The ALMA Pipeline

Introduction
Pipeline Subsystem
Heuristics
Task Interface
Tests / Commissioning
Results Demos

D. Muders, MPIfR
L. Davis, NRAO

Max-Planck-Institut
für
Radioastronomie

Introduction

- ALMA must be available to all of the astronomical community, i.e. also interferometry ***non***-experts
- Hence an automatic pipeline is required to provide science quality data products
- The automatic processing will be performed for all ***standard reduction modes***
- The pipeline scripts will be made available for further offline processing by the PIs and the ARCs

Max-Planck-Institut für Radioastronomie

ALMA Pipelines

- ALMA actually features 3 pipelines:
  - Telescope Calibration Pipeline:
    - Online at AOS/OSF
    - Real-time processing; Scan based
    - Provides calibration results
  - Quicklook Display Pipeline:
    - Online at AOS/OSF
    - Slightly decoupled from observing system; Session based
    - Provides quick overview for AoDs and operators
  - Science Pipeline:
    - Offline at Santiago
    - Decoupled from observing system; Science goal based
    - Delivers final data products

ALMA Pipelines

- ALMA actually features 3 pipelines:
  - Telescope Calibration Pipeline:
    - Online at AOS/OSF
    - Real-time processing; Scan based
    - Provides calibration results
  - Quicklook Display Pipeline:
    - Online at AOS/OSF
    - Slightly decoupled from observing system; Session based
    - Provides quick overview for AoDs and operators
  - Science Pipeline:
    - Offline at Santiago
    - Decoupled from observing system; Science goal based
    - Delivers final data products

This talk

Max-Planck-Institut für Radioastronomie

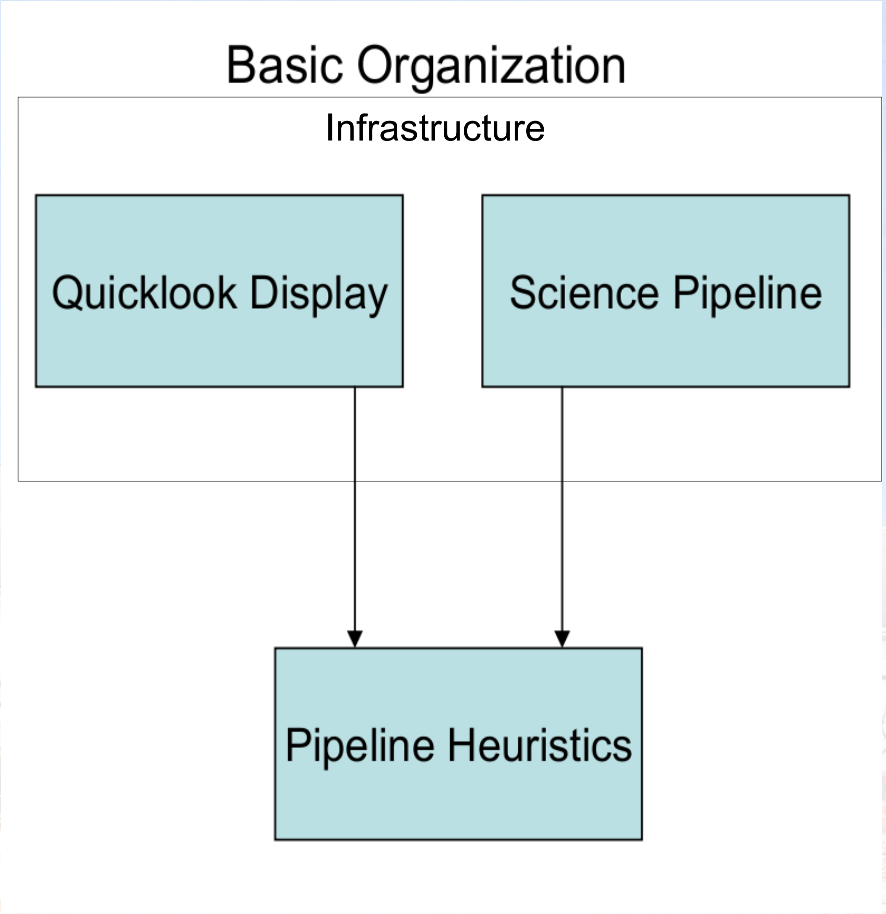
Pipeline Subsystem

The Pipeline Subsystem is divided into 2 operational systems: Quicklook Display and Science Pipeline.

The Pipeline Subsystem software is divided into 2 parts: the infrastructure and the heuristics packages.

The infrastructure packages provide the supporting interfaces, libraries, and services required to operate Quicklook Display and Science Pipeline.

The pipeline heuristics packages encapsulate the radio interferometry and single dish expertise required to pipeline process ALMA data. They support both Quicklook Display and Science Pipeline.



Pipeline Team

- Quicklook Display Pipeline
  - Lindsey Davis (50%)
  - Honglin Ye (50%)
  - Jonas Larsen (as needed)
- Science Pipeline
  - Lindsey Davis (50%)
  - Nick Elias (50%)
- Heuristics
  - John Lightfoot (100%)
  - Dirk Muders (30%)
  - Stewart Williams (50%)
  - Friedrich Wyrowski (20% in kind)
  - George Kosugi (10%)
  - Takeshi Nakazato (30%)
- Subsystem scientists
  - Chris Wilson
  - Baltasar Vila Vilaro

Max-Planck-Institut für Radioastronomie

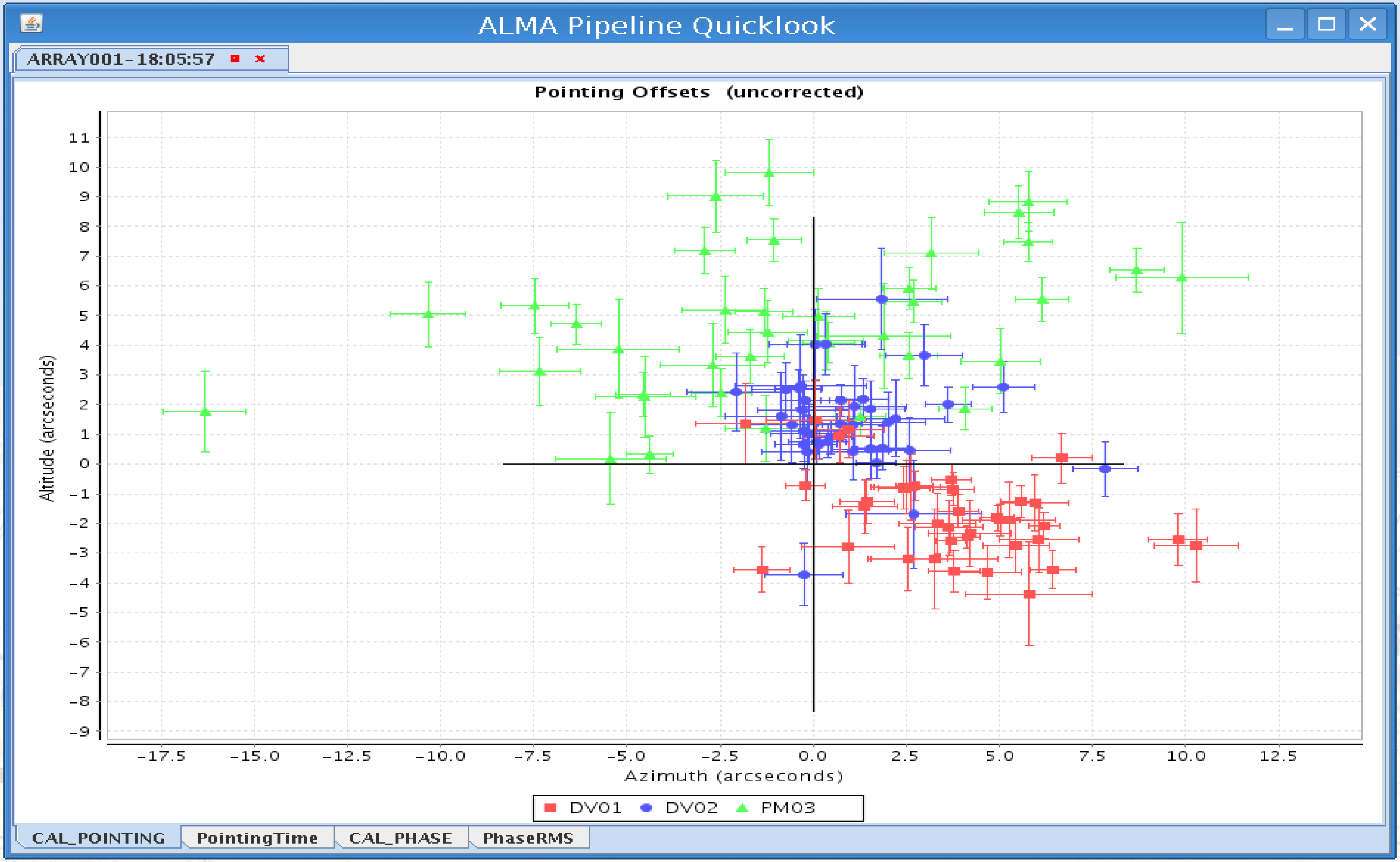
Infrastructure

- The ALMA Pipeline is based on CASA using data in ASDM / MS format
- The processing is (meta-) data driven
- CASA team provides engines as tools and tasks
- Pipeline team uses these engines to create heuristics tasks for Quicklook and Science Pipelines
- Pipeline will run on HPC cluster environment
- Parallelization via parallel tasks and Open MP / Open MPI

Quicklook Display System

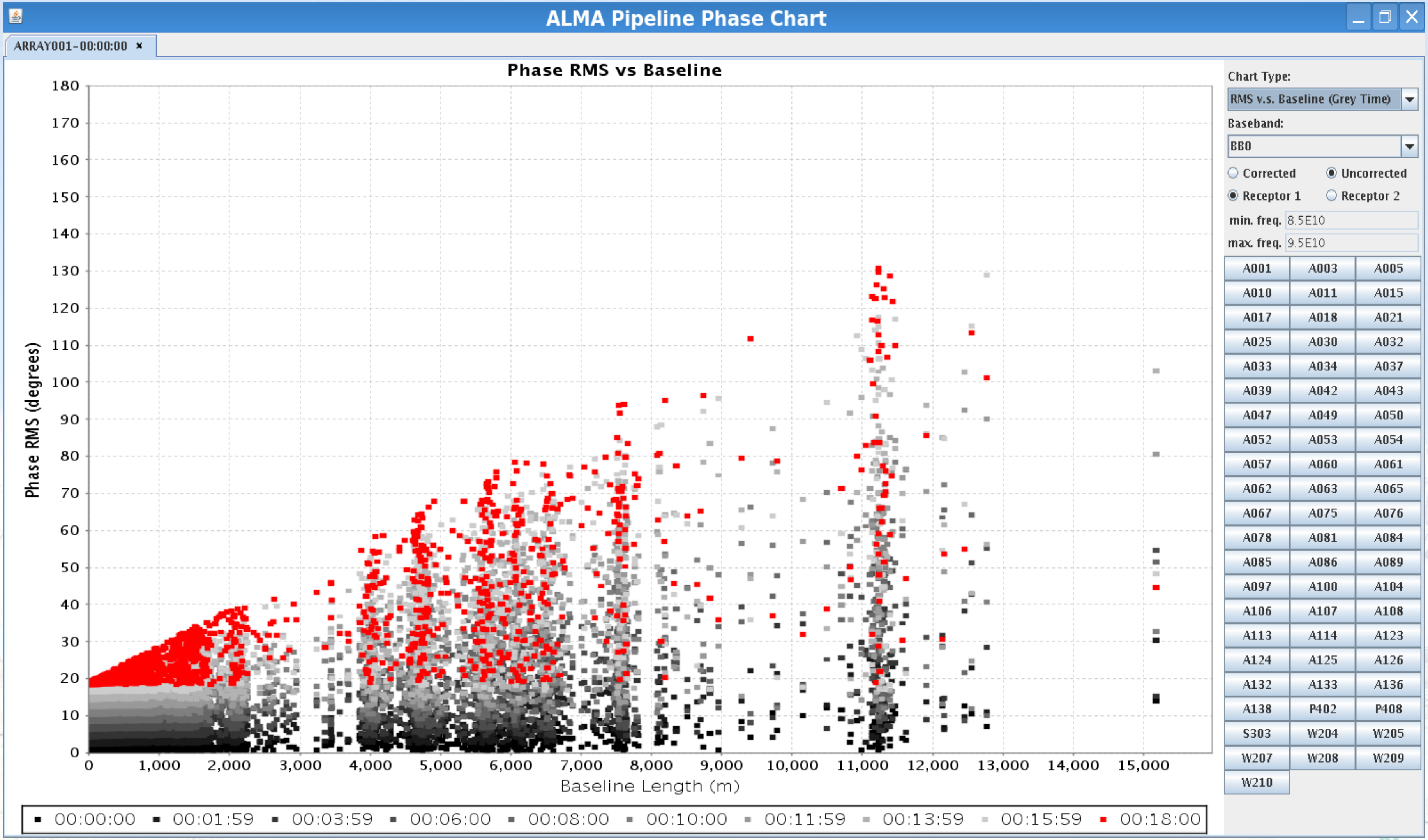
- Monitor / display on-line calibration results, e.g. pointing offsets, phase rms
- Monitor / display incoming data, e.g. raw amplitudes and phases
- Perform quick look reductions for standard observing modes, e. g. single field
- Alert observing staff to deteriorating data quality, e.g. bad system temperature

