The University of Manchester
Jodrell Bank
Observatory

MANCHESTER
1824

Long Baselines at Low Frequencies: high-resolution surveys with LOFAR

Neal Jackson, for the Long Baseline Working Group

Leah Morabito, Adam Deller, Alexander Drabent, Stephen Bourke, Eskil Varenius, Colm Coughlan, Olaf Wucknitz, Tobia Carozzi, Javier Moldon, Alexander Kappes, Marco Iacobelli, Sean Mooney, Tim Shimwell, Maaijke Mevius, Ger de Bruyn, Anna Kapinska, Kaspars Prusis, Atvars Nikolajevs, Naim Ramirez, Etienne Bonnassieux, Paul Burd, Marcin Hajduk, Gatis Gaigals, Evita Vavilina

The University of Manchester
Jodrell Bank
Observatory

MANCHESTER
1824

Long Baselines at Low Frequencies: high-resolution surveys with LOFAR

Neal Jackson, for the Long Baseline Working Group

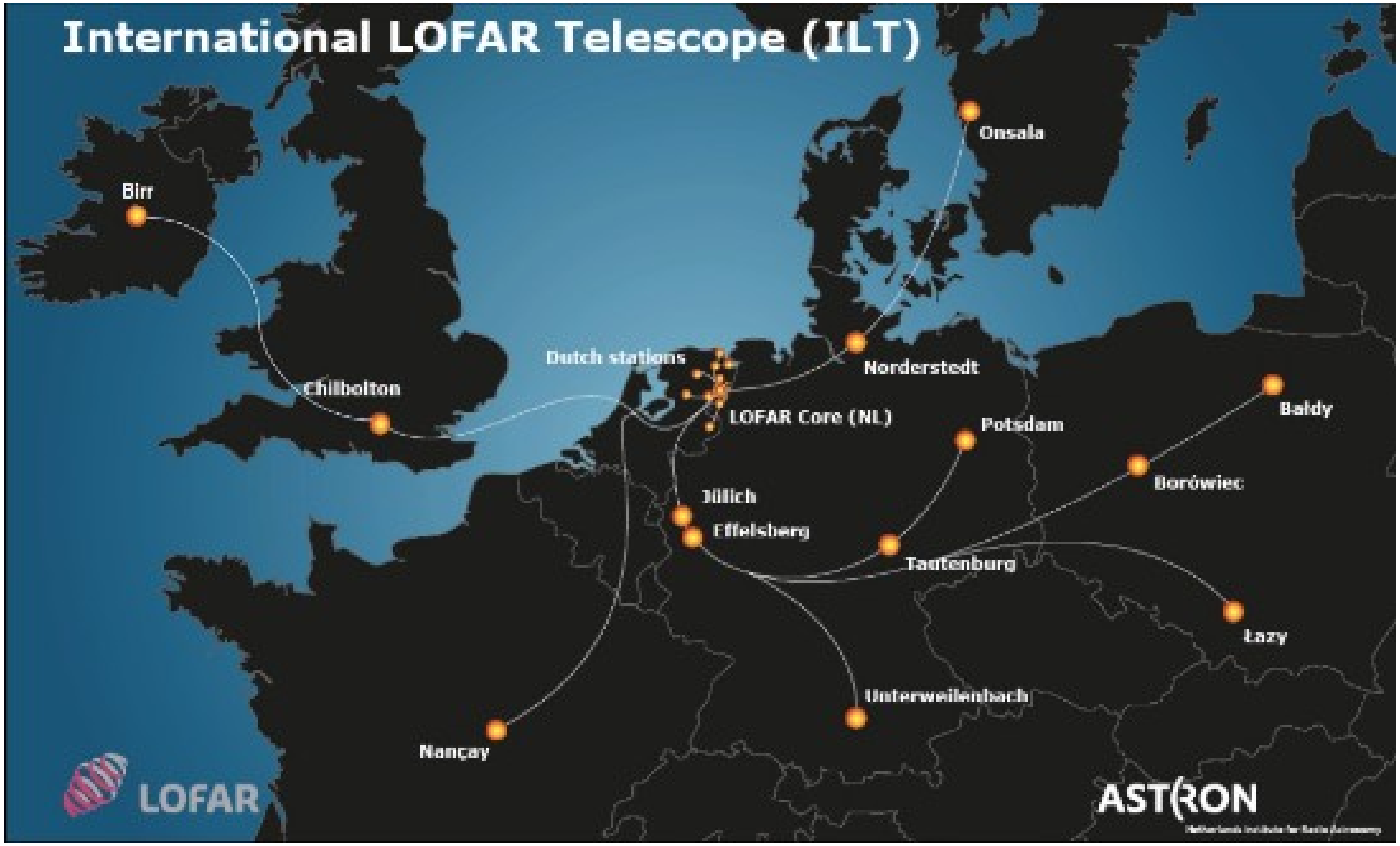
Leah Morabito, Adam Deller, Alexander Drabent, Stephen Bourke, Eskil Varenius, Colm Coughlan, Olaf Wucknitz, Tobia Carozzi, Javier Moldon, Alexander Kappes, Marco Iacobelli, Sean Mooney, Tim Shimwell, Maaijke Mevius, Ger de Bruyn, Anna Kapinska, Kaspars Prusis, Atvars Nikolajevs, Naim Ramirez, Etienne Bonnassieux, Paul Burd, Marcin Hajduk, Gatis Gaigals, Evita Vavilina

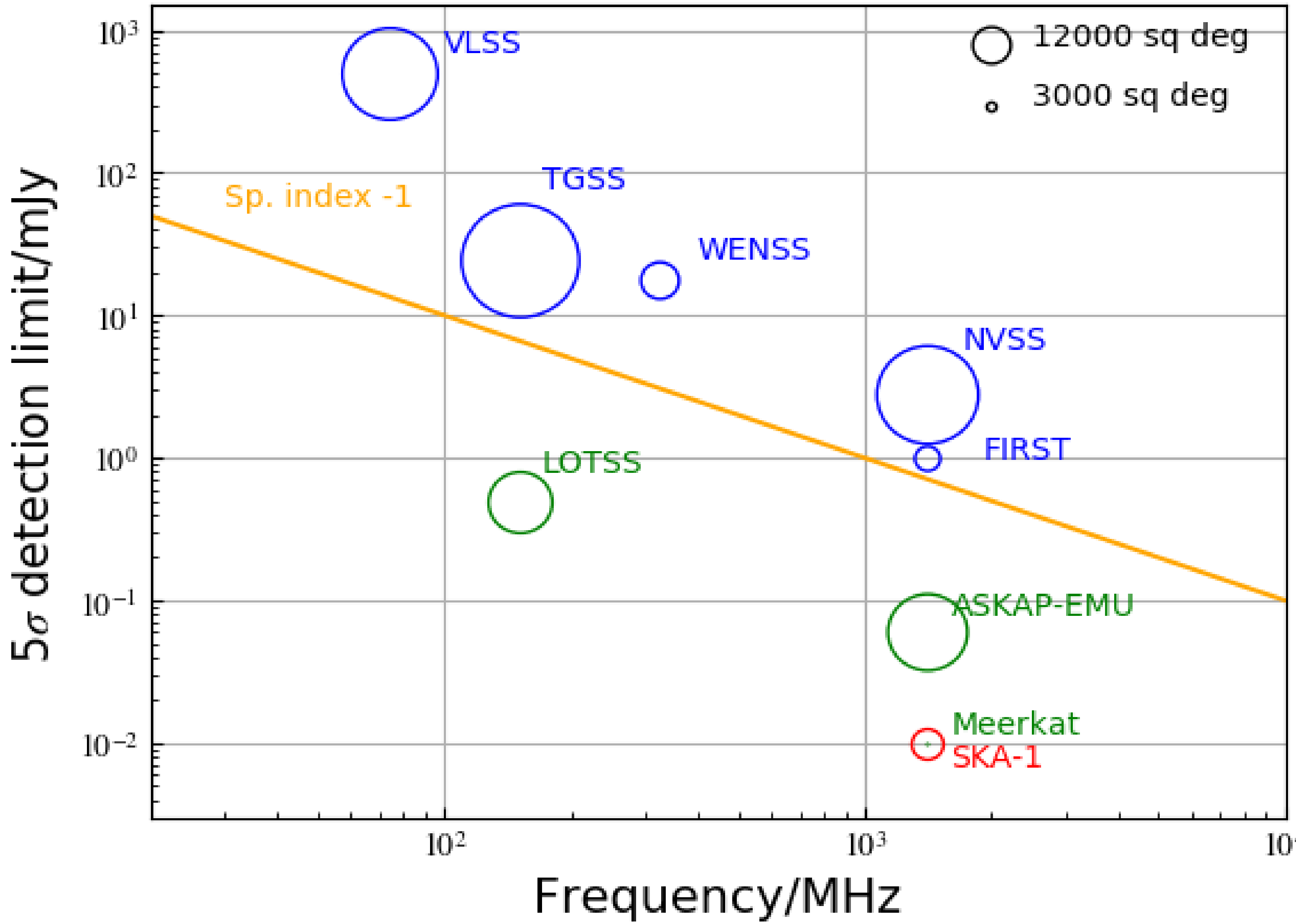
1. Introduction: survey parameter space
2. Scientific motivation: what can be done
3. Engineering: where we are

The University of Manchester
Jodrell Bank Observatory

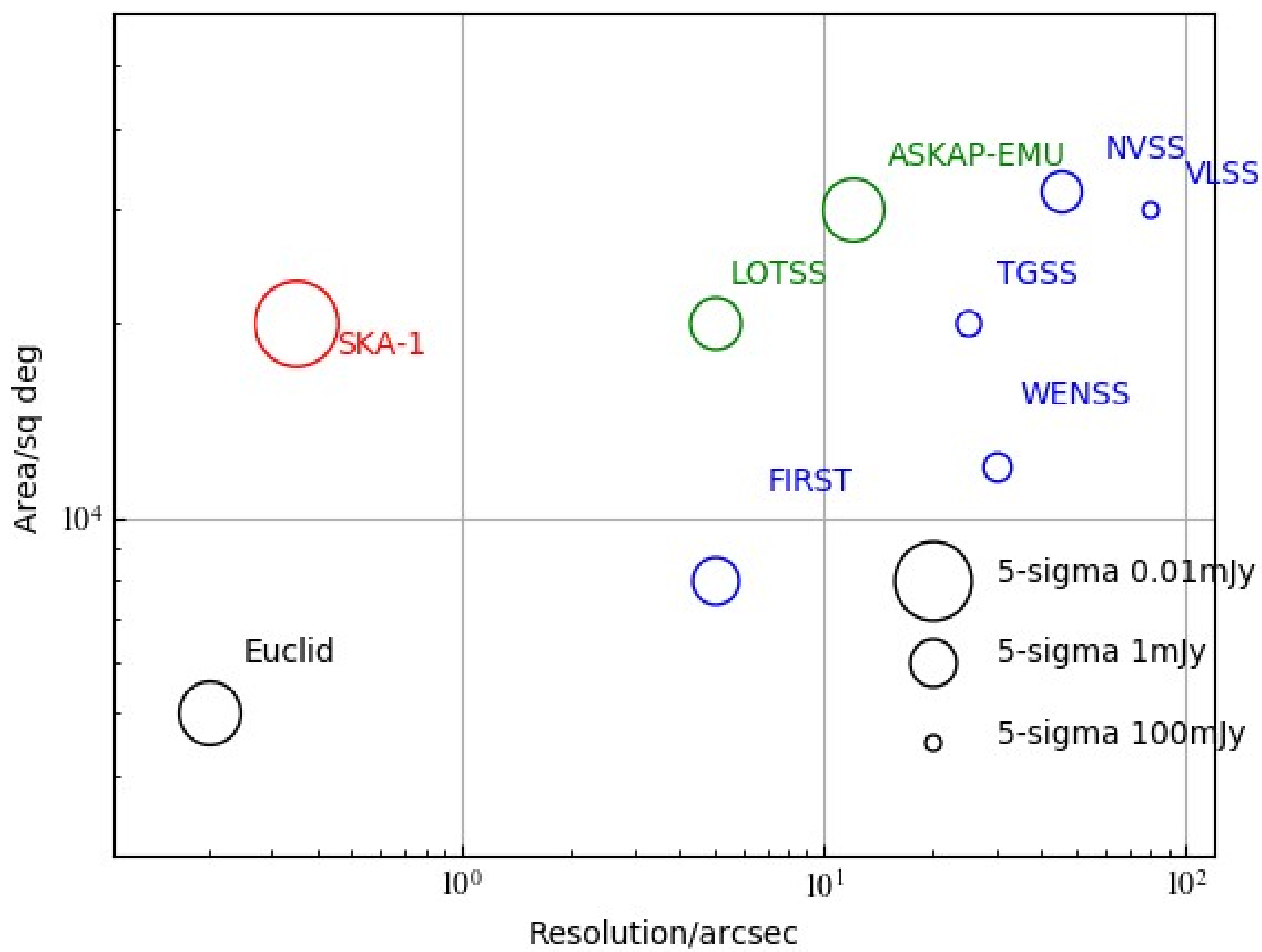
MANCHESTER 1824

Long Baselines at Low Frequencies: high-resolution surveys with LOFAR

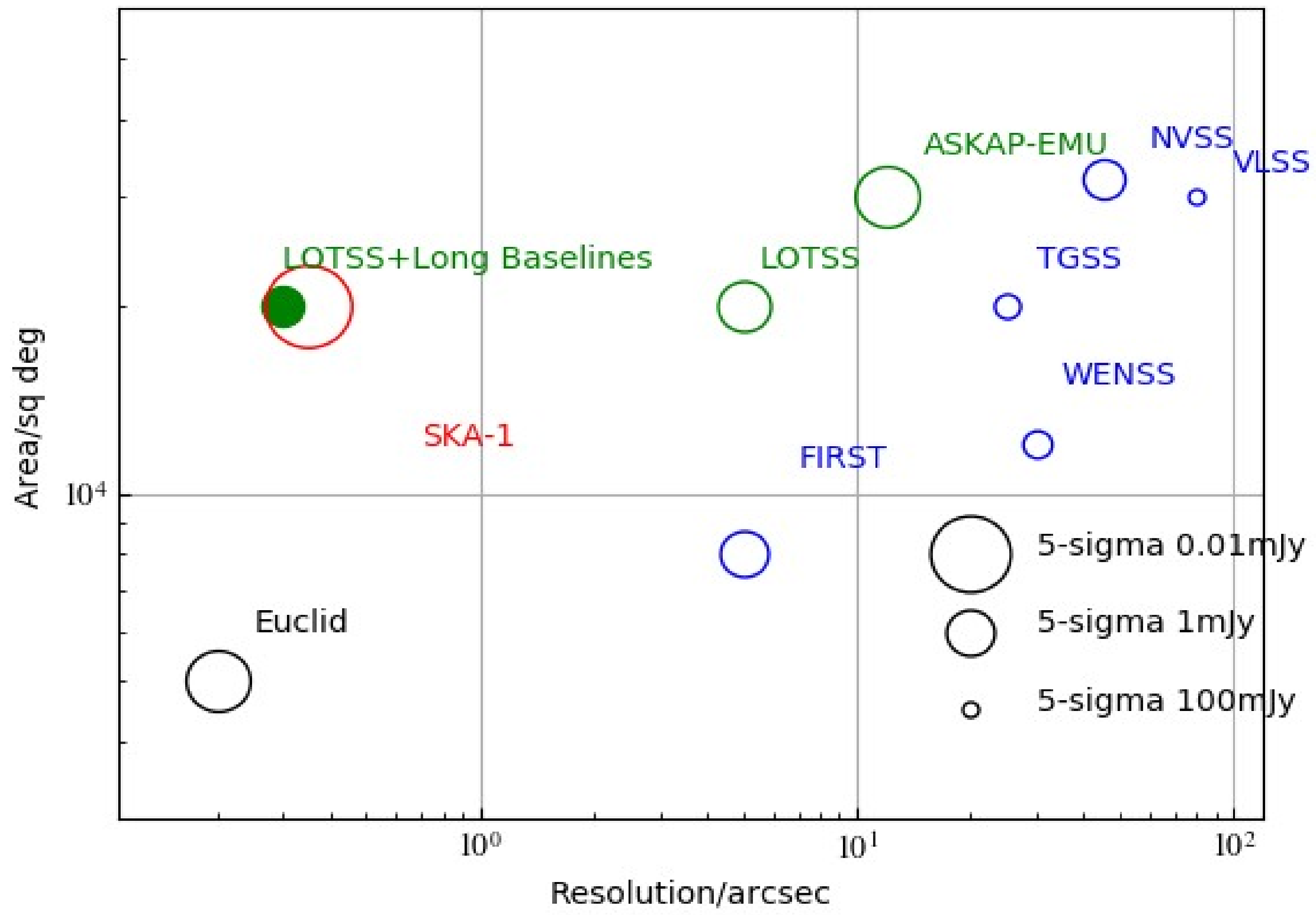




Conventional survey space: depth vs frequency
Traditional 'wedding-cake' structure: sacrifice area for depth
Current, in-progress and future surveys separated by factor ~10 in depth
Combination of depth and multiple frequencies gives science on individual objects



