5

Inferring Currents from the Zeeman Effect at the Solar Surface

Graham Barnes and K. D. Leka

ABSTRACT

The brightness and morphology of coronal structures and the sharply bound morphology of photospheric magnetic structures indicate that electric currents permeate the solar atmosphere. Quantifying the strength of these currents is, however, difficult. In this introduction to measuring electric currents in the atmosphere of the Sun, we present an overview of the present methods for estimating them near the solar surface, including a brief overview of the steps required to infer a photospheric current density from remotely sensed observations of polarized light. We focus on results from the Zeeman effect at the photosphere, and highlight the limitations of spatial resolution and the challenges of sampling the solar atmosphere at multiple heights, both of particular interest for understanding the true nature of the solar electric currents.

5.1. INTRODUCTION

The solar atmosphere is threaded with magnetic fields and, by extension, electric current systems. Many of the ubiquitous features of the Sun, from the confined character of the magnetic structures to the energy storage and release required for the energetic explosive events, point to the existence of solar electric currents.

However, electric currents in the solar atmosphere are not measured directly: “What we actually measure are electrons on the CCD” [J. Schou, private communication]. Slightly less literally, what are measured are photons from the Sun that have been influenced by the magnetized plasma in which they originate. Multiple additional steps are needed just to estimate the magnetic field vector, before any component of the electric current density can be inferred.

In the high-density plasma close to the Sun’s surface in the photosphere, the most commonly measured effects are the Zeeman and Hanle effects; the former impacts energy-level splitting and the polarization state in absorption lines and is most relevant for the stronger fields and higher densities of the solar photosphere, while the Hanle depolarization of scattered light is the primary mechanism by which magnetic fields in the solar chromosphere are now being investigated, where the field strength and density are both lower on average than at the photosphere (see details and reviews in del Toro Iniesta and Ruiz Cobo [2016]; Stenflo [1994]). Readers interested in advances in solar polarimetric observing and analysis techniques are directed to the proceedings of the solar polarization meetings series, published by the Astronomical Society of the Pacific [e.g., Kuhn et al., 2011].

For a typical absorption line sampling the Zeeman effect in the photosphere, the height of formation covers a relatively small range compared to the scale height of the field, so the field is considered to be measured at a single height, and local variations in this height are generally ignored [Faurobert et al., 2009, 2012]. In the low density of the corona, the Sun’s outer atmosphere, thermal bremsstrahlung and thermal gyroresonance emission are observed in radio wavelengths, and can be used to determine a line-of-sight integrated value of the magnetic field [Miyawaki et al., 2016; White and Kundu, 1997]. The direct analysis of Stokes polarization signals
from coronal forbidden lines is also now being used to infer coronal magnetic structures [Dalmasse et al., 2016; Dove et al., 2011; Judge and Casini, 2001], although the analysis is challenging due to weak fields, low signal to noise, and again the corona is optically thin [Dima et al., 2016; Kramar et al., 2013, 2016; Merenda et al., 2006; Schad et al., 2016]. The line-of-sight integral for coronal emission in both wavelength regimes means that these observations are generally not useful for determining the current because the lack of information about the spatial variation of the field precludes applying Ampère’s law. Still, modeling efforts have demonstrated the possibility of differentiating between current-filled and current-free coronal streamer-type structures when Zeeman and Hanle diagnostics are combined [Judge et al., 2006].

Between the corona and the photosphere, the chromospheric atmosphere can be optically thin or thick depending on the structure being observed and where in the spectral line (wings versus core) the observation is made, providing an expectedly nonplanar surface. This review focuses on the electric current as inferred using the Zeeman effect primarily in the photosphere. Many of the same principles may apply for the solar chromosphere/corona magnetic field as inferred from the Hanle effect, although the geometrical considerations pose additional challenges in that regime.

