

# Using ALMA archival data - A Primer



[www.almascience.org](http://www.almascience.org)

## User Support:

For further information or to comment on this document, please contact your regional Helpdesk through the ALMA User Portal at [www.almascience.org](http://www.almascience.org) . Helpdesk tickets will be directed to the appropriate ALMA Regional Center at ESO, NAOJ or NRAO.

## Revision History:

---

Version	Date	Editors
1.0	March 2022	Kazi L. J. Rygl & George J. Bendo

---

## Contributors

George J. Bendo, Sandra Burkutean, Yanett Contreras, Elisabetta Liuzzo, Marcella Massardi, Adele Plunkett, Kazi L. J. Rygl, MCarmen Toribio, Andrew Lipnicky, Luke Maud, Emily Moravec, Rosita Paladino, Yoshito Shimajiri, Felix Stoehr, Alvaro Hacar, Ciro Pappalardo, Anita Richards, Alvaro Sanchez-Monge

In publications, please refer to this document as:

**Rygl K. L. J. & Bendo G. J., 2022, Using ALMA archival data - A Primer, ALMA Doc. 9.22, Version 1.0**

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	Purpose	9
1.2	The ALMA Interferometer	9
1.3	Science with ALMA	10
1.4	The ALMA Science Archive	10
<b>2</b>	<b>Querying the ALMA Science Archive</b>	<b>11</b>
2.1	ALMA data structure	11
2.2	The ASA query and results interface	11
2.3	Querying the ASA	13
2.3.1	The search box	13
2.3.2	Searching using the image panel	13
2.3.3	Refining search results in the results table	15
2.3.4	Preview of archived ALMA observations within the ASA interface	15
2.4	Example: star-forming regions on the spiral arms in M83	15
2.5	Producing query scripts	15
2.5.1	PyVO	18
2.5.2	Astroquery	19
2.5.3	ALMA Science Archive Notebooks	20
2.6	Results tables	21
2.6.1	Release dates, proprietary periods and data accessibility	21
2.6.2	ALMA quality assurance	21
2.6.3	Exporting the result tables	22
2.7	Additional considerations when using archival data	22
<b>3</b>	<b>Downloading data from the ASA</b>	<b>25</b>
3.1	The Request Handler	25
3.1.1	Data packages available and their content	25
3.1.2	Download selection in the Request Handler	28
3.2	Download options	28
3.3	External download options	29
3.3.1	PyVO	29
3.3.2	Astroquery	29
3.3.3	ALMA Science Archive Notebooks	30
3.4	Download completion checks	30
<b>4</b>	<b>Overview of downloaded archival data</b>	<b>31</b>
4.1	Unpacking archival data	31
4.2	Archival data structure	31
4.3	Archival product files	32

<b>5</b>	<b>Inspecting archival data</b>	<b>35</b>
5.1	QA reports for archival data	36
5.1.1	QA0 report	38
5.1.2	QA2 report	38
5.2	Inspecting manually calibrated archival data	40
5.2.1	Polarisation data	47
5.2.2	Total power data	48
5.3	Inspecting pipeline calibrated data	48
5.3.1	Polarisation data	55
5.3.2	Total power data	56
5.4	Inspecting calibrated data using CASA <code>plotms</code>	58
<b>6</b>	<b>Checking the quality of the archived images and image cubes</b>	<b>61</b>
6.1	Images available in the ASA	61
6.2	Tools for inspecting ALMA imaging products	61
6.3	Properties of ALMA imaging products	62
6.3.1	Spatial and spectral coverage	62
6.3.2	Sensitivity	62
6.3.3	Beam sizes and <i>uv</i> -coverage	63
6.3.4	Residual features and artefacts	66
<b>7</b>	<b>Requesting and recreating calibrated visibilities for archival data</b>	<b>69</b>
7.1	Requesting calibrated visibilities for archival data	69
7.2	Recreating calibrated visibilities from/for archival data	69
7.2.1	Initial inspection of the data quality and calibration procedure	69
7.2.2	Applying the ALMA calibration tables	70
7.2.3	Files created by the calibration script	71
7.2.4	Possible issues and caveats	72
7.2.5	Modifying the calibration script	72
7.3	Solar System flux calibrators	72
7.4	Data averaging	73
<b>8</b>	<b>Re-imaging archival data</b>	<b>75</b>
8.1	The <code>datacolumn</code> parameter	75
8.2	Spectral line images and continuum subtraction	76
8.3	Continuum imaging	78
8.4	Weighting and angular resolution	78
8.5	Other commonly-used parameters	79
8.6	Self calibration	80
<b>9</b>	<b>Data combination</b>	<b>83</b>
9.1	Preparation steps before array combination	83
9.2	Combining different 12-m Array configurations	84
9.3	Combining 7-m and 12-m Arrays	84
9.4	Combining 7-m/12-m and TP Arrays	85
9.4.1	Feathering	85
9.4.2	Joint deconvolution	85
9.4.3	Joint reconstruction ( <code>sdintimaging</code> )	85
9.4.4	Model-assisted cleaning	85
9.5	Combining ALMA data with other single dish or interferometer data	86
<b>10</b>	<b>Using the calibrators as science targets</b>	<b>87</b>
10.1	ALMA calibrator targets	87
10.2	Identifying flagged spectral features in calibrators	87

<b>11 ALMA Calibrator Source Catalogue</b>	<b>89</b>
11.1 ALMA calibrator (grid) sources . . . . .	89
11.2 Checking the flux calibration of archival data . . . . .	89
11.2.1 Obtaining calibrator flux densities . . . . .	90
11.2.2 Comparing the flux densities of the archival data to the catalogued values . . . . .	92
<b>12 Useful additional CASA packages</b>	<b>93</b>
12.1 Analysis Utilities . . . . .	93
12.2 CASA Toolkit . . . . .	93
<b>13 Frequently encountered issues</b>	<b>95</b>
13.1 Errors when running the download*.sh script . . . . .	95
13.2 Finding the CASA version used for the data reduction . . . . .	95
13.3 Problems with running *.scriptForPI.py . . . . .	96
13.4 Issues related to the CASA version used for the data reduction . . . . .	97
13.5 Unit conversion . . . . .	98
13.6 Reference frames . . . . .	98
13.7 Leap second table TAI_UTC is out-of-date . . . . .	99
13.8 Problems with delivered ALMA data and QA3 . . . . .	99
<b>14 Assistance with Archive mining and science exploitation</b>	<b>101</b>
<b>A Acronym List</b>	<b>103</b>
<b>B Further reading</b>	<b>105</b>



# List of Figures

2.1	The ASA web interface. The top panel shows the initial view. The bottom panel shows the expanded search box, which is marked by a rectangle with a magnifying glass at the top left of the page. . . . .	12
2.2	The image panel of the ASA interface zoomed towards the Sgr A* region. The top panel shows the footprints of ALMA observations in the region on the sky. The bottom panel shows the footprint of the Sgr B2 star-forming region that was selected for download. . . . .	14
2.3	The top panel shows the preview window that appears when hovering over the postcard icon. The bottom panel shows the interactive preview of a single spectral window after having clicked on the map/spectrum of that spectral window. . . . .	16
2.4	An example of using the search box to query the archive for observations at the position of M 83 within the 230-232 GHz frequency range. . . . .	17
2.5	The results listed in the Observations table after performing a query for observations at the position of M 83 within the 230-232 GHz frequency range. Each observation is shown as a yellow footprint on the sky map and as yellow bars in the spectral coverage panel. Hovering over an observation makes the corresponding row in the search results table, the corresponding footprint and the corresponding frequency coverage all appear bright yellow. . . . .	17
2.6	Another example view of the search results for M 83 observations in the 230-232 GHz frequency range that illustrates how the results can be refined by selecting only observations from project 2015.1.00121.S. When an observation is selected, the corresponding row in the search results table, footprint and frequency coverage are all coloured orange. Hovering over the postcard icon will show the interactive preview for that observation. . . . .	18
2.7	An example of the Jupyter Notebooks available for illustrating how to use pyVO and astroquery to search the ASA. . . . .	20
2.8	The ASA results table. Hovering over the publication column displays the publications related to the M 83 project 2015.1.00121.S. . . . .	21
3.1	The top panel shows the default view of the Request Handler page for project 2013.1.01215.S. The bottom panel shows the expansion of the product tarball of the first Member OUS. . . . .	26
3.2	An example of the CARTA image viewer displaying the CS(5–4) image cube of the envelope of AGB star IRC+10216. . . . .	27
5.1	Examples of good and bad bandpass calibration coloured by correlation (XX in purple, YY in orange). In the left panel, the amplitudes vary smoothly versus channel (and versus frequency) for all antennas and for both correlations. In the right panel, the YY data for a few baselines jump in amplitude for channels 480-909 and should be flagged. (These plots were produced from manually calibrated ALMA data and are not from the ASA.) . . . . .	36
5.2	Examples of good and bad phase calibration shown for two separate baselines used during the same EB. The XX correlation is shown in purple and the YY correlation is shown in orange. In the left panel, the phases for the CM03 & CM06 baseline show little scatter, which is normal. In the right panel, the phases for the CM03 & CM07 baseline exhibit a large amount of scatter in the first three scans, which is going to cause problems when calibrating the data, and these noisy data should be flagged. (These plots were produced from manually calibrated ALMA data and are not from the ASA.) . . . . .	37

