Lofar data reduction of a wide field of view

Emanuela Orrú (ASTRON)

ERIS 2013 Friday 13 September

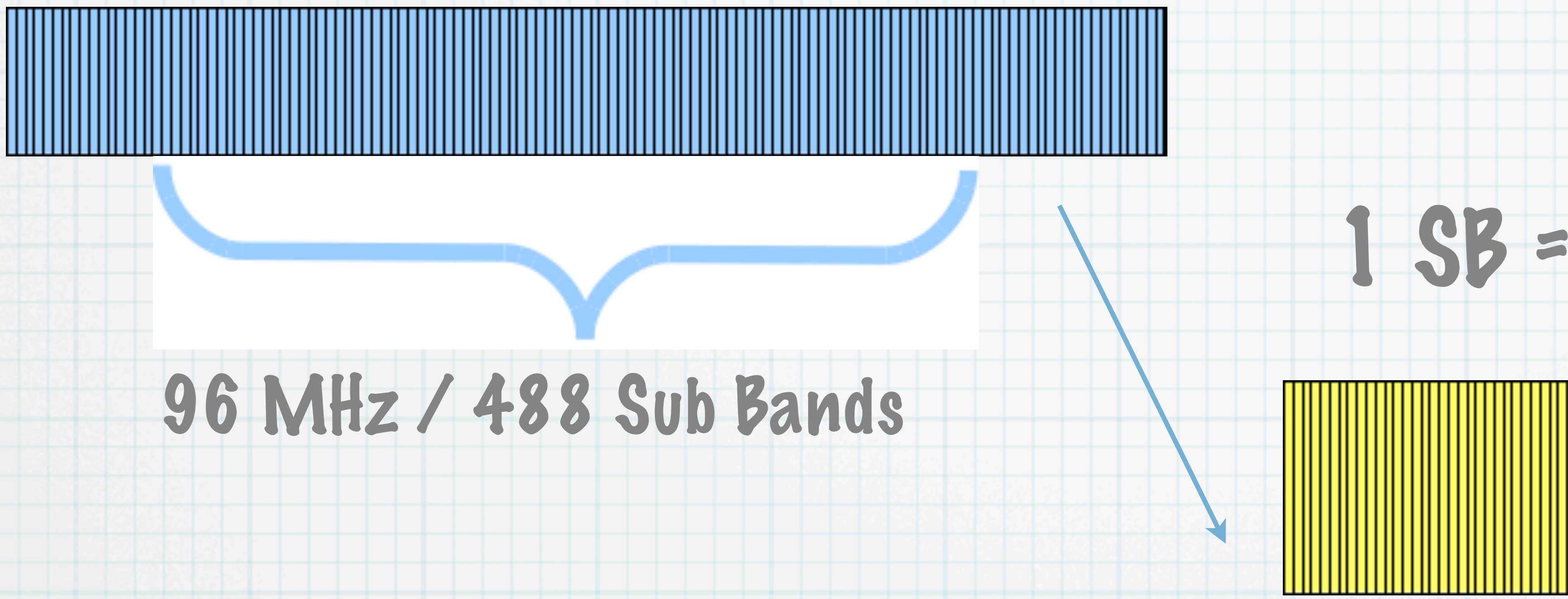
outline

* data size and structure
* RFI
* A-team removal
* Calibration
* DDE
* Imaging

1: Data size and structure

* From the station the signal is correlated by the software correlator BP/P soon COBALT.
* observing modes: 16, 8, and 4 (still under test) bit allow to record respectively 48, 96 and 196 MHz bandwidth.
* e.g 6 hrs observation run ILT in 8 bit is about 50 TBs.

Data size and structure



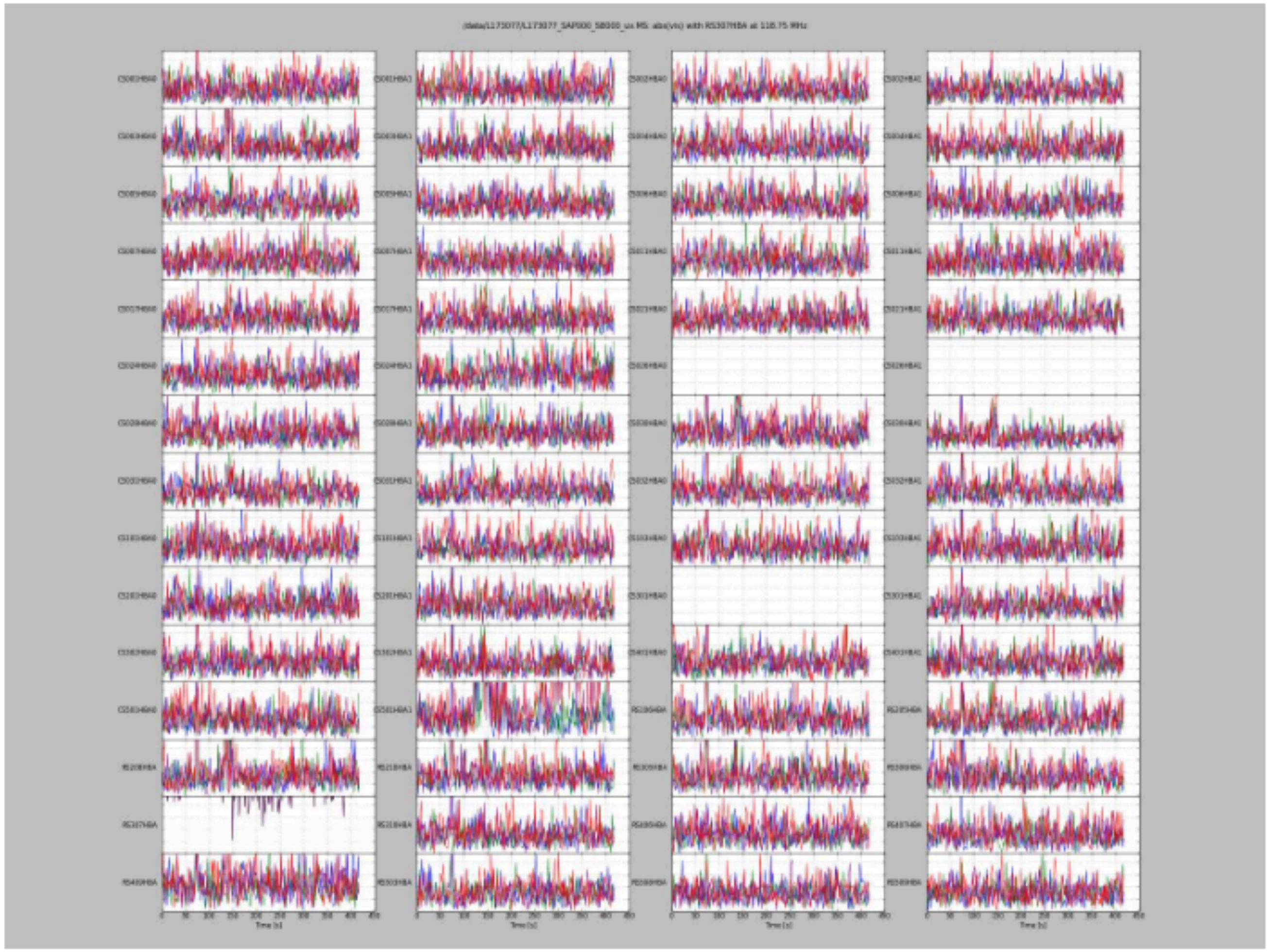
* the size depends on: n of bits, n of SBs, n of channels, integration time
* to be usable the data need to be compressed: this happens in the pre-processing phase.
* the calibration takes place on single SBs or in groups of SBs there is not software or computer in our hand able to handle 50 TB of data

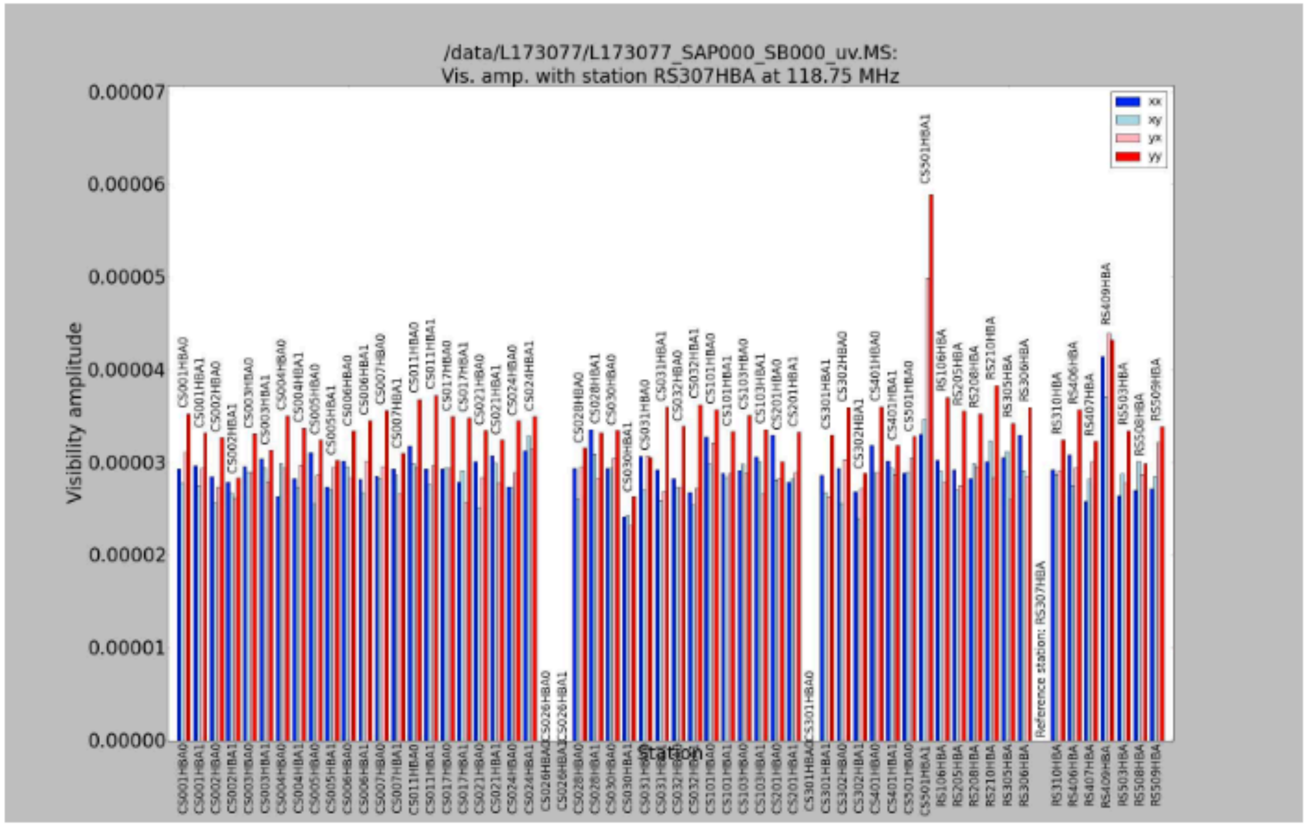
Observing strategy

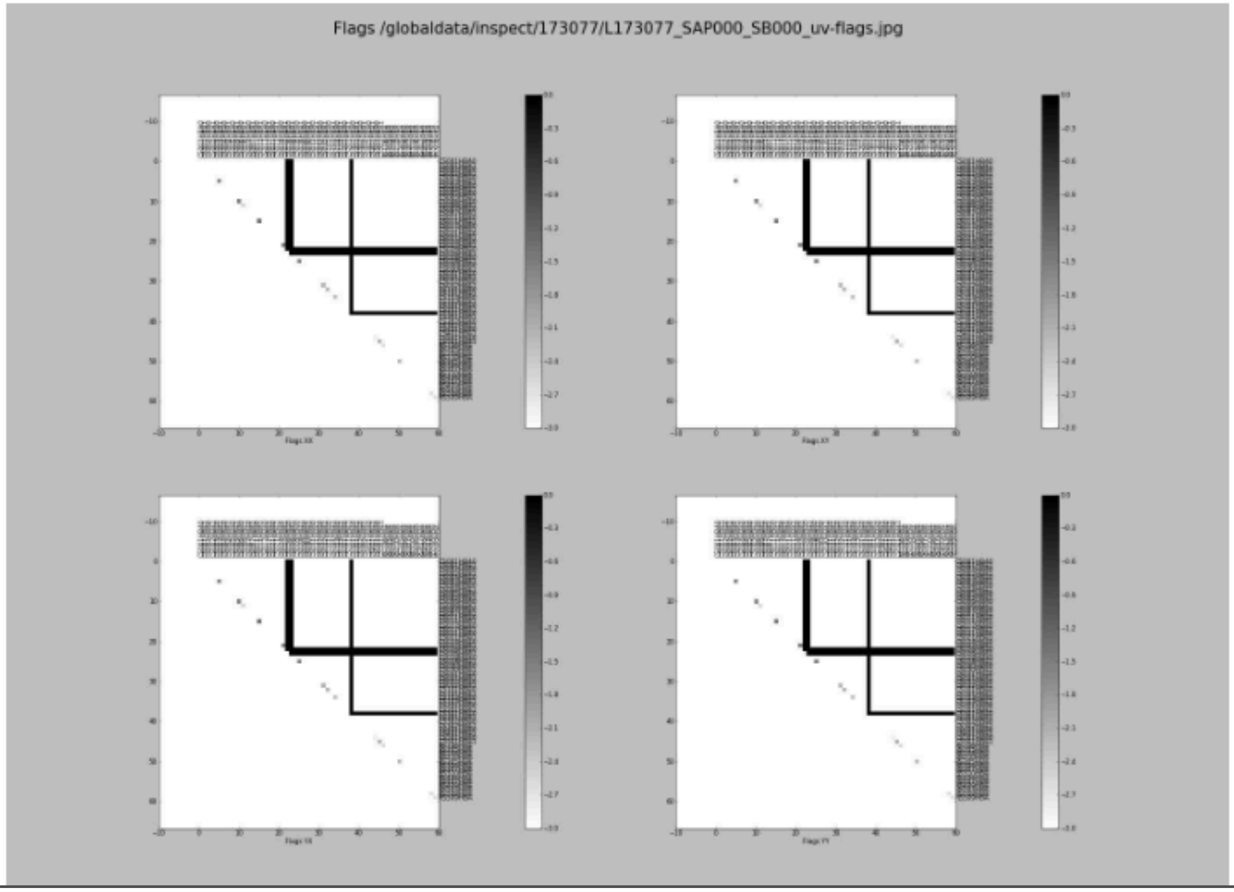
* **LBA** Single observations that are continuous in time/Hour Angle Half of the available bandwidth on the target field (BW<=48 MHz, <=244 subbands) and half on a calibrator. Any direction on the sky since the LBA dipoles are sensitive to the full sky.

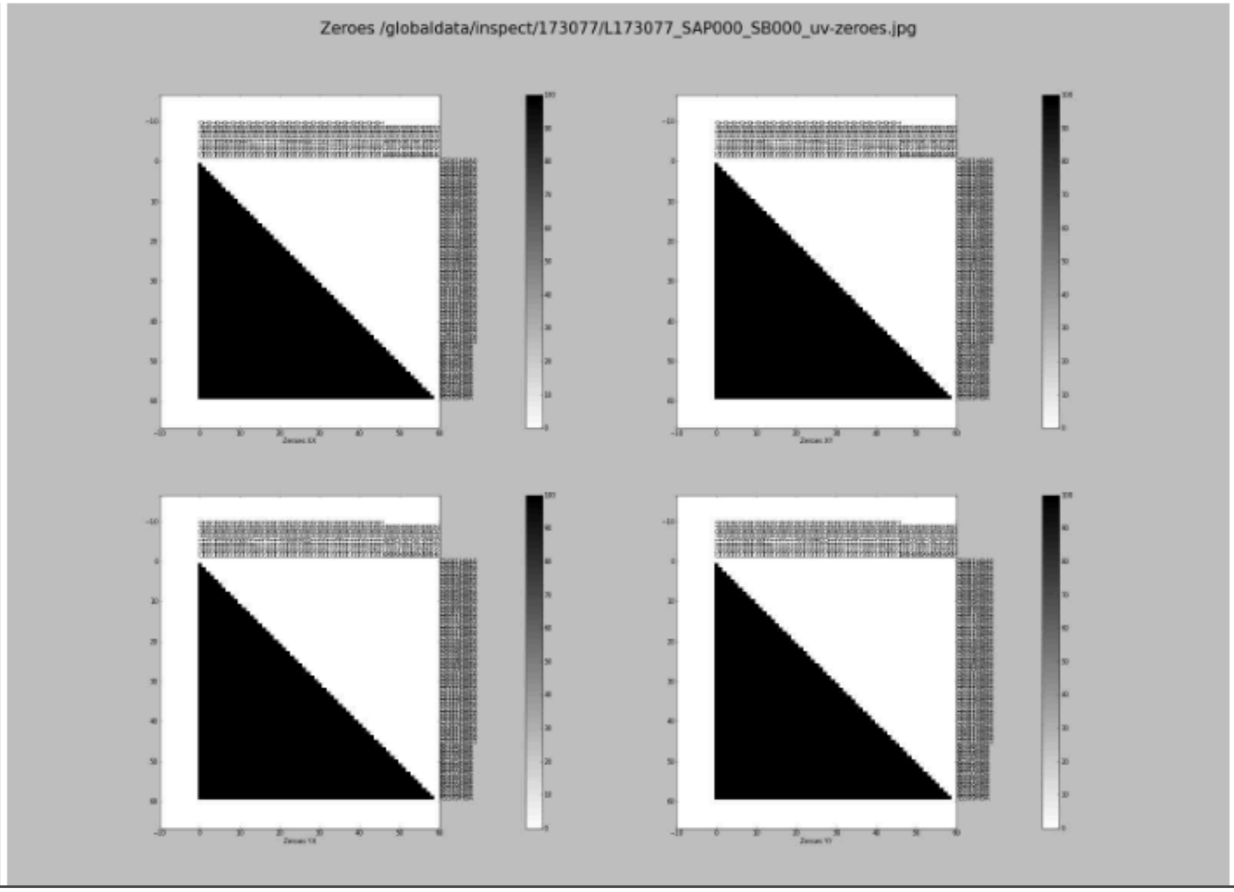
* **HBA** Interleaved short calibrator observations (eg. 2 min) with target field (eg. ~ 30 min), quasi-continuous in HA. Because the tiles at station level are analogically combine and the station is sensitive to a limited area of the sky (about 10 degrees). So if your calibrator is not within that region you can not use the dual beam strategy used for the LBA.

Inspection plots

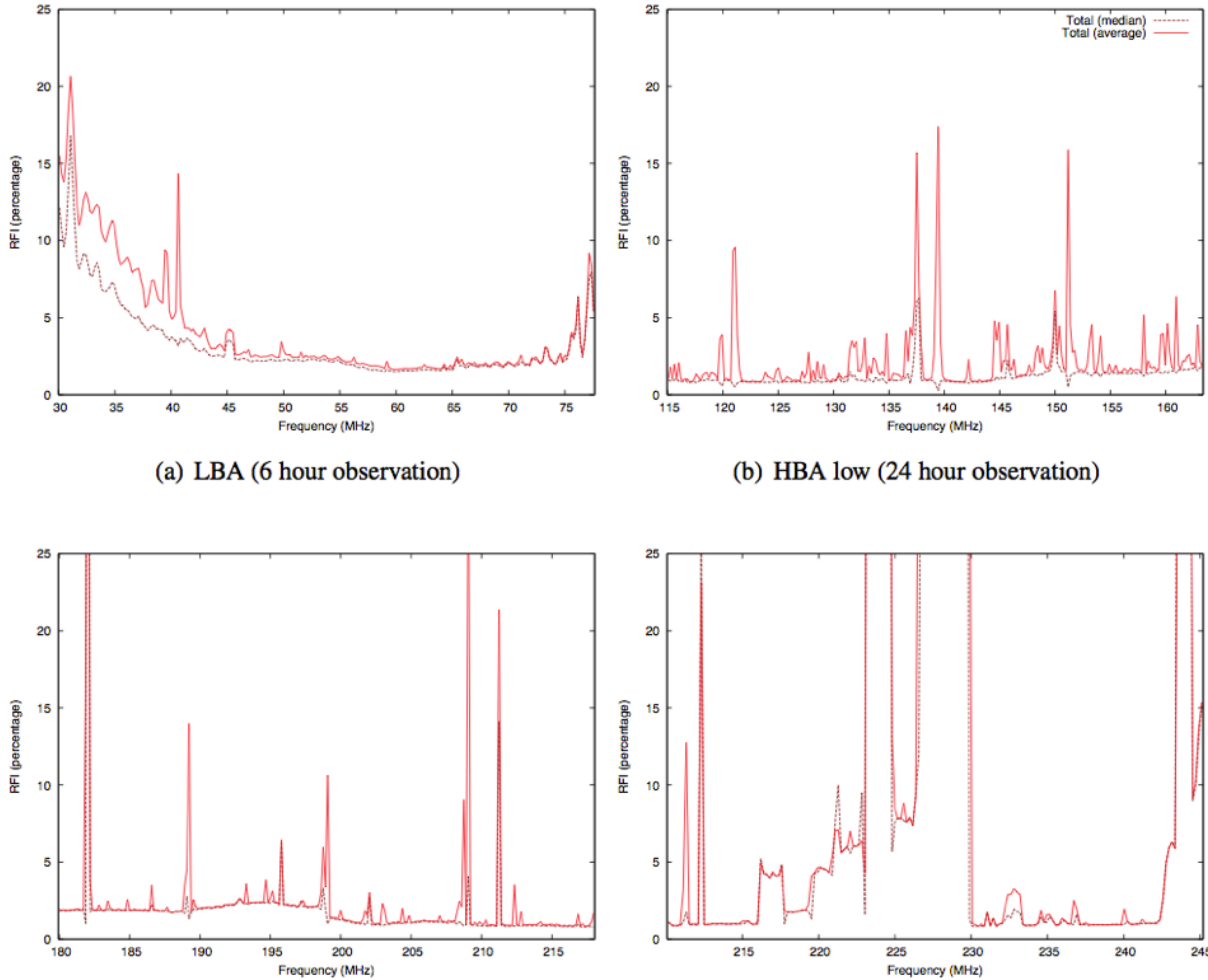








2: Radio Frequency Interference in the LOFAR bands



(a) LBA (6 hour observation)

(b) HBA low (24 hour observation)

(c) HBA mid (1 hour observation)

(d) HBA high (6 hour observation)

NDPPP
(NEW DEFAULT PRE-PROCESSING PIPELINE)

* Flagging (automatic or manual)
* Averaging in time and/or frequency
* Phase shift to another phase center
* Count flags and writing the counts into a table for plotting purposes.
* Combine subbands into a single MeasurementSet
* Demix and subtract A-team sources
* Add stations to form a superstation
* Filter out baselines and/or channels

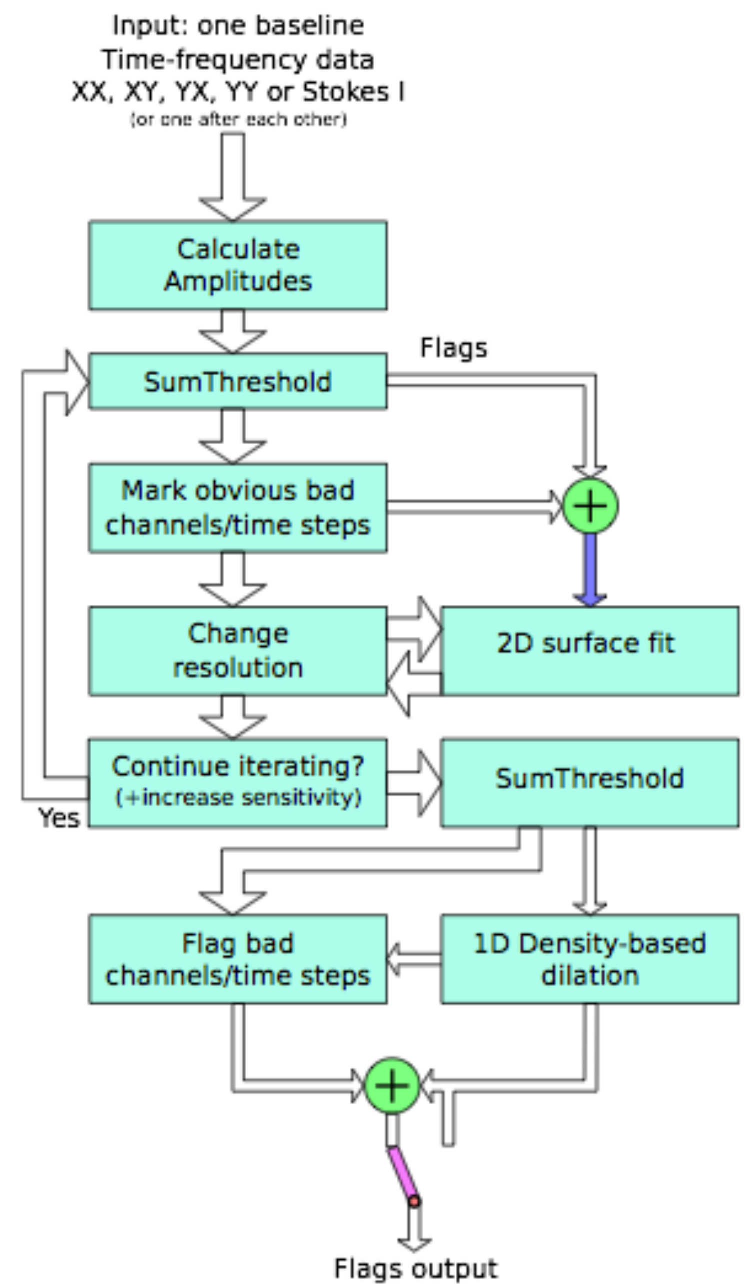
```
msin = in.ms
msin.startchan = nchan/32
#1
msin.nchan = nchan*30/32
msin.autoweight = true
msout = out.ms
steps = [flag,avg]          #2
flag.type = aoflagger        #3
flag.memoryperc = 25
avg.type = average           #4
avg.freqstep = 60
avg.timestep = 5
```

