HMI ring diagram analysis II. Data products

R. S. Bogart, C. Baldner, S. Basu, D. A. Haber and M. C. Rabello-Soares

1 Stanford University, Stanford, Calif. USA
2 Yale University, New Haven, Conn. USA
3 JILA, University of Colorado, Boulder, Colo. USA
E-mail: rbogart@stanford.edu

Abstract. The combination of high resolution, spatial coverage, and continuity of photospheric Doppler and other data from HMI has allowed us to embark on a program of systematic exploration of solar subsurface flows and thermal structure variations using the technique of ring-diagram analysis on an unprecedented scale. There are two ring-diagrams pipelines, as described in [1]. In this paper we discuss the synoptic pipeline execution and describe the data being processed and produced.

1. Data products
Input data for the ring-diagram pipelines are at present taken from the published and mirrored data series hmi\_test.V\_45s in the Data Record Management System (DRMS) as described in [2]. This is an unarchived prototype series for the forthcoming definitive HMI Level 1 observable data product hmi.V\_45s. The structure of the definitive series is not expected to change significantly from that of the prototype series. The various output data series are correspondingly being put in the DRMS namespace hmi\_test; when the input series is migrated to its official namespace the output series will be as well. The output data series are available for mirroring as requested by NetDRMS sites [2].

All ring-diagram pipeline data products are named according to the rule {NS}.rd\{OBS\}.{PROD}{{SER}}, where NS is the DRMS namespace (hmi for definitive products based primarily on HMI data), OBS is a code for the input observable (V for Doppler data), PROD is a product name describing the module producing the data series, and SER is a name specific to a particular pipeline. The product names for data produced by the various modules are: track for output of module mtrack; pspec for output of pspec3; fitsf for output of rdfitf; and flows for output of rdvinv. The implemented pipeline series codes are: fd05 for the synoptic 5° tiles; fd15 for the synoptic 15° tiles; fd30 for the synoptic 30° tiles; and targ for the target pipeline. The various products created in the synoptic 15° tile set, for example, are:

- hmi\_test.\_rdVtrack\_fd15
- hmi\_test.\_rdVpspec\_fd15
- hmi\_test.\_rdVfitsf\_fd15
- hmi\_test.\_rdVfitsf\_fd15
- hmi\_test.\_rdVflows\_fd15\_frame
A DRMS data series is organized by its metadata prime keys: the union of values of the prime keys is taken to uniquely define a record. Multiple records with the same values of their prime keys are assumed to refer to the same data, and only the one with the highest internal record number (the one most recently added to the database) is ordinarily provided by data requests. For an image series such as the input Dopplergrams it is normally sufficient to use a single prime key, the observation time; but for the pipeline data sets described here it is obviously necessary to provide at least three prime keys, designating both the time and the location of the tracked cubes or those to which the analysis products refer. In principle all of the products could be included in a single data series if we used more prime keys to designate for example the product type and the relevant pipeline or the parameters uniquely defining it, but that would be cumbersome, which is why we have chosen to create the 20 or more different series described above. Each of those series is also uniquely described by a number of non-prime keys whose values are constant in the database. This prevents for example a power spectrum from accidentally being ingested into a tracked cube series or a 5° tile from being ingested into a series reserved for 30° tiles.

Because the phasing of ring-diagram pipeline analyses is keyed to Carrington times, the product data series use two prime keys to distinguish the time: Carrington Rotation and Central Meridian Longitude at the midpoint of the analysis interval. Three prime keys are used to describe the spatial locations of the analysis regions: Heliographic Latitude, Carrington Longitude, and Central Meridian (Stonyhurst) Longitude. Although the three longitude keys are redundant, they are chosen for the convenience of being able to select certain groupings easily, for example all analyses for a given Carrington longitude, or all analyses on the central meridian for a certain time range.

For the products of the target pipeline, the prime keys are an Identifier (the active region number with an optional longitude offset for comparison regions), plus the Carrington Rotation and Carrington coordinates of the region. This assumes that all regions will be analyzed for intervals centred on the times of their central meridian crossings. If that constraint is relaxed at least one additional prime key will be required.

