The Standard Imaging Pipeline
Generic pipeline
MSSS

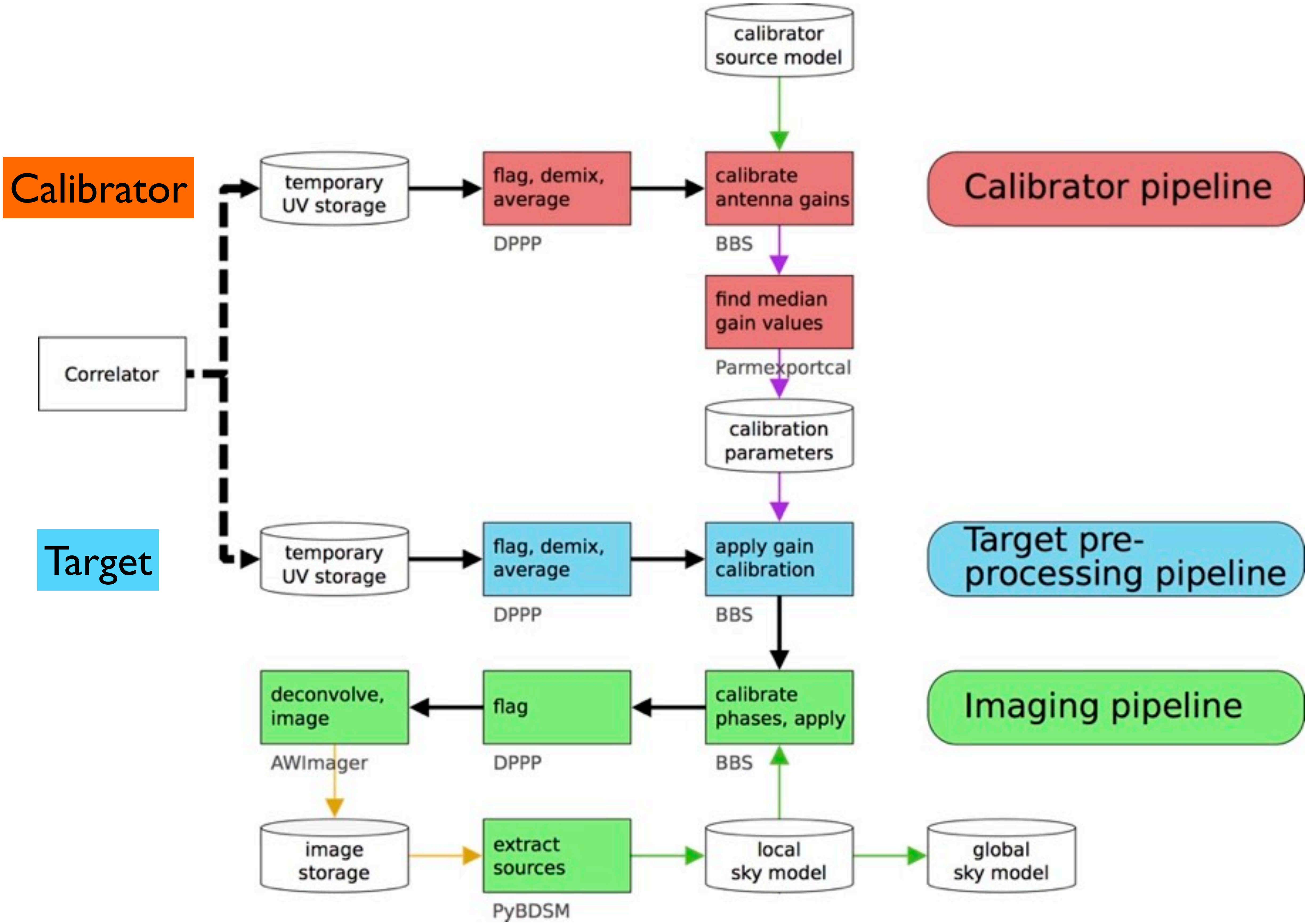
Emanuela Orru' (Astron)

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- now outdated new pipelines coming soon



Calibrator

Target

Correlator

temporary UV storage

flag, demix, average

DPPP

calibrate antenna gains

BBS

calibrator source model

find median gain values

Parmexportcal

calibration parameters

Calibrator pipeline

temporary UV storage

flag, demix, average

DPPP

apply gain calibration

BBS

Target pre-processing pipeline

deconvolve, image

AWImager

flag

DPPP

calibrate phases, apply

BBS

Imaging pipeline

image storage

extract sources

PyBDSM

local sky model

global sky model

https://www.astron.nl/radio-observatory/astronomers/array-configurations/3-telescope-parameters-and-array-configurations

