Introduction to Data Usage and Error Analysis for HMI Magnetic Field Data

Dr. K. D. Leka

October 2017

1 Introduction

This is a brief document outlining how to *begin* to estimate the uncertainties in magnetic field data from the *Solar Dynamics Observatory*'s Helioseismic and Magnetic Imager. It does not include many sources of uncertainty, and thus what is retrieved should be considered a *lower bound*.

There are two aspects covered here: (1) how to determine which pixels have well-measured estimates *vs.* which are dominated by noise (and thus should be ignored or processed in a special manner), (2) for well-measured pixels, how to calculate an estimate of the uncertainty in the measurement and quantities derived from it.

2 Sources of Uncertainty

Uncertainty or error arises from many sources, starting from the statistical fluctuations of observing photons, through instrumental impacts, to the influence of the data reduction algorithms. For HMI the primary sources of errors include:

- Limited spatial resolution: blurs/mixes true signal [see 6].
- Limited temporal sampling: underlying evolution blurs / mixes signals; signals impacted by solar oscillations.
- Limited exposure: decreases photon signal/noise ratio.
- Limited spectral sampling (across the line) and spectral resolution: decreases information available for inversion.
- Limited spectral coverage (single line): limited information available for inversion [13].
- Incomplete characterization and mitigation of instrumental & optical elements: introduces bias to signals, sometimes of unknown character / magnitude [10, 3, 11].
- Spacecraft motion: imparts bias into signals [12].
- Milne-Eddington Inversion: assumes constant field vector and thermodynamics across the formation height of the spectral line (and other assumptions) which is a strong simplifying assumption [9, 15].
- HMI implementation of ME Inversion: additionally assumes unity magnetic fill fraction; all pixels independent [2, 5].
- Disambiguation of the 180° ambiguity in $B_\perp$: optimizes an imposed functional over strong-field areas, different scheme (3 options) for weak-field areas [7, 5].
- Coordinate transforms and "deprojection" can make simplifying assumptions that impart a bias or error in the result [4, 1].

Many algorithms are designed and implemented to minimize or at least characterize the impacts of the above, but many are simply taken "as is" (for example: the assumption that the "magnetogram" estimates the field on a planar surface, where in fact it follows the curved surface of the optical depth, τ).

3 Estimating Uncertainty

"There is no measurement that is valid without an estimate of its uncertainty."

While it is almost impossible to account for every possible source of error, tools are in place to provide estimates of the largest impacts. This enables the researcher to not only provide an estimate of the uncertainty in their derived quantities, but to avoid including data where the signal is dominated by noise – thus avoiding spurious measurements.

For HMI data, the largest sources of the uncertainty in the magnetic field vector arise from assumptions included in the inversion, limited spatial resolution, the fits to the model within the inversion, the impacts of spacecraft orbital velocity, and the disambiguation.

3.1 Relevant Approximations

To understand some of the impacts, it is helpful to keep these "back-of-the-envelope" approximations in mind:

- Poisson statistics; noise $\propto \sqrt{n_\gamma}$.
- The polarization signal for the line-of-sight component $\mathrm{B}_\parallel \propto d\mathrm{I}/d\lambda$. The polarization signal for the component of the field in the plane-of-the-sky $\mathrm{B}_\perp \propto d^2\mathrm{I}/d\lambda^2$.
- Solar plasma velocities are *generally* well-matched to the 12-minute cadence, but in emerging flux regions and dynamic areas near the solar disk, there will be signal smearing across pixels.

These mean that better signals are obtained in bright areas (more photons), but this isn't always where the polarization (hence the magnetic field) is strongest. Where it is strongest (umbrae, for example), there are fewer photons and the spectral line can be more poorly sampled, leading to a worse fit by the inversion. In very bright areas, the spectral line is well characterized, but the polarization signals will be low due to: low field strength and most likely small fill fraction, leading again to a worse fit in the inversion (and higher uncertainty in the estimation of the field).