Quicklook Pointing Display



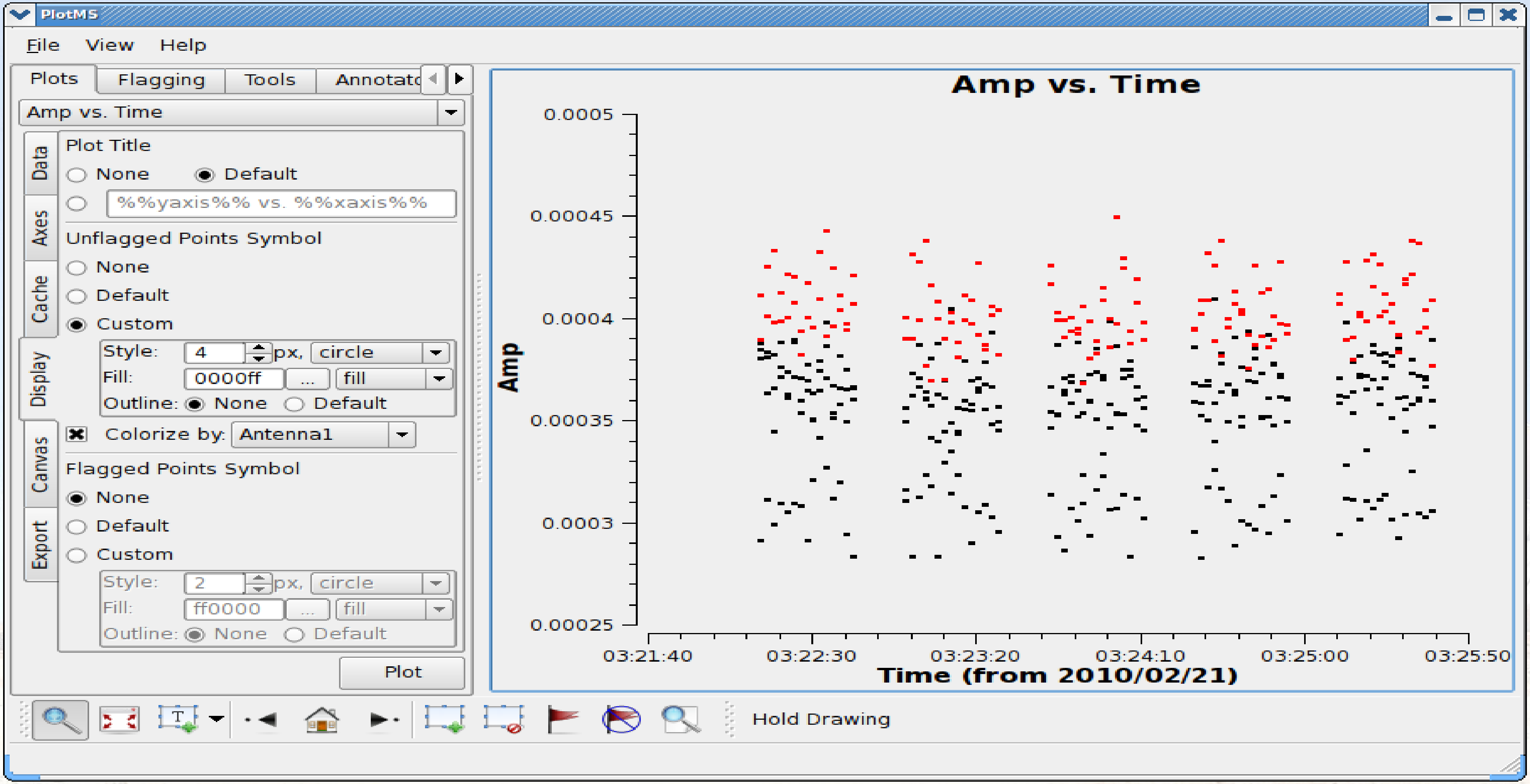
Max-Planck-Institut für Radioastronomie

Quicklook Phase RMS Display



Max-Planck-Institut für Radioastronomie

Quicklook Real Time Display



OMC-1. Plot of channel averaged data vs. time color coded by antenna1 for scan 10

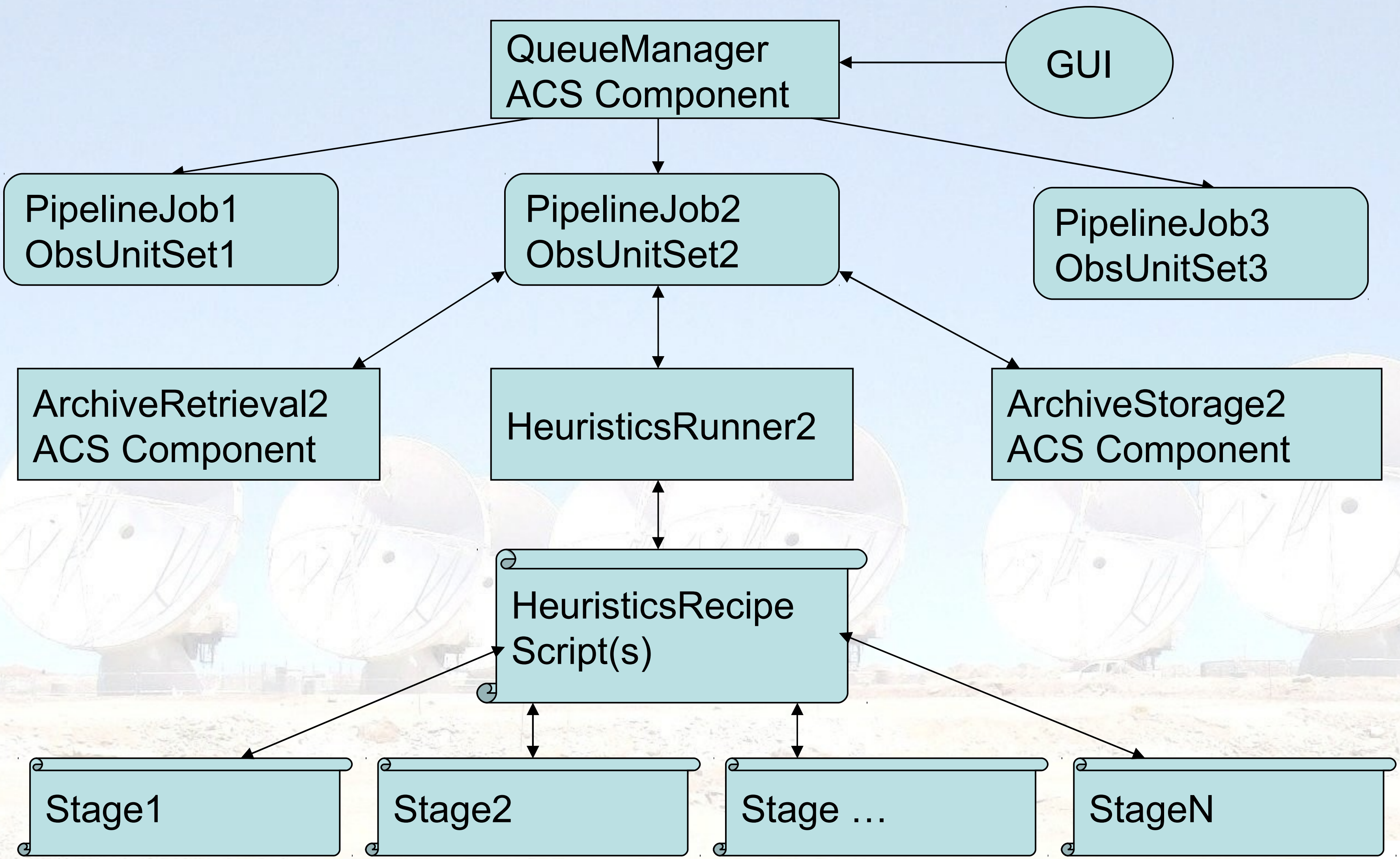
Max-Planck-Institut für Radioastronomie

Science Pipeline System

- Automatically reduce data sets taken in standard observing / reduction modes, e.g. single field, mosaic, single dish
- Store processing flags, calibration solutions, and reference images in the archive
- Store quality assessment metrics in the archive
- Store summary and detailed logs in the archive

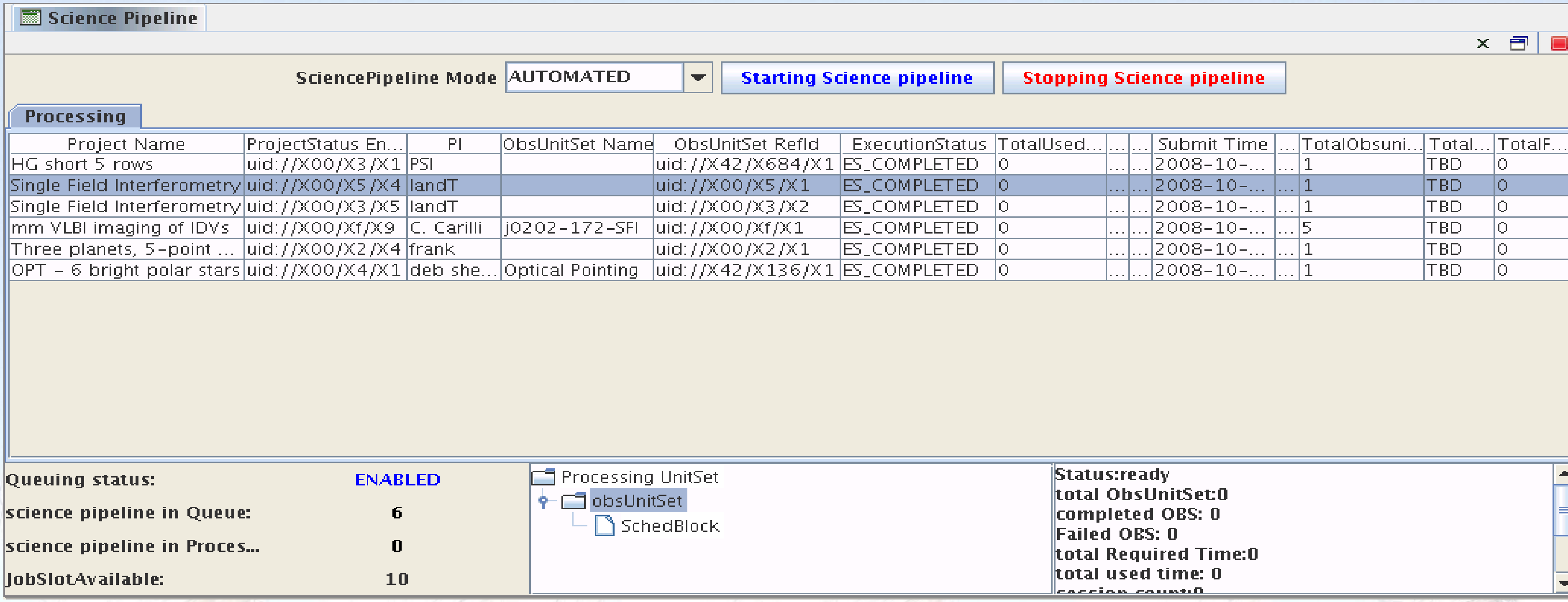
Max-Planck-Institut für Radioastronomie

Science Pipeline Structure



Max-Planck-Institut für Radioastronomie

Science Pipeline Queue Manager and GUI



The development version of the science pipeline queue manager and operator GUI. The GUI displays information about the projects in the queue and their execution status. The displayed projects are artificially constructed so some of the required information is missing. The queue shown here has been displayed after 6 projects have successfully completed.

Max-Planck-Institut für Radioastronomie

Heuristics

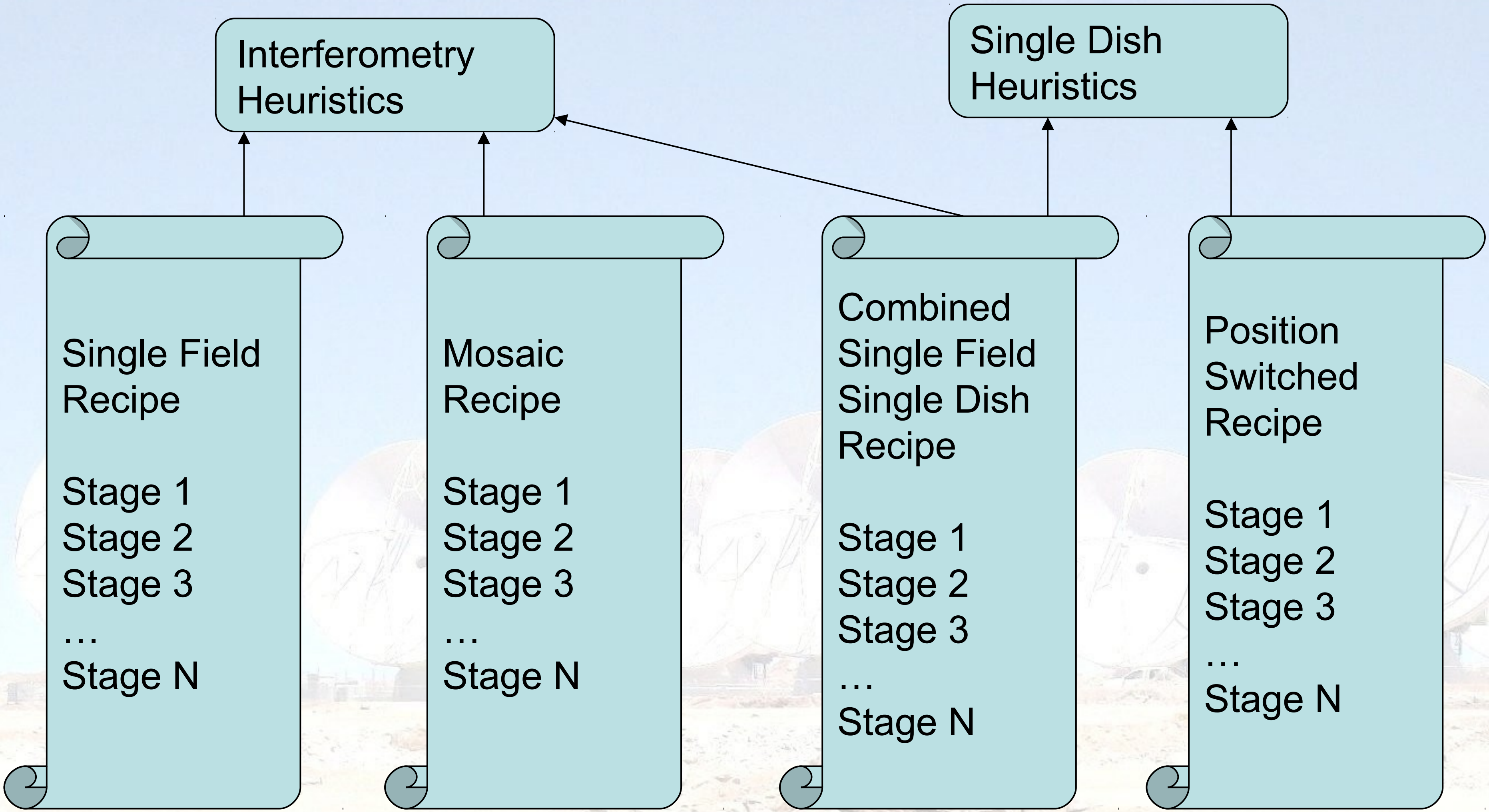
- A Heuristic is an experienced-based technique that helps in problem solving, learning and discovery.
- A heuristic method is particularly used to rapidly come to a solution that is hoped to be close to the best possible answer, or 'optimal solution'.
- Heuristics are "rules of thumb", educated guesses, intuitive judgments or simply common sense. A heuristic is a general way of solving a problem.