Pre-processing pipeline

- Flags the data in time and frequency

- flags the autocorrelations

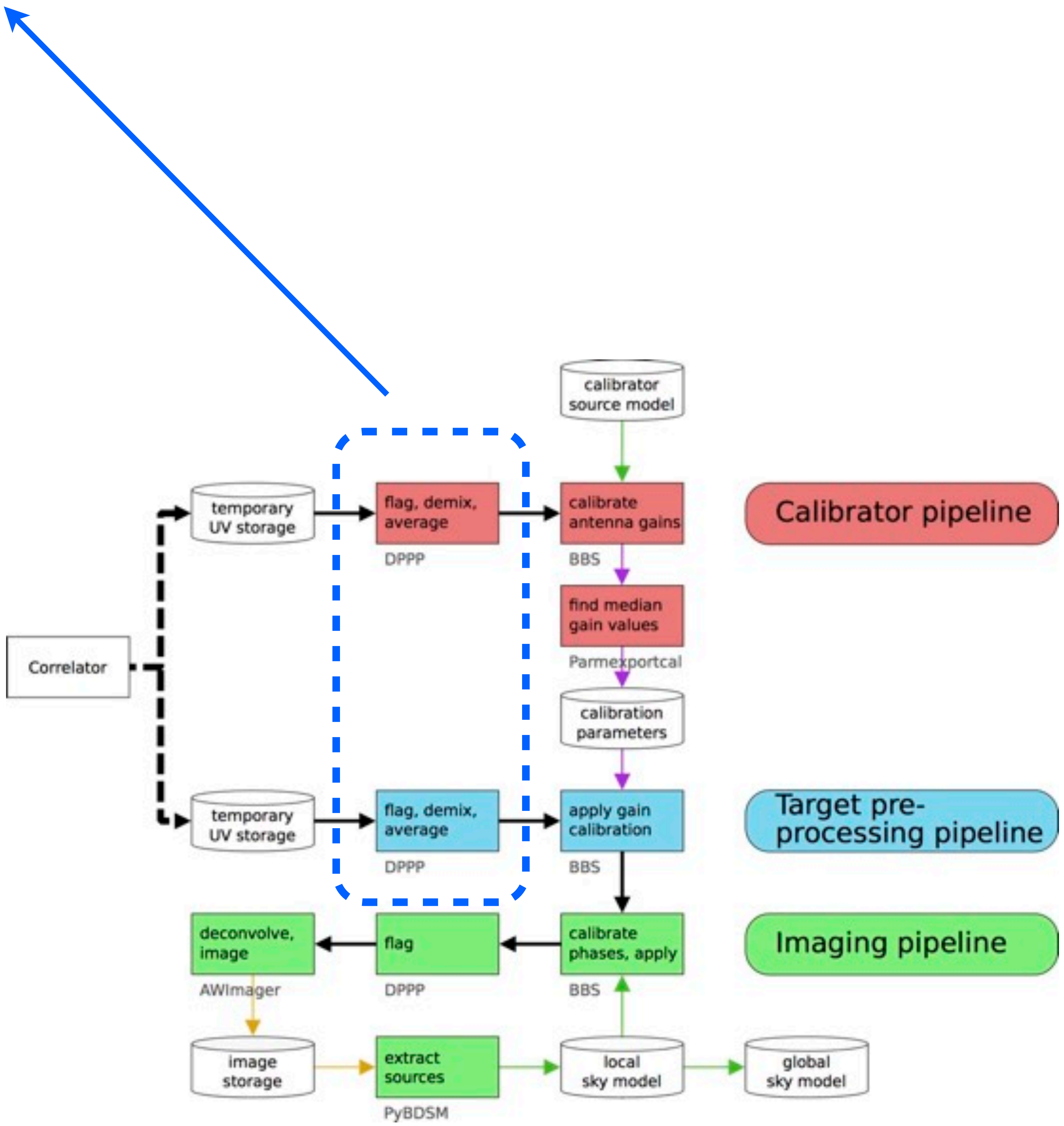
- flags the first and last two channels

- 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

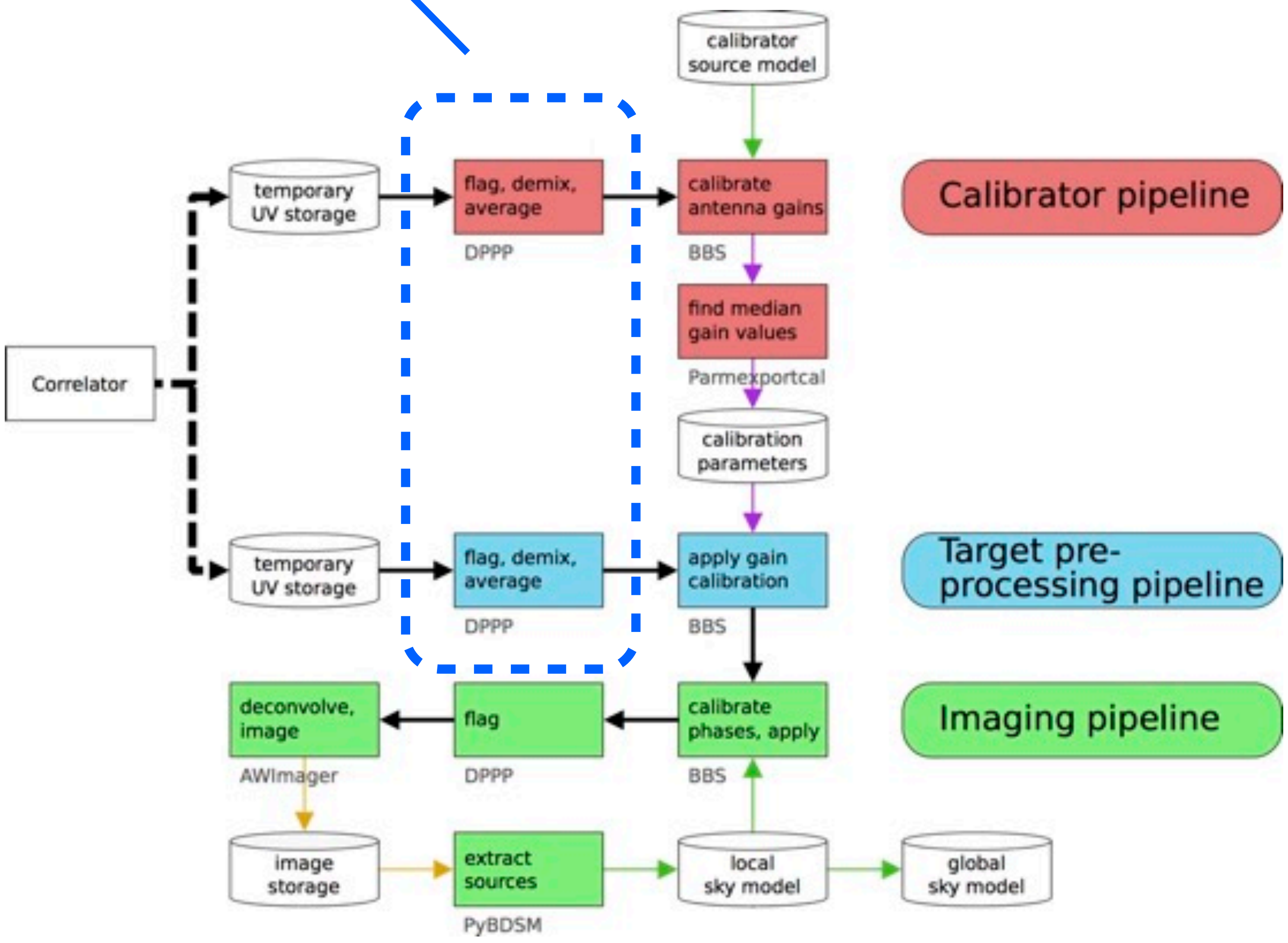
- optionally “demix” subtraction of the contributions of the brightest sources in the sky (the so called "A-team

