

CAUSES OF SOLAR ACTIVITY

A Science White Paper for the Astro 2010 Decadal Survey
Submitted for consideration by the Stars and Stellar Evolution Panel

Mark S. Giampapa
National Solar Observatory
950 N. Cherry Ave.
POB 26732
Tucson, AZ 85726-6732

giampapa@nso.edu
520-318-8236

Sarah Gibson, High Altitude Observatory/NCAR
J. W. Harvey, National Solar Observatory
Frank Hill, National Solar Observatory
Aimee A. Norton, James Cook University (Australia) & NSO
A. Pevtsov, National Solar Observatory

1. Introduction

Our nearest star, the Sun, exhibits several processes that are common to other stars: magnetic cycles, flares, prominences, and stellar winds. Recent studies suggest that many of these processes are governed by universal laws applicable to both the Sun and other stars^[1, 2]. This solar-stellar connection can be used to extend our understanding of stellar phenomena via the study of solar processes. Thus, for example, studies of the solar dynamo will yield a better understanding of stellar dynamos: for example, what are the necessary conditions for dynamo action and how are magnetic cycles maintained? The study of the magnetic topology of *solar* flares and prominences will contribute significantly to the understanding of the nature of *stellar* flares, prominences, and winds. Herein we argue that a sustained program of long-term, high-continuity observations of the solar magnetic field is required to reach a better level of understanding of the universal processes taking place throughout the Universe.

The magnetic field is a corner stone of solar activity. Sunspots, flares and coronal mass ejections (CMEs), coronal heating and solar wind acceleration, all bear on the topological properties of magnetic fields. The magnetic field itself originates in a dynamo operating deep inside the Sun. Recent progress in dynamo theory and numerical modeling brought us closer to understanding the dynamo on the Sun to the point of attempting to predict the magnitude of the next solar cycle. Nevertheless, there remain many unresolved questions about the solar dynamo(s). For example, we still know very little about the subphotospheric meridional circulation that is a key ingredient for solar subphotospheric dynamos. The magnetic field in the solar polar regions is an untapped territory. Both meridional circulation and polar magnetic fields are critical in resolving the enigma of the dynamo. Also, it appears that there are two separate dynamos operating in the Sun: a deep-seated helical dynamo, responsible for the generation of the strong magnetic fields of active regions, and a near-surface chaotic dynamo, producing weak network fields. How these two dynamos interact with each other and what their roles are in the overall solar cycle are unknown. With respect to other stars, helical dynamos are typical for solar-type dwarfs, while non-helical (distributed) dynamos can operate on other, e.g., T Tauri and low-mass, fully convective dwarf stars. Since our observational capabilities are still a long way from resolving the fundamental scales of magnetic fields on the surfaces of other stars, the Sun offers the unique opportunity to study these dynamo processes directly.

The current state of solar physics indicates an important convergence between numerical modeling and observations. The numerical capabilities are now reaching the state when realistic 3D models of solar processes can be constructed at a level allowing the direct comparison with high resolution observations. One should expect that within the next decade the progress in both computational capabilities and the observations will come close to solving the nature of many solar phenomena. With these numerical models, developed to explain solar phenomena, one can use the Sun as a giant astrophysics laboratory to study processes occurring on other stars.

To achieve the goals described above requires two key ingredients: a sustained program of long-term, high-continuity observations of the solar magnetic field coupled with a strong program of topical numerical simulations aimed on modeling solar phenomena. This white paper concentrates on the former.

2. Observational Imperatives

Briefly, what is needed to achieve a better understanding of the causes of solar activity is a long-term, comprehensive study based on high and medium resolution full-disk, vector magnetic field observations at several layers in the solar atmosphere. To understand what triggers flares/CMEs, we need to observe the reconnection site at the spatial resolution and time cadence that allows us to resolve the processes taking place at the time of the event. Neither flare locations nor times are known *a priori*, implying the need for high resolution observations of the entire solar disk. Similarly, the early stages of active region emergence (and all processes related to this) are usually missed. On longer time scales, systematic long-term observations are required to understand the change in magnetic field topology during the course of a solar cycle. The key lack on the ground is poor observational continuity with regard to measurements of vector magnetic fields at different heights in solar atmosphere.

