Netherlands Institute for Radio Astronomy

LOFAR: Imaging Pipeline and MSSS

George Heald

CALIM2011, Manchester

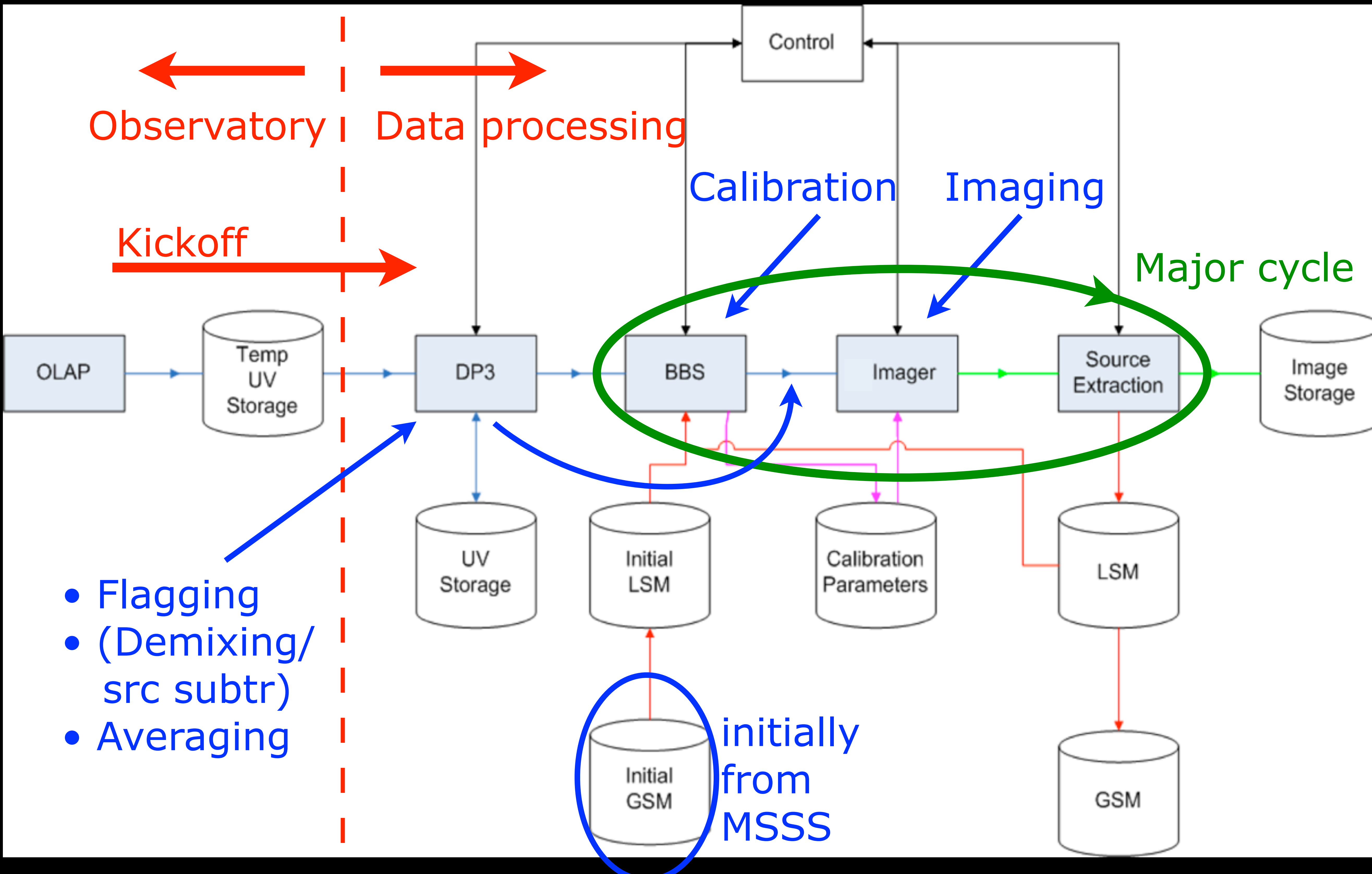
26 July 2011

Outline

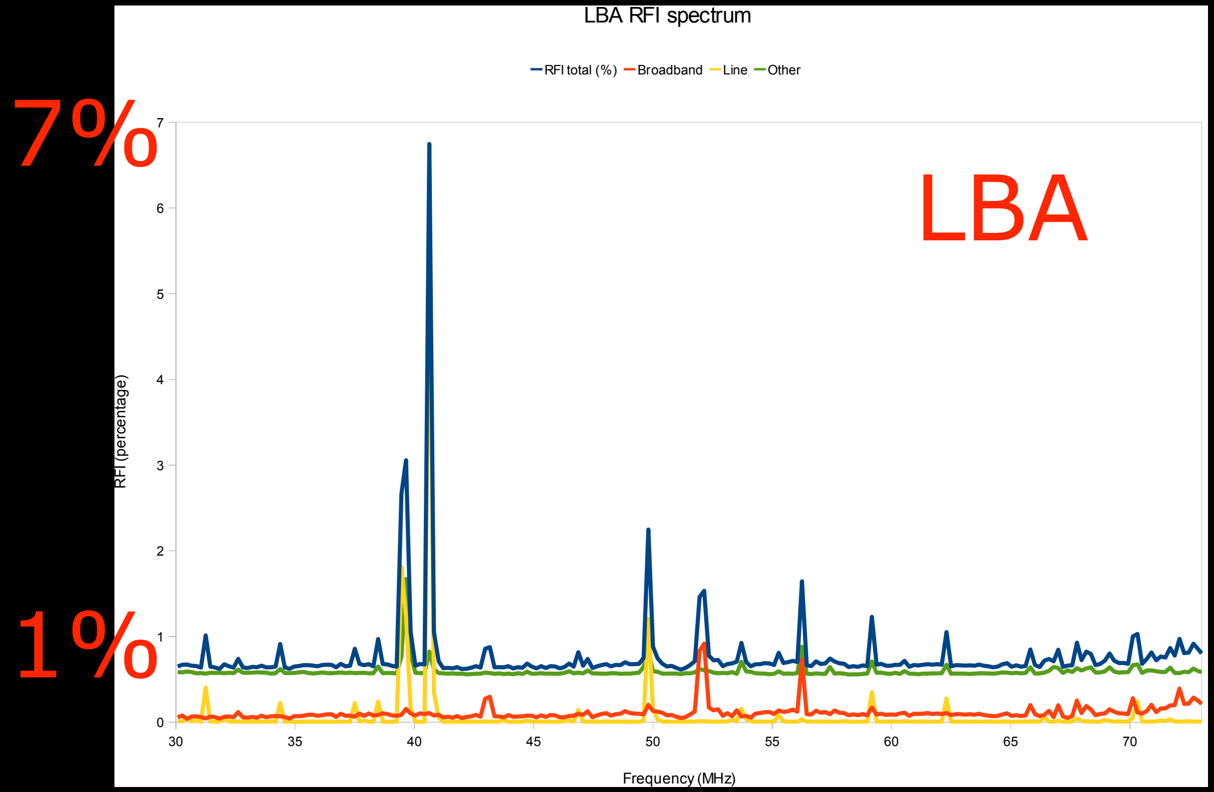
- Recap of the LOFAR Imaging Pipeline flow
  - Practical look at the current status of the components and pipeline, with illustrative examples
  - Plans for short-term improvements

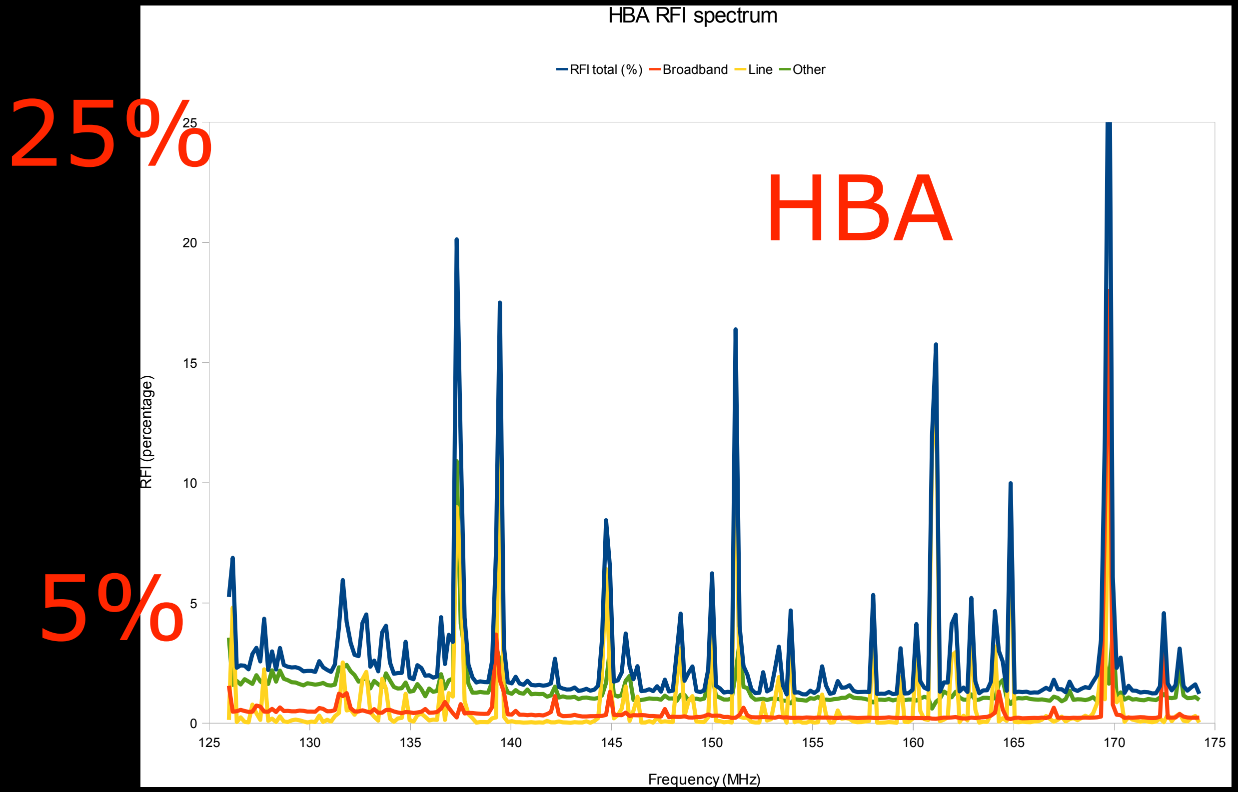
- MSSS - the LOFAR Commissioning (LOCo) Survey
  - Review of capabilities
  - Needs and constraints
  - Overview of plans for MSSS

LOFAR standard imaging pipeline



- Flagging with AOFlagger is now standard, and included in the NDPPP step
- Flag results and statistics seem satisfactory, e.g.:

