HMI global helioseismology data analysis pipeline

Tim Larson, Jesper Schou
Hansen Experimental Physics Lab (HEPL), Stanford University, Stanford, CA
E-mail: tplarson@sun.stanford.edu

Abstract. The HMI global helioseismology data analysis pipeline is based largely on the
MDI medium-l program. All of the modules that ran in the SOI Science Support Center have
been ported for use in the SDO Joint Science Operations Center (JSOC) and given greater
functionality. Many errors and approximations which are present in the standard MDI pipeline
have been corrected and improvements have been added. Scripts have been written to automate
the submission of compute jobs to our local cluster; it is now possible to go from dopplergrams
to mode parameters with the push of a button. JSOC dataseries have been created to hold
all intermediate data products, timeseries, window functions, and mode parameters. Here we
discuss the operation of the pipeline, the structure of the data it generates, and access to the
same.

1. Overview
The pipeline begins with dopplergrams as input. These are apodized in fractional solar radius
and remapped to a regular grid in longitude and sin(latitude). Each image is then decomposed
into its spherical harmonic components of degree $l$ and azimuthal order $m$, although it must
be noted that since we cannot see the entire surface of the Sun, the modes cannot be perfectly
separated by this operation. A critical component of the pipeline is thus the generation of a
leakage matrix which quantifies how much each mode leaks into its neighbors. The next step is
to construct timeseries of the complex spherical harmonic amplitudes. These can be chunked in $l$
and adjusted in length to suit our I/O and memory restrictions, but they will eventually be retiled
into longer timeseries containing only a single $l$ each. At this point gaps and discontinuities in
the timeseries are identified. Next the timeseries are detrended by polynomial subtraction and
gaps are filled using an autoregressive algorithm. Then a Fourier transform is performed, and
the peaks in the transforms themselves are fit using a maximum likelihood minimization, which
also takes into account the leakage matrix. The result is the peak frequency, amplitude, and
linewidth for each $l$ and radial order $n$ the code was able to fit. Also fit for are the $a$-coefficients,
which specify a polynomial giving the $m$ dependence of the peak frequency. Modes with the
same $l$ and $n$ are assumed to have the same amplitude and linewidth for all $m$. Once we have
these mode parameters they can be inverted for sound speed and differential rotation.

2. Improvements
One advantage of this pipeline is that the same software can also process the MDI data that
has been ingested into JSOC. In this case we are able to correct in the remapping for change in
plate scale, p-angle error, cubic distortion from optics, and alleged CCD tilt. In the HMI data,
these type of corrections have already been applied to the dopplergrams, but the remapping
still benifits from a correction to the Carrington inclination and an improved algorithm for interpolation. The gapfilling implemented increases the duty cycle and improves the results at high frequencies relative to the standard MDI gapfilling. The peakbagging can take into account distortion of eigenfunctions by differential rotation, the horizontal component of oscillations, and asymmetric line profiles. In this last case another parameter is fitted for each \( l \) and \( n \) and assumed to be the same for all \( m \). For the effect of these various corrections on mode parameters, see our previous work in [1].

3. Data
We have the option of outputting all intermediate data products, but the pipeline will normally only produce data corresponding to timeseries. The keywords describing the time period that a particular data record refers to are given in table 1. \( T_{\text{STEP}} \) will be a constant in any given data series, and equal to the CADENCE keyword of the input dopplergrams. Therefore a time period can be uniquely specified by \( T_{\text{START}} \) and \( NDT \), and its length will be \( NDT*T_{\text{STEP}} \). \( T_{\text{START}} \) can be specified as a date string or as an offset from the MDI epoch of 1993.01.01 TAI. When a data record additionally refers to a range of spherical harmonic degrees, these must be additionally specified in order to select a unique record. When a data record contains only a single \( l \) then we will have \( \text{LMIN} = \text{LMAX} \), so these two are usually redundant.

| Keyword     | Type | Definition                                                                 |
|-------------|------|-----------------------------------------------------------------------------|
| \( T_{\text{START}} \) | time | the beginning of the time interval a record corresponds to.                 |
| \( NDT \)   | int  | number of time points represented.                                          |
| \( T_{\text{STEP}} \) | float | length of a time step in seconds, usually a constant.                      |
| \( T_{\text{STOP}} \) | time | the beginning of the following timeseries, or \( T_{\text{START}} + NDT * T_{\text{STEP}} \). |
| \( T_{\text{OBS}} \) | time | the midpoint of a timeseries as given by \( (T_{\text{START}} + T_{\text{STOP}}) / 2 \).  |
| \( \text{LMIN} \) | int  | minimum spherical harmonic degree represented.                             |
| \( \text{LMAX} \) | int  | maximum spherical harmonic degree represented.                             |

4. Processing plan
Our current plan is to run the first part of the pipeline once per day, producing one day timeseries with \( l = 0 \) to \( l = l_{\text{max}} \). Currently \( l_{\text{max}} = 1800 \), and these timeseries shall be archived. Every 36 or 72 days the one day timeseries will be retiled into longer timeseries with one \( l \) each. These are detrended and gapfilled and the resulting timeseries shall also be archived. Peakbagging and whatever other algorithms are in the pipeline will be run, and Fourier transforms and power spectra will be created as needed. The resulting mode parameters shall be archived, and finally will be inverted. All of these data products, as well as window functions, shall be available for download. The source code and scripts that create them are also available online.

5. Other codes and future work
The pipeline described above processes spherical harmonic degrees in the range \( l = 0-300 \). Higher than this individual modes blend into ridges and other algorithms must be employed. One such algorithm that we have implemented in the pipeline is the WMLTP method of Reiter and Rhodes, which can fit \( m \)-averaged power spectra to above \( l = 1000 \) [2]. Another high-\( l \) algorithm pending
integration with JSOC is the ridge-fitting method of Korzennik and Rabello-Soares [3]. More generally, any code that processes timeseries, power spectra, or Fourier transforms and that takes input from fits files can easily be incorporated into the global pipeline.

Another module under development shall simulate MDI medium-l data, which is convolved with a gaussian and subsampled, from HMI dopplergrams, as well as perform arbitrary binning. This has the dual advantage of reducing I/O and memory usage and also of giving HMI better continuity with MDI. We also plan to add corrections for difference in light travel time from different points on the disk and geometric errors related to the height of formation of spectral lines.

Acknowledgments
This work has been supported by NASA contract NAS5-02139.

References
[1] Larson, T. & Schou, J. 2008, J. Phys.: Conf. Series, 118, 012083
[2] Reiter, J. et al. 2010, in preparation
[3] Rabello-Soares, M. C., Korzennik, S. G. & Schou, J. 2008, Solar Phys., 251, 197