Although some of these shortcomings could be remedied by space-borne instruments (e.g. the Solar Dynamics Observatory, SDO), reliance on a single spacecraft for synoptic data may lead to significant data gaps, as has happened in the past (e.g., the SOHO spacecraft failure). Furthermore, ground-based observational capabilities often serve as a springboard for the development of space-borne instrumentation and as a training ground for students and engineers. Maintaining a ground-based network of identical magnetographs will ensure the continuity of synoptic studies of the solar cycle. The current anomalous solar minimum teaches us that we need more than a single example of a cycle in order to not draw unjustified conclusions.

3. Origins of Solar Magnetic Activity

The broad goal of understanding of solar magnetic activity at all its spatial and temporal scales requires observations of the Sun with enough spatial resolution, sensitivity and sufficient temporal frequency and continuity to identify what processes produce the variations. This broad goal is best addressed through synoptic network observations of the following interlinked processes:

- **Origin and evolution of sunspots, active regions and complexes of activity;**
- **Sources and drivers of solar activity and disturbances;**
- **Properties of the solar dynamo on short and long time scales.**

A network of instruments with vector polarimetric capabilities, along the lines of the NSO SOLIS (Synoptic Long-term Optical Investigations of the Sun) Vector Spectromagnetograph (VSM), is required for the acquisition of high-precision magnetic field data across the solar disk through the emerging science of vector polarimetry.

Many recent scientific advances have been made through the application of vector polarimetry using, for example, the Advanced Stokes Polarimeter (ASP)/Diffraction Limited Spectropolarimeter (DLSP), Spectropolarimeter for Infrared and Optical Regions (SPINOR), SOLIS VSM and the Hinode Solar Optical Telescope (SOT). Horizontal magnetic fields in the photosphere were found to have a higher-than-expected flux density distribution in quiet-Sun

regions^[4,5] while granular sized horizontal fields are observed to be associated with granular convection in plage region^[6]. Meanwhile, the study of active region geometries has shown small-scale complexities such as field-wrapping around penumbral filaments^[7] and the observed emergence of a helical flux rope under an active region prominence that suggests that helicity assists in maintaining prominences^[8]. The magnetic topology of the quiet-Sun with respect to the height-distribution of null points is also being examined using spectropolarimeters to better understand energy deposition and heating. New results show that the majority of null points are expected in the lower photosphere (54%), less in the chromosphere (44%) and with nearly insignificant frequency (2%) in the corona^[9]. Simultaneous efforts by Deluca et al.^[10] examine data from coordinated observations with Solar Optical Telescope, X-Ray Telescope, and the SOLIS VSM to determine the characteristics of the magnetic field topology that create bright points.

The last few years of solar spectropolarimetric measurements have furthered interactions between theory, modeling and data analysis. For example, observations of kG field strengths appearing in the presence of strong downflows are evidence for the convective instability mechanism of flux tube creation in the photosphere^[11]. Observed topologies of Stokes asymmetries in network cell interiors may be the first evidence of boundary layers predicted years ago by numerical simulations^[12]. Also, spatially resolved spectropolarimetric observations in the quiet Sun support the downward revision in the solar oxygen abundance^[13], in conflict with the success of traditional composition used by helioseismologists in solar interior models. These observations have re-invigorated discussions between modelers and observers. In brief summary to this point vector magnetic field data are an asset to the solar physics and solar-terrestrial communities, and extensive databases are in short supply. A global network of vector instruments will produce for the community an archive that is long overdue that contains temporally and spatially extensive coverage of photospheric and chromospheric vector magnetic field data.

4. Why a Global Network?

The Sun varies on a wide range of time scales from billions of years to milliseconds and on spatial scales from the entire heliosphere to less than a kilometer. The variations are manifested in electromagnetic radiation from radio frequencies of a few kHz to powerful gamma rays. Particle emission ranges from the relatively gentle solar wind to blasts of charged and neutral particles that reach Earth at virtually the speed of light. A synoptic program concentrates on observations of the Sun over time scales of minutes to decades and spatial scales from about 1 Mm to the full solar disk. Why are synoptic network observations needed? Fundamentally, we cannot understand the Sun by single or sporadic observations, no matter how good they are. Here are some of the reasons:

1. Observations from a single site on Earth are a poor match to the frequency and duration of many forms of important solar activity such as magnetic field evolution leading to flares and coronal mass ejections. A network of similar instruments is the solution to this problem. Sustained observations from a global network are required to catch short-lived but infrequent phenomena such as flares and coronal mass ejections.