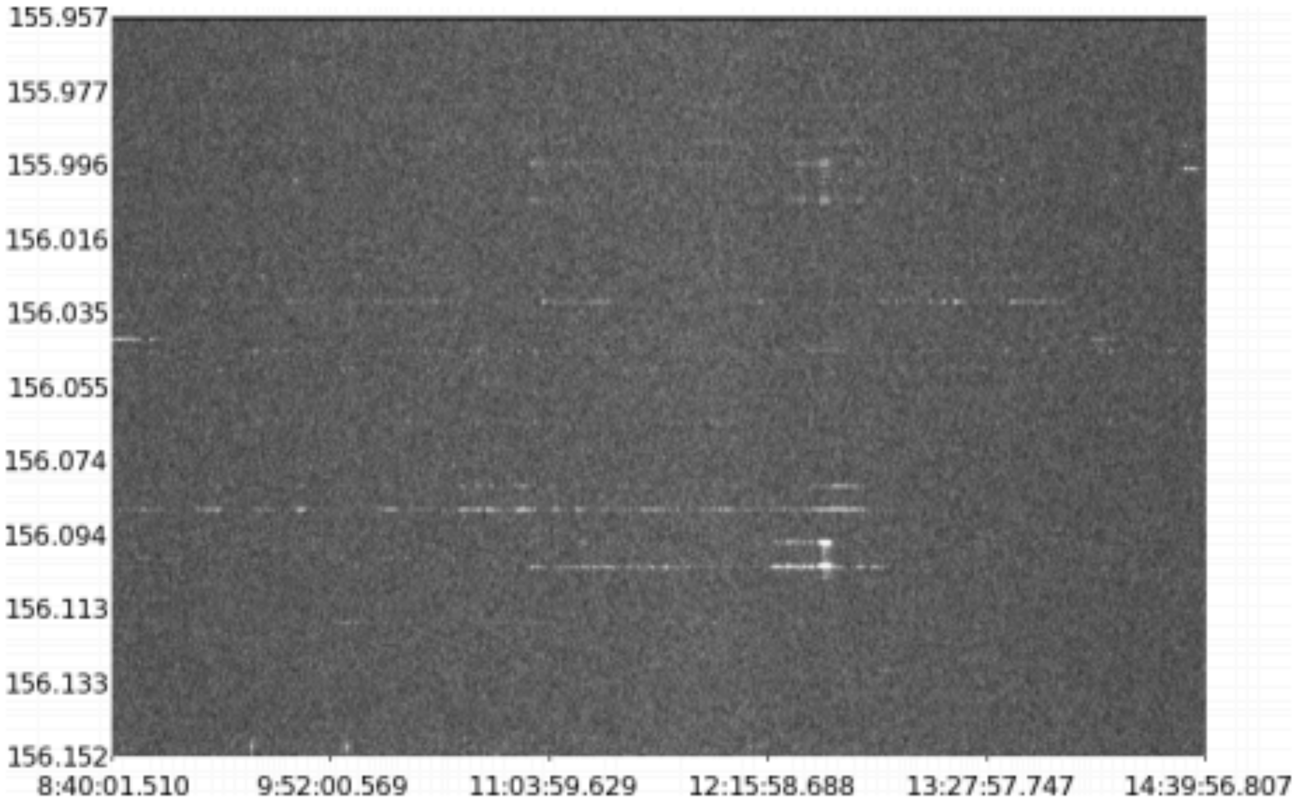




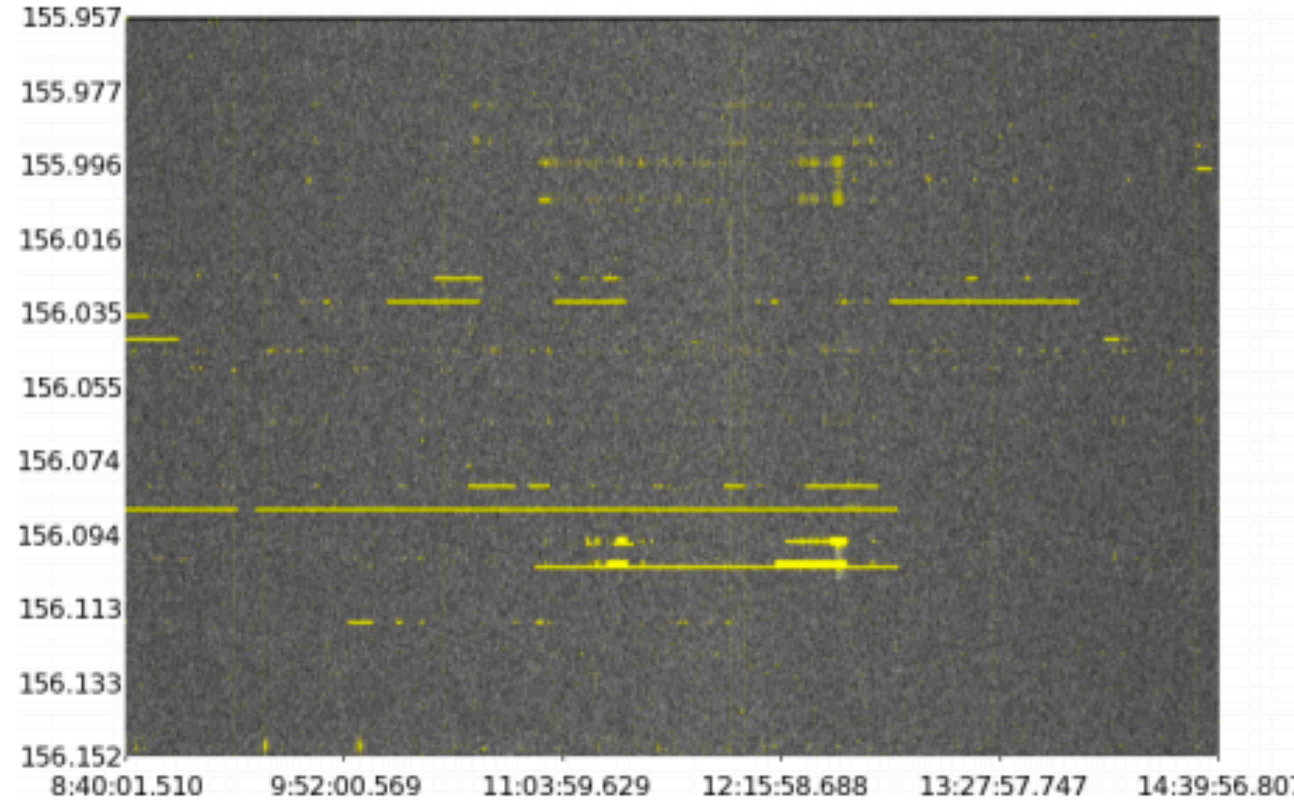
Offringa

- Processing speed with the new cluster for a 6hr observation:
  - 15 minutes LBA (flagging, no averaging)
  - 22 minutes HBA (flagging and averaging)

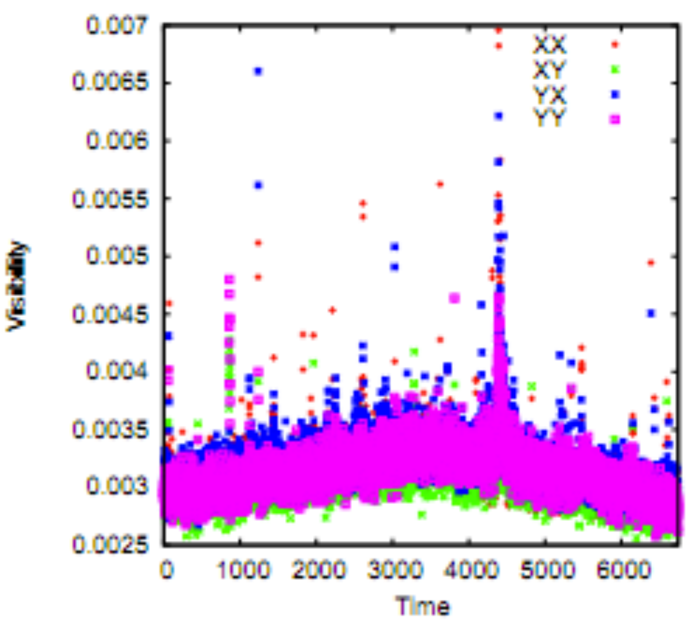
NDPPP (example 6hr LOFAR HBA)

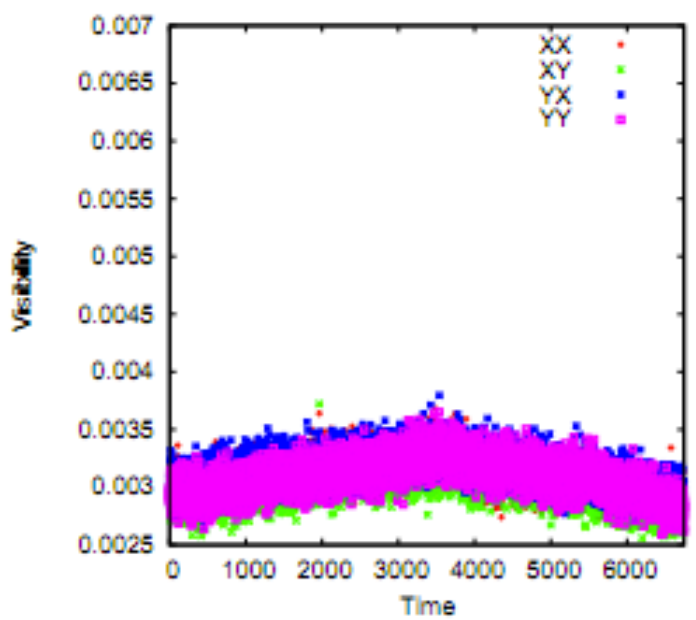


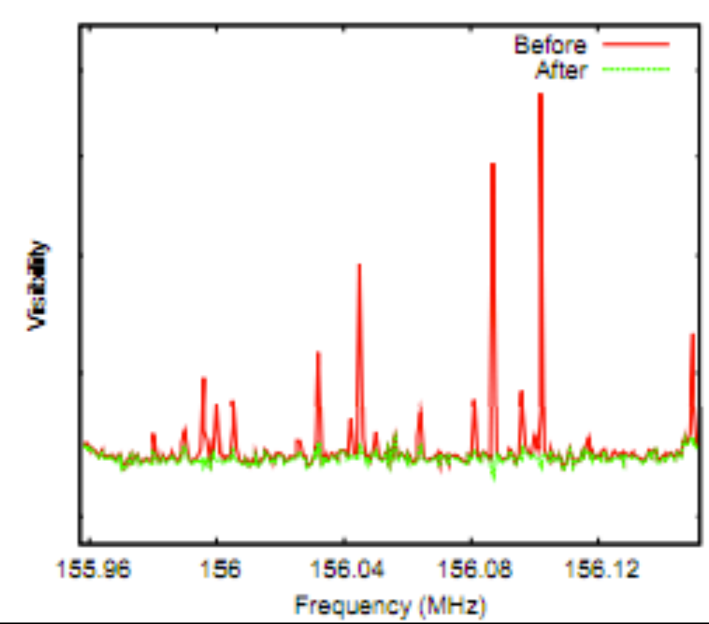
(a) Time-frequency plot before flagging



(b) Time-frequency plot after flagging







Offringa et al.

Demixing

- Described by van der Tol et al. (2007) IEEE TSP, 55, 4497
- Used as an alternative to direction-dependent gain solutions
- Measured visibility (where $\mathbf{a}_n$ contains phase shift etc):

$$\hat{\mathbf{v}} = \mathbf{a}_1 v_1 + \mathbf{a}_2 v_2 + \mathbf{a}_3 v_3 = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{A}\mathbf{v}$$

$$\mathbf{v} = \mathbf{A}^\dagger \hat{\mathbf{v}} = (\mathbf{A}^{\mathrm{H}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{H}}\,\hat{\mathbf{v}}$$

$$\overline{v}_i = \frac{1}{N_t N_f} \sum_{j=1}^{N_t} \sum_{k=1}^{N_t} a^*_{ijk} \hat{v}_{jk}$$

- Introducing "mixing matrix" **M**:

$$\mathbf{M} = \left(\frac{1}{N_t N_f} \mathbf{A}^{\mathrm{H}}\mathbf{A} \right)^{-1}$$

- Estimated visibility function of source 1, "demixing" 2 and 3:

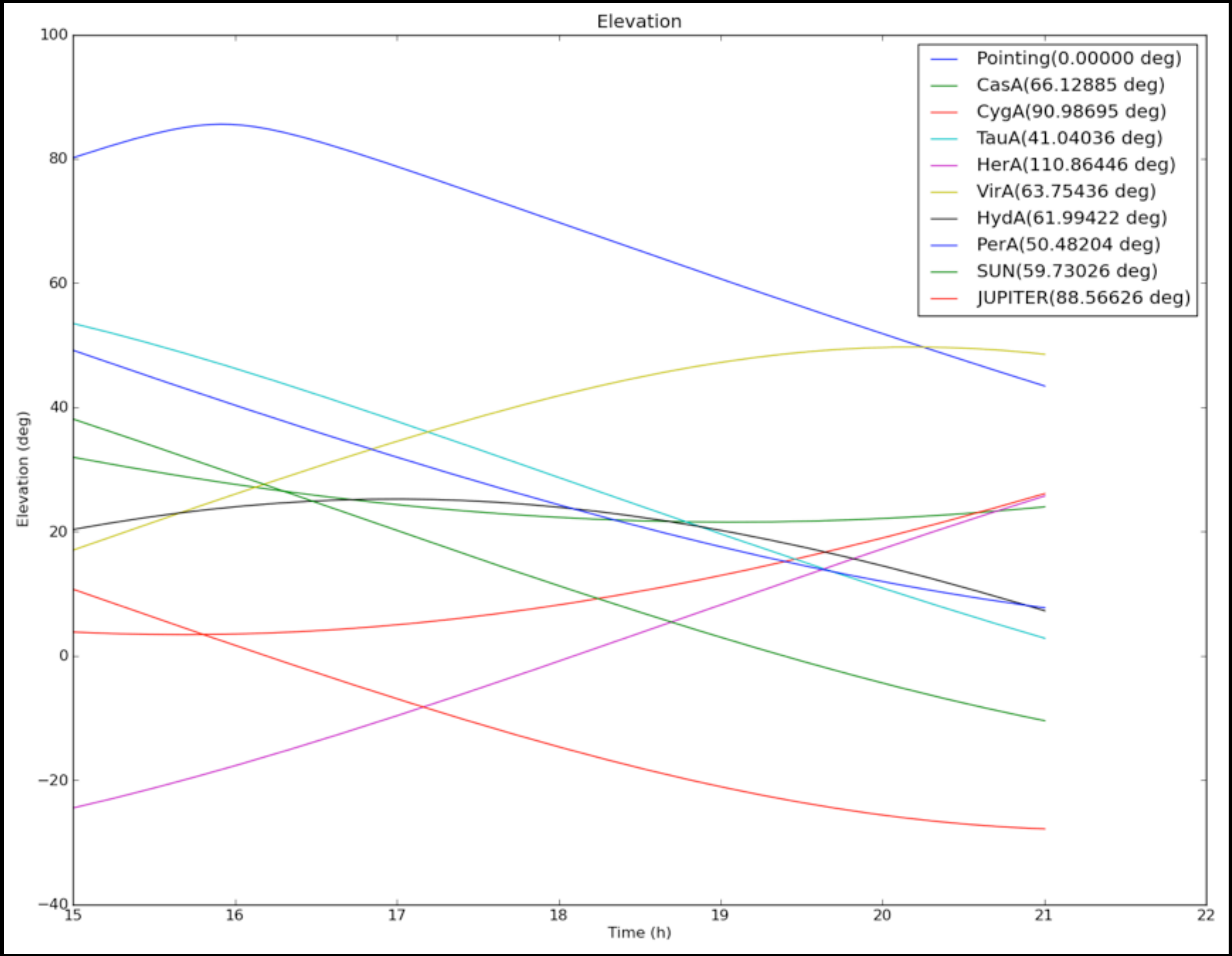
$$v_1 = \overline{v}_1 - M_{1,2}\tilde{v}_2 - M_{1,3}\tilde{v}_3$$

