

LOFAR UV-DATA PROCESSING TUTORIAL

Marco Iacobelli (ASTRON)

European Radio Interferometry School, 19th Oct 2017





Overview

Tutorial topics

- General notes about low frequency data (see also L4,L9 by JC,VM)
- Introduction to main LOFAR software tools
- Basic processing of LOFAR (HBA) uv-data
- Introduction to LOFAR advanced processing strategies

Tutorial goals

- Inspection of (raw) data
- Editing (both manual and automatic) of data (e.g. flagging)
- Direction independent effects (DIE) calibration of data
- Wide field imaging of data
- Sky model extraction
- Self calibration
- Correction for ionospheric Faraday rotation
- Subtraction of bright sources

General notes about low frequency data

- Observing at low frequency with LOFAR
 - large data sizes: typical size of a raw/preprocessed uv-dataset is 100/30TB
 - Direct inspection by eye prohibitive => proper data visualization needed to identify trends (e.g. in time / frequency) and systematics in data
 - multi-beam capabilities: solutions transfer, innovative calibration strategies
 - Many beams at different levels => complex response of the side lobes pattern
 - wide field of view: Direction Dependent (DDE) vs independent (DIE) calibration
 - calibration and/or imaging computational intensive
 - DIE, DDE, or DIE+DDE approaches to be evaluated based on different factors and needs (e.g. the presence of bright off-axis sources to be subtracted, the need for the deepest map ..)
 - the impact of the ionosphere on both continuum and polarimetry
 - Both flux of *all* Stokes parameters and position of sources affected; also correction for ionospheric Faraday rotation is needed
- Large data sizes + computational intensive processing
 - lots of challenges for developing innovative, dedicated, fast and accurate algorithms and software tools
 - 1 MS to be used in this tutorial !

LOFAR software processing tools

- In this tutorial we will use some of the main software tool developed for LOFAR and other low frequency telescopes data processing, such as:
 - msoverview => casacore based tool for quick metadata inspection
 - DPPP & (Gaincal) => pipeline to perform some operations on data
 - BBS => a software package designed for the calibration (both DIE and DDE) and simulation of LOFAR data.
 - RMextract => python based tool to extract TEC, vTEC, Earth magnetic field and Rotation Measures from GPS and WMM data for radio interferometry observations and applying Ionospheric RM corrections
 - pyBDSF => python based source finder
 - WSClean => a fast generic widefield imager

LOFAR software processing tools

DPPP & Gaincal:

- The only program that can read raw correlated LOFAR data and writes it out as the standard (CASA) MS-format.
- Command-line tool, input as a parset, output as feedback on screen. E.g. > DPPP reduction.parset
- Each step operates on the output of the previous step. Typical steps are:
 - Averaging: time and frequency domain; also combine stations
 - Demixing: subtraction of calibrated model visibilities of bright sources
 - Flagging
 - Direction independent calibration (amplitudes, phases)
 - Phase shift

The tutorial data set & CEP3 usage

- HBA calibration of a field of view dominated by a bright source near the field center
 - The case of a calibrator field; no transfer of solution and 1MS only



CEP3 login and usage

- 3 working nodes reserved: lof015 lof016 lof017
- 15 active user accounts: lods01...lods15
- 1 active reservation with id=eris2017_55
- Work in couples!
 - Username=lodsXX | working node=lof01X
 - Accounts from lods01 to lods05 => lof015
 - Accounts from lods06 to lods10 => lof016
 - Accounts from lods11 to lods15 => lof017

Log in to the LOFAR portal: > ssh -Y lodsXX@portal.lofar.eu Log in to the head node: > ssh -Y lodsXX@lhdhead.offline..lofar Activate a dummy session using the reservation id and the your node: srun -A eris2017 --reservation=eris2017_55 -N 1 -w lof01X -u bash -i

CEP3 login and usage

- 3 working nodes reserved: lof015 lof016 lof017
- 15 active user accounts: lods01...lods15
- 1 active reservation with id=eris_55
- Work in couples!
 - Username=lodsXX | working node=lof01X
 - Accounts from lods01 to lods05 => lof015
 - Accounts from lods06 to lods10 => lof016
 - Accounts from lods11 to lods15 => lof017