5.3	Examples of good and bad phase calibration shown for the same baseline but measured during two different EBs in the same program. The XX correlation is shown in purple and the YY correlation is shown in orange. In the left panel, the phases vary smoothly from scan to scan, which is normal. In the right panel, the phases jump in the fifth phase calibrator scan and should be flagged. (These plots were produced from manually calibrated ALMA data and are not from the ASA.) . . . . .	37
5.4	An example from SB Z_CMa_a_06_TM2 (2018.1.01131.S) of the system temperature plotted per antenna in the QA0 report, which can be used to assess the antenna performance. . . . .	38
5.5	An example of the phase RMS behaviour for various baseline lengths, showing calibrator J0654-1053 in Band 6 from SB Z_CMa_a_06_TM2 (2018.1.01131.S) in the top panel and SB Z_CMa_a_06_TM1 (2016.1.00110.S) in the bottom panel. . . . .	39
5.6	An example of the RMS and beam size, EB summary and pipeline execution sections from the QA2 report of SB Z_CMa_a_06_TM2 (2018.1.01131.S). . . . .	40
5.7	Part 1 of the manual calibration QA2 plots from SB Ephemer_i_c_06_TE (2015.1.01084.S). . . . .	41
5.8	Part 2 of the manual calibration QA2 plots from SB Ephemer_i_c_06_TE (2015.1.01084.S). . . . .	43
5.9	Part 3 of the manual calibration QA2 plots from SB Ephemer_i_c_06_TE (2015.1.01084.S). . . . .	45
5.10	An example of the QA2 text file created during the manual calibration of the SB Ephemer_i_c_06_TE (2015.1.01084.S). This example shows the observation summary, source coordinates and spectral window information. . . . .	46
5.11	Examples of two polarisation calibration plots for SB BRI_0952_a_07_TM1 (2018.1.01536.S). The left panel shows the gain ratio RMS for the polarisation calibrator before and after applying the polarisation calibration, and the right panel shows the cross-phase solutions per channel. . . . .	47
5.12	An example of a SD sampling plot from SB M83_b_03_TP (2013.1.01312.S) showing the observed footprints on the sky including the off-target positions used for measuring the sky signal for calibration. . . . .	49
5.13	An example of a WebLog Home page for SB Z_CMa_a_06_TM1 (2018.1.01131.S). . . . .	49
5.14	An example of a WebLog measurement set overview webpage for SB Z_CMa_a_06_TM1 (2018.1.01131.S). . . . .	50
5.15	An example of a WebLog “By Topic” page for SB Z_CMa_a_06_TM1 (2018.1.01131.S). . . . .	51
5.16	An example of a WebLog “By Task” page for SB Z_CMa_a_06_TM1 (2018.1.01131.S). . . . .	52
5.17	An example of a WebLog <code>hifa_tsysflag</code> task page for SB Z_CMa_a_06_TM1 (2018.1.01131.S) showing the diagnostic information of this calibration step. . . . .	53
5.18	An example of the plot from the WebLog task <code>hifa_wvrflag</code> showing the deviation from median phases versus time for the calibrator targets in SB Z_CMa_a_06_TM1 (2018.1.01131.S). . . . .	53
5.19	Examples of the <code>hsd_baseline</code> plots for SB NGC_4698_a_06_TP (2019.1.00763.L) showing spectral window 19 before (left panel) and after (right panel) the baseline subtraction. . . . .	57
5.20	An example of the <code>plotms</code> GUI showing amplitude versus $uv$ -distance within spectral window 31 for the bandpass calibrator J1517-2422. . . . .	59
5.21	An example of the <code>plotms</code> GUI showing amplitude versus channel for the bandpass calibrator J1517-2422 for spectral window 31. The atmospheric transmission curve is shown in violet. . . . .	59
5.22	An example of the <code>plotms</code> GUI showing phase versus time for the phase calibrator J1610-3958 for spectral window 31. . . . .	60
6.1	ALMA <code>simobserve</code> simulations showing how $uv$ -coverage affects the angular resolution and recovery of extended structure. Panel a shows the “true” sky brightness, while panels b, c and d show images from the extended 12-m Array configuration C43-6, the intermediate configuration C46-3, and the compact configuration C43-1. (Image credit: ALMA Technical Handbook.) . . . . .	64
6.2	An example of the antenna position plot made with the <code>plotants</code> task . . . . .	65
6.3	Three images of a Gaussian source model produced with the same <code>tclean</code> command and shown with the same flux scaling. The images are created from the same simulated measurement set, but the amplitude gain has been corrupted for one of the inner antennas in the centre image, and the amplitude gain has been corrupted for one of the outer antennas in the right image. No primary beam correction has been applied. . . . .	66
6.4	Three images of a Gaussian source model produced with the same <code>tclean</code> command and shown with the same flux scaling. The images are created from the same simulated measurement set, but antenna A25 has corrupted amplitude gains in the centre image and corrupted phase gains in the right image. No primary beam correction has been applied. . . . .	67



11.1	Top panel: The query interface to the ALMA Calibrator Source Catalogue. Bottom panel: An example of a plot from the ALMA Calibrator Source Catalogue for J1256-0547. . . . .	91
12.1	The two Gaussian model image created using the CASA Toolkit commands from the Images, Coordsys and Componentlist modules. . . . .	94

This page was intentionally left almost blank

# Chapter 1

## Introduction

### 1.1 Purpose

This document provides basic introductory information of how to use the Atacama Large Millimeter/submillimeter Array (ALMA) data available through the ALMA Science Archive (ASA), including how to query, extract and scientifically exploit the content of the ASA. The aim of this document is to assist the reader with the exploitation of archived ALMA data by providing a concise summary of the data and practical examples that include CASA (Common Astronomy Software Applications, [McMullin et al., 2007](#)) tasks. While this primer is intended for archival researchers, principal investigators (PIs) of ALMA data may also find it useful. The document starts with a description of querying the ASA (Chapter 2) and a description of selecting and downloading data (Chapters 3 and 4). This is followed by examples of how to inspect and evaluate the quality of the data and images in the archive (Chapters 5 and 6), how to recreate the calibrated visibilities (Chapter 7) and how to re-image the calibrated data (Chapter 8). The document then gives an outline of how to proceed with data combination (Chapter 9), how to use calibrators as science targets (Chapter 10) and how to use the ALMA Calibrator Source Catalogue (Chapter 11). The document concludes with a description of useful CASA packages (Chapter 12), a list of frequently encountered issues (Chapter 13) and additional information on obtaining assistance with archive mining and science exploitation (Chapter 14).

### 1.2 The ALMA Interferometer

ALMA is one of the largest multinational science projects to date. The ALMA interferometer consists of the 12-m Array, which is made up of fifty 12 meter diameter antennas, and the Atacama Compact Array (ACA), also known as Morita Array, which itself comprises the 7-m Array, consisting of twelve closely-spaced 7 meter diameter antennas, and the Total Power or TP Array, consisting of four 12 meter antennas used for total power observations. ALMA is a complete imaging and spectroscopic instrument currently operating at frequencies of 84 to 950 GHz (wavelengths of 3.6 mm to 0.320  $\mu\text{m}$ ). It can perform polarimetry and mosaicking observations, and data from the three different antenna arrays of ALMA – the 12-m Array, 7-m Array, and TP Array – can be combined to examine the observed source structure on various spatial scales. The 12-m Array antennas are transportable and can be arranged into a number of configurations with baseline<sup>1</sup> lengths from 0.16 km to 16 km.

The ALMA antennas are located at roughly 5000 m altitude on the Llano de Chajnantor in northern Chile, one of the driest locations on earth. The atmospheric stability and transparency at submillimetre wavelengths are essential for ALMA, and this, combined with the large number of antennas and wide bandwidth, make ALMA the most sensitive millimetre/submillimetre observatory on Earth.

The location of the telescope itself is referred to as the Array Operations Site (AOS). The telescope is operated from the Operations Support Facility (OSF), and the operations are the responsibility of the Joint

---

<sup>1</sup>A baseline is the separation between two antennas.

ALMA Observatory (JAO), which has a central office in Santiago, Chile. The observatory and the global astronomical community interact through the ALMA Regional Centers (ARCs) spread across East Asia, Europe and North America.

More information on the ALMA interferometer as well as interferometry in general can be found in [Observing with ALMA - A Primer](#) and the [ALMA Technical Handbook](#).

### 1.3 Science with ALMA

The extensive capabilities of ALMA have inspired a very wide range of scientific studies ranging from Solar System objects to star formation in the Milky Way to interstellar gas and dust in galaxies at high redshifts. The technical requirements of ALMA were initially motivated by three main science aims.

- I The ability to detect spectral line emission from C<sup>+</sup> in a normal galaxy like the Milky Way at a redshift of 3 in less than 24 h.
- II The ability to image gas kinematics in a solar-mass protostellar disk at 150 pc so as to study physical, chemical and magnetic field structures and detect tidal gaps from planet formation.
- III The ability to provide high dynamic range (up to 1000) images at 0.1 arcsec resolution for targets above 20 degree elevation.

More information on science with ALMA can be found in [Observing with ALMA - A Primer](#) and on the ALMA website<sup>2</sup>. In addition, the ALMA Science Archive can be queried for publications associated with archival ALMA data through the Publications results table (See Section 2.6).

### 1.4 The ALMA Science Archive

The first ALMA science observations were performed on 2 March 2011. Since then, ALMA observations have been stored in the ALMA Science Archive (ASA)<sup>3</sup>. The ASA is publicly accessible and can be queried using many different criteria. It contains all the science observations of PI projects, including regular projects, projects that use ALMA as a phased array for VLBI, Large Programs<sup>4</sup> and Director's Discretionary Time (DDT) projects. The archive also contains calibrator data and Science Verification data. Data taken for regular ALMA projects are proprietary for 12 months after a Member Observing Unit Set (Member OUS, see Section 2.1) has been delivered to the PI. Data for DDT projects are proprietary for 6 months after delivery of the Member OUS. After the proprietary period, the data become public. All released Science Verification projects are also publicly available. A thorough description of the ASA and how it works can be found in the [ALMA Science Archive Manual](#).

---

<sup>2</sup>A (non-exhaustive) list of ALMA science highlights organized under six broad themes can be found at <https://almascience.org/alma-science>.

<sup>3</sup>The ALMA Science Archive is accessible at <http://almascience.org/asax/>.

<sup>4</sup>A complete list of Large Programs as well as abstracts and links to the primary science articles, team websites and data deliveries can be found at <https://almascience.org/alma-data/lp>.

## Chapter 2

# Querying the ALMA Science Archive

### 2.1 ALMA data structure

The ALMA Science Archive is the repository of the data collected with ALMA and contains data packages for each proposed project.

Each scientific project is split at the proposal stage into Science Goals (SG) according to the purpose, spectral setting and sensitivity goals of the observation. An SG consists of a set of fields that are observed using a specific set of spectral tunings to achieve a set of specified sensitivity and angular resolution goals. Each SG is translated at the observing stage into one or more Group Observing Unit Set (Group OUS).

Each Group OUS may contain one or more Member OUSs. Data from different ALMA configurations and with different spectral tunings may be placed within separate Member OUSs. A single Member OUS may contain multiple science targets. Typically, one Member OUS corresponds to one Scheduling Block<sup>1</sup> (SB) and contains the data from all repetitions of the observations. Each of these repetitions is called an Execution Block (EB). A Member OUS is the minimal searchable and downloadable unit of archived science observation data.

