Self-calibration

Joshua Marvil

8th VLA Data Reduction Workshop

Introduction

This talk is based on the VLA self-calibration CASA guide:

casaguides.nrao.edu/index.php?title=VLA_Selfcalibration_Tutorial

Please see the casaguides wiki page for links to download the data, the complete set of CASA commands and a more in-depth discussion

Goals

— learn the basic self-calibration procedure

— understand the parameters and how to choose them

— recognize common pitfalls that cause selfcal to fail

— quantify how well selfcal has worked

Calibration in CASA

1) Set model visibilities (column: MODEL_DATA) for the field being calibrated— **setjy, ft, tclean**

   (if no model is defined, a 1 Jy point source is assumed)

2) Compare observed visibilities (column: DATA) to the model— **gaincal, bandpass, polcal, blcal, fringefit**

3) Calibration tables are the solutions that make DATA more closely resemble MODEL_DATA

> **Both standard and self-calibration are consistent with this definition**

Calibration in CASA

Standard calibration:

- input fields are marked as calibrators

- models from setjy or default 1 Jy

- solutions applied to target fields

Self calibration:

- input fields are marked as target

- models are typically from tclean

- solutions applied to same fields

When to Selfcal might help

There are artifacts from the source of interest due to calibration errors

There are artifacts from a background source due to direction-independent calibration errors

Typical sources of these errors:
- **atmospheric density fluctuations**
- **antenna-based gain changes**

When Selfcal will not help

- Errors in creating the image (e.g., ignoring the W-term)

- Errors in deconvolution (e.g., ignoring wide-band effects)

- Insufficient UV coverage (e.g., missing short spacings, snapshot synthesis)

- Unflagged RFI in the target visibilities

- Source(s) are too weak to achieve sufficient signal-to-noise in the self-calibration solutions

- There is a bright outlying source with artifacts due to direction-dependent errors

The Observation

- Target is z~1 galaxy cluster with high radio activity

- B-configuration (project 17B-197; October 2017)

- C-band continuum (8 bit, 2 GHz, 16 SPWs)

- 3 target scans of ~10 minutes each

- Multiple radio sources, peak flux ~ 7 mJy

- Calibrated with the VLA CASA pipeline

- Strong artifacts remained after cleaning

obj.prelim_clean.3arcmin.image.tt0−raster

8th VLA Data Reduction Workshop
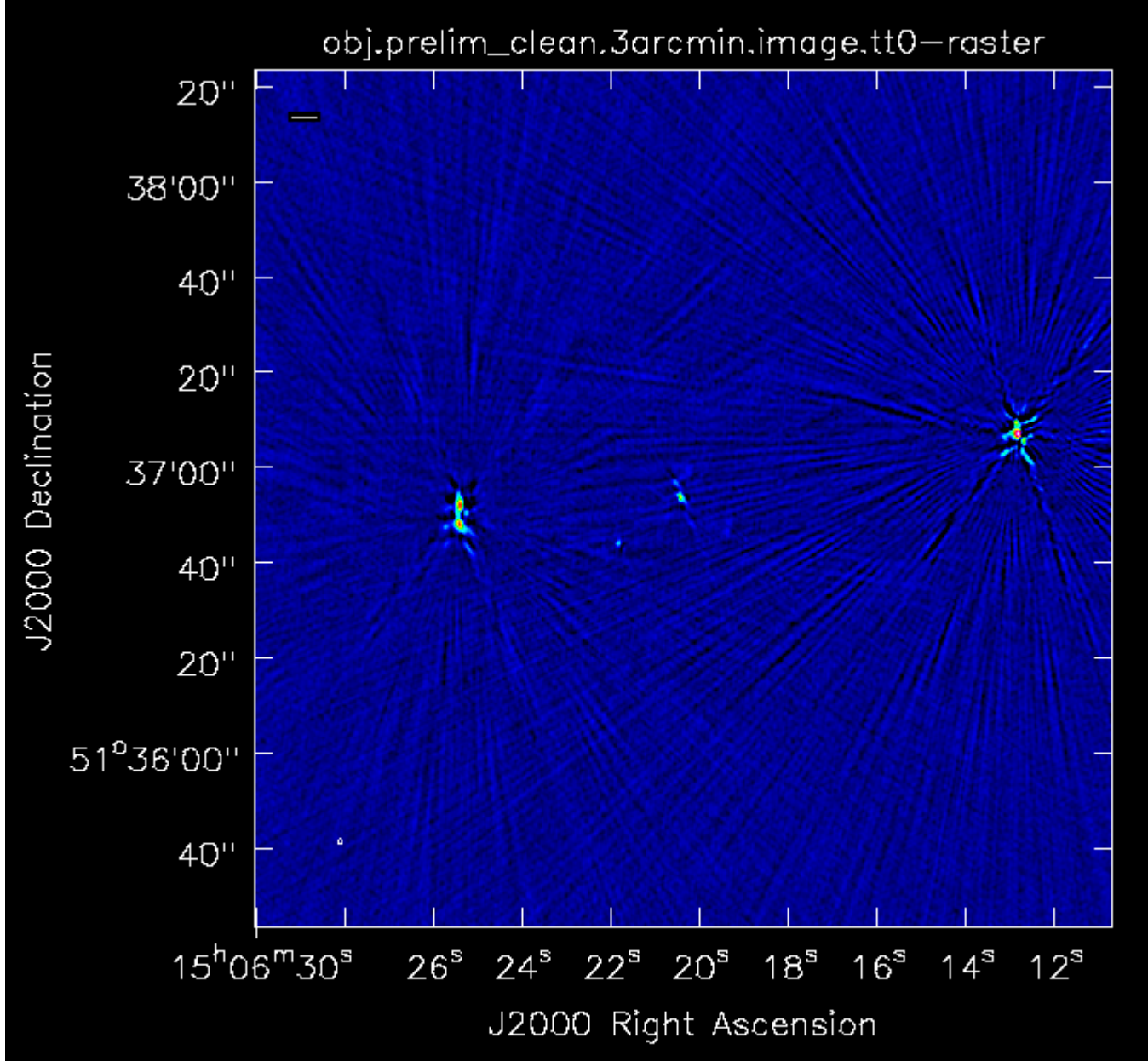
Phase Stability

Our first goal with self-calibration will be to improve the image by correcting phase errors in the target visibilities

Before we do this, it may be insightful to learn more about the observing conditions (i.e., phase stability)

One way to do this is to identify a long scan on a bright calibrator (e.g., bandpass) and plot the phase vs. time for a single baseline

Phase Stability

```
plotms(vis='17B-197.sb
34290063.eb34589992.58
039.86119096065.ms',
xaxis='time',
yaxis='phase',
ydatacolumn=
'corrected',
field='2',
antenna='ea05',
correlation='RR,LL',
avgchannel='64',
iteraxis='baseline',
coloraxis='spw')
```
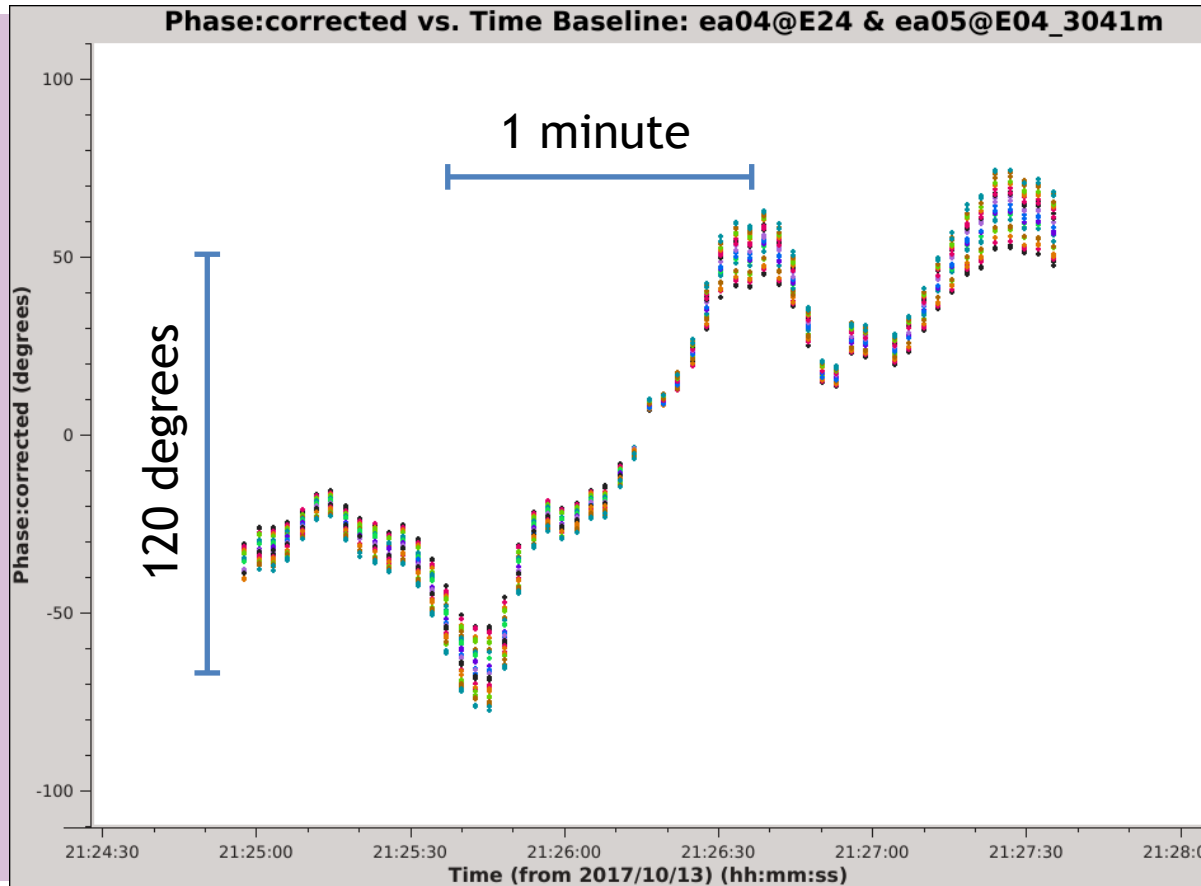