Alternative 'survey pixel volume' space
Higher resolution makes huge areas very difficult (even for SKA)
Relevant survey space for e.g. gravitational lens surveying, parallel studies of many objects



Alternative 'survey pixel volume' space
Higher resolution makes huge areas very difficult (even for SKA)
Relevant survey space for e.g. gravitational lens surveying, parallel studies of many objects
Extension of LOTSS to long baselines can be a leading survey in this space (even with SKA1)

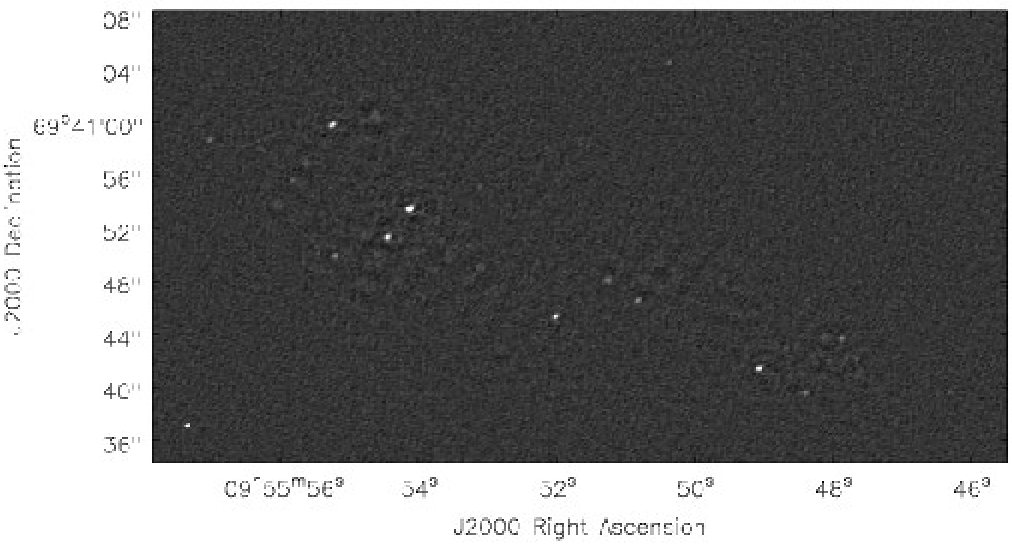
LOFAR LB survey gets you a wide-area survey of sources

At sub-galactic resolution for most z

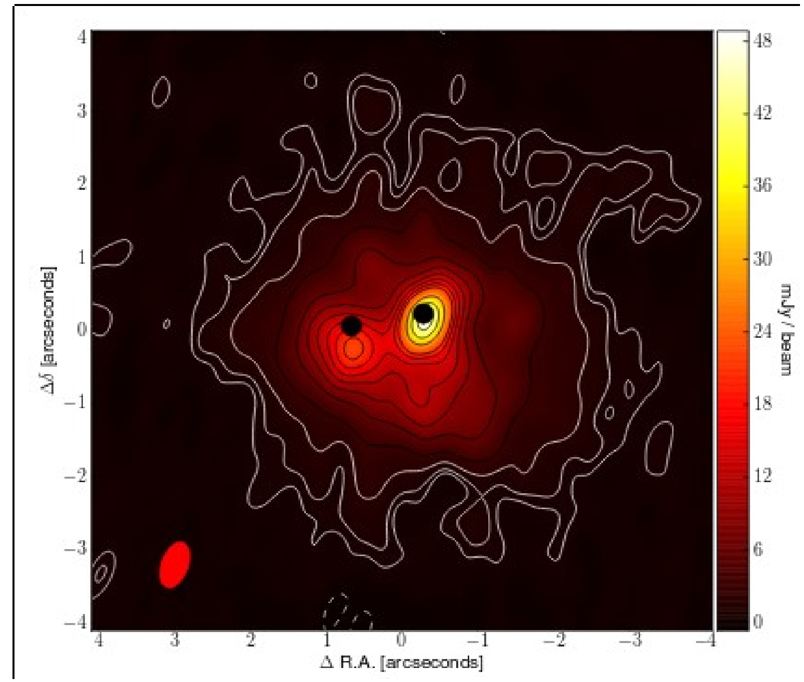
To highly competitive (though <SKA-1) sensitivity for steep-spectrum sources (but with complementary science, and earlier)

And 1-1.2 times resolution of SKA-1

Science with the long baselines: star-forming galaxies



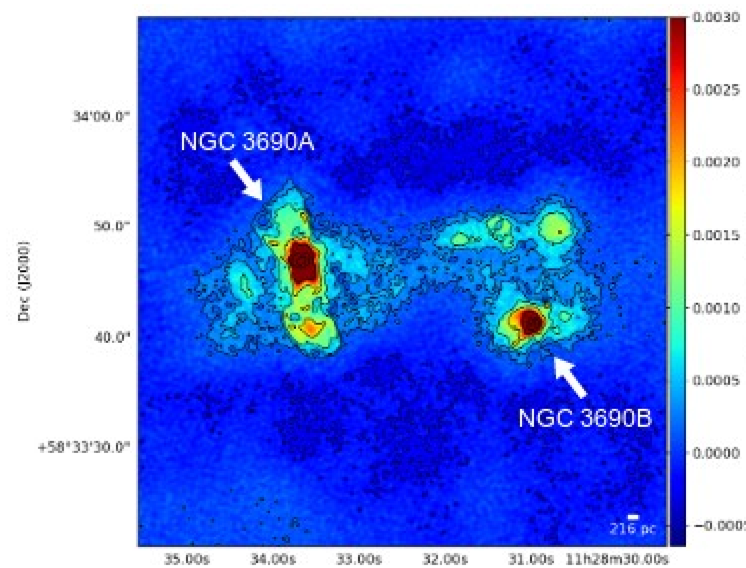
M82 (Varenius et al. 2014) – LOFAR 150MHz



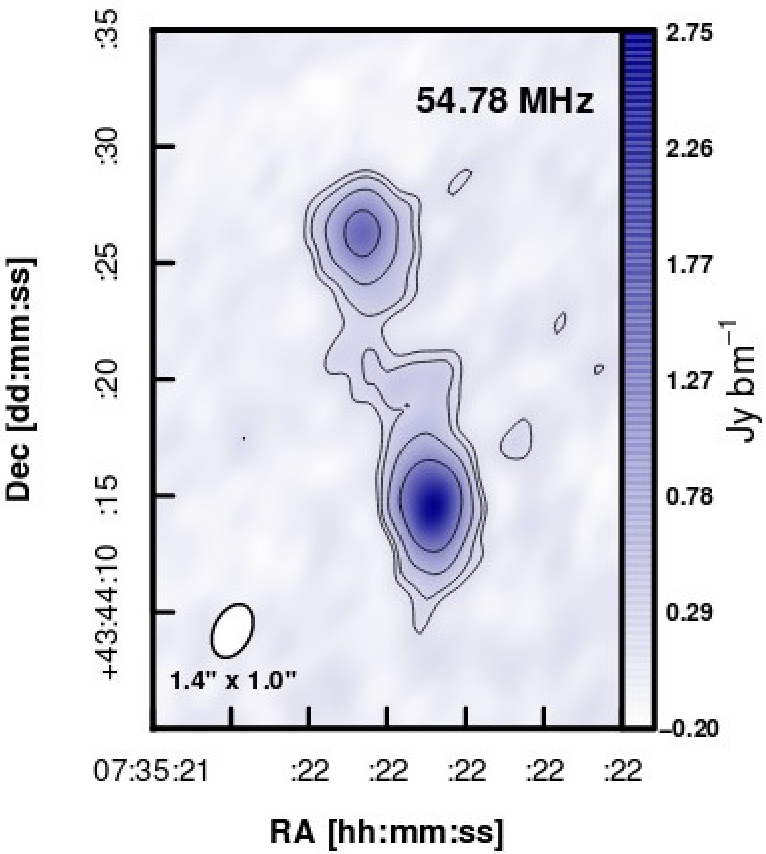
Arp220 (Varenius et al. 2016)

Subarcsecond resolution crucial for SF regions
Extended steep-spectrum emission from outflows visible at low frequencies
Spectral indices – SNRs vs HII regions

Arp299 (Ramirez-Olivencia 2017 in prep)
See poster by Varenius et al.



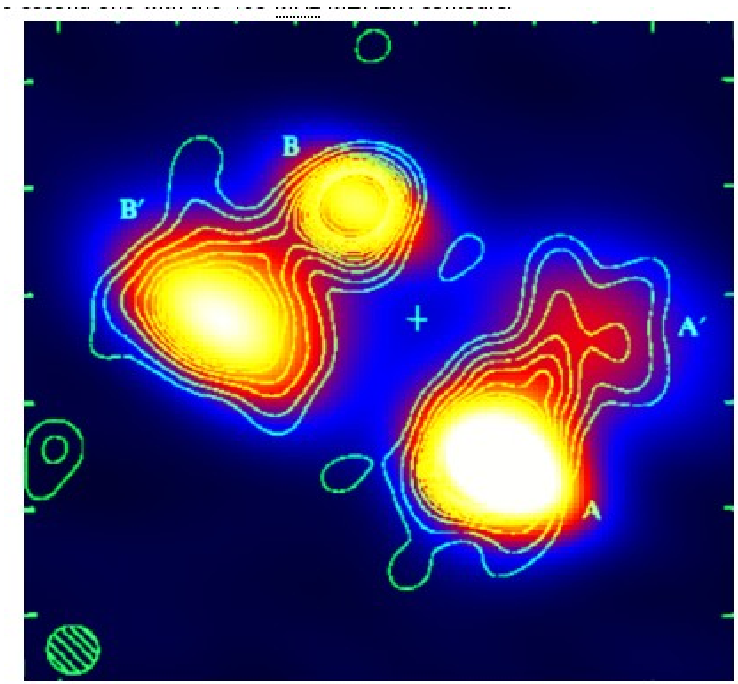
Science with the long baselines: radio galaxies



Morabito et al. 2016 (LBA! 50MHz)

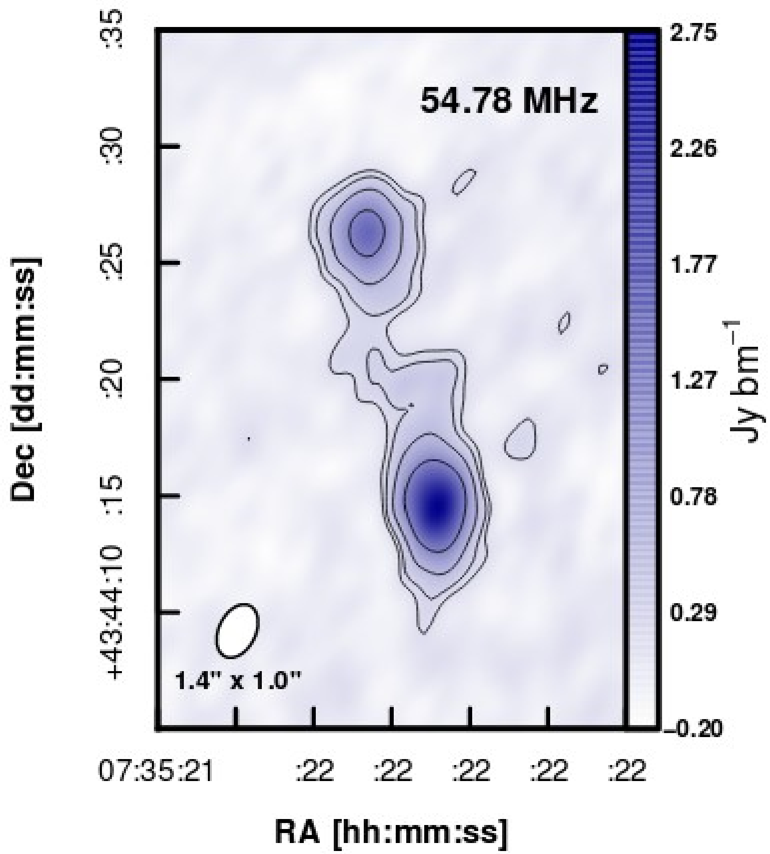
Steep-spectrum emission:

- nature of USS radio galaxies
- spectral indices/physics of extended radio emission



3C196 (Wucknitz 2012)

Science with the long baselines: radio galaxies



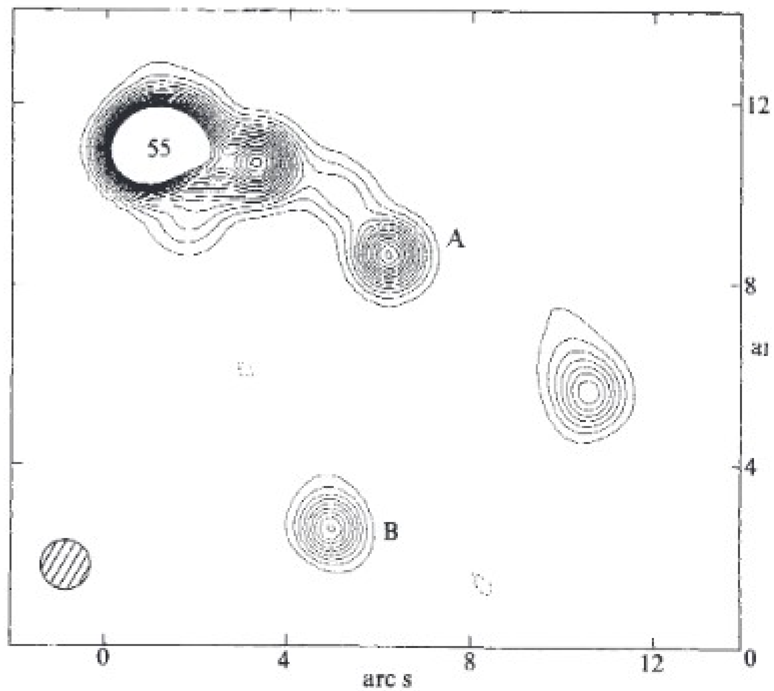
Steep-spectrum emission:

- nature of USS radio galaxies
- spectral indices/physics of extended radio emission