- 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

Familiar?! All steps performed in T1



Calibrator pipeline

- Pre-processing
- Calibrate the calibrator The antenna

gain solutions are estimated by BBS

- Solutions stored in a parmdb

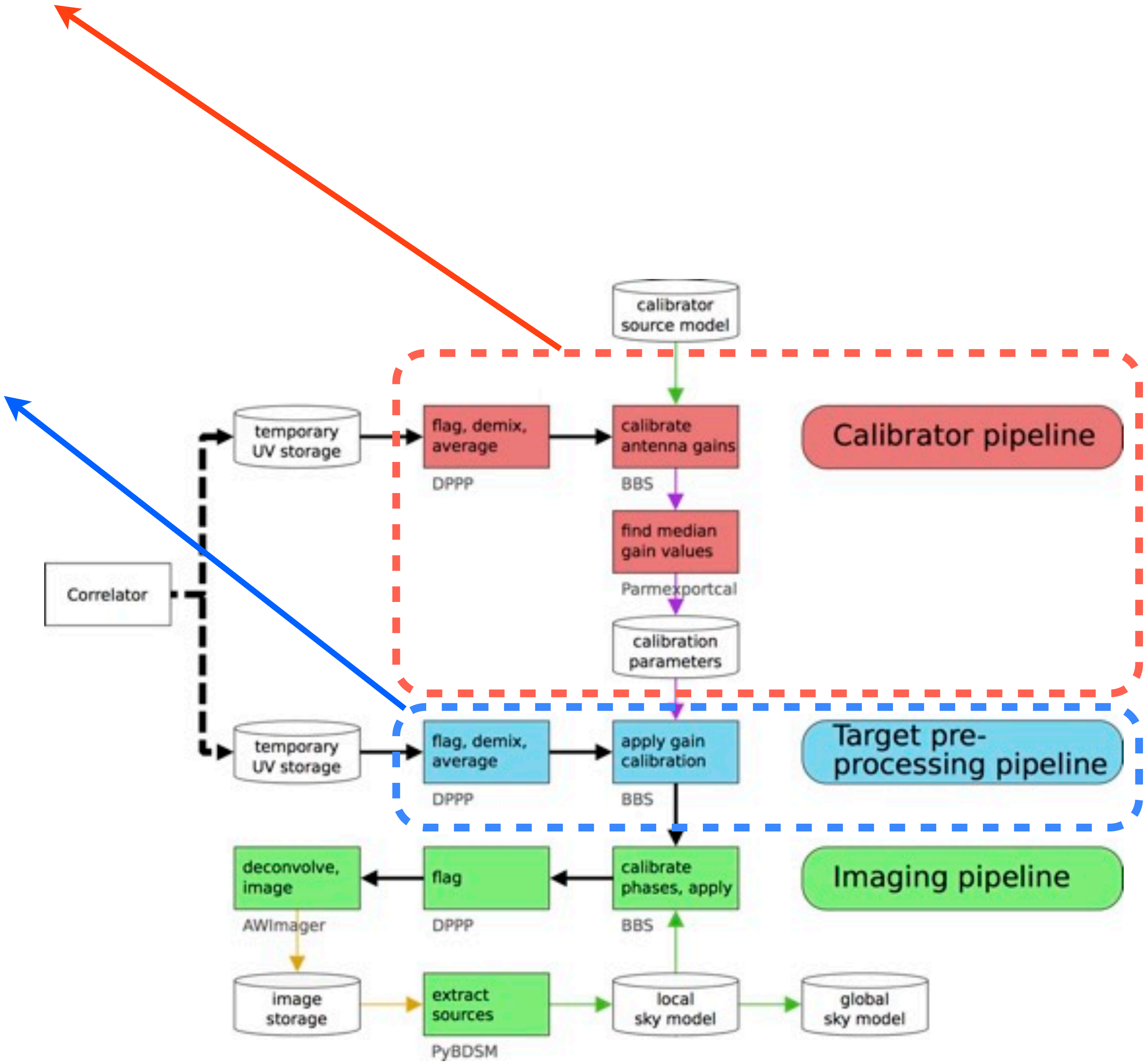
Target pipeline

- 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

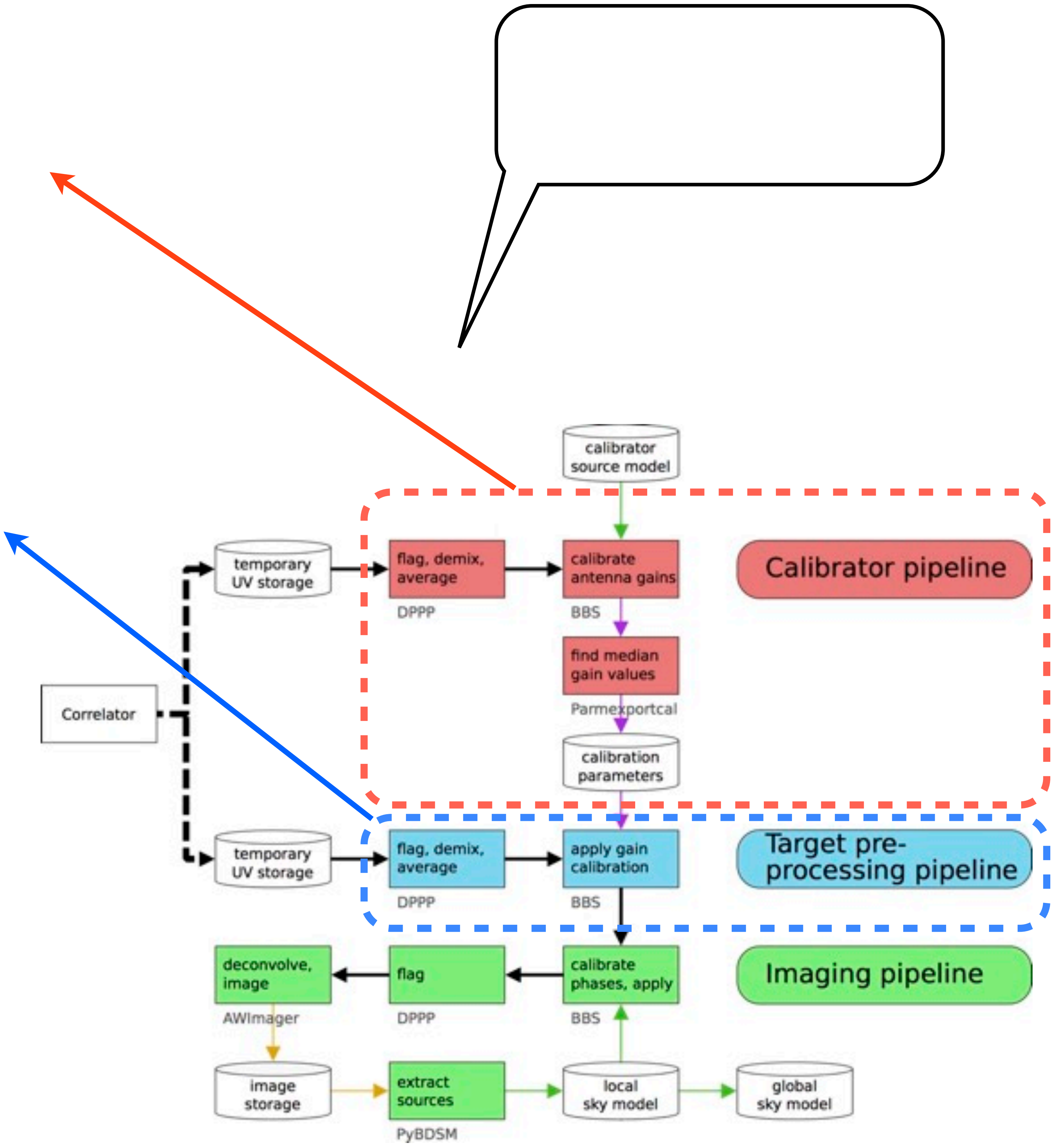
- Pre-processing
- Calibrate the calibrator The antenna gain solutions are estimated by BBS
- Solutions stored in a parmdb

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

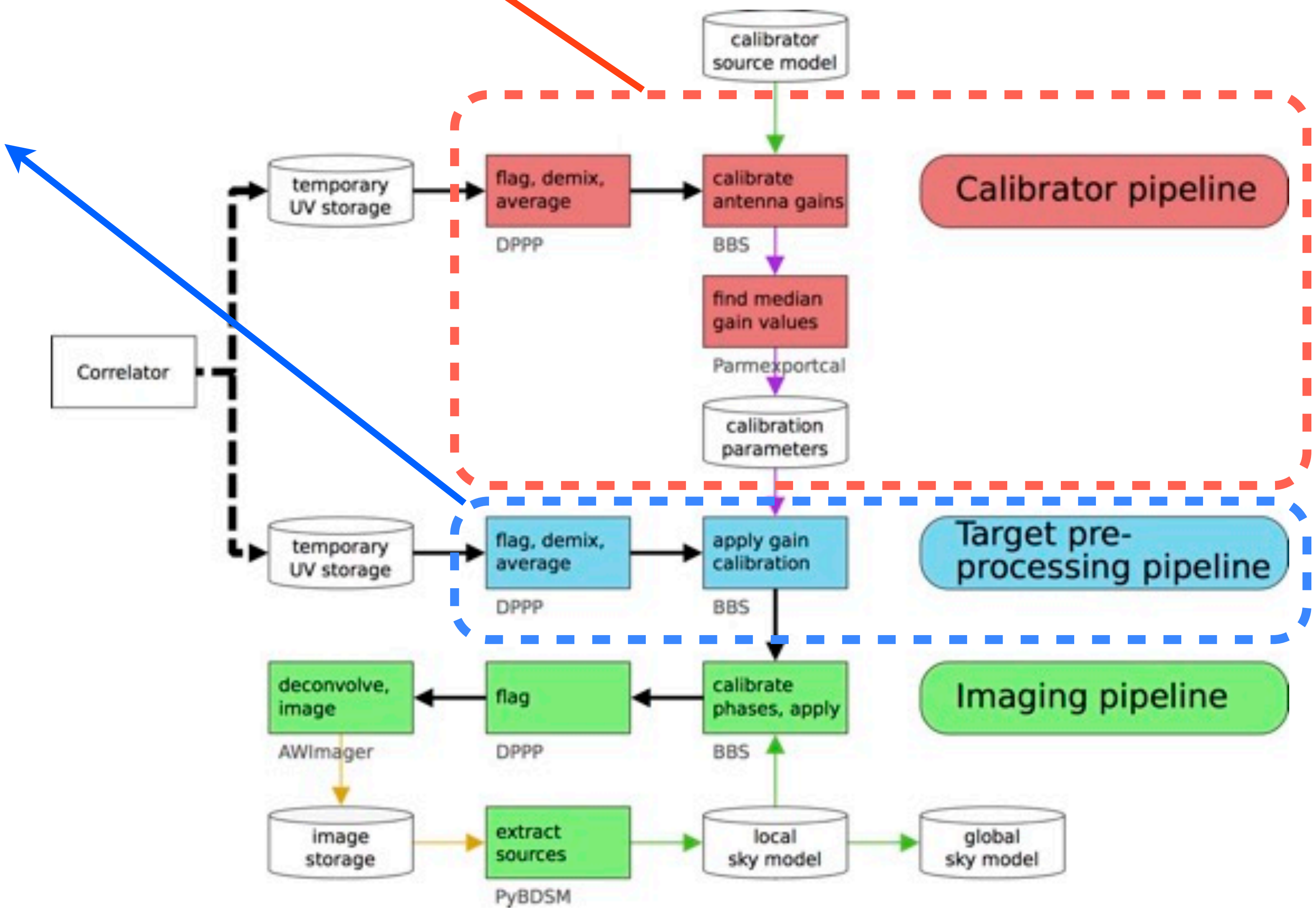
Target pipeline

- 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.