Demixing

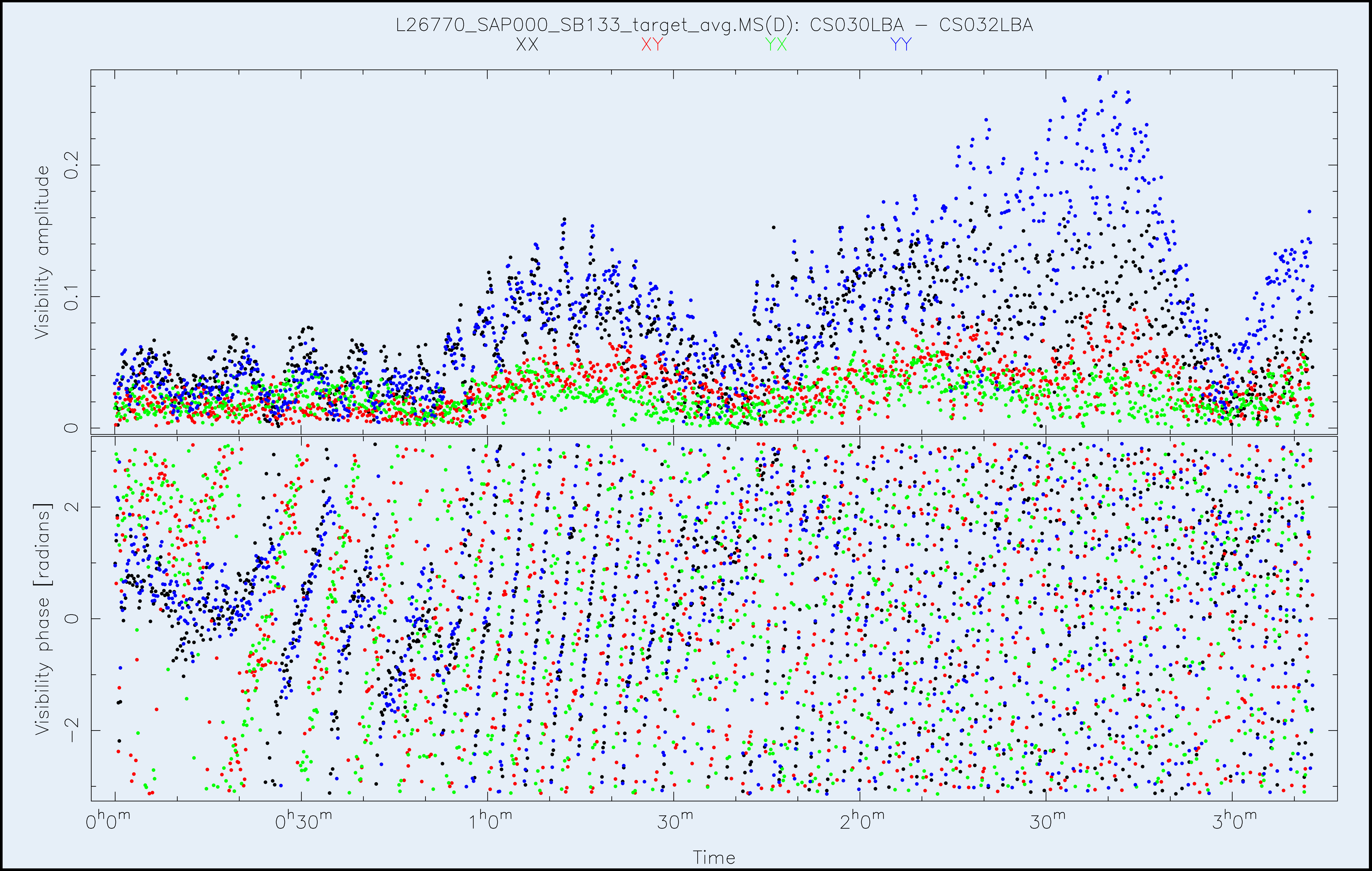
- Demixing algorithm developed by Bas van der Tol is working well
  - Now ~standard for LBA datasets
  - Not typically used (but may be important later?) for HBA
- Speed is a problem:
  - good models of the offaxis sources (though NOT of the target field) are required
  - a significant runtime factor is the time to produce model visibilities [FFT-based model prediction in BBS now a high priority]
- Results are good, when bright offaxis sources are >20 degrees from target field (at ~60 MHz) - more test beams available
- Now implemented as a demonstrator python script, but will soon be added as part of DPPP/BBS for speed

Demixing: 3C196 LBA_INNER

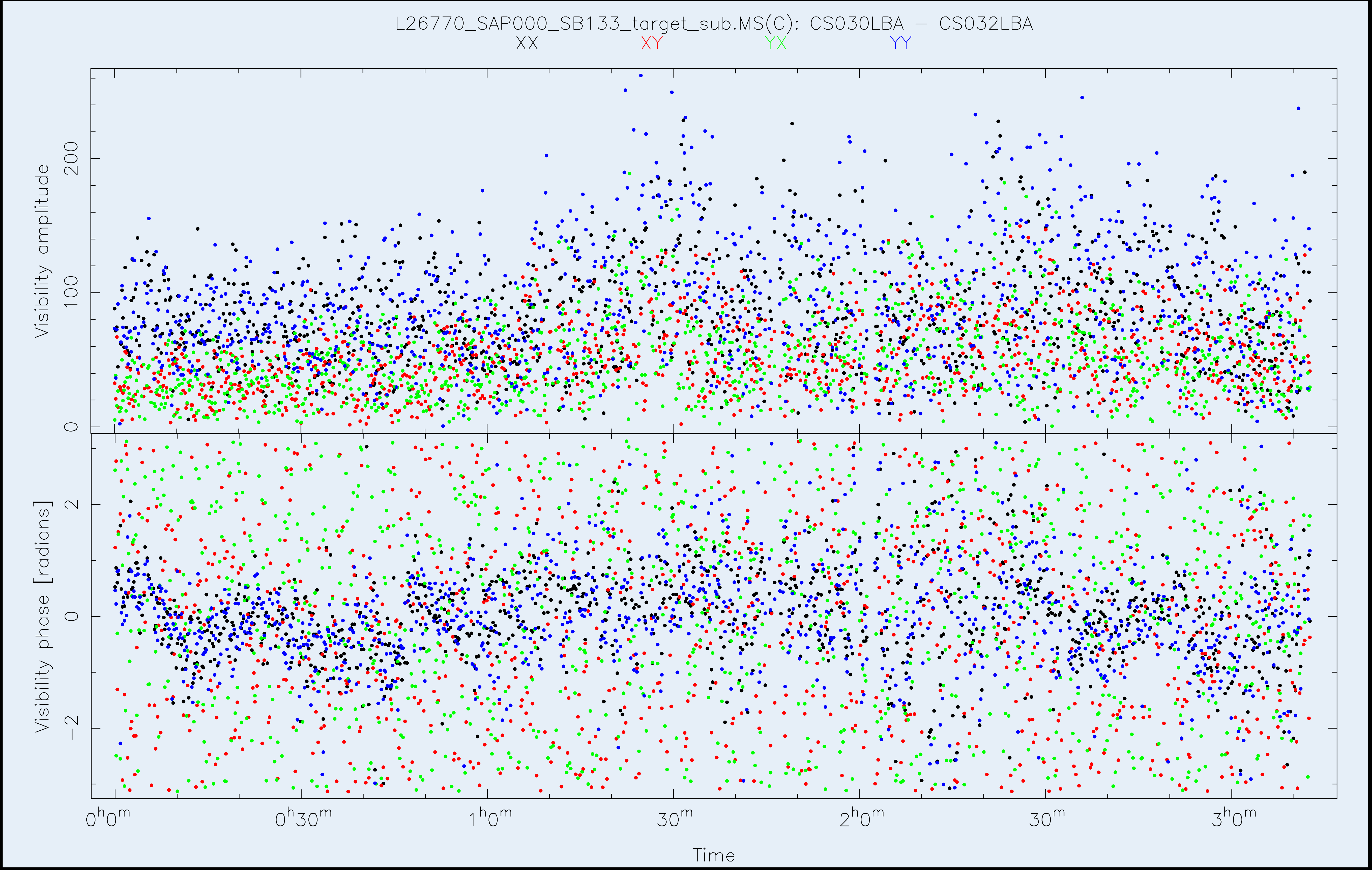
- Demixed CygA, CasA, VirA (latter not necessary in hindsight)



