MASSACHUSETTS INSTITUTE OF TECHNOLOGY
HAYSTACK OBSERVATORY
WESTFORD, MASSACHUSETTS 01886
February 27, 2024

To: VGOS Group
From: Brian Corey
Subject: Data reduction for measurements of orientation-dependent phasecal delay at GGAO

1. Introduction

A limiting factor on the accuracy with which a geodetic VLBI station position can be determined is antenna-orientation-dependent variations in the electrical delay of the cable (or fiber) that carries the reference signal to the phase calibration signal generator in the receiver. In the absence of a dedicated cable delay measurement system (CDMS), a series of specialized measurement sessions has been carried out over the past decade to measure these variations for the 12-m antenna at the Goddard Geophysical and Astronomical Observatory (GGAO). This memo describes the data acquisition strategy during a session and the data reduction process for generating plots of delay versus antenna azimuth and elevation.

2. Phase calibration system

VLBI systems typically include a phase calibration (a.k.a. phasecal or pcal) system designed to measure temporal and frequency variations in instrumental phase and delay response that, if left uncorrected, would degrade the precision of VLBI observables. Most such systems inject a continuous series of narrow pulses with fast rise/fall times (< 50 ps) into the signal path as close as possible to the antenna feed. If the delay in the signal path that follows the feed (i.e., receiver, cable/fiber to the control room, and control room electronics) varies, those variations can be detected as changes in the epoch when the pulses arrive in the control room. In turn, those variations can be used to apply instrumental phase and delay corrections to the signal received from the radio source being observed with the antenna.

The pulse rate of such systems is typically one pulse every 1, 0.2 or 0.1 μs ($1x10^6$, $5x10^6$ or $10x10^6$ pulses/s). In the frequency domain, such a pulse train corresponds to a set of tones (sinusoids) at harmonics of 1, 5, or 10 MHz, respectively, extending above 15 GHz.

3. What's the problem? What's the solution?

In order to generate a calibration signal coherent with the rest of the VLBI electronics, the phasecal signal generator is driven by a reference signal coherent with the frequency standard (usually a hydrogen maser) used to drive the other electronics. Because the phasecal generator is located on the antenna and the maser on the ground, the reference signal must be sent via coaxial cable or optical fiber up the antenna. If the propagation time through that cable or fiber varies over time, it will change the epochs at which the pulses are generated. In turn, the pulse epochs seen in the VLBI signal in the control room will also change, and those changes may be mistaken for variations in the VLBI instrumental delay, leading to erroneous "corrections" to the radio source signal.

Of particular concern in geodesy are systematic variations in the reference signal delay that correlate with antenna orientation. If these variations are not accounted for by subtracting them from the pulse delays observed in the control room, they can shift the apparent location of the antenna in geodetic solutions.

Many antenna systems have a CDMS to measure the delay in the reference signal between maser and phasecal generator. Because GGAO does not have one, however, an alternative method has been adopted

of determining the dependence of reference signal delay on antenna orientation in special-purpose measurement sessions.

The basic measurement strategy is to slew the antenna as rapidly as possible in azimuth or elevation while logging phasecal data, then look for systematic behavior in plots of phasecal delay vs. azimuth or elevation. Rapid movement is essential to reduce the effect on the measurements of thermally driven, slow drifts in delay that originate in the cables or electronics. At GGAO, the 5 MHz reference signal is carried to the phasecal generator over coax, and VLBI signals above 5 GHz are brought from the receiver to the control room over optical fiber. Because delay variations induced by twisting or bending as the antenna moves are generally much larger for coax than for fiber, orientation-dependent delay variations in phasecal signals above 5 GHz should therefore be dominated by delay variations in the coax carrying the reference signal up to the phasecal generator.

