

# AUTOMATING THE QUALITY ASSURANCE PROCESS FOR GONG+ DATA

This document describes the operation of the Automated Image Rejection (AIR) program package and how it interfaces with the GONG+ VMBICAL image calibration software. Information useful to the scientist needing to understand how AIR performs its tests, and for the programmer wanting to utilize it, or modify or extend its operation is included. A basic familiarity with the structure of the VMBICAL data tree (wa0 and wa1 directories and their contents) will be required to understand much of this discussion. AIR is normally triggered and run automatically by scripts within VMBICAL, but unusual circumstances (e.g. errors caused by unusual data events and testing modifications to the programs) may arise making it useful to run some components manually. Information is given for the manual operation of the components of AIR, including the use of options not needed in normal use by VMBICAL.

## 1. Introduction

An Automated Image Rejection (AIR) program for image Quality Assurance (QA) has been developed for use in processing GONG+ data. It includes checks for detecting bad images due to thin or partial cloudiness, obstructions in the field of view (e.g. low altitude aircraft at Big Bear), and abnormal instrument operation. Data from a test period were reprocessed through our image calibration pipeline using the AIR program for image QA decisions. The dataset produced with AIR had a slightly higher duty cycle than the archived data. Comparison of the resulting l-nu diagrams, peak mode-frequency distributions, and flow maps derived from ring diagrams show good agreement with the archived data products.

The current QA process for the GONG+ image calibration and processing pipeline requires much human interaction. AIR was developed to automate the bulk of this QA task. The goals for this program include:

- reducing the level of human interaction needed to process data
- use objective, repeatable criteria for image selection/rejection
- to improve duty cycle

Image quality tests performed by AIR include

- checks of instrument status keyword values in the FITS file header of each image to verify the instrument is configured for normal data collection and is operating properly.
- checks for active tracking and image stability.
- checks for image signal and noise levels.
- checks for large temporal variations between images.

## 2. AIR Interface and ALGORITHMS

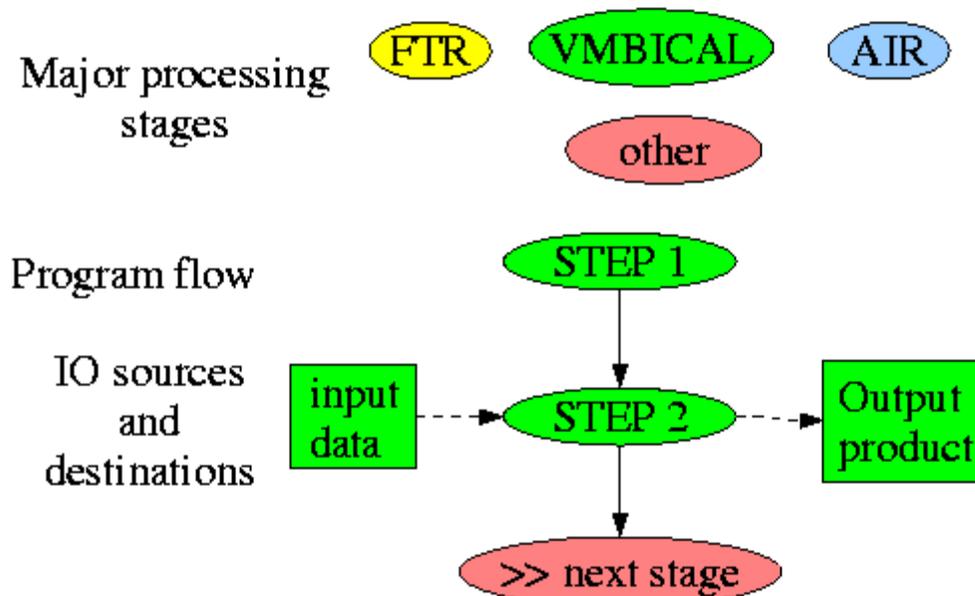
The components of AIR are normally run in an automatic mode from scripts that are part of VMBICAL. To run modules of AIR manually, or to construct the input files needed to run AIR from the scripts, the following information will be useful.

VMBICAL obtains information from image headers extracted during the Field Tape Read (**FTR**) phase of data ingest and creates a directory structure for a set of site days. Header parameter files for several items in the image headers are created within this tree in the PREPDISK step of VMBICAL. VMBICAL then gets compressed image files and places them in a <SITEDAY>.stz directory. They are uncompressed and moved to <SITEDAY>.raw directories for each site day by the STZ2RAW script of VMBICAL.

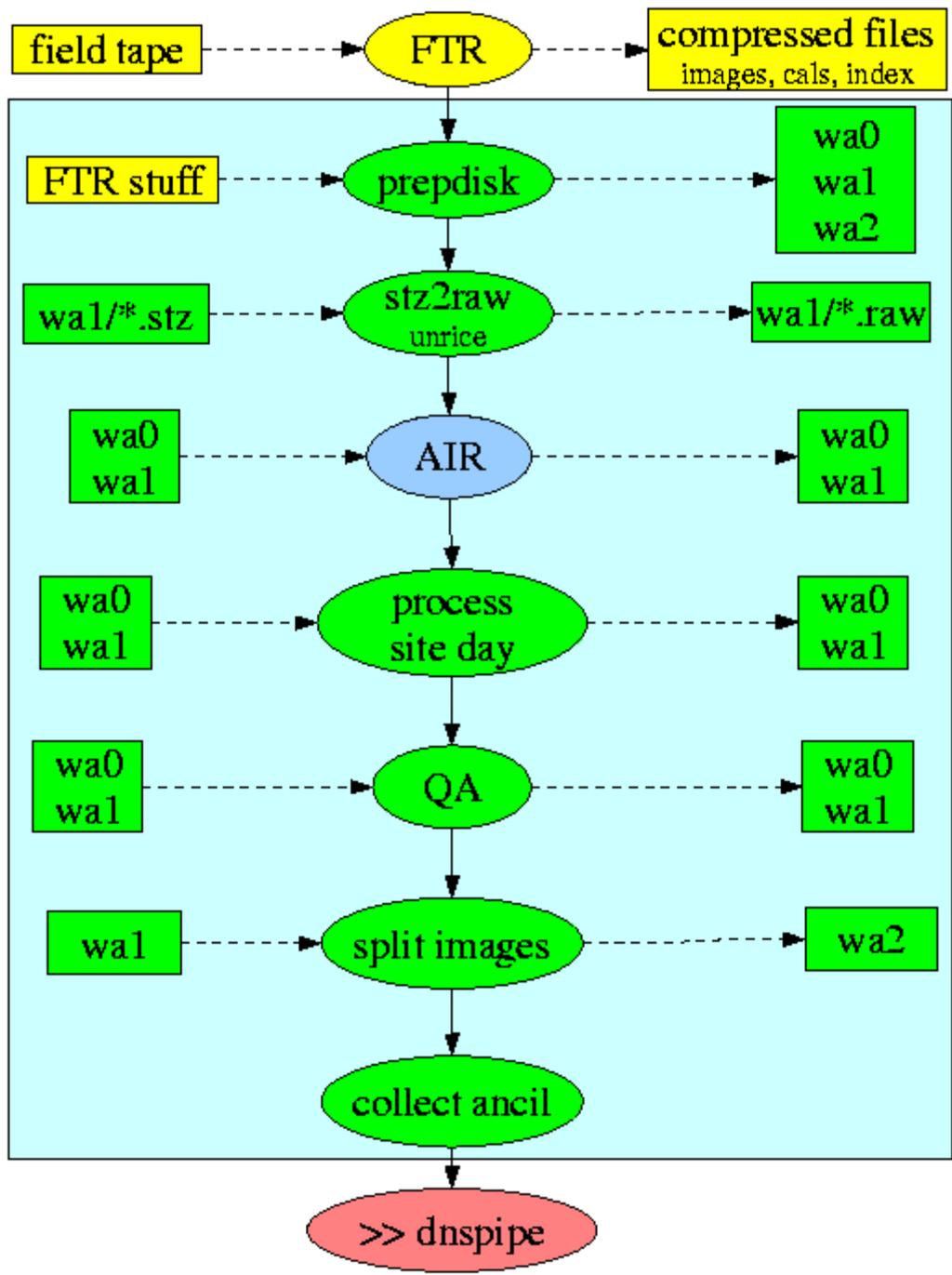
After filling the .raw directories VMBICAL stops to run AIR.

AIR consists of four components.

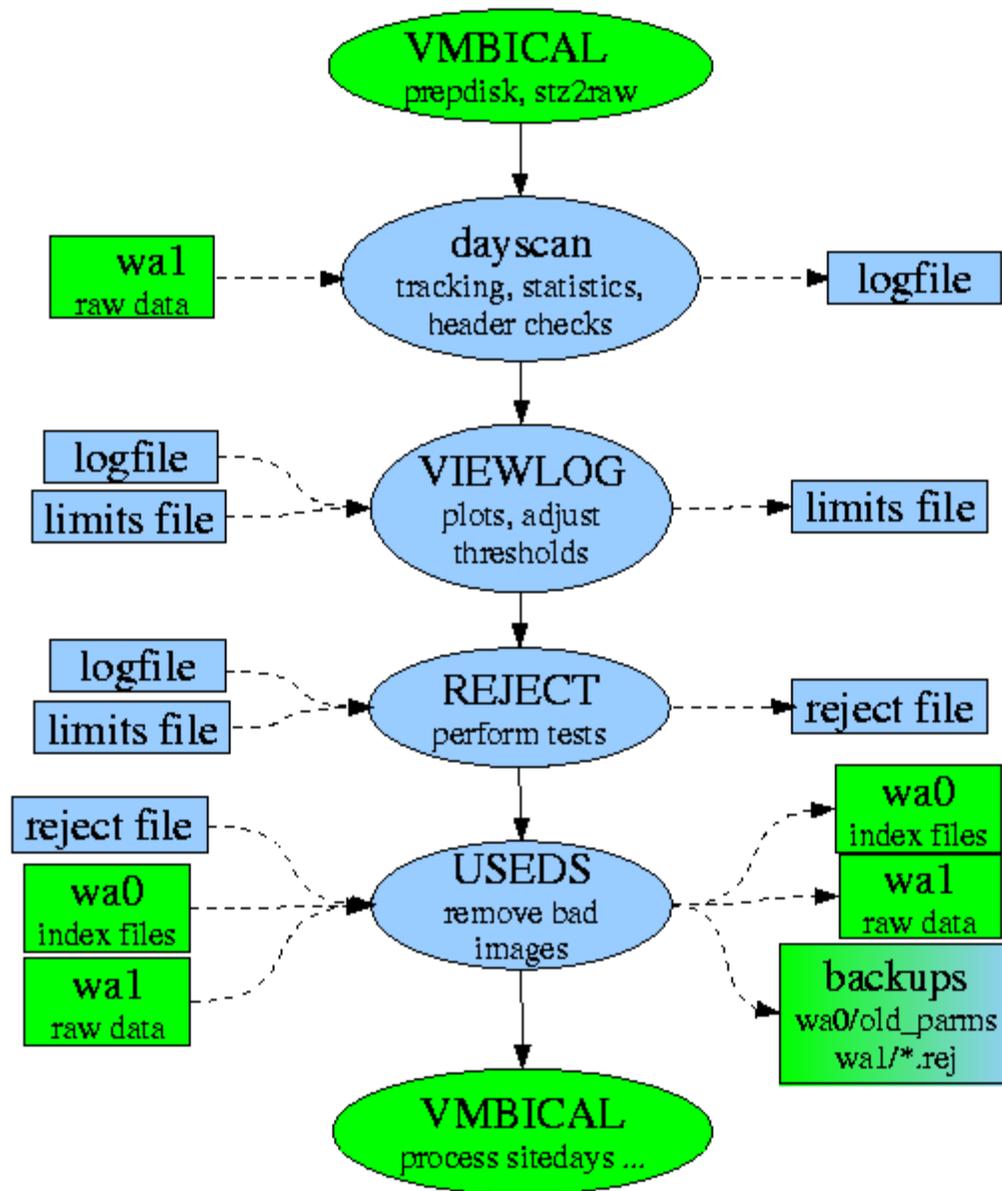
- DAYSCAN examines raw images and creates a logfile with image statistics
- VIEWLOG reads the logfile and produces plots showing selected image properties and the threshold values used for image acceptance. An opportunity for human intervention is allowed here.
- REJECT reads the logfile and applies several tests (described below) for acceptance/rejection and produces an image disposition report file.
- USEDS reads the REJECT report file and applies the decisions to the data in the VMBICAL data tree.



