SPACEINN

D3.19) 1st report on HMI data products and usage statistics

This deliverable is provided by the Max-Planck-Institut für Sonnensystemforschung (MPG).

4th December 2013

HMI Data Products

The Solar Dynamics Observatory (SDO) was launched 11th February 2010 and is NASA's most important solar physics mission of this decade. The three primary instruments on board SDO are the Helioseismic and Magnetic Imager (HMI), the Atmospheric Imaging Assembly (AIA) and the Extreme ultraviolet Variability Experiment (EVE). The data will be predominantly used to learn about solar magnetic activity.

The most important instrument for helioseismology is the Helioseismic and Magnetic Imager (HMI). HMI data are key for studying the Sun's interior and various components of magnetic activity at the photospheric level. HMI observes the full solar disk all year round and provides state-of-the-art 4096x4096 pixel images. The solar scientific community is currently inundated with a flood of large-volume/high-resolution data and corresponding high-level science data products.

HMI data are described in detail on the Joint Science Operations Center (JSOC) website http://jsoc.stanford.edu.

The main HMI observables are available at cadences of 45s and 720s, except for the continuum intensity data without limb darkening that is only available at a cadence of 720s. The HMI observables are

-   Line-of-sight magnetograms
-   Line-of-sight Dopplergrams
-   Continuum intensity
-   Continuum intensity without limb darkening (only at 720s cadence)
-   Line widths
-   Line depths

The data volume generated for each of the HMI observables (45s cadence) amounts to 1 TB per month. At the time of writing the SDO satellite had been collecting data for 45 months and hence there is 45 TBs for each of the lDopplergrams, magnetograms,

continuum intensity, line-widths, and line-depths; a total of 225 TBs for the five main HMI (45s cadence) observables.

There are also a variety of magnetic field data products available at a cadence of 720s

- Stokes vectors parameters (I, Q, U, V)
- Milne-Eddington inversion results
- Vector magnetic field in HMI Active Region Patches

There are also several helioseismic science data products resulting from the JSOC pipeline including

- Ring-diagram analysis products for tracked HMI Doppler cubes of size 5, 15, and 30 degrees, full 3D power spectrums, multi-parameter model fit parameters. In addition, sub-surface flow maps are available from inversions of the fits to the power spectrum.
- Time-distance products including travel-times maps using both Gabor wavelet fitting and the method from Gizon and Birch (2002). Also, inverted flows and sound-speed results are available for either Gabor or Gizon and Birch travel-times as well as using either Born of ray kernels.
- Global helioseismology products such as mode parameters, power spectra.

Lower level products such as HMI Level-1 filtergrams are also available.

Most of the HMI data described above are available through the Data Record Management System (NetDRMS, http://jsoc.stanford.edu/netdrms) software that was developed by Stanford University. NetDRMS uses "publication" and "subscription" procedures for sharing SDO to users worldwide. The Joint Science Operations Center (JSOC), in the USA, has published (in the NetDRMS sense) all the HMI data, and therefore it's possible for participating sites that also have NetDRMS installed to obtain that data by subscribing to the relevant data series. The subscribing site will automatically receive the HMI meta-data of interest, and furthermore, the site can then choose to transfer all the corresponding HMI images or a subset of the images. The meta-data is always transferred

directly from the publishing site (e.g. JSOC for HMI data) so that there is always a unique source for the meta-data. The meta-data is typically on the order of GBs so there are no bandwidth problems when many subscribing sites obtain the meta-data from a single source. However, the corresponding images will be transferred from a site that is geographically closest to the subscribing site. The images are on the order of TBs, so it's important to spread the load and reduce the bandwidth on JSOC.

GDC-SDO Data Products

The German Data Center for the Solar Dynamics Observatory (GDC-SDO) is hosted by the MPG and funded by DLR, and is dedicated to the acquisition and distribution of SDO data. It is a European master distribution center for HMI data and hence fosters the European scientific community. The GDC-SDO will ensure that HMI images are available for the SPACEINN Project. The GDC-SDO has a raw online storage capacity of 660 TB plus a tape library/robot with 200 SDLT-4 slots that provides 160 TBs of near-line archiving capacity. The incredibly high-volume of HMI data requires such a dedicated center within Europe to facilitate rapid and transparent data access and scientific analysis to the European community.

The German Data Center for the Solar Dynamics Observatory (GDC-SDO) is subscribed to all the HMI observables through the Data Record Management System (NetDRMS) infrastructure; i.e. the GDC-SDO has all the HMI meta-data for the following data series

- Line-of-sight magnetograms (45s and 720s cadences)
- Line-of-sight Dopplergrams (45s and 720s cadences)
- Continuum intensity (45s and 720s cadences)
- Continuum intensity without limb darkening (only 720s cadence)
- Line widths (45s and 720s cadences)
- Line depths (45s and 720s cadences)

Furthermore, all the 4096x4096 images for the line-of-sight Dopplergrams, line-of-sight magnetograms, and continuum intensity maps are online at the GDC-SDO. This amounts to 135 TBs of data, which is about 20% of the raw online storage capacity of the GDC-SDO.

In addition, much smaller subsets of images for the line-widths and line-depths are also online due to individual user requests.

Summarily, the GDC-SDO is currently providing HMI line-of-sight Dopplergrams, line-of-sight magnetograms and intensity images for the SPACEINN project; through the NetDRMS infrastructure. When the SPACEINN community requests additional data products, the GDC-SDO will subscribe to them immediately and make them available. The NetDRMS infrastructure is capable of very efficient transfer rates of up to several TB per day.

Usage Statistics

Several users and user groups have already made considerable use of, or expressed interest in, the GDC-SDO since the start of the SPACEINN project.

The Kiepenheuer-Institut für Sonnenphysik (KIS) in Freiburg, Germany recently expanded their available storage resources and they require, in the very least, all HMI Dopplergrams for their scientific research. They intend to start the transfer of all the Dopplergrams in the coming months, which at the time of writing amounts to over 45 TBs of data. Then the data series will be automatically and continuously updated (through NetDRMS), amounting to an extra 1 TB every month. KIS has already successfully completed the initial testing that is required prior to installing NetDRMS. When KIS subscribes to the HMI Dopplergram data series (in the coming months), the meta-data is on the order of 10 GB and will come directly from JSOC, since JSOC is the unique source for the meta-data for that data series. However, the HMI Dopplergram 4096x4096 images will be transferred from the GDC-SDO due to its close proximity to KIS.

Users from IRAP in Toulouse, Paris, regularly made requests to the SPACEINN GDC-SDO Help Desk for HMI data in order to run a Coherent Structure Tracking routine. The GDC-SDO Help Desk provided the data manually by using export utilities that come packaged with the NetDRMS system. The data amounted to a few days worth or about 200 GBs.

Finally, the MPG also made extensive use of the GDC-SDO, which is very useful because the GDC-SDO disks are connected locally via InfiniBand to the MPG's computing resources. The MPG uses the Pegasus Workflow Management System (http://pegasus.isi.edu) for running helioseismology workflows. The main inputs to the GDC-SDO helioseismology workflows are HMI line-of-sight Dopplergrams, line-of-sight magnetograms, or continuum intensity images. The Pegasus workflows were executed about 20 000 times over the past year for a variety of different projects.