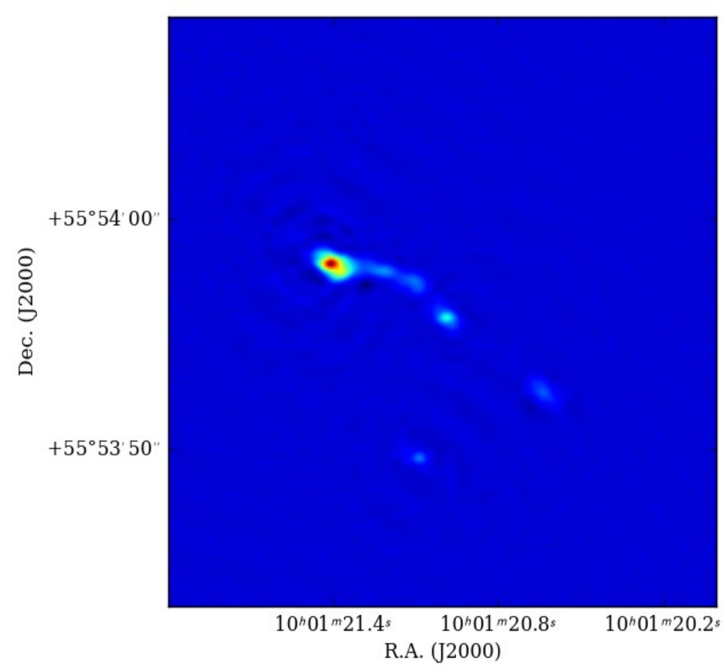
Morabito et al. 2016 (LBA! 50MHz)

Technically very difficult – extreme ionosphere/calibration problems
But seriously unexplored part of parameter space at high resolution

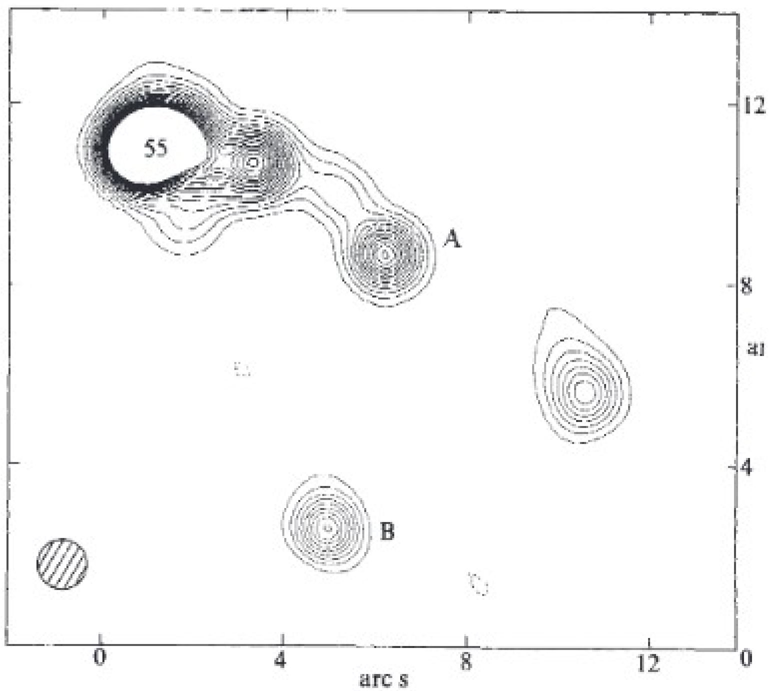
Science with the long baselines: gravitational lenses



Q0957+561 (Noble et al. 1988, 408MHz)
LOFAR LB (Hartley et al. in prep, 151 MHz)



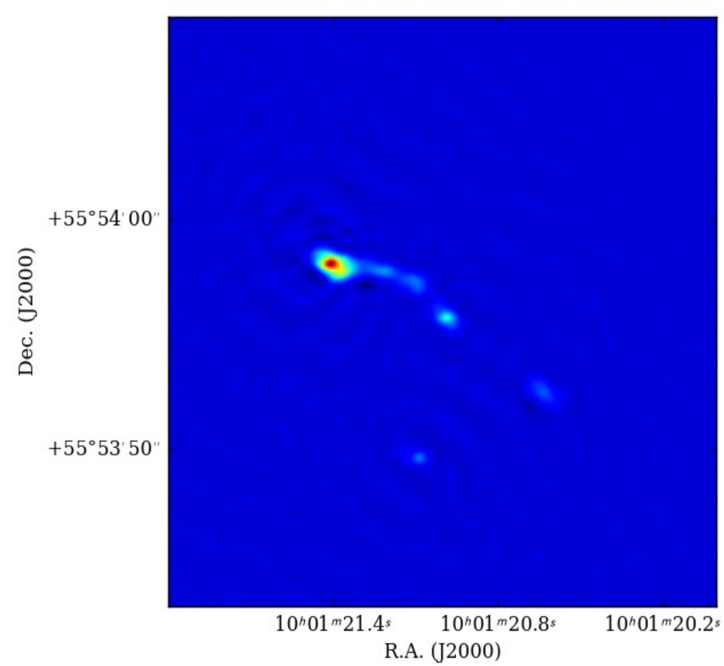
Science with the long baselines: gravitational lenses



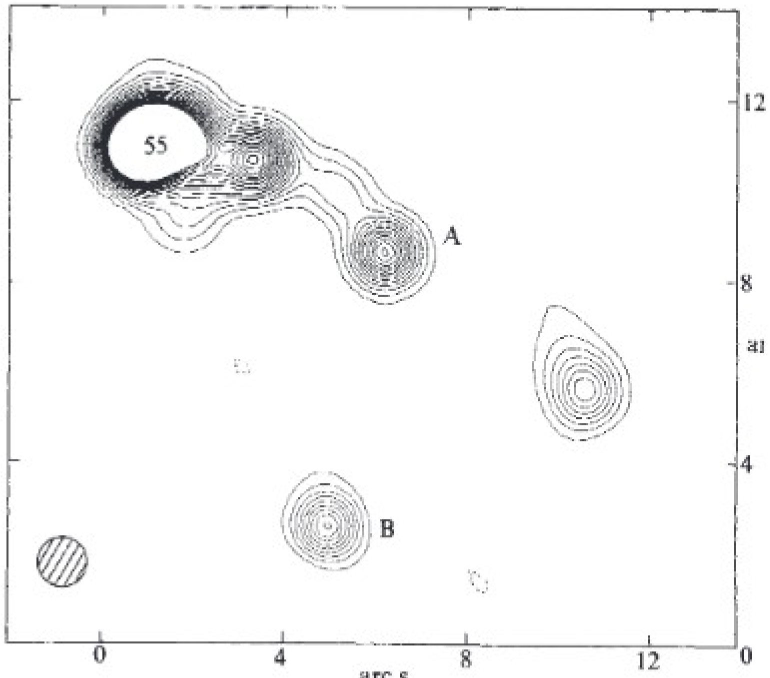
Q0957+561 (Noble et al. 1988, 408MHz)
LOFAR LB (Hartley et al. in prep, 151 MHz)

Science from comparison of 2 sightlines, 1 object

- foreground effects / scattering in lens elliptical
- potential for surveys with rare objects



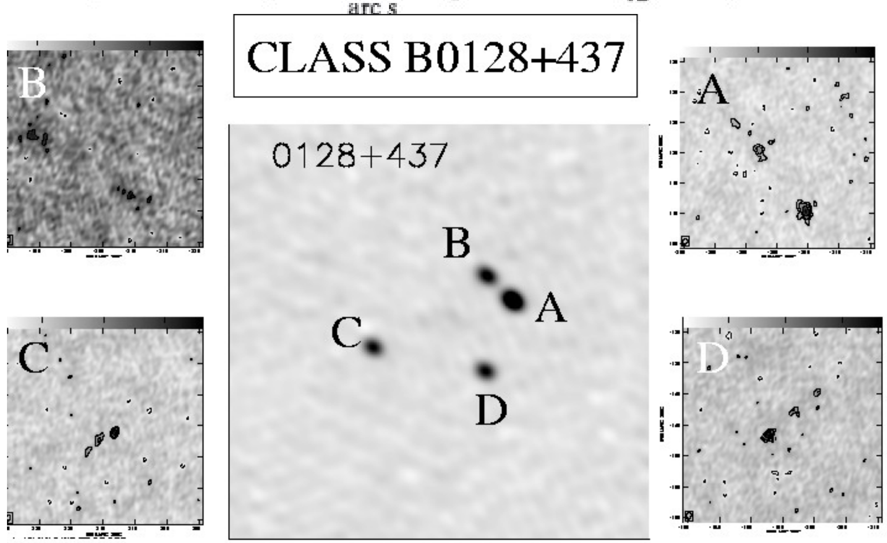
Science with the long baselines: gravitational lenses



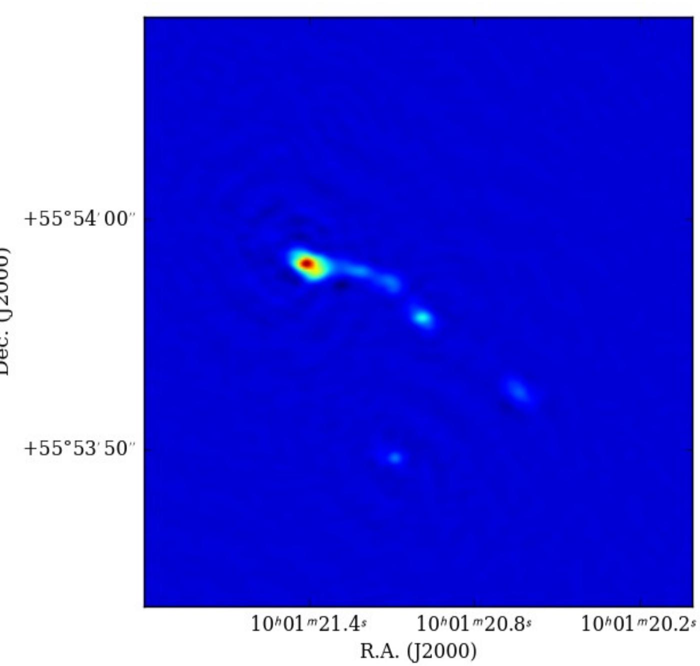
Q0957+561 (Noble et al. 1988, 408MHz)
LOFAR LB (Hartley et al. in prep, 151 MHz)

Science from comparison of 2 sightlines, 1 object

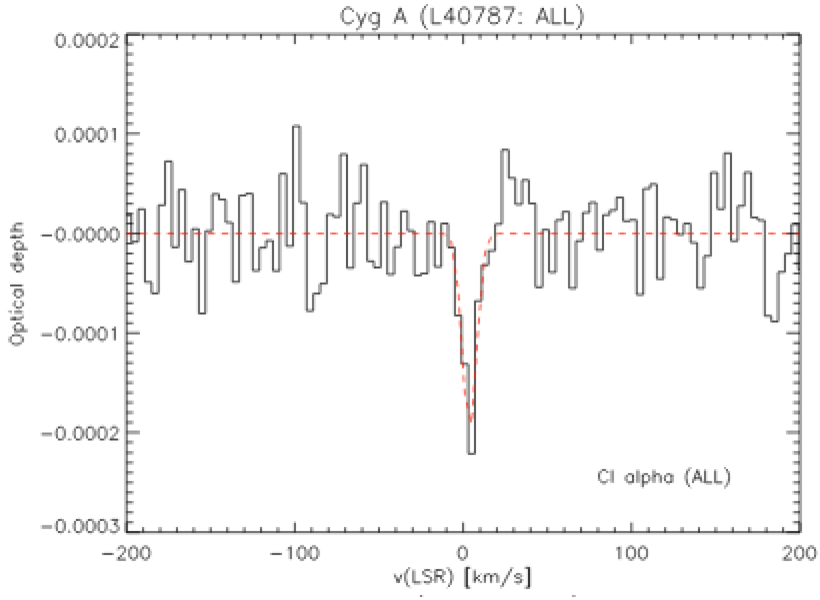
- foreground effects / scattering in lens elliptical
- potential for surveys with rare objects



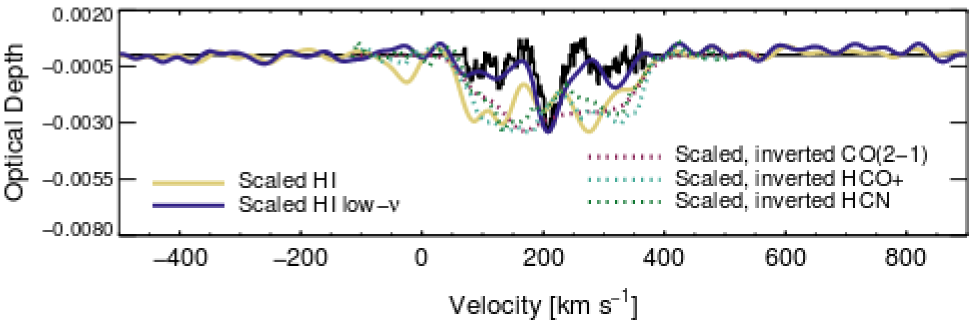
MERLIN full-track+VLBA 2hrs (all at 5GHz)



Science with long baselines: radio recombination lines



CRRL Cyg A (Oonk et al. 2014)



CRRL M82 (Morabito et al. 2014)

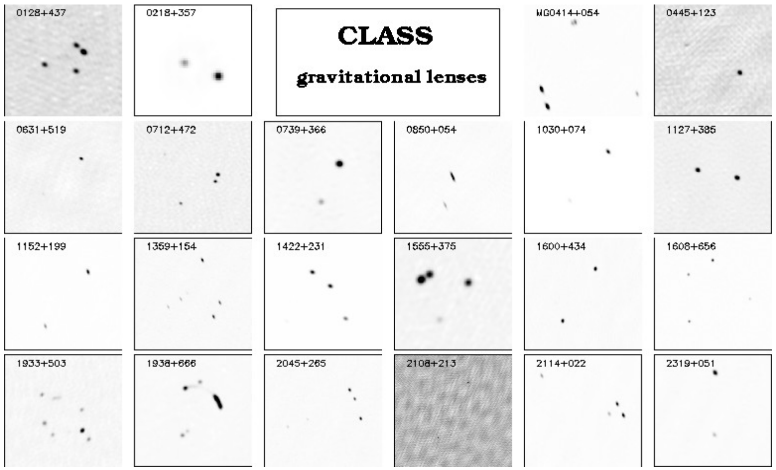
Studies of cold neutral gas → geometry, physics of CNM
Current work with Dutch array
With LB, spatially resolved absorption (cf. current HI)

Important to go from single-object science to large samples

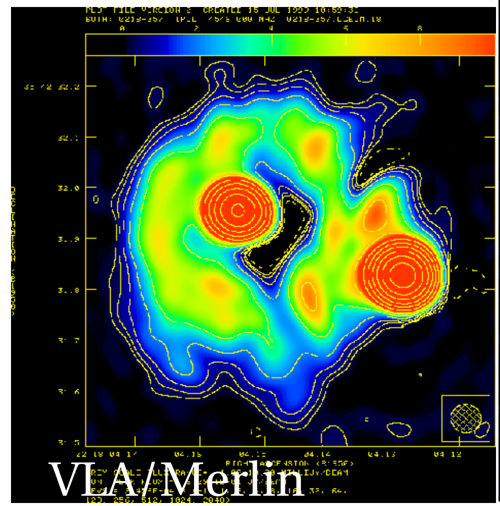
- Discovery of rare objects (e.g. strong lensing surveys)
- Census of AGN at high resolution (census of BH, resolution important for compact sources)
- Important to have large samples for statistics of radio source evolution

