CHILES

The COSMOS HI Large Extragalactic Survey: Pipeline Development

D.J. Pisano (WVU/UCT)
on behalf of the CHILES collaboration

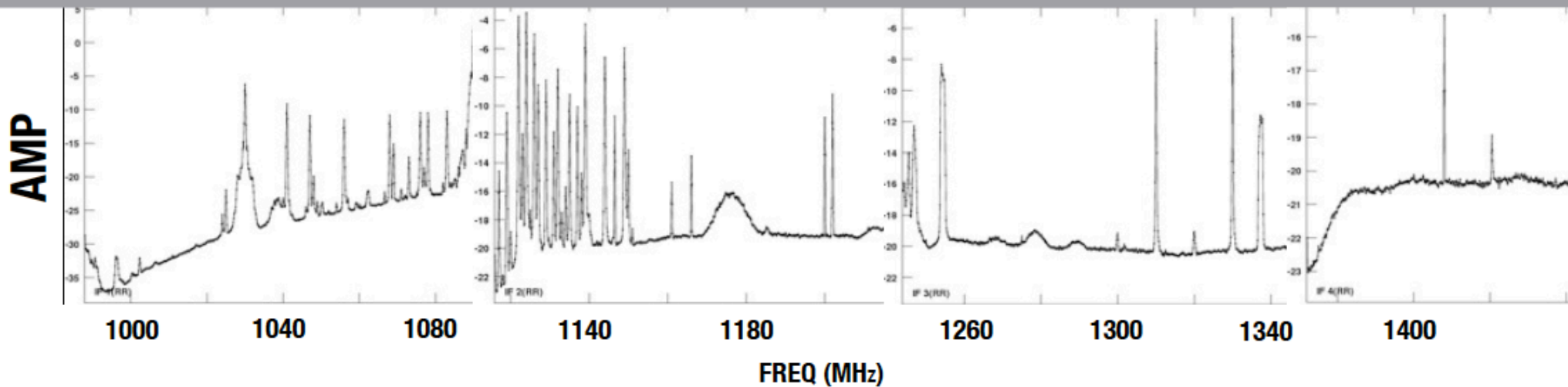# CHILES

- Deep HI survey (1002 hr) with VLA in B array.
- Final $\sigma \sim 50$ μJy per 31 kHz channel.
- 30720 channels spread across 15 spectral windows (~950-1420 MHz).
- Beam: 6"-8", FoV: ~ 30'-45'
- So far: 178 hrs in 2013-2014, 210 hrs in 2015.
- In 16A: 54/60/120 hrs allocated at priority A/B/C.
- K. Hess will provide more details on survey, science goals.

Challenges for CHILES

- Data volume: 6 hours of data is ~800 GB (in CASA). Too much to process by hand.

- RFI: Lots of strong and weak RFI throughout the band. Affects calibration & image quality.

- Together these require novel algorithms to efficiently reduce CHILES data.

Challenges:
RFI in L-band

PILOT

AMP

FREQ (MHz)

AIRPORT

VLA
MODEM

RADARS

GLONASS &
RADARS

* Long spacings (5 > km) are not as affected by RFI

Challenges:

RFI in L-band

The biggest challenge for the survey is manmade radio frequency interference from radar and navigation satellites.
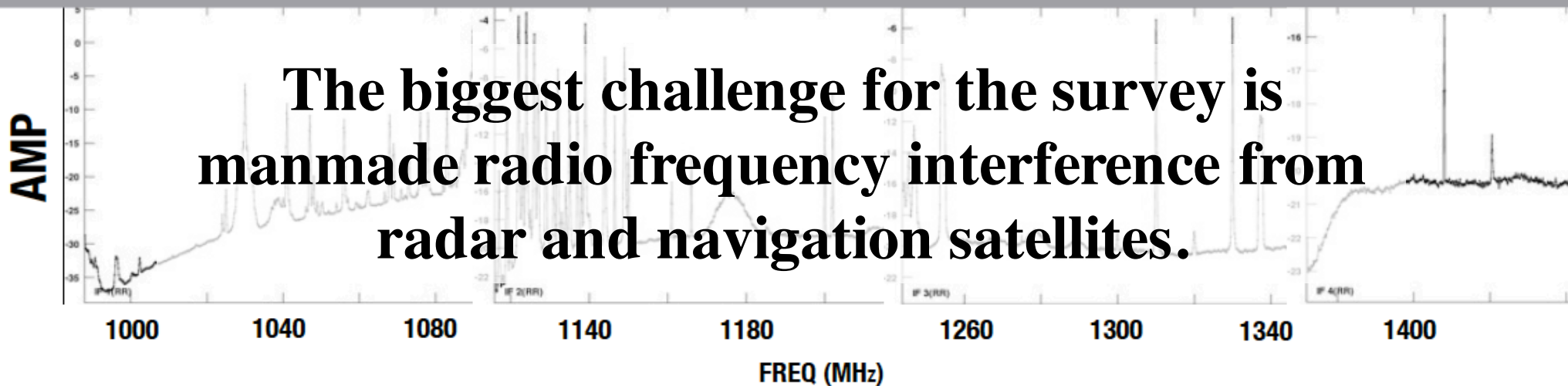
PILOT

AIRPORT   VLA MODEM   RADARS

GLONASS & RADARS

☀ Long spacings (5 > km) are not as affected by RFI

Pilot Survey & Pipeline

- 60 hours of pilot data taken in 2011 (Fernandez et al. 2013).

- ν = 1190-1426 MHz; z=0-0.19

- 16384x31.2 kHz channels spread over 32 spectral windows.