Both the $\mathrm{B}_\parallel$ (B_{los}) and $\mathrm{B}_\perp$ (B_{trans}) components, including the plane-of-sky azimuthal angle ϕ are used to derive the magnetic field vector. From above, the noise in $\mathrm{B}_\perp$ can be an order of magnitude larger than the noise in $\mathrm{B}_\parallel$. Propagation of the uncertainty in both components, and ϕ, is needed to estimate the uncertainty in the full vector or any of its physically-relevant components (*e.g.*, B_r, the normal component). In that context, because more $\mathrm{B}_\perp$ goes in to B_r as the observing angle goes away from disk center, the uncertainty in B_r increases toward the limb.

3.2 Resources

Generally speaking, the largest components of the uncertainty in the HMI magnetic field estimates are the impact of the spacecraft orbit, the inversion, and the disambiguation. The first of these is very difficult to quantify for any given magnetogram, although it is possible to estimate the variation of the field components as a function of time [5, 12].

- *KEYWORDS* and what they mean, *e.g.* `QUALITY`, `OBS_VR`
- *Auxiliary files:*
  - ⊙ In `hmi.sharp_cea_720s` (Table1, 2):
  - ⊙ In `hmi.sharp_720s` (*not* CEA:
  - ⊙ Note there is no `magnetogram_err.fits or Dopplergram_err.fits`, although there is `Br_err.fits` in CEA and the components to propagate the errors in `hmi.sharp_720s`, or `chisq.fits` in `hmi.sharp_cea_720s`:
    Note that there is more information regarding the processing results in this `hmi.sharp_720s` series, but the data are in the plane-of-sky coordinates.
- The noise masks as f(`obs_vr`) can be computed using the Chebychef coefficients in the `hmi.lookup_ChebyCoef_BNoise` series; probably not needed for daily use - they are used

Table 1: `hmi.sharp_cea_720s` segments (fits files) including relevant errors / uncertainty files

`Br_err.fits`	uncertainty from ME inversion propagated for radial component
`Bp_err.fits`	uncertainty from ME inversion propagated for poloidal component (north/south)
`Bt_err.fits`	uncertainty from ME inversion propagated for toroidal component (east/west)
`bitmap.fits`	map of "active pixels" of the HARP. 0: offlimb, 1: quiet pixel outside HARP, 2: active pixel outside HARP, 33: quiet pixel inside HARP, 34: active pixel inside HARP
`conf_disambig.fits`	90: pixels were annealed; 60: annealing buffer, uses smoothed nearest neighbor; 50: using a radial-field solution (CEA; for non-CEA and full disk, option are included for potential and random solutions); 0: not disambiguated (limb).

Table 2: `hmi.sharp_720s` segments (fits files) including some relevant errors / uncertainty files

`field_err.fits, azimuth_err.fits`	uncertainty from ME inversion
`inclination_err.fits, vlos_err.fits`	" " " "
`field_inclination_err.fits`	cross-term errors
`field_az_err.fits`,	" " " "
`inclination_azimuth_err.fits`	" " " "
`info_map.fits`	quality map (see [table A.6 in ref. 5]
`confid_map.fits`	confidence index of inversion output, *ibid.*
`conv_map.fits`	flag/index of ME convergence process, *ibid.*
`chisq.fits`	χ^2 value of overall ME-inversion fit
`bitmap.fits, conf_disambig.fits`	see above

to determine the disambiguation thresholds, so are reflected in `conf_disambig.fits`.

•Documentation websites:

⊙ `http://jsoc.stanford.edu/jsocwiki/ReleaseNotes`
(specifically for a very early release of AR11158 data, there are helpful exlanations and links available).

⊙ `http://jsoc.stanford.edu/doc/data/hmi/sharp/sharp.htm`.
I suggest read thoroughly, all the way through.

•The HMI documentation papers, especially tables of keyword explanations, *etc.*.
See references [10, 11, 5, 3, 2, 1].

3.3 Propagating Errors

Assuming independent variables (neglecting cross-terms, correlations in the errors) and assuming that all of the error measurements of x ("σ_x") are reasonably small, gives the basic formula for error propagation of a function $f(x, y, z...)$:

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 \sigma_z^2 + ...} \quad (1)$$

The cross-terms needed for a more rigorous error estimation are provided for the `hmi.sharp_720s` series, however for most of the purposes of most research, using the above and neglecting the covariance terms should be acceptable.