FLAG

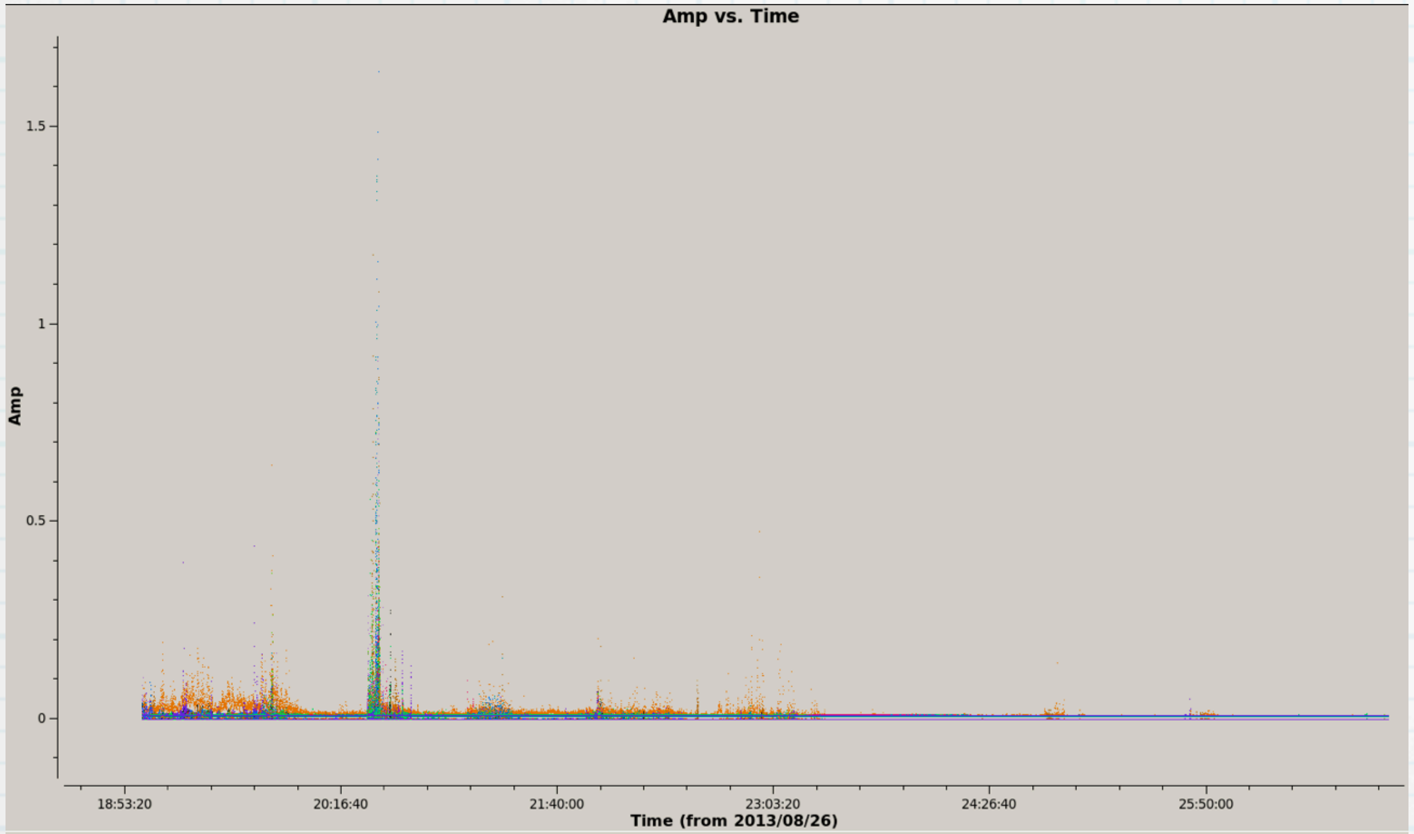
- in the NDPPP module: preflagger, aoflagger
- preflagger: flags autocorrelations: e.g CS001&CS001, or single antennas
- AOflagger: automantic

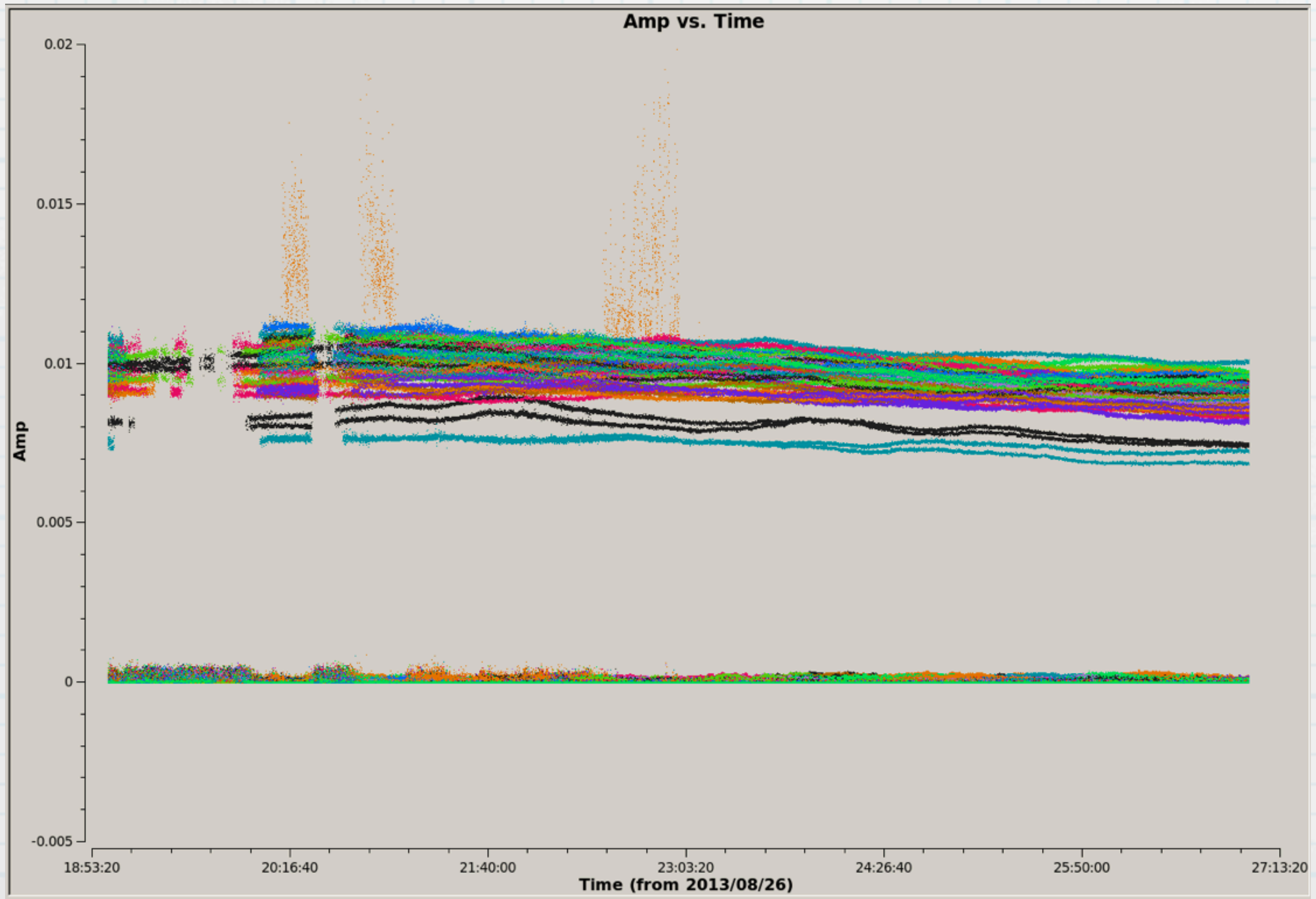
interactions: time, freq, corr....

$$\Omega_d^{n+1}(x) = \begin{cases} 0 & \text{if } \Omega_d^n(x) = 0 \lor \exists i \in \{0 \dots M_n - 1\} : \\ & \sum_{j=0}^{M_n-1} \left| \rho_d \odot \Omega_d^n(x+i-j) \right| > \frac{\chi_{M_n}}{\sum_{j=0}^{M_n-1} \Omega_d^n(x+i-j)} \\ 1 & \text{otherwise,} \end{cases}$$



before & after RFI editing

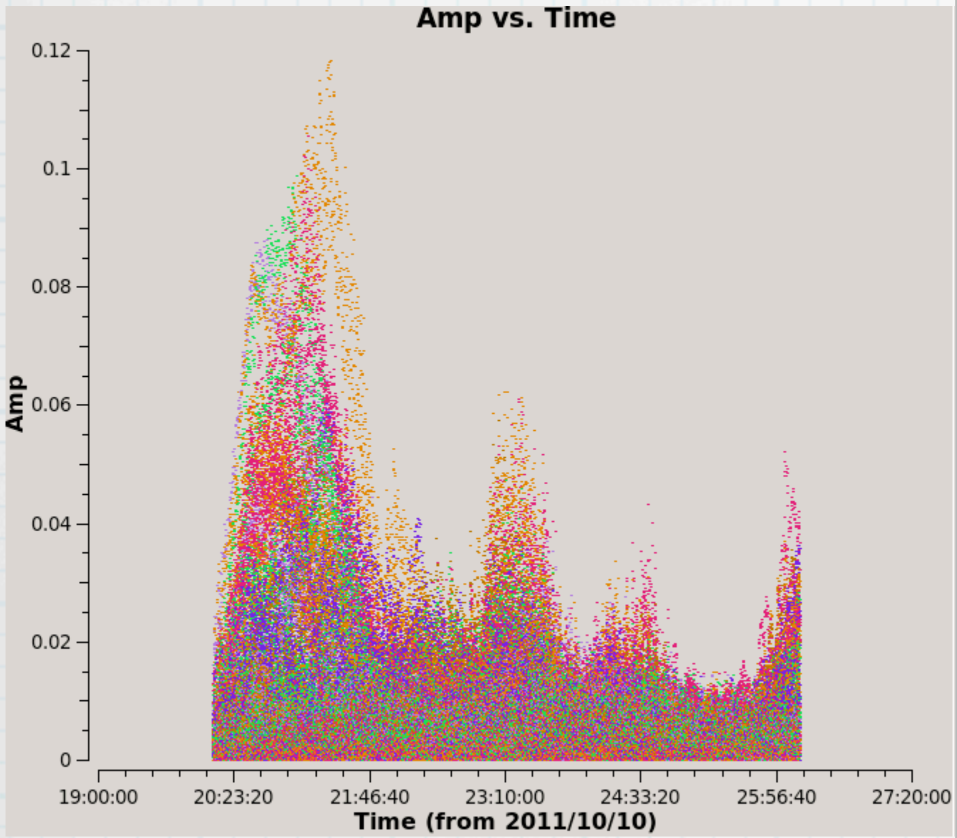


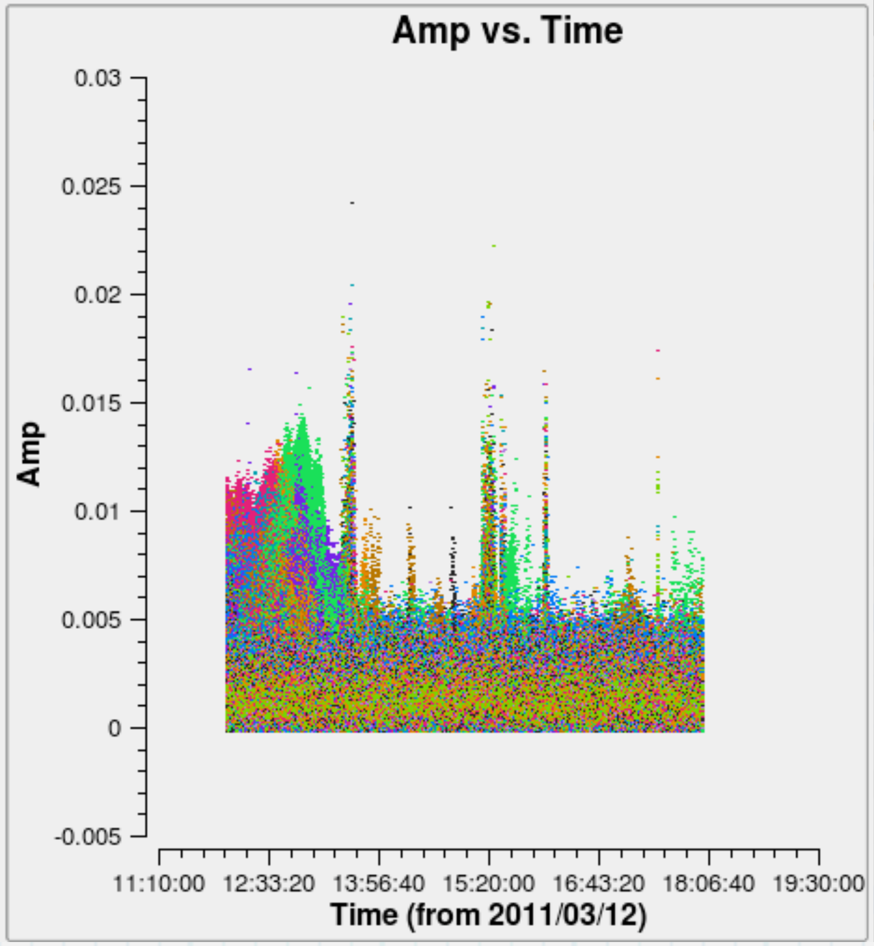


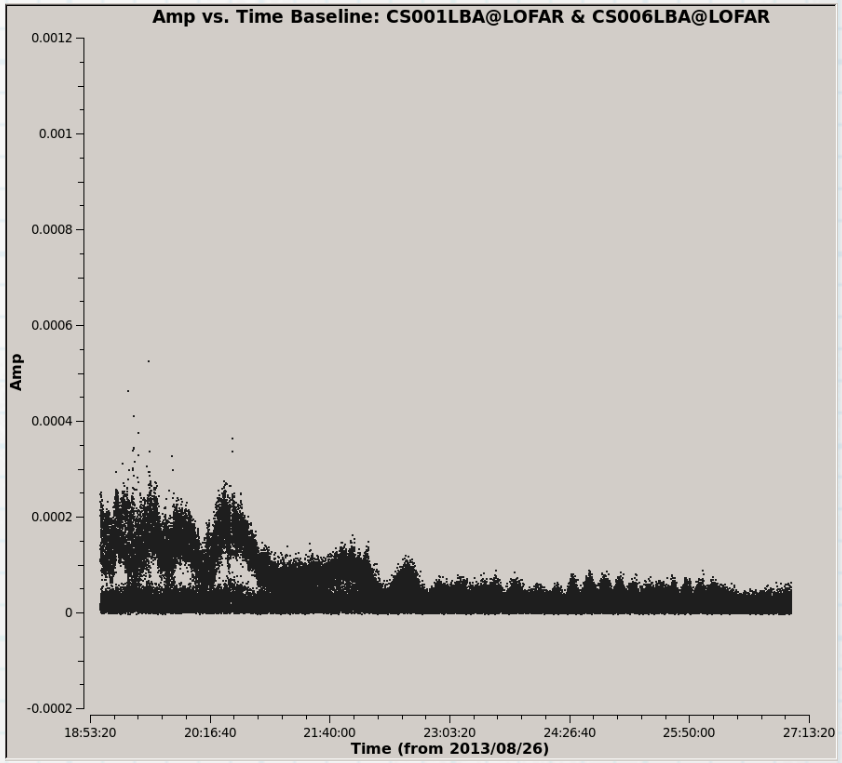
* works well
* optimized for LOFAR HBA EoR observations
* flag RFI that is (1) detectable and (2) not both broadband and continuous
* Not flagged broadband in situ RFI (RFI generated by hardware near the telescope) can not be removed by flagging and might cause calibration failures and/or sidelobes while imaging. Other techniques should be used in such cases.
* LBA strategy is needed since HBA strategy flags A-team signal

3: Data inspection

what are these bumps?

