The Standard Imaging Pipeline
Generic pipeline
MSSS

Emanuela Orru’ (Astron)
Outline

• RO standard imaging pipeline
• MSSS the survey and products
• Generic pipeline and its applications (Pre-factor FACTOR)
• RO planning for the future
Standard Imaging Pipeline

- An automatic pipeline, RO can run for users
- Preprocessing, calibrate, image, long baseline pre-processing (see Moldon lecture), PSR
- used for MSSS as testbed
- now outdated new pipelines coming soon
Pre-processing pipeline

- Flags the data in time and frequency
- Flags the autocorrelations
- Flags the first and last two channels
- Optionally “demix” subtraction of the contributions of the brightest sources in the sky (the so called "A-team"
- Currently, users should specify if demixing is to be used, and which sources should be demixed.
- It is the pipeline most commonly used

DATA PRODUCTS:
For each observing beam a set of uncalibrated correlated visibilities is provided
Pre-processing pipeline

• Flags the data in time and frequency
• flags the autocorrelations
• flags the first and last two channels
• optionally “demix” subtraction of the contributions of the brightest sources in the sky (the so called "A-team"

• Currently, users should specify if demixing is to be used, and which sources should be demixed.

• It is the pipeline most commonly used

DATA PRODUCTS:
For each observing beam a set of uncalibrated correlated visibilities is provided
Calibrator pipeline

- Pre-processing
- Calibrate the calibrator. The antenna gain solutions are estimated by BBS
- Solutions stored in a parmdb

Target pipeline

- Pre-processing
- Applies the antenna gains (solutions) obtained in the calibrator pipeline

Used to calibrate MSSS

DATA PRODUCTS:
For the calibrator and target beams a set of uncalibrated and calibrated correlated visibilities are provided, written respectively in the DATA and CORRECT columns of the MS.
Calibrator pipeline

- Pre-processing
- Calibrate the calibrator
- The antenna gain solutions are estimated by BBS
- Solutions stored in a parmdb

Target pipeline

- Pre-processing
- Applies the antenna gains (solutions) obtained in the calibrator pipeline

Used to calibrate MSSS

DATA PRODUCTS:
For the calibrator and target beams a set of uncalibrated and calibrated correlated visibilities are provided, written respectively in the DATA and CORRECT columns of the MS.
Calibrator pipeline

- Pre-processing
- Calibrate the calibrator
- The antenna gain solutions are estimated by BBS
- Solutions stored in a parmdb

Target pipeline

- Pre-processing
- Applies the antenna gains (solutions) obtained in the calibrator pipeline
- Used to calibrate MSSS

DATA PRODUCTS:
For the calibrator and target beams a set of uncalibrated and calibrated correlated visibilities are provided, written respectively in the DATA and CORRECT columns of the MS.
Imaging pipeline

- Sub-bands are concatenated in frequency. Bandwidth size set according to user specification

- Flag of outliers

- A run of phase calibration with BBS is performed using a GSM model.

- Sub-bands are concatenated in time when observations are performed in snapshot

- Images are produced in hdf5 format (using awimager)

- Source finding software is used to identify the sources detected in the image, and generate an updated local sky model.

DATA PRODUCTS:

Calibrated images.
Imaging pipeline

- Sub-bands are concatenated in frequency. Bandwidth size set according to user specification
- Flag of outliers
- A run of phase calibration with BBS is performed using a GSM model.
- Sub-bands are concatenated in time when observations are performed in snapshot
- Images are produced in hdf5 format (using awimager)
- Source finding software is used to identify the sources detected in the image, and generate an updated local sky model.

DATA PRODUCTS:

Calibrated images.

Same concept will be used in T2 - T3 with different software.
Multifrequency Snapshot Sky Survey

Goals: obtain broadband sky model, shakedown LOFAR operations

MSSS-LBA

Frequency: 30-75 MHz
(8 x 2 MHz bands)
Resolution: ≤100 arcsec
Sensitivity: ≤15 mJy/beam
Area: 20,000 square degrees
Number of Fields: 660
Simultaneous ~10° beams: 5
Test observations resuming

MSSS-HBA

Frequency: 120-160 MHz
(8 x 2 MHz bands)
Resolution: ≤120 arcsec
Sensitivity: ≤5 mJy/beam
Area: 20,000 square degrees
Number of Fields: 3616
Simultaneous ~4° beams: 6
Observations 100% complete
Observations

7min 7min

2hrs

Processing

RO SIP + Script analogue to imaging pipeline

Tuesday, 6 September 16
MSSS-HBA catalog (v0.1): 130000 sources

• Final catalog still under development (need flux correction..)

