

The ALMA Pipeline



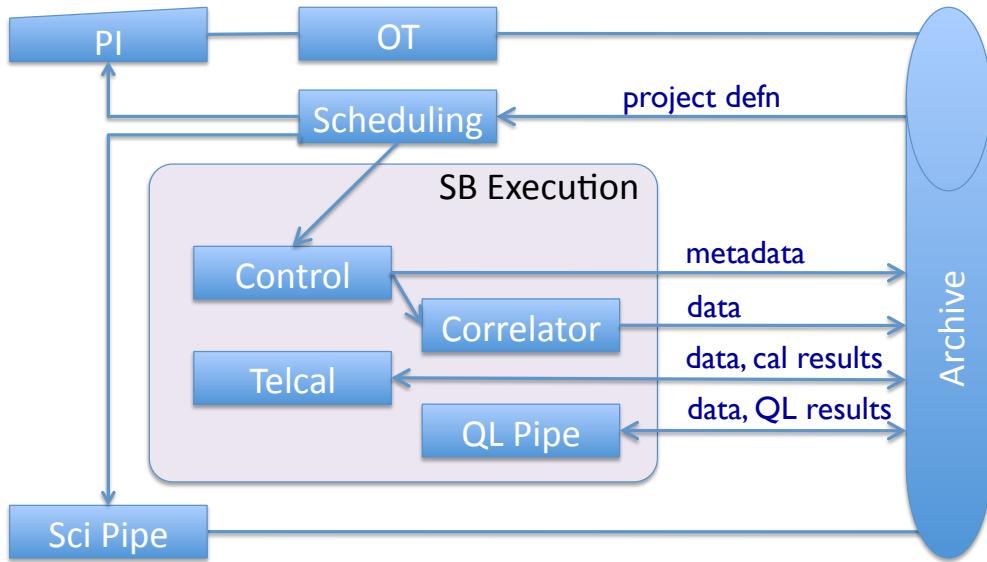
Caveat Emptor

This represents Remy's interpretations of a lot of statements like

“The factorization that will result from the analysis phase needs to be translated into internal APHS Use Cases and corresponding Sequence Diagrams. We will use a UML Tool such as Rational Rose.”



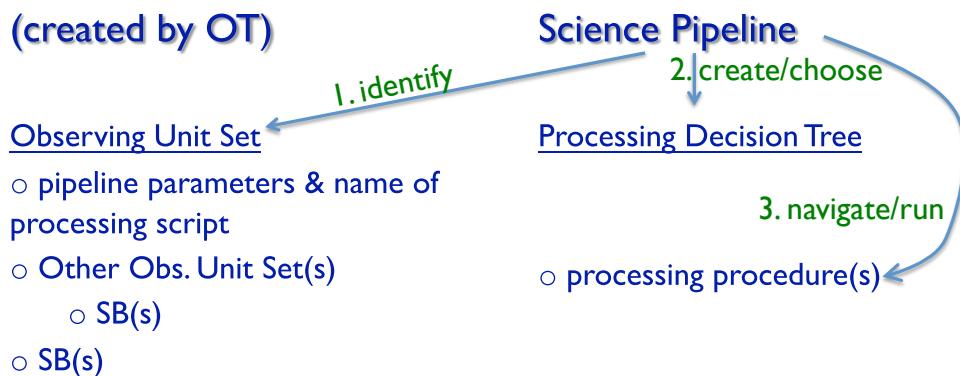
System Context : ACS



RI Oct 2010 NAASC Pipeline Intro

3

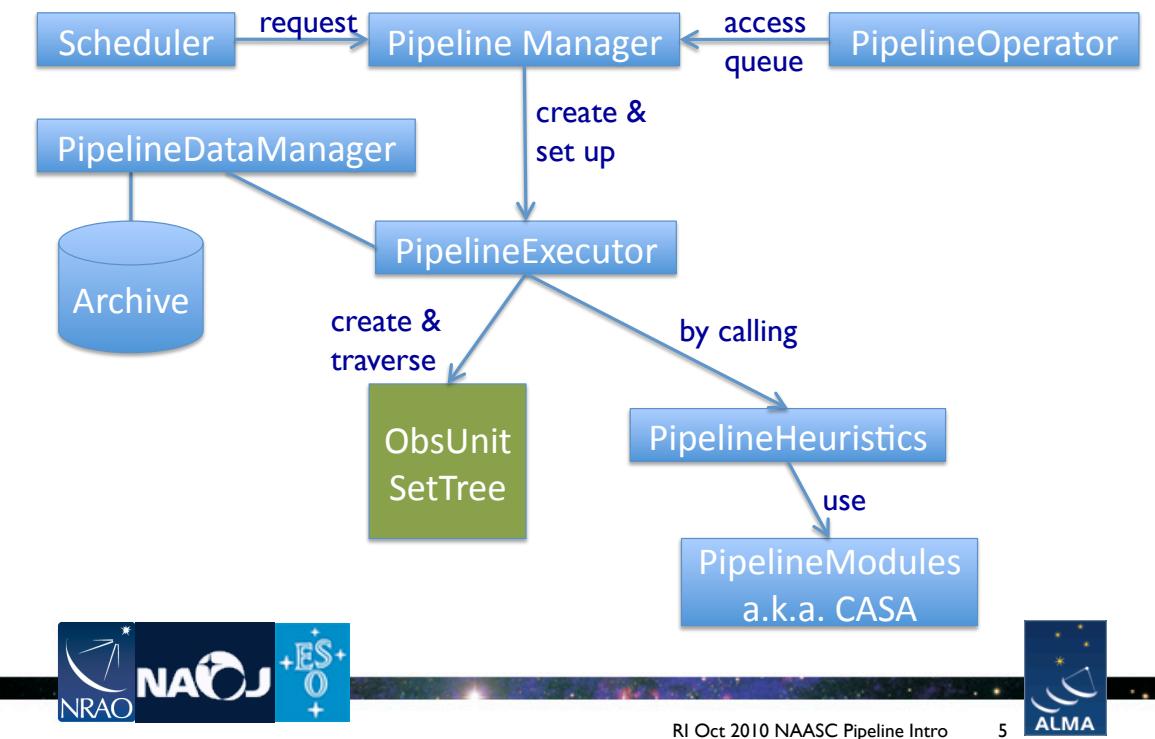
Process and relation to SBs



RI Oct 2010 NAASC Pipeline Intro

4

Architecture & Work Flow



PipelineHeuristics

Requirements

- o define data processing decision tree
- o deliver tree(s) as scripts
- o compute QA metrics
- o navigate the tree based on
 - o user intents (e.g. OT)
 - o input data, &
 - o previous processing results

Implementation

- one recipe per standard mode
 - SingleFieldInterferometry, SingleDish
 - ?
- 5-6 params, decide after ES (CDR7)
 - Intent, stored in APDM (maybe)
 - ? e.g. SD+mosaic ?

