The Sun-Earth Connection as a Single System: Data Analysis and Processing Needs of Current and Future Missions

Alzate, N., Seaton, D., Kirk, M., Morgan, H., & Di Matteo, S.

Increasingly, diverse data sets are being used to understand the Sun-Earth connection. Data analysis tools applied to remote-sensing solar measurements (e.g., GOES x-ray flux[1], RHESSI[2], SDO[3]) and *in situ* heliospheric measurements (e.g., ACE[4], STEREO[5], SOHO[6]) contribute crucial information to our understanding of the Sun/Corona/Heliosphere as a single system. The availability of multiple datasets is matched by a plethora of analysis and processing techniques, all of which extract useful information from the observations. These include advanced image processing techniques to reveal structural detail, inversion methods for physical values (e.g. density, temperature), feature recognition and tracking, and processing of data for use as boundary conditions for models. Increases in processor speed and massively parallel computational techniques have enabled improvements in the possible spatiotemporal resolution of analysis methods and models. Improvements in physics-based numerical simulations have improved our understanding of the underlying solar physical processes. However, current datasets and tools are not always used to their full potential often as a result of the limitations summarized in Table 1.

Table 1. Current data and computational limitations.

Data Limitations
• Different datasets and catalogs are in different formats and require different analysis tools • Different datasets often cover different regions (see Figures 1 and 2) and/or properties • Observations limited by spacecraft and ground-based instrument configuration, which limits cross-mission investigations (e.g. *in situ* and remote sensing observations)
Computational Limitations
• Numerous analysis approaches (e.g. image processing, time series analysis) not fully optimized • Various programming languages used, some of which require expensive licensing • Lack of browsing, cataloging and public availability of software tools to the community

One of the first steps is the cataloging of observations and simulation outputs in standardized formats of both data and metadata. The exchange of information included in the metadata is fundamental for data integration in models, comparison of model output with data, and, most importantly, for the development of new tools, as needed. Extraction of useful information from remote sensing observations is essential for exploring the physical properties of the plasma to be then compared with models and, when possible, *in situ* measurements. For example, emission from spectral lines from different charge states of the same element serves as a diagnostic for electron temperature[7]. Other examples include the use of polarimetric observations in select spectral lines for coronal magnetometry[8,9].

Fully extracting useful information from observation data requires advanced image processing techniques. This is true even for high spatial resolution white light/EUV observations due to the steep decrease in plasma density/emission with heliocentric distance from the Sun. Additionally, coronagraph observations are limited by scattered light and EUV observations are usually made over distances up to only a fraction of a solar radius above the limb.

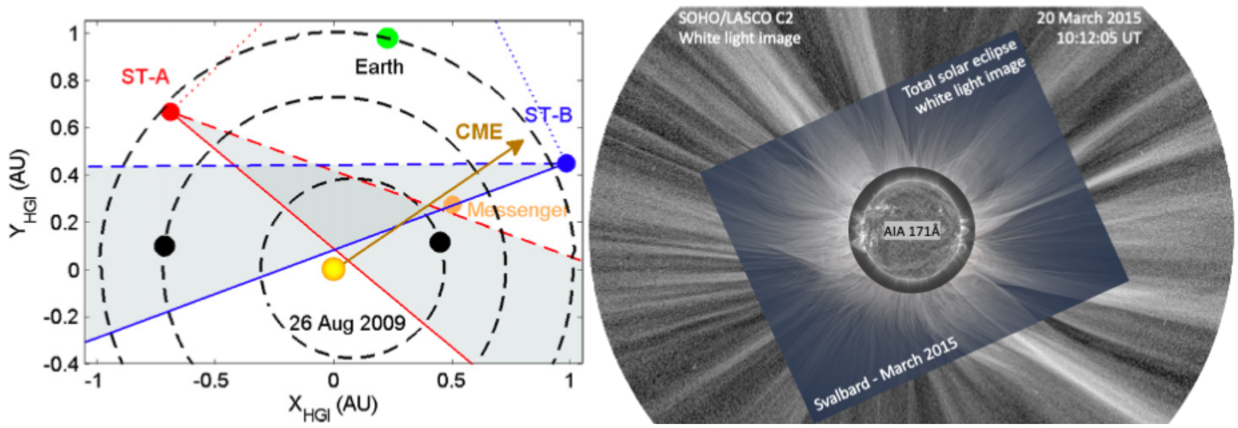

Figure 1. *Left) Location of STEREO A, STEREO B, MESSENGER, Earth, and a CME in the ecliptic plane on 26 August 2009*[10]*. Right) Composite WL LASCO C2 (outer), March 2015 total solar eclipse over Svalbard, Norway (middle; M. Druckmüller & S. R. Habbal) and* SDO/AIA *171 Å (insert).*

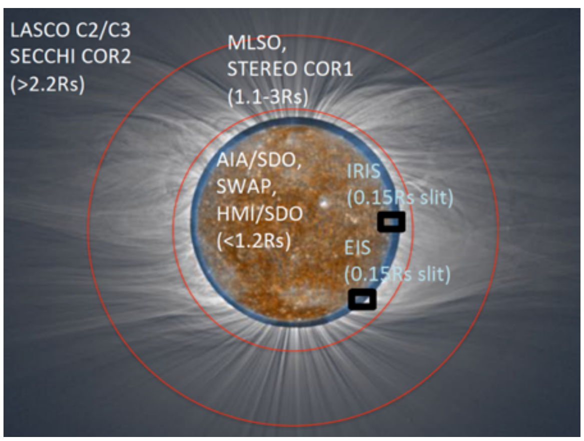

Figure 2*. Composite white light image of July 2009 total solar eclipse over Enewetak, Marshall Islands (outer) and SOHO/EIT 304 Å (insert; M. Druckmüller).*

Previous studies[11,12,13] have focused on the application of advanced image processing techniques to EUV and coronagraph data to reveal fine-scale signatures in the low corona that have an impact on the structure of the extended corona, which would otherwise go unnoticed. One way to achieve a continuous coverage of the corona is to create `maps' using instrument suites to reveal the connection among activity at various heights in the corona and the heliosphere[14]. Figure 1 (right) and 2 are examples of the radial span of various instruments, which independently do not span an uninterrupted FOV from the low corona out to several solar radii. Other commonly used tools for extracting information from dynamic events include, for example, Huygens plotting[15] and persistent mapping[16].