Same concept will be used in T2 with different software



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.

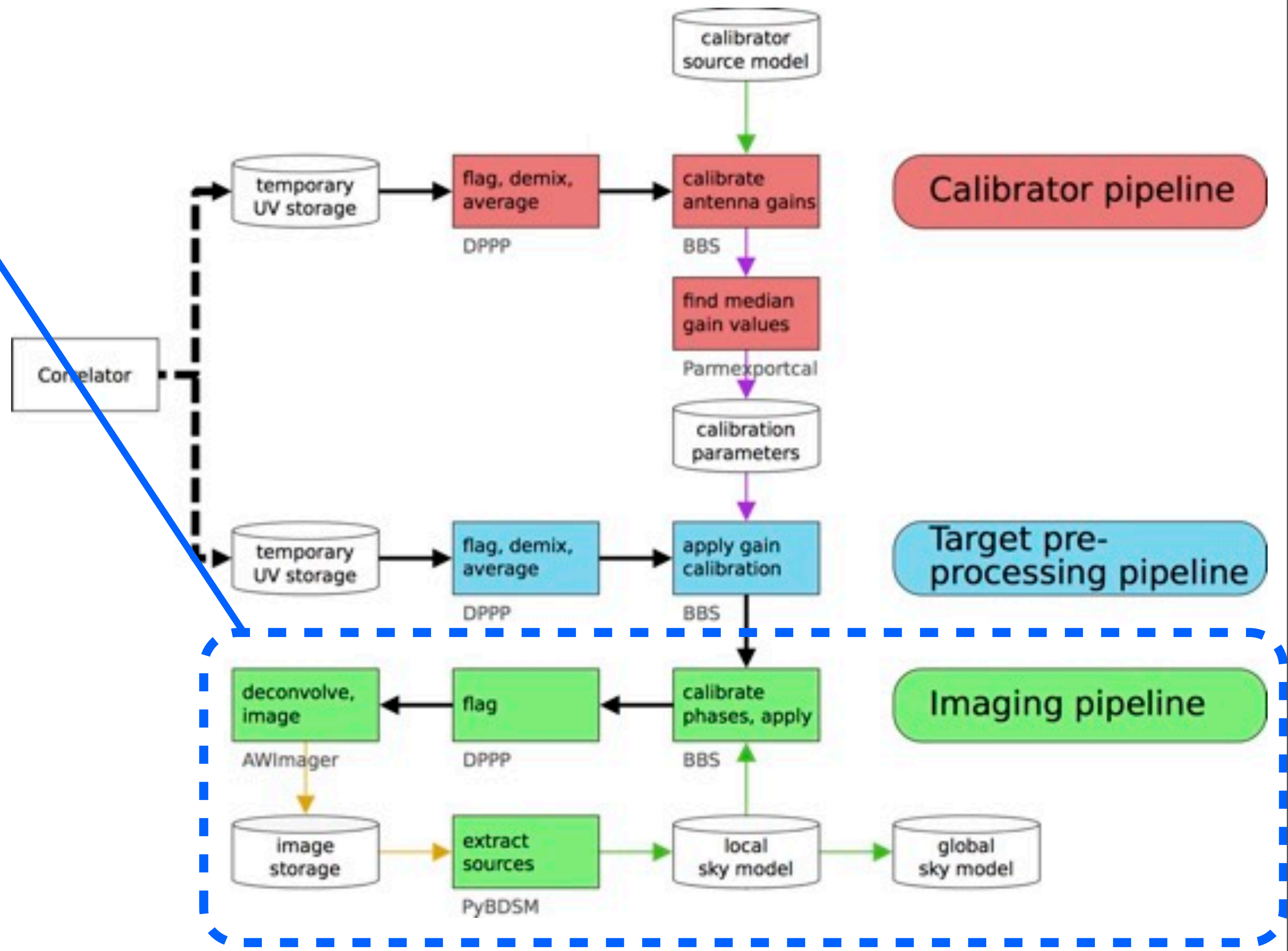
- Sub-bands are concatenated in time when observations are performed in snapshot

