GMRT HI Imaging Pipeline

Manisha S Samble

Guided by: Jayaram N Chengalur

Outline

- Introduction

- Packages used in pipeline

- The task 'gautoclean'

- Pipeline flowchart

- Input parameters to pipeline

- Timing

- Results

- Summary

Introduction

- To design a HI Imaging Pipeline for GMRT data.

- Script language : python

- Input : Datafile, Output : Spectral Cube.

- Reduce analysis time, less manual error.

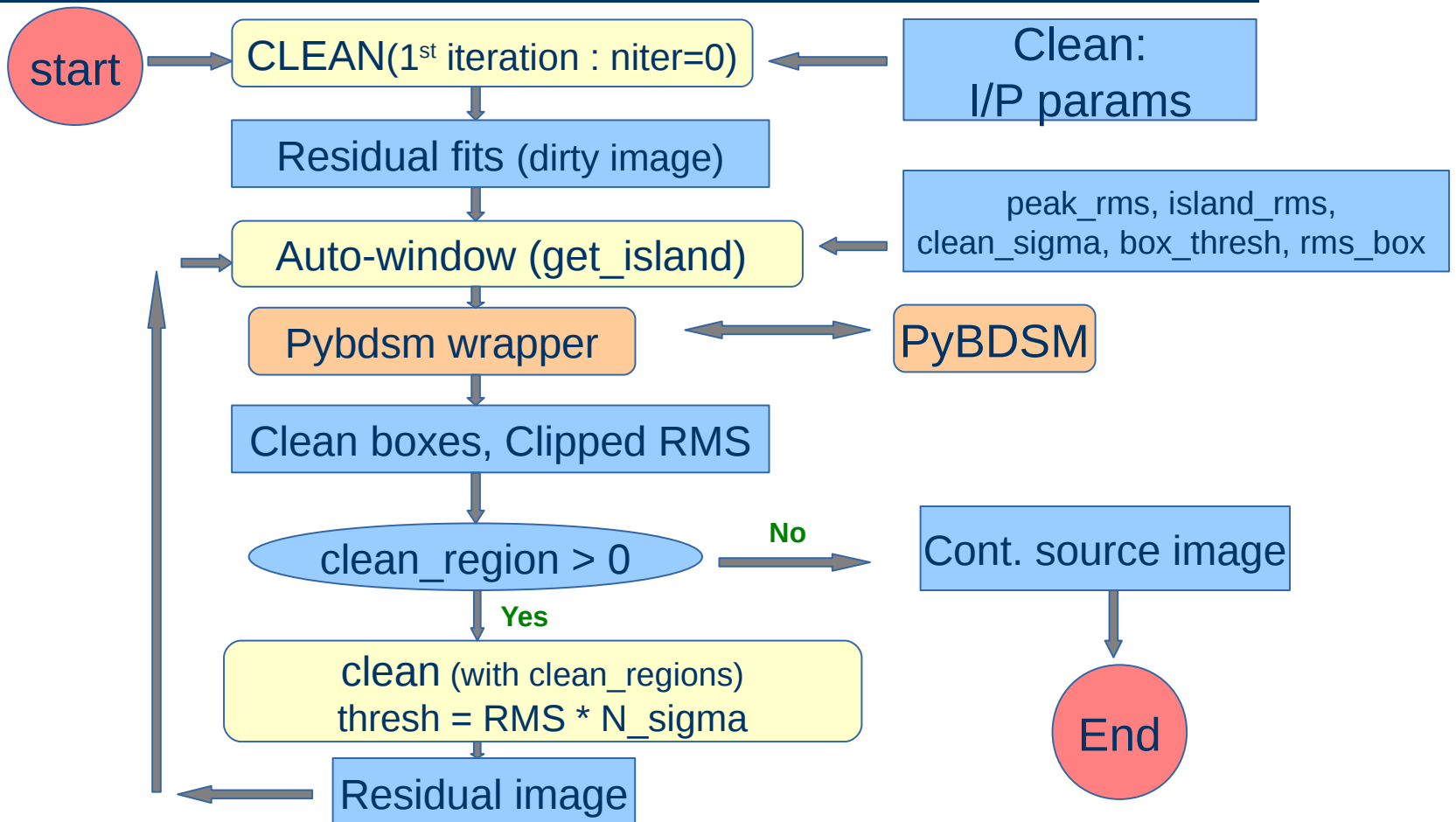- Implementation & preliminary results from the pipeline

Packages used in Pipeline

- flagcal, CASA, PyBDSM, SoFiA

- flagcal : package to flag & calibrate GMRT data.(Chengalur, 2014)

- CASA : Imaging and self calibration (https://casa.nrao.edu/)

- PyBDSM : Identify continuum sources (Mohan, 2009)

  - Sources identified using local rms in image.

  - Output clean region written in CASA CRTF format.