### 2.2 The ASA query and results interface

Using the ASA web interface<sup>2</sup>, the archive can be searched for Member OUSs in several ways. The ASA manual, ASA news, tips and tricks and FAQs can be found in the Archive documentation website<sup>3</sup>.

The top panel of Figure 2.1 shows the ASA search and results interface as it appears when it is first opened. The top left part shows images of the sky using Aladin Lite. When zoomed in, this panel shows the sky coverage of individual ALMA observations. At the top left of the page above the image panel is the search box, which is a white rectangle with a magnifying glass icon. When hovering the pointer over the search box, it expands to show a number of search keys as shown in the bottom panel of Figure 2.1.

The lower panel of the ASA interface shows the results table. It displays all ALMA observations by default when no search keys are specified.

The top right panel of the ASA interface shows, in yellow, the spectral coverage of the data in the results table. Hovering over a specific observation in the results table will highlight it in bright yellow within the spectral plot. Selected data from the results table will appear orange. The panel also shows by default a typical ALMA atmospheric transmission curve and a selection of molecular spectral line frequencies shifted to the median estimated redshift of the displayed observations. Specific spectral lines can be selected using either the “Molecules” drop-down box or by entering the molecular line of interest or part of the molecular line’s name

---

<sup>1</sup>For commissioning projects, a Member OUS may contain more than one SB, but most Member OUSs in the archive will contain only one SB.

<sup>2</sup>The ASA web interface is located at <http://almascience.org/asax/>.

<sup>3</sup>The ASA documentation is accessible at <https://almascience.org/alma-data/archive/archive-documentation>.

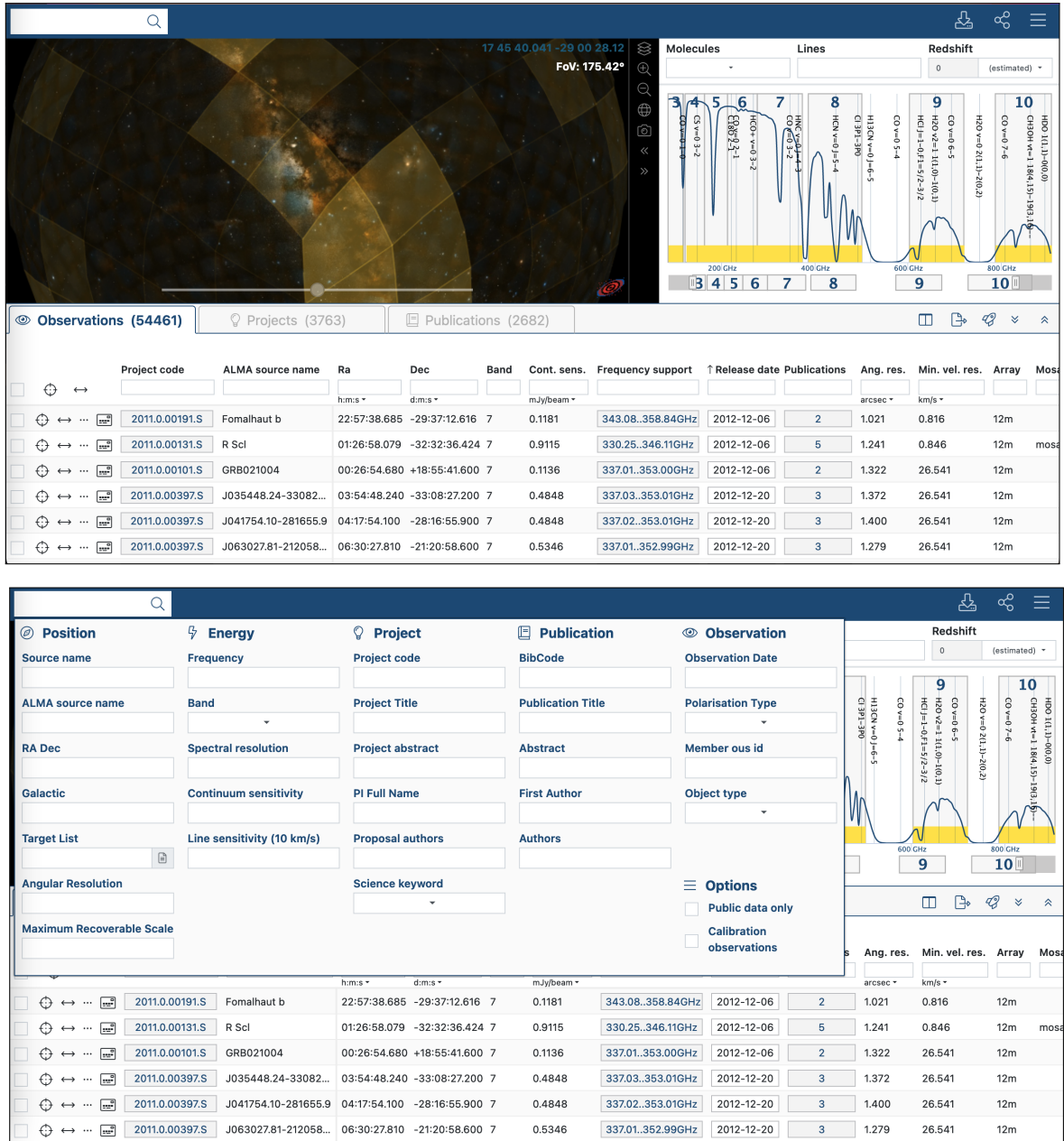


Figure 2.1: The ASA web interface. The top panel shows the initial view. The bottom panel shows the expanded search box, which is marked by a rectangle with a magnifying glass at the top left of the page.

in the “Lines” box. For example, entering “HCO” in the “Lines” box shows all line transitions of molecules with “HCO” in their name, such as HCO<sup>+</sup>, t-HCOOD, ect. The placement of the lines can be adjusted by manually inserting a redshift.

## 2.3 Querying the ASA

### 2.3.1 The search box

The principal way to search the archive using the ASA is to enter one or more search keywords in the search box (Figure 2.1, right panel). Most fields are self-explanatory, and hovering over an entry field activates a pop-up window with an explanation. A few notes about the search box are listed below.

- **Source name versus ALMA name:** The field “Source name” can be used to search for a Sesame (NED/SIMBAD/Vizier) recognised name, but no wildcards are allowed. The field “ALMA name”, which does accept wildcards, refers to the list of names given by the PI (which could be, for example, “Mystery\_Object”). The difference between “Source name” and “ALMA name” is best illustrated using Centaurus A as an example. Typing “Cen” into the ALMA name field shows a list of possible completions, including “Centaurus\_A”, “Proxima\_Centauri”, “alpha\_centauri\_A”, “CenA”, “NGC6357\_center”, and “AM\_Cen”. Some of these entries refer to the radio galaxy Centaurus A, others to (different) stars in the Centaurus constellation and one to the center of the nebula NGC 6357. In this case, when using the “ALMA name” selection (without wildcards to avoid mixing different targets) a number of observations of the Centaurus A radio galaxy will be excluded from the results table. Inserting “Cen A” or “Centaurus A”, which are the Sesame recognised names, into the “Source name” field will immediately yield all the observations of Centaurus A regardless of the name used by the PI.
- **Lists of sources:** Multiple sources may be searched for by clicking on the “Target List” option and uploading an ASCII file with one source name or position per line.
- **Multiple Member OUs or Projects:** It is possible to search for multiple Member OUs or project codes by connecting them with the pipe (|) symbol. For example: “uid://A002/X639a2a/X2a|uid://A002/X5a9a13/X528”.
- **Ranges of observing properties:** In addition to specifying single values for position, angular resolution, spectral resolution, sensitivity, observing frequency or polarisation properties, ranges can be specified with Boolean operators (e.g. >, <, !, |, and ..) to filter the search results.

### 2.3.2 Searching using the image panel

The Aladin Lite<sup>4</sup> image panel can be used to browse the sky in multiple bands. Zooming in on a location of interest shows the area mapped with ALMA, including areas from different projects. The top panel of Figure 2.2 shows the fields observed in the Sgr A\* region. Hovering the cursor over one of the footprints will cause an information window to appear with data about the observations, including the project code, the release date and the ALMA band. By clicking on the footprint, the Member OUS corresponding to that footprint is then selected for download, and both the footprint in the image pane and the corresponding row in the results table become orange. Additionally, a yellow box appears in the top right of the interface with the number of observations selected for download. The bottom panel of Figure 2.2) shows an example of this. Note that the image panel only shows those ALMA observations that meet the search criteria set in the search box or the subfilters in the results table. Search criteria are treated logically as a Boolean ‘AND’ when both search box and results table subfilters are used. When no constraints are set, the panel shows ALMA’s complete coverage of the sky.

---

<sup>4</sup>The Aladin sky atlas (in Lite version) is available at <https://aladin.u-strasbg.fr/AladinLite/>.