Conventions used in the flowcharts



A high level view of the processing stages of VMBICAL and AIR.



## AIR internal structure and interaction with VMBICAL

The processing stages of AIR and their interaction with VMBICAL.

### 3. DAYSCAN

DAYSCAN is intended to process one site day at a time. It reads each image file and calculates a variety of statistics for the image as described below. These statistics and several values read from the image headers are written in a logfile for each site day. This logfile is ascii and can be examined if desired, but its structure has grown over time and it should not be considered a human oriented document.

#### 3.1 User Interaction for Running DAYSCAN

- Start IRAF. Be sure the DAYSCAN task is defined.
- cd to a working directory. This is where the logfile will be produced. When prompted for directory locations the path can be absolute or relative to this working directory. It is recommended not to use the raw image directory itself as the working directory. If there is a need to use the current directory as a source or destination directory for data or output it should be referenced explicitly as './' rather than '.'. Similarly, a subdirectory within the current directory should be given as './' rather than 'directory'. Any directory name can always be given by its absolute path.
- Start DAYSCAN.
- Give the path to the raw image directory when prompted. It doesn't matter whether or not there is a trailing '/'. The number of fits files found in that directory is reported.
- When prompted for first image enter 0 to select the entire set of images in the directory. If a number > 0 is given, it is taken as the beginning of a subrange. You are then prompted for the file number of the last image to be included.
- The plane pair prompt selects how to obtain the difference between modulator states. An image is produced as the difference between corresponding planes (4-1, 5-2, 6-3) in the two modulator states. These form 'pseudomagnetograms' (sometimes just 'magnetograms' in the source code comments) for their strong resemblance to actual magnetograms. Selecting '0' uses the averages of planes 1:3 and 4:6. This reduces the sensitivity to actual solar magnetic activity, which can masquerade as image noise in the whole disk statistics, and is the preferred option. The pseudomagnetograms are sometimes called modulator (difference) images in the source code. This should not be confused with the modulation images produced later in VMBICAL.
- The logfile prompt controls the location of the logfile. The 'p' option will take the base name from the raw image directory to form the logfile name with an extension of '.l###'. ## is the plane pair combination used for the pseudomagnetograms (00, 41, 52, or 63). You can also give the filename explicitly (any response > 2 characters in length). You can also specify that no logfile be generated, such as when making a special run to generate the intermediate images (discussed next) for examining a period of interest within the site day after the main logfile has been produced. This option was frequently used during early development and the capability was left in the program.
- The 'save images' prompt refers to the pseudomagnetograms and 1-minute difference images produced for each raw image. This has been useful in the past for testing, debugging, and looking at interesting images. But you will not want to save them during normal processing so enter 'n', or just press return. If you do want to save them, enter the path to the directory where they are to be saved as either '/absolute\_path' or './relative\_path'. A response of 'y' will cause them to be saved in the current directory. You are then prompted for which type of image to save-- the magnetograms, the 1-minute differences, or both.
- If a subrange of images was selected earlier, or if intermediate images are to be saved, it is assumed that testing, debugging, or detailed examination of data is being done and additional options may be offered.

Most recently this was used for debugging some issues with the limbfinder. These issues have now been addressed. But the logic has been left in place as a starting point for the day that something else may need to be investigated. If you find yourself getting this prompt 'limb debugging: none, (y)es, (v)erbose' enter 'n', or just RETURN to avoid triggering a flood of output and frequent pauses during processing to allow you to examine that output.

- DAYSCAN starts running and needs no further input.

### 3.2 The Opening Dialog of a Normal DAYSCAN Session Run Manually

User input is underlined.

```

cl> dayscan
dayscan/AIR v 1
full pathname to raw image directory? /data2/wa1.le040803/LE040827.raw
516 files found
first image (or 0 for all)? 0
lane pairs for pseudomagnetogram? 4-(1), 5-(2), 6-(3) or 0 0
logfile or use (p)ath or (n)one for results?
p
opening log file le040827.raw
save images: don't, (y)es, or [/absolute or ./relative path]) n
intermediate images will not be saved
processing level: raw= yes, lmbfit= no
LE 2004/08/26 23:50
/data2/wa1.le040803/LE040827.raw/LE040826235016.fits    5    3    900 1024    6
  1    1  2350 1430   16.0   60.1    0.0    0.0  441.4  522.6  373.3 S.....
  2    2  2351 1431   16.1   59.1   -0.1   30.1  441.4  522.6  373.4 .....
  3    3  2352 1432   18.5   62.9   -2.4   31.5  441.3  522.5  373.4 .....
...

```

These responses can be collected into a 'dayscan\_in' file to be redirected to STDIN, as is done with the VMBICAL scripts. [An IRAF script 'runds.cl' performs a similar function and can be used for testing purposes. Though possibly obsolete, it is still included with the AIR package.]

### 3.3 A typical input file

```

/data2/wa1.le040803/LE040827.raw
0
0
p
n

```

With this input DAYSCAN will produce two files in the working directory: LE040827.raw.l00\_a  
LE040827.raw.l00

They are almost identical. LE040827.raw.l00\_a is a temporary copy of the logfile. LE040827.raw.l00 is the final copy containing information inserted at the beginning (for later steps of AIR) that isn't readily available until DAYSCAN completes operation. The \_a file should be deleted.

### 3.4 Checks and Calculations Performed in DAYSCAN

#### 3.4.1 Image Header Parameters Read by DAYSCAN

FITS Parameter Name	IRAF name or data type and comment
IRAFNAME?	(i_pixfile)
BITPIX?	(i_pixtype)
NAXIS	(= i_naxis)

```

NAXIS1      (= i_naxis1)
NAXIS2      (= i_naxis2)
NAXIS3      (= i_naxis3)

SITE        (str)
DATE-OBS    (str)
TIME-OBS    (str)

TYPE        (i) (-> flags[2], (used with ITIME == 60))
VELOCITY    (r) (-> logfile)
ITIME       (r) (+.001 before truncate to integer)
GUIDE       (b) (-> flags[3])
CENTER      (b) (-> flags[4])
VALID       (b) (-> flags[5])
MAGNETO     (b) (-> flags[6])
X_CENTER    (r/i) (-> (flags but ignored), logfile)
Y_CENTER    (r/i) (-> (flags but ignored), logfile)
ZENITH      (r) (-> flags[5] if bad (with VALID))
PA          (r) (-> logfile)
BETA        (r) (-> logfile)
ROLL, EROLL (r) (-> logfile)
PITCH, EPITCH (r) (-> logfile)
ROTATOR, EROTATOR (r) (-> logfile)
INSTTIME    (str) (delta TIME-OBS -> logfile) (not yet in field)

```

### 3.4.2 Image Calculations

A pseudomagnetogram (modulator difference image) is produced. This image is the difference between either a specified pair of planes or the average of all 3 planes from each modulator state.

An intensity image (the image representing the first modulator state from the pseudomagnetogram production) is passed to a simple limb finder procedure.

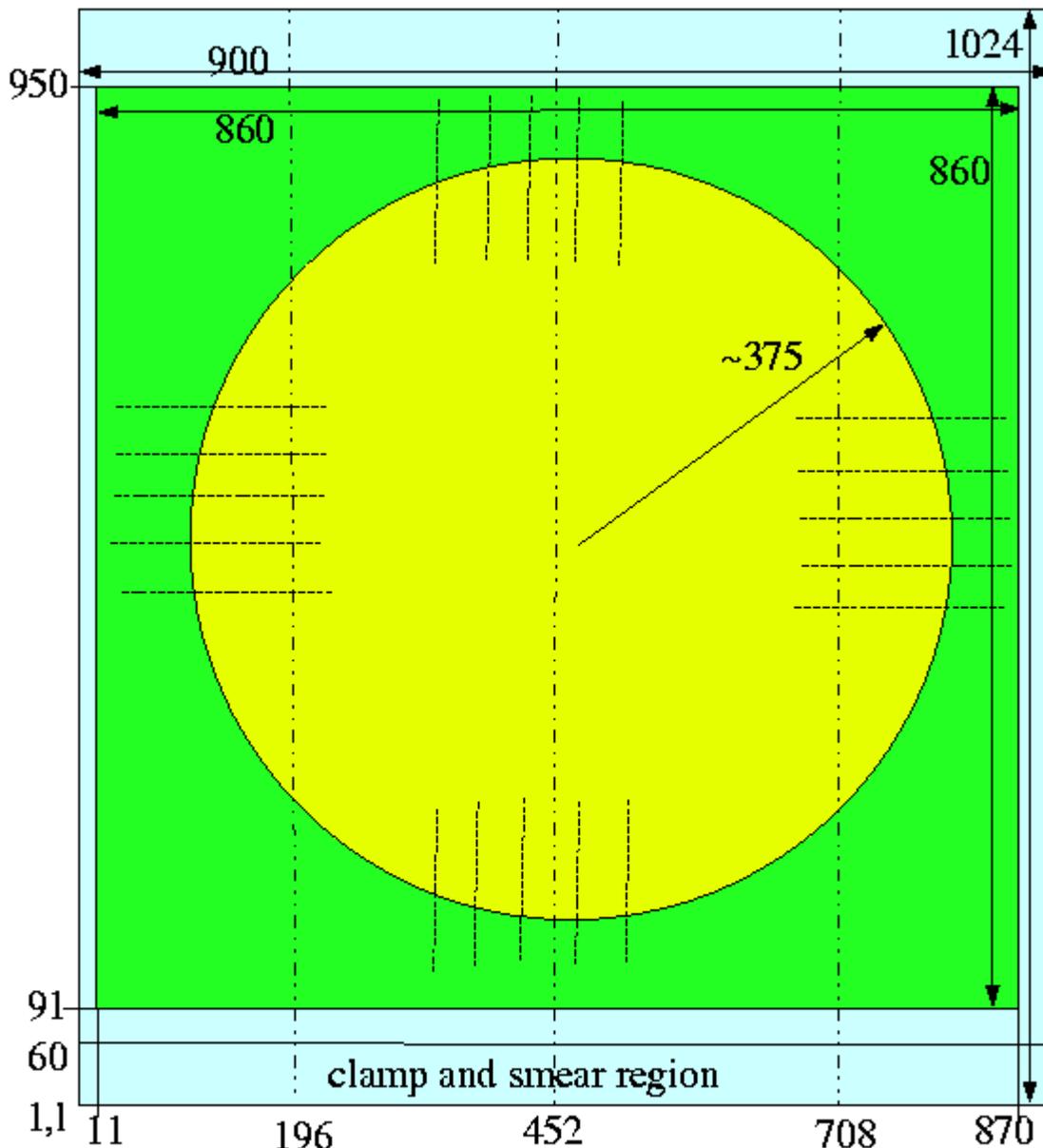
### 3.4.3 The Limbfinder

The figure shows the layout of the raw GONG images on the CCD. The raw image (900x1024) is cropped from the camera CCD (1024x1024) at each GONG site before ever being saved on tape. The leftmost column of the raw image is column 61 on the CCD. The CCD has 4 channels, each 256 pixels wide. The channel boundaries can be seen in columns [196, 452, 708] of the 900x1024 raw images because of decreased sensitivity in the first column of each channel.

VMBICAL crops a fixed region from the 900x1024 raw image to produce an 860x860 image. The camera axis, optical axis, and guider sweet spot are kept sufficiently well aligned that the solar disk will be fairly well centered within the final 860x860 VMBICAL images whenever the instrument is guiding on the sun. This places significant constraints on the location of the disk center in a good image. There is a small amount of image scale variation from site to site and the angular size of the sun varies with season. The solar radius in GONG+ images should always be within the range 365 - 390 pixels.

Several rows and columns passing through the neighborhood of the expected disk center are sampled. For each line, and on both sides of the disk, a maximum gradient search (based on a 6 point neighborhood) is used to locate the limb. A cubic polynomial is then fit on the neighborhood to get the precise location for the limb. Each line scanned produces a pair of points on the limb. The mean of the x values of the limb on each row gives an estimate of the x location of the center, similarly the limb points from the columns give estimates for the y location of the center. Once the center is found from the average of all the independent estimates, each sampled limb point can produce an estimate of the disk radius. As these scans are clustered in 4 regions on the solar circumference (~25% of the total circumference is sampled), only a circular approximation to the sun's disk is attempted. This is quite adequate for the purpose at hand. The location of the center and solar radius obtained from this method routinely give sub pixel agreement with the more rigorous limb geometry used in later stages of processing.