(Target scans are 10 minutes long)

Initial Inspection

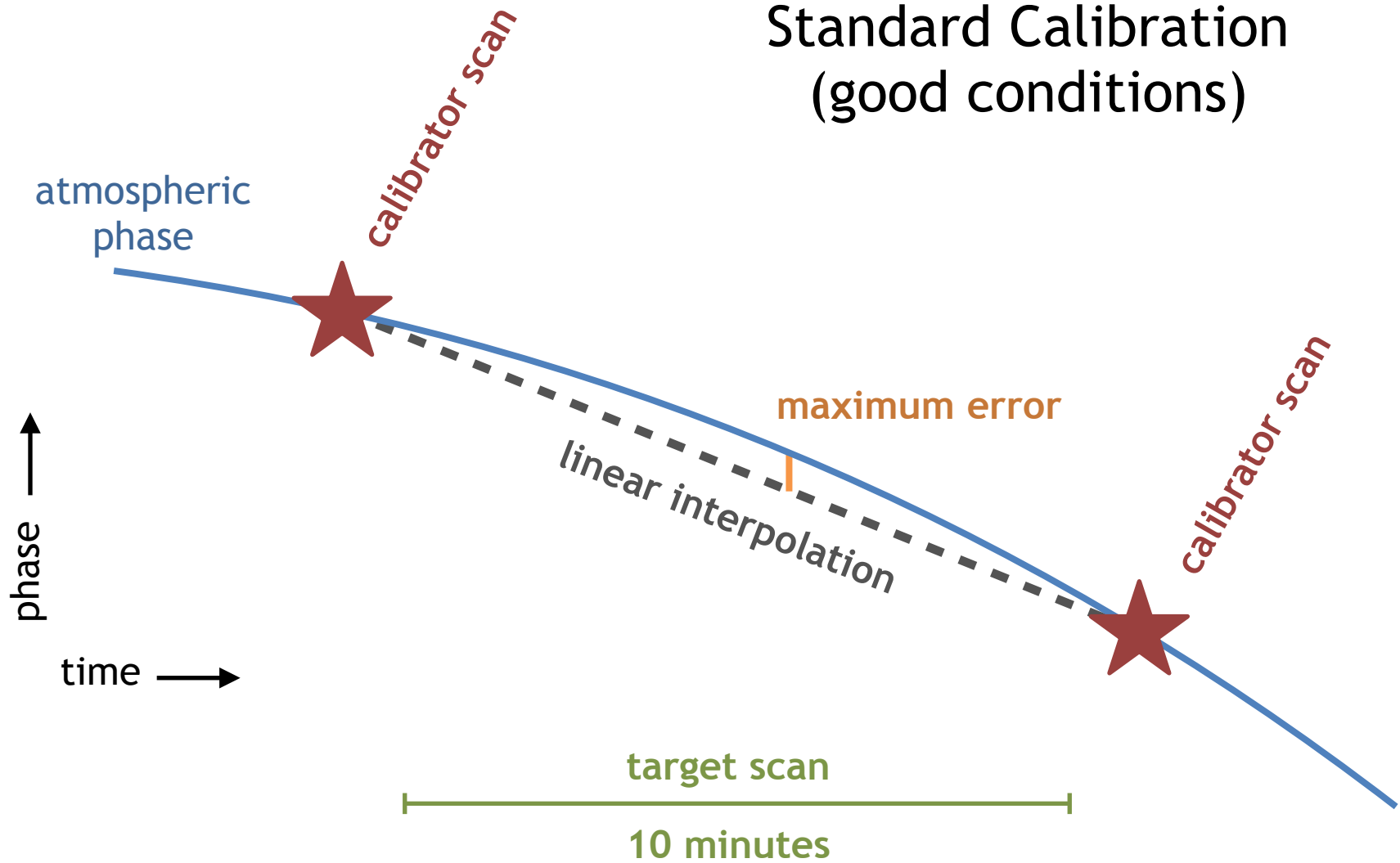The atmospheric phase stability is not good

The time it takes for the visibility phase to change by a radian, i.e., the *coherence time*, is ⪅ 1 minute

This is much shorter than the calibrator cycle time of 10 minutes
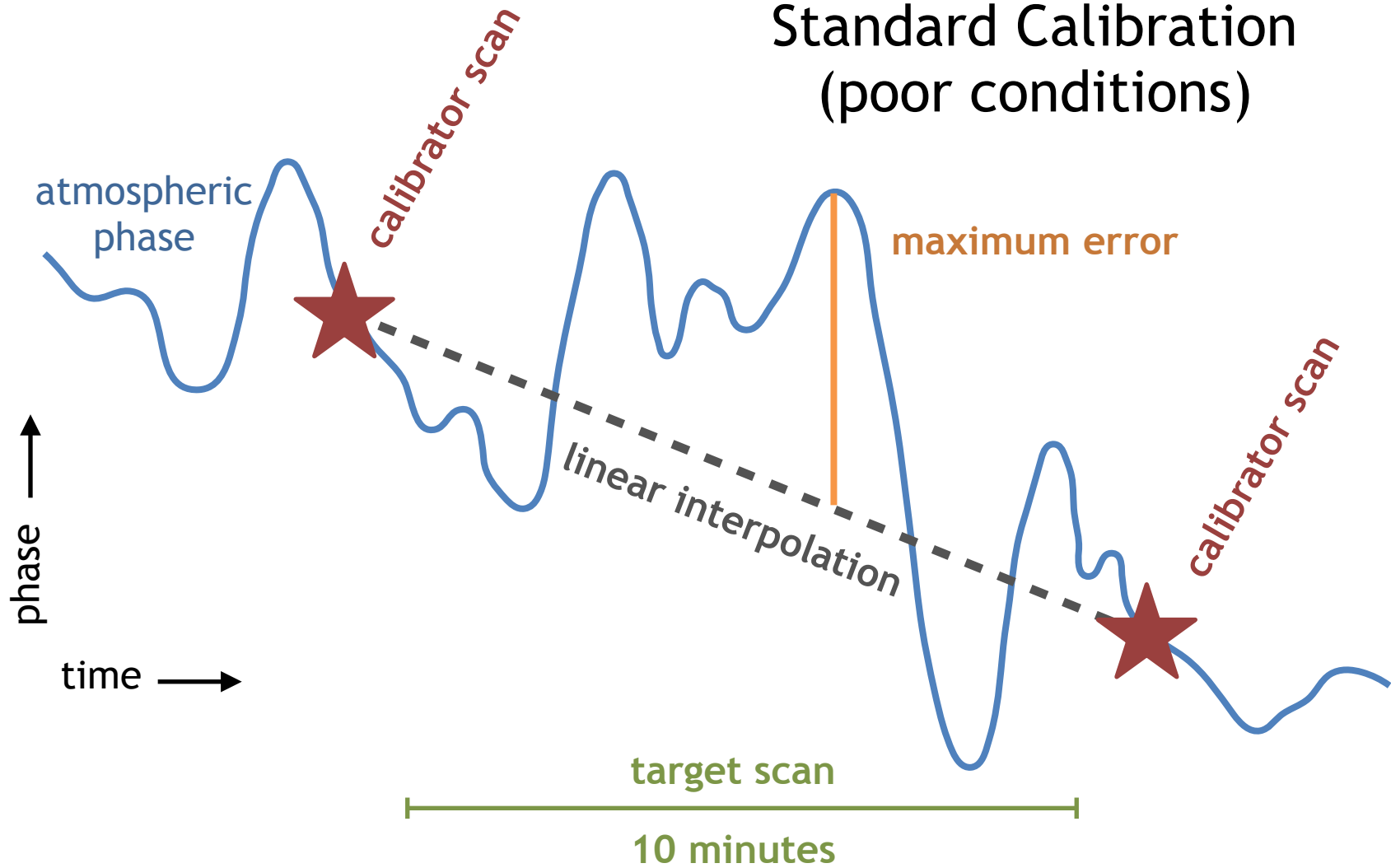
Standard calibration may leave large residual phase errors in the target data