4. GGAO block diagram

Important features of the GGAO electronics relevant to the measurement of phasecal delay are shown in Figure 1. The signal flow is as follows:

a. 5 MHz from the hydrogen maser is sent over coax from the control room to the phasecal signal generator in the receiver on the antenna.
b. From the 5 MHz, the phasecal generator creates harmonics of 5 MHz (i.e., pure tones at N x 5 MHz) to above 15 GHz.
c. The phasecal signal is inserted into the signal path carrying the horizontal (H) and vertical (V) polarization signals out of the antenna feed.
d. Following amplification in a low-noise amplifier (LNA), each polarization signal is split into two frequency ranges: low range from 2 to 5 GHz, high range from 4 to 12 GHz.
e. The low frequencies are carried to the control room over coax, the high over optical fiber.
f. The RF distributor splits the high frequency signal for each polarization into 3 identical signals.
g. A UDC (up-down converter) translates a 512-MHz-wide portion of each RF signal down in frequency to 512-1024 MHz.
h. The RDBEs (ROACH Digital Back End) apply to each signal a 2nd-Nyquist-zone bandpass filter 512 MHz wide and then convert it from analog to digital at a sample rate of 1024 MS/s. Note that, because the sampling is done in the 2nd Nyquist zone, the frequency order in the digitized frequency domain is flipped relative to the order in the original analog domain. For example, a frequency of 513 MHz in the analog signal appears at 1024 – 513 = 511 MHz in the sampled data, and 1023 MHz analog appears at 1024 – 1023 = 1 MHz sampled.
i. One of the digitized signals in each RDBE is accumulated in 1024 time bins spaced 1 sample apart, with binning done modulo 1 μs. The accumulated totals are stored once per second in a multicast log. The multicast data are the information source for estimating phasecal delay.

5. Brief overview of data reduction process

For each polarization, four 512-MHz-wide signals are digitized and then binned, modulo 1 μs, in the RDBEs – see item (i) in the preceding section. The binned time series in the RDBE multicast logs are then transformed, via Fast Fourier Transform (FFT), to produce estimates of the amplitudes and phases of the 103 phasecal tones in each band/polarization for each second. A delay model is then fit to each set of phases vs. frequency. These delays are plotted against the azimuth and elevation of the antenna.

For convenience, the four bands are labeled A-D, with band A having the lowest RF frequency range and D the highest. Table 1 lists their RF frequency ranges and ancillary information for all measurements performed since 2014.