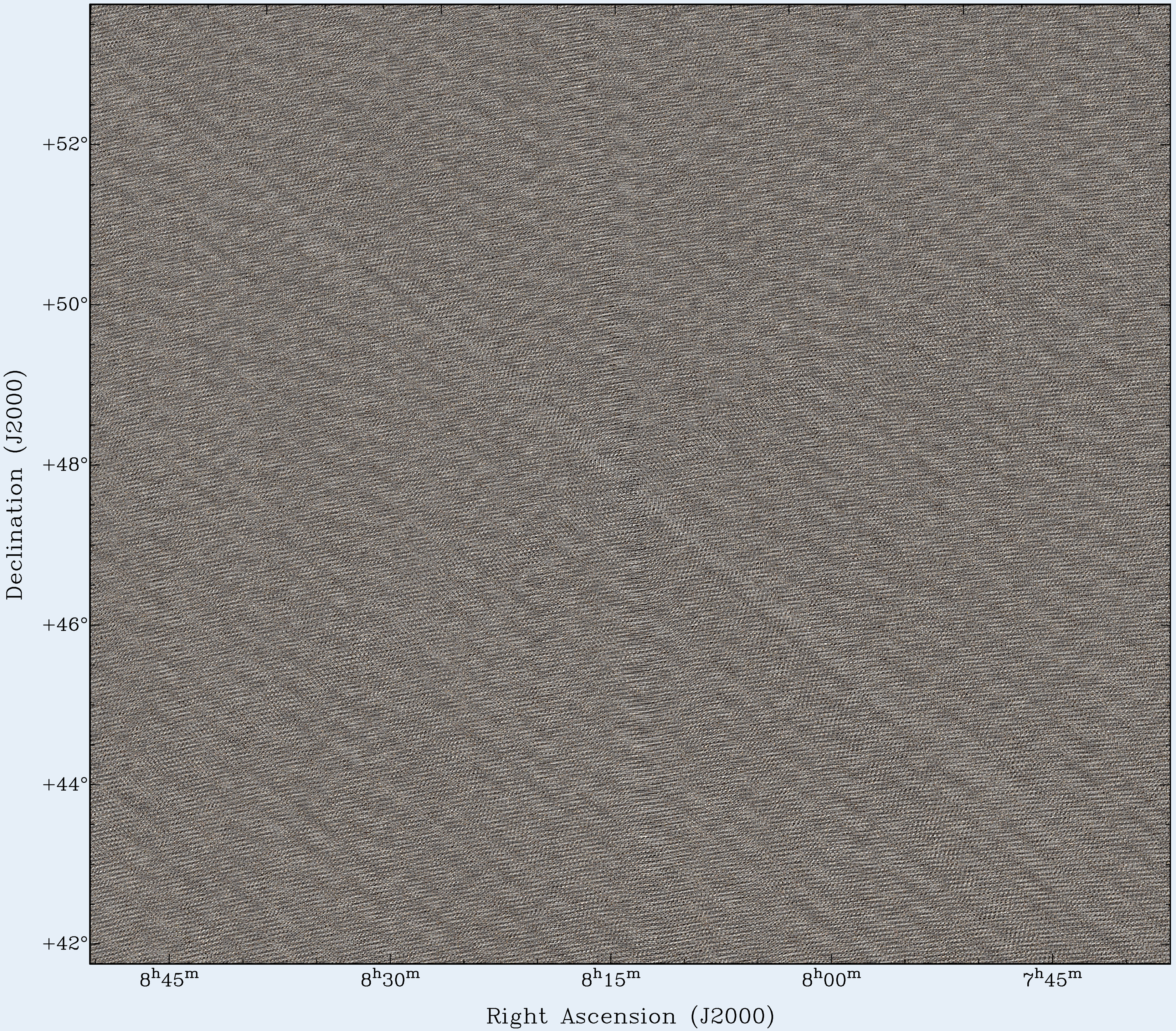
3C196 LBA_INNER: visibilities



3C196 LBA_INNER: visibilities



3C196 LBA_INNER: images

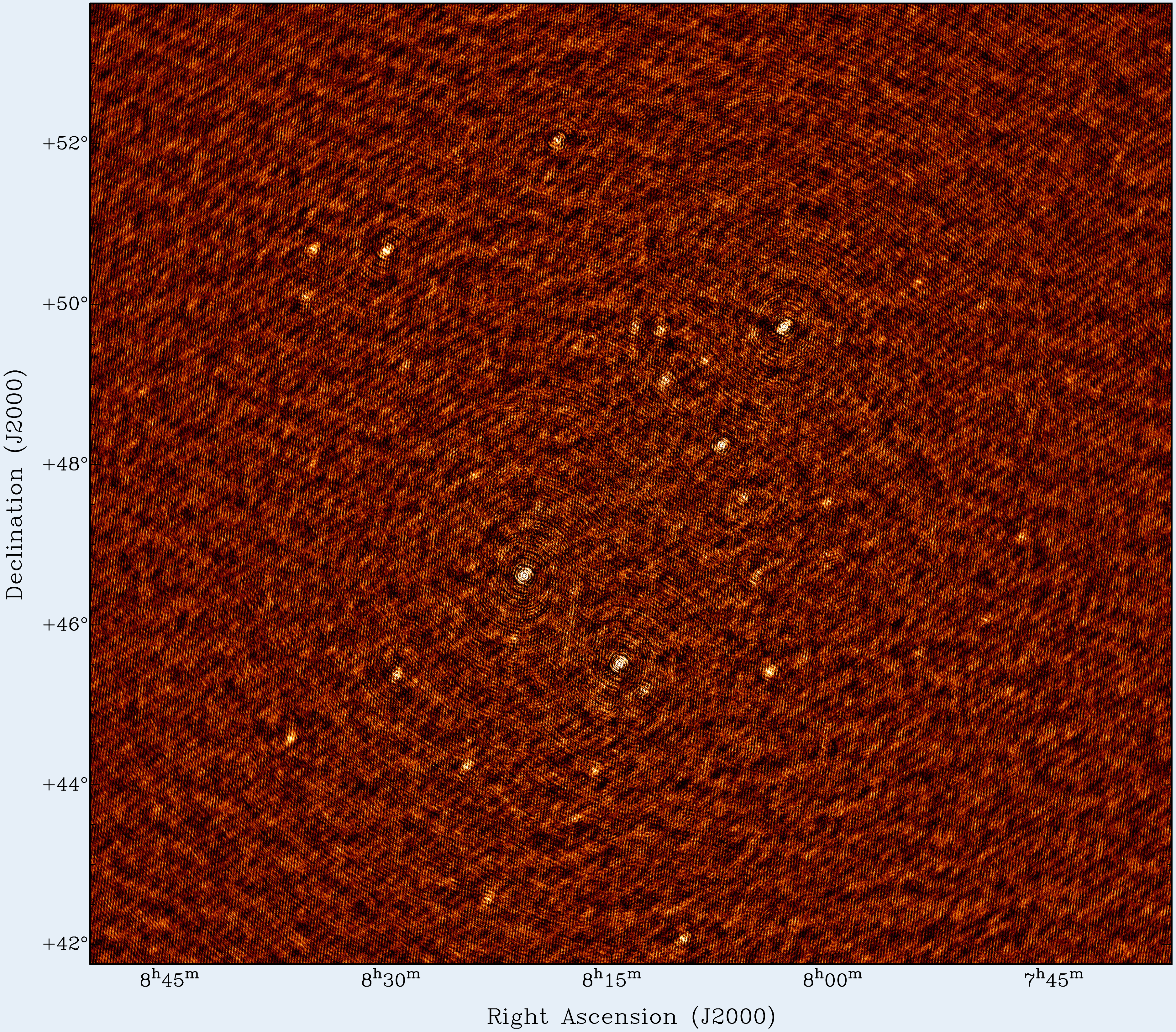


no demixing

averaged to
- one channel
- 10 seconds

calibrated on
3C196
and imaged

3C196 LBA_INNER: images

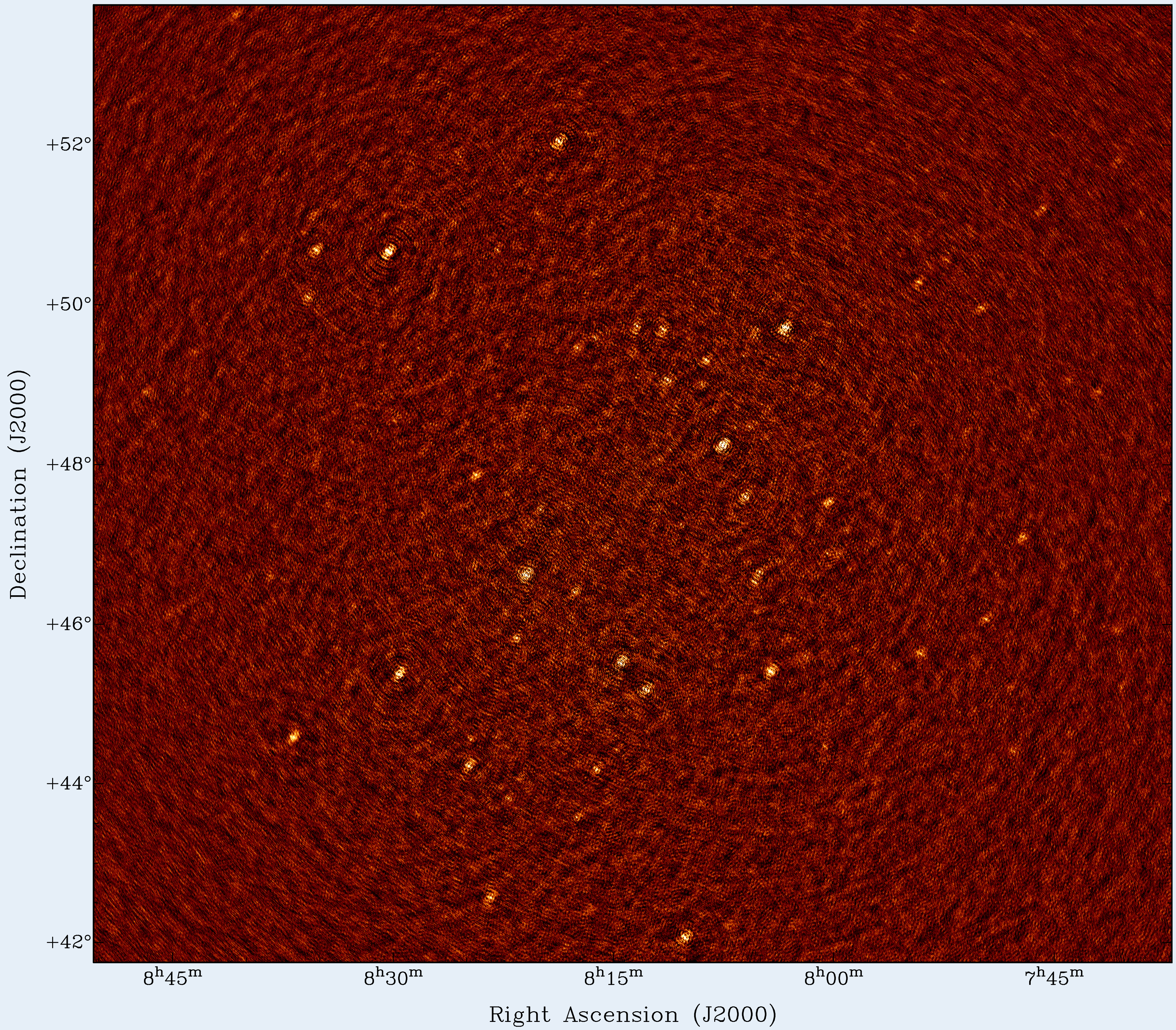


no demixing

averaged to
- one channel
- 10 seconds

post-demixing calibration applied before imaging

3C196 LBA_INNER: images



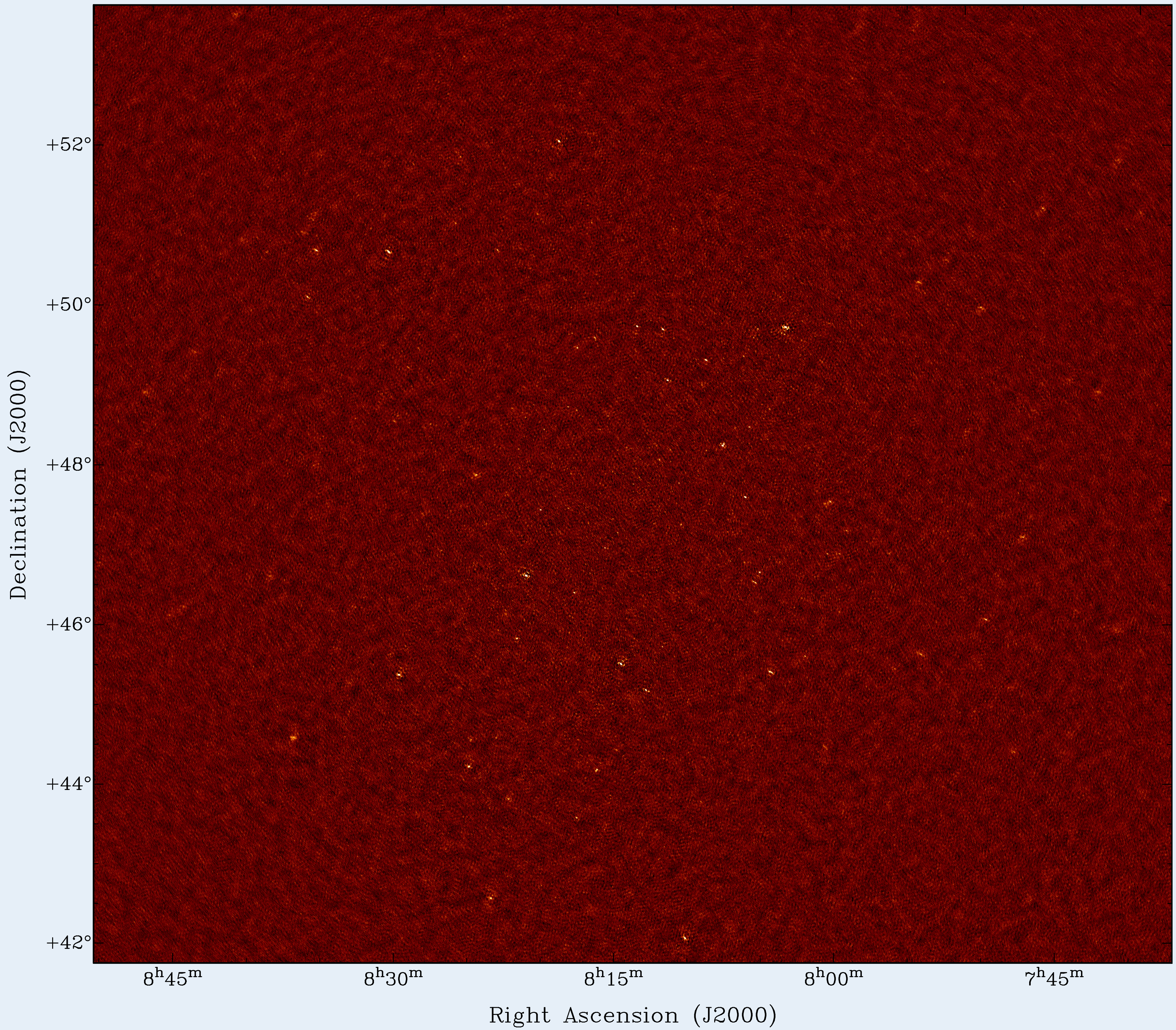
with demixing

averaged to
- one channel
- 10 seconds

post-demixing calibration applied before imaging

no cleaning

3C196 LBA_INNER: images



with demixing

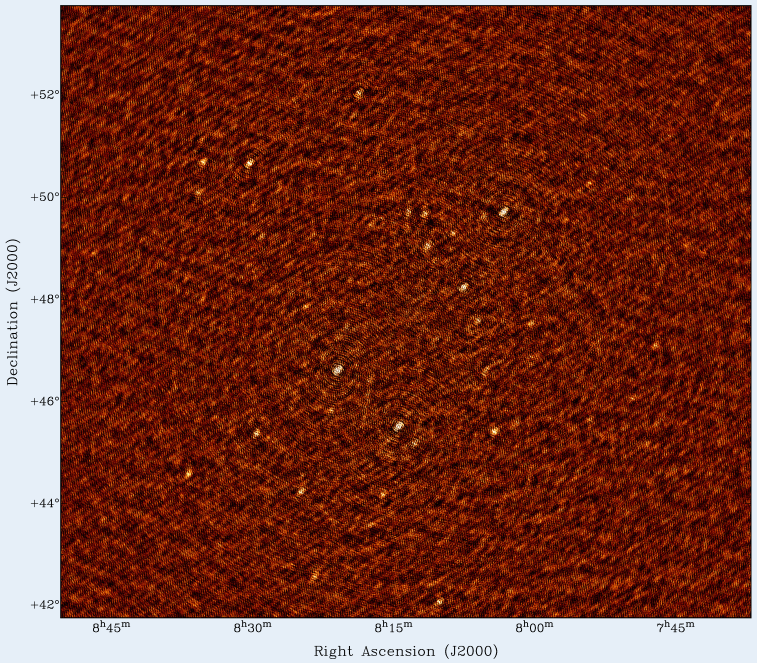
averaged to
- one channel
- 10 seconds

post-demixing calibration applied before imaging

with cleaning

~100 SQ DEG

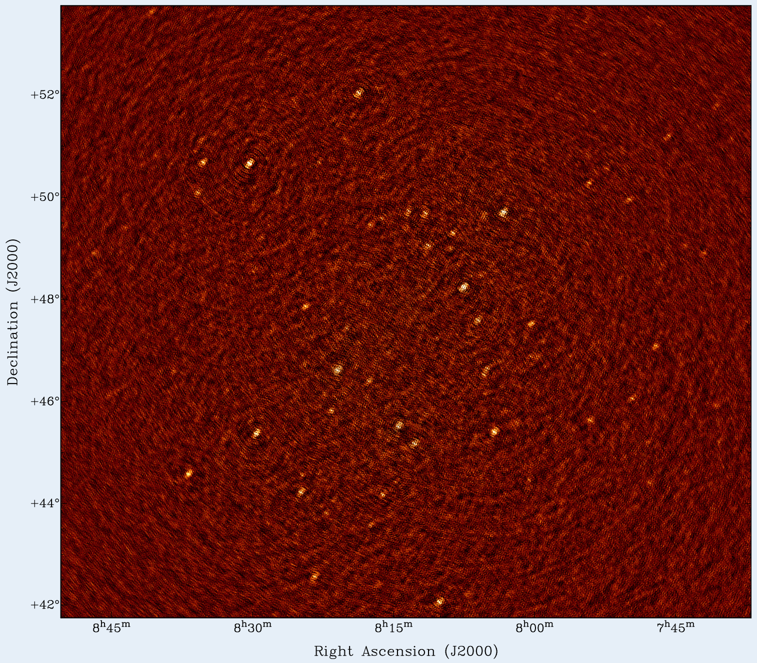
3C196 LBA_INNER: images



no demixing
(but using gains obtained post-demixing)

no cleaning

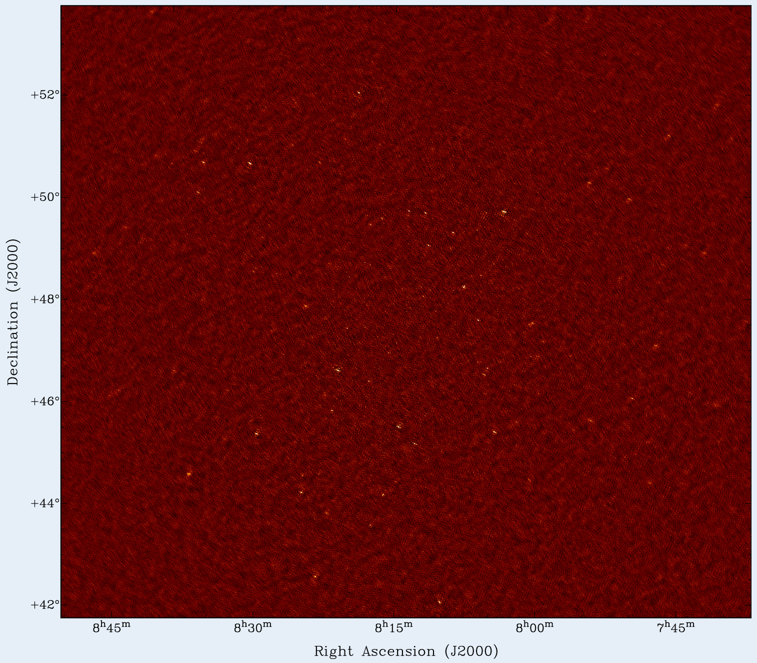
image rms
122 mJy



with demixing
(using gains obtained post-demixing)