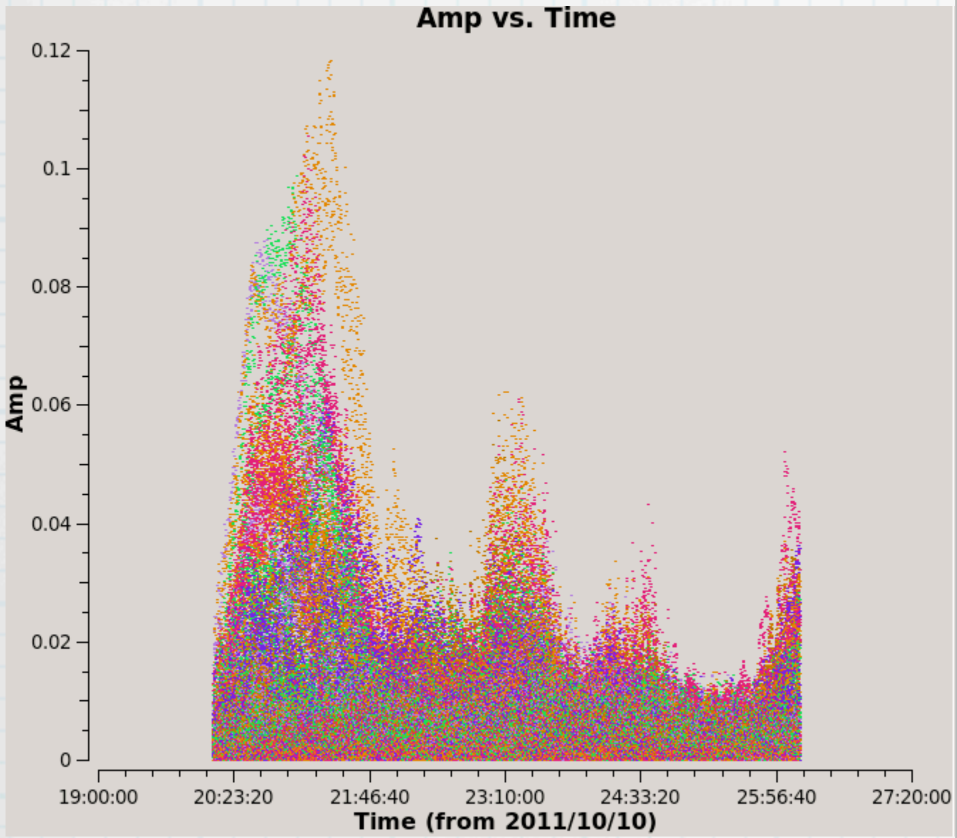


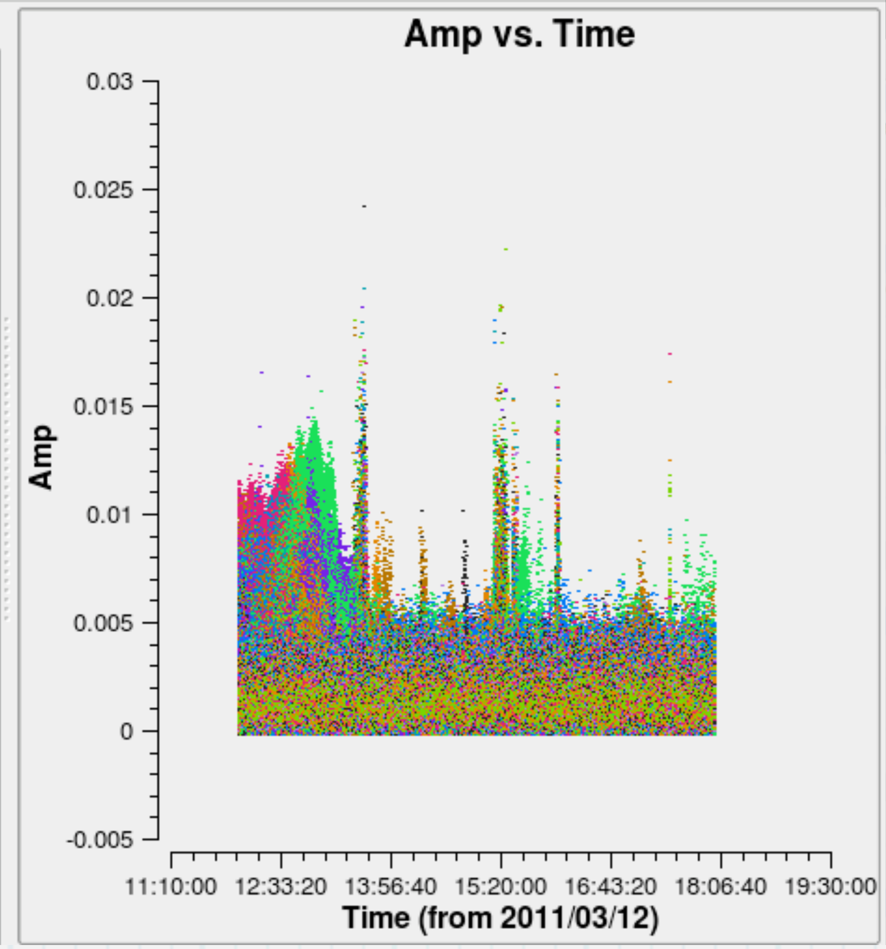


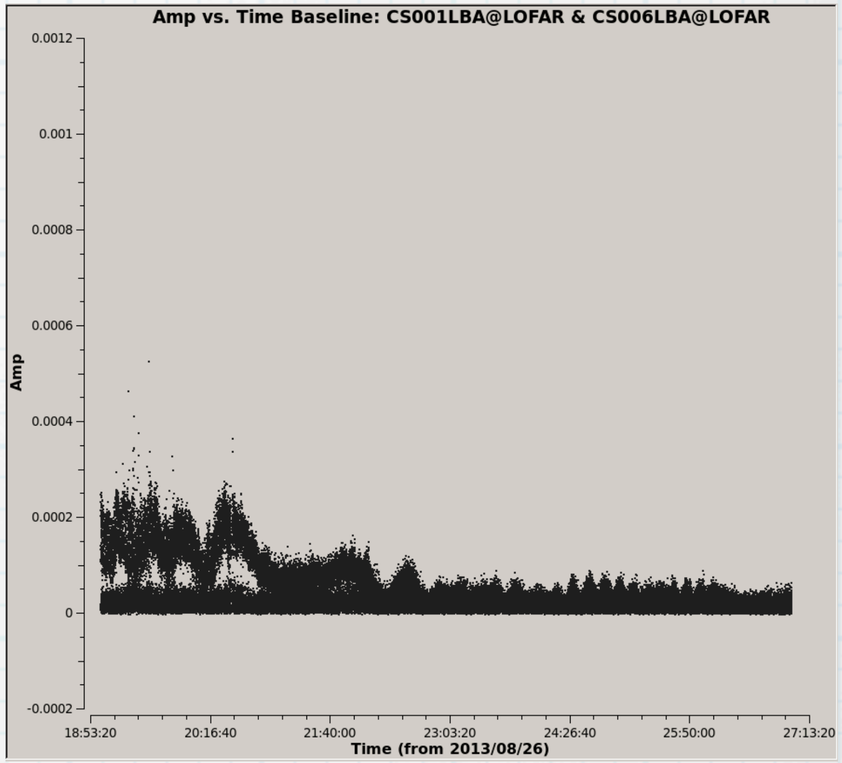


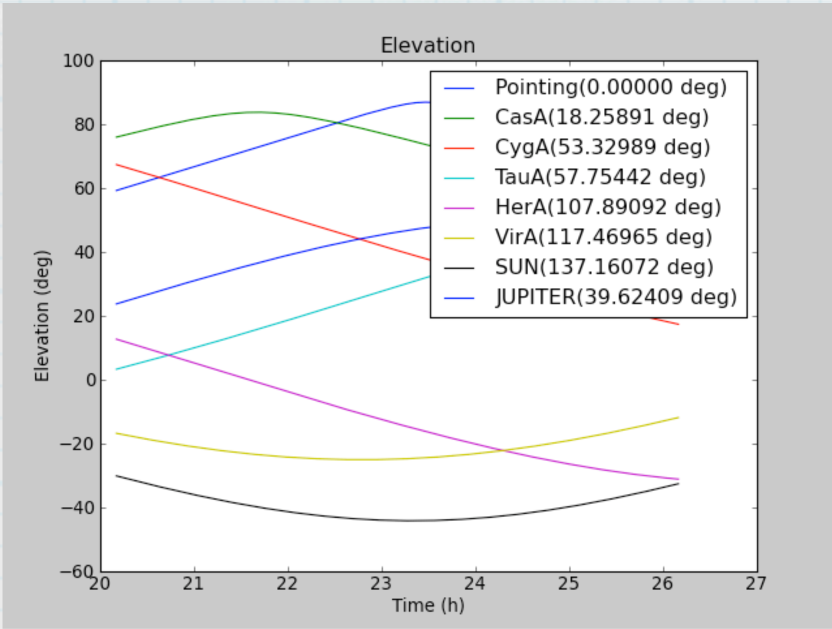
3: Data inspection

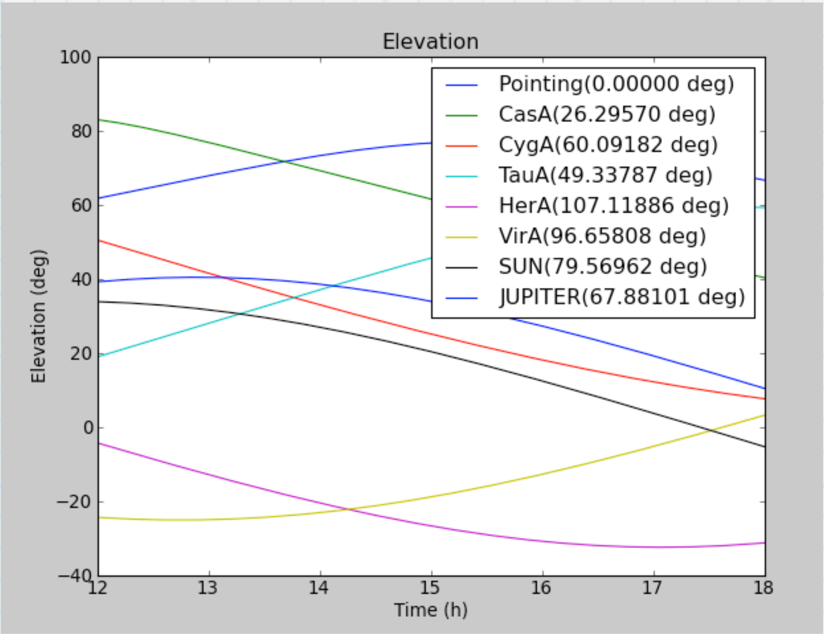
what are these bumps?

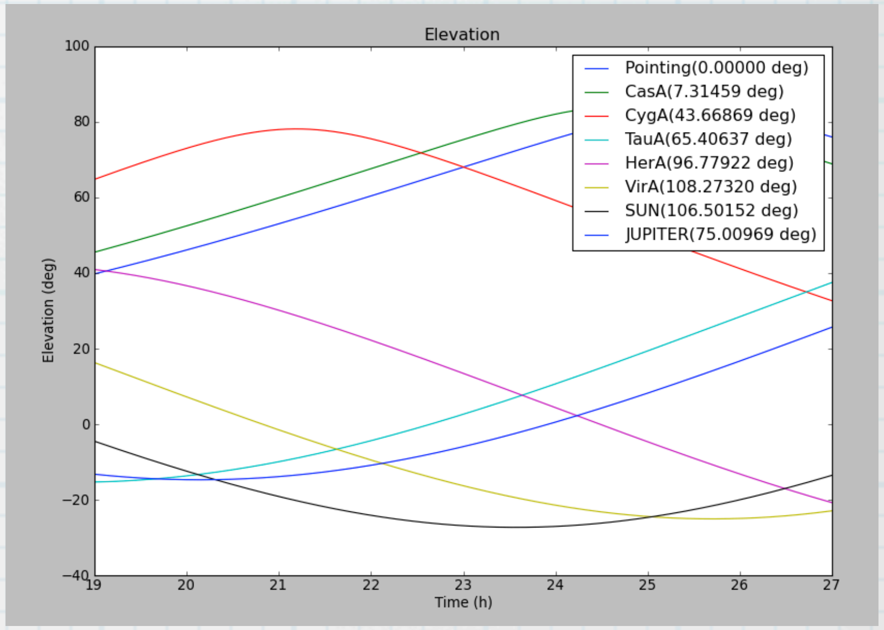






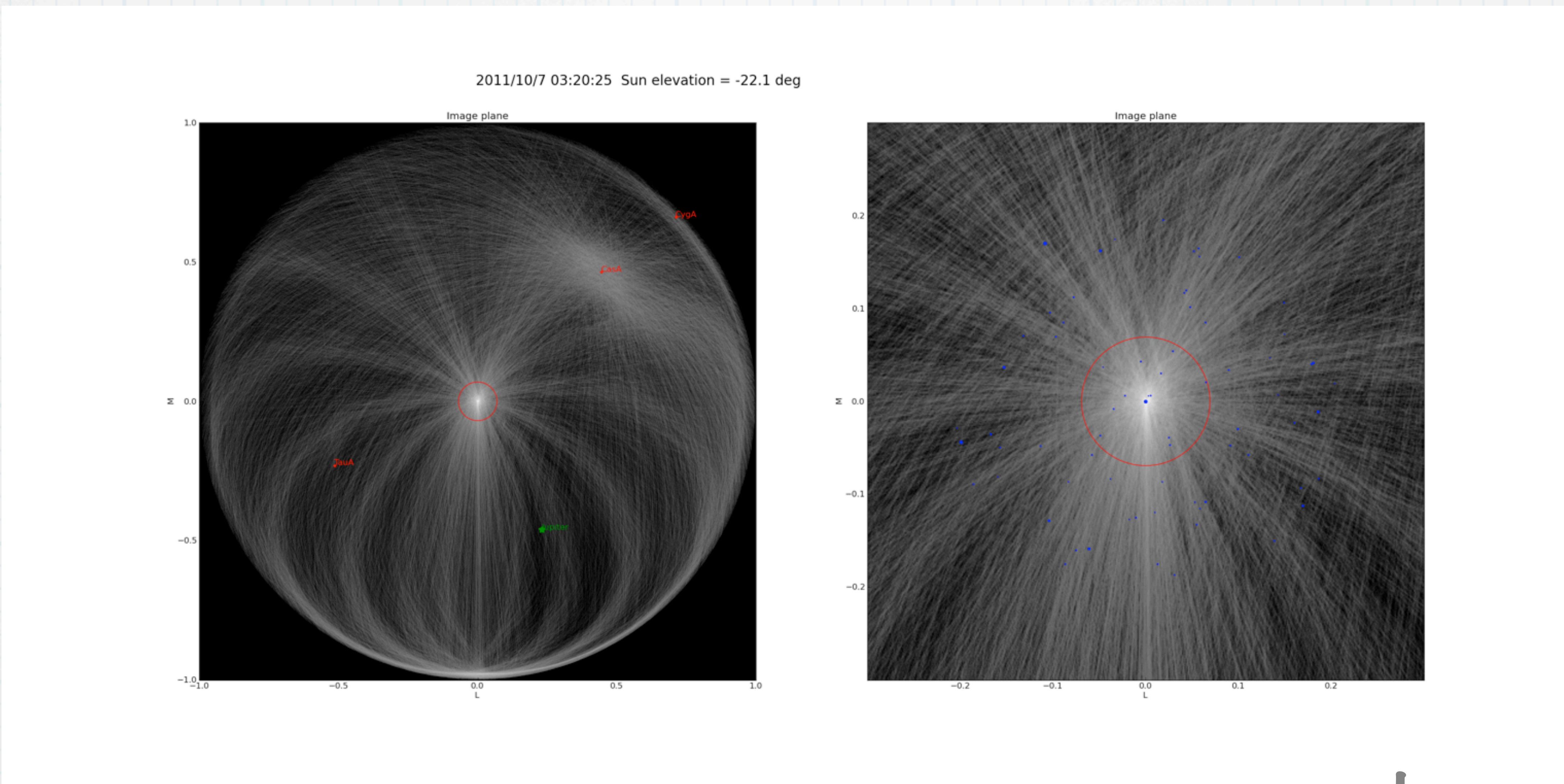






drawer: which sources contribute to the visibilities

* The software automatically converts the fringes seen in the visibilities to locations in the sky
* As the fringes "produced" by each individual baseline are rotating on the sky, each source modulates the visibility, depending on its distance from the phase center (far away sources give a higher fringe rate).



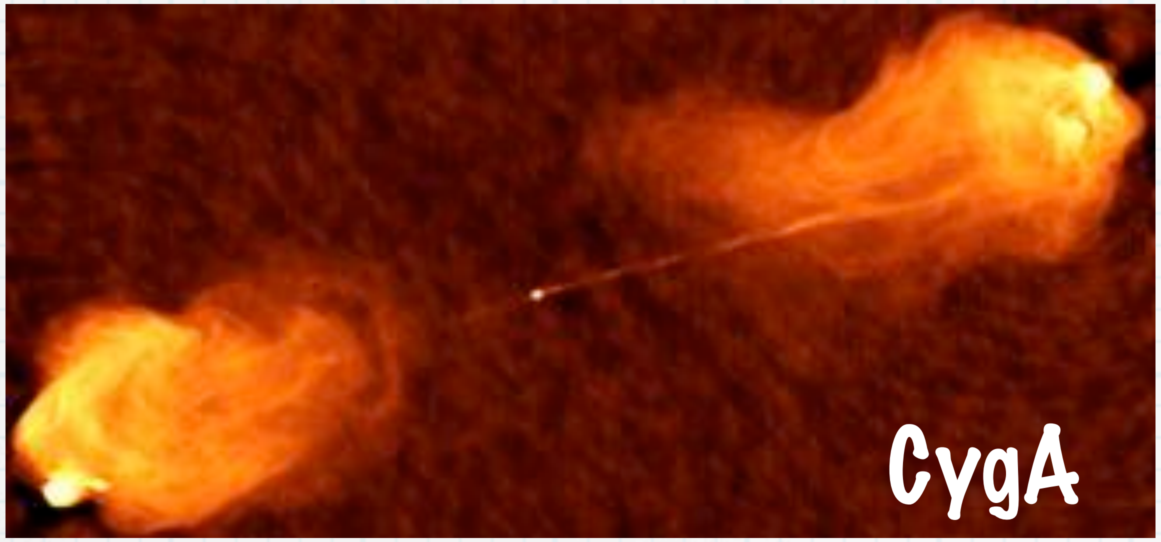
...show movie

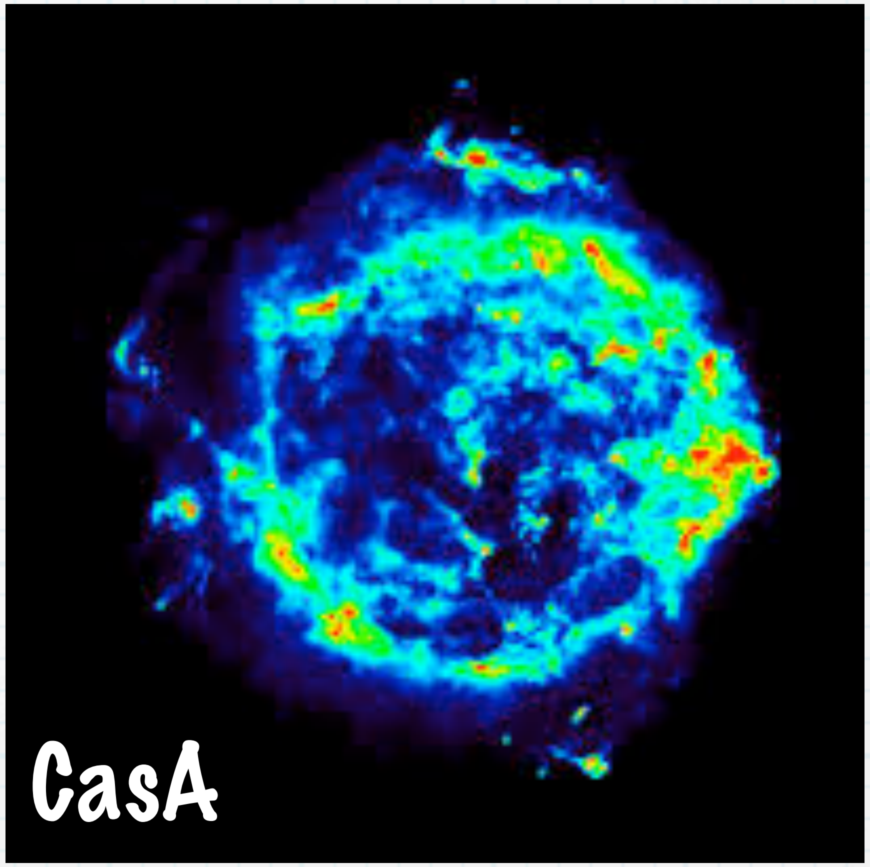
4: Removal of A-team

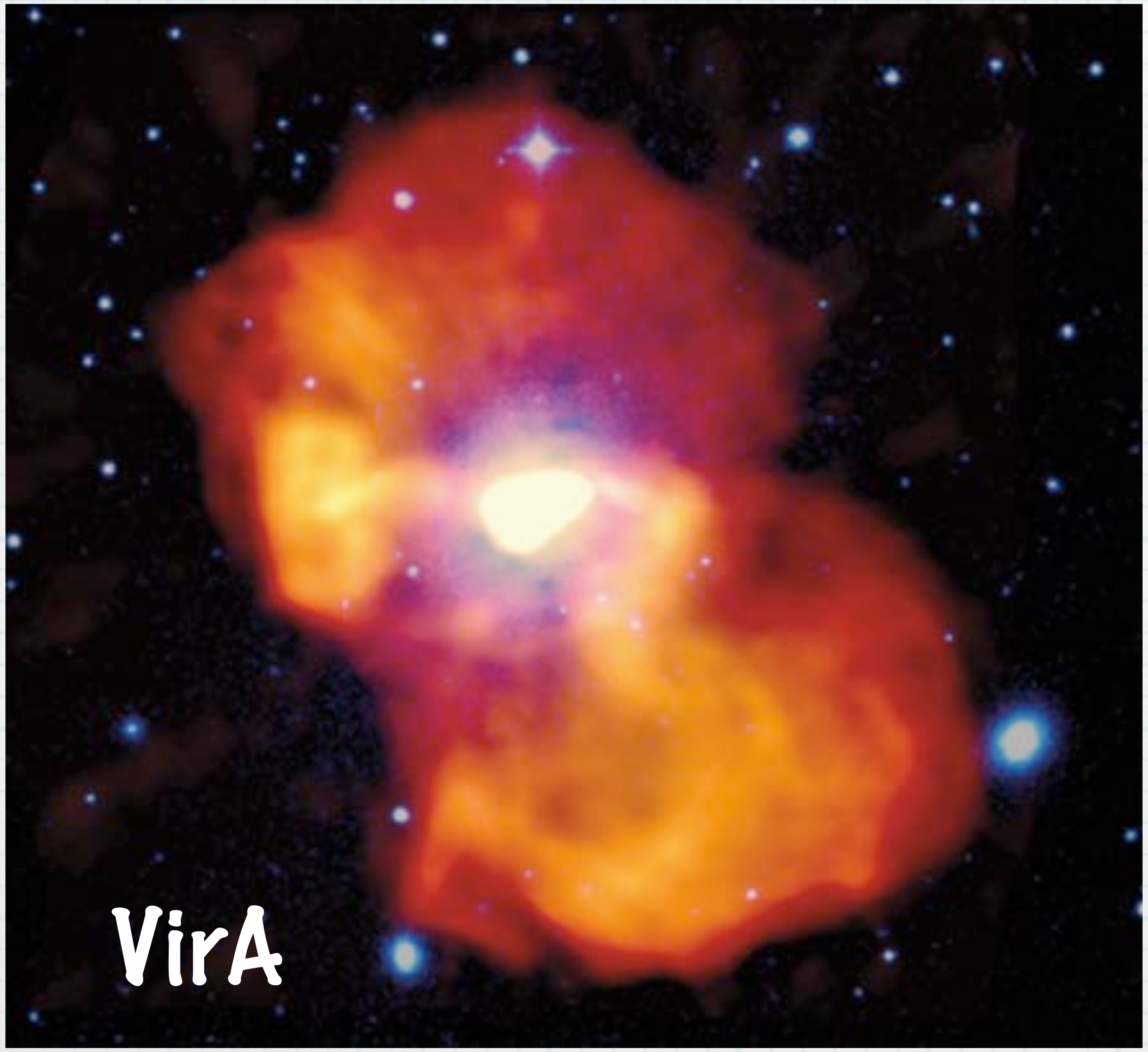
* **The A-Team** is an American action-adventure television series, running from 1983 to 1987

