Processing LOFAR VLBI data: the Long Baseline pipeline

Alexander Drabent (TLS Tautenburg) & Marco Iacobelli (ASTRON)

5th LOFAR Data School, 20th-21st Sept 2018

ASTRON
Netherlands Institute for Radio Astronomy

LOFAR

Overview

Tutorial topics

- General notes about calibration strategy (see also L9 by LB)
- Pipeline workflow
- Diagnostic plots
- Long baseline imaging

Tutorial goals

- Finding good (primary & secondary) calibrators
- Assessing data quality
- Calibrating and imaging long baseline uv-data

T7 part I

Tutorial topics

- **General notes about calibration strategy (see also L9 by LB)**
- **Pipeline workflow**
- Diagnostic plots
- Long baseline imaging

Tutorial goals

- **Finding good (primary) calibrators**
- Assessing data quality
- Calibrating and imaging long baseline uv-data

CEP3 login and use

6 working nodes reserved: lof006,lof010,lof012,lof015,lof016,lof017
25 active users accounts: lods01, . . ., lods25
1 active Slurm reservation with id=lofar_school2018_114

Work in couples!

Username=lodsXX | working node=lof01X
Accounts from lods01 to lods05 => lof006
Accounts from lods06 to lods10 => lof012
Accounts from lods11 to lods15 => lof015
Accounts from lods16 to lods20 => lof016 **T7**
Accounts from lods21 to lods25 => lof017 **T7**

General notes about the needed software

How to

Set up needed data & software on CEP3

```
> module load dysco lsmtool ; module unload lofar ; module load lofim; module load rmextract
```

Configure parset and configuration files

Where

Parset and configuration files

```
> cp -f /home/iacobelli/T7/*.parset /data/scratch/<your wd>/
> cp -f /home/iacobelli/T7/*.cfg /data/scratch/<your wd>/
> geany data/scratch/<your wd>/long_baseline_pipeline.parset data/scratch/<your wd>/lb_pipeline.cfg &
```

Raw data

```
> cp -rf /data009/scratch/iacobelli/L665012_3C147raw/*.MS_dysco /data/scratch/<your wd>/DATA/
```

Pipeline and other scripts (=>/home/iacobelli/T7/)

General notes about the data

Calibrator 3c295 and target 3c147 scans

- the bandwidth is 120-187MHz
- the duration is 10min

Raw data inspection

- data size is 656 GB . . multiple copies will take some time so . .
- typically, a long synthesis (i.e. duration ≥6 hours) observation consists of (at least) 243 sub-bands:

What is the MS size for a long synthesis observation ?

What is the total data volume ?

How does this compare with data from other (radio) telescopes ?

General notes about the data

Raw data inspection

to show details of the observation type:

```
> msoverview in=L665012_SAP000_SB000_uv.MS_dysco verbose=T
```

Note the message 'This is a raw LOFAR MS (stored with LofarStMan)'

i.e. the data cannot be handled with CASA software! (=> DPPP)

What array configuration was used and what it mean ?

What was the duration of the observation ?

Which field was observed ?

How many channels (frequencies) are in the data set ?

What was the centre frequency of this particular sub-band ?

What is the number of time slots ?

What is the integration time per time step ?

How many stations (core and remote), and how many baselines ?

What is the relation between number of stations and baselines (no autocorrelations) ?

