Workflow for GASKAP imaging at DUG
Based on WSCLEAN version 3.0

Scripts for WSClean 3.2 workflow are in /data/csiro_od207510/casda/scripts
Folder /data/csiro_od207510/casda/38466 has deployed config for imaging with 3.2

Scripts for WSClean 3.0 workflow are in /data/csiro_od207510/casda/scripts-3.0
Folder /data/csiro_od207510/casda/38466-was has deployed config for imaging with 3.0

Please refer to Pingel, N. M., et al., PASA 39, e005. doi:10.1017/pasa.2021.59, arXiv:2111.05339 for a description of the logic and purpose of this workflow.

Please refer to Kemp, I. P., et al., Astronomy and Computing (paper submitted Oct 2024) for information on this workflow which is designed for DUG Technology supercomputer, and how it differs from the workflow described by Pingel et al. which was implemented for Avatar at the Australian National University.

DETAILED STEPS

1 Download
- 1.1 Confirm the scheduling block id for the observation sbid
- 1.2 Create folder casda/<sbid>
- 1.3 Log into casda using OPAL account that has project access
- 1.4 Search on sbid, tick 'unreleased'
- 1.5 Change page length to max (50), select 'all' on every page
- 1.6 Click the download icon, select option from external website
- 1.7 After fetch complete – get the links file
- 1.8 Make <sbid>download.schema from links file
  - 1.8.1 Remove links to averaged data
  - 1.8.2 Remove links to checksum
  - 1.8.3 Add a row number at the start of each row (should be 108 rows)
- 1.9 Transfer the schema file to DUG
- 1.10 Check 01-download.job to ensure schema file name is correct
- 1.11 Submit 01-download.job

2 Bin
- 2.1 update scripts 02-binA.py, 02-binB.py, 02-binC.py with correct folder names for the measurement sets – the naming can vary a bit between observations
- 2.2 Submit 02-bin-all.job
- 2.3 Delete the original un-binned measurement sets to conserve storage (can do this progressively while the script is in progress)

3 Listobs
- 3.1 Update the scripts 03-listobs.sh and 03-listobs.py with the correct folder and measurement set names
- 3.2 Run 03-listobs.sh on the head node

4 Rotate phase centre
- 4.1 Update the script 04-pcr-all-check.job with the correct measurement set names
- 4.2 Run 04-pcr-all-check.job to verify phase centres
- 4.3 Update the script 04-pcr-all.job with the chosen phase centre (usually whatever is reported in the previous step for the A interleaf)
- 4.4 Run 04-pcr-all.job
- 4.5 Run 04-pcr-all-check.job to verify that all beams are now on the same phase centre

5 Continuum subtraction
- 5.1 Tar one example ms (for example Beam 16 interleaf A) and scp it to an environment where casa plotms is available (and untar)
- 5.2 update script 05-plotms.py with the correct path to the example measurement set
- 5.3 run script 05-plotms.py and screen capture the emission spectrum which is plotted
- 5.4 select the channel numbers which defined the 'continuum' – ie the background away from the emission peaks in the chart, for example 0~600,800~1450,1850~2111
- 5.5 Retain the screen shot as it will be used later
- 5.6 Update the scripts 05-ContSub1.py and 05-ContSub2.py with the correct paths, file names and continuum specs
- 5.7 Submit 05-ContSub1.py and 05-ContSub2.py (The two jobs will use both processors on one Dual CLX node)
- 5.8 Delete the non-contsub measurement sets after they have been processed to conserve storage
- 5.9 Tar one ms (corresponding to the one used in 5.1) and scp it to an environment where casa plotms is available
- 5.10 Update 05-plotms.py with the new path, run and capture the spectrum as before – confirm that continuum subtraction has given the expected results

6 Split By Channel
- 6.1 Update scripts 06-SplitByChannel.job and 06-SplitByChannel.py with the updated paths and file names
- 6.2 Run 06-SplitByChannel.job Note that the channelised data this creates is about 3x the volume of the source measurement sets, so ensure storage is available.
- 6.3 Count files in the chan folder using ls -l | wc to ensure that all channels were split – should be 228098 (108x2112+2)

7 Make Clean Mask
- 7.1 Update script 07-wsclean_cleanmask.job with the correct output file name (output =)
- 7.2 Update 07-wsclean_cleanmask.schema with the correct base path and select a channel number near the emission peak in the continuum-subtracted spectrum (obtained in step 5.10)
- 7.3 Submit 07-wsclean_cleanmask.job

7.4 Take one of the beam files from the imaging run – images/<chan>/*beam.fits and move it along with the scripts 07-makemask.py and 07-run-makemask.sh to an environment casa ipython and astropy are available.
7.5 Rename the beam file to beam.fits
7.6 Run casa -c 07-makemask.py
7.7 Move the output file mask.fits back to the base folder for the observation an rename it JD-mask.fits