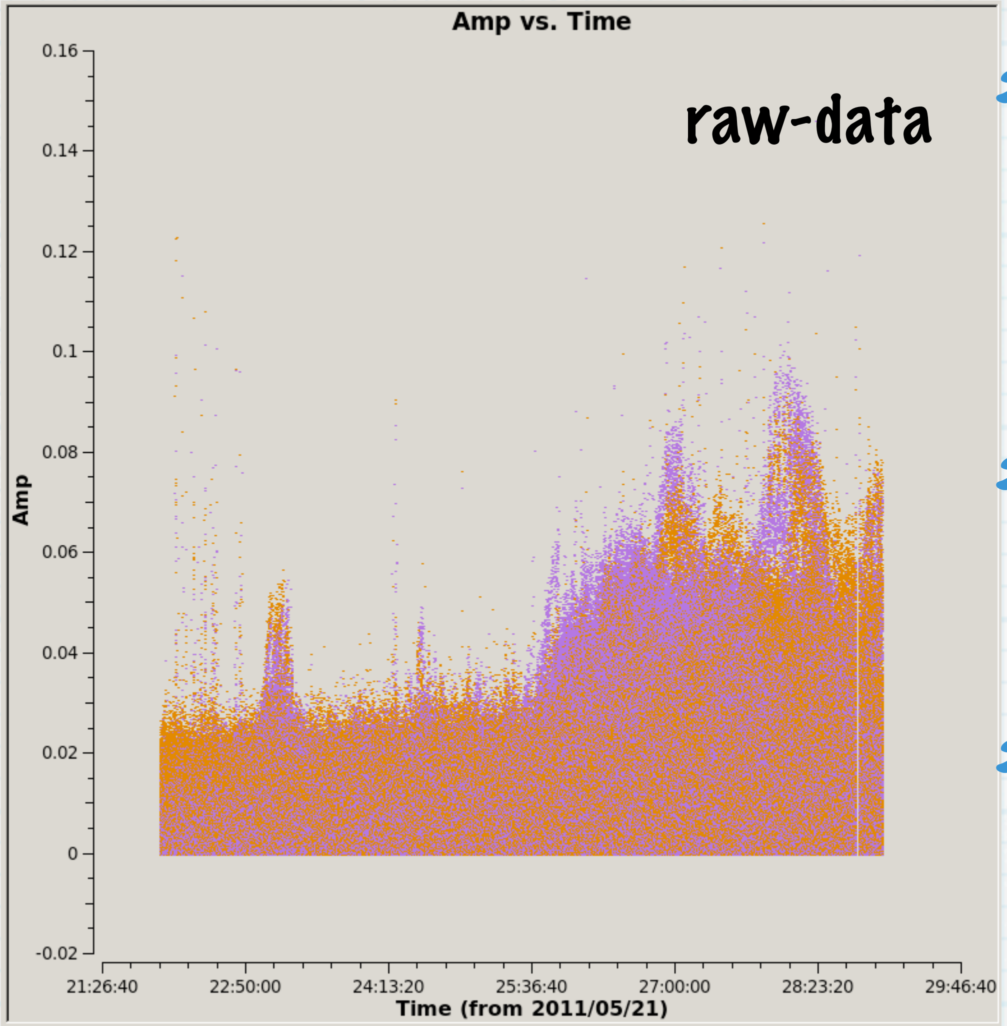
4: Removal of A-team

4: Removal of A-team

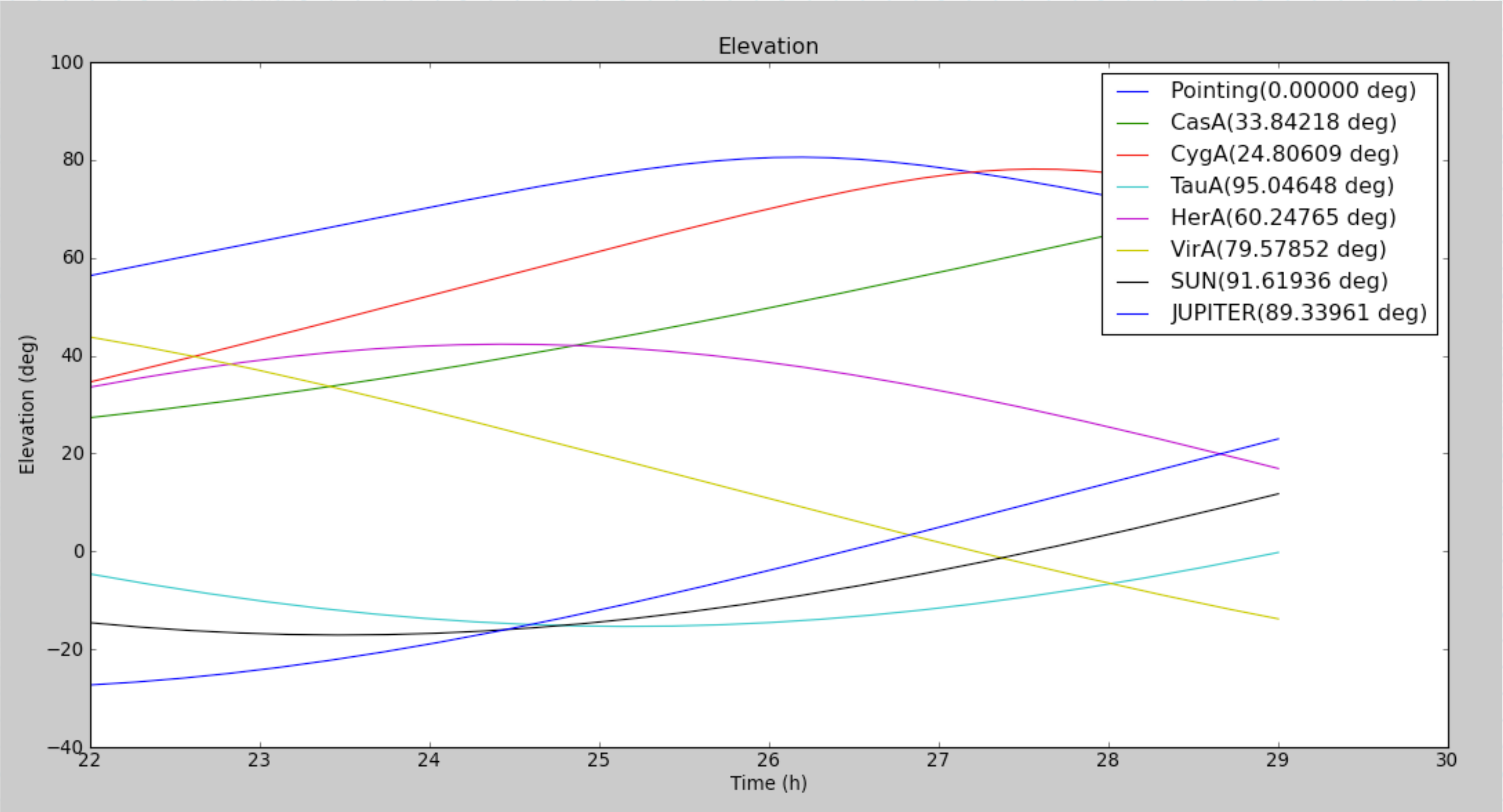






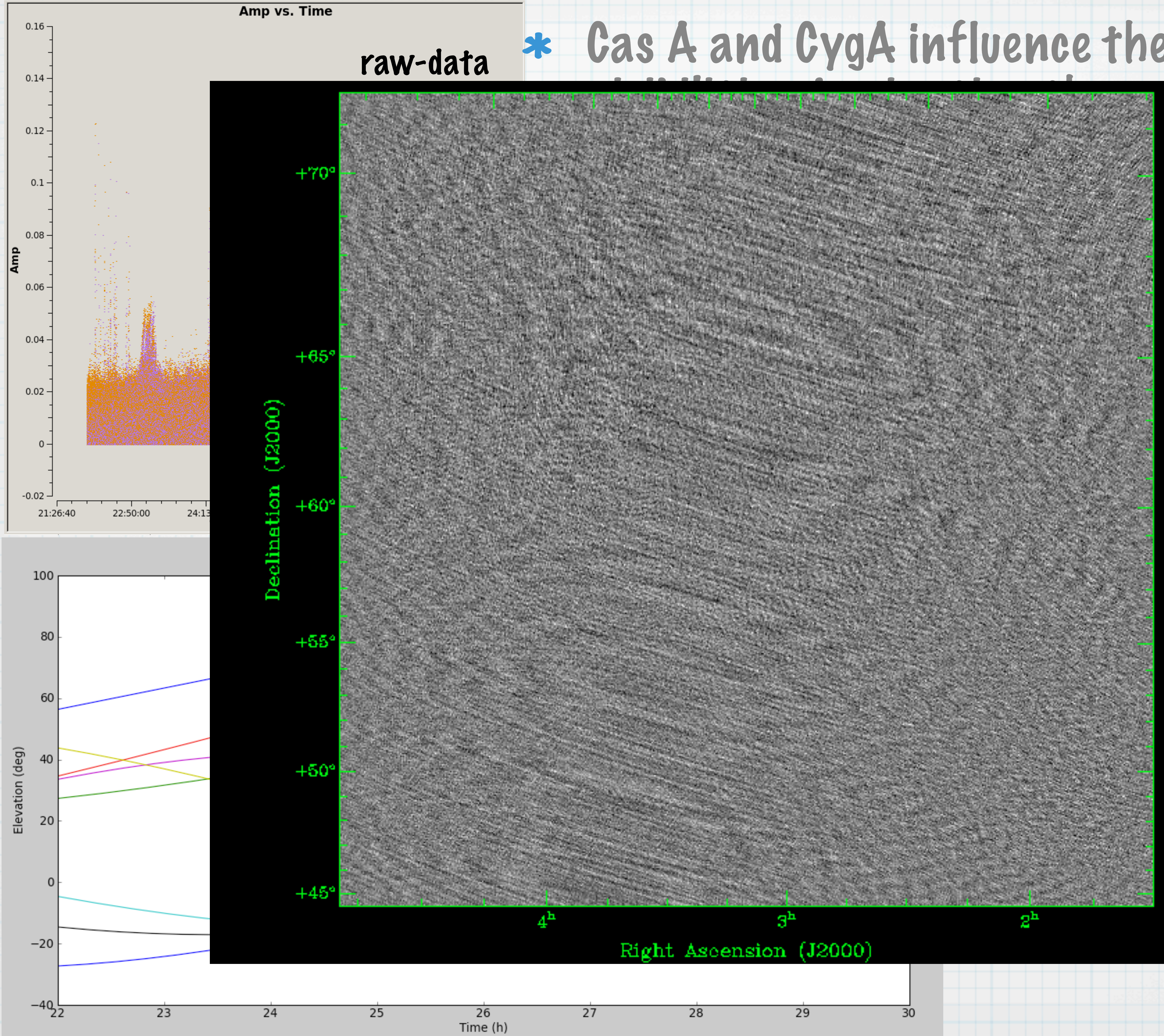


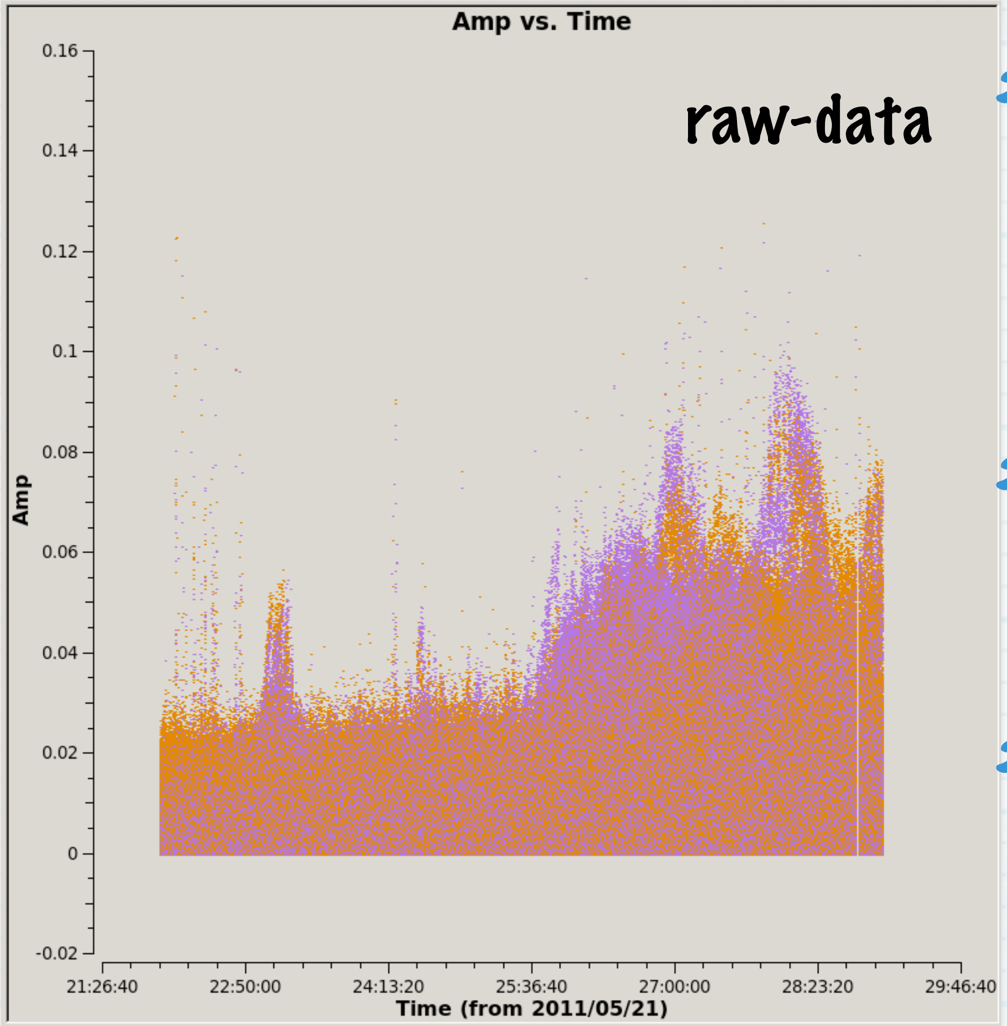
- Cas A and CygA influence the visibilities dominating the second half of the observation
- need to subtract them before calibration. Why?
- 3 options: Demix & 2 types of Direction dependent calibration (BBS and SAGECAL) but...



* Cas A and CygA influence the
…cond half

…f Direction
…nd

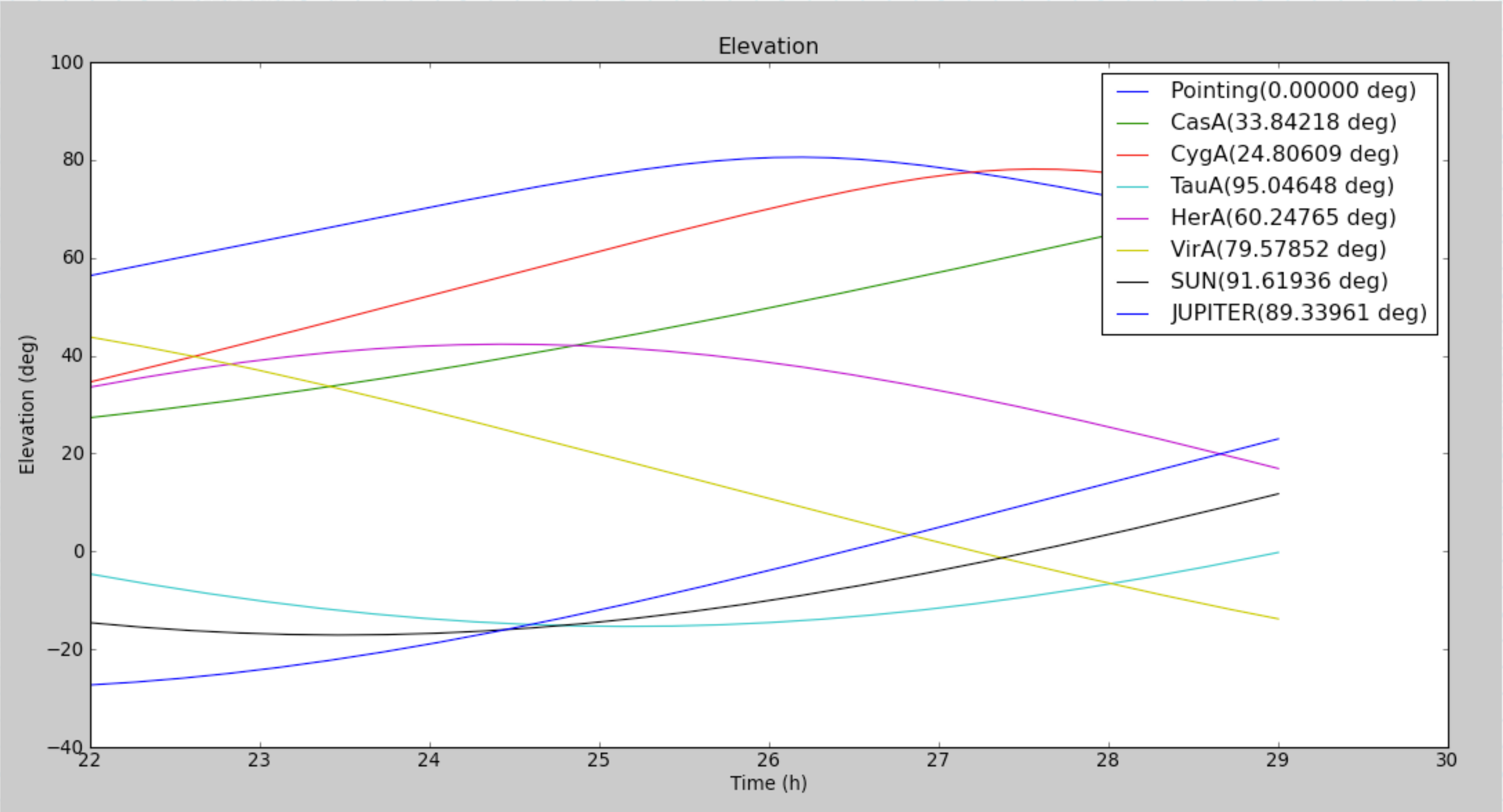




* Cas A and CygA influence the visibilities dominating the second half of the observation

* need to subtract them before calibration. Why?

* 3 options: Demix & 2 types of Direction dependent calibration (BBS and SAGECAL) but...

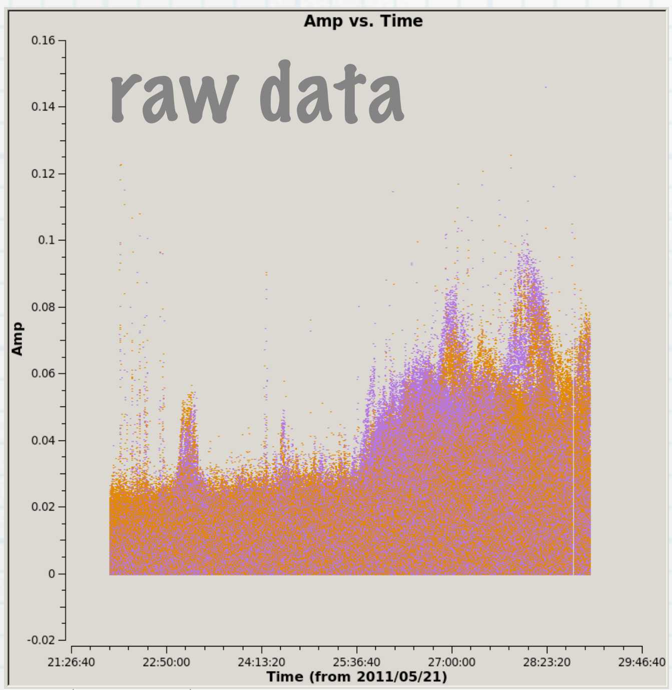


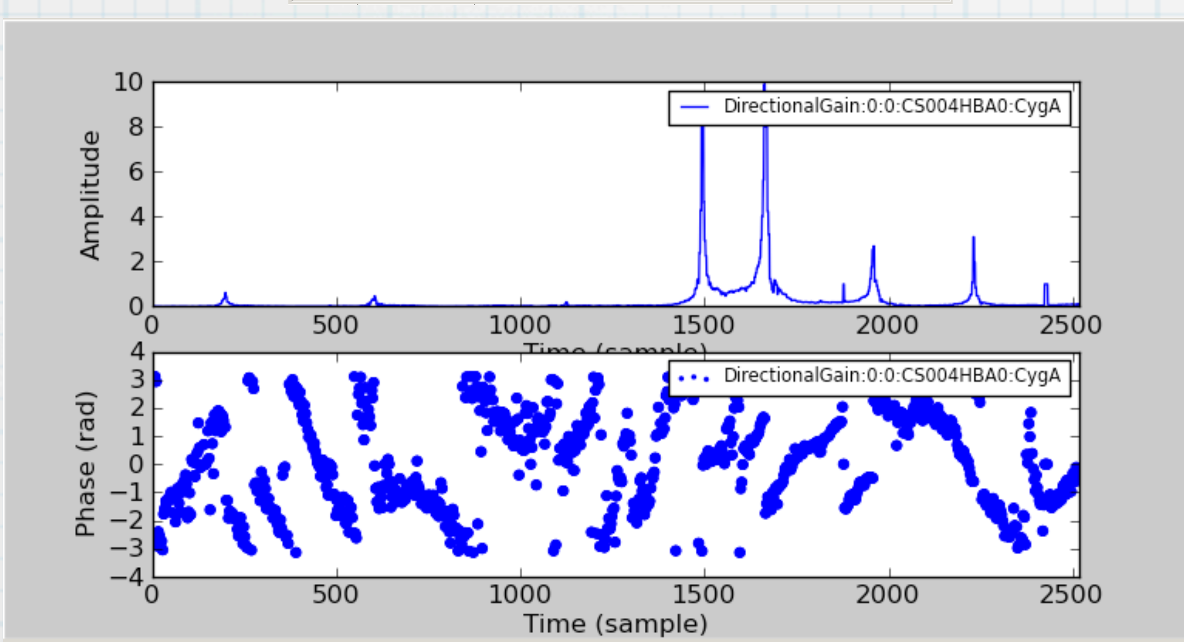
Direction Dependent Calibration

* After the drawer we know which sources contribute to the visibilities so if we create a model of the full field of view we can address the problem
* In principle the contribution of the A-team sources can be removed by a directional gain calibration and subtract step both BBS or SAGECAL.
* The high time and frequency resolution needed to prevent smearing makes DDC technique computationally expensive
* to prove that is is possible I did it for a averaged dataset

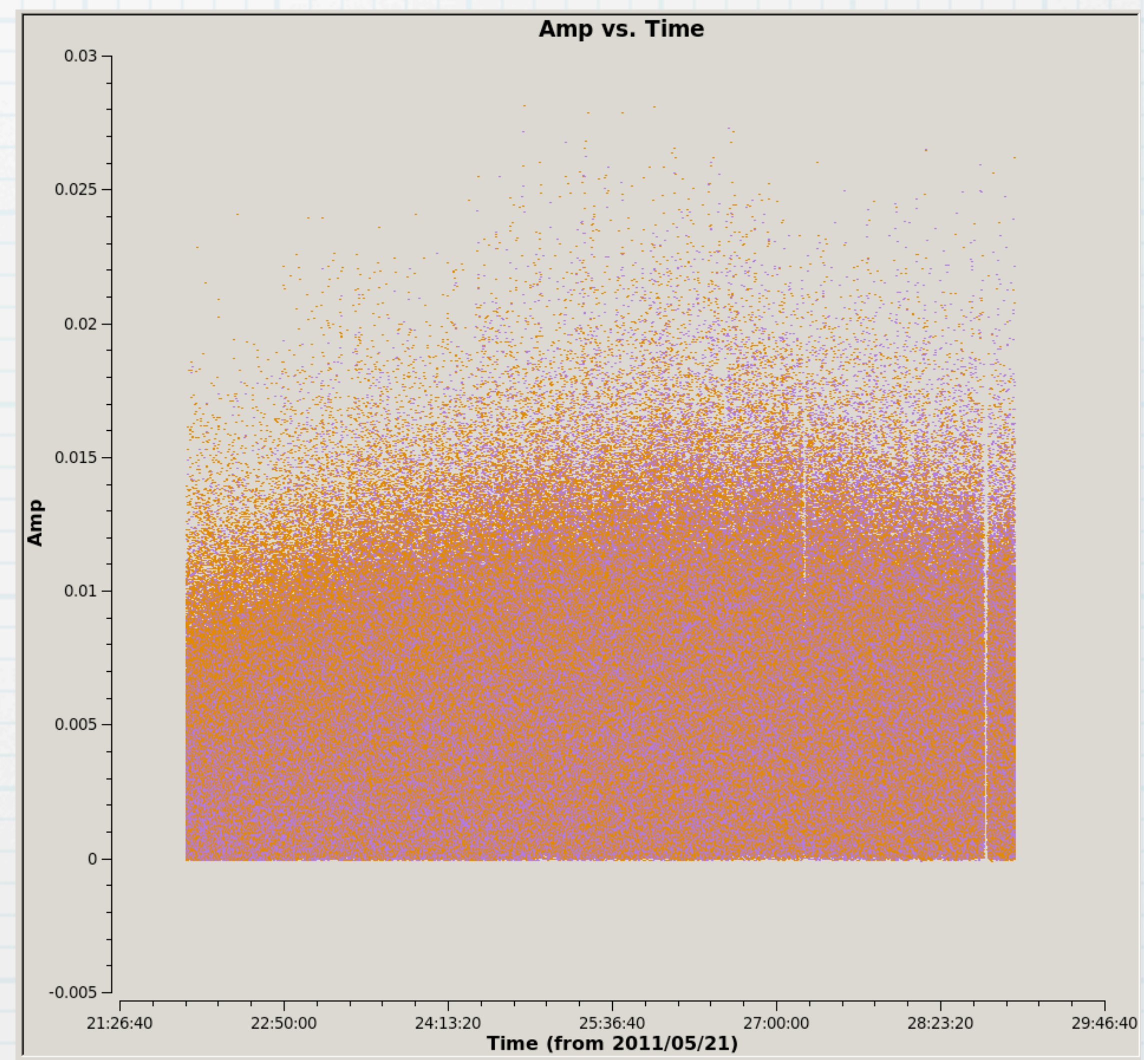
BBS 2 step

* Directional gain on CasA and CygA
* subtract & correct towards the target



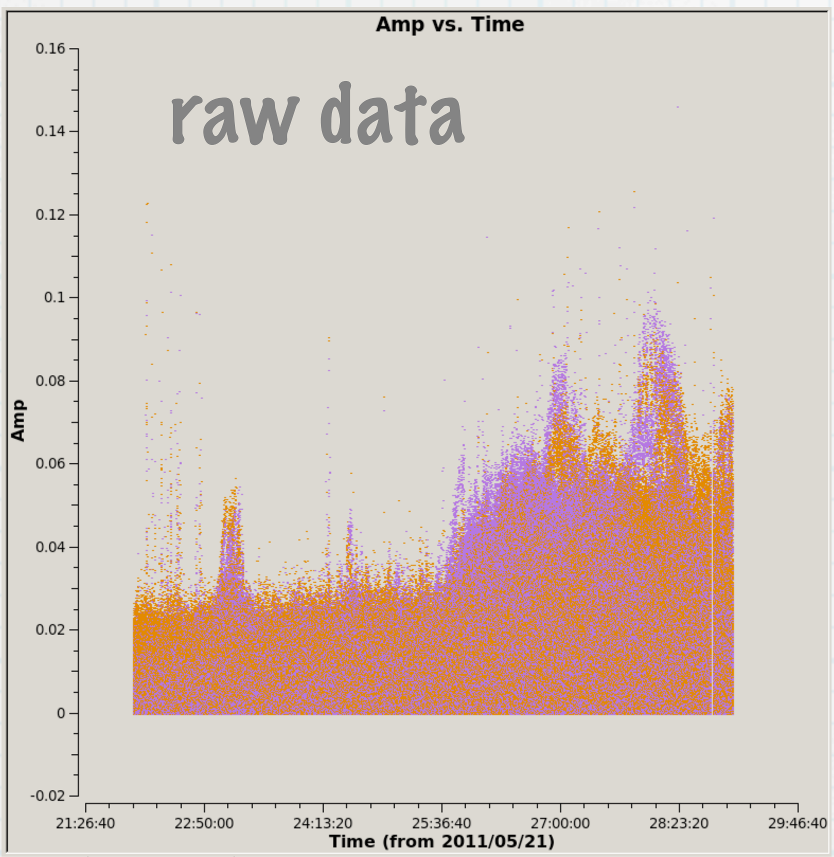


after CasA&CygA subraction with BBS

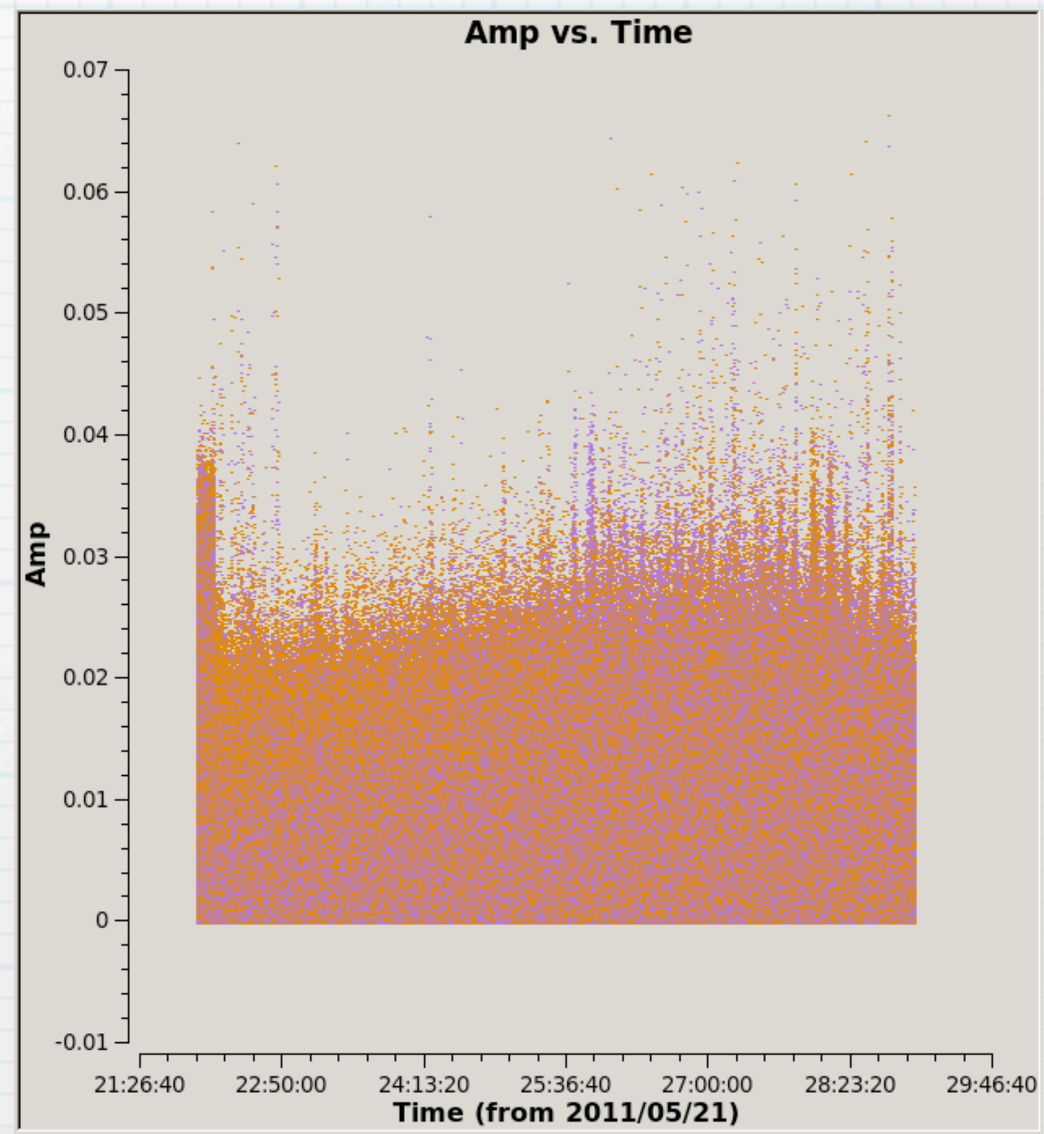


SAGECAL 1 step

* is typically used for selfcalibration on pre-calibrated data, is still under test for uncalibrated data.
* highly performing in terms of computing time



after CasA&CygA subraction with SAGECAL



Demix

van der Tol, S. ., Jes, B. D., & van der Veen, A.-J. . 2007, IEEE Transactions on Signal Processing, 55, 4497

* The observed data is the sum of the contribution of the different fields (target and other sources)
* ignore the other sources and phase shift and average. The contributions of the other sources are attenuated, because they add incoherently.
* Compares data to a model using the mixing matrix. Demixing works per baseline. Each station gain is estimated using the baselines to all other stations. Each station participates in short and long baselines.
* Subtract predicted visibilities from averaged visibilities.