no cleaning

image rms
66 mJy



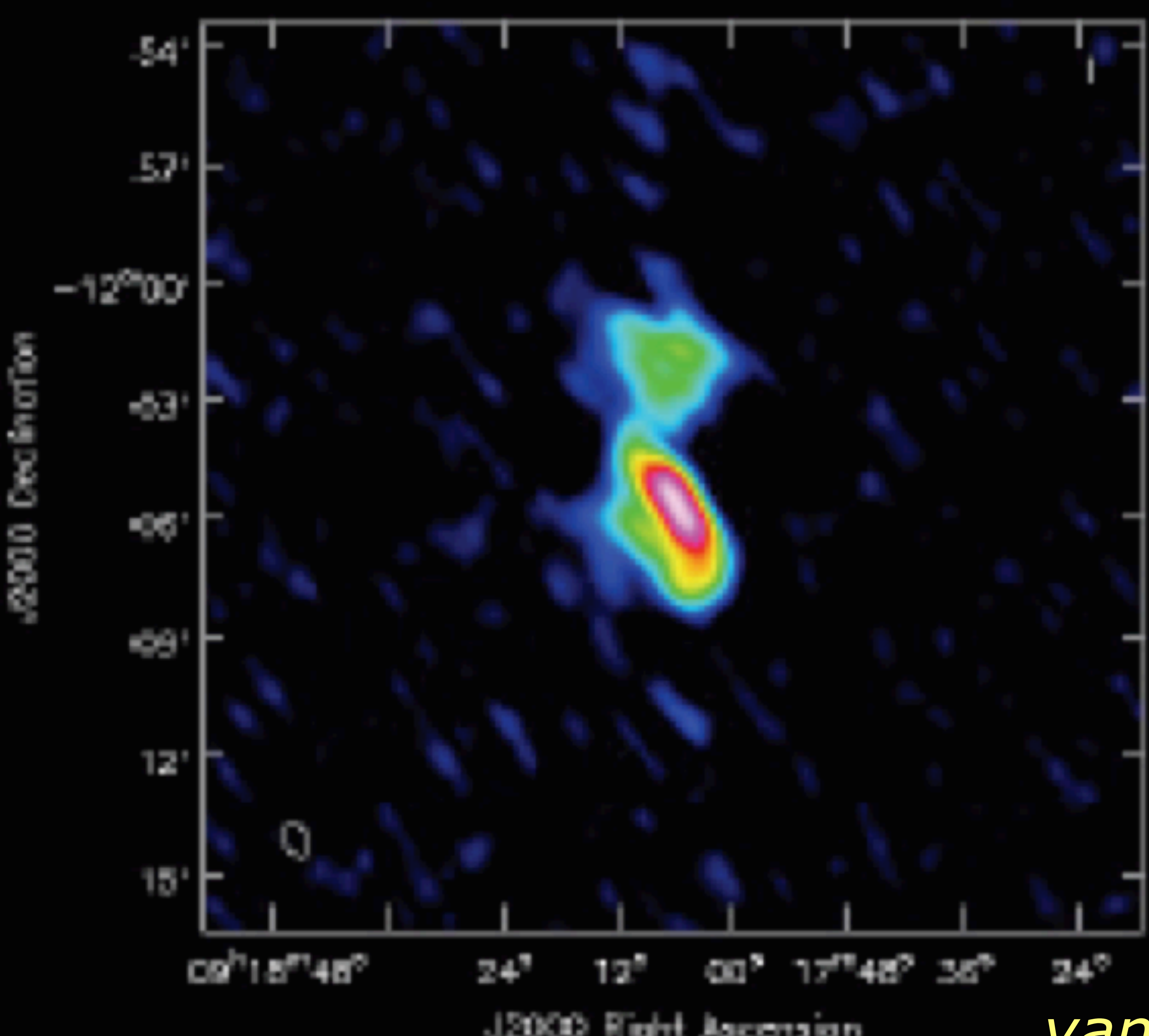
with demixing
(using gains obtained post-demixing)

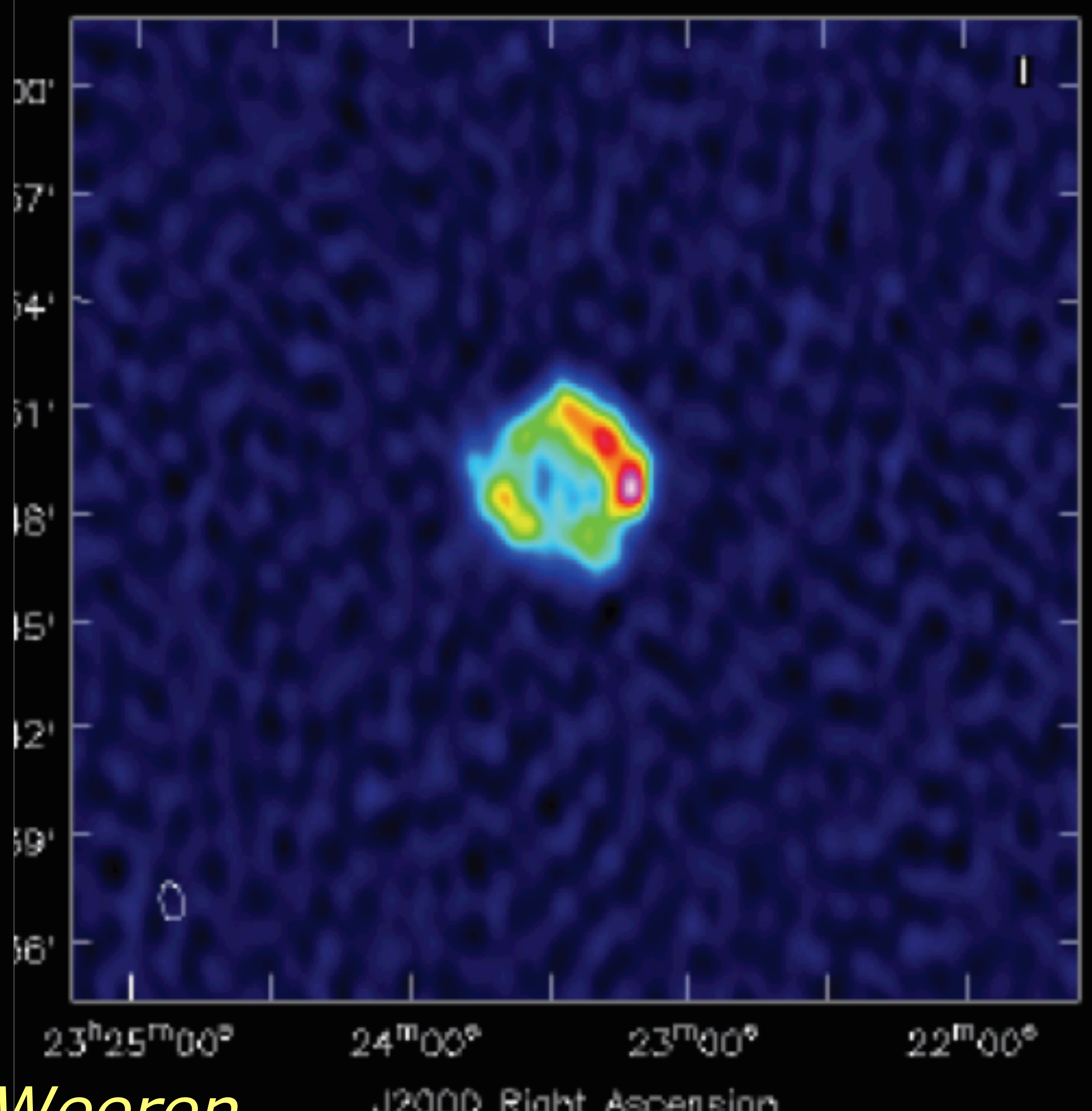
(unguided) cleaning

image rms
40 mJy

Demixing: imaging off-axis sources

- Hydra A - Cas A distance ~ 127 degrees
- Post-demixing images (target = Hydra A):





van Weeren

Averaging

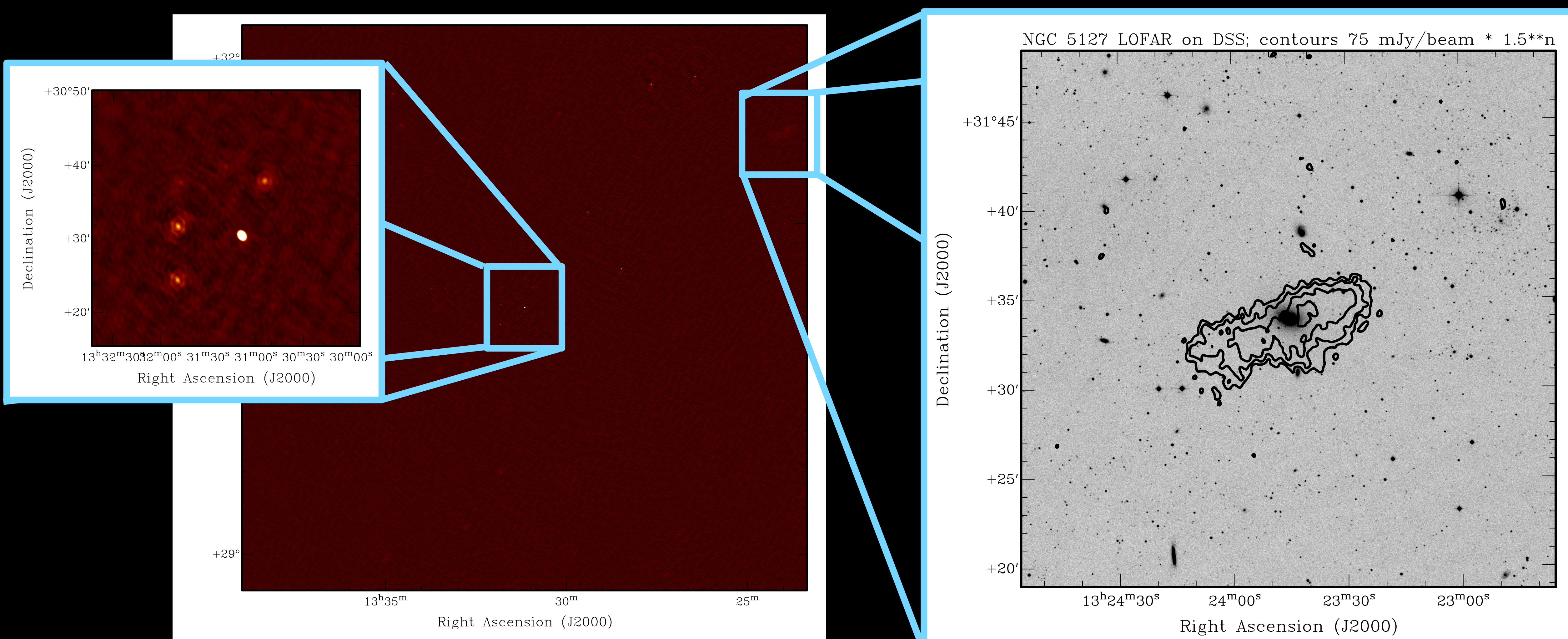
- After demixing, averaging has been done and the data volume is managable
  - each subband typically reduces to about Δt=10sec, N_{ch}=1 (note that such frequency compression will not be suitable for polarization work ...)
  - the compression factor is up to 640 (LBA data compressed to 1ch/SB), but extra columns are added for calibration etc
  - yielding (LBA) data volume decrease from ca. 28 GB to ~214 MB - note HBA_DUAL is ~4x bigger (both the raw data and the processed data)
  - (for 244 subbands, this means currently ~ 52 GB/6 hr LBA observation, or ~200 GB/6 hr HBA_DUAL observation)

Calibration: models

- To begin calibration, either a good field model, or a calibrator gain transfer, is required: correct for clock delays, flux scale
- Currently, catalog extractions (WENSS, VLSS, NVSS) are used to kickstart data calibration. High-resolution images from e.g. VLA (at higher frequencies) are used when available
  - Low resolution models are insufficient to calibrate remote stations in the first pass
    - clean components work better than shapelets or gaussians for extended sources
  - Self-calibration does not work (properly) at this stage due to lack of beam correction in the imaging stage
- Matching resolution is crucial to jumpstarting the calibration cycle on remote baselines

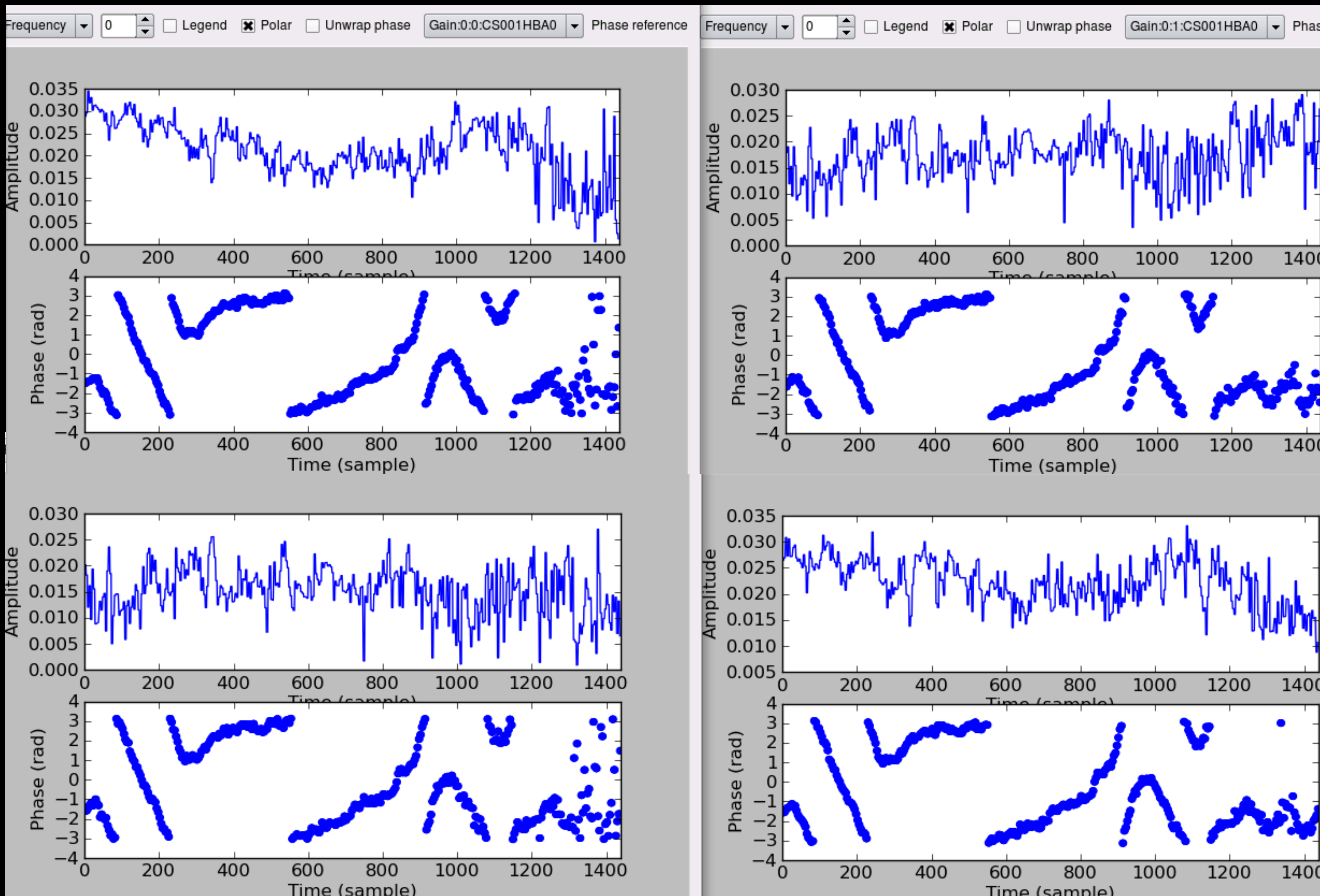
Calibration: solution transfer

- Example: test transfer from 3C286 to NGC 4631
  - Calibrator-source angular separation ~15 degrees
  - Observation at 150 MHz
  - Bandwidth evenly divided between calibrator and target field - same frequencies observed in two directions simultaneously