- Pipeline written in AIPS since CASA did not have needed tools yet.
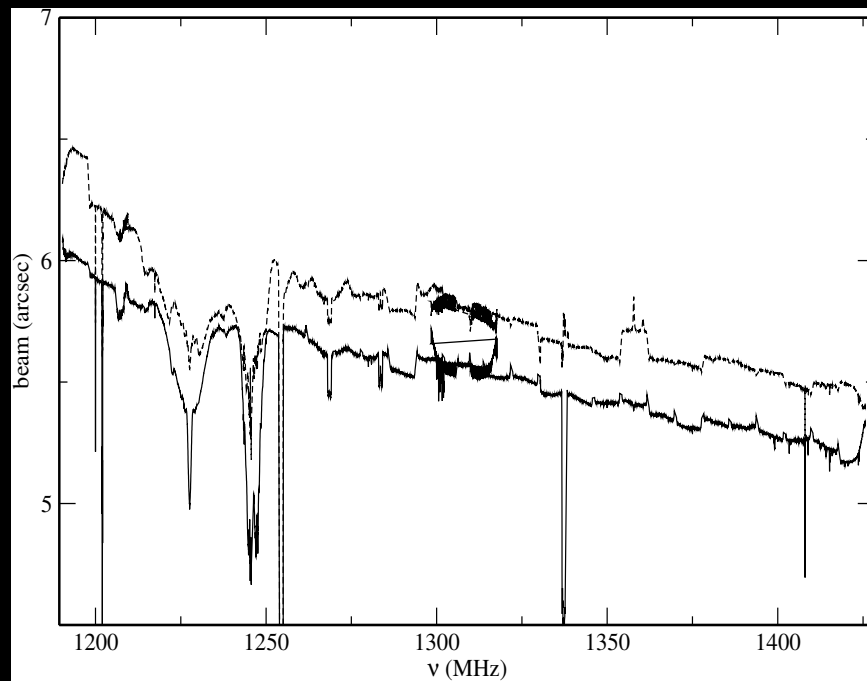
AIPS pipeline
(written by X. Fernandez)

- Each spectral window reduced independently.
- Does initial calibration (bandpass, complex gain) in order to run RFLAG. Then final calibration is done and source is flagged.
- RFLAG was written by E. Greisen to do automated RFI flagging mimicking how you would do it by hand. Allows you to flag based on time-averaged data and extend flags.
- Took about 48 hours to reduce 6 hours of data.
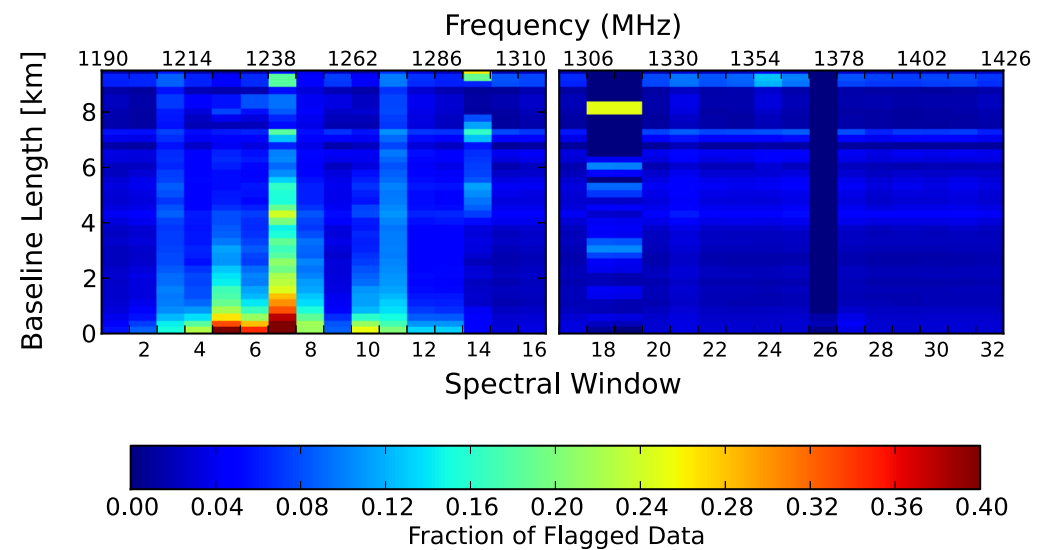- Some manual flagging was necessary requiring pipeline to be re-run on some windows.

Rms noise as function of frequency looks really good
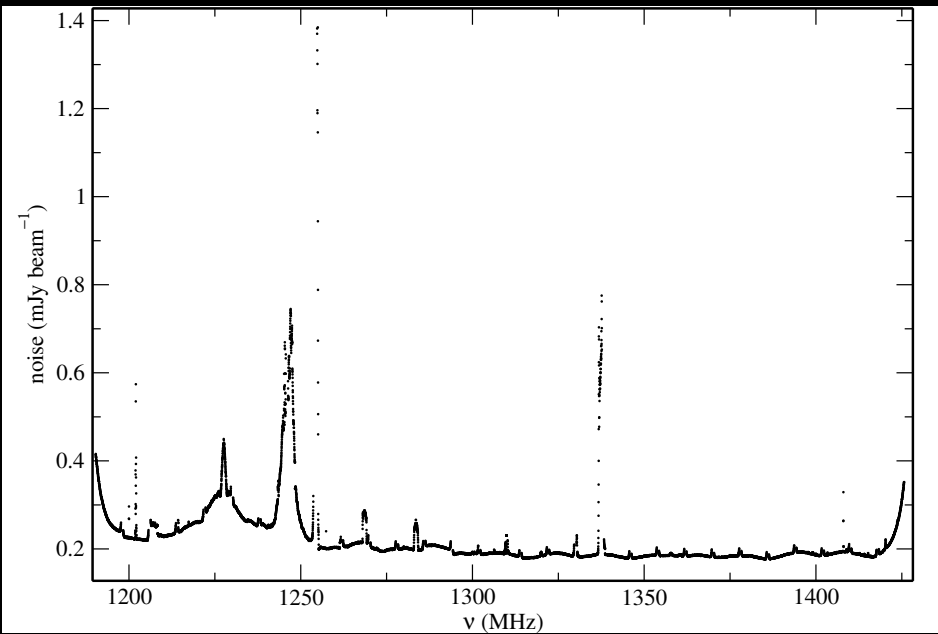