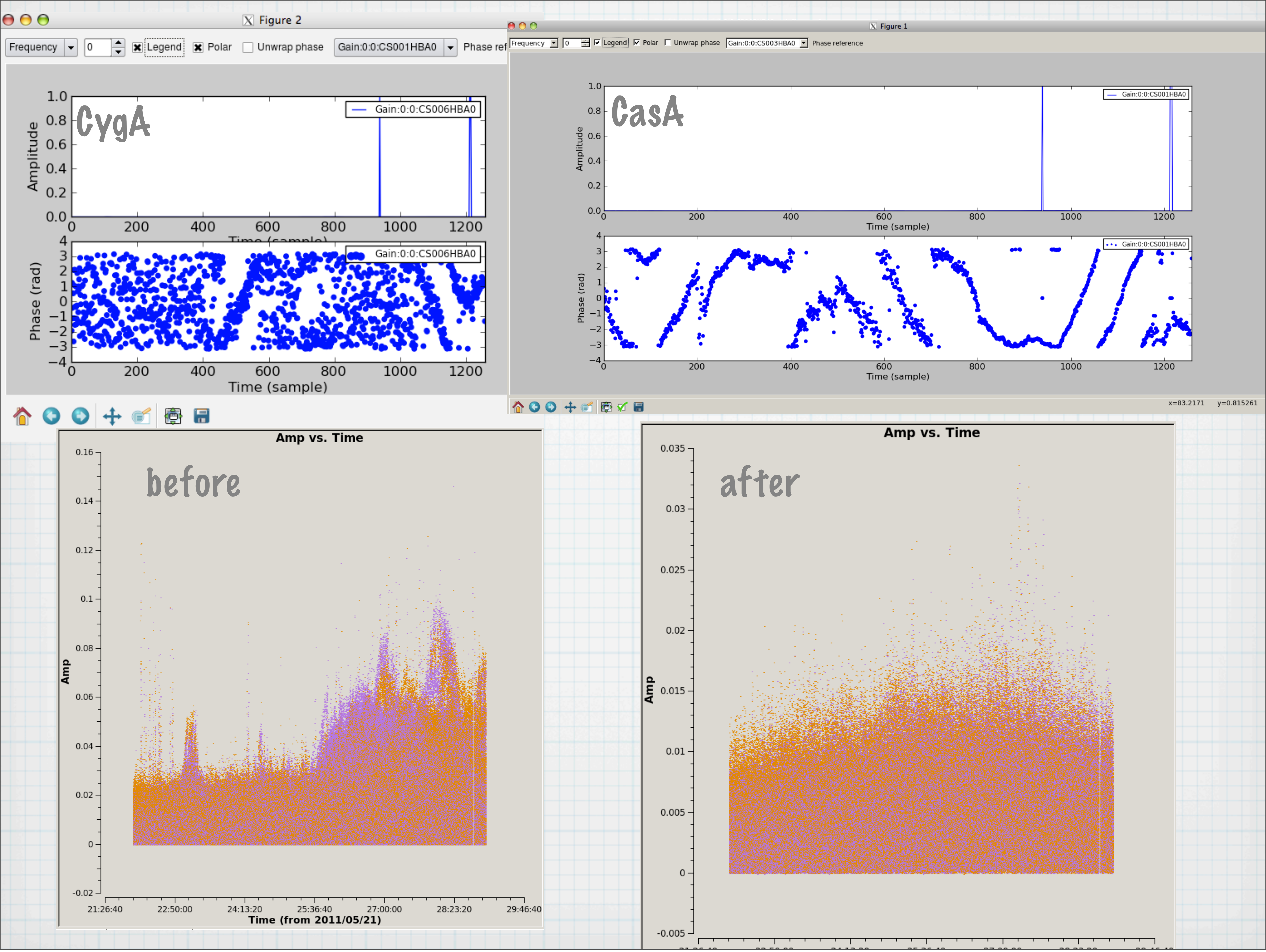


Figure 2
Frequency
Legend
Polar
Unwrap phase
Gain:0:0:CS001HBA0
Phase ref
CygA
Amplitude
Phase (rad)
Time (sample)
Gain:0:0:CS006HBA0
Figure 1
Phase reference
Gain:0:0:CS003HBA0
CasA
Gain:0:0:CS001HBA0
x=83.2171 y=0.815261
Amp vs. Time
before
Amp
Time (from 2011/05/21)
21:26:40
22:50:00
24:13:20
25:36:40
27:00:00
28:23:20
29:46:40
Amp vs. Time
after
Amp

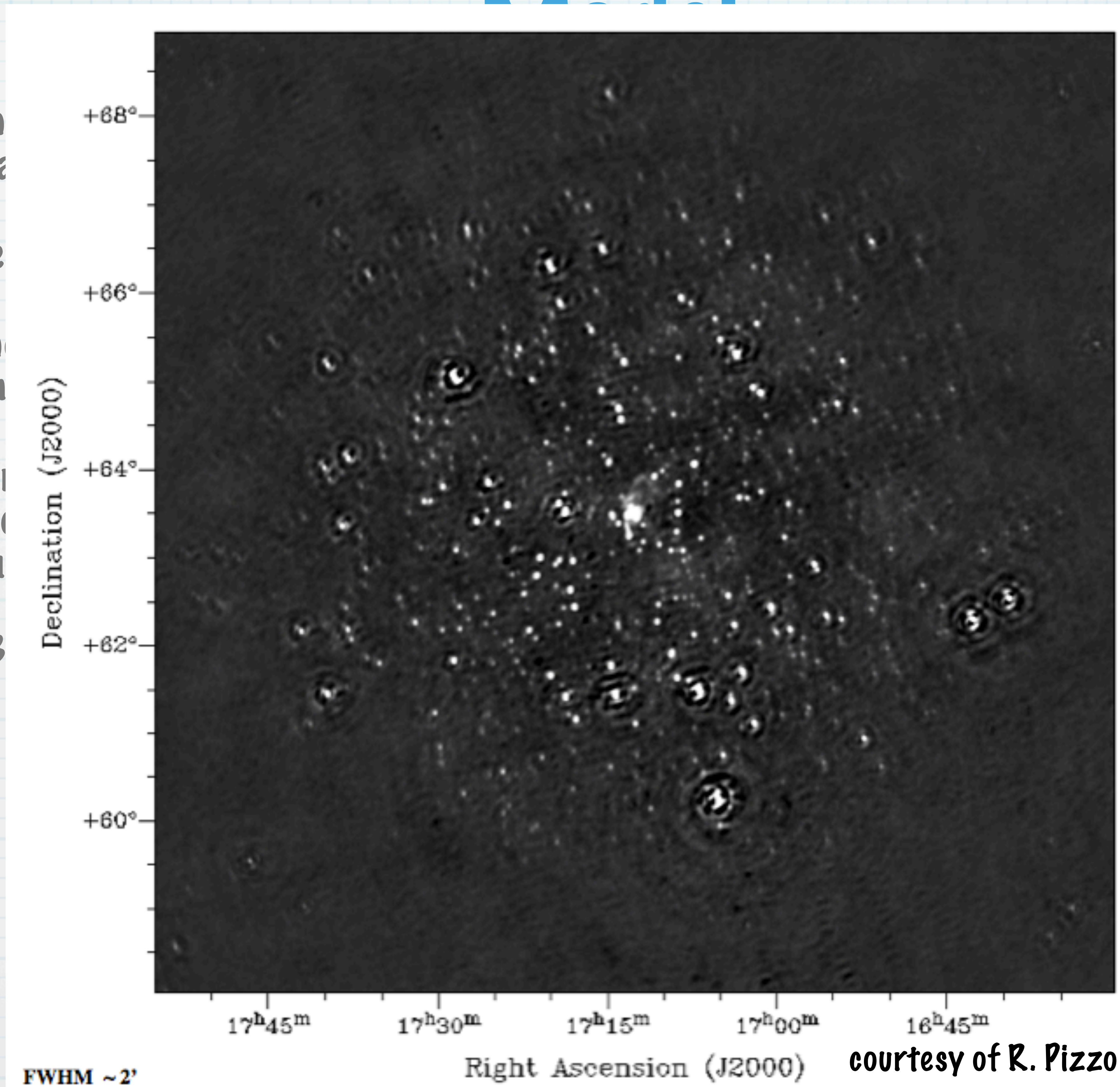
5: Calibration. Model

* The model should represent as much as possible the data, LOFAR sensitive to big areas so make sure you include all the fields of view and further.
* Get a good skymodel: CC components, WENSS, VLSS, pyBDSM, shapelets
* the quality of the calibration depends by match between the RESOLUTION of the MODEL and the BASELINE LENGHT!
* for complex sources CC models give best results but it slows down the process. You can use also pyBDSM and shapelets. For complex sources possible check --> to simulate and image your model
* BBS correct in one direction

5: Calibration.

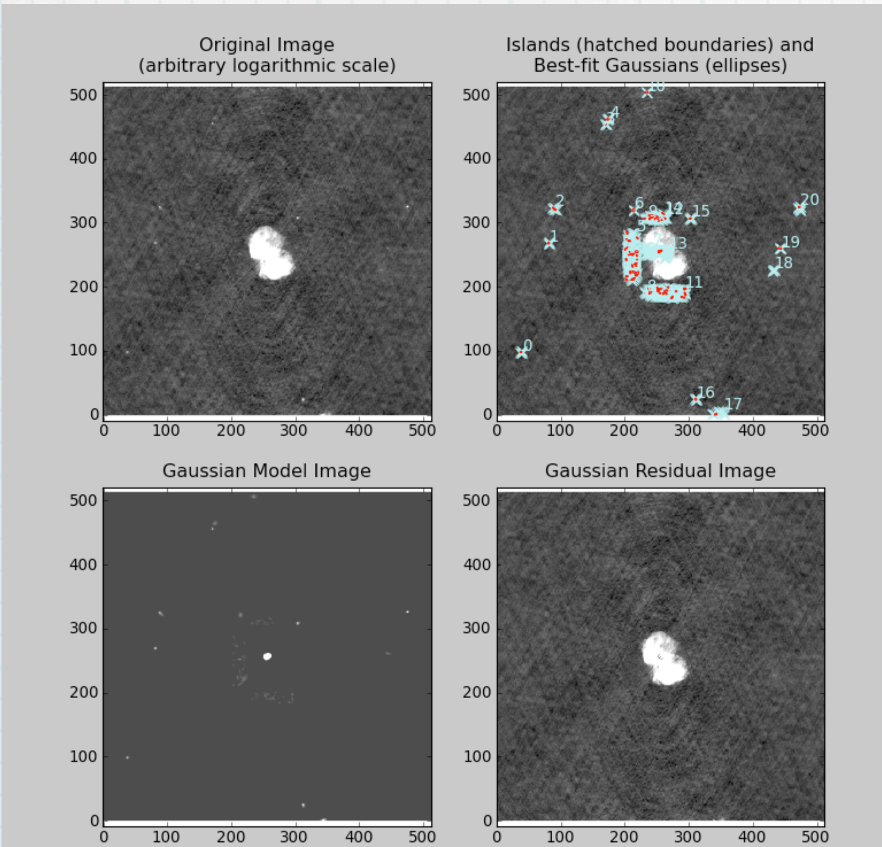
Model

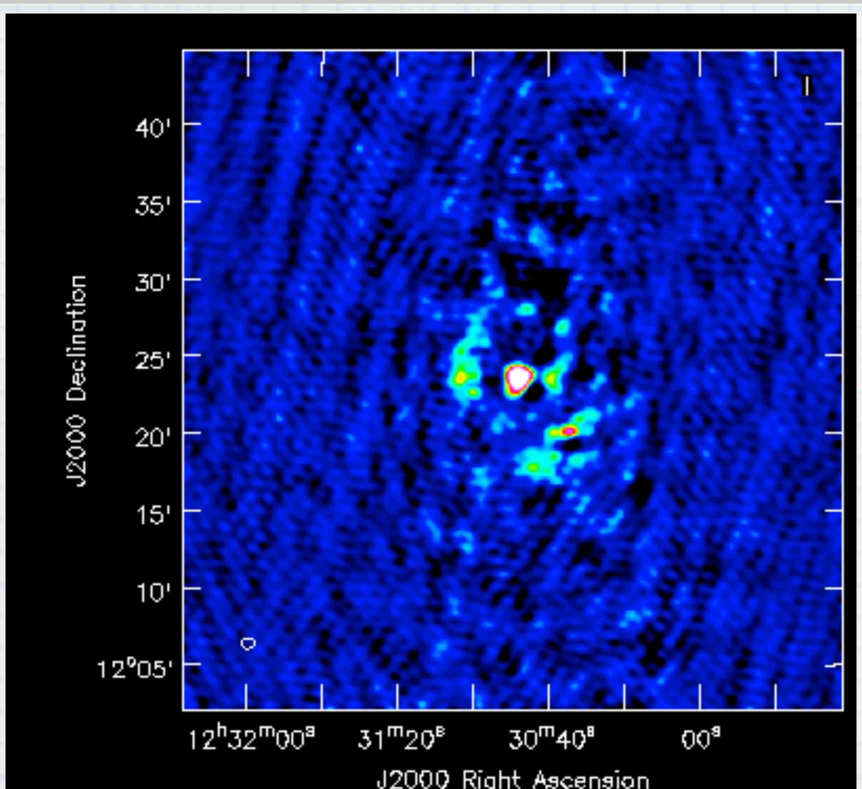
* Th… to big areas so ma…
* Ge…
* th… l of the MODEL an…
* for… cess. You can use… simulate and im…
* BB…



courtesy of R. Pizzo

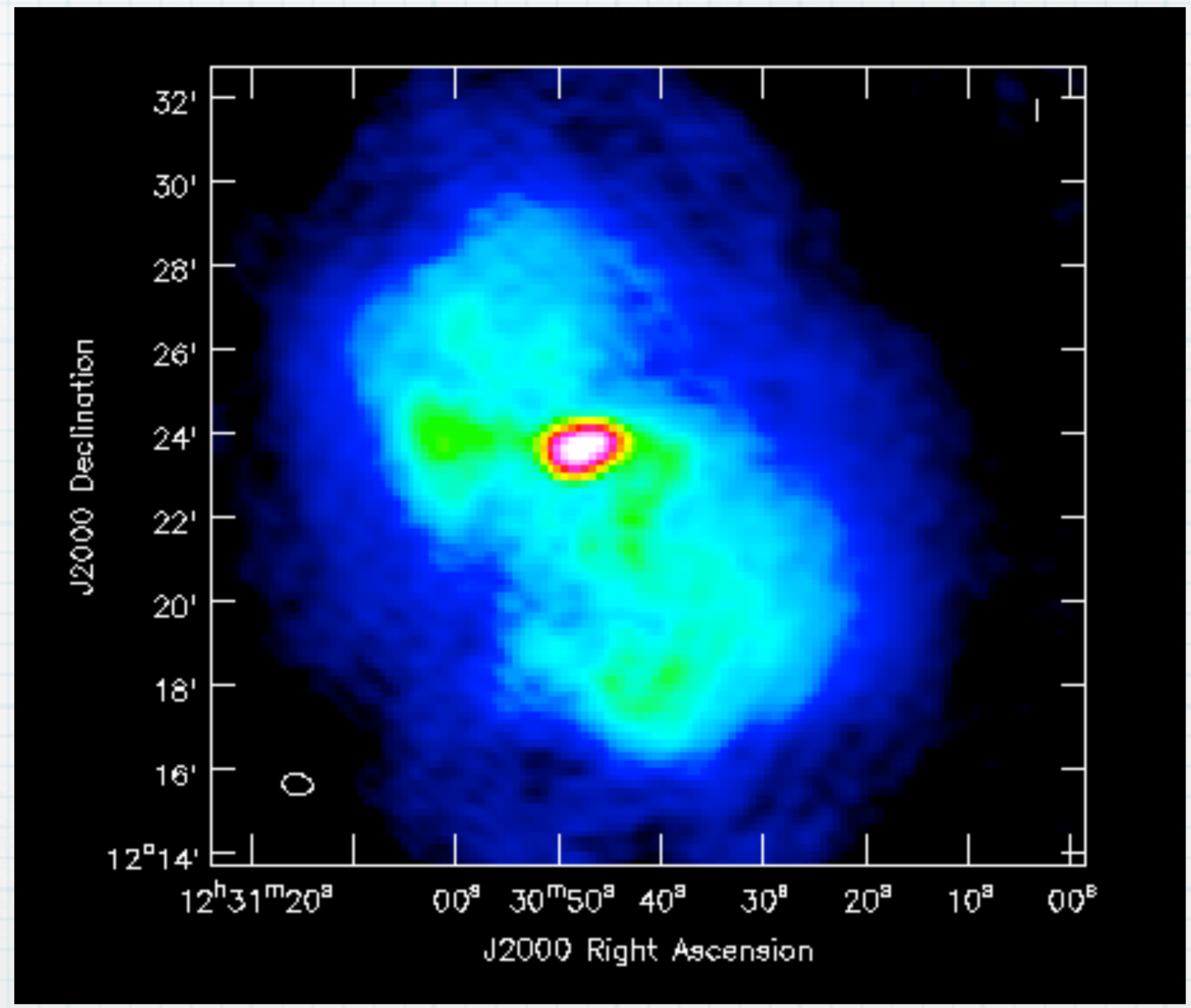
wrong model
Gaussian





good model

CC



BBS(Black board Selfacl)

$$V_{pq} = \int_{4\pi} \mathrm{J}_p(\sigma) B(\sigma) J_q^H(\sigma) d\Omega$$

$$J_p(l) = G_p E_p(t,l) K_p(l)$$

M.E.

Smirnov 2011

- -G direction independent term
- -E direction dependent term

Strategy

Solve

Correct

Parset

Strategy.ChunkSize = 0
Strategy.UseSolver = F
Strategy.Steps = [solve, correct]

Step.solve.Operation = SOLVE
Step.solve.Model.Sources = [target1]
Step.solve.Model.Gain.Enable = T
Step.solve.Model.Cache.Enable = T
Step.solve.Model.Beam.Enable = T
Step.solve.Solve.Parms = ["Gain:0:0:*","Gain:1:1:*"]
Step.solve.Solve.CellSize.Freq = 0
Step.solve.Solve.CellSize.Time = 1
Step.solve.Solve.CellChunkSize = 10
Step.solve.Solve.Options.MaxIter = 20
Step.solve.Solve.CalibrationGroups = []
Step.solve.Solve.Options.EpsValue = 1e-9
Step.solve.Solve.Options.EpsDerivative = 1e-9
Step.solve.Solve.Options.ColFactor = 1e-9
Step.solve.Solve.Options.LMFactor = 1.0
Step.solve.Solve.Options.BalancedEqs = F
Step.solve.Solve.Options.UseSVD = T

Step.correct.Operation = CORRECT
Step.correct.Model.Sources = [target1]
Step.correct.Model.Gain.Enable = T
Step.correct.Model.Beam.Enable = T
Step.correct.Output.Column = CORRECTED_DATA

Parset

Strategy

Strategy.ChunkSize = 0
Strategy.UseSolver = F
Strategy.Steps = [solve, correct]

Solve

Step.solve.Operation = SOLVE
Step.solve.Model.Sources = [target1]
Step.solve.Model.Gain.Enable = T
Step.solve.Model.Cache.Enable = T
Step.solve.Model.Beam.Enable = T
Step.solve.Solve.Parms = ["Gain:0:0:*","Gain:1:1:*"]
Step.solve.Solve.CellSize.Freq = 0
Step.solve.Solve.CellSize.Time = 1
Step.solve.Solve.CellChunkSize = 10
Step.solve.Solve.Options.MaxIter = 20
Step.solve.Solve.CalibrationGroups = []
Step.solve.Solve.Options.EpsValue = 1e-9
Step.solve.Solve.Options.EpsDerivative = 1e-9
Step.solve.Solve.Options.ColFactor = 1e-9
Step.solve.Solve.Options.LMFactor = 1.0
Step.solve.Solve.Options.BalancedEqs = F
Step.solve.Solve.Options.UseSVD = T

Correct

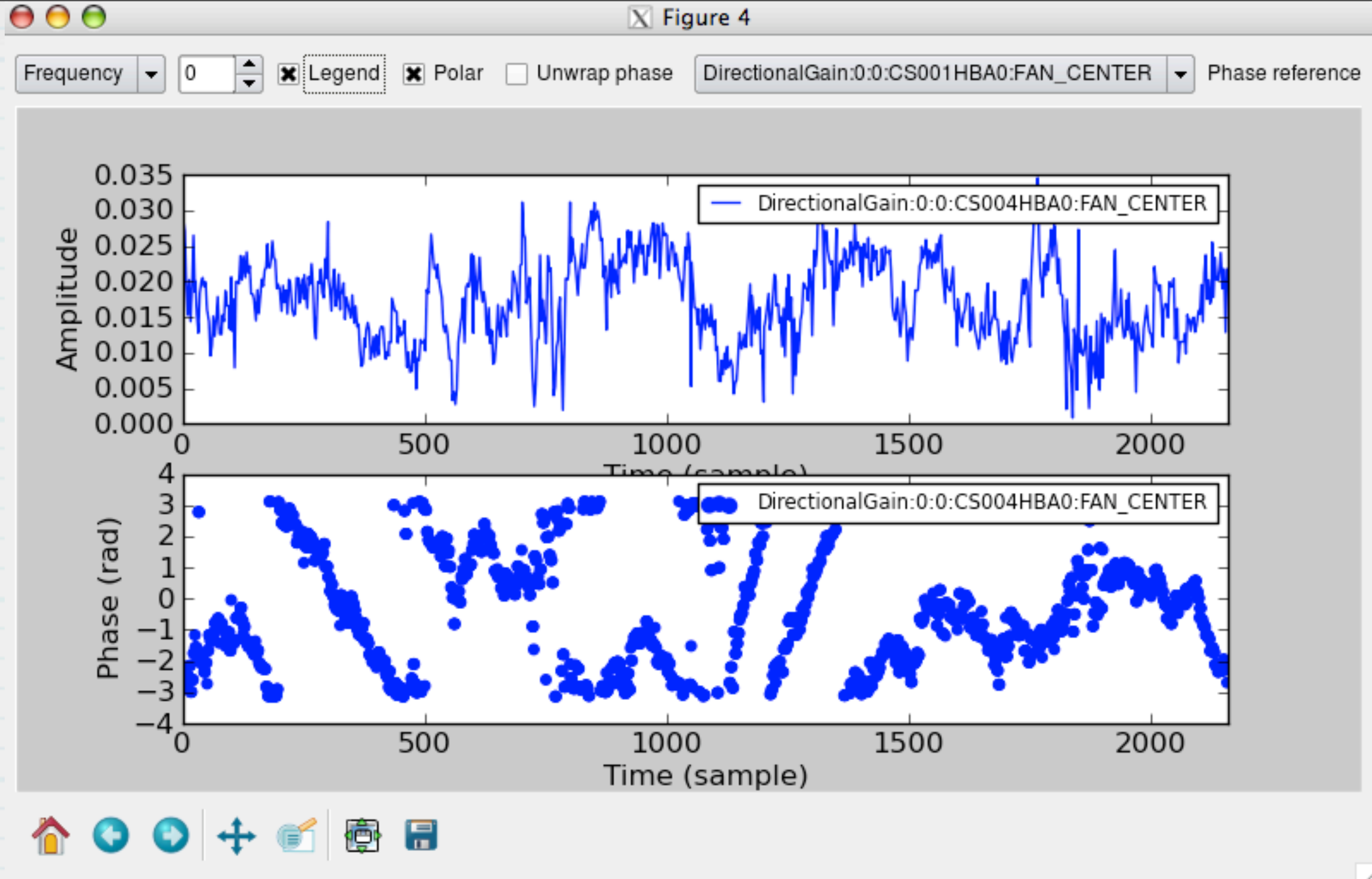
Step.correct.Operation = CORRECT
Step.correct.Model.Sources = [target1]
Step.correct.Model.Gain.Enable = T
Step.correct.Model.Beam.Enable = T
Step.correct.Output.Column = CORRECTED_DATA