General notes about the LB pipeline

How to

Configure parset and configuration files

```
########################################################################
## PARAMETER SETUP
## SECTION 1: things that absolutely need to be configured

## global information
! base_directory                = input.output.working_directory  ## this is taken from the config file -- DO NOT CHANGE
! job_directory                 = input.output.job_directory      ## this is taken from the config file -- DO NOT CHANGE

! long_baseline_pipeline_dir    = /home/iacobelli/long_baseline_pipeline
! prefactor_dir                 = /home/iacobelli/tutorialprefactor
! losoto_executable             = /home/drabent/losoto/bin/losoto

## target data information
! target_input_path             = /data/scratch/iacobelli/L665012_3C147raw
! target_input_pattern          = L*dysco

## Prefactor solution information
! transfer_amp_clock_sols_store = /data/scratch/iacobelli/L665004_3C295_RESULTS ##/data/scratch/iacobelli/prefactor_sols/cal_values/ ## directory with amplitude and clock solutions of p
! amp_sols_basename             = caldata_transfer
! phase_sol_input_path          = /data/scratch/iacobelli/L665012_3C147results/ ## directory with measurement sets containing direction-independent phase solutions of the target field

! image_cat                     = {{ job_directory }}/lotss_catalogue.csv
! delay_cat                     = /data/scratch/iacobelli/delay_calibrators.csv
! subtract_cat                  = {{ job_directory }}/subtract_sources.csv
! do_download                   = False

## Averaging parameters
! cal_shift_avg_timestep        = 8    ## Calibrator time averaging: output should be 8 seconds
! cal_shift_avg_freqstep        = 8    ## Calibrator freq averaging: output should be 2 channels per subband
! tgt_shift_avg_timestep        = 16   ## Target time averaging: output should be 16 seconds
! tgt_shift_avg_freqstep        = 16   ## Target freq averaging: output should be 1 channel per subband

## Stations to flag
! flag_baselines                = [] ## for HBA data before October 2015, should set to: [ CS013HBA* ]
```

General notes about the LB pipeline

How to

Configure parset and configuration files

```
[DEFAULT]
lofarroot = /opt/cep/lofim/daily/Mon/lofar_build/install/gnucxx11_opt/
casaroot = /opt/cep/casacore/current
pyraproot = /opt/cep/pyrap/current
hdf5root =
wcsroot = /opt/cep/lofar/external/wcslib
aoflaggerroot=/opt/cep/aoflagger/current
pythonpath = /opt/cep/lofim/daily/Mon/lofar_build/install/gnucxx11_opt/lib64/python2.7/site-packages
runtime_directory = /data/scratch/iacobelli/Pipeline_prefactor
recipe_directories = [%(pythonpath)s/lofarpipe/recipes,/home/iacobelli/long_baseline_pipeline/,/home/iacobelli/tutorialprefactor/]
working_directory = %(runtime_directory)s
task_files = [%(lofarroot)s/share/pipeline/tasks.cfg]

[layout]
job_directory = %(runtime_directory)s/%(job_name)s

[cluster]
clusterdesc = /data/scratch/lb_tutorial/pipeline.clusterdesc

[deploy]
engine_ppath = %(pythonpath)s:%(pyraproot)s/lib:/opt/cep/pythonlibs/lib/python/site-packages
engine_lpath = %(lofarroot)s/lib:%(casaroot)s/lib:%(pyraproot)s/lib:%(hdf5root)s/lib:%(wcsroot)s/lib

[logging]
log_file = %(runtime_directory)s/%(job_name)s/logs/%(start_time)s/pipeline.log
xml_stat_file = %(runtime_directory)s/%(job_name)s/logs/%(start_time)s/statistics.xml

[feedback]
# Method of providing feedback to LOFAR.
# Valid options:
#    messagebus    Send feedback and status using LCS/MessageBus
#    none          Do NOT send feedback and status
method = none

[remote]
method = local
max_per_node = 8
```

General notes about running the pipeline

How to run it

In a screen session (this will take a while ..)

```
> genericpipeline.py -d -c pipeline.cfg long_baseline_pipeline.parset
```

Performance up to 1st loop:

