Working with data from the NASA STEREO Heliospheric Imager an essential guide

Chris Davis, Danielle Bewsher, Jackie Davies, Steve Crothers,

Rutherford Appleton Laboratory,

First issued: April 2007, last updated: July 2008.

Introduction

A large and ever-increasing number of documents exist that describe the inner workings of the NASA STEREO mission in some detail. Those documents pertinent to HI are available via the UK STEREO website (www.stereo.rl.ac.uk) and so no attempt will be made to explain the instruments in that level of detail. Instead, this document is intended for those of you who know what the STEREO Heliospheric Imagers are and want to know how to work with HI images. This document will hopefully provide you with enough information to enable you to interpret the data in a meaningful way.

1. What to expect in an HI image

Images from the two cameras in each HI instrument are quite distinct. The HI-1 camera images the heliosphere closest to the Sun. Each HI-1 camera has a 20˚ field of view centred 14˚ from Sun centre. In contrast, the HI-2 cameras have a much wider (70˚) field of view centred 53.7˚ from Sun centre. These images are distinct from the HI-1 images in that the sunward edge is limited by the curved baffles assembly while the (now redundant) earth-occulter plate covers a rhombic region on the opposite side of the field of view. The Sun is on the right hand side of HI-A images and the left hand side of HI-B images. Solar North is always upwards and HI-A looks towards the eastern limb of the Sun while HI-B looks towards the western limb of the Sun.

Both cameras see many stars, visible down to around 13th magnitude and, without processing, the signal in both is dominated by the scattered light of the F-corona. Cosmic rays are removed onboard the spacecraft and should not appear in the final summed images. Any bright star or planet will saturate the detector and the signal will spill into adjacent rows, forming a bright column. This is further exaggerated during read-out since neither HI camera has a shutter.
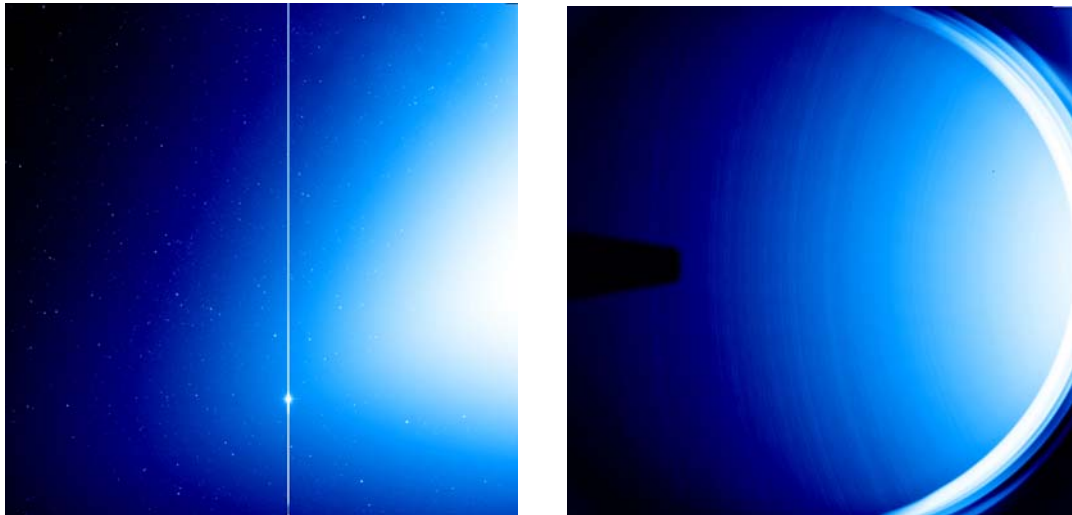
Figure 1. Raw data images from HI-1 (left) and HI-2 (right). The planet Venus in HI-1 is saturating the detector resulting in a vertical line in the image. HI-2 images are quite distinct, with the Earth occulter and baffles both visible in the field-of-view. The two images are not to scale. HI-1 has a 20˚ field of view compared with 70˚ for HI-2.
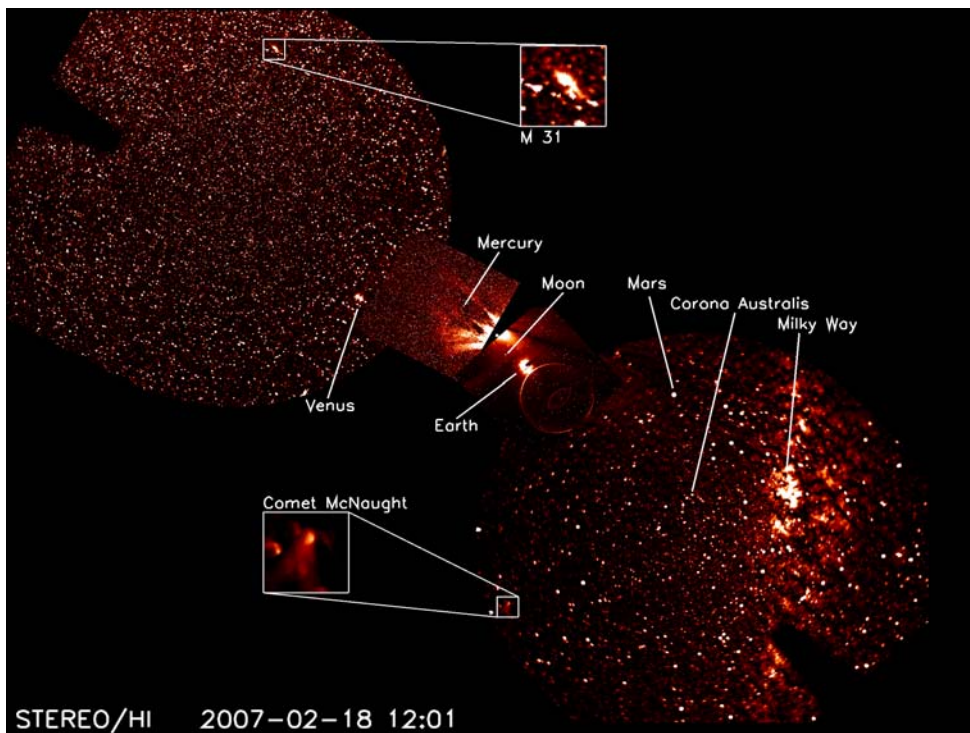