Wikipedia definition of Heuristic

Pipeline Heuristics

- The Pipeline Heuristics tries to capture the - sometimes diffuse - expert knowledge and encode it as data reduction ***recipes***
- There is a recipe per ***reduction mode***
- By now there are heuristics recipes for
  - Single field interferometry
  - Pointed Mosaics
  - Single dish data
- IF/SD combination recipe is under development

Heuristics Design Overview



Max-Planck-Institut für Radioastronomie

Requirements

- There are also some explicit requirements for algorithms from experience with existing interferometers:
  - The SSR requirements document defines the range of observing (→ *reduction*) modes and algorithms to be implemented
  - Granular requirements derived from the original SSR document are the main input for our developments
  - ALMA SciOps added a few new items in 2009

Developments since 2003

- Initial heuristics scripts in glish using AIPS++
- Porting to proto-pipeline (AIPS++ / ACS) and later to CASA
- Development of interferometry (IF) and single dish (SD) ***recipes***
- Several external User Tests
- Switch from monolithic scripts to modules, commands, and finally CASA tasks
- Adjustment to ALMA SD (2009) and IF (2010) data

Max-Planck-Institut für Radioastronomie

Expert Knowledge

- Began with data reduction scripts from OFFLINE (i.e. CASA) subsystem user tests
- Talked a lot with experts to learn about their approaches to data reduction
- Condensed this expertise into algorithms
- Tested and refined them using datasets of existing observatories (PdB, VLA, SMA, IRAM 30m, APEX, HHT, Effelsberg, GBT). Since 2009 also with ALMA SD and IF data.

Interferometry Heuristics

- The recipe is laid out in a series of 'stages'.
- The idea is that as the reduction of a dataset moves through the stages the bad data are gradually removed and the best methods for calibrating the data are found.
- In the final stages the cleaned images and other data products are calculated.

Interferometry Flagging

- Mainly flag calibrators
- Go from coarse to fine flagging
- Working on several kinds of data slicing / averaging, mostly antenna based
- Flag noisy phases / amplitudes
- Flag time series outliers
- Using MAD (Median Absolute Deviation) as a more robust measure of variability

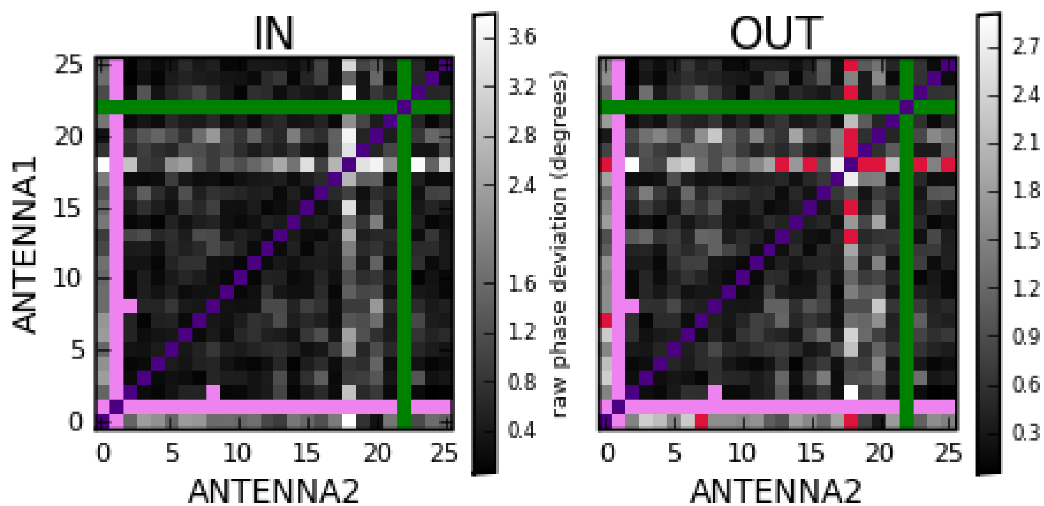
$$\mathrm{MAD} = \mathrm{median}_i\left(\left|X_i - \mathrm{median}_j(X_j)\right|\right),$$

Max-Planck-Institut für Radioastronomie

IF Flagging Example

STAGE: 8 Flag calibrator baselines with noisy phases
Field:1331+305 (BANDPASS+FLUX)
SpW: 0 Pol:RR raw phase deviation

Flagging noisy calibrator phases. Each data point in the display shows a measure of the phase noise for that baseline.



No data
Flagged at previous stages:
cannot calculate on recipe entry
cannot calculate at stage entry
cannot calculate now
Flagged here:
rules:
high outlier
ANTENNA2 axis - too many flags

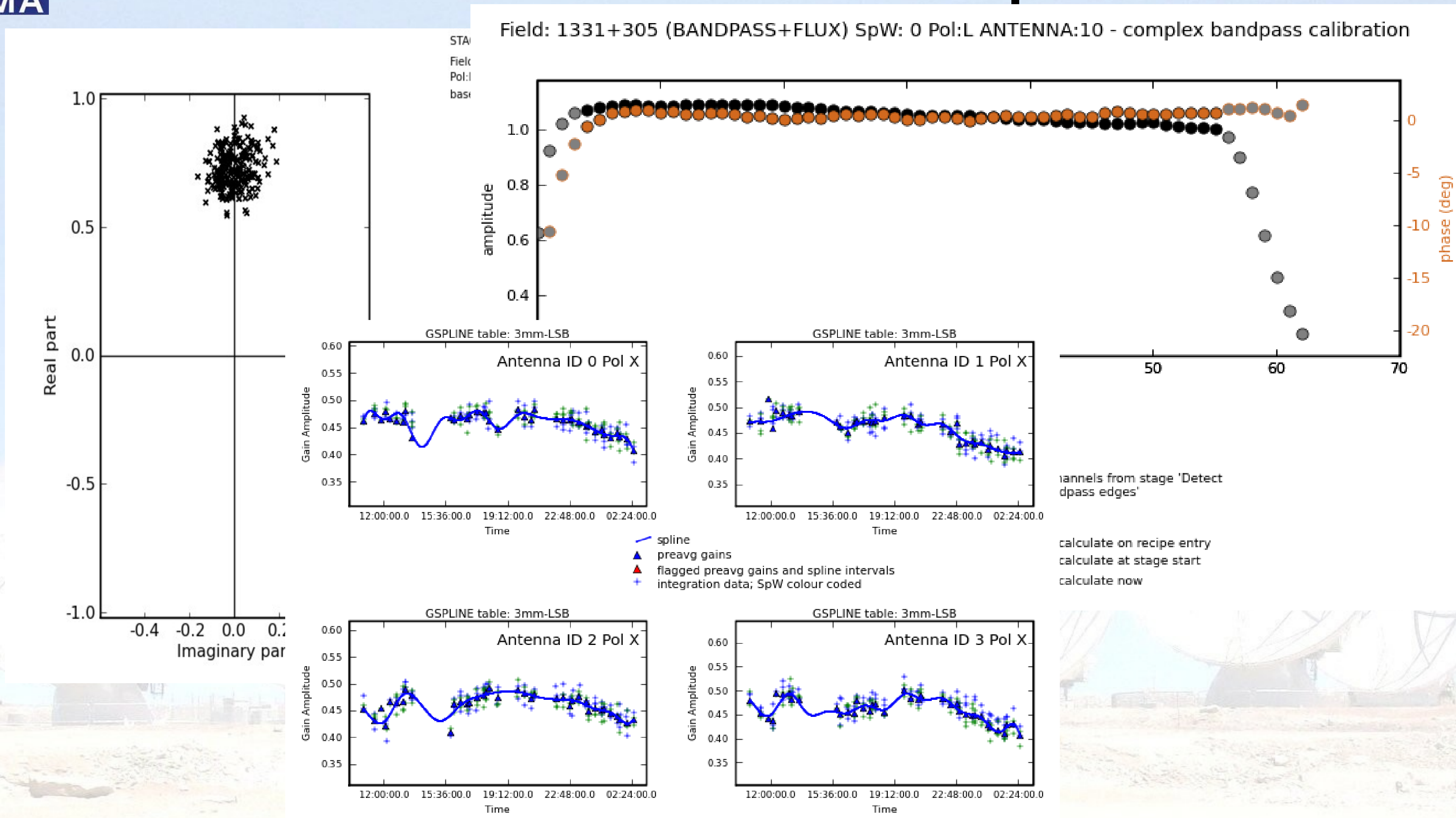
Max-Planck-Institut für Radioastronomie

Interferometry Calibration

- Find best bandpass solution based upon flatness choosing from different algorithms (channel-wise, polynomial, different times,...)
- Flag gain calibrations with bad S/N
- Flag antennas with large phase jumps
- Flag large residual deviations
- Spectral window grouping to increase S/N
- Fit splines to time series of gain calibrator measurements

Max-Planck-Institut für Radioastronomie

IF Calibration Examples



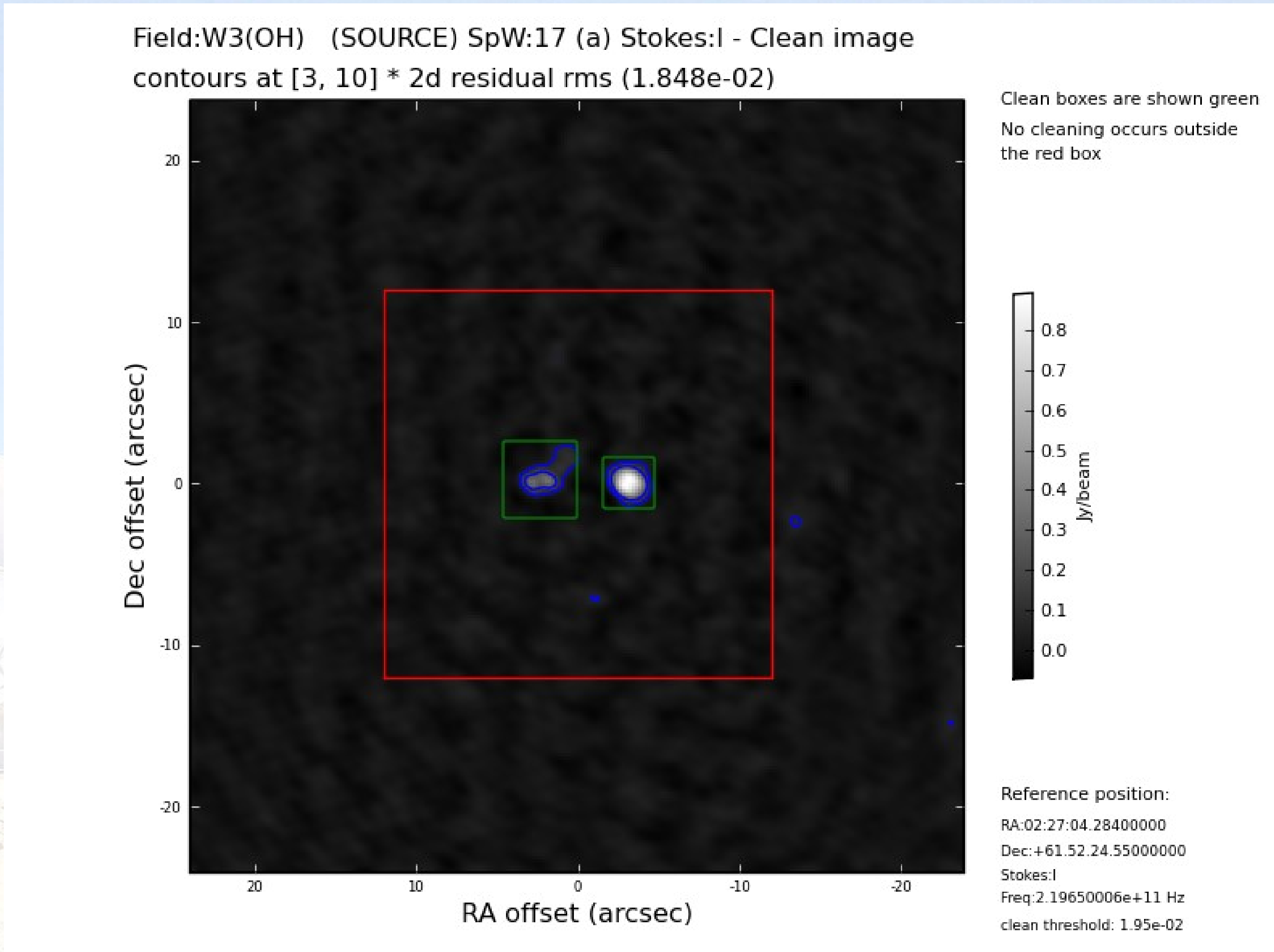
Max-Planck-Institut für Radioastronomie

Interferometry Imaging

- First make 'pilot' integrated image to find sources and define clean boxes / polygon
- Iterative clean loop (currently Hogbom, natural weighting)
- Continuum subtraction based on predefined line windows; later automatic line finding from SD recipe
- Pointed Mosaic cleaning