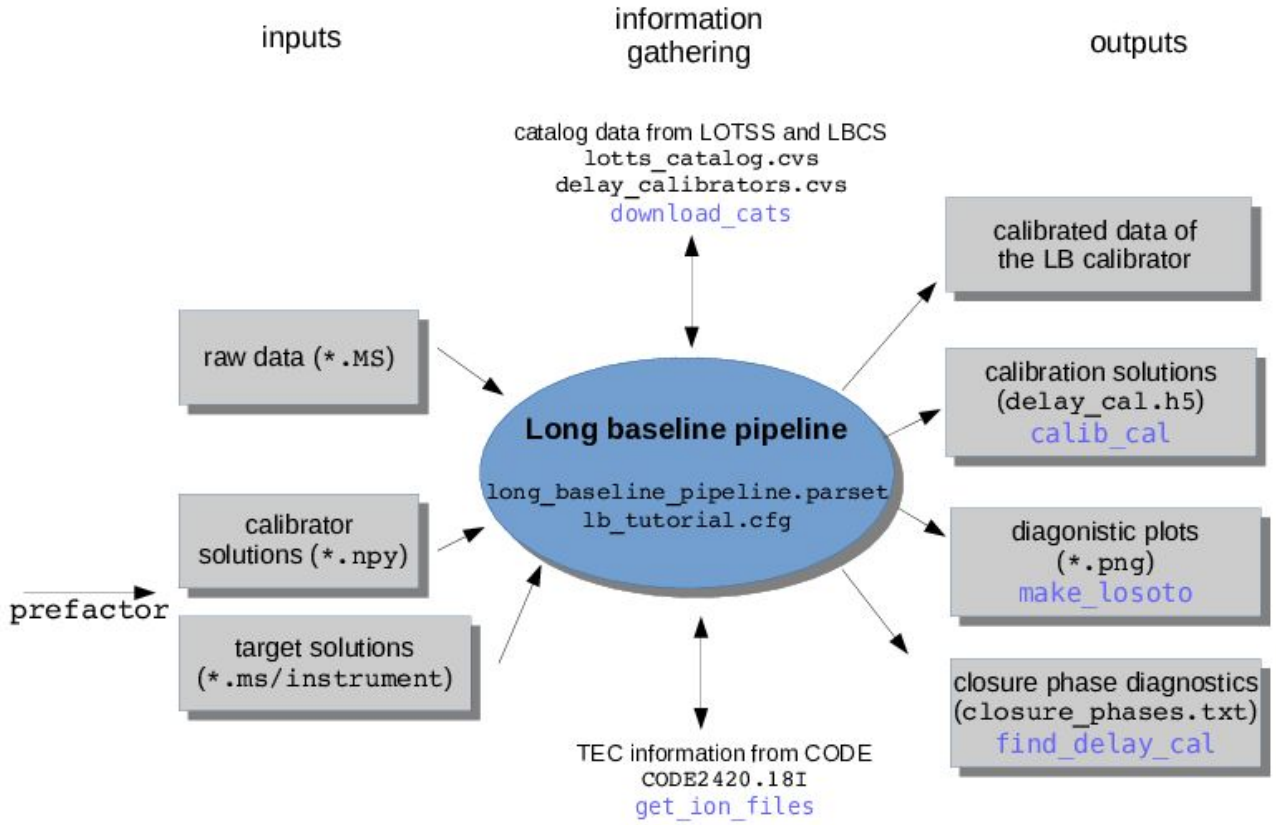
running time is 6.0 hrs data size increases of a factor 1.6

preparation section: 3hrs

find_delay_cal: 0.5 hrs

ndppp_apply_cal: 2.5 hrs

General notes about the LB pipeline



LOFAR advanced processing strategies

State of art pipelines to automatically reduce and calibrate LOFAR data are now available (and in progress)

- Pipeline for LOFAR LBA data* > https://github.com/lofar-astron/PiLL
- Pipelines for LOFAR HBA data:
  - Pre-Facet (i.e. DIE) calibration* > https://github.com/lofar-astron/prefactor
  - Initial-Subtract imaging* > https://github.com/lofar-astron/prefactor
  - Factor (i.e. DDE) calibration > https://github.com/lofar-astron/factor
  - Long Baseline calibration* > . . . Coming soon !

PiLL: tool for producing low-noise, mid-resolution wide-field images

Pre-Facet & Initial-Subtract: tools for producing moderate-noise, mid-resolution wide-field images

FACTOR: tool for producing low-noise, high-resolution wide-field images

LOFAR advanced processing strategies

Pipeline for LOFAR LB data > https://github.com/lmorabit/long_baseline_pipeline

a generic pipeline implementation of the LOFAR long baseline reduction pipeline

Pipeline workflow

What is the pipeline doing?

1. applying the prefactor solutions to the unaveraged data (1s,16ch/sb)
2. identify good delay calibrators . . . via closure phases
3. begin the calibration with “best” calibrator

Pipeline workflow

What is the pipeline doing?

1. applying the prefactor solutions to the unaveraged data (1s,16ch/sb)

Let us have a look at the prefactor solutions: (plots & h5 parm)

```
> python
>>> import h5py
>>> filename = 'instrument.h5imp_cal'
>>> data = h5py.File(filename, 'r')
>>> for key in data.keys(): print(key)
Sol000
>>> data['sol000'].keys()
[u'RMextract', u'XYoffset', u'XYoffset_notimes', u'amplitude000', u'antenna',
u'bandpass', u'bandpass_notimes', u'clock', u'clock000', u'phase000', u'phase_offset000',
u'rotation000', u'source', u'tec000']
```

Tomorrow you will have a new h5 file as output of the pipeline: check the content . .