Figure 5.1 shows photospheric images of a group of sunspots as seen in continuum light and with the line-of-sight component of the magnetic field; three areas of particular solar structure are indicated: quiet Sun (where the field is weak), a sunspot umbra (where the field is strong), and penumbra (where the field is both strong and highly inclined). Figure 5.2 shows an example of the polarization signals in the common Stokes formalism observed by the Hinode Solar Optical Telescope/SpectroPolarimeter (SOT/SP) [Kosugi et al., 2007; Lites et al., 2013; Tsuneta et al., 2008] in these structures. In the quiet Sun, two magnetically sensitive spectral absorption lines are clearly evident in the unpolarized component, while the polarized states are dominated by noise. In the umbra and penumbra, there is a clear signal present in both polarized and unpolarized states.

**Figure 5.1** The photospheric surface of the Sun as seen on 15 February 2010 at 10:12 TAI: (left) The line-of-sight component of the magnetic field, with the field of view of SOT/SP encompassing NOAA Active Region 11158 outlined by the blue box; (right) NOAA AR 11158 as observed by SOT/SP in continuum light (top) and the line-of-sight component of the magnetic field (bottom) inverted using the MERLIN inversion and available online (http://sot.lmsal.com/data/sot/level2d/). A quiet Sun point is indicated with a blue diamond, a plage point is labeled with a green square, and an umbral point is labeled with a red star. Spectra from these locations are shown in Figure 5.2.
To leading order, a magnetic field can be considered to influence the polarized states through the Zeeman effect according to

\[
B_{\parallel} \propto V, \quad (5.1)
\]

\[
B_{\perp} \propto \left( Q^2 + U^2 \right)^{1/2}, \quad (5.2)
\]

\[
\phi \approx \tan^{-1}(U/Q), \quad (5.3)
\]

where \( B_{\parallel} \) is the component of the field directed along the line of sight, \( B_{\perp} \) is the magnitude of the field transverse to the line of sight, and \( \phi \) is the direction of the transverse component of the field, while \( Q, U, V \) are the Stokes parameters with \( Q \) and \( U \) being the linear polarization with orientation differing by 45°, and \( V \) being the circular polarization.

With the initial observations of the linear polarization component of magnetically sensitive spectral lines and its interpretation as the transverse component of the solar magnetic field, the existence of electric current systems in the solar atmosphere was inferred [Rayrole and Semel, 1970; Severnyi, 1965]. Indeed it was quickly recognized that the observed sharp (at 6") gradients in field strength, especially near the inferred magnetic polarity inversion lines, could be indications of strong current systems. It was also quickly realized that chromospheric fibril structures reflected the nonpotential current-carrying character of the photospheric magnetic field [Hale, 1927; Nakagawa et al., 1971]. As spatial resolution, polarimetric sensitivity, and data reduction algorithms improved, including algorithms for implementing coordinate transforms to fully consider the heliographic magnetic components [Canfield et al., 1993; Gary and Haggard, 1990], the understanding of the crucial role that electric current systems play in solar physics increased as well. This understanding spans the role of currents as energy storage and possible trigger
mechanisms for solar energetic events, as indications of helicity transfer, and in the organization of solar structures from the smallest penumbral fibrils to the largest coronal loops. Even with the recent advances in observational and interpretive capability, there are many limitations to these measurements; we will highlight here only some of the effects of limited spatial resolution.

5.2. FROM PHOTONS TO CURRENT

There are four main steps needed to infer the (vertical) current density at a given location on the solar surface: (1) Instrumental calibration, (2) inversion, (3) disambiguation, and (4) Ampère’s law. Calibration is extremely important, but is instrument-specific and so will not be covered in this review. Common approaches to methods of inversion and disambiguation are outlined below, including the possibility of additional information becoming available from more sophisticated methods that are presently not routinely employed. The various ways to compute the current from Ampère’s law are then discussed.

5.2.1. Inversion

Magnetic fields on the Sun are inferred from the field strength and direction needed to reproduce the observed polarization signals. As alluded to above, very rough approximations can relate the two directly. However, in the context of inferring electric currents, these approximations and other approaches that do not rely on good spectral information can be misleading. The magneto-optical rotation of the plane of linear polarization that occurs when the light passes through a magnetized medium can introduce apparent vertical current systems where there are none. Hence, a full inversion of the polarized spectra of magnetically sensitive spectral lines, which accounts for these effects, is required [del Toro Iniesta, 2003; del Toro Iniesta and Ruiz Cobo, 2016]. The goal of the inversion is to determine the set of parameters describing the field and plasma that result in emergent spectra that best match the observations. The model for the magnetized plasma used in this procedure generally includes simplifying assumptions.