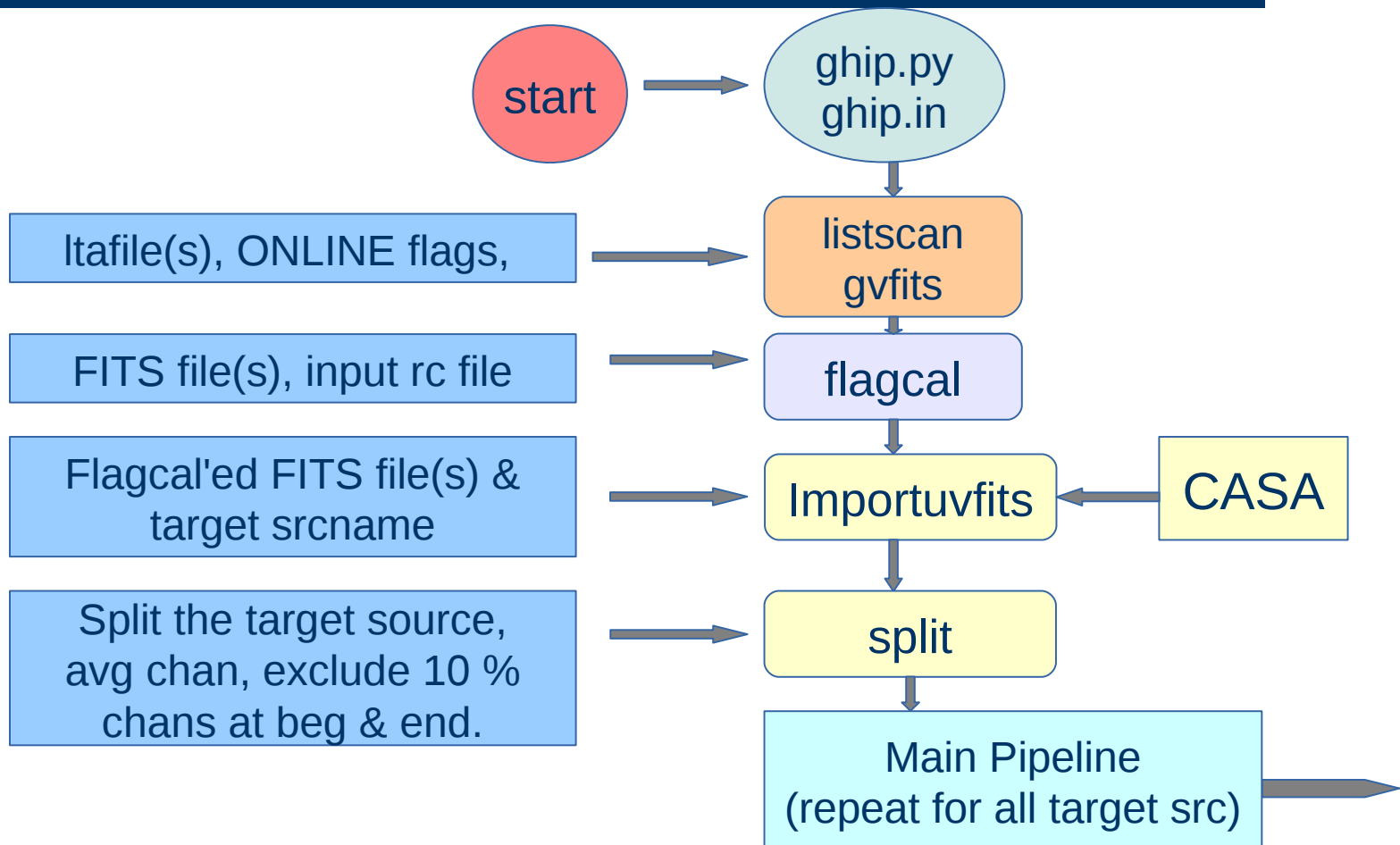- SoFiA : to identify spectral emission (https://github.com/SoFiA-Admin/SoFiA)

The task 'gautoclean'

- Created iterative cleaning task 'gautoclean'

  - based on CASA task 'autoclean'