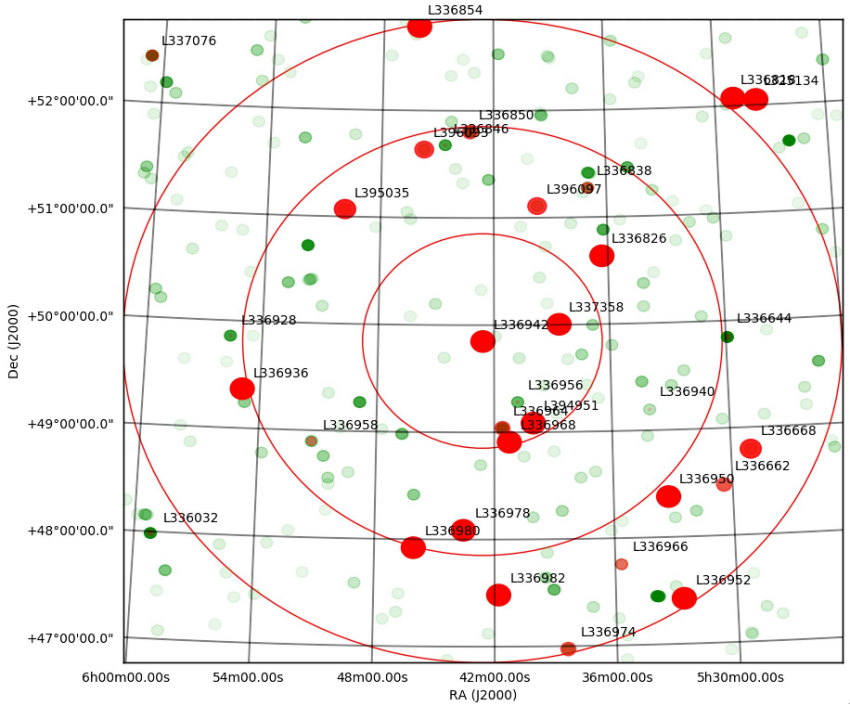
Pipeline workflow

What is the pipeline doing?

2. Identify good delay calibrators . . .
   inspection of sources in the target field

```
> cp -f /data/scratch/LDS2018/lbcs_plot.py .
> python lbcs_plot.py 85.650575 49.852009
```

What do we see here?
To which lbcs source corresponds 3C147?



Pipeline workflow

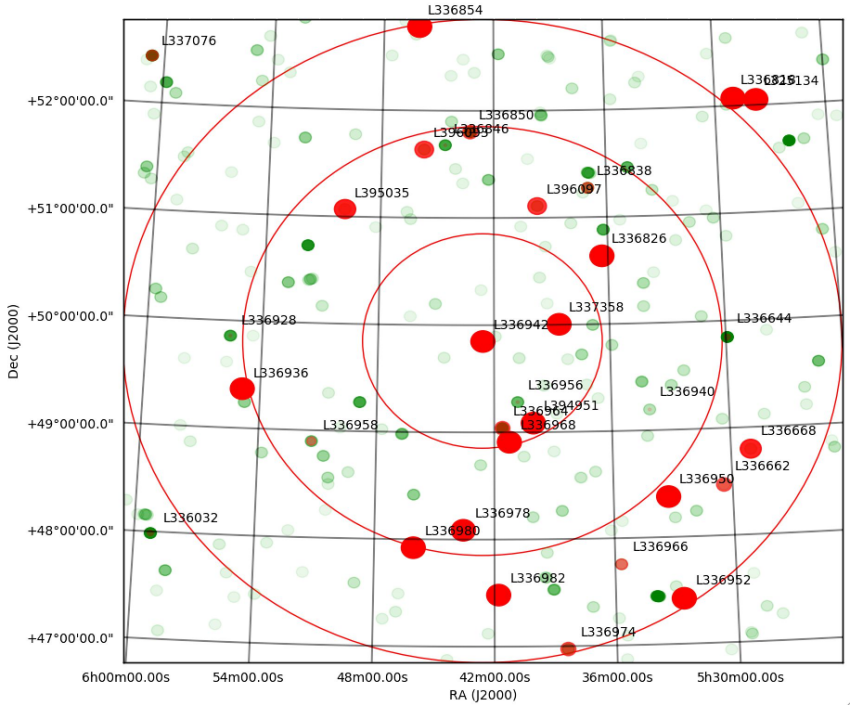
What is the pipeline doing?