Now, open a new terminal tab (keep the previous one open too!) and type: > ssh -Y lodsXX@portal.lofar.eu > ssh -Y lodsXX@lhdhead.offline.lofar > ssh -Y lodsXX@lof01X

- 3 working nodes reserved: lof015 lof016 lof017
 - Data and scripts are in /data/scratch/ERIS2017 so first make your working directory:
 - > mkdir /data/scratch/lodsXX
 - > cd /data/scratch/lodsXX
 - To source main software packages, type:
 - > module load wsclean
 - Which modules you loaded ?
 - Let us copy the data:

> cp -rf /data/scratch/ERIS2017/L74759_SAP000_SB000_uv.MS .

• Data size is 11 GB . . multiple copies will take some time so . .

- Data size is 11 GB . . multiple copies will take some time so . .
 - Typically, a long synthesis (i.e. duration ≥6 hours) observation consists of (at least) 244 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 ?
 - The MS (sub-band) naming convention:
 - Lxxxxxx is the observation/pipeline ID
 - SAPxxx is the sub-array pointing (beam)
 - SBxxx is the sub-band number.

- To show details of the observation type:
 > msoverview in=L74759_SAP000_SB000_uv.MS 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 (excluding autocorrelations) ?

 Check pointing elevation and sky configuration of bright sources (A-team) wrt our beam:

> python plot_Ateam_elevation.py L74759_SAP000_SB000_uv.MS

- What are closest sources?
- Which are their fluxes around the MS reference frequency?
- In case of close A-team sources we can simulate (i.e. "predict") the associated visibilities and subtract ("demix") them using DPPP. But this is a time consuming processing step.



Flagging, averaging and inspection

- Data flagging and compression (to 4ch,4sec) carried out using DPPP.
 - Max averaging in frequency set by the size of the field to be imaged and the amount of bandwidth smearing at the edges of the field.
 - Time averaging limited by the amount of time smearing and by the changes in the ionosphere.
 - A simple DPPP parset is required (edit the msin and msout fields to point at your working directory):
 - > cat DPPP_HBA_preprocess.parset
 - > cp -rf /data/scratch/ERIS2017/DPPP_HBA_preprocess.parset .
 - > geany DPPP_HBA_preprocess.parset &
 - > DPPP DPPP_HBA_preprocess.parset
 - DPPP will take about ~5 minutes to flag and average the data. The screen lists input / output parameters, the level of flagging at each step and the total amount of data flagged.
 - The total data flagged for each of the flagging steps is 4.7%, 3.4% and 2.6% respectively.
 - The flagged and compressed MS has a much more manageable size of . . . ?

Flagging, averaging and inspection

- Inspection of flagged & averaged data done using CASA:
 - > module purge
 - > module load casa
 - > casaplotms



- Plot Amplitudes vs. time and UV distance (wavelengths) by selecting
 "Data Column: data" and plot only the XX and YY correlations.
 - Which station shows lower/ higher amplitudes?
 - From the screen info of initial DPPP preprocessing step you will notice that ~40-50% of the data for this station was flagged by aoflagger.

Flagging, averaging and inspection

- Removal of CS302HBA0 with DPPP. This can be done either by:
 - flagging it with a preflagger step
 - filtering it out. In this way it is completely removed from the MS so we have to specify a new output MS.
 - > module purge
 - > module load wsclean
 - The parset for this quick operation is:

> cp -rf /data/scratch/ERIS2017/DPPP_filter_baseline.parset
> DPPP DPPP_filter_baseline.parset

- Plot again visibility amplitude against UV distance in wavelengths
- What is the size of the new MS ? Does filtering help to optimize disk space?