Recipes & Use Cases

<http://www.aoc.nrao.edu/~dshepher/alma/usecases/>

Pipeline.UC.ProcSciData.html	Process Science Data (C. Wilson)
Pipeline.UC.SnglFld.html	Science Pipeline: Process Single Field Interferometric Data (without single dish data) (J. Pety, F. Gueth, D. Shepherd)
Pipeline.UC.Mosaic.html	Science Pipeline: Process Mosaic, no short spacings (C. Wilson)
Offline.UC.SnglFldReduce.html	Reduce & Image Single Field Data (with and without single dish data) (D. Shepherd, S. Myers)
Offline.UC.MosaicReduce.html	Reduce & Image Multi-Field Mosaic Data (with and without single dish data) (D. Shepherd, S. Myers)
Offline.UC.TotPower.html	Offline: Reduce & Image Auto-Correlation (Single Dish) Data (D. Shepherd)
Offline.UC.SolarSystem.html	Solar System - comets and asteroids (D. Shepherd, P. Palmer)
<u>Polarization</u>	- TBD
<u>Astrometry</u>	- TBD
<u>Sun</u>	- TBD
<u>Pulsar</u>	(phased array, gated observations) - TBD

Software Design Appendices COMP-75.45.00.00-002-L-DSN also has:

- Self-Cal
- Complete mapping (combine one field with single dish)
- Wide field mapping (mosaics, with and without ACA and TP)
- Deep Field
- Weak line detection
- Survey of emission lines
- Survey of absorption lines



Input Parameters

<http://almaweb.hq.eso.org/almaweb/bin/view/PIPELINE/HeuristicsInputParameters>

(in particular the PDF attached there *** 1/2005)

Science Goals	Required / Optional	ASDM	Specific Data Reduction Parameters	Required / Optional
Full source identification and coordinates	R	✓ASDM	Flagging criteria	O
Continuum project	yes/no	R	Phase calibrator fitting parameters	O
Polarization project	yes/no	R	Flux calibrator fitting parameters	O
Desired angular resolution	Angular dimensions	R	Amplitude calibrator fitting parameters	O
Desired spatial dynamic range	Angular dimensions min/max	O	Bandpass calibrator fitting parameters	O
Line identification or frequencies	Transition name or GHz	R	Self-cal parameters	O
Desired velocity resolution (line)	km/s	R	Imaging parameters	O
Desired brightness temperature sensitivity limit	K	R	Single-dish interferometry combinations	O
Desired S/N or rms	Number or Jy	R		
Desired intensity dynamic range	DB / linear / min/max (?)	O		
Source Properties	Required / Optional			
Largest source structure (line/cont)	Angular dimensions	O		
Source extent	Point source / several compact sources / extended source	O		
Source shape	Circular, elongated, etc.	O		
Source type	Outflow, planetary nebula, HII region, etc.	O		
Estimated source flux (line/cont)	Jy	O		
Estimated line widths	km/s	O		
Source velocity structure	Gradients, velocity fields	O		



Regression Server

<http://alma-heuristics.mpifr-bonn.mpg.de/>
(don't publicize outside the project)

ALMA Pipeline Heuristics Server at the MPIfR

User Test Data and Results Regression Results Releases Heuristics Wiki Impressum

Max-Planck-Institut für Radioastronomie

Regression Results: Single Field Interferometry Mosaic Interferometry Single Dish

SFI Regression Results

- 2010-10-06
- uid_A002_X136360_Xa3.ms
- 2010-09-27
- 2010-07-26
- 2010-07-23
- 2010-07-13-patched
- 2010-07-13
- 2010-07-08
- 2010-07-02
- 2010-06-21
- 2010-06-15
- 2010-05-20
- 2010-05-06
- 2010-05-01-CL10S
- 2010-05-01
- 2010-04-23
- 2010-04-18
- 2010-04-15
- 2010-04-07
- 2010-03-20
- 2010-03-15
- 2010-02-20
- 2010-02-14
- 2010-02-12
- 2010-02-07

Local flagging of uid_A002_X136360_Xa3.ms.

- 1 Dataset summary

Flag data already known to be bad or invalid. These stages flag data that are known beforehand to be bad or inappropriate for inclusion in the reduction.

- 2 Flag autocorrelations
25368 rows flagged

Compute initial calibration solutions

- 3 Initial Bandpass Calibration
- 4 Initial Gain Calibration

Combine and image all datasets.

