WALLABY Data Validation Metrics

Version: 1.1

1. Observation

SBID
Observation scheduling block identification number for the science target. It can be used to search for an observational footprint in CASDA.

No. of Antennas
Maximum number of antennas being utilised. This omits the number of antennas being 100% flagged.

Obs Start Date/Time
Start of observing date and time in UT.

Obs End Date/Time
End of observing date and time in UT.

Duration (hr)
Integration time on target in hour.

Field
Observed target field name following A or B to indicate which footprint.

R.A.
Right ascension (J2000) of the observational centre – hr:min:ss

Decl.
Declination (J2000) of the observational centre – deg.arcmin.arcsec

Total Bandwidth
Total observed bandwidth in MHz. ASKAP observation makes up of two bands, i.e. band 1 and band 2.

2. Processed Image Cube

ASKAPsoft version
The last ASKAPsoft version being used for the data reduction.

Cal SBID
Observation scheduling block identification number for the calibrator. It can be used to search for the calibrator data in CASDA.

Frequency Range
Frequency range of the processed data.

Central Frequency
Central frequency of the processed data.

Channel Width
ASKAP channel width in kHz.

Synthesised Beam
Synthesised beam of major and minor axes -- arcsec x arcsec. Values are extracted from the image cube header.

Beam Logs
A traffic light plot is used to evaluate if the measured restoring beams pass or fail a certain threshold or fall onto an acceptable threshold. Given that the flux calculation depends on the beam area, the threshold is determined based on the acceptable level of flux error. This metric considers threshold: ≥ 5% bad (red), ≥ 3-5% OK (yellow) and < 3% good (green). The threshold is calculated using:

$$threshold = \frac{(\theta_{maj} + \sigma_{maj}) \times (\theta_{min} + \sigma_{min})}{\theta_{maj} \times \theta_{min}} - 1.0$$

where θ_{maj} and θ_{min} are beam major and minor axes recorded in the image cube header. σ_{maj} and σ_{min} are standard deviation of restoring beam major and minor axes calculated using:

$$\sigma_{maj} = \sqrt{\frac{\sum_{i=0}^{n}(\theta_{maj,i} - \theta_{maj})^2}{N_{maj}}} \text{ and } \sigma_{min} = \sqrt{\frac{\sum_{i=0}^{n}(\theta_{min,i} - \theta_{min})^2}{N_{min}}}$$

where $\theta_{maj,\,I}$ and $\theta_{min,\,I}$ are restoring beam major and minor axes of each channel recorded in the beam log ascii file. Channels with a zero value are excluded from the summation term. N is the total number of data for the standard deviation calculation, which also excludes the channels with zero values. θ_{maj} and θ_{min} are used instead of mean $\theta_{maj,\,I}$ and $\theta_{min,\,I}$ because we want to compare how much of the measured values deviate from the header values.

Flagged Visibilities
A colour coded plot for visualising the flagged data fractions of 36 beams. A flagging log is provided in an ascii format for each beam, which includes the total flagged percentage, F_{total}, i.e. data + auto-correlation, and total number of *uv* visibilities. The flagged auto-correlation percentage can be calculated using:

$$x_{ij} = 100\% \times \frac{N_{ant} \times N_{rec} \times N_{chan}}{N_{uv}}$$

where N_{ant} is the total number of antennas (i.e. 36 for ASKAP), N_{rec} is the total number of spectra feeds into the synthesis image, N_{chan} is the total number of channels and N_{uv} is the total number of *uv* visibilities. Hence, the flagged data fraction in percentage is $F_{\text{total}} – x_{ij}$.

Flagged Antennas
A colour coded plot for visualising the number of antennas being 100% flagged for each of the 36 beams. This is useful for data processing that adopts beam-by-beam antenna flagging strategy. An ascii file is also provided to indicate which antennas have been 100% flagged for the corresponding beam.

Expected RMS
A colour coded plot for visualising the expected RMS noise of 36 beams. First, to get the theoretical RMS of 36 beams, a natural weighting for the visibilities data is assumed and the fraction of flagged data is omitted. The theoretical RMS given by:

$$RMS_{theo}\,(Jy) = \frac{SEFD}{\eta_c\sqrt{n_{pol} \times N_{ant}(N_{ant}-1) \times t_{obs} \times \Delta\nu}}$$

where SEFD is the System Equivalent Flux Density, η_c = 1.0 is the correlator efficiency, n_{pol} is the number of polarisations, n_{pol} =2 for images in Stokes I, Q, U or V, N_{ant} is the total number of antennas, t_{obs} is the integration time and $\Delta\nu$ is the channel width. The SEFD is a function of frequency (1150-1450 MHz) and its value per channel is calculated within the ASKAPSoft pipeline. The median SEFD of Stoke I is given by:

$$SEFD_{med} \approx \frac{1}{2} \times \sqrt{(SEFD_{med,XX})^2 + (SEFD_{med,YY})^2}$$

where $SEFD_{med,XX}$ and $SEFD_{med,YY}$ are the median SEFD of polarisations XX and YY. Refer Hotan et al. (2021) for a detail description of the ASKAP system.