Figure 2. The complete field of view from the two Heliospheric Imagers showing solar wind structures, planets, comets, asteroids and stars. The F-corona has been subtracted from these images. On this scale, the Sun is the size of the small white dot in the centre of the picture.

2. Accessing the data

Data from the STEREO imagers are made available via the UK Solar System Data Centre (UKSSDC). Other outlets for the data are available but may not contain the most up-to-date calibration information. The data and analysis tools referred to in this document assume that the data has been obtained from the UKSSCD and, in the case of processed data, has passed through the analysis described below.

With the many cameras and instruments on board the twin STEREO spacecraft, it is important to identify a specific time interval as there is a limit to the amount of data that can be transferred via the UKSSDC website. If you require large sequences of data (for long-term studies for example) please contact the UK STEREO team (details below).

In order to help identify periods of interest, we make daily and monthly movies of the HI data available via our website as well as an event list which picks out periods of interest.

Once you know what instrument or date interval you are interested in obtaining data from, it can be obtained on-line from the UKSSDC website at;

http://ukssdc-www.stp.rl.ac.uk/solar/stereo/data.html

The data are stored as fits files (.fts) and can be selected by instrument and date interval before being retrieved via ftp. Two types of HI data are currently available, known as L0 and L1 data. L0 files contain raw, uncalibrated data while the L1 data has been processed using the secchi_prep routine detailed below in order to apply a flat-field, alignment and shutterless corrections.

Data from all STEREO instruments can be down-loaded from the UKSSDC and this document will endeavour to cover the calibration and viewing of all data in a generic fashion. However, since this document is tailored for users of the Heliospheric Imager It should be assumed, that if no specific instrument is named, we are talking about HI.

Figure 3. Access to STEREO data via the UKSSDC website

3. The HI event list

If you would like to know when Coronal Mass Ejections (CMEs) or Corotating Interaction Regions (CIRs) have been observed in the data, an event list for HI is available at www.stereo.rl.ac.uk in which the most significant outflows observed in the HI cameras are presented.