2. Identify good delay calibrators . . .
   inspection of sources in the target field

```
> cp -f /data/scratch/LDS2018/lbcs_plot.py .
> python lbcs_plot.py 85.650575 49.852009
```

What do we see here?
To which lbcs source corresponds 3C147?

LBCS sources in the 3C147 field.
LBCS sources are plotted larger & redder the more coherence is seen on the longer baselines. WENSS sources are plotted in green, with brighter sources in darker green. Note that a bright WENSS source is not necessarily a good long-baseline calibrator.
Plotting software is available on GitHub https://github.com/nealjackson/lofar-lb

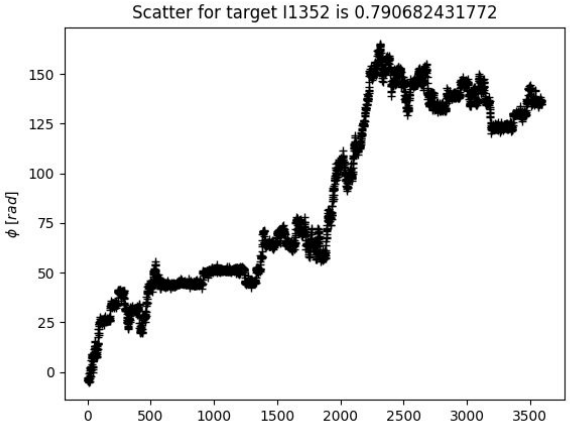


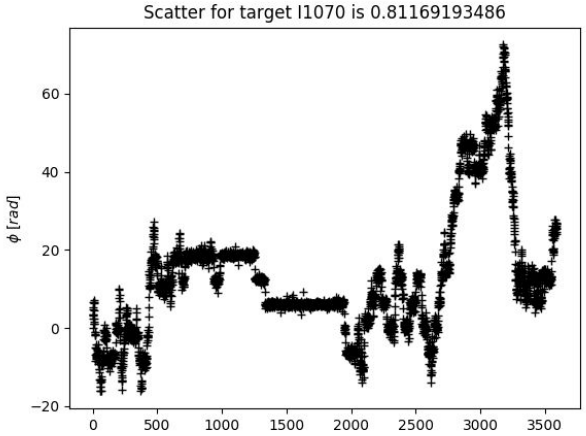
Pipeline workflow

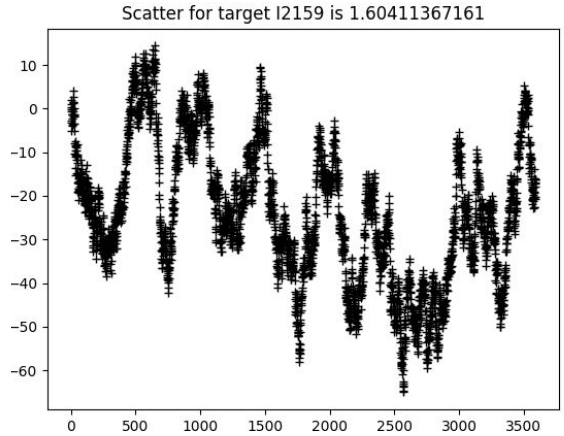
What is the pipeline doing?

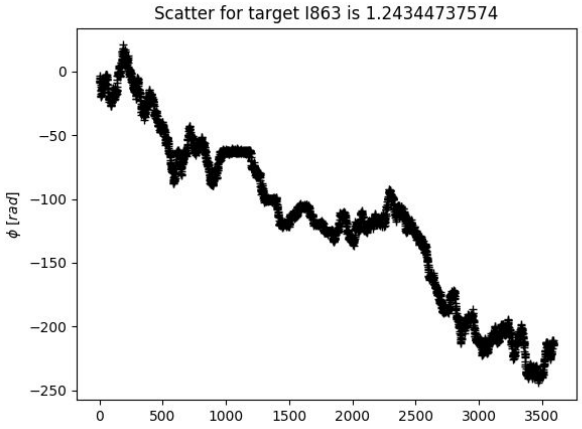
2. Identify good delay calibrators . . .
   Closure phase plots for single calibrator sources !