2. Solar activity is not simply the endless repetitions of the same events. The statistical properties of the various phenomena convey vital information about their origins and can be utilized to forecast energetic events.
3. Solar data at a regular cadence is often critical in helping to interpret short-lived campaign observations, i.e., it is important to put high-resolution observations into context.
4. Sampling that is too isolated in time can give misleading results. An example is the observation made from Skylab for several months in the 1970s, from which it was concluded that coronal holes rotate nearly rigidly. Later observations showed that this was true only for that particular, brief phase of the solar cycle.
5. Some phenomena are intrinsically periodic such as the magnetic activity cycle. We still do not have enough well-observed periods or the proper kinds of observations to understand the degree to which the cycle is quasi-periodic or chaotic.

4.1. Origins of Solar Activity

Our understanding of solar magnetic activity is far from complete. It is commonly believed that the principal driver of solar activity is stressed magnetic field. The stresses are released in the solar corona, producing explosive flare events. Flares usually occur in areas where the magnetic configuration is complex, with strong shears, high gradients, long and curved neutral lines, etc.^[16] This implies that the trigger mechanisms of flares are controlled by critical properties of the magnetic field that lead eventually to MHD instabilities. But what kinds of instability actually govern, and under what conditions they are triggered, are unknown^[17] Vector magnetic field measurements are necessary to infer field topology and vertical electric current, both of which are essential to understand the flare process^[18]. Observations are required that can continuously track changes in the magnetic field and electric current with sufficient spatial resolution to reveal changes of field strength and topology before and after flares. The potential appearance of recurrent patterns in the local vector field evolution is the first step in the development of empirical predictors of flare activity. A vector network can provide these unique measurements of the vector magnetic field in specific flare active regions with high accuracy and at high cadence.

In addition to flares, coronal mass ejections (CMEs) are energetic eruptive events associated with heliospheric and geomagnetic disturbances. CMEs and flares were at first thought to be directly connected, with the flare driving the CME. However, only 60% of larger flares are associated with CMEs^[21]. Similarly, many CMEs are not associated with flares. It is now thought that CMEs and associated flares are caused by a common event (i.e., the CME peak acceleration and the flare peak radiation often coincide). In general, all of these events (including the CME) are thought to be the result of a large-scale restructuring of the magnetic field. In particular, CMEs apparently interact with the global-scale magnetic field^[22], but many CMEs, especially fast CMEs, are associated with flares, which are considered local phenomena. Model-based reconstruction of 3D magnetic structure is one way to estimate the field from observations^[25]. Models using vector field data in active regions provide the best match to the observations. More realistic MHD coronal models^[26] based on high-cadence vector-field maps as boundary conditions will greatly enhance our understanding of how the corona responds to evolving, non-potential active regions.

The origin of solar transient events and quasi-static structures ranging from active complexes to bipolar spot groups is intimately related to the emergence of magnetic flux. Flares are associated with the emergence of magnetic flux within active regions. Emerging magnetic regions near filaments leads to the eruption of the filaments^[27]. CMEs are also found to accompany emerging flux regions. However, emergence of isolated active regions can proceed without any eruptive events. This suggests that magnetic flux emerging into the atmosphere interacts with pre-existing fields leading to a loss of magnetic field stability. Observations of electric current and magnetic topology differences between emerging and existing fields may lead to an understanding of why emerging flux causes transient events^[28].

The high sensitivity, for example, of the SOLIS VSM instrument allows us to see much of this flux, extending from the ubiquitous small regions of mixed polarity, termed the 'magnetic carpet', to the larger magnetic structures. In addition, a network of such instruments would allow global-scale observations of the small-scale element distribution, their interactions on short and long time scales, and the resulting transformation of the large-scale field.