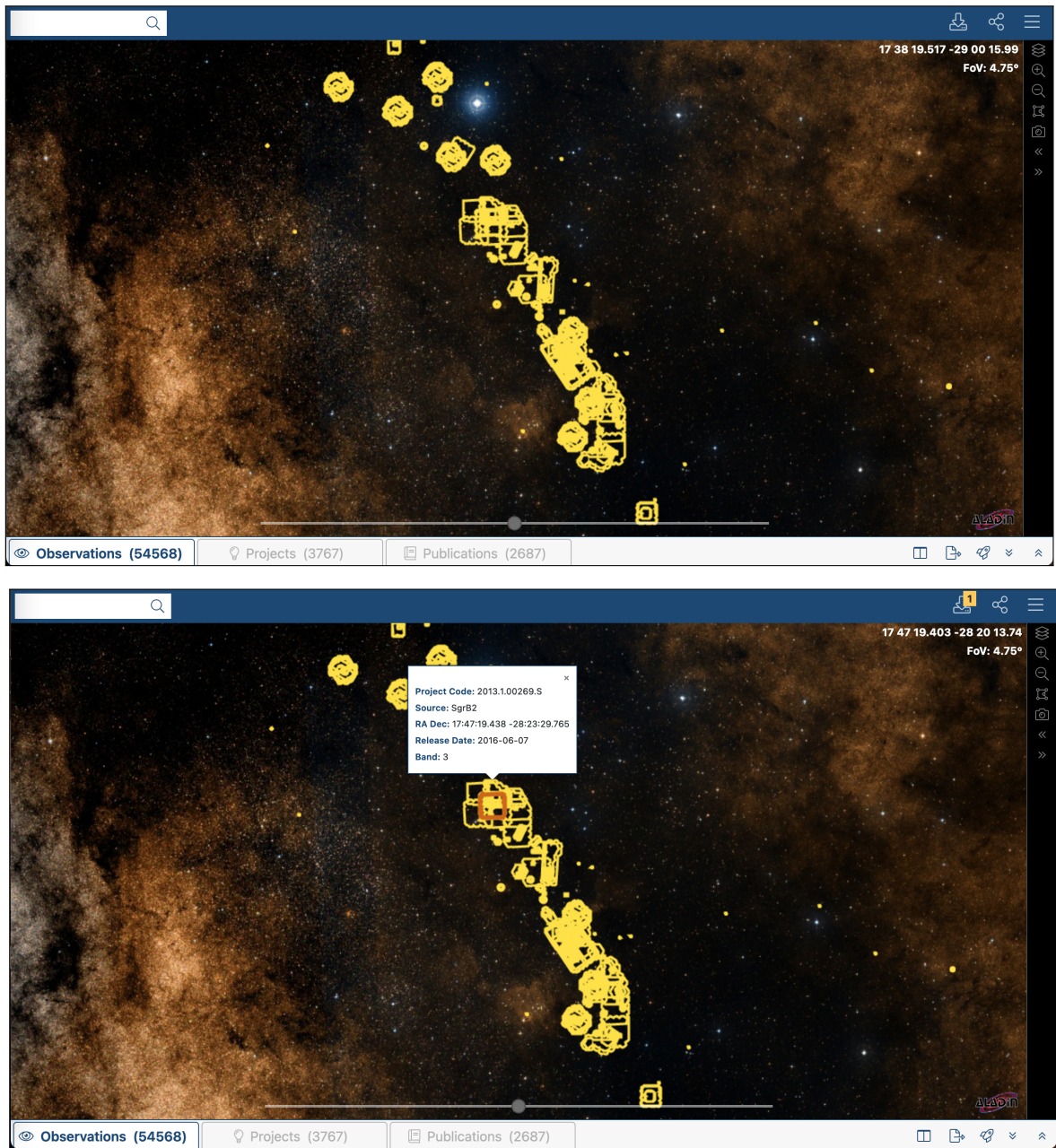


Figure 2.2: The image panel of the ASA interface zoomed towards the Sgr A\* region. The top panel panel shows the footprints of ALMA observations in the region on the sky. The bottom panel shows the footprint of the Sgr B2 star-forming region that was selected for download.



### 2.3.3 Refining search results in the results table

The bottom panel in the ASA interface (Figure 2.1, left) shows the results table. Without any search constraints, it will list all ALMA observations, with each observation (i.e. each source within a Member OUS) listed per row. The list can be subfiltered using the search boxes below each column name. For example, typing “>0” in the box under Dec will list all Member OUSs in the northern half of the sky, which would be useful for identifying ALMA data that could be combined with data from a telescope in the Northern Hemisphere. As another example, typing “0” in the box under Publications will list all data that have not yet been published.

### 2.3.4 Preview of archived ALMA observations within the ASA interface

Interactive previews of the archived ALMA datasets can be viewed by hovering over one of the postcard buttons in results table. An example is shown in the top panel of Figure 2.3. The preview makes use of QA2-generated images cubes as well as external image cubes, such as those from the ARI-L project (see Section 3.1.1). For each spectral window, the previews will show, when available, a continuum image, a maximum flux (moment 8) image and, for a strong spectral line, moment 0 and moment 1 images. In addition, three spectra (from the peak in the continuum emission, the peak in the integrated line emission, and from the entire imaged area) are shown. Line identification in these spectra is performed by the ALMA Data Mining Toolkit (ADMIT)<sup>5</sup>.

The results for the different spectral windows can be viewed by scrolling upwards or downwards in the pop-up window. Clicking on a map or spectrum will open an interactive preview window for that spectral window as shown in the bottom panel of Figure 2.3.

## 2.4 Example: star-forming regions on the spiral arms in M 83

Consider an example study into the spatial distribution of star-forming molecular clouds within the nearby face-on spiral galaxy M83. Figure 2.4 shows an example of a search using “M83” as the source name and a frequency range of 230-232 GHz (using “230..232” in the Frequency box as explained in Section 2.3.1). This range covers the <sup>12</sup>CO(2–1) line at 230.530 GHz, which is a strong spectral line that can be used as a tracer of these molecular clouds. Once M83 is selected, the redshift will be indicated at the top of the top right panel and applied to the spectral line transitions shown in the spectral coverage plots.

This initial search returns 34 observations belonging to 5 projects. The image pane will show the footprint of each observation. For investigating star formation in the spiral arms, the large mosaics covering the inner spiral arms of M83 may be useful. Hovering over one of the outer footprints will show that they belong to project 2015.1.00121.S and at the same time, highlight the spectral line coverage in bright yellow in the spectral coverage panel (see Figure 2.5).

Typing “2015.1.00121.S” under Project code in the results table will filter the search results to show just the observations from this project as shown in Figure 2.6. Once this footprint is selected by clicking on it, the observation is selected for download, and the corresponding row in the results table, the corresponding footprint, and the corresponding spectral line coverage will be highlighted in orange. The search results also show in the Publications tab that multiple paper have already been published for these data. The preview of the image and image-cubes can be inspected by hovering over the postcard icon. These previews will show, for example, the bright CO(2–1) line emission from the central parts of the galaxy.

## 2.5 Producing query scripts

The ASA can also be accessed using scripts. The tools access the ASA through standardized Virtual Observatory (VO) services. ALMA provides the VO services [Table Access Protocol](#), [Simple Image Access](#) as well as [DataLink](#).

---

<sup>5</sup>The ADMIT software package is available at <https://admit.astro.umd.edu/index.html>.

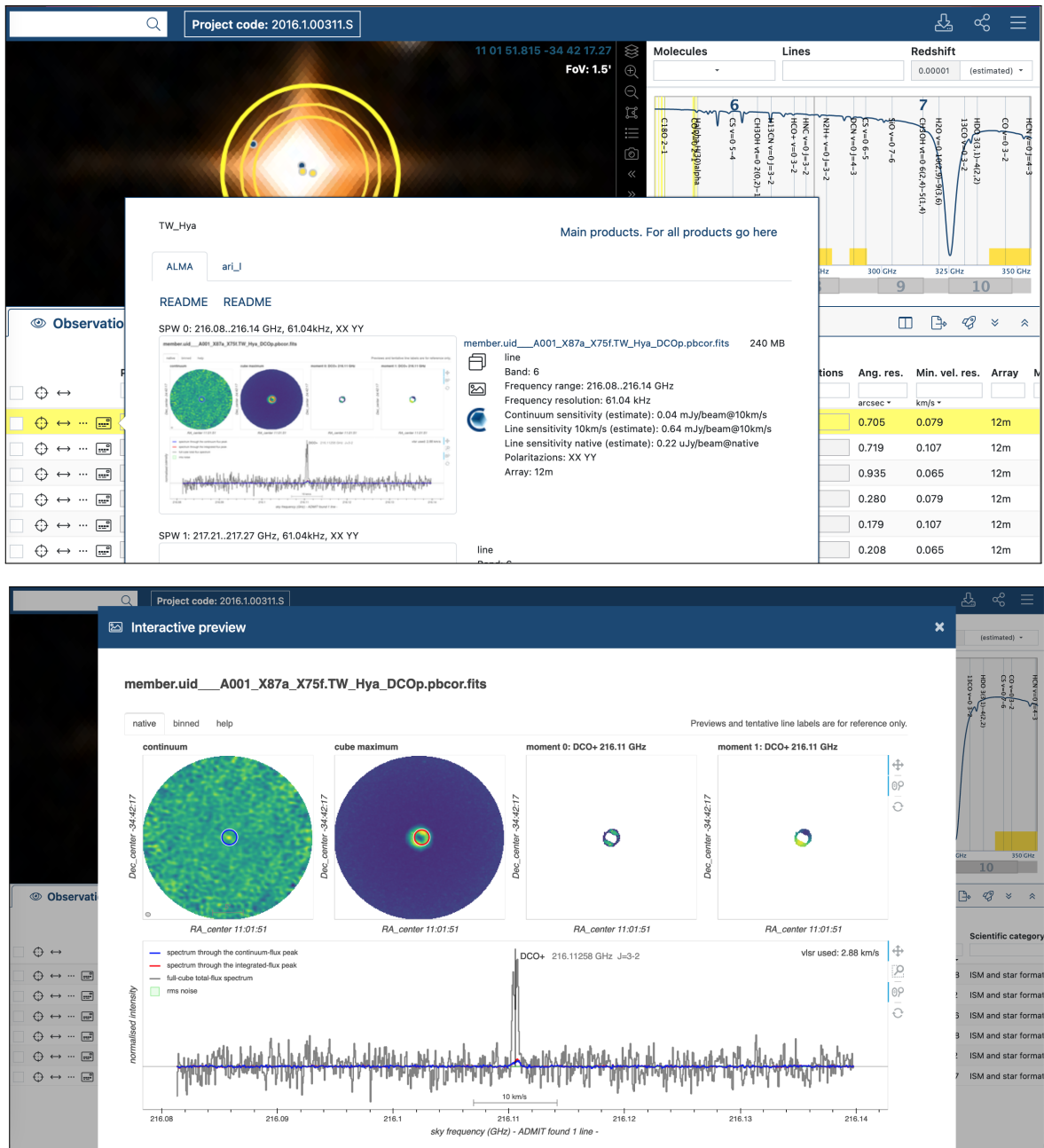


Figure 2.3: The top panel shows the preview window that appears when hovering over the postcard icon. The bottom panel shows the interactive preview of a single spectral window after having clicked on the map/spectrum of that spectral window.