An RMS error estimate for each of the disk center coordinates (x and y) and the radius is calculated. This has proven to be a very strong check on guiding and tracking. Correct tracking, even in early morning or late afternoon, yields RMS values no greater than .25 pixels for each of the three quantities.



GONG image centering and VMBICAL image cropping

### 3.4.4 Image Statistics

Once the location and size of the solar disk are known, the sky background can be masked out and further image calculations and manipulations can be performed.

Whole disk statistics (mean, standard deviation) are obtained for the intensity and pseudomagnetograms. The whole disk intensity statistics are repeated for each of the 6 image planes.

CCD channel intensity averages and standard deviations are calculated. These are taken from a single representative column of pixels from each channel. The columns are selected to be near the channel 1-2 and

3-4 boundaries to provide a more direct comparison between channels. No attempt is made to correct for limb darkening. A significant margin (50 pixels) from the limb is omitted from the each channel's sample column.

A difference image is made from consecutive pseudomagnetograms and its whole disk statistics calculated.

### 3.5 Logfile Structure

The logfile consists of 11 header lines identifying: the site, logfile version, plane pairs, date and time of first and last images in the dataset and number of images, location and size of original dataset, etc. The most important information is collected together in the first line, though most of it is repeated elsewhere in the header. Lines 6-11 are obsolete and should be ignored.

line 1: version number, site id, plane pairs selected, beginning date and time, end date and time, number of images included in the logfile

line 2: a dataset identifier, the number of fits files in the image directory. Generally this will be the same as the number of images reported in the first line.

line 3: site, date, and time taken from the first image included in the logfile.

line 4: processing levels- raw/VMBICAL? FINDLMB already run? Initially there were no specifications or understanding of where AIR should fit in the processing pipeline. DAYSCAN now works only with raw images.

line 5: path to first image included, data type stored in the fits file, number of axes, size of each axis

lines 6-11: obsolete. Threshold values for some of the image selection tests are reported. These tests are now done in REJECT which has its own reject.dat file to contain these values. These lines are needed as filler.

The rest of the logfile is a table of summary statistics, one line for each image.

```
V22 CT p00 20040822 1201 20040822 2129 560
data set ct040822.raw.100 560
CT 2004/08/22 12:01
processing level: raw= yes, lmbfit= no
/sandbox/ladd/wal.ct040714/CT040822.raw/CT040822120116.fits 5 3 900 1024 6
course center limit: 40
limbfind and radius sigma limit: 0.50
pseudomagnetogram mean limit: 150
pseudomagnetogram stdev limit: 200
differencegram mean limit: 100
differencegram stdev limit: 100
 1  1 12 1  721 79955  8781 3249.5 489.2  0.0  0.0 437.2 520.4 374.1 496
521 4789.0 0.205 0.229 0.056 81372 3301 82182 3337 76514 2621 76231
2471 76676 8225 80793 9183 82163 9238 79686 8637 84109 9679 85575
9738 -9999.00 0.32306 0.52862 1.57081 3.14184 0.00032 1.57080 3.14159 0.00000
S.....pP..
 2  2 12 2  722 120253 13927  93.9  83.8 3155.7 508.6 437.3 520.5 374.1 499
521 4785.4 0.206 0.213 0.048 122284 5237 123818 5309 114792 4140 114646
3900 115061 13041 121580 14566 123749 14654 115144 13039 121692 14576 123836
14654 -9999.00 0.32307 0.52819 1.38122 -0.34023 0.87269 1.38170 -0.34053
0.87269 .....dD
 3  3 12 3  723 120774 13995  94.5  87.2  -0.6  46.3 437.4 520.4 374.1 498
521 4785.1 0.239 0.189 0.045 122765 5244 124357 5390 115286 4134 115127
3882 115558 13104 122106 14637 124285 14726 115642 13102 122219 14645 124372
14725 -9999.00 0.32307 0.52775 1.37734 -0.34238 0.87135 1.37780 -0.34268
0.87135 .....
```

### 3.5.1 The data type, internal name, format, and meaning of entries in each statistics line:

```

INT iline          %4d      file counter from pattern (skipped in REJECT)
INT jline          %4d      file counter for files processed
REAL hour          %3.0f    obs hour from header
REAL tmin          %2.0f    obs minute, (referred to in REJECT as hhmm)
REAL obstim        %5.0f    time in minutes, continuous across 1440, (tmnt)
REAL iavgar        %7.0f    whole disk intensity average
REAL istdvar       %7.0f    whole disk intensity stdev
REAL avgar         %7.1f    whole disk psuedomag average
REAL stdvar        %7.1f    whole disk psuedomag stdev
REAL davgar        %7.1f    1 minute difference image average
REAL dstdvar       %7.1f    1 minute difference image stdev
REAL fxcr          %7.1f    x of disk center
REAL fyrcr         %7.1f    y of disk center
REAL radar         %7.1f    disk radius

INT hxcr           %5d      header x_center
INT hycr           %5d      header y_center
REAL vel           %8.1f    header diode velocity
REAL fxsg          %7.3f    x center sigma
REAL fysg          %7.3f    y center sigma
REAL frsg          %7.3f    rad sigma
ndx=1,4 {
  REAL swavg[ndx]  %7.0f    channel sample average
  REAL swstd[ndx]  %7.0f    channel sample stdev
}
ndx=1,6 {
  REAL plavg[ndx]  %7.0f    image plane disk average
  REAL plstd[ndx]  %7.0f    image plane disk stdev
}
REAL cctoff        %7.2f    data - instrument clock offset
REAL pang          %9.5f    PA
REAL beta          %9.5f    BETA
REAL rol           %9.5f    ROLL
REAL pit           %9.5f    PITCH
REAL rot           %9.5f    ROTATOR
REAL erol          %9.5f    EROLL
REAL epit          %9.5f    EPITCH
REAL erot          %9.5f    EROTATOR

STRING hdrstr      _%s\n    header flags (first 6 used)
hdrstr[1]          discontinuity in time [SB] (start, break)
[2]               daily cal or other non standard image [C]
[3]               GUIDE not on [G]
[4]               CENTER not on [N]
[5]               VALID not on [V] or bad ZENITH [U]
[6]               MAGNETO not on[M]
[7 - 12]          ignored, superseded in reject.x

```

## 4. VIEWLOG

VIEWLOG provides a look at some of the image statistics recorded in the DAYSCAN logfile for a site day. These plots provide an overview of the data quality for a site day before beginning the main image calibration phase of VMBICAL. There is an option to flag a day for deletion, so it can be omitted from further processing. The other purpose for running VIEWLOG during VMBICAL is to monitor, and adjust if necessary, one of the rejection thresholds used by AIR.

### 4.1 User Interaction with VIEWLOG

VIEWLOG must be run from within an IRAF session. For its normal use with VMBICAL, cd to the site day directory within the current wa0 directory. For example  
`cd /sandbox/ladd/wa0.le040803/LE040827`

Dialog with VIEWLOG: The user input for the opening dialog with VIEWLOG is normally supplied from an input file 'viewlog.in' created by a perl script.

Unlike the input files created for the other components of AIR, this one should not be redirected to the process standard input (STDIN). It should be named 'viewlog.in' and placed in the current directory where VIEWLOG will recognize it and read its initial input strings. If 'viewlog.in' does not exist it is assumed that this is not the standard automated VMBICAL setting and VIEWLOG will enter its more detailed 'verbose' mode.

(Programmers: STDIN must be left attached to the keyboard for actual human interaction. Contrast this with the handling of input in USEDS. The technique developed there turned out to be unnecessary, but very useful for VIEWLOG.)

The required input:

- the name of the DAYSCAN logfile  
ct030120.raw.100
- The first line of the logfile identifies the site, date and time span covered in the logfile. The site ID is taken from this line and confirmation is requested.  
Found station CT, OK? or 'xx' to specify limits  
y  
Any response other than 'xx' will accept the site and the threshold values for AIR's image selection criteria read from a file 'reject.dat'. [VIEWLOG and REJECT use a shared procedure to process the logfile header and read threshold values from the reject.dat file. This confirmation is for REJECT which has the option (for testing purposes) of ignoring the values in reject.dat and letting the user provide them.]
- The location of the reject.dat file is requested. Enter the path to wherever it may be located. For normal VMBICAL operation it was decided to place a local copy of VMBICAL's default reject.dat in each site day directory  
reject.dat
- The tape VSN and day number for the site day are requested. The contents of this line are used for labeling the plots and are not processed or parsed.  
VSN and day #?  
v012345 1

No further input is taken from viewlog\_in.

The remainder of the logfile is then read. Summary statistics for most of the columns in the logfile are calculated and printed. Each line shows: parameter, mean, standard deviation, minimum, maximum.

fxcr:	445.85	1.097	444.6	448.5	findlimb center x
fycr:	517.88	0.608	516.2	519.1	findlimb center y
rad:	370.18	0.099	369.5	370.3	findlimb radius
vel:	5423.46	839.337	1133.7	12667.4	diode velocity
I mean:	138784	15232.1	36783	152074	whole disk intensity mean
mag mean:	178.46	426.009	-1635.5	2603.5	difference between
mag stdv:	263.44	366.396	85.0	2559.7	modulator states
dif mean:	-0.00	596.684	-2794.5	2644.4	1 minute differences
dif stdv:	335.63	540.986	0.0	3877.8	1 minute differences
I stdv:	16895.1	1965.26	3181.0	18628.0	whole disk intensity RMS
swavg: 1	136086.23	14802.010	37180.0	149099.0	CCD channel 1 mean
swstd: 1	5613.82	650.861	1133.0	6554.0	CCD channel 1 std dev
swavg: 2	137250.41	14950.679	37310.0	150324.0	CCD channel 2 mean
swstd: 2	5917.95	678.964	1182.0	6997.0	CCD channel 2 std dev
swavg: 3	135652.30	15005.001	34784.0	148739.0	CCD channel 3 mean
swstd: 3	5036.99	603.434	769.0	6997.0	CCD channel 3 std dev
swavg: 4	133950.61	14867.713	33975.0	146865.0	CCD channel 4 mean
swstd: 4	4692.60	561.047	694.0	6588.0	CCD channel 4 std dev
plavg: 1	133863.77	14627.976	35848.0	146634.0	image plane 1 mean
plstd: 1	15943.73	1849.084	3001.0	17567.0	image plane 1 std dev
plavg: 2	141103.89	15477.006	37194.0	154574.0	image plane 2 mean
plstd: 2	17785.10	2082.779	3338.0	19627.0	image plane 2 std dev
plavg: 3	140942.70	15552.042	37227.0	154527.0	image plane 3 mean
plstd: 3	17578.19	2050.774	3321.0	19407.0	image plane 3 std dev
plavg: 4	134029.06	14491.537	37895.0	146767.0	image plane 4 mean
plstd: 4	15961.37	1827.574	3206.0	17568.0	image plane 4 std dev
plavg: 5	141269.87	15331.519	39357.0	154713.0	image plane 5 mean
plstd: 5	17805.58	2060.579	3578.0	19643.0	image plane 5 std dev
plavg: 6	141146.12	15412.532	39389.0	154705.0	image plane 6 mean
plstd: 6	17596.19	2028.637	3545.0	19417.0	image plane 6 std dev

- The next few lines of output are related setting the limits for the whole disk pseudomagnetogram averages and will shortly be discussed at some length.
- If input has been coming from viewlog\_in the plots will automatically be produced in a graphics window on the screen. In verbose mode the graphics output device can be selected. Choices are the screen (s), your default printer (p), a device independent metacode file (v or l) whose name is based on the name of the logfile (xyzy.l00 -> xyzy.v00), or you can explicitly specify the name for the metacode file. (If you specify the name for the metacode file it should be at least 3 characters long.) Or you can quit before producing any plots.  
graphics device? [spvlq] or name  
s

With verbose OFF a single page of plots will be produced showing the whole disk average of the modulator difference images (also called pseudomagnetograms), standard deviation (rms) of these modulator images, whole disk average of 1-minute difference between images, and rms of the 1-minute difference images. The threshold limits taken from the reject.dat file are also shown (in red if your using IRAF's xgterm)

After the plots are produced the text window shows the prompt ENTER if ok, (r)edraw, (c)hange mod limits, (p)rint, (v)erbose

An option is selected by entering the letter in parentheses.