- 5 calibrator cleaned

BANDPASS GAIN FLUX TARGET POINTING UNKNOWN

FIELD_ID

FIELD_ID	BANDPASS	GAIN	FLUX	TARGET	POINTING
0	0	0	~10	~10	0
1	~10	~10	0	0	0
2	0	0	0	0	~10
3	0	0	0	~10	0



SFI Offline Use

- ?GUI? : input intents ~ OT, control tree navigation ~ PipelineExecutor
- CLI: run casapy scripts for each tree node, make own decisions

<http://almawh.hq.eso.org/almawh/bin/view/PIPELINE/PipelineHeuristicsTaskInterface>

unix> cd /Users/ri3e/casa/pipeline/PIPELINE-R1/Heuristics/src/heuristics/SFIPipelineTaskInterface/
unix> buildmytasks

```
CASA> sys.path.insert(0, '/Users/ri3e/casa/pipeline/PIPELINE-R1/Heuristics/src')  
CASA> execfile("/Users/ri3e/casa/pipeline/PIPELINE-R1/Heuristics/src/heuristics/SFIPipelineTaskInterface/  
mytasks.py")
```

```
CASA> sfi_setupPipelineFramework  
CASA> sfipipeline_object. <TAB>
```

creates a SFIPipelineStagesInterface
list methods

```
CASA> inp sfi_configureHeuristicParameters  
CASA> inp sfi_configurePaths
```

set observatory, maxPixels
* you have to start casapy in the output dir

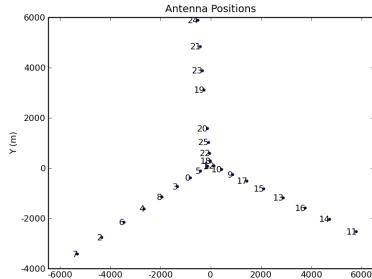
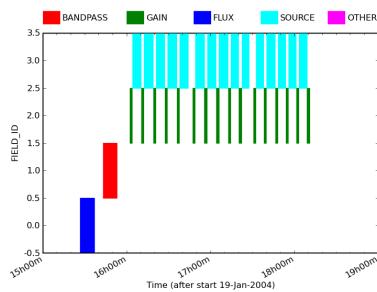
```
CASA> sfi_readData()
```



SFI

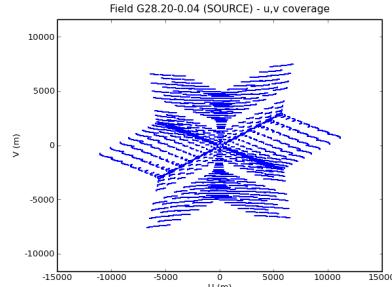
CASA> sfi_createSummary()

file:///opt/casa/pipeline/out1/html/AAAROOT.html



+ Field and spw summary tables

gets updated after each step,
but be careful about restarts



RI Oct 2010 NAASC Pipeline Intro

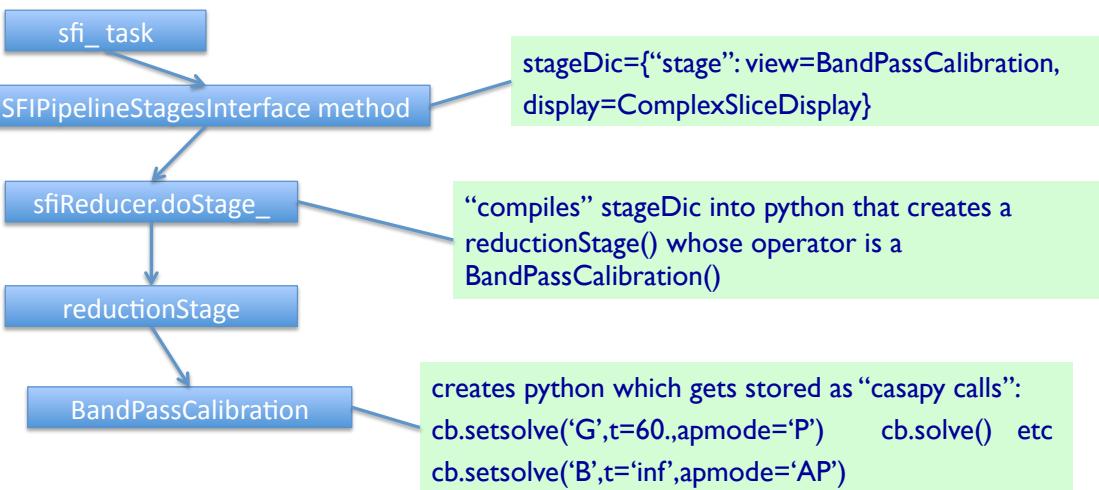
13

SFI

CASA> sfi_flagKnownBadData()

Autocorrelations and Gibbs channels

CASA> sfi_computeInitialCalCoefficients(dataType="BANDPASS")

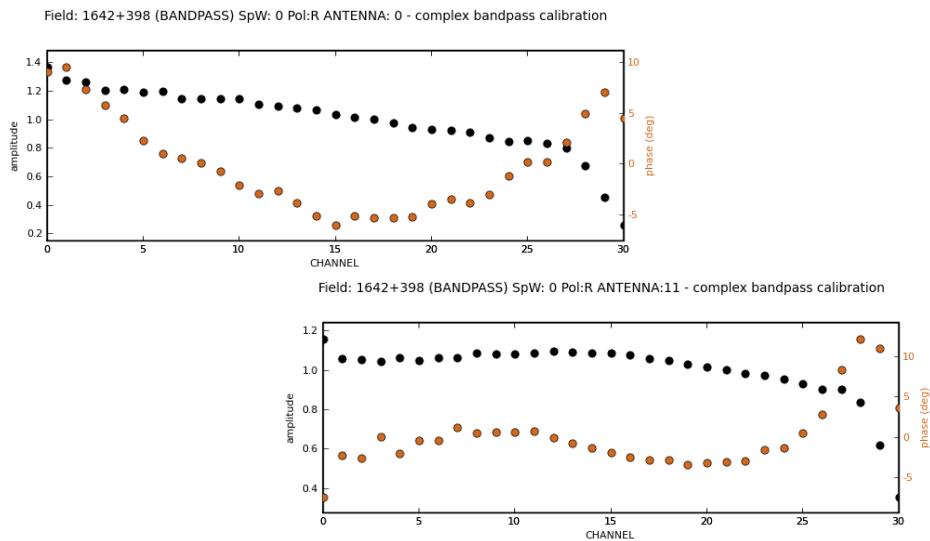


RI Oct 2010 NAASC Pipeline Intro

14

SFI

CASA> sfi_computeInitialCalCoefficients(dataType="BANDPASS")