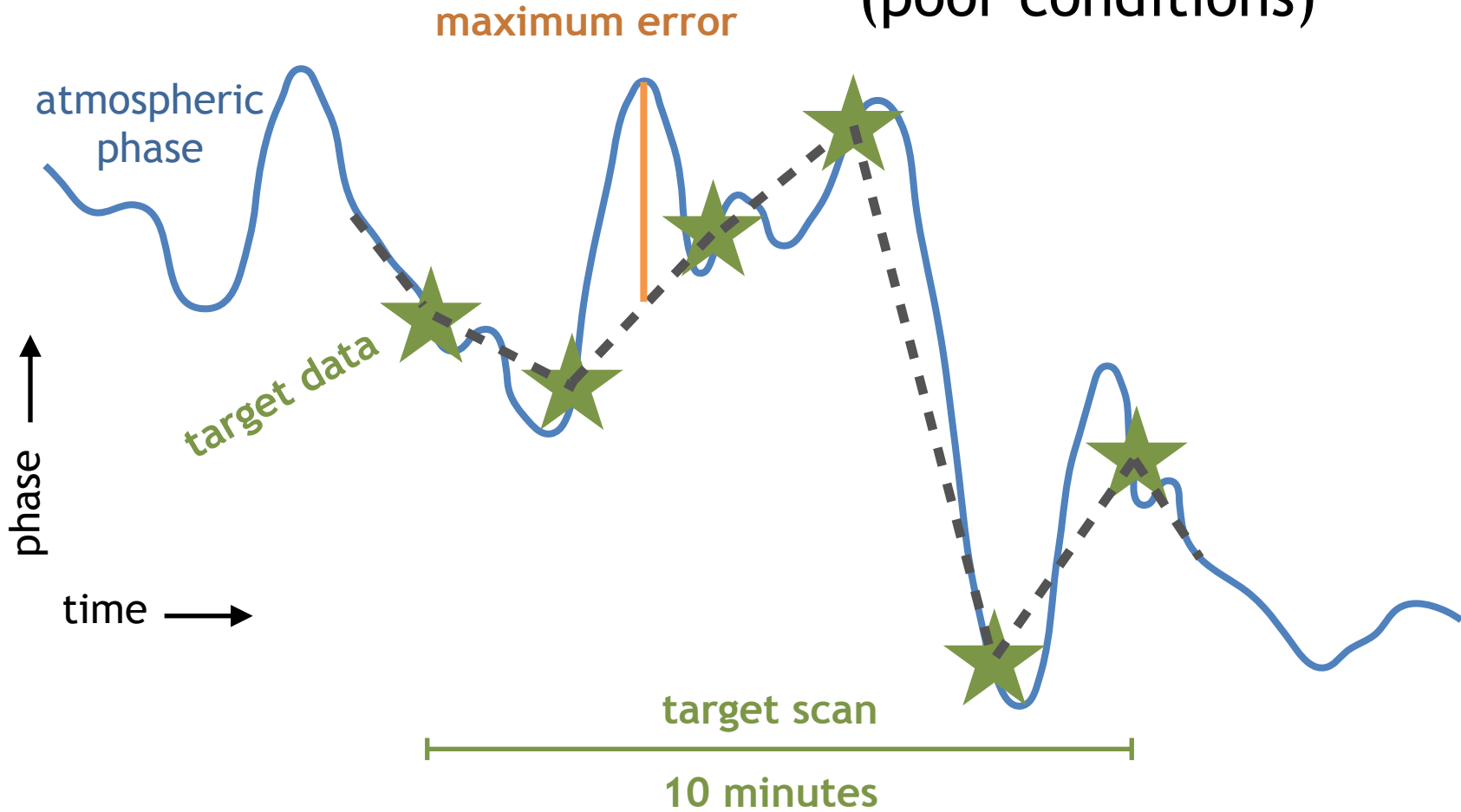We should aim for self-calibration solutions ⪅ 1 minute if possible

Standard Calibration (good conditions)

atmospheric phase

calibrator scan

maximum error

linear interpolation

calibrator scan

phase

time

target scan

10 minutes

Standard Calibration (poor conditions)

atmospheric phase

calibrator scan

maximum error

calibrator scan

linear interpolation

phase

time

target scan

10 minutes

Self Calibration
(poor conditions)

maximum error

atmospheric phase

target data

phase

time

target scan

10 minutes

Preparing the Data

At the end of the standard calibration procedure, the applycal task is run to create the CORRECTED_DATA column

To prepare the data for self-calibration, we will copy the corrected visibilities to a new measurement set ** selfcal will not work without this step **

Specifically, the task split will take the CORRECTED_DATA column of the input MS and place it in the DATA column of the output MS

You can also use data selection to copy only the target data of interest, e.g., field(s), SPWs, correlations

```
split(vis='17B-197.sb34290063.eb34589992.58039.861
19096065.ms', datacolumn='corrected', field='1',
correlation='RR,LL', outputvis='obj.ms')
```

Selfcal Requires an Initial Model

Now that we have prepared a measurement set to start the self-calibration process, the first step is to create a model of our target field

This is typically done by imaging the target visibilities, but if you have prior knowledge of the field you may be able to use that instead

In this demonstration, we will run the task **tclean** to image and deconvolve the target field

Creating the Initial Model

It is important to ensure that proper imaging techniques are used (e.g., wide-band, wide-field) so as not to create additional image errors

We want to model all sources that contribute significantly to the visibilities, even those far away from the target of interest

We want to clean conservatively to avoid artifacts (e.g., sidelobes, negative regions) becoming part of the model

Choosing the Field-of-View

We create a dirty image of the entire primary beam and its first sidelobe to see where the bright sources are

We see two bright sources near the center of the image

Four much fainter sources are circled in magenta
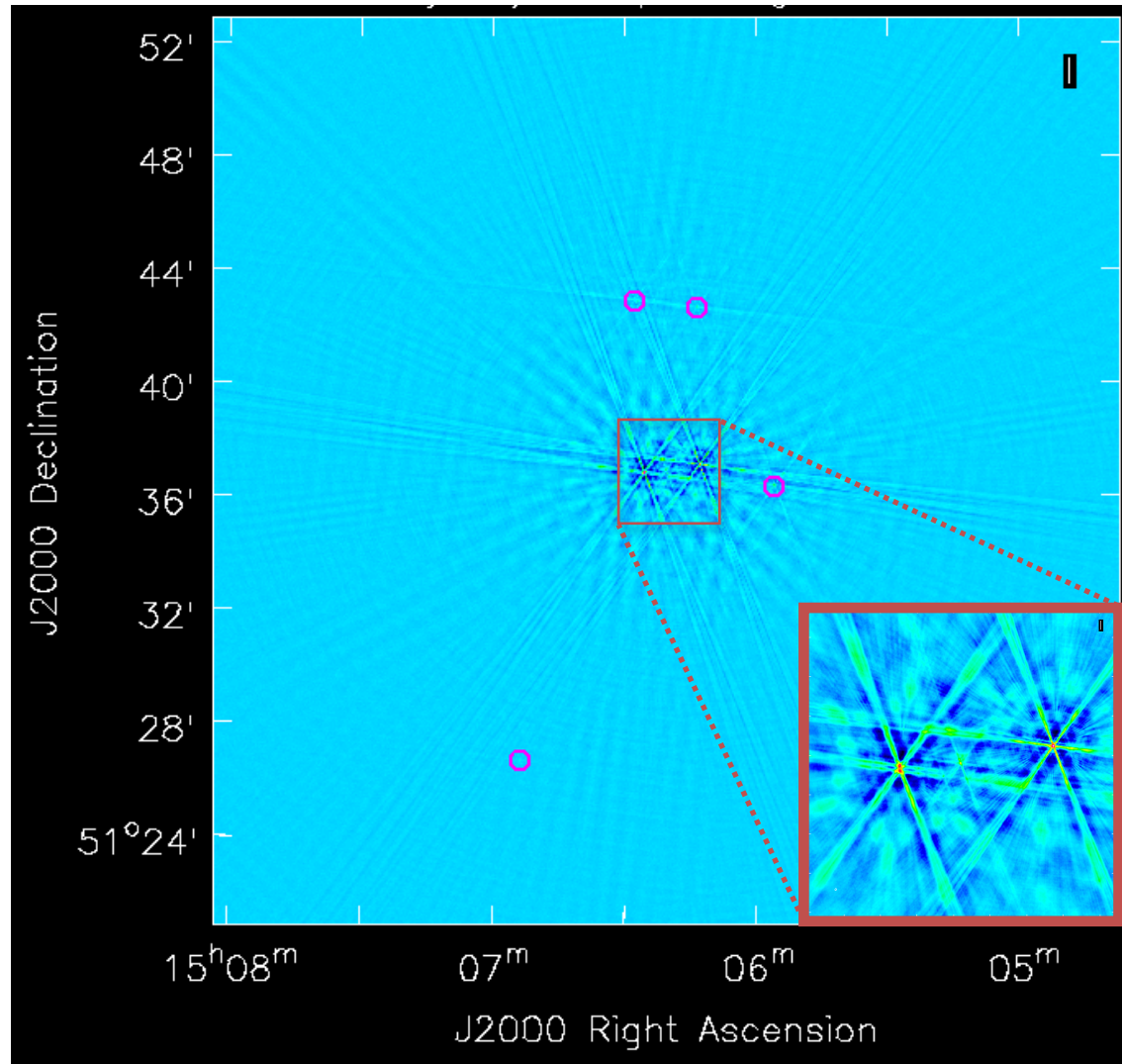
Handling Outlier Sources

Outlying sources can be modeled by cleaning the large image or by placing an outlier field on each source.