- The redraw option- When IRAF initially produces the plot, the graphics window it creates will probably be smaller than desired. You can manually resize the window by whatever method you normally use. Then enter 'r' and the plot will be redrawn in the new window. The new graphics window size will remain in effect for the remainder of the IRAF session. This also clears up some strange shading that is initially present on the plots.

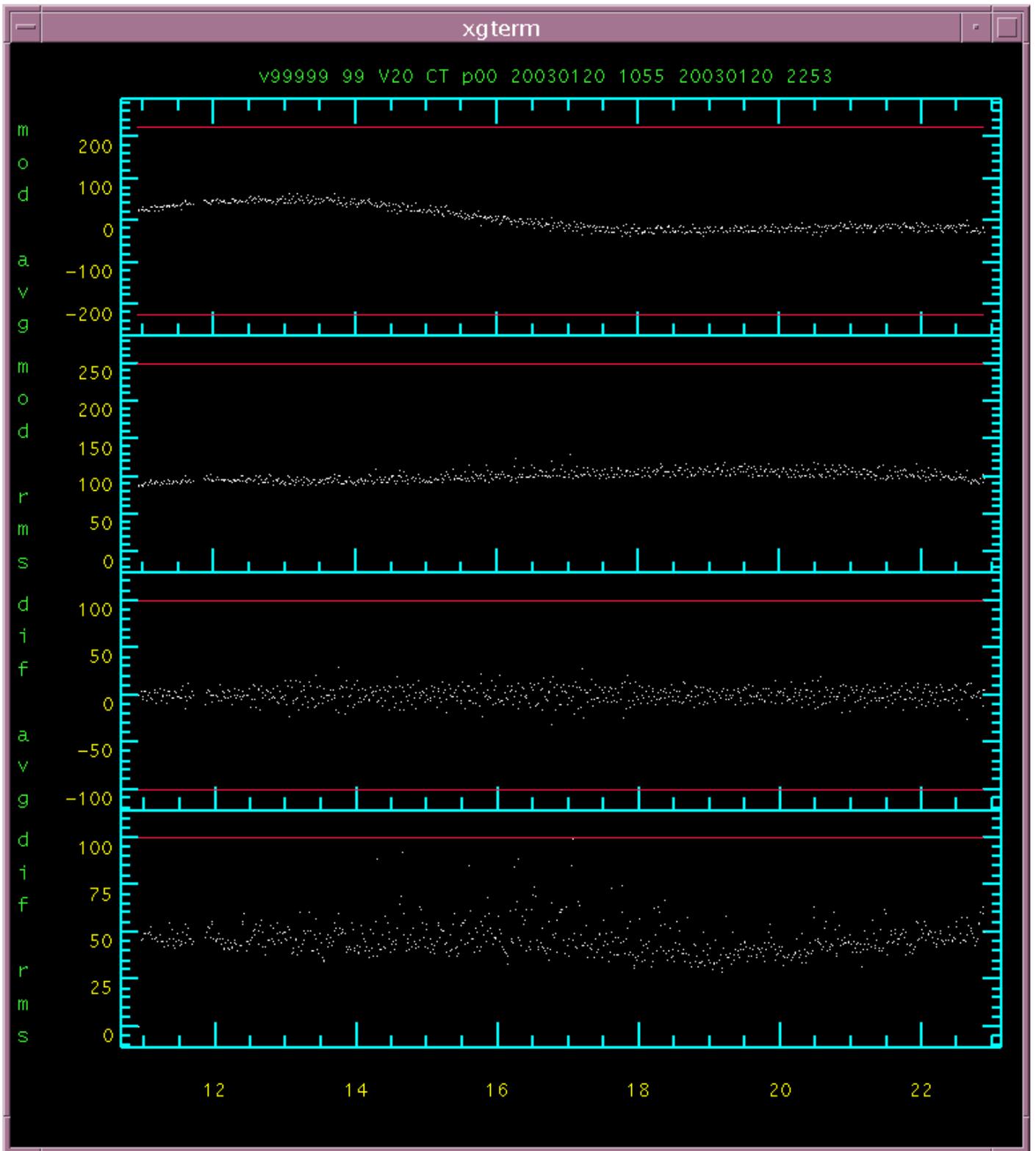
- The print option- The displayed plots will be printed. IRAF buffers the print job. It will not appear in the print queue until the IRAF session is closed.
- The verbose option- Plots of additional parameters from the logfile can be produced by selecting the verbose option. This also jumps back up to the graphics device selection prompt so hardcopy or a graphics metacode file can be produced if desired. (The print option described above applies only to the modulator and difference image plots.) [As each parameter is being plotted some numbers may be printed in the text window. These are related to the automatic scaling of the plots. IRAF's automatic scaling doesn't work very well for these plots so VIEWLOG does its own scaling.] Among the other parameters that will be shown are the limb finder results for the sun's center and radius, intensity values for each of the CCD's four channels, intensity for each of the six image planes, and ephemeris/rotater values.
- The change modulator limits option- This will allow entering new values for the minimum and maximum threshold values for the pseudomagnetogram whole disk average. The plots will be redrawn with the new values shown. This step can be repeated multiple times. Upon final acceptance, the local copy of reject.dat will be updated with the new values. A multi-pass fitting procedure is now applied to each day of pseudomagnetogram averages to estimate appropriate values for the minimum and maximum limits. This reduces, but does not eliminate, the need for the manual *change* option. On rare occasions the fit may not produce valid results, so this option is still necessary.

## 4.2 Setting the Modulator Thresholds

The values of the image rejection thresholds in reject.dat were determined empirically. For most of these values there is little variation from site to site, nor are they expected to vary significantly over time. The exception is the whole disk modulator average. This can change over time, either from periodic seasonal changes (usually fairly small), or due to changes in the instrument. An example of this from CT in early 2003 is shown in the figures below. Over a period of 11 days the modulator average shifted from near zero to below -200. It remained in this state for a few weeks before returning to 'normal'. The overall data quality was not affected by this modulator deviation. Had the rejection thresholds not been adjusted, large amounts of good data would have been rejected.

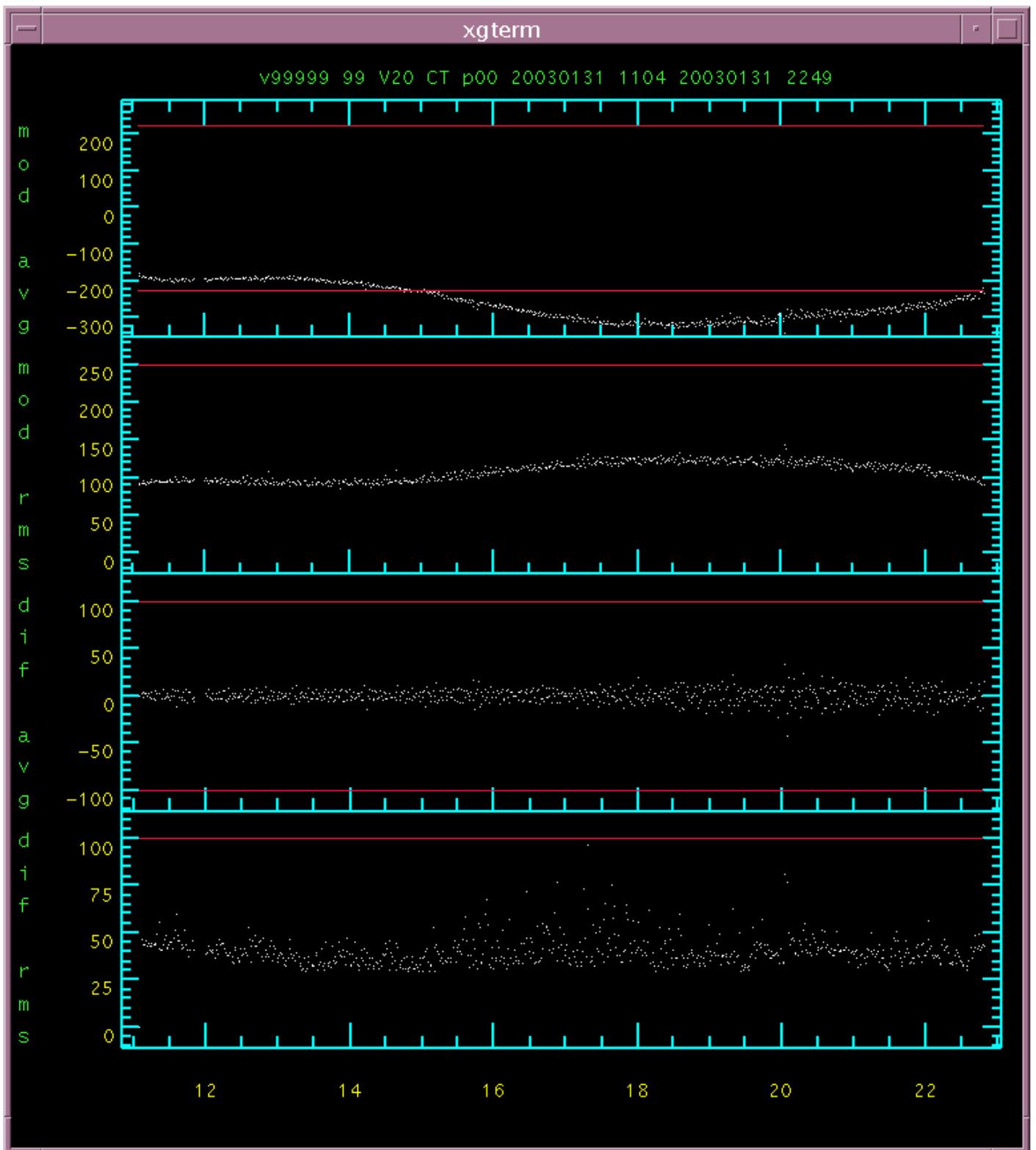
During good observing, the plot of whole disk average modulator state difference should be a smooth line, although there can be gradual variations through the course of the day (turret polarization, etc). This trend should approximately repeat from day to day. The width of the fair weather noise envelope about this trend curve due to normal seeing and instrument noise is generally no more than  $\pm 30$  units or so. The pseudomagnetograms are highly sensitive to even very thin clouds, and also respond to variations in seeing--factors which do not necessarily degrade the velocity images, so images should not be rejected too aggressively. Deviations from the whole disk modulator average trend line as great as 100-150 may be acceptable. If not, one of the other tests will flag the image as bad. There is some redundancy between the various tests so it should be ok to allow a limit to relax to as much as  $\sim 300$  units from the modulator trend line before it is desirable to adjust it.

When one of these long term shifts occurs it may be preferable to modify the master copy of reject.dat in the VMBICAL installation rather than manually change the limits for each day.



Screen shot of CT030120 VIEWLOG plot

This was a typical day of good data just before the modulator shift occurred. The default values for the pseudomagnetogram thresholds comfortably bracket the whole disk average values.



Screen shot of CT030131 VIEWLOG plot

Eleven days later... The modulator now shows a significant bias in the whole disk averages, with most of the day falling outside of the default thresholds in use at the time. This shift is also weakly reflected in the pseudomagnetogram RMS. The data collected this day was nearly all good. The bias does not seem to adversely effect velocity or magnetogram quality. The cause of this modulator drift is not currently known and does not seem to correlate with any anomalous instrument behavior.

After using the *c* option, or being updated by the fitting procedure described below, the local copy of reject.dat for the 31st might look like:

```
center -- modulator --      -- 1 minute --
off rms min max rms min max rms rot
CT 40 0.5 -438.2 -56.8 250.0 -100.0 100.0 100.0 100.0 0.010
LE 40 .5 -225.0 225.0 260.0 -100.0 100.0 100.0 .01
UD 65 .5 -225.0 225.0 250.0 -100.0 100.0 100.0 .01
TD 40 .5 -200.0 200.0 250.0 -100.0 100.0 100.0 .01
CT 40 .5 -250.0 250.0 250.0 -100.0 100.0 100.0 .01
TC 40 .5 -225.0 225.0 250.0 -100.0 100.0 100.0 .01
BB 40 .5 -225.0 225.0 290.0 -100.0 100.0 120.0 .01
ML 40 .5 -225.0 225.0 240.0 -100.0 100.0 100.0 .01
```

The reject.dat file contains two header lines identifying the columns followed by the table of threshold values for each station. A new line for CT has been written after the header. The remainder of the original reject.dat file, including the entry for CT, is then copied. The modified version replaces the original copy to provide the updated values to REJECT-- hence the use of a local copy for each site day rather than the master copy. When the reject.dat file is read to get the limits, the first line that matches the current site is used and everything else ignored. If a permanent change to the official copy is desired the CT line in the default copy of reject.dat could be edited in place, or a new line with updated values could be inserted ahead of the original line. It is also ok to include comments in the reject.dat file, such as the dates of changes. The only restriction is that the first two letters of a comment must not match a station ID. A protocol for maintaining and documenting the history of changes to the default values in the reject.dat file has not been established.