- 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

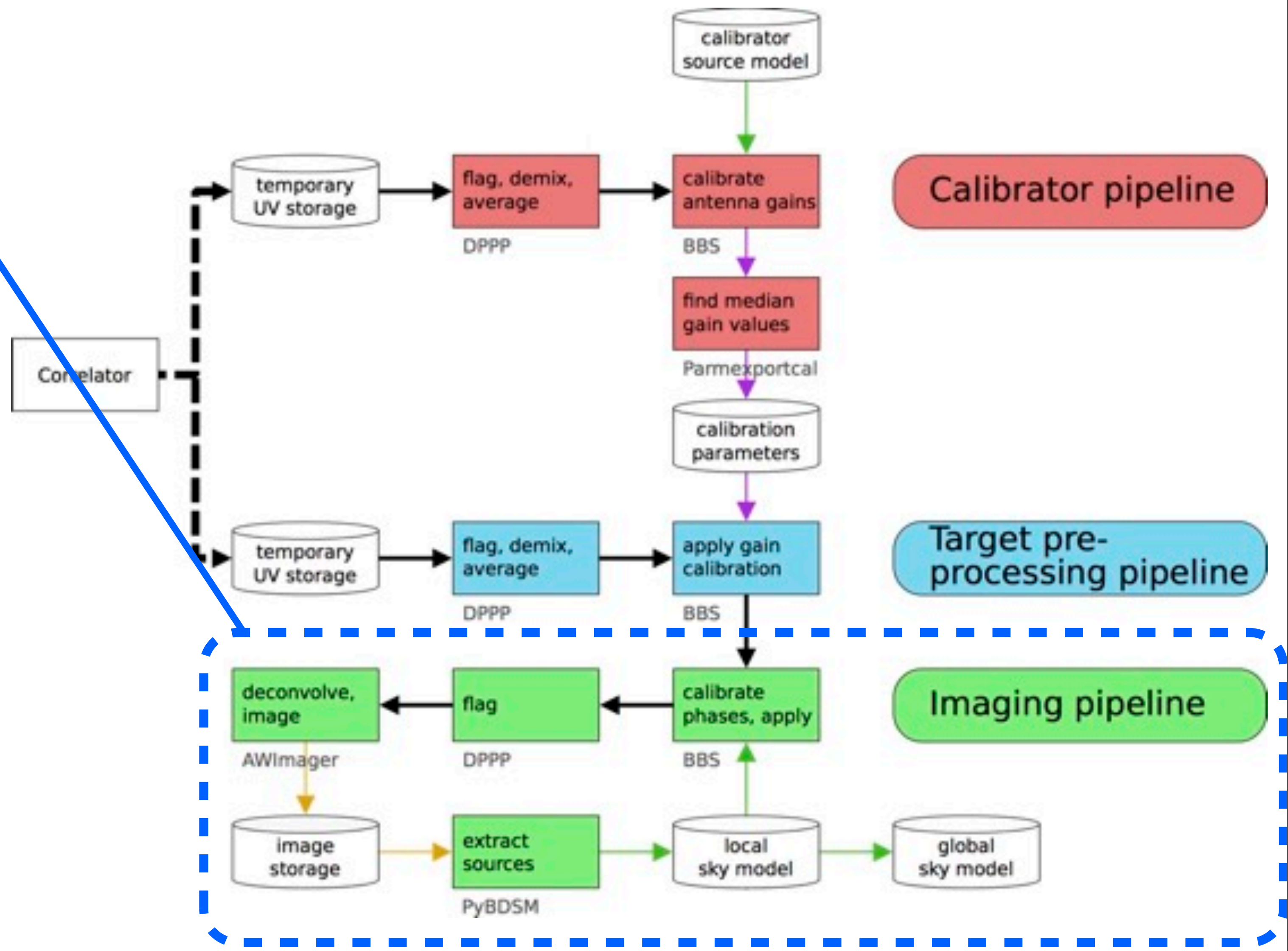
- flag of outliers

- 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.



Same concept will be used in T2 - T3 with different software

DATA PRODUCTS:

Calibrated images.

Multifrequency **S**napshot **S**ky **S**urvey

Goals: obtain broadband sky model, shakedown LOFAR operations

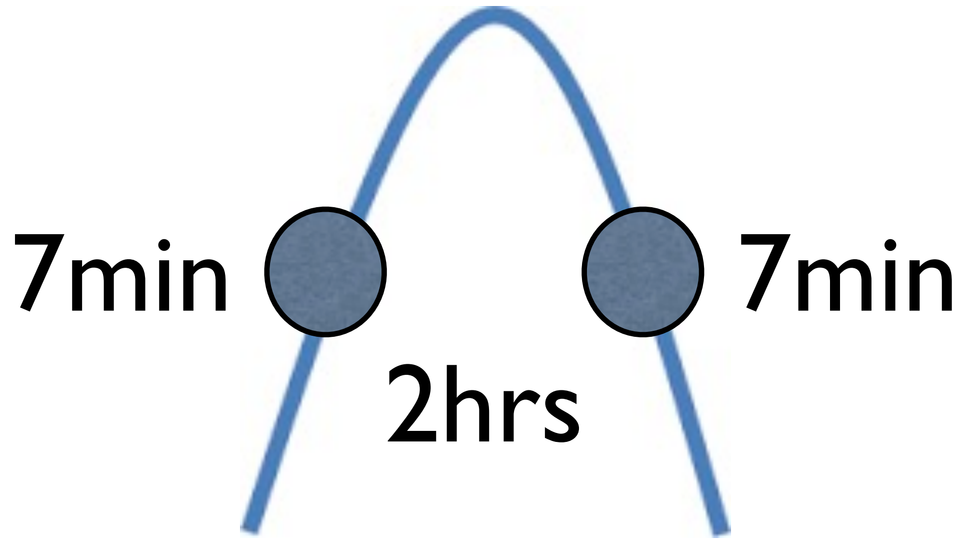
MSSS-LBA

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

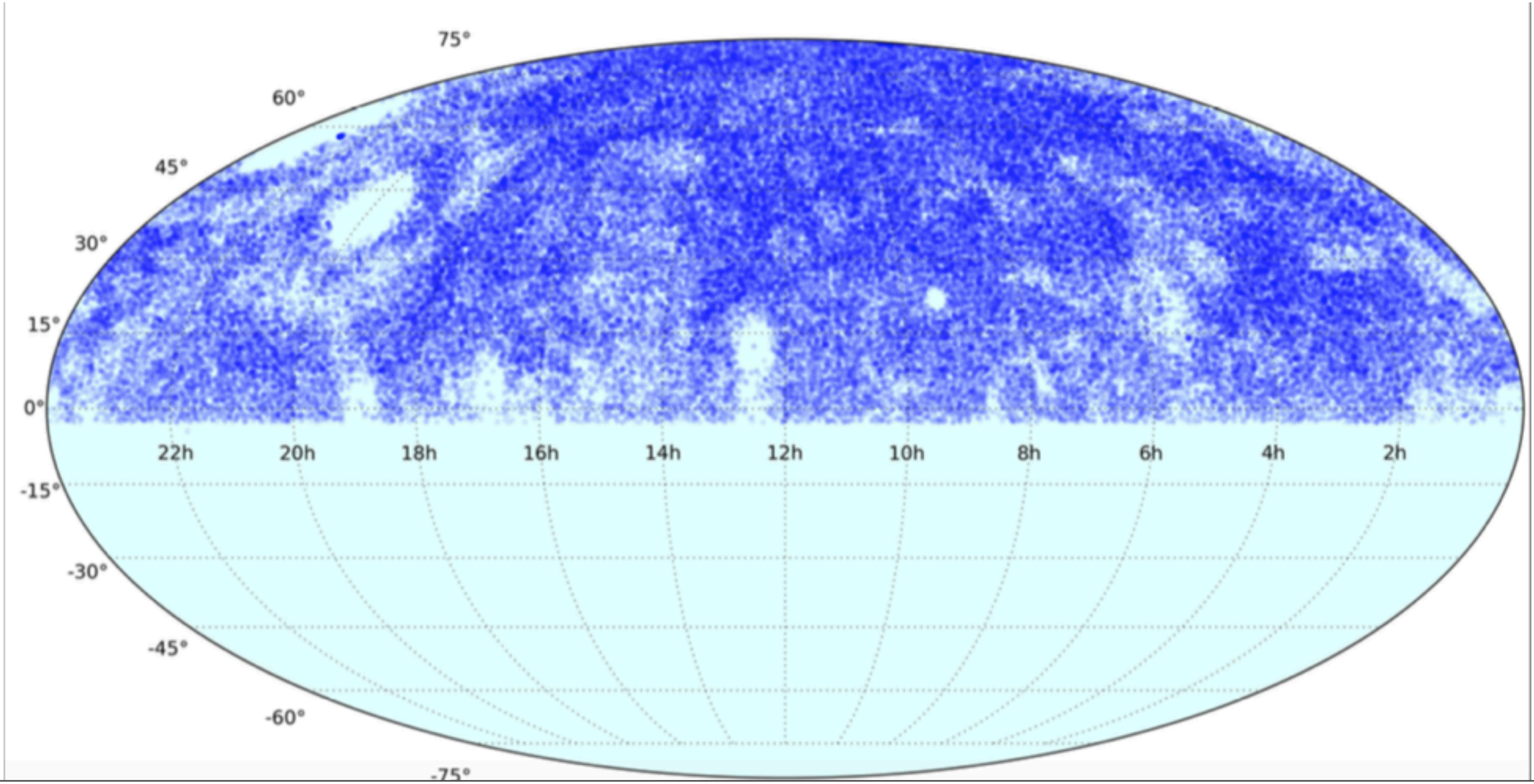
MSSS-HBA

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



Processing
RO SIP
+ Script analogue to imaging pipeline



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

More info verification field paper: *Heald et al. 2015, A&A, 582, A123*

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

```
pipeline.steps      = [step1,step2,step3,...]
pipeline.pluginpath = plugins                 # optional
pipeline.mapfile    = /path/to/your_data.mapfile  # optional
```