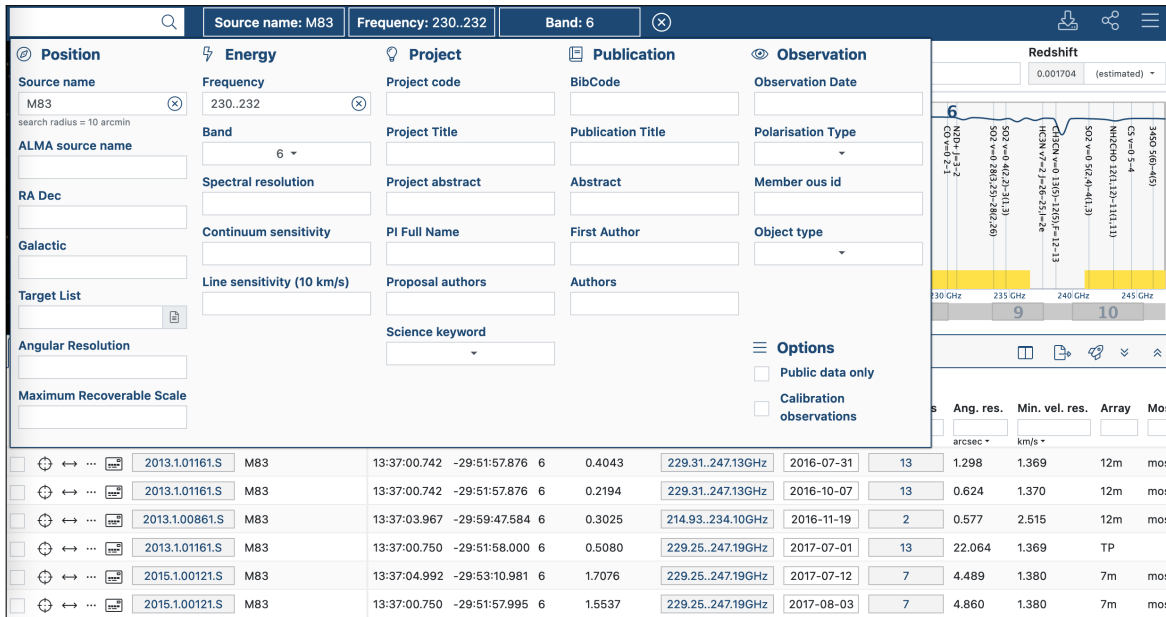


Figure 2.4: An example of using the search box to query the archive for observations at the position of M83 within the 230-232 GHz frequency range.

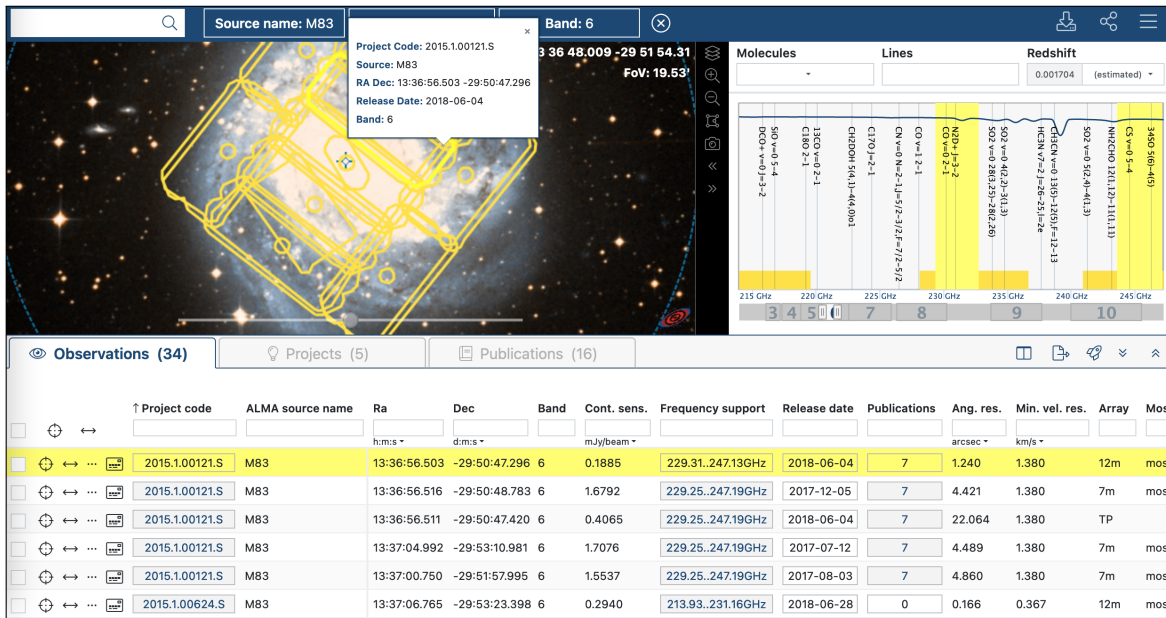


Figure 2.5: The results listed in the Observations table after performing a query for observations at the position of M83 within the 230-232 GHz frequency range. Each observation is shown as a yellow footprint on the sky map and as yellow bars in the spectral coverage panel. Hovering over an observation makes the corresponding row in the search results table, the corresponding footprint and the corresponding frequency coverage all appear bright yellow.

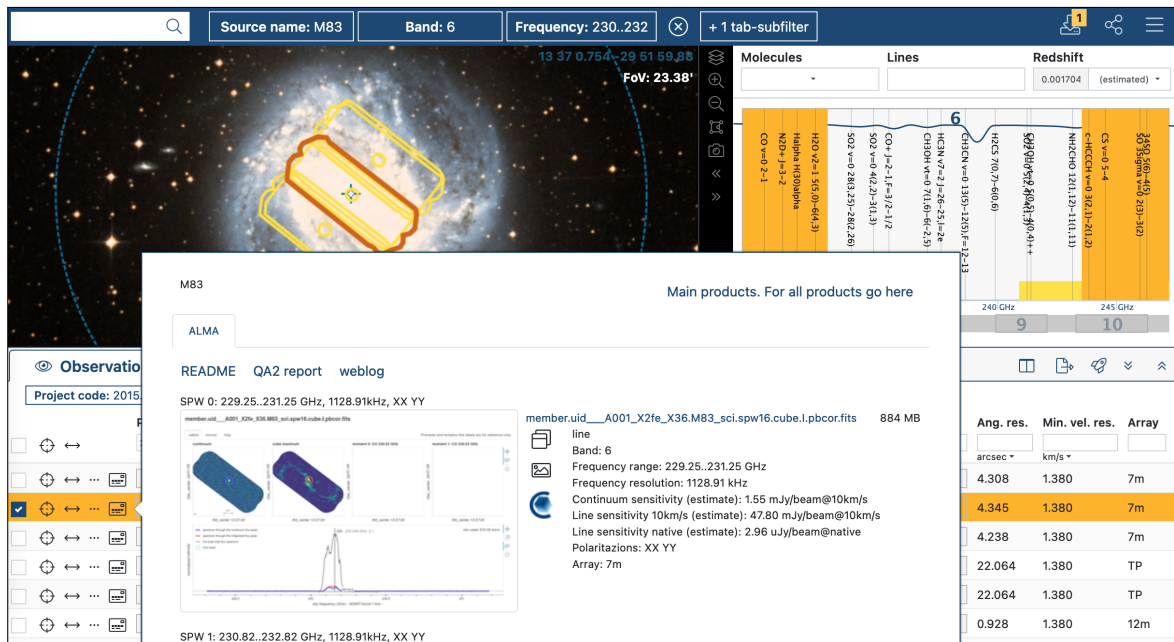


Figure 2.6: Another example view of the search results for M 83 observations in the 230-232 GHz frequency range that illustrates how the results can be refined by selecting only observations from project 2015.1.00121.S. When an observation is selected, the corresponding row in the search results table, footprint and frequency coverage are all coloured orange. Hovering over the postcard icon will show the interactive preview for that observation.

## 2.5.1 PyVO

One convenient way of accessing the ASA programmatically is the pyVO package<sup>6</sup> To use pyVO, it first needs to be imported into a python session<sup>7</sup> as shown below.

```
>>> import pyvo
```

One of the most common searches is by coordinate. The example below shows how to perform a cone search of radius 0.006 degrees around the coordinates  $RA = 201.365063$  and  $DEC = -43.019112$ .

```
>>> service = pyvo.dal.TAPService("https://almascience.eso.org/tap")
>>> query = f""" \
        SELECT * \
        FROM ivoa.obscore \
        WHERE INTERSECTS(CIRCLE('ICRS',201.365063,-43.019112,0.006),s_region)=1 \
        """
>>> output = service.search(query).to_table().to_pandas()
```

It is possible to search the archive based on the properties of the individual datasets. The example below shows a search for all data from the program 2018.1.01131.S.

<sup>6</sup>The pyVO documentation is available from <https://pyvo.readthedocs.io/en/latest/>.

<sup>7</sup>In this document a python terminal is shown as starting with >>>.

```
>>> service = pyvo.dal.TAPService("https://almascience.eso.org/tap")
>>> query = f""" \
          SELECT * \
          FROM ivoa.obscore \
          WHERE obs_publisher_id like '%2018.1.01131.S%'\
          """
>>> output = service.search(query).to_table().to_pandas()
```

The results from one of these searches can be printed using the line below.

```
>>> output.head
```

The output is a data table in the form of a `pandas.core.frame.DataFrame` object<sup>8</sup>.

## 2.5.2 Astroquery

Astroquery<sup>9</sup> is an externally-developed python package that allows for querying and retrieving data from many astronomical archives including the ASA (see [Ginsburg et al. 2019](#)). Installation instructions and examples are available in the [astroquery documentation](#), but we show three examples below of querying the archive with astroquery version 0.4.4<sup>10</sup>.

To begin with, an import command needs to be executed in a python session before Astroquery can be used.

```
>>> from astroquery.alma import Alma
```

To search for data covering an object, `astroquery.query_object` can be used. Note that this will resolve the position associated with the input name and return data that use alternate names for that object; this search method is not equivalent to searching based on the field name specified in the original proposal.

```
>>> almaQuery=Alma.query_object('M83')
```

Searching for data at a given position requires importing additional packages from `astropy` and defining coordinates as input, but it is otherwise straightforward.

```
>>> import astropy.units as u
>>> from astropy import coordinates
>>> coordQuery=coordinates.SkyCoord('12 37 43.597', '+11 49 05.12',
          frame='icrs', unit=(u.hourangle, u.deg))
>>> almaQuery=Alma.query_region(coordQuery, radius=1*u.arcmin)
```

More complex queries are also possible. The query below will search the ALMA archive for all Cycle 5 programs (which have proposal IDs that begin with the digits 2017).

```
>>> almaQuery=Alma.query_tap("select * from ivoa.obscore where proposal_id like '2017%'")
```

<sup>8</sup>See <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html> for more information on `pandas.DataFrame`.