- After any redrawing, modulator changes, and printing, an OK response will be given. The next prompt will be  
mark day for rejection

Sometimes it may be immediately obvious to the operator that the day contains no useful data. This could be because of very poor data (all points fall outside the threshold lines on the plots), or because too little data is present to begin with. This option allows signaling to the top level perl scripts and r\_all script that the day should be discarded.

VIEWLOG creates a file called viewlog.done. The existence of this file is used as a semaphore to indicate that any necessary changes to the reject.dat file have been made before continuing with REJECT and USEDS. It is also used to pass the DELETE DAY signal if so chosen.

The creation of viewlog.done and modification of reject.dat are the only changes caused by VIEWLOG that would affect GONG data. It could be run additional times without interference-- for instance switching to verbose mode to see plots of additional parameters or produce hard copy of the plots. Changes to reject.dat will be cumulative but only the most recent one will have any affect. It also creates a new viewlog.done each time it runs, so if the day is to be marked for deletion, be sure to mark it again the last time you use VIEWLOG.

An additional tool, readlog.pro, is available for examining the logfiles if a day is particularly interesting. It is written in IDL and is more interactive than its SPP equivalent (which is a direct ancestor of VIEWLOG). It does not read or modify the reject.dat file. It is more experimental and meant to be customized for specific investigations as they arise.

#### 4.2.1 A Multi-Pass Fit to Set the Pseudomagnetogram Limits

A procedure to dynamically determine reasonable values for the pseudomagnetogram limits using a multi pass polynomial fitting technique has been devised. Each site day is handled individually. The whole disk pseudomagnetogram averages for each image during the day form a time series.

- Pass 1: All points having 'reasonable' values (e.g. Between -600 and 600) are fit using a cubic polynomial.

- The residues for each data point relative to the fit from the first pass are calculated. Points with small residues ( $\pm 75$ ) are included in a new cubic polynomial fit.
- Points having small residues ( $\pm 75$ ) relative to the second fit are included to produce a final quartic fit.
- The extreme values of this polynomial between the start and stop times of the original time series are found. An additional margin (150) is added to each side to get the limits image acceptance or rejection.

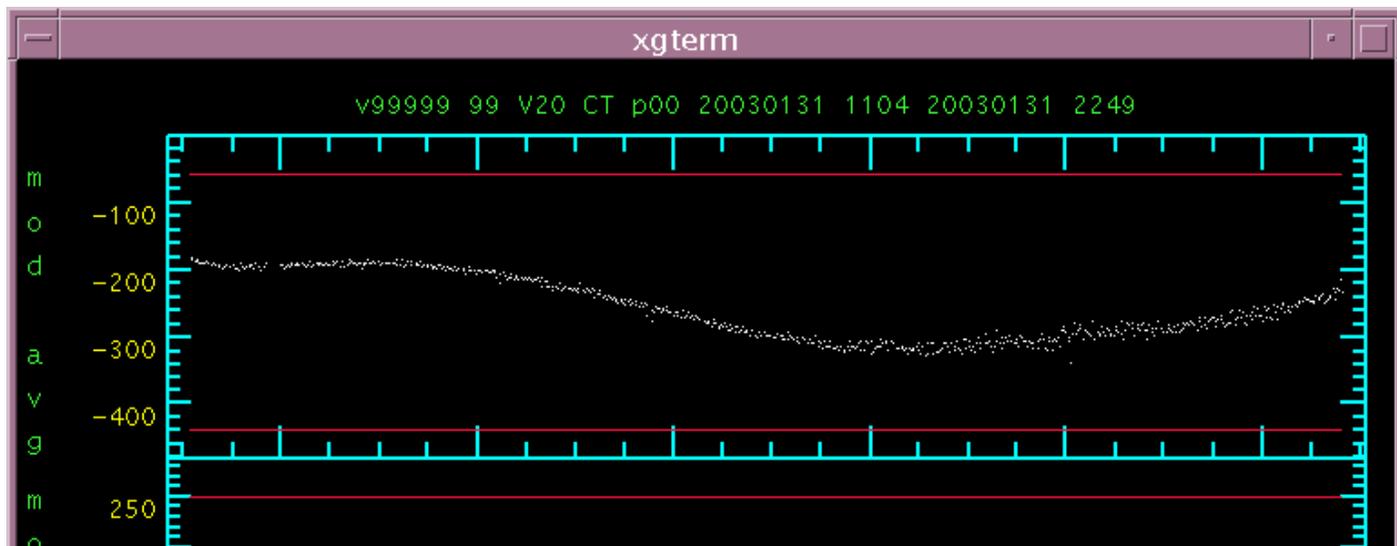
The purpose of this exercise is not to directly use the fit for comparisons against a threshold for image selection, but to determine the range of values in which good data should fall. BB and TD both show a rather complex daily trend. To produce a fit that closely approximates the daily trend of whole disk averages would require a polynomial of fairly high order, which could be prone to large oscillations giving bad fits to days with significant amounts of bad data.

The progress and results of the fit are displayed.

	Total	used	fitting coefficients				
	N	n	C <sub>i</sub>				
p1:	561	516	-3239.004395	463.194305	-20.674019	0.298059	
p2:	561	451	-1659.902222	231.233597	-9.394986	0.116809	
p3:	561	461	-7742.413574	1499.778687	-107.475861	3.449541	-0.042004
range:		59.0	141.0				
mod limits-- default:			-225.0	225.0			
		calculated:	59.0	141.0			
updating limits...			-66.0	266.0			

The minimum and maximum values of the fitting polynomial over the timespan covered in the time series, the default values of the thresholds taken from the reject.dat file, and the values (with some additional margin added) used for updating the reject.dat file are report in VIEWLOG's output.

On days with small amounts of good data the fitting procedure may fail, or produces unreasonable results, though this is rare due to the constraining effects of the first pass. If the calculated threshold values fall outside of a reasonable range, the default values from reject.dat will be retained. The *c* option for manually updating the threshold values remains available if necessary.



A section of the VIEWLOG plot for the 31st after the fitting procedure has calculated new thresholds.

## 5. REJECT

REJECT reads the logfile, compares the statistics against threshold values and performs other tests to produce the image disposition file. Elsewhere within AIR documentation and source code this file is also referred to as a reject file or a report file. It generally uses the name of the DAYSCAN logfile with '\_rpt' appended. This report file includes a record of the threshold values used for image rejection and is included with the qac ancillary files saved and archived on the DSDS.

- Start IRAF. Be sure the REJECT task is defined.
- cd to a working directory where the reject file is to be produced. This is generally the same working directory where the logfile resides.
- Run REJECT.
- When prompted, give the name of the site day logfile created by DAYSCAN. The header from that logfile is displayed. The site ID is taken from the logfile and shown for confirmation. Normally the image selection threshold values for this station are taken from the reject.dat file. Entering 'xx' in response to this confirmation will disregard the reject.dat file and let the user specify threshold values. This option may be useful during testing but there normally should be no need to modify the default values except for the whole disk modulator average limits which may be adjusted in VIEWLOG.
- Finally, the pathname of the reject.dat file is requested. This can be absolute or relative to the current directory. For normal use it will be in the current site day control directory.

The reject file will be produced having a name of <logfile>\_rpt.

A sample input file that might be used to run REJECT in VMBICAL:

```
/sandbox/ladd/wa0.bb040708/BB040726/bb040726.raw.100
y
/sandbox/ladd/wa0.bb040708/BB040726/reject.dat
```

The report file produced by REJECT includes the summary line from the beginning of the logfile and a line giving the threshold values from the reject.dat file.

### 5.1 How the logfile entries are used in REJECT

```
INT obstim          used for alignment
REAL iavgar
REAL istdvar
REAL avgar          checked against threshold
REAL stdvar         checked against threshold
REAL davgar         checked against threshold
REAL dstdvar        checked against threshold
REAL fxcr
REAL fycr
REAL radar

INT hxcr            checked against limits
INT hycr            checked against limits
REAL vel            not used
REAL fxsg           checked against threshold
REAL fysg           checked against threshold
REAL frsg           checked against threshold
ndx=1,4 {
  REAL swavg[ndx]   under investigation
  REAL swstd[ndx]   under investigation
}
ndx=1,6 {
```

```

    REAL plavg[ndx]      checked against threshold
    REAL plstd[ndx]     checked against threshold
}
clock offset           not used
rotation/ephemeris quantities  delta = EOTATOR - (ROLL - PITCH - BETA - PA)

STRING hdrstr   %s\n   header flags (first 6 used)
[1]   discontinuity in time [SB] (start, break)
[2]   daily cal not yet done [C]
[3]   GUIDE not on [G]
[4]   CENTER not on [N]
[5]   VALID not on [V] or bad ZENITH [U]
[6]   MAGNETO not on [M]
[others]           not used, superseded in reject.

```

## 5.2 The reject.dat file used by REJECT and VIEWLOG

The default threshold values for the various tests are stored in this file. It has a one line entry for each GONG site giving the threshold values. The columns are:

```

site ID
maximum offset of solar image center from image center (applied to each axis)
maximum RMS for center and radius estimates of solar image
minimum value for whole disk pseudomagnetogram average
maximum value for whole disk pseudomagnetogram average
maximum standard deviation for whole disk pseudomagnetogram
minimum value for whole disk 1 minute difference image average
maximum value for whole disk 1-minute difference image average
maximum standard deviation for whole disk 1 minute difference image
rotator threshold

```

	center	--	modulator	--	--	1 minute	--		
	off	rms	min	max	rms	min	max	rms	rot
LE 40	.5	-225.0	225.0	260.0	-100.0	100.0	100.0	100.0	.01
UD 65	.5	-225.0	225.0	250.0	-100.0	100.0	100.0	100.0	.01
TD 40	.5	-200.0	200.0	250.0	-100.0	100.0	100.0	100.0	.01
CT 40	.5	-250.0	250.0	250.0	-100.0	100.0	100.0	100.0	.01
TC 40	.5	-225.0	225.0	250.0	-100.0	100.0	100.0	100.0	.01
BB 40	.5	-225.0	225.0	290.0	-100.0	100.0	120.0	100.0	.01
ML 40	.5	-225.0	225.0	240.0	-100.0	100.0	100.0	100.0	.01

Each line of the reject.dat file is read and the first two characters are checked against the requested site ID for a match. Lines are read until a match is found. The threshold values are taken from the first line matching the site ID. The remainder of the file is ignored. This allows adding column labels or comments as necessary. It also allows keeping a history of modifications to the default threshold values as only the first line matching a site will contain the current values and other lines for that site later in the file are ignored. The *change modulator limits* option in VIEWLOG uses this feature, though there is currently no provision to propagate such changes from a local site day copy of reject.dat to the default copy in the VMBICAL installation. With the new dynamic fitting procedure for determining appropriate values for the pseudomagnetogram thresholds, the need for *change limits* should be less frequent and the need for updating the default values may not be an issue.

## 5.3 Tests

REJECT has several sets of tests, each applied independently. Failure to pass any test will get an image marked for rejection.

### 5.3.1 Header Checks (first pass)

The first set of tests is based directly on values read from the raw image fits headers. These are to assure that

the instrument is in its normal operating and tracking mode. The conditions that will cause one of these flags to be set are:

- first image of day or discontinuity in time. It has been observed that upon resuming normal operation after a period of downtime the first image is nearly always bad.
- the image is from a calibration sequence, noon drift scan, or other non standard observation and is not to be processed by VMBICAL. (normally FTR diverts such images) This includes a check that the image integration time is 60 seconds. (the first observation following the daily calibration sequence often violates this condition)
- GUIDE not on
- CENTER not on
- not in a valid observing mode
- MAGNETOgram mode not on

These last two appear to be relics from GONG classic and may be redundant.

Flag fields [7 - 12] give the results of comparisons performed by DAYSCAN of image statistics against early, hardwired threshold values. They are ignored. These tests are now performed in REJECT.

### 5.3.2 Guiding and Tracking (second pass)

- The location of the center of the sun's image is checked for proper location on the CCD. The threshold for maximum displacement is given in reject.dat
- The RMS values of the sun's radius and location of its center are checked against a threshold. The same threshold is used for all three.