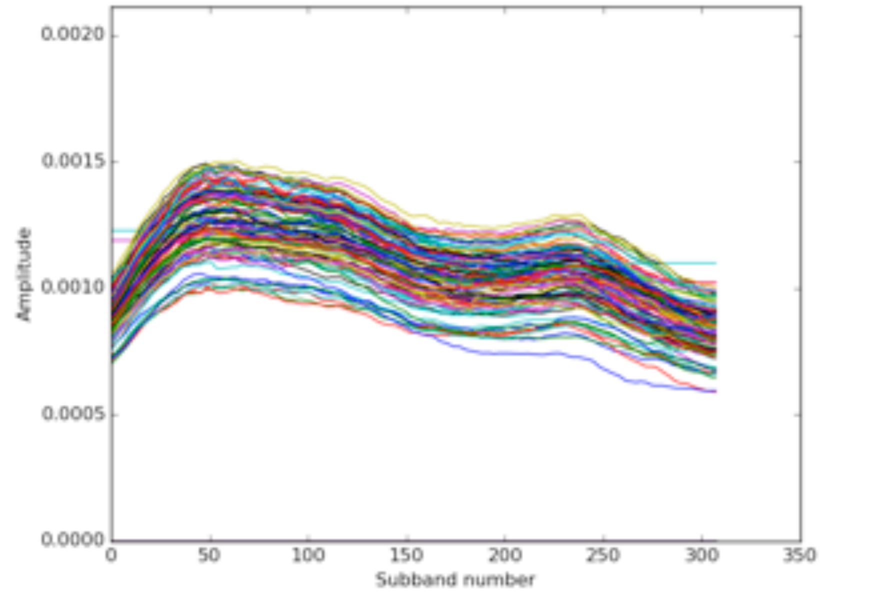
- The step list determines the order of execution.
- Optional path to a plugins directory
- Mapfiles are the data descriptors, the path to your input data

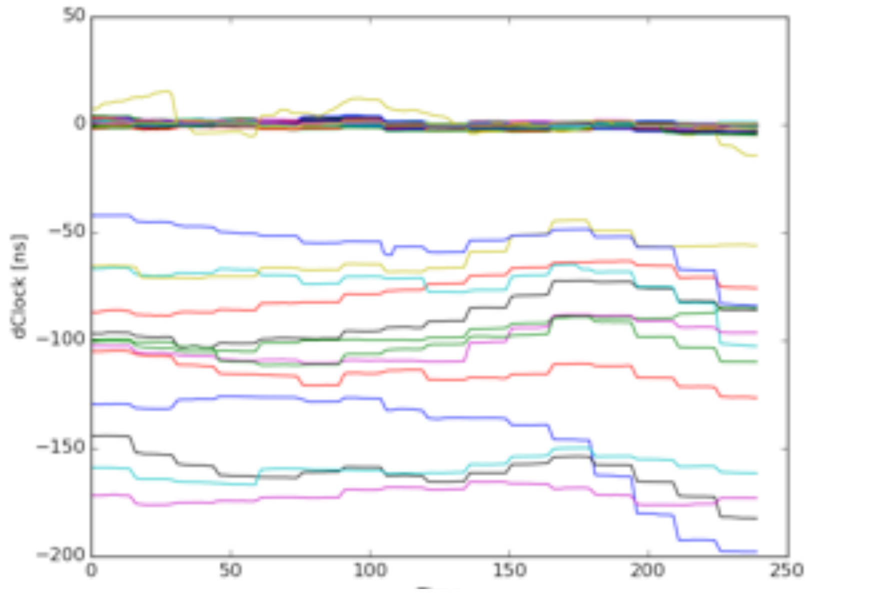
```
# Pipeline for running NDPPP on all files in a directory

#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
```

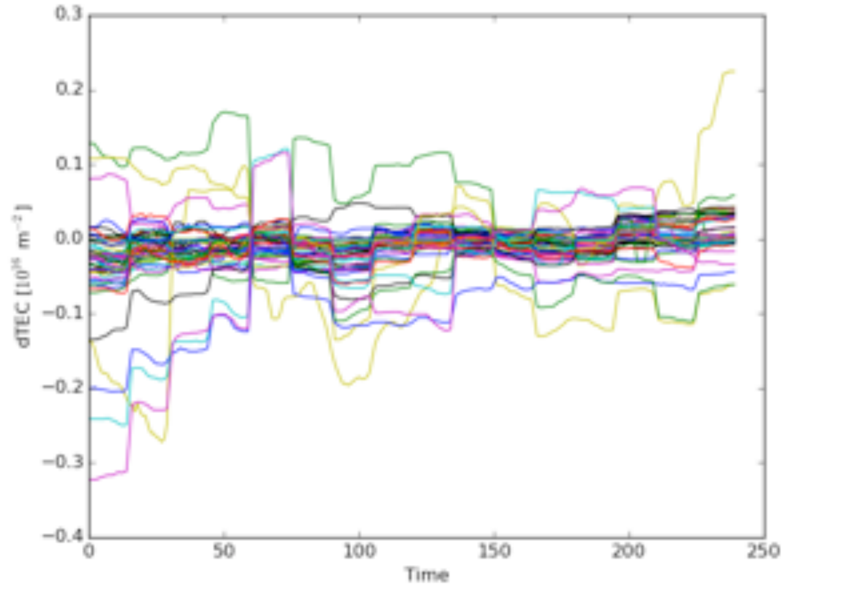
More info: http://www.astron.nl/citt/genericpipeline/#

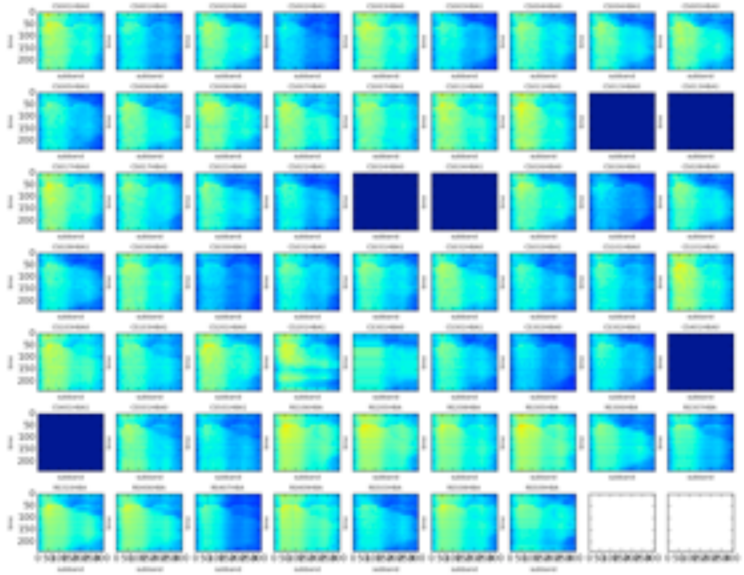
Pre-factor (new calibrator pipeline)

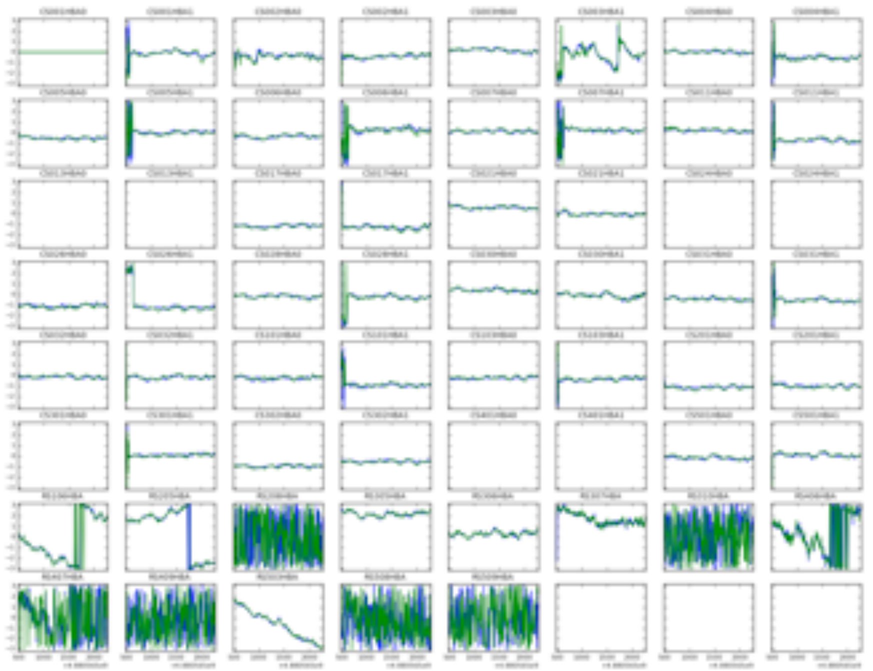
- In the calibrator field: solves for bandpass and separates the instrumental (clock) from the ionospheric (TEC) contribution to the phase delays.
- Transfers to the target field the bandpass, clock and phase offset.
- Combines SBs and image the blocks at high and low resolution. Produces residual images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

