The ALMA pipeline

Eric Villard (ALMA JAO)

Extracts from presentation by
Dirk Muders and Lindsey Davis

There are 3 pipelines

- Telescope calibration pipeline
  - online, to provide calibration results
- Quicklook display pipeline
  - online, to provide quick overview to AoD and operators
- Science pipeline
  - offline

The science pipeline

- 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

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

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.

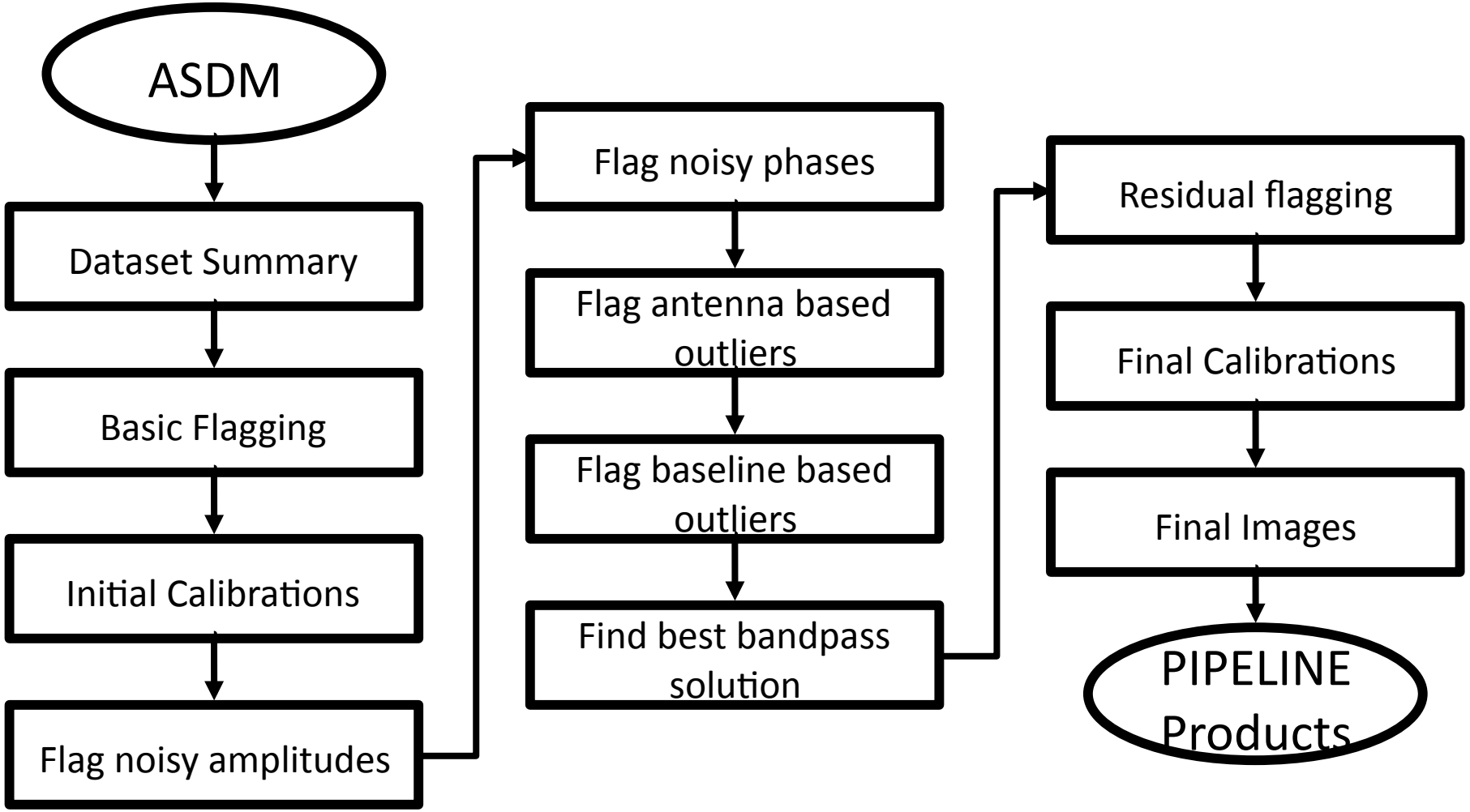
Pipeline heuristics

- 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

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 heuristics workflow



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

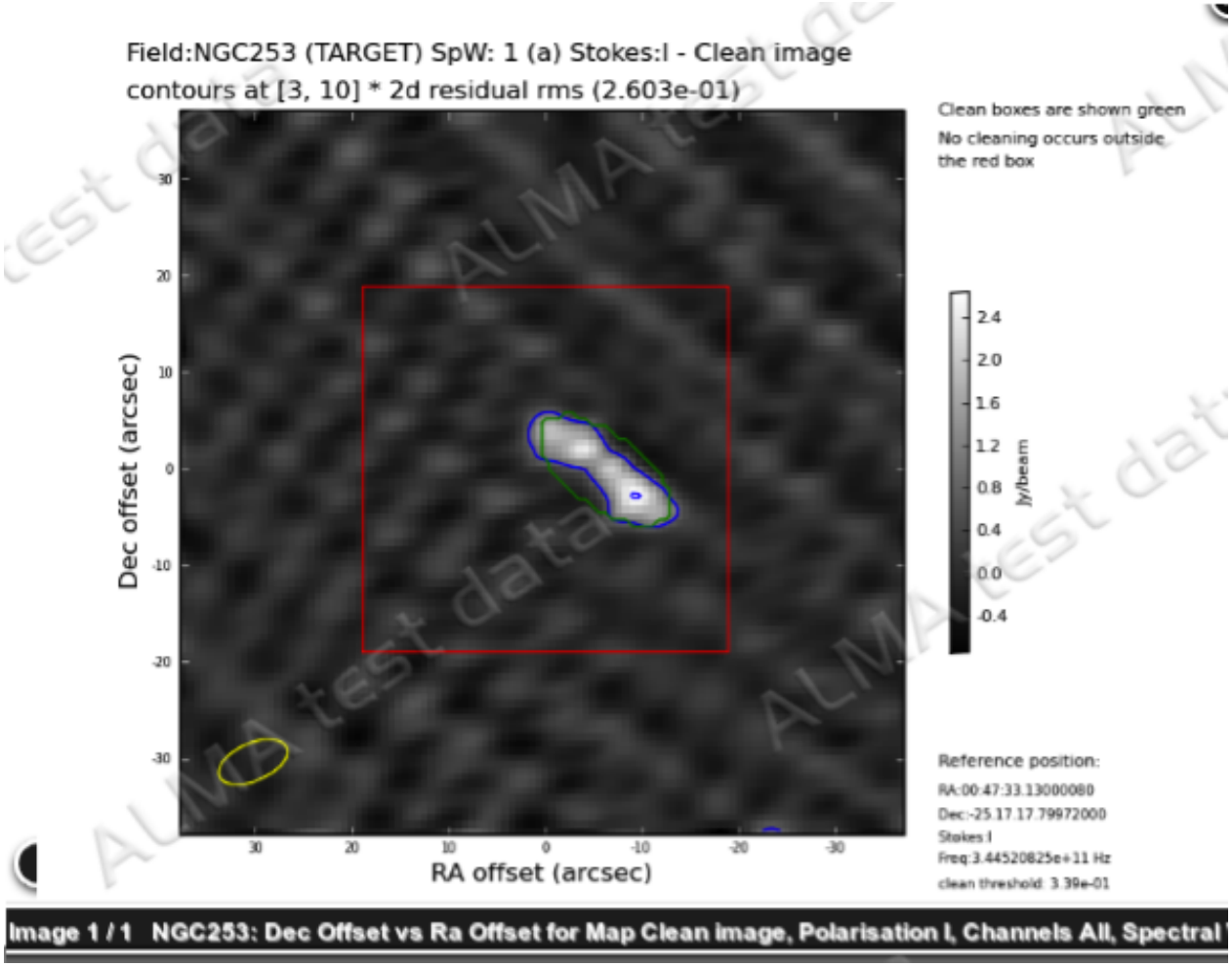
Calibration