<sup>9</sup>The Astroquery website is located at <http://www.astropy.org/astroquery/>.

<sup>10</sup>Astroquery versions < 0.4.4 may have different command syntax.

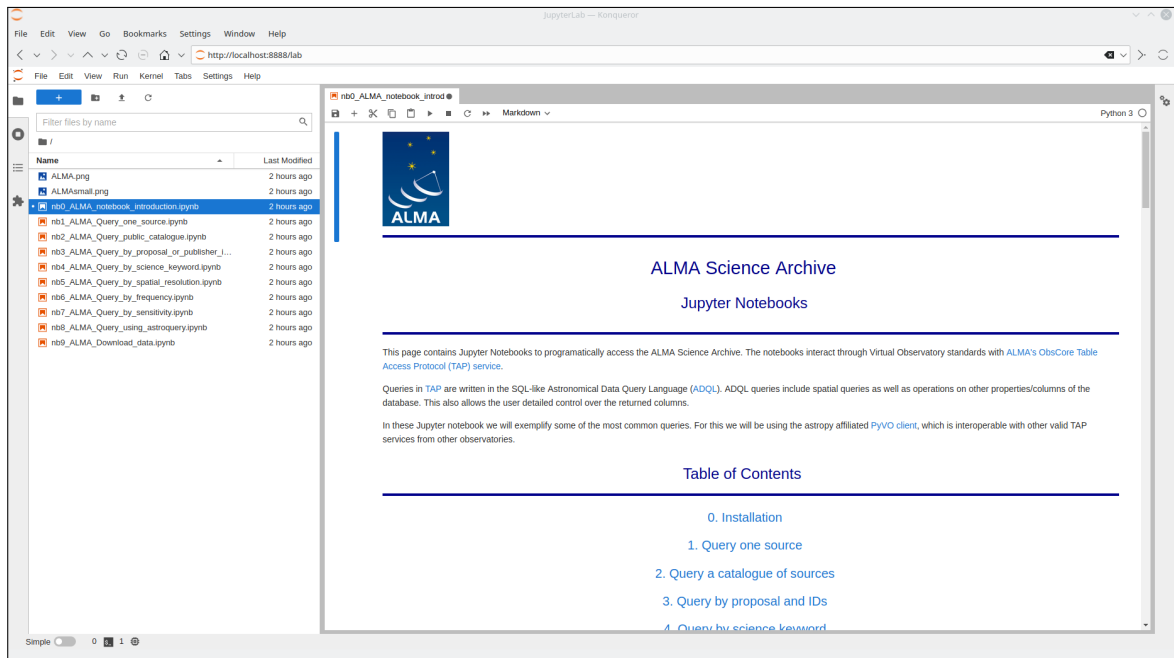


Figure 2.7: An example of the Jupyter Notebooks available for illustrating how to use pyVO and astroquery to search the ASA.

The results are returned as tables (or, more specifically, `pyvo.dal.tap.TAPResults` objects<sup>11</sup>). These results can be used in multiple ways. The example code below will print the proposal IDs and SB names returned by one of these searches (although this may be a very long list for very broad searches, such as the third example above).

```
>>> for i in range(len(almaQuery)):
      print(almaQuery[i]['proposal_id'], almaQuery[i]['schedblock_name'])
```

More information is available from [Ginsburg et al. \(2019\)](#) and the [astroquery documentation](#).

### 2.5.3 ALMA Science Archive Notebooks

To illustrate how to query the archive within Python, the ASA provides a set of Jupyter Notebooks on the ALMA Science Portal that explain in detail how to query the archive in different ways<sup>12</sup>. This includes relatively straightforward queries like searching for all observations associated with a single object but also includes examples based on using other parameters, such as spatial resolution, frequency, or publication, as well as examples of querying for all data associated with a public catalogue.

Most of the notebooks are based on using pyVO, although one notebook shows how astroquery can be used as an alternative. Installation instructions for using both the notebooks as well as astroquery and pyVO are included. See Figure 2.7 for an example of what the Jupyter Notebooks look like.

<sup>11</sup>More information about TAPResults tables is available at <https://pyvo.readthedocs.io/en/latest/api/pyvo.dal.TAPResults.html>.

<sup>12</sup>The notebooks are accessible from <https://almascience.org/alma-data/archive/archive-notebooks>.

BibCode	Author	Journal	Year	Title
2020ApJ...890L.10K	Koda, Jin	ApJL	2020	Systematic Variations of CO J = 2-1/1-0 Ratio and Their Implications in The Nearby Barred Spiral Galaxy M83
2021A&A...653A.172S	Stuber, Sophia K.	A&A	2021	Frequency and nature of central molecular outflows in nearby star-forming disk galaxies
2021A&A...656A.133Q	Querejeta, M.	A&A	2021	Stellar structures, molecular gas, and star formation across the PHANGS sample of nearby galaxies
2021ApJS...255...19L	Leroy, Adam K.	ApJS	2021	PHANGS-ALMA Data Processing and Pipeline
2021ApJS...257...43L	Leroy, Adam K.	ApJS	2021	PHANGS-ALMA: Arcsecond CO(2-1) Imaging of Nearby Star-forming Galaxies
2021PASP..133c4504S	Sawada, Tsuyoshi	PASP	2021	Offline Correction of Atmospheric Effects on Single-dish Radio Spectroscopy
2022MNRAS.509.1303W	Williams, Thomas G.	MNRAS	2022	The 2D metallicity distribution and mixing scales of nearby galaxies

Publications	Ang. res.	Min. vel. res.	Array	Mosaic
7	4.489 arcsec	1.380 km/s	7m	mosaic
7	4.860	1.380	7m	mosaic
7	4.421	1.380	7m	mosaic
7	22.064	1.380	TP	mosaic
7	22.064	1.380	TP	mosaic
7	1.030	1.380	12m	mosaic
7	1.019	1.380	12m	mosaic
7	1.240	1.380	12m	mosaic
7	22.064	1.380	TP	mosaic

Figure 2.8: The ASA results table. Hovering over the publication column displays the publications related to the M83 project 2015.1.00121.S.

## 2.6 Results tables

Any search with the ASA interface always returns three tables: the Observations table, the Projects table and the Publications table. One but only one of these tables are always visible in the bottom panel of the ASA (see Figures 2.1 and 2.5). The Observations table lists one entry per observation (i.e. per source for each Member OUS) satisfying the search criteria. Multiple EBs within an SB and multiple pointings for a mosaic are grouped together in each individual entry. The Projects table lists one row per project as identified by the project code. The Publications table lists one paper per row, which could be associated with more than one project.

As already mentioned above, the result tables can be subfiltered by adding criteria in the box at the top of each column. Furthermore, the table contains a wealth of metadata for each ALMA observation. Hovering the mouse over the table results enclosed by a box will display more details of the content in a pop-up window. For example, hovering over the project code will display the abstract and an acknowledgement statement, and hovering over the publication column will display the Bibcode, Author, Journal, Year and Title of the publication as shown in Figure 2.8.

### 2.6.1 Release dates, proprietary periods and data accessibility

The release date shows when data become public. Data for which the proprietary periods have expired can be anonymously downloaded from the archive. Data that are still within their proprietary periods can be selected for download. However, only PIs and their delegates can actually download proprietary data, so login details need to be provided before the data can be downloaded. When selecting a row in the Observations table that is not yet public, the checkbox will be red. Hovering the cursor over the release date box will reveal a pop-up window with options to be notified of the release using one of several calendar applications.

To allow other users to download data, PIs can delegate access by first logging onto the ALMA Science Portal<sup>13</sup> and then clicking on the link to their user profiles, which appears when hovering the pointer over the username at the top right of the website. The profile page has a number of tabs. The Project delegation tab is the one that PIs should use to delegate access to other users. The “Next” and “Previous” buttons can be used to switch between these tabs. The operation has to be finalised in the Confirm tab for the update to take effect.

### 2.6.2 ALMA quality assurance

In ALMA, a layered quality assurance (QA) process is applied to all data. These layers are given numbers from 0 to 3.

QA0 is a series of checks performed when the data are initially acquired by the telescope; these checks

<sup>13</sup>The ALMA Science Portal is located at <https://almascience.org>.

ensure that the observations were performed correctly and within constraints set by the PI. Some EBs may be incomplete (for example, missing calibration observations or containing too little science target data) and will not be used for QA2 and the final data quality assessment. These data are nevertheless delivered to the archive as QA0 Semi-pass data because the data may still be useful for other scientific purposes, and they appear in the Request Handler as “raw (semipass)” (see Section 3.1.1).

QA1 is a long-term monitoring program not tied to any specific Member OUS that ensures that the telescope is performing at a nominal level.

In QA2, which follows QA0, the data are fully calibrated, and then images (which may cover only part of or all of the available science target data) are created to verify that the resolution and sensitivity requested by the PI are achieved. Data that satisfy these criteria are delivered to PIs, and a 12-month proprietary period begins for the data<sup>14</sup>. After this, the data become public through the archive. If the data do not meet the proposed specifications, additional EBs are acquired if possible. Otherwise, the data are tagged as either “QA2 Semi-pass” or “QA2 Fail”. QA2 Semi-pass data are considered usable for science but did not meet the PI requirements. These data will have undergone the QA2 process, where efforts were made to meet the goals set by the PI, and they may be distributed with the same imaging products, calibration scripts and auxiliary files that are distributed with data that pass QA2. QA2 Fail data are not usable for science, are not delivered to users and thus do not appear in the ASA. The QA2 result is displayed in the QA2 Status column in the Observations table (see Section 2.6). QA2 Pass and most QA2 Semi-pass data are distributed with imaging products, calibration scripts and auxiliary files (see Section 3.1.1). Information about the QA2 status of the Member OUS, including the reason why it was labelled Semi-pass, may be given in the Member OUS README file or in the QA2 report (see Section 5.1).