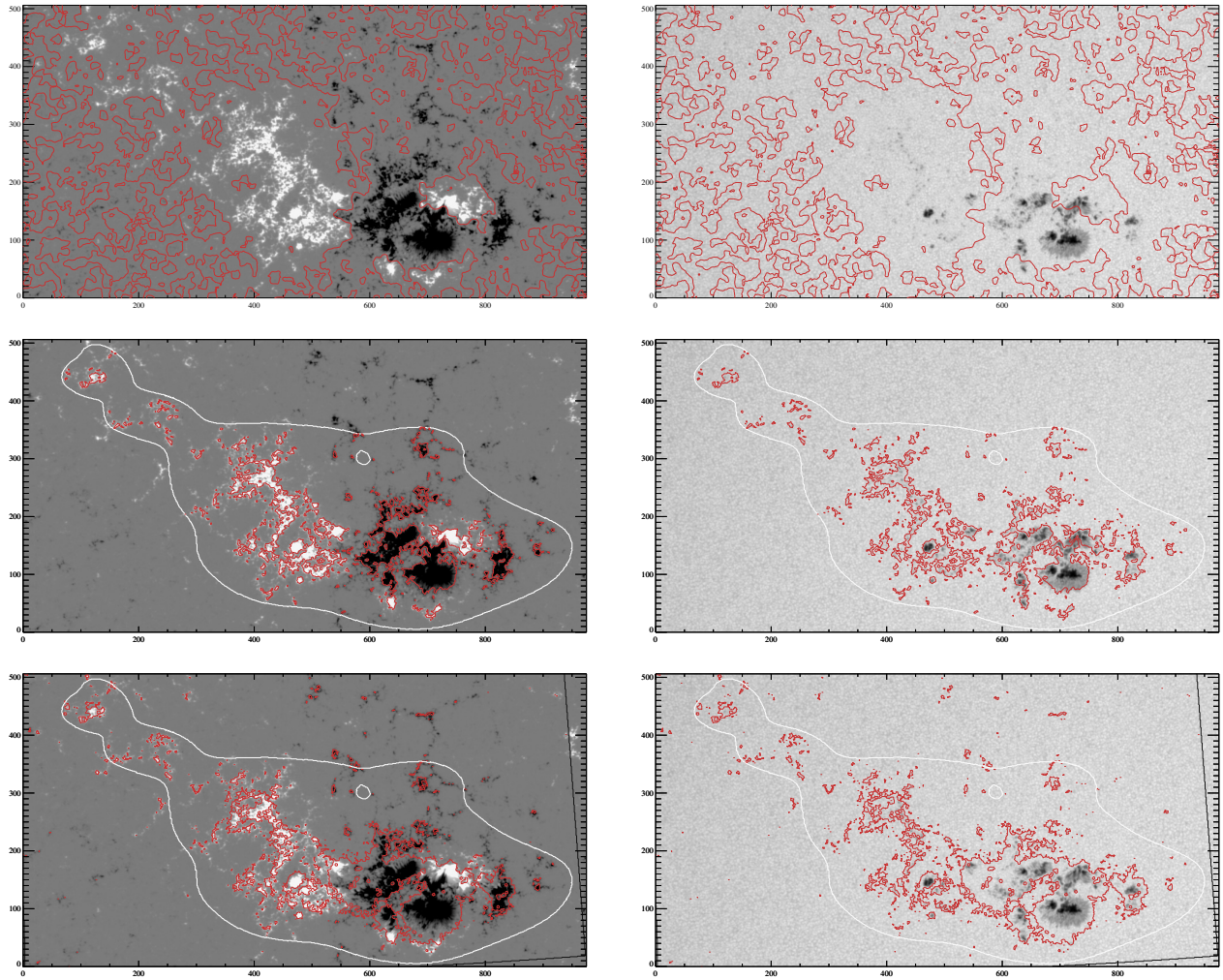

Figure 1: Images from HARP 833 (NOAA AR#11283) at 2011.09.05_00.48.00 from the `hmi.sharp_cea_720s` series. **Left:** magnetogram.fits, **Right:** continuum.fits. Contours are **Top:** a very smoothed polarity inversion line, **Middle:** the "active region pixels" maps, at contour levels 32 (showing full extent of HMI "blob") and 34 (indicating only "active pixels" – see Table1), **Bottom:** disambiguation confidence map at level=90, which are pixels that are annealed. the "blob" is shown, too (Table1). Note the subtle differences between strong-field "active" pixels (middle) and pixels that were annealed for the disambiguation (bottom).

When applying the above to formulae that include angles, it is helpful to perform Taylor expansions or similar simplifications. The Wikipedia Page `https://en.wikipedia.org/wiki/Propagation_of_uncertainty` includes some useful tips.

As an example, applying the above to computing the Total Magnetic Flux

$$\Phi = \int |B_r| d\mathrm{A} \tag{2}$$

we assume that the integral can be approximated as a summation, and that there is no uncertainty in the area in each pixel, $d\mathrm{A}$. Then,

$$\Phi = \mathrm{Br}_1 d\mathrm{A}_1 + \mathrm{Br}_2 d\mathrm{A}_2 + \mathrm{Br}_3 d\mathrm{A}_3 + ... \tag{3}$$

$$= \sum_{n=0}^{N} |\mathrm{Br}_n| d\mathrm{A}_n \tag{4}$$

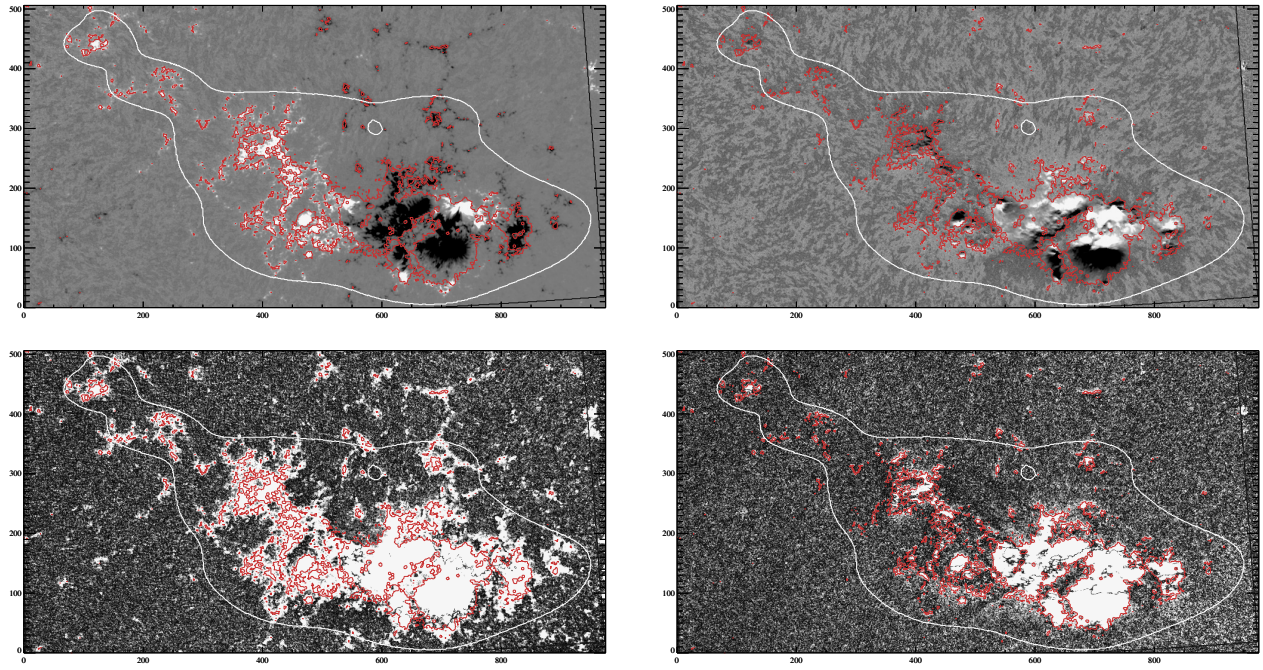

Figure 2: Same as Figure 1 but now **Top:** showing **Left** the radial component `br.fits`, **Right** the toroidal component `bt.fits` (see Table 2); both also include the "blob" and the disambiguation confidence=90 contour. **Bottom**: the first use of the uncertainty files. **Left:** image of the signal-to-noise ratio of the radial component $\mathcal{S}/\mathcal{N} = |\mathrm{Br}|/\mathrm{Br\_err}$, scaled [0, 3], **Right:** image of the signal-to-noise ratio of the toroidal component $\mathcal{S}/\mathcal{N} = |\mathrm{Bt}|/\mathrm{Bt\_err}$ with the same scaling. KDL advocates for generally using only pixels with the relevant component having $\mathcal{S}/\mathcal{N} > 3$ (at least 2).

---

where the subscript denotes the pixel number. With $\partial\Phi/\partial \mathrm{Br}_1 = d\mathrm{A}_1$ *etc.*, and the "uncertainty" in each pixels' B_r measurement from `Br_err.fits` and other sources we have simply:

$$\sigma_\Phi = \left(d\mathrm{A}_1^2 \sigma_1{}^2_{\mathrm{Br}} + d\mathrm{A}_2^2 \sigma_2{}^2_{\mathrm{Br}} + d\mathrm{A}_3^2 \sigma_3{}^2_{\mathrm{Br}} + ...\right)^{1/2} \quad (5)$$

$$= \left(\sum_{n=0}^{N} d\mathrm{A}_n^2 \sigma_n{}^2_{\mathrm{Br}}\right)^{1/2} \quad (6)$$

3.4 Disambiguation Uncertainty

One step of data reduction is to solve the inherent 180° ambiguity in the direction of $\mathrm{B}_\perp$. This is an ambiguity in the *plane of the sky* (like the `hmi.sharp_720s` series), not the heliographic components (like the `hmi.sharp_cea_720s` series) – it is solved using physics of the latter, but the solution is in the $\mathrm{B}_\perp$ component of the former. The solution can change the apparently magnetic polarity of an area (Figure 3) and thus relocate magnetic polarity inversion lines.

In the HMI pipeline, the "Minimum Energy" algorithm is used [8, 7], where the "cooling" of the simulated annealing (the optimization method used) is rapid enough (a solution is "found" rapid enough) to keep up with the pipeline data. This works well for most areas within most strong-$\mathrm{B}_\perp$ regions. The areas with `conf_disambig`= 90 have been annealed, the areas with `conf_disambig`= 60 are a buffer which was part of the annealing and is then solved with a nearest-neighbor smoothing.

There is no further reporting of uncertainty for the disambiguation. It is difficult with angles to assign a "10%" error, however that can be assigned to the image-plane components prior to the coordinate transform to (*e.g.*) CEA coordinate system. This error is *not* presently included,

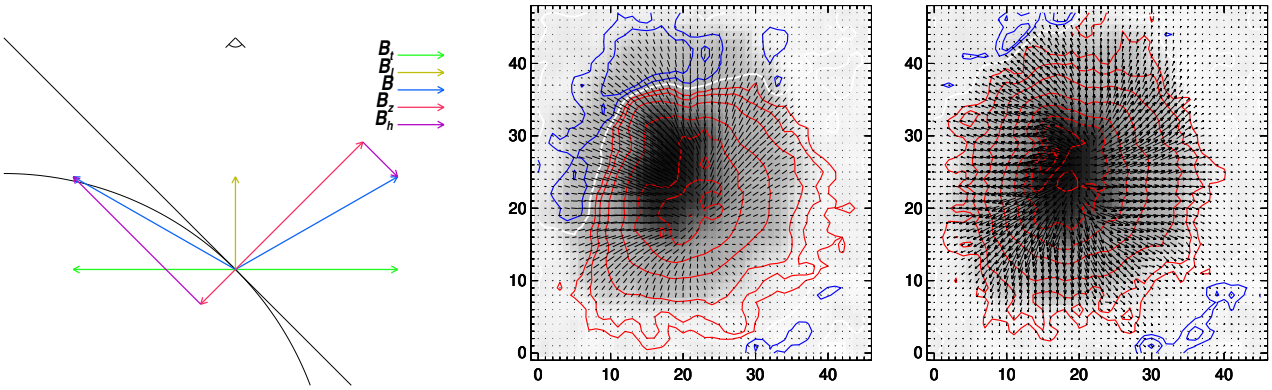

Figure 3: Disambiguation for beginners: **Left:** curved surface (solar surface), viewing angle (eyeball) at $+y$, observed is $B_{\parallel} \equiv B\text{los}$ (yellow) and $B_{\perp} \equiv B_{\text{trans}}$ (green), with two possible directions. The resulting heliographic components B_h (purple) and B_z (red) along with the full vector $\boldsymbol{B}$ (blue) for the two possible solutions are then shown, and as indicated by the red B_z arrows, they result in different magnetic polarities. **Center:** a sunspot and the plane-of-sky components B_ℓ (red/blue for positive/negative polarity) and $B_{\perp}$ line segments (no directionality implied), white contour is the implied B_ℓ polarity inversion line. (**Right:**) heliographic components B_z red/blue for positive/negative polarity), the B_h vectors, and the true polarity inversion line.

---

however, in the uncertainties reported in the `hmi.sharp_cea_720s` files (*e.g.* `Br_err.fits`, `Bp_err.fits`).

Custom disambiguation (with slower cooling, additional visits and 10 or 100 random number seeds) can be run in order to: disambiguate weaker-field areas, disambiguate a very difficult area, to assign more realistic errors to the disambiguation results. The results must then be transformed to heliographic components by the user. If you are interested in any of these options, please contact me.

Of note, in the pipeline, each time is disambiguated independently from the other times, occasionally resulting in areas "flip-flopping" between solutions [see Fig. 15 in ref. 5]. This can be caused by strong plasma dynamics which also produce non-classical Stokes profiles, thus a high error in the inversion results (examine the χ^2!), and a "garbage in, garbage out" situation for the disambiguation. NWRA is working on time-series disambiguation; a selection of popular HARPs will be available late 2017 / early 2018 with the disambiguation algorithm optimized to include temporal behavior, as well. Again, contact me if you are interested in this product and how to use it.

3.5 Strategies for Challenging Estimations

3.5.1 Orbital velocity impacts (on field strength, angles, etc.):

As SDO moves in its orbit, the target spectral line shifts and (due to reasons we do not yet fully understand), the resulting inverted magnetic field varies, too [see 5]. It is not a linear-function variation and somewhat difficult to model well [see 12]. I suggest *against* trying to *remove* the orbital variations. Instead I suggest quantifying the variations in nearby quiet- and sunspot-areas, where evolution is expected to be minimal, and adding that percentage error to your error propagation (see § 4, below).

3.5.2 Orbital velocity impacts (on noise levels:)

Noise levels vary as a function of the orbital velocity [5]. The noise masks that characterize this variation can be reconstructed from Chebyshev coefficients (`hmi.lookup_ChebyCoef_BNoise` series), but in practice - know that this varying noise mask and noise level is used to construct the confidence maps, so by using those, you are by default including the varying noise levels in some manner.

3.5.3 Distortions from CEA transform:

The cylindrical equal area transform does not preserve angles perfectly. There is a distortion near the edges of the field of the horizontal field components. It is at most a few degrees [see 1] but can influence, for example, the estimation of shear or vertical current in a sunspot near the edge of a CEA SHARP cut-out. I suggest – especially for very large SHARPS – comparing the azimuthal angles of the horizontal field between the `hmi.sharp_cea_720s` and `hmi.sharp_720s` series, transformed using the heliospherical transforms [14, 4]. I don't suggest "correcting" the data, but adding an estimate of the error into the error propagation.

4 Example

With these tools and descriptions, let us examine how to best estimate the evolution of the total magnetic flux of an active region over the course of a time t, using the HARP CEA cut-outs:

1. Obtain all of the `hmi.sharp_cea_720s` segments over the course of time t.
2. Identify those pixels in the `Br` segment which
   (a) are within the "blob" of active pixels,
   (b) either have `conf_disambig` > 60 or `bitmap`$= 34$ or where $|B_r|/\sigma(B_r) > N$ (I like $N \geq 3$, but at the minimum, $N = 2$)
3. Create a mask or other selection tool using these criteria, and perform your analysis only using these pixels.
4. Estimate root-mean-squared variation of $B_r(t)$ as $f(|Br|)$, for example "for $|B_r| > 1000$G there is a 10% variation in a pixel's field strength over (t)" or similar estimation. Add that amount to the values for $\sigma(B_r)$.
5. Calculate $\Phi(t), \sigma_\Phi(t)$.
   (a) The equation above will simplify somewhat because the area of each pixel, in Mm^2, is the same. This would not be the case for the `hmi.sharp_720s` series data.
6. Fit $\Phi(t)$ *and* $\sigma_\Phi(t)$ with *no higher an order of function* than physical models, the data, and common sense allow. Generally, use the lowest order function which will fit the majority of the data including the error bars (Figure 4).
7. Quote the fit *and the uncertainty in the fit* as your answer.

For other variables, such as $J_z = 1/\mu_0 \nabla \times B_h$ or the total magnetic shear, other terms enter into the error propagation, and the calculations may become algebraically cumbersome. I suggest breaking things up into terms, and simplify as you proceed.

For looking at correlations, again use both the values and each point's uncertainty when fitting for the correlations, or use the uncertainties to perform simple Monte Carlo estimates of the variation in the correlation coefficients.

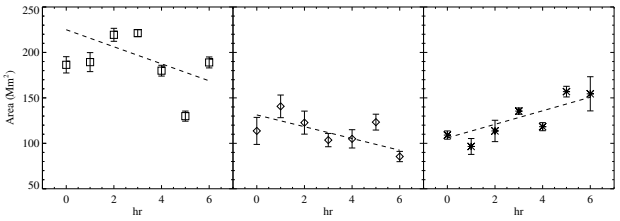

Figure 4: Variation over three separate 6 hr periods of the area of strong magnetic shear in the vicinity of the magnetic neutral line. Error bars are shown. The data do not require a higher-order function than linear, especially over this short a time-period and this large a cadence.

5 *Caveat Emptor*

These notes are meant to guide, not be a final answer to anything. I take responsibility for any errors herein and will correct them as best I become aware of them. There are a large number of error sources not discussed and as such, it is the duty of the user to determine whether the data and the uncertainty in the data values are adequate for the scientific question being asked.

Dr. Leka welcomes comments, concerns, and corrections, and can be contacted:

Senior Research Scientist
NorthWest Research Associates
3380 Mitchell Lane
Boulder CO 80301 USA
leka@nwra.com
+1 303 415 9701x219
`www.nwra.com`

Designated Professor
Nagoya University / ISEE
Furo-cho Chikusa-ku
Nagoya, Aichi 464-9601 JAPAN
kdleka@isee.nagoya-u.ac.jp
+81 52 747 6349
`www.isee.nagoya-u.ac.jp`

References

[1] M. G. Bobra, X. Sun, J. T. Hoeksema, M. Turmon, Y. Liu, K. Hayashi, G. Barnes, and K. D. Leka. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs - Space-Weather HMI Active Region Patches. *Sol. Phys.*, 289:3549–3578, September 2014. doi: 10.1007/s11207-014-0529-3.

[2] R. Centeno, J. Schou, K. Hayashi, A. Norton, J. T. Hoeksema, Y. Liu, K. D. Leka, and G. Barnes. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: Optimization of the Spectral Line Inversion Code. *Sol. Phys.*, 289:3531–3547, September 2014. doi: 10.1007/s11207-014-0497-7.