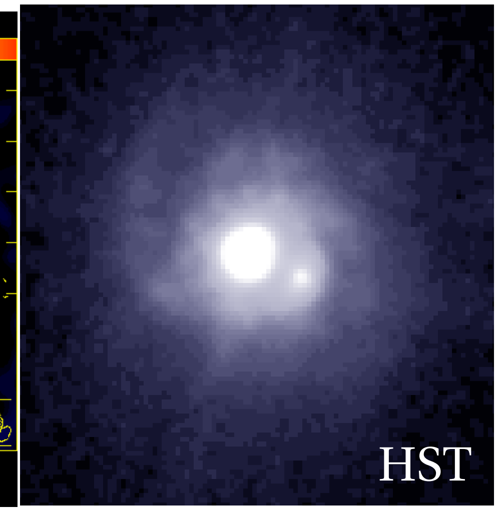
Strong gravitational lens surveys

- Important for studies of dark matter/substructures
- Radio lenses especially useful for substructure with VLBI/central images
- Discovery is difficult – objects are rare

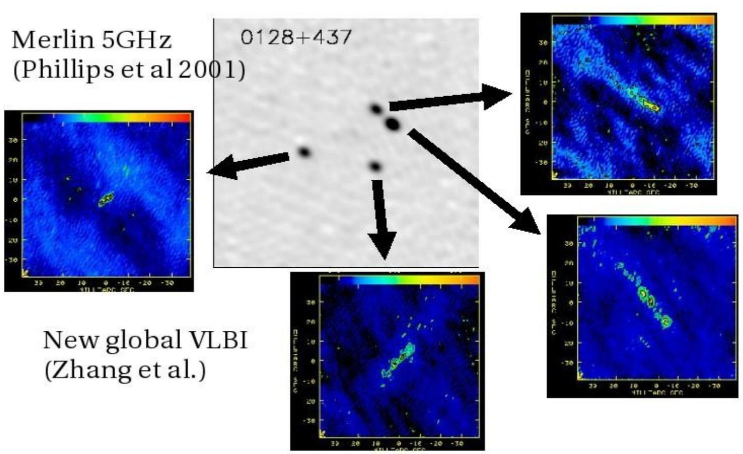


Samples still dominated by 1990's surveys (MG/CLASS – total of about 40)



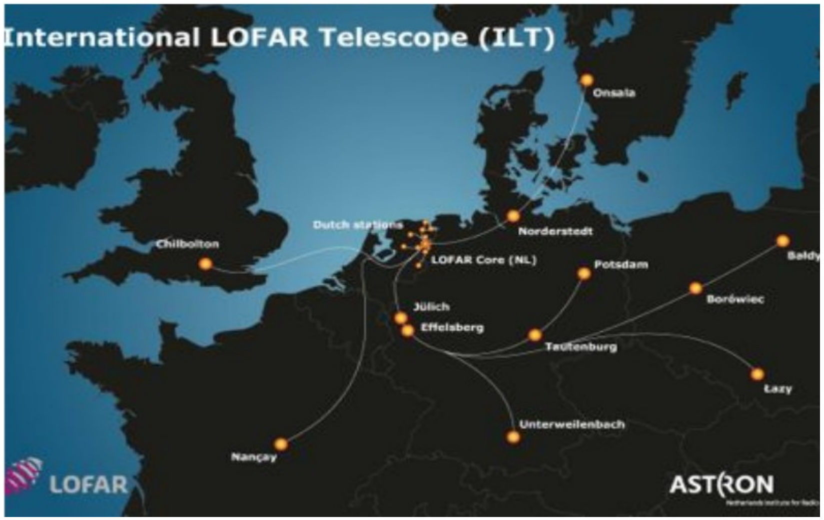


Science example: CLASS0218+357 (H0 measurement/mass models)



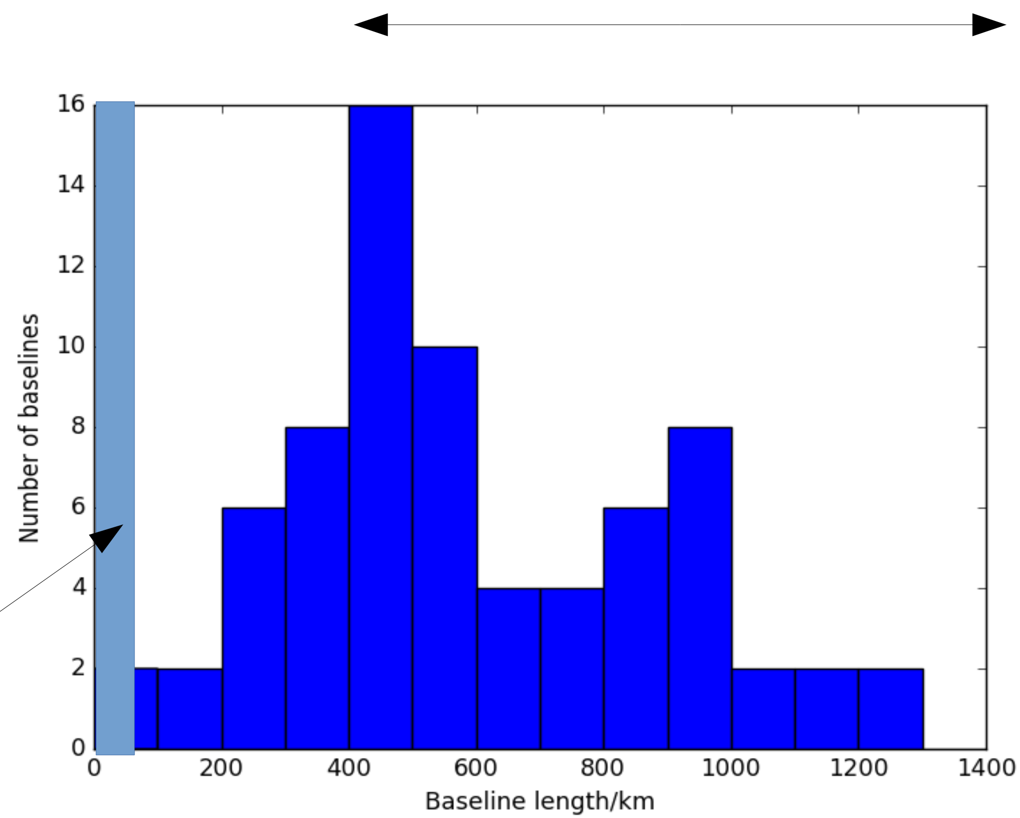
0128+437 (substructure/foregrounds)

Technical challenges: imaging/uv coverage



u-v coverage: 80-300km baselines vital

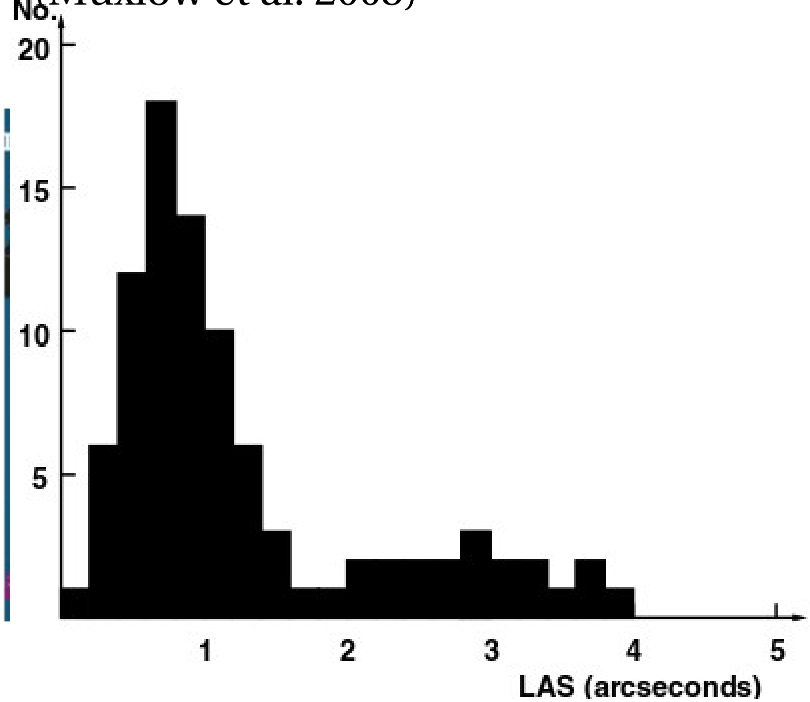
PL/IR baselines



NL baselines

Technical challenges: imaging/uv coverage

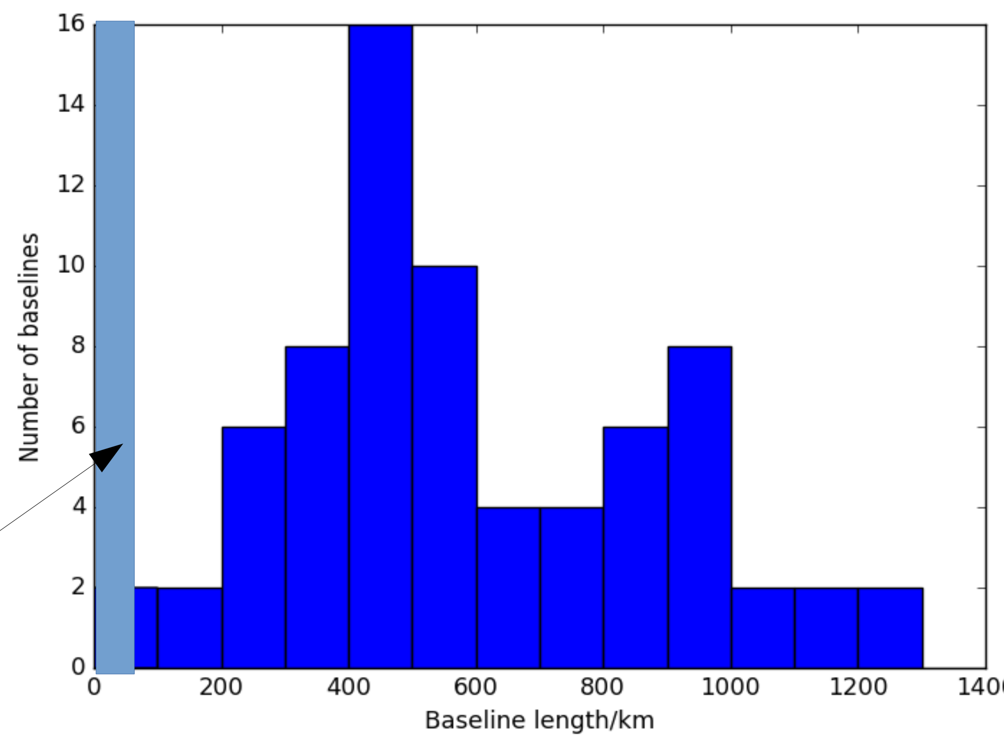
LargAngSizes of faint GOODS-N radio sources (Muxlow et al. 2005)



u-v coverage: 80-300km baselines vital

Sources have much structure on 1” scale (150-200km); reproducing this faithfully v. important for calibration
Currently heavily reliant on a few stations for this (DE601,605,609)

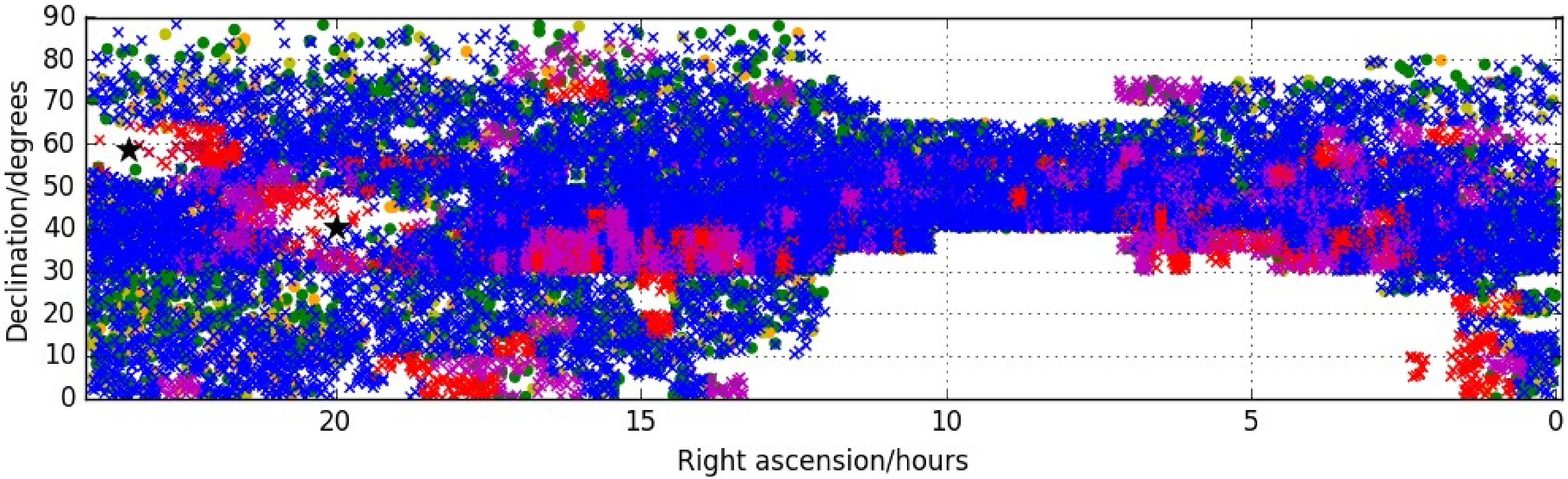
PL/IR baselines



NL baselines

Technical challenges – calibration

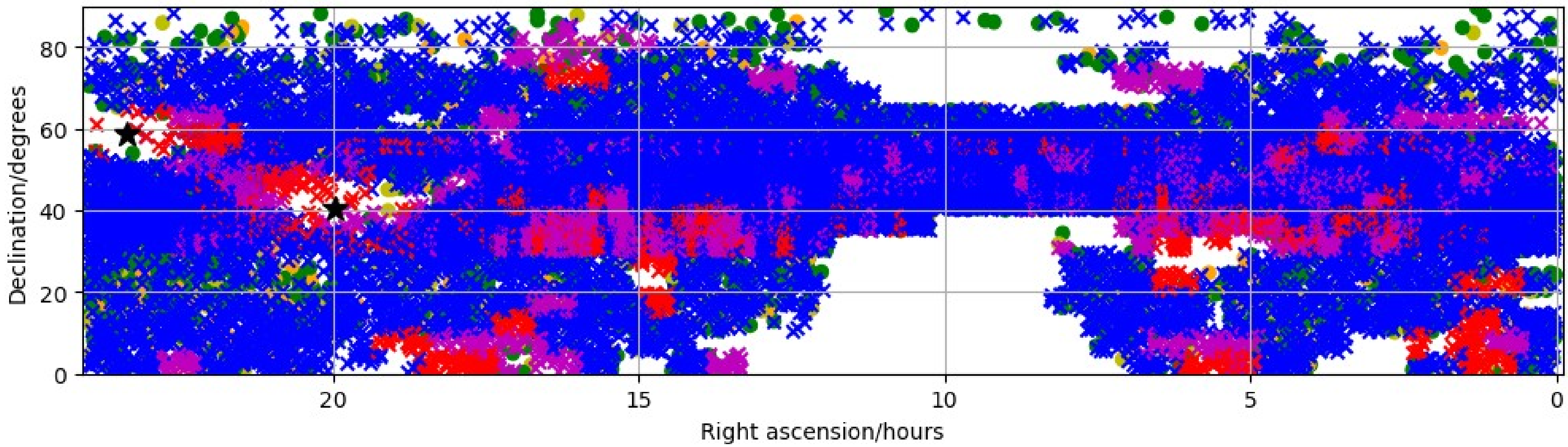
- Calibrator sources necessary for ionospheric phase calibration
- More compact sources for long-baseline calibration are rarer
- LBCS survey under way (LBWG/Jackson et al. 2016)