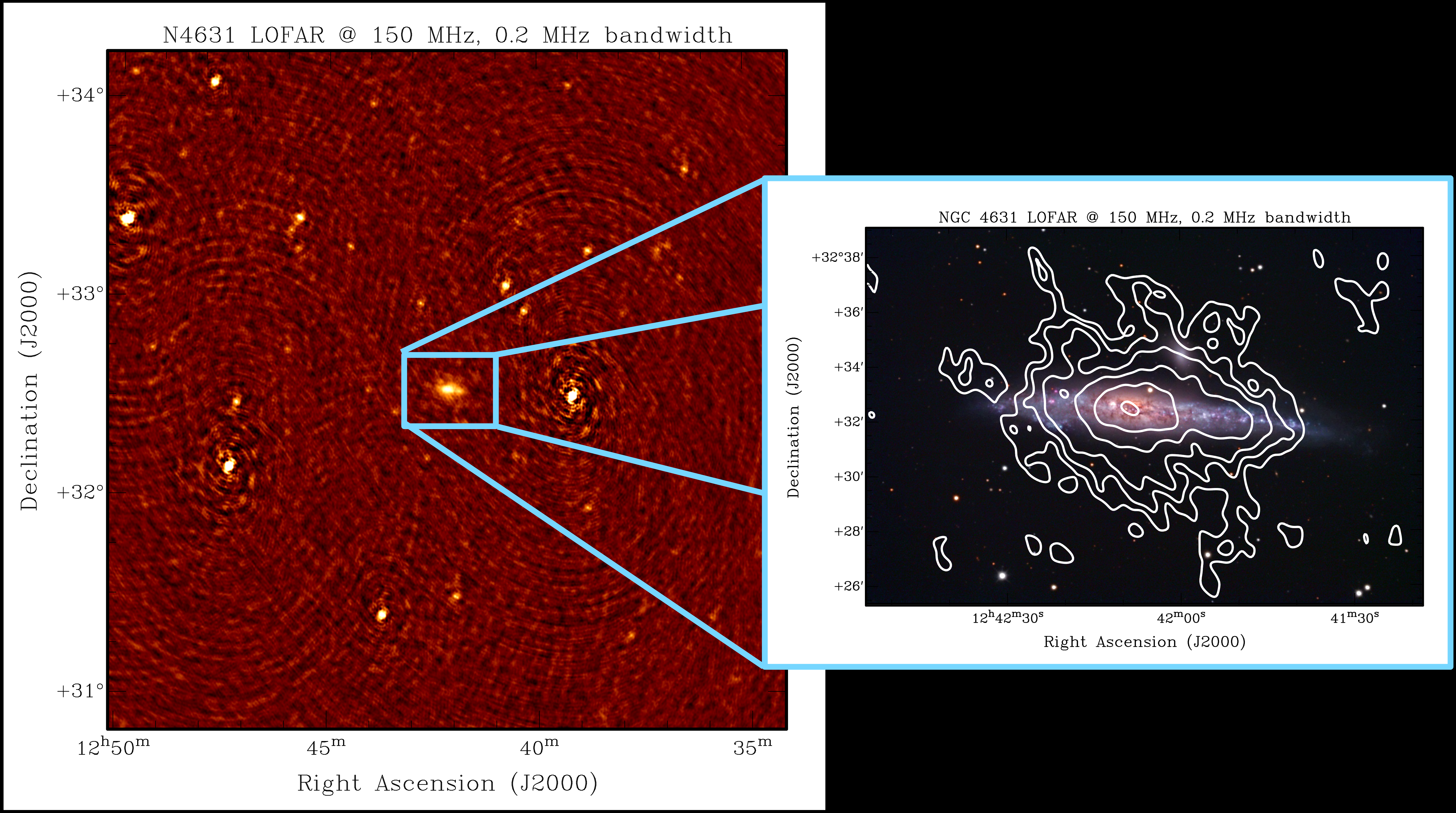
Calibration: solution transfer

- Gain solutions from 3C286:

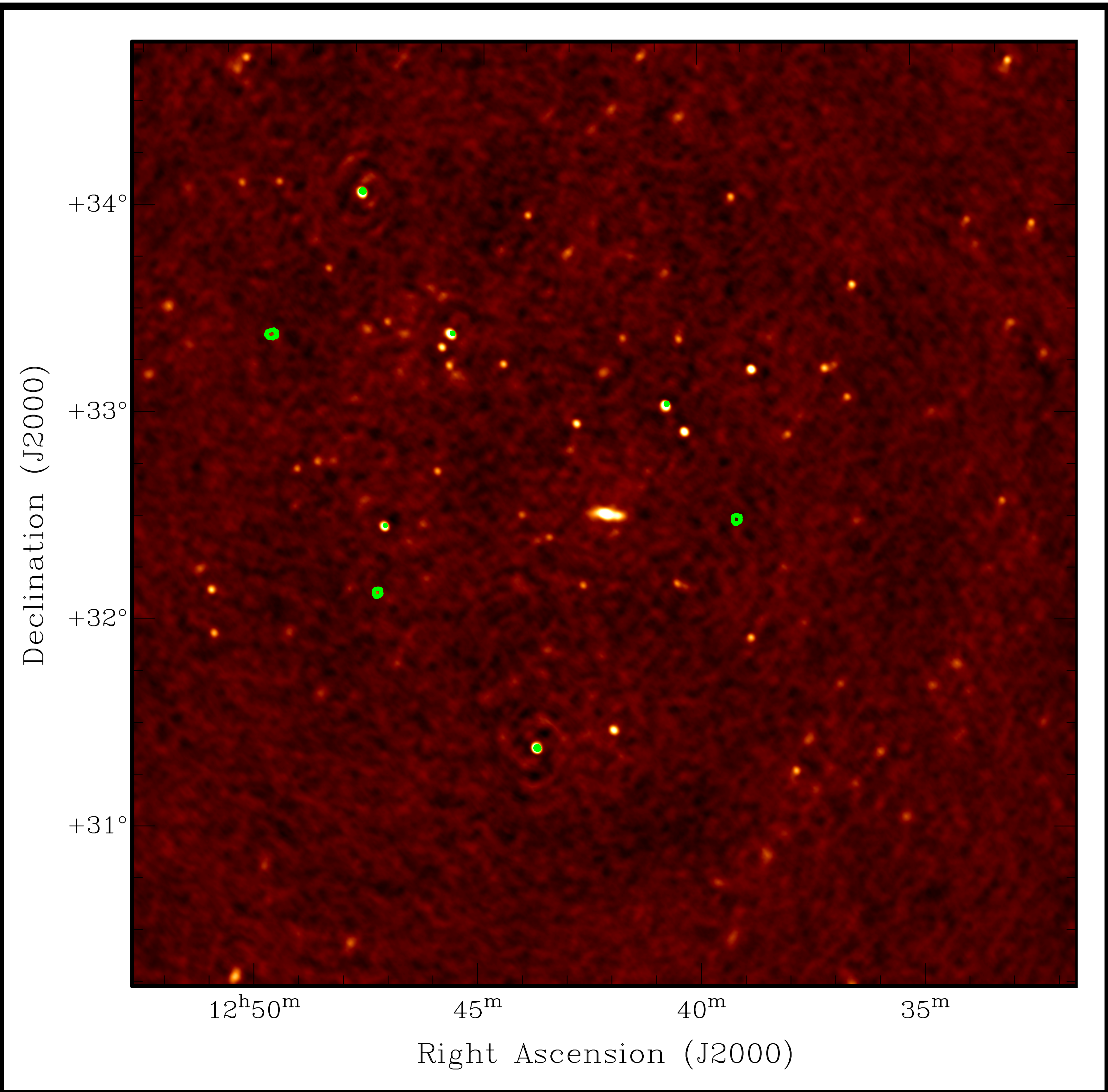


Calibration: solution transfer

- Image of NGC 4631 field after applying gains (and nothing else)



Removing the brightest sources

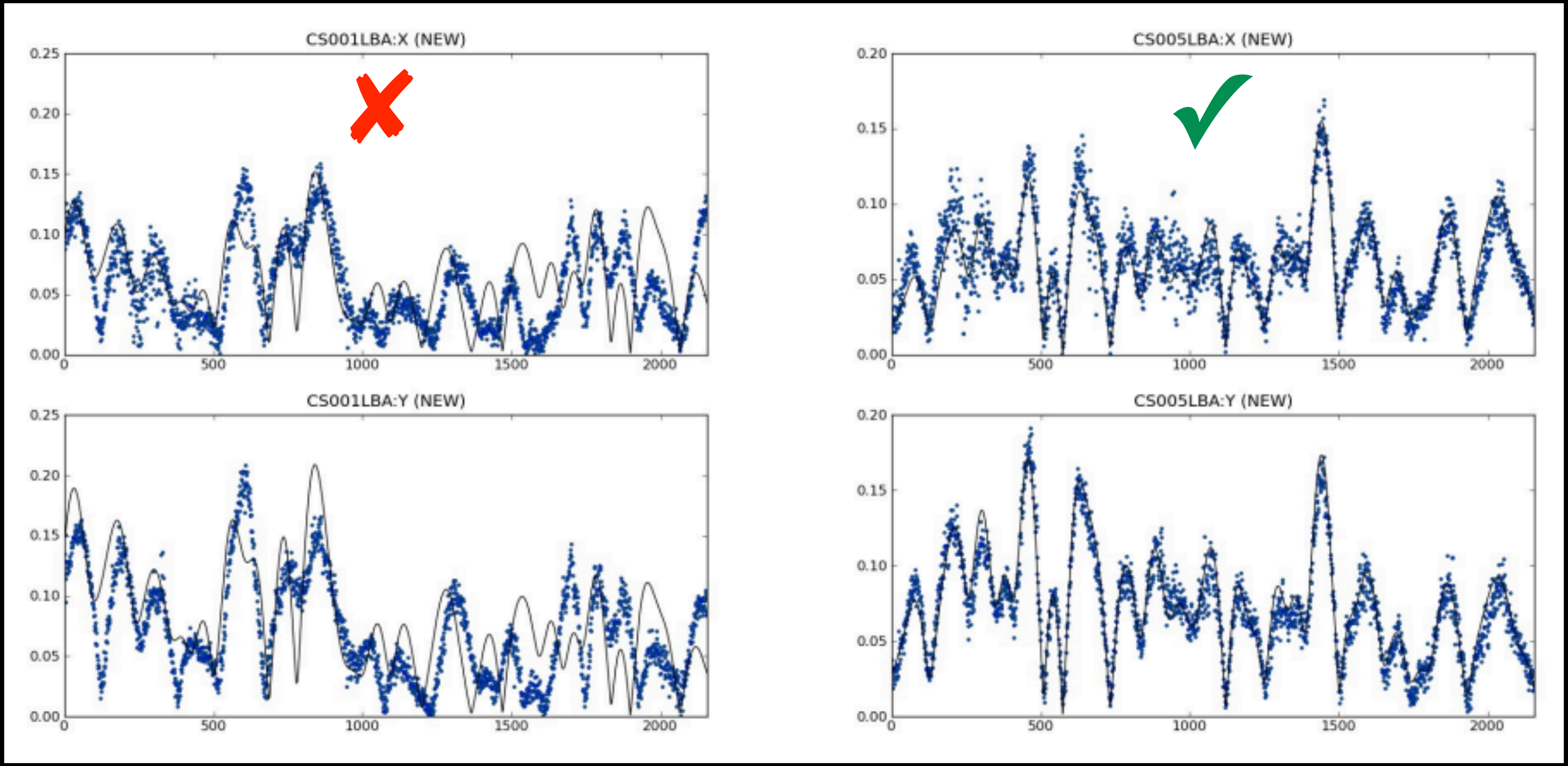


Calibration: solution transfer

- Instrumental polarization in the calibrator field reduced from several percent (strongest in Stokes U) to ~noise level
- Application of these gains to target field did *not* eliminate the instrumental polarization
  - found to be due to incorrect HBA tile beam prediction in beam module (used in BBS and awimager)
- Tile beam prediction has been updated, new test planned to check that instrumental polarization can be calibrated out...

Calibration: LOFAR beams

- For the most part, the beam predictions seem decent - but not on all stations - (related to polarization swaps?)



van Zwieten

Calibration: timing

- Runtimes (on *old* cluster) typically ~30 min / 6hr dataset (after averaging to 10s/1ch; doing (predict), solve, subtract, correct)
- The runtimes increase strongly with extra features
  - Complexity of models: reiterates need for FFT-based predict step!
  - Direction-dependent gain solutions
  - Beam prediction

- Note the need for benchmarks relevant to MSSS-type observations and major cycle characteristics: one reason for MSSS-Test1,2,3 as described later

Calibration: results

- After demixing and calibration, clipping and flagging are usually required to eliminate bad corrected visibilities
  - S/N flagging as in casa or AIPS is probably needed
  - Solver statistics are being investigated
- Output images post-calibration are getting closer to thermal noise estimates, but in many fields we are limited to ca. 10-20 times the theoretical thermal limit: (better with awimager??)

Imaging

- Current development is in the casa suite - working on ‘awimager’.
  - this is an updated version of the lwimager, taking LOFAR beams into account and doing A-projection
  - development still concluding, with initial test program already in place and being analyzed with simulated data
  - time required for imaging should not much exceed ‘normal’ widefield imaging in casa, since the calculation of the convolution functions needed for A-projection is not a dominant additional time

Major cycle

- To close the major cycle we need a trusted version of the imager that will provide us with confident source positions and fluxes (pre- and post-deconvolution)
- Other basic tools are in place:
  - source finding routines (pybdsm and pyse)
  - database I/O in place and being tested

Pipeline times (and unknowns...)