Occasionally, data defects may be identified after the data are delivered. These data will then undergo another round of quality assurance labelled QA3. Any Member OUSs undergoing QA3 will not be available for download from the archive until the process is complete. After the problems with the data have been resolved, the newly created products are ingested into the ASA and become available. In some cases, an extension of the proprietary period is granted to the PI if the data were not public when the QA3 process started. Details on ALMA’s proprietary period policy can be found in the [ALMA Users’ Policy document](#). More information on ALMA quality assurance processes can be found in Chapter 11 of the [ALMA Technical Handbook](#)

### 2.6.3 Exporting the result tables

Any of the three result tables can be exported in a number of formats. To do this, select the table to be exported, then click on the “Export table” symbol, which is located at the top right of the results table and looks like a page with an arrow pointing to the right. A pop-up window will appear with options for the format of the downloaded table. When one of these options is selected, another window will appear at the bottom right corner of the web browser with buttons to either download or delete the file. Either the complete table or just the selected rows may be exported in VOTable, CVS or TSV formats.

## 2.7 Additional considerations when using archival data

ALMA is a telescope that aims to deliver the data within the angular resolution and sensitivity limits requested in the proposal. However, different science goals may have different requirements. This is why, in archive mining, users should consider the sensitivity, angular resolution, spectral resolution and maximum recoverable scale that are needed for the realisation of their science goals. For explanations of these quantities, consult the [Observing with ALMA - A Primer](#) or the [ALMA Technical Handbook](#).

Chapter 5 provides more information on checking the quality of the data calibration, and Chapter 6 provides information on checking the images from the archive. In general, the calibrated visibility data are considered science-ready for projects labelled as QA2 Pass. However, not all projects, even grade A projects, may be fully observed. Even for completed observations, it is possible that specific characteristics of the data, such as the sensitivity or angular resolution, do not exactly match what was specified in the original proposals.

---

<sup>14</sup>DDT observations have a 6-month proprietary period.



Most ALMA projects have images available from the ASA, and these are generally ready for scientific analysis. However, these images may not necessarily be optimal for all scientific applications. The data may need to be re-imaged to meet the needs for various scientific applications, such as, for example, using a different spectral smoothing. For early ALMA cycles, re-calibration may be necessary to update the Solar System Object calibration model (see Section 7.3). In the rare case where significant problems are found with the dataset after data delivery, the data set may need to pass through QA3 (see Section 13.8).

Users are always welcome to contact their ARC via the Helpdesk (see Chapter 14) to ask for help with assessing the quality of archival ALMA data and with re-imaging the data.



## Chapter 3

# Downloading data from the ASA

### 3.1 The Request Handler

After selecting one or more entries in one of the Results Tables and pressing the Download request button in the top right of the page, the ALMA Request Handler will open in a new browser tab. See the example below in Figure 3.1. If proprietary data were selected, a login screen will appear, and users will need to input their ALMA credentials to access the Request Handler.

The files that can be downloaded are organised by Project, Science Goal OUS, Group OUS, and Member OUS as explained in Chapter 2. The structure can be seen in the top panel of Figure 3.1. As many data sets are relatively large, it may take a few seconds for the ALMA Request Handler to construct the list of the selected ALMA data. The bottom panel shows the expansion of the product tar ball for SB IRC\_+102-\_a\_06\_TE. All of these files be downloaded separately. Additionally, all primary beam corrected FITS files (`*.pbcor.fits`) can be viewed using the Cube Analysis and Rendering Tool for Astronomy (CARTA)<sup>1</sup> (without downloading the files) by clicking on the CARTA symbol in the Actions column. Figure 3.2 shows the CARTA preview of the CS(5–4) image cube of the target source, AGB star IRC+10 216, when the CARTA symbol is clicked.

#### 3.1.1 Data packages available and their content

For each selected Member OUS, several different kinds of packages may be available for download.

- **Member OUS README:** This ASCII file contains a description of the main actions, categories and files of data at the Member OUS level. It may also contain a brief description of the project and the data quality.
- **product:** These files are the FITS images produced in the QA2 process, including both calibrator and science target images.
- **auxiliary:** This includes the calibration tables, logs, quality assessment plots, and scripts.
- **raw data:** These are the ALMA Science Data Model (ASDM) files. Each file contains the raw data for one EB. These files include QA2 Pass and QA2 Semi-pass data.
- **raw (semipass):** ASDM files for QA0 Semi-pass data.
- **external:** These are files produced by external groups and could include enhanced visibility data or advanced user-created images.

The archive shows all of the raw data for Member OUSs that are labelled either QA2 Pass or QA2 Semi-pass. These data are distributed with a script for calibration, calibration tables and quality assurance imaging products. A Member OUS may also contain “raw (semipass)” data, which are QA0 Semi-pass EBs that did not

---

<sup>1</sup>The CARTA website is accessible at <https://cartavis.github.io/>.

**ALMA Request Handler** Login

Anonymous User: Request #2154970674126 ✓  
Request Title: [click to edit](#)

Download Selected

readme  product  auxiliary  raw  raw (semipass)  external

Project / OUSet / Executionblock	File	Size	Accessible	Actions
Request 2154970674126		8 GiB		
Project 2013.1.01215.S				
Science Goal OUS uid://A001/X13a/Xc3				
Group OUS uid://A001/X13a/Xc4				
Member OUS uid://A001/X13a/Xc5				
SB IRC_+102_a_06_TE				
<input checked="" type="checkbox"/> readme	member.uid_A001_X13a_Xc5.README.txt	13 KiB	✓	
<input type="checkbox"/> product	2013.1.01215.S.uid_A001_X13a_Xc5_001_of_001.tar	6 GiB	✓	
<input checked="" type="checkbox"/> auxiliary	2013.1.01215.S.uid_A001_X13a_Xc5_auxiliary.tar	381 MiB	✓	
<input type="checkbox"/> raw	2013.1.01215.S.uid_A002_Xa48b1f_X1122.asdm.sdm.tar	24 GiB	✓	
<input type="checkbox"/> raw	2013.1.01215.S.uid_A002_Xa82459_Xb6c.asdm.sdm.tar	21 GiB	✓	
<input type="checkbox"/> raw	2013.1.01215.S.uid_A002_Xa82459_Xe2e.asdm.sdm.tar	10 GiB	✓	
<input type="checkbox"/> raw (semipass)	2013.1.01215.S.uid_A002_Xa5d2c_X4680.asdm.sdm.tar	10 GiB	✓	
Member OUS uid://A001/X13a/Xc7				
SB IRC_+102_a_06_TC				
<input checked="" type="checkbox"/> readme	member.uid_A001_X13a_Xc7.README.txt	11 KiB	✓	
<input type="checkbox"/> product	2013.1.01215.S.uid_A001_X13a_Xc7_001_of_001.tar	832 MiB	✓	
<input checked="" type="checkbox"/> auxiliary	2013.1.01215.S.uid_A001_X13a_Xc7_auxiliary.tar	871 MiB	✓	
<input type="checkbox"/> raw	2013.1.01215.S.uid_A002_X9630c0_X1532.asdm.sdm.tar	17 GiB	✓	

**ALMA Request Handler** Login

Anonymous User: Request #2154970674126 ✓  
Request Title: [click to edit](#)

Download Selected

readme  product  auxiliary  raw  raw (semipass)  external

Project / OUSet / Executionblock	File	Size	Accessible	Actions
Request 2154970674126		8 GiB		
Project 2013.1.01215.S				
Science Goal OUS uid://A001/X13a/Xc3				
Group OUS uid://A001/X13a/Xc4				
Member OUS uid://A001/X13a/Xc5				
SB IRC_+102_a_06_TE				
<input checked="" type="checkbox"/> readme	member.uid_A001_X13a_Xc5.README.txt	13 KiB	✓	
<input checked="" type="checkbox"/> product	2013.1.01215.S.uid_A001_X13a_Xc5_001_of_001.tar	6 GiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_C4H_24-23.flux.fits.gz	351 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_C4H_24-23.image.pbcor.fits	515 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CN_2-1.flux.fits.gz	330 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CN_2-1.image.pbcor.fits	515 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CO_2-1.flux.fits.gz	407 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CO_2-1.image.pbcor.fits	515 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CS_5-4_BB3.flux.fits.gz	415 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CS_5-4_BB3.image.pbcor.fits	515 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CS_5-4_BB4.flux.fits.gz	400 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CS_5-4_BB4.image.pbcor.fits	515 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CS_v_1.flux.fits.gz	406 MiB	✓	
<input type="checkbox"/> product	member.uid_A001_X13a_Xc5.IRC_p102_a_06_TE_CS_v_1.image.pbcor.fits	515 MiB	✓	

Figure 3.1: The top panel shows the default view of the Request Handler page for project 2013.1.01215.S. The bottom panel shows the expansion of the product tarball of the first Member OUS.

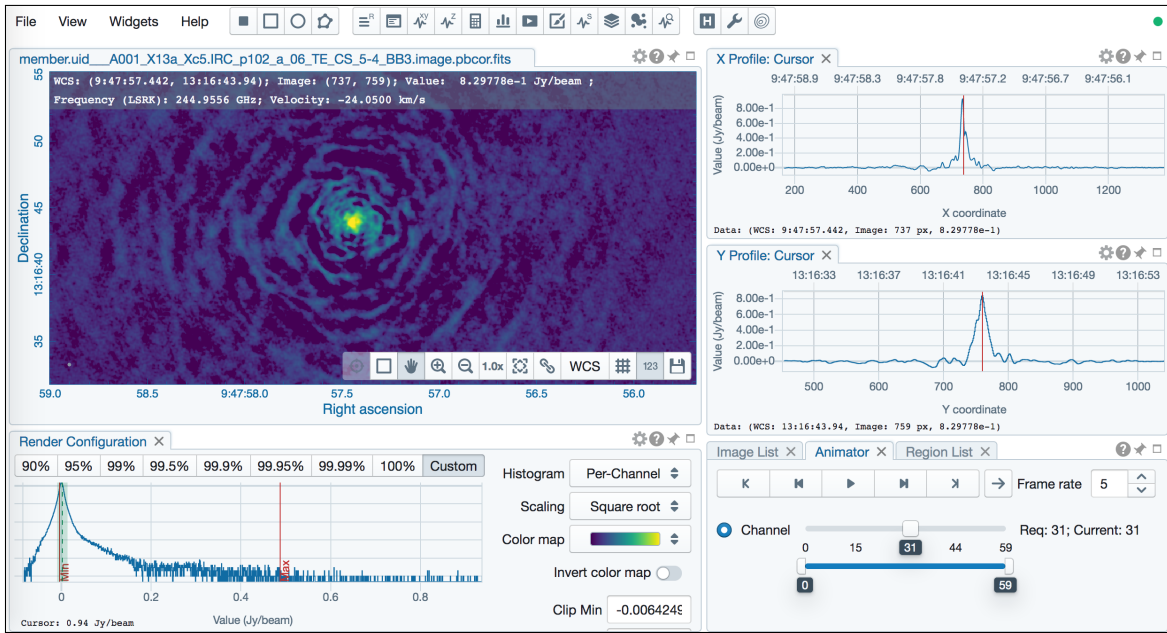


Figure 3.2: An example of the CARTA image viewer displaying the CS(5–4) image cube of the envelope of AGB star IRC+10216.

fulfil the quality assurance requirements but could still be at least partially calibrated and potentially used for science. Hence, QA0 Semi-pass raw data are also included in the archive, although no scripts or image products are generated for these data.