What do we see here?
Which is the best source ?
Tomorrow check the pipeline output closure_phase.txt









Pipeline workflow

What is the pipeline doing?

3. begin the calibration with “best” calibrator
   inspection of Losoto diagnostic plots

```
> display /your/job/directory/results/inspection/*.png
```

What do we see here?

Pipeline workflow

What is the pipeline doing?

3. begin the calibration with “best” calibrator
   inspection of Losoto diagnostic plots

```
> display /your/job/directory/results/inspection/*.png
```

What do we see here (per station)?

delay_cal_amp_polXX.png => amplitude solutions(XX polarization) in colorcode (time vs. freq)

delay_cal_ph_polXX.png => phase solutions(XX polarization) in colorcode (time vs. freq)

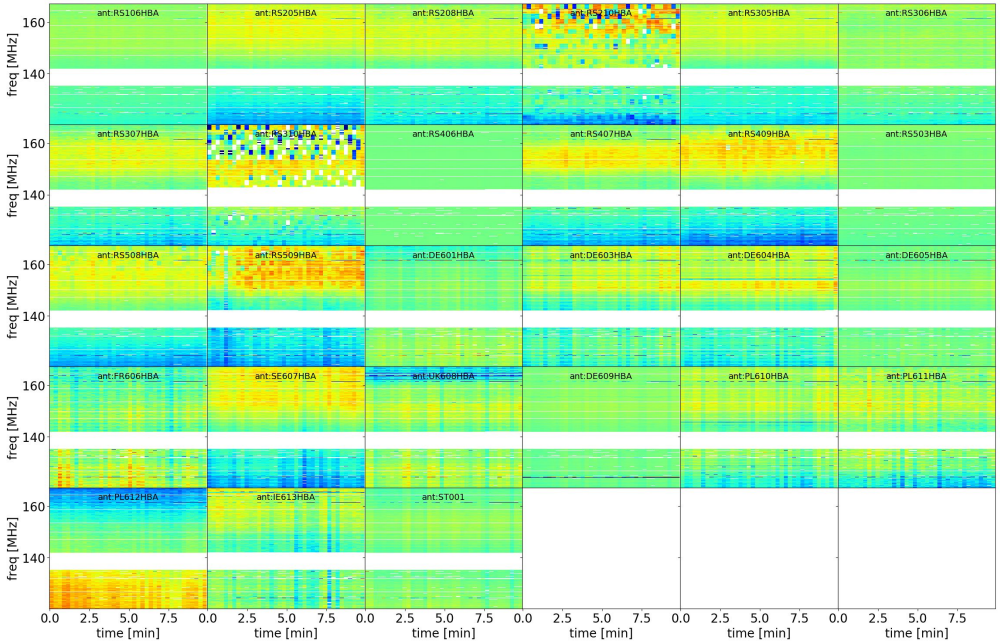
delay_cal_clock.png => clock offset (in seconds) with time

delay_cal_tec.png => differential TEC (in TECU) with time

Pipeline workflow

What is the pipeline doing?