LOFAR ASTRON

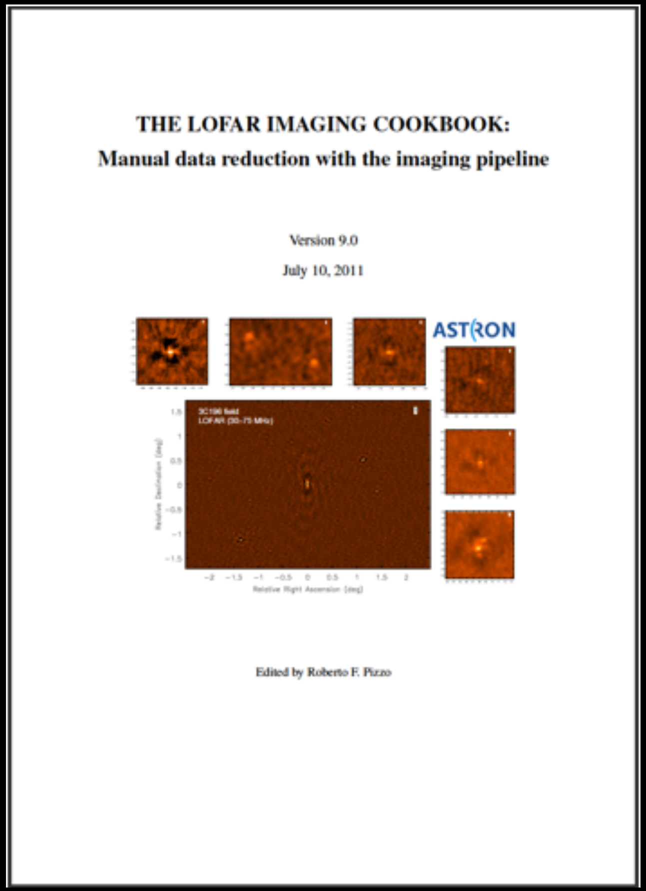
- Pipeline overview, and timings of components (example 6hr LBA)

Number of major cycle loops to be determined
Needs to be minimized...

flagging	demixing	calibration	imaging	source extract & catalog
works and fast	works but slow (fast implementation in the works)	works and (can be) fast	does not work and slow (new imager in the works)	works and fast
~15min	~5hr (demixing Cyg,Cas)	~30min	?????	~5min (?)

People working on the SIP

- Many students and postdocs are taking active roles in commissioning the imaging through Busy Weeks and Busy Wednesdays (which are ***very*** useful...)
- The KSPs have many students and postdocs available now working on imaging developments
- LOFAR Imaging Cookbook in good shape and available online:
  `http://www.astron.nl/radio-observatory/lofar/lofar-imaging-cookbook`



THE LOFAR IMAGING COOKBOOK:

Manual data reduction with the imaging pipeline

Version 9.0

July 10, 2011

Edited by Roberto F. Pizzo

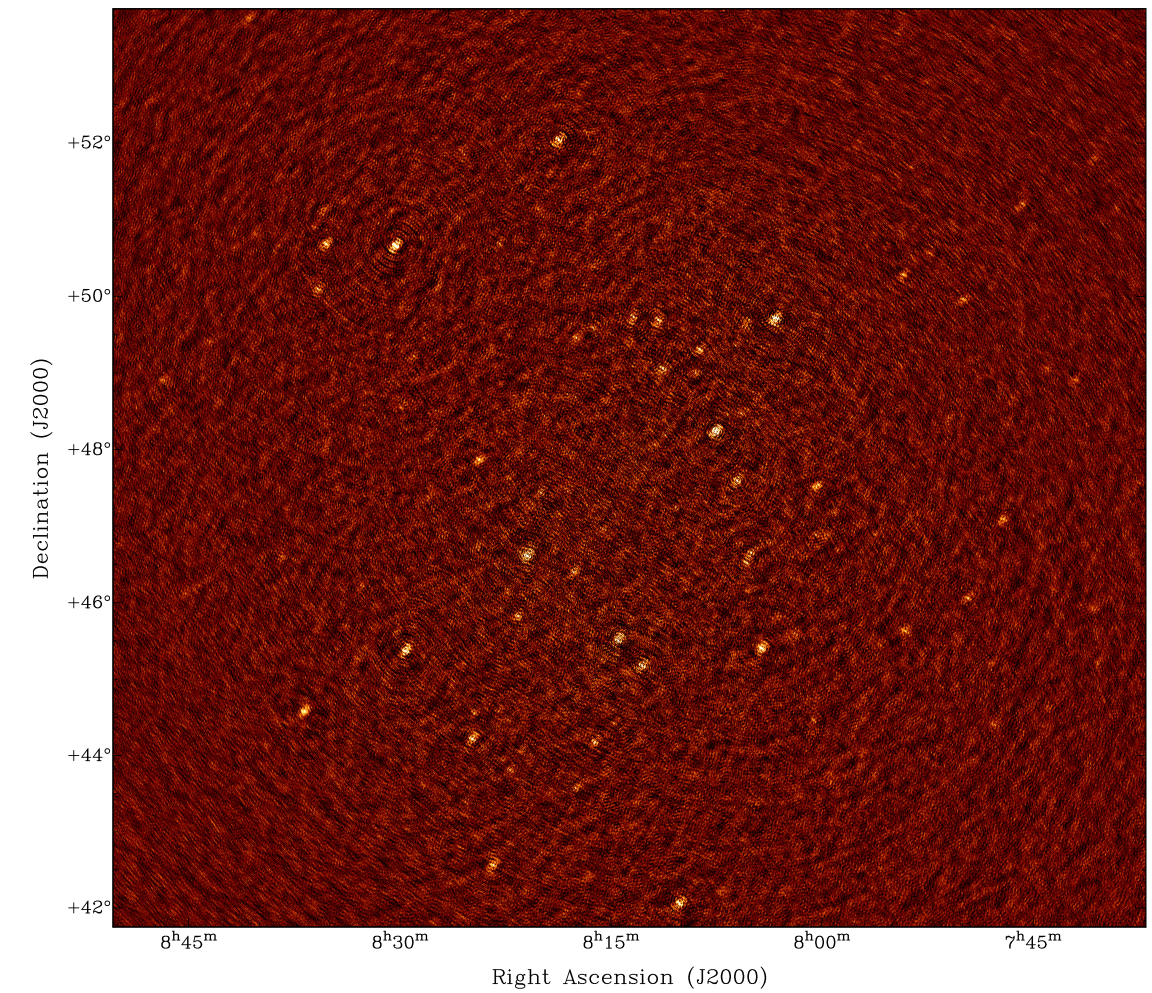
Next (major) steps

- Understand limits of calibrator gain solution transfer
- Continue work on understanding, and speed up, demixing
- Speed up model prediction in BBS
- Commission new imager
  - Give feedback to Beam Team, update beam models, iterate
- Exercise source finders, and test GSM/LSM I/O
- Define major cycle stopping criteria (*for abitrary fields...*)
- Practice with early MSSS-like observations

MSSS, AKA LOCo

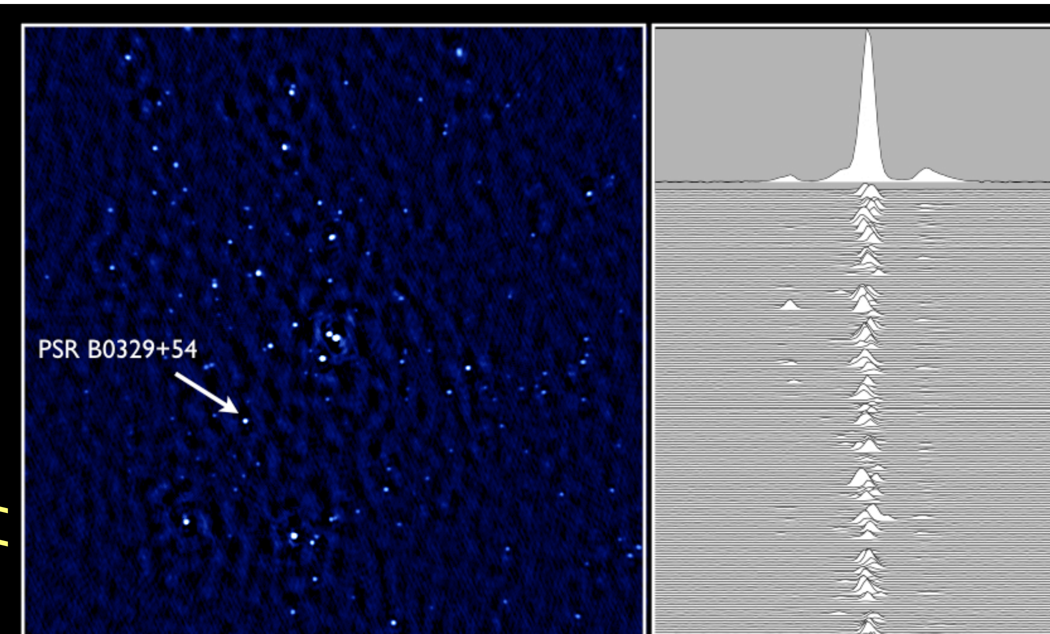
- MSSS: the LOFAR Commissioning Survey
- Key roles:
  - Fill the initial GSM for calibration of arbitrary fields at arbitrary frequency in LOFAR bands
  - Guide development of & exercise:
    - observatory operations
    - processing software
    - imaging pipeline
    - piggyback applications
- Input from KSPs is being folded into the planning
- Expected survey time ~2-3 months depending on efficiency

Heald et al.



3C196 LBA field, ~60 mJy
(MSSS observations ~4x deeper)

Stappers et al.



Motivations for "core-only" MSSS

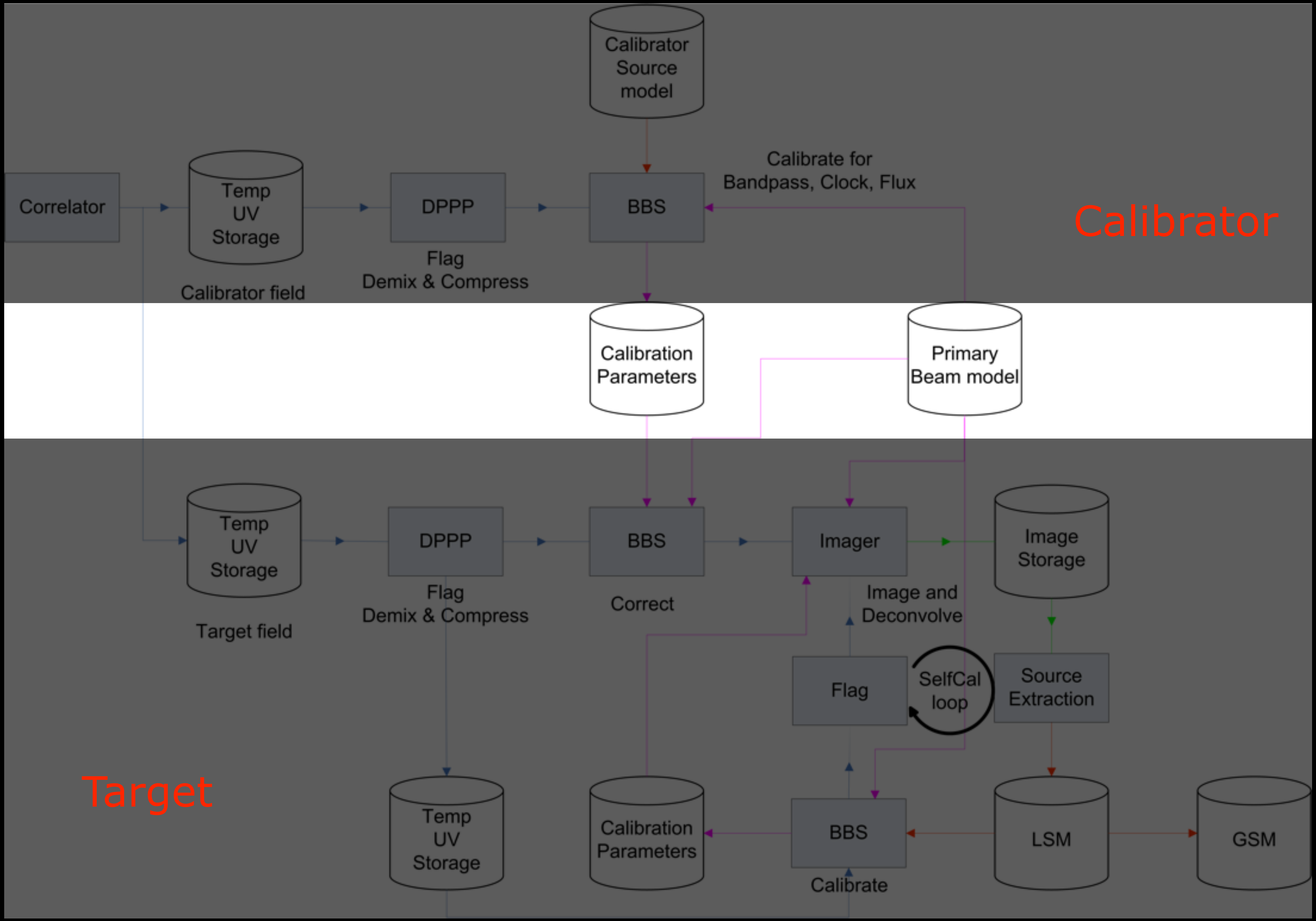
- This really means 3km in HBA; 10km in LBA, which would yield similar characteristic beamsizes in both bands, of ~1.5-2 arcmin (@ 60,150 MHz)
- It also refers more to the (initial) *processing* than to the observations - the plan is to use all (at least all Dutch) available stations

- Key reasons (more on next slide):
  - Beam information is not available (yet) in the imager, so differing HBA station sizes could not be taken into account (this may be a moot point in the near future...)
  - Processing time increases with baseline length, so using a compact array makes real-time processing more realistic

Benefits of a core-only MSSS

Band & setup	Pros	Cons
MSSS-LBA (CS+RS; baselines ≤ 10km, with multiple snapshots; LBA_INNER)	•Same angular resolution as MSSS-HBA •Within a snapshot, the beam variation is minimal •No station size (RS-CS) difference •Ionosphere minimized	•Requires multiple snapshots for *uv* coverage •LBA_INNER needs testing
MSSS-HBA (CS only; single snapshot; HBA_DUAL)	•Obviates need for beam-corrections in imager •Good snapshot uv coverage •Relatively rapid pipeline processing (***needs benchmarking***) •Within a snapshot, the beam variation is minimal	•Resolution and sensitivity not competitive for scientific use?