  - modfied to in-corporate PyBDSM's boxing algorithum

  - Used to identify and clean continuum sources

- Sources detected in the residual image based on

  - User input parameters (peak_rms, island_rms, clean_sigma, boxthresh, rmsbox)
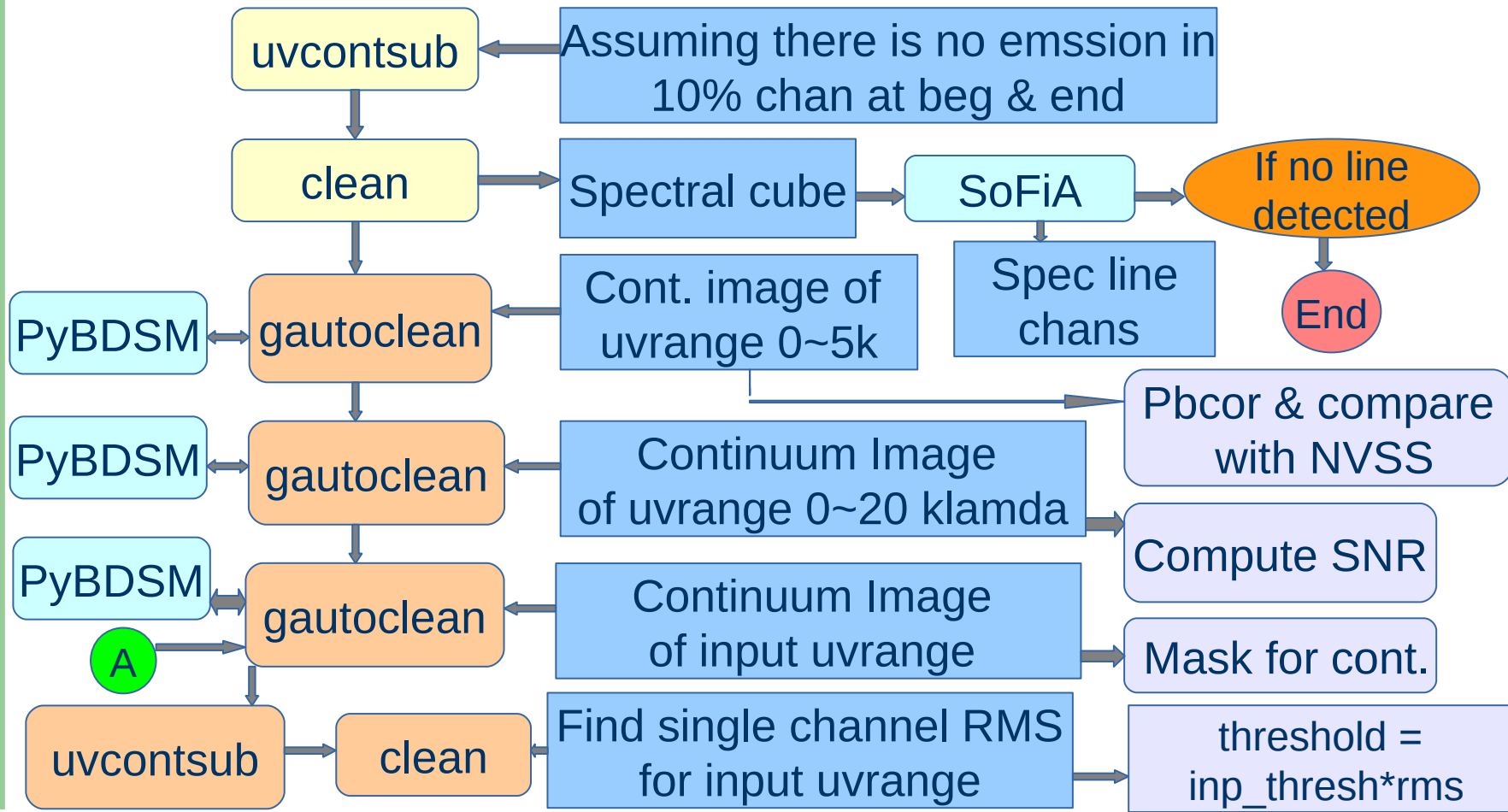
  - local rms in dirty image