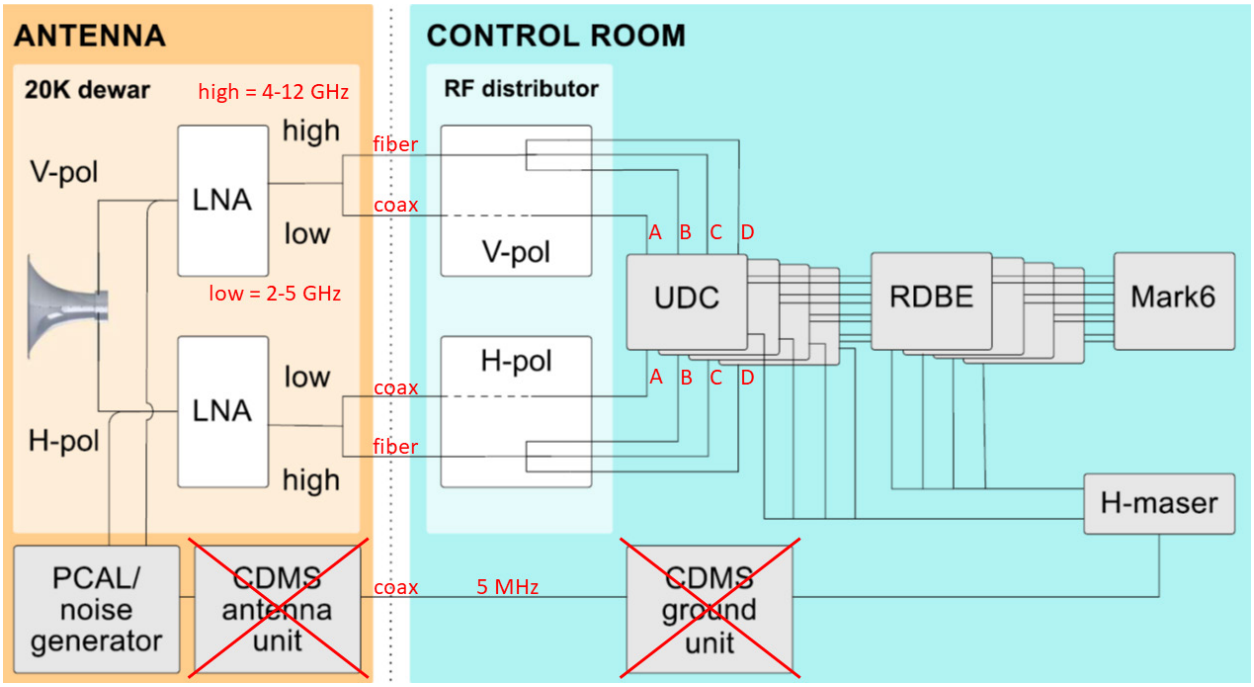

Figure 1. Simplified block diagram of antenna and control room electronics at GGAO. Original drawing is figure 2 in https://doi.org/10.1029/2018RS006617 and depicts the electronics at Westford. The main difference between GGAO and Westford is that GGAO has no CDMS ground or antenna unit. Annotations in red are added to indicate the frequency ranges of the low and high frequency bands sent from the receiver to the control room over coaxial cable and optical fiber, respectively, and the labels (A-D) of the four 512-MHz-wide bands recorded on the Mark6.

Table 1. Four frequency bands from which phasecal signal is extracted

band label	*RF frequency span (MHz)*	*conduit from rec.*	*RDBE IP address*	*mid-band tone RF freq. (MHz)*
A	2984.4-3496.4	coax	...120	3240
B	5224.4-5736.4	fiber	...122	5480
C	6344.4-6856.4	fiber	...124	6600
D	10184.4-10696.4	fiber	...126	10440

6. Deriving phasecal tone phases and amplitudes from multicast data

An RDBE multicast log can contain only one binned phasecal time series, so it is not possible to log the time series for both polarizations simultaneously. The workaround is to alternate the input stream to the binning process every second between H and V, with even-numbered seconds containing H data (a.k.a. IF0) and odd-numbered containing V (a.k.a. IF1).

For each measurement session there are 4 multicast files, one for each band. Default file names are *rdbe30_mon-lite_dat_192.168.1.XXX-YYYYMMDD.log*, where *192.168.1.XXX* is the IP address of the RDBE that generated the file (values of *XXX* are in column 4 of Table 1) and *YYYYMMDD* is the date when data acquisition began (e.g., 20210302 = 2021 March 2).

Up until 2021, it was possible for a Haystack user to log into the 'monkey' server at GGAO and upload the multicast logs (along with the Field System log – see section 8) to Haystack. For security reasons, this procedure was changed in 2021 to the current arrangement, where GGAO personnel upload the logs to an intermediate repository at https://www.box.com/, from which Haystack then downloads them.

Once the four multicast logs are copied to Haystack, script *make_ggao_pcalVSazel_files.script* is run on them. It has multiple functions:

a. Rename logs to a more compact form *rdbe.G.[A-D].YYYYMMDD.log*.
b. Run *pcalfft_noplot.py* on each log to generate, via FFT, phasecal tone amplitude and phase time series files, with one set of 103 tone phases and one set of 103 tone amplitudes per second.
c. Run fortran program *even_odd_epochs* on each amplitude and phase file to create files containing only epochs with even- or odd-numbered seconds, and hence only H or V data.

Script output files that are subsequently used to estimate phasecal delays have names *rdbe30.G.[A-D].YYYYMMDD[even/odd].log_[amp/phase]* (e.g., *rdbe30.G.B.20210302odd.log_amp*).

7. Calculation of phasecal delays

The desired quantity for plotting against azimuth or elevation is the time delay of the phasecal signal, i.e., how much the arrival time of the signal as recorded in a multicast log varies as the antenna is moved. There are two ways to calculate the delay, both of which are based on the measured phases of the phasecal tones:

- Group delay = d(phase)/d(frequency)
- Phase delay = phase/frequency at a reference RF frequency (i.e., frequency before down-conversion to lower frequency in the UDC)

Some tones are known to be corrupted by "spurious" signals that interfere with the true phasecal tone. At GGAO they are harmonics of 100 MHz as well as individual frequencies 3090, 3360, and 5625 MHz. These tones are flagged by assigning a weight of zero to them in the delay fitting process.