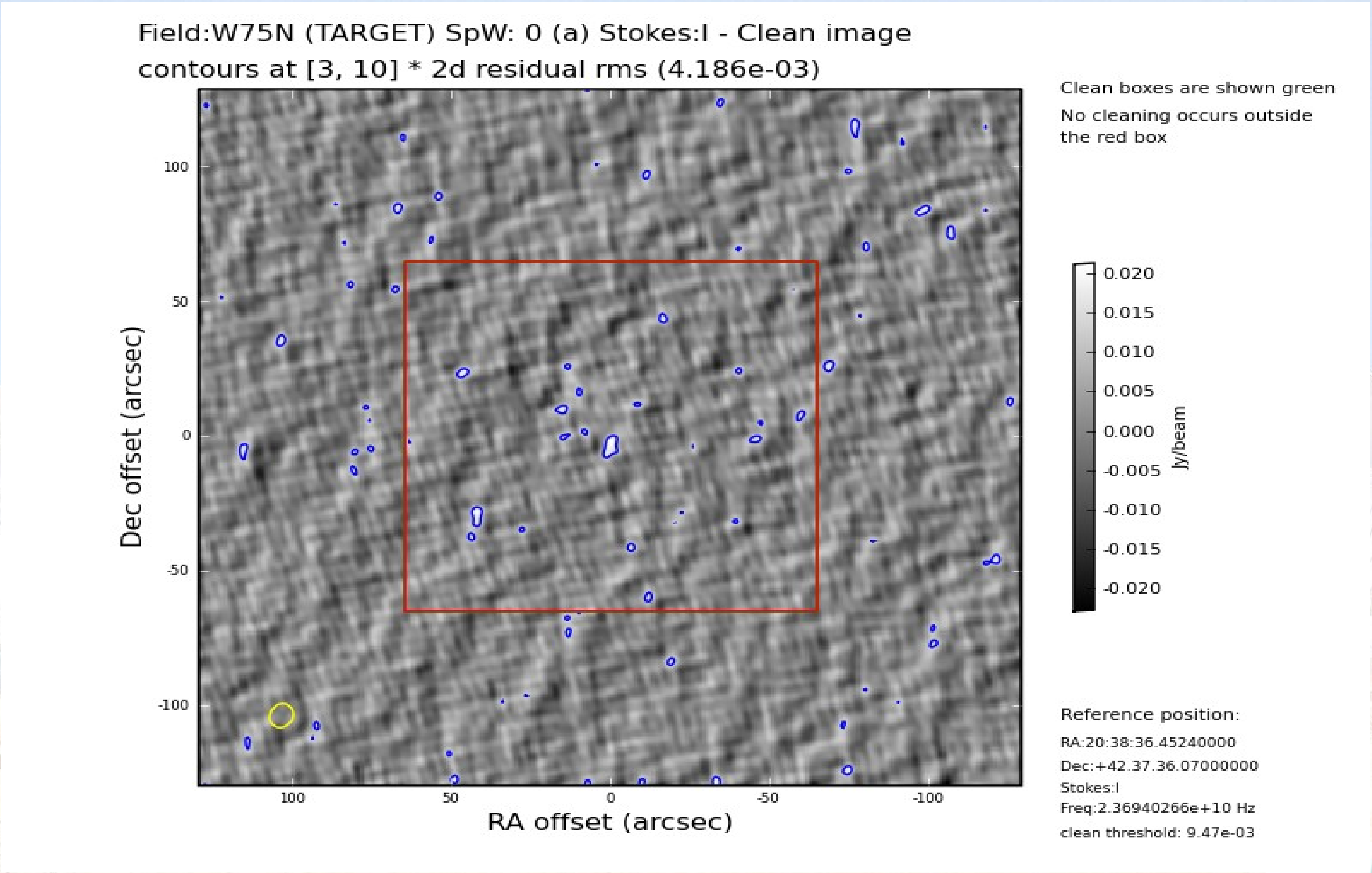
Max-Planck-Institut für Radioastronomie

Single Field Imaging Example



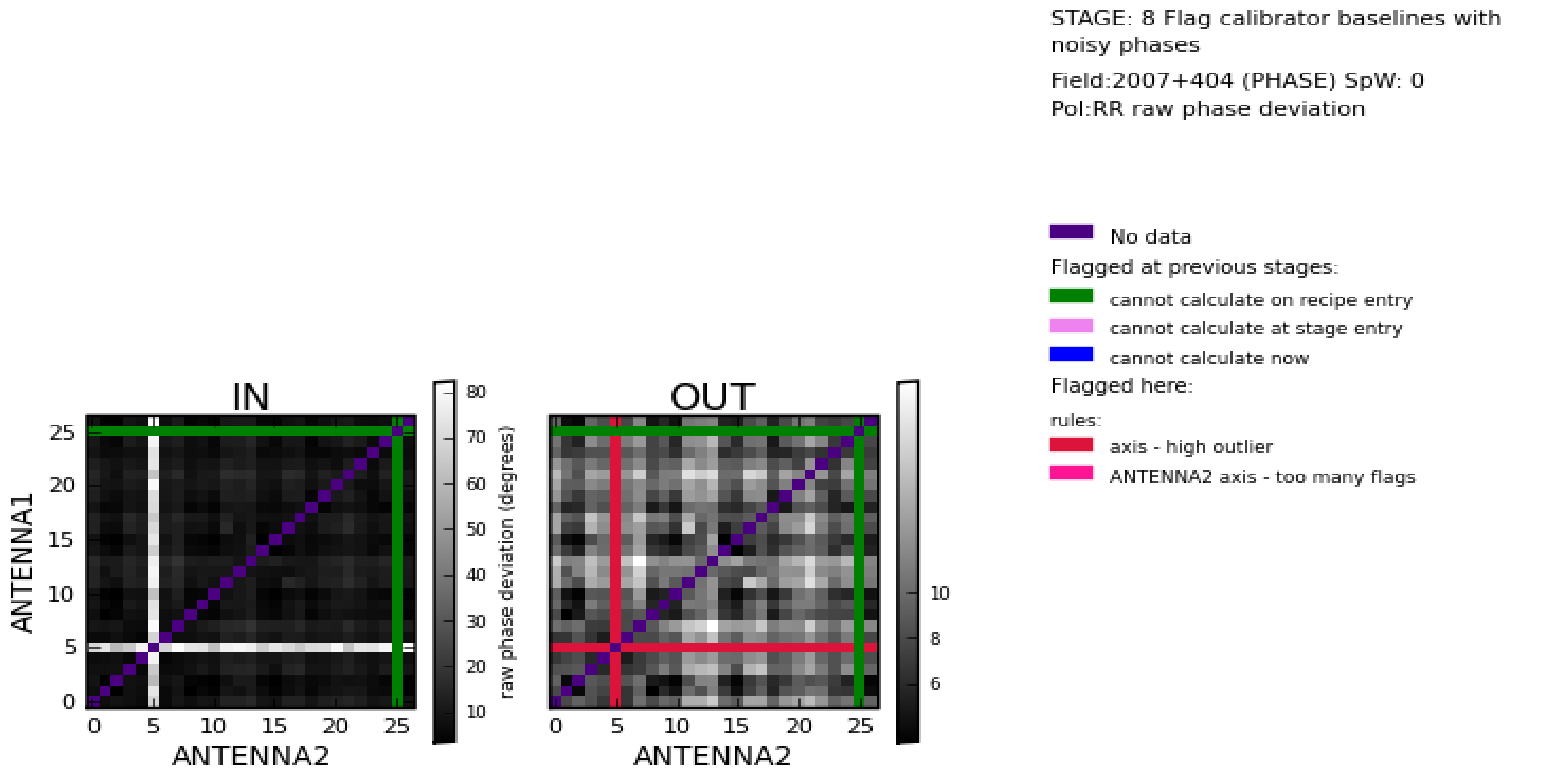
Max-Planck-Institut für Radioastronomie

Single Field VLA Initial Target Image

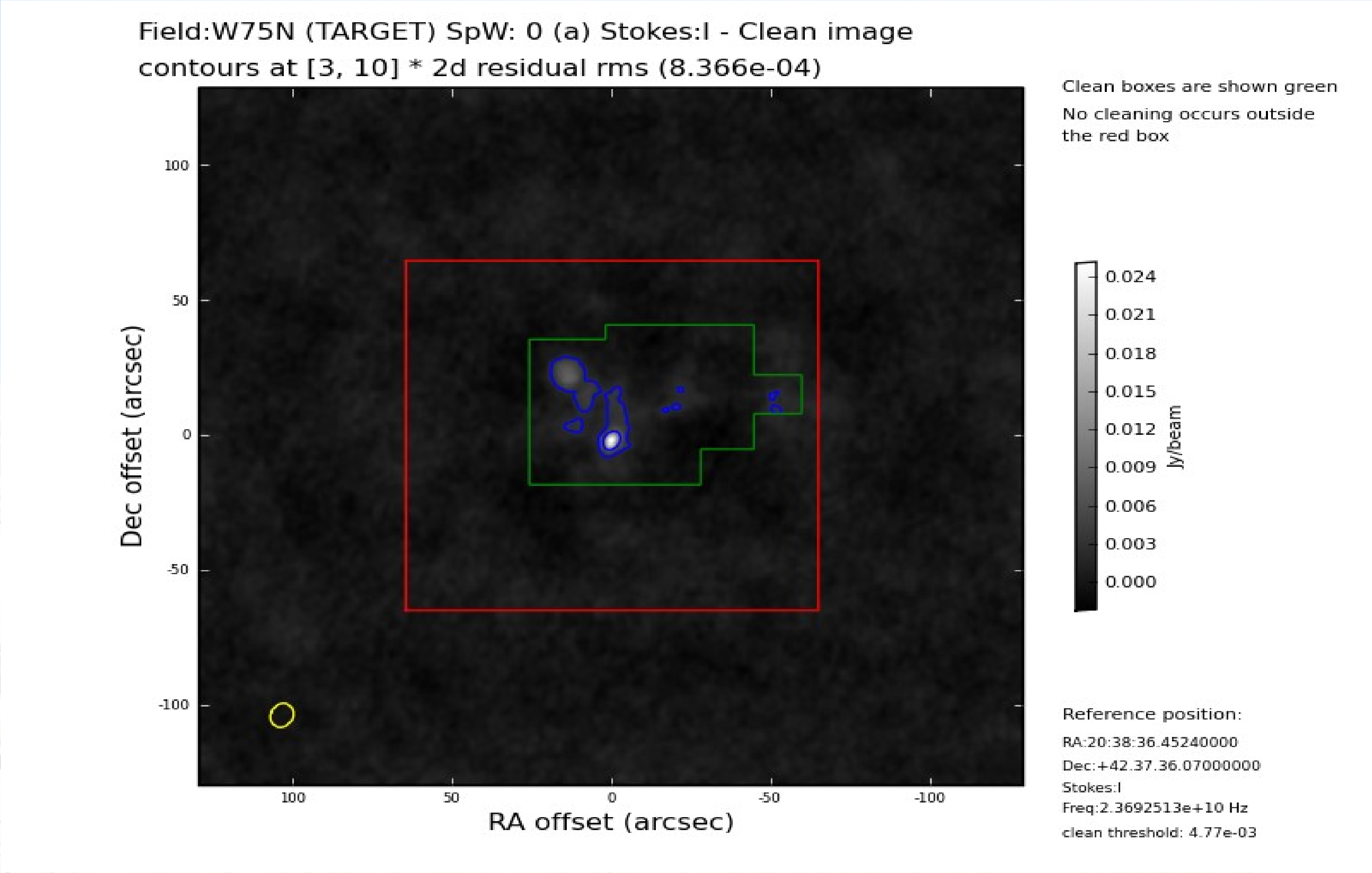


Max-Planck-Institut für Radioastronomie

Single Field VLA Noisy Phase Flagging

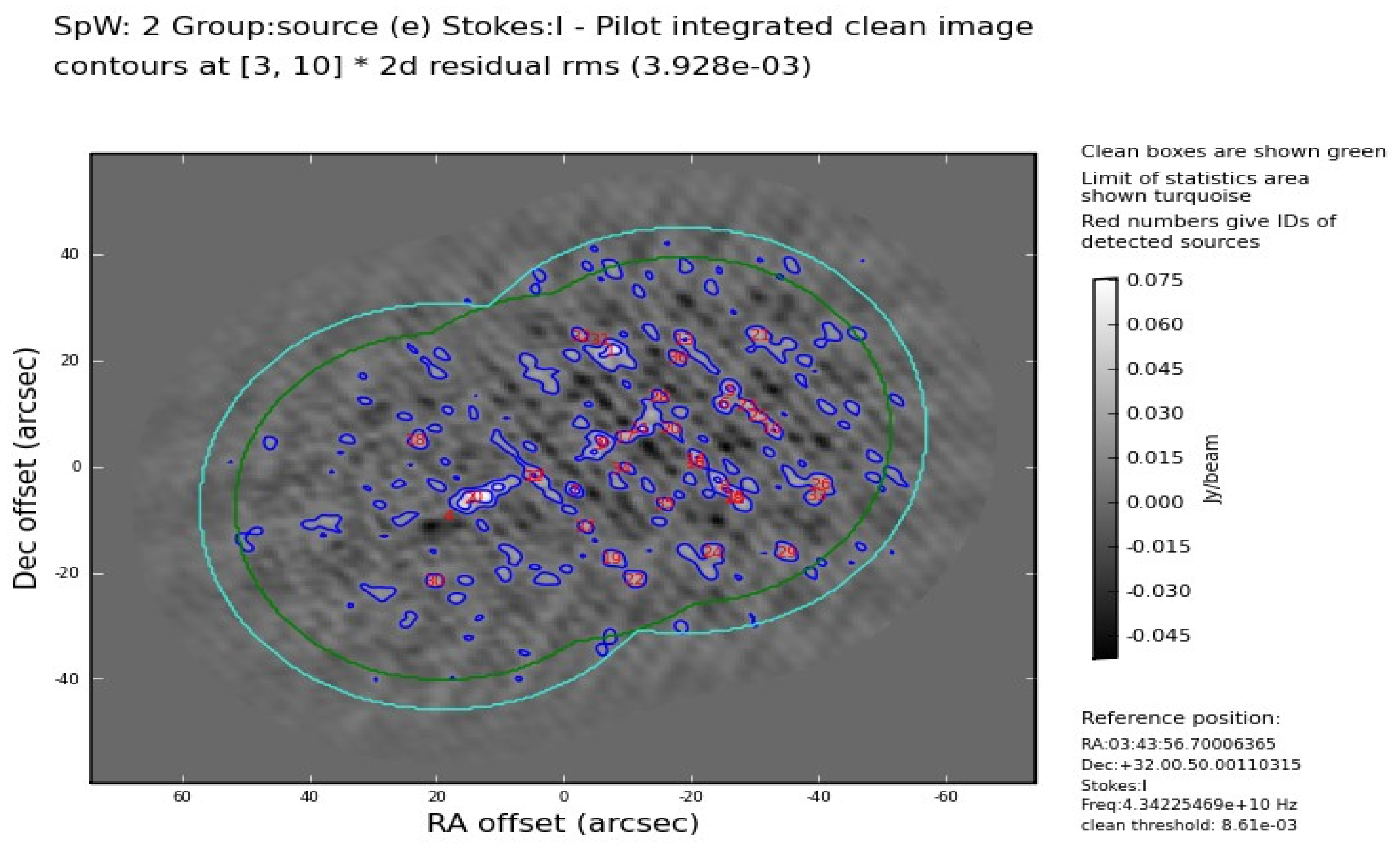


Single Field VLA Final Target Image



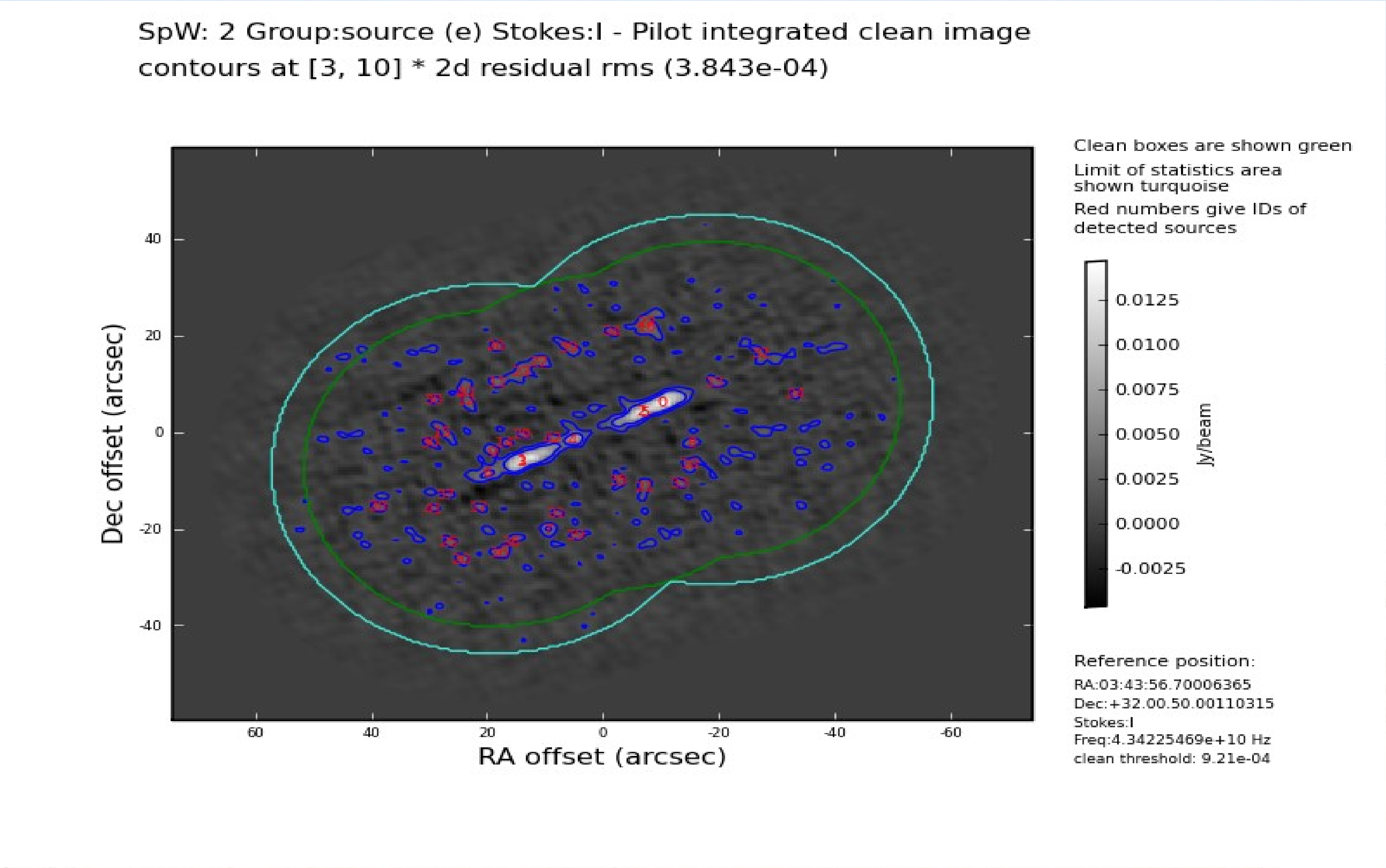
Max-Planck-Institut für Radioastronomie

VLA Mosaic Initial Continuum Image



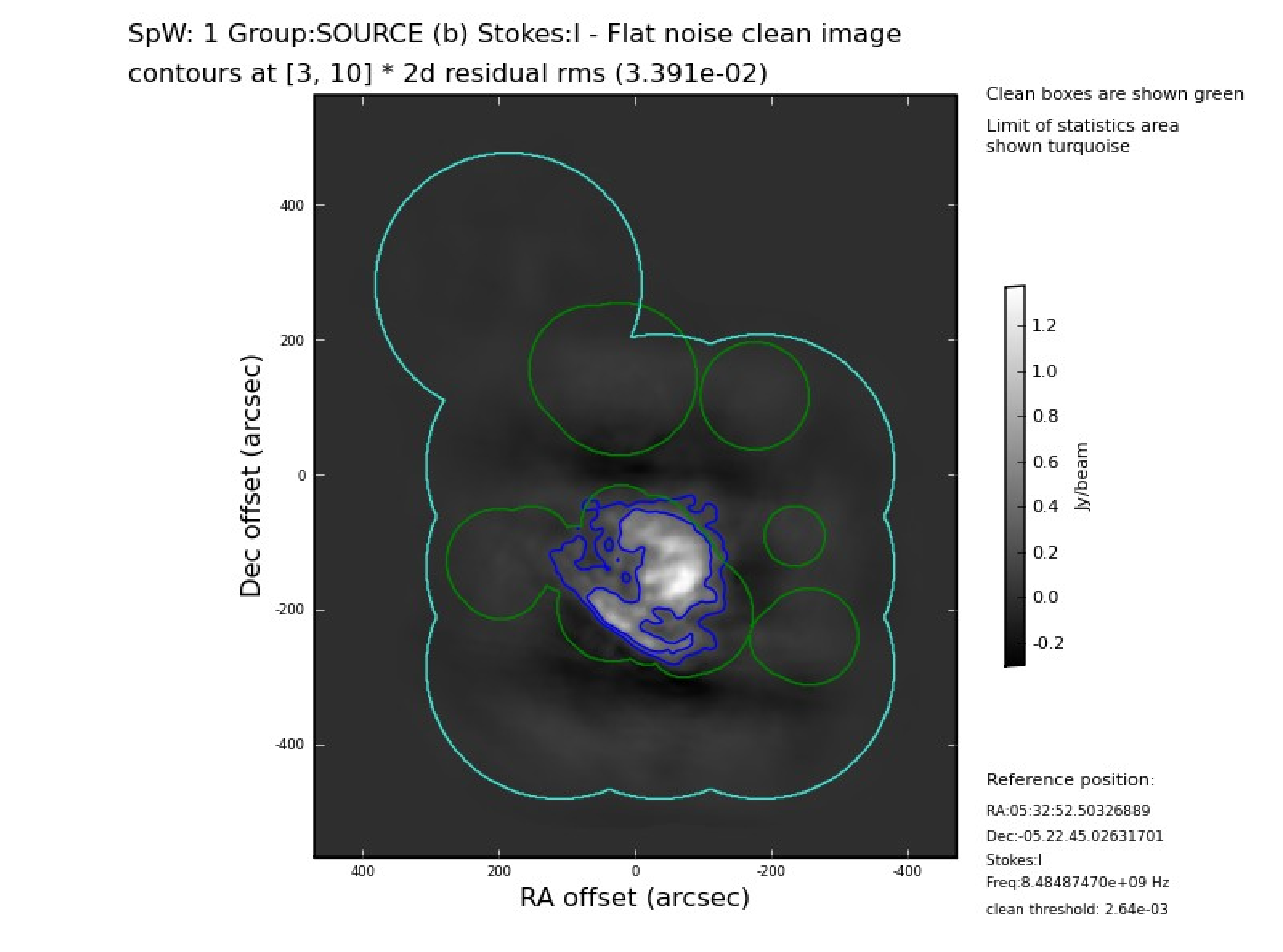