One very common set of assumptions used for producing vector magnetic field maps is the Milne-Eddington atmosphere (ME) [del Toro Iniesta, 2003; Mihalas, 1978; Skumanich and Lites, 1987]: the field strength and direction are constant with depth over the line-forming region, as are the governing thermodynamic parameters, such as temperature, density, velocity, as also manifest in the radiative transfer parameters such as the Voigt-function damping parameter and the line-to-continuum opacity ratios. Only the source function is allowed to vary (linearly) with optical depth. An unmagnetized component of plasma can also be included in this atmospheric model, occupying a fraction of the volume specified by a filling factor determined as part of the inversion. Typically, about 10 parameters are thus needed to characterize the field and the plasma. For forward-model-based inversion methods, the magnetic and thermodynamic behavior of the atmospheric model are incremented by means of some nonlinear least-squares fitting technique to optimize the agreement between the model and the observed Stokes profiles [del Toro Iniesta and Ruiz Cobo, 2016]. The resulting maps of the vector magnetic fields on the solar surface have been shown to be a reasonable average of the true magnetic and thermodynamic structures over the ≈150 km in height that most photospheric lines are formed [Westendorp Plaza et al., 1998]. Inversions based on the ME approximation can be used for data that are comparatively sparsely sampled spectrally, and are sufficient fast and robust to be routinely applied to large numbers of pixels [e.g., Borrero et al., 2011; Centeno et al., 2014], but contain no information on the variation of the magnetic field along the line of sight.

Without information about the gradient along the line of sight, the full current vector cannot be computed (see section 5.2.3). One way to obtain gradient information is to apply a ME inversion to two spectral lines with different heights of formation. This has been tried by Bommier [2014] using the UNNOFIT inversion [Bommier et al., 2007] for the Fe I 630.15 nm and 630.25 nm lines whose height of formation differs by approximately 100 km. The difference in the field from the two inversions in principle can be used to determine the gradient, but when \( \mathbf{\nabla}\cdot \mathbf{B} \) is estimated this way, it gives a systematically nonzero result, so work remains to be done to make this approach reliable for estimating the current.

For instruments with high spectral resolution, more sophisticated inversions that relax some of the assumptions in the ME approximation to allow for gradients along the line of sight are possible. For example, the Stokes Inversion based on Response functions (SIR) [Ruiz Cobo and del Toro Iniesta, 1992] includes gradients while assuming local thermodynamic equilibrium (LTE). The Non-LTE Inversion COde using the Lorien Engine (NICOLE) [Socas Navarro et al., 2000; Socas-Navarro et al., 2015] goes a step further and relaxes the assumption of local thermodynamic equilibrium in its multilevel treatment of Zeeman line transfer. This inversion code was developed for application to deep lines that span the photosphere and chromosphere, such as the infrared triplet lines of Ca II, but is well suited for retrieving the gradients for suitable photospheric lines such as the Fe I pair at 630.15 and 630.25 nm used by the SOT/SP. Indeed, an early application of non-LTE codes provided one of the first maps of the current vector in a sunspot [Socas Navarro, 2005].
5.2.2. Disambiguation

There is a degeneracy in the Zeeman regime in the polarization states for transverse fields that are oppositely directed (recall that the direction is given by \( \phi \approx \tan^{-1}(U/Q) \)), so there is an inherent 180° ambiguity in the direction of the transverse field (the Hanle effect and forbidden-line transitions also suffer from different degeneracies). Figure 5.3 schematically illustrates how the choice of the direction of the transverse component of the field affects the horizontal components, and hence the vertical current density. For a single height inversion, some assumption(s) and/or approximation(s) must be made to remove this ambiguity.

There are two main approaches to resolving this inherent ambiguity in the direction of the transverse field. One is to construct a reference field, typically either potential or linear force-free, and pick the selection closest to this reference field. The second is to approximate the vertical derivative needed to compute the divergence of the field, then minimize the value of \(|\nabla \cdot \mathbf{B}|\). Again, a potential or linear force-free field is generally used to approximate the vertical derivative. There are many implementations of these methods. An overview including comparison of their performance on synthetic data is given in Metcalf et al. [2006], while the effects of noise, spatial resolution, and method assumptions are evaluated in Leka et al. [2009].