Current status of LBCS calibrator survey (early 2017)

Technical challenges – calibration

- Calibrator sources necessary for ionospheric phase calibration
- More compact sources for long-baseline calibration are rarer
- LBCS survey under way (LBWG/Jackson et al. 2016)

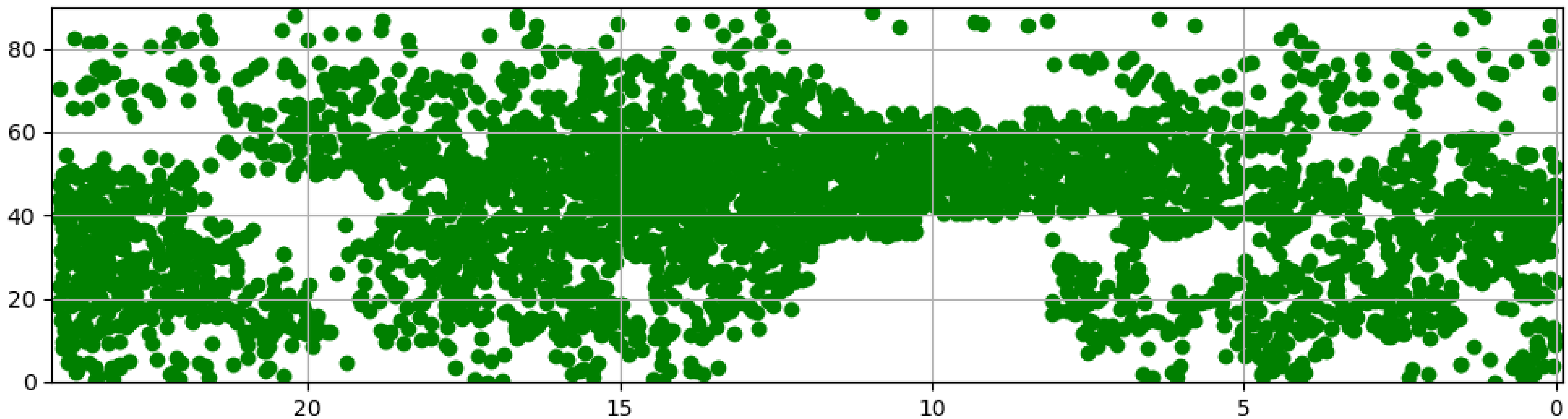


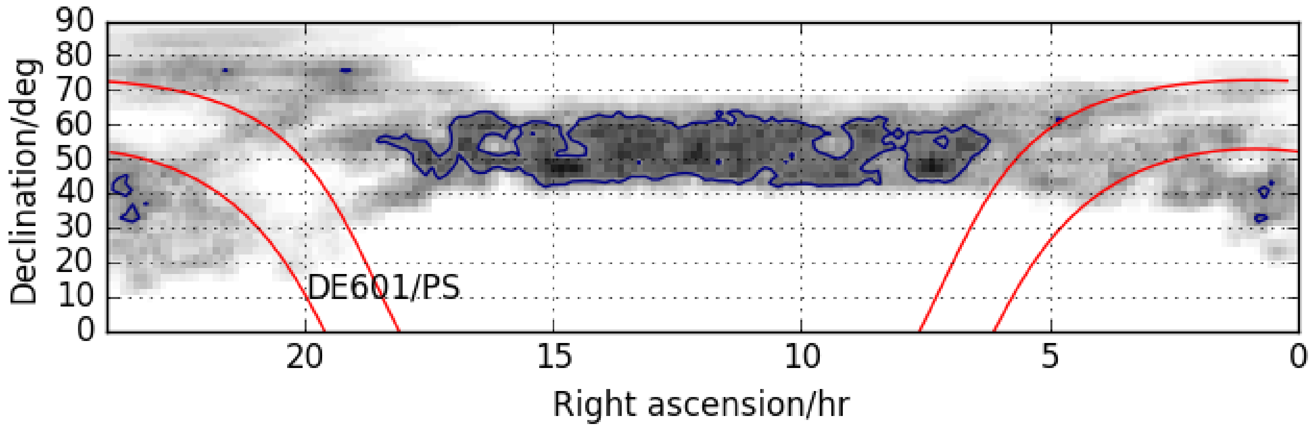
Current status of LBCS calibrator survey (mid 2017)

Calibrator list available on vo.astron.nl
Updated as-it-happens version on http://www.jb.man.ac.uk/~njj/lbcs_stats.sum

Technical challenges – calibration

- Calibrator sources necessary for ionospheric phase calibration
- More compact sources for long-baseline calibration are rarer
- LBCS survey under way (LBWG/Jackson et al. 2016)

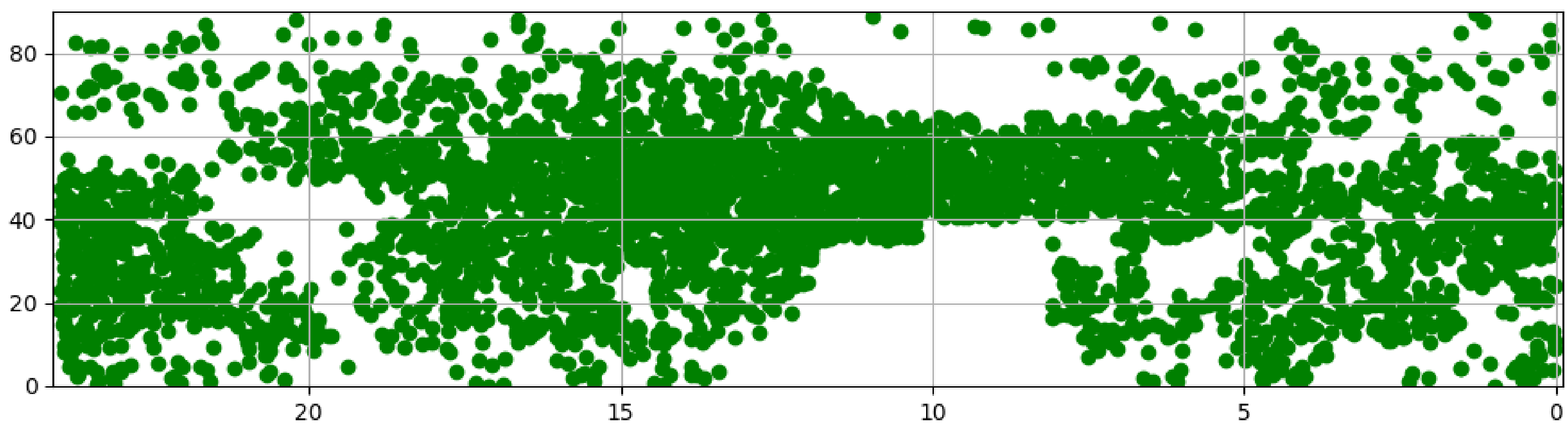


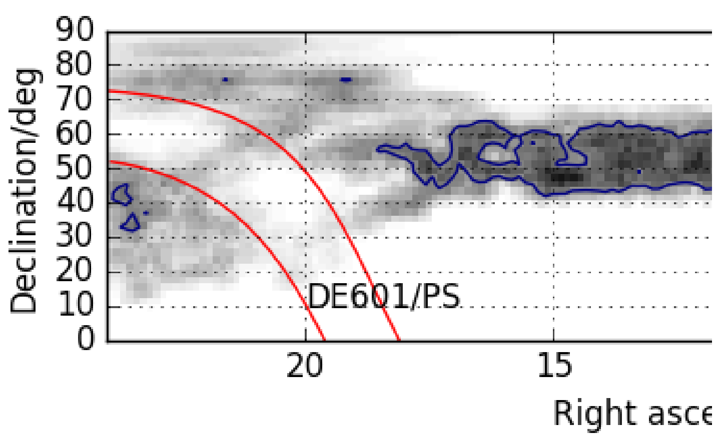


200-km baselines
1 source/sq deg

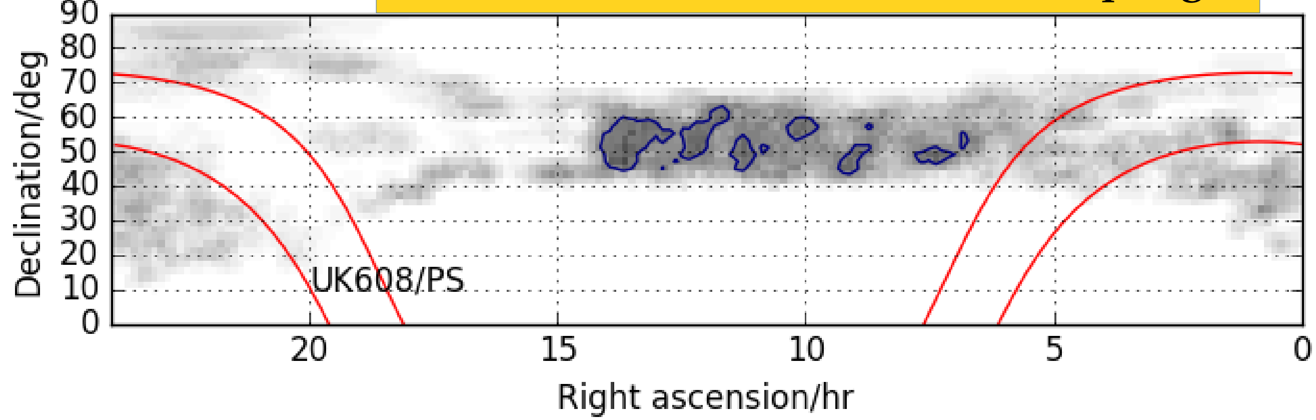
Technical challenges – calibration

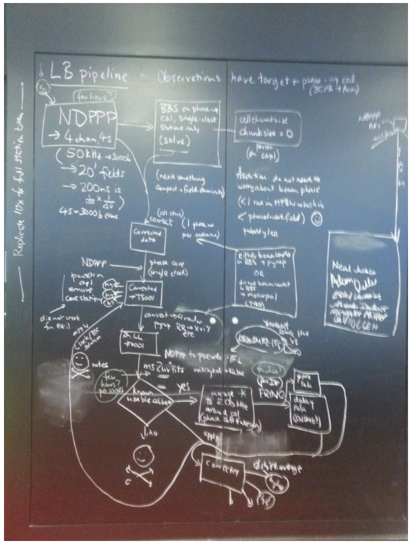
- Calibrator sources necessary for ionospheric phase calibration
- More compact sources for long-baseline calibration are rarer
- LBCS survey under way (LBWG/Jackson et al. 2016)





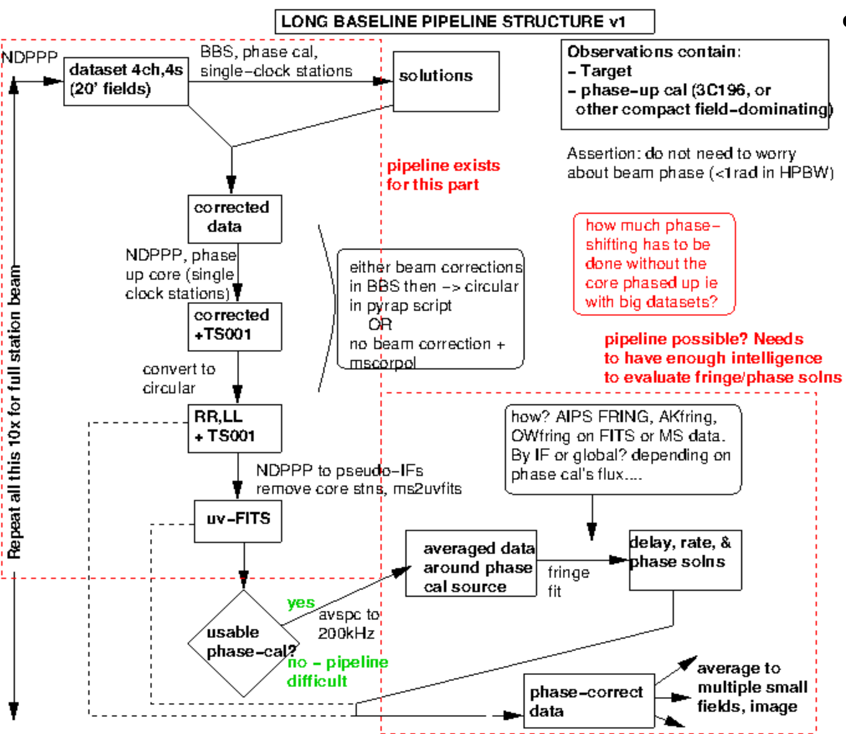
800-km baselines - 0.5 source/sq deg





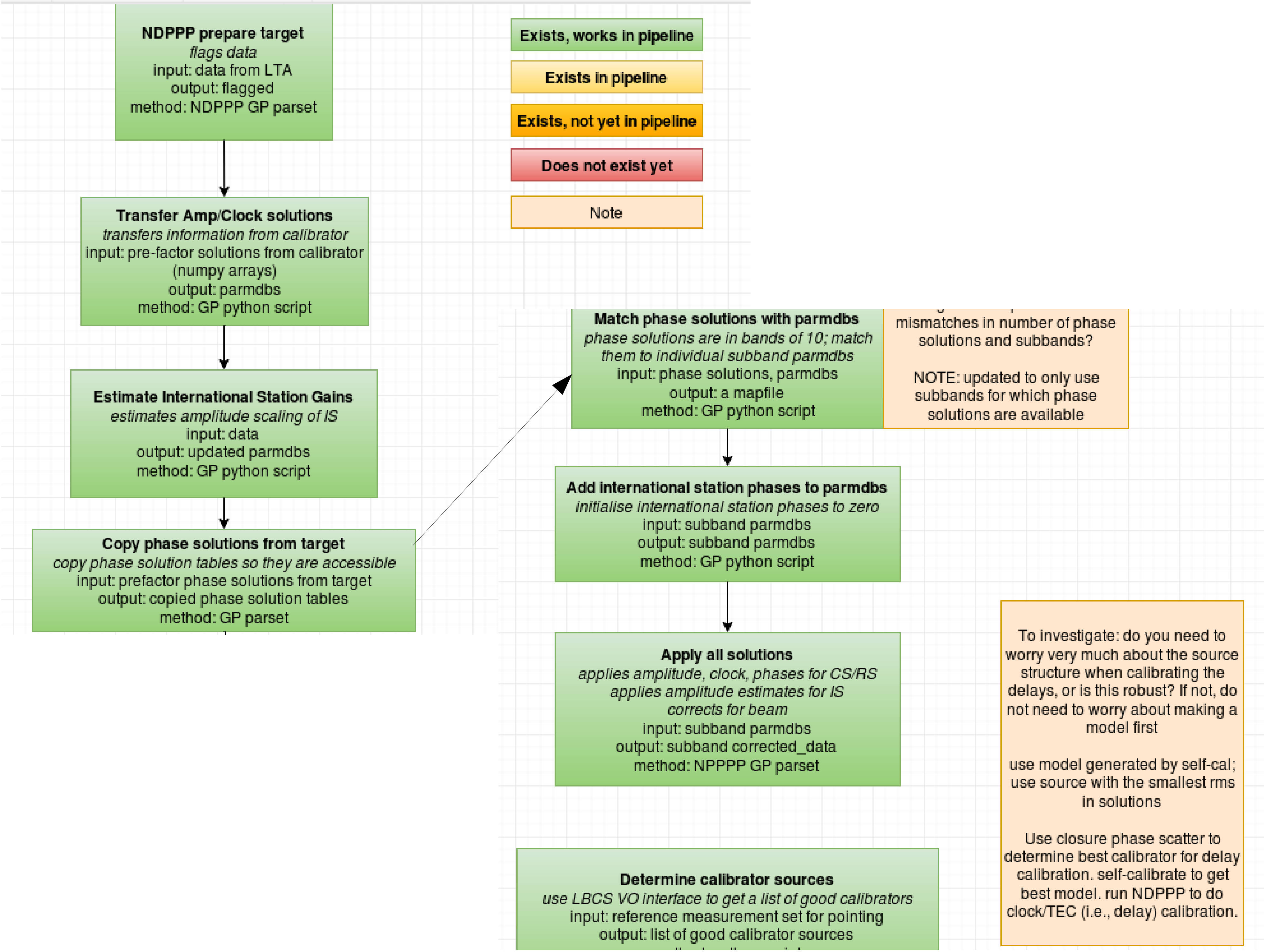
Current technical capabilities: pipeline processing