This event list was created by taking the central rows of differenced HI images and plotting them as a function of time. In this way, it is possible to track the propagation of features through the HI cameras. Figure 4 shows an example of such a plot.
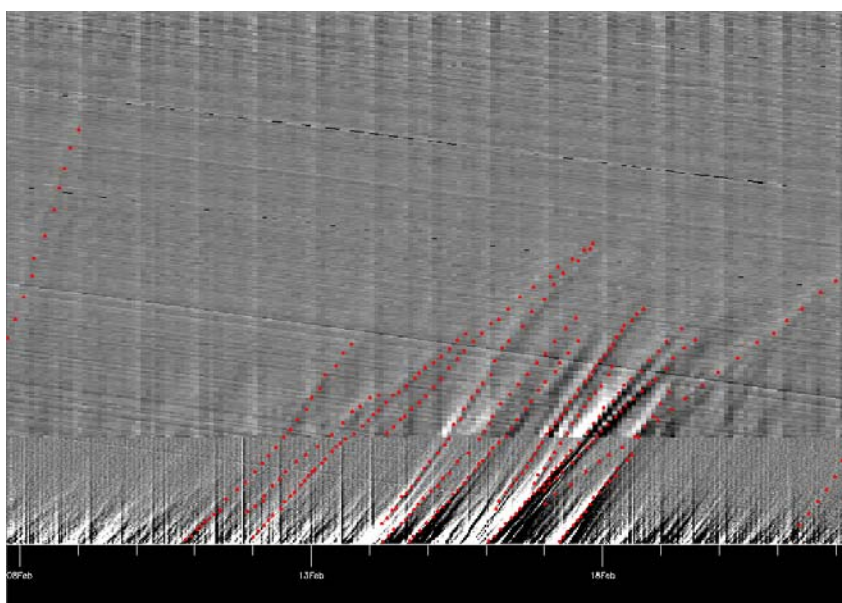

Figure 4. An example of significant outflows identified in STEREO HI data
By plotting the distance from the Sun against time, it is possible to estimate
the speed and direction of features of interest. The red points identify features
of interest

A CME travelling across the HI field of view will have an apparent acceleration depending on its direction of propagation, as illustrated in figure 5. If it is assumed that the CME is travelling at a constant speed, the shape of the track can be used to estimate the speed and direction a CME.
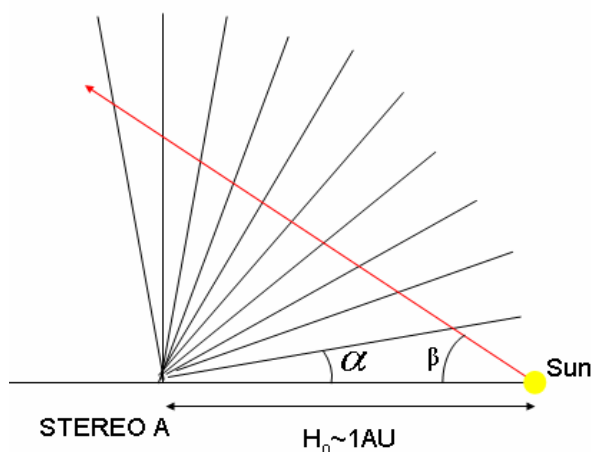
Figure 5. A CME travelling through the HI field of view at a constant speed (represented as a red line in this diagram) will spend different amounts of time in each pixel (of constant angular extent, represented by the black lines). In this example for instance, the CME will appear to accelerate as it crosses into the centre of the field of view and decelerate as it moves to the opposite side of the image. This apparent acceleration can be used to estimate the speed and direction of all transient features crossing the HI fields of view.

The apparent elongation of a CME from the Sun, α, is related to the speed of the CME, υ, and the spacecraft-sun-CME angle, β, by the formula

$$\alpha(t) = \arctan\left[\frac{vt\sin(\beta)}{H_\circ - vt\cos(\beta)}\right]$$