Flowchart : 'gautoclean'

```
start  →  CLEAN(1st iteration : niter=0)  ←  Clean:
                      ↓                        I/P params
          Residual fits (dirty image)
                      ↓
     →    Auto-window (get_island)  ←  peak_rms, island_rms,
     ↑                                 clean_sigma, box_thresh, rms_box
     ↑         Pybdsm wrapper  ↔  PyBDSM
     ↑                ↓
     ↑    Clean boxes, Clipped RMS
     ↑                ↓
     ↑       clean_region > 0  — No →  Cont. source image
     ↑                ↓ Yes                      ↓
     ↑    clean (with clean_regions)           End
     ↑    thresh = RMS * N_sigma
     ↑                ↓
     ←──  Residual image
```
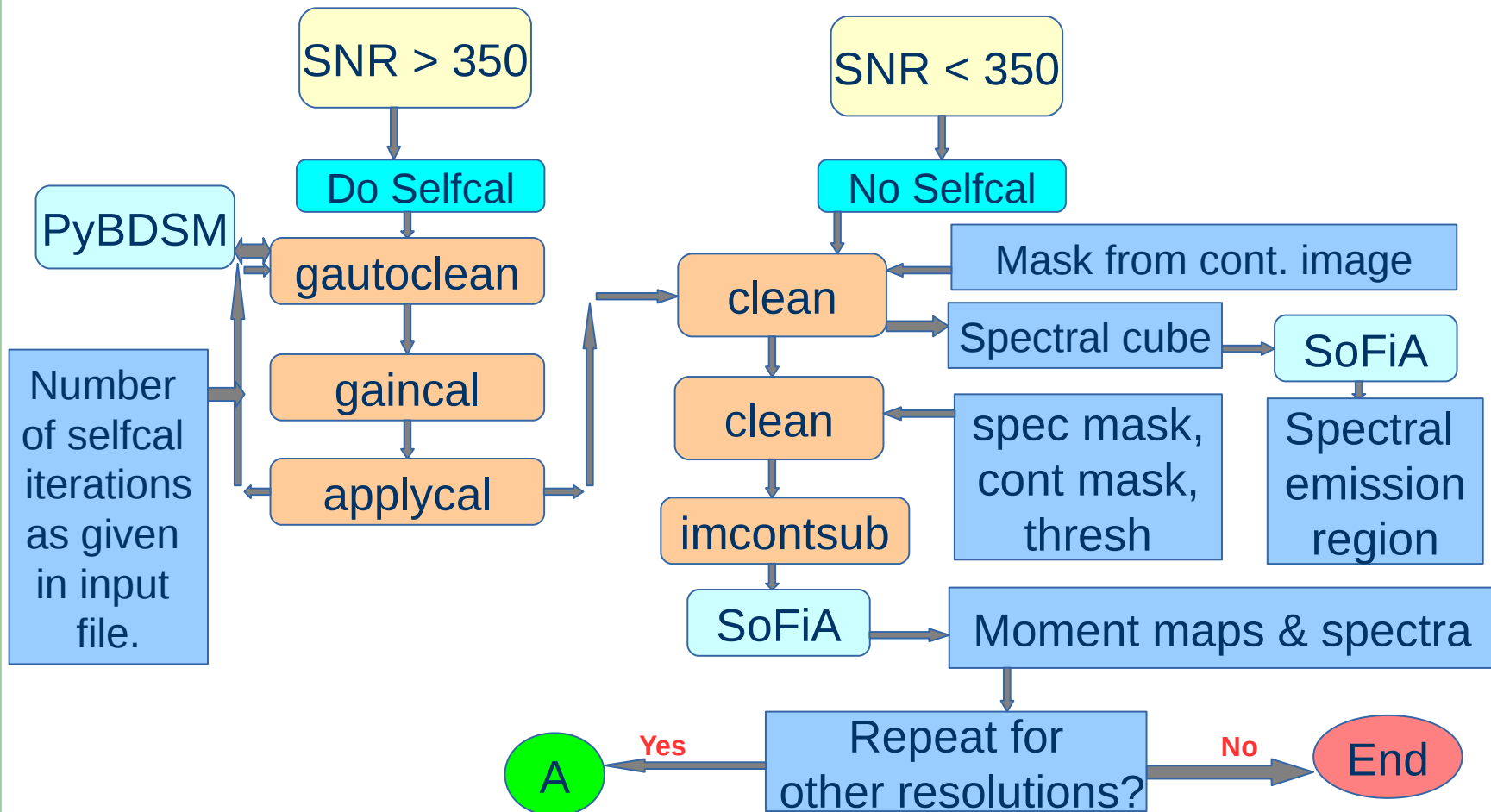
Pipeline flowchart - 1

start → ghip.py ghip.in

ltafile(s), ONLINE flags, → listscan gvfits

FITS file(s), input rc file → flagcal

Flagcal'ed FITS file(s) & target srcname → Importuvfits ← CASA

Split the target source, avg chan, exclude 10 % chans at beg & end. → split

Main Pipeline (repeat for all target src)

Pipeline flowchart -2

```
uvcontsub  ←  Assuming there is no emssion in
                10% chan at beg & end

clean  →  Spectral cube  →  SoFiA  →  If no line detected
                              ↓              ↓
                         Spec line         End
                           chans

gautoclean  ←  Cont. image of
                uvrange 0~5k
PyBDSM
                                    →  Pbcor & compare
                                          with NVSS

gautoclean  ←  Continuum Image
                of uvrange 0~20 klamda  →  Compute SNR
PyBDSM

gautoclean  ←  Continuum Image
                of input uvrange  →  Mask for cont.
PyBDSM
    A

uvcontsub  →  clean  →  Find single channel RMS   →  threshold =
                          for input uvrange            inp_thresh*rms
```

Pipeline flowchart -3

```
                SNR > 350                              SNR < 350

                 Do Selfcal                             No Selfcal

PyBDSM                                                                    Mask from cont. image
                 gautoclean                     clean
                                                                   Spectral cube          SoFiA
Number
of selfcal       gaincal                        clean              spec mask,            Spectral
iterations                                                         cont mask,            emission
as given                                                           thresh                region
in input         applycal                      imcontsub
file.
                                                SoFiA               Moment maps & spectra


                               Yes      Repeat for            No
                         A           other resolutions?            End
```

User Input Parameters

- CHAN_AVG_SPLIT = 1 # Default : 1 (channel average in split)
- CHAN_AVG_CLEAN = 1 # Default : 1 (channel average in CLEAN)
- BCHAN_SPLIT = 0 # Default : 10% of the total_channel
- ECHAN_SPLIT = 0 # Default : total_channel - 10% of the toal_channel.
- MULTISCALE = [0,4,8] # Default : 0 (multiscale parameters in pixels)
- NPIXPERBEAM = 4 # Default : 3 (Number of Pixels per beam)
- UVMAX = 5 # Default : 5 (list of UVMAX in klambda)
- IMSIZE = 20 # Default : 40 (corresponding image size in arcmin)
- CELLSIZE = 10 # Default : 10 (Cellsize in arcsec)
- IMSIZE_PIXEL = 256 # Default : 256 (Imsize in Pixel)
- OUTERTAPER = 4 # Default : 4 (Outer Taper in klambda)
- THRESHOLD_CLEAN = 1.5 # Default : 1.5 (cleaning thresh : thresh_clean*rms)
- NO_SELF_CAL = 0 # Default : 0 (no. of selfcal with phaseonly)
- DO_CUBE = 0 # Default : 0 (1 - make cubes.)

Advanced Input Parameters

- PEAK_RMS   =   7.0   # Default : 7.0 (island should contain at least one pixel >= PEAK_RMS*sigma)
- ISLAND_RMS   =   5.0   # Default : 5.0 (island is marked with all continuous pixels which are above ISLAND_RMS*sigma)
- BOXTHRESH   =   70   # Default : 70 (In percentage. All islands whose total flux is more than boxthresh/100 times maximum flux is included)
- CLEAN_SIGMA   =   3.0   # Default : 3.0 (for clean clean_sigma*clip_rms =thresh)
- RMSBOX   =   0   # Default : 0 (0 = calculated in script, boxsize to calculate RMS at a given pixel in image)
- SMALLSCALEBIAS =   0.6   # Default : 0.6 (if multiscale ON)
- CONT_ROBUST   =   0.0   # Default : 0.0 (between -2 to 2)
- CUBE_ROBUST   =   0.0   # Default : 0.0 (between -2 to 2)
- SOLINT_P   =   5   # Default : 5 (solint for phaseonly gaincal in mins)
- SOLINT_AP   =   8   # Default : 8 (solint for amplitude and phase gaincal in mins)

Timing

- Observation parameters
    - 5.5 hrs on target source, 8 sec integ., 512 spectral channels. (file size ~ 12 GB)
- Machine specs (standard desktop)
    - 6 cores, 16 GB RAM
- Software versions
    - CASA version 4.5
    - flagcal : version – 0.989
- Total analysis time ~ 3.5 hrs (for one resolution)
    - ~ 30m for flagcal
    - Most of the rest of the time is spent in the CASA 'clean' task
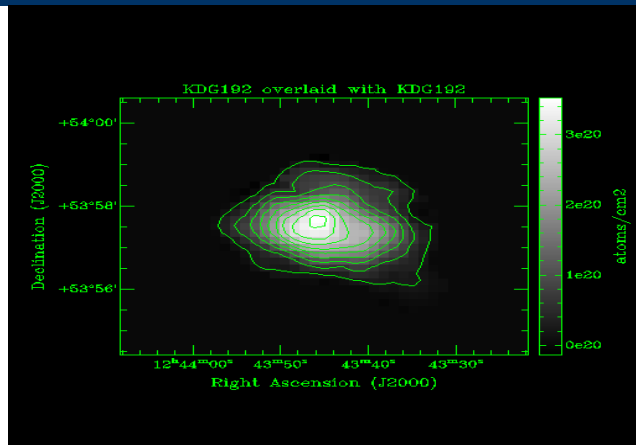- Analysis time increases with image size.

Results : Spectra
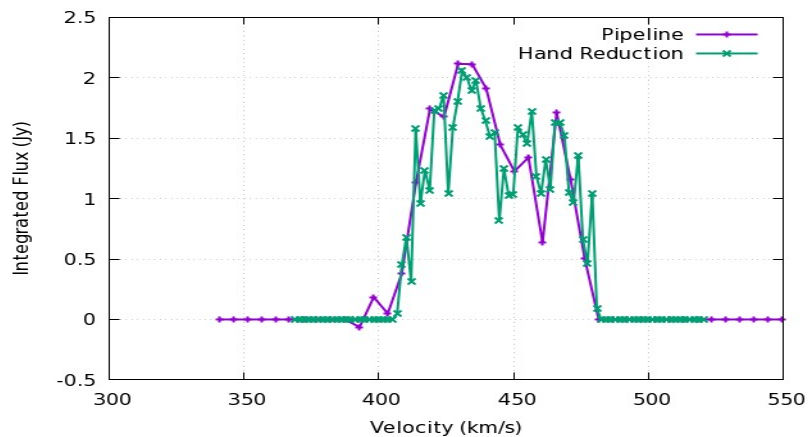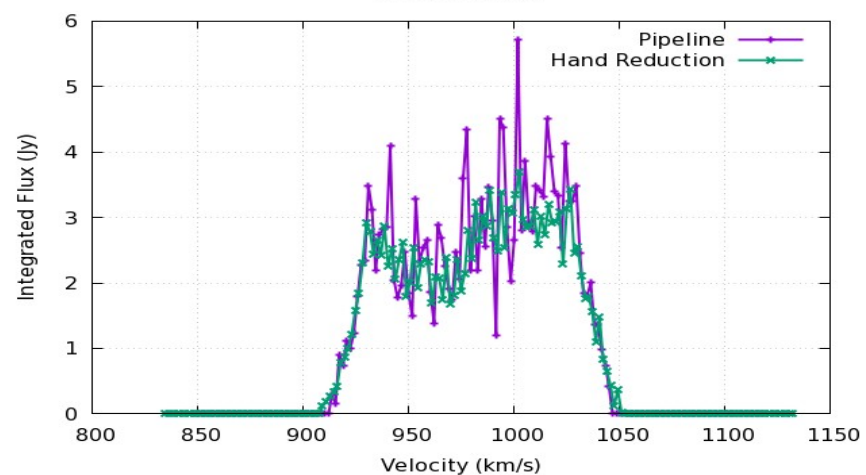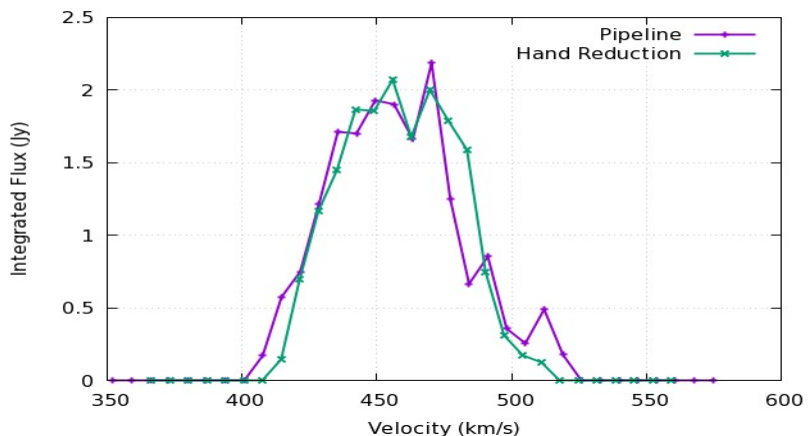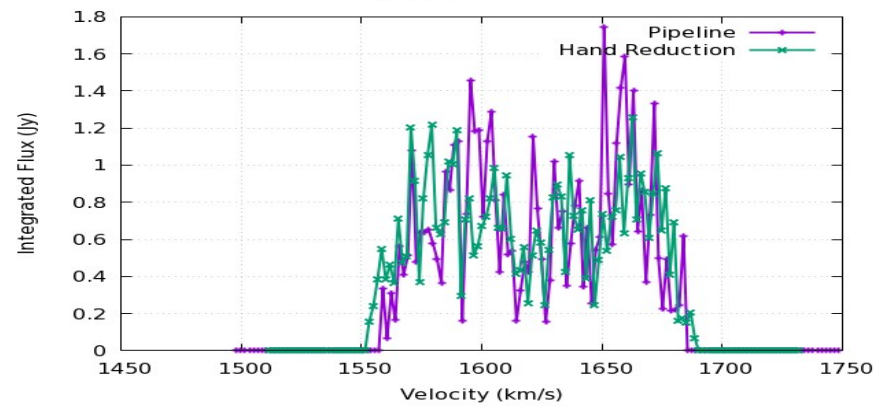
Results : Moment Maps

Source KDG192

Hand Reduction, Beam size ~ 39"
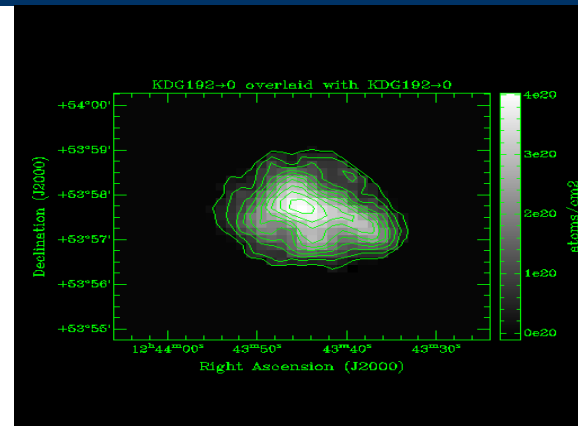
Source : KDG192
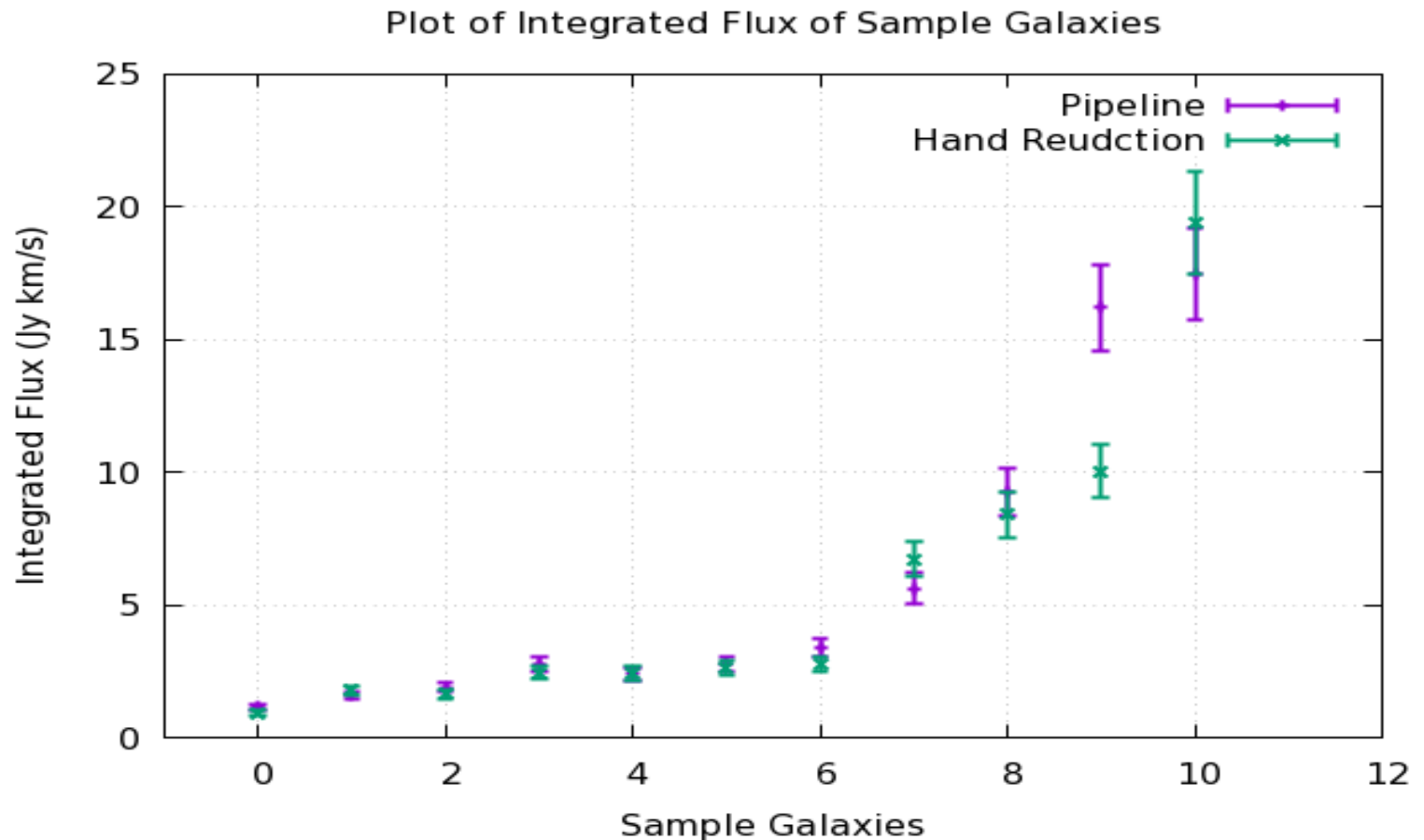
Pipeline Reduction, Beam size ~ 34"
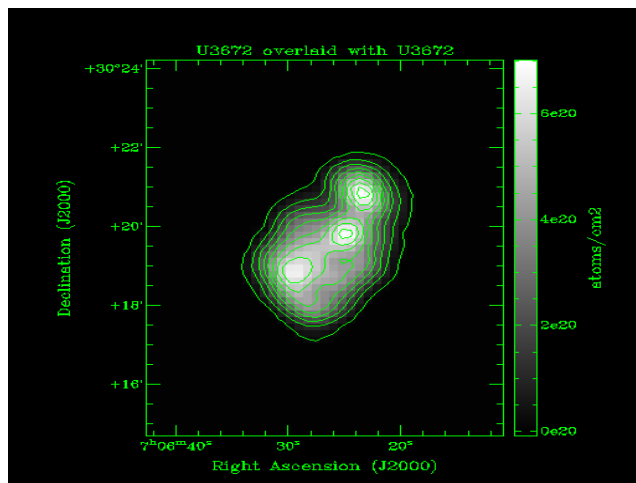
Source : UGC3672

Hand Reduction, Beam size ~ 39 "
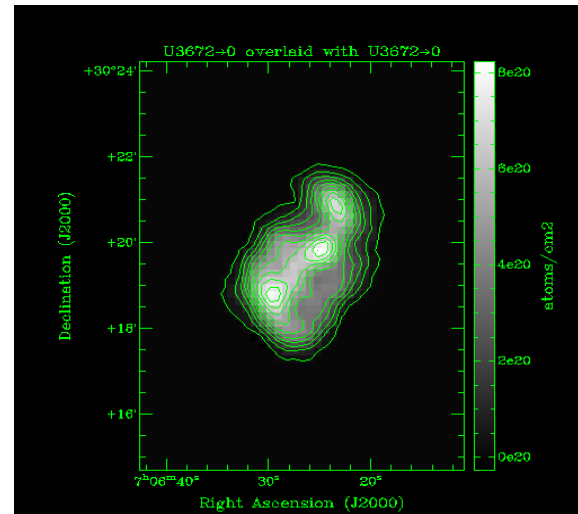
Source : UGC3672

Pipeline Reduction, Beam size ~ 35"

Comparison of integrated Fluxes


Plot of Integrated Flux of Sample Galaxies

Summary

- A pipeline has been developed for analysis of GMRT spectral data

- Based on a number of different packages
  - Flagcal for flagging and calibration
  - CASA for the core imaging
  - PyBDSM for continuum source identification
  - SoFiA for identification and parameterization of spectral emission