- To calibrate we need:
 - an initial sky model for correcting the data
 - a parset file to direct the calibration
- Sky model generation can be done in different ways:
 - In case the bright dominant source is a well-known calibrator source (like in this tutorial), good model for it generally are available.
 - > cp -f /data/scratch/ERIS2017/3C295.skymodel .
 - In general, an initial sky model is based on what we think the sky looks like at the frequency and resolution we are interested in. This means constructing a model from:
 - good image we have at a different frequency/resolution
 - in the case of self-calibration, the image we have just made.
 - Alternatively, a sky model can be created browsing astronomical catalogues and databases, e.g.:
 - You can use the TGSS ADR source catalog for generating a sky model to calibrate your LOFAR data:

> wget -O GSM3C295field.skymodel 'http://tgssadr.strw.leidenuniv.nl/cgi-bin/ gsmv3.cgi?coord=3c295&radius=5&unit=deg&deconv=y&cutoff=0.3'

- To calibrate we need:
 - an initial sky model for correcting the data
 - a parset file to direct the calibration

	Shape	Calibration type	Free parameters
	$\mathbf{G}_{p} = \begin{pmatrix} A_{xx}^{(p)} e^{\phi_{xx}^{(p)}} & A_{xy}^{(p)} e^{\phi_{xy}^{(p)}} \\ A_{yx}^{(p)} e^{\phi_{yx}^{(p)}} & A_{yy}^{(p)} e^{\phi_{yy}^{(p)}} \end{pmatrix}$	fulljones	8
a	$\mathbf{G}_{p} = \begin{pmatrix} A_{xx}^{(p)} e^{\phi_{xx}^{(p)}} & 0\\ 0 & A_{yy}^{(p)} e^{\phi_{yy}^{(p)}} \end{pmatrix}$	diagonal	4
	$\mathbf{G}_{p}=\left(egin{array}{cc} e^{\phi_{ ext{xc}}^{(p)}} & 0 \ 0 & e^{\phi_{ ext{yy}}^{(p)}} \end{array} ight)$	phaseonly	2
	$\mathbf{G}_p = \left(egin{array}{cc} e^{\phi^{(p)}} & 0 \ 0 & e^{\phi^{(p)}} \end{array} ight)$	scalarphase	1

 To be used by DPPP the sky model needs to be converted to a sourcedb format:

> makesourcedb in=3C295.skymodel out=3C295.sourcedb format='<'</p>

- The parset file for DPPP solves and corrects the data.
 > cp -rf /data/scratch/ERIS2017/DPPP_calibrate.parset .
 > DPPP DPPP_calibrate.parset
- The calibration process will be completed in about 5 minutes.

- It is useful to look at the solutions of calibrated data. Several tools are available.
 - The LOFAR Solution Tool (LoSoTo): a Python package to handle LOFAR solutions in many ways:
 - ABS: absolute value
 - CLIP: flag above/below a certain sigma
 - CLOCKTEC: clock/tec separation
 - FLAG: remove trends and flag outliers
 - INTERP: interpolate solutions along (even multiple) axis
 - NORM: normalize solutions
 - PLOT: 1D, 2D, movie
 - RESIDUALS: subtract a table from another
 - REWEIGHT: used to flag antennas, channels...
 - RESET: to default values
 - SMOOTH: smooth along (even multiple) axis
 - TECFIT: fit TEC values to phase solutions
 - TECSCREEN: fit TEC screens to TEC values
 - The data files used by LoSoTo are not in the standard parmdb format used by BBS/ NDPPP (e.g. the "instrument" table). LoSoTo uses instead an innovative data file, called H5parm, which is based on the HDF5 standard.

- It is useful to look at the solutions of calibrated data.
 Several tools are available.
 - parmdbplot.py (basic LOFAR default tool); from the GUI select stations and look at solutions (eventually changing the resolution that is used to display solutions).
 - > parmdbplot.py L74759_SAP000_SB000_uv.MS.avg.dppp.filt/instrument/

•••	🗙 instrument	
Gain:0:0:CS001HBA0 Gain:0:0:CS001HBA1 Gain:0:0:CS002HBA0 Gain:0:0:CS002HBA1 Gain:0:0:CS003HBA0 Gain:0:0:CS003HBA0		
Gain:0:0:CS004HBA0 Gain:0:0:CS004HBA1 Gain:0:0:CS005HBA0 Gain:0:0:CS005HBA1 Gain:0:0:CS006HBA0 Gain:0:0:CS006HBA1		_
Use resolution	1953.125 Hz	36.010 s
Plot	DefValues	Close figures