MSSS Pipeline



Plans for specifications

- MSSS-LBA
  - Array configuration: 24 CS + 11 RS (LBA_INNER)
  - Bandwidth: 16 MHz over 30-78 MHz [coverage uncertain]
  - Number of independent beams: 3
  - Time per field: 90 minutes (9x10 minutes; δHA~0.5h)
  - Resulting sensitivity (approximate): ~15 mJy
  - Required number of fields: 619 (covering 2π sr)
  - Required on-source observing time for full survey: 619/3x90min = 309.5 hr

Plans for specifications

- MSSS-HBA
  - Array configuration: 24 CS (HBA_DUAL?)
  - Bandwidth: 16 MHz over 120-168 MHz [coverage uncertain]
  - Number of independent beams: 3
  - Time per field: 15 minutes
  - Resulting sensitivity (approximate): ~5 mJy
  - Required number of fields: 3522 (covering 2π sr; see above)
  - Required on-source observing time for full survey: 3522/3x15min = 293.5 hr

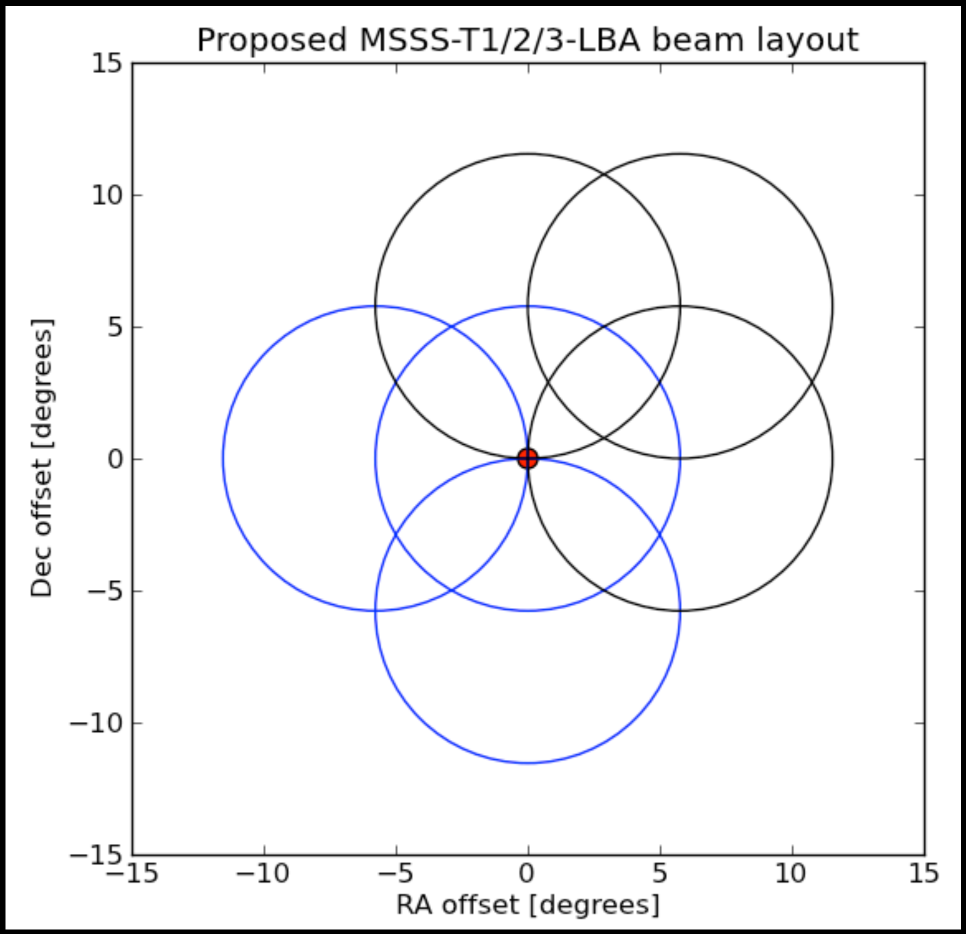
Comparison of relevant surveys

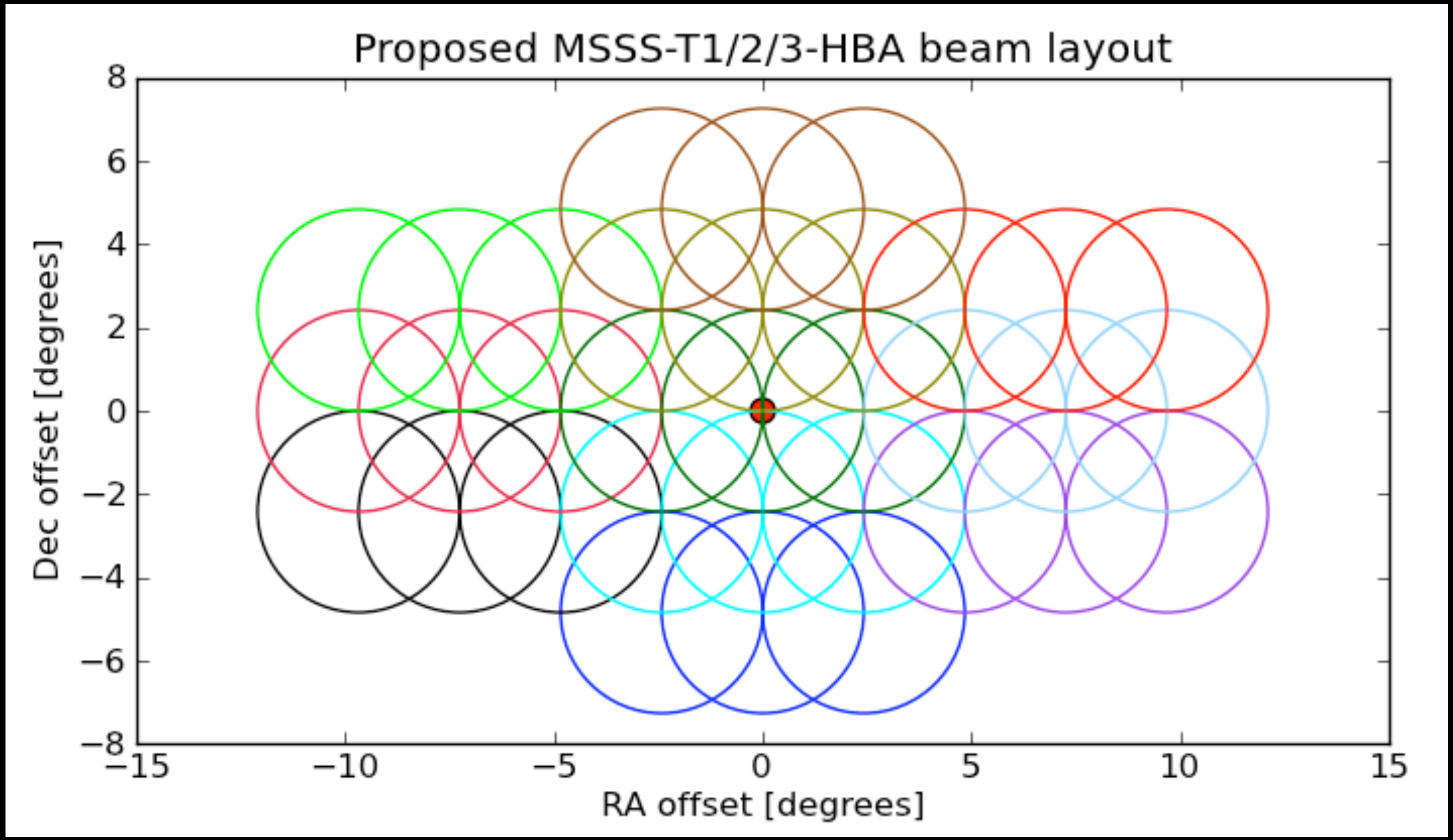
Survey	Frequency	Sensitivity	Resolution	Sky coverage
MSSS-LBA (core only)	~30-78 MHz	≲ 15 mJy/beam [1,2]	≲ 100” [1]	20,000 square deg (dec > 0)
VLSS	74 MHz	100 mJy/beam	80”	30,000 square deg (dec > -30)
MSSS-HBA (core only)	~120-170 MHz	≲ 5 mJy/beam [1,2]	≲ 120” [1]	20,000 square deg (dec > 0)
TGSS	140-156 MHz	7-9 mJy/beam	20”	(2100 of) 32,000 square deg (dec > -30)
WENSS	330 MHz	3.6 mJy/beam	54”	10,000 square deg (dec > 30)
NVSS	1400 MHz	0.45 mJy/beam	45”	35,000 square deg (dec > -40)

[1] Conservatively based on short-baseline-only processing; longer baselines will be available
[2] MSSS sensitivities to be verified during test observations .... !

Proposed test pointings

- Each set of pointings covers approx 200 square degrees
- Key is to make these as realistic as possible
  - test, understand, and optimize: data taking, handling, processing, major cycle algorithm, pipeline runtime, catalog creation, etc etc etc



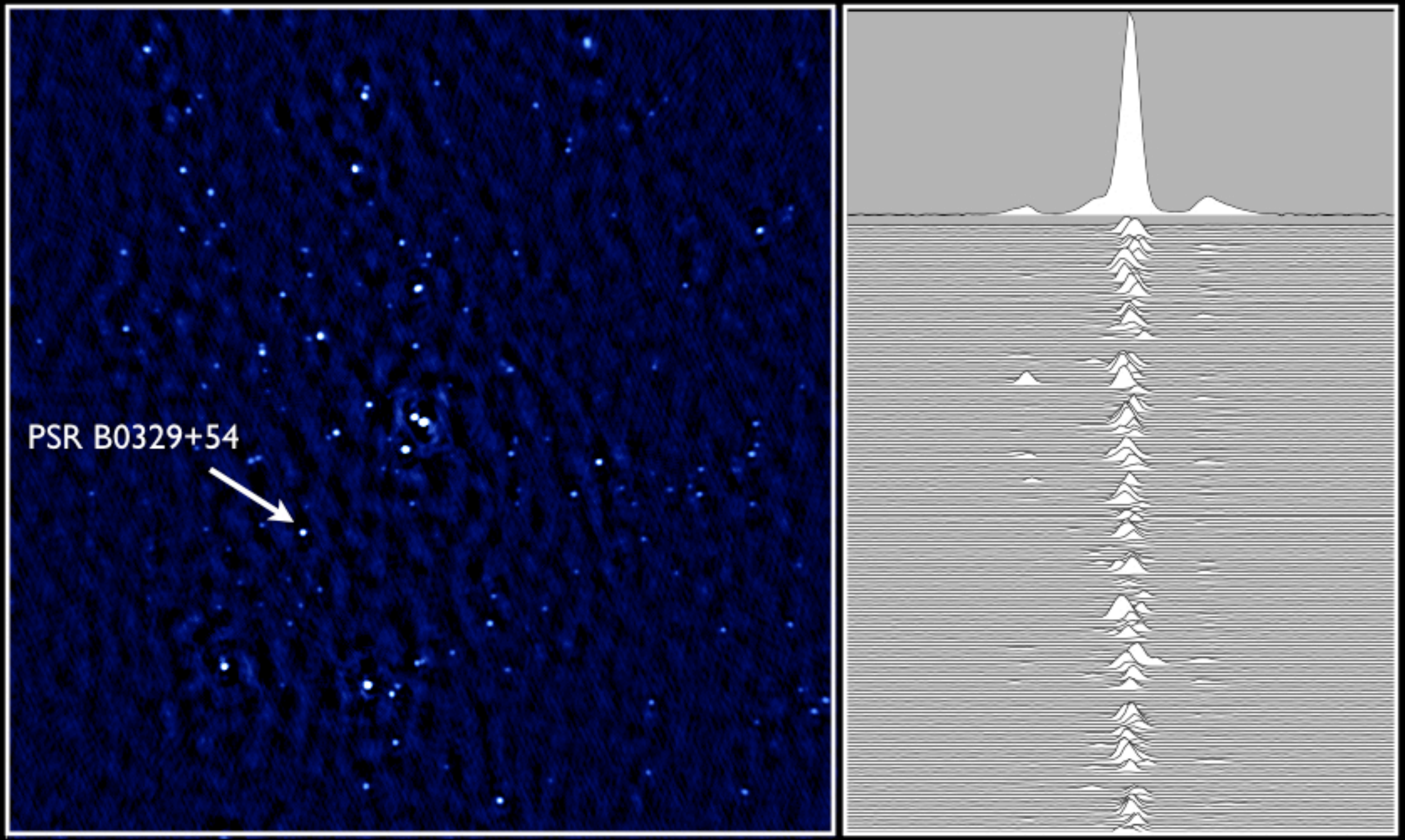


Proposed test pointings

- Test data processing and output catalogs in three regimes:
  - Test1: near bright, simple source (3C196)
    - field already well understood
    - it won’t get any easier than this ....
    - also contains another mid-complicated 3C source (3C219) about 10 deg away, so leads to Test2
  - Test2: near bright, complicated source (3C465)
    - bright in-field calibrator, but not as simple
    - intermediate case for closing the major cycle
  - Test3: blank field (where??)
    - no bright in-field calibrator
    - must understand (blindly) closing the major cycle

Piggybacking

- Piggybacking welcomed for at least part of the MSSS area
- Assessment of impact on observing and processing strategy required to ensure that the core MSSS goals are not affected...
- As much as possible, aim to include piggybacking applications in MSSS test observations



Stappers et al.

Constraints and output

- Constraint: minimize survey time (on-telescope and processing)
  - allow maximal amount of other project time
  - avoid tying up resources (cpu, storage) on cluster
  - provide MSSS skymodels for later observations of same area

- Output
  - key deliverable is a wideband catalog of the brightest sources in the LOFAR sky
  - positions, fluxes, source extent and orientation

Summary

- Calibration of individual LOFAR fields is well in hand
  - Individual components of standard imaging pipeline being commissioned and integrated into automated version
  - New imager is the next big step for us ... it allows proper imaging and to close the (calibration) major cycle
- Major effort now on preparing for MSSS/LOCo
  - Test observations starting
  - Automated pipeline (and kickoff at observatory level) now in active development
  - LOCo will be the first LOFAR survey, and will provide a 20,000 square degree catalog of the 30-180 MHz sky