- Preliminary comparisons shows that the pipeline results match reasonably well with the results of manual analysis
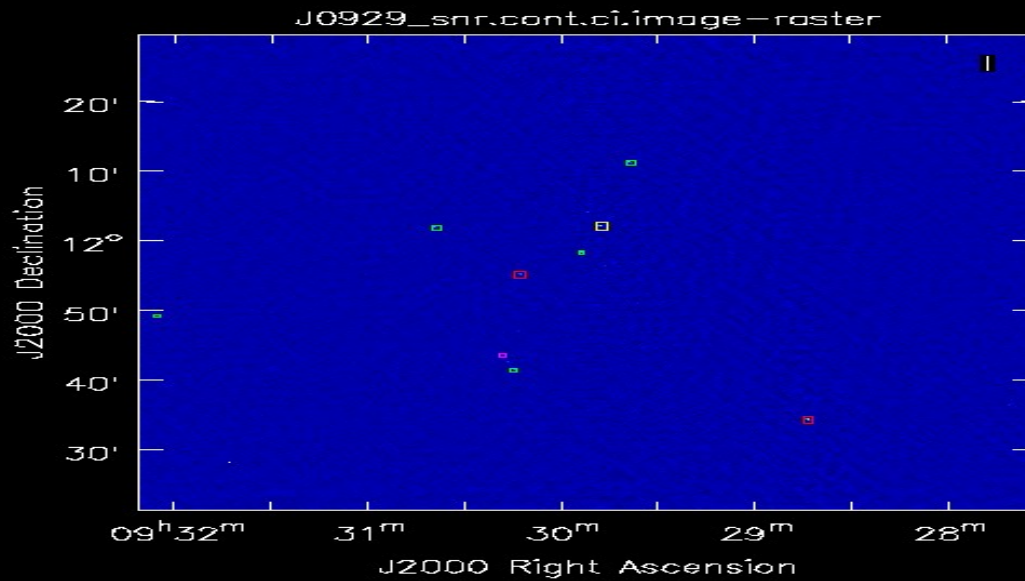
Acknowledgement

- I would like to thank Niruj Mohan, Sanjay Kudale, Narendranath Patra, Peter Kamphuis.

THANK YOU
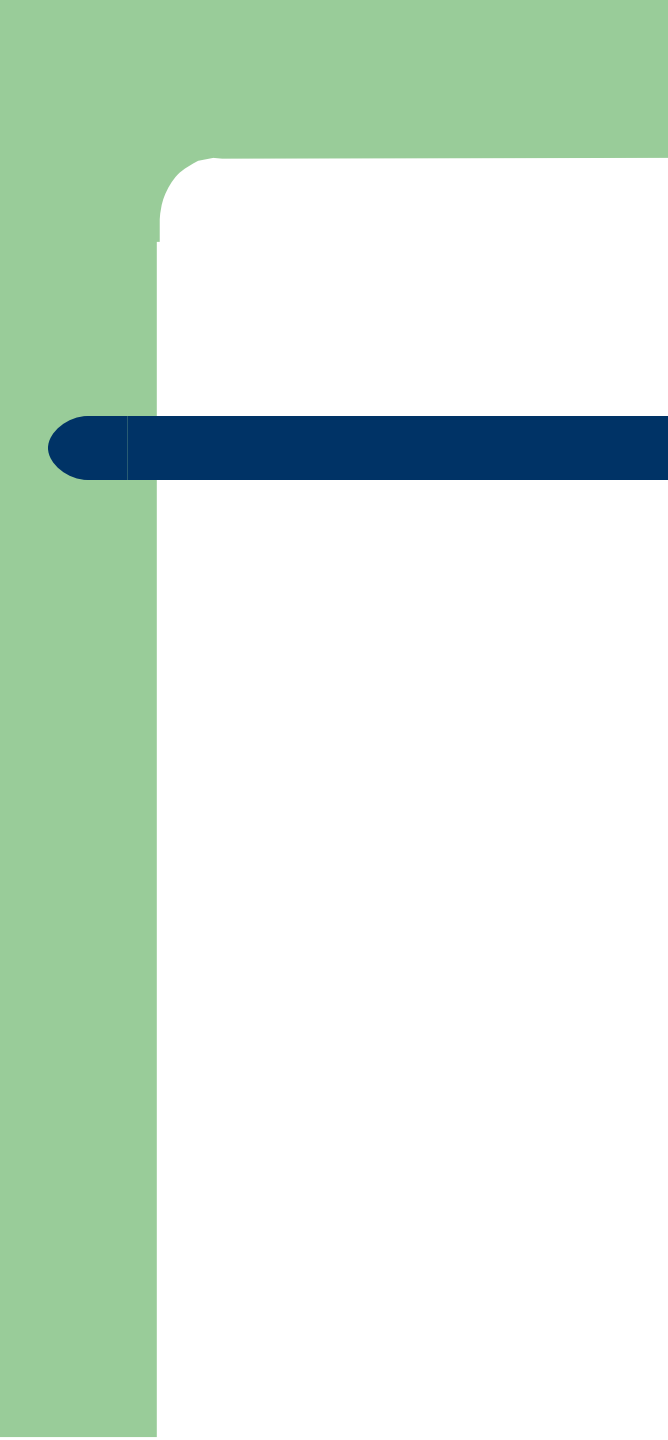
Continuum image & boxes

Comparision with NVSS Image