

eSDO: UK Access to the Solar Dynamics Observatory

Leveraging the UK Virtual Observatory for AIA and HMI Data and Algorithms

Elizabeth Auden¹, J.L. Culhane¹, Y.P. Elsworth², A. Fludra³, M.J. Thompson⁴

¹Mullard Space Science Laboratory, University College London.

²Physics Department, University of Birmingham.

³Rutherford Appleton Laboratory

⁴Department of Applied Mathematics, University of Sheffield

Abstract

The UK-based SDO e-Science consortium has been funded by PPARC as a three year project to work with US SDO investigation teams to extract data products and enable data access by the UK solar physics community. Following a year of research into meeting the UK solar community's needs through algorithms, data centre access and quicklook products, two years of eSDO development and implementation will begin in October, 2005.

Introduction

PPARC has funded the eSDO project to make data and algorithms from the Solar Dynamic Observatory (SDO) available to the UK solar community using the virtual observatory. The Atmospheric Imaging Assembly (AIA) and Helioseismic and Magnetic Imager (HMI) instruments on board SDO will produce ~ 2 TB of raw data per day, and the high-level pipeline data products will be approximately one tenth of that volume. Four UK Solar Physics groups have Co-Investigator involvement in the SDO AIA and HMI investigations: MSSL, Rutherford Appleton Laboratory, and the Universities of Birmingham and Sheffield.

The huge SDO data volume requires special measures to ensure effective data handling. First, eSDO scientists and programmers will develop local and global helioseismology specialist algorithms as well as feature recognition procedures for AIA and HMI images. These algorithms will be made available to the UK virtual observatory through web service access to systems in the US and UK. Next, development of quicklook and summary data will allow high speed searches of the SDO databases. SDO data will be made accessible through installation of AstroGrid software at the US SDO data centre hosted at Stanford University as well as deployment of a UK SDO data centre. The UK data centre will use a high volume storage facility, such as the ATLAS petabyte store at RAL, to provide access to archived SDO science products, quicklook products and visualization software.

Algorithms

Three categories of specialist algorithms will be prepared by eSDO developers. Feature recognition algorithms will detect solar features present in AIA and HMI images. Global helioseismology algorithms will analyse HMI high level data from deep inside the solar interior while local helioseismology algorithms will deal with solar activity origins within the Sun's convection zone. Many algorithms will be installed at Stanford University and Lockheed Martin as optional user-selected "add-ons" for generating on-the-fly high level science products; these algorithms would be available through "grid user" access to the Stanford and Lockheed data centre

systems. In addition, eSDO algorithms will be hosted in the UK as AstroGrid CEA tools. A third method of access to eSDO algorithms is also under consideration: each algorithm could be encased in an IDL wrapper and distributed through SolarSoft using the MSSL gateway.

Image Feature Recognition Tools (AIA / HMI)

- Loop recognition:
- Magnetic field extrapolation
- Magnetic helicity computation
- CME dimming region recognition
- Differential emission measure computation
- Small event detection

Developers and scientists and MSSL will produce three image processing algorithms: coronal loop recognition, magnetic field extrapolation, and helicity measurement. Coronal loops observed by SDO will be identified in AIA images. Currently, an examination of curvelet analysis techniques is being appraised for suitability in identifying loops without the use of accompanying magnetic field data. A second approach would use HMI magnetograms in conjunction with AIA images to match identified loops with magnetic field line extrapolations. The magnetic field extrapolation algorithm will compute magnetic connectivities between opposite magnetic polarities and draw characteristic field lines using nonlinear force field techniques. Using a series of SDO magnetograms as input, this procedure will compute the flux of magnetic helicity through the photospheric boundary. Magnetic field extrapolations and the derived magnetic connectivities will be used to improve the helicity flux results.