But, baseline distribution matters

Synthesized beam

Full Survey Pipeline 2013-4
(E. Momjian)

- Full survey has wider bandwidth, more channels so AIPS was not a practical solution. Spws reduced together.
- For first 178 hours, we used a modified version of the NRAO continuum pipeline.
- Alternates calibration and flagging in 4 stages.
- Takes 60 hours to reduce 6 hours of data!
- Must be run in entirety; no way to assess data quality before complete.
- Inspection/additional flagging/re-running pipeline takes roughly 1 astronomer week (few hours a day).
- CASA tasks lacked same functionality of AIPS tasks.

Noise Variation for 6 hour Observation
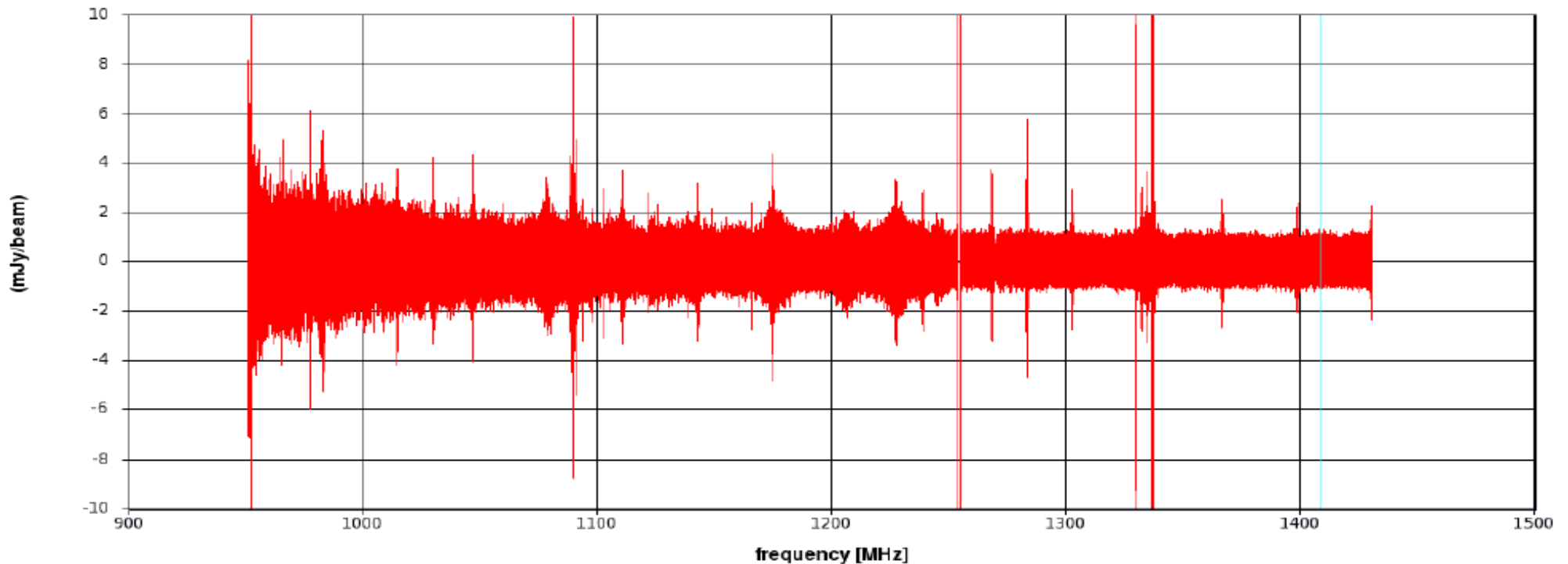
- First 31000 channel cube ever made!
- 15 x 32 MHz dual polarization spectral windows covering 941 – 1421 MHz (z = 0-0.509)
- 2048 channels per spectral window (~ 3.3 km/s at z=0)
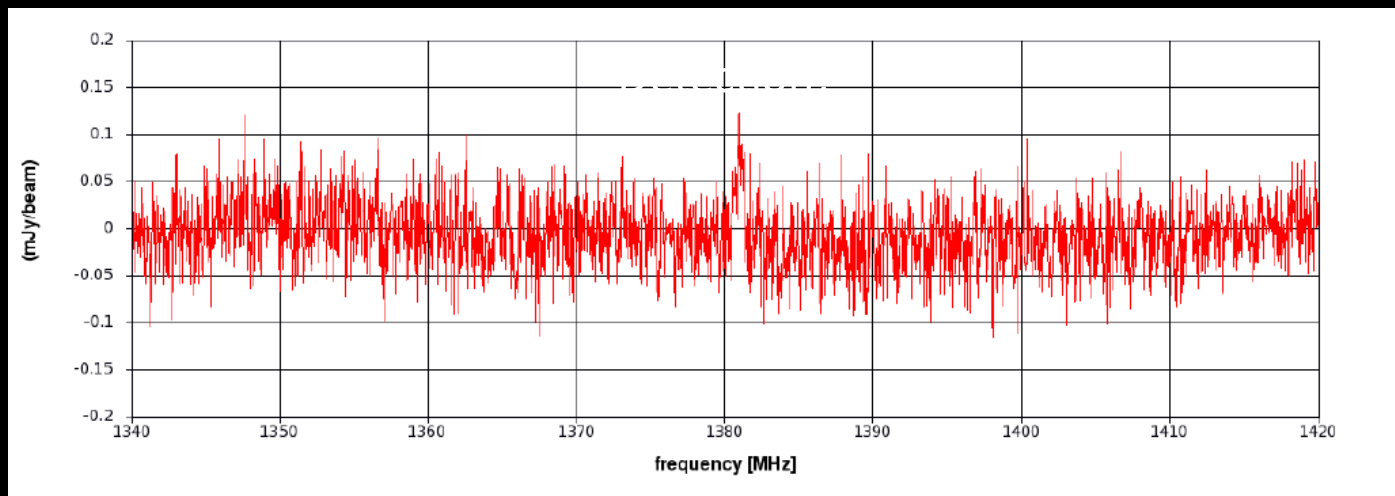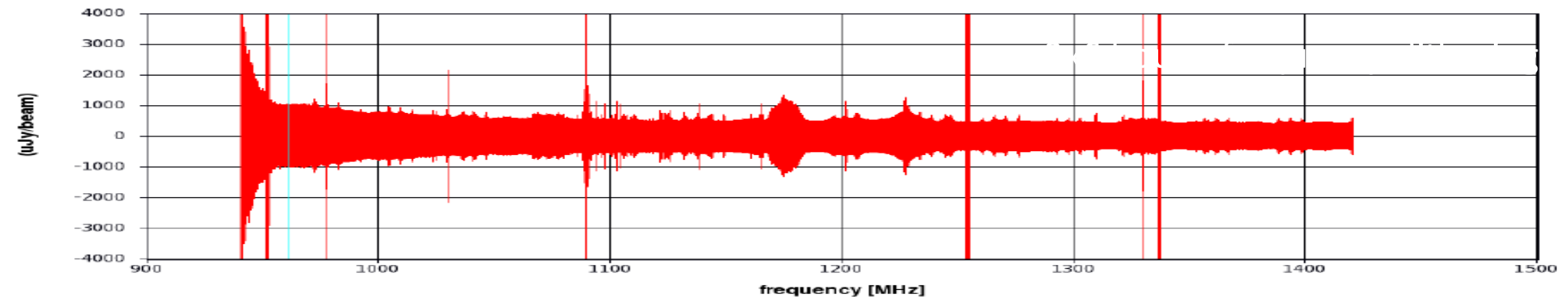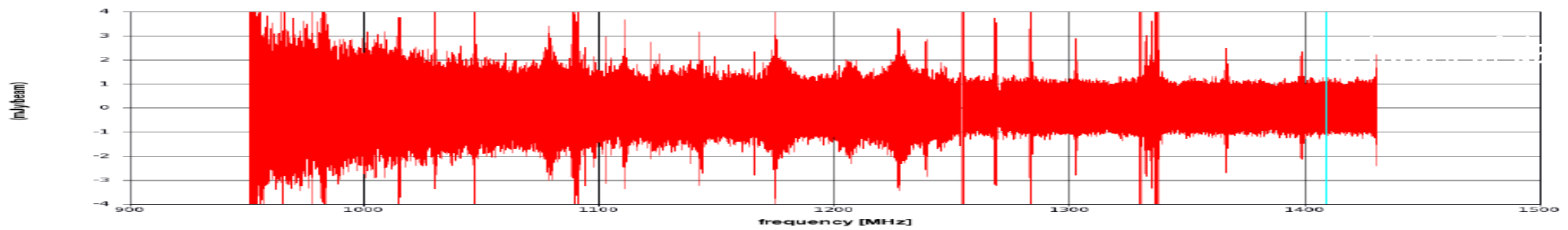
The good news is.. RFI less below 1190 MHz, frequency dithering works beautifully