For each second's worth of phasecal amplitudes and phases for a given band and polarization, the delays are calculated by these steps:

a. Subtract from each tone phase the phase for the first epoch. This removes a large phase curvature vs. frequency which, if left in place, would complicate the delay estimation.
b. With uniform weights except for "spurious" frequencies assigned a weight of zero, fit a straight line to tone phase vs. RF frequency via least-squares. (In the current incarnation of the data reduction, phasecal amplitudes are not used.)
c. Group delay = slope of line,
   Phase delay = (line intercept at reference frequency) / (reference frequency),
   where 'reference frequency' is the RF frequency of the phasecal tone closest to the middle of the band (these frequencies are listed in column 5 of Table 1).

Because each epoch in the multicast logs corresponds to the end of the second over which the phasecal data were accumulated, the epochs associated with the group and phase delays are shifted earlier by a half second so that they correspond to the mid-point of the accumulation intervals.

The delay calculation is done in Supermongo (SM) macros *setup* and *mcpcal*, which is called by *setup*. These macros, and all other macros listed in this memo, are in file *command_GGAOpcalVSazel.sm*.

8. Antenna azimuth and elevation time series

Antenna orientation is controlled via the Field System. Up until late 2023, antenna commands were entered manually by the operator at GGAO. Starting with the measurement session on 2023 Nov 1, control was completely automated, courtesy of Ed Himwich.

Information about antenna pointing direction is contained in the Field System log. The current file name convention is *cdmYYDDD.log*, where *YY* is the last two digits of the year and *DDD* is the day of year.

At Haystack, Fortran program *decodeGGAOazel* is run on a Field System log to create a more compact file containing just the epochs, azimuths and elevations. The actual (as opposed to, e.g., commanded) azimuth and elevation are contained in log entries like

```
2021.061.17:52:30.03#trakl#<ReadGroup: 0(0x0) 59275(0xe78b)
64349964(0x3d5e70c) 450547(0x6dff3) 753381(0xb7ee5) 821323776(0x30f46800)
4096(0x1000) 0(0x0) 0(0x0) -10(0xfffffff6) -38(0xffffffda) 49(0x31) 0(0x0)
0(0x0)>
```

The epoch is given in the first 20 characters, the elevation in units of 100 micro-degrees in the 5th field (the underlined '450547' in this example), and the azimuth in units of 100 micro-degrees in the 6th field (underlined '753381'). The resulting output record from *decodeGGAOazel* for this epoch is

```
#DDD HH MM SS.SS   az(deg)  el(deg)
 061 17 52 30.03   75.3381  45.0547
```

In order to have matching epochs between the azimuth/elevation (hereafter abbreviated 'az/el') positions and phasecal delays, the az/el data are interpolated to the phasecal delay epochs. This is done with SM macro *interp_azel2pcal*, which is called by macro *setup*.

9. Time series plots

Useful for diagnosing measurement (or software!) problems are time series plots of various calculated phasecal quantities and of azimuth and elevation. Figure 2 is an example generated by SM macro *timeseries*.

From top to bottom, the panels are:

1. RMS (root-mean-square) of post-fit phase residuals (in degrees) for band A
2. Band A group delay (ps)
3. Band A phase delay (ps)
4. RMS of post-fit phase residuals (deg) for bands B (red), C (green), and D (blue)
5. Group delays (ps) for bands B (red), C (green), and D (blue)
6. Phase delays (ps) for bands B (red), C (green), and D (blue)
7. Azimuth (deg), with limits $-180° \leq$ azimuth $\leq 360°$
8. Elevation (deg), with limits $7° \leq$ elevation $\leq 87°$

Because the tone phases are initialized to zero prior to the least-squares fit, the initial value in panels 1-6 is always zero.