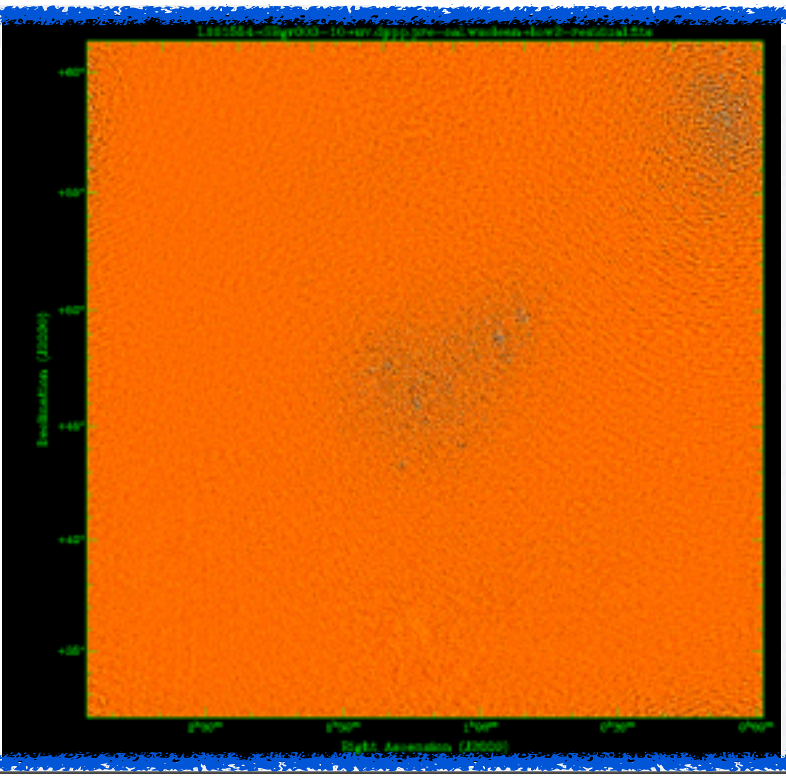

0.0020
0.0015
0.0010
0.0005
0.0000
Amplitude
0
50
100
150
200
250
300
350
Subband number

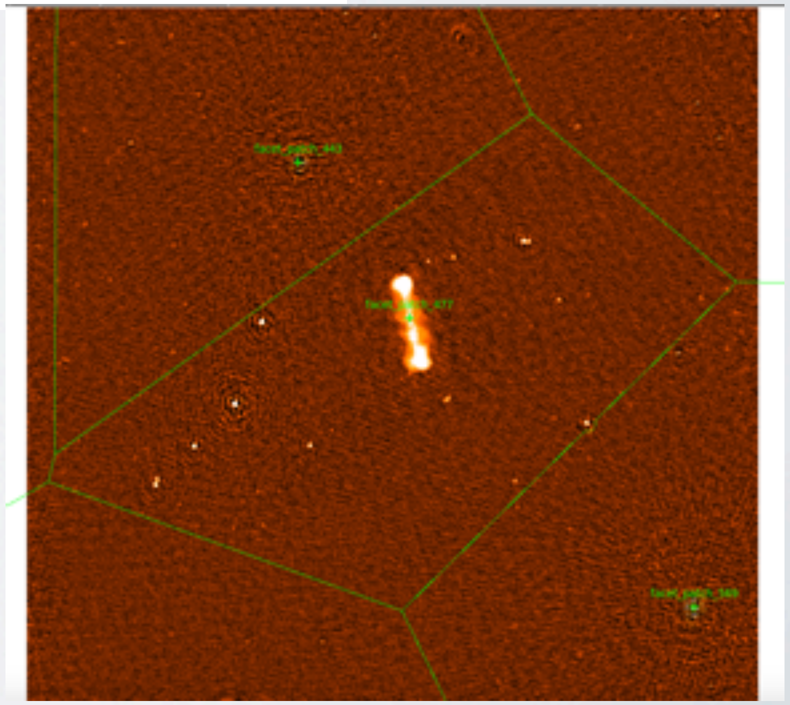

50
0
−50
−100
−150
−200
dClock [ns]
0
50
100
150
200
250


0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
dTEC [10^16 m^-2]
0
50
100
150
200
250
Time









