

The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: Overview and Performance

J. Todd Hoeksema · Yang Liu · Keiji Hayashi · Xudong Sun · Jesper Schou ·
Sebastien Couvidat · Aimee Norton · Monica Bobra · Rebecca Centeno · K.D. Leka ·
Graham Barnes · Michael Turmon

Received: 1 October 2013 / Accepted: 7 March 2014 / Published online: 25 March 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The *Helioseismic and Magnetic Imager* (HMI) began near-continuous full-disk solar measurements on 1 May 2010 from the *Solar Dynamics Observatory* (SDO). An automated processing pipeline keeps pace with observations to produce observable quantities, including the photospheric vector magnetic field, from sequences of filtergrams. The basic vector-field frame list cadence is 135 seconds, but to reduce noise the filtergrams are combined to derive data products every 720 seconds. The primary 720 s observables were released in mid-2010, including Stokes polarization parameters measured at six wavelengths, as well as intensity, Doppler velocity, and the line-of-sight magnetic field. More advanced products, including the full vector magnetic field, are now available. Automatically identified HMI Active Region Patches (HARPs) track the location and shape of magnetic regions throughout their lifetime.

The vector field is computed using the Very Fast Inversion of the Stokes Vector (VFISV) code optimized for the HMI pipeline; the remaining 180° azimuth ambiguity is resolved with the Minimum Energy (ME0) code. The Milne–Eddington inversion is performed on all full-disk HMI observations. The disambiguation, until recently run only on HARP regions,

The Many Scales of Solar Activity in Solar Cycle 24 as seen by SDO
Guest Editors: Aaron Birch, Mark Cheung, Andrew Jones, and W. Dean Pesnell

J.T. Hoeksema (✉) · Y. Liu · K. Hayashi · X. Sun · J. Schou · S. Couvidat · A. Norton · M. Bobra
W.W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, CA,
USA
e-mail: todd@sun.stanford.edu

R. Centeno
High Altitude Observatory (NCAR), Boulder, CO, USA

K.D. Leka · G. Barnes
Northwest Research Associates, Inc., Boulder, CO, USA

M. Turmon
Jet Propulsion Laboratory, Pasadena, CA, USA

J. Schou
Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg 3, 37077 Göttingen, Germany

is now implemented for the full disk. Vector and scalar quantities in the patches are used to derive active region indices potentially useful for forecasting; the data maps and indices are collected in the SHARP data series, `hmi.sharp_720s`. Definitive SHARP processing is completed only after the region rotates off the visible disk; quick-look products are produced in near real time. Patches are provided in both CCD and heliographic coordinates.

HMI provides continuous coverage of the vector field, but has modest spatial, spectral, and temporal resolution. Coupled with limitations of the analysis and interpretation techniques, effects of the orbital velocity, and instrument performance, the resulting measurements have a certain dynamic range and sensitivity and are subject to systematic errors and uncertainties that are characterized in this report.

Keywords Magnetic fields, photosphere · HMI: vector field · Solar active regions

1. Introduction

The *Helioseismic and Magnetic Imager* (HMI, Schou *et al.*, 2012b) provides the first uninterrupted time series of space-based, full-disk, vector magnetic field observations of the Sun with a 12-minute cadence. This report summarizes the HMI vector magnetic field analysis pipeline, provides an assessment of the quality of the vector data, and characterizes some of the known uncertainties and systematic errors.

HMI was launched on 11 February 2010 as part of the *Solar Dynamics Observatory* (SDO, Pesnell, Thompson, and Chamberlin, 2012) and began taking continuous solar measurements on 1 May. The HMI investigation (Scherrer *et al.*, 2012) studies the origin and evolution of the solar magnetic field and links to the dynamics of the corona and heliosphere by producing long uninterrupted time series of the global photospheric velocity, intensity, and magnetic fields.

HMI measures full-disk scalar quantities – Doppler shift, longitudinal magnetic field, continuum intensity, line depth, and line width – using a repeating sequence of narrow-band images recorded every 45 seconds with a 4096×4096 camera called the Doppler camera. This paper focuses primarily on the analysis of data from a second, identical Vector camera that measures linear and circular polarization with a 135 s cadence frame list (Schou *et al.*, 2012a). From these filtergrams the vector magnetic field and thermodynamical parameters are determined. See Figure 1.

Scientific observables are produced by a series of software modules operating in an analysis pipeline. HMI collects filtergrams almost continuously. The Level-0 reconstruction of the instrument data and computation of most quick-look observables is completed within a couple of minutes. Some preliminary analysis is completed using this near-real-time (NRT) data. Validation and filling of gaps in the scientific and housekeeping data stream typically takes a little more than a day, after which the final calibration and processing is initiated to produce definitive Level-1 filtergrams.

The filtergrams are calibrated further to correct a variety of instrument distortions and remove trends. Temporal and spatial interpolations are applied to remove the effects of cosmic rays and data gaps, and to correct for temporal evolution and solar rotation. Combining the calibrated filtergrams produces the scientific observables. The primary vector magnetic observable is the Stokes vector data series, `hmi.S_720s`, in which the I, Q, U, and V polarimetric quantities are determined every 12 minutes at six wavelengths (Section 2.2). Table 1 lists relevant HMI data series.

Active regions emerge quickly and span a large range of size and complexity. An automated code identifies and tracks HMI Active Region Patches (HARPs) using the HMI

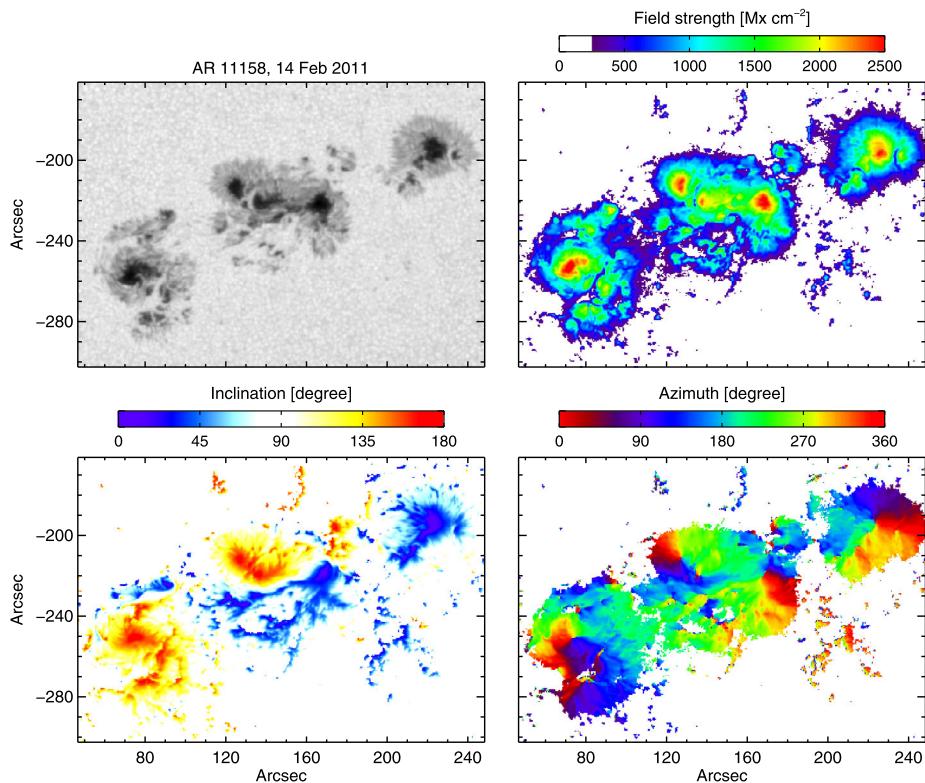


Figure 1 HMI active region patch 377 (NOAA AR #11158) produced the first X-class flare of Solar Cycle 24 early on 15 February 2011. This figure shows the intensity and magnetic-field components observed near disk center a few hours earlier, at 19:00 TAI on 14 February. The panel in the upper left shows the continuum intensity in CCD coordinates labeled in arc seconds from disk center. The image has been rotated 180° to put solar North near the top. The upper right panel shows the total field strength for pixels with total field strength above 250 Mx cm^{-2} , saturating red at 2500 Mx cm^{-2} . The angle of inclination relative to the line of sight appears in the lower left panel, with the field directed toward the observer shown in blue, the transverse field in white, and the field directed away from the observer in red. The lower right panel shows the azimuth angle, adjusted to give the angle relative to the direction of rotation (*i.e.* West). Only pixels above a 250 G threshold are plotted. Data are from the series `hmi.sharp_720s`.

720-second line-of-sight (LoS) magnetic field and computed continuum intensity images (Turmon *et al.*, 2014 and Section 3). HARPs are analogous to NOAA Active Regions, but there are many more smaller patches. HARPs aim to identify complete flux complexes, and so may be larger and incorporate more than one NOAA AR. HARPs follow a region during its entire lifetime or disk passage, beginning one day before the first flux emerges and continuing one day after the region has decayed. The HARP data series, `hmi.Msharp_720s`, defines the geometry of a rectangular bounding box on the CCD that encloses the maximum heliographic area achieved by a patch during its lifetime. At each time step a rectangular bitmap identifies which CCD pixels lie within the patch.

To recover useful information about the Sun's vector magnetic field an inversion of the Stokes vector must be performed that requires specific assumptions about the solar atmosphere. Such an inversion can be computationally intensive. Section 4 briefly describes how the `fd10` version of the VFISV code (Borrero *et al.*, 2011; Centeno *et al.*, 2013) that is

Table 1 HMI magnetic field and related data series available in JSOC.

| Data series ^a name | HMI data product |
|-------------------------------------|---|
| <i>Scalar observables</i> | |
| hmi.S_720s | Level 1 Stokes-polarization images – 720 s averages 24 images: I Q U V at six wavelengths |
| hmi.M_720s ^b | Full-disk line-of-sight magnetic field |
| hmi.M_45s | 45-second full-disk line-of-sight magnetic field |
| hmi.V_720s | Full-disk line-of-sight Doppler velocity |
| hmi.Ic_720s | Full-disk computed continuum intensity |
| hmi.Ic_noLimbDark_720s ^b | Ic with limb darkening removed |
| hmi.Lw_720s | Full-disk computed line width |
| hmi.Ld_720s | Full-disk computed line depth |
| hmi.Mharp_720s ^b | HMI active region patch (HARP) geometry information tracked bitmap computed from LoS magnetic field |
| <i>Vector observables</i> | |
| hmi.ME_720s_fd10 ^b | Full-disk VFISV fd10 Milne–Eddington inversion magnetic and plasma parameters with uncertainties |
| hmi.B_720s | Disambiguated full-disk vector magnetic field |
| hmi.sharp_720s ^b | Vector and scalar field HARPs + active-region indices region tracked for entire disk passage every 720 s |
| hmi.sharp_cea_720s ^b | Remapped field patches and AR indices cylindrical equal area projection of B_r , B_θ , B_ϕ |
| hmi.lookup_ChebyCoef_BNoise | Velocity-dependent noise masks for vector field |

^aA description of JSOC data series is provided in Appendix A.

^bSeries available in both definitive and near-real-time versions, *e.g.* hmi.sharp_720s_nrt.

based on a relatively simple Milne–Eddington atmosphere has been adapted and optimized for use in the HMI pipeline. All images since 1 May 2010 have been processed using a version of fd10 (data series: hmi.ME_720s_fd10).

The inversion applied to each pixel cannot resolve the inherent 180° azimuth ambiguity in the transverse field direction (Harvey, 1969), so a nonlocal minimization is applied (Metcalf, 1994; Barnes *et al.*, 2014). The disambiguation is sensitive to noise in the inverted values, which varies spatially and temporally in the HMI data. Section 5 describes the scheme in more detail and Section 7.1 explains the method by which the noise threshold is determined. The disambiguation requires computing resources that increase with the size of the region and when spherical geometry is required; so the near-real-time pipeline routinely processes smaller active region patches in rectilinear coordinates. Disambiguation in spherical coordinates of larger patches has been accomplished using the same module as the definitive data. Pipeline full-disk disambiguation of definitive data has begun for data collected after 19 December 2013.

The disambiguated field is being computed for each active region patch every 12 minutes. From the vector field a representative set of indices is computed that may be of interest for space weather forecasting purposes. The indices include total unsigned flux, current helicity, and mean horizontal field gradient, to name just three; see Section 6. The SHARP (Space-weather HARP) data series (hmi.sharp_720s) includes for each numbered HARP at each time record all of the indices as keywords, as well as data arrays providing maps of nearly all the HMI scalar and vector observables and uncertainties. The pipeline produces a version

of SHARPs in the standard CCD coordinates and another that is remapped and projected onto heliographic cylindrical equal area coordinates centered on the HARP, (data series: `hmi.sharp_cea_720s`). Figure 1 shows the vector field components and derived continuum intensity for HARP 377 (NOAA AR #11158) observed at 19:00 TAI on 14 February 2011. HMI times are given in International Atomic Time (TAI), which is 35 s ahead of UTC in 2013.

The HMI vector field data are of good and consistent quality and the random noise characteristics (about 100 Mx cm^{-2} in the total magnetic field strength) in quiet regions are consistent with the design specifications of the instrument and the uncertainties intrinsic to the VFISV Milne–Eddington inversion code. The dominant time-varying systematic errors arise from the $\pm 3 \text{ km s}^{-1}$ daily velocity shift of the spectral line due to the geosynchronous orbit of SDO. The effects can be mitigated; nevertheless, daily variations that depend on velocity, field strength, and disk position remain in the data. See Section 7.

Comparison of the LoS magnetic field strength derived from HMI LCP and RCP observations analyzed using the simpler MDI-like method (see Section 2.3 and Couvidat *et al.*, 2012) with field strengths obtained using the `fcd10` inversion show differences at values as low as 1000 G. These arise from difficulties in precisely modeling the spectral line and the detailed instrument spectral characteristics. The dynamic range limitations of the instrument do not become significant until nearly 3000 G (lower in high radial-velocity regions). Comparison with higher $0.3''$ -resolution *Hinode*/SP data qualitatively match the $1''$ -resolution HMI observations in AR #11158. However, the flux density and intensity contrast reported by HMI are lower for a variety of reasons, as may be expected, see Section 7.3.2.

1.1. Paper Overview

Section 2 gives more details about the HMI pipeline processing steps relevant to the quality and interpretation of the vector field data through to the production of the Stokes observable. The active region identification and tracking analysis is summarized in Section 3. Section 4 describes the vector field inversion, specifically, improvements to the VFISV code. Section 5 explains the disambiguation processing. The SHARPs are described in Section 6. In Section 7 the sensitivity of the instrument, systematic errors, and other limitations of the data are discussed. The final section provides a summary. Three appendices provide details of the data segments in the final vector field data series, describe differences between the pipeline processing and an earlier released version of the HMI vector field, and give more details about the space weather quantities in SHARPs.

2. The HMI Magnetic Field Pipeline Processing

2.1. The HMI Instrument and Data Flow

This section briefly describes the HMI instrument and data flow for the basic observables. The HMI instrument design and planned processing scheme are described by Schou *et al.* (2012b) and Scherrer *et al.* (2012). The SDO mission, launched 11 February 2010, is described by Pesnell, Thompson, and Chamberlin (2012). The Joint Science Operations Center (JSOC) serves both the HMI and *Atmospheric Imaging Assembly* (AIA) instruments on SDO, providing pipeline processing of the incoming data products as well as archival and retrieval services for investigators who want to use the data. Extensive online documentation for the JSOC data center can be accessed at jsoc.stanford.edu; see also Table 2 in Section 8.

The HMI telescope feeds a solar image through a series of bandpass filters onto two 4096²-pixel CCD cameras. Each camera records a full-disk image of the Sun every 3.75 seconds in a 76 m Å wavelength band selected by tuning the final stage of a Lyot filter and two Michelson interferometers across the Fe I 6173.34 Å absorption line. Six wavelengths are measured in different polarizations in a sequence defined by a continuously repeating frame list. One camera measures right and left circular polarization at each wavelength, completing a 12-filtergram set every 45 s from which the Doppler velocity, LoS magnetic field, and intensities can be determined. The second Vector camera measures six polarization states every 135 s: nominally $I \pm V$, $I \pm Q$, and $I \pm U$, where I Q U V are the Stokes polarization parameters. From these the vector magnetic field and other plasma parameters can be derived.

The filtergrams are cropped, effectively truncated to the photon noise limit using a look-up table, and then losslessly compressed before being downlinked in real time to a ground station at White Sands, NM from geosynchronous orbit with a nominal bit rate of 55 Mb/s. The data are decoded, buffered and transmitted to the JSOC at Stanford University with a delay of about a minute. As the first step of the JSOC SDO data analysis pipeline, the telemetry are validated and checked for completeness; they can be retransmitted within 60 days from data retained at White Sands. Level 0 filtergrams are reconstructed and used together with the housekeeping and flight dynamics data to produce calibrated Level 1 filtergrams. The primary HMI observables are computed from these definitive Level 1 filtergrams.

The HMI observables are available since 1 May 2010. There are very few gaps in coverage. However, because of SDO's inclined orbit, twice yearly, during two-week intervals around March and September, the Earth eclipses the Sun for up to 72 minutes per day near local midnight. Calibrations briefly interrupt the observing sequences each day at 6 UT and 18 UT, and there are occasional disruptions due to eclipses, station keeping, special calibrations, intense rain, ground equipment problems, etc. As one measure, through the end of 2012 98.44 % of the possible 1.876 million 45 s Dopplergrams are available. Instrument performance is constantly monitored and adjustments are made periodically to account for changes in instrument performance, such as focus, filter drift, and optical transmission.

2.2. Stokes Vector Processing Description

The Stokes-vector observables code produces averaged I Q U V images at six wavelengths on a regular 12-minute cadence centered at the time given in the data series keyword T_REC. Figure 2 summarizes the processing steps. The vector field observing sequence captures six polarizations at each wavelength according to a repeating 135 s framelist. The code uses 360 Level 1 filtergrams taken by the Vector camera of HMI. Level 1 filtergrams have had a dark frame subtracted, been flat-fielded, and had the overscan rows and columns of the CCD removed. A list of permanently and temporarily bad pixels is stored in the BAD_PIXEL_LIST segment of each hmi.lev1 record. (See Appendix A for more about JSOC data series nomenclature.) A limb-fitting code finds the solar disk center coordinates on the CCD and the observed solar radius. Because the formation height of the signal changes, the limb position measured by HMI changes with wavelength away from the Fe I line center. Consequently, the measured radius (corrected for the SDO–Sun distance) varies as a function of the difference between the target wavelength and the wavelength corresponding to the known SDO–Sun velocity, OBS_VR. We regularly verify our model of this variation from the measured radius. Since the velocities at the east and west limbs differ due to solar rotation, an offset also appears in the center position of the image. The keywords for the disk center location (CRPIX1 and CRPIX2) and solar radius at the SDO distance (RSUN_OBS) returned by the limb finder are corrected by up to about half a pixel for the atmospheric height sampled

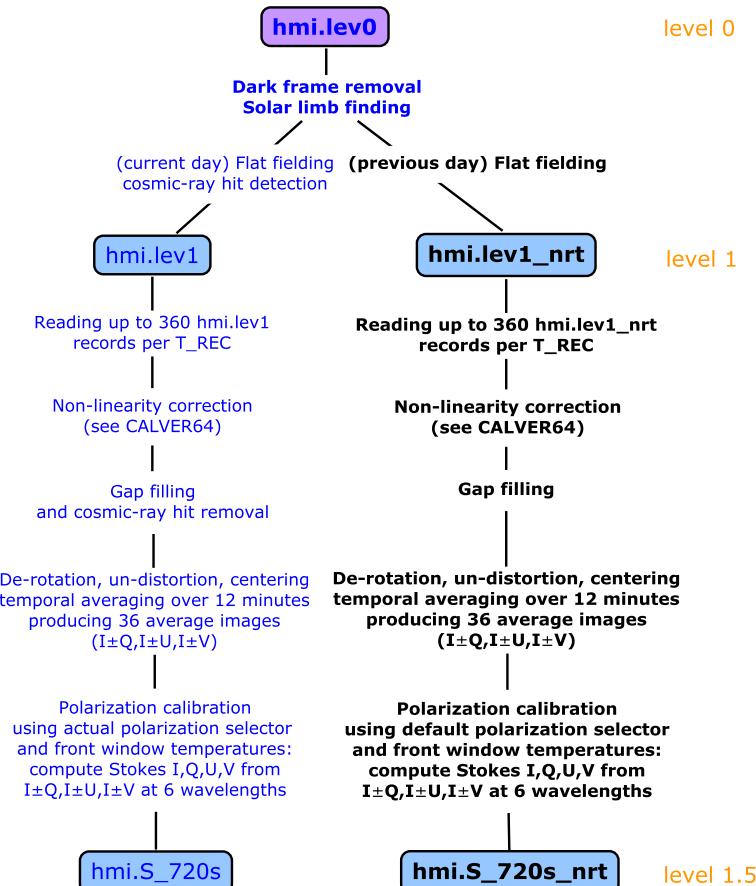


Figure 2 The HMI vector field pipeline from Level 0 filtergrams to the Level 1.5 Stokes observable. Reconstructed Level 0 filtergrams are corrected for dark current, the disk location is determined, and the images are flat fielded. For each time record, the resulting Level 1 images are corrected for temporally and spatially dependent instrumental nonlinearities; gaps in the time series are interpolated; bad pixels are corrected; and each filtergram is undistorted, co-registered, and derotated to compensate for solar rotation. A tapered temporal average is performed every 720 seconds using observations collected over 22.5 minutes (1350 s) to reduce noise and minimize the effects of solar oscillations. The averaged values for each wavelength and polarization tuning are corrected for known spatially and temporally varying polarization effects. The filtergrams are linearly combined to determine the Stokes I Q U V values at each wavelength and archived in `hmi.S_720s`. The near-real-time pipeline, right, uses preliminary calibration values and not all the recovered data may be available.

at each wavelength; the reported image scale (`CDELT1=CDELT2`) is consistent with these values.