- Pre-processing – shared with surveys
- LBWG pipeline – superterp formation, circular conversio
- Delay calibration – manual, being pipelined
- Shift/average to sources – being pipelined
- Phase calibration/imaging – being pipelined

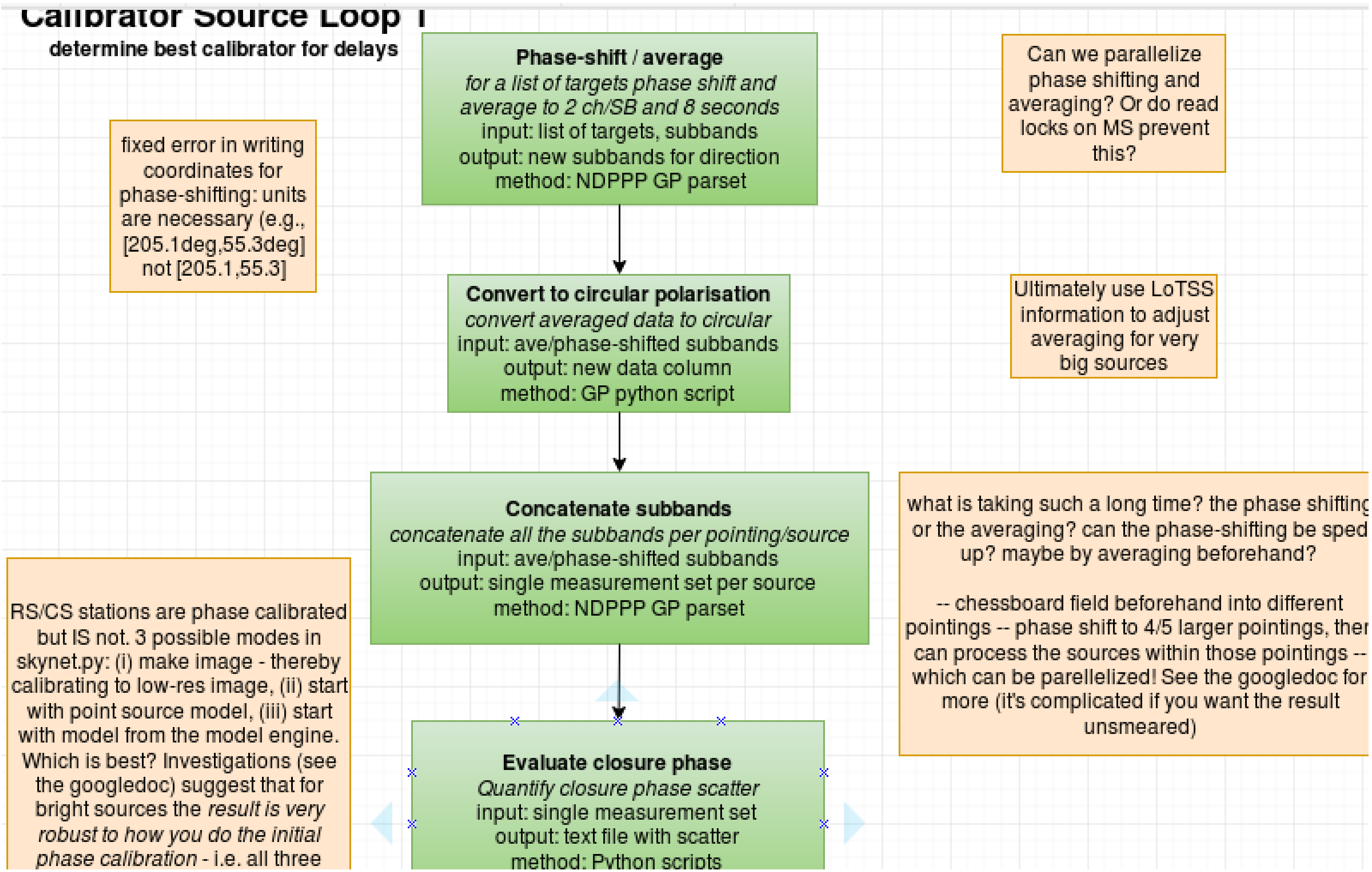


Limitations

- Storage, processing power (IB datasets 50Gb/sb 16ch/1s
- Ionospheric calibration
- Unclear how faint we can go
- Close to thermal noise demonstrated however

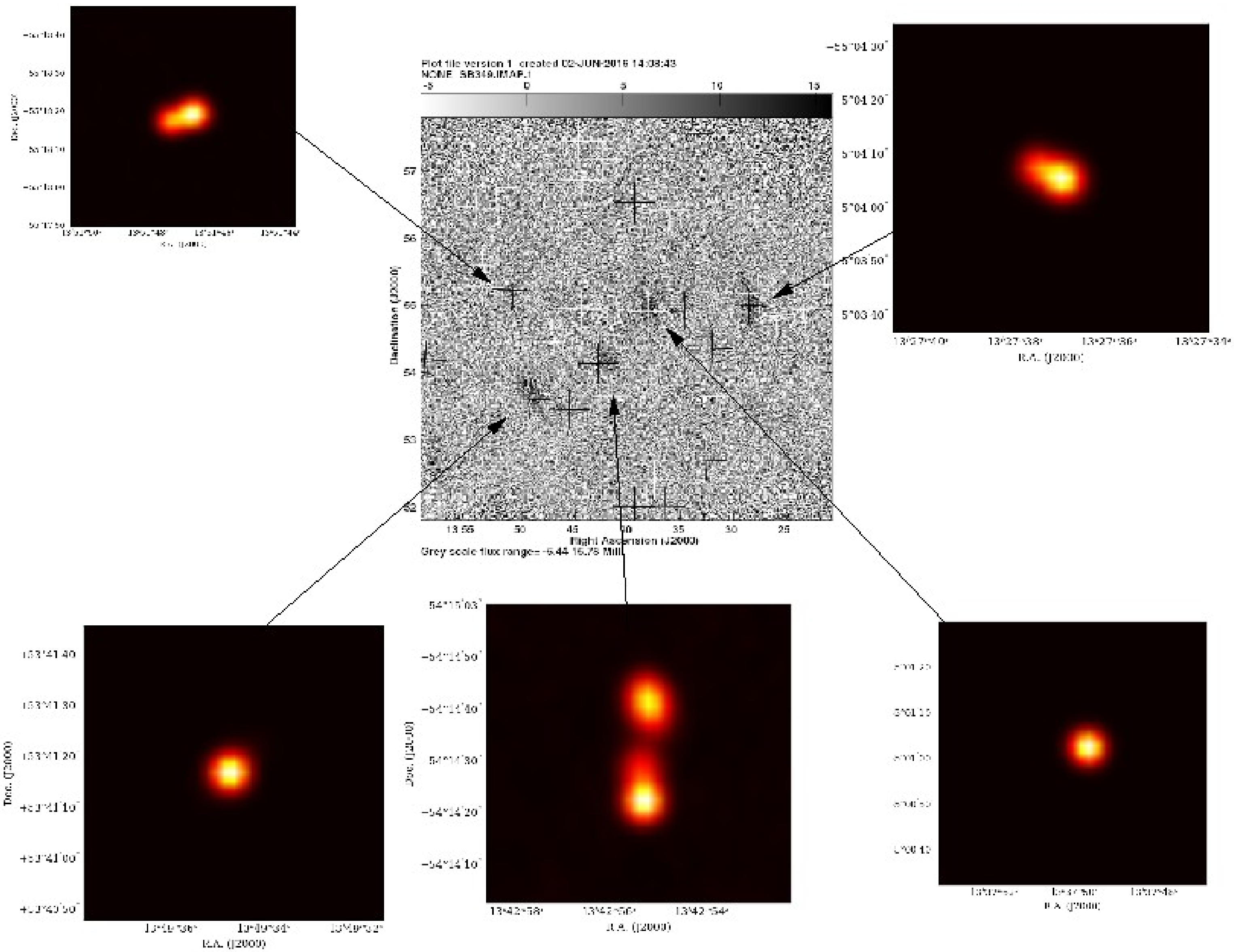


NDPPP prepare target
flags data
input: data from LTA
output: flagged
method: NDPPP GP parset
Exists, works in pipeline
Exists in pipeline
Exists, not yet in pipeline
Does not exist yet
Note
Transfer Amp/Clock solutions
transfers information from calibrator
input: pre-factor solutions from calibrator (numpy arrays)
output: parmdbs
method: GP python script
Estimate International Station Gains
estimates amplitude scaling of IS
input: data
output: updated parmdbs
method: GP python script
Copy phase solutions from target
copy phase solution tables so they are accessible
input: prefactor phase solutions from target
output: copied phase solution tables
method: GP parset
Match phase solutions with parmdbs
phase solutions are in bands of 10; match them to individual subband parmdbs
input: phase solutions, parmdbs
output: a mapfile
method: GP python script
mismatches in number of phase solutions and subbands?
NOTE: updated to only use subbands for which phase solutions are available
Add international station phases to parmdbs
initialise international station phases to zero
input: subband parmdbs
output: subband parmdbs
method: GP python script
Apply all solutions
applies amplitude, clock, phases for CS/RS
applies amplitude estimates for IS
corrects for beam
input: subband parmdbs
output: subband corrected_data
method: NPPPP GP parset
To investigate: do you need to worry very much about the source structure when calibrating the delays, or is this robust? If not, do not need to worry about making a model first
use model generated by self-cal; use source with the smallest rms in solutions
Use closure phase scatter to determine best calibrator for delay calibration. self-calibrate to get best model. run NDPPP to do clock/TEC (i.e., delay) calibration.
Determine calibrator sources
use LBCS VO interface to get a list of good calibrators
input: reference measurement set for pointing
output: list of good calibrator sources



Calibrator Source Loop 1
determine best calibrator for delays
Phase-shift / average
for a list of targets phase shift and average to 2 ch/SB and 8 seconds
input: list of targets, subbands
output: new subbands for direction
method: NDPPP GP parset
Can we parallelize phase shifting and averaging? Or do read locks on MS prevent this?
fixed error in writing coordinates for phase-shifting: units are necessary (e.g., [205.1deg,55.3deg] not [205.1,55.3]
Convert to circular polarisation
convert averaged data to circular
input: ave/phase-shifted subbands
output: new data column
method: GP python script
Ultimately use LoTSS information to adjust averaging for very big sources
Concatenate subbands
concatenate all the subbands per pointing/source
input: ave/phase-shifted subbands
output: single measurement set per source
method: NDPPP GP parset
what is taking such a long time? the phase shifting or the averaging? can the phase-shifting be sped up? maybe by averaging beforehand?
-- chessboard field beforehand into different pointings -- phase shift to 4/5 larger pointings, then can process the sources within those pointings -- which can be parellelized! See the googledoc for more (it's complicated if you want the result unsmeared)
RS/CS stations are phase calibrated but IS not. 3 possible modes in skynet.py: (i) make image - thereby calibrating to low-res image, (ii) start with point source model, (iii) start with model from the model engine. Which is best? Investigations (see the googledoc) suggest that for bright sources the result is very robust to how you do the initial phase calibration - i.e. all three
Evaluate closure phase
Quantify closure phase scatter
input: single measurement set
output: text file with scatter
method: Python scripts

Current technical capabilities – illustration on LBWG surveys field



Greyscale: fringe-rate/delay map on ST001 – DE605 baseline
Cutouts: FIRST images of some of the brighter sources in the field

Aside: fringe-frequency/delay mapping

$$\phi = \frac{2\pi L}{\lambda} \left(\sin d \sin D + \cos d \cos D \cos(H - h)\right), \qquad (2)$$

where L is the baseline length and λ the observing wavelength.

We can write the time and frequency derivatives of ϕ as functions of sky coordinates, expressed as offsets ΔH and ΔD, as

$$\begin{pmatrix} \partial\phi/\partial t \\ \partial\phi/\partial f \end{pmatrix} = \begin{pmatrix} \frac{1}{\cos D}\frac{\partial^2\phi}{\partial H \partial t} & \frac{\partial^2\phi}{\partial D \partial t} \\ \frac{1}{\cos D}\frac{\partial^2\phi}{\partial H \partial f} & \frac{\partial^2\phi}{\partial D \partial f} \end{pmatrix} \begin{pmatrix} \Delta H \\ \Delta D \end{pmatrix}, \qquad (3)$$

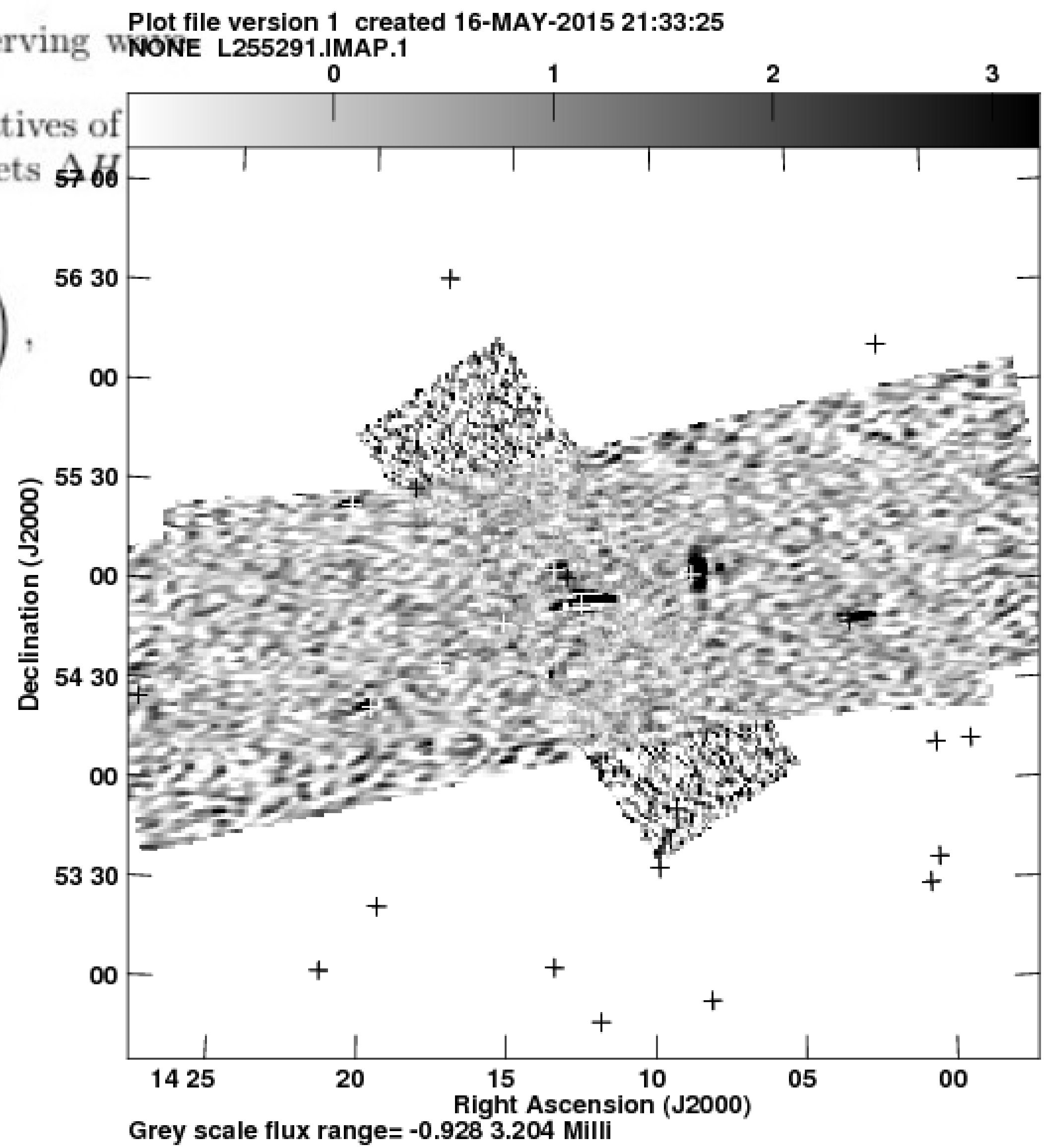
Aside: fringe-frequency/delay mapping

$$\phi = \frac{2\pi L}{\lambda}\left(\sin d \sin D + \cos d \cos D \cos(H - h)\right), \qquad (2)$$

