

# Global Helioseismology Pipeline

Tim Larson  
tplarson@sun.stanford.edu

HMI Team Meeting  
September 2009  
Stanford University

# Modules

- jv2helio – remaps dopplergrams to heliographic coordinates. input can be of any resolution, output are maps equally spaced in longitude and sin latitude.
- jhelio2mlat – performs fourier transform in longitude and transposes data.
- jqdotprod – performs projection onto spherical harmonics. output are timeseries chunked in l.
- jv2ts – combines the above 3 modules to save on I/O.
- jv2hmod – jv2helio modified to output dopplergrams corrected for distortion and interpolated to any resolution. useful for making HMI-like data from MDI data.
- jretile – input and output are timeseries, changes tiling in time and chunking in l.
- jtsfiddle – detrends and gapfills timeseries. output can be timeseries, fourier transforms, power spectra, etc.
- jpkbgn – extracts mode parameters from fourier transforms. currently input are timeseries, output are ascii.

# Status

- All modules are written and tested. Most are already in cvs.
- Shell scripts have been written to automate submission of jobs to computing cluster.
- JSOC dataserries have been created to hold all intermediate data products as well as timeseries.
- All modules automatically save processing metadata in an ancillary dataserries.

# To Do

- Complete testing of last 2 modules and add to cvs.
- Write script to tie all modules together, also to recreate data that has aged off disk. Replace shell scripts with perl scripts as appropriate. Add scripts to cvs.
- Measure load and run times to determine how to store data.
- Automate the detection of gaps and discontinuities in timeseries. We hope to derive this information from the housekeeping data, if not from QUALITY keyword.
- Write a module to simulate MDI vw\_V data from HMI dopplergrams, as well as arbitrary binning.
- Leakage matrix
- Document all modules and scripts.

# More Codes

- Ridge fitting – Cristina & Sylvain
- MPTS & WMLTP codes – Ed & Johann
- Sylvain's long timeseries code
- GONG PEAKFIND – help me Frank!
- 
- David Salabert's collapsogram code?
- Stuart Jeffries' metrology algorithm?

We need to make some artificial timeseries for intercomparison.  
It would also help to agree on an output format for mode parameters.

# Tentative Processing Plan

- Run the first part of the pipeline once a day, producing 1 day timeseries with  $l=0$  to  $l=l_{\max}$ . Do not make any intermediate data products except by request. What is  $l_{\max}$ ?
- Every 36 or 72 days, retile the above timeseries into longer timeseries containing one  $l$  each. Detrend and gapfill these timeseries. Should these and/or the raw ones be saved or archived?
- Run peakbagging and whatever other algorithms are in the pipeline. Create fourier transforms or power spectra as needed. Save mode parameters.
- Run inversions for sound speed and rotation.

FYI:

$l=0-1500 \Rightarrow 1,127,251$  m's

1 day = 1920 records at 45 sec cadence

storing floats  $\Rightarrow 17,314,575,360$  bytes/day  $\sim 17$ GB/day

36 days = 612 GB