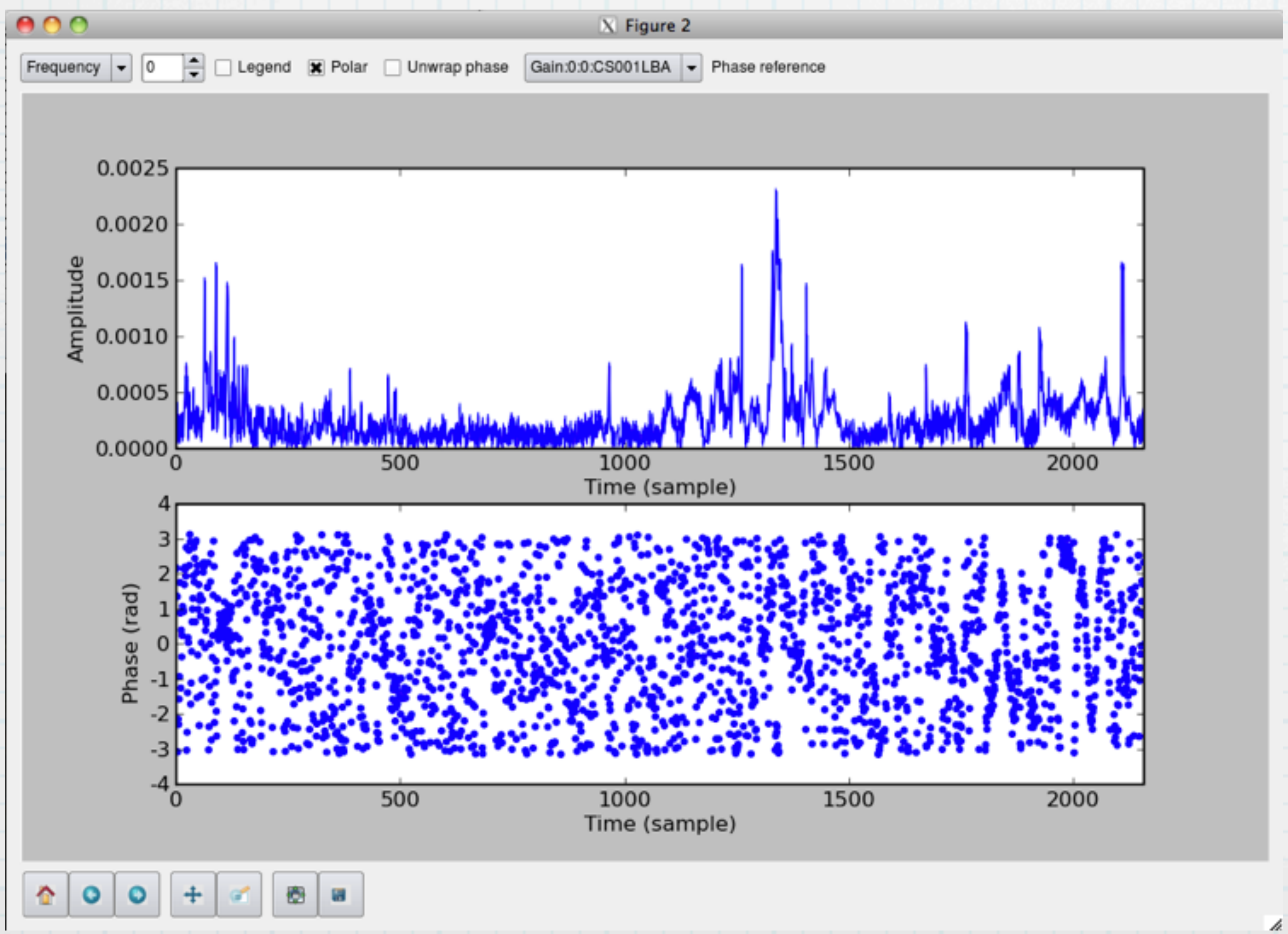
Beam

$$\begin{bmatrix} XX & XY \\ YX & YY \end{bmatrix}$$

Antenna Gain

Calibration solutions





which solution I like?

BBS(Black board Selfacl)

$$V_{pq} = \int_{4\pi} \mathrm{J}_p(\sigma) B(\sigma) J_q^H(\sigma) d\Omega$$

M.E.

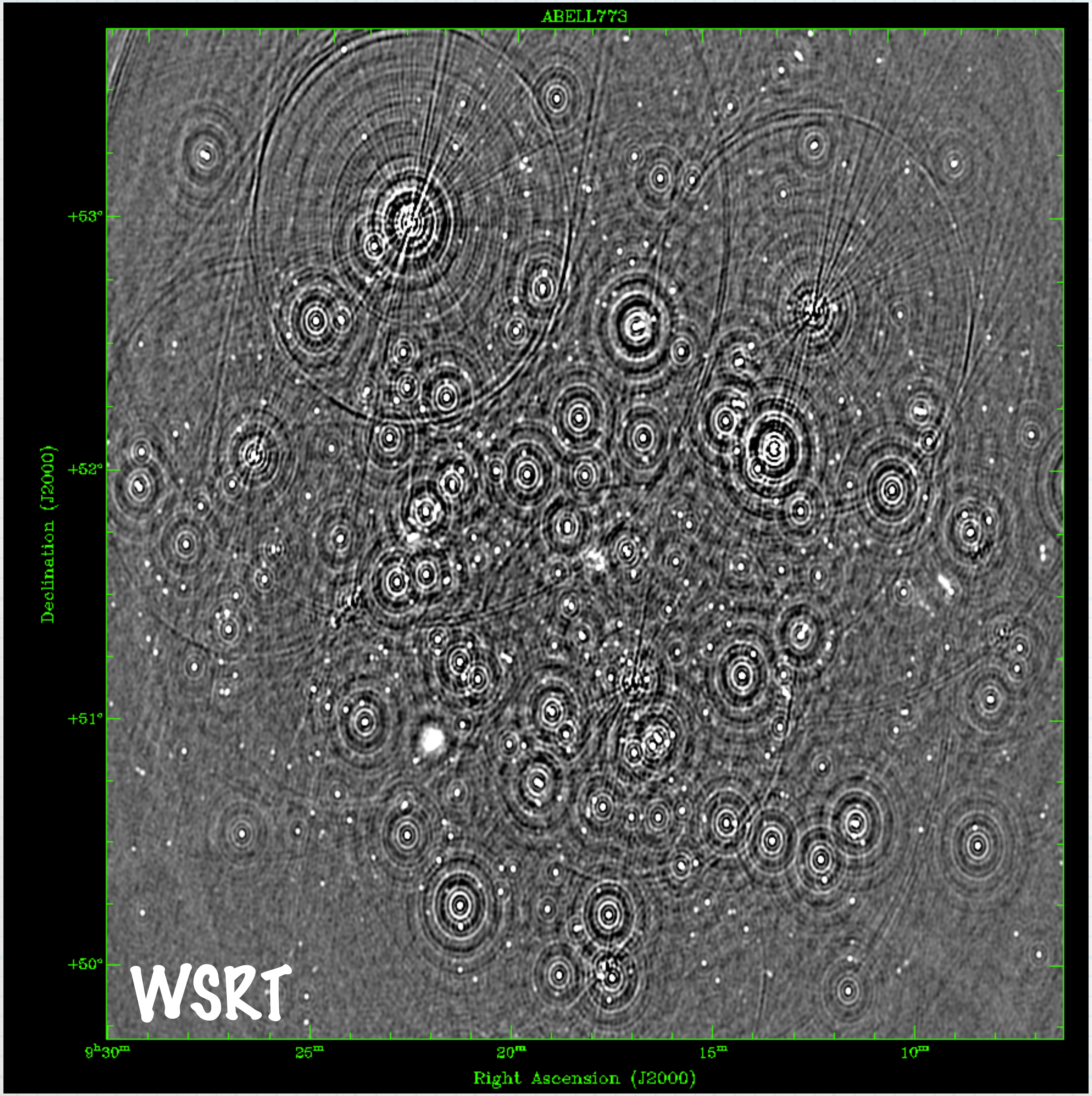
Smirnov 2011

$$J_p(l) = G_p E_p(t, l) K_p(l)$$

-G direction independent term
-E direction dependent term

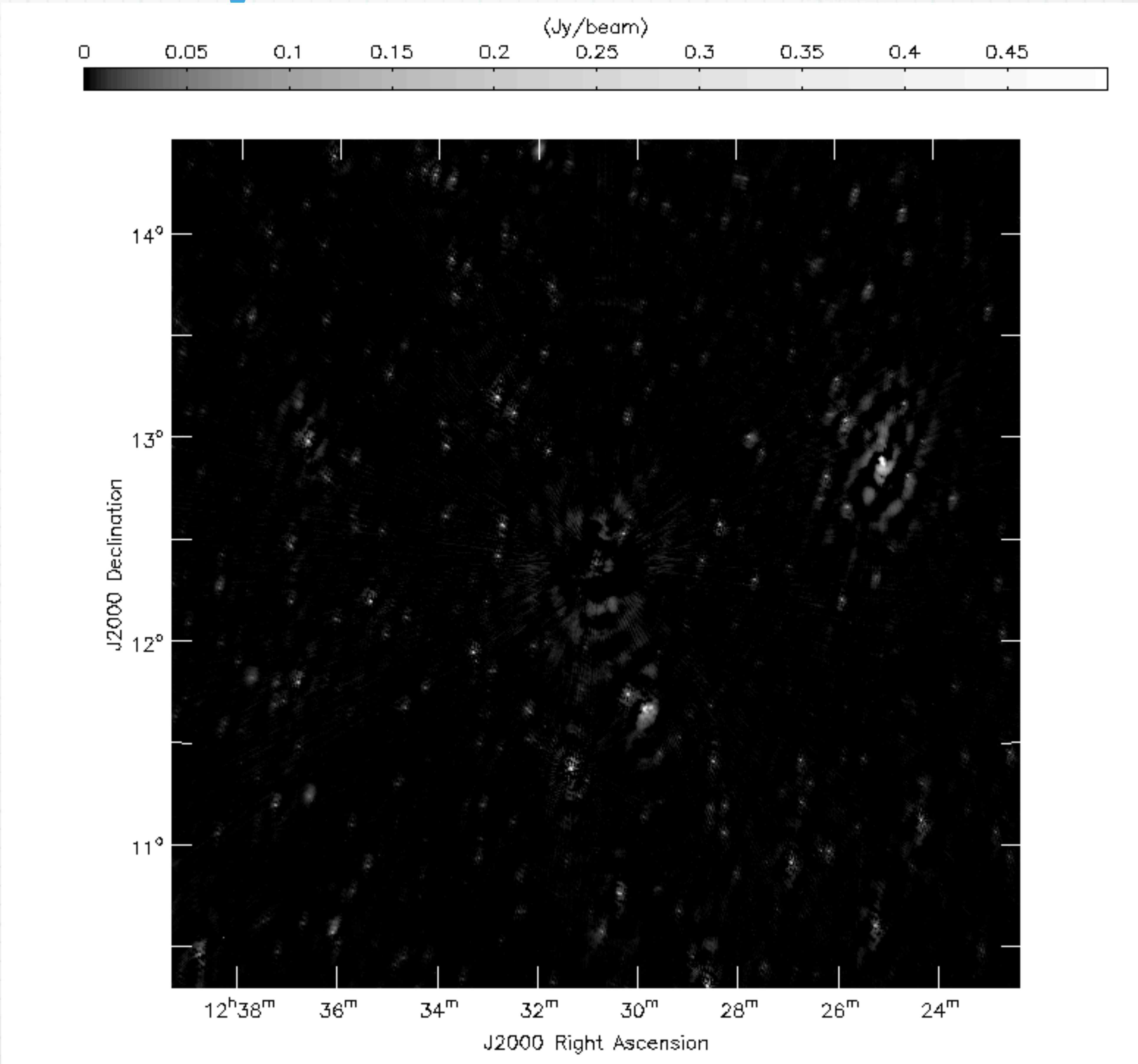
* ionosphere, variable beam
* effects: artifacts in offset sources when applying for the full FOV solutions obtained in single direction
* need to solve in directions of bright offset sources
* as showed before A-team removal

6: Direction Dependent Errors (DDEs)



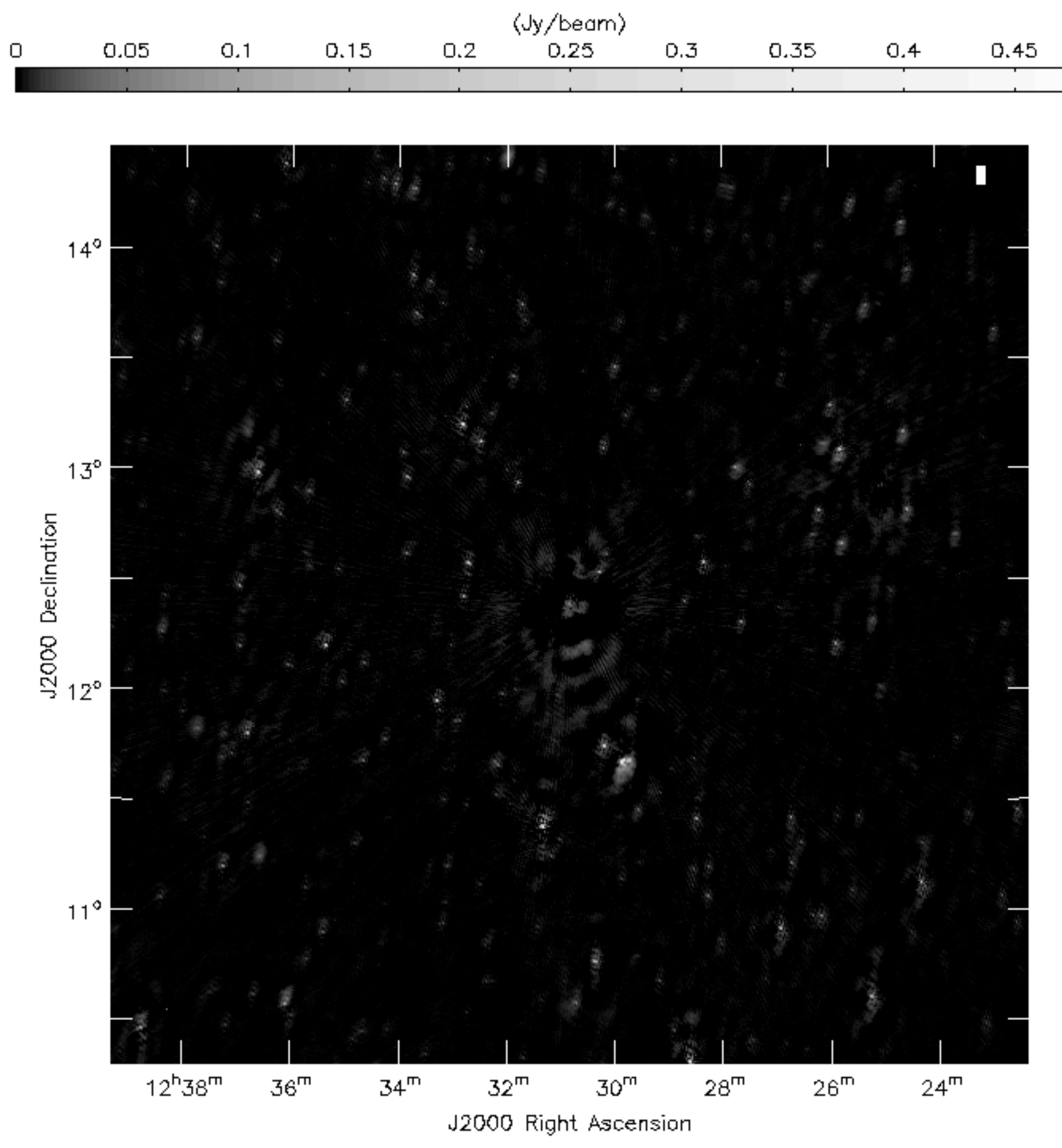
6: Direction Dependent Errors (DDEs)

6: Direction Dependent Errors (DDEs)



DDEs with BBS

```
Step.solve.Model.DirectionalGain.Enable = T
Step.solve.Model.Sources       = [src1, src2, CygA]
Step.solve.Model.Cache.Enable = T
Step.solve.Model.Beam.Enable = T
Step.solve.Correlation.Selection = CROSS
Step.solve.Correlation.Type = []
Step.solve.Solve.Parms = ["DirectionalGain:*"]
```



- computationally expensive, limited number of directions
- in avg data smearing effects

DDEs with SAGECAL

Space Alternating Generalized Expectation Maximization Calibration

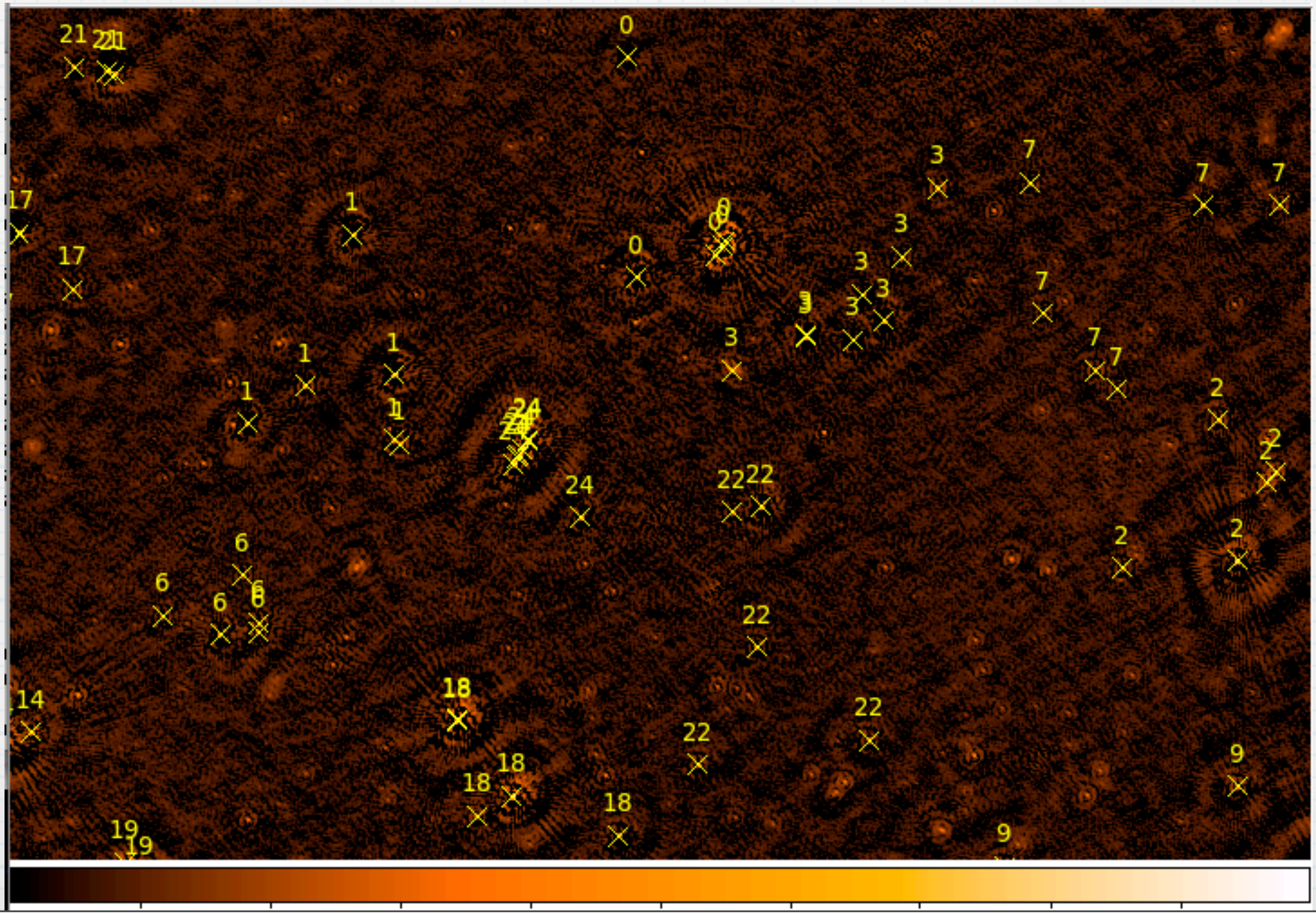
Yatawatta at al. 2009 and Kazemi et al. 2011

* data averaged to one freq. channel and pre-calibrated with BBS: solve for G and beam
* create an updated model for the selfcal
* create the cluster file where the directions you will solve for are grouped

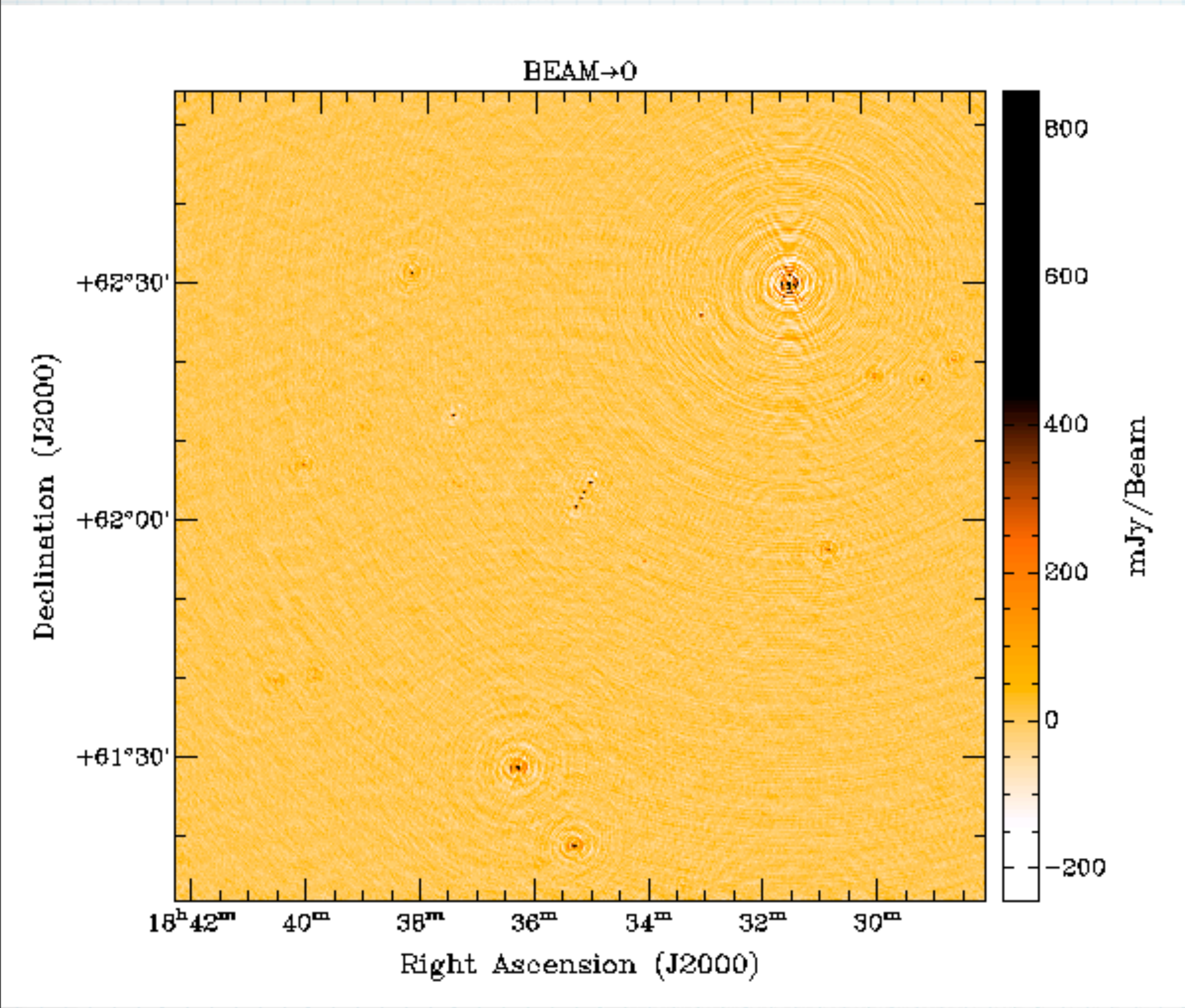
model

MODEL
this is an LSM text (hms/dms) file
fields are (where h:m:s is RA, d:m:s is Dec):
name h m s d m s I Q U V spectral_index RM extent_X(rad) extent_Y(rad) pos_angle(rad) freq0
P2C1 18 35 38.540 59 39 56.535 0.563165 0 0 0 -1.984837 0 0 0 0 140171967.455621
P2C2 18 35 39.513 59 39 47.352 0.571942 0 0 0 -1.697620 0 0 0 0 140171967.455621
P1C1 18 45 31.251 59 38 12.570 0.125407 0 0 0 -2.019737 0 0 0 0 140171967.455621
P3C1 18 29 56.454 59 45 4.239 0.260833 0 0 0 -3.134929 0 0 0 0 140171967.455621
P4C1 18 46 37.449 59 46 5.129 0.115548 0 0 0 -0.908031 0 0 0 0 140171967.455621
P5C1 18 39 44.730 59 52 36.108 0.108609 0 0 0 -1.835421 0 0 0 0 140171967.455621.........