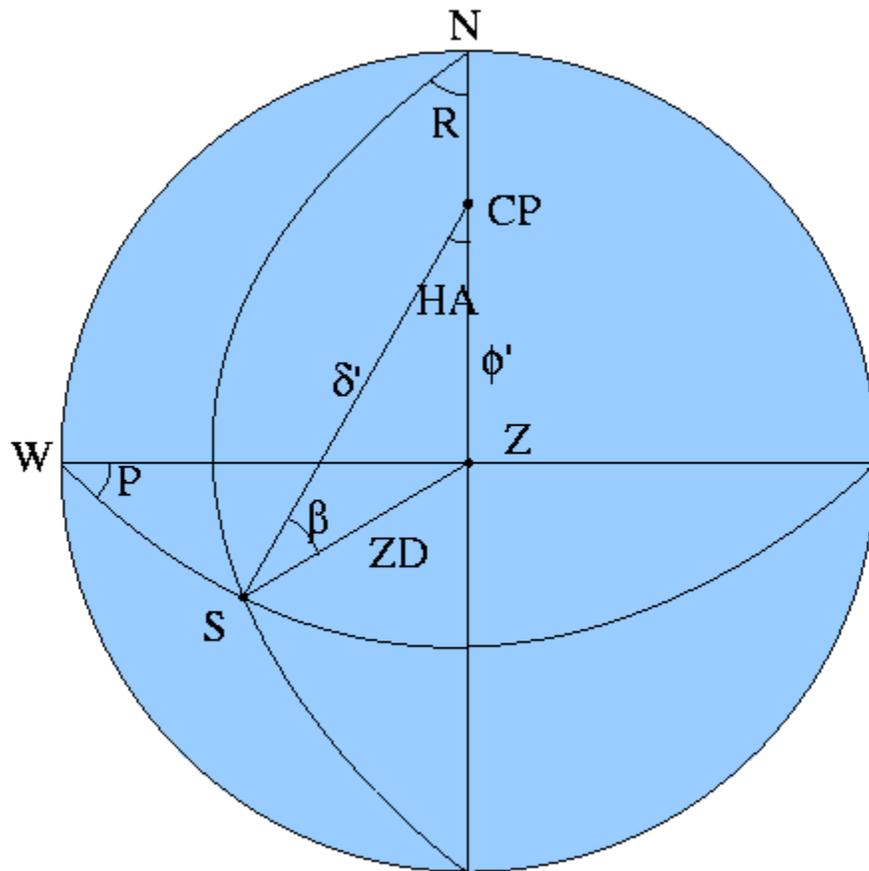
### 5.3.3 Ephemeris and Rotator (third pass)

This quality test and was activated in January 2005. It was added to check for glitches in the camera rotator position. These can arise from the rotator being inappropriately commanded to other than the normal tracking position or from glitches in the site instrument computer clock. This clock is periodically synchronized with the data computer clock which is slaved to a GPS based reference. A low precision ephemeris model is used to calculate the sun's position as turret roll and pitch angles. These are used for initial pointing and acquisition of the sun. An active guider then controls precise turret pointing and tracking. The parallactic angle (BETA) and solar pole position angle (PA) are also calculated from the ephemeris. The camera rotator angle is calculated from these ephemeris values from the relation

$$\text{EROTATOR} = \text{EROLL} - \text{EPITCH} - \text{BETA} - \text{PA}$$

The raw image files contain keyword entries for each of these in the fits header. There are also entries for ROTATOR, ROLL, and PITCH. The 'E' values are based on ephemeris calculations. The other set contains the values read from encoders that sense the actual position. The turret angles (roll, pitch, beta) are shown below in the figure.

## TURRET POINTING AND CAMERA ROTATOR PARAMETERS



Reference points -  
 N: Geo North  
 W: East or West  
 CP: North Celestial Pole  
 Z: Zenith  
 S: Sun

Angles -  
 $\delta'$ : 90-Declination  
 $\phi'$ : 90-Latitude  
 HA: Hour Angle  
 + afternoon  
 ZD: Zenith Distance  
 R: Roll  
 P: Pitch  
 $\beta$ : Parallactic Angle  
 + afternoon north  
 - afternoon south  
 PA: Pole Position Angle

$$\text{ROTATOR} = R - P - \beta - \text{PA}$$

This rotator angle is dependent on time as read from the instrument computer clock. Unlike the turret pointing angles, which are controlled by the active guider, there is no easily observable real world quantity to override the calculated rotator angle should it be in error. However, a sanity check has been devised using the actual turret pointing angles in the above expression. This result can be compared with the recorded rotator value. Even under normal operation the match is not as close as might be hoped due to mechanical considerations (gear periodicities, axis misalignment) and because the various calculated and observed quantities recorded in the image headers do not seem to form an internally consistent (all referring to the same instant) set.

The threshold used in the Ephemeris section is taken from reject.dat. In the absence of mechanical errors, a value of .01 (.57 degrees) corresponds to an error of a couple of minutes in the instrument clock. If this error were switched on abruptly while observing (e.g. CT 031226), the resulting dataset would not cause any program aborts in the DNS pipeline but would corrupt the COPIPE angles and likely produce a large washing machine effect.

### 5.3.4 Image Statistics (fourth pass)

The whole disk statistics for the pseudomagnetograms and 1-minute difference images are checked here. The disk averages are compared against both a minimum and maximum threshold. The standard deviations are checked against an upper limit threshold. The values for these six thresholds are read from the reject.dat file.

The pseudomagnetogram averages for CT, BB, and TD have all exhibited a significant non-zero bias which varies with time. The values for min and max can be reset individually in VIEWLOG for each site day. If an intermediate to long term shift in the pseudomagnetogram bias is observed it may be preferable to modify the values in the primary copy of reject.dat

The 1 minute difference averages have an expected value of zero. The values in reject.dat are min = -max.

If any one of these four quantities is outside of its threshold range the image is marked as bad.

#### **5.3.4.1 Coherent Noise test**

Generally the three image planes from each modulator state are averaged to produce intensity images for locating the limb and to produce the pseudomagnetograms. On rare occasions a pattern of coherent noise can be introduced by the instrument. When the image planes are averaged the noise subtracts out, resulting in seemingly good images on which the other tests are performed. However, when the image planes are examined individually, they are clearly corrupted and will not produce valid data. The mean and RMS for all six planes are checked against thresholds. In particular, the RMS values will be strongly affected by the presence of this noise.

This test is currently performed and indicated in the report file during the third pass with the rotator test rather than with the other image statistics tests. The causes of the coherent noise, and of some rotator glitches are not well understood. When either of these conditions is detected a special message can be generated and emailed to software development and instrument operations personnel.

#### **5.3.5 Merged**

Finally the results of all the sets of tests are merged. Failure of any of the tests will result in an image being marked as bad.

### **5.4 The report file from REJECT**

Here is a report file produced by DAYSCAN/REJECT for a site day of marginal quality. The first line is a copy of the summary line from the DAYSCAN logfile. The second line gives the thresholds from the reject.dat file used for the tests.

- The 1st value (40) is allowable distance from the CCD center of the x or y coordinate of the sun's image to be considered proper tracking.
- The 2nd (.5) is the maximum acceptable value of the RMS from the determination of the x or y location of the image center, or the image radius.
- The 3rd and 4th (-225, 225) are the min and max values for the pseudomagnetogram average.
- The 5th (290) is the maximum value for the pseudomagnetogram standard deviation.
- The 6th-8th (-100, 100, 120) are the corresponding values for the 1 minute difference images.
- The 9th (.01) is the maximum offset (in radians) between its reported and expected value.

The results for each of the individual sections (passes) are shown. The first number is the UT hour (+24 after a UT date change during the day at BB, ML, or LE) There is a character for each minute of the hour, grouped into blocks of 10.

'.' indicates the image has passed the test for acceptance.

'-' indicates no image is present for that minute.

Any other character indicates the image did not pass a test. Initially 'R' for Reject was used, but the 2nd, 3rd, and 4th passes include multiple tests. It is possible to encode the results of up to four tests as a hexadecimal digit. This allows monitoring the performance of individual tests for future adjustments or fine tuning. The last two numbers are the number of images present for that hour, and the number marked for acceptance. The line at the end of each section gives the total number of images and the number that were accepted. The logfile summary line is also tacked on again.

```
V21 BB p00 20040716 1817 20040717 0126
BB limits: 40 0.500 -225.0 225.0 290.0 -100.0 100.0 120.0 0.010
Header flags
18 -----R.. -----R ..... R----R--- 25 21
19 -----R--- -----R.. .-R.... 10 7
20 ---R..... -----R-R.. -R..... -R..... 38 33
21 ..... -R. ...--R-R.. --R..... 32 28
22 -----R----- -R.....- R--R..-- R--R....-R 20 13
23 ..-R..... -----R..... 31 29
24 ..... 60 60
25 ..... 27 27
file count_Head 243 218 V21 BB p00 20040716 1817 20040717 0126
```

```
guiding and tracking
18 -----54. -----4 4444444444 4444444444- 5-----4--- 25 1
19 -----5-----544 4-44444- 10 0
20 ---545444 -----F-.44--- -4.4444444 4-544444... .....4 38 14
21 4CC4444444 44444444-.4 ...--4-.1. --5..... 32 11
22 -----1----- -1.....- 1---...- 1--.....- 20 16
23 .1-.1..... -----1. .... 31 28
24 ..... 60 60
25 ..... 27 27
file count_guid 243 157 V21 BB p00 20040716 1817 20040717 0126
```

```
ephemeris and rotator
18 -----... ----- 25 25
19 -----... .-..... 10 10
20 -----F-. -F-.FC- -ACFEFC... -FCCFCF8. .... 38 38
21 ..... -F-.FC- -ACFEFC... -FCCFCF8. .... 32 32
22 -----F-. -F-.FC- -ACFEFC... -FCCFCF8. .... 20 20
23 ..-..... -----F-. -F-.FC- -ACFEFC... -FCCFCF8. .... 31 31
24 ..... 60 60
25 ..... 27 27
file count_ephe 243 243 V21 BB p00 20040716 1817 20040717 0126
```

```
stats
18 -----1C. -----F FEADC...A8 ...EEEDC- F----F--- 25 8
19 -----F-----FC. .-FEFCF- 10 2
20 ---FEEBBC -----F-FC- -ACFEFC... -FCCFCF8. .... 38 16
21 ..... A8.C-FF C...-F-FFF --FCAEEF- 32 16
22 -----F- -FC..88.F- F--FC- F--DC454-F 20 4
23 EF-FD8CC8. -----F- -FC..88.F- F--FC- F--DC454-F 31 21
24 ..... 4 4.C.D4CC8C 60 51
25 C.88CC.... 27 22
file count_stat 243 140 V21 BB p00 20040716 1817 20040717 0126
```

```
merged
18 -----RR. -----R RRRRRRRRRR RRRRRRRRR- R----R--- 25 1
19 -----R--- -----RRR R-RRRR- 10 0
20 ---RRRRRR -----R-RRR- -RRRRRRRRR R-RRRRRR. ....R 38 10
21 RRRRRRRRRR RRRRRRR-RR R..--R-RRR --RRRRRR- 32 2
22 -----R----- -RR..RR.R- R--RR- R--RRRRR-R 20 4
23 RR-RRRRRR. -----RR ..... 31 21
24 ..... R R.R.RRRRRR 60 51
25 R.RRRR.... 27 22
file count_merg 243 111 V21 BB p00 20040716 1817 20040717 0126
```

## 6. USEDS

USEDSED reads the image disposition file and applies the decisions indicated there. Rejected images in the raw image directory are moved to a quarantine area in a backup directory. The header parameter index files are rewritten to reflect the new contents of the raw image directory, with the originals moved into quarantine in a backup directory. USEDSED also includes a short day filter and can generate a signal for a site day to be omitted from further processing.

In a manner similar to VIEWLOG, USEDSED creates a semaphore file 'useds.done'. The existence of this file signals to the controlling VMBICAL scripts that it has been run and moved the designated bad images aside. useds.done is also used to send the message to delete the site day. However, if the day is flagged for deletion, the bad image quarantine is skipped.

For normal use with VMBICAL USEDSED is run from the same site day directory under the wa0 control directory as are DAYSCAN, VIEWLOG, and REJECT.