Many groups are actively engaged in research utilizing vector magnetogram data to gain further insight, and therefore predictive capability, concerning eruptive solar phenomena. It is generally accepted that the source of the energy released in solar transient events such as flares and CMEs results from the stored magnetic energy in an active region, where the solar magnetic field departs from a simple potential configuration. The best measure of magnetic non-potentiality can be provided by vector magnetograms. Global MHD models that reconstruct the 3D nature of the corona show improved ability to reproduce the eruption and evolution traits of specific CME events^[30]. These models are sufficiently comprehensive to incorporate the non-potentiality observed by vector magnetograms into the initial configuration of the magnetic field prior to eruption.

The application of vector magnetograms specifically to active regions that could be the sites of coronal mass ejections (CMEs) is especially crucial to understanding the origins of these powerful events. The high sensitivity of SOLIS/VSM-equivalent synoptic observations will enable the investigation of whether the magnetic evolution of an active region as represented by the time-rate-of-change of the total non-potentiality can provide a higher CME-prediction success rate^[34]. Therefore, extending these pioneering studies with a VSM-like network as the cycle is rising to solar maximum becomes particularly timely. In brief summary, we expect that the combination of global MHD modeling, CME and flare prediction capability, and routine vector magnetogram observations from a vector magnetic field network will result in more accurate models of these explosive solar events.

4.2. Sources and Mechanisms of Solar Irradiance Variations

Besides sustaining life on Earth, it is well known that spectral irradiance from the Sun has strong effects on the upper and middle terrestrial atmosphere. Effects of variations of total solar irradiance are harder to detect in the troposphere. The ultimate source of solar spectral and total irradiance variations is the changing solar magnetic field. Most of the variations can be explained by magnetic fields in the form of cool sunspots and hot faculae and plages with occasional transients from flares^[35]. Attempts have been made to model solar irradiance variations, starting

with the sunspot and facular symptoms of the magnetic field, with impressive success^[36]. But what works for one cycle does not work for another^[37], pointing to a need to better understand how the Sun distributes its magnetic field between sunspots and faculae and plages. The long-term availability of high-quality vector magnetograms of solar activity will help solve this dilemma.

5. Estimated Usage

A global network of vector instruments will serve a large international community of users. In order to estimate the usage of the VSM network data, consider the history of the line-of-sight magnetic field instruments. Mapping of the solar magnetic field started with one instrument and a few investigators. More instruments and scientists rapidly followed with sustained observations of the LOS field, starting at NSO in 1973 and continuing today. These observations are freely available with a result that from 1974–2004 more than 1,206 research papers and theses (involving more than 983 unique users from every country with a solar research program, including a Nobel Laureate and a current member of Congress) used data from the NSO synoptic magnetic field program. The program has been supported by both NOAA’s Space Weather Prediction Center and NASA for nearly 30 years in order to provide near-real-time data in support of those agencies’ missions. We see a similar path emerging for the full vector field.

Observations of the vector magnetic field are now moving from the exploratory to sustained effort phase. We project a large growth in users of vector data as it becomes available on a regular basis. We studied the usage of three exploratory instruments to set a baseline for future growth. These include the spectrograph-based Advanced Stokes Polarimeter located at NSO and operated by the High Altitude Observatory, and filter-based vector magnetographs of the Solar Flare Telescope (SFT) located at the National Astronomical Observatory of Japan, and the Imaging Vector Magnetograph (IVM) located at Haleakela and operated by the Institute for Astronomy of the University of Hawaii. Other notable filter-based instruments were operated by Big Bear Solar Observatory and NASA’s Marshall Space Flight Center. Only the SFT has a readily available public archive.

Table 1. Usage of Three Exploratory Vector Magnetographs (Source: ADS)

<i>Instrument</i>	<i>Service Dates</i>	<i>Refereed Papers</i>	<i>Citations</i>	<i>Unique Authors</i>	<i>Countries</i>
ASP (spectro)	1991 – 2006	57	1626	81	9
SFT (filter)	1991 - 2007	42	491	67	5
IVM (filter)	1992 - 2006	50	577	92	9

As expected, these usage values are small fractions of those of the sustained NSO LOS synoptic program. The SOLIS VSM is essentially a full-disk version of the ASP. We estimate that a VSM network would have usage at a rate up to an order of magnitude greater than ASP.