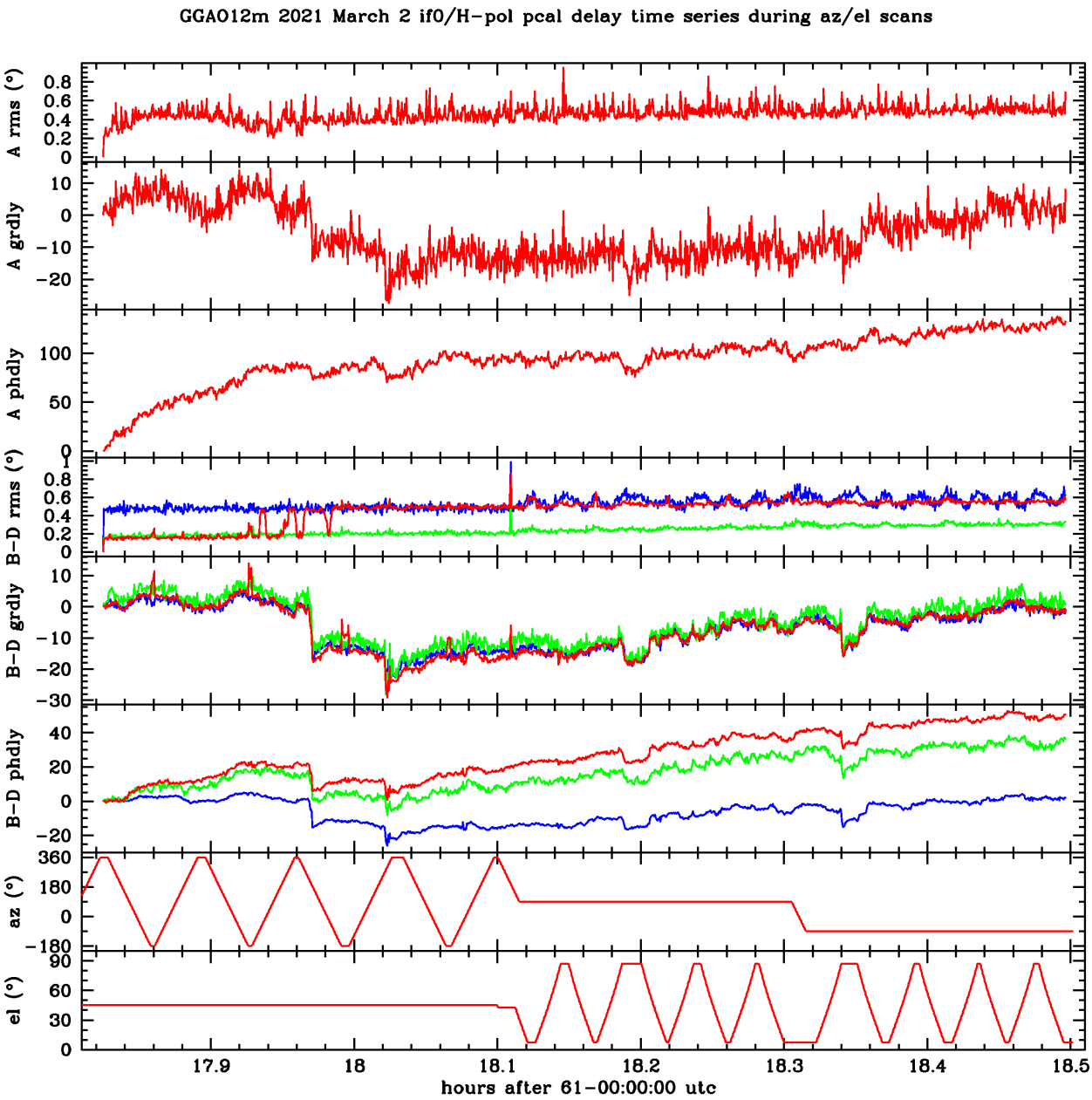

Figure 2. Example of time series plot for H-polarization data, created with SM macro *timeseries*.

10. Delay drift removal and plots of delay vs. az/el

As can be seen in Figure 2, there is no obvious correlation between delay and az/el. The reason is a combination of the orientation-dependent variations being small in March 2021, significant noise in the measured delays, a few jumps of order 10 ps of unknown cause, and large drifts on time scales of tens of seconds to minutes that are uncorrelated with antenna motion. These drifts are mainly thermally driven and arise in the cables and electronics. To the extent possible, they need to be removed in the data analysis, to make any underlying dependence on az/el more visible.