Weekly flat fields are obtained by offsetting the HMI field of view (FOV) with the piezo-electric transducers (Wachter and Schou, 2009). Periodically observables calculated for a `T_REC` close to midnight use Level 1 filtergrams corrected with different flat fields. Flat fields are stored in the `hmi.flatfield` series. Every three months offpoint flat fields are obtained by moving the legs of the instrument.

The HMI cameras are affected by a small nonlinearity in their response to light exposure (see Figure 19 of Wachter *et al.*, 2012). This correction, on the order of 1 % for intensi-

ties below 12 000 DN/s, is implemented separately for Doppler and Vector camera data. A third-order polynomial, based on intensities obtained from ground-calibration sequences and averaged over the four quadrants of the CCDs, is applied to Level 1 values. The version of this calibration is indicated by the CALVER64 keyword.

After collecting the appropriate Level 1 records for the entire time window, a gap-filling routine deals with bad pixels and cosmic-ray hits in each filtergram (Schou and Couvidat, 2013). The observables code then calls a temporal averaging routine that performs several operations:

- De-rotation: each filtergram is re-registered to correct for the pixel shift caused by solar differential rotation. This correction is made at subpixel accuracy using a Wiener spatial-interpolation scheme. The time difference used to calculate the pixel shift is the precise observation time of the filtergram, T_{OBS} .
- Un-distortion: each filtergram is corrected for optical distortions that can be up to a few pixels in magnitude. The instrumental distortion as a function of field position is reconstructed from Zernike polynomials determined during pre-launch calibration using a random-dot target mounted in the stimulus telescope (Wachter *et al.*, 2012). The solar disk center and radius of the undistorted images are recalculated. This center and radius are slightly different from the Level 1 values calculated using distorted images.
- Centering: each filtergram is spatially interpolated to a common solar-disk center and radius that are averages of the input Level 1 filtergrams used to produce the I Q U V observables at T_{REC} . These first three steps are performed in a single operation.
- Temporal averaging: de-rotated, un-distorted, and re-centered filtergrams are combined with a weighted average for the target time T_{REC} .

Conceptually, the temporal averaging is performed in two steps. First a temporal Wiener interpolation of the observed filtergrams onto a regular temporal grid with a cadence of 45 s is performed; this results in a set of 25 frames for each wavelength/polarization state constructed using the ten original 135 s framelists. The full time window over which the interpolation is performed is 1350 s, which is wider than the averaging window; a wider window is required as the interpolation needs filtergrams before and after the interpolated times. That is followed by the averaging of the frames using an apodized window with a FWHM of 720 s; the window is a boxcar with \cos^2 apodized edges that nominally has 23 nonzero weights, of which the central nine have weight 1.0. Temporal gap filling is also performed if needed.

The pipeline next calls the calibration routine that converts the six polarizations taken by the observables sequence into a Stokes [I Q U V] vector. To perform this conversion the polarimetric model described in Schou *et al.* (2012b) is used, including the corrections that depend on the front window temperature and the polarization selector temperature. At each point in the image a least-squares fit is then performed to derive I Q U V from the six observed polarization states.

Two additional corrections based on post-launch analysis are applied to the model described in Schou *et al.* (2012a). The first compensates for what looks like telescope polarization, a spatially dependent term proportional to I that appears in the demodulated Q and U at the level of about a part in 10^4 . The dependence of Q/I, U/I, and V/I on distance from disk center was determined using the good-quality images from 3 May – 3 September 2010. The effect on V is negligible, so no correction is performed on V. The coefficients of proportionality for Q and U are given as fourth-order polynomials in the square of the distance from the center of the image, and this allows for a correction to better than a few parts in 10^5 . It may be noted that this is not strictly a telescope polarization term, because it depends on the polarization-selector setting.

While the need for the first correction was anticipated, the second was not. A perceptible (relative to the photon noise) granulation-like pattern appears in Q and U (again, V is largely unaffected). This signal appears to be caused by a point spread function (PSF) that differs with polarization state. The consequence is that a contamination from I convolved with a different PSF is added to the two linear polarization signals. This is corrected by convolving I with a five by five kernel and subtracting the result. At present a spatially independent kernel is used. Details can be found in the pipeline code (see [PipelineCode](#) in Table 2).

The final result of the Stokes-vector observables code are the four Stokes parameters at the six wavelengths in `hmi.S_720s`.

2.3. Line-of-Sight Field Processing

The line-of-sight observables code is called to compute the Doppler velocity and LoS magnetic field strength, as well as the Fe I line width, line depth, and continuum intensity, from the Stokes parameters at each time step. The LoS observables are calculated with the MDI-like algorithm detailed in Couvidat *et al.* (2012). Briefly, discrete estimates of the first and second Fourier coefficients of the solar neutral iron line are calculated from the six wavelengths, separately for the $I \pm V$ polarizations. The Doppler velocity is proportional to the phase of the Fourier coefficients; currently, only the first Fourier coefficient is used to compute the velocity. Certain approximations made in this calculation introduce errors in the results: the HMI filter profiles are not delta functions, the discrete estimate of the Fourier coefficients is not completely reliable due to the small number of wavelength samples, and the spectral line profile is not Gaussian (an assumption made to relate the phase to the velocity). Hence Doppler velocities returned by the basic algorithm must be corrected. This is accomplished in two steps.

The first relies on tabulated functions saved in the `hmi.lookup` data series. These tables are based on calibrated HMI filter transmission profiles and on a more realistic Fe I line profile. They vary slightly across the HMI CCDs and with time, primarily due to drifts in the Michelson interferometers. The values also depend on the tuning of the instrument and must be re-calculated after each HMI re-tuning. However, uncertainties remain in the filter transmission profiles and, perhaps more importantly, the Fe I line profiles; consequently, the look-up tables do not completely correct the Doppler velocity.

A second correction step has also been implemented. Every day the relationship between the median Doppler velocity RAWMEDN (measured over 99 % of the solar disk) minus the accurately known Sun–SDO radial velocity (OBS_VR) is fitted as a function of RAWMEDN with a third-order polynomial. A linear temporal interpolation of the polynomials closest in time to T_REC is used to further adjust the Doppler velocities. The HMI LoS magnetic field measurement is proportional to the difference between the corrected $I+V$ and $I-V$ Doppler velocities. Only Doppler velocities and field strengths are corrected in this way; no correction algorithm has been implemented for the other LoS observables. Table 1 lists the names of data series for the HMI pipeline quantities.

2.4. Quicklook/Near-Real-Time Pipeline Processing

Definitive data products are computed after all downlinked data have been retrieved and all housekeeping, calibration, and trend data have been analyzed. Typically it takes several days from the time data are observed until the definitive observables are complete. For instrument health and safety monitoring and for quick-look space-weather purposes, much of the data are also processed in a near-real-time mode. During the last quarter of 2012 96 % of the

NRT observables were available within 60 minutes and an additional 1 % were processed in the second hour. Because of disk and data base issues, 3 % of the NRT products became available more than 2 hours after the observations were recorded. Note that the higher level vector products derived from the `hmi.S_720s_nrt` data series require up to three hours of additional processing time.

The algorithms for processing the NRT and definitive 720 s data products are essentially the same. The differences arise because of the lack of robust cosmic ray detection, stale flat-fielding maps, and the use of default or predicted instrument parameters and housekeeping information. Differences in the data are generally small or localized. Note that in the parallel 45 s reduction pipeline, the processing of the definitive and NRT observables are less alike because fewer images are used to perform the temporal interpolation and gap filling. Higher level NRT products are generally not archived, with the exception of the SHARP product described in Section 6.

3. The Geometry of HMI Active Region Patches – HARPs

Magnetic regions of various sizes emerge, evolve, and sometimes disappear rapidly. HARPs provide primarily spatial information about long-lived, coherent magnetic structures at the size scale of a solar active region. The data series `hmi.Mharp_720s` catalogs HARPs that are automatically identified using HMI 720 s LoS magnetograms and intensity images. HARPs are given an identifying number, HARPNUM, and followed during their entire disk passage. Many HARPs can be associated with NOAA active regions (ARs). A typical set of definitive HARP patches inside their bounding boxes is shown in Figure 3 for 1 July 2012. Since we could not afford to fully process all of the vector magnetic field data, we concentrate some of the HMI vector pipeline analysis on the patches surrounding active regions, particularly for the NRT observations (see Section 6). Definitive full-disk vector data collected after 19 December 2013 are being fully processed. Depending on available resources and performance, the complete full-disk vector field may ultimately be computed at a lower cadence or more intensively for selected earlier intervals.

There is no 1:1 correspondence between HARPs and NOAA ARs. Each HARP is intended to encompass a coherent magnetic structure, so two or more NOAA ARs may all belong to a single HARP region. Coherent regions that are small in extent or have no associated sunspot can be detected and tracked by our code; such faint HARPs often have no NOAA correspondence. In the interval from May 2010 through January 2012 there were 350 NOAA ARs and 1187 definitive HARPs.

HARPs are identified at each time step in the 720 s HMI data series according to the following steps: the module first identifies the magnetically active pixels: approximately, the pixels with an absolute LoS field greater than 100 G. Then, taking spherical geometry into account, active pixels are grouped into instantaneous activity patches, which generally have smooth contours, but may consist of several regions separated by quiet Sun (Turmon, Pap, and Mukhtar, 2002; Turmon *et al.*, 2010). The instantaneous patches, which are the colored blobs in Figure 3, are combined into a temporal track by stringing together image-to-image associations of the patches. A simple latitude-dependent motion model is part of the association metric, so data gaps of less than two days do not cause association problems. Note that there is no requirement that the net flux be zero, so a HARP may be somewhat unipolar.

An individual HARP is therefore a list, over a time index, of all the associated instantaneous patches, plus geometric metadata and selected summary statistics. The instantaneous

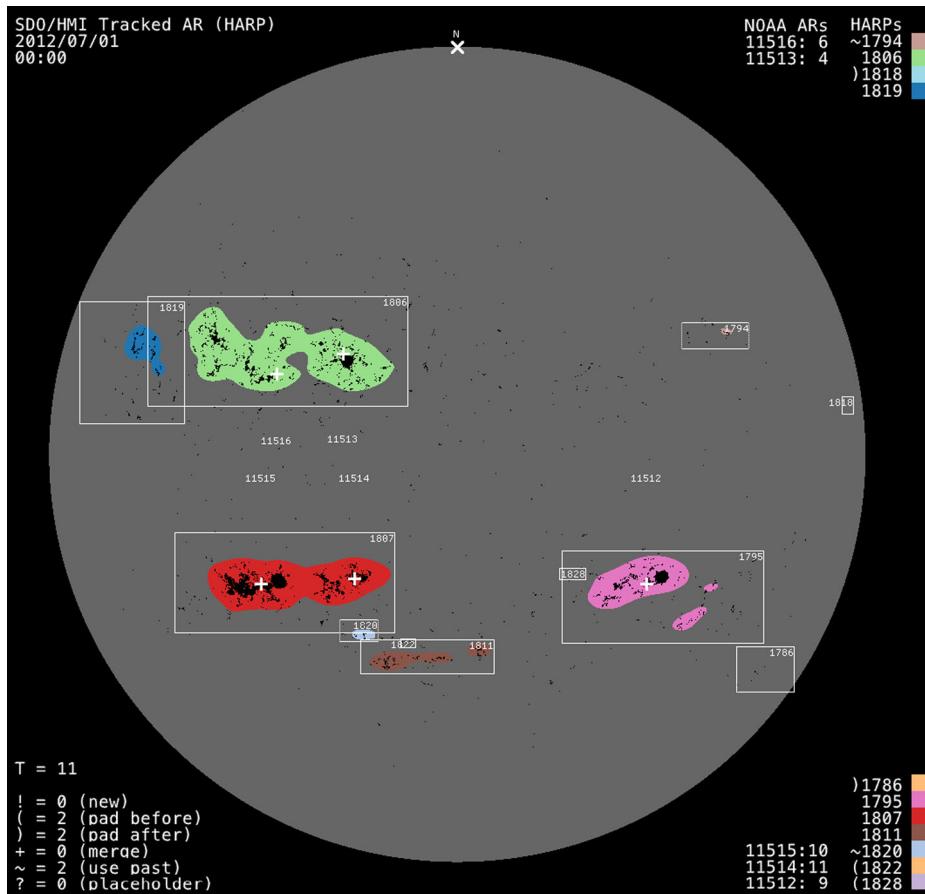


Figure 3 The eleven HARPs present on the Sun at 00:00 TAI on 1 July 2012, four in the North and seven in the South, are shown with their lifetime maximum bounding boxes. Each box is labeled with the HARP number in the upper right corner. The colored patch encloses the active region (with active pixels shown in black) associated with the HARP bitmap at the time of the image. White + signs mark the reported locations of NOAA Active Regions; the NOAA number appears at the same longitude near the equator for identification purposes. HARPs may be associated with one or more NOAA active regions (e.g. HARP 1807 is associated with NOAA AR #11515 and #11514), but most are not. HARP bounding boxes may overlap, as 1819 and 1806 do in this image, or a small HARP may even be completely enclosed inside another, e.g. 1822 inside 1811. Colored patches denoting unique HARPs never overlap. Definitive HARPs are tracked both before emergence (e.g. 1822) and after decay (e.g. 1786), as indicated by the parenthesis character to the left of the HARP number in the legends in the two right corners. The key for the HARP information appears in the lower left. Additional information can be found at jsoc.stanford.edu/data/hmi/HARPs_movies/movie-note.

patches are stored as coded bitmaps coinciding with a rectangular bounding box within the CCD plane. The bitmap's numerical code identifies the pixels that are part of the active region patch and also indicates whether each pixel is magnetically active or quiet. See Section A.6 for details. The potential exists to introduce additional encodings, such as sunspot umbra/penumbra.

The geometric information in the HARP is used to determine how parts of the solar disk are processed in the vector magnetic-field pipeline. In particular, the rectangular bounding

box (shown in white in Figure 3) is the smallest pixel-aligned box enclosing an inner latitude/longitude box. The latter box has constant width in heliographic latitude and longitude, and rotates across the disk at a constant angular velocity (defined for each HARP). This latitude/longitude box is just big enough to enclose all instantaneous patches during the lifetime of the HARP. All these geometric parameters are stored in HARP keywords. Some elementary summary characteristics of each HARP at each time are also determined, *e.g.* the total LoS flux, net LoS flux, total area, and centroid, as well as any corresponding NOAA region number (or numbers). The keyword NOAA_AR gives the number of the first NOAA active region associated with a HARP. If more than one NOAA region is associated with a HARP, the total number is given by NOAA_NUM and a comma-separated list of regions appears in NOAA_ARS. Maps of the observed solar quantities are not part of the HARP data series; the observables are collected and additional parameters are provided as part of the space-weather HARPs (SHARPs) described in Section 6.

There are in fact two types of HARP, definitive and near-real-time (NRT). The main distinction is that the definitive HARP geometry is not finalized until after the associated magnetic region has decayed or rotated off the disk, so all geometric information referred to above is consistent for the lifetime of the region. In particular, the bounding box of a definitive HARP encloses the same heliographic area during its entire lifetime. The temporal life of a definitive HARP starts when it rotates onto the visible disk or one day before the magnetic feature is first identified in the photosphere. The HARP expires one day after the feature decays or when it rotates completely off the disk. Steps are taken to ensure that activity fluctuating around the threshold of detection is not separated into multiple HARPs with temporal breaks. Additionally, active regions that grow over time, eventually overlapping other regions, are merged into a single definitive HARP.

A near-real-time HARP is similar, but does not begin until the feature can be identified (there is no one-day pad at the beginning). Moreover, the heliographic size of the region may change during its life. NRT HARPs may also merge, but the merger is not seamless because it results in termination of one or more NRT HARPs and continuation of the larger, merged object. Flags are provided to identify mergers in the NRT product, because the sudden amalgamation of an NRT HARP with another region is an artifact of the grouping and not of physical origin. In the definitive product, regions do not merge because the entire history of the HARP is known before the bounding box is determined. Note that definitive and NRT HARP numbers are not the same.

HARPs are available in the data series hmi.Mharp_720s_nrt and hmi.Mharp_720s. More information about HARPs is available online at jsoc.stanford.edu/jsocwiki/HARPDataSeries (see Table 2) and in Turmon *et al.* (2014).

4. Milne–Eddington Inversion

Spectral line inversion codes are tools used to interpret spectro-polarimetric data. Given a set of observed Stokes spectral profiles, the purpose of an inversion code is to infer the physical properties of the atmosphere where they were generated.

To interpret HMI data, a relatively simple spectral line synthesis model, based on a Milne–Eddington (ME) atmosphere, is used. The ME model assumes that all of the physical properties of the atmosphere are constant with height, except for the source function, which varies linearly with optical depth. The generation of polarized radiation is described by the classical Zeeman effect. Under these assumptions, the polarized radiative transfer equation has an analytical solution, known as the Unno–Rachkovsky solution (Unno, 1956;

Rachkovsky, 1962, 1967). In the ME context, the forward model for spectral line synthesis traditionally has 11 free parameters: three magnetic-related quantities (magnetic field strength, its inclination with respect to the line of sight, and its azimuth with respect to an arbitrarily chosen direction in the plane perpendicular to the LoS), five thermodynamical ones (Doppler width, line core-to-continuum opacity ratio, Lorentz-wing damping, and the source function and its gradient), two kinematic variables (Doppler velocity and macroturbulent velocity), and a geometrical parameter, the filling factor, which quantifies the fraction of the pixel that is occupied by a magnetic structure.

The Very Fast Inversion of the Stokes Vector (VFISV) is a ME spectral line inversion code specifically designed to infer the vector magnetic field of the solar photosphere from HMI Stokes measurements (Borrero *et al.*, 2011; Centeno *et al.*, 2013). VFISV uses a Levenberg–Marquardt least-squares minimization (Press *et al.*, 1992) of a χ^2 function. χ^2 is a metric of the difference between the observed and synthetic Stokes profiles (*i.e.* the goodness of the fit). Given an initial guess for the model atmosphere, the algorithm produces synthetic Stokes profiles and compares them with the observations. The model parameters are then modified in an iterative manner until the synthetic data are deemed to be a good-enough fit to the observations. ‘Good enough’ is defined by the convergence criteria that determine when to exit the iteration loop.

Several limitations arise as a consequence of using the ME approximation. The thermodynamical parameters are complicated, nonlinear combinations of actual physical properties (density, temperature, pressure, etc.), so one cannot extract the corresponding real thermodynamical information using only the ME assumption. In addition, due to the lack of gradients in the model, asymmetries present in the Stokes profiles cannot be properly fit by the inversion code. However, the retrieved values of magnetic field and LoS velocity are representative of the average conditions throughout the spectral region of formation and averaged over temporal and spatial resolution (Westendorp Plaza *et al.*, 1998; Leka and Barnes, 2012). HMI does not have an absolute wavelength reference. The zero point of the derived velocity depends on filter profile calibrations determined relative to the full-disk line center at the nominal wavelength of 6173.3433 (Norton *et al.*, 2006); no correction is made for the effect of convective blueshift on the observations. See Figure 16 in Section 7.2.3 for information about the daily variation of the velocity. Welsch, Fisher, and Sun (2013) have developed a scheme to determine the velocity zero point in active regions.