Alternatively, they can be UV-subtracted if they are not of scientific interest.

Another option is to simply ignore them if they are too weak to contribute to the calibration solutions

We conclude that the outlying sources do not need to be modeled because they are much fainter than the two bright sources in the center

Running tclean

By studying the large field-of-view image we were able to make a decision about the outlying sources and to choose an appropriate image size to use going forward

We will now run **tclean** again to save the initial model. We choose the following key parameters:

savemodel='modelcolumn' -- To enable writing the MODEL_DATA column to the MS. **** selfcal will not work without this step ****

weighting='briggs'; robust=0 -- to to suppress PSF sidelobes and thereby reduce certain types of artifacts

interactive=True -- to avoid artifacts by careful masking and by deciding when to stop cleaning

Interactive Masking

This is one of the two bright sources in the center of the field.

We draw tight masks around the emission that we believe is real

This is the same source after cleaning for 3 major cycles

We are suspicious that the residual structures are artifacts so we do not include them in the mask
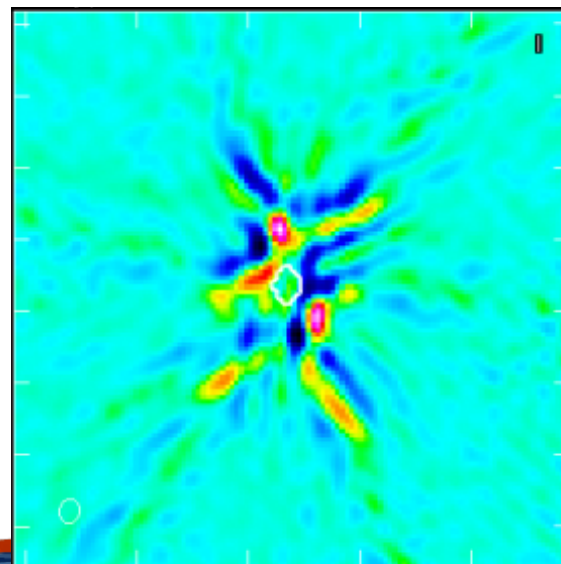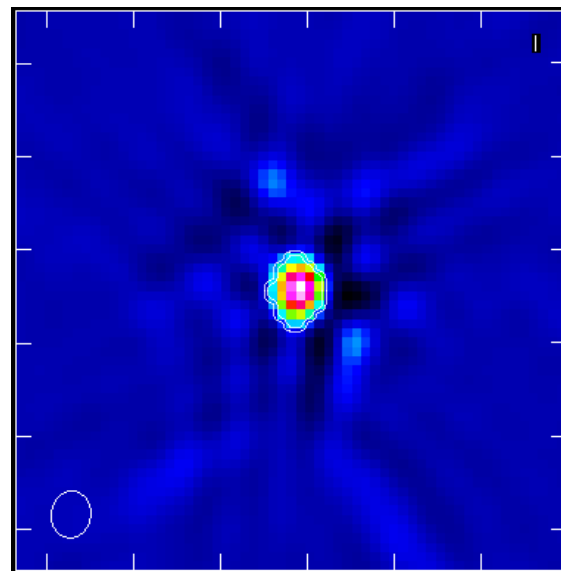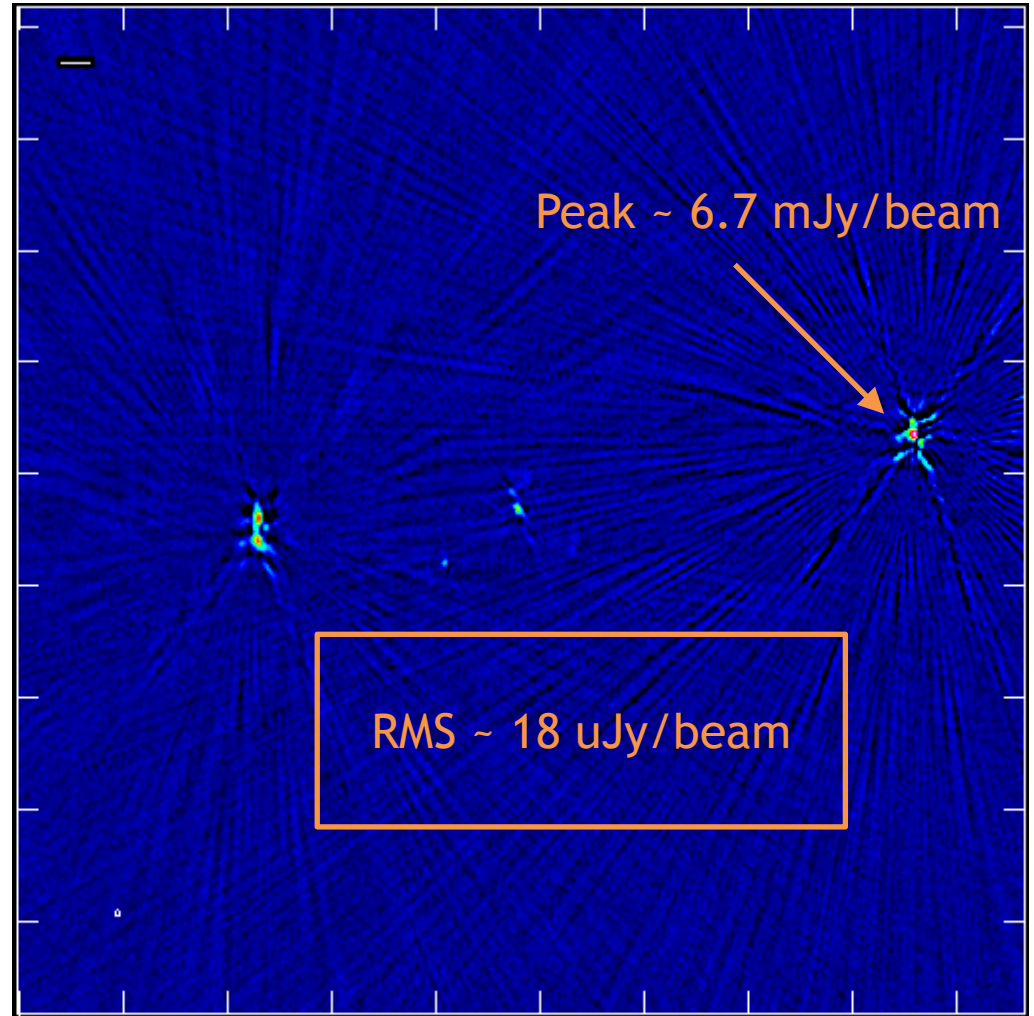
The Initial Image

This is the image we get after cleaning

The sensitivity is not thermal noise-limited

Artifacts from the bright sources dominate the noise floor

We measure a **dynamic range** of:

   Peak / RMS ~ 370

Peak ~ 6.7 mJy/beam

RMS ~ 18 uJy/beam

Verifying the Initial Model

The initial model now exists in two forms: the model image used during deconvolution, and the model visibilities in the MODEL_DATA column of the measurement set