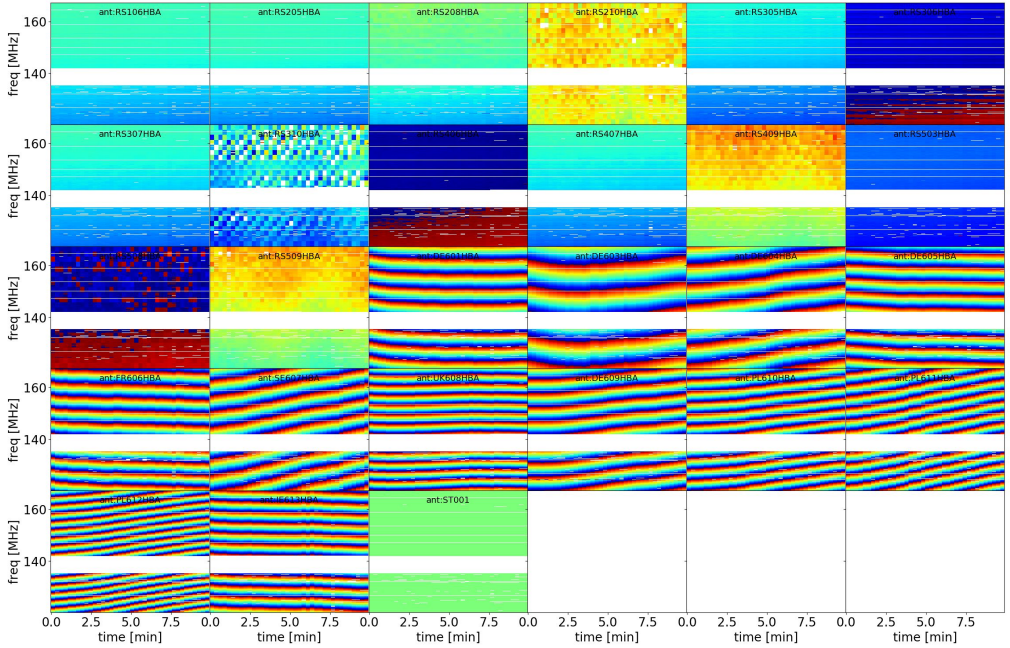
3. begin the calibration with “best” calibrator
   inspection of LoSoTo diagnostic plots



Pipeline workflow

What is the pipeline doing?

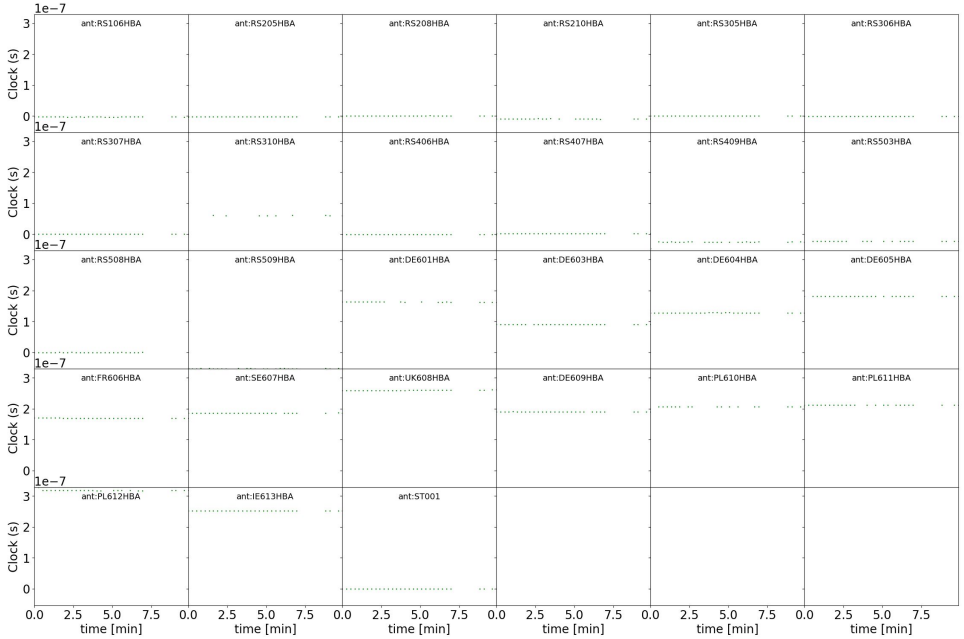
3. begin the calibration with “best” calibrator
   inspection of LoSoTo diagnostic plots



Pipeline workflow

What is the pipeline doing?