- It is useful to look at the solutions of calibrated data. Several tools are available.
- It is useful to look at the solutions of calibrated data with:
 - Iofar.parmdb (python module) to read and plot solutions in Python. Use this example script to plot all the phase and amplitude solutions in single images
 - > cp -f /data/scratch/ERIS2017/plot_solutions_all_stations_v0.py .
 - > python plot_solutions_all_stations_v0.py -s -n
 - L74759_SAP000_SB000_uv.MS.avg.dppp.filt/instrument/ SB000_gains
 - > display SB000_gains_amp.png
 - > display SB000_gains_phase.png
 - Which are suspicious solutions (antennas)?

- Finally let us inspect the calibrated with CASA.
 - > module purge
 - > module load casa
 - > casaplotms



- Plot Amplitudes vs. time and UV distance (wavelengths) by selecting "Data Column: corrected" and plot only the XX and YY correlations. Use color scheme 'Antenna1'
 - Which station(s) baseline(s) shows bad amplitudes?

- Let us go back, flag bad station and baselines and re-do the calibration
 cp -f /data/scratch/ERIS2017/DPPP_flag.parset.
 - > module purge
 - > module load lofar
 - > DPPP DPPP_flag.parset
 - > DPPP DPPP_calibrate.parset



- Plot Amplitudes vs. time and UV distance (wavelengths) by selecting "Data Column: corrected" and plot only the XX and YY correlations. Use color scheme 'Antenna1'
- Plot Uwave vs Vwave by selecting "Data Column: corrected" and plot only the XX and YY correlations. Use color scheme 'Antenna1':
 - How are distributed visibilities ?
 - What does it mean for imaging ?

- To image the CORRECTED_DATA let us copy them to a new MS
 - > cp -f /data/scratch/ERIS2017/DPPP_newms.parset .
 - > DPPP DPPP_newms.parset
- Deconvolution of 3C295 field of view done with Wsclean.
 - The list of parameters and a brief description of each is obtained with: > wsclean
 - At 120 MHz the LOFAR field of view is 4.5 deg (FWHM) and the theoretical resolution should be around 8". Because of the limited uv-coverage we will limit to a resolution of 25-30".
 - Which is the corresponding uv cut in units of λ ?

- We limit the usage of computing resources (-j 10 -mem 30)
- Also we adopt:
 - a square image with size of 4704 pixels (-size 4704 4704)
 - a slightly uniform weighting (-robust -0.5)
 - a pixel size of 7.5" (-scale 0.00208) to properly sample the beam
 - to pad by a factor of 1.4 we set the size and trim parameters appropriately to avoid aliasing effects
 - Given the scale, size and pad values which is the size in degrees expected for the FoV ?

• Putting all together we get:

> wsclean -name 3C295field_SB000 -size 4704 4704 -mgain 0.65 auto-mask 3 -pol I -minuv-I 80 -maxuv-I 7000 -weighting-rank-filter 3 auto-threshold 0.3 -j 10 -mem 30 -weight briggs -0.5 -scale 0.00208 niter 40000 -trim 3360 3360 -apply-primary-beam -use-differentiallofar-beam -local-rms -multiscale -no-update-model-required reorder -fitbeam L74759_SAP000_SB000_uv.MS.avg.dppp.flag.dppp

This will take <u>about 15 minutes</u> and give you a very large image, with sources all over. It is better to run it on a screen session, i.e. type:

- > screen (next press ENTER 2 times)
- > module load wsclean
- > wsclean . . .
 - To detach from your screen session press: CTRL+A+D
 - To attach again to your session type: > screen -r

- WSClean will write several images as output:
 - *-beam-*.fits => the 8 components of the beam Jones matrix
 - *-dirty.fits => the image without cleaning
 - *-image.fits => the cleaned image
 - *-image-pb.fits => the cleaned primary beam corrected image
 - *-model.fits => the clean component image
 - *-psf.fits => the point spread function image
 - *-residual.fits => the image of residuals
- To check fits images:
 - > module load ds9
 - > ds9 -scale limits -0.1 1 3C295field_SB000-image.fits
 - What is the noise level and the PSF size of the cleaned image ?
 - What is the flux density of 3C 295?