The expected RMS noise takes into account the flagged percentage, which is given by:

$$RMS_{exp}(Jy) = \frac{1}{\sqrt{1-f}} \times RMS_{theo}$$

where f is the non-flagged fraction and RMS_{theo} is the theoretical RMS.

Ratio RMS (actual/expected)

A colour coded plot for visualising the ratio of RMS noise (actual/expected) of 36 beams. The actual RMS is the average RMS of all channels per beam.

3. Beam Statistics

Two plots are provided for the beam image cube, continuum subtracted beam cube and residual beam cube. One plot is showing the minimum, maximum and 1 percentile flux levels of 36 beams of all processed channels in frequency. The other plot is showing standard deviation (Stdev) and median absolute deviation of median flux (MADMF) noise levels of 36 beams of all processed channels in frequency. All in unit of mJy/beam. These plots are generated by the ASKAPSoft pipeline, and these statistical values are recorded in an ascii file. Further metrics are also provided below for the continuum subtracted beam cube.

MAD Max Flux Density

A colour coded plot for visualising the median absolute deviation (MAD) of max flux density of 36 beams in mJy/beam. This metric is sensitive to broadband artefacts, such as the solar interference. Values are calculated based on the max flux density recorded in the ascii file. The maximum scale of 3.0 is a conservative threshold based on the WALLABY early science M83 field data.

One Percentile Noise Rank

A plot showing 1-percentile noise rank histograms for each beam. The red line represents a Gaussian fit to the histogram, which outliers have been excluded from the fitting. The bin width of the histogram is determined via the Freedman-Diaconis rule.

A traffic light plot is used to evaluate if the 1-percentile noise rank distribution of each beam is good, ok or bad. The variance of the Gaussian fit is used as a metric. This metric considers variance: $\leq$ 0.2 good (green), 0.2—0.25 OK (yellow) and $\geq$ 0.25 bad (red).

Restoring Beam Stdev

Plots show standard deviation values of major and minor axes of each restoring beam in arcsec. Channels with a zero value have been excluded from the standard deviation calculation.

Restoring Beam Area Ratio

Plots show maximum and minimum ratios of restoring beam area. The ratio is given by:

$$ratio = \frac{\theta_{maj,i} \times \theta_{min,i}}{\theta_{maj} \times \theta_{min}}$$

where $\theta_{maj,\,I}$ and $\theta_{min,\,I}$ are restoring beam major and minor axes of each channel recorded in the beam log ascii file, θ_{maj} and θ_{min} are beam major and minor axes recorded in the image cube header.

4. Mosaic Statistics

A plot is provided with minimum, maximum, 1 percentile flux levels, standard deviation and scaled median absolute deviation flux median (MADFD) noise levels of all process channels in frequency for each of the LINMOS mosaicked image cube, continuum subtracted cube and residual cube. All in the unit of mJy/beam. These plots are generated by the ASKAPSoft pipeline, and the statistical values are recorded in an ascii file. Further metrics are also provided below for the mosaicked continuum subtracted cube.

No. of Bad Channel

Checking for bad data channels in the mosaicked continuum subtracted cube. To statistically define a bad channel, we use the following definition:

$$|\sigma_{rms} - \mu| > C \times \tau \times \varepsilon$$

where σ_{rms}, μ and ε are RMS noise level, median of RMS noise level and median absolute deviation of RMS noise level, respectively. Assuming the underlying scatter in RMS values follows a Gaussian distribution, a constant $C \approx 1.4826$ is used to convert the median absolute deviation into a regular standard deviation. τ is set to 1.2 as a threshold criterion, which is consistent with the SoFiA flagging criterion (Serra et al. 2015; Westmeier et al. 2021). A total number of bad channels is provided in the coloured box and a list of bad channels is recorded in an ascii file. The list can be used in SoFiA as part of their pre-conditioning filter.

Missing Data

Checking any missing data channels in the mosaicked continuum subtracted cube. This helps to check for the correlator drop out and/or insufficient RFI flagging. Each correlator block is 4MHz, which corresponds to ~216 channels. Doppler correction can affect the channel number for the correlator block, as such a more conservative threshold is adopted. A traffic light coloured box is used to indicate if there is no issue (green), missing channel number < 100 (yellow) and missing channel number ≥100 (red).

5. HIPASS Sources within Field of View

A list of HIPASS sources within the 6x6 sq degree field of view is provided. The search uses HICAT (Meyer et al. 2004), HIPASS BGC (Koribalski et al. 2004) and NHICAT (Wong et al. 2006) via vizier server. Rotated footprint cannot be taken into account.