Max-Planck-Institut für Radioastronomie

VLA Mosaic Final Continuum Image



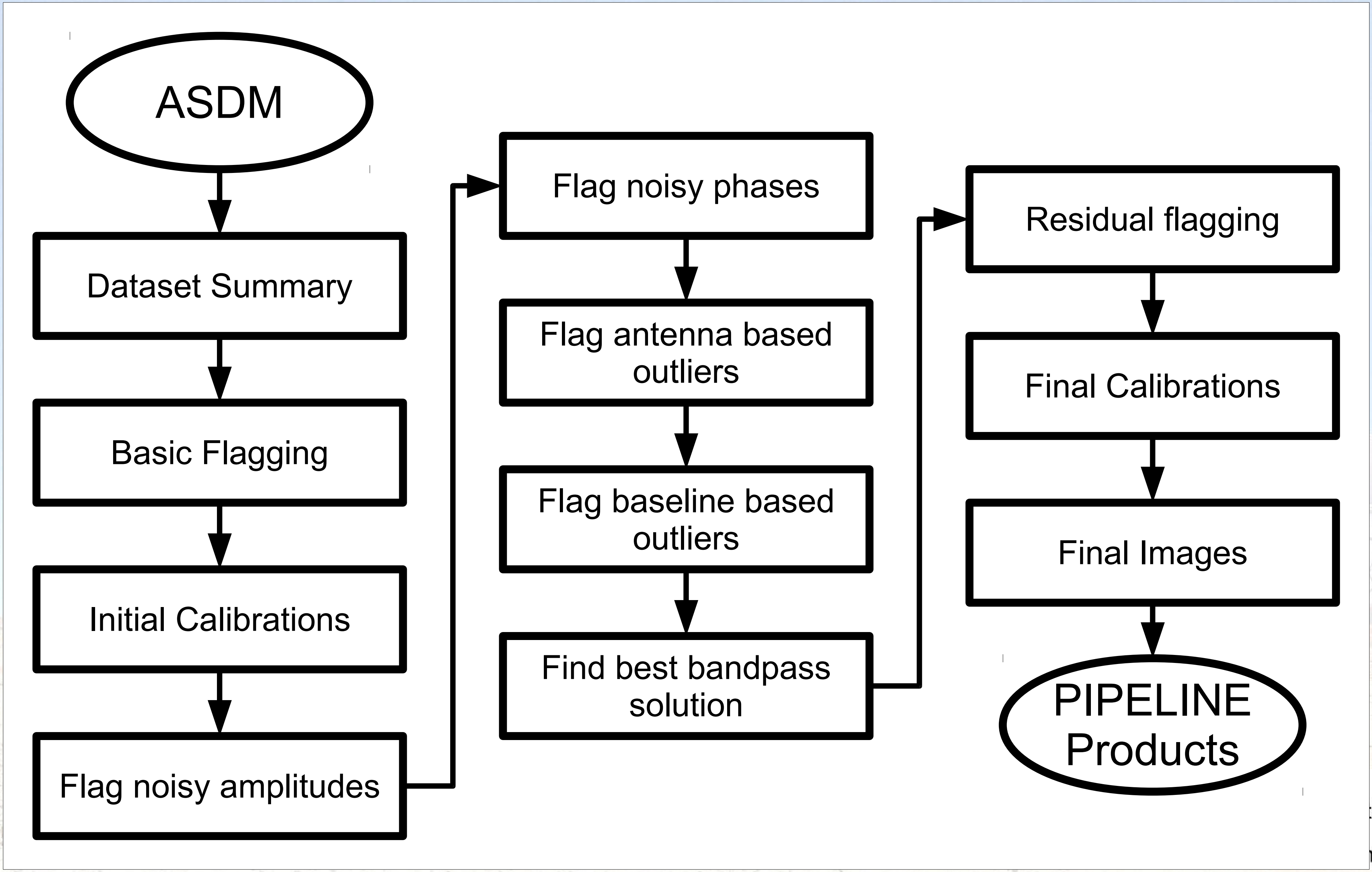
Max-Planck-Institut für Radioastronomie

Larger Mosaic Example



Max-Planck-Institut für Radioastronomie

IF Heuristics Workflow



Single Dish Heuristics

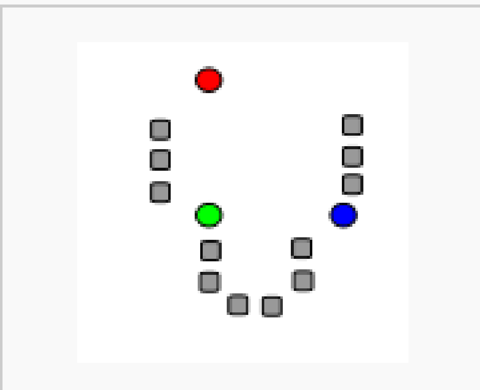
- The SD recipe implements stages to
  - Calculate calibrated spectra for different observing switch modes (on-off, nutator, frequency switching)
  - Automatically identify lines
  - Fit spectral baselines
  - Flag bad data
  - Image the data
  - Line finding, baselining and flagging are iterated in a loop