DIE phase only calibration with DPPP

Shape

- To calibrate we need:
 - an initial sky model for correcting the data $_{\mathbf{G}_{p}=\left(A_{xx}^{(p)}e^{\phi_{xy}^{(p)}}\right) }$
 - we can use the TGSS ADR one but remember to convert it to a sourcedb format:



Calibration type

Free parameters

> makesourcedb in=GSM3C295field.skymodel out=GSM3C295field.sourcedb format='<'</p>

- a parset file to direct the calibration
 - > cp -f /data/scratch/ERIS2017/DPPP_calibrate_phaseonly.parset .
 - Check differences against the previous calibration step . . .
- before calibrating let us split the RS from the data and create a new MS
 - > cp -f /data/scratch/ERIS2017/DPPP_csonly.parset .
 - > DPPP DPPP_csonly.parset
 - How effective is this step to reduce the MS size ?

DIE phase only calibration with DPPP

- To calibrate we need:
 - an initial sky model for correcting the data
 - a parset file to direct the calibration



Let us re-image CORRECTED_DATA and check maps

> wsclean -name 3C295fieldCSONLY_SB000 -size 1960 1960 -mgain 0.65 -auto-mask 3 -pol I -minuv-I 80 -weighting-rank-filter 3 -auto-threshold 0.3 -j 10 -mem 30 -weight briggs -0.5 -scale 0.01 -niter 30000 -trim 1400 1400 -apply-primary-beam -usedifferential-lofar-beam -local-rms -multiscale -no-update-model-required -reorder fitbeam L74759_SAP000_SB000_uv.MS.avg.dppp.flag.dppp.csonly

Shape

 $\mathbf{G}_{p} = \begin{pmatrix} A_{xx}^{(p)} e^{\phi_{xx}^{(p)}} & A_{xy}^{(p)} e^{\phi_{xy}^{(p)}} \\ A_{yy}^{(p)} e^{\phi_{yx}^{(p)}} & A_{yy}^{(p)} e^{\phi_{yy}^{(p)}} \end{pmatrix}$

 $0 \\ A_{yy}^{(p)} e^{\phi_{yy}^{(p)}}$

0

 $\mathbf{G}_p = \begin{pmatrix} A_{xx}^{(p)} e^{\phi_{xx}^{(p)}} \\ 0 \end{pmatrix}$

The importance of having a good sky model . . .

CASE CS ONLY

Converged: 899, stalled: 0, non converged: 0, failed: 0 Iters converged: 1, stalled: 0, non converged: 0, failed: 0

CASE CS+RS

Calibration type

fulljones

diagonal

phaseonly

Free parameters

8

4

2

Converged: XXX, stalled: X, non converged: 0, failed: 0 Iters converged: X, stalled: X, non converged: 0, failed: 0

Try to merge sky models

Extract a sky model and a mask

- To this aim we use the LOFAR source detection tool PyBDSF (Python Blob Detection and Source Finder)
 - > module load pybdsf

> pybdsf

- BDSF [1]: process_image(filename='<*-image.fits>', rms_box=(60, 20))
- To inspect results

BDSF [2]: show_fit (+ Click on a Gaussian source to get info)

 For self-calibration and imaging purposes it is useful to save files BDSF [3]: export_image(img_type='island_mask', mask_dilation=0) BDSF [4]: write_catalog(format='<ascii/fits>', catalog_type='srl') BDSF [5]: exit (+ ENTER)

Ionospheric RM correction (with BBS)

•	То	cali	brate	we	need:
---	----	------	-------	----	-------

- an aux sky model for correcting the data
 cp -f /data/scratch/ERIS2017/dummy.skymode
- a parset file to direct the calibration
 cp -f /data/scratch/ERIS2017/BBS_RMcorr.parset .