CLUSTER
0 P86C1 P88C1 P89C1 P90C1 P104C1 P105C1
1 P66C1 P69C1 P67C1 P72C1 P73C1 P92C1
2 P52C1 P52C2 P53C1 P55C1 P56C1 P61C1 P62C1 P70C1 P76C1 P85C1
3 P74C1 P77C1 P79C1 P79C2 P80C1 P83C1 P87C1 P97C1
4 P129C1 P130C1 P131C1 P132C1 P134C1 P135C1 P136C1 P137C1
5 P34C1 P34C2 P49C1
6 P47C1 P50C1 P48C1 P51C1 P54C1
7 P71C1 P75C1 P81C1 P96C1 P95C1 P98C1
8 P13C1 P15C1 P19C1 P23C1 P23C2 P25C1 P26C1 P28C1
9 P17C1 P24C1 P31C1 P30C1 P33C1 P39C1.......

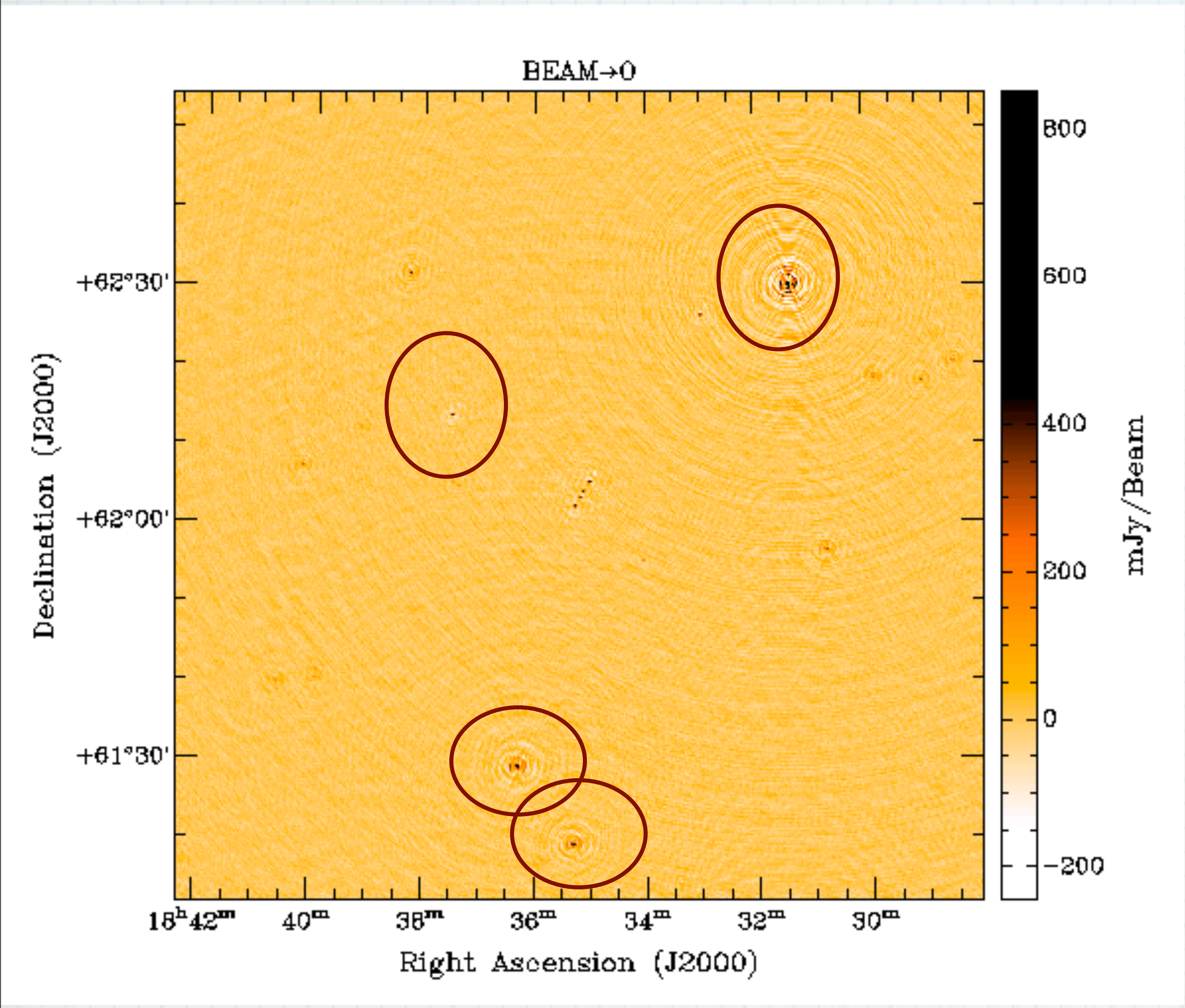


before and after



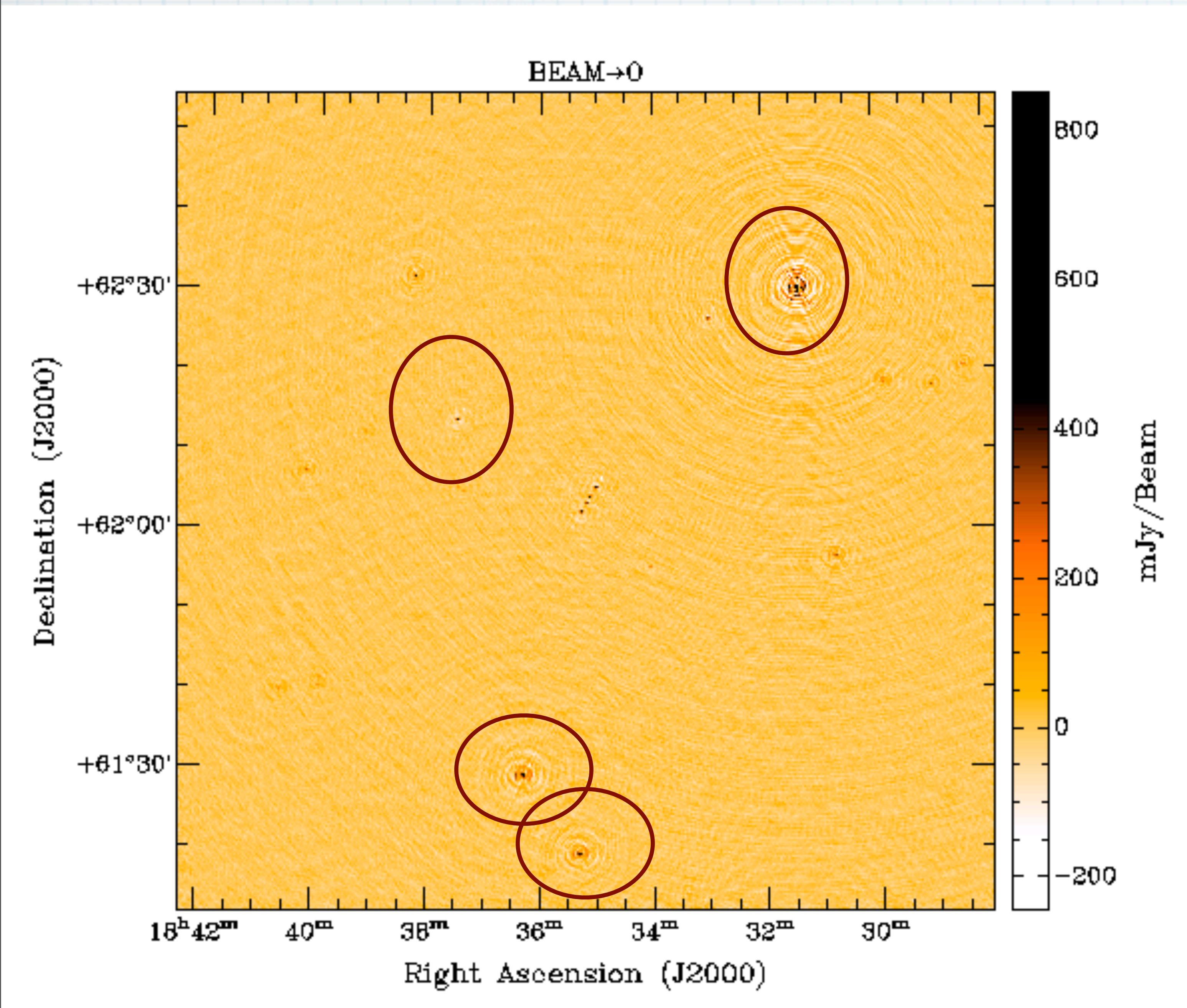
for which directions would you correct for?

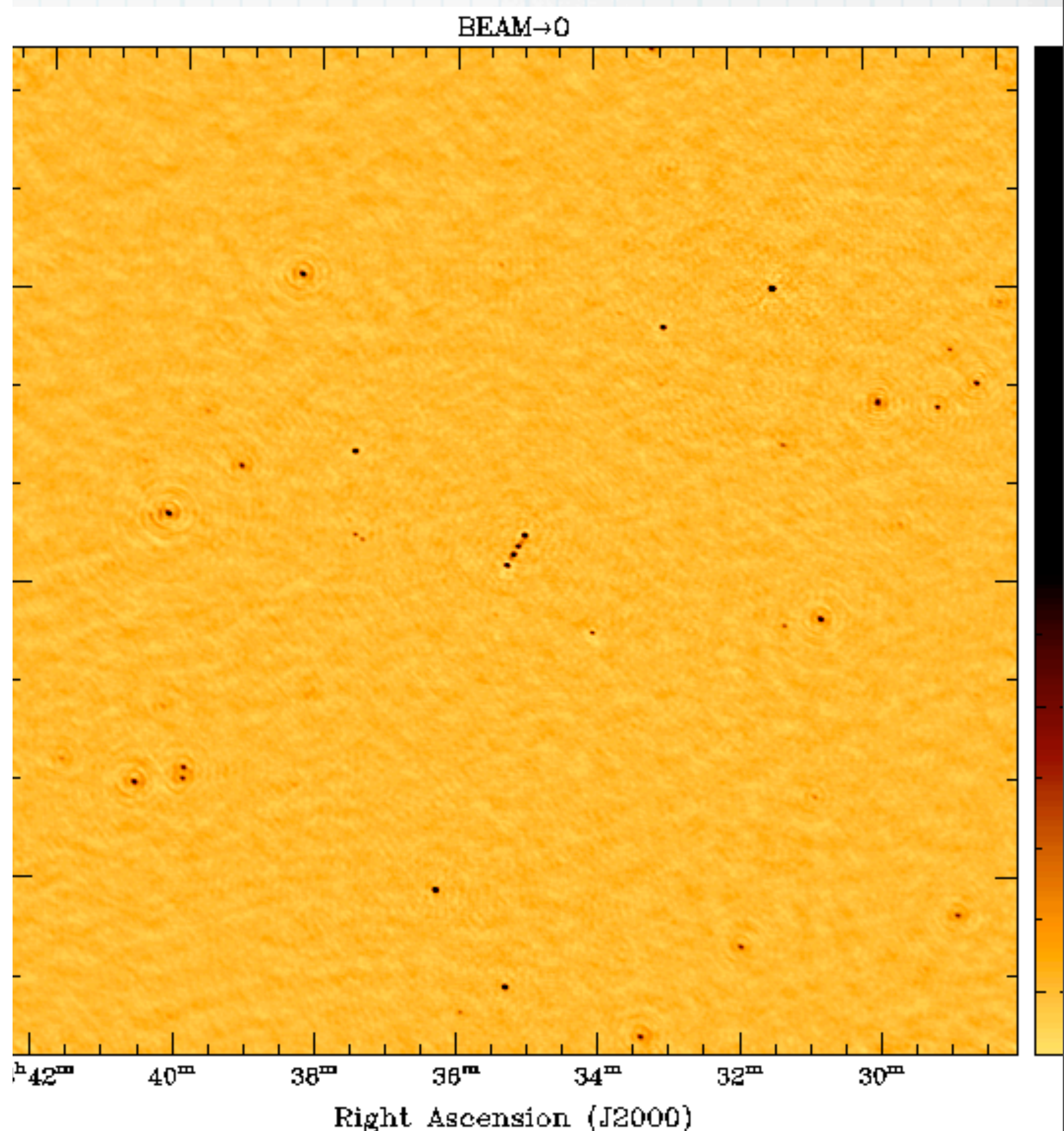
before and after



for which directions would you correct for?

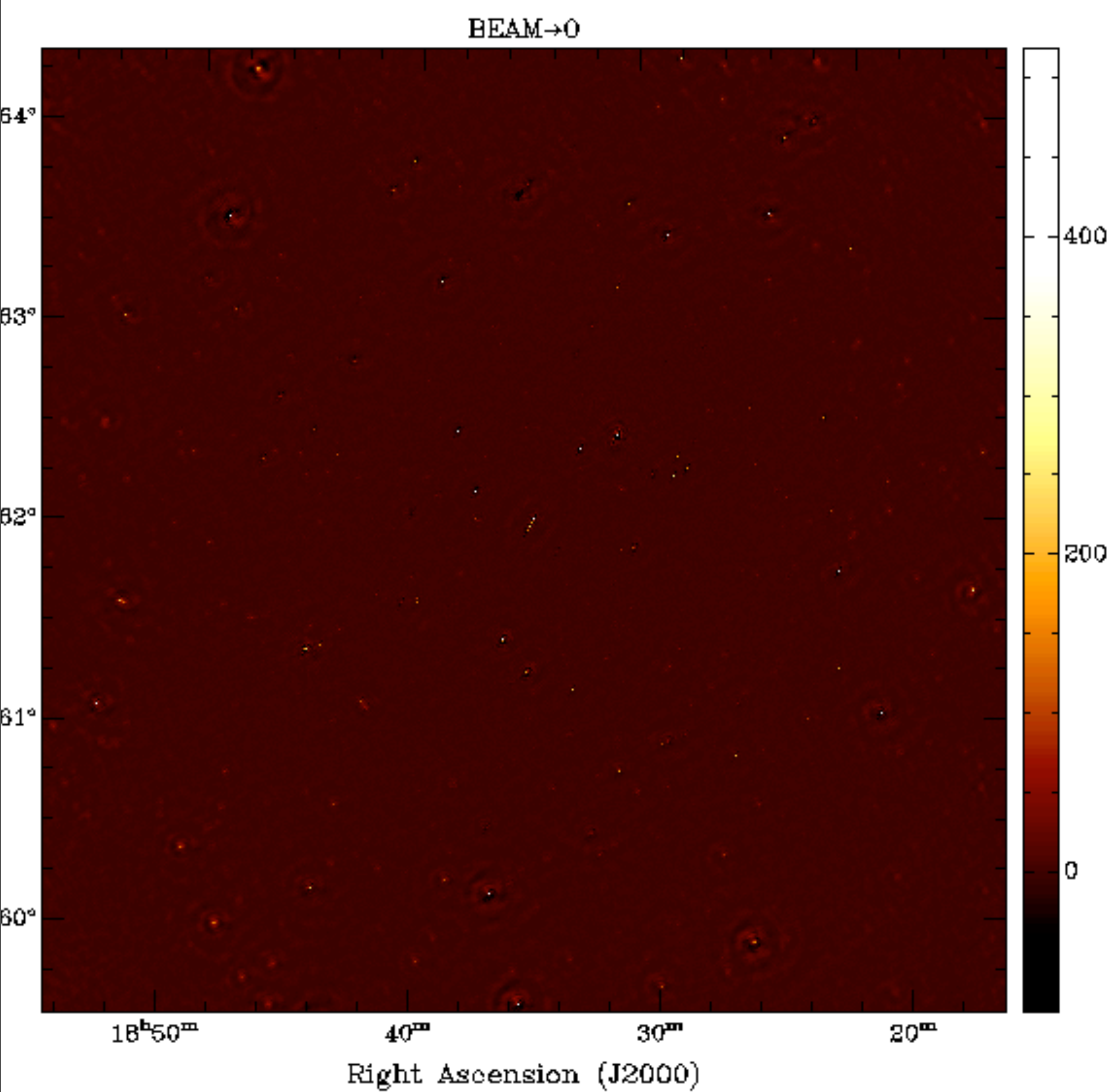
before and after

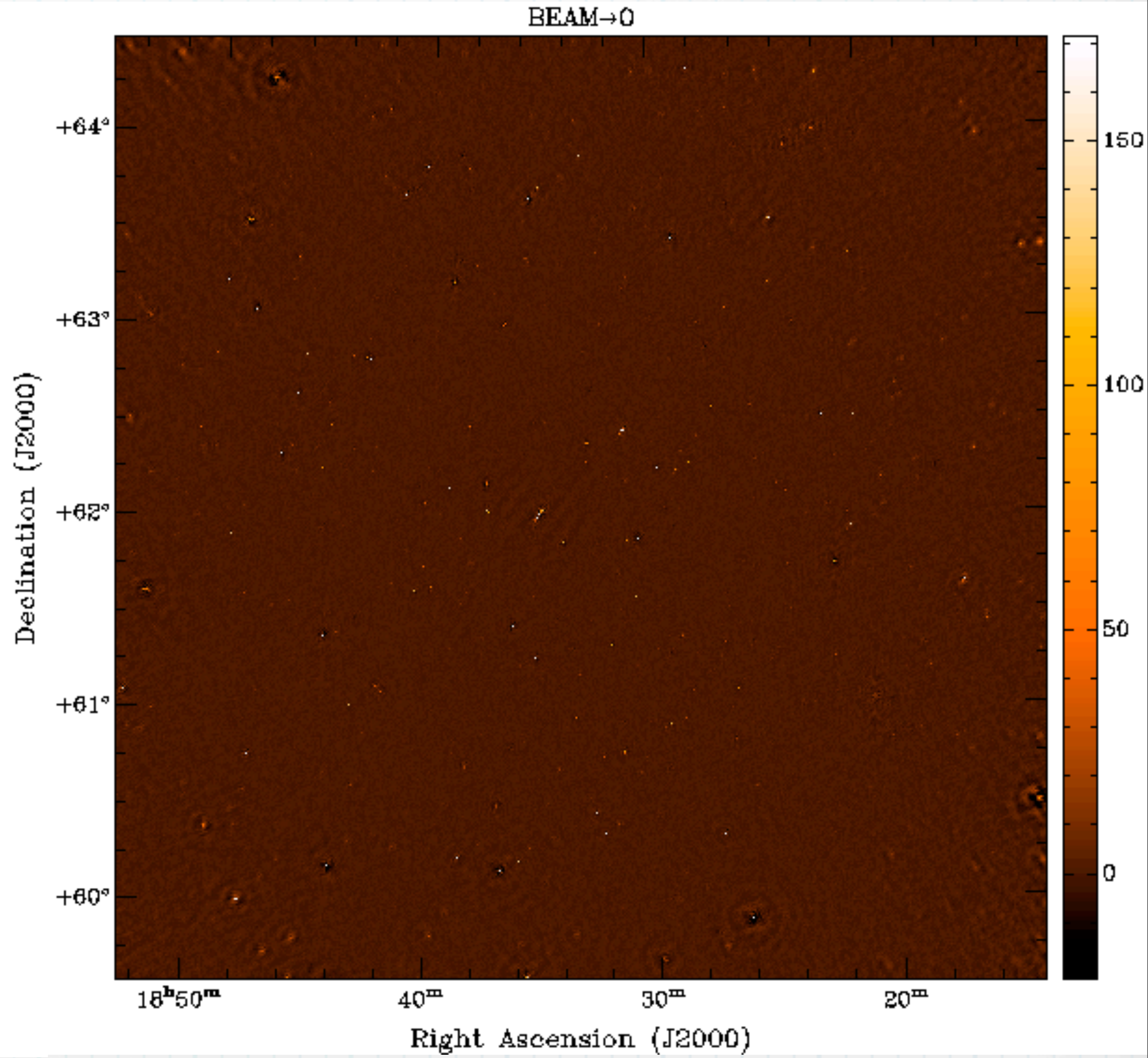




for which directions would you correct for?