It is recommended to verify that the later has been filled properly
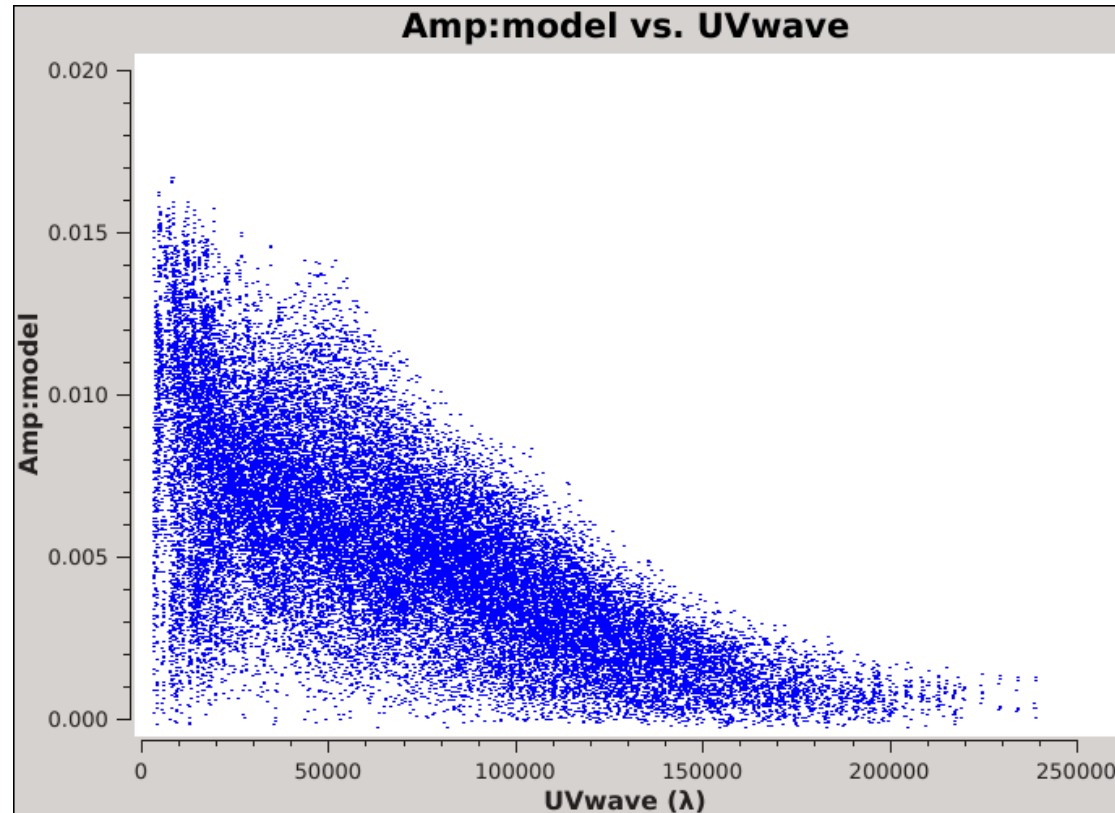

```
plotms(vis='obj.ms', xaxis='UVwave', yaxis='amp',
ydatacolumn='model', avgchannel='64', avgtime='300')
```

Choosing Parameters

Now that we have filled the MODEL_DATA column with visibilities we are ready to solve for calibration solutions (i.e., compare with DATA)

It is recommended to begin with **gaincal** in <u>phase-only</u> mode because phase corrections can boost the amplitude by reducing decorrelation

We will have several **gaincal** parameters to consider, in particular:

- solution interval

- data combination (i.e., pol, spw)

- solution signal-to-noise ratio

Choosing Parameters

type of solutions: set **calmode** = 'p' for phase-only solutions (recommended for the first round), or 'ap' for amplitude+phase

solution interval: as short as possible while maintaining an acceptable signal-to-noise. **solint** = 'int', 'inf' or time quantity

data combination: only if demonstrated to show the same underlying pattern, and only if needed to achieve the desired signal-to-noise. **gaintype** = 'T' will combine polarizations, 'G' will solve each polarization separately. **combine** = 'spw' will combine all selected SPWs

minimum signal-to-noise: default **minsnr** = 3.0, but can be raised to reject noisy solutions; can be set to zero for inspection purposes

Initial Solutions

Our first calibration table will use the highest time interval and will not flag any solutions for low signal-to-noise.

This will be a temporary table that we will inspect to help determine the optimal **gaincal** parameters.

```
gaincal(vis='obj.ms', caltable='selfcal_initial.tb',
solint='int', refant='ea24', calmode='p', minsnr=0,
gaintype='G')
```
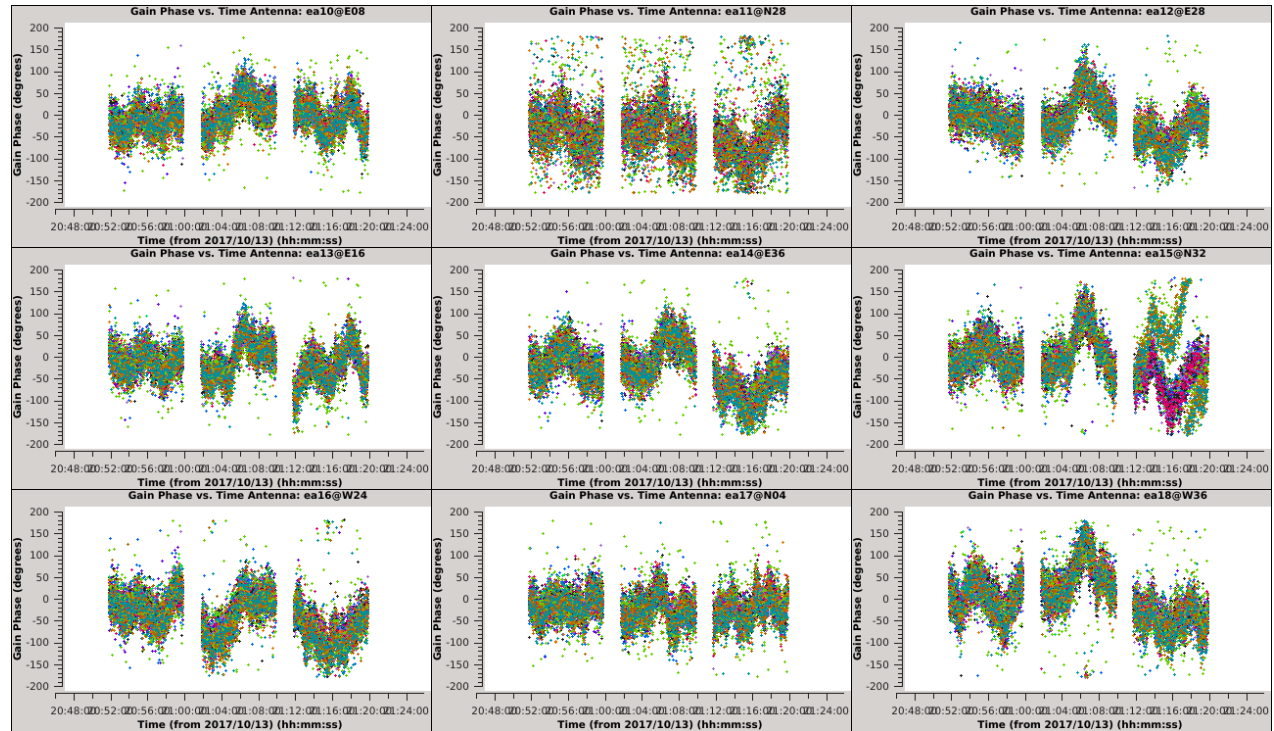
Inspecting Solutions

Both polarizations (red and black) show the same pattern, and therefore could be combined if necessary
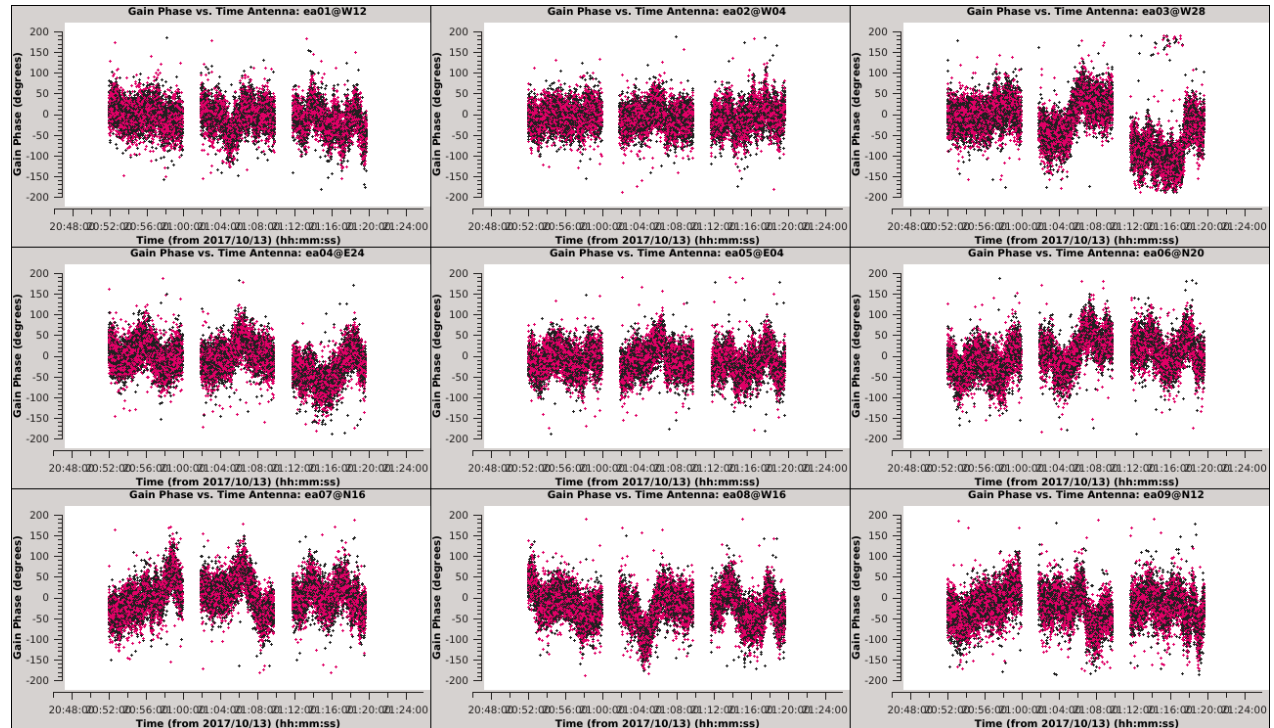
```
plotms(vis='selfcal_initial.tb', xaxis='time',
yaxis='phase', iteraxis='antenna', gridrows=3,
gridcols=3, coloraxis='corr')
```