- > module load rmextract
- > createRMParmdb <MS> --out <source name>_RMParmdb --IONprefix CODG
- let us plot the RM correction values with a simple script
 - > cp -f /data/scratch/ERIS2017/plotRMcorr.py .
 - > python plotRMcorr.py <source name>_RMParmdb
 - How large is the variation of RM over the observation duration ?
 - What is the corresponding rotation angle ?
- next we apply the RM correction to a new MS with BBS

> calibrate-stand-alone -f -t 2 --parmdb <source name>_RMParmdb <MS> BBS_RMcorr.parset dummy.skymodel

Can you write a parset for doing it with DPPP ?

Sł	nape	Calibration type	Free parameters
	$\mathbf{G}_{p} = \begin{pmatrix} A_{xx}^{(p)} e^{\phi_{xx}^{(p)}} & A_{xy}^{(p)} e^{\phi_{xy}^{(p)}} \\ A_{yx}^{(p)} e^{\phi_{yx}^{(p)}} & A_{yy}^{(p)} e^{\phi_{yy}^{(p)}} \end{pmatrix}$	fulljones	8
_	$\mathbf{G}_{p} = \left(\begin{array}{cc} A_{xx}^{(p)} e^{\phi_{xx}^{(p)}} & 0\\ 0 & A_{yy}^{(p)} e^{\phi_{yy}^{(p)}} \end{array}\right)$	diagonal	4
) .	$\mathbf{G}_p = \left(\begin{array}{cc} e^{\boldsymbol{\phi}_{xx}^{(p)}} & 0 \\ 0 & e^{\boldsymbol{\phi}_{yy}^{(p)}} \end{array}\right)$	phaseonly	2
	$\mathbf{G}_p = \left(egin{array}{cc} e^{\phi^{(p)}} & 0 \ 0 & e^{\phi^{(p)}} \end{array} ight)$	scalarphase	1
reat			

Imaging Stokes I,Q,U,V

- To check magnitude of polarization leakages towards the centre (no cleaning applied)
 - Stokes V (i.e. circular polarization) useful to check noise level, because the map should be empty.
 - Stokes Q,U (i.e. linear polarization)

> wsclean -apply-primary-beam -use-differential-lofar-beam -name 3C295QUV -no-update-model-required -pol IQUV -channelsout 1 -localrms -auto-threshold 0.3 -auto-mask 3 -niter 0 -mgain 0.65 -multiscale size 1960 1960 -scale 0.01 -trim 1400 1400 -apply-primary-beam -usedifferential-lofar-beam -weight briggs -0.5 -minuv-l 0.0 -fitbeam L74759_SAP000_SB000_uv.MS.avg.dppp.flag.dppp.csonly.rmcorr

Subtraction of sources with NDPP

- Subtraction of bright sources is sometimes needed and having a good model is mandatory for obtaining good results. Let us subtract 3C 295 from our data.
- To calibrate we need:
 - an initial sky model for the source
 3C295.sourcedb
 - a parset file to direct the subtraction

> cp -f /data/scratch/ERIS2017/DPPP_subtract.parset .

> DPPP DPPP_subtract.parset

• Finally let us image the subtracted data . . .

- State of art pipelines to automatically reduce and calibrate LOFAR data are now available (and in progress)
 - Pipeline for LOFAR LBA data* > <u>https://github.com/lofar-astron/PiLL</u>
 - Pipelines for LOFAR HBA data:
 - Pre-Facet (i.e. DIE) calibration* > <u>https://github.com/lofar-astron/prefactor</u>
 - Initial-Subtract imaging* > <u>https://github.com/lofar-astron/prefactor</u>
 - Facet (i.e. DDE) calibration > <u>https://github.com/lofar-astron/factor</u>
 - 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, midresolution wide-field images

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

Pipeline for LOFAR LBA data > <u>https://github.com/lofar-astron/PiLL</u>

It works on multi-beam observations calibrator + target(s) and is divided into steps:

- calibrate faraday rotation, cross-delay, bandpass, clock, tec, tec III order.
- transfer of bandpass (including ionospheric effects).
- selfcal: calibrate TEC, FR, cross-delay, station-averaged beam
- remove of 1st side lobe sources.
- DDE cal: uses new solver code for simultaneous TEC calibration, facet imaging

Toothbrush cluster @ 20" resolution. From DIE (rms~5 mJy/psf) to DDE (rms~1.5 mJy/psf) cal.