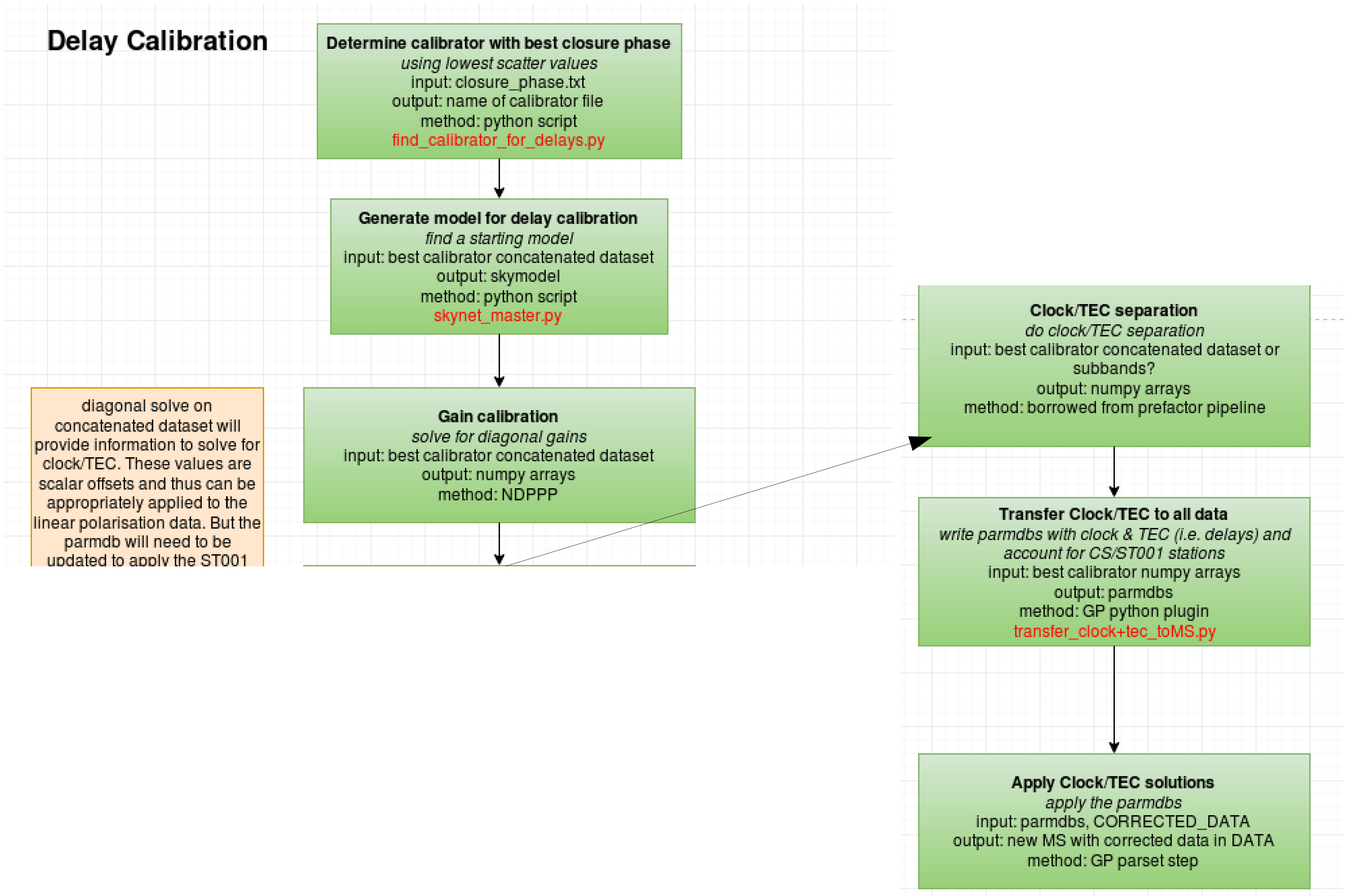
where L is the baseline length and λ the observing w
length.

We can write the time and frequency derivatives of functions of sky coordinates, expressed as offsets ΔH ΔD, as

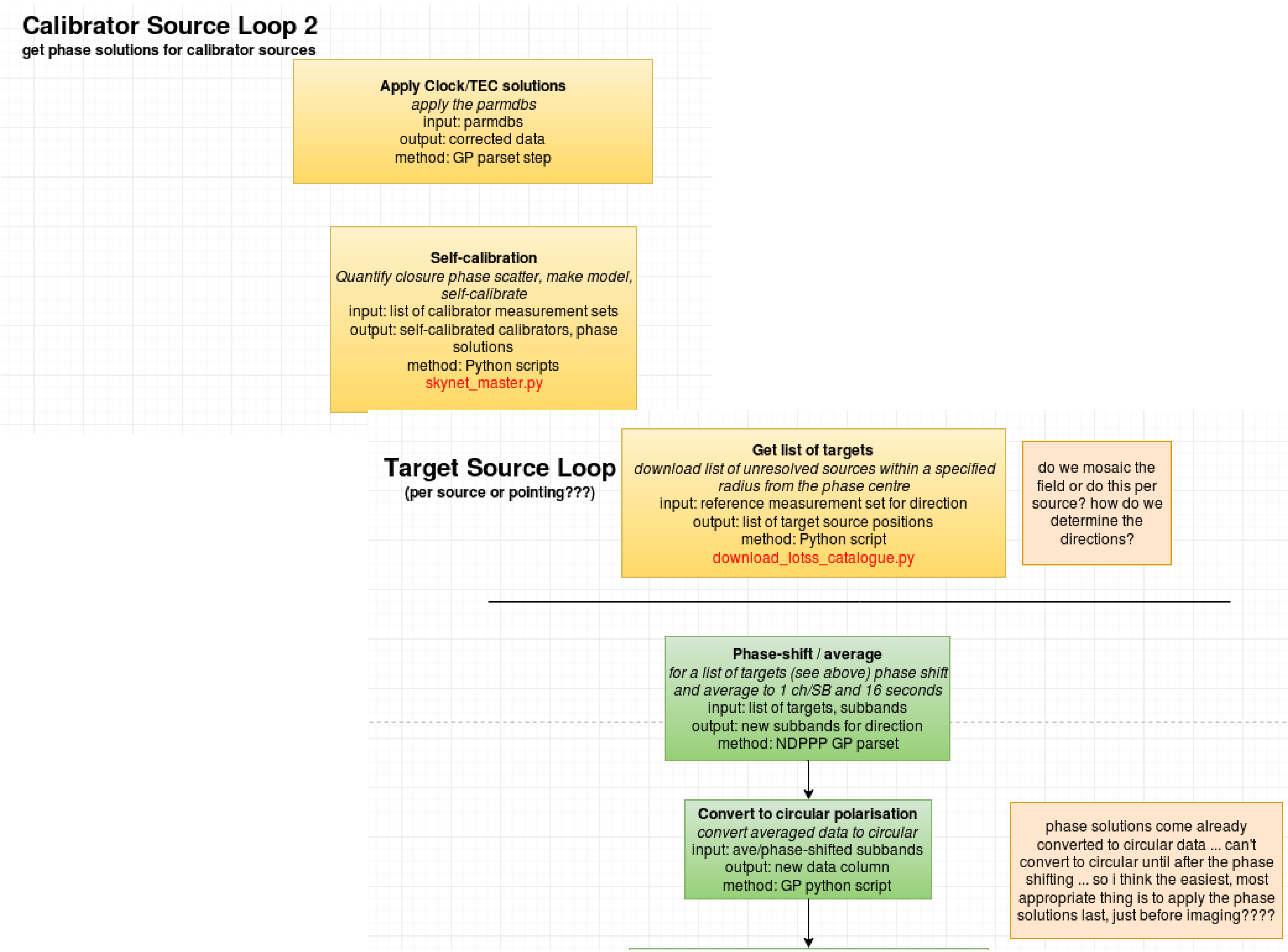
$$\begin{pmatrix} \partial\phi/\partial t \\ \partial\phi/\partial f \end{pmatrix} = \begin{pmatrix} \frac{1}{\cos D}\frac{\partial^2\phi}{\partial H \partial t} & \frac{\partial^2\phi}{\partial D \partial t} \\ \frac{1}{\cos D}\frac{\partial^2\phi}{\partial H \partial f} & \frac{\partial^2\phi}{\partial D \partial f} \end{pmatrix} \begin{pmatrix} \Delta H \\ \Delta D \end{pmatrix},$$

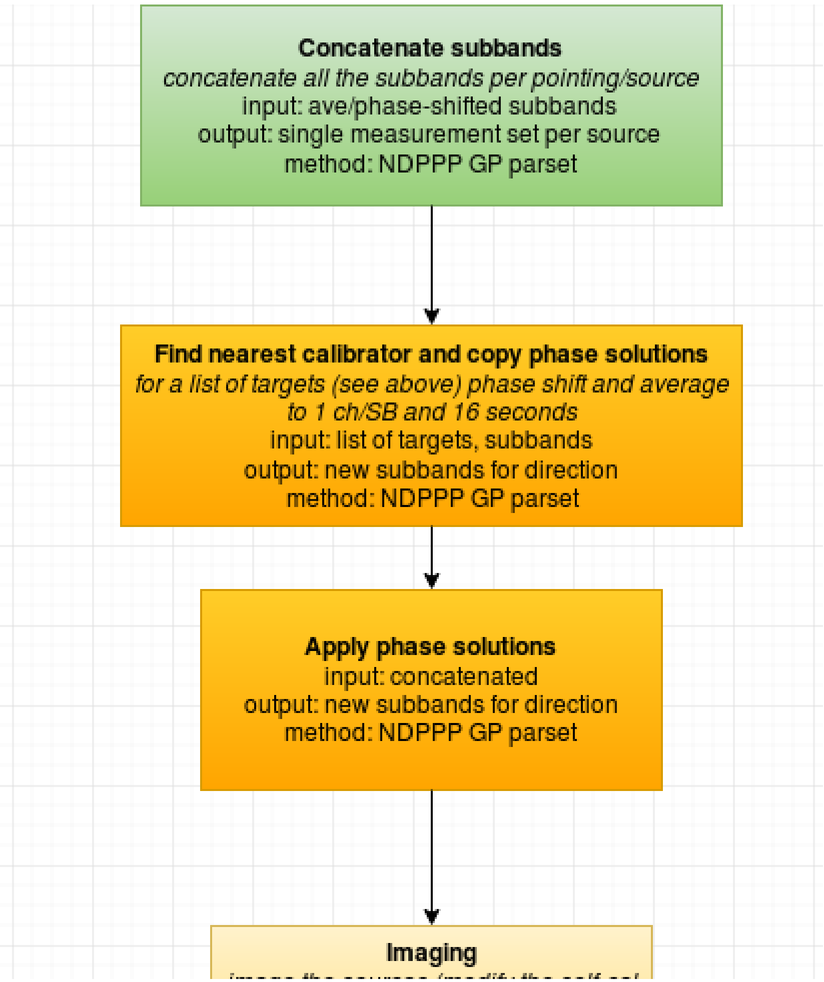


Very quick way to determine which sources in the field have correlated flux on a particular baseline

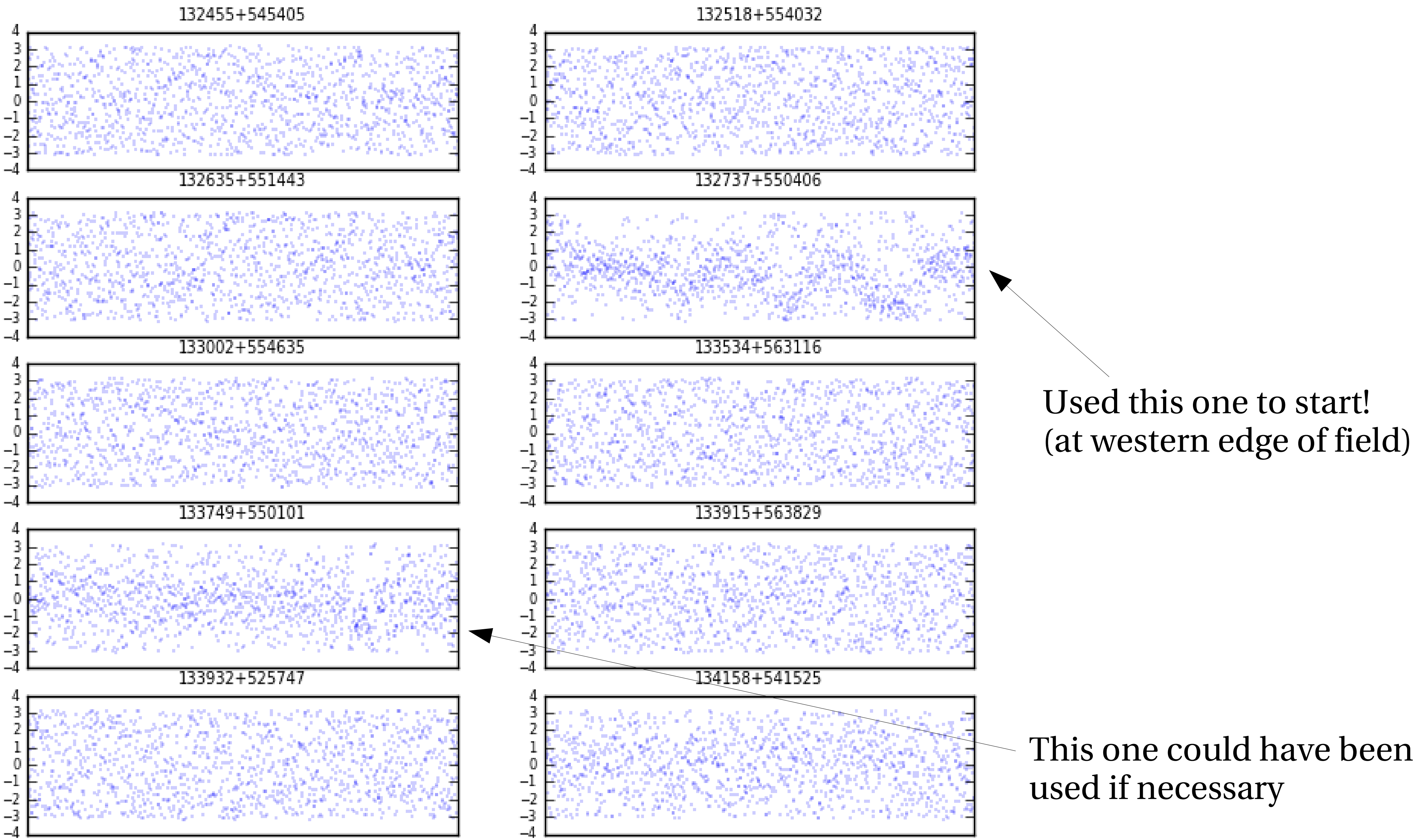


Delay Calibration
Determine calibrator with best closure phase
using lowest scatter values
input: closure_phase.txt
output: name of calibrator file
method: python script
find_calibrator_for_delays.py
Generate model for delay calibration
find a starting model
input: best calibrator concatenated dataset
output: skymodel
method: python script
skynet_master.py
diagonal solve on concatenated dataset will provide information to solve for clock/TEC. These values are scalar offsets and thus can be appropriately applied to the linear polarisation data. But the parmdb will need to be updated to apply the ST001
Gain calibration
solve for diagonal gains
input: best calibrator concatenated dataset
output: numpy arrays
method: NDPPP
Clock/TEC separation
do clock/TEC separation
input: best calibrator concatenated dataset or subbands?
output: numpy arrays
method: borrowed from prefactor pipeline
Transfer Clock/TEC to all data
write parmdbs with clock & TEC (i.e. delays) and account for CS/ST001 stations
input: best calibrator numpy arrays
output: parmdbs
method: GP python plugin
transfer_clock+tec_toMS.py
Apply Clock/TEC solutions
apply the parmdbs
input: parmdbs, CORRECTED_DATA
output: new MS with corrected data in DATA
method: GP parset step