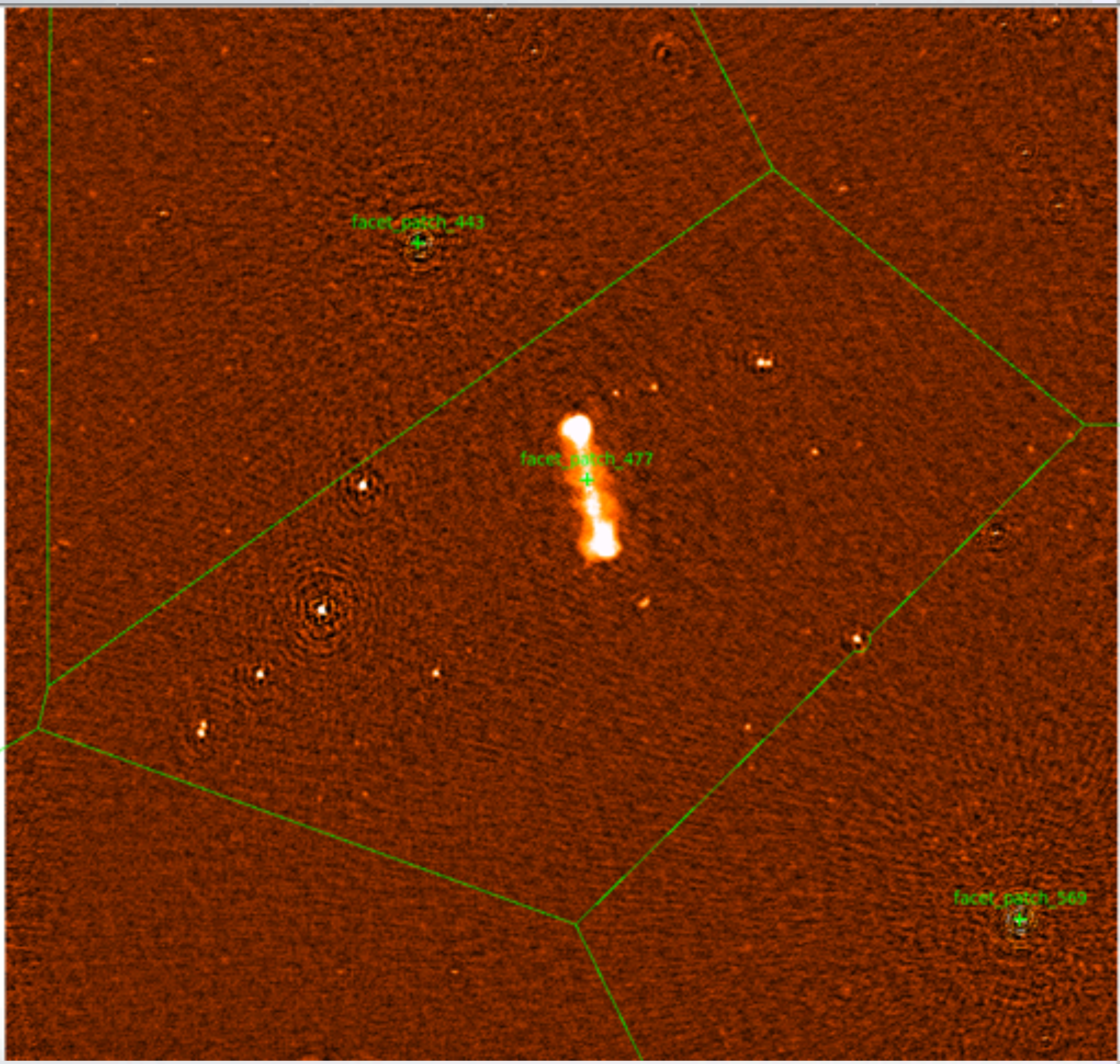
FACTOR: facet self calibration pipeline

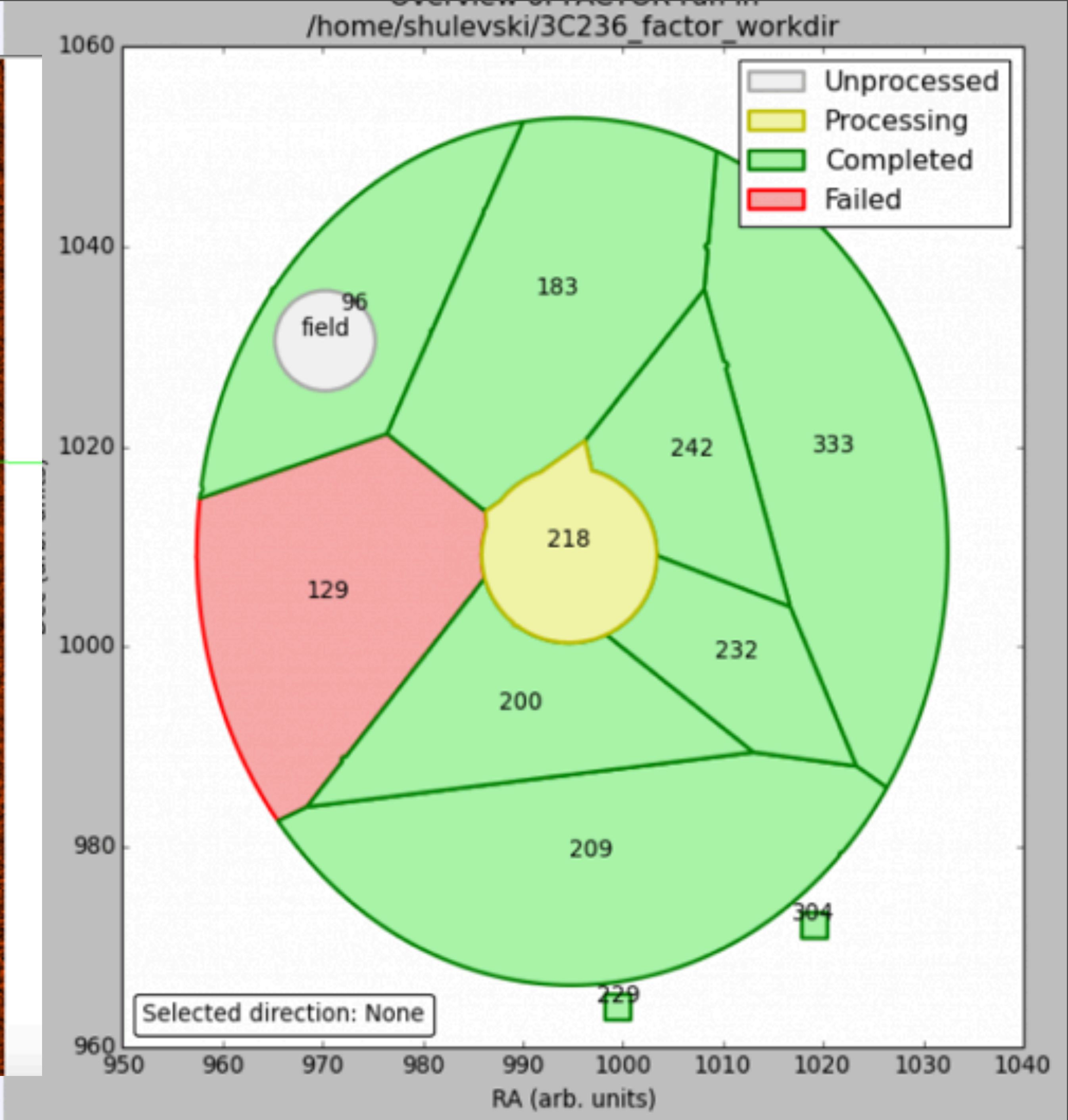
- By dividing up the field into many facets and solving for the direction-dependent corrections in each facet using the "peeling"

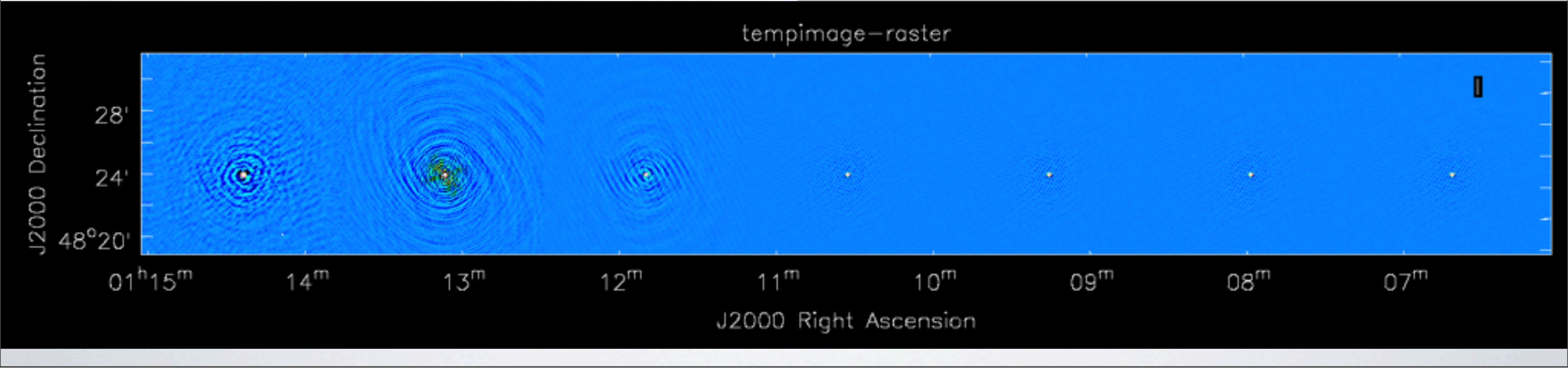
- phase calibration on short time scale >> ionospheric effects amplitude calibration long time scale >> residual beam errors

- 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.


facet_patch_443
facet_patch_477
facet_patch_569


/home/shulevski/3C236_factor_workdir
Unprocessed
Processing
Completed
Failed
96
field
183
242
333
218
129
232
200
209
304
229
Selected direction: None
1060
1040
1020
1000
980
960
950
960
970
980
990
1000
1010
1020
1030
1040
RA (arb. units)


tempimage−raster
J2000 Declination
28'
24'
48°20'
01h15m
14m
13m
12m
11m
10m
09m
08m
07m
J2000 Right Ascension

future development

- RO plans to implement the generic pipeline ad use it as platform for all the current pipelines -- highly flexible easy for users to provide different set up wrt 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.