• All fields and catalog loaded into MSSS VO server for use by initial science and checks.

• Group of testers (active MSSS participants) ... identifying bugs in system. MSSS Forum being used to collect issues.

• Data products are password-protected before data release

Generic pipeline

- The Pipeline Framework is used by RO for automated processing on the CEP cluster systems (e.g. SIP). Writing these pipelines is complicated and requires a lot of knowledge of the framework itself and some programming skills.

- The generic pipeline is a pipeline based on that system. It is a simplified layer of software that helps users with the design and execution of their own workflow without the need to understand the underlying system.

- For a pipeline the user should organize the work into building blocks e.g. preprocessing-calibration-imaging....

- Parameters can be reduced to a minimum with different sets of default parsets for every step. Integration of other peoples steps and reusing existing work are one of the primary goals of the generic pipeline.

- A users pipeline can run on a single workstation and in a cluster environment without the need to change the pipeline parset or to program process management.

- Steps and arguments to the steps are defined in a parset file.

- This parset is then the argument for the genericpipeline.py.

- The pipelines name will be the first part of the parsets name mypipeline.parset. Log files will be tracked under that name.
Parset

- The step list determines the order of execution.
- Optional path to a plugins directory
- Mapfiles are the data descriptors, the path to your input data

```python
# Pipeline for running NDPPP on all files in a directory

#variable parameters
#path to the directory where we are looking for the input data
! input_path = /data/scratch/dummyuser/test-in
# path to the parset
! ndppp_parset = /home/dummyuser/parsets/NDPPP-preproc-parset.proto

pipeline.steps=[createmap,ndppp]

#Step 1: search for all measurement sets in one directory and generate a mapfile
createmap.control.kind            =   plugin
createmap.control.type            =   addMapfile
createmap.control.cmdline.create  =   mapfile_from_folder
createmap.control.mapfile_dir     =   input.output.mapfile_dir  #this is the name that the mapfile will have
createmap.control.filename        =   input_data.mapfile
createmap.control.folder          =   {{ input_path }}  #this references to the path we defined above

#Step 2: run NDPPP with a given parset on all files that the previous step found
ndppp.control.type                   =  dppp
ndppp.control.parset                 =  {{ ndppp_parset }}  #this references to the parset we defined above
ndppp.control.max_per_node           =  4  #run 4 instances of NDPPP in parallel
ndppp.control.environment            =  {OMP_NUM_THREADS: 6}  #tell NDPPP to use only 6 threads
ndppp.argument.msin                  =  createmap.output.mapfile
```

Pre-factor (new calibrator pipeline)

- In the calibrator field: solves for bandpass and separates the instrumental (clock) from the ionospheric (TEC) contribution to the phase delays.
- Transfers to the target field the bandpass, clock and phase offset.
- Combines SBs and image the blocks at high and low resolution. Produces residual images and MS to be fed to the FACTOR pipeline.
FACTOR: facet self calibration pipeline

- By dividing up the field into many facets and solving for the direction-dependent corrections in each facet using the “peeling”
- Phase calibration on short time scale >> ionospheric effects amplitude calibration long time scale >> residual beam errors
- Instrumental-noise limited images (~ 0.1 mJy/beam for an 8-hour observation), high-resolution images (~ 5 arcsec FWHM) and high-fidelity images
- Designed to distribute of jobs over multiple nodes of a cluster and for the processing of facets in parallel.
future development

• RO plans to implement the generic pipeline and use it as a platform for all the current pipelines -- highly flexible, easy for users to provide different setup with respect to the standard.

• RO plans to substitute the calibrator, target, and imaging pipeline with the pre-factor pipeline.

• FACTOR needs more automatization and performance optimization in order to be taken into account as part of the SIP.