The synoptic pipeline has been run for all target intervals from Carrington time 2096:030 (17 May 2010) through 2100:035 (3 Sep 2010) — four full rotations. Since the data product series were for testing, no data were archived to tape, and the data on disc from times prior to 2098:330 (18 June 2010) have aged off. Continuous coverage in the analysis data series of the synoptic pipeline is currently (26 Sep 2010) available for the following Carrington frames:

- hmi_test.rdV*_fd30 : 2098:330 – 2100:060; 33 frames
- hmi_test.rdV*_fd15 : 2098:330 – 2100:045; 67 frames
- hmi_test.rdV*_fd05 : 2098:330 – 2100:035; 204 frames

Each frame includes a large number of analysis data cubes: 69 30° tiles, 284 or 281 15° tiles, and 2748 or 2727 5° tiles. The number and distribution of tiles change with $B_0$ as described in [1] and illustrated schematically in Figure 1.

Only one record set of the target pipeline exist so far — AR 11072 (2097:315.5,15.0S, 2010.05.23) — due to the lack of strong active regions at central meridian in the data to date.

2. Pipeline execution

In order to keep cadence, it is clear that the three components of the synoptic pipeline must be started once every 9, 27, and 54 hours. The pipeline has been running long enough to gain some
Figure 1. Schematic locations of the centres of the target grid points for the 15° and 5° synoptic data series. The figure on the left illustrates the 15° locations when $|B_0| < 3°.625$, and the one in the center for times when $B_0 \geq 3°.625$. The 5° grid on the tight is for $B_0 \leq -3°.625$. The shaded dots correspond in aspect ratio to their projected appearance, but are smaller than the actual extent of the regions, which overlap their neighbors by one half their width.

experience of its running times under various load conditions. The typical latency (the time to complete the pipeline for a single analysis time) of the 30° pipeline is about 20 hours, meaning that under normal conditions it is running only about 1/3 of the time. The 15° pipeline has a typical latency of about 18 hours, so it must be running about 2/3 of the time. The 5° pipeline has a minimum latency (without the inversions) of about 12 hours, which is longer than its required cadence; an average of 1.3 instances of it must be running at all times. During times of extreme system load the timings for the 5° pipeline have increased by as much as a factor of two. At these times the system load was so severe that all activities were affected and problems identified and corrected. The ring-diagrams pipeline itself has occasionally run for over a week at triple cadence with no noticeable impact on the system load nor on its own latencies.

The real latency of availability of data products of course involves more than the processing times for the ring-diagrams pipeline itself. There is a delay of at least one day for production of the definitive observables. Because the pipeline requires background subtraction of a 9-day mean, it cannot commence until at least 5 days after the target time. If full disc crossings are used for the initial tracking sets, this will require a latency of about 12 days (although the latency for the background mean will be subsumed in that).

The various elements of the pipeline contribute very differently to the overall processing time for the different synoptic pipelines. For the 30° pipeline, the **rdfitc** module (for just the nine regions on central meridian or seventeen regions on meridian or equator, running on parallel processors) takes more than half the time, while the flow inversions take a negligible fraction. For the 5° pipeline on the other hand, the inversions (if they were being carried out) would consume the largest fraction, while the **rdfitc** fitting can be run for all 2700+ regions in only half the tracking time. Relative timings for the individual module components of the different synoptic pipelines are given in Table 1. The relative timings for the alternate 5° pipeline are for ones in which **rdfitc** fits are done only on the central meridian, but for which the flow inversions are performed.

3. Other data sets
The pipeline elements have been run on Doppler data concurrently observed with both HMI and MDI for Carrington times 2096:150–2097:225 (8–30 May 2010). A sample comparison, of the averaged $U_x$ and $U_y$ parameters from the ring fits, is shown in Figure 2. The agreement of
Table 1. Relative timings for individual modules in the synoptic pipelines.

| Module   | 30° tiles | 15° | 5°   | 5°*  |
|----------|-----------|-----|------|------|
| mtrack   | 0.29      | 0.55| 0.44 | 0.28 |
| pspec3   | 0.07      | 0.15| 0.14 | 0.08 |
| rdfitf   | 0.08      | 0.15| 0.20 | 0.13 |
| rdfitc   | 0.54      | 0.10| 0.22 | < 0.01 |
| rdvinv   | 0.02      | 0.05| —    | 0.50 |

Figure 2. Average values of the \( U_x \) and \( U_y \) parameters (from rdfitc) for 15° tiles along the central meridian, averaged over all modes, as a function of latitude. The red symbols are for MDI data and the black symbols for HMI data for the same 10 time intervals. The difference in zonal velocity values reflects the difference in tracking rates for the MDI and HMI pipelines [1].

Values and dispersion is excellent, as we would expect from the similarity of size and scale of the regions analyzed.

References
[1] Bogart R S, Balchmer C, Basu S, Haber D A and Rabello-Soares M C 2011 J. Phys. Conf. Proc. this issue
[2] Bogart R S, 2007 Astron. Nachrichten 328 352