(as derived by Sheely at al, JGR, 1999). Once the speed and direction of an outflow have been calculated in this way, estimates of the launch time (assuming instant acceleration) and predicted time of arrival at the Earth are made and all this information is summarised in an event list, containing links to the relevant movies for each event. This list is constantly being updated and should not be considered final. If you find an event in the data which you feel should be covered by the list, please contact us.
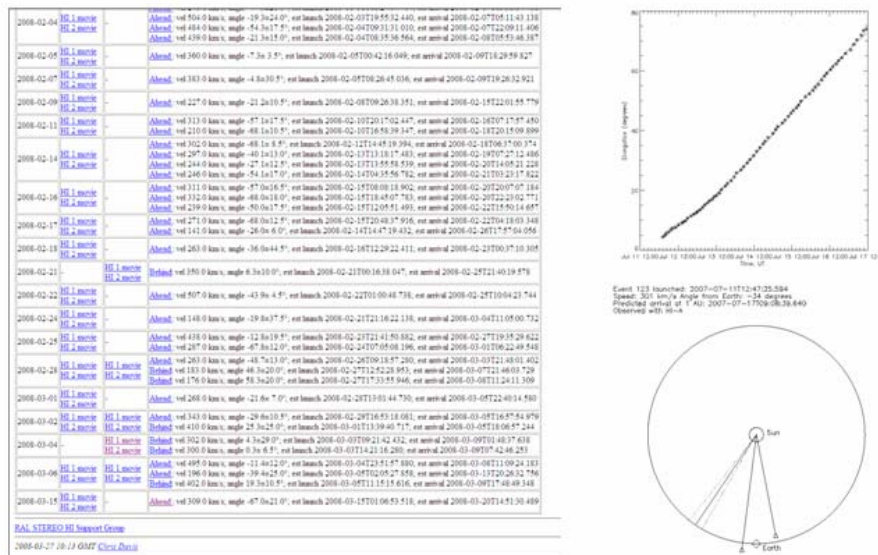
Figure 6. A section of the event list showing an example of the best-fit to a series of points and the resulting CME direction with errors (solid line flanked by dotted lines), the position of the Earth and the two STEREO spacecraft in relation to the Sun.

5. Quick-look viewing of the data

If there is a particular time interval you are already studying, you may wish to look at the quick-look HI movies to see if there is anything worth further study. This section lists the various types of movies that are available via the UK website and ways of examining the data in more detail once you have identified the region of interest.

i) Web movies

Daily, monthly and entire movie sequences are available at;

http://www.ukssdc.ac.uk/solar/stereo/movies/MOVIES/

and these can be accessed via the UK stereo web-pages (www.stereo.rl.ac.uk). These movies provide a means of having a quick-look at the data before embarking on a more detailed analysis. In general, there are three types of movie for each time interval;

difference movies (which use a running difference between frames). These grey-scale movies are characterised by filenames containing the string 'diff'. Changes between frames in such movies (such as the movement of a CME) show up as light/dark patches.

Background subtracted movies In the daily movies, the background is characterised by taking the minimum value in each pixel for all frames in the movie. This background is then subtracted from each frame. In this way, moving objects such as CMEs are retained in the image while the much less variable (but far brighter) F corona is removed. Monthly movies are the concatenation of several daily movies. The different backgrounds subtracted throughout these movies may cause the apparent intensity of events to vary if they span several days.

Histogram equalised movies These movies are the same format as difference movies but displayed with the colour contrast displayed using 'histogram equalised' binning. The advantage of this method is that it enhances faint features but the contrast between objects is less easy to determine since the scale is no longer linear.

All images undergo flat-fielding and shutterless correction as applied by the solarsoft routine *secchi_prep.pro*.

ii) Stand-alone fits file viewers

Analysis of the data is supported in the IDL programming language through the freely available solarsoft libraries. While these provide a comprehensive suite of routines for analysing data, other stand-alone applications may prove useful in obtaining further information about a particular image.

The software most favoured by the UK HI team for this purpose is 'FV' (fits viewer). This software is freely available from the Goddard Space Flight Centre at;

http://heasarc.nasa.gov/lheasoft/ftools/fv/

which supports windows, unix and mac OS-X platforms. Details on how to use this software will not be covered in this document since the above website contains a comprehensive user guide.