Developing these techniques for application on current datasets will prove valuable in interpreting future datasets. In fact, new missions present new challenges to the data analysis capabilities of the solar and heliospheric community. The resolutions of observations have also improved with new instrumentation. This leads to very high volumes of data, which test the limits of current computational capabilities, requiring compromises between amount of information able to be gathered in time and space versus the amount of information actually utilized for science. For example, Parker Solar Probe (PSP)[17] and Solar Orbiter (SO)[18] are "encounter" missions that strictly limit the duration of data collection. The Daniel K. Inouye Solar Telescope (DKIST)[19] provide huge amount of information but with respect to a very small region of the Sun. All three of these missions require contextual imagery and modeling to fully utilize their capabilities and meet their science goals. Current missions in development, such as PUNCH[20], extend remote sensing data at distances from the Sun out to Earth through multi-point observations, which must be supported by numerical simulations, models, and robust image processing software.

An example of successfully merging multi-view remote observations and multi-point *in situ* measurements analysis was described by [10] who reconstructed a `stealth'[21] CME from *in situ* measurements (see Figure 1 left). Accessibility and Interoperability of various tools is essential to extract all possible information available about the Sun-Earth system, which serves the larger heliophysics community including those developing modeling and theory. For example, detailed information of the Sun serves as boundary conditions for models like WSA[22] and ENLIL[23], while improved information *in situ* serves as validation of their output. The emerging data science and parallel computing capabilities, combined with the large amount of data that will be available from near future and future missions, will provide unprecedented opportunities. However, future missions in isolation will not fully solve science and problems, but developing tools to seamlessly merge output from complementary efforts will greatly enhance the performance of future missions.

In the next decades, we should pursue the following:

- The standardization of tools and techniques for the analysis of heliospheric data
- The establishment of web services for collecting and cataloging the tools themselves and their outputs. Current examples include the CCMC[24], VSO[25] and HEK[26]
- Implementation and coupling of tools/models via directed support of software development
- Targeted investment in the full use of computational resource capabilities
- Combination of observations and tools capable of recovering the necessary information needed for a global view of the Sun-Earth system[27]
- Tech companies - scientist communication to develop new technol and related science tools
- The coordination between international working groups of these subjects, including computer scientists dedicated to optimizing software tools, under the guidance of an international dedicated core team

By 2050, we envision a library of opensource software tools and techniques that would be the basis for the extensive and optimized use of archived datasets, which set the standard for the way new observations are utilized. The documentation and examples from this library would be freely hosted through a comprehensive web interface (equivalent to the current ADS for publications).

References

[1] Hanser, F.A. *et al. (1996), Proc. SPIE,* 2812, GOES-8 and Beyond
[2] Lin, R.P., *et al.* (2002), *SoPh.,* 210, 3–32
[3] Lemen, J. R., *et al.* (2012), *SoPh*, 275, 1
[4] McComas, D., *et al.* (1998), *SpScRev.,* 86, 563
[5] Howard R. A., *et al. (2008), SpScRev, 136, 67*
[6] Brueckner, G.E., *et al.* (1995), *SoPh,* 162, 357
[7] Habbal, S.R., *et al.* (2010), *ApJ,* 719, 1362
[8] Judge, P.G. (1998), *ApJ*, 500, 1009–1022
[9] Lin, H., *et al.* (2000), *ApJ*, 541, L83
[10] Nieves-Chinchilla, T., *et al.* (2013), *ApJ,* 779, 55
[11] Alzate, N., & Morgan, H. (2016), *ApJ,* 823, 129
[12] Alzate, N., & Morgan, H. (2017), *ApJ,* 840, 103
[13] Morgan, H., *et al* (2013), *ApJS,* 206, 19
[14] Alzate, N., *et al.* (2020), under review at *ApJ.*
[15] Wills-Davey, M. J., & Thompson, B. J. (1999), *Sol. Phys.*, 190, 467
[16] Thompson, B. J., & Young, C. A. (2016), *ApJ,* 825, 27
[17] Fox, N.J., *et al.* (2016), *SpScRev.,* 204, 7–48
[18] Müller, D. *et al.* (2013), *Sol. Phys.,* 285, 25–70
[19] Tritschler, A. et al. (2016), *AstrNachr*, 337, 1064
[20] PUNCH https://punch.space.swri.edu/science
[21] Robbrecht E., *et al.* (2009), *ApJ,* 701, 283
[22] Arge, C. N., & Henney, C. J. (2013), *AIP Conf. Proc.,* 1539:1, 11-14
[23] Odstrcil, D. (2003), *Adv. Space Res.,* 32, 497
[24] CCMC https://ccmc.gsfc.nasa.gov
[25] VSO https://sdac.virtualsolar.org
[26] HEK https://www.lmsal.com/hek/
[27] Seaton, D. et al., A Strategy for Coherent and Comprehensive Observations of the Middle Corona, Helio2050 white paper