full band
before and after

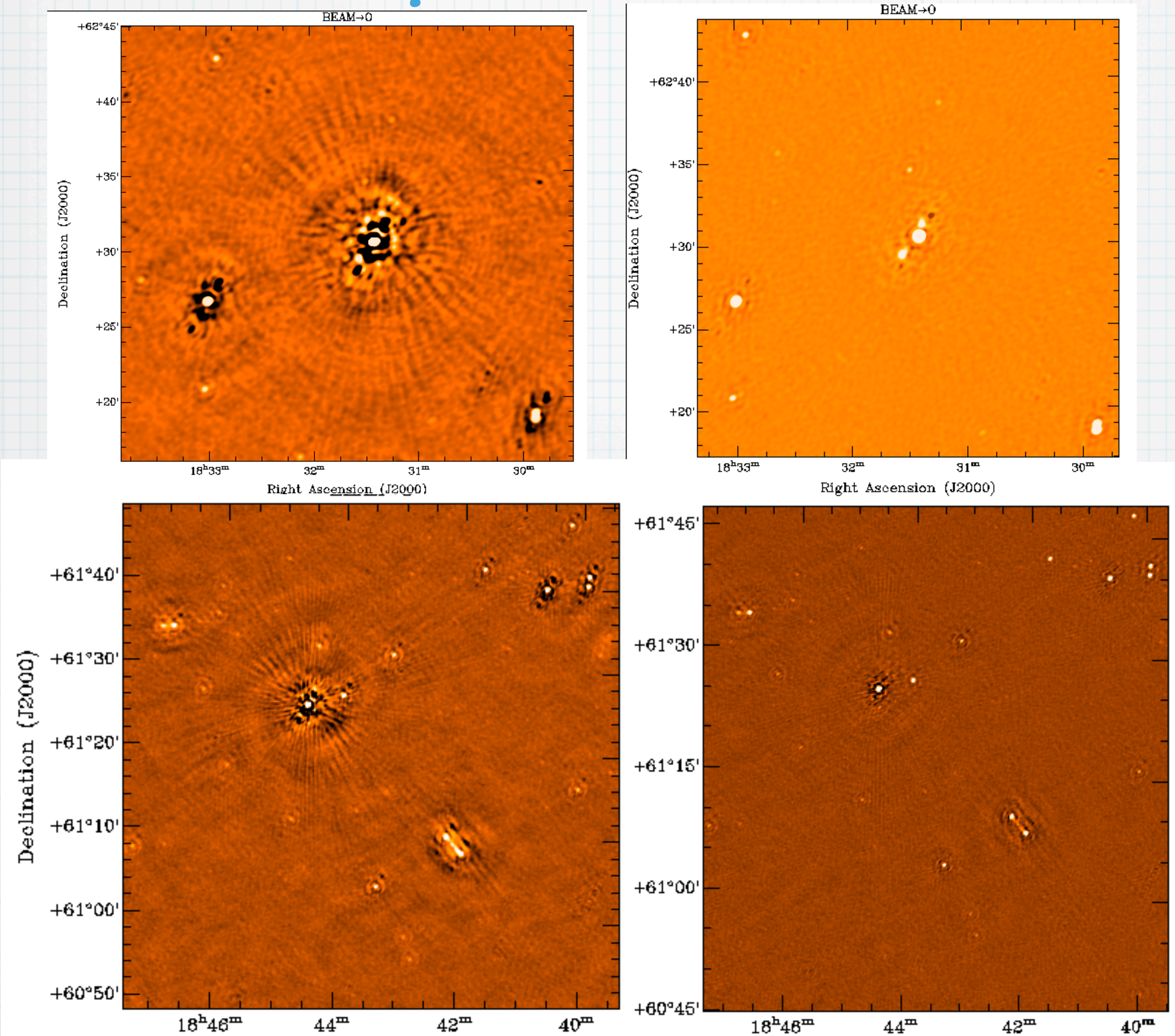




rms=1.2 mJy/beam

rms=0.8 mJy/beam

ionosphere+deconvolution errors



7: Imaging

* **w-term** takes into account of the non-coplanar baselines

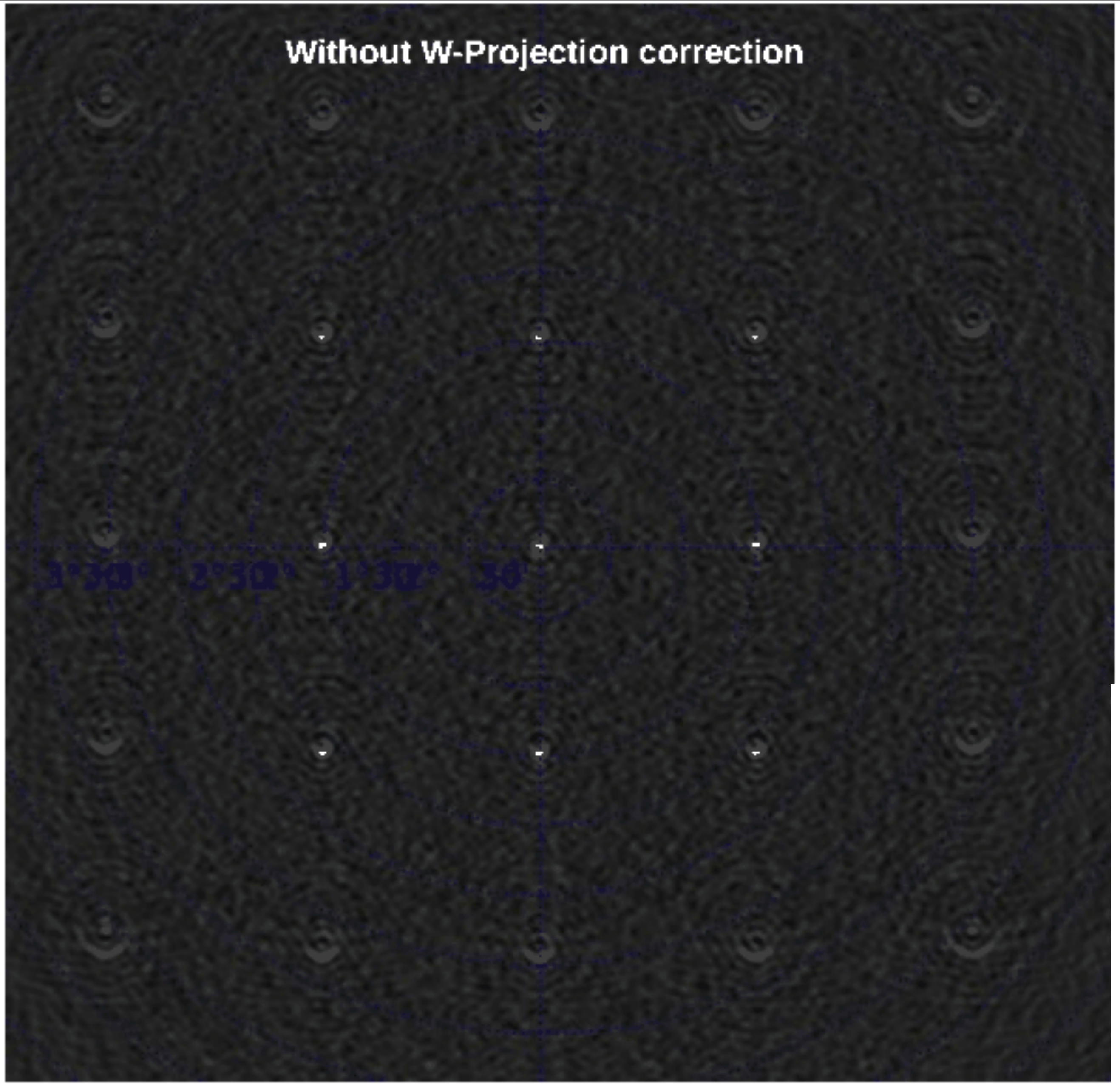
$$V(u,v,w) = \int \frac{I(l,m)}{\sqrt{1-l^2-m^2}} e^{2\pi i[ul+vm+w(\sqrt{1-l^2-m^2}-1]} dl dm$$

$$2\pi w(\sqrt{1-l^2-m^2}-1) < 1$$ can be ignored, 2D FT

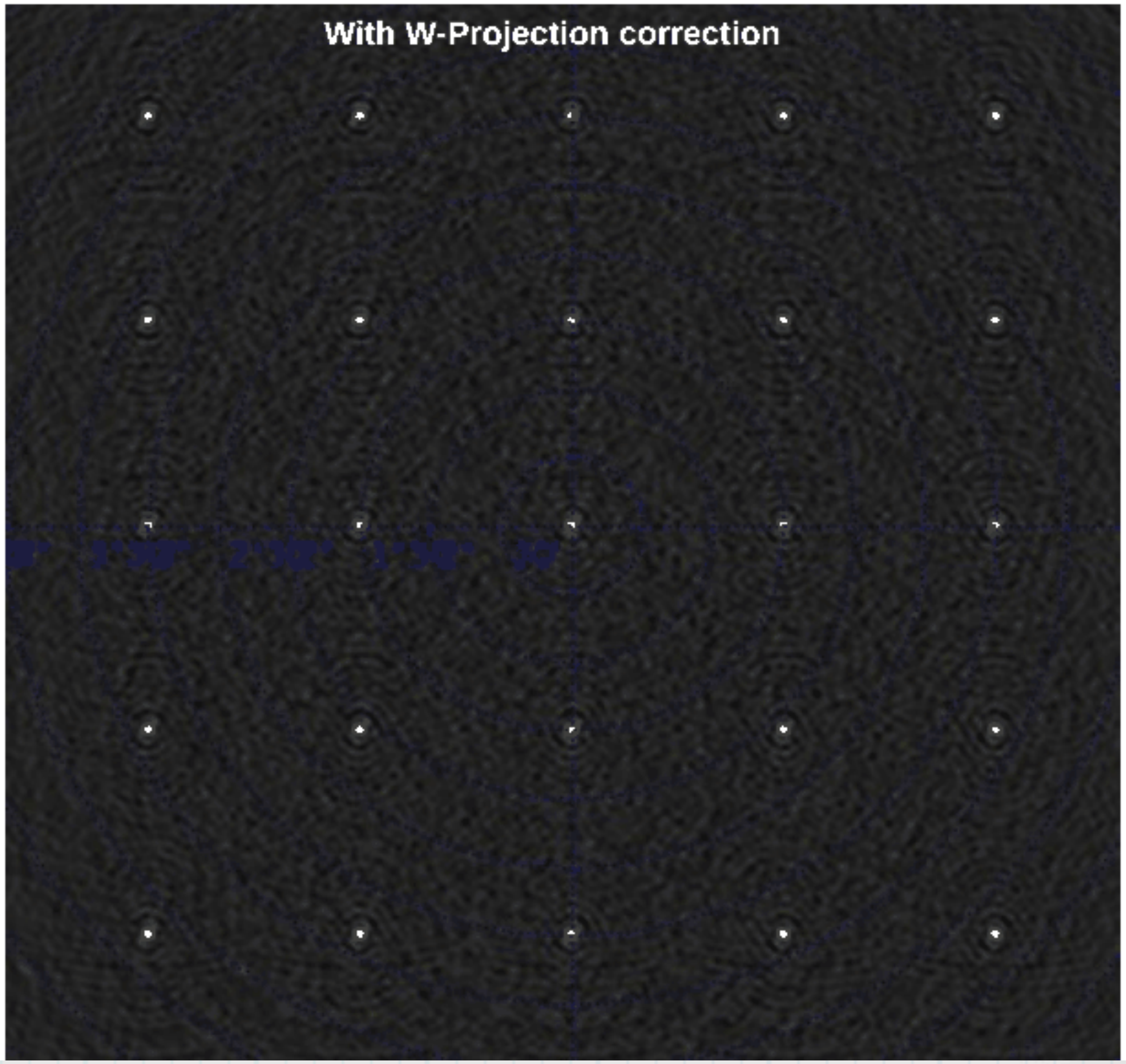
$$2\pi w(\sqrt{1-l^2-m^2}-1) > 1$$ can not be ignored, 3D FT

this happens when $N_F = \frac{D^2}{B\lambda} < 1$

* occurs for small apertures, long baselines, or long wave- lengths.
* methods proposed so far such as faceting computationally expensive, w-projection is the most promising



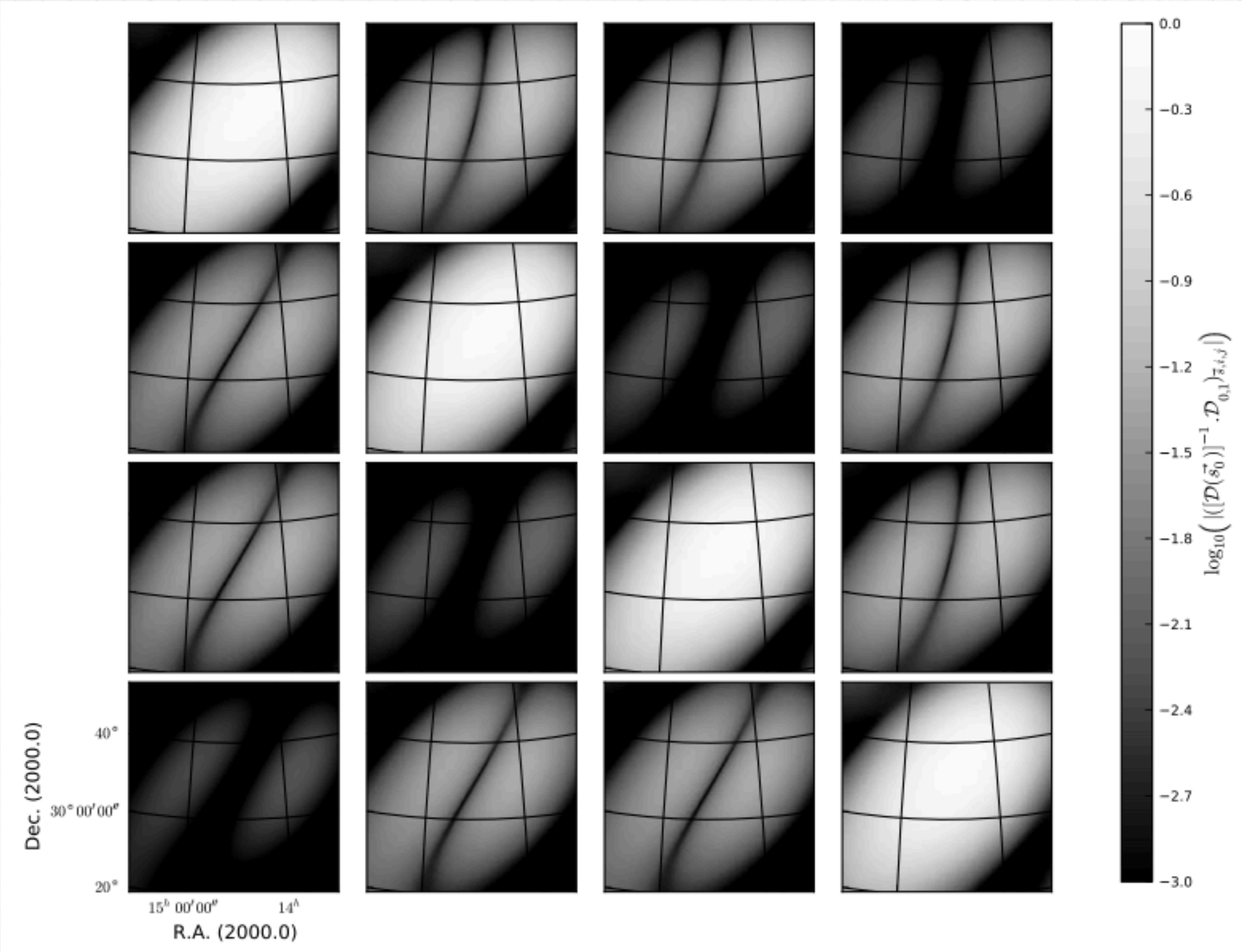
courtesy of C. Tasse



* A-term takes into account all the effects associated to DDE

* dishes-based interferometers, where the beam shape and polarization angle are affected by pointing errors and rotated on the sky by the parallactic angle (depending on the dish mount) no influence

* LOFAR is based on phased arrays that have very wide fields of view (up to ~ 12 degrees), non- trivial and quickly varying beams, thereby driving complicated polarization effects.



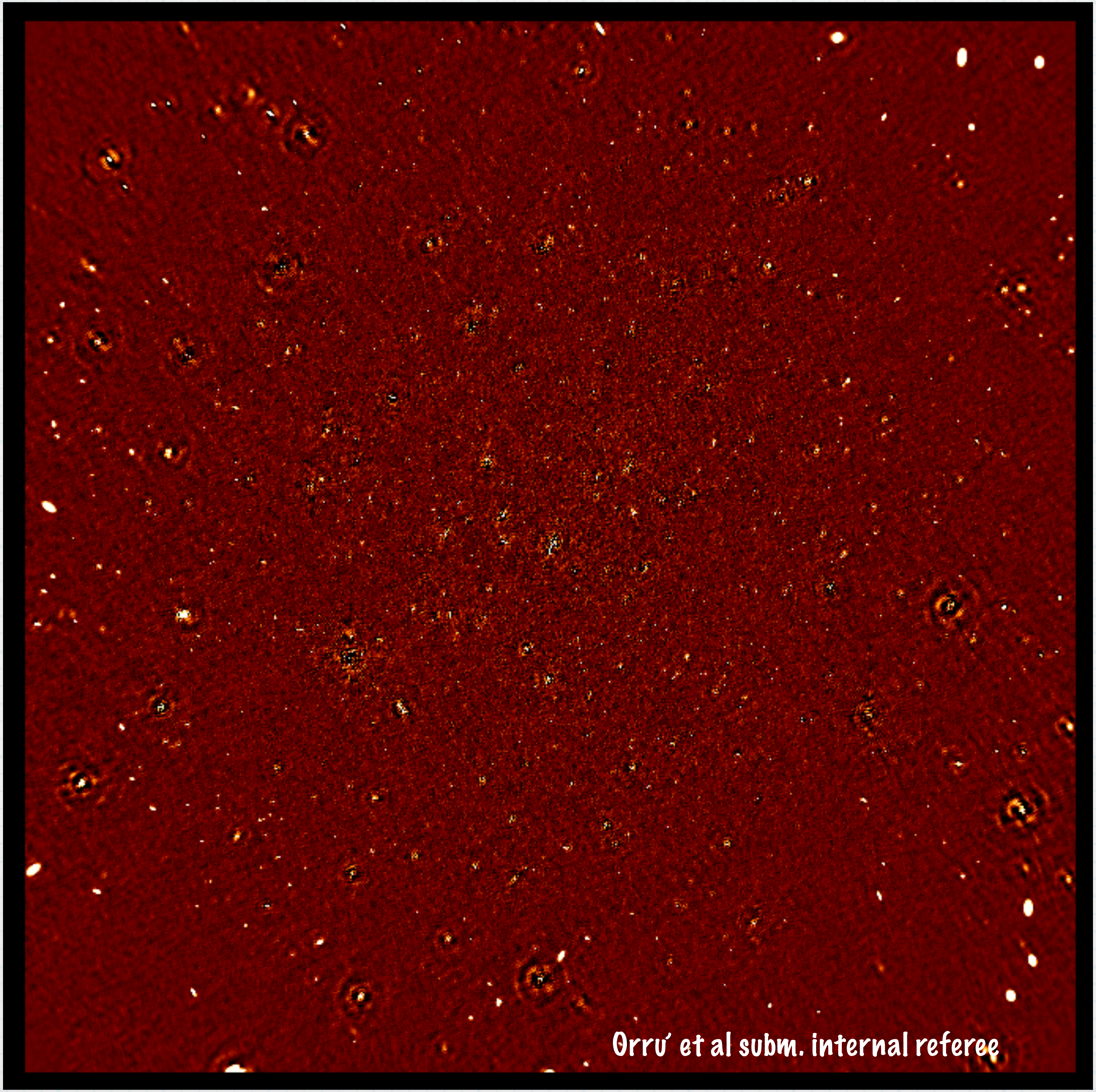
Non-unitary, time, frequency, baseline dependent Muller Matrix

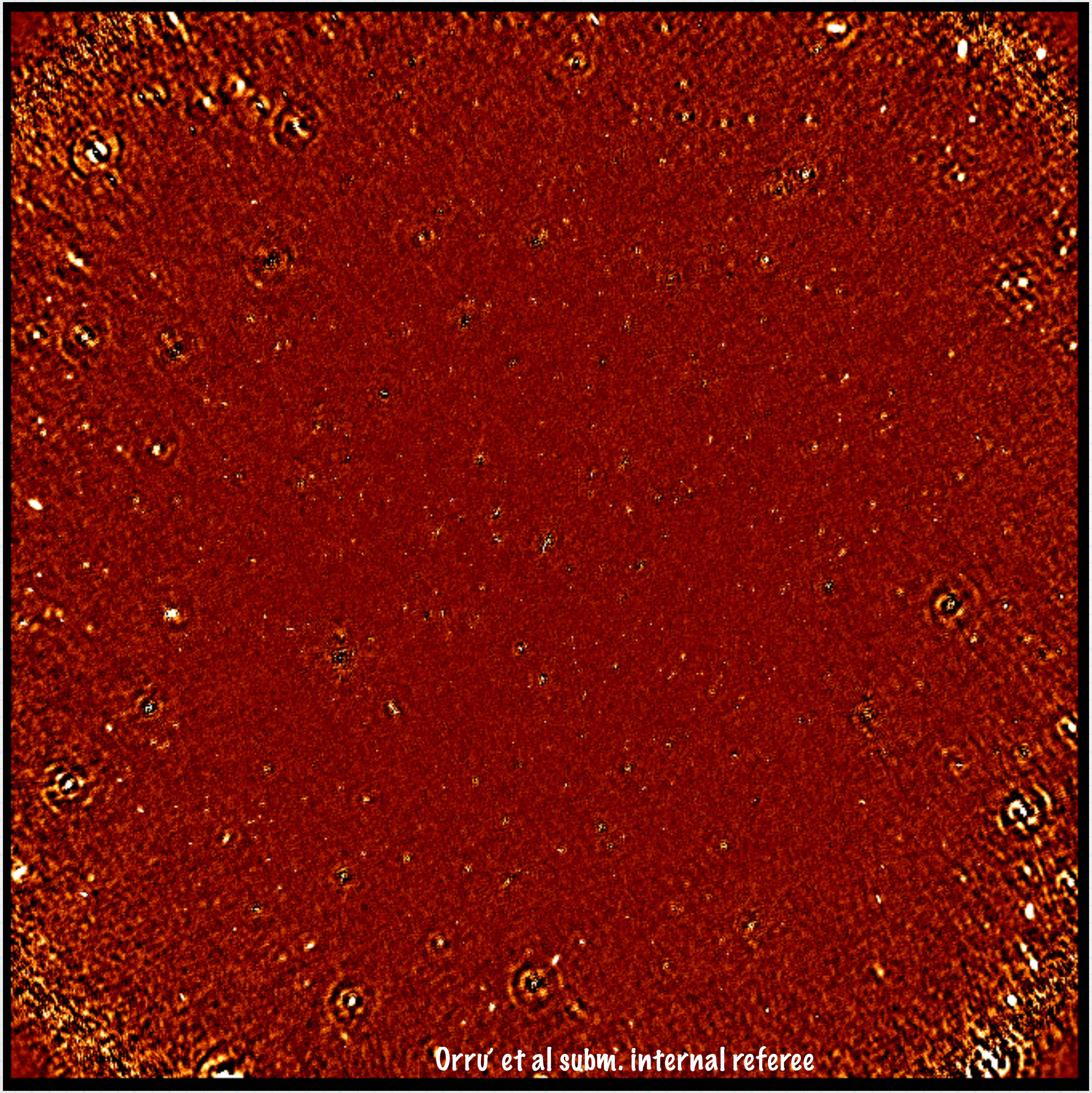
each individual polarization cannot be treated independently from the others

Tasse et al. 2013

* AW-imager is the imaging algorithm implemented for LOFAR
* takes into account the variable beam so we have flux corrected images also at the edge of the field
* in the future it will apply while imaging phase-screen in order to correct for the ionosphere

Can you find the difference?




Orru' et al subm. internal referee

conclusion

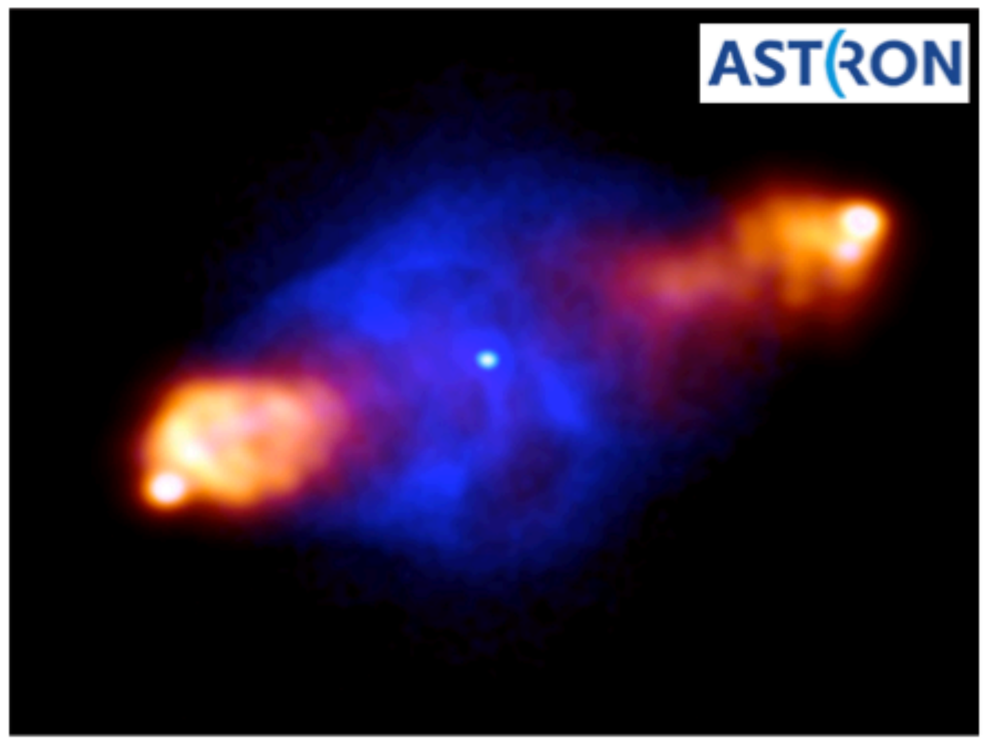
* LOFAR data are big so the effort to produce a nice image!
* we still miss to add ionospheric correction but the infrastructure is in place
* If you want to become a LOFAR commissioner please contact Roberto (pizzo@astron.nl) or me (orru@astron.nl).
* The LOFAR cookbook can be found in:http://www.lofar.org/operations/doku.php?id=commissioning:cookbooks

THE LOFAR IMAGING COOKBOOK:

Manual data reduction with the imaging pipeline

Version 13.0

August 30, 2013



Edited by Roberto F. Pizzo