RAL will also take responsibility for image processing and feature recognition algorithms: recognition of coronal mass ejection (CME) dimming regions, small event detection, and differential emission measure (DEM) computation. Time series of difference images in coronal lines will be examined for the change of brightness, or dimming. Visual detection of the coronal dimming has proved successful in the past, but this automated algorithm will detect varying spatial scales of the dimming region from approximately 45 heliographic degrees down to a fraction of an active region area. Calculation of differential emission measure (DEM)¹ requires calibrated line intensities and line emissivities as a function of temperature. SDO developers at Lockheed Martin will be optimizing multiple DEM solutions, but the eSDO algorithm will initially focus on DEM program included in the CHIANTI package. The small event detection algorithm will detect small-scale brightenings in AIA images of the corona, transition region and photosphere. Based on the EIT brightness detection work of Berghmans, Clette, and Moses², this procedure will define a reference background emission, identify events through a scan of light curves of all pixels, and determine event dimensions in the temporal and spatial domains.

Helioseismology Tools (HMI)

- Subsurface flow analysis
- Perturbation map generation
- Local helioseismology workbench

- Mode frequency analysis
- Mode asymmetry analysis
- Solar rotation profiling
- Oscillation source analysis

The University of Sheffield will concentrate on local helioseismology algorithms: helioseismic inversions, subsurface flow analysis, and perturbation map generation. First, the inversion algorithm will be used to obtain subsurface flow maps under tracked active regions in HMI dopplergrams. Subsurface flows will also be measured in the upper convection zone, resulting in travel-time difference maps for different skip distances and orientations. Flow maps under tracked regions will be combined into synoptic maps. These data products will promote understanding and predicting active region evolution and evolution of atmospheric magnetic structures. The perturbation algorithm will measure wavespeed anomalies in the upper convection zone and, in particular, under active regions to understand active region and sunspot subsurface structures and flux emergence. This analysis will generate travel-time anomaly maps for different skip distances, subsurface wavespeed anomalies under the tracked region obtained by inversion, and combinations of wavespeed anomaly maps under tracked regions into synoptic perturbation maps.

The University of Birmingham will use HMI data to focus on global helioseismology algorithms, mode frequency analysis and mode asymmetry analysis. The mode frequency analysis algorithm will calculate mode parameters such as frequency, power, line width and variability. The analysis procedure will filter HMI time series data to isolate single oscillation modes or a family of modes. After a Fourier analysis is applied to these modes, fitting procedures will determine the mode parameters. The mode asymmetry algorithm will measure the departure of the mode from a symmetric profile by applying simultaneous fitting over several modes to identify the noise contributions.

If time permits, scientists and developers from the universities of Sheffield and Birmingham will collaborate on two further algorithms, solar rotation profiling and oscillation source analysis. HMI data will be used to develop solar rotation profiles as a function of depth and latitude for the internal Sun between the solar surface and the convective zone. The oscillation source analysis algorithm will begin with the assumption that the mode asymmetry in velocity can be related to the location of the oscillation source. This algorithm will augment the oscillation source profile using noise analysis in both velocity and intensity.

SDO Quicklook Formats

The huge volume of SDO data collected every day makes it imperative to have an efficient method of searching the data archive. Quicklook products, also referred to as "browse data", allow scientists to identify specific datasets more efficiently. For some SDO products such as internal rotation and sound speed maps, quicklook products will take the familiar form of image thumbnails. Most thumbnails will be generated and archived at Stanford University and Lockheed Martin. However, other quicklooks will display a synopsis of solar behaviour over long periods. For instance, the HMI instrument will produce a magnetogram every 45 seconds, but the UK data centre can request one such image every hour and then generate a movie showing a month of

solar activity. Movies of AIA data can achieve further compression by overlaying EUV images from different channels onto the same images.

Interaction with the UK solar community has shown that future users will be interested in both thumbnail images and movies of low level AIA and HMI data. Two storage options will be explored during Phase B of the eSDO project. In the first case, thumbnail images will be generated from AIA calibrated filtergrams and HMI line-of-sight magnetograms at a cadence of one per hour. MPEG movies will be generated from one month's worth of thumbnail images and stored. Alternatively, A second visualization solution would involve the generation of thumbnail images from all AIA calibrated filtergrams and HMI magnetograms; rather than automatically generate MPEGs from these images, users could select the start time, end time, and cadence of filtergrams and magnetograms through a CEA tool that would encode these images as an MPEG on the fly.

US and UK SDO Data Centres

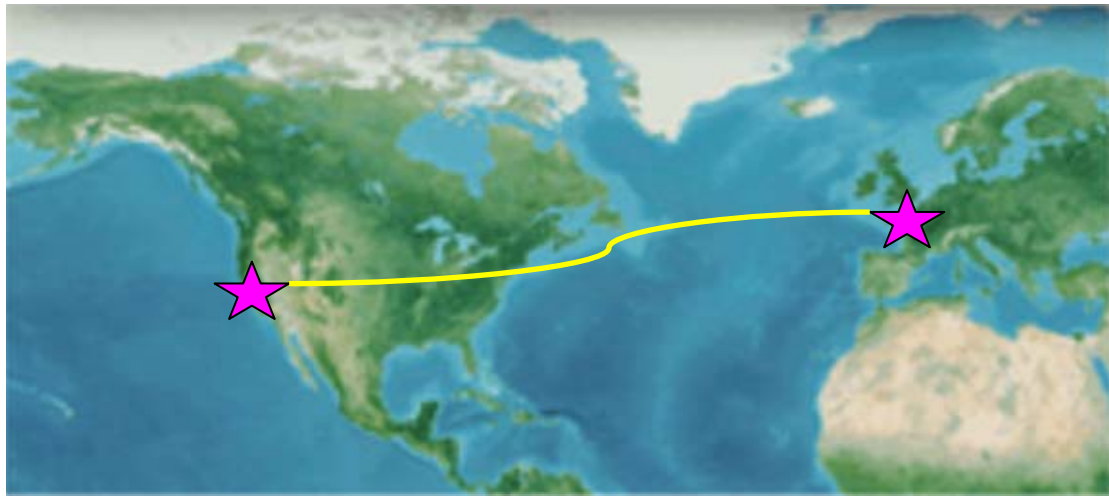


Figure 1: SDO data centre at the RAL ATLAS facility will request data from the primary SDO data centre at Stanford University.

The event-driven nature of many solar research projects means that demand for data and algorithms will peak after flares, coronal mass ejections, and other solar events. This will place stress on data centres, workflow engines and data conduits. The eSDO project will develop methods to balance data request loads between US and UK SDO data centres.

US Data Centre

The primary SDO data centre will be hosted at Stanford University. Installation of the AstroGrid³ DSA (Data Set Access) software at this facility will allow UK “grid user” access to HMI and AIA images along with maps, movies, magnetograms, and other high level data products. While all low level and some high level data products will be held online and then archived in a grid-accessible tape store, some high level products will be generated on-the-fly by user requests.

Accessing data products through AstroGrid involves two software components: DSA (data set access) and CEA (common execution architecture). Metadata and keywords

describing data products are held in a database – the data centre at Stanford will store metadata in an Oracle database. Users can search the database by submitting ADQL (astronomical data query language) queries to the DSA software via a web service. DSA translates the ADQL into an SQL statement appropriate to the backend database. Once the metadata results have been returned by the database, DSA converts the tabular information to VOTable formatted XML file. The VOTable can be returned to the user or used as input to the data centre’s CEA tool for data access. This CEA tool is a script that transfers copies of requested data products to a user’s online storage area. The tool can be configured to use the data centre’s preferred transfer protocol, such as FTP, gridFTP, wget, curl, and http.

All data stored in the SDO data centre at Stanford will be available for public access. Most data products will be automatically generated and stored in a disc cache or tape archive, but certain high level data products will be generated on the fly from user requests and not stored online⁴. The AstroGrid DSA software at Stanford will allow grid access to most level 1 products, such as calibrated AIA filtergrams and HMI line-of-sight magnetograms, along with level 2 HMI products like dopplergrams and intensity grams. Level 2 AIA data products such as thermal maps will be stored in online disc caches at Lockheed Martin; these products will be accessible through VSO (the Virtual Solar Observatory). The level 2 and 3 HMI products that are not automatically generated will be created on-the-fly through user requests in the form of database calls. Access to these database calls will be available to the Stanford and Lockheed Martin teams as well as their co-investigator teams. However, “grid user” access will be negotiated during Phase B of the eSDO project, particularly as AstroGrid matures through the potential use of digital certificates in conjunction with community accounts.

UK Data Centre

A secondary SDO data centre will be installed in the UK to reduce traffic on the SDO data centre at Stanford and to reduce the need for transatlantic data calls. The ATLAS storage facility at RAL has been identified as a candidate for the UK SDO data centre. This facility has the ability to provide temporary storage through disc caches as well as long term online tape library storage. It is anticipated that three types of storage will be required: a partial mirror of HMI and AIA data products generated at Stanford and Lockheed Martin, a cache of popular data products generated by UK algorithms, and storage of quicklook products such as thumbnails, movies, and catalogues. This will allow faster access to UK users through bypassing traffic on the Stanford network if data products are available in the UK data centre. Second, a permanent tape store will sample magnetograms and other images at an hourly or daily cadence and convert them to movies and catalogues stored over the duration of the SDO mission.

One model for the UK data centre will provide a mirror of HMI magnetograms, dopplergrams and intensitygrams over the lifetime of the SDO mission. In addition, the UK data centre would keep a rolling store of the most recent 2 months’ worth of common AIA science products:

- 1000 by 1000 low resolution full disk images
- 500 by 500 high resolution images tracking active regions
- Thermal maps: 4 to 10 images / day

- DEM measure: 4 – 10 images / day
- Irradiance estimates: tabular data several times / day
- Potential force field magnetic extrapolations: 4 images / day
- Non-linear force field magnetic extrapolations: 1 image / day

Although file sizes are only estimated at the moment, it is anticipated that the three HMI products under consideration for archiving – magnetograms, dopplergrams and intensitygrams – will require around 35 TB per year (uncompressed), while the accumulated AIA low and high level products would approximate storage requirements of 55 TB per year (uncompressed). An archive of thumbnail images created from low level AIA and HMI products would occupy less than 1 TB of storage, but a collection of monthly MPEGs generated from hourly sampled thumbnail images would add a negligible amount of storage requirements to the overall archive.

The UK data centre will store metadata describing these data products in a MySQL database searchable through ADQL queries to an AstroGrid DSA installation. A CEA tool will be configured to use either wget or gridFTP as a transfer mechanism between the Atlas data store facility and a user's grid or local storage. The UK DSA and CEA tools will be accessible through the AstroGrid portal, but any software infrastructure (such as other virtual observatories, IDL's ION software, or standalone applications using web service containers like Axis) that can make web service calls to these tools will be able to access their functionality.

Conclusions

The Solar Dynamic Observatory will produce substantial amounts of raw data every day that will be further processed into high level products. Solar communities in both the US and the UK will be interested in high and low level products; to assist demand for data and algorithms, the eSDO project will deploy VO-accessible SDO data centres using AstroGrid DSA software on both sides of the Atlantic. The US data centre at Stanford University will provide a back end to grid user accessible databases and file systems. In addition, a UK data centre will provide long term storage of certain AIA and HMI science products in addition to a cache of high level data products generated by eSDO algorithms in the UK. UK access to SDO data and algorithms will be improved through visualisation using thumbnail images of low level products, event catalogues and movies of solar behaviour over the lifetime of the SDO mission.

References

- [1] Withbroe, 1975, Solar Phys., 45, 301.
 [2] Berghmans, Clette and Moses, 1998, A&A, 336, 103.
 [3] Nicholson et al, 2005, AstroGrid Version 1.0,
<http://wiki.astrogrid.org/bin/view/Astrogrid/ReleaseV1Pages> (viewed 18 July 2005)
 [4] HMI AIA JSOC Ground Data System Plan Overview, HMI-S019, 10 Nov 2004,
http://hmi.stanford.edu/doc/SOC_GDS_Plan/JSOC_GDS_Plan_Overview.pdf
 (viewed 18 July 2005)