RI Oct 2010 NAASC Pipeline Intro

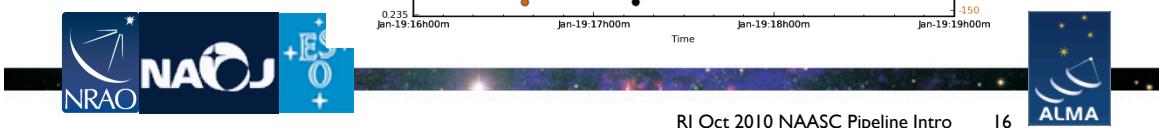
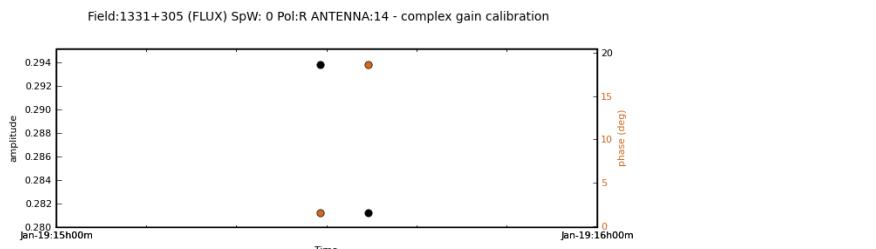
15

SFI

CASA> sfi_computeInitialCalCoefficients(dataType="FLUX")

```
stageDic={"stage": view=GroupSplineFluxCalibration,
gainCalMethod={'caltype':'G', 't':300., 'Min SNR': 0.0}
bandpassCal=UnNormalizedBandPassCalibration}
```

```
cb.setapply('B')
cb.setsolve('G', t=300., apmode='P')
```



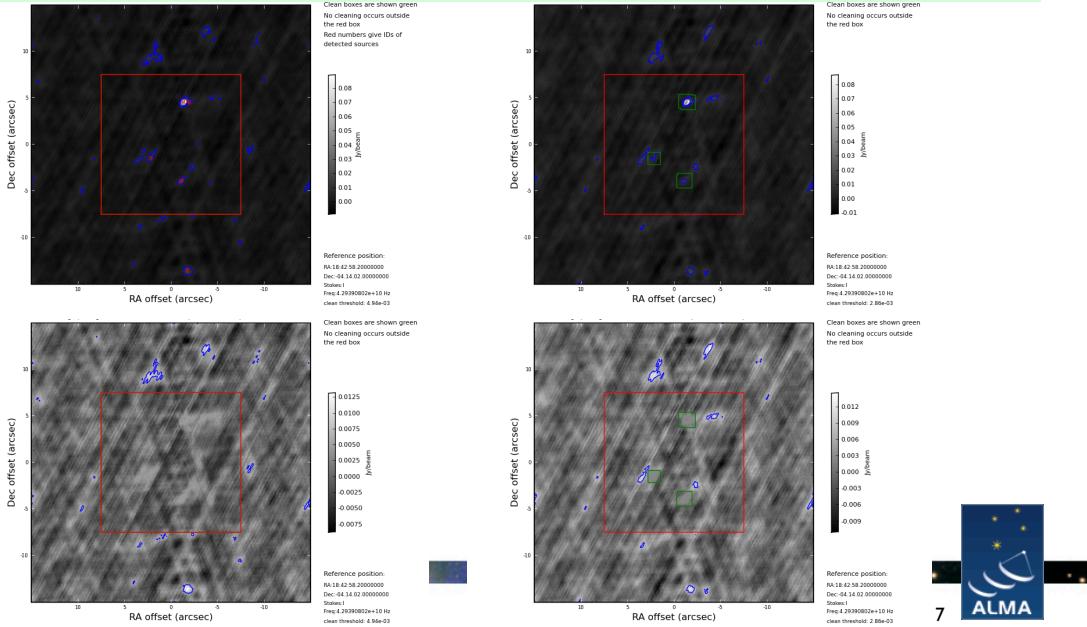
RI Oct 2010 NAASC Pipeline Intro

16

SFI

sfi_plotMSCleanedImages([gain,source,flux,bandpass])

- cell = 0.5/max_uvdist
- imsize = PB = 1.22 lambda/D [round up to composite #]
- clean inner 1/4 of image to 0.1Jy, then down to 2*rms(outside) OR convergence OR divergence
- search for sources w/total flux > 10*rms/(flux of brightest source)
- reclean with boxes



7

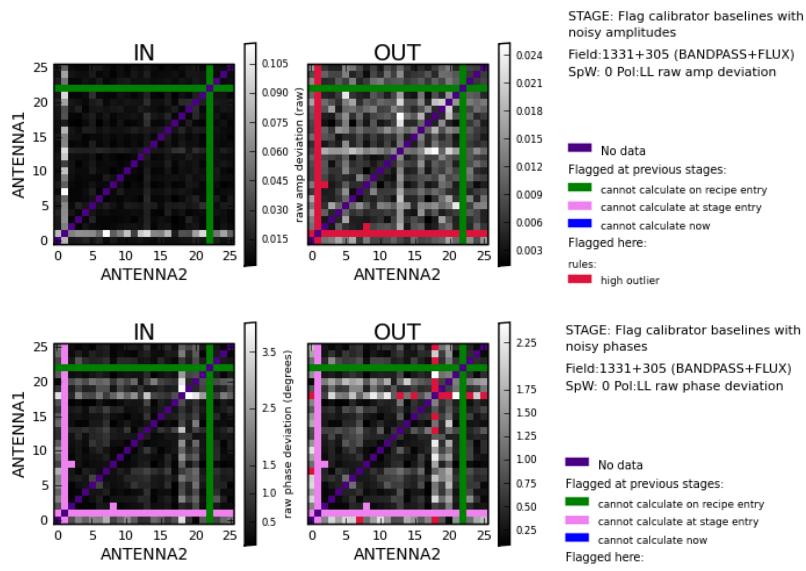
SFI

*This is a different dataset from UT6 *

CASA> sfi_flagCalibratorNoisyAmplitudes(), sfi_flagCalibratorNoisyPhases()

- flag baselines > 10*MAD

MedianAbsDeviation= median(|x_i – median(x_j)|)