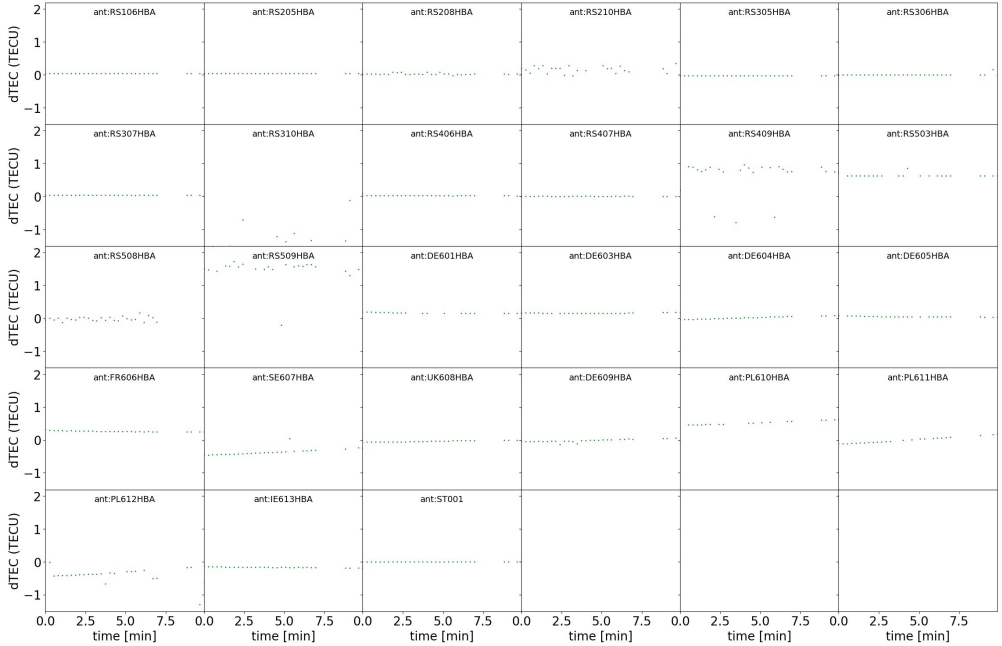
3. begin the calibration with “best” calibrator
   inspection of LoSoTo diagnostic plots



Pipeline workflow

What is the pipeline doing?

3. begin the calibration with “best” calibrator
   inspection of LoSoTo diagnostic plots



T7 part II

Tutorial topics

- General notes about calibration strategy (see also L9 by LB)
- Pipeline workflow
- **Diagnostic plots**
- **Long baseline imaging**

Tutorial goals

- Finding good (primary) calibrators
- **Assessing data quality**
- **Calibrating and imaging long baseline uv-data**

Pipeline workflow

What has the pipeline done?

3. begin the calibration with “best” calibrator
   inspection of LoSoTo diagnostic plots

```
> display /your/job/directory/results/inspection/delay_cal_tec.png
> display /your/job/directory/results/inspection/delay_cal_clock.png
> display /your/job/directory/results/inspection/delay_cal_ph_polXX.png
> display /your/job/directory/results/inspection/delay_cal_amp_polXX.png
```

Imaging

How does the source look like ?

For imaging see L9 and T4

```
> wsclean -j %s -mem %s -use-differential-lofar-beam -no-update-model-required
-reorder -local-rms -auto-threshold 1 -auto-mask 3 -niter %s -mgain 0.65 -multiscale -name
%s -size %s %s -scale %sasec -weight briggs %s -fit-beam -pol I -channels-out %s
-data-column %s <a sub-set of your calibrated *MS>
```

To prepare your data make a parset file to average and concatenate MS (see T2 & T3 and https://www.astron.nl/lofarwiki/doku.php?id=public:user_software:documentation:ndppp

Which averaging factors ?

```
> DPPP <avg & concat>.parset
```

Imaging

How does the source look like ?

For imaging see L9 and T4

```
> wsclean -j %s -mem %s -use-differential-lofar-beam -no-update-model-required
-reorder -local-rms -auto-threshold 1 -auto-mask 3 -niter %s -mgain 0.65 -multiscale -name
%s -size %s %s -scale %sasec -weight briggs %s -fit-beam -pol I -channels-out %s
-data-column %s <a sub-set of your calibrated *MS>
```

```
nthreads 6
fraction_memory 30
niter 10000
image_name /you/can/choose/it
npixels 3600
cellsize 0.5asec
robustBRIGGS 0.0
numCHAN 1
datacol CORRECTED_DATA
inMS /your/data/scratch/dir/where/calibrated/L665012_SAP000_SB*_uv.ndppp_prep_target
```

Imaging

How does the source look like ?

For imaging see L9 and T4

Useful online resources

The LOFAR documentation at
https://www.astron.nl/radio-observatory/lofar-documentation/resources/resources

The LOFAR imaging cookbook at
https://www.astron.nl/radio-observatory/lofar/lofar-imaging-cookbook

Software processing tools at
https://www.astron.nl/radio-observatory/lofar-data-processing/software-processing-tools/software-processing-tools and https://github.com/lofar-astron

ASTRON

Netherlands Institute for Radio Astronomy

LOFAR