New CHILES pipeline
(D.J. Pisano et al.)

- Inspired by AIPS pipeline with some code taken from initial CASA pipeline.

- Designed specifically to deal with CHILES data (not general).

- Broken into 4 modules:
    - Import data, apply online flags
    - Calibrate & flag flux cal., derive BP solutions.
    - Calibrate & flag phase cal., derive complex gain solutions.
    - Flag target.

- Initial data assessment done via logs from NRAO's default pipeline (run on all data).

- All modules produce output to verify data quality.

- Still developing, latest version available at github.com/djpisano3/chiles-pipeline.git

- To process 2 hours of data takes less than 5 hours!
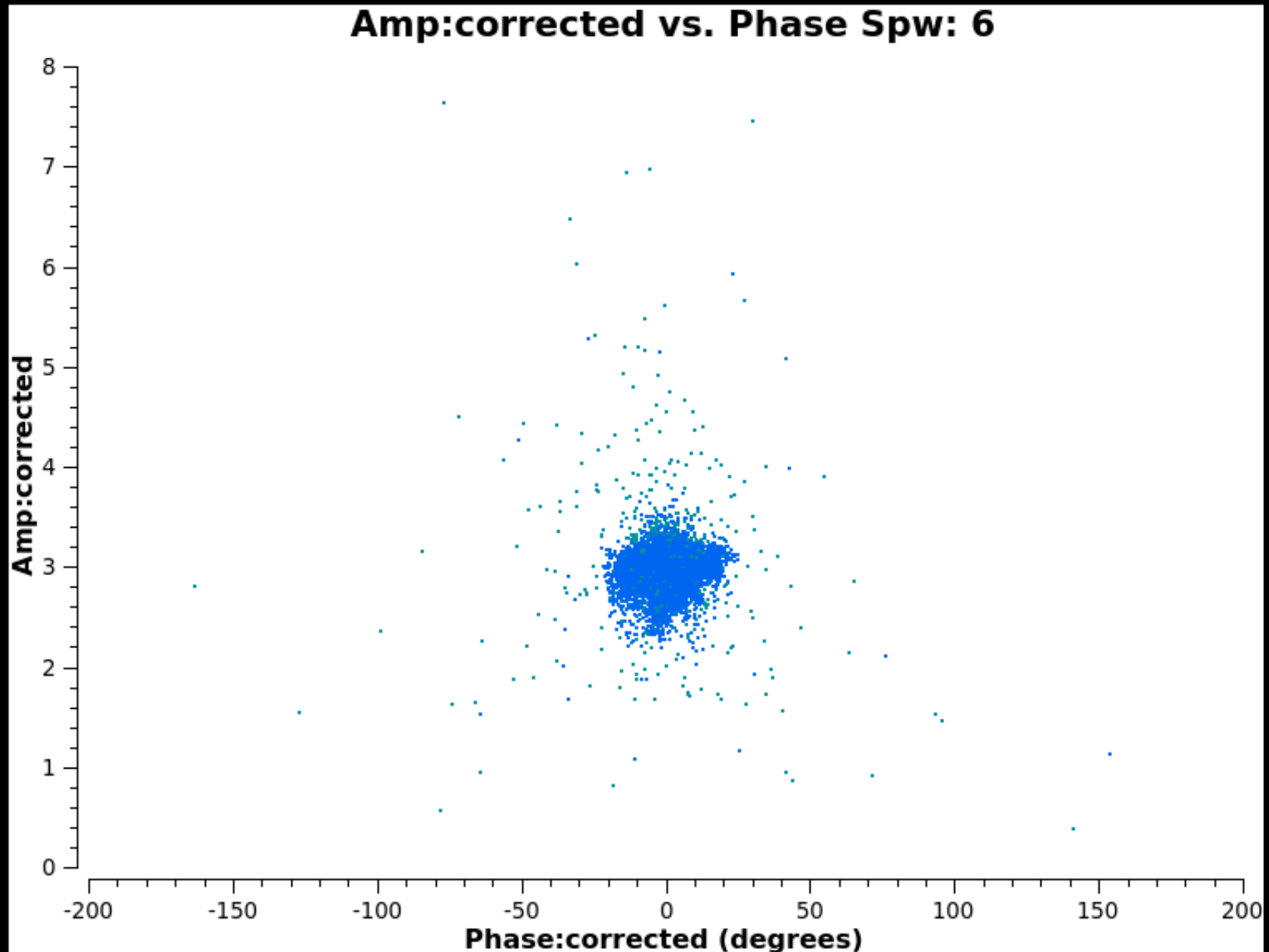
Data Quality Assessment
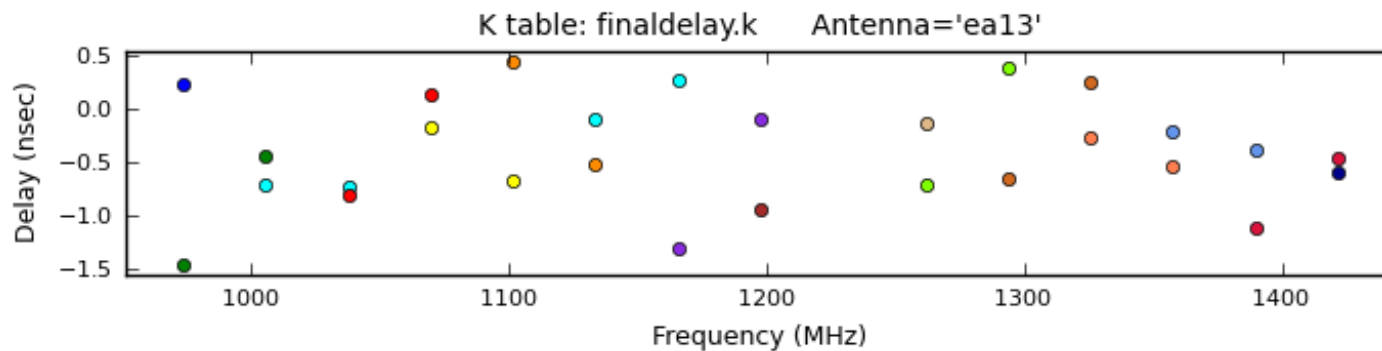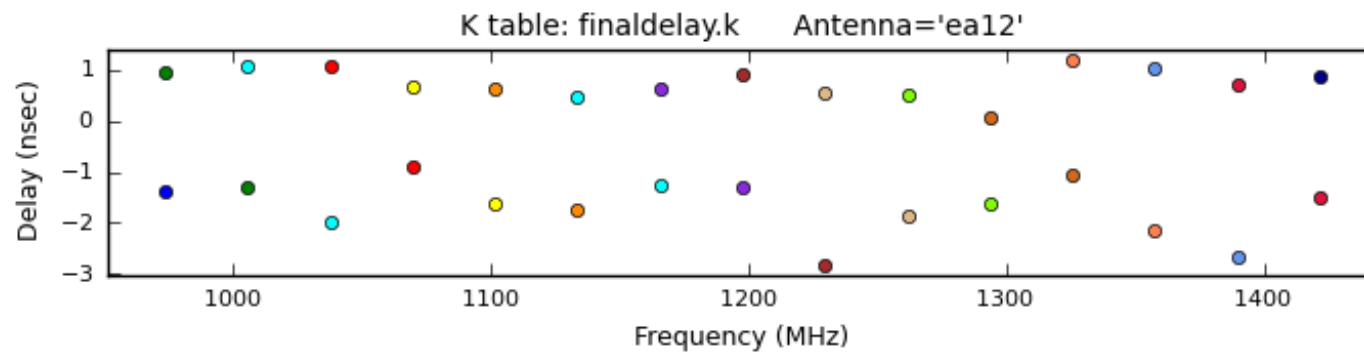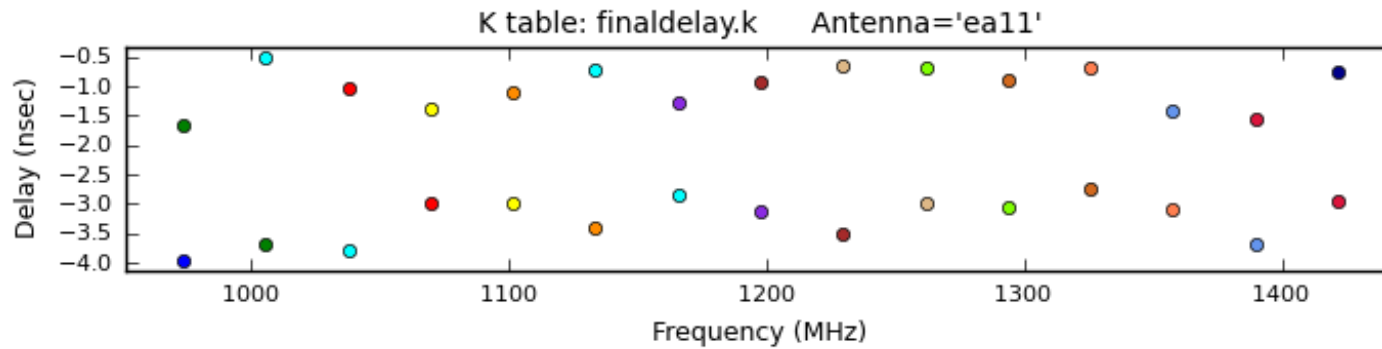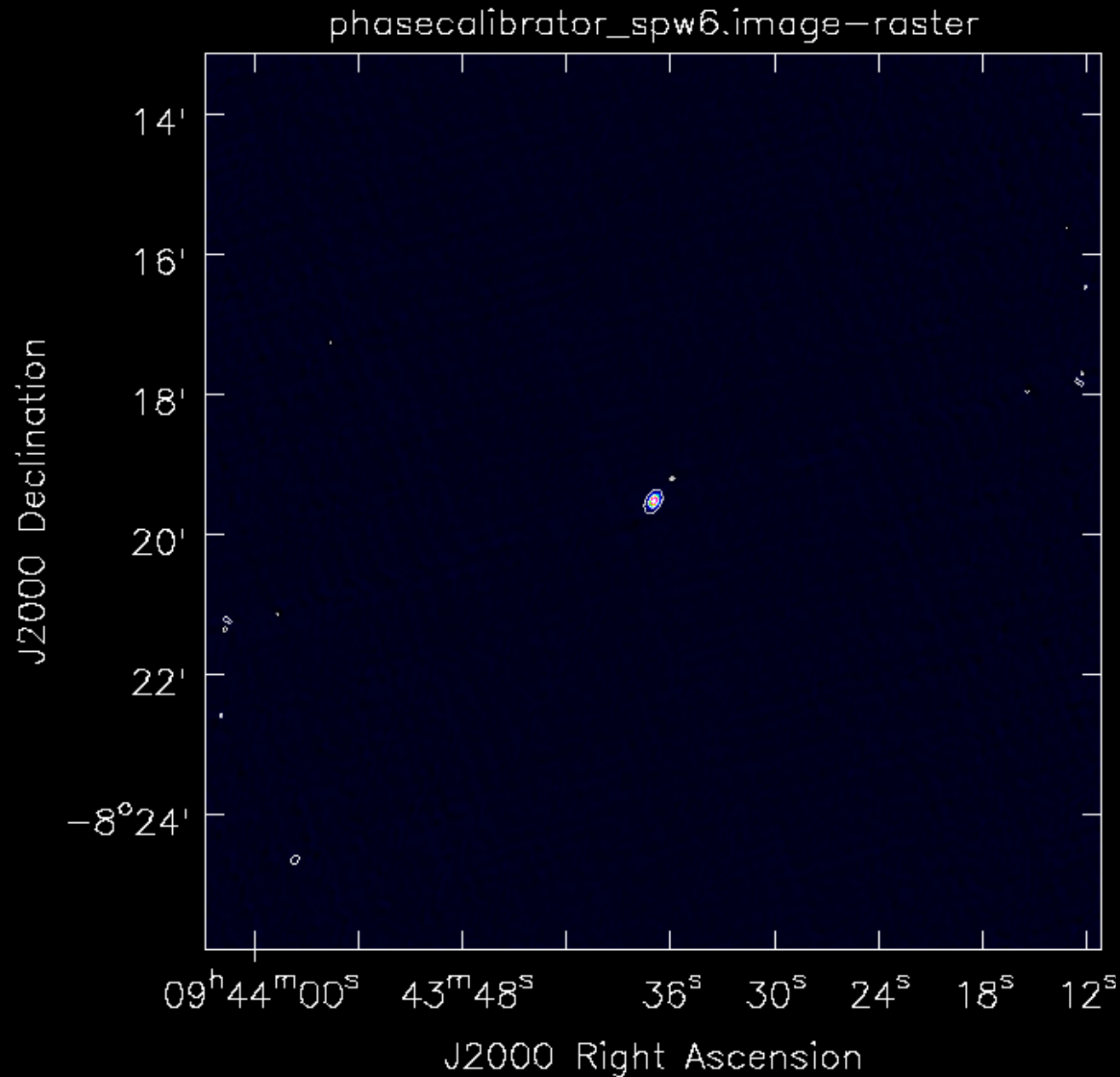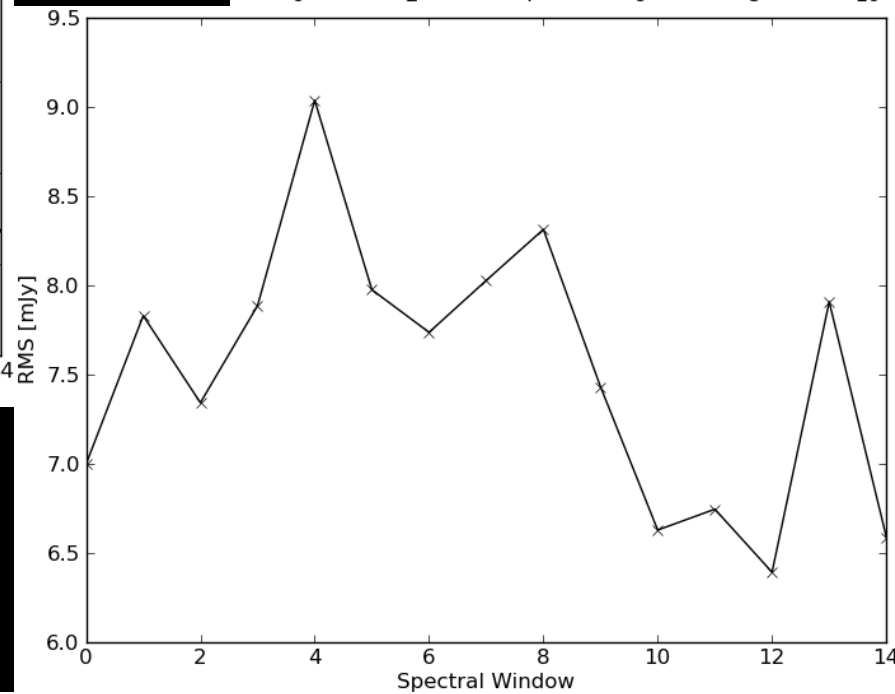
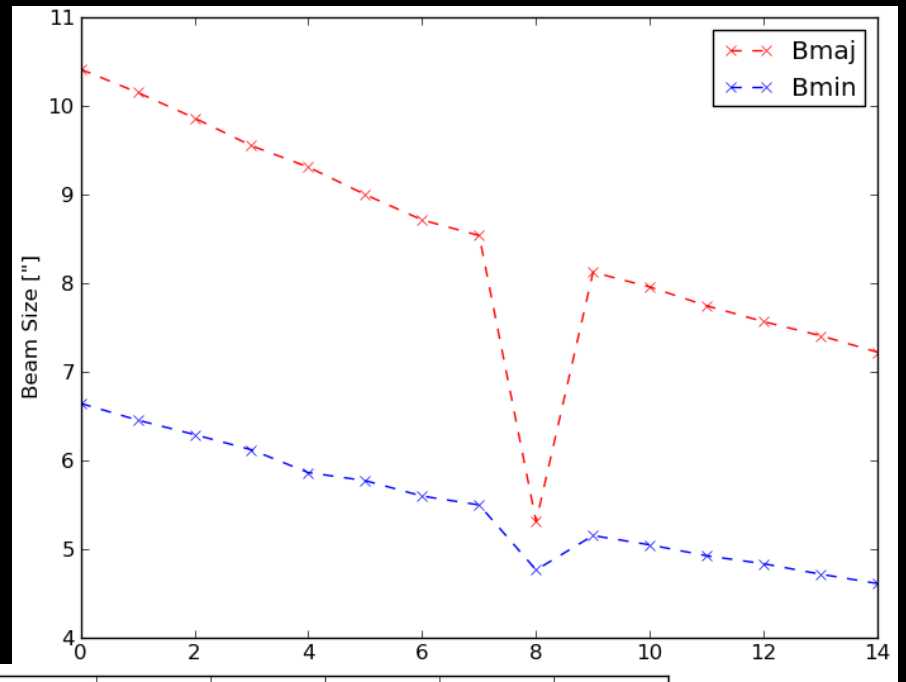# Data Quality Assessment