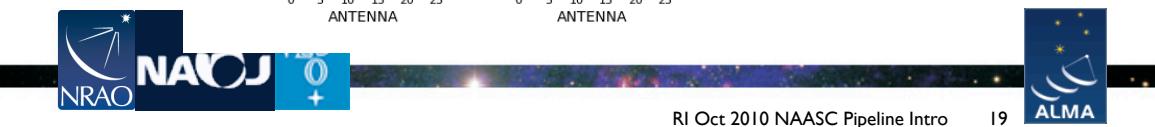
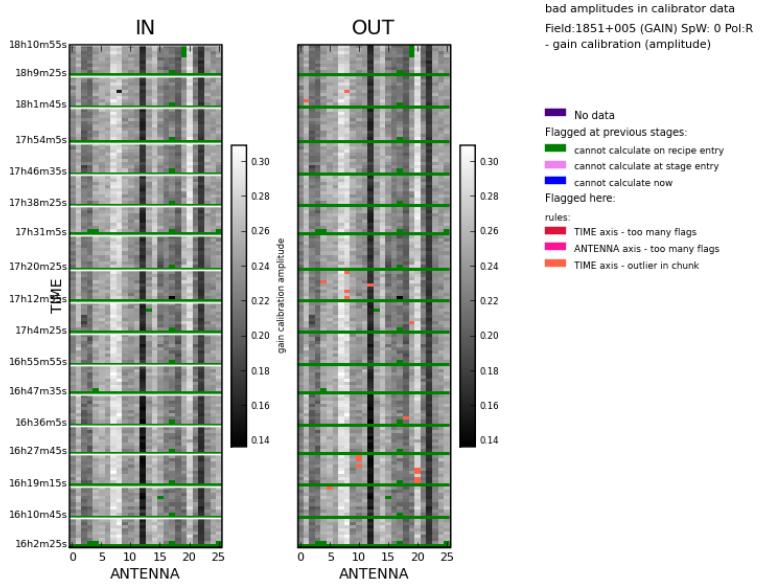


SFI

- (recalculate gain soln)

CASA> sfi_flagCalibratorAntennaSingleAmplitudes(), sfi_flagCalibratorBaselineAmplitudes()

- flag ants, then baselines >< 10* MAD of time sequence in amp



RI Oct 2010 NAASC Pipeline Intro

19

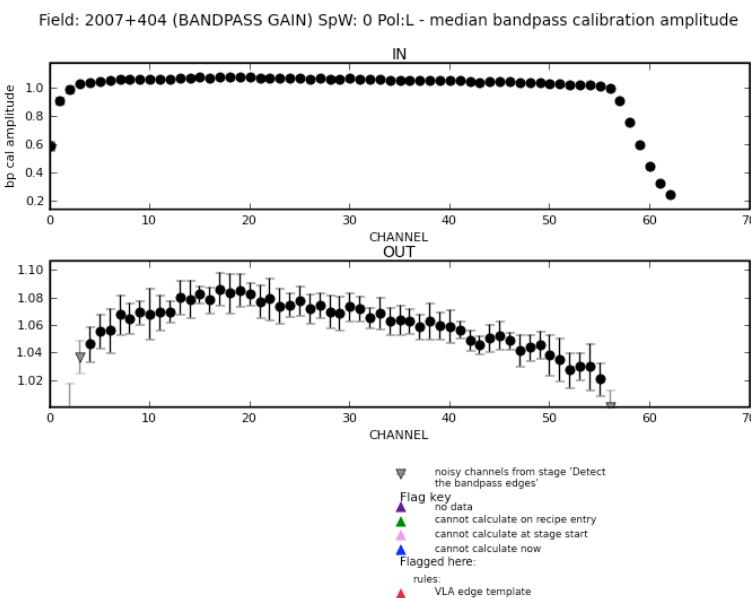
**SFI**

- (recalculate BP)

*This is a different dataset from UT6 *

CASA> sfi_detectBandpassEdge()

- best chisquared fit to a template of what the BP amp should be



RI Oct 2010 NAASC Pipeline Intro

20



SFI

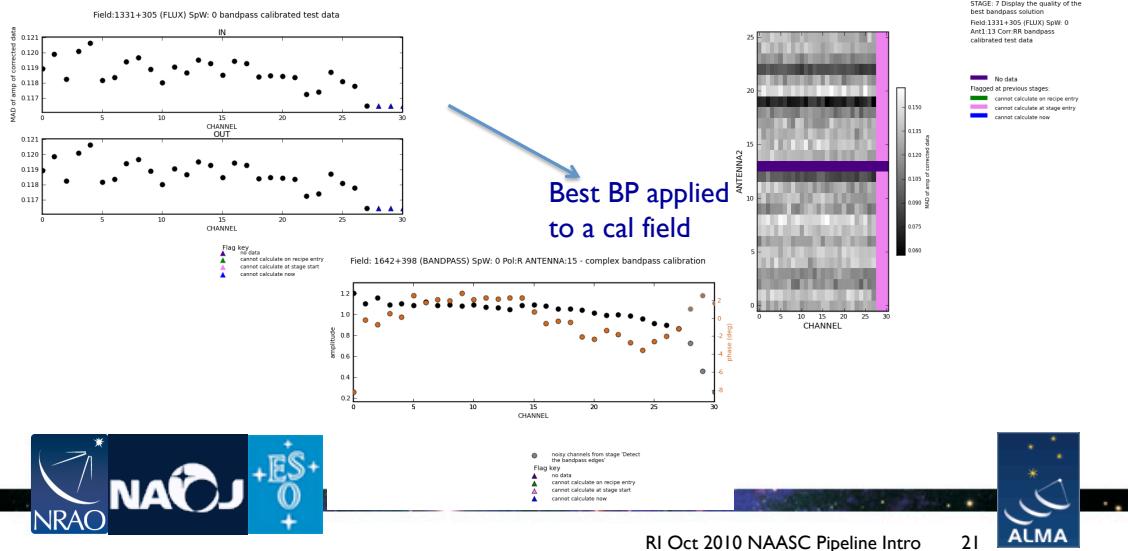
*This is a different dataset from UT6 *

sfi_findBestBandpassSolution(phaseUp_tList=[20.0,60.0,160.0,320.0,3600.0])

for different methods (here, different phase-only solint times), calculate and apply BP solution to a calibrator, choose based on spectral flatness after application, metric per baseline = sum(|data-median(data)|)/sum(sigma)

sfi_flagNoisyChannels()

recursively flag chans whose median is >10*MAD from the median across all channels