Inspecting Solutions

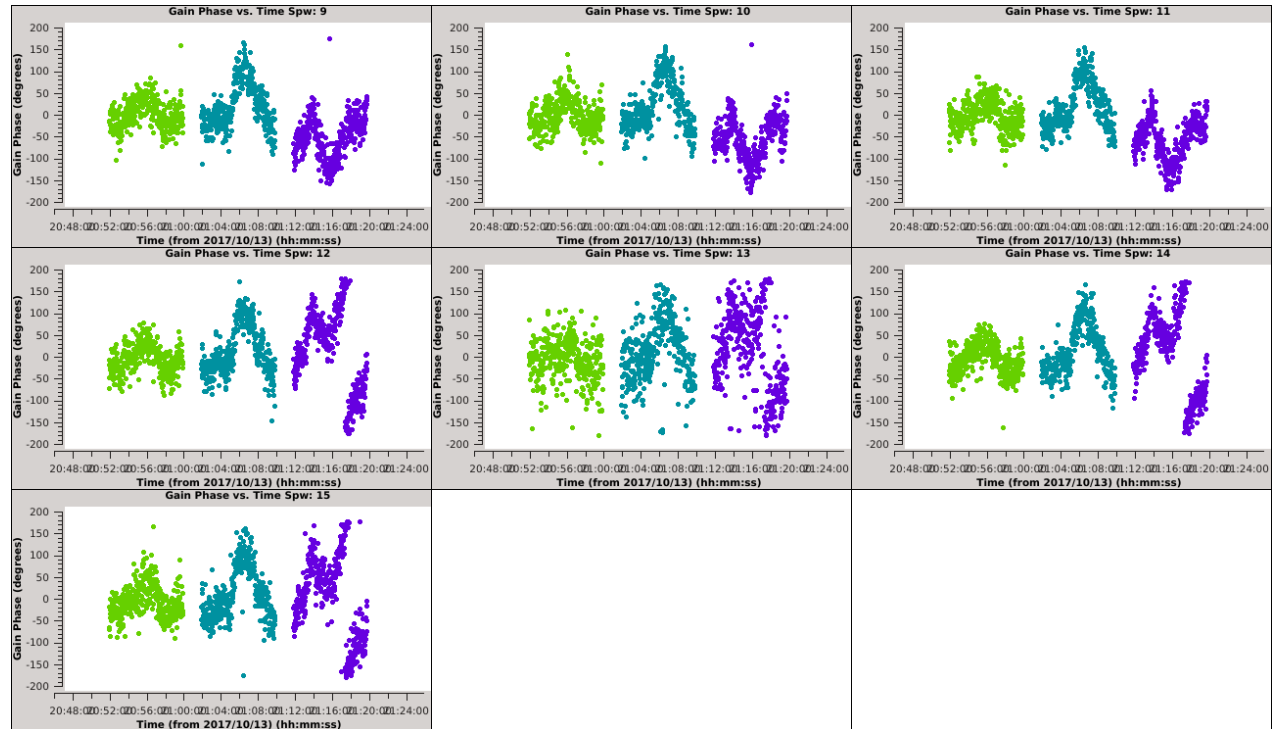All SPWs (colored) do not always show the same pattern, e.g., antenna ea15

```
plotms(vis='selfcal_initial.tb', xaxis='time',
yaxis= 'phase', iteraxis='antenna', gridrows=3,
gridcols=3, coloraxis='spw')
```

Inspecting Solutions

The last four SPWs show a phase wrap (2π) while the other SPWs do not.
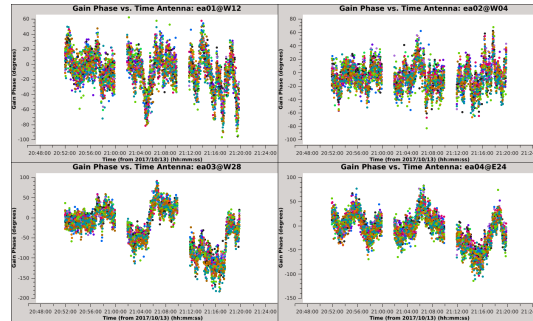
This means that combining SPWs will not be trivial


```
plotms(vis='selfcal_initial.tb', xaxis='time',
yaxis= 'phase', antenna='ea15', iteraxis='spw',
gridrows=3, gridcols=3, coloraxis='scan')
```
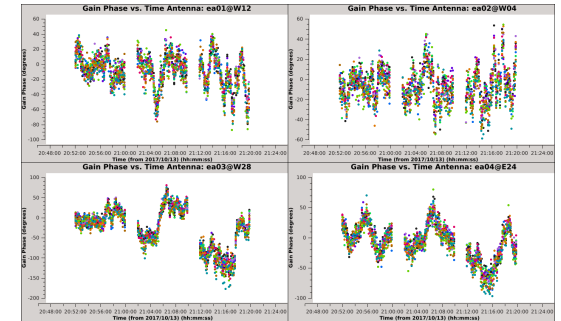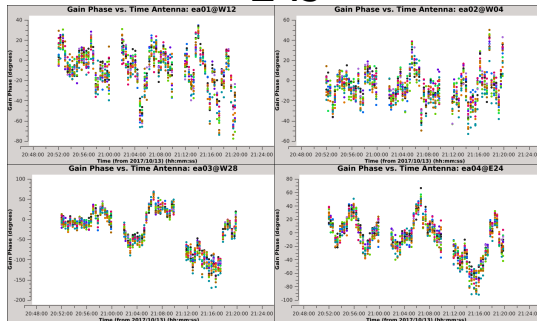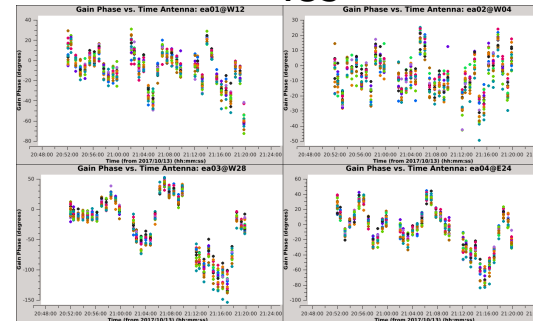
Comparing Solution Intervals
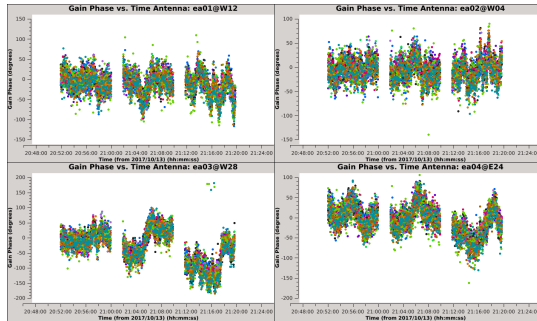

```
for solint in ['3s','6s','12s','24s','48s','96s']:
    gaincal(vis='obj.ms', solint=solint, refant='ea24',
    caltable='selfcal_combine_pol_solint_'+solint+'.tb',
    calmode= 'p', gaintype='T', minsnr=0)
```

Inspecting Solutions

Comparison of 3s (blue) to 96s (red) solution intervals.

The 3s solutions show significant structure but also a lot of noise.