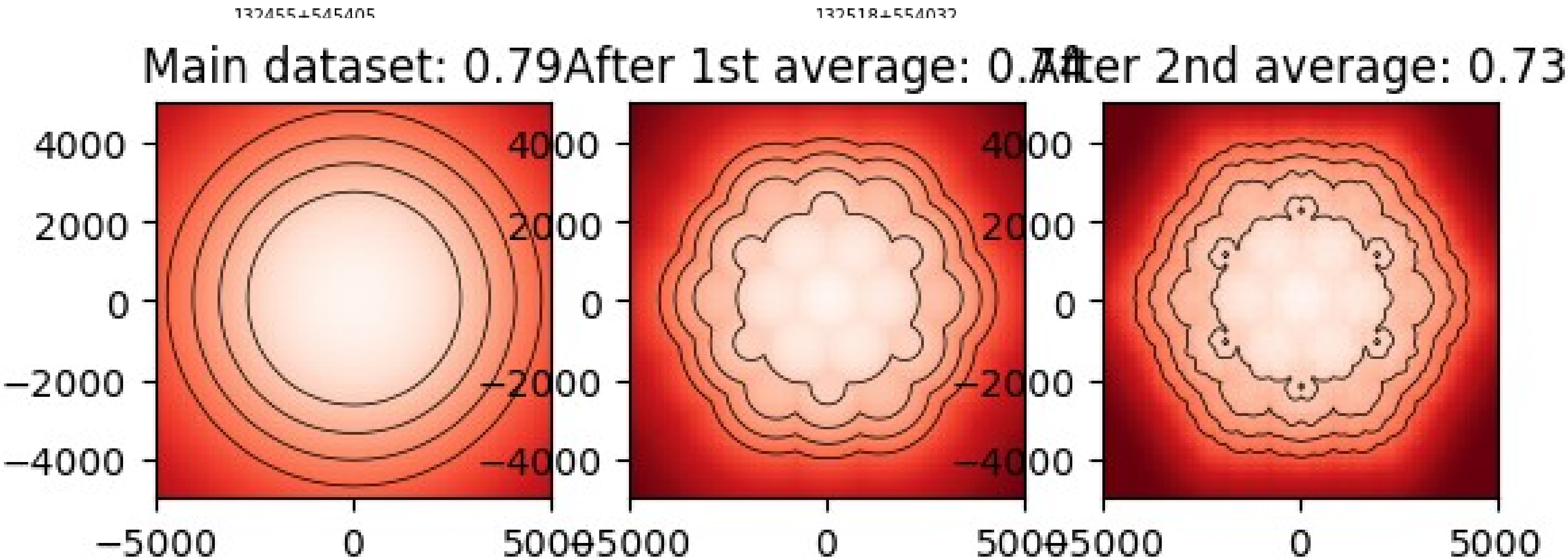

Calibrator Source Loop 2
get phase solutions for calibrator sources
Apply Clock/TEC solutions
apply the parmdbs
input: parmdbs
output: corrected data
method: GP parset step
Self-calibration
Quantify closure phase scatter, make model, self-calibrate
input: list of calibrator measurement sets
output: self-calibrated calibrators, phase solutions
method: Python scripts
skynet_master.py
Target Source Loop
(per source or pointing???)
Get list of targets
download list of unresolved sources within a specified radius from the phase centre
input: reference measurement set for direction
output: list of target source positions
method: Python script
download_lotss_catalogue.py
do we mosaic the field or do this per source? how do we determine the directions?
Phase-shift / average
for a list of targets (see above) phase shift and average to 1 ch/SB and 16 seconds
input: list of targets, subbands
output: new subbands for direction
method: NDPPP GP parset
Convert to circular polarisation
convert averaged data to circular
input: ave/phase-shifted subbands
output: new data column
method: GP python script
phase solutions come already converted to circular data ... can't convert to circular until after the phase shifting ... so i think the easiest, most appropriate thing is to apply the phase solutions last, just before imaging????


Concatenate subbands
concatenate all the subbands per pointing/source
input: ave/phase-shifted subbands
output: single measurement set per source
method: NDPPP GP parset
Find nearest calibrator and copy phase solutions
for a list of targets (see above) phase shift and average to 1 ch/SB and 16 seconds
input: list of targets, subbands
output: new subbands for direction
method: NDPPP GP parset
Apply phase solutions
input: concatenated
output: new subbands for direction
method: NDPPP GP parset
Imaging

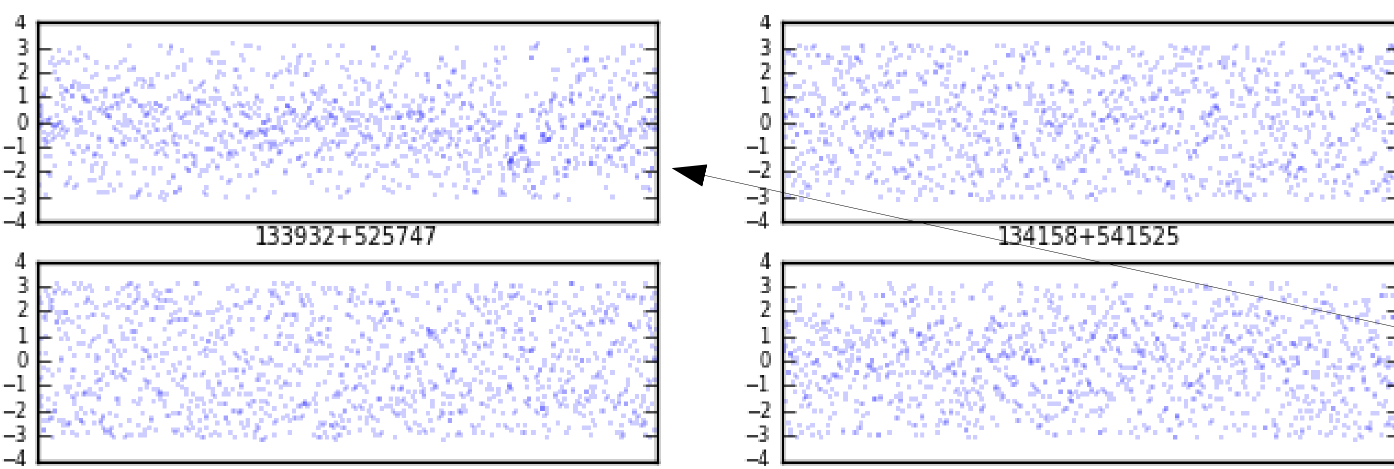
Large dataset is then shifted and averaged around particular sources



Large dataset is then shifted and averaged around particular sources
(if many sources, need to do chessboard averaging to avoid huge runtimes and smearing)

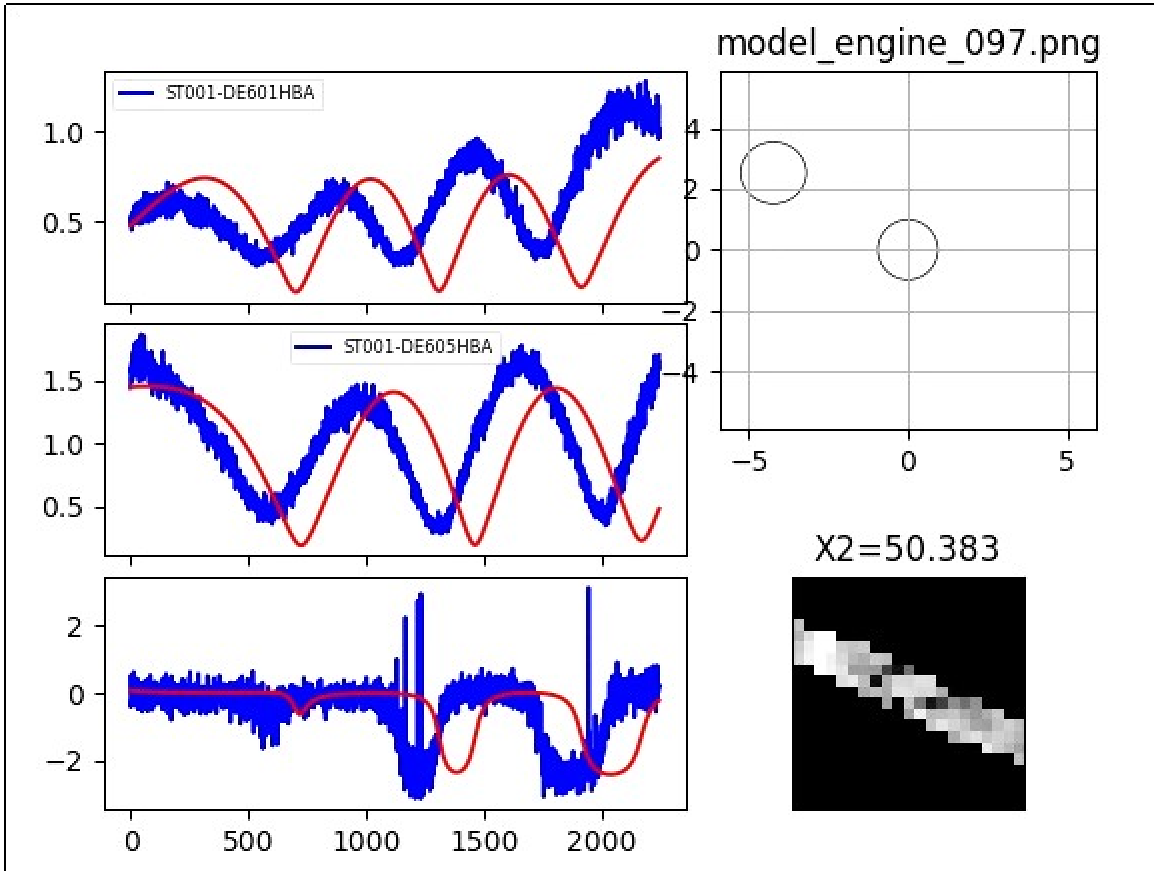


d)



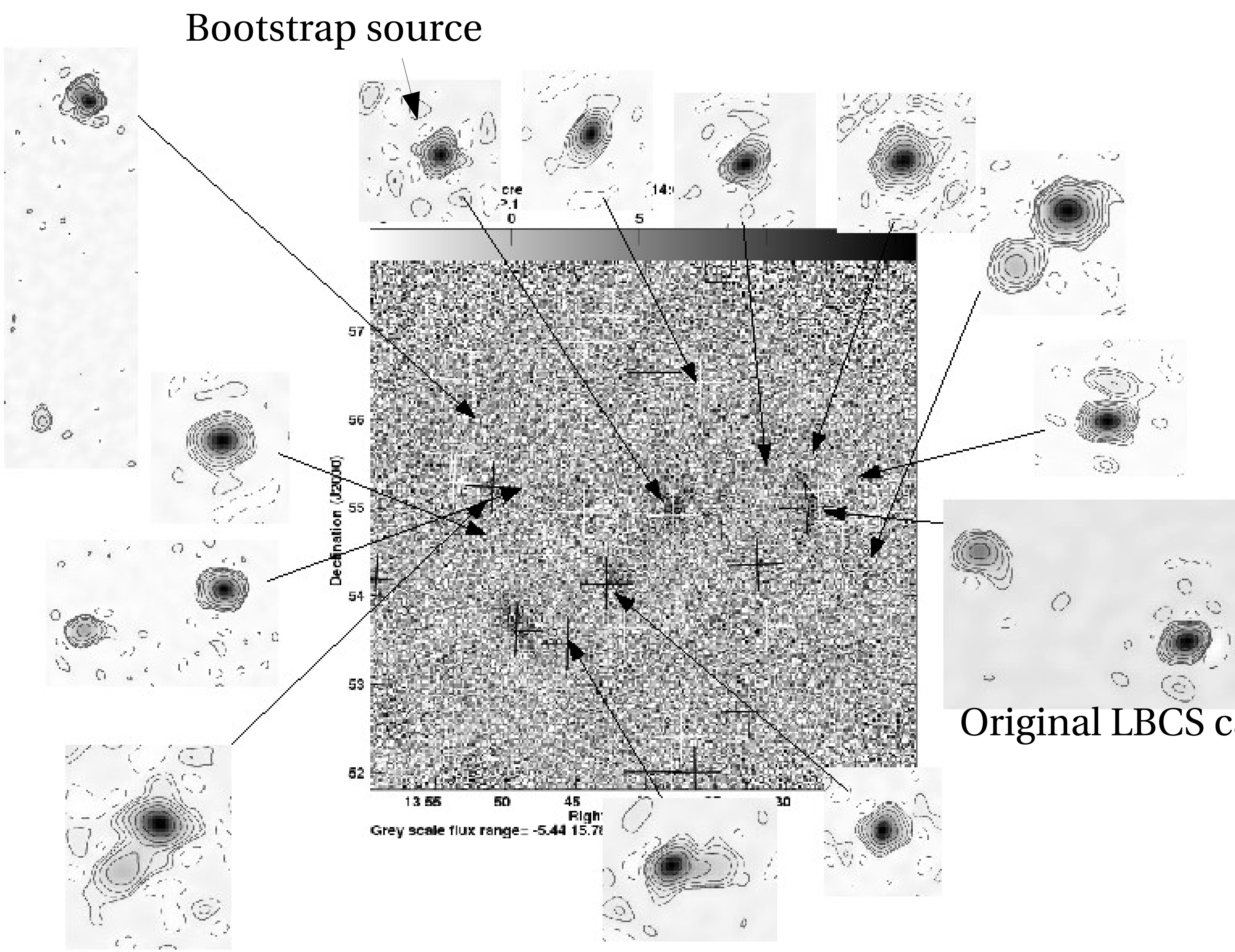
This one could have been used if necessary

Starting models can be made using closure phases



Bright sources relatively robust against starting model
Faint sources are not – major issue in avoiding selfcaling sources into existence
- here important to examine closure phases

Then propagate phase solutions across field from initial calibrator to bootstrap sources



Are we nearly there yet?

- Ideal: reprocess all the LOTSS data with international stations
- Plan: busy week in ASTRON September (test full pipeline)
  wider meeting March/April for community input/finalise parameters
- Still to determine: final calibration parameters
  timings/resource reduction
  where to run the pipelines