RI Oct 2010 NAASC Pipeline Intro

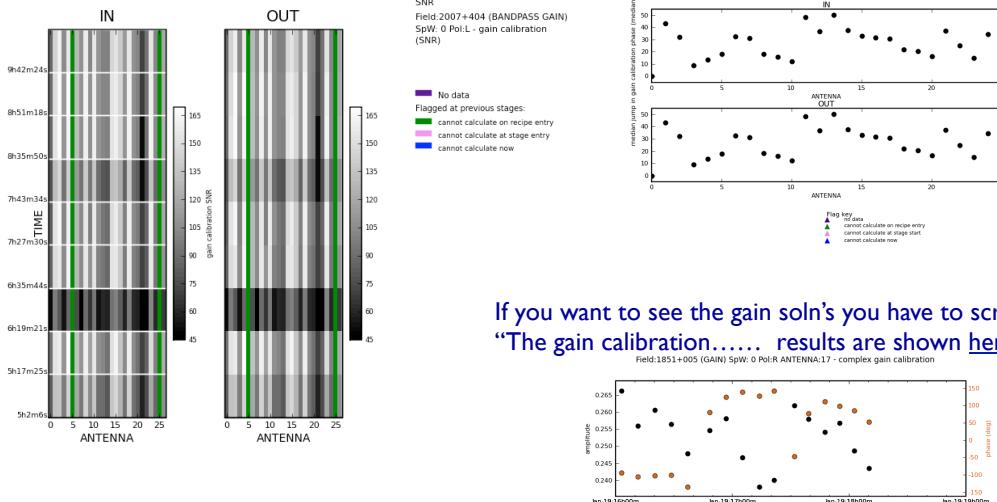
21

SFI
sfi_flagLowSNRCalibration()

- (recalculate Gain)
- based on SNR per ant reported by Calibrator
- (recalculate Gain)

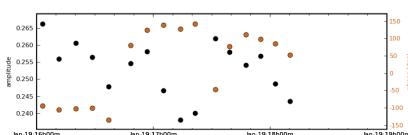
sfi_flagPhaseJumpCalibration()

- median absolute jump in phase between G calibrations



If you want to see the gain soln's you have to scroll down to
"The gain calibration..... results are shown [here](#)"

Field:1851+005 (GAIN) SpW: 0 Pol:R ANTENNA:17 - complex gain calibration



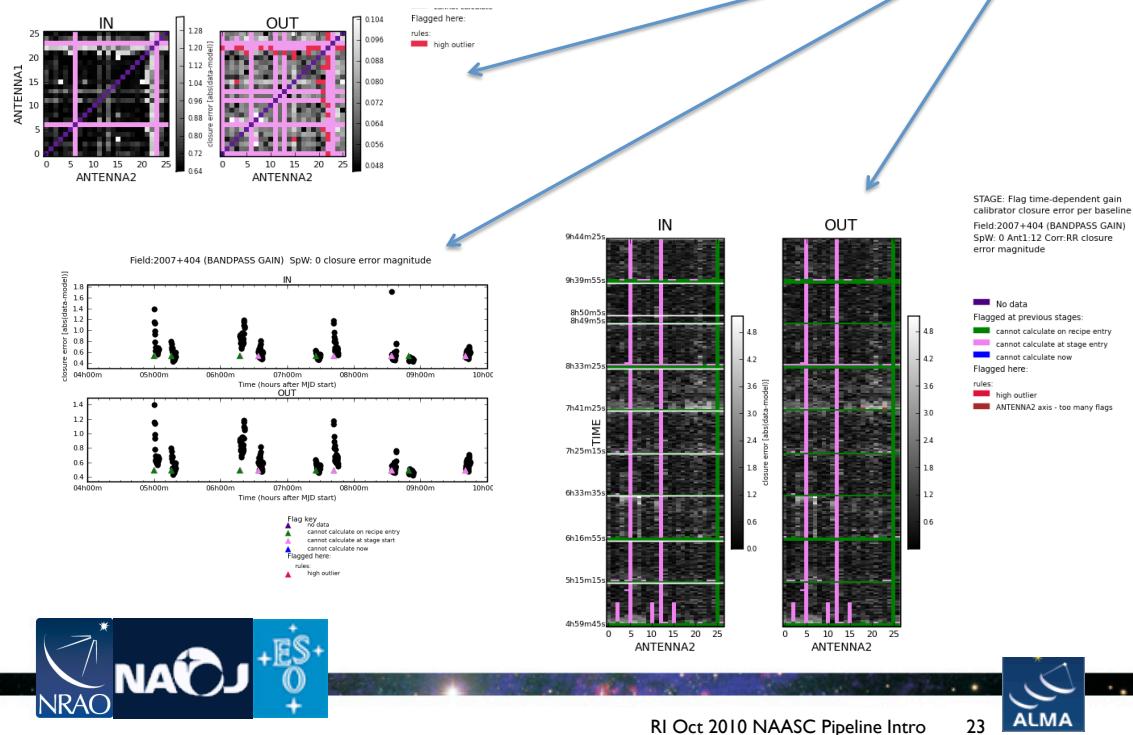
RI Oct 2010 NAASC Pipeline Intro

22



SFI ○ (recalculate Fluxcal, reimagine calibrators) *These are from different UT6 datasets*

sfi_flagLargeClosureErrors (in calibrators) PerAntenna, PerBaseline, inTime, inTime/Base