The method used to remove drifts is based around identifying pairs of consecutive scans in opposite directions in azimuth or elevation that start and end at the same az/el position. Because the position is the

same at the start and end, the az/el contribution to the delay should likewise be the same at the start and end. Any difference in the measured delays at these two epochs is then ascribed to a linear drift in delay over time. For example, in Figure 2, there are four pairs of azimuth scans that begin and end at 360°. The third pair starts at 17.96 hr with a counter-clockwise (CCW) scan and finishes with a clockwise (CW) scan that ends at 18.025 hr. The model for the temporal drift is a line connecting the delays at those two epochs. This model is then subtracted from the delays between the two epochs.

The standard sequence of antenna movements in a measurement session is as follows:

- Four CCW+CW azimuth scan pairs, with elevation fixed at 45° and with each scan pair starting at 360° azimuth, spinning CCW to -180°, then returning CW to 360°.
- Four down+up elevation scan pairs, with azimuth fixed at 90° and with each scan pair starting at 87° elevation, dropping to 7°, then returning to 87°.
- Same as the previous four elevation scan pairs except with azimuth fixed at -90°.

(In the data taken on 2021 March 2, this sequence was inadvertently not followed precisely, as can be seen in Figure 2: instead of the elevation scan pairs starting at 87°, they started at 7°**.** This was accommodated in the data analysis by discarding the first 'up' and last 'down' scan at each of the two azimuths, with the result that there were only three down+up scan pairs at each azimuth.)

The process of identifying epochs at the beginning and end of an azimuth or elevation scan pair makes use of so-called turning points. A turning point is defined as an epoch, or set of consecutive epochs, when the antenna is nearly stationary in both axes and either the azimuth is 360° (an azimuth turning point) or the elevation is 87° (an elevation turning point). Per the preceding paragraph, one azimuth (elevation) turning point will immediately precede each azimuth (elevation) scan pair, and another will immediately follow. Turning points are identified in the az/el time series by SM macro *turningpts*, which writes their epochs into file *turningspts.dat*. The delays at the turning points define, for each scan pair, the linear drift model. The list of turning points is compared manually with the az/el plots, to determine which turning points bracket good azimuth or elevation scan pairs. (A "bad" scan may be one in which something untoward occurs in the phasecal delay or antenna pointing, such as a glitch in the delays or a too-long pause in data-taking between the CCW and CW scans of an azimuth pair. An example of the former is the jumps in group and phase delays at 17.97 and 18.02 hr during the third azimuth scan pair of Figure 2. The problem with a too-long (minutes) pause is that it makes the drift model less reliable.) Once the turning points for good scans have been identified, macro *delayvsazel* plots group and phase delay, corrected for drift via macro *driftmodel*, vs. azimuth and elevation for one band and one polarization.

This turning point method is admittedly cumbersome. Better approaches to identifying the start and end of a good scan pair could be explored as part of a future project to automate the data reduction process. If possible, any such project should use a single, more modern software platform than the current mishmash of Fortran, SM, and Python.

11. Median delays for bands B-D

Because the delay through an optical fiber is generally much more stable than through a coaxial cable, especially when subject to torsion or bending as in an antenna wrap, the most reliable estimate of delay variations in the 5 MHz cable is likely to be the median delay over bands B-D (which are brought down the antenna on fiber) in azimuth and elevation angular bins. Plots of the delays for these three bands, together with solid points denoting median delays over all scans at selected angles in the three bands, are created with SM macro *BCDmedians*. Figure 3 is an example of the output from this macro. Note that medians are calculated separately for CW and CCW azimuth scans and for up and down elevation scans because hysteresis effects may cause the delay at a given az/el to depend on the direction in which the

antenna moved to reach that position. The macro also writes text files listing median group delay vs. azimuth and elevation. These files are named *mean_delay_vs_[az/el].gs.YYMMDD.if[0/1].txt*.

While the scan-to-scan scatter is typically comparable for group and phase delays (e.g., Figure 3), the scatter in phase delay is occasionally much larger than in group delay. The larger phase delay scatter is due to larger phase delay drifts on time scales of minutes that are not linear in time and hence are not fully removed by the procedure outlined in Section 10. Drifts in the UDC local oscillator phase, which affect phase delay but not group delay, are believed to be the cause of the excess phase delay drifts. Note that in panels 3 and 6 of Figure 2, there are significant differences among the four bands in the long-term phase delay drifts. These differences arise from different drift rates (degrees/second) in the four local oscillator phases and from the different reference frequencies (column 5 of Table 1) used in calculating phase delay.