- Pre-Facet (i.e. DIE) calibration: processing the (amplitude-)calibrator, gain transfer to the target and direction-independent calibration of the target. The pipeline consists of 2 parts: for calibrator and target data
 - pre-process the input MSs
 - calibrate the pre-processed data with BBS (against an existing skymodel)
 - collect all calibration values in a LoSoTo file
 - do clock/TEC separation on the phases
 - flag and filter the amplitudes
 - determine the phase difference of the X- and Y-dipoles
 - generate diagnostic plots
 - create parmDBs with the RM correction from RMextract and the gain values from the calibrator observation for each input MS
 - pre-process the input MSs which also applying the values from the parmDBs
 - predict and flag the A-team contamination
 - sort the MSs by time and frequency into groups that are to be concatenated
 - concatenate the groups in frequency
 - filter out MSs with too little unflagged data
 - do phase-only calibration on the given skymodel
 - create some diagnostic plots

Marco Iacobelli – LOFAR Data Processing Tutorial – ERIS, 19th Oct 2017

LOFAR advanced processing strategies

Pre-Facet (i.e. DIE) calibration: <u>diagnostic plots</u>

Amplitude solutions



Good amplitude solutions: no stations that are clear outliers, no spikes (or dips) in frequency, and the bandpass of the high-band antennas is nicely visible.

Pre-Facet (i.e. DIE) calibration: <u>diagnostic plots</u>

Differential-clock and differential-TEC



d-clock and d-TEC values are solved for at the same time => useful to inspect these plots together. For CS both d-clock and d-TEC should be close to zero. RS can have different values, but the d-clock values should be reasonable constant in time while the d-TEC values can change significantly.

Pre-Facet (i.e. DIE) calibration: <u>diagnostic plots</u>

XX-YY phase offset



Plot of the median difference of the phase solutions for the X, Y antennas (referenced to a station).

During station calibration the phases of X and Y dipoles are not aligned with each other (to this aim a polarized calibrator is needed).

The resulting phase offsets is typically linear in frequency.

Pre-Facet (i.e. DIE) calibration: <u>diagnostic plots</u>

Amplitude / phases solutions per station



Waterfall plots to overview the RFI situation at the different stations and can be used to identify which stations are bad, if bad are stations found in other plots.

Pre-Facet (i.e. DIE) calibration: <u>diagnostic plots</u>

Structure functions (see Mevius et al. 2016)



X axis = baseline length Y axis = var(phases x,y)

Blue circles = data Red circles = outliers Grey circles = threshold outliers

Red line = linear fit; low extreme estimates the <u>diffraction scale</u> r_{diff}

rms 🔰 if r_{diff} 켜

Useful plots to overview the ionospheric situation at each frequency band and can be used to identify the diffraction scale (=> max angular resolution achievable).

- Initial-Subtract imaging: imaging the full field of view (and 1st sidelobe), generating a sky-model, and subtracting it from the visibilities
 => needed for facet calibration.
- Facet (i.e. DDE) calibration: FACTOR pipeline corrects for DDE effects in HBA LOFAR data, including ionospheric effects and beam-model errors.

- Facet (i.e. DDE) calibration: selecting the DDE calibrators and facets
- The DDE calibrators can be provided to FACTOR in a file or can be selected internally.

Green regions show the facets.

Yellow boxes show the regions to be imaged during self calibration.

The inner box is the region over which sources are added back to the empty datasets.