6. Working with the data in more detail

i) File naming conventions

Each HI image is the sum of many shorter full-resolution (2048x2048 pixel) exposures that have each been scrubbed for cosmic-rays (as otherwise the final summed image would be dominated by cosmic ray tracks). These summed images are then binned to 1024x1024 pixels before being transmitted back to Earth. These summed images can be identified by the string 's4' in their names. A smaller number of full resolution (2048x2048) single exposures have been downloaded for calibration purposes. These short time exposures (of the order of a few seconds) contain very

little signal from solar wind transients and so are less useful for solar science. They can be identified by the string 'n4' in their filenames. In general, s4 images should be used for all science applications.

ii) Solarsoft

While other data analysis packages such as MATLAB now support fits files, STEREO analysis is supported under IDL as part of the comprehensive solarsoft libraries. These can be downloaded from;

http://www.lmsal.com/solarsoft/index_old.html

where comprehensive instructions can also be found on how to install this software. While it is unnecessary to download everything within solarsoft, you will need to download the STEREO libraries if you want to calibrate STEREO data and have access to the latest analysis tools (such as creating 3D images from EUVI data). The Heliospheric Imager forms part of the Sun Earth Coupled Coronal Heliospheric Investigation (SECCHI) instrument package and so the calibration of HI data is carried out with secchi_prep.pro. This will be detailed below.

iii) Loading the data into IDL

While individual images can be read into IDL using the solarsoft routine readfits.pro, it is more often the case that multiple images need to be loaded into IDL. This can be done with either sccreadfits.pro or mreadfits.pro each of which takes a filelist argument and loads all the listed files into IDL in one step, storing the data in a data cube in the workspace. The command;

IDL> files=file_list('/data/hi_1/20070111')

lists all available fits files in the directory /data/hi_1/20070111 in the variable 'files'

This can then be read into IDL using the command;

IDL> data_cube = sccreadfits(files,hdrs)

This reads in multiple fits files listed in 'files' and stores the image data in the variable 'data_cube' which contains a cube of data of dimensions 1024x1024xn (for 'n' standard 1024x1024 images)

If your file list contains both s4 and n4 images (1024x1024 and 2048x2048 arrays), a useful sequence of commands to ensure only the s4 images are considered would be;

Read the file headers only using the command

IDL> junk=sccreadfits(files, hdrs, /nodata)

Find which files contain arrays of 1024 pixels using the command

IDL> ss=where(hdrs.naxis2 eq 1024)

Read this subset of images into the final data cube

IDL> data_cube=sccreadfits(files[ss],hdrs)

These data can then be viewed with IDL routines such as tvimage.pro or specific viewing tools contained within solarsoft such as plot_image.pro. For example;

IDL> plot_image, data_cube(*,*,1)

will display the first image in the data cube. Subtle features in the data can be drawn out by plotting the data in a histogram equalised form. To do this, the above command can be modified to;

IDL> plot_image, hist_equal(data_cube(*,*,1))

Subsets of each image can be plotted by replacing the asterisks with specific pixel values (remember IDL counts from 0 to 1023 for a 1024x1024 image).

iv) Calibrating the data using secchi_prep.pro

Like most datasets, HI images need to be calibrated before they can be used for detailed science. The raw (L0) data needs to have corrections applied to account for;

flat-field – a gain-map is applied to compensate for any uneven responses of pixels across the CCD

shutterless readout – Without a shutter, some parts of the image are exposed more than others due to the finite time taken to read the information from the CCD array. As the image is read out, the image is smeared vertically. The effect is known and can be compensated for mathematically.

missing data – Missing data blocks are identified and replaced with NaNs

bad files – known bad files are identified from a 'bad files' list and excluded from further analysis unless specifically requested

saturated columns – Each pixel in the array can only receive a finite signal before it overflows. If this happens the CCDs have been designed to bleed signal vertically, creating vertical columns of saturated pixels. In extreme cases, some of this signal can bleed across columns via the readout register. If a saturated column is detected therefore, it is replaced with NaNs, as is a small number of columns either side of it in order to remove any columns potentially polluted with the saturated signal.

updated pointing information – The raw data files contain nominal pointing information derived from calibrated average instrument offsets from the spacecraft pointing. Updated pointing information has subsequently been calculated using the starfields contained within images. Running a file through secchi_prep replaces the nominal values with the best information on the instrument pointing (see also Section X HI calibrated pointing for more information).

updated optics parameters – Nominal information about the camera optics (estimated from pre-flight calibration) are contained within the raw data file headers. Updated values for these parameters have been calculated from the star-field. Running a file through secchi_prep replaces the nominal values with

the most up-to-date information on the instrument optics (see also Section X HI calibrated pointing for more information).

 To calibrate raw (L0) HI images, use secchi_prep.pro (in the STEREO tree of the solarsoft installation). The command;

IDL> secchi_prep,files,headers,images, savepath='save_directory',/write_fts

will take the list of files contained in the filelist 'files', load the data into IDL, carry out the default corrections and write the calibrated data back out (in this case) into a directory named 'save_directory'. In the absence of the 'write_fts' switch, the calibrated data is stored in a data cube (data_cube = secchi_prep,files,images,headers)

v) Subtracting the background

The F-corona produces a more-or-less constant background which dominates the images and prevents transient features such as CMEs from being observed. To remove the background a sequence of images (at least three) is needed. A good approximation to the F-corona is to find the minimum value occurring in each pixel over the sequence of images. Since stars move across the image from frame to frame, these are removed. If larger scale features, such as the Milky Way, occur in the images, a longer sequence of images may be required to correctly remove the background. The command;

IDL> new=min_filter(data_cube,bkgd=bkgd)

will calculate the minimum value for each pixel, storing the background in the variable 'bkgd' while subtracting this from the data to form the data cube 'new'. This routing takes any bad data (e.g. missing blocks, saturated columns) into account. Since most of these have now been trapped and corrected for in secchi_prep.pro, min_filter becomes the equivalent of simply taking the minimum value along the time dimension of a data cube, as with;

IDL> bg = min(data_cube, dimension=3)

This will take the minimum value along the third (counting from zero) dimension and store these values in the variable 'bg'.

A more stable, but slightly less accurate background can be obtained for reasonable sized data tables is;

IDL> bg=median(data_cube, dimension=3)

Subtracting either of these background values from each image in a data cube will reveal transient features such as CMEs. To do this, first, find out the size of the data cube;

IDL> sz=size(data_cube)

Then replicate the background so that it is the same size as the data cube and can be subtracted;

IDL> data_cube_nobg = data_cube – rebin(bg,sz[1],sz[2],sz[3])

One unexpected finding of the STEREO spacecraft moving in heliocentric orbits is that the pointing of HI instrument on the Behind spacecraft contains many slight movements (most likely due to the impact of dust particles on the instrument as it is facing the ram-direction on the Behind spacecraft). These offsets range in size from a fraction of a pixel to tens of pixels and are seen most obviously as 'jumps' in the position of the background starfield. These offsets make a simple background subtraction less effective, since gradients in the F-corona mean that an offset of around 1/3 of a pixel, during a simple background subtraction such as described above, is enough to leave a residual background signal that is of an equivalent intensity to the transient solar wind features.

By using the background starfield to obtain accurate pointing information for each image, it is possible to create backgrounds that take this mis-alignment into account before calculating the minimum value seen in each pixel. We are in the process of calculating corrected backgrounds (averaged over periods varying from one day to twenty seven days) that will be stored in a database alongside the HI data. It is our intention that the best available background will then be retrieved using a command such as;

```
IDL> bg=scc_getbkgimg(header,/daily)
```
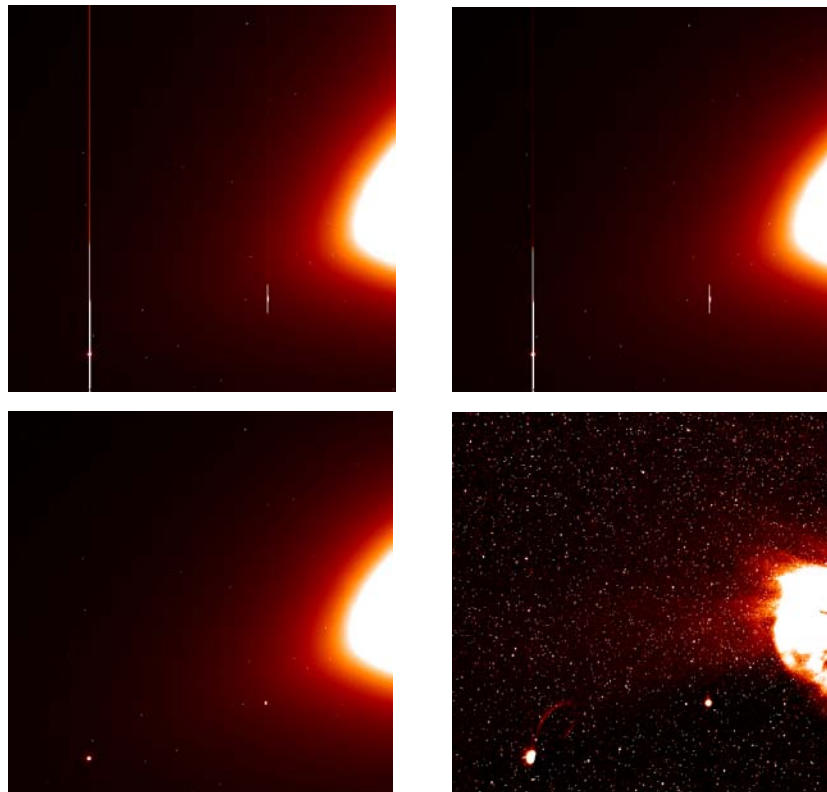
Figure 7. A sequence of images from HI-1A showing the effects of various stages of image processing. The raw image (top left) is corrected for shutterless operation and flat-field using secchi_prep, producing the image top right. The residual 'blooming' caused by saturation around the planets Venus and Mercury is removed (bottom left) and finally the background is subtracted to reveal the CME (bottom right)

In the images shown in figure 7, the blooming caused by saturated pixels around the planets has been removed before the background was subtracted. To do this, the following command can be used;

images = rm_bloom(filelist,headers, images, /fill)

We recommend that this is only done for aesthetic reasons and not for scientific data.

The baffle system is visible in HI-2 images. To mask out these regions of the images use the following commands;

```
IDL> mask=get_smask(header)
IDL> outside=where(mask eq 0, noutside)
IDL> if noutside gt 0 then image[outside]=!values.f_nan
```

You can use get_smask for all instuments that support masks.

secchi_prep introduces NaNs (Not a Number) into the data arrays. If you wish to replace these with a 'bad' value of your own (say, -999), this can be done with the following set of commands;

```
IDL> w=where( not finite(data), ctr)
IDL> if ctr gt 0 then data[w] = -999
```

vi) Identifying stars and other features in your data.

Once the files have had their header information updated during the secchi_prep process, this pointing information can be used to identify objects within the HI fields of view. The most convenient way to do this is using the world coordinate system (WCS), the routines for which are contained within solarsoft. WCS is a method of embedding coordinate information within the header of a FITS file. There are three principal IDL routines;

FITSHEAD2WDC – this forms a WCS structure

WCS_GET_COORD – which converts pixels into coordinates

WCS_GET_PIXEL – which converts coordinates into pixels.

So, for example if you wanted to extract coordinates over an entire image;

```
IDL> wcs=fitsheader2wcs(header)
IDL> coord=wcs_get_coord(wcs)
```

where coord has dimensions (2, Nx, Ny)

To extract celestial coordinates of a particular star (R.A., Dec.)

```
IDL> wcsa = fitshead2wcs(header, system='A')
IDL> coorda = wcs_get_cood(wcsa)
```

vii) Determining the relative positions of the STEREO spacecraft

The position and orientation of the two STEREO spacecraft, as well as the Sun and the planets can be determined using a software package called SPICE. SPICE is a software package for ephemerides and spacecraft attitude data. SPICE is distributed as part of the STEREO tree in solarsoft, along with the ephemerides. To work with SPICE, you will need to start IDL in 32 bit mode;

csh> idl -32

There are three main routines;

GET_STEREO_COORD - retrieves the position of an object within a specified coordinate system.

GET_STEREO_LONLAT – retrieves the position of an object as angles

CONV_STEREO_COORD – converts between systems

CONV_STEREO_LONLAT – converts angles

So, for example, to find the position of the behind spacecraft on the HI-2 image from Ahead;

IDL> wcs = fitsheader2wcs(header)
IDL> coord = get_stereo_lonlat(header.date_obs,'Ahead', /degrees $,
            system='HPC',target='STEREO Behind')
IDL> pix = wcs_get_pixel(wcs, coord[1:2])

viii) Creating movies

Daily and monthly movies are routinely generated and made available via the web. The software used to do this can be applied to generating movies of your own devising. Contact a member of the UK STEREO support team if you wish to use this facility.

ix) Star removal

The routine hi_remove_starfield.pro can be used to remove the starfield from HI images. It is an image processing technique, no prior knowledge of exactly which stars are in the field-of-view is necessary, i.e. there is no input from a star catalogue. Peaks (stars) in the data are identified by analysing the del^2 field of the image. A successive overrelaxation (SOR) method is then used to reconstruct only the pixels where the stars were identified.