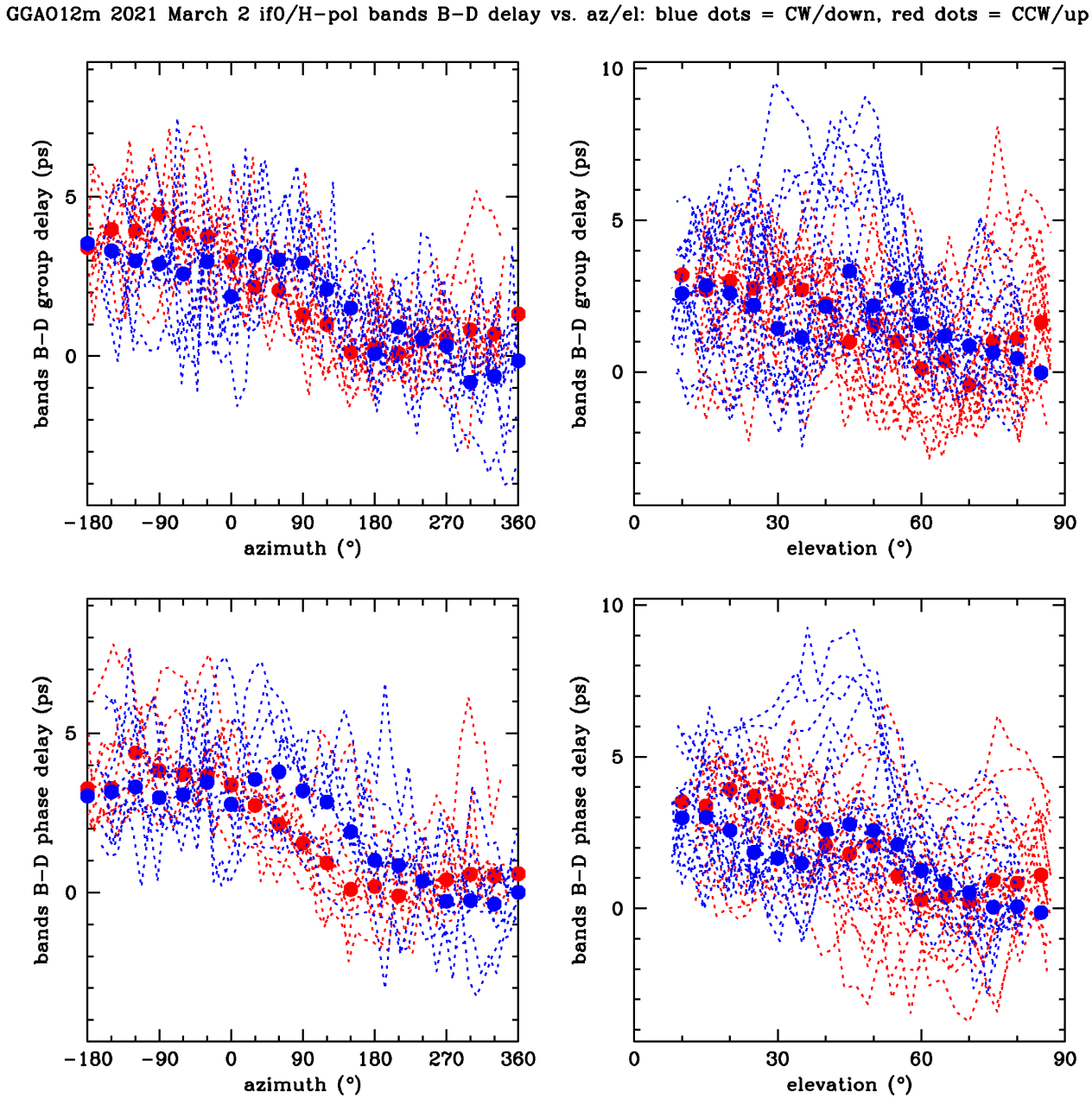

Figure 3. Median H-polarization delays for bands B-D vs. az/el, created with SM macro *BCDmedians*.