Despite being a general ME code with customizable settings, VFISV has been optimized to work within the HMI data processing pipeline. The inversions are typically applied to the hmi.S_720s series (full Stokes vector, averaged over 720 seconds, for the full disk of the Sun, with 4096×4096 pixels), which requires being able to invert more than 20 000 pixels per second. To meet the speed requirements, certain aspects of VFISV have been hard-coded for performance purposes; for instance, VFISV uses the HMI transmission filter profiles for each pixel in the computation of the synthetic spectral line. One important facet of this inversion code is that it forces the filling factor, α , to be unity. This constraint, enforced because of the sparse spectral sampling and moderate spatial resolution of the HMI data, assumes that the plasma is uniformly magnetized in each pixel. In the weak-field regime, there is a strong degeneracy between the filling factor and the magnetic field strength. However, the product of the two (αB) is very well constrained by the observed Stokes profiles. Therefore, assuming $\alpha = 1$ means that the quantity B should be interpreted as the average magnetic field of the pixel. The stray light, *i.e.* light scattered from other parts of the Sun that contaminates any given pixel, is currently not being considered either. This generally leads to lower values of the inferred magnetic field strength, especially in strong-field regions (LaBonte, 2004a).

Several changes and optimizations have been introduced in VFISV since its release (Borrero *et al.*, 2011). The HMI team has derived optimal settings for the pipeline processing to improve the speed and general performance of the code. These changes to the original VFISV code are described in detail in Centeno *et al.* (2013). Some of the significant changes include a) a regularization term added to the χ^2 minimization merit function to bias the solution towards the more physical solution when encountering two nearly degenerate minima, b) an adjustment to the relative weighting applied to the Stokes parameters I Q U V, c) implementation of tailored iteration and exit criteria that results in an average of ~ 30 iterations per pixel, d) use of a spectral line synthesis module that computes the full radiative transfer solution only for the inner wavelength range of the HMI spectral line, and e) a simpler initial-guess scheme that does not use a neural net. In the HMI pipeline context, the inversions produced with these settings are referred to as fd10.

Two keywords indicate the version of the fd10 code: INVCODEV and CODEVER4. A significant improvement was incorporated on 30 April 2013, when the code was enabled to use time-dependent information about the HMI filter profiles. The differences are spatially dependent and systematic. Over the full disk, a typical difference map between the two versions shows a mean of 0.1 G and an rms of 4 G, with a distinct East/West asymmetry and greater bias in strong-field regions. The filter profiles must be updated when the instrument tuning is adjusted, as indicated by the INVPHMAP keyword. Inversions of data collected between 1 August 2011 and 23 May 2013 use earlier, less accurate filter phase maps. Notable changes to the pipeline modules and processing notes are documented at jsoc.stanford.edu/jsocwiki/PipelineCode as listed in Table 2.

For each pixel in the field of view, VFISV returns its best estimate of the atmospheric model together with the standard deviation associated with each model parameter and the normalized covariances in the errors of selected pairs of parameters. Appendix B describes differences between fd10 and a previous version of the inversion associated with an earlier data release.

5. The Disambiguation Algorithm

Inversion of the Zeeman splitting to infer the magnetic-field component transverse to the line of sight results in an ambiguity of 180° in its direction (Harvey, 1969); some additional assumption(s) or approximation must be made to completely determine the magnetic-field vector. A plethora of methods have been proposed to resolve this ambiguity (for an overview, see Metcalf *et al.*, 2006). For disambiguation of HMI data, a variant on the “Minimum Energy” method (Metcalf, 1994) called ME0 has been implemented.

The original Minimum Energy algorithm minimizes the sum of the squares of the magnitude of the field divergence and the total electric-current density, where the vertical derivatives in each term are approximated by the derivatives of a linear force-free field. Minimizing $|\nabla \cdot \mathbf{B}|$ gives a physically meaningful solution, and minimizing \mathbf{J} provides a smoothness constraint. For computational efficiency, the algorithm applied to HMI data uses a potential field to approximate the vertical derivative in $|\nabla \cdot \mathbf{B}|$ and minimizes only the magnitude of the normal component of the current density. Note that now we minimize $\sum |\nabla \cdot \mathbf{B}| + |J_z|$ in place of $\sum |\nabla \cdot \mathbf{B}|^2 + |J|^2$ as in the original Minimum Energy method.

This method has been shown to be among the best performing on a wide range of tests on synthetic data for which the answer is known (Metcalf *et al.*, 2006; Leka *et al.*, 2009). However, Leka *et al.* (2009) showed that the minimum energy state is not always the correct ambiguity resolution in the presence of noise. Furthermore, convergence of the optimization

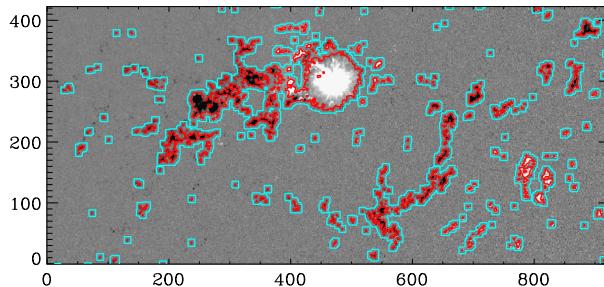


Figure 4 The radial component of the field for HARP 1795 (NOAA AR #11512) on 1 July 2012 at 00:00 TAI, saturated at ± 500 G. Axes are in pixels. The disambiguation result from simulated annealing is returned for pixels inside the red contour. SHARPs observed before January 2014 are disambiguated in patches, and all pixels outside the red contour are annealed. In the case of full-disk disambiguation, pixels between the blue and red curves are annealed, then smoothed to determine the disambiguation; pixels outside the blue contour have three disambiguation results returned: potential field acute angle, random, and most radial.

routine is slow in pixels where the signal is dominated by noise, particularly noise in the transverse component of the field. Therefore, the final disambiguation for those pixels that fall below a specified noise threshold is made with several simpler and quicker methods in the pipeline.

The steps needed for implementing the algorithm are:

- Determine spatially dependent noise mask;
- Compute potential field derivative;
- Determine confidence in result;
- Minimize energy function;
- Disambiguate noisy pixels.

These basic steps apply to disambiguation of both HARPs and full-disk images. The code currently uses a planar approximation when computing the energy for small HARPs ($< 20^\circ$ extent in both latitude and longitude), but uses spherical geometry for larger HARPs or full disk. More details of the algorithm as used in the HMI pipeline are presented in Barnes *et al.* (2014).

5.1. The Noise Mask

The HMI noise level varies with location on the disk and with relative velocity between the spacecraft and Sun (see Section 7.1 for a discussion of the temporal and large-scale spatial variations found in the data). A constant is added to the derived noise value, and pixels with a transverse field strength above this noise estimate are assigned to the disambiguation mask. The mask is then eroded to eliminate isolated pixels above the noise, and this defines the high-confidence pixels. Intermediate confidence is assigned to a buffer zone surrounding the clusters of pixels with a well-determined field. All pixels within that grown mask are included in minimizing the energy.

Figure 4 shows an example of the mask determined for a subarea of a full-disk disambiguation. The actual solution from the energy minimization is returned only for pixels within the eroded mask, *i.e.*, where the transverse field is well determined. As described in Section 5.4, in the surrounding five-pixel buffer area the annealed solution is smoothed, and in weak-field regions alternatives to the minimum-energy disambiguation are provided.

When only patches are disambiguated *all* pixels outside the high-confidence region are considered to be intermediate. Currently, all SHARPs observed before 15 January 2014 rely on such patch-wise disambiguation; consequently, all pixels within the HARP bounding box have been annealed, and those outside strong-field high-confidence areas have also been smoothed. This changed when the definitive pipeline switched over to computing the full disk disambiguation at each time step. The type of solution returned is recorded in the CONF_DISAMBIG segment; see Section A.4 for more information.

5.2. The Potential Field

To determine the vertical derivative used in approximating $\nabla \cdot \mathbf{B}$, the potential field that matches the observed LoS component of the field is computed in planar geometry (Alissandrakis, 1981). By matching the line-of-sight component, the potential field only needs to be computed once, and by using planar geometry, fast Fourier transforms can be used. When planar geometry is used for the energy calculation, a single plane is used in the potential field calculation. For spherical geometry, the observed disk is divided into tiles by taking a regular grid in an area-preserving Mollweide projection and mapping it back to the spherical surface of the Sun. Each tile is treated as a separate plane for computing the potential field, although neighboring tiles are overlapped, and a windowing function is applied to reduce edge effects.

5.3. Minimum Energy Methods

The ambiguity is resolved by minimizing an energy function

$$E \propto \sum (|\nabla \cdot \mathbf{B}| + |(\nabla \times \mathbf{B}) \cdot \mathbf{n}|), \quad (1)$$

where the sum runs over pixels in the noise mask, \mathbf{n} is a unit vector normal to the surface, and the second term on the right-hand side is proportional to the normal component of the current density. The calculation of the normal component of the curl, and the horizontal terms in the divergence is straightforward, requiring only observed quantities in the computation and a choice of the ambiguity resolution; a first-order forward difference scheme is used to calculate all the horizontal derivatives. However, these finite differences mean that the computation is not local. To find the permutation of azimuthal angles that corresponds to the minimum of E , a simulated annealing algorithm (Metropolis *et al.*, 1953; Kirkpatrick, Gelatt, and Vecchi, 1983), as described in Barnes *et al.* (2014), is used. This method is an extremely robust approach when faced with a large, discrete problem (there are two and only two possibilities at each pixel) with many local minima (Metcalf, 1994).

5.4. Treatment of Areas Dominated by Noise

The utility of minimizing the energy given by Equation (1) depends on being able to reliably estimate the divergence and the vertical current density. Where noise dominates the retrieved field, alternate approaches are used.

A neighboring-pixel acute-angle algorithm based on the University of Hawai'i Iterative method (UHIM; Canfield *et al.*, 1993; Metcalf *et al.*, 2006) that locally maximizes the sum of the dot product of the field at one pixel with the field at its neighbors is used. In areas close to pixels with well-determined fields, the method performs well, as it quickly produces a smooth transition to noisy pixels. While the result is always a smooth solution, the algorithm

can become trapped in a smooth but incorrect result, which tends to be propagated outwards from the well-measured field (Leka *et al.*, 2009). Therefore this method is currently used for SHARPs, and in small intermediate areas around well-determined fields for the full-disk.

Farther away from the well-measured field in full-disk mode, where the annealing method is not applied, the pipeline module disambiguates the azimuth in three ways, each with its own strengths and weaknesses. Method 1 selects the azimuth that is most closely aligned with the potential field whose derivative is used in approximating $\nabla \cdot \mathbf{B}$. Method 2 assigns a random disambiguation for the azimuth. Method 3 selects the azimuth that results in the field vector being closest to radial. Methods 1 and 3 include information from the inversion, but can produce large-scale patterns in azimuth. Method 2 does not take advantage of any information available from the polarization, but does not exhibit any large-scale patterns and reflects the true uncertainty when the inversion returns purely noise. The results of all three methods are recorded in DISAMBIG; see Section A.4 for specifics.

6. Space-Weather HARPs – SHARPs

The Space-weather HARP (SHARP) data series collects most of the relevant observables in HMI active-region patches and computes from them a variety of summary active-region parameters that evolve in time. The SHARP data series, `hmi.sharp_720s`, is quite complete, comprising 31 maps of the active patch at each time step with all of the disambiguated field and thermodynamic parameters determined in the inversion, along with uncertainties and error cross-correlation coefficients, as well as the continuum intensity, Doppler velocity, and line-of-sight magnetic field (Bobra *et al.*, 2014).

There is an extensive literature describing potentially useful active-region indices (*e.g.* Leka and Barnes, 2003a, 2003b, 2007; Barnes and Leka, 2006, 2008; Falconer, Moore, and Gary, 2008; Schrijver, 2007; Mason and Hoeksema, 2010; Georgoulis and Rust, 2007). The SHARP series computes many of these quantities that are averaged or summed over the entire region. A list of computed indices can be found in Appendix C and includes total unsigned flux, mean inclination angle (relative to vertical), mean values of the horizontal gradients of the total, horizontal, and vertical field, means of the vertical current density, current helicity, twist parameter, and excess magnetic energy density, the total excess energy, the mean shear angle, and others. See Bobra *et al.* (2014) for details.

The alternate SHARP series `hmi.sharp_cea_720s` includes a smaller number of maps (eleven) remapped to cylindrical equal area (CEA) heliographic coordinates centered on the HARP center point. The coordinate remapping and vector transformation are described in Section 6.1. The CEA series includes maps of the three vector magnetic field components and the expected standard deviation of each component, along with the LoS magnetic field, Dopplergram, continuum intensity, disambiguation confidence, and HARP bitmap (see Appendix C for details).

6.1. Coordinate Remapping and Vector Transformation

Two coordinate systems are involved in the representation of the vector magnetograms. The first refers to the observation’s physical location on the Sun or in the sky, for which HMI uses World Coordinate System (WCS) standards for solar images (Thompson, 2006). The second refers to the direction of the three-dimensional field vector components at a particular location. In this article, we call the conversion from one WCS format to another “remapping” in the first geometrical case and “vector transformation” in the second.

6.1.1. Remapping

Currently, SHARP data are available in CCD image coordinates or in a cylindrical equal-area projection centered on the patch with the WCS projection type specified in CTYPE1 and CTYPE2. When data are requested from the JSOC, users will eventually have the option to customize the data using various remapping and vector transformation schemes.¹ For a more detailed account of different map projections, we refer to Calabretta and Greisen (2002) and Thompson (2006).

Care must be taken with image coordinates. The standard SHARP image segment cut-outs, like most HMI image data series, are stored in helioprojective CCD image coordinates, registered and corrected for distortion, but without any remapping. Each pixel represents the same projected area on the plane of the sky, but corresponds to a different area on the solar surface. We note that the +y direction of the HMI CCD array in the data series does not generally correspond to solar North. The counterclockwise (CCW) angle needed to rotate the image to the “North-up” format is recorded in the keyword CROTA2. For most standard HMI images CROTA2 is close to 180°, *i.e.* raw HMI CCD images appear upside down.

The vector magnetic field in SHARPs is also provided in a cylindrical equal-area projection data series, hmi.sharp_cea_720s. A thorough discussion is available as a technical note (Sun, 2013). Each pixel in the remapped CEA image represents the same surface area on the photosphere. For the CEA remapping, the vector magnetogram is first converted into three images, each frame representing one component of the three-dimensional field vector (see next section). A target coordinate grid is defined for a CEA coordinate system in which the origin of the map projection is the center of the HARP. The CEA grid is centered on the region of interest to minimize geometric distortion in the result. The remapping is made individually for each vector component, as well as on the scalar observables. The uncertainties, BITMAP, and confidence maps are derived a little differently: first, the center of each pixel in the remapped coordinate system is located in the original image; then the nearest neighboring pixel in that original image is identified and the value for that nearest original pixel is reported.

6.1.2. Vector Transformation

In the “native format” the field vector at each pixel is represented by three values derived from the inversion (Section 4): the total field strength (B), the inclination (γ), and the azimuth (ψ). Inclination is defined with respect to the HMI line of sight, so the longitudinal (B_l) and transverse field (B_t) components can be easily separated as $B_l = B \cos \gamma$ and $B_t = B \sin \gamma$.

As mentioned in the previous section, we decompose the vector magnetogram into three component images before remapping using the following basis vector: $(\mathbf{e}_\xi, \mathbf{e}_\eta, \mathbf{e}_\zeta)$, where \mathbf{e}_ξ refers to the +x direction of the native-format image, \mathbf{e}_η to +y, and \mathbf{e}_ζ to +z out of the image. Note that \mathbf{e}_ζ coincides with the LoS direction. Remapping is performed subsequently on the three component images.

The vector transformation is performed as the last step, following Equation (1) of Gary and Hagyard (1990) after determining the heliographic latitude and longitude at each pixel. For the CEA map-projected data we transform the vector into standard heliographic spherical coordinates with a basis vector $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\phi)$, where \mathbf{e}_r is the radial direction normal to the

¹ Relevant modules have been developed and tested. However, their integration with the JSOC export web tools has yet to be implemented.

solar surface, e_θ points southward along the meridian at the point of interest, and e_ϕ points westward in the direction of solar rotation along a constant latitude line. The r component corresponds to the vertical component, θ and ϕ to the horizontal components.

6.2. A Comparison of Definitive and Quick-look SHARP Data

Definitive SHARPs provide the most complete and best-calibrated data. Because the extent of the definitive bounding box is determined only after a region completes its disk passage to ensure uniform geometric coverage, there is a delay of ~ 35 days in data availability. Fortunately, quick-look data are available in NRT, and the SHARP indices are computed as data become available, usually in less than three hours. Occasional data gaps longer than an hour are simply bypassed. Apart from this, the processing of the definitive and NRT data differs in three significant ways:

- As described in Section 2.2, a preliminary flat-field is used, gap-filling is less sophisticated, cosmic-ray correction is not performed, and calibrations are based on default or preliminary instrument parameters. All of the NRT 720 s observables (vector and scalar) are calculated from the hmi.S_720s_nrt Stokes series.
- The NRT HARP bounding box is based on the data available at that point in time, therefore the heliographic size may change and regions may even be merged by the time the definitive data are created. Obviously, no NRT HARP data are available before the region emerges, and no association with a NOAA AR can be made until NOAA assigns a number. The NRT and definitive HARP numbers may be different.
- The NRT fcd10 VFISV inversion is computed only for identified NRT HARPs. The inverted region extends beyond the NRT HARP bounding box by 50 pixels in each direction. Consequently, for NRT the size of the padding region outside the HARP used to compute the initial field parameters is reduced from 500 to 50 (or 20 in some intervals, check keyword AMBNPAD). NRT disambiguation parameters are adjusted to speed up the calculation by increasing the effective temperature scale factor, AMBTFCTR, and decreasing the number of reconfigurations, AMBNEQ (see Barnes *et al.*, 2014 for details). The results for particularly small and (until September 2013) particularly large NRT SHARPs are not computed to limit the computing load. All NRT SHARPs are disambiguated using planar geometry rather than spherical. The same noise-threshold map is used in both pipelines, subject to some delays in updating. The SHARP parameters are computed using only pixels both within the HARP (the colored blobs in Figure 3) and above the noise threshold (see Section 7.1.1); the number of pixels is given in the keyword CMASK.

Figure 5 illustrates some of the differences between SHARP 2920 and NRT SHARP 1407 during the interval 1–14 July 2013. The region was associated with NOAA ARs #11785, #11787, and #11788. The total unsigned flux, USFLUX, is computed by integrating the vertical flux, B_r , in the high-confidence areas of the active region. For computing the SHARP indices, high-confidence regions are those above the noise-mask, where the confidence in the disambiguation is highest (CONF_DISAMBIG = 90, see Appendix C). As the region first rotates onto the disk on 1 July and subsequently grows, the USFLUX increases, as shown by the overlying purple and green points in the upper panel. A second region emerges nearby on 2 July that is initially classified as an independent NRT SHARP that does not contribute to the USFLUX of 1407. That region grows quickly on 2–3 July. An M 1.5-class flare was observed at 07:18 UT on 3 July. On 4 July the two NRT SHARPs merge. The definitive SHARP evolves smoothly because its heliographic boundaries are determined only after the region has completed its disk transit and so includes both regions. The SHARP eventually rotates off the disk on 14 July.