• Facet (i.e. DDE) calibration: checking the progress of a run



Useful online documentation

- The LOFAR documentation at <u>https://www.astron.nl/radio-observatory/lofar-</u> <u>documentation/resources/resources</u>
 - The LOFAR imaging cookbook at <u>https://www.astron.nl/radio-observatory/lofar/lofar-imaging-</u> <u>cookbook</u>
 - Software processing tools at <u>https://www.astron.nl/radio-observatory/lofar-data-processing/</u> <u>software-processing-tools/software-processing-tools</u> and <u>https://github.com/lofar-astron</u>
- LOFAR data processing school at <u>http://www.astron.nl/lofarschool2016/</u>

Answers

- Raw data inspection
 - Which modules you loaded ? wsclean/2.4 lofar/2.21.5 casacore/2.3.0 casarest/1.4.2 aoflagger/2.9.0 python-casacore/2.1.2 pybdsf/1.8.12
 - what is the MS size for a long synthesis observation ? 11GB x 6hrs ≥ 66GB
 - what is the total data volume ? 11GB x 6hrs x 244SBs > 16TB
 - how does this compare with data from other (radio) telescopes ? ~3 orders of magnitude larger !
 - What array configuration was used and what it mean ? HBA DUAL INNER mode used, i.e. the core stations are split in two HBA sub-fields
 - What was the duration of the observation ? 3599 sec
 - Which field was observed ? 3c295
 - How many channels (frequencies) are in the data set ? 64 spectral channels
 - What was the centre frequency of this particular sub-band? 118.849 MHz total bandwidth ~0.2 MHz
 - · What is the number of time slots ? 3594 timestamps
 - What is the integration time per time step? 1s
 - How many stations (core and remote), and how many baselines ? 54=nstations ,1485=nbaselines
 - What is the relation between number of stations and baselines (excluding autocorrelations)? nstations=autocorr and [nstations * (nstations 1)] / 2 = nbaselines
 - What are closest sources? VirA, HerA, CygA . .
 - Which are their fluxes around the MS reference frequency? 1.1E3 Jy, 3.7E2 Jy, 8.8E3 Jy . .
- Flagging, averaging and inspection
 - The flagged and compressed MS has a much more manageable size of . . . ? 333MB
 - Which station shows lower/higher amplitudes? CS302HBA0
 - What is the size of the new MS ? Does filtering help to optimize disk space? Size decreased by ~3% (i.e. 11MB): 11MB*244=2684MB~3GB modest impact ! Far better to remove intermediate products . .

Answers

- DIE Calibration with DPPP
 - Which are suspicious solutions (antennas) ? Amp: RS208,508,509 scattered; phase: RS305(bad phases at the end),RS508 and RS509 fast phases (XX,YY phases offset)
 - Which station(s) baseline(s) shows bad amplitudes ? amplitudes of data from CS302HBA1 are also high. Baselines RS508HBA&RS509HBA (visible in orange) and RS208HBA&RS509HBA (in green) look bad.
 - How are distributed visibilities ? Strong asymmetric UV distribution out of 7000λ . .
 - What does it mean for imaging ? An asymmetric PSF towards high angular resolution . .
- Imaging
 - Which is the corresponding uv cut in units of λ ? ~7000
 - Given the scale, size and pad values which is the size in degrees expected for the FoV ? 7.5*4704/3600/1.4=7deg
 - What is the noise level and the PSF size of the cleaned image ? ~20 mJy/beam
 - What is the flux density of 3C 295 ? 110Jy
- DIE phase only calibration with DPPP
 - How effective is this step to reduce the MS size ?
- Ionospheric RM correction (with BBS)
 - How large is the variation of RM over the observation duration ? 0.25 rad/m²
 - What is the corresponding rotation angle ? 120 MHz = 2.5m, $\lambda^2 = 6.25 \text{ m}^2 \text{RM}_{ion}^* \lambda^2 = \text{rot.angle [rad]} = 1.57 => 90 \text{ deg }!!!$
 - Can you write a parset for doing it with DPPP ? msin=<xxx> msout=<xxx> steps=[applycal] applycal.parmdb=<RMParmdb> applycal.correction=rotationmeasure

Marco Iacobelli – LOFAR Data Processing Tutorial – ERIS, 19th Oct 2017



RadioNet has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730562



LOFAR



Netherlands Institute for Radio Astronomy