An example of how to use hi_remove_starfield is:

IDL> recon = hi_remove_starfield(data)

where recon is the HI image with the stars removed, and data is the original HI image with the stars in.

An optional threshold keyword can be used to stipulate the minimum intensity at

which a peak is identified. The default is 1.0, which is reasonable for data converted to DN/s by secchi_prep.pro. An example of how to use hi_remove_starfield with the optional threshold keyword is given below.

IDL> recon = hi_remove_starfield(data, threshold=0.5)

Note, the lower the threshold, the fainter the removed stars will be, however, this will require more pixels to be reconstructed, so the process will take longer. Furthermore, a particularly sensitive threshold runs the risk of nibbling away of actual structure. This can be analysed by looking at the removed field with high contrast, e.g.,

plot_image, data-recon, min=-0.1, max=0.1

Another optional keyword is /edgefix. This removes peaks/stars from pixels around the edge of the image. This is generally not a problem with full frame HI images, but if the starfield is removed on a smaller cutout region, then unremoved stars at the edge of the frame can bleed into the frame. An example of how to use hi_remove_starfield with the edgefix keyword is given below.

IDL> recon = hi_remove_starfield(data, /edgefix)

Obviously, both keywords can be used in conjunction, e.g.,

IDL> recon = hi_remove_starfield(data, threshold=0.5, /edgefix)

x) HI calibrated pointing

In the L0 HI data, the headers contain the HI pointing derived from applying calibrated average instrument offsets to the spacecraft pointing. As of August 2008, the HI headers are updated with the best calibrated pointing available to the HI team, when the data are prepared using secchi_prep. If a user has HI data that have already been prepared prior to August 2008 - and want to update the information in the headers without redoing all of the data preparation, then they can use hi_fix_pointing.pro. An example of which can be found below.

If 'header' is a secchi header structure (or an array of secchi header structures) then you can call hi_fix_pointing like this

IDL> hi_fix_pointing,header

As well as updating the relevant keywords with calibrated pointing and optical parameters, it also populates a new keyword 'ravg'. ravg is a measure of how good the calibration is. It is the average radial error (in pixels) between positions of stars in the image and the position of the stars in the star catalogue. We have decided that a ravg of less than one is good, and ravg between 1 and 5 is acceptable and a ravg greater than 5 is bad. hi_fix_pointing applies calibrations where ravg is less than or equal to 5. If you need better accuracy, then you can specify your own criteria for ravg in the following manner:

hi_fix_pointing,header,ravg=1

Where (for what ever reason) there is no calibrated pointing information for an HI fits file, the calibrated instrument offsets are used to derive the HI pointing. In these cases, the following values of ravg will be used for the reasons stated.

-999 - No pointing calibration has been applied. No attempt to find a pointing claibration has been made. This is the value that will be found in the L0 data

Once the data has been prepared using secchi_prep or the header have been run through hi_fix_pointing, then the following error codes will be used.

Error codes from the calibration minimisation code
-993 - Can't match any of the stars in the catalogue to any peaks (stars) found in the data
-992 - No peaks (stars) found in the data
-991 - No catalogue stars found in the field-of-view

Error codes from the application of the pointing calibration (hi_fix_pointing)
-883 - ravg is greater than 5 (or different criteria that you have specified)
-882 - There is no pointing calibration for the FITS file
-881 - There is no pointing calibration file found for the day of your FITS file

In the last 2 cases, either (a) you are using up to date data and the pointing calibration file hasn't yet made it into solarsoft or (b) the data has been deemed to be 'bad' by the HI team and there is no calibration available.

A paper (Calibrating the pointing and optical parameters of the STEREO Heliospheric Imagers; Brown, Bewsher & Eyles) has just been submitted to Solar Physics which describes the methodology and results of the calibration.

8. Contacting the UK STEREO support team

For all help and advice with using STEREO HI data, please contact the UK STEREO support team by emailing stereo_support@ukssdc.ac.uk.