RI Oct 2010 NAASC Pipeline Intro

23

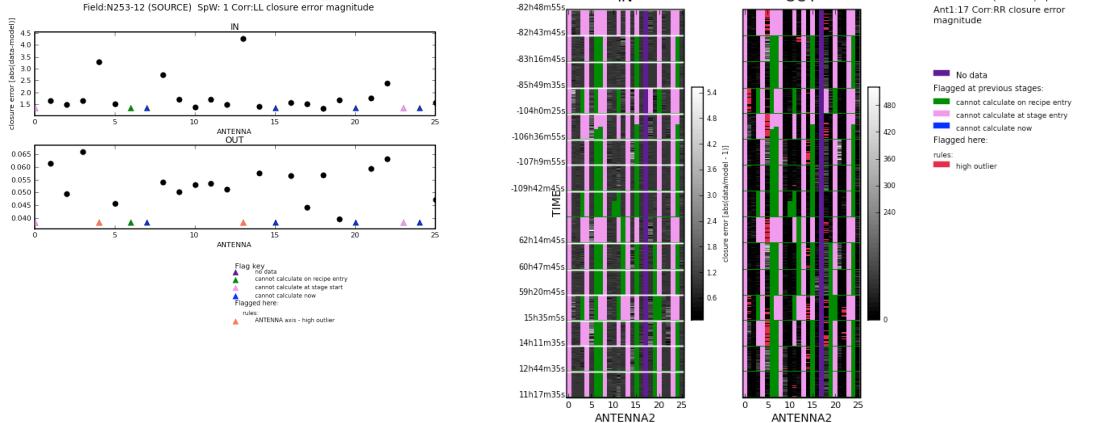
SFI

These are from different UT6 datasets

sfi_calculateFinalCalibrationSolutions

BP and Gain

sfi_flagLargeClosureErrors(*SOURCE*) ○ (re) image and clean the source per antenna, per baseline, in time, and per base in time



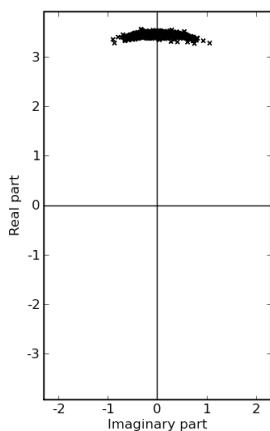
RI Oct 2010 NAASC Pipeline Intro

24

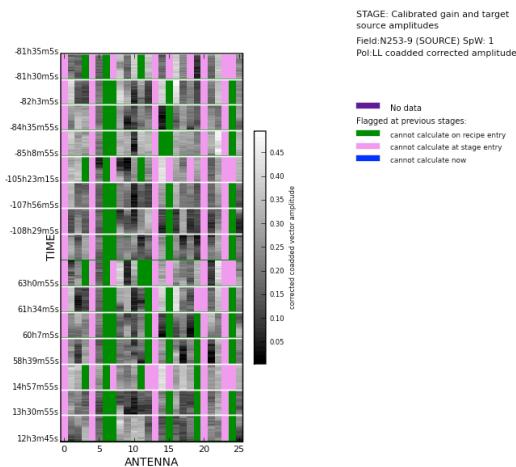
SFI

* These are from different UT6 datasets *

sfi_plotVisibilities(),
calibrated gaincal



sfi_plotVisibilityAmplitudes(): calibrated source



And then clean once again...



SFI Summary

- Pre-flagging BP & gain
- Image: pilot clean, find sources, auto-box, clean with boxes
- first Flagging:
 - high deviant baselines over entire chunk, amp & phase
 - high or low deviant times, amplitude antenna-based, then BL-based
- first BP soln & flagging:
 - find edge chans by fitting a template
 - find best p-only solint for phase-up of BP calibrator
 - flag chans that remain deviant when BP applied to a test source
- first Gain soln & flagging:
 - Flag low SNR from Calibrator
 - Flag antennas with large phase jumps between cal observations
 - Flag closure errors; /ant, /base, /time, and /time/base
- final BP and Gain soln
- flag closure errors in SOURCE, /ant, /base, /time, and /time/base
- image and clean again



QuickLook v. Science Pipelines

Share some features, but not identical:

Science

- Pre-flagging BP & gain
- Image: pilot clean, find sources, box & clean
- first Flagging
- can do manually
- first BP soln & flag deviant chans, ants
- first Gain soln & flag phase jumps, closure
- final BP and Gain soln
- flag closure errors in Source
- image and clean again

QuickLook

- Flag extremely bad data
- decide whether to use WVR?
- apply BP from Telcal
- apply Gain from Telcal
- Dirty image

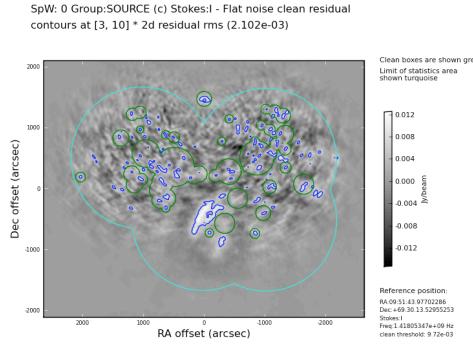
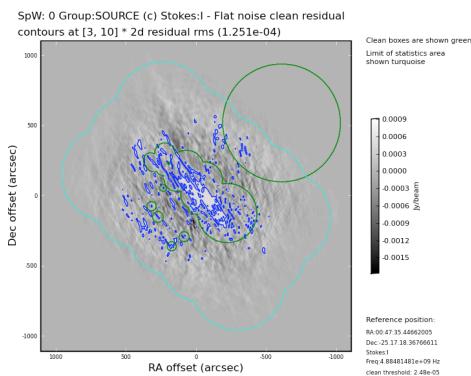
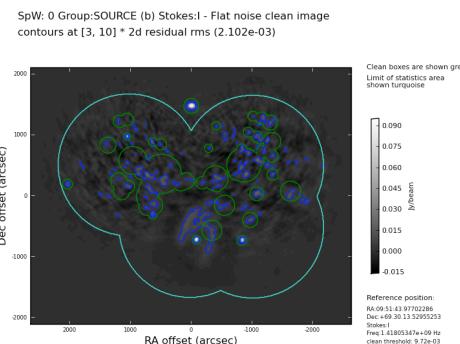
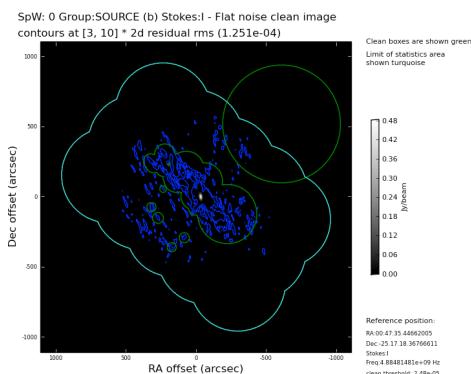
It is not currently envisioned that one can rerun the QL Pipeline afterwards Offline