Max-Planck-Institut für Radioastronomie

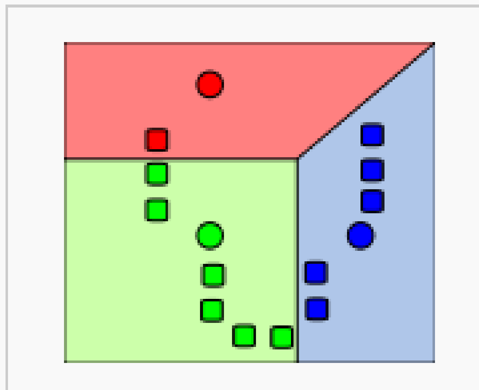
SD Line Finding / Baselining

- Line finder searches for Gaussian components
- K-means algorithm is used to cluster line windows spatially

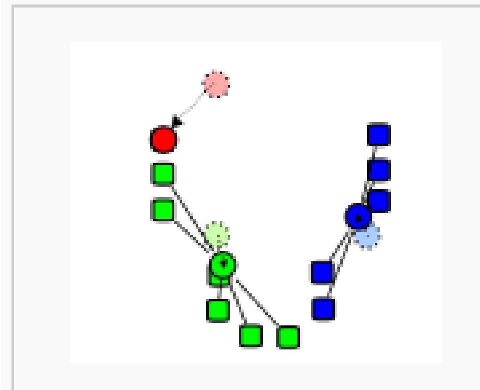
K-Means Clustering



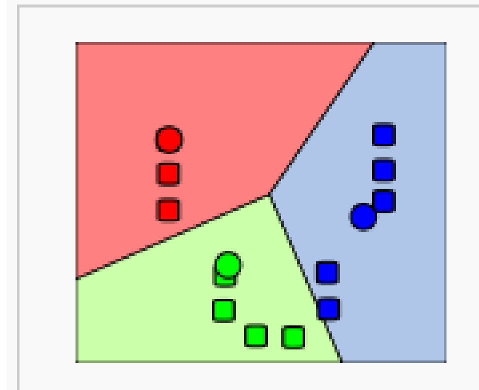
1) *k* initial "means" (in this case *k*=3) are randomly selected from the data set (shown in color).



2) *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.



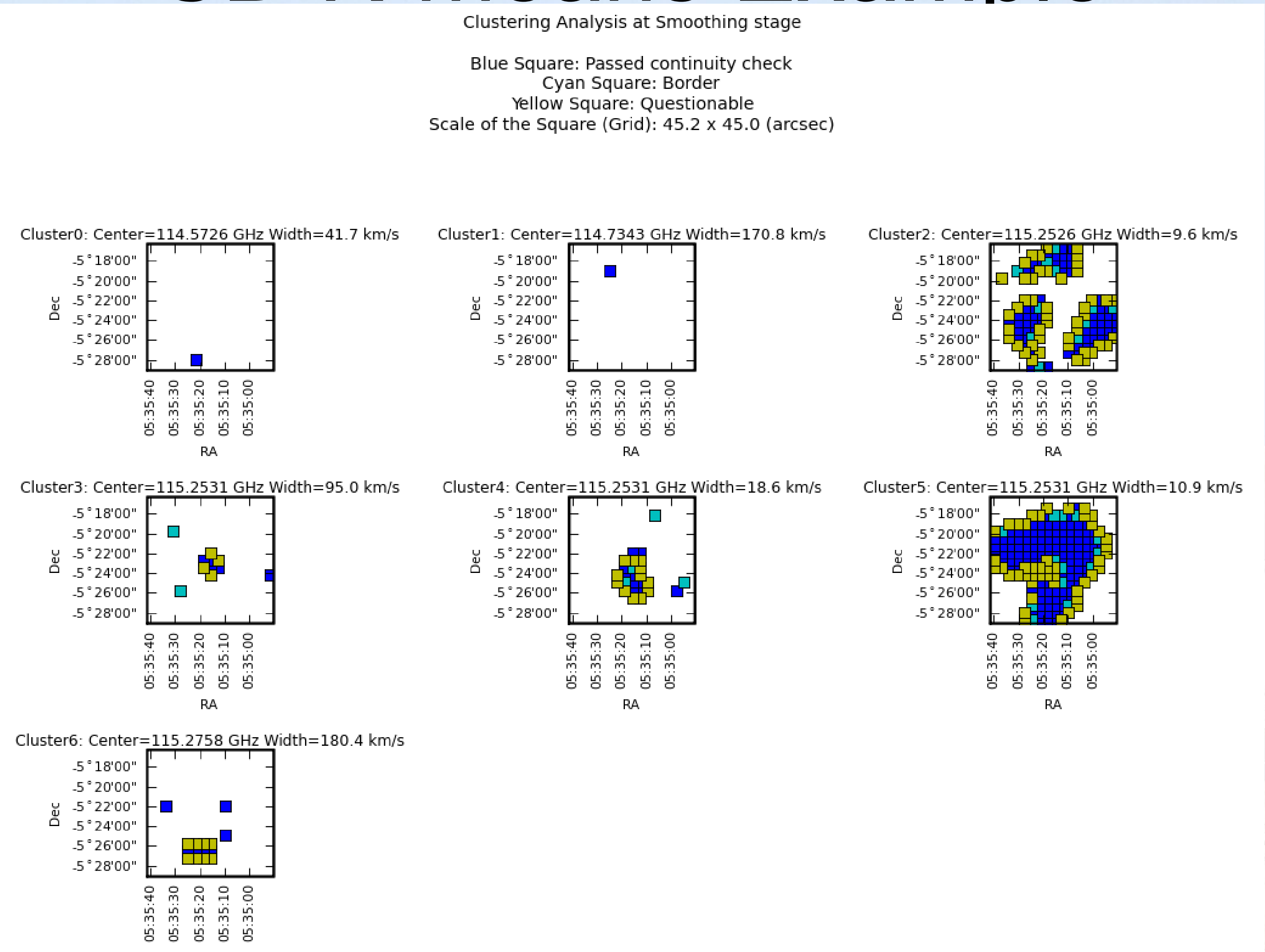
3) The centroid of each of the *k* clusters becomes the new means.



4) Steps 2 and 3 are repeated until convergence has been reached.

Wikipedia

SD K-means Example



Max-Planck-Institut für Radioastronomie

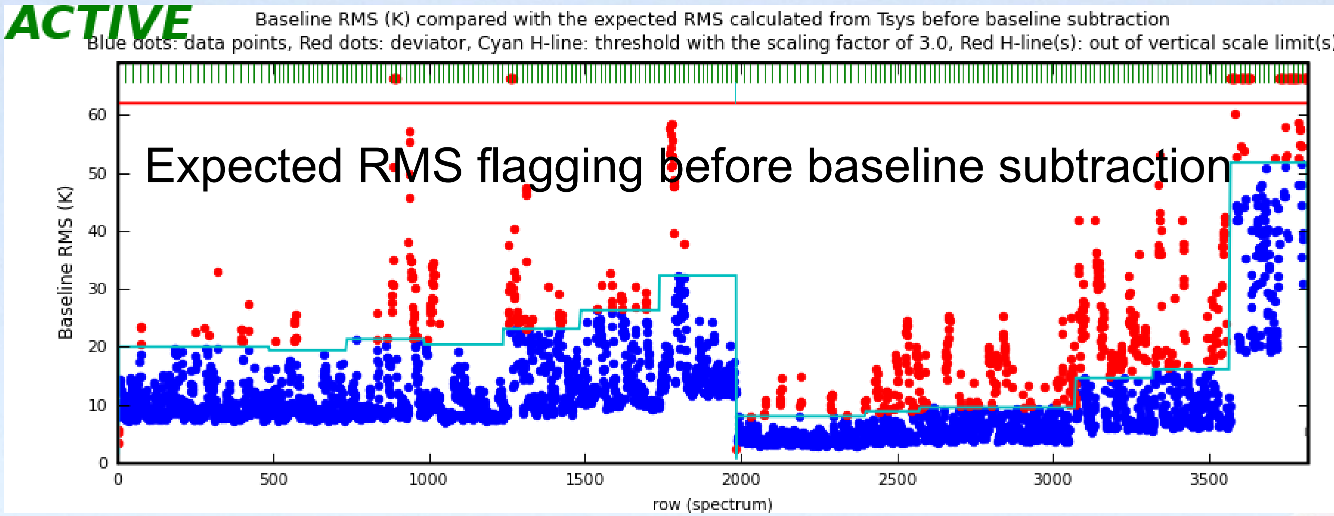
SD Line Finding / Baselining

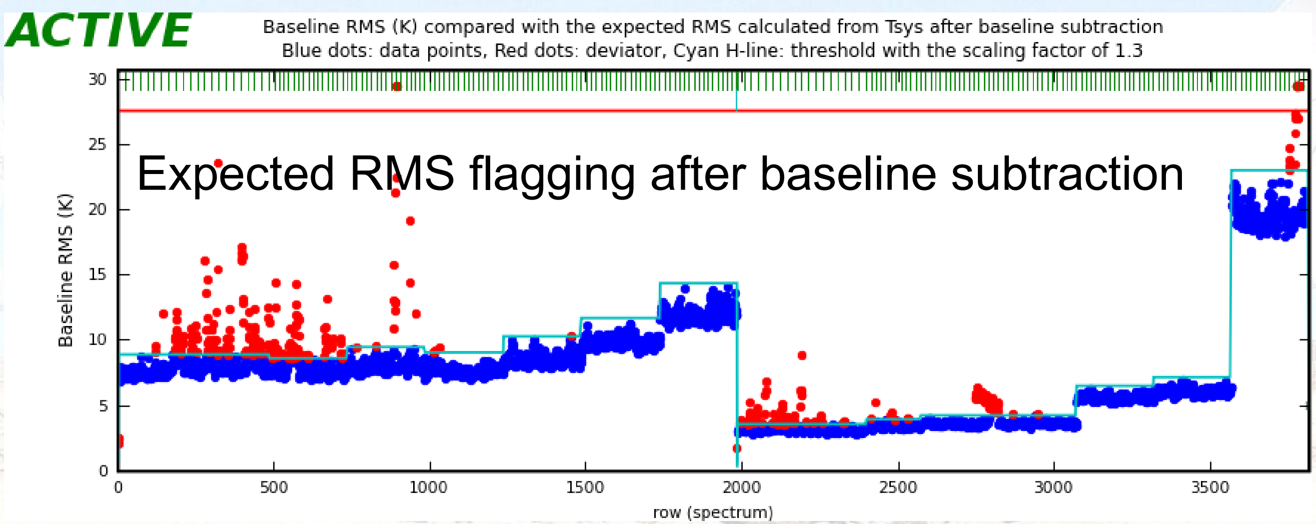
- Line finder searches for Gaussian components
- K-means algorithm is used to cluster line windows spatially
- Spectral baseline order is determined by principal component analysis of line-free region
- Baseline order is capped for broad lines
- Polynomial and Spline fit baselines are available

Single Dish Flagging

- Several flagging rules are implemented to flag:
  - High spectral baseline RMS
  - Temporal running mean outliers
  - Expected RMS (radiometer formula) outliers
  - Weather / PWV conditions
  - Tsys outliers