The 96s solutions are unable to sample the large, fast phase changes.
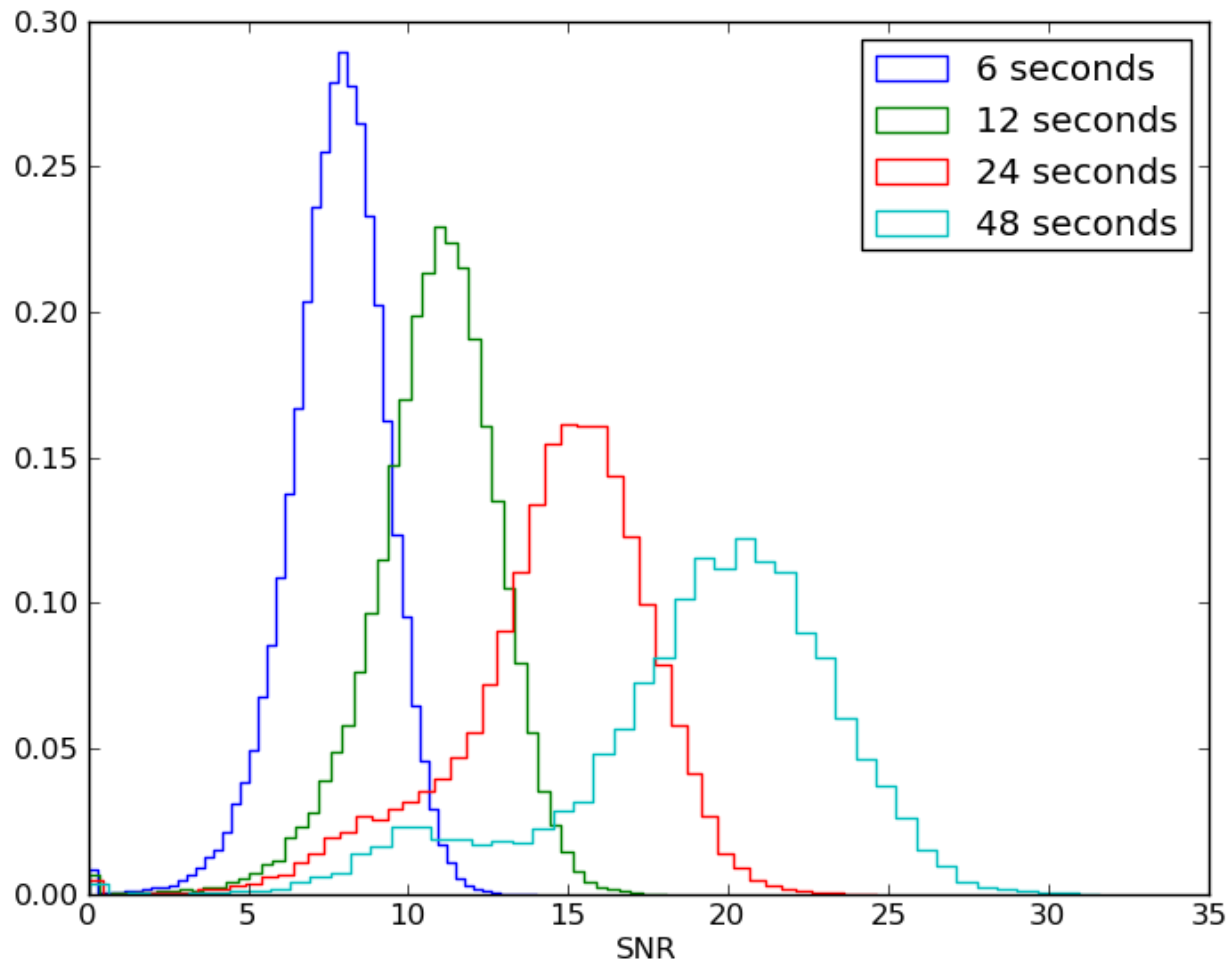

Gain Phase vs. Time
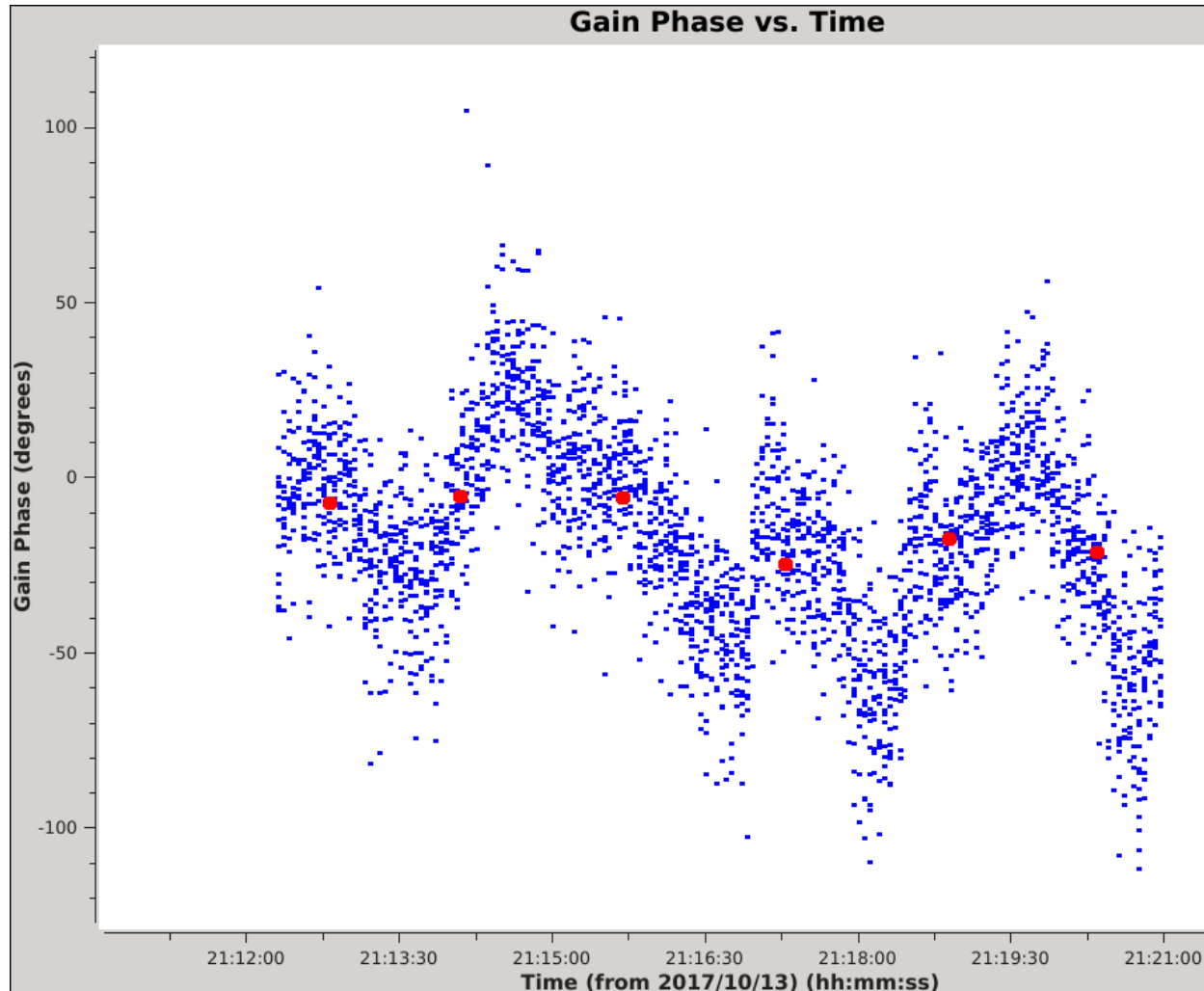
Choosing Parameters

We suggest that the optimal parameters will use the shortest possible interval for which the signal-to-noise is also sufficient.

Low SNR solutions can bias the results, so we adopt a conservative value of **minsnr**=6

This leads us to choose the 12s solutions since <2% of solutions will be flagged

The Final Solutions

We generate the final calibration table by imposing the **minsnr** of 6 and then applying the solutions.

Then we re-run **tclean** using the same parameters except we set datacolumn='corrected' to use the self-calibrated visibilities
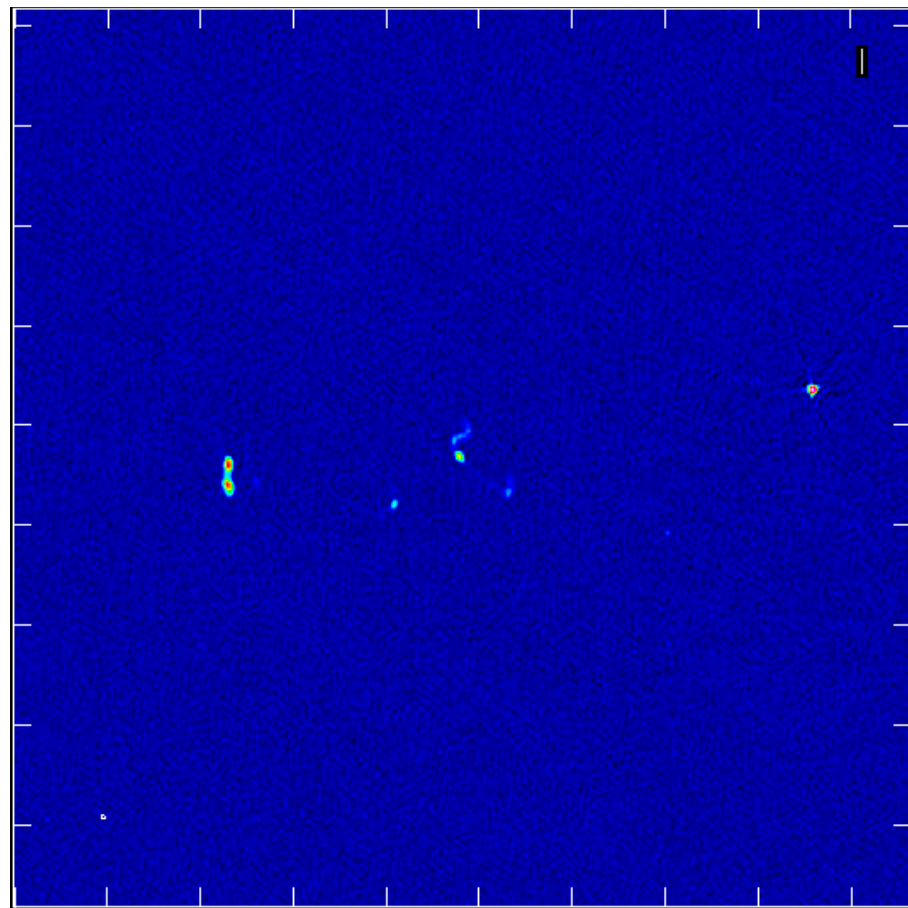
```
gaincal(vis='obj.ms', solint='12s', refant='ea24',
caltable='selfcal_combine_pol_solint_12_minsnr_6.tb',
calmode='p', gaintype='T', minsnr=6)


applycal(vis='obj.ms',gaintable='selfcal_combine_pol_
solint_12_minsnr_6.tb')
```