For inversions that retrieve information about the variation of the field along the line of sight, it is no longer necessary to approximate the vertical derivative from a reference field. Instead, the divergence can be estimated using only inferred values of the field from the inversion [Crouch and Barnes, 2008]. Care must be taken in how this is done, since the use of finite differences or similar methods to approximate the horizontal derivatives makes this a nonlocal problem. When applied locally [Cuperman et al., 1992; Li et al., 2007; Wu and Ai, 1990] or sequentially [Boulmezaoud and Amari, 1999], an incorrect solution can still result. The use of a global optimization method to minimize the magnitude of the divergence is therefore most appropriate [Crouch et al., 2009], although even this can be susceptible to the influence of noise [Crouch, 2013]. Nevertheless, the use of gradient information removes the necessity for one assumption in disambiguating the data, and is thus preferable when the necessary information is available.

5.2.3. Ampère’s Law

Once the magnetic field vector has been determined at a discrete set of locations, Ampère’s law is used to infer the current density. This is most commonly done using the differential form

\[ \mathbf{J} = \frac{c}{4\pi} \nabla \times \mathbf{B}. \]  (5.4)

Note that for a single height inversion, only the horizontal variation of the field is known, and the vertical derivatives needed to determine the horizontal components of \( \mathbf{J} \) cannot be computed. One is left with simply the radial component of the current (shown in spherical coordinates here, but often approximated in planar geometry by ignoring the curvature of the solar surface):

\[ J_r = \frac{c}{4\pi r \sin \theta} \left[ \frac{\partial}{\partial \phi} (B_\phi \sin \theta) - \frac{\partial B_\theta}{\partial \phi} \right]. \]  (5.5)

It is frequently of interest to know the total current through a solar structure, such as a sunspot. In this case, the integral form of Ampère’s law is also sometimes used (e.g., Georgoulis et al. [2012]):

\[ I = \oint_S \mathbf{J} \cdot d\mathbf{A} = \frac{c}{4\pi} \oint_{\partial S} \mathbf{B} \cdot d\mathbf{r}, \]  (5.6)

where the closed curve \( \partial S \) must lie on the surface of the Sun for a typical inversion, so again it is the radial current that is being computed.

There are multiple ways in which the derivatives (integral) needed to compute the vertical current can be approximated. Generally, these either take the form of finite differences or finite elements. However, within each category, different orders of differencing can be used, along with different stencils. When the surface of the Sun

![Figure 5.3 Illustration of the ambiguity in vector magnetograms. The direction of the component of the field transverse to the line of sight is ambiguous by 180°, as indicated by the two green arrows. When transforming to heliographic components, the two choices of direction can result in not just a different direction and magnitude for the horizontal field (purple) but even a change in sign of the vertical field (red).](image)
is approximated as a plane (a substantial approximation, especially in sunspot areas), and the grid is regular in both directions, one common approach is to use a four-point stencil to compute the current on a staggered grid where it is accurate to second order [Canfield et al., 1993].

To illustrate the impact that the implementation has on the result, Figure 5.4 shows the radial current density computed using forward and backward first-order differences from SOT/SP observations of NOAA AR 11158 on 15 February 2011 at 10:11 TAI. Qualitatively, the results are very similar, with, for example, predominantly negative current in the sunspot in the lower left corner of the image. However, when the results of the two differencing schemes are compared point by point (Fig. 5.5), there are substantial differences in the magnitude of the current density at any particular pixel. From this we conclude that it is possible to infer the existence of strong currents at the solar surface, but the magnitude of the current is not generally precisely known.