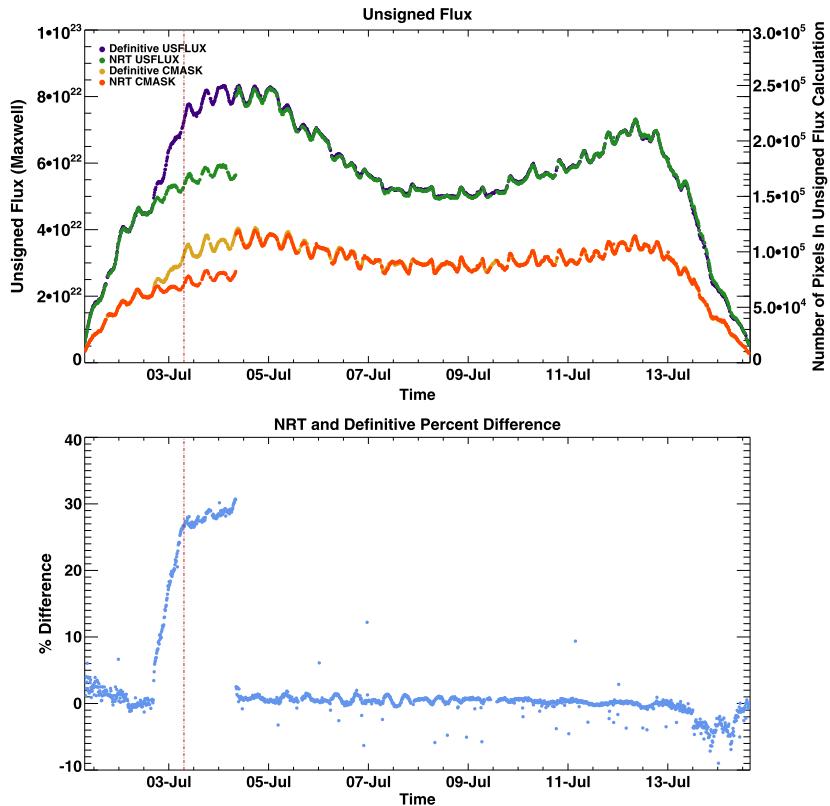


Figure 5 The top panel shows the total unsigned flux, USFLUX, during the disk passage of SHARP 2920 in purple and the same region, NRT SHARP 1407, in green from 1–14 July 2013. The region crossed disk center ∼ 07:48 UT on 8 July. The region is ultimately associated with NOAA ARs #11785, #11787, and #11788. The vertical line at 07:18 UT on 3 July indicates the GOES X-ray flux peak associated with an M 1.5-class flare in NOAA AR #11787. Differences between the definitive and NRT processing are described in the text. The heliographic area inside the bounding box of the definitive SHARP does not change, whereas the green line jumps on 4 July when two NRT SHARP regions merge. Also shown are the number of high-confidence pixels contributing to the USFLUX for the definitive (yellow) and NRT (red) SHARP. Variations in USFLUX track the changes in the area of the region due to evolution and to systematic effects, see text. The total flux increases as the region moves away from the limbs. The broad peaks in total unsigned flux ∼60° from central meridian are attributed to increased noise in the field away from disk center. The bottom panel shows the percentage difference between the USFLUX computed for the definitive and NRT SHARPs. The differences between the two are small after accounting for the merger. Outliers in the values correspond to times when the QUALITY of the images was not good. The poor quality images were excluded in the upper panel.

The number of high-confidence pixels is shown by the yellow (definitive) and red (NRT) points in the upper panel of Figure 5. The number reflects both the size of the evolving active region and systematic variations. The clear 12-hour periodicity is associated with the spacecraft orbital velocity, V_r , relative to the Sun and is discussed in Section 7.1.2. The broad, roughly symmetric peaks centered ∼60° from central meridian are due to the spatially dependent sensitivity of the instrument (see Section 7.1.1). The noise level in low and moderate field strength regions changes as a function of center-to-limb angle and V_r , and consequently the number of pixels contributing to the USFLUX index changes. Histograms of two representative active regions show a shift in the peak of the distribution of low-to-

moderate field pixels toward higher values as the regions move away from central meridian; the width of the distribution also broadens. An increase of a few tens of Gauss in the field strength distribution at 60° increases the number of pixels in the 250–750 G range by a few tens of percent. The field strength in higher field regions is more stable, but because the area covered by the strong field is much smaller, the variation in USFLUX is significantly affected.

The lower panel shows the percentage difference in the definitive and NRT USFLUX index as a function of time. The two match within a few percent except during the 2–4 July interval when there were two NRT SHARP regions. In the upper panel we have excluded all images where the QUALITY keyword exceeded 0x010000. Bits in the QUALITY keyword identify potential problems with the measurements (in this case typically missing filtergrams because of calibration sequences). For comparison, the lower panel includes all of the points, and the isolated points lying well away from the line warn of the consequences of using less-than-ideal measurements.

7. Uncertainties, Limitations, Systematics, and Sensitivities

The HMI vector field data are of high quality and more than meet their original performance specifications. Nevertheless, the measurements are far from perfect and proper care must be taken when using the data. This section discusses various uncertainties, limitations, and known systematic issues with the data produced in the HMI vector field pipeline. Sources of error include imperfections and limitations of the instrument and observing scheme, as well as imperfections and limitations of the analysis methods. Variances of the inverted variables and covariances between their errors are obtained during the VFISV inversion (see Section 4). These are recorded as estimated errors (square roots of the variances) for these quantities and their correlation coefficients (related to the covariances). We characterize the known uncertainties and errors as best we can; the user must take this information into account when performing research.

7.1. Temporal and Spatial Variations of the Inverted Magnetic Field

The dominant driver of time-varying errors in the magnetic field is the Doppler shift of the spectral line. By far the largest contributor is the daily period due to the geosynchronous orbit of SDO. On longer time-scales there are other orbital and environmental variations, along with changes in the instrument itself, changes in the way the instrument is tuned or operated, and variations in the calibration. Most of the short-term variation from solar oscillations is eliminated by the 12-minute averaging. This section concentrates primarily on the daily variability.

7.1.1. Time-Varying Noise Mask

The noise level of the inverted magnetic field shows large-scale spatial variations over the entire disk caused by effects of irregularities in the HMI instrument and viewing angle. This pattern changes with the orbital velocity of the geosynchronous satellite relative to the Sun, V_r (keyword OBS_VR). V_r ranges $\pm 3 \text{ km s}^{-1}$ from local dusk ($\sim 1 \text{ UT}$) to dawn ($\sim 13 \text{ UT}$). The corresponding Doppler shift moves the spectral line back and forth by about one tuning step every 12 hours. Combined with the fixed velocity pattern of solar rotation, this leads to the temporal and spatial variations of the inverted magnetic field over the Sun's disk every 24 hours.

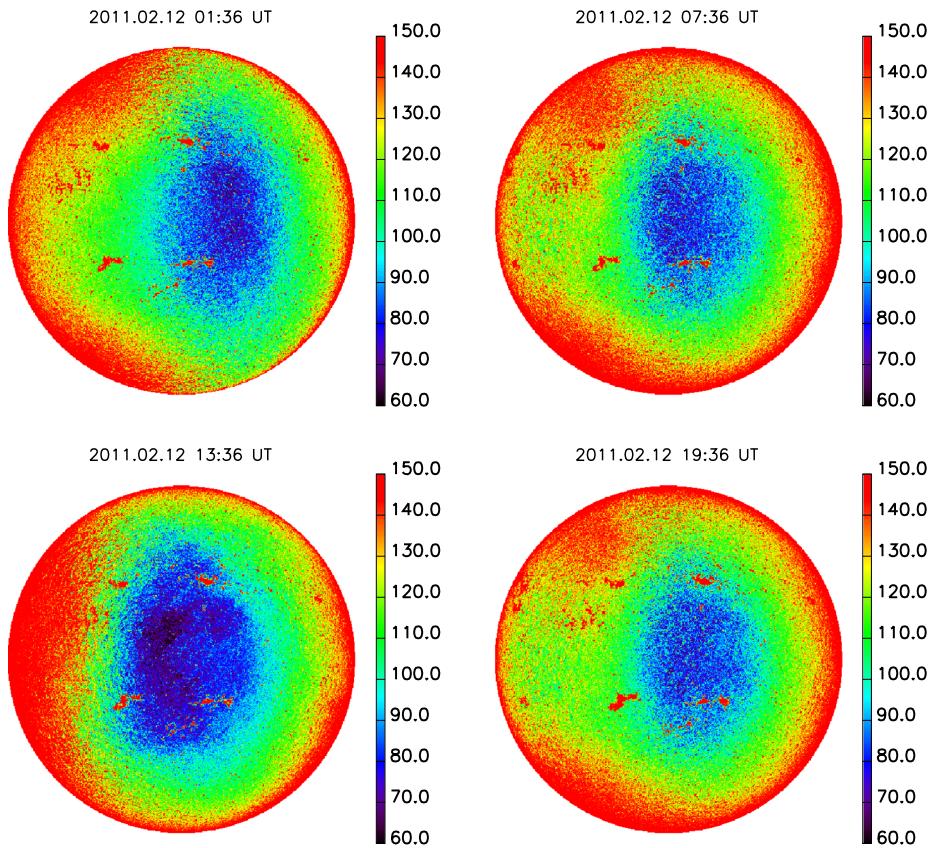


Figure 6 From left to right, then top to bottom: the total magnetic field strength at 01:36 TAI (local dusk, maximum relative radial velocity, $V_r = 2980 \text{ m s}^{-1}$), 07:36 TAI (noon, near zero $V_r = 360 \text{ m s}^{-1}$), 12:36 TAI (dawn, minimum $V_r = -2372 \text{ m s}^{-1}$), and 18:36 TAI (midnight, near zero $V_r = 263 \text{ m s}^{-1}$) on 12 February 2011. The field strength should be uniform in the quiet Sun if it is measured perfectly. The large-scale quiet-Sun noise pattern varies with the time of day, *i.e.* relative velocity. The color table shows the field strength in Mx cm^{-2} .

Figure 6 shows the field strength determined at 01:36 TAI (top left), 07:36 TAI (top right), 13:36 TAI (bottom left), and 19:36 TAI (bottom right) on 12 February 2011, at which times SDO was close to local dusk, midnight, dawn, and noon. At these times, the relative orbital velocity reaches a maximum, minimum, or is close to zero. The large-scale patterns differ with V_r . However, when the SDO–Sun velocities are approximately the same, the patterns are very similar, as shown in the right column of Figure 6 and in Figure 7. The spatial pattern in the magnetic-field strength noise is stable in time but varies with V_r .

As described in Section 5.4, the disambiguation requires an estimate of the noise, as may researchers analyzing the data. To describe this temporally and spatially varying magnetic strength noise we construct a noise mask. Each mask characterizes the noise level in the magnetic-field strength for a particular relative velocity, V_r . Reference noise masks are first determined at 50 m s^{-1} steps using sets of full-disk magnetic field strength data (~ 50 magnetograms) observed in 100 m s^{-1} intervals. The median of the field strength for each on-disk pixel is calculated; excluded from the fit are strong field pixels where

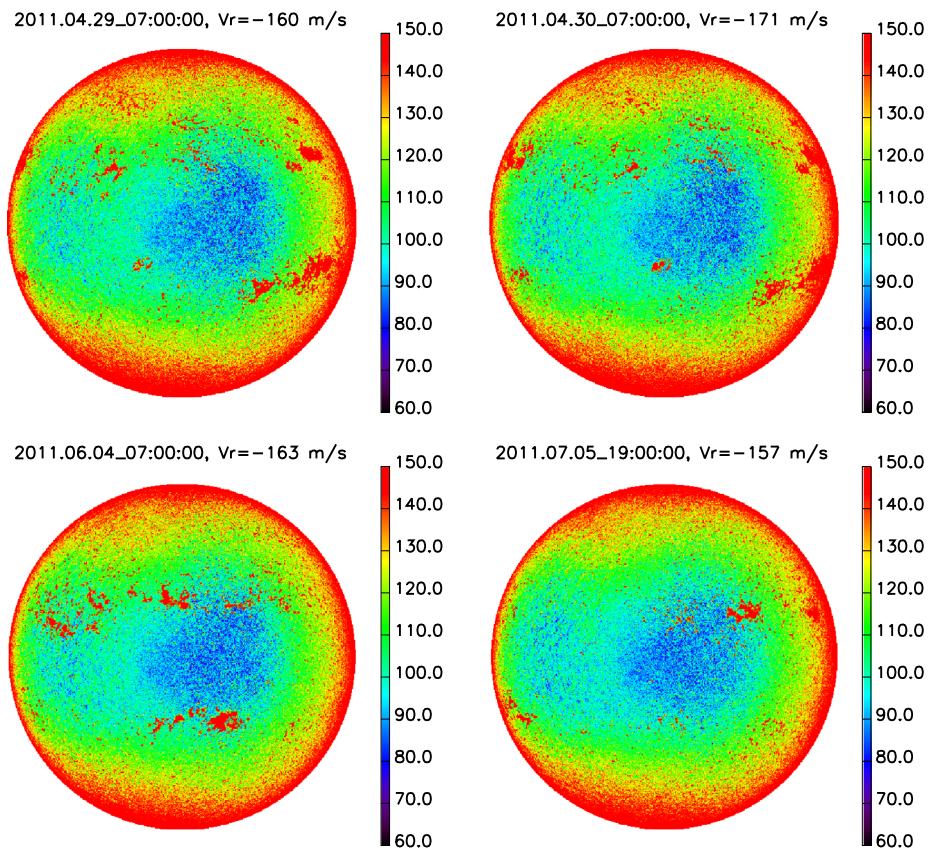


Figure 7 Field strengths observed at 07:00 TAI on 29 April 2011 (top left), 30 April 2011 (top right), 4 June 2011 (bottom left), and at 19:00 TAI on 5 July 2011 (bottom right). SDO's radial velocity relative to the Sun, V_r , is close to -165 m s^{-1} at these four times. Excluding strong-field regions, the patterns are similar. The color table shows field strength in Mx cm^{-2} .

$B > 300 \text{ Mx cm}^{-2}$, roughly 3σ of the nominal field strength noise. To that map of median values a 15th-order two-dimensional polynomial is fit. The basis functions are Chebyshev polynomials of the first kind. The reference coefficients are saved in the data series `hmi.lookup_ChebyCoef_BNoise`. When a noise mask for a specific velocity is required, the coefficients are linearly interpolated for that V_r and used to reconstruct a full-disk noise mask. Figure 8 shows the reference noise masks for a range of orbital velocities.

The noise pattern changes somewhat with instrument parameters, for example when the front window temperature is adjusted. Such instrument changes have occurred at a few known times: 13 December 2010, 13 July 2011, 18 January 2012, and 15 January 2014. See references in jsoc.stanford.edu/jsocwiki/PipelineCode, Table 2. Reference noise mask coefficients are generated specifically for each time interval.

7.1.2. Periodicity in the Inverted Magnetic-Field Strength

The daily variation manifests itself differently in strong and weak fields. To quantitatively describe this V_r -dependent variation in the inverted magnetic field we consider Active Re-

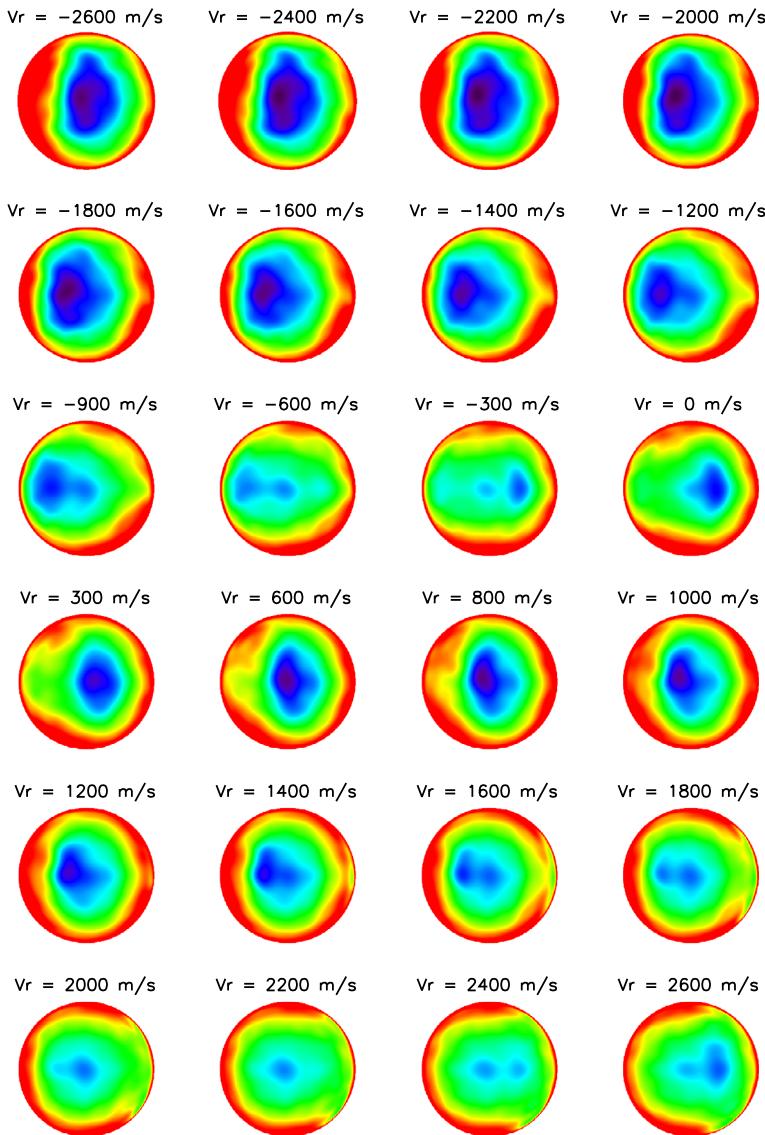
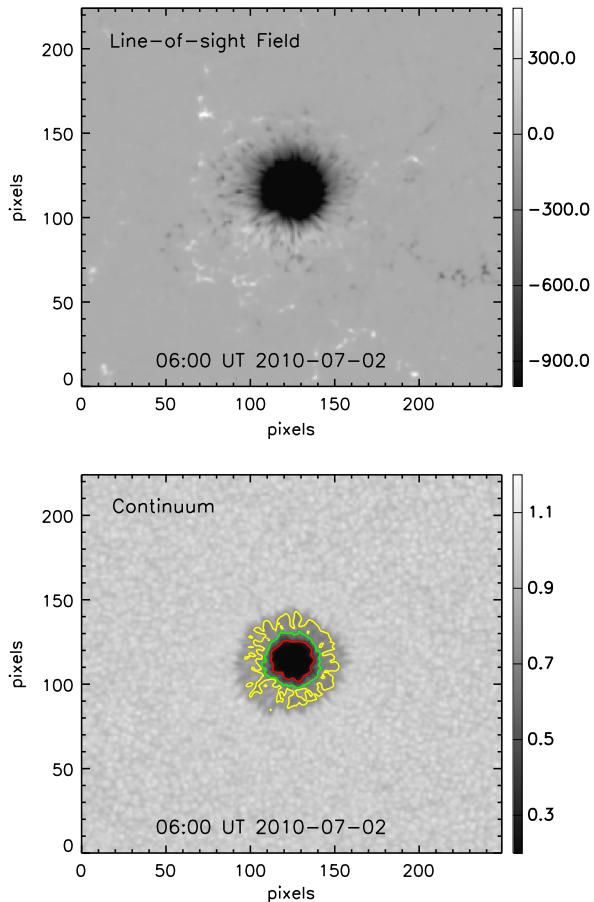


Figure 8 Derived noise masks for a range of orbital velocities, V_r . The data used for these reference masks were collected between 01 February and 11 March 2011. The colors represent values between $60-150 \text{ Mx cm}^{-2}$, as in Figure 7.

gion AR #11084. This is a mature early-cycle southern hemisphere active region with a single, stable, circular, negative-polarity sunspot (see Figure 9). The umbra and penumbra are clearly seen. The active region was tracked for four days, from 01–04 July 2010. Shown in black curves in the left column of Figure 10 are the mean magnetic-field strength as a function of time in the umbra of AR #11084 (top panel), the penumbra (second), and a quiet-Sun region (third); for reference the mean LoS velocity in a quiet-Sun region appears at the bottom of both columns. The quiet-Sun region is a 40×30 pixel rectangle in the bottom-right

Figure 9 HMI LoS magnetic field, B_{LoS} (top), and continuum intensity (bottom) of AR #11084 at 06:00 TAI on 2 July 2010 when the region was at S19 E01. The pixel size is 0.504 arc sec. The units of field are Mx cm^{-2} and intensity is shown in relative units of the mean nearby quiet-Sun continuum intensity. Intensity contours are drawn at 0.35, 0.65, and 0.75 of the quiet-Sun continuum intensity in red, green, and yellow, respectively.



corner of the images in Figure 9 where no evident magnetic features appear. The fixed-size quiet-Sun region moves with AR #11084. The red curves in the three upper panels on the left are a third-order polynomial fit to the evolving mean field strength. The residuals are plotted in the right column. The variations in the residuals show a strong correlation to the relative velocity.