[3] S. Couvidat, J. Schou, R. A. Shine, R. I. Bush, J. W. Miles, P. H. Scherrer, and R. L. Rairden. Wavelength Dependence of the Helioseismic and Magnetic Imager (HMI) Instrument onboard the Solar Dynamics Observatory (SDO). *Sol. Phys.*, 275:285–325, January 2012. doi: 10.1007/s11207-011-9723-8.

[4] G. A. Gary and M. J. Hagyard. Transformation of Vector Magnetograms and the Problems Associated with the Effects of Perspective and the Azimuthal Ambiguity. *Sol. Phys.*, 126: 21–36, 1990.

[5] J. T. Hoeksema, Y. Liu, K. Hayashi, X. Sun, J. Schou, S. Couvidat, A. Norton, M. Bobra, R. Centeno, K. D. Leka, G. Barnes, and M. Turmon. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: Overview and Performance. *Sol. Phys.*, 289:3483–3530, September 2014. doi: 10.1007/s11207-014-0516-8.

[6] K. D. Leka and G. Barnes. Modeling and Interpreting the Effects of Spatial Resolution on Solar Magnetic Field Maps. *Sol. Phys.*, 277:89–118, March 2012. doi: 10.1007/s11207-011-9821-7.

[7] K. D. Leka, G. Barnes, and A. Crouch. An Automated Ambiguity-Resolution Code for Hinode/SP Vector Magnetic Field Data. In B. Lites, M. Cheung, T. Magara, J. Mariska, & K. Reeves, editor, *The Second Hinode Science Meeting: Beyond Discovery-Toward Understanding*, volume 415 of *Astronomical Society of the Pacific Conference Series*, pages 365–+, December 2009.

[8] T. R. Metcalf. Resolving the 180-degree Ambiguity in Vector Magnetic Field Measurements: The 'Minimum' Energy Solution. *Sol. Phys.*, 155:235–242, 1994. doi: 10.1007/BF00680593.

[9] D. Mihalas. *Stellar Atmospheres*. W. H. Freeman & Co., New York, 1978.

[10] J. Schou, J. M. Borrero, A. A. Norton, S. Tomczyk, D. Elmore, and G. L. Card. Polarization Calibration of the Helioseismic and Magnetic Imager (HMI) onborad the Solar Dynamics Observatory (SDO). *Sol. Phys.*, pages 177–+, October 2010. doi: 10.1007/s11207-010-9639-8.

[11] J. Schou, P. H. Scherrer, R. I. Bush, R. Wachter, S. Couvidat, M. C. Rabello-Soares, R. S. Bogart, J. T. Hoeksema, Y. Liu, T. L. Duvall, D. J. Akin, B. A. Allard, J. W. Miles, R. Rairden, R. A. Shine, T. D. Tarbell, A. M. Title, C. J. Wolfson, D. F. Elmore, A. A. Norton, and S. Tomczyk. Design and Ground Calibration of the Helioseismic and Magnetic Imager (HMI) Instrument on the Solar Dynamics Observatory (SDO). *Sol. Phys.*, 275:229–259, January 2012. doi: 10.1007/s11207-011-9842-2.

[12] P. W. Schuck, S. K. Antiochos, K. D. Leka, and G. Barnes. Achieving Consistent Doppler Measurements from SDO/HMI Vector Field Inversions. *ApJ*, 823:101, June 2016. doi: 10.3847/0004-637X/823/2/101.

[13] A. Skumanich and B. W. Lites. Stokes Profile Analysis and Vector Magnetic Fields. I - Inversion of Photospheric Lines. *ApJ*, 322:473–482, November 1987. doi: 10.1086/165743.

[14] W. T. Thompson. Coordinate Systems for Solar Image Data. *A&A*, 449:791–803, April 2006. doi: 10.1051/0004-6361:20054262.

[15] C. Westendorp Plaza, J. C. del Toro Iniesta, B. Ruiz Cobo, V. Martinez Pillet, B. W. Lites, and A. Skumanich. Optical Tomography of a Sunspot. I. Comparison Between Two Inversion Techniques. *ApJ*, 494:453–471, 1998.