For a sufficiently spatially resolved field, the forward and backward differences should result in the same value for the current. In the case shown here, the lack of the same value indicates that the grid size is too large for the differences to accurately represent the derivatives. While a higher order differencing scheme will converge faster as the grid size is decreased, in the observational case, the grid size is set by the instrument and cannot be varied, so a higher order scheme is not necessarily more accurate. The presence of unresolved structure becomes apparent in many inversion codes in the value of the dimensionless filling factor, α. In the center of sunspots, where the magnetic field is less variable, inversions generally find α ≈ 1. For a pixel containing unresolved structure, the best fit to the spectral line consists of a magnetic field that only occupies a fraction of the volume, and α is not unity. When this is the case, there is no physical possibility for any finite difference (or finite element) scheme to accurately determine the spatial derivative of the field.

An appropriate test is to evaluate the differences in inferred current by comparing the forward and backward differences for all strong field pixels compared with only those with a large filling factor. The correlation coefficient when including all strong field pixels shown in Figure 5.5 is 0.57, but restricting to α > 0.9, the correlation coefficient rises to 0.80. The increase in correlation

**Figure 5.4** The vertical current density from the inversion of SOT/SP data shown in Figure 5.1 for NOAA AR 11158 on 15 February 2011 at 10:11 TAI. The grey scale in both panels is saturated at ±0.1 A m⁻²: (left) first-order forward differences; (right) first-order backward differences. Only pixels for which the field is well measured are shown. The results are qualitatively similar.

**Figure 5.5** The vertical current density from SOT/SP data of NOAA AR 11158 on 15 February 2011 at 10:11 TAI comparing first-order forward differences (horizontal axis) with first-order backward differences (vertical axis). Only pixels for which the field is well measured are used in this scatter plot.
supports the idea that a lack of spatial resolution is impacting the estimated radial current magnitude in these data.

### 5.3. THE EFFECTS OF RESOLUTION ON ESTIMATES OF THE CURRENT

One of the main impacts on measuring the current is the resolution of the instrument being used, both spectral and spatial. In contrast to the SOT/SP, the Helioseismic and Magnetic Imager (HMI) [Hoeksema et al., 2014; Scherrer et al., 2012; Schou et al., 2012] on board NASA’s Solar Dynamics Observatory (SDO) [Pesnell et al., 2012] has lower resolution. It has a spatial resolution of approximately 1.0” compared with SOT/SP’s roughly 0.3” resolution. For HMI, and in some cases for SP, the observations are critically sampled, resulting in a pixel size that is half the spatial resolution of the telescope. HMI has a spectral resolution of approximately 8 pm and samples at six points across a single Ni I line at 676.8 nm, as compared with a spectral resolution of approximately 2 pm for SP, which samples at 112 points across two Fe I lines at 630.15 nm and 630.25 nm. The reduced resolution of HMI is a trade-off for the improved cadence and a much larger field of view than SOT/SP: HMI typically produces a vector magnetogram of the full solar disk every 12 min as compared to a single active region sized scan observed over roughly an hour for SOT/SP. In many cases, this makes HMI a more suitable instrument for studying solar energetic events like flares, which happen on timescales short compared with a typical SOT/SP scan.

Figure 5.6 shows the Stokes polarization spectra measured by HMI at approximately the same locations as shown in Figure 5.2 for SOT/SP. Qualitatively, the same features are seen in both. In the quiet Sun, there is an obvious absorption line in the unpolarized signal from HMI, but the polarized states are dominated by noise. In the sunspot umbra, in each line in the SOT/SP data and in the HMI data, the $V$ signal shows the same negative-to-positive trend with wavelength, and the $Q$ and $U$ signals show the same patterns between the two data sets. However, the relative magnitudes of the polarization signals are significantly different between the two, as is the level of spectral detail available; SOT/SP data clearly constrain any model atmosphere and field more strongly. For HMI, only a ME inversion is routinely performed, using the Very Fast Inversion of the Stokes Vector (VFISV) [Borrero et al., 2011; Centeno et al., 2014].

![Figure 5.6](image-url)
The data from SOT/SP and HMI are different in spectral and spatial resolution, spectral sampling method, as well as the line that is being observed. To disentangle the effect of spatial resolution from the other effects, consider SOT/SP data that have been spatially degraded without changing the other properties. Figure 5.7 shows the magnetic field seen at two spatial resolutions inferred from SOT/SP for NOAA AR 10953 observed at 18:35 UT on 30 April 2007, inverted using the NCAR/HAO Milne-Eddington inversion [Lites et al., 1993; Lites and Skumanich, 1990; Skumanich and Lites, 1987]. To construct the lower spatial resolution map, the spectra were spatially binned and then inverted to mimic the effect of a lower spatial resolution instrument [Leka and Barnes, 2012]. As expected, the smallest spatial scales are no longer resolved in the binned data.