Scatter plots of the mean field strength residuals and the relative velocity better illustrate this dependence. The left panels of Figure 11 show that a $\sim \pm 2 \text{ km s}^{-1}$ relative radial velocity variation causes a $\pm 1\%$ change in the umbral field strength through the day, or if a linear fit is made, the field strength varies $\sim 15 \text{ G/km s}^{-1}$. In the penumbra the daily variation has slightly smaller magnitude, but is a larger fraction of the mean value, $\sim \pm 2\%$. A linear fit to velocity gives about 6 G/km s^{-1} , though there seems to be a concentration of high positive-value pixels below -1.5 km s^{-1} that may be an artifact of the polynomial fitting. When these points are excluded, the linear fit for the penumbra is $\sim 8 \text{ G/km s}^{-1}$ (dashed line). An analysis of the LoS and transverse fields separately confirms that this velocity-dependent variation is only seen in the line-of-sight component of the strong field (in the penumbra and umbra). The strong field shifts either left or right circular polarization away from one

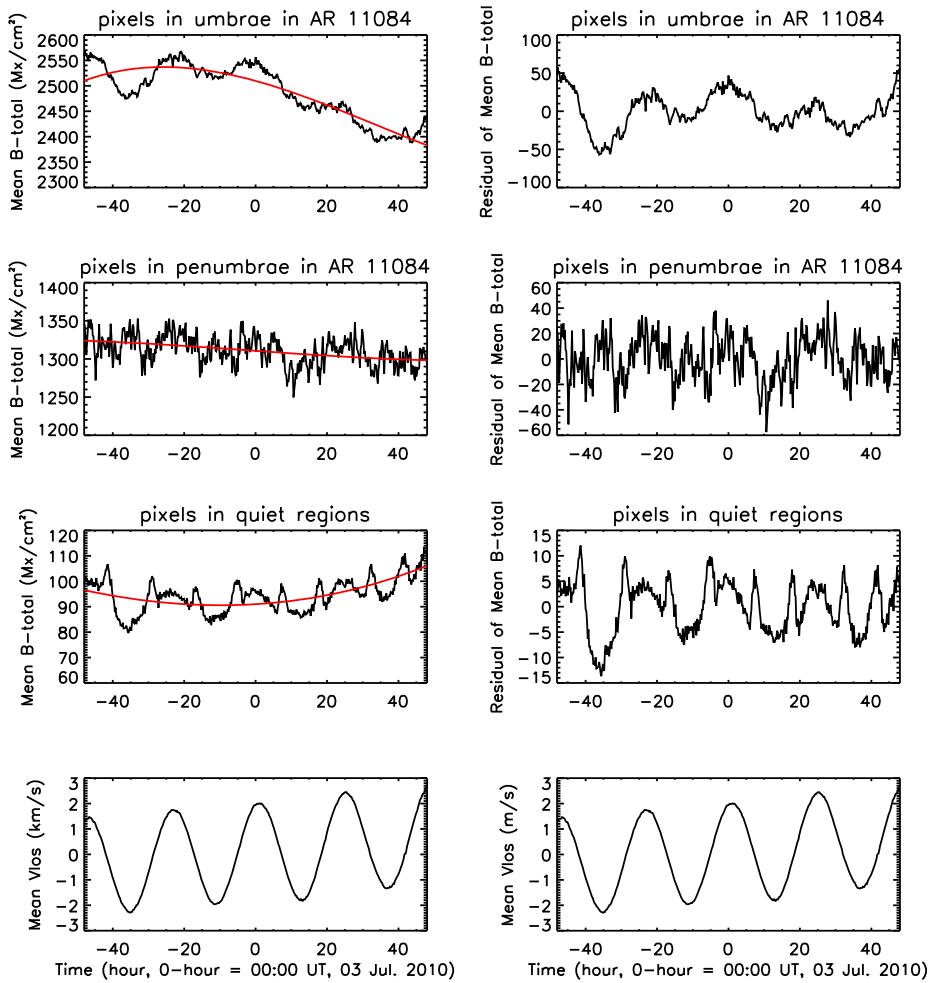


Figure 10 The black curves in the left column in the top three panels are four-day temporal profiles for AR #11084 of the mean magnetic-field strength determined for the umbra, penumbra, and quiet Sun, respectively. The red curves are third-order polynomial fits. Residuals are plotted in the top three panels on the right. The residual is the difference between the mean field strength and the polynomial fit. For reference, the mean LoS velocity observed in the quiet-Sun region is plotted in the two bottom panels. For this analysis the umbra includes pixels where, compared with the quiet Sun, the continuum intensity, $I_c < 0.35$ when corrected for limb darkening. In the penumbra $0.65 < I_c < 0.75$.

of the wavelength-tuning positions, whereas it only broadens the linear polarization that contributes to Q and U. This may provide a partial explanation for the velocity-dependent variation in the strong field.

Variability of the mean field strength in the quiet Sun is as much as 5 % of the typical weak-field magnitude and is dominated by daily variations and a reproducible twice-daily excursion centered near 0 km s^{-1} . That excursion appears only in the transverse field, *i.e.* Stokes Q and U, and is attributed to sensitivity that depends on the line position relative to the HMI tuning wavelengths. The linear trend is just 2.3 G/km s^{-1} . When the range -1.0 to $+0.5 \text{ km s}^{-1}$ is excluded, a linear fit gives 2.6 G/km s^{-1} (dashed line). Note that the ve-

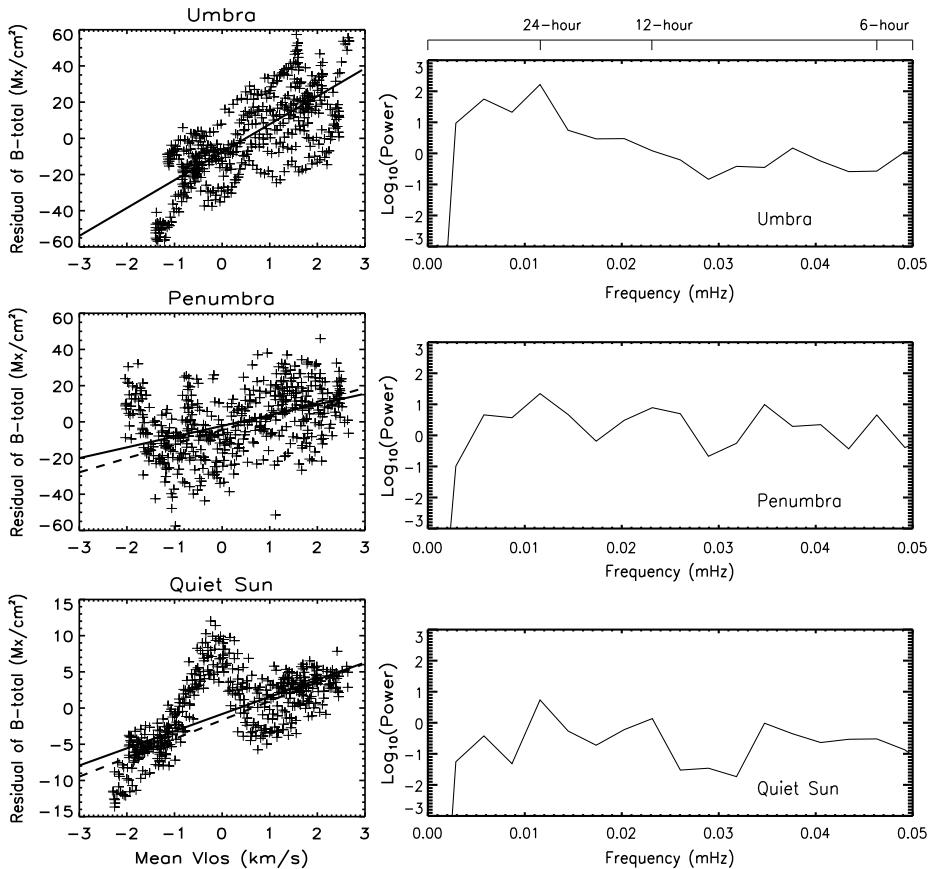


Figure 11 Scatter plots on the left show the relationship between the residual of the mean field strength and LoS velocity in the AR #11084 umbra (top), penumbra (middle), and nearby quiet Sun (bottom). Note the scale change for the quiet Sun. Linear fits are described in the text. The right panels show the power spectra of the mean field strength residuals for the sunspot umbra (top), penumbra (middle), and quiet Sun (bottom).

lacity range does not extend to as highly negative velocity values in the strong-field regions because of the suppression of the convective blue shift. Power spectrum analysis shows a peak at 24 hours in all three region classes (see right panels of Figure 11). The 24-hour peak is strongest in the umbra and less so in weaker field regions, which also have a significant 12-hour periodicity caused by the 0 km s^{-1} excursion.

Does the sensitivity of the magnetic field strength measurement to radial velocity itself vary with field strength? The dependence of the magnitude of the V_r -dependent variation of the magnetic field on the measured magnetic-field strength is shown in Figure 12. The linear dependence on Doppler velocity of the residual magnetic field is determined for a sample of 20 simple, stable sunspots. For each sunspot the data are sorted using the continuum intensity (`hmi.lc_720s`) as a proxy for the magnetic-field strength. As above, the average magnetic field measured in each intensity bin forms a several-day time series from which a third-order polynomial fit is removed. The magnetic-field residual is fit to the Doppler velocity, as in Figure 11, to determine the sensitivity of the magnetic measurement to velocity. The y-axis in Figure 12 is the slope of this linear fit in units of $\text{Mx cm}^{-2}/\text{km s}^{-1}$. The

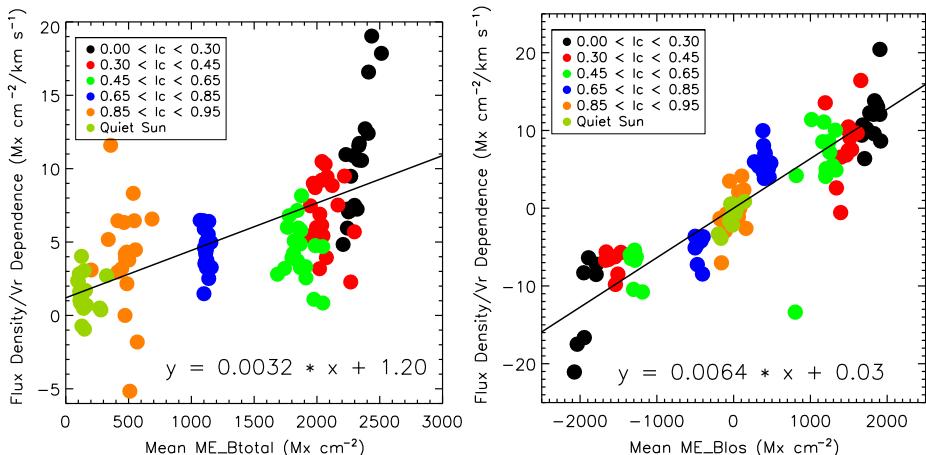


Figure 12 The change in measured magnetic-field strength with radial velocity is plotted as a function of field strength for a selection of 20 simple sunspots. Each spot is divided into continuum intensity bins to segregate by field strength. The bins are tracked for several days and the dependence of measured field strength on V_r is determined. The slope is plotted for each sunspot in each bin as a function of the measured field strength. The left panel shows the slope versus total magnetic-field magnitude. There is a significant zero offset and a relationship of $0.0032 \text{ Mx cm}^{-2}/\text{km s}^{-1}$ per Mx cm^{-2} . The LoS component of the vector magnetic field, B_{LoS} , can be determined, and the relationship is shown in the right panel. Note the scale change. There is a smaller zero offset and a relationship of $0.0064 \text{ Mx cm}^{-2}/\text{km s}^{-1}$ per Mx cm^{-2} .

x-axis indicates the mean magnetic field strength of the region in each intensity bin. The left panel shows the slopes as a function of total field strength. Colors denote bins of continuum intensity with the indicated range. “Quiet Sun” reflects data in a fixed 40×30 pixel rectangle at the lower right corner of the HARP bounding box where there is no significant flux concentration. The solid black line is a linear fit to the points. The variation increases linearly from $1.2 \text{ G}/\text{km s}^{-1}$ at 0 field strength to $4.4 \text{ G}/\text{km s}^{-1}$ at 1000 G . From the vector we can also determine the LoS component of the magnetic vector, B_{LoS} . The right panel shows how the velocity dependence of B_{LoS} changes with B_{LoS} . A relationship between B_{LoS} and the magnitude of the V_r -dependent variation is clearly shown. There is virtually no variation with velocity at low field values, but the variation increases linearly to $6.4 \text{ G}/\text{km s}^{-1}$ at 1000 G .

7.2. Known Issues

7.2.1. Bad Pixels

Occasionally, poor inversion results are obvious in the inverted magnetic-field data. They are particularly noticeable in sunspots, where pixels with abnormal values differ significantly from those adjacent to them or at nearby time steps. This usually happens when the inversion code fails to converge.

Some examples are shown in Figure 13. In the top left panel, the field strength in HARP 1638, AR #11476 at 12:00 TAI 8 May 2012 at N10 E34 is shown. Pixels with magnetic fields significantly stronger than the surroundings are denoted by an arrow. A horizontal cut through this image is plotted in the lower left panel, where it is easy to see an abrupt jump from 2000 Mx cm^{-2} to 5000 Mx cm^{-2} (the VFISV maximum). The panels in

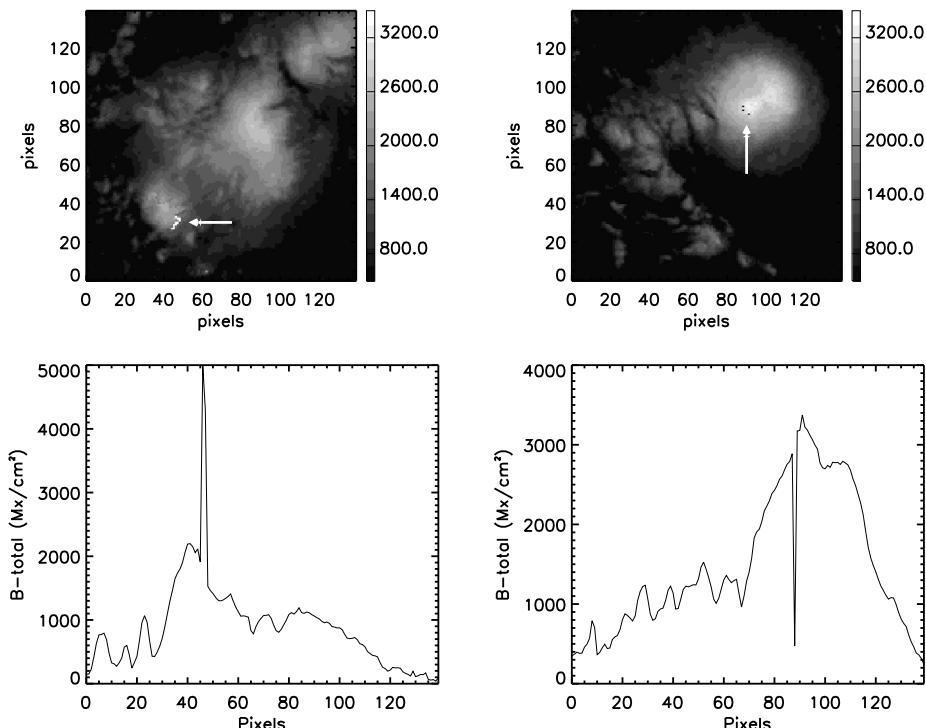


Figure 13 Two types of bad pixels. Top left: magnetic-field strength of AR #11476 at 12:00 TAI 8 May 2012 at N10 E34. The erroneous values are denoted by an arrow. The field strengths are much higher than in adjacent pixels. Bottom left: field strength along a horizontal row including a bad pixel. Note that 5000 G is the highest possible value allowed by the VFISV code. Top right: field strength of AR #11515 at 02:24 TAI 1 July 2012 at S17 E29. Bad pixels are denoted by an arrow. The field strengths at bad pixels are much lower than in adjacent pixels. Bottom right: field strength along a horizontal row including a bad pixel.

the right column display another instance of a failed convergence. In this case, HARP 1807, AR #11515 at 02:24 TAI 1 July 2012 at S17 E29 presents a few pixels with rather low field strengths. Again, a horizontal cut through this image in the lower right panel shows a drop from 3000 Mx cm^{-2} to 400 Mx cm^{-2} for an adjacent pixel.

A number of reasons can lead the inversion code to fail to converge. Inside very dark sunspots, where the magnetic field is very strong, the Zeeman splitting of the spectral line can be strong enough to push one (or both) of the Stokes V (Q and U) lobes outside of the measured spectral range observed by HMI. This problem is exacerbated by the Doppler shift that arises as a combination of the orbital velocity of SDO, the solar rotation, and the intrinsic plasma motions at the photosphere. In areas with strong gradients of the magnetic field or the LoS velocity (such as the penumbra of a sunspot, an emerging region, or a flaring region) the Stokes profiles tend to show large asymmetries. The inversion code often fails to converge in these cases because the ME approximation is strongly violated.

In addition to having abnormal field strengths, these pixels typically have out-of-the-ordinary values in other quantities computed with VFISV. This provides a way to identify them. Figure 14 shows a cluster of anomalous inversion values in AR #11476. They are clearly seen in the error of the field strength, and are also visible in the errors of inclina-

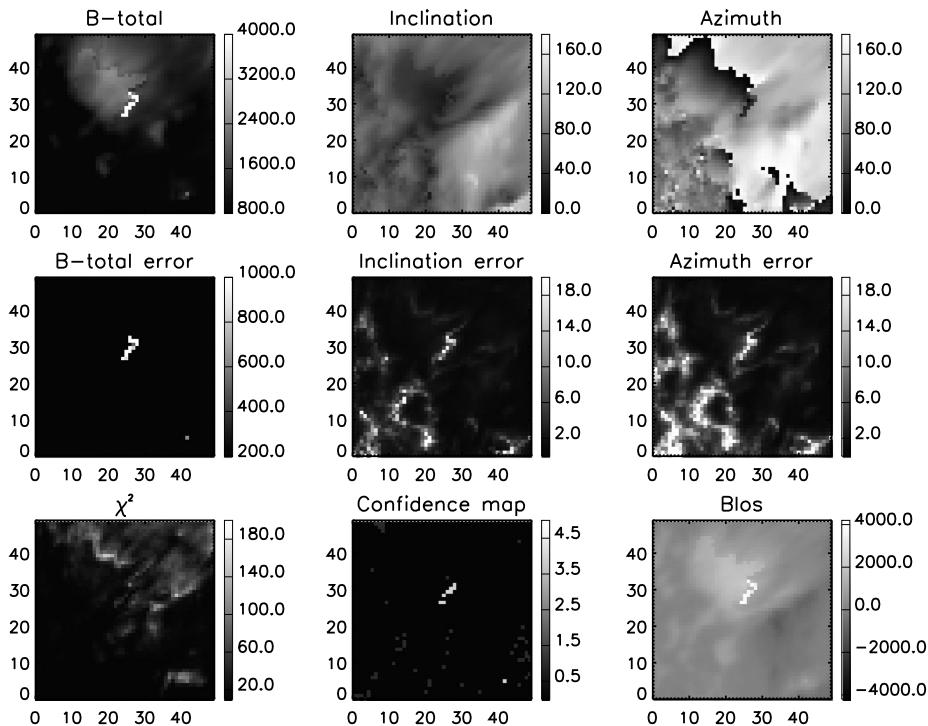


Figure 14 CCD cut-outs of inverted data for a small area in HARP 1638, AR #11476 showing bad pixels. Panels from left to right and top to bottom are field strength, inclination, azimuth of magnetic field, their errors, χ^2 for the least-squares fit, the confidence map, and LoS magnetic field. Erroneous values can be clearly seen in field strength, error of field strength, and LoS field. They are also visible in errors of inclination and azimuth. Units along the x- and y-axis are pixels.