Before selfcal

After selfcal

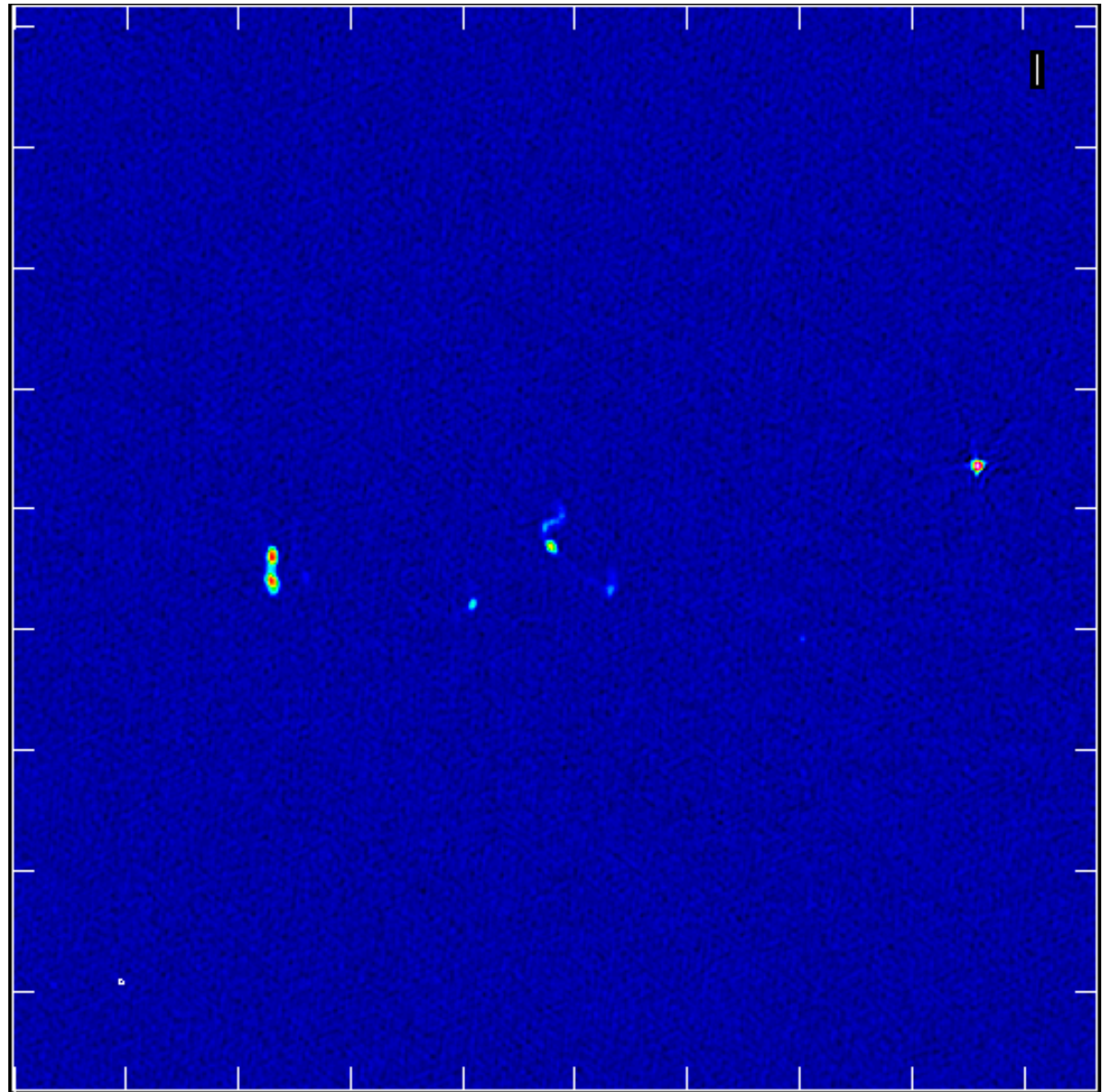**8th VLA Data Reduction Workshop**
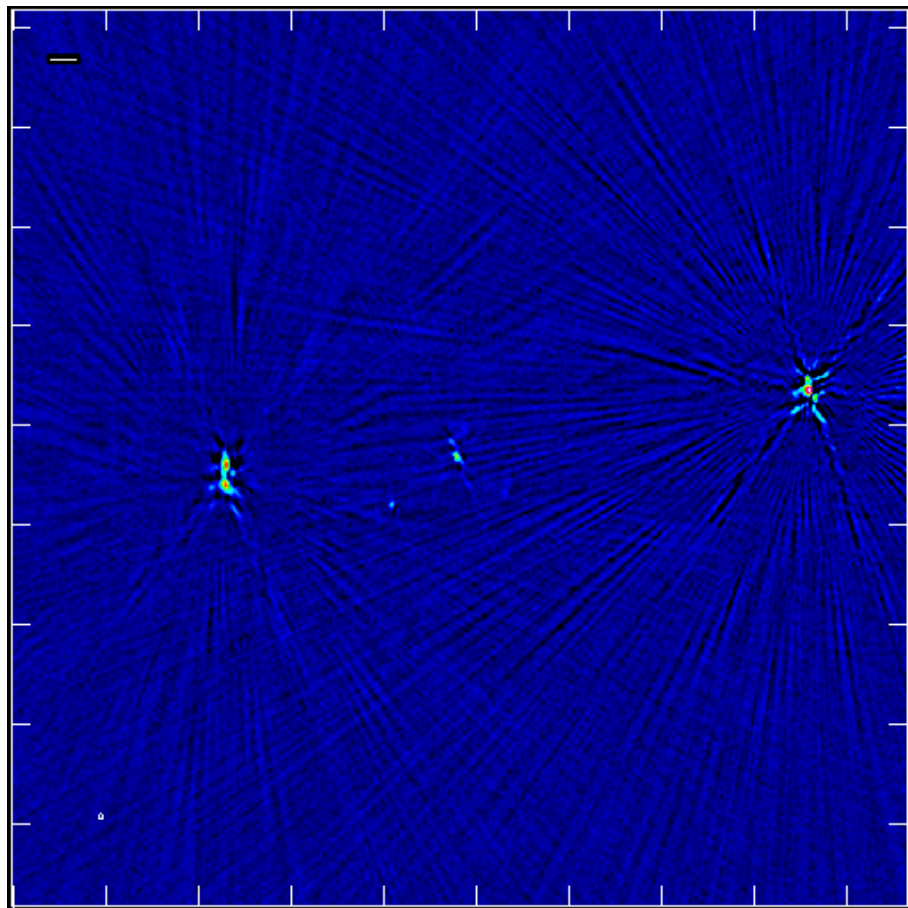
Peak flux has **increased** from 6.7 to 9.1 mJy

Background RMS has **decreased** from 18.2 to 8.2 uJy and noise looks more Gaussian

Dynamic range has **improved** by a factor of 3, from 370 to 1100!

Science target (central source with bent tail) is more clearly visible

Next Steps

Decide to stop: the new image may now be good enough to meet your science goals, so there may be no need to go further

Modify Image Parameters: the new image may now be limited by one or more issues that self-cal can not improve

Iterate on Selfcal Parameters: it's easy to undo selfcal and try again, just remember to restore any flags written by applycal

Additional Rounds of Phase-only Self-calibration: sometimes it may take several iterations to reach a satisfactory result

Next Steps

Amplitude Self-calibration: after taking phase-only selfcal as far as possible, you may want to try including amplitude corrections. It will be important to clean very deep and/or normalize the amplitude solutions in order to preserve the flux-density scale.

Bandpass Self-calibration: running the bandpass task instead of gaincal will trade frequency resolution for signal-to-noise. The solutions can be a function of time or just a single solution interval

Direction-dependent Self-calibration: peeling is possible in CASA but requires several steps. Scripting the peeling process is highly recommended. Other more advanced techniques require 3rd party software

References

CASA Guides:
casaguides.nrao.edu

CASA Docs:
casa.nrao.edu/casadocs

Synthesis Imaging in Radio Astronomy II:
ui.adsabs.harvard.edu/abs/1999ASPC..180.....T

www.nrao.edu
science.nrao.edu
public.nrao.edu

The National Radio Astronomy Observatory is a facility of the National Science Foundation
operated under cooperative agreement by Associated Universities, Inc.