To see the impact that (only) decreasing spatial resolution has on the current, Figure 5.8 shows the total unsigned current

\[ I_{\text{tot}} = \int |J \cdot \mathbf{n}| dA, \]  

in several different solar structures as a function of the pixel size (inverse of resolution). There is a general tendency for the total current to decrease as the spatial resolution decreases, with slightly different trends seen in different solar structures, presumably as a result of the current due to the smallest scale features being lost.

So is the current at the highest resolutions observed by SOT/SP correct? Parker [1996] showed that the mean current that would be inferred for unresolved magnetic fibrils bears little resemblance to the actual current. When the net current in a fibril is zero, yet current is present, a nonzero net current will still be inferred if the fibril is not resolved. Conversely, Leka and Barnes [2012] considered a potential (current-free) field model for an active region, constructed to include spatial scales smaller than can be resolved by SOT/SP but otherwise include solar-like magnetic field characteristics. Figure 5.9 shows the total current that would be inferred to be present in this potential field as a function of the pixel size. The current at the highest spatial resolution (smallest pixel size) is the lowest, thus the closest to being correct, but then increases, corresponding to the inability of finite differences to accurately recover the horizontal derivatives of the field, especially in the unresolved plage area. For some structures, there is then a decrease in the total current at the lowest resolution, similar to what is seen in the SOT/SP data. Note, however, that the magnitude of the current in the potential field case is substantially smaller than what is seen for the Sun. Thus, it seems likely that real currents are being detected, and indeed the magnitude of the currents...
could be substantially larger than what is inferred if there are spatial scales smaller than those presently resolved.

5.4. SUMMARY AND FUTURE DIRECTIONS

Energy stored and then released from the coronal magnetic field, in the form of both major solar energetic events and the continual background brightenings of coronal loops, make it clear that electric currents must be running through much of the solar atmosphere. However, to determine the magnitude of that current at any particular location is very challenging.

Inferring electric currents in the solar atmosphere requires a number of intermediate steps, each of which presents its own challenges. In general, the magnetic field in the solar atmosphere is determined by observing its effect on photons passing through magnetized plasma. From the field estimated at discrete locations, Ampère’s law is used to infer the current.

Most often, it is the impact of the Zeeman effect on an absorption line formed in the photosphere that is inverted to determine the magnetic field. For instruments observing this effect, there is a trade-off between measuring the line with high spectral resolution, having high spatial resolution, a large field of view, and high cadence. The instruments presently in use each have perhaps two of these properties.

Knowledge of the full current vector through the solar atmosphere may be the ultimate goal, and steps to achieve it are under way now. As demonstrated, unresolved structure poses problems for estimating even the normal component of the current; hence high-resolution observations (both in the spectral and spatial domains) are need to fully resolve the magnetic field vector and allow finite-difference methods to be employed with better confidence on discretely sampled data. This combination of high spectral and spatial resolution typically limits the cadence and field of view over which the current density can be inferred, but this challenge should be accessible with new facilities such as the Daniel K. Inouye Solar Telescope presently under construction. If the full current density vector is desired, then an inversion method that includes information about gradients in the field along the line of sight is required. Further advances would be made by understanding solar structures to the point of extending all calculations to treat nonplanar surfaces.

Not all topics in solar physics require quantitative absolute knowledge of the solar electric currents.
Well-characterized estimates of the vertical current density, as presently obtained with myriad assumptions applied to data, which do not fully resolve all solar structures, continue to provide insights into general trends and statistical understanding for topics where the strength of the current density is not necessary except in a relative sense.

ACKNOWLEDGMENTS

We thank both referees for clarifying several of the topics presented here. This material is based upon work supported by NASA under award No. NNX16AF83G.

REFERENCES


Severny, A. B. (1965), The nature of solar magnetic fields (the fine structure of the field), *Sov. Astron.*, 9, 171.