The raw data files are distributed as tar files, with each tar file containing one individual ASDM file. For all ALMA cycles except Cycle 0, the product and auxiliary files are distributed as separate tar files. These tar files have a triangle next to their listing, as shown in the example in Figure 3.1. Clicking the triangle will display a list of all of the files contained within the tar file. After doing this, it is then possible to select individual files to download instead of downloading the whole products package.

## QA2 imaging products

When QA2 imaging has been performed manually (especially for Cycle 4 and earlier), it is possible that only one continuum image of one source or one spectral line image cube of one source was created as the minimal amount of data needed to assess the data quality. Any additional images would have been produced on the basis of the assessor’s ability to do so. From Cycle 5 onward, most ALMA data were passed through the ALMA Calibration and Imaging Pipeline, which automates the image creation process to produce images and image cubes for all of the data when possible. However, when images may exceed 30 GB in size (as may be the case for image cubes of long baseline data), the pipeline may not generate image cubes for these data.

While all archive images are intended to be science-ready, they are not necessarily optimized for all scientific applications, and PIs or archive data miners may want to only use them as indicators of data quality and coverage. Hence, it may be desirable to re-image the data using *tclean* parameters optimized for specific datasets or science projects. Additional improvements can be made by, for example, applying self-calibration on bright science targets (see Section 8.6) or performing user-driven interactive cleaning on images with complex, extended emission.

## External product packages

Externally-provided products are typically images, image cubes or advanced image products. Large Programs<sup>2</sup> are required to provide their imaging products as external products in the ASA.

The Additional Representative Images for Legacy<sup>3</sup> (ARI-L, [Massardi et al. 2021](#)) project is an approved ALMA Development Project that aims to increase the legacy value of the ASA by bringing the reduction level of ALMA data from Cycles 2-4 close to the level of data from more recent cycles. As such, these data are being recalibrated and re-imaged using more recent versions of the ALMA Pipeline. The project has produced a uniform set of full data cubes and continuum images covering at least 70% of the data from Cycles 2-4 that have been ingested into the ASA. These cubes complement the much smaller QA2-generated image products, which cover only a small fraction (< 10%) of the observed data for those cycles.

### 3.1.2 Download selection in the Request Handler

In some circumstances, the Request Handler may display data for multiple Member OUSs within a Science Goal, even if they were not selected in the Results Table. This could occur, for example, when a source is observed with multiple arrays or multiple array configurations within the same program.

By default, the products, auxiliary files (if present), and README file are automatically selected for download. The raw data files (including QA0 Semi-pass data) and any external files (such as images provided by ALMA Large Projects or ARI-L) are not selected by default. It is possible to either select individual files for download or use the checkboxes at the top of the screen to select or deselect all versions of a specific file type (such as all auxiliary files). It is also possible to select or deselect all files within a Project, Science Goal, Group OUS, or Member OUS by clicking on the box next to the item in the list. The Request Handler shows the file size of each package to download, which can be used for preparing disk space for the files.

If the data have not yet been publicly released, this is indicated by a red cross in the Accessible column of the table. It will not be possible to check any boxes next to the files for these data without having permission to access the data either through being PI of the project or by being delegated access to the project. Users who want to select and download any proprietary data but who have not yet logged in will need to do so by using the link at the top right corner of the Request Handler. Upon logging in, an email is sent to the user with a link to the Request Handler, and users with permission to access the data will be able to do so. Note that the PI will never get an email if someone is trying to download their data.

## 3.2 Download options

After selecting one or more files, they can be downloaded by clicking on the Download Selected button. This will provide two options for downloading the files: a download script and a file list. The download script is typically the most reliable method to use for downloading files. Each single file can also be downloaded directly from the Request Handler itself by clicking on the filename.

### Download Script

When this option is selected, the web browser will download a csh script containing the commands<sup>4</sup> for accessing and downloading the archival data. This file needs to be made executable using the terminal command<sup>5</sup>

```
> chmod u+x download*.sh
```

and then can be executed using

<sup>2</sup>A complete list of Large Programs including their data deliveries can be found at <https://almascience.org/alma-data/lp>.

<sup>3</sup>The ARI-L website is accessible at <https://almascience.org/alma-data/aril>.

<sup>4</sup>The download script will first attempt to use `wget`, and when this is not available it will use `curl`.

<sup>5</sup>In this document a Linux/Unix terminal is shown as starting with `>`.

```
> ./download*.sh
```

When the file is executed, it will download all of the requested files. Notifications about the download progress as well as any errors will be shown in the terminal. The script may ask for the user's password before downloading any data, especially if the data are proprietary or if the user was logged in when using the Request Handler. The script will also ask the user whether to untar the files after downloading them.

Multiple files can be downloaded at once, and if a download is interrupted, the script automatically tries to resume. Should this fail, the script can be restarted later to continue the download.

The most common download errors encountered when using the download script are listed in Section 13.1.

## File List

When this option is selected, the web browser will open a new window with links to each of the requested files. The web browser itself can then be used to download the files.

## 3.3 External download options

### 3.3.1 PyVO

PyVO (see Section 2.5.1) can be used not only to search the archive based on various input parameters but also to download data. The example below can be used to download data associated with a specific Member OUS.

```
>>> datalink = pyvo.dal.adhoc.DatalinkResults.from_result_url(
    f"https://almascience.eso.org/datalink/sync?ID={'uid__A001_X135b_X6b'}")
>>> for dl in datalink:
    dl.cachedataset(filename=os.path.basename(dl['access_url']))
```

Note that the slashes (/) in the MOUS name need to be replaced with underscores (\_) when using this pyVO command. This example will retrieve all of the files associated with the Member OUS.

### 3.3.2 Astroquery

Astroquery (see Section 2.5.2) can also be used to query the archive using standard parameter searches (source name, source position, ALMA band, etc.). Both public and proprietary data can be downloaded using python packages described below, but users will need to provide authentication before downloading proprietary data.

The script below will download all of the files associated with a Member OUSs from one of the almaQuery tables created using the commands in Section 2.5. (The username and directory in the example below should be replaced with an actual username and directory before the lines are executed. Also, the login is optional and only needed for proprietary data.)

```
>>> mous=almaQuery['member_ous_uid'][0]
>>> linkList=get_data_info(mous,expand_tarfiles=True)
>>> downloadAlma=Alma()
>>> downloadAlma.login('username')
>>> downloadAlma.cache_location='/directory/'
>>> downloadAlma.retrieve_data_from_uid(mous)
```

This example will retrieve the whole data tarball associated with a Member OUS. However, additional techniques can be used to download just the data that contain the images. See the [Astroquery documentation](#) for additional information.

### 3.3.3 ALMA Science Archive Notebooks

The ASA has provided Jupyter Notebooks that primarily illustrate how to use either pyVO or astroquery to search the archive (see Section 2.5.3). However, one of the notebooks includes examples of how both pyVO and astroquery can be used to download data and could be consulted for further information on how to do this.

## 3.4 Download completion checks

The most straightforward way to check whether the files were completely downloaded from the archive is to re-execute the download script. The script will check whether the download for all files has been completed. If so, the script will print a message indicating that download was successful and ask whether to untar the data. If a prior download is interrupted, then re-executing the download script should automatically restart the download. If the download script prompts the question whether to the untar the downloaded files, the download should have been successful.



## Chapter 4

# Overview of downloaded archival data

### 4.1 Unpacking archival data

Product, auxiliary, raw data and external product files are downloaded from the ASA as tar files<sup>1</sup>. If the download script has not already been used to untar the tar files (see Section 3.2), the command below can be used to do so.

```
> for i in *.tar; do tar -xvf $i; done
```

In tcsh or csh, multiple tar files can be unpacked using a script containing the lines below.

```
#!/bin/csh
foreach i (*.tar)
    tar -xvf $i
end
```

### 4.2 Archival data structure

A typical untarred package will consist of a directory structure organised by project code, Science Goal OUS, Group OUS and Member OUS. This follows the ALMA directory structure as explained in Section 2.1.

The various tar files associated with a single project (e.g the products files, the auxiliary file and the raw data tar files) should be unpacked in the same parent directory so that the files will be placed in the correct place within the ALMA directory structure. For example, after downloading all that is available for scheduling block DK\_Tau\_a\_06\_TM1 (Member OUS uid\_\_\_A001\_X885\_X1f6) from project 2016.1.01164.S and unpacking it, the commands below can be used to inspect the contents of the Member OUS directory, which should list the following output.

```
> cd 2016.1.01164.S/science_goal.uid___A001_X885_X1ee/group.uid___A001_X885_X1f5/
    member.uid___A001_X885_X1f6/
> ls
calibration/                qa/          log/          raw/
member.uid___A001_X885_X1f6.README.txt  script/     product/
```

In this example, the auxiliary tar file would have been unpacked into the `calibration`, `log`, `qa` and `script`

---

<sup>1</sup>Most data tar files can also be downloaded as single files; see also Section 3.1.1.

directories, while the product and raw tar files would have been unpacked into their namesake directories. The README has been placed in the Member OUS directory itself. The Member OUS directories contain the following files.<sup>2</sup>

- **calibration:** This directory contains the calibration tables, the flag tables and, for Member OUSs that are pipeline calibrated, the so-called helper files `flux.csv` and `antennapos.csv`.
- **log:** This directory contains the CASA logs from the reduction and pipeline runs.
- **member.\*.README.text:** The Member OