). The second iteration is then done in the same way as the first. At no point do we ever attempt to fit modes for which there are estimated to be other modes within ± 2 in l and within twice the line width. These typically occur at the ends of ridges and do not converge in any case.

For subsequent iterations, the modes which have not converged to within $0.1\sigma(\nu)$ or for which there exist unconverged modes with the same n and $\Delta l = \pm 1$ are always fitted (occasionally more modes would be fit in the original analysis). In the original analysis the convergence of the modes would be examined to determine the total number of iterations, which would usually be from 9 to 11. All modes would be fit in at least the last two iterations. In the new analysis, for the sake of automation, the peakbagging would always be performed for 10 iterations with all modes being fit during the last three. In both cases, the final fits are repeated with both 18 and 36 a-coefficients.

After the final iteration, the resulting set of modes is automatically weeded one last time. For the fits with six a-coefficients, modes differing by more than $0.25\sigma(\nu)$ from their input guesses are rejected. Additionally, any mode with a large error on its frequency given its width is suspect. If there were no background noise, we would expect a frequency error given by

$$(2l + 1)\sigma^2(\nu) = \frac{w}{4\pi T} \tag{15}$$

where T is the length of the timeseries (Libbrecht, 1992). Any mode with a frequency error greater than 6 times this prediction is rejected. The same theoretical error estimate is the motivation for identifying modes for which the line width is smaller than the width of a frequency bin. These modes have the error estimates on their frequencies and a-coefficients increased by a factor of $\sqrt{1/(wT)}$. This prevents underestimates of the error caused by low estimates of the widths in the region where they cannot be reliably estimated. The resulting set of mode

parameters is then compared to those of a model obtained from an inversion of fits to a 360 day timeseries at the beginning of the mission. The median difference between the fit and the model of the odd a-coefficients is taken for the f-mode to account for their change throughout the solar cycle. The differences for all the modes are compared to this median; any that differ by more than 10σ are rejected.

To weed the 18 and 36 a-coefficient fits, their error estimates are adjusted as above. Frequencies and a-coefficients are then compared to the six a-coefficient case. Any mode for which the error estimates on any of these parameters increased by more than a factor of 2, or for which any of these parameters changed by more than 2σ (estimated from the 18 or 36 a-coefficient fits) is rejected. Any mode which was rejected in the six a-coefficient fits is also removed from the 18 a-coefficient fits, and any mode which was rejected in the 18 a-coefficient fits is also removed from the 36 a-coefficient fits.

3.3.1. Leakage Matrix

For this work, the leakage matrix elements are computed by generating artificial images containing components of vector spherical harmonics projected onto the line of sight for a subset of the modes we wish to fit. A mode on the Sun has a velocity at the surface with components proportional to the real parts of [6]

$$
\begin{aligned}
u_r &= Y_l^m(\phi,\theta) = P_l^m(x)e^{im\phi} \\
u_\theta &= -\frac{1}{L}\frac{\partial Y_l^m}{\partial \theta} = \frac{1}{L}\frac{dP_l^m}{dx}e^{im\phi}\sin\theta \\
u_\phi &= -\frac{1}{L}\frac{1}{\sin\theta}\frac{\partial Y_l^m}{\partial \phi} = -\frac{1}{L}\frac{im}{\sin\theta}P_l^m(x)e^{im\phi}
\end{aligned}
\tag{16}
$$

where $x = \cos\theta$ and $L = \sqrt{l(l+1)}$. A mode with oscillation amplitude V_{lm} will then have a total velocity of

$$
\boldsymbol{V} = \boldsymbol{V}_{lm}^r + c_t \boldsymbol{V}_{lm}^h
\tag{17}
$$

where $\boldsymbol{V}_{lm}^r = V_{lm}u_r\hat{r}$, $\boldsymbol{V}_{lm}^h = V_{lm}(u_\theta\hat{\theta} + u_\phi\hat{\phi})$, and

$$
c_t = \frac{\nu_0^2(0,l)}{\nu_0^2(n,l)}
\tag{18}
$$

is the ratio of the mean multiplet frequency of the f-mode squared to the mean multiplet frequency of the given mode squared at that l (Rhodes et al., 2001). Therefore $c_t = 1$ for the f-mode and $c_t < 1$ for the p-modes. However, since the n's are not separated by the spherical harmonic decomposition, we create a separate matrix for the vertical and horizontal components; the effective leakage matrix will be computed during the fitting by combining them according to Equation (17). We project each component onto the line of sight separately using

---

[6]The sign of u_r relative to u_θ and u_ϕ depends on the convention for the sign of m.

projection factors calculated for a finite observer distance. In the approximation of an infinite observer distance this would become

$$u_{vertical} = V_{lm} P_l^m(x) e^{im\phi} \sin\theta \cos\phi$$

$$u_{horizontal} = \frac{V_{lm}}{L} \left(\frac{\mathrm{d}P_l^m}{\mathrm{d}x} \sin\theta \cos\theta \cos\phi + \frac{im}{\sin\theta} P_l^m(x) \sin\phi \right) e^{im\phi} \quad (19)$$

where we choose $V_{lm} = 1000$ m/s to give us roughly the same order of magnitude as the observations. As with the real data, these images are only calculated for $m \geq 0$. The resulting leakage matrix will be divided by 1000.

These images are first generated as they would appear to MDI, assuming an observer distance of 1 AU, a P-angle and B0 both equal to zero, and that the image is centered on the CCD. They are then convolved with a gaussian in each dimension with a width of $\sigma = 4/\sqrt{2}$, just as they are onboard, but they are not subsampled at this point. Rather they are also convolved with a function that takes into account the interpolation errors made during the subsequent spherical harmonic decomposition, at which time they will be remapped to the same resolution in longitude and sin(latitude) as the real data. The higher resolution images are used to simulate an average over different pixel offsets; we have verified the accuracy of this technique by generating lower resolution images and actually performing the average. After the remap, the artificial data are processed exactly like the real data. For each image, we take its inner product with a set of target spherical harmonics in the range $\Delta l = \pm 6$ given above. The results are the coefficients c^{RR} and c^{II} given in Equation (8). The values for the modes we did not compute directly are found by interpolation. The values for negative m given by Equations (9).

In the original analysis, only the vertical component of the leakage matrix was used, meaning that the horizontal component was assumed to be zero. Although this is not a bad approximation for high-order p-modes, it becomes worse as one approaches the f-mode ridge, where the horizontal and vertical components have equal magnitude. In the new analysis, our first improvement to the peakbagging is to include both components.

For a spherically symmetric Sun, the horizontal eigenfunctions would be spherical harmonics. Although the presence of differential rotation breaks this symmetry, the true eigenfunctions can still be expressed as a sum over spherical harmonics. In the new analysis, this is accounted for in the peakbagging by appropriately summing the leakage matrix. We use the prescription given by Woodard (1989) with the differential rotation expanded as

$$\Omega(x) = B_0 + B_1 x^2 + B_2 x^4 \quad (20)$$

where again $x = \cos\theta = \sin(\text{latitude})$. We first used constants derived from surface meaurements, with values of $B_1 = -75$ nHz and $B_2 = -50$ nHz as given by Woodard (the value of B_0 is not used). However, this has the drawback of distorting every mode in the same way, even though they sample different depths where the differential rotation has a different dependence on latitude. Following Vorontsov (2007), we use the estimated splitting coefficients to calculate B_1 and

B_2 for each mode separately. Using only six a-coefficients, in the asymptotic limit we will have

$$B_1 = -5a_3 - 14a_5$$
$$B_2 = 21a_5 \qquad (21)$$

so that B_1 and B_2 change as the iteration proceeds. Fortunately this did not disrupt the convergence of the a-coefficients. This change made only a modest difference in the mode parameters, as discussed below.

3.3.2. Asymmetry

In addition to the symmetric line profiles described by Equation (4), we have also used asymmetric profiles to fit the data. Although it is common to use the profile derived by Nigam and Kosovichev (1998), their equation has the undesirable property that its integral over all frequencies is infinite. Also, there is no clear way to evaluate it between the modes. To derive a more well behaved profile, we begin with Equation (3) of Nigam and Kosovichev (1998), which was derived for a one-dimensional rectangular potential well model, and generalize it by replacing their βX with an arbitrary function of frequency $h(\nu)$. Since β is generally very small, we drop the second term in the numerator to arrive at a variance given by

$$v(\nu) = \frac{P_D(\nu)\cos^2\!\big[h(\nu)+\delta(\nu)\big]}{g(\nu)+\sin^2\!\big[h(\nu)\big]} \qquad (22)$$

where P_D is the power spectrum of the excitation, δ is a measure of the asymmetry, and g is related to the damping. The function h is constrained to be $n\pi$ at the mode frequencies, and in the numerator we have changed sin to cos so that $\delta = 0$ implies a symmetric profile. Considering a single l and m, we can expand in terms of profiles given by Equation (4) to get

$$v_{lm}(\nu) = \cos^2\!\big[h_{lm}(\nu)+\delta_{lm}(\nu)\big]\sum_n \frac{1}{\cos^2\!\big[\delta_{lm}(\nu_{nlm})\big]}\frac{2wA^2}{w^2+4(\nu-\nu_{nlm})^2} \qquad (23)$$

where the factor $1/\cos^2(\delta_{lm}(\nu_{nlm})$ has been added so that to lowest order, A retains its original meaning. To find a function to use for h, we note that from the Duvall law (Duvall, 1982) we can define $h_0(\nu) = \nu F(\nu/(l+1/2)) - \pi\alpha(\nu) \approx n\pi$, where F and α are known functions. These we have tabulated from a fit to a 360 day timeseries at the beginning of the mission, and interpolate them as needed during the peakbagging. We then choose $h = h_0 + h_1$ where h_1 is a piecewise linear function chosen to make h exactly $n\pi$ at the mode frequencies as required. The function δ can likewise be interpolated using a piecewise linear function derived from the input guesses. Above the frequency of the maximum n and below the frequency of the minimum n, we assign constant values to h_1 and δ.

Equation (23) is valid for all frequencies. Restricting ourselves to single mode, we can now replace the variance in Equation (11) with

$$v_{nlm}(\boldsymbol{p}, \nu) = \frac{\cos^2[h_{nl}(\nu') + \delta_{nl}(\nu')]}{\cos^2[\delta_{nl}(\nu_{nlm})]} \frac{2w_{nl}A_{nl}^2}{w_{nl}^2 + 4(\nu - \nu_{nlm})^2} \qquad (24)$$

where $\nu' = \nu - \nu_{nlm} + \nu_0(n, l)$, ν_{nlm} is given by Equation (7), and we have implicitly assumed that the asymmetry is the same for all m.

To form the initial guess for the asymmetric fits, we examined the frequencies and asymmetry parameters resulting from a preliminary fit. We then fit the frequency shift by fitting a sixth order polynomial in frequency, which we now add to the initial guess for the frequency. For the asymmetry parameter, we use a third degree polynomial in (adjusted) frequency directly for the initial guess.

[handwritten margin note: what is "adjusted Frequency"?]

When we tried the iteration scheme described above for the 15 periods we analyzed in detail, we found that for some of them very few f-modes were fitted. We therefore added an automatic rejection of fits with negative asymmetry parameters in the range $\nu < 2000\mu Hz$ between iterations of the peakbagging. This solved the problem for these 15 periods, but when we reanalyzed the entire mission, a small number of periods still had few f-modes fit. We were able to improve the coverage of those periods by adding a further criterion to reject modes that had an extremely high asymmetry parameter, but this caused other periods to lose modes. We therefore reverted to the initial rejection criteria. Clearly, the asymmetric fits are much less stable than those using symmetric profiles.

[handwritten margin note: which? ... FiTS n.]

4. Results

4.1. Mode Parameters

[handwritten margin note: 15 periods of 72 days each]

We applied 11 different analyses to 15 72 day periods beginning in January 2004 (see Table 2). To quantify the effect of each change in the processing on the mode parameters, for each one of the periods, we found the modes that were common between each analysis and the preceding one. We then took an average in time over whatever intervals had each mode successfully fit. In so doing, we are assuming that the difference in mode parameters resulting from the difference in the analysis is much more significant than their relative change over time. In the following figures, we plot the difference in several mode parameters normalized by their error estimates. For these plots, we calculated the average error estimates, rather than the error on the average, and for any given comparison between two analyses, we use the larger error estimate of the two. Thus the significance we have plotted is the least that one might expect from a single 72 day fit. The range of some plots excludes a few outliers; this is always less than 2% of the data [7]. The sense of subtraction is the later analysis minus the earlier one.

[handwritten margin note: nice!!]

---

[7]Here we have plotted all parameter differences as a function of frequency. Full listings of all mode parameters for all periods and all analyses we performed are provided as ASCII tables in the electronic supplementary material.

plate scale

detrending

cubic dist.

gapfilling

p-angle

horiz. disp.

carrington inc.

woodard effect

ccd tilt

asymmetry

ν (mHz) ν (mHz)

Figure 4. Change in mean multiplet frequency resulting from each change as a function of frequency, in units of standard deviation.
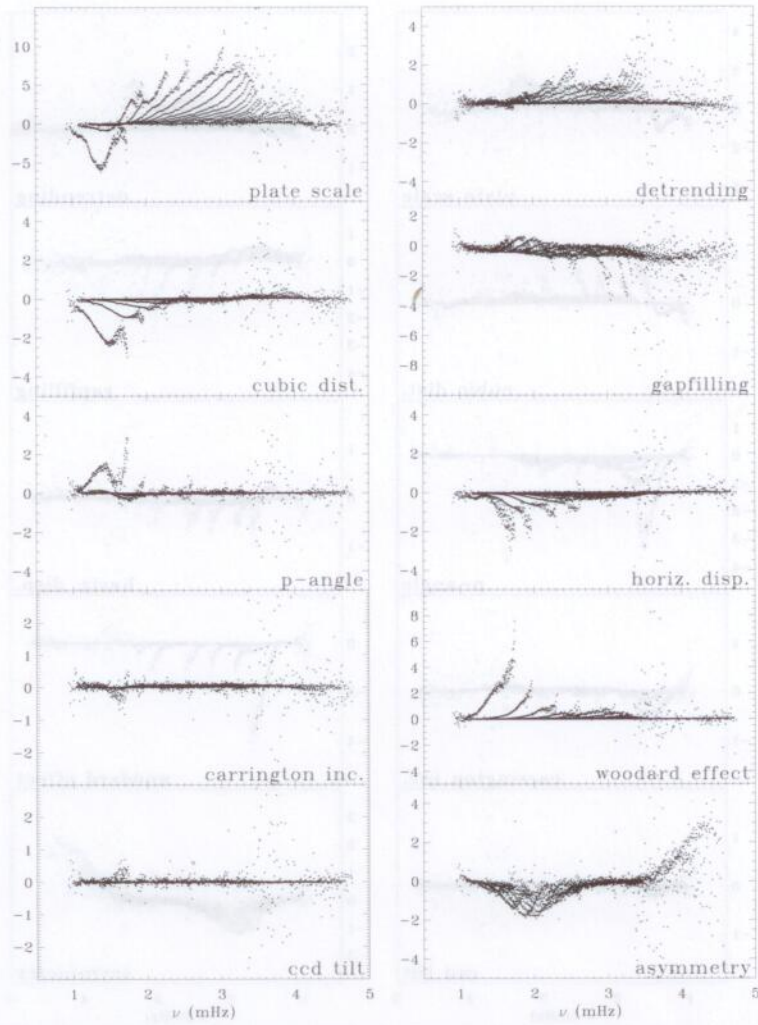
Figure 5. Change in amplitude resulting from each change as a function of frequency, in units of standard deviation.
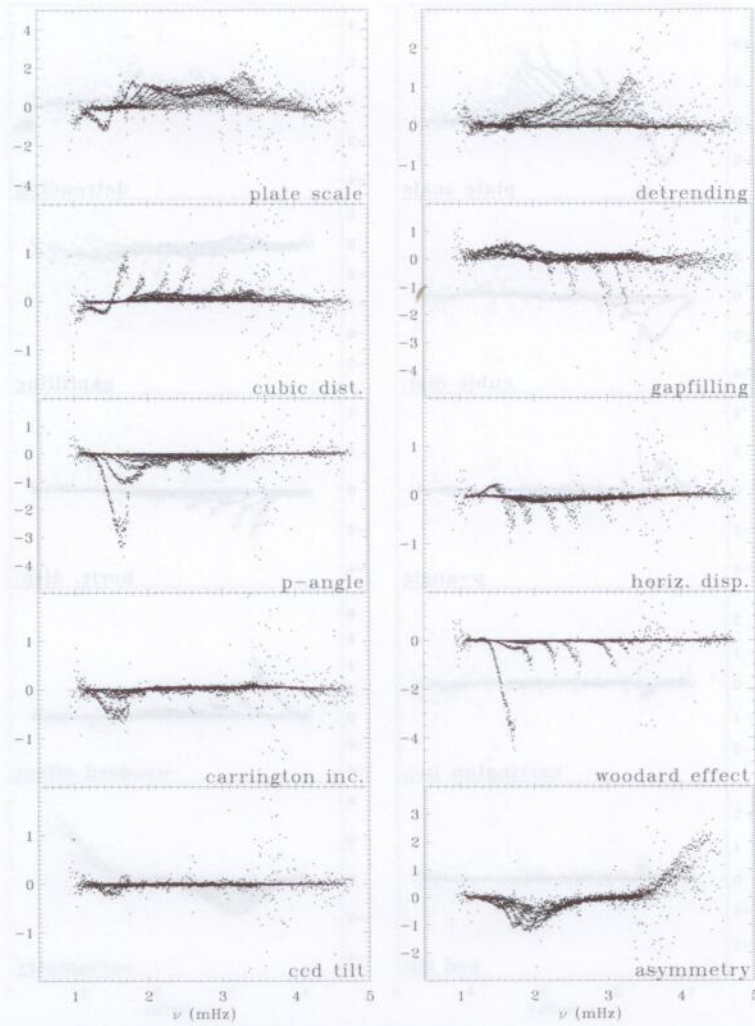
Figure 6. Change in width resulting from each change as a function of frequency, in units of standard deviation.
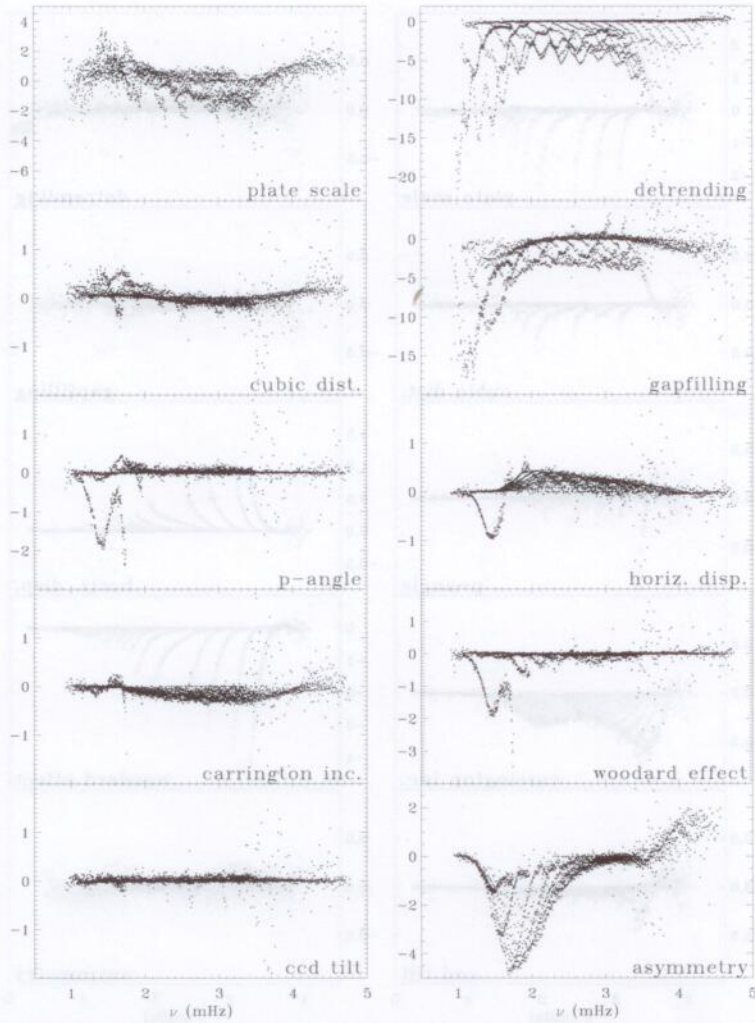
Figure 7. Change in background parameter resulting from each change as a function of frequency, in units of standard deviation.
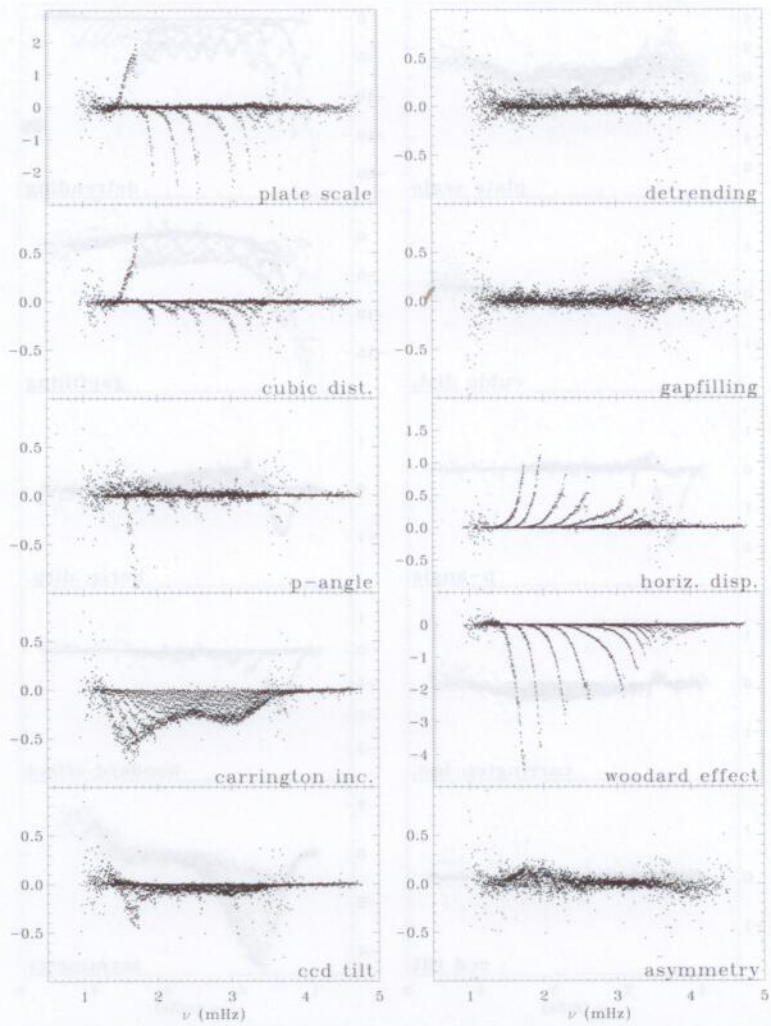
Figure 8. Change in a_1 resulting from each change as a function of frequency, in units of standard deviation.

As can be seen in Figure 4, the change in frequency was most significant for the plate scale correction and asymmetric fits. Including the horizontal displacement and correcting for distortion of eigenfunctions made the next most significant changes, followed by correcting for cubic distortion, in agreement with our previous work (Larson and Schou, 2008). Differences in detail between these and our previous results can mostly be attributed to the different method we have used for computing mode averages; by first taking the common mode set for each 72 day interval, the calculation of the averages becomes much more straightforward. For the plate scale correction, some of the difference in magnitude of the change in mode frequency can be attributed to the different epoch we reanalyzed. Previously we studied the two years beginning in January 2003, whereas in this work we study the three years beginning in January 2004, and the plate scale error is the only problem with the original analysis that is known to become worse over time. For the asymmetric fits, we used an improved iteration scheme for the asymmetry parameter, which seems to have resulted in a smaller change in frequency. The p-angle correction made a signficant difference for the *f*-mode, but otherwise this, the correction for the Carrington inclination error, and improved detrending and gapfilling did not result in much difference to the mode frequencies. We have also used a different method for calculating the Woodard effect, as described above, but we found this made less than a 0.5σ difference in all the parameters for a vast majority of modes. Therefore in all plots we show only the results of using the second method.

We find similar results for the amplitude and width (Figures 5 and 6), although for both of these parameters the detrending and gapfilling made much more significant differences. This is likely because these two changes in the processing made the dominant changes to the background parameter (Figure 7), as one might expect. We also point out that the large scatter of all three of these parameters just above 3.5 mHz indicates an instability of the fits in this frequency range, which may perhaps relate to the bump as well.

The changes in a_1 (Figure 8) have relative magnitudes roughly similar to the changes in frequency, the most notable exception being that correcting for the Woodard effect caused the dominant changes to this parameter. For the *f*-mode, correcting for the plate scale, cubic distortion, and P-angle error resulted in changes with the same sign as the frequency changes, but for the *p*-modes, and all modes when correcting for horizontal displacement and the Woodard effect, the changes had opposite sign. The changes in a_1 resulting from the Carrington inclination correction were more significant than the frequency changes, and show an interesting frequency dependence not seen in other parameters for this correction. The effect of the various changes on inversions of a_1 are discussed below (see Figure 15).

To see the effect of all the changes in the processing taken together, we examine the mission averages, formed as described above. Figure 9 shows the result for various mode parameters. For the *p*-modes, the error estimates were mostly unaffected. However, the set of all improvements up to and including the correction for the Woodard effect resulted in substantially lower error estimates for the *f*-modes, as shown in Figure 10. Unfortunately, using asymmetric line
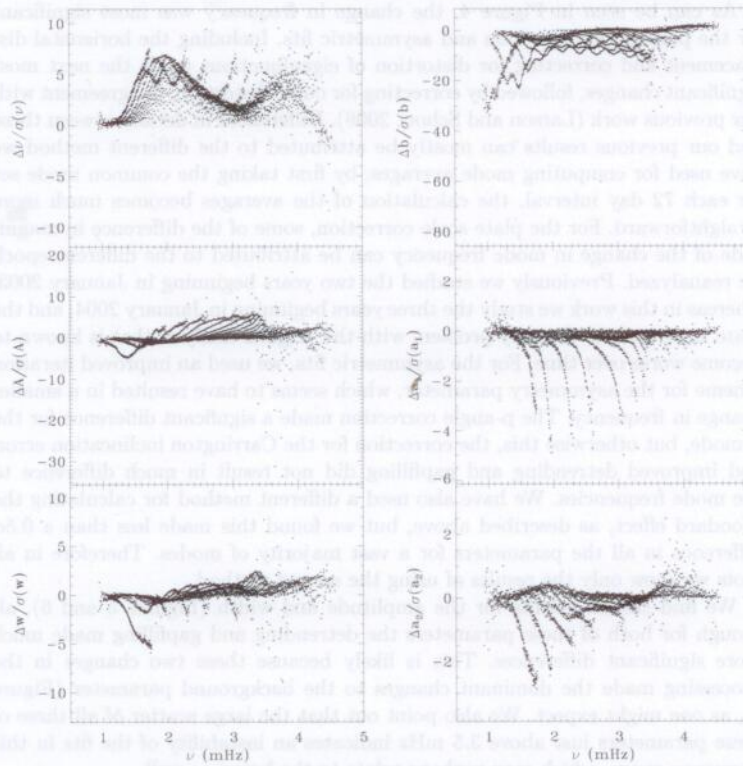
Figure 9. Change in frequency, amplitude, width, background parameter, a_1, and a_2 resulting from all changes as a function of frequency, in units of standard deviation. In these plots all points are included.

Say mission average + all effects

profiles resulted in substantially higher error estimates for the mode frequencies and background parameters, as shown in Figure 11.

One easy check of the robustness of our results is the comparison of the 72 day and one year analyses. Even without examining any mode parameters, one can see that the 360 day analysis was more successful in the sense that it was able to fit more modes, as shown in Figure 12. To compare the mode parameters, we averaged the results of five 72 day analyses (three for the third year) for the modes that were present in all of them. The errors used are the errors on the average. Then we formed common modesets as described above, this time taking the average error. The differences in mode parameters using asymmetric line profiles are shown in Figure 13 and the corresponding error ratios are show in Figure 14. The results were mostly similar using symmetric line profiles. To compare the background parameters, we subtracted log(5) from the 360 day fits.
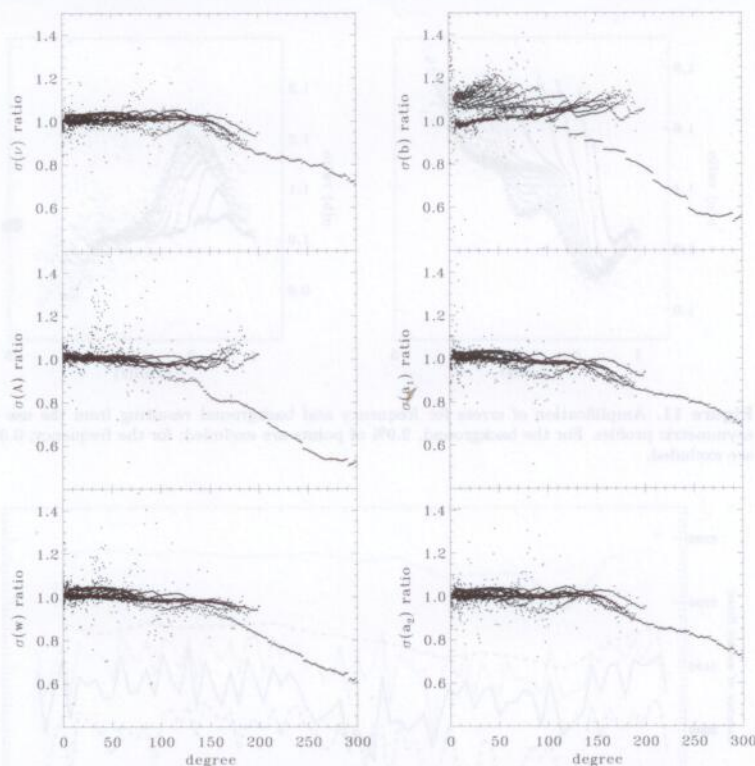
Figure 10. Ratio of the improved error estimates to the original error estimates as a function of degree for the parameters shown in Figure 9. The improved estimates do not include fitting asymmetric profiles. For the background, 2.1% of points do not fall within the range shown on the plots. For the other parameters, at most 0.5% of points are not shown.

Although the change in frequency seems to show a weak systematic dependence on frequency, the changes are mostly not significant. The change in frequency was slightly more significant using symmetric line profiles, especially at low frequencies. The changes in amplitude show ridge structure; although the majority of modes show reduced amplitude, the mean change is actually positive. The changes in width show ridges as well, but here the width is almost always less for the 360 day fits, and more so at lower frequencies, as one might expect, where the lower width will be better characterized by the increased frequency resolution. The background parameter shows the most significant changes (an increase except for the *f*-mode), but centered on the *p*-mode band, where the noise is nearly nonexistant. The changes in a_1 are the flattest, although a feature is discernible around 3.5 mHz. The asymmetry parameter was in general greater for the 360 day fits, with a peak around 1.8 mHz. For the frequency, width, and
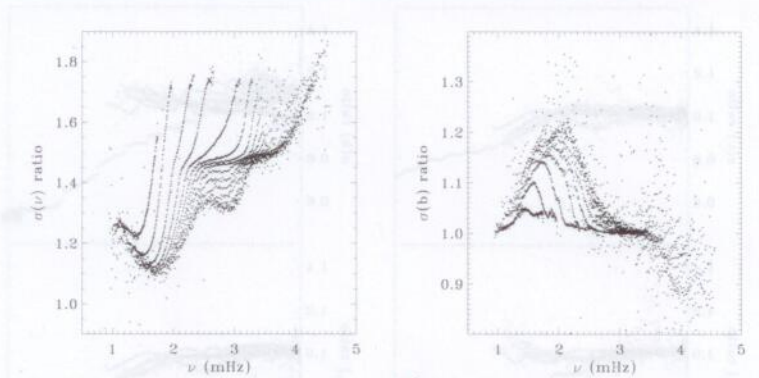
Figure 11. Amplification of errors for frequency and background resulting from the use of asymmetric profiles. For the background, 2.9% of points are excluded; for the frequency, 0.3% are excluded.
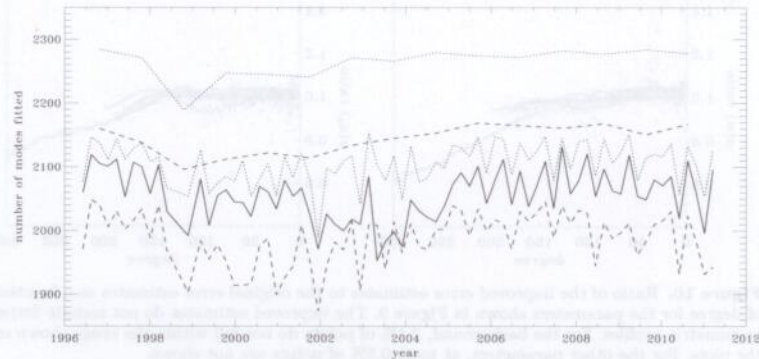
Figure 12. Number of modes fitted as a function of time for the five different ways we analyzed the entire mission. Dotted lines show the set of all changes in the processing up to correcting for the Woodard effect; dashed lines show the result of also using asymmetric line profiles. In both cases the higher line is for the 360 day fits, the lower line is for the 72 day fits. The solid line shows the original analysis.

a_1, the estimated errors were much lower for the 360 days fits at low frequencies, again as one might expect. Harder to understand is why the error on the asymmetry parameter increased in the same frequncy range. The background parameter also had lower errors, but again in the center of the frequency range.

4.2. Systematic Errors

In this section we begin to refer to the changes in processing by the order in which they were applied. This is summarized in Table 2.
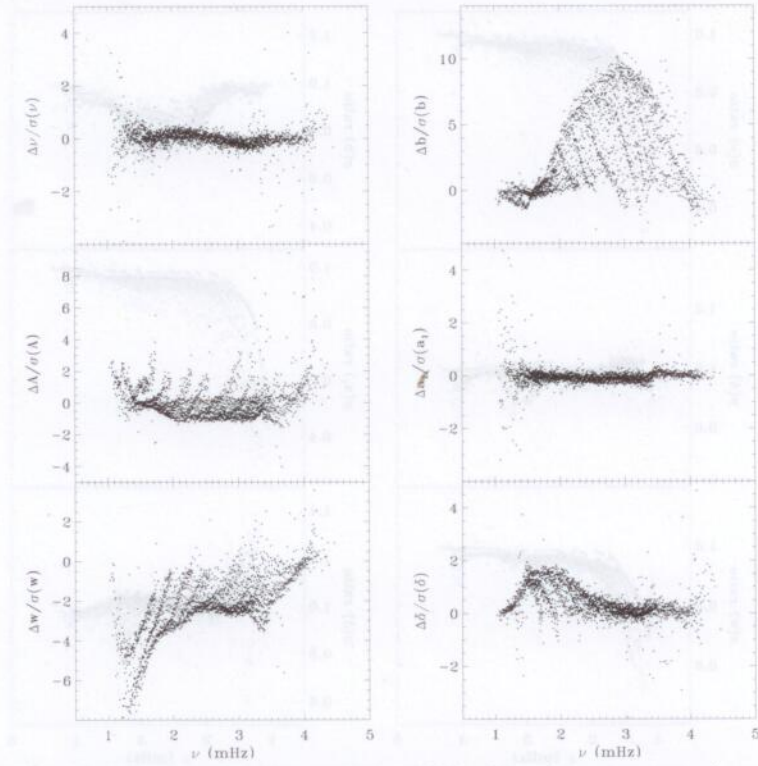
Figure 13. Difference in frequency, amplitude, width, background parameter, a_1, and asymmetry parameter between 360 day fits and an average of 72 day fits as a function of frequency, in units of standard deviation from the 360 day fits. The sense of subtraction is 360 day minus 72 day. At most 0.8% of points have been excluded.

To see the effect of the various changes on our systematic errors, we begin by performing simple one-dimensional regularized least squares inversions of the a_1 coefficient only. An RLS inversion seeks to minimize the the sum of normalized residuals squared plus a penalty term that serves to constrain rapid variations in the solution. In particular, we have chosen to minimize

$$\sum_{nl} \left[\frac{1}{\sigma(a_1(n,l))} \left(\int_0^1 K_{nl}(r)\bar{\Omega}(r)dr - a_1(n,l) \right) \right]^2 + \mu \int_0^1 \left(\frac{\mathrm{d}^2 \bar{\Omega}}{\mathrm{d}r^2} \right)^2 dr \quad (25)$$

where $\bar{\Omega}$ is the inferred rotation rate, the K_{nl} are (known) kernels calculated from the mode eigenfunctions that relate the rotation rate to a_1, σ is the standard error on a_1, r is fractional radius, and μ is the tradeoff parameter that controls the relative importance of the two terms. A low value of μ will fit the data better,
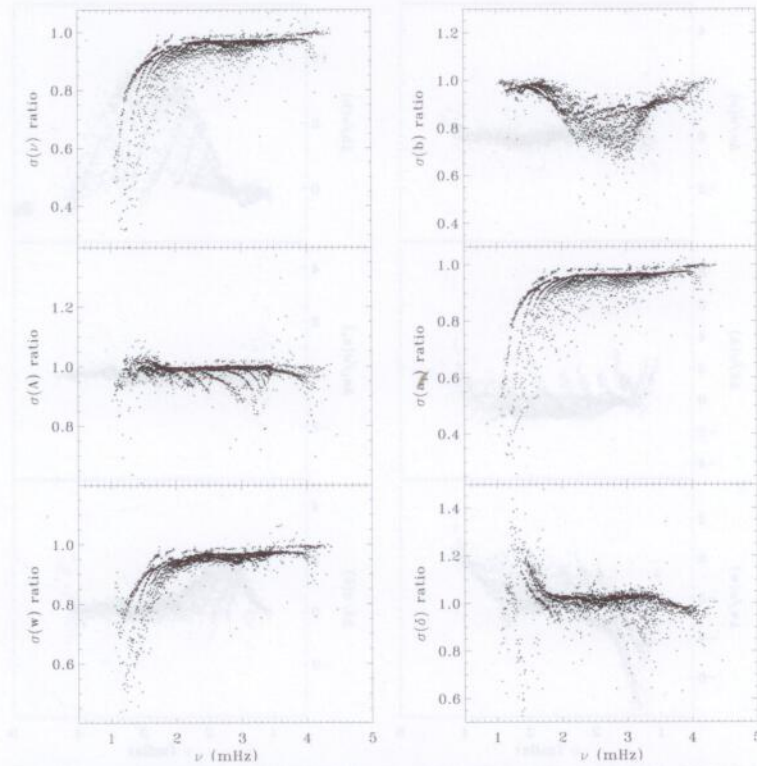
Figure 14. Ratio of the errors estimated from 360d fits to the errors estimated from an average of 72 day fits as a function of frequency for the parameters shown in Figure 13. At most 1.8% of points have been excluded.

but the solution may oscillate wildly as a function of radius. A higher value of μ will remove this feature (the solution will be more regularized) at the cost of increased residuals (Schou, Christensen-Dalsgaard, and Thompson, 1994). To choose a value of μ we have examined tradeoff curves, which are constructed by varying μ and plotting the rms of the residuals against the magnitude of the penalty term. The changes in a_1 that underlie the difference in the tradeoff curves for the different analyses were shown in Figure 8. The tradeoff curves themselves (shown in Figure 15) were computed using a modeset constructed by finding the modes common to all eleven analyses for each time period and taking the average in time over whatever modes were present; in this case the errors used are the errors on the average.

As one can see, the plate scale correction made a substantial difference to the tradeoff curve. The curve for cubic distortion is nearly indistinguishable. The P-angle correction made another significant reduction in the residuals, but the

Table 2. Sequence of changes made to the analysis; each analysis includes the changes made in all previous ones.

0	original analysis
1	plate scale
2	cubic distortion
3	P-angle error
4	Carrington inclination error
5	CCD tilt
6	window functions and detrending
7	gapfilling
8	horizontal displacemnt
9	distortion of eigenfunctions (Woodard effect)
10	asymmetric line profiles

curves for the next four changes to the analysis all lie between the previous two. Accounting for the horizontal displacement caused a substantial increase in the residuals, but accounting for the Woodard effect resulted in the lowest curve shown. The use of asymmetric profiles made no change to the tradeoff curve. This is basically in line with what one might expect from Figure 8. To choose a value of μ, one typically looks for the "elbow" in the tradeoff curve: the place where the residuals stop decreasing sharply, so that further decreases of μ will be of little benefit. Unfortunately, this "elbow" is not very pronounced in the curves shown in Figure 15. For the initial and final analyses, we have marked
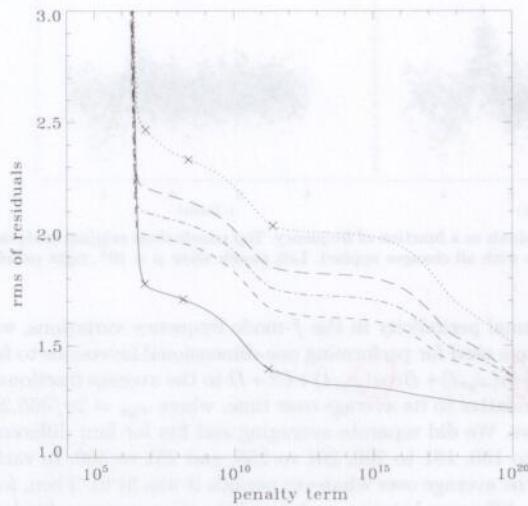
Figure 15. Tradeoff curves for several analyses. Dotted curve is for original analysis. Dash-dot curve shows first correction. Short-dashed curve shows first three corrections. Long-dashed curve shows first eight corrections (note this curve is above the one for only the first correction). Solid line is for all corrections. Symbols, from left to right, indicate tradeoff parameters of $\mu = 10^{-4}$, $\mu = 10^{-6}$, and $\mu = 10^{-9}$.

the point corresponding to the highest reasonable value of μ (10^{-4}), the lowest value one might possibly use (10^{-9}), and a value that is typically used in RLS inversions (10^{-6}).

In Figure 16 we show the normalized residuals of the inversions for the original and final analyses and for the smallest and largest values of μ given above. As one can see, the bump was mostly unaffected by all the changes in the analysis. A smaller value of μ decreases the size of the bump, but as Figure 17 shows, the resulting rotation profile is difficult to believe. The fact that the bump is only marginally present in the residuals for $\mu = 10^{-9}$ suggests that this systematic error is responsible for the "knee" in the tradeoff curves. Notably, even this small value of μ was not able to fit the "horns" in the original analysis, which are greatly reduced in the final analysis. This is likely the cause of the overall reduction in χ^2.
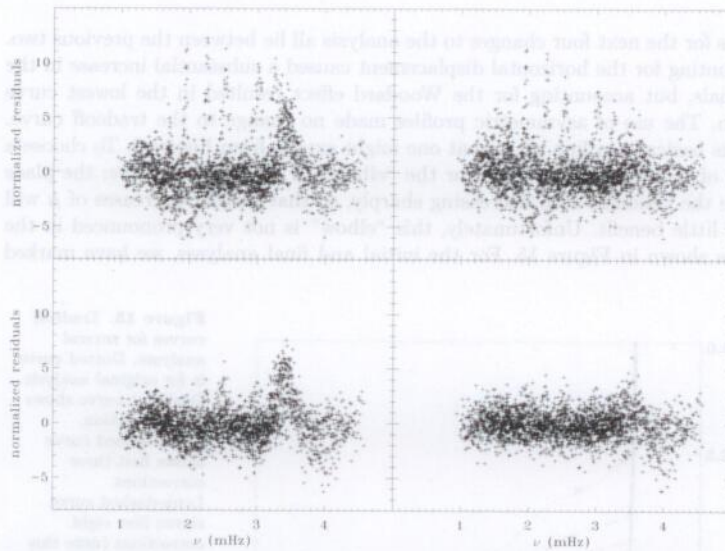
Figure 16. Normalized residuals as a function of frequency. Top panels show original analysis, bottom panels show analysis with all changes applied. Left panels show $\mu = 10^4$, right panels show $\mu = 10^{-9}$.

To investigate the annual periodicity in the f-mode frequency variations, we used the common modesets used for performing one-dimensional inversions to fit a function of the form $A\sin(\omega_{yr}t) + B\cos(\omega_{yr}t) + Ct + D$ to the average fractional f-mode frequency shift relative to its average over time, where $\omega_{yr} = 2\pi/365.25$ and t is measured in days. We did separate averaging and fits for four different ranges in degree l: 101 to 150, 151 to 200, 201 to 250, and 251 to 300. In each case, for each l we took the average over whatever periods it was fit in. Then, for each period, we took the difference between each l and the time average, divided
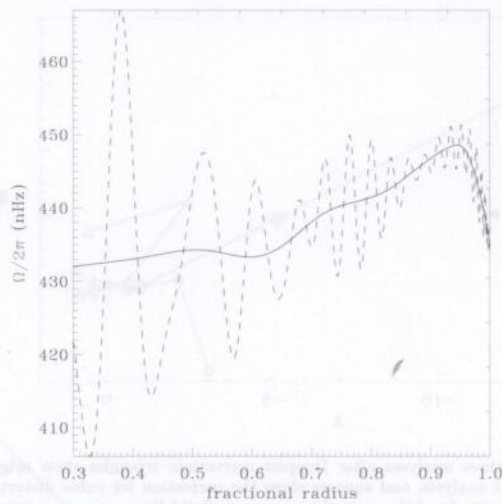
Figure 17. Internal rotation as a function of radius for the final analysis; curves for original analysis are similar. The solid line is the inversion result using $\mu = 10^{-4}$; the dashed line uses $\mu = 10^{-9}$.

by the time average, and then averaged over the range in l. We performed a weighted least squares fit to this data, which yielded values for the parameters A, B, C, and D and their corresponding errors.

The images produced by MDI, however, are taken at equal intervals of time on the spacecraft, whereas they should be taken at equal intervals of time on the Sun. To correct for this effect, we applied the relativistic Doppler shift due to the motion of the spacecraft. That is, we multiplied each frequency and its error by $\sqrt{(c+v)/(c-v)}$ where c is the speed of light and v is the average velocity of the spacecraft away from the Sun, as derived from the OBS_VR keyword of the input dopplergrams for each 72 day period. The resulting fits are shown in Figure 18, as well as the shift caused by the Doppler correction.

The amplitude of the annual component has a large variation between the different analyses, but in general it is always greater for the higher ranges in l. The point in the plot for l=251-300 of the original analysis contradicts that trend, but it must be noted that the fit represented by that point was an extremely poor one, which is likely due to the horns in the original analysis. For the lower two ranges in l, the amplitude was only marginally significant. Although not shown here, we note that the slope C was zero for the lowest range in l, and become steadily more negative as l increased, in agreement with previous findings (Antia *et al.*, 2001).

Finally, to explore the high-latitude jet, we used the 36 *a*-coefficient fits to perform two-dimensional RLS inversions for internal rotation. We formed common mode sets and averaged them using the same method described above for inversions, and used tradeoff parameters of $\mu_r = 10^{-6}$ and $\mu_\theta = 10^{-2}$ for the radial and latitudinal regularization terms respectively. Using a relatively
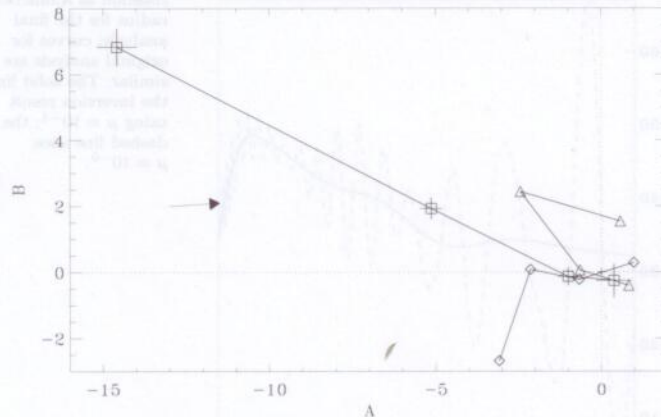
Figure 18. Phase plot for three analyses after Doppler correction: triangles show original analysis, diamonds show final analysis, and squares show the correction for cubic distortion, which yielded the largest amplitude of the annual component. Solid lines connect points for different ranges in l, beginning with the lowest range on the lower right. The arrow shows the shift resulting from the Doppler correction. The errors on A and B were similar for all analyses; the error bars show an average value. All values have been multiplied by 10^6 to match the units in Figure 3.

high value for μ_θ should dampen variations in latitude. The results are shown in Figure 19; the jet is more pronounced in this plot than in Figure 2, which can be attributed both to the different mode set and to the smaller errors. Although every change actually increased the magnitude of the polar jet, the gapfilling resulted in a reduced rotation rate in the lower convection zone, which brings our result closer to agreement with inferences drawn by the GONG analysis (Schou *et al.*, 2002).

5. Discussion and Future Prospects

We have found that the various changes we made to the processing of medium-l data from MDI resulted in significant changes in mode parameters. In summary, changes in width were overall the least significant, followed by the changes in a_1, which mostly resulted from applying the Woodard effect. The background was largely unaffected by most changes except the improved detrending and gapfilling. The plate scale correction made the dominant changes to the amplitudes and frequencies. For the latter, large changes also resulted from accounting for asymmetry, horizontal displacement, the Woodard effect, and cubic distortion, in decreasing order of significance.

Not only is one led to believe these changes represent an improvement as a matter of principle, but some of the systematic errors in the analysis have been
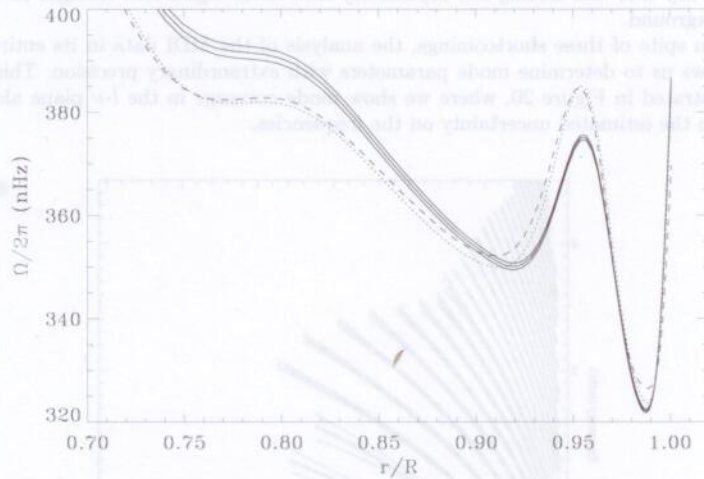
Figure 19. Internal rotation as a function of radius at 75° latitude for three analyses. Solid lines show the original analysis and its error bars; errors on the other analyses are similar. The dotted curve is the analysis that includes the improved gapfilling, and the dashed curve is the final analysis.

reduced as well. In particular, the "horns" have been greatly reduced, resulting in overall lower residuals from rotation inversions. A more stubborn systematic error is the "bump" in the odd *a*-coefficients, which seems to be reflected in the anomalous shape of the tradeoff curve. This remained almost completely unchanged in all analyses. Nor did any change to the analysis make a reduction in the high-latitude jet just below the solar surface, although there is an indication of improvement in the lower convection zone.

Regarding the annual periodicity in the *f*-mode frequencies, we found that the first change we applied, the plate scale correction, resulted in a drastically increased magnitude of the annual component for the higher two ranges in *l*. The correction for cubic distortion resulted in an even higher amplitude. After correcting for the P-angle error, however, the amplitude was reduced and did not vary much for later changes. We conjecture that the original fits were so poor at high *l* (the "horns") that the one-year period was not discernible there. The first correction was the most significant one for the frequencies, and hence uncovered the one-year period. Then the first correction that itself has a one-year period greatly reduced it.

Of concern to us is the discrepancy between the 360 day analysis, which in principle should be more accurate, and the 72 day analysis. Most notably, it indicates a problem with our model of the background. Interestingly, the asymmetry was the only parameter for which the error was greater for the

360 day fits, and adding the asymmetry also made significant changes to the background.

In spite of these shortcomings, the analysis of the MDI data in its entirety allows us to determine mode parameters with extraordinary precision. This is illustrated in Figure 20, where we show mode coverage in the l-ν plane along with the estimated uncertainty on the frequencies.
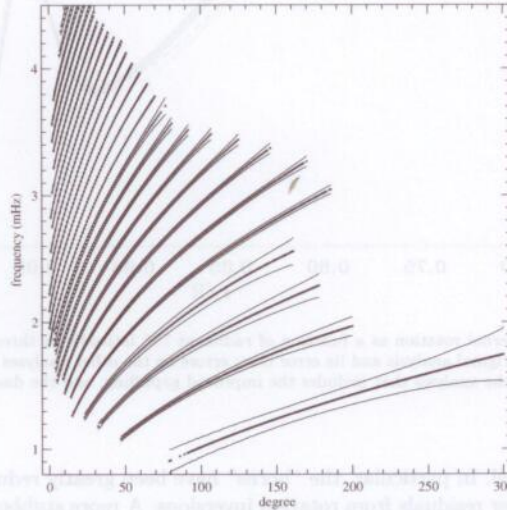
Figure 20. An l-ν diagram with magnified errors. Dots represent a mode that was fitted in at least 12 of the 15 years we analyzed using symmetric profiles. Solid lines show the errors: for the f-mode, these have been multiplied by 100000. For $n = 1, 2, 3$ the errors have been multiplied by 30000. The next eight ridges (n=4-11) have errors multiplied by 10000. The remaining ridges have errors multiplied by 1000.

Although our analysis has in general been very successful, the core peak-bagging routines were written at a time when computational capabilities were far less than now. A number of approximations which were necessary 20 years ago could now be lifted. The current work is an attempt to remove some of these limitations. Over the years, other workers in the field have also made contributions to the problem of inferring physical properties of the Sun from medium-l MDI data, among them XXX.

A potential drawback facing these efforts is that they utilize the leakage matrix calculated by us. In general, the use of a leakage matrix should increase the stability of fits, but the results will then depend upon the assumptions that went into its calculation. In particular, one might consider using leakage matrices calculated for different observer distances and values of B0. This has been attempted by XXX. The assumption that the real-to-imaginary and the imaginary-to-real leaks are zero could be lifted, but at the cost of considerable more complexity in the fitting codes. XXX have proposed doing so.

For us, there are a number of ways to move forward. The most obvious is the extension of this work to other datasets. First and foremost of these must be the MDI full disk data, which will allow us to determine how systematic errors and mode parameters might depend on the smoothing of the medium-*l* data and its apodization. Because of its duty cycle the full disk data cannot be used to study the annual periodicity in our results, but now the Helioseismic and Magnetic Imager (HMI) onboard the *Solar Dynamics Observatory (SDO)* has taken a long enough span of data for it to be suitable for this purpose. It has been suggested that the one year period may be related to the variable (in solar coordinates) width of the gaussian used for smoothing the medium-*l* data; an analysis of the MDI medium-*l* proxy from HMI should elucidate the issue. Finally, a repetition of the comparison with GONG results is long overdue. The original comparisons all used GONG classic data; now that GONG+ has been in place for over 13 years and software pipelines in both projects have been updated, the time has come to renew an investigation of the systematic differences between the two.

There also still remain possibilities for progress with the MDI medium-*l* data itself. One that is suggested by the results of this paper is to correct the timeseries for the relative motions of *SOHO* and the Sun. Although we can correct the frequencies after the fitting by Doppler shifting them, there is no obvious way to correct the other mode parameters. Another change in the analysis that suggests itself is to the width of the fitting window, since this is one of the things most notably different in the GONG analysis and is also known to affect the shape of the bump in the *a*-coefficients. During the remapping performed prior to spherical harmonic decomposition, we could implement an interpolation algorithm that takes into account the correlation between points introduced by the gaussian smoothing. We have also considered the common practice of zero-padding our timeseries before performing Fourier transforms. Lastly, the parameter space of the detrending and gapfilling remains almost entirely unexplored.

Appendix

Data Access

Detailed information on how to access MDI data from the global helioseismology pipeline can be found at http://jsoc.stanford.edu/MDI/MDI_Global.html. This page contains documentation describing how the datasets used in this paper were made and how they can be remade. In this appendix we describe how to access the relevant archived data. In what follows we must assume some familiarity with the Data Record Management System (DRMS), detailed documentation for which is linked from the above website.

Mode parameter files for every analysis discussed in this paper are available in the electronic supplementary material. For the original analysis, they (and a helpful Readme file) can also be found at http://sun.stanford.edu/~schou/anavw72z/. For all other analyses, they can also be retrieved from JSOC. The fields of a mode parameter file are the following: l, n, ν_0, A, w, b, x, $\{\delta\}$, $\sigma(\nu_0)$, $\sigma(A)$, $\sigma(w)$, $\sigma(b)$, $\sigma(x)$, $\{\sigma(\delta)\}$, a_1, a_2, ... a_N, $\sigma(a_1)$, $\sigma(a_2)$, ... $\sigma(a_N)$. The parameter δ and its error will not be present for fits done with symmetric profiles. The value of N is either 6, 18, or 36. Any parameter with zero error has not been fit for. The parameter x is not fit for in these analyses and is retained for historical reasons.

The data for the different "corrections" are labelled by the strings "corr1" to "corr9" corresponding to the numbering scheme in Table 2. The final correction in this set refers to the first way of applying the Woodard effect (holding B_1 and B_2 constant). These data have all been generated in the first author's name space, with mode parameters found in su_tplarson.corr_vw_V_sht_modes. The primekeys are T_START, LMIN, LMAX, NDT, and TAG, where T_START is the beginning of the corresponding timeseries, most easily specified by the MDI day number suffixed by "d" (see Table 1). For all records in this series, LMIN=0, LMAX=300, and NDT=103680, so these primekeys need never be specified. The TAG keyword is the label string, so TAG and T_START uniquely specify every record.

The second way of applying the Woodard effect, as well as the asymmetric fits, are both represented in the official MDI name space (mdi). For the former, mode parameters can be found in mdi.vw_V_sht_modes and for the latter in mdi.vw_V_sht_modes_asym. The primekeys are the same as given above, with the exception that these series do not have the TAG keyword and that NDT=518400 for the 360 day fits. In addition, the results used in this paper have the VERSION keyword (not a primekey) in these series set to "version2". If these data are reprocessed in the future, VERSION will get a new value.

The dataseries containing timeseries and window functions in the mdi name space have also been archived and can be retrieved; details on these data products are given on the above website. The corresponding data in the su_tplarson name space have not been archived, but can be recreated if needed. Again, the procedure for doing so can be found on the website. The original timeseries and window functions have been archived in the dsds namespace, but have not yet been ported to the standard DRMS format for global helioseismology data products. They can, however, still be retrieved by request.

References

Anderson, E.R., Duvall, T.L. Jr., Jefferies, S.M.: 1990, Modeling of solar oscillation power spectra. *Astrophys. J.* **364**, 699 – 705. doi:10.1086/169452.

Antia, H.M., Basu, S., Pintar, J., Schou, J.: 2001, How correlated are f-mode frequencies with solar activity? In: Wilson, A., Pallé, P.L. (eds.) *SOHO 10/GONG 2000 Workshop: Helio- and Asteroseismology at the Dawn of the Millennium*, ESA Special Publication **464**, 27 – 32.

Beck, J.G., Giles, P.: 2005, Helioseismic Determination of the Solar Rotation Axis. *Astrophys. J. Lett.* **621**, L153 – L156. doi:10.1086/429224.

Duvall, T.L. Jr.: 1982, A dispersion law for solar oscillations. *Nature* **300**, 242. doi:10.1038/300242a0.

Fahlman, G.G., Ulrych, T.J.: 1982, A New Method for Estimating the Power Spectrum of Gapped Data. *Mon. Not. Roy. Astron. Soc.* **199**, 53.

Korzennik, S.G., Rabello-Soares, M.C., Schou, J.: 2004, On the Determination of Michelson Doppler Imager High-Degree Mode Frequencies. *Astrophys. J.* **602**, 481–516. doi:10.1086/381021.

Larson, T.P., Schou, J.: 2008, Improvements in global mode analysis. *Journal of Physics Conference Series* **118**(1), 012083. doi:10.1088/1742-6596/118/1/012083.

Libbrecht, K.G.: 1992, On the ultimate accuracy of solar oscillation frequency measurements. *Astrophys. J.* **387**, 712–714. doi:10.1086/171119.

Nigam, R., Kosovichev, A.G.: 1998, Measuring the Sun's Eigenfrequencies from Velocity and Intensity Helioseismic Spectra: Asymmetrical Line Profile-fitting Formula. *Astrophys. J. Lett.* **505**, L51–L54. doi:10.1086/311594.

Rhodes, E.J. Jr., Reiter, J., Schou, J., Kosovichev, A.G., Scherrer, P.H.: 2001, Observed and Predicted Ratios of the Horizontal and Vertical Components of the Solar p-Mode Velocity Eigenfunctions. *Astrophys. J.* **561**, 1127–1143. doi:10.1086/323260.

Schou, J.: 1992, On the Analysis of Helioseismic Data. PhD thesis, , Aarhus University, Aarhus, Denmark, (1992).

Schou, J.: 1999, Migration of Zonal Flows Detected Using Michelson Doppler Imager F-Mode Frequency Splittings. *Astrophys. J. Lett.* **523**, L181–L184. doi:10.1086/312279.

Schou, J., Bogart, R.S.: 2002, Reduction of Systematic Errors in MDI Measurements. In: *American Astronomical Society Meeting Abstracts #200, Bulletin of the American Astronomical Society* **34**, 645.

Schou, J., Brown, T.M.: 1994, Generation of artificial helioseismic time-series. *Astron. Astrophys. Suppl.* **107**, 541–550.

Schou, J., Christensen-Dalsgaard, J., Thompson, M.J.: 1994, On comparing helioseismic two-dimensional inversion methods. *Astrophys. J.* **433**, 389–416. doi:10.1086/174653.

Schou, J., Howe, R., Basu, S., Christensen-Dalsgaard, J., Corbard, T., Hill, F., Komm, R., Larsen, R.M., Rabello-Soares, M.C., Thompson, M.J.: 2002, A Comparison of Solar p-Mode Parameters from the Michelson Doppler Imager and the Global Oscillation Network Group: Splitting Coefficients and Rotation Inversions. *Astrophys. J.* **567**, 1234–1249. doi:10.1086/338665.

Vorontsov, S.V.: 2007, Solar p modes of high degree l: coupling by differential rotation. *Mon. Not. Roy. Astron. Soc.* **378**, 1499–1506. doi:10.1111/j.1365-2966.2007.11894.x.

Woodard, M.F.: 1989, Distortion of high-degree solar p-mode eigenfunctions by latitudinal differential rotation. *Astrophys. J.* **347**, 1176–1182. doi:10.1086/168206.