SD Flagging Examples



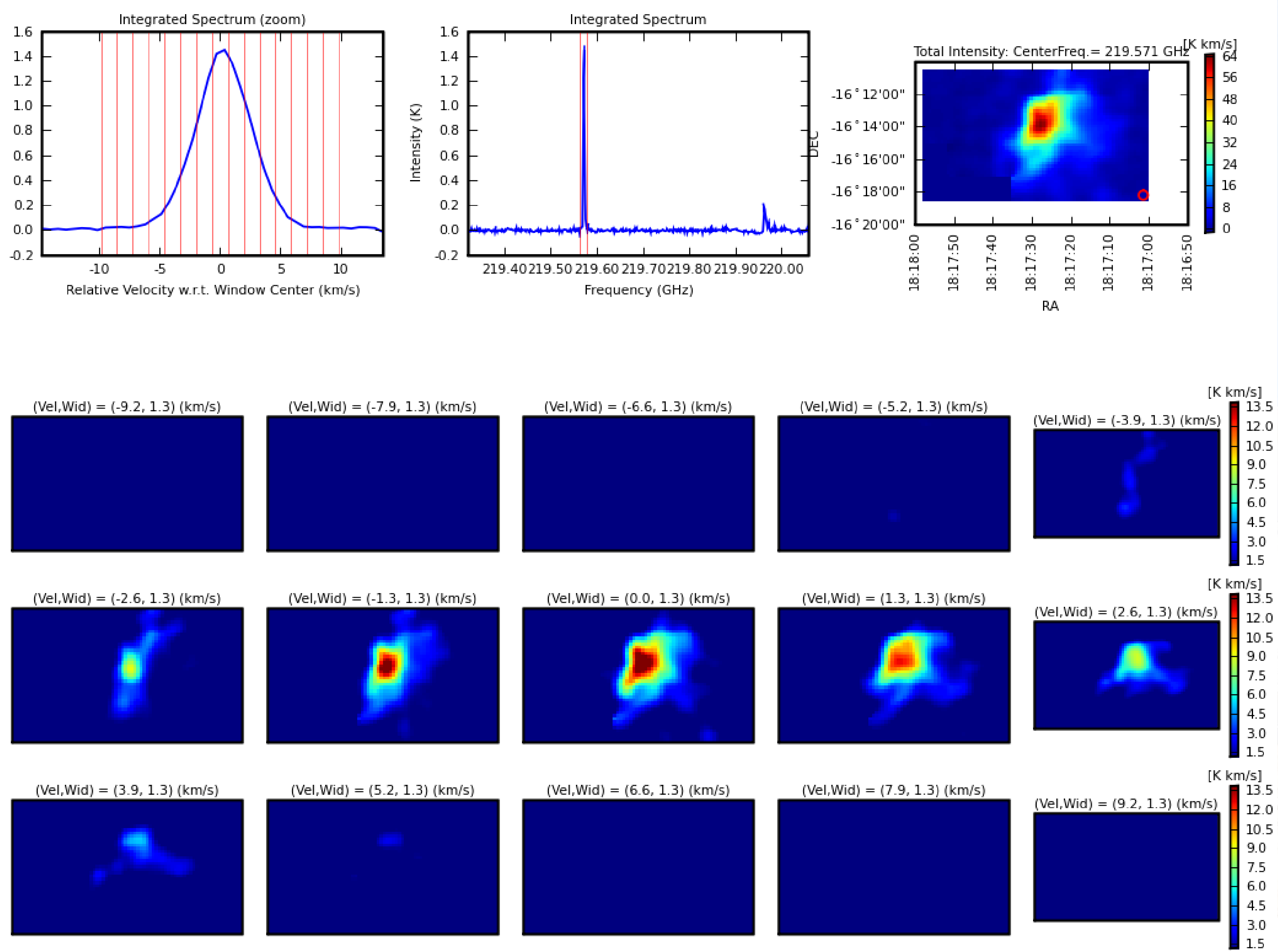


Max-Planck-Institut für Radioastronomie

Single Dish Imaging

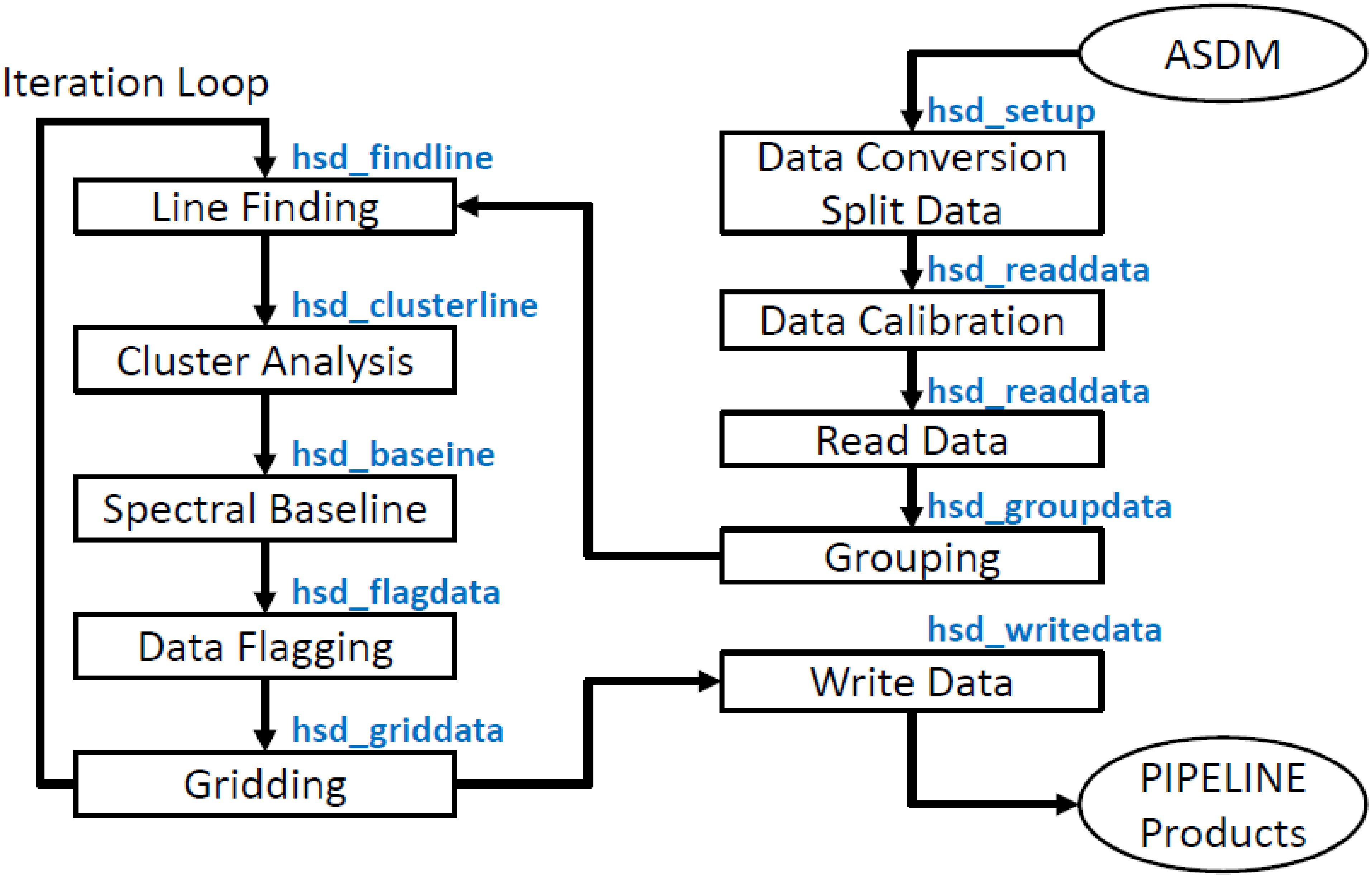
- Data cubes are gridded and channel maps are made per found line window
- Line-free region is used to produce (pseudo-) continuum images; optionally with spatial baseline

SD Imaging Example



Max-Planck-Institut für Radioastronomie

(Simplified) SD Heuristics Workflow



Results Presentation

- The heuristics recipes produce plots and logs for each stage
- A hierarchy of web pages with sidebar navigation is provided to easily browse the results
- Dynamic thumbnail pages with filters facilitate access to result details

Interferometry Web Results

flagbackup=False)

- 15 Initial Bandpass Calibration
- 16 Delay Calibration
- 17 Corrected Bandpass Calibration
- 18 Initial Gain Calibration

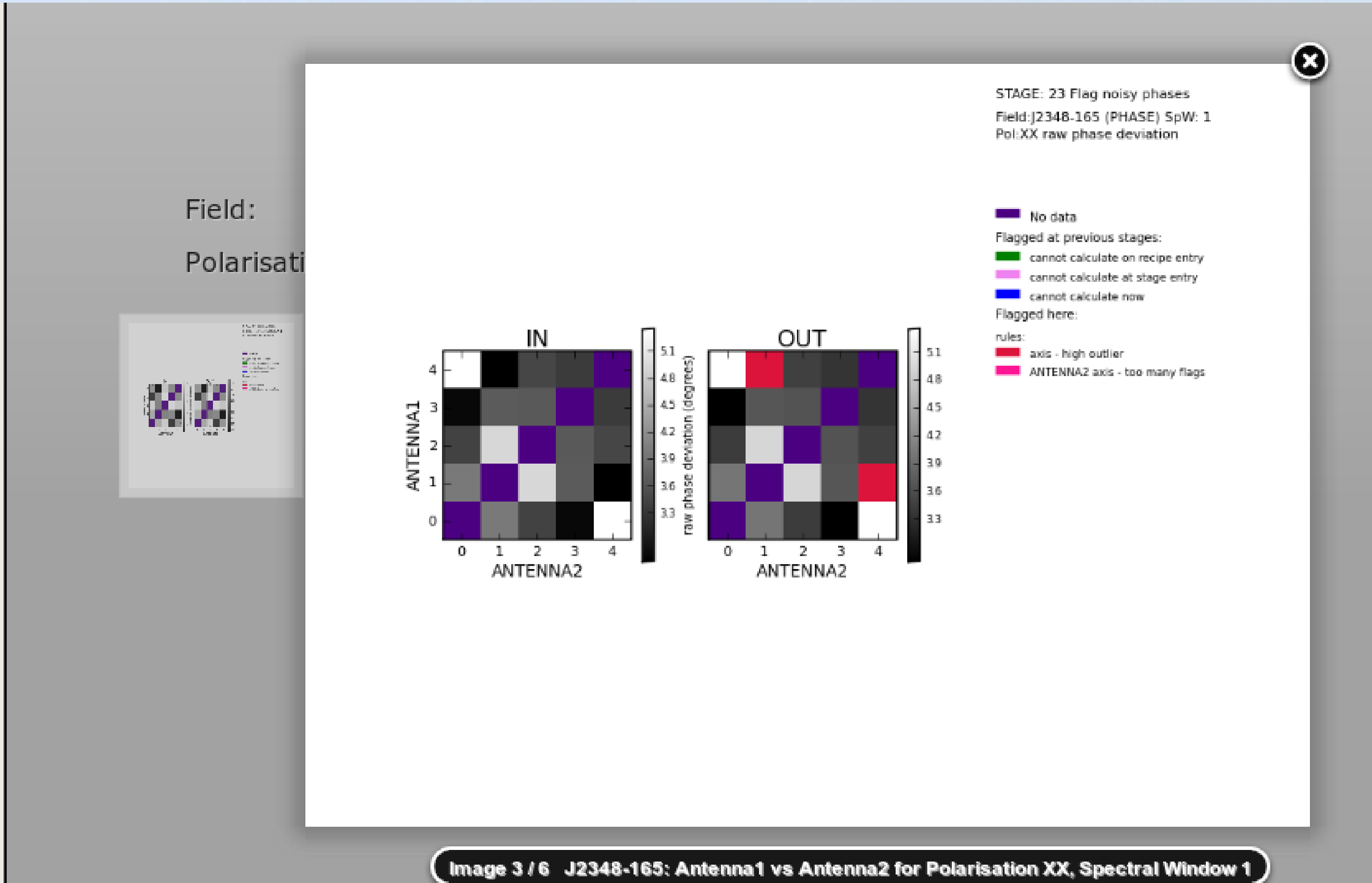
Combine and image all datasets.

- 19 calibrator cleaned integrated maps
- 20 calibrator cleaned integrated maps
- 21 Target cleaned integrated maps

Flag Raw Calibrator Data

vis=uid__A002_X10e085_X

- 22 Flag noisy amplitudes
  90 rows flagged
- 23 Flag noisy phases
  24 rows flagged
- 24 Flag antenna



Single Dish Web Results

SD Pipeline

- Summary
- Clustering
- BaselineFit
- FitStatistics
- MultiSpectra
- Gridding
- SparseSpectraMa
- Log