- 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

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

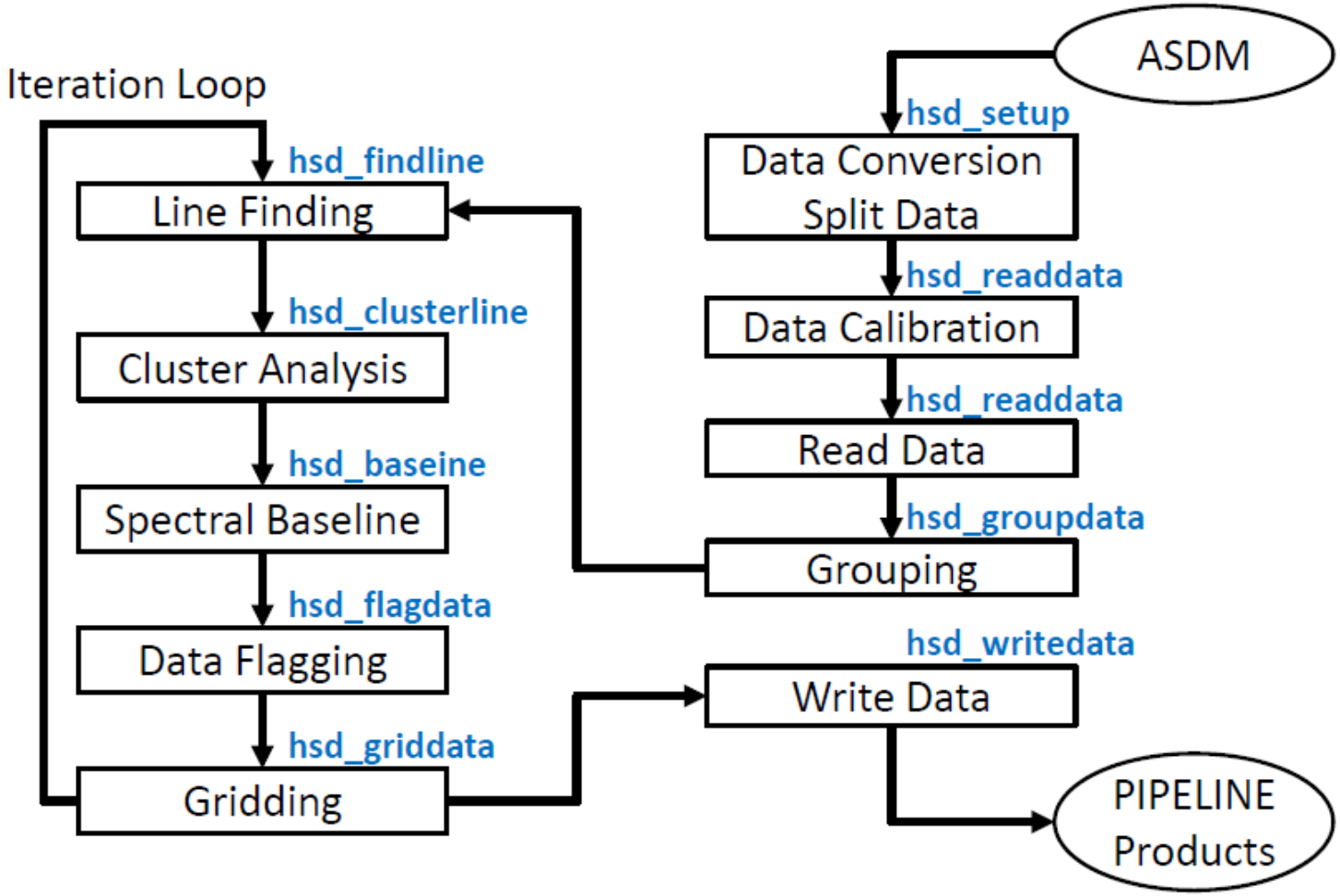
Example: NGC253 at band 7



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

Single dish heuristics workflow



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

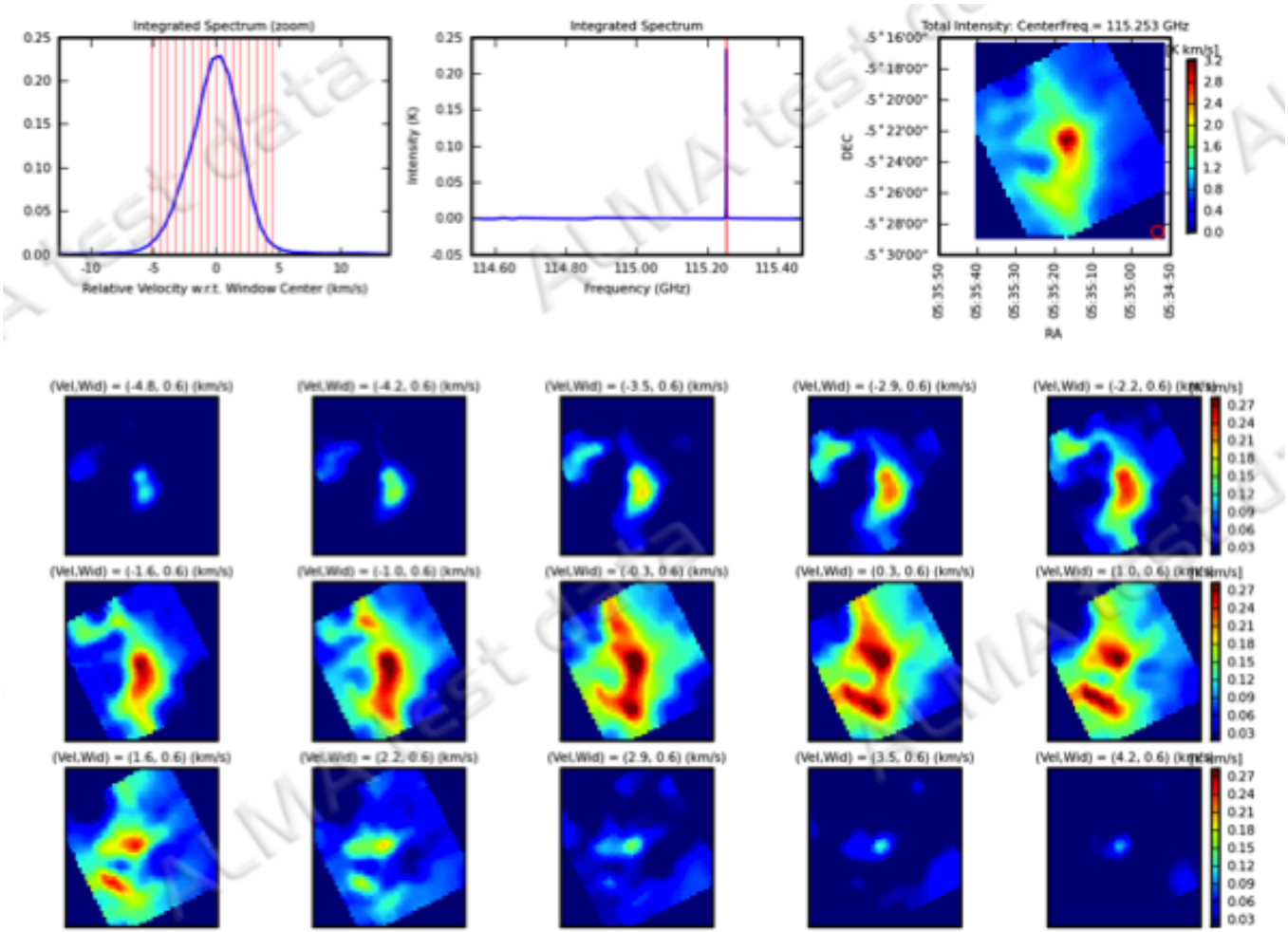
Flagging

- 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

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

Example: Orion



Data products

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

Commissioning of the pipeline

- 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

Current and future developments

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

For more information:

http://almasw.hq.eso.org/almasw/bin/view/PIPELINE/WebHome