8 Imaging
8.1 Update the pathname to the project, in the file 08-wsclean_parallel.schema
8.2 Submit the job 08-wsclean_parallel.job [estimated run time 1400node-hours]

9 Not used

10 Collect image files & PB files
10.1 Update the path to the 2 collection scripts, in 10-run-collect.job
10.2 Submit the job 10-run-collect.job
10.3 count files in images/all [use ls -l | wc] to ensure that all channels were imaged – should be 2113 (no. channels +2)

11 Import images & PB files into CASA format
11.1 Ensure script and data paths are correct in script 11-run-import.job
11.2 Submit 11-run-import.job
11.3 delete the fits images in the folders all and all-PB

12 Update metadata in images & PB headers
12.1 Ensure data paths & output file names are correct in script 12-update-headers.job
12.2 Submit 12-update-headers.job

13 Not used

14 Concatenate images into cubes
14.1 Ensure paths and cube file names are correct in script 14-run-concat.job
14.2 Submit 14-run-concat.job

15 Normalise the PB cube
15.1 update path and file name in the scripts 15-normalise1.py, 15-normalise2.py, 15-run-normalise1.job and 15-run-normalise2.job
15.2 submit the job 15-run-normalise1.job
15.3 pull the value of 'maxi' from the logfile and update the value in script 15-normalise2.py (ensure that the value is rounded up, eg round 1.455294 as 1.4553)
15.4 submit the job 15-run-normalise2.job

16 Primary Beam correction
16.1 update the path and file names with the new obs id in scripts 16-pb-correct.py and 16-run-pb-correct.job

17 Get the Parkes cube

17.1 Request the GASS data cube as described in the imaging guide. Note that the coordinates should be space delimited, not colon delimited as given in the imaging guide

17.2 update the file name in the script 17-wget-parkes.sh

17.3 rename the output file from gass****.fits to [sbid]_gass.fits

18 Drop degenerate stokes axes from the cubes

18.1 update the file names and paths in 18-dropdeg.py and 18-run-dropdeg.job

18.2 submit the job 18-run-dropdeg.job

19 update/replace the beam metadata

19.1 update the file names and path name in 20-update-beam-info.py and 20-update-beam-info.job

19.2 submit the job update-beam-info.job

20 Smooth the ASKAP data to put it all on a common beam

20.1 update the file name and path name in 19-smooth.py and 19-run-smooth.job

20.2 submit the job 19-run-smooth.job

21 Feather the pb corrected ASKAP cube and the Parkes cube

21.1 update the path and file names in 21-feather.py and 21-run-feather.job

21.2 submit 21-run-feather.job

Cube image names

	Input	Output
Step 14 – concatenate		
	[sbid]_cube_lsrk.combImage	[sbid]_cube_pb_lsrk.combImage
Step 15 – normalise		
		[sbid]_cube_pb_lsrk_norm.combImage
Step 16 – normalise & clip		
	normalise	
	[sbid]_cube_lsrk_corrected.image	
	Clip	
	[sbid]_askap_lsrk.image	
Step 17 – get Parkes data		
		[sbid]_gass.image
Step 18 – drop degenerate axes		
	[sbid]_askap.image	[sbid]_gass_cube.image
Step 19 – update beam info		
		[sbid]_gass_corrected.image
Step 20 – smooth		
	[sbid]_askap_smooth.image	
Step 21 – feather		
	SB[sbid]_askap_parkes_PBC_JyB.comb	
	SB[sbid]_askap_parkes_PBC_JyB.fits	

Indicative Run Times – See Kemp et al., to appear in Astronomy & Computing, for further explanation

38466 - late 2023									
		Time per job	# jobs	Elapsed time (hr)	Total time (node-hours)				
01	download	2443	108	6.2	73.3	KNL			
02	bin	65314	1	18.1	18.1	KNL			
03	listobs	0	1		0.0				
04	pcr check	32	1		0.0				
04	pcr	451	36	0.7	4.5	CLX			
04	pcr check	32	1	0.0	0.0				
05	cont sub	211980.5	2	59.9	117.8	CLX			
06	split channel	256283	36	38.9	2562.8	KNL			
07	make mask	5473	1		1.5	CLX			
08	imaging	9419	528	97.6	1381.4	CLX | ILX	39.2	minutes per channel	
10	collect	115	1		0.0	KNL			
11	import	5203	1		1.4	KNL			
12	update headers	2137	1		0.6	KNL			
14	concatenate	71104	1	19.8	19.8	CLX			
15	normalise PB	407	1		0.1	CLX			
16	PB correction	732	1		0.2	CLX			
17	get Parkes cube	0	0		0.0	head node			
18	dropdeg	263	1		0.1	CLX			
19	update headers	7	1		0.0	CLX			
20	smooth	35195	1	9.8	9.8	CLX			
21	feather	43185	1	12.0	12.0	CLX			
TOTAL KNL				63.2	2656.3				
TOTAL CLX | ILX				199.7	1547.1				