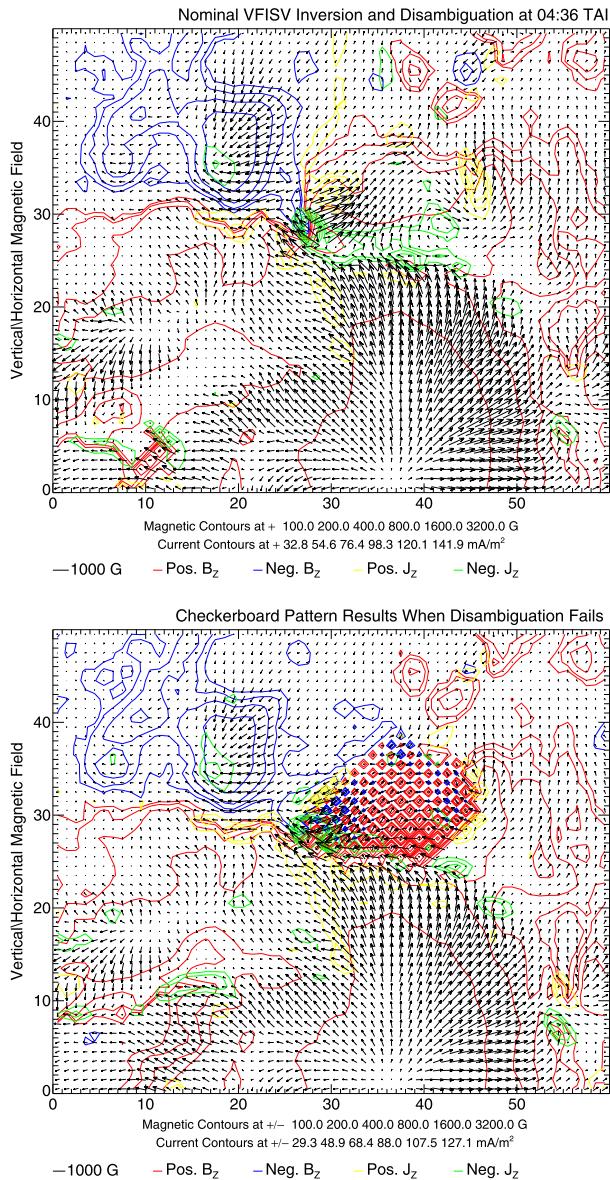
tion and azimuth and the LoS field. Some also appear in the confidence maps, although they do not always show up in χ^2 . A description of the data confidence map is given in Section A.1.

7.2.2. Reliability of the Disambiguation

There are two main failure modes for the simulated annealing part of the disambiguation: 1) the minimization algorithm fails to find the global minimum of the energy, or 2) the global minimum does not correspond to the correct disambiguation. In the first case, the result can be improved by changing the cooling schedule used for the annealing, but it is computationally impractical to use a cooling schedule in the pipeline that will find the global minimum. Typically, however, the pixels that differ from the global minimum configuration occur in areas where the noise in the transverse field is high. The second case can occur for several reasons. As in the first case, when the transverse field is dominated by noise or is not spatially resolved, then finite differences do not accurately represent the derivatives of the field. Alternatively, if the field is sufficiently far from potential, then the approximation to the divergence of the magnetic field will not be accurate.

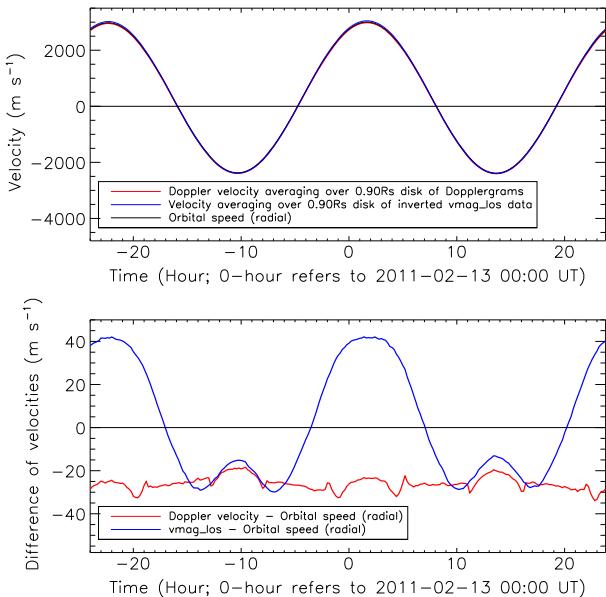
One common manifestation of a failure of the annealing to reach the global minimum is the “checkerboard” pattern illustrated in Figure 15. In this case, the azimuthal angle al-

Figure 15 Closeup of magnetic field and currents in AR #11158 for two consecutive time steps on 16 February 2011 at 04:36 and 04:48 TAI. The red and blue contours show the vertical magnetic field and the arrows indicate the strength and direction of the horizontal field component. Yellow and green contours show where the computed vertical current is strong. Axes are in pixels. The upper panel at 04:36 TAI shows a fairly normal field configuration. The central region of the lower panel at 04:48 TAI with diamond-shaped contours shows a case where the disambiguation module gives conflicting results in adjacent pixels. Certain areas of an active region may be more susceptible to the instability, which will come and go from one time step to the next. See text for details.



ternates between neighboring pixels in being in the range $0^\circ \leq \psi < 180^\circ$ and being in the range $180^\circ \leq \psi < 360^\circ$. This particular pattern is an artifact of the four-point stencil used for the finite differences, but similar artifacts occur when other stencils are used. It occurs because the energy is very similar in the checkerboard configuration and in the configuration in which all the azimuths are in one or the other of the above ranges. Strong-field regions where the disambiguation is less stable are often regions where the inversion also gives variable results from frame to frame, suggesting that the magnetic field is complex in these regions.

Figure 16 Top: 48-hour temporal profiles of the known orbital radial speed relative to the Sun, V_r (black); the average over the inner 0.9 of the solar disk of Dopplergrams, V_{Dop} (red); and a similar disk-average of the LoS velocity from inverted VFISV data, V_{Inv} (blue). The highest redshift of $\sim 3000 \text{ m s}^{-1}$ occurs at dusk, a little after 1 UT. Bottom: offset difference $V_{Dop} - V_r$ (red) and the offset difference $V_{Inv} - V_r$ (blue) as a function of time. The average Doppler velocity, V_{Dop} , is blueshifted by about 30 m s^{-1} . V_{Inv} shows a clear daily variation correlated with V_r .



7.2.3. Comparing HMI Velocity and Magnetic Field Using VFISV and the MDI-Like Algorithms

A comparison is made between the precisely known satellite speed relative to the Sun, V_r , the calibrated HMI Doppler velocity determined using the standard MDI-like algorithm, V_{Dop} , and the LoS velocity obtained with the VFISV inversion V_{Inv} . The three curves in the top panel of Figure 16 show 48-hour temporal profiles of V_r (black), the velocity averaged over the inner $0.90 R_s$ from HMI Dopplergrams computed with the same algorithm and corrections used for helioseismology (V_{Dop} , red), and the LoS velocity averaged over the inner $0.90 R_s$ determined with VFISV (V_{Inv} , blue). R_s refers solar disk radius. These three measurements are nearly identical. The bottom panel of Figure 16 displays differences of averaged $V_{Dop} - V_r$ (red) and of averaged $V_{Inv} - V_r$ (blue). The calibrated Doppler velocity, V_{Dop} , is blueshifted $\sim -30 \text{ m s}^{-1}$ relative to V_r ; the calibration depends on a smoothed daily fit to the 45 s V_{Dop} observed with the other HMI camera. The offset in the fd10 velocity, V_{Inv} , has a daily periodicity with amplitude $\sim 40 \text{ m s}^{-1}$ associated with the wavelength shift caused by V_r . The error in V_{Inv} changes character at dawn ($\sim 14 \text{ UT}$), when the spacecraft is moving most rapidly toward the Sun. Welsch, Fisher, and Sun (2013) discuss the sources of offsets associated with instrumental effects and the convective blue shift.

Figure 17 shows a more detailed comparison between the velocity and magnetic fields. The inverted velocity, V_{Inv} , and inverted LoS magnetic field, B_{Los} , are determined using the fd10 inversion. The Doppler velocity, V_{Dop} , and LoS magnetic field, M_{Los} , are computed from LCP and RCP using the MDI-like algorithm (see Section 2.3). The data are all observed with the same HMI camera. Four pairs of observations, taken 12 February 2011 at 00:48, 06:48, 12:48, and 18:48 TAI are used. The magnetic comparison is shown in the left column of Figure 17 and the velocity comparison is shown in the right column. The MDI-like LoS magnetogram and Dopplergram values are shown on the x-axis versus the fd10 magnetic and velocity values on the y-axis. Results for different parts of the disk and for strong-field regions are shown. Note that the color scales in the left and right columns are different to

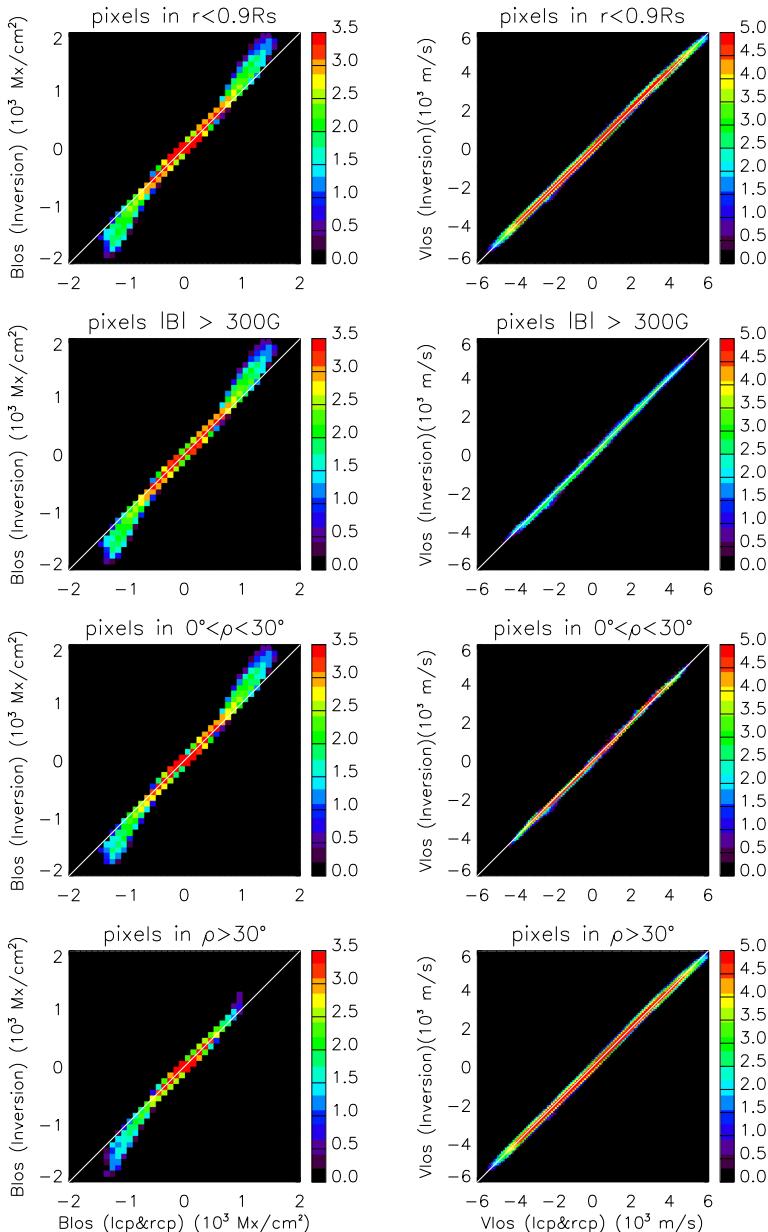


Figure 17 Comparison of VFISV inversion and the data derived using the MDI-like algorithm. The data used are the 720 s `f0d10` inverted data (ME), Dopplergrams and LoS magnetograms taken 12 February 2011 at 00:48, 06:48, 12:48, and 18:48 TAI. Comparison is made between LoS magnetic fields from the MDI-like algorithm (x-axis in left column) and from the ME inversion (y-axis in left column). The right column compares LoS velocities from calibrated Dopplergrams (x-axis in right column) with velocity from the ME inversion (y-axis in right column). The color table shows the logarithmic density of the number of scatter plot points. From top to bottom panels show comparisons for pixels within 0.90 R_s , strong-field pixels with $B_{\text{Tot}} > 300 \text{ Mx cm}^{-2}$, disk center pixels ($\rho < 30^\circ$), and limb pixels ($\rho > 30^\circ$), where ρ is the center-to-limb angle.

better show the details of the distribution of scatter-plot points. For reasons discussed in Section 2.3 and by Couvidat *et al.* (2012) and Liu, Norton, and Scherrer (2007), the MDI-like algorithm is not as accurate in strong-field regions and underestimates the field strength. The LoS velocities from the two algorithms agree very well.

7.3. Limits and Validation

The inference of solar photospheric magnetic-field vector maps is at best an estimate that represents the average field over height, time, and space, as deduced using a record of photons averaged in the spectral domain as well. In practice, vector magnetogram maps are the result of inverting data that only poorly constrain an optimization method based on physical assumptions known to be incorrect. There are additional instrumental artifacts that may or may not be mitigated fully. In this section, some rudimentary evaluation of the evolution of AR #11158 features are discussed in light of the limitations of the data, and a brief qualitative comparison is presented with simultaneous observations from the *Hinode/SpectroPolarimeter* (Kosugi *et al.*, 2007; Tsuneta *et al.*, 2008; Lites *et al.*, 2013). Comparisons beyond the qualitative presentations made here are beyond the scope of this paper.

7.3.1. Validation and Limitations of HMI Processing and Data

Characteristics of both the HMI instrument and the processing can lead to systematic and possibly subtle effects of which users need to at least be aware.

First, there is no treatment of scattered light, either as part of the data reduction or the inversion procedure. This can lead to systematically lower inferred field strengths and incorrect inferred angles of the magnetic vector (LaBonte, 2004b). It will also provide simply incorrect relative photometry between different solar structures. In the future a treatment of scattered light may be performed; this is a feature available in the original VFISV scheme.

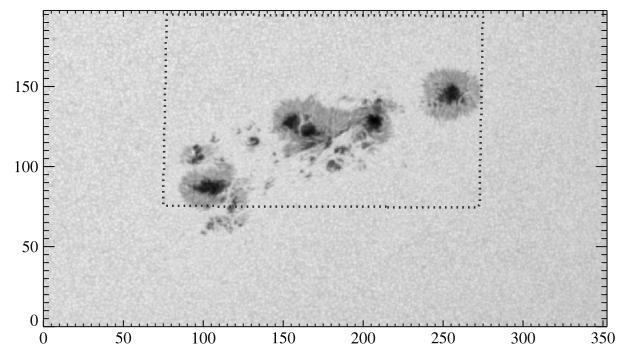
Second, the sparse and limited spectral sampling precludes a clean continuum point from being sampled consistently in the same part of the spectrum. For segments of the daily orbit, either the first or the last filter position should sample near the spectral continuum, but rarely far enough from the spectral line to be uncontaminated, and never for more than a few hours at a time. Again, this can impact the interpretation of relative photometry especially for analysis of temporally varying features.

Third, as summarized in Section 2.2, there is overlap in the contribution of filtergrams to each 12-minute average, such that a single magnetogram is not a completely independent measurement from its temporal neighbors. The effects of this may be subtle, but may be detectable in the velocity-tracking codes during times of strong temporal evolution (*e.g.* flares or emergence), and may contradict some assumptions made by statistical analysis.

7.3.2. HMI Vector Magnetograms in Context

Every instrument and observing scheme makes compromises to balance the limitations imposed by a finite number of photons and the performance of real optics. HMI's vector magnetogram capability, as with data from other vector-spectropolarimetric instruments, enables a broader spectrum of physical interpretation than instruments such as MDI that provide solely LoS measurements. Compared with other vector magnetogram data, HMI observations have much better spatial and temporal coverage than *Hinode/SP*, whereas the spatial and spectral resolution, sampling, and polarization sensitivity of *Hinode/SP* are significantly

Figure 18 Continuum image from hmi.sharp_720s centered on NOAA AR #11158, from 15 February 2011 at 10:36, with an outline of the field of view covered by the *Hinode*/SP fast-scan obtained between 10:11 – 10:55 TAI, which had approximately the same center time as the HMI data. Axes are in arc seconds.



better. HMI data have much better spatial and temporal coverage than the *Imaging Vector Magnetograph* (IVM) archive, but poorer spectral sampling, in addition to the advantages and disadvantages of space-based vs. ground-based observing. HMI's full-disk imaging capability makes its global spatial coverage superior to any data except the SOLIS/SPM and Huairou/HSOS. The HMI spatial resolution of $1''$ is similar to that of the IVM and the optical potential of SOLIS/SPM, but significantly poorer than that of *Hinode*/SP and certain high-resolution ground-based instruments. HMI's combination of cadence, coverage, and continuity are unmatched.

7.3.3. HMI Vector Magnetograms vs. *Hinode*/SP Vector Magnetograms

A simple statistical comparison was performed between co-temporal HMI and *Hinode*/SP data. The HMI data from hmi.sharp_720s were co-aligned with the *Hinode*/SP fast-scan in 20110215_101126.fits², which had a center-scan time approximately two minutes different from the time-stamp in the HMI data. The *Hinode*/SP map was inverted using the MERLIN inversion scheme as the data are available from the Level-2 production; no additional processing was performed. Figure 18 shows the HMI cutout area, with the field-of-view covered by the *Hinode*/SP scan indicated.

A comparison between continuum intensity (normalized by the median continuum intensity) and the inferred field strength for the two datasets is shown in Figure 19. The HMI data are reported in their native $0.5''$ sampling, reflecting the $\approx 1.0''$ spatial resolution of the HMI telescope. The *Hinode*/SP fast-scan data are sampled to match the $\approx 0.3''$ resolution of that telescope. The flux density is reported here for HMI data because the magnetic fill fraction is not a free parameter in the inversion and is fixed at unity. The field and magnetic fill fraction are fit and returned separately for *Hinode*/SP data inverted with MERLIN; for these comparisons, these two quantities are multiplied. No other manipulation of the data has been performed for this comparison.

The distribution plots demonstrate that the data are qualitatively very similar, but there are quantitative and systematic differences. The HMI flux density generally falls short of that reported by *Hinode*/SP, as does the contrast and its variation. Both of these effects are expected with lower spatial resolution (Leka and Barnes, 2012) and the presence of scattered light (LaBonte, 2004a). In these plots the flux-density effects are clearest in the sunspot umbrae, whereas the contraction of the contrast distribution is easiest to see in the quiet-Sun or plage areas. The lack of a scattered light correction may also contribute to

²sot.lmsal.com/data/sot/level2d.

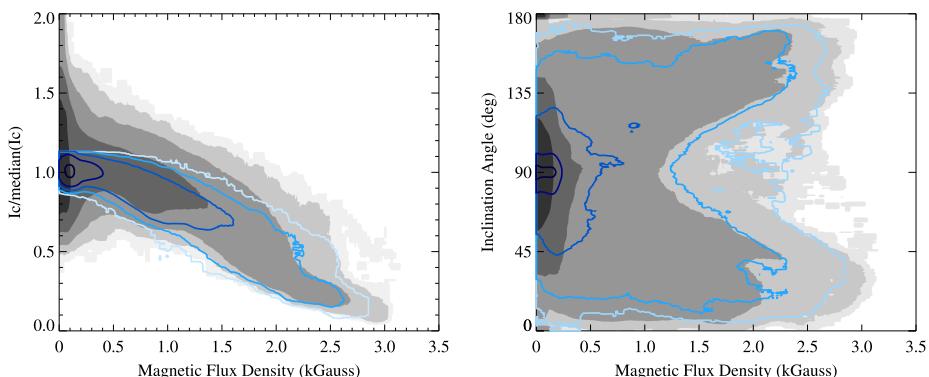


Figure 19 Left: density histogram of the scatter plots of the flux density (field times the inferred magnetic filling factor for *Hinode*/SP) vs. continuum intensity (normalized by the median). Right: flux density vs. the magnetic vector inclination relative to the LoS. Gray scale: *Hinode*/SP data, blue contours: HMI data. Contour levels are logarithmic in density, normalized for the number of pixels considered, at 0.01, 0.1, 1, 10, 100 (except that the highest contour is 50 for the intensity plot). Only the area co-observed by the two instruments is considered. The fraction of pixels inside each contour decreases from over 99.9 % within contour level 0.01 to between roughly 75 %–99 % at level 1.0 and to 30 %–50 % at the highest contour. If the distributions were identical, the contours and color-boundaries would coincide. The figure shows that the flux density measured by HMI is generally lower than that of *Hinode*/SP.

both the field strength and continuum-intensity differences. The inclination angles display the same morphology in the distribution plots, but the HMI data generally fail to achieve the extreme inclination angles at all flux densities and tend to lie closer to the plane of the sky. This again may be a consequence of scattered light (LaBonte, 2004b) or noise (Borrero and Kobel, 2011), or it may be attributable to the better wavelength sampling of *Hinode*/SP, which effectively leads to an enhanced signal-to-noise ratio.