- The first prompt is rather cryptic:  
take inputs from (s)tdin or (filename)

Enter 's' in response to this, even if the inputs have been placed in an input file that is being redirected to the job's standard input as is done by the VMBICAL scripts. The explicit use of an input file (by specifying its name in response to the prompt) switches into a test mode in which some actions of USEDSED can be modified if necessary. (programmers: contrast this with the handling of the input file in VIEWLOG)

- When prompted for the rejection report file enter the name of the report file (usually <logfile>\_rpt) produced by REJECT. This file is then read and decoded. Screen output shows line by line progress while the reject file is being parsed and a summary of site day start and stop times, total number of images, number of images to retain and reject is output.
- The next prompt is for VSN and day. This is the VSN of the field tape and the day number of the current site day on that tape. The information entered here is not used by USEDSED but is for identification purposes only. USEDSED includes a short day filter which checks the number of images present (with more sophisticated tests to be added as we gain experience). If the site day does not pass the short day filter, then a message to delete the day is inserted into the useds.done file and USEDSED stops its processing here. The VSN and day number are used for identification in this message.
- The next prompt is 'path to data directory (wa1):'. This is the location of the raw data (generally <SSYYMMDD>.raw) under the wa1 directory.
- This is followed by 'path to header parameters (wa0):', requesting the location of the header parameters index files in the SITEDAY directory under the wa0 control directory. This is normally a subdirectory in the current directory called 'hdr\_parms' but the full path can be explicitly specified if need be.
- The next prompt is 'path to rejected data directory:'. This is a location to put the images marked for rejection. The rejected images are not deleted, but moved to this location. It has been decided to create a new directory under the wa1 directory for each site day <SSYYMMDD>.rej. A response of '>' will cause this directory name to be automatically generated and used for the rejected images directory. The directory path can be explicitly specified if need be. If this directory already exists it will be used. Otherwise (normally the case) a new directory is created.
- This next prompt is 'path to updated header parameters:'. The index files must be modified to reflect changes in the .raw directory as rejected images are removed. As with the rejected images directory, enter

'>' to use the standard naming conventions we have worked out. The new index files are written in a directory called 'new\_parms'. Upon successful completion, the original 'hdr\_parms' directory is renamed 'old\_parms' and 'new\_parms' is renamed 'hdr\_parms'.

The names of all of the original and new directories, and all of the index files are displayed.

USEDSDS starts going through all the images in the raw directory. For each image it reports whether it is retained or moved to the rejected directory.

Upon completion it reports the number of images originally reported by DAYSCAN, found in the REJECT report file, and processed in the raw image directory. These three counts should all be equal.

Here is a sample input file for USEDSDS as created by the VMIBCAL control scripts:

```
S
/sandbox/ladd/wa0.le040803/LE040823/le040823.raw.100_rpt
v012345 02
/sandbox/ladd/wa1.le040803/LE040823.raw
hdr_parms
>
>
```

## 6.1 Short Day Filter

The Short Day Filter checks the total number of images accepted. More elaborate tests, such as checking that useful numbers of consecutive runs of accepted images are present, can be added in the future. Tests to determine the sensitivity of either global or local helioseismology products to high frequency structure in the window function have not been performed.

## 7. Tests with Calibrated Images

The main function of AIR is performed on raw images prior to any calibration or processing. It is possible that additional QA checks that can only be applied to processed images will be included in AIR in the future. An interactive QA stage is currently included between the PROCESS\_SITE\_DAYS and SPLIT\_IMAGES stages of VMBICAL. While AIR produces satisfactory data quality in its present form, there is still room for improvement by adding automated tests or retaining a low level of interactive supervision at this stage. Currently the only such check performed by AIR is concerned with the value of VELSCALE.

The GONG instrument does not have provision for an absolute velocity reference. A time series is produced using the whole disk average velocity for each image from a site day. This is fit to the topocentric radial velocity of the center of the sun's disk according to the model:

$$V_{\text{eph}} = V_{\text{avg}} * \text{VELSCALE} - \text{VEL\_BIAS}$$

VELSCALE and VEL\_BIAS are found for each site day by a least squares fit.

Checks are performed on the value of VELSCALE. The first check verifies that the sign of VELSCALE is consistent with the solar rotation velocity visible in the image. The average (median) velocities from sample regions near the east and limbs are taken from an image. Sign consistency is indicated if the test

$$(V_{\text{east}} - V_{\text{west}}) * \text{VELSCALE} > 0$$

is true.

It is also required that VELSCALE be fairly close to  $\pm 1$ . This condition is satisfied if the test

$$.8 < |\text{VELSCALE}| < 1.25$$

is true.

If either test fails a signal is set to eliminate the day from further processing.

## 8. Error Recovery and Cleanup

It is, of course, impossible to anticipate the unforeseen situations that will cause software to fail, or to anticipate the exact manner in which it will fail. Nevertheless, the following discussion is offered. While it may be of limited use for the actual recovery and resumption of processing of the site day dataset causing the problem, it should be helpful with cleanup during tests to re validate the software after modifications to deal with future occurrences of the situation. Recreating the VMBICAL data tree for the troublesome site day or VSN between tests should not be necessary.

### 8.1 DAYSCAN, VIEWLOG, and REJECT

Only minimal cleanup should be required if an error condition should occur during DAYSCAN, VIEWLOG, or REJECT.

The only residue from an aborted DAYSCAN job might be a partial logfile (name ending with '\_a'). After dealing with whatever caused the error condition delete this partial logfile and run it again.

The local copy of reject.dat, which contains threshold values use by REJECT for image selection, can be modified by VIEWLOG. The limits within which good whole disk pseudomagnetogram averages should fall can vary over time and human intervention is still used to adjust these limits. A technique showing some promise for automatically setting these limits has been developed and may be included in the future.

A file called 'viewlog.done' is also created as a signal that VIEWLOG has been run and the pseudomagnetogram limits adjusted if necessary. As explained elsewhere, viewlog.done and is also used to signal that the operator has decided that day should be deleted from further processing.

Before rerunning VIEWLOG after correcting an error condition, it may be desirable to get a fresh copy of reject.dat from the default copy in VMBICAL. It isn't really necessary to delete viewlog.done as VIEWLOG itself does this each time it is run.

If REJECT aborts, it may have created a partial report file (filename ends with '\_rpt'). If so, delete it before rerunning.

Unlike the other modules, USEDS actually interacts with and modifies VMBICAL's working directories. Cleanup can be nontrivial.

### 8.2 USEDS

When USEDS operates on a site day it uses a directory in the wa1 tree called <SITEDAY>.rej. Images in <SITEDAY>.raw designated for rejection are moved-- not copied!! -- into the .rej directory. Any files in this .rej directory will need to be moved back into the site day's .raw directory.

Also, a new set of copies of the header parameter index files in the site day directory under the wa0 tree are created. The original files are in a directory called 'hdr\_parms'. A new directory called 'new\_parms' is created to hold the new copies. After successful operation, the original 'hdr\_parms' directory is renamed 'old\_parms' and 'new\_parms' is renamed 'hdr\_parms' to take the place of the original. The original directory and its contents are never altered, just renamed.

But before attempting any actual cleanup, look around and make sure that everything can be accounted for. In the wa0 site day directory perform these checks:

```
wc hdr_parms/*
wc old_parms/*
wc new_parms/*
```

To get the line counts of the index files. There should be one line per image.

Check the first line of the logfile for the number of images it contains. (wc should give a line count equal to the number of images plus 11.)

Check the file count reported on the last line of the report file.

In the wal directory check the file counts in the .raw and .rej directories:

```
ls .raw/* | wc  
ls .rej/* | wc
```

The total number of image files should match the line counts of the index files.

If everything is accounted for (great care has been taken to prevent the actual loss of any data by USEDS) then the following procedure should restore the VMBICAL files to their original state:

If hdr\_parms and old\_parms both exist (the USEDS output will probably indicate normal termination) then delete hdr\_parms and rename old\_parms -> hdr\_parms.

If new\_parms and hdr\_parms both exist (the USEDS output should indicate an error state or abnormal termination) delete new\_parms.

Move the files in .rej into .raw.

When finally satisfied, rerun useds in its normal 'automated' mode used with VMBICAL to perform its image rejection and to make sure the useds.done file is correctly created.

The next test can now be conducted or VMBICAL can be resumed.

# APPENDIX A

## Comparison of Archived Dataset with AIR

GONG data from a 108 day period (GONG months 77-79: 021102 - 031117) were processed using AIR to select images for inclusion. Helioseismology products generated from this dataset were compared with the archived products.

During month 77 UD was not operating and ML was having a scheduled PM during the first two weeks. However, LE was operating normally and experienced generally good weather, so there were no major gaps in coverage. LE was not operating for much of months 78 and 79. UD resumed operation early in month 78.

The duty cycles based on the window functions from the merged monthly time series are:

Month	Archived	AIR
77	87	92
78	86	88
79	86	87

Helioseismology products generated from this dataset were compared with products from the original, normally processed dataset

The AIR dataset had a slightly higher net duty cycle: ~89% vs ~86%. The difference in duty cycle at individual sites was generally slightly larger before allowing for station coverage overlap.

The results from comparing mode frequency peaks (shown in Fig. 1 and 2), l-nu diagrams, and a Regularized Least Squares inversion (Fig. 3) were all very similar for the two datasets.

Ring diagram derived flows are shown in Fig. 4.