Data Quality Assessment

Data Quality Assessment

phasecalibrator_spw6.image—raster
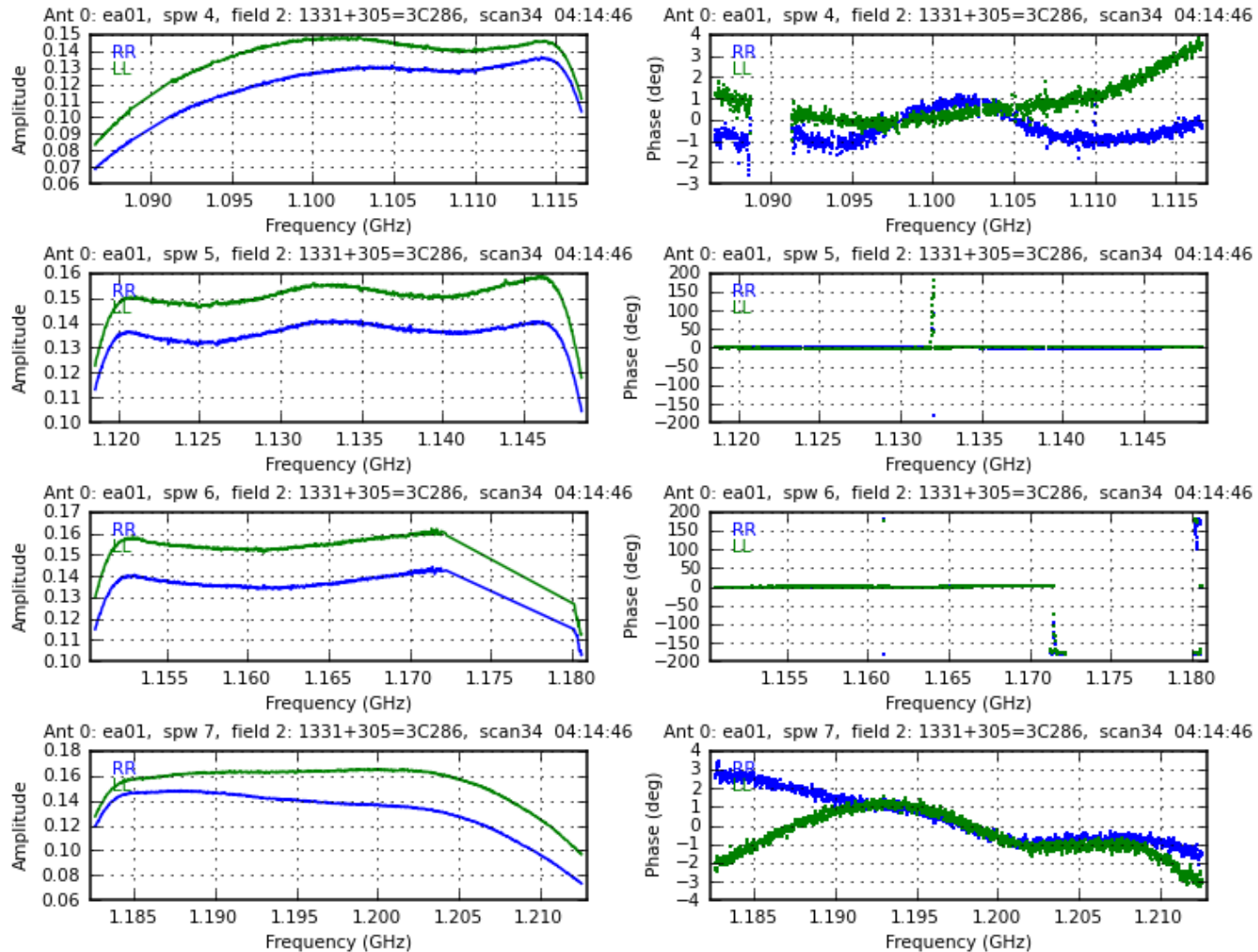
Data Quality Assessment
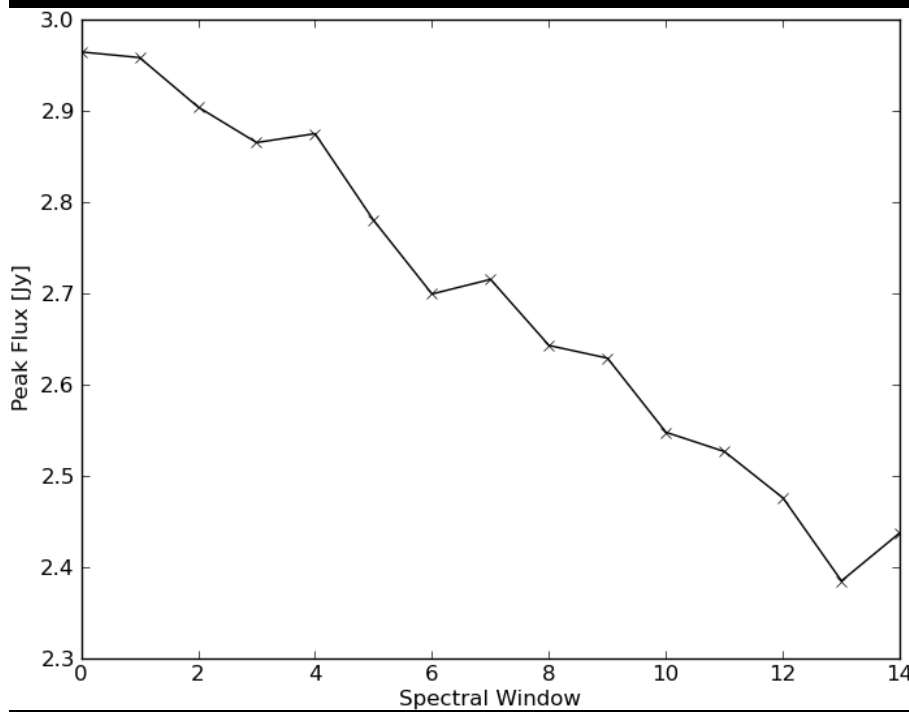
Outstanding Pipeline Issues

- Pipeline structure much simpler and faster!
- Modular structure makes it easier to discriminate between poor calibration and excessive RFI and fix manually.
- Diagnostic plots very helpful for identifying issues.
- *Excessive flagging occurring on all sources (~60-70% vs. ~30%). Still trying to figure out why. Change in behavior of CASA?*
- Test version of CASA now allows time-averaging for RFLAG. Not yet ready for pipeline, but should help with flagging.

After the reduction pipeline…

- After calibration, we ship data to Perth.. (both calibrated data and raw data with tables)
- Imaging is done with computers/Amazon cloud in Perth.
- Very useful new algorithm being developed by Kumar Golap:
  - Store all UV data in a gridded UV plane, then do FT. This reduces ultimate data base size by factor 100.
  - Also does good job flagging RFI
  - Brilliant idea, task almost ready
- Can image data in traditional manner or after doing uvgridding.

Lessons

- CASA is slowly developing tools needed to flag, calibrate, and image large datasets, but it is not there yet.
- Need pipelines optimized for data being taken. Beware general usage pipelines: one size does not fit all!
- Data volumes matter! Need to test pipelines with full datasets!
- It is clear that even with manual intervention, faint RFI is still appearing in our final data cubes. RFLAG with time-averaging should help, but this problem will get worse with time.
- Imaging has its own challenges. Talk to A. Popping or M. Meyer for details.
- Stick around for talks by K. Hess & X. Fernandez to hear about the great science we are doing with CHILES.