8. Summary

HMI vector magnetic field observations have been collected continuously since 1 May 2010. Examples have been shown to demonstrate the capabilities of the HMI instrument and explain how the observables are generated routinely in the vector field pipeline analysis. HARP 377, AR #11158 was highlighted in part because it produced the first X-class flare of Cycle 24 and in part because an earlier version of the HMI vector field data were released for this region.

A sequence of full-disk, polarized, narrow-passband filtergrams are collected every 135 seconds. A Level 1 data analysis pipeline corrects the filtergrams for a variety of instrumental and observational effects, co-aligns them, and averages data from ten sequences together on a 720 s cadence (Schou and Couvidat, 2013). The basic magnetic observables, Stokes [I Q U V], are determined at six wavelengths. Other scalar quantities provided at the same cadence include Doppler velocity, line-of-sight magnetic field, continuum intensity, line depth, and line width. The data series are listed in Table 1.

HMI active region patches are identified automatically and tracked during their entire disk passage (Turmon *et al.*, 2014) based on analysis of the HMI LoS magnetic field and continuum intensity. Each HARP is assigned a number and may in many cases be associated with one or more NOAA active regions. There are also many HARPs that contain no sunspot

and therefore have no NOAA AR number. The geometric information in the HARP data series is used to select regions for further processing.

VFISV, a very fast Milne–Eddington inversion code, is applied to determine the magnetic field and other plasma parameters. The fcd10 version of the code described in this article and in more detail in Centeno *et al.* (2013) has been run on all HMI full-disk vector field data since 1 May 2010.

The inverted fcd10 data require disambiguation to resolve the 180° azimuthal uncertainty in the transverse field direction. We have applied the method described by Barnes *et al.* (2014) to a growing selection of active region patches. The disambiguation minimizes a quantity proportional to the sum of the absolute values of the field divergence and the electric current density. The pipeline disambiguation code is being run routinely on active-region patches observed before 15 January 2014. Full-disk disambiguated vector magnetograms are being used for more recent observations.

The SHARP (Space-weather HARP) data series gathers together for each HMI active-region patch at each time step the disambiguated vector field, other inversion outputs, and most of the HMI scalar observables. The SHARP data series also provides a number of active-region quantities computed at each time step for each patch, such as total flux, mean current helicity, and weighted measures of the field gradient (Bobra *et al.*, 2014). Maps of select inversion outputs are remapped into heliographic coordinates using a cylindrical equal area projection centered on the HARP.

The pipeline is applied to definitive HMI data that become available several days after the observations are collected. The definitive HARP analysis is not completed until after the region has rotated off the visible solar disk, introducing delays as long as several weeks. For most scientific purposes these high-quality data are the most useful. However, for space-weather forecasting and evaluative purposes, quick-look near-real-time data for many observables are generally available within minutes or at most a few hours of observation. These data include the full-disk scalar quantities and everything required to make SHARPs.

In Section 7 and elsewhere many of the known limitations, uncertainties, and issues associated with the vector field data were described. Prudence must be exercised when using the data quantitatively. HMI vector field measurements compare favorably with other observations, but of course have been optimized for the capabilities and constraints of the instrument, satellite, and processing.

Table 2 Useful web sites with information about HMI magnetic field data.

| Web address | HMI data product |
|--|---|
| hmi.stanford.edu/magnetic | HMI magnetic field images and data products |
| jsoc.stanford.edu | Introduction to the data in the JSOC |
| jsoc.stanford.edu/jsocwiki | Introduction to JSOC data series and access |
| jsoc.stanford.edu/jsocwiki/MagneticField | HMI magnetic product information |
| jsoc.stanford.edu/jsocwiki/HARPDataSeries | Description of HARP data series |
| jsoc.stanford.edu/jsocwiki/sharp_coord | SHARP coordinates and CEA remapping |
| jsoc.stanford.edu/jsocwiki/PipelineCode | Guide to HMI pipeline code and processing notes |
| jsoc.stanford.edu/ajax/lookdata.html | Data exploration and request tool for JSOC |
| jsoc.stanford.edu/ajax/exportdata.html | Data export tool for JSOC data |
| www.lmsal.com/sdouserguide.html | Comprehensive guide to SDO data analysis |
| sdo.gsfc.nasa.gov | SDO mission web site |

Table 2 provides links to additional documentation relevant to the HMI magnetic field data.

Acknowledgement We thank the numerous team members who have contributed to the success of the SDO mission and particularly to the HMI instrument. This work was supported by NASA Contract NAS5-02139 (HMI) to Stanford University. Some of the research described here was carried out by staff of the Jet Propulsion Laboratory, California Institute of Technology. HAO/NCAR is supported by the National Science Foundation. Efforts at NWRA were also supported through NASA Contract NNN09CF22C and by PO#NNG12PP28D/C#GS-23F-0197P from NASA/Goddard Space Flight Center. J. Schou acknowledges support from EU FP7 Collaborative Project *Exploitation of Space Data for Innovative Helio- and Asteroseismology* (SPACEINN).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix A: Explanation of Certain Vector Field Data Segments

A JSOC *data series* consists of *keywords*, *segments*, and *links* for a set of *records*. A *record* holds the information for specific *primekeys*. The *primekey* is typically a time, or in the case of HARPs and SHARPs, the two *primekeys* are HARP number and time. A *record* can have one or more *segments* that generally contain image data. A number of *keywords* contain metadata about the data series or about particular *records*. For example, in most data series the keyword T_REC gives the time of the *record* and the keyword QUALITY holds information about possible issues in the collection or processing of the data (see jsoc.stanford.edu/doc/data/hmi/Quality_Bits/QUALITY.txt for details).

In the vector field data series the segment FIELD is the array of total magnetic field strength. Various *links* provide pointers to segments in other data series; for example, the link HMI.MHARP_720S in the hmi.sharp_720s series is a link to the full disk line-of-sight magnetogram at time T_REC. The elements of the series are stored independently until they are assembled on export. Detailed explanations of how data are stored and accessed can be found at the JSOC wiki page, jsoc.stanford.edu/jsocwiki/DataSeries. Brief definitions of keywords, segments, and links for each particular series are available using the lookdata utility at jsoc.stanford.edu/ajax/lookdata.html in the “Series Content” tab.

Most data segments in the vector field data series are self-explanatory, *i.e.* they contain maps of an observable or inverted physical quantity, an uncertainty, or the correlation of the errors in two inverted parameters. For example, the data segment INCLINATION contains a map of the magnetic-field inclination angle relative to the line of sight; the data segment AZIMUTH_ERR is the formal error of the magnetic-field azimuth value, and FIELD_INCLINATION_ERR is the cross correlation of the field strength and inclination errors as determined by the fd10 VFISV inversion.

Certain other data segments that are more obscure in their meaning (see Table 3) are explained in the following sections. For the most part they are the same in all of the vector field data series, but in some cases (as noted) the name is changed or the values are updated. Key-words generated by upstream modules in the analysis pipeline are generally not described here.

A.1 Data segment CONFID_MAP

Data segment CONFID_MAP quantifies the data quality of each pixel in a very simple way. Values range from 0 to 6 with lower numbers indicating higher confidence in the reported quantities resulting from the VFISV inversion and disambiguation (see Table 4).

Table 3 Vector data segments requiring additional explanation.

| Segment name | Description |
|---------------|---|
| CONFID_MAP | Quantitative estimate of data quality |
| INFO_MAP | Encoded confidence indicator for data quality in <i>B</i> and SHARP |
| QUAL_MAP | Same as INFO_MAP for ME series |
| CONV_FLAG | Status of the VFISV inversion code result |
| DISAMBIG | Code for azimuthal disambiguation |
| CONF_DISAMBIG | Confidence indicator for disambiguation |
| BITMAP | HARP region map of active pixels inside and outside patch |

Table 4 Meaning of CONFID_MAP values.

| Value | Meaning |
|-------|--|
| 0 | Pixel with no known problems |
| 1 | One of Stokes [Q U V] or LoS field was lower than noise threshold |
| 2 | Two of Stokes [Q U V] or LoS field were lower than noise threshold |
| 3 | Stokes [Q U V] and LoS field all lower than noise threshold |
| 4 | VFISV inversion failed to converge properly, see CONV_FLAG |
| 5 | Bad pixel, defined as aberrant compared with adjacent pixels |
| 6 | Missing data in at least one of input Stokes [I Q U V] |

Table 5 Description of QUAL_MAP and INFO_MAP Bits.

| Bit | Description |
|-------|--|
| 0–7 | Reserved for disambiguation information (TBD – to be defined) |
| 8–23 | TBD |
| 24–26 | Binary code indicating status of VFISV computation and inputs See CONV_FLAG table in next section for bit assignments |
| 27 | Bad pixel; 0 = NO, 1 = YES (VFISV module) |
| 28 | LoS field lower than noise threshold; 0 = NO, 1 = YES (VFISV module) |
| 29 | Stokes Q or U lower than noise threshold; 0 = NO, 1 = YES (Stokes module) |
| 30 | Stokes V lower than noise threshold; 0 = NO, 1 = YES (Stokes module) |
| 31 | Any missing Stokes [I Q U V] data; 0 = NO, 1 = YES (Stokes module) |

A.2 Data Segments QUAL_MAP and INFO_MAP

The segment QUAL_MAP or INFO_MAP describes qualities of the vector magnetic field data with defined codes as described in Table 5. QUAL_MAP is used for fd10 products and INFO_MAP is used for disambiguated vector products. The contents are equivalent. These segments are 32-bit arrays that contain bitwise encoded information about the VFISV fd10 inversion and disambiguation for each pixel.

Bit 27 indicates a questionable pixel, which is defined as one having inversion values that are abnormal compared with adjacent pixels. For example, the field strength value in a bad pixel is often much lower than that of the adjacent pixels. Currently, we use an algorithm

Table 6 VFISV CONV_FLAG error codes.

| Code | VFISV error code |
|------|--|
| 000 | VFISV inversion module completed nominally |
| 001 | Continuum intensity below required threshold; pixel not inverted |
| 010 | Maximum number of iterations reached without converging |
| 100 | Value of χ^2 too high; exited loop |
| 101 | Singular Value Decomposition (SVD) failed to compute errors |

to identify these bad pixels. We evaluate $|M_{\text{LoS}}| - |B_{\text{LoS}}|$, where M_{LoS} refers to the LoS field derived using the MDI-like algorithm (see Section 2.3), and B_{LoS} is LoS field from the VFISV inversion. If the difference is greater than a threshold value, *i.e.* the LoS inverted field is much lower than the MDI-like field, it is likely to be a bad pixel. The threshold value is currently set to 500 Mx cm^{-2} , which successfully identifies bad pixels in sunspot umbrae, but sometimes fails to find bad pixels outside of sunspots.

A.3 Data Segment CONV_FLAG

The status of the fd10 VFISV inversion module processing is recorded in the character array CONV_FLAG, which is of the same size as the magnetic-field map (see Table 6). The value is a three-bit code that indicates if there is a problem in the inversion of the pixel. This information is repeated in bits 24–26 of the QUAL_MAP segment.

A.4 Data Segment DISAMBIG

The segment DISAMBIG is a character array the size of the magnetic-field map that appears in disambiguated data series in native CCD coordinates, *e.g.* hmi.sharp_720s, hmi.Bsharp_720s, and hmi.B_720s. If 180° must be added to the result of the fd10 inversion reported in AZIMUTH, a bit is set in the DISAMBIG segment. Only in the four SHARP data series are the AZIMUTH or field component data actually modified.

Pixels are considered to be high confidence, intermediate confidence, or weak confidence as reported in the CONF_DISAMBIG segment described in the next section. In high-confidence pixels the minimum-energy method is always used. In intermediate pixels the minimum energy result is smoothed using the acute angle method. In weak pixels the results of three disambiguation methods are provided.

For the patch-wise SHARP analysis applied to HMI data collected before 15 January 2014, the pixel disambiguation is generally considered high-confidence if the transverse field strength exceeds the velocity and position-dependent HMI noise mask by a threshold, DOFFSET = 20 G. The threshold is currently set to 50 G for full-disk disambiguation. Adjustments are made by eroding the region to eliminate isolated pixels. See Section 7.1.1 for a more detailed description of the noise mask.

Pixels below the threshold but near strong-field regions are considered to be intermediate. In patch-wise disambiguation analysis all pixels that are not strong are considered to be intermediate. In the full-disk disambiguation the intermediate pixels are those within five pixels of the high-confidence region. The minimum-energy disambiguation is smoothed using the nearest-neighbor acute-angle method in intermediate regions.

Table 7 DISAMBIG Bit codes.

| | | |
|--------------|-----|---|
| Confidence | Bit | If set, add 180° to value in AZIMUTH |
| High | All | Add 180° to azimuth per minimum-energy method |
| Intermediate | 0 | Add 180° per smoothed minimum-energy method |
| Weak | 0 | Add 180° to azimuth per potential-field method |
| Weak | 1 | Add 180° to azimuth per random-assignment method |
| Weak | 2 | Add 180° to azimuth per radial-acute angle method |

The minimum-energy disambiguation is reported for all high-confidence pixels above the noise threshold. All pixels are considered to be either strong or intermediate in patch-wise disambiguation.

In weak-field regions the full-disk disambiguation module reports the results of three alternative methods by setting bits as described in Table 7. The minimum-energy method is not applied in weak regions. Instead, we provide the disambiguation consistent with a potential field, a random assignment, or a radial-field acute-angle method. The SHARP azimuth reported for full-disk disambiguation in weak regions is the radial-field acute-angle determination.

In some hmi.sharp_720s series processed before August 2013 additional information is provided for intermediate-strength pixels. Bit 0 is the standard smoothed minimum energy result, bit 1 is the result of a random assignment, and bit 2 is the radial acute-angle determination. Use of this information is deprecated.

The CONF_DISAMBIG segment indicates the pixel type.

A.5 Data Segment CONF_DISAMBIG

Data segment CONF_DISAMBIG is an indicator of the confidence in the disambiguation result reported in DISAMBIG. The value depends on field strength and proximity to strong-field regions. The CONF_DISAMBIG segment is a character array the size of the magnetic-field map. If the pixel was not disambiguated, the value is 0, indicating no confidence in the DISAMBIGUATION segment. The value is set to 90 if the annealed results of the minimum energy method are used directly, *i.e.* in strong-field pixels. In intermediate confidence regions, where smoothing is applied to the annealed result, the value is 60. When disambiguation is performed on HARP patches, all on-disk pixels within the rectangular HARP bounding box have a value of 90 or 60. When full-disk disambiguation is performed, locations more than five pixels from a high-confidence region are considered to be lower-confidence and a value of 50 is assigned. Weak pixels are currently identified only in full-disk disambiguation. In principle, other values may be assigned if another method is used.

A.6 Data Segment BITMAP

The HARP module defines a rectangular array on the CCD called a bounding box. The bounding box encloses the identified active region. The active region consists of one or a few contiguous regions with smooth borders within the bounding box. The bounding box may enclose an area substantially larger than the actual active region at some time steps. Within the bounding box the BITMAP also distinguishes between active and quiet pixels. Active pixels are those above a certain defined threshold in field strength. The BITMAP array has the same dimensions as the bounding box (see Table 8).

Table 8 Meaning of HARP BITMAP values.

| Value | Meaning |
|-------|--|
| 0 | Off-limb pixel anywhere inside HARP bounding box |
| 1 | Quiet pixel outside the HMI active-region patch |
| 2 | Active pixel outside the HARP |
| 33 | Quiet pixel inside the HARP |
| 34 | Active pixel inside the HARP |

Appendix B: The Earlier Version of HMI Vector Field Data: e15w1332

The first HMI vector field data released for scientific analysis in late 2011 provided information about NOAA AR #11158, the first X-class flare producing region of Solar Cycle 24. Subsequently, data for several other regions were released, all processed with the same earlier version of the pipeline code, designated e15w1332. These data, described at jsoc.stanford.edu/jsocwiki/ReleaseNotes and in [ReleaseNotes2](#), are archived at JSOC in the data series hmi.B_720s_e15w1332. and hmi.B_720s_e15w1332_CEA. The current data that are generated are labeled fd10. Section 4 of this article gives a brief description of the fd10 pipeline processing, but an in-depth description can be found in Centeno *et al.* (2013) and Barnes *et al.* (2014). This Appendix briefly summarizes the differences between the processing that generated the e15w1332 data and the current pipeline that generates the fd10 and SHARP data:

- The input HMI observables, the hmi.S_720s Stokes parameters, are unchanged between the e15w1332 and fd10 data.
- The VFISV code that generated the e15w1332 data ran relatively slowly and did not converge as reliably. The code was subsequently optimized for speed as described below.
- First, the forward calculation of the spectral line was changed to be a non-explicit calculation for wavelengths beyond $\pm 0.65 \text{ \AA}$ from the core of the line. This was made under the assumption that there are no significant spectral features beyond the central core region of the line. This change sped up the inversion by $\sim 64\%$.
- Second, instead of inverting the source function (S_0) and its gradient (S_1) separately, S_0 was fit and ($S_0 + S_1$) were fit together in the production of fd10 data, improving the efficiency of the inversion algorithm since these parameters are strongly coupled. Similarly, eta0 (η_0) and Doppler width ($\Delta\lambda_D$) were found to be degenerate, in the sense that different combinations of the two parameters resulted in very similar goodness of fit. The earlier inversion code fit η_0 and $\Delta\lambda_D$ separately, but fd10 fits $(\Delta\lambda_D \times \sqrt{\eta_0})$ and $\Delta\lambda_D$. These variable changes account for a $\sim 10\%$ reduction in computing time for a full-disk map.
- Third, the convergence criteria were altered to reduce the average number of iterations from 200 for e15w1332 data to ~ 30 for fd10. Ordinarily, when the χ^2 values of two consecutive successful iterations (*i.e.* χ^2 is decreasing) differ by less than a tolerance value, the algorithm exits the iteration loop. Although for most pixels the inversion results do not change after 30 iterations, a small percentage ($< 1\%$) benefit from increasing the number of iterations to 200. The convergence criteria for e15w1332 were set very conservatively, $\epsilon < 10^{-15}$, so that at each point the maximum 200-iteration limit was reached; this is the origin of the e15 in the series name. A new diagnostic procedure was implemented for the fd10 data that identified the more difficult pixels (the $< 1\%$ of pixels) and enforced restarts on the problematic pixels while reducing the average number of iterations overall.

- A set of weights for the Stokes profiles that provided the smoothest possible solution inside active regions (and performs less well in weak field regions) was chosen for both the e15w1332 and fd10 data. The values are [1:3:3:2] for [I Q U V] for the e15w1332 data. A normalization of the weights by relative photon noise of the Stokes profiles was implemented in the fd10 data so that the equivalent weights are [1:2.89:2.89:2.02] for [I Q U V], as described in Centeno *et al.* (2013).
- A regularization term that penalizes high values of η_0 was added to the merit function in the fd10 version of VFISV. This minimizes a double-minima problem in the parameter space and helps prevent the code from converging to unphysical solutions. This was not present in the production of the e15w1332 data.
- Limits on the range of each atmospheric parameter are set to prevent the algorithm from probing extreme unphysical values. These limits were set to different values for the production of e15w1332 data.
- The disambiguation for e15w1332 data is essentially as described in Section 5, but used different annealing parameters. The cooling speed was 0.998 for e15w1332 and 0.98 for fd10. The number of configurations attempted at each cooling speed was 50 for e15w1332 and 100 for the SHARPs. The e15w1332 data relied on a simpler (not time-varying) treatment of the background noise.

The value of the threshold field strength is primarily determined by the noise level in the field. Due to the challenges in determining a realistic noise value from the inversion (see Section 7.1.1 for a discussion of the temporal and large-scale spatial variations found in the data), the following simple expression was used for the threshold value:

$$B_{\text{thresh}} = B_1 + (B_0 - B_1)\sqrt{1 - \mu^2}, \quad (\text{B.1})$$

where $\mu = \cos\theta = r/r_s$ measures the distance from disk center; thus, the parameter B_0 determines the threshold at disk center, while the parameter B_1 determines the threshold at the limb. For the e15w1332 data, we set $B_0 = 200$ G and $B_1 = 400$ G. Section 7.1.1 explains the derivation of the threshold set for fd10.

- No SHARP data were generated using the e15w1332 processing. The geometry for NOAA AR #11158 generated with the e15w1332 processing was based on HARP analysis with the bounding-box padding adjusted manually to include a slightly greater area than was automatically generated.