RI Oct 2010 NAASC Pipeline Intro 27

UT6

I didn't ask permission, so you didn't see this here 😊



RI Oct 2010 NAASC Pipeline Intro

28

Conclusion End (for now)

SFI, SD heuristics scripts available to run and evaluate
Mosaic test results available online (from UT6)
Other Use Cases will determine future Heuristic development

The team will be here Nov 15-18,

<http://almasw.hq.eso.org/almasw/bin/view/PIPELINE/FaceToFaceMeeting2010>

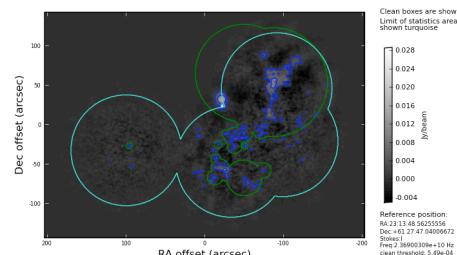
In particular, we need to be prepared to give useful feedback on Thurs, Nov 18



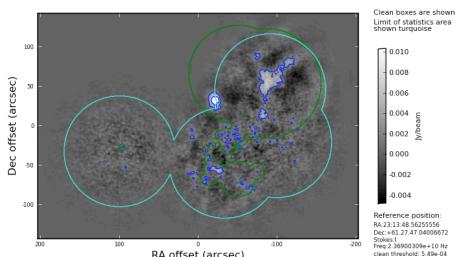
UT6

I didn't ask permission, so you didn't see this here ☺

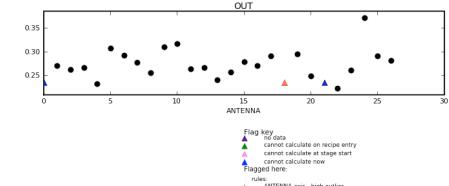
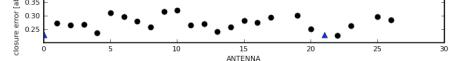
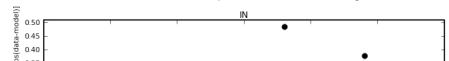
SpW: 0 Group:SOURCE (b) Stokes:I - Flat noise clean image
contours at [3, 10] * 2d residual rms (7.231e-04)



SpW: 0 Group:SOURCE (c) Stokes:I - Flat noise clean residual
contours at [3, 10] * 2d residual rms (7.231e-04)



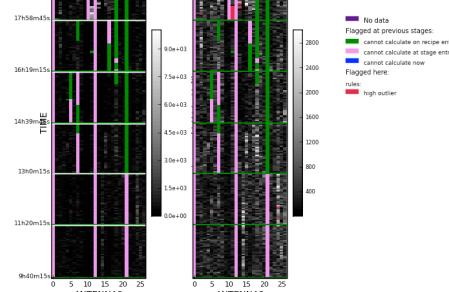
Field:NGC7538D (SOURCE) SpW: 1 Corr:LL closure error magnitude



STAGE: Flag time-dependent target closure error per baseline

Field:NGC7538C (SOURCE) SpW: 1

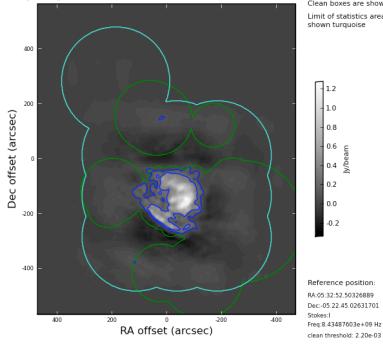
Ant: 12 Corr:LL closure error magnitude



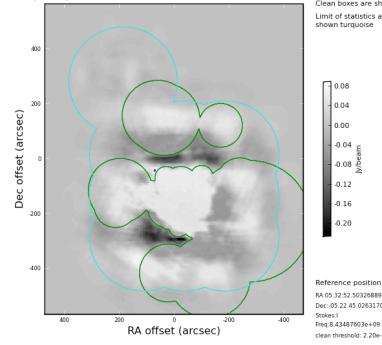
UT6

I didn't ask permission, so you didn't see this here 😊

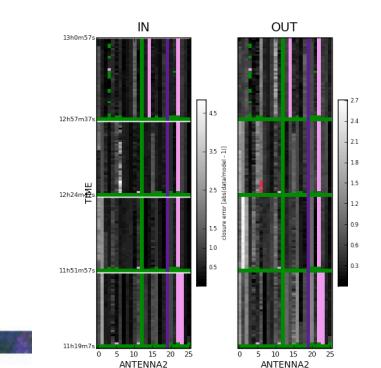
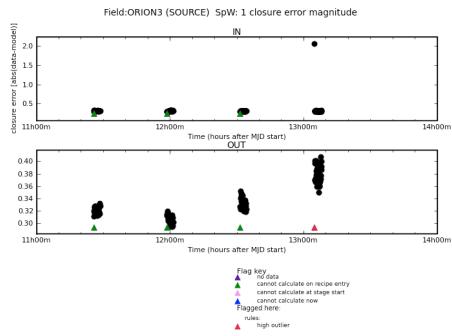
SpW: 0 Group:SOURCE (b) Stokes:I - Flat noise clean image
contours at [3, 10] * 2d residual rms (2.862e-02)



SpW: 0 Group:SOURCE (c) Stokes:I - Flat noise clean residual
contours at [3, 10] * 2d residual rms (2.862e-02)



Clean boxes are shown green
Limit of statistics area shown turquoise



13 OCT 2019 14:55:00 | Pipeline Image