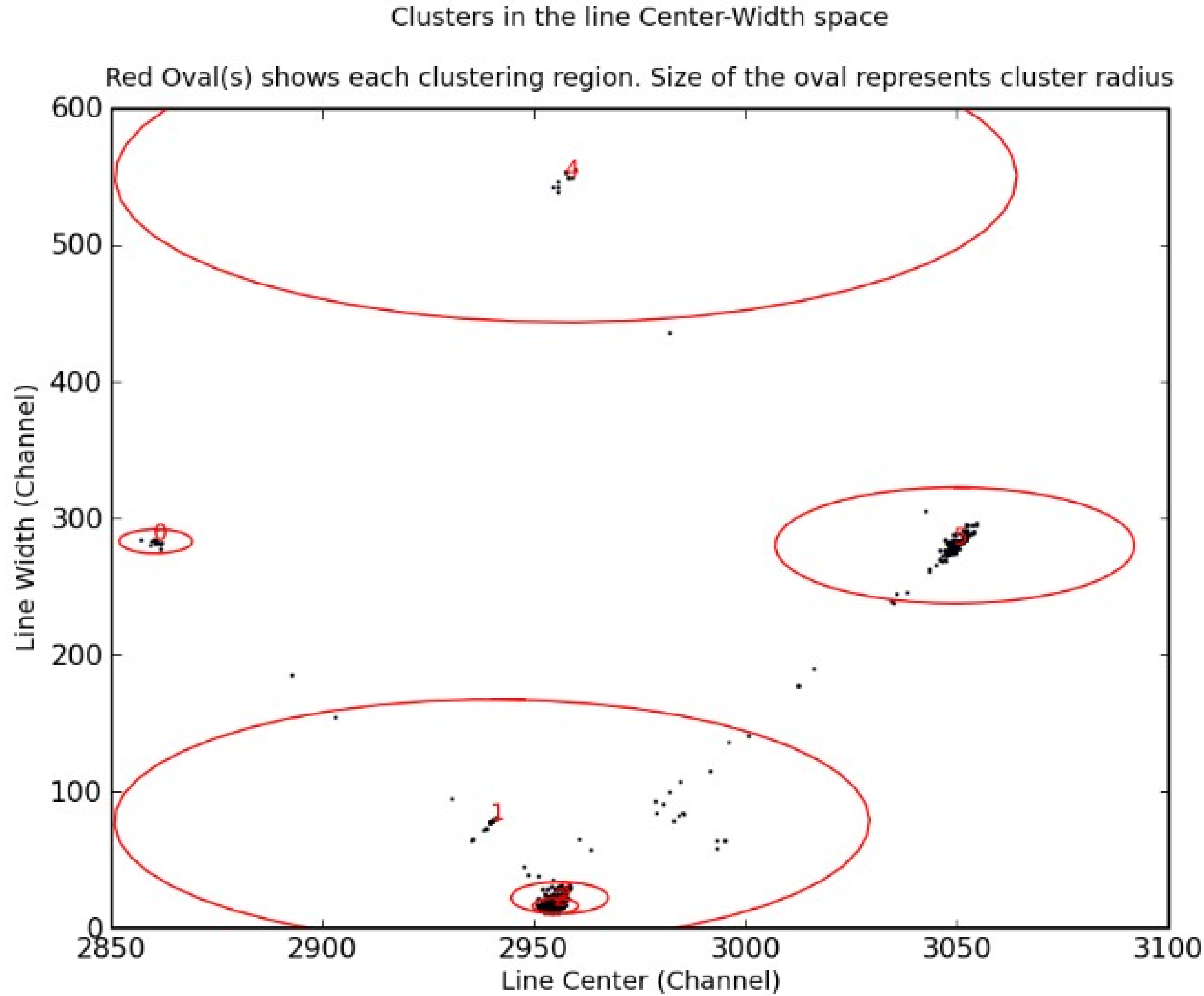
Clustering Analysis

Description

Itr: Iteration Cycle
IF: Frequency ID
Pol: Polarization
Stage
  Map: Cluster Space
  D: Detection
  V: Validation
  S: Smoothing
  R: Regions

< Prev   Next >

2	0	Regions	4
2	0	Map	5



Data Products

- The final data products stored in the archive comprise
  - Raw data
  - Flagging tables
  - Calibration tables
  - Images / Data cubes (FITS)
  - Quality Assessment measures
  - CASA logs
  - Web results pages
  - Heuristics scripts

Heuristics Task Interface

- The heuristics stages commands are implemented as additional CASA tasks
- Naming prefixes are used to distinguish the tasks:
  - hif_ – heuristics interferometry
  - hsd_ – heuristics single dish
  - hco_ – heuristics IF/SD combination
- Currently the tasks are distributed as an add-on
- Later they will be part of CASA

Max-Planck-Institut für Radioastronomie

SD Task Sample

```
CASA <24>: inp hsd_configflag
---------> inp(hsd_configflag)
#  hsd_configflag :: Activate and configure flag rules
tsys                =       True         #  Tsys Flag rule
     ttsys          =        3.0         #  Tsys Flag rule Threshold

weather             =      False         #  Weather Flag rule
preexprms           =      False         #  RmsExpectedPreFit Flag rule
postexprms          =       True         #  RmsExpectedPostFit Flag rule
     tpostexprms    =     1.3333         #  RmsExpectedPostFit Flag rule
                                         #   Threshold

prerms              =      False         #  RmsPreFit Flag rule
postrms             =       True         #  RmsPostFit Flag rule
     tpostrms       =        4.0         #  RmsPostFit Flag rule Threshold

prerunmean          =      False         #  RunMeanPreFit Flag rule
postrunmean         =       True         #  RunMeanPostFit Flag rule
     tpostrunmean   =        5.0         #  RunMeanPostFit Flag rule Threshold
     npostrunmean   =          5         #  RunMeanPreFit Flag rule Nmean

userflag            =      False         #  User Flag rule
```


Max-Planck-Institut für Radioastronomie

Pipeline User Tests

- The heuristics developments are being tested regularly by external astronomers
- The test reports provide feedback for further heuristic developments
- The single dish heuristics was judged to produce science quality data products in UT5/UT5A (2009)
- The interferometry heuristics passed UT6 (2010)

Pipeline Commissioning

- We already began adjusting the heuristics to available ALMA data
- The pipeline will be commissioned during ALMA Early Science
- The automatic results will be compared against manual reductions and parameters and algorithms will be tuned

Ongoing Work

- Interferometry
  - Resolved calibrators
  - Phase transfer
  - Polarization
  - Self calibration
  - Different imaging algorithms
- Single Dish
  - More complex continuum processing
  - Using line catalogs for feature identification
- 12m - 7m - SD combination

Releases

- First release in September 2010, mainly aiming at ALMA staff and the ARCs; based on UT6 scripts; ALMA SD OK; Interferometry only for PdB and VLA
- After iterations of ALMA recipes and task names the next release is planned for February 2011
- During commissioning there will be further releases with improvements and bug fixes
- Releases will be aligned with CASA cycles
- The JIRA system will be used for bug reports

Pipeline Links

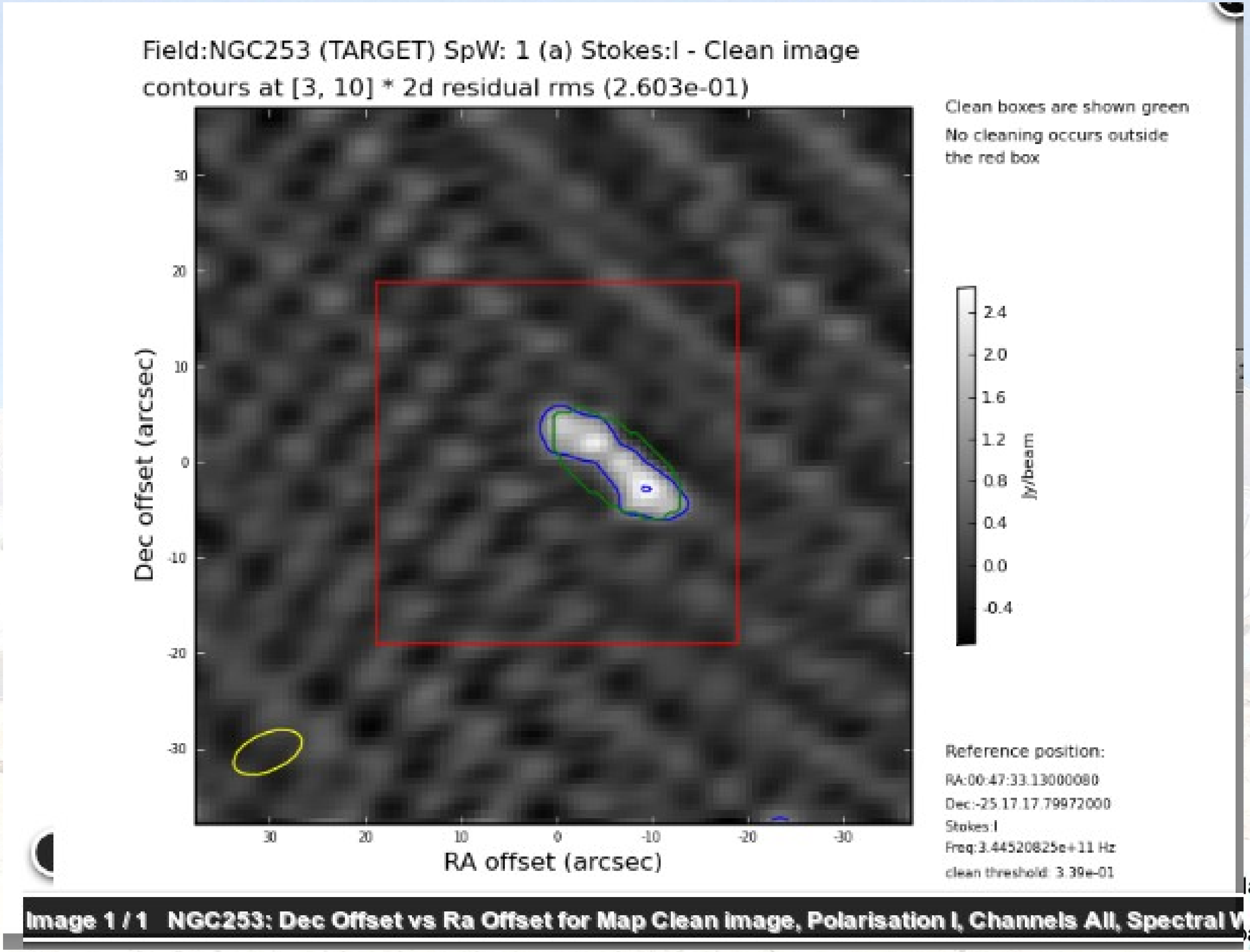
- ALMA Pipeline Wiki
  - http://almasw.hq.eso.org/almasw/bin/view/PIPELINE/WebHome
- ALMA Heuristics Server at the MPIfR (User Test Data, Regression Results, Releases)
  - http://alma-heuristics.mpifr-bonn.mpg.de
- CASA
  - http://casa.nrao.edu
- References

  - ALMA Pipeline Heuristics, Lightfoot, J.; Kosugi, G.; Wyrowski, F.; Zapata, L.; Muders, D.; Boone, F.; Tsutsumi, T.; Davis, L.; Wilson, C.; Shepherd, D.; ADASS XVII, ASP Conference Series , Vol 394, p. 573, 2008
  - ALMA Pipeline Heuristics, Muders, D.; Boone, F.; Wyrowski, F.; Lightfoot, J.; Kosugi, G.; Wilson, C.; Davis, L.; Shepherd, D.; ADASS XV, ASP Conference Series, Vol. 376, p 241, 2007

Max-Planck-Institut für Radioastronomie

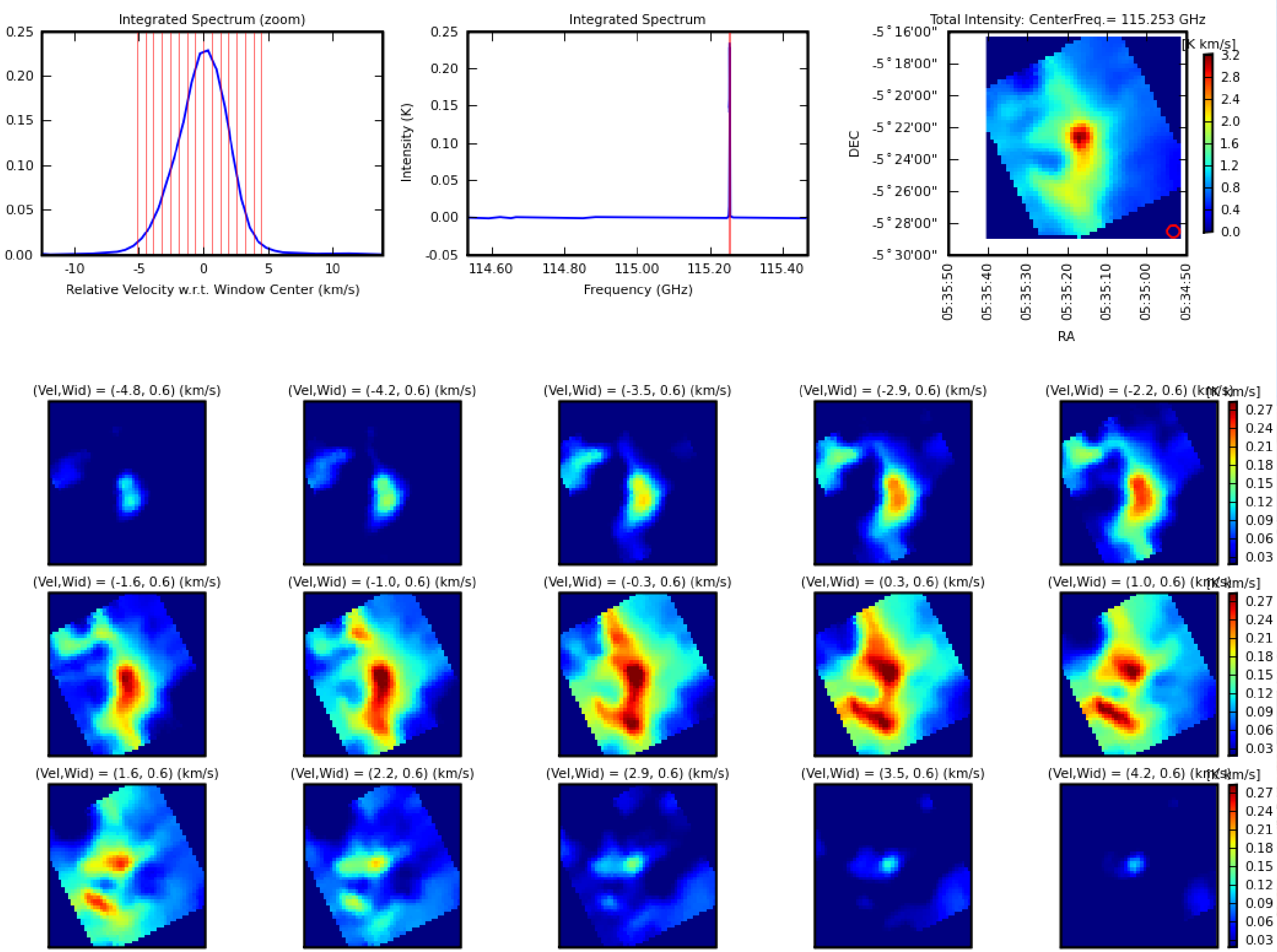
Web Results Demos

ALMA NGC 253 Band 7



lanck-Institut für astronomie

ALMA Orion SD



Max-Planck-Institut für Radioastronomie