Appendix C: SHARP Space-Weather Quantities

With a regular time series of vector magnetic-field data, the time evolution of each active region can be characterized. Table 9 lists the indices computed every 12 minutes that are currently available as SHARP keywords. These quantities are generally mean values, sums, or integrations computed for the tracked HMI active-region patches using the high-confidence values in the remapped CEA series. The high-confidence pixels are those with CONF_DISAMBIG = 90 (see Section A.5) and the number of pixels is given in keyword CMASK. The formal uncertainty associated with each parameter is given its own keyword. WCS-standard FITS header keywords CDELT1, RSUN_OBS, and RSUN_REF were used to convert to the units specified in the units column. NRT data can be used for forecasting and definitive data for studies of how to make predictions. Details are provided in Bobra *et al.* (2014).

Table 9 SHARP active-region parameter formulae.

| Keyword | Description | Units | Basic formula ^a |
|---------|---|--------------------------------|--|
| USFLUX | Total unsigned vertical flux | Mx | $\sum B_z dA$ |
| MEANGAM | Mean deviation of field from radial | Degrees | $\tan^{-1}(\frac{B_h}{B_z})$ |
| MEANGBT | Mean horizontal gradient of total field | G Mm ⁻¹ | $\sqrt{(\frac{\partial B}{\partial x})^2 + (\frac{\partial B}{\partial y})^2}$ |
| MEANGBZ | Mean horizontal gradient of vertical field | G Mm ⁻¹ | $\sqrt{(\frac{\partial B_z}{\partial x})^2 + (\frac{\partial B_z}{\partial y})^2}$ |
| MEANGBH | Mean horizontal gradient of horizontal field | G Mm ⁻¹ | $\sqrt{(\frac{\partial B_h}{\partial x})^2 + (\frac{\partial B_h}{\partial y})^2}$ |
| MEANJZD | Mean vertical current density, J_z | mA m ⁻² | $J_z = (\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y})$ |
| TOTUSJZ | Total unsigned vertical current | A | $\sum J_z dA$ |
| MEANALP | Characteristic twist parameter, α | Mm ⁻¹ | $\frac{\sum J_z \cdot B_z}{\sum B_z^2}$ |
| MEANJZH | Mean vertical current helicity | G ² m ⁻¹ | $h_c = B_z J_z$ |
| TOTUSJH | Total unsigned vertical current helicity | G ² m ⁻¹ | $\sum h_c $ |
| ABSNJZH | Absolute value of the net vertical current helicity | G ² m ⁻¹ | $ \sum h_c $ |
| SAVNCP | Sum of the absolute value of the net vertical currents per polarity | A | $ \sum^{B_z+} J_z dA + \sum^{B_z-} J_z dA $ |
| MEANPOT | Mean photospheric excess magnetic energy density proxy | erg cm ⁻³ | $\rho = \frac{(B^{\text{Pot}} - B^{\text{Obs}})^2}{8\pi}$ |
| TOTPOT | Total photospheric magnetic free energy density proxy | erg cm ⁻¹ | $\sum \rho dA$ |
| MEANSHR | Mean shear angle | Degrees | $\cos^{-1}(\frac{B^{\text{Pot}} \cdot B^{\text{Obs}}}{B^{\text{Pot}} B^{\text{Obs}}})$ |

^aSummation and normalization over the CMASK active pixels for mean quantities are implied; fundamental constants and constant multiplicative factors are generally not shown in the formulae.

C.1 SHARP Data Segments

SHARPs are provided in standard CCD image coordinates and also remapped and projected onto cylindrical equal area (CEA) coordinates. Table 10 lists all of the data segments provided in `hmi.Sharp_720s` and `hmi.Sharp_720s_nrt`. The 31 quantities include line-of-sight observables as well as results of the VFISV inversion and disambiguation. These data are CCD pixel cut-outs that include the entire HARP bounding box in the original solar-disk pixels. The inversion code provides estimates of uncertainties as well, including χ^2 , the computed standard deviations (σ), and the correlation coefficients of the errors in the derived parameters.

It is often useful to have the data remapped to heliographic coordinates and the vector field projected onto three orthogonal components with uncertainties for each component. The CEA versions of the SHARP data series provide data remapped onto a cylindrical equal-area coordinate system centered on the tracked center of the HARP. For a definitive SHARP the bounding box at every time step encloses the minimum and maximum Carrington latitude and longitude attained by the HARP during its entire lifetime. The heliographic center of the HARP at central meridian passage is uniformly tracked at a nominal differential rotation rate appropriate for that latitude as given in $\text{OMEGA_DT} = 13.6144 - 2.2 \sin^2(\text{lat})$ degrees per day. The NRT series are computed before the full life-cycle of the region is

Table 10 The 31 maps included as SHARP series segments.

| Segment name | Units | Data (CCD coordinates) |
|-----------------------|---------------------|---|
| MAGNETOGRAM | Gauss | Line-of-sight magnetic field (MDI method) |
| BITMAP | Code | Bounding box mask for the HARP patch |
| DOPPLERGRAM | m s^{-1} | Line-of-sight velocity (MDI method) |
| CONTINUUM | DN/s | Continuum intensity (MDI method) |
| INCLINATION | Degrees | Inclination angle of the field to the LoS |
| AZIMUTH | Degrees | Azimuth angle CCW from up on the CCD |
| FIELD | Mx cm^{-2} | Total magnetic field flux density |
| VLOS_MAG | cm s^{-1} | LoS velocity of plasma with field |
| DOP_WIDTH | m \AA | Doppler width |
| ETA_0 | | Source function, η_0 |
| DAMPING | Doppler width units | Damping function |
| SRC_CONTINUUM | DN/s | Continuum intensity |
| SRC_GRAD | DN/s | Gradient |
| ALPHA_MAG | | Field-filling factor, currently 1.0 |
| CHISQ | | χ^2 of the VFISV solution |
| CONV_FLAG | Code | Flag and index of VFISV convergence |
| INFO_MAP | Code | Updated quality map |
| CONFID_MAP | Code | Updated confidence index |
| INCLINATION_ERR | Degrees | Computed standard deviation (σ) of inclination angle |
| AZIMUTH_ERR | Degrees | σ of azimuth |
| FIELD_ERR | Mx cm^{-2} | σ of field |
| VLOS_ERR | cm s^{-1} | σ of V_{LoS} |
| ALPHA_ERR | | σ of alpha |
| FIELD_INCLINATION_ERR | | Correlation of field and inclination errors |
| FIELD_AZ_ERR | | Correlation of field and azimuth errors |
| INCLIN_AZIMUTH_ERR | | Correlation of inclination and azimuth errors |
| FIELD_ALPHA_ERR | | Correlation of field and alpha errors |
| INCLINATION_ALPHA_ERR | | Correlation of inclination and alpha errors |
| AZIMUTH_ALPHA_ERR | | Correlation of azimuth and alpha errors |
| DISAMBIG | Code | Flag for 180° change in disambiguated azimuth |
| CONF_DISAMBIG | Code | Confidence of disambiguation result |

known, so the size and heliographic center of the NRT HARP bounding box will generally change as the region evolves, even if regions do not merge.

Table 11 lists the 11 segments provided in the `hmi.Sharp_cea_720s` and the `hmi.Sharp_cea_720s_nrt` series. The r , θ , and ϕ components of the field are given along with the scalar field, velocity, and continuum intensity. The line-of-sight field in MAGNETOGRAM is remapped to the proper location, but the value is not corrected for projection. The uncertainty reported for each vector field component is the one computed at the CCD pixel in the un-remapped SHARP nearest to the center of the CEA pixel location. Similarly, the BITMAP and CONF_DISAMBIG values refer to the nearest neighbor. Because of the remapping from a sphere to a plane and the conversion to a CEA coordinate system, there is a slight difference between the direction of the magnetic-field vector coordinates

Table 11 The 11 SHARP cylindrical equal area (CEA) data segment maps.

| Name | Units | Data (CEA coordinates) |
|---------------|--------------------|---|
| MAGNETOGRAM | Gauss | Line-of-sight magnetogram (MDI method) |
| BITMAP | Code | Mask for the HARP patch |
| DOPPLERGRAM | m s^{-1} | Line-of-sight velocity (MDI method) |
| CONTINUUM | DN s^{-1} | Continuum intensity (MDI method) |
| BP | Gauss | ϕ Component of the magnetic field, positive westward |
| BT | Gauss | θ Component of the field, positive southward |
| BR | Gauss | Radial component of the field, positive upward |
| BP_ERR | Gauss | Computed σ of B_ϕ |
| BT_ERR | Gauss | Computed σ of B_θ |
| BR_ERR | Gauss | Computed σ of B_r |
| CONF_DISAMBIG | Code | Confidence in disambiguated result |

and the coordinate system unit vectors because the CEA coordinates do not strictly follow standard spherical heliographic directions r , θ , and ϕ except at the center of the window (Sun, 2013).

References

- Alissandrakis, C.E.: 1981, On the computation of constant alpha force-free magnetic field. *Astron. Astrophys.* **100**, 197–200.
- Barnes, G., Leka, K.D.: 2006, Photospheric magnetic field properties of flaring versus flare-quiet active regions. III. Magnetic charge topology models. *Astrophys. J.* **646**, 1303–1318. [DOI](#).
- Barnes, G., Leka, K.D.: 2008, Evaluating the performance of solar flare forecasting methods. *Astrophys. J. Lett.* **688**, L107–L110. [DOI](#).
- Barnes, G., Leka, K.D., Crouch, A.D., Sun, X., Wagner, E.D., Schou, J.: 2014 The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: disambiguation. *Solar Phys.*, in preparation.
- Bobra, M., Sun, X., Hoeksema, J.T., Turmon, M., Liu, Y., Hayashi, K., Barnes, G., Leka, K.D.: 2014, The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: SHARPs – Space-weather HMI Active Region Patches. *Solar Phys.*, in press.
- Borrero, J.M., Kobel, P.: 2011, Inferring the magnetic field vector in the quiet Sun. I. Photon noise and selection criteria. *Astron. Astrophys.* **527**, A29. [DOI](#).
- Borrero, J.M., Tomeczyk, S., Kubo, M., Socas-Navarro, H., Schou, J., Couvidat, S., Bogart, R.: 2011, VFISV: Very Fast Inversion of the Stokes vector for the helioseismic and magnetic imager. *Solar Phys.* **273**, 267–293. [DOI](#).
- Calabretta, M.R., Greisen, E.W.: 2002, Representations of celestial coordinates in FITS. *Astron. Astrophys.* **395**, 1077–1122. [DOI](#).
- Canfield, R.C., de La Beaujardière, J.-F., Fan, Y., Leka, K.D., McClymont, A.N., Metcalf, T., Mickey, D.L., Wülser, J.-P., Lites, B.W.: 1993, The morphology of flare phenomena, magnetic fields, and electric currents in active regions I. Introduction and methods. *Astrophys. J.* **411**, 362–369. [DOI](#).
- Centeno, R., Schou, J., Hayashi, K., Norton, A., Hoeksema, J.T., Liu, Y., Leka, K.D., Barnes, G.: 2013, The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: optimization of the spectral line inversion code. *Solar Phys.* [DOI](#).
- Couvidat, S., Schou, J., Shine, R.A., Bush, R.I., Miles, J.W., Scherrer, P.H., Rairden, R.L.: 2012, Wavelength Dependence of the Helioseismic and Magnetic Imager (HMI) Instrument Onboard the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 285–325. [DOI](#).
- Falconer, D.A., Moore, R.L., Gary, G.A.: 2008, Magnetogram measures of total nonpotentiality for prediction of solar coronal mass ejections from active regions of any degree of magnetic complexity. *Astrophys. J.* **689**, 1433–1442. [DOI](#).
- Gary, G.A., Hagyard, M.J.: 1990, Transformation of vector magnetograms and the problems associated with the effects of perspective and the azimuthal ambiguity. *Solar Phys.* **126**, 21–36.

- Georgoulis, M.K., Rust, D.M.: 2007, Quantitative forecasting of major solar flares. *Astrophys. J. Lett.* **661**, L109–L112. [DOI](#).
- Harvey, J.W.: 1969, Magnetic fields associated with solar active-region prominences. Ph.D. thesis, University of Colorado.
- Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: 1983, Optimization by simulated annealing. *Science* **220**, 671–680. [DOI](#).
- Kosugi, T., Matsuzaki, K., Sakao, T., Shimizu, T., Sone, Y., Tachikawa, S., Hashimoto, T., Minesugi, K., Ohnishi, A., Yamada, T., Tsuneta, S., Hara, H., Ichimoto, K., Suematsu, Y., Shimojo, M., Watanabe, T., Shimada, S., Davis, J.M., Hill, L.D., Owens, J.K., Title, A.M., Culhane, J.L., Harra, L.K., Doschek, G.A., Golub, L.: 2007, The hinode (Solar-B) mission: an overview. *Solar Phys.* **243**, 3–17. [DOI](#).
- LaBonte, B.: 2004a, The imaging vector magnetograph at Haleakala: III. Effects of instrumental scattered light on Stokes spectra. *Solar Phys.* **221**, 191–207. [DOI](#).
- LaBonte, B.: 2004b, The imaging vector magnetograph at Haleakala: III. Effects of instrumental scattered light on Stokes spectra. *Solar Phys.* **221**, 191–207. [DOI](#).
- Leka, K.D., Barnes, G.: 2003a, Photospheric magnetic field properties of flaring versus flare-quiet active regions. I. Data, general approach, and sample results. *Astrophys. J.* **595**, 1277–1295. [DOI](#).
- Leka, K.D., Barnes, G.: 2003b, Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis. *Astrophys. J.* **595**, 1296–1306. [DOI](#).
- Leka, K.D., Barnes, G.: 2007, Photospheric magnetic field properties of flaring versus flare-quiet active regions. IV. A statistically significant sample. *Astrophys. J.* **656**, 1173–1186. [DOI](#).
- Leka, K.D., Barnes, G.: 2012, Modeling and interpreting the effects of spatial resolution on solar magnetic field maps. *Solar Phys.* **277**, 89–118. [DOI](#).
- Leka, K.D., Barnes, G., Crouch, A.D., Metcalf, T.R., Gary, G.A., Jing, J., Liu, Y.: 2009, Resolving the 180° ambiguity in solar vector magnetic fields: evaluating the effects of noise, spatial resolution, and method assumptions. *Solar Phys.* **260**, 83–108. [DOI](#).
- Lites, B.W., Akin, D.L., Card, G., Cruz, T., Duncan, D.W., Edwards, C.G., Elmore, D.F., Hoffmann, C., Katsukawa, Y., Katz, N., Kubo, M., Ichimoto, K., Shimizu, T., Shine, R.A., Streander, K.V., Suematsu, A., Tarbell, T.D., Title, A.M., Tsuneta, S.: 2013, The hinode spectro-polarimeter. *Solar Phys.* **283**, 579–599. [DOI](#).
- Liu, Y., Norton, A.A., Scherrer, P.H.: 2007, A note on saturation seen in the MDI/SOHO magnetograms. *Solar Phys.* **241**, 185–193. [DOI](#).
- Mason, J.P., Hoeksema, J.T.: 2010, Testing automated solar flare forecasting with 13 years of Michelson Doppler imager magnetograms. *Astrophys. J.* **723**, 634–640. [DOI](#).
- Metcalf, T.R.: 1994, Resolving the 180-degree ambiguity in vector magnetic field measurements: the ‘minimum’ energy solution. *Solar Phys.* **155**, 235–242. [DOI](#).
- Metcalf, T.R., Leka, K.D., Barnes, G., Lites, B.W., Georgoulis, M.K., Pevtsov, A.A., Gary, G.A., Jing, J., Balasubramaniam, K.S., Li, J., Liu, Y., Wang, H.N., Abramenko, V., Yurchyshyn, V., Moon, Y.-J.: 2006, An overview of existing algorithms for resolving the 180° ambiguity in vector magnetic fields: quantitative tests with synthetic data. *Solar Phys.* **237**, 267–296. [DOI](#).
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: 1953, Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092. [DOI](#).
- Norton, A.A., Graham, J.P., Ulrich, R.K., Schou, J., Tomczyk, S., Liu, Y., Lites, B.W., López Ariste, A., Bush, R.I., Socas-Navarro, H., Scherrer, P.H.: 2006, Spectral line selection for HMI: a comparison of Fe I 6173 Å and Ni I 6768 Å. *Solar Phys.* **239**, 69–91. [DOI](#).
- Pesnell, W.D., Thompson, B.J., Chamberlin, P.C.: 2012, The Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 3–15. [DOI](#).
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: 1992, *Numerical Recipes in c*, 2nd edn. Cambridge University Press, New York. ISBN 0-521-43108-5.
- Rachkovsky, D.N.: 1962, Magnetic rotation effects in spectral lines. *Izv. Krym. Astrofiz. Obs.* **27**, 148.
- Rachkovsky, D.N.: 1967, The reduction for anomalous dispersion in the theory of absorption line formation in a magnetic field. *Izv. Krym. Astrofiz. Obs.* **37**, 56.
- Scherrer, P.H., Schou, J., Bush, R.I., Kosovichev, A.G., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Zhao, J., Title, A.M., Schrijver, C.J., Tarbell, T.D., Tomczyk, S.: 2012, The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 207–227. [DOI](#).
- Schou, J., Couvidat, S.: 2013, HMI on-orbit calibration analysis. Private communication.
- Schou, J., Borrero, J.M., Norton, A.A., Tomczyk, S., Elmore, D., Card, G.L.: 2012a, Polarization Calibration of the Helioseismic and Magnetic Imager (HMI) onboard the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 327–355. [DOI](#).
- Schou, J., Scherrer, P.H., Bush, R.I., Wachter, R., Couvidat, S., Rabello-Soares, M.C., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Allard, B.A., Miles, J.W., Rairden, R., Shine, R.A., Tarbell,

- T.D., Title, A.M., Wolfson, C.J., Elmore, D.F., Norton, A.A., Tomczyk, S.: 2012b, Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) instrument on the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 229–259. [DOI](#).
- Schrijver, C.J.: 2007, A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *Astrophys. J. Lett.* **655**, L117–L120. [DOI](#).
- Sun, X. (HMI Team): 2013, On the coordinate system of Space-weather HMI Active Region Patches (SHARPs): A technical note. http://jsoc.stanford.edu/jsocwiki/sharp_coord.arXiv.
- Thompson, W.T.: 2006, Coordinate systems for solar image data. *Astron. Astrophys.* **449**, 791–803. [DOI](#).
- Tsuneta, S., Ichimoto, K., Katsukawa, Y., Nagata, S., Otsubo, M., Shimizu, T., Suematsu, Y., Nakagiri, M., Noguchi, M., Tarbell, T., Title, A., Shine, R., Rosenberg, W., Hoffmann, C., Jurcevich, B., Kushner, G., Levay, M., Lites, B., Elmore, D., Matsushita, T., Kawaguchi, N., Saito, H., Mikami, I., Hill, L.D., Owens, J.K.: 2008, The solar optical telescope for the hinode mission: an overview. *Solar Phys.* **249**, 167–196. [DOI](#).
- Turmon, M., Pap, J.M., Mukhtar, S.: 2002, Statistical pattern recognition for labeling solar active regions: application to SOHO/MDI imagery. *Astrophys. J.* **568**, 396–407. [DOI](#).
- Turmon, M., Jones, H.P., Malanushenko, O.V., Pap, J.M.: 2010, Statistical feature recognition for multidimensional solar imagery. *Solar Phys.* **262**, 277–298. [DOI](#).
- Turmon, M., Hoeksema, J.T., Sun, X., Bobra, M., Sommers, J.: 2014, The Helioseismic and Magnetic Imager (HMI) magnetic field pipeline: HARPs – HMI Active Region Patches. *Solar Phys.*, in preparation.
- Unno, W.: 1956, Line formation of a normal zeeman triplet. *Publ. Astron. Soc. Japan* **8**, 108.
- Wachter, R., Schou, J.: 2009, Inferring small-scale flatfields from solar rotation. *Solar Phys.* **258**, 331–341. [DOI](#).
- Wachter, R., Schou, J., Rabello-Soares, M.C., Miles, J.W., Duvall, T.L., Bush, R.I.: 2012, Image Quality of the Helioseismic and Magnetic Imager (HMI) onboard the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 261–284. [DOI](#).
- Welsch, B.T., Fisher, G.H., Sun, X.: 2013, A magnetic calibration of photospheric Doppler velocities. *Astrophys. J.* **765**, 98. [DOI](#).
- Westendorp Plaza, C., del Toro Iniesta, J.C., Ruiz Cobo, B., Martinez Pillet, V., Lites, B.W., Skumanich, A.: 1998, Optical tomography of a sunspot. I. Comparison between two inversion techniques. *Astrophys. J.* **494**, 453. [DOI](#).