5.1. Community Support. A VSM network was the first-ranked, ground-based project in the Small Projects category in the 2000 NAS/NRC report *Astronomy and Astrophysics in the New Millennium*, i.e., the Decadal Survey. Specifically, it was the primary recommendation for ground-based initiatives in the small size category: “*The expansion of SOLIS to a 3-station network around the globe - \$4.8 million. To gain near continuous coverage in full-disk solar*

vector magnetic field monitoring, as the backbone of an assessment of the solar magnetic field flux budget over the solar cycle.”

The NSF Division of Astronomical Sciences (AST) Senior Review Committee report (issued in 2006) commissioned by the NSF Mathematical and Physical Sciences Advisory Committee and entitled, *“From the Ground Up: Balancing the NSF Astronomy Program,”* recommended the following with respect to the SOLIS VSM: *“The SR [Senior Review] supports the proposal to construct two additional copies of SOLIS [VSM] in order to ensure more complete temporal coverage.”*

6. Summary

The advantages of a VSM-like global network over the single-site currently in operation include:

- **Nearly continuous data on magnetic field evolution.**
- **Two-three times better chance of capturing rare events.**
- **Improved potential for short-term activity forecasts.**
- **Improved coverage for comparison with helioseismology.**
- **Better contextual data for space missions such as Hinode, Solar Orbiter, Solar Dynamics Observatory, etc.**
- **Ability to detect and correct systematic data errors.**
- **Robustness against a single-site failure.**
- **Fosters international scientific collaboration.**
- **Stimulates stronger research programs on solar activity.**
- **Better constraints on theoretical models of activity.**
- **As a ground-based system, has a potentially much longer lifetime than any space experiment can ever offer.**

7. References cited

1. Pevtsov A.A. et al, 2003, ApJ, 598, 1387-1391.
2. Schwadron, N. A. & McComas, D.J, 2008, ApJ, 686, L33.
4. Harvey, J.W. et al., 2007, ApJ, 659, L177-L180.
5. Lites, B.W. et al., 2008, ApJ, 672, 1237-1253.
6. Ishikawa, R. et al. 2008, A&A, 481, L25-28.
7. Borrero, J.M. et al. 2008, A&A, 481, L13-16.
8. Okamoto, T.J. et al. 2008, ApJ, 673, L215-218.
9. Regnier, S. et al. 2008, in press.
10. Deluca, E.E. et al. 2007, AGU, SH53A, 1058.
11. Nagata, S. et al. 2008, ApJ, 677, L145-147.
12. Rezaei, R. et al. 2007, A&A, 476, L33-36.
13. Socas-Navarro, H. and Norton, A.A., 2007, ApJ, 660, L153-156.
16. Sakurai, T. and Hiei, E., 1996, AdSpR, 17, 91-100.
17. Sammis, I., Tang, F. and H. Zirin, 2000, ApJ, 540, 583-587.
18. Canfield, R.C. et al. 1993, ApJ, 411, 362-369.
21. Andrews, M. D., 2003, Sol. Phys., 218, 261-279.
22. Luhmann, J.G. et al. 1998, JGR, 103, 6585-6594.
25. Falconer, D.A., Gary, G.A., Moore, R.L., Porter, J.G., 2000, ApJ, 528, 1004-1014.
26. Lee, J., White, S.M., Kundu, M.R., Mikic, Z., and McClymont, A.N., 1999, ApJ, 510, 413- 421.
27. Wang, Y.-M. and Sheeley, N.R., Jr., 1999, ApJ, 510, L157-L160.
28. Priest, E.R. and T.G. Forbes, 2002, A&ARv, 10, 313-377.
30. Riley, P., Lionello, R., Mikic, Z., and Linker, J., 2008, ApJ, 672, 1221-1227.
34. Falconer, D. A., Moore, R. L. and Gary, G. A. 2007, BAAS, 210, 135.
35. Woods, T.N. and Lean, J., 2007, EOSTransactions, AGU, 88, 44, 457.
36. Fröhlich, C. and Lean, J., 2004, A&ARv, 12, 273-320.
37. de Toma, G. et al. 2004, ApJ, 609, 1140-1152.