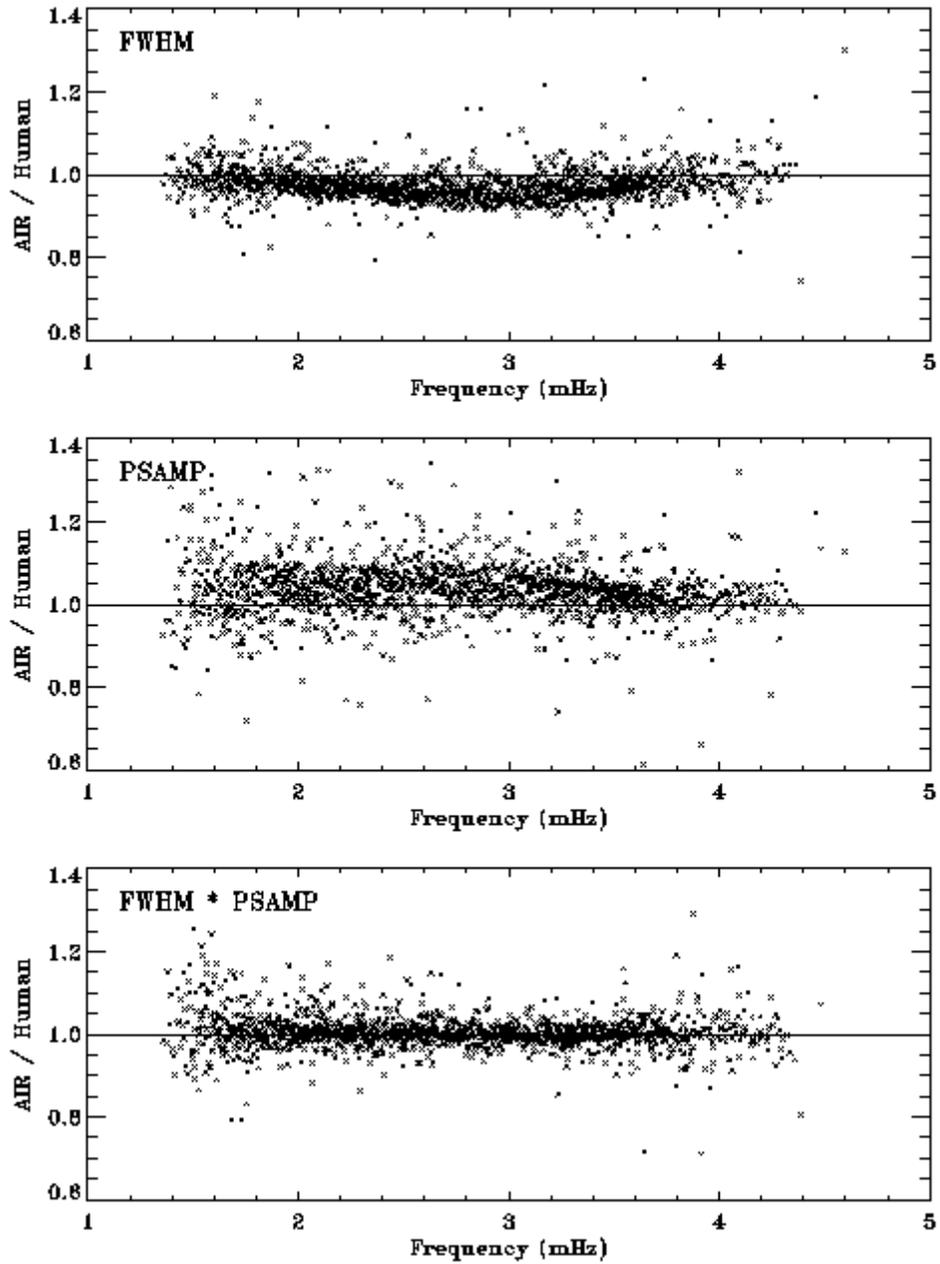
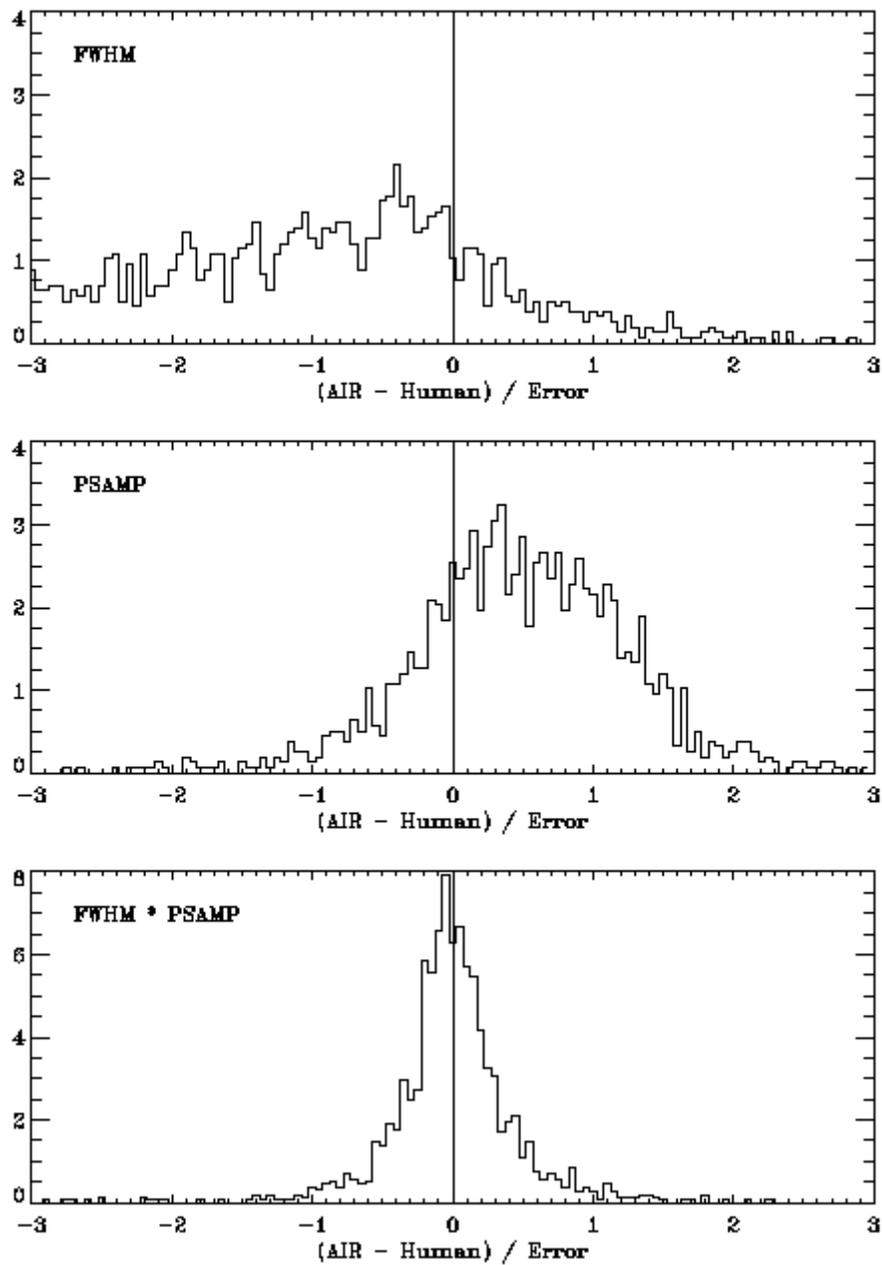
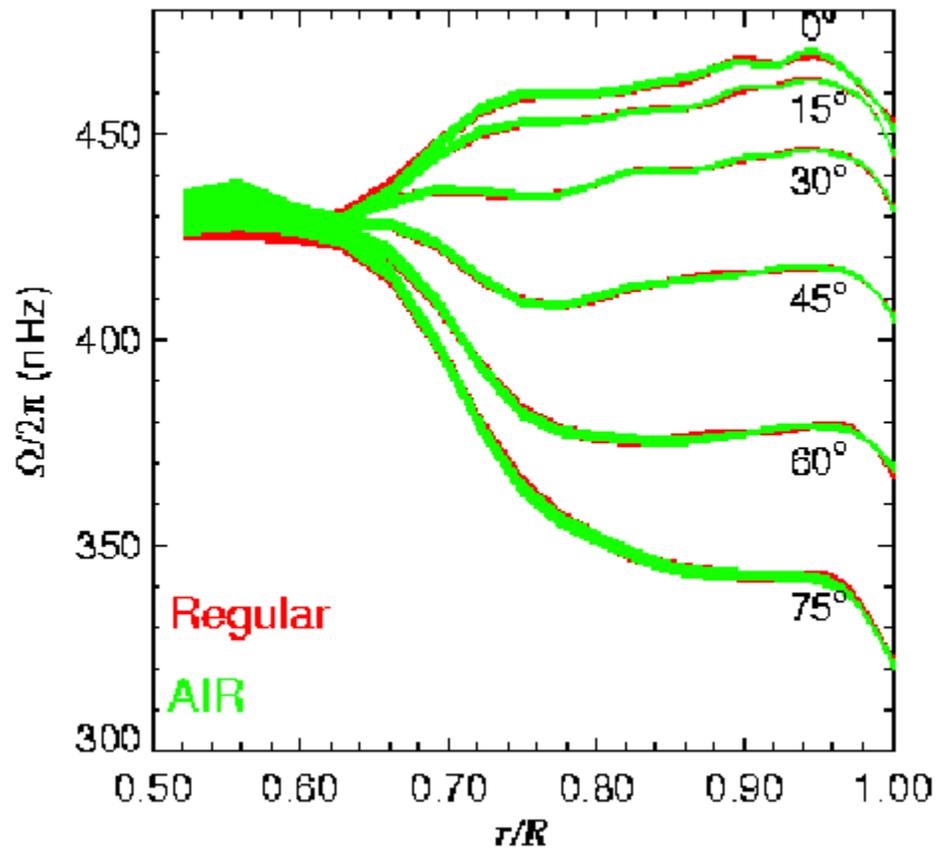


Figure 1 PEAKFIND was used to identify p-modes from the time series produced from each version of the dataset. The ratios of mode widths, amplitudes, and strengths of individual modes are shown.



*Figure 2* Here are histograms of the distribution of mode widths, amplitudes, and strengths. There are systematic differences in the widths and amplitudes reflecting the different duty cycles but using the product of width \* amplitude as a measure of mode strength shows very little difference between the two datasets.



*Figure 3* The results of a regularized least squares inversion to find rotation rate vs depth for several latitudes. The line width is proportional to the uncertainty in the inversion.

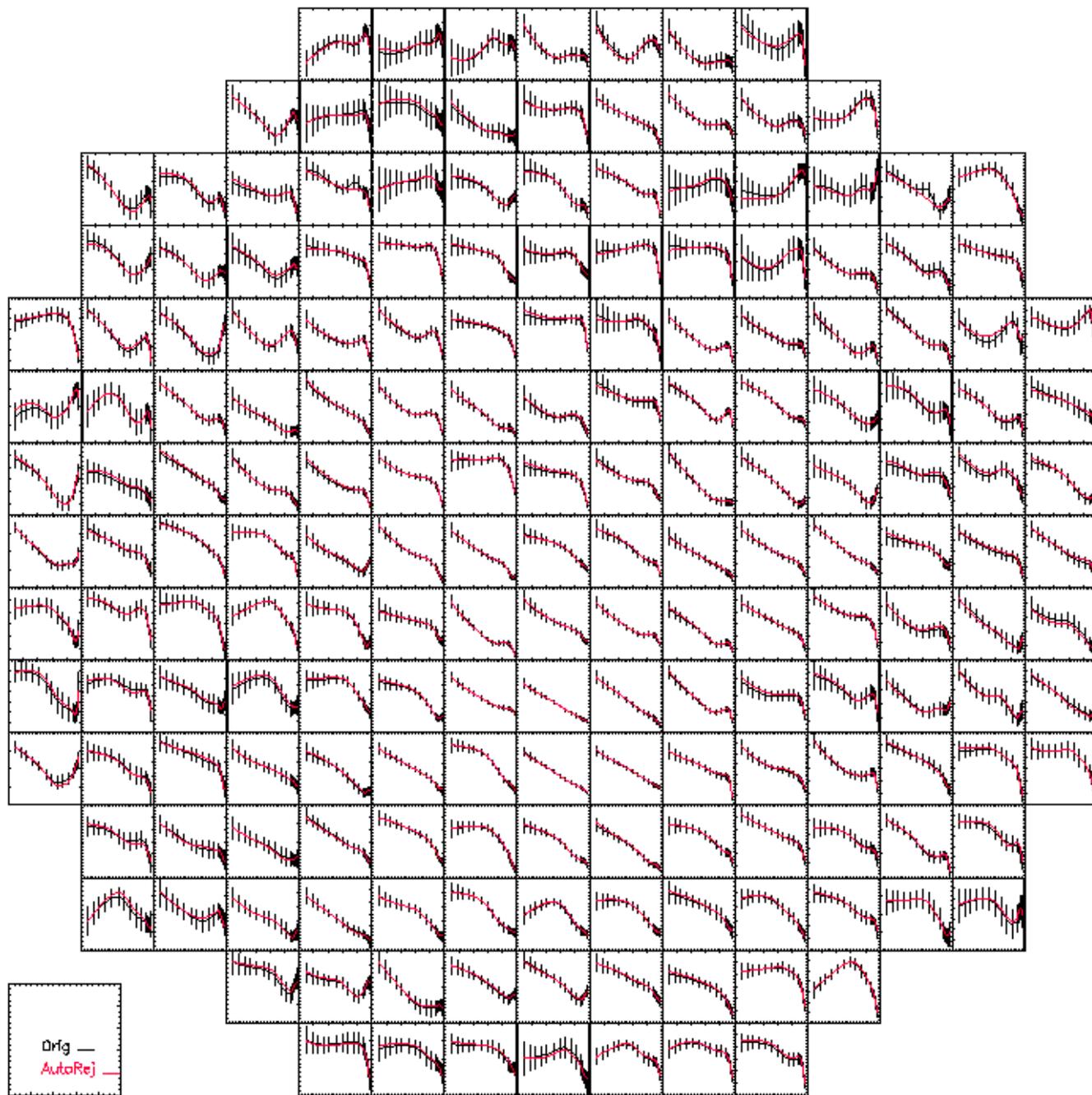


Figure 4 A ring diagram analysis was used to obtain flow velocities. Here the x component of velocity is shown. The velocities and error bars from the standard dataset are plotted with the velocities from the AIR dataset over plotted in red.

# APPENDIX B

## Program Source Files

dayscan (uses oscmd, and the Bevington library (polfit))  
mk\_dayscan  
dayscan.x stats.x  
center.x mnxsig.x  
swath.x strmid.x  
mask.x ringbell.x  
difmask.x

viewlog mk\_viewlog  
viewlog.x vwp45.x  
getrjd.x vwp6.x  
fitmod.x vwpltp.x  
vwp1.x pltv.x  
vwp2.x mnxsig.x  
vwp3.x strmid.x

reject mk\_reject  
reject.x strmid.x  
getrjd.x  
reject.dat

useds mk\_useds  
useds.x strmid.x  
rdrjct.x doymd.x  
sdfilt.x

Other utilities and scripts have played a major role in the development of AIR but are not formally part of it.

readlog mk\_readlog  
readlog.x strmid.x  
pltv.x mnxsig.x  
pltp.x

readlog.pro

runds.cl, runrj.cl, runuseds.cl

showfill mk\_showfill  
showfill.x ungz7  
strmid.x ungzhdr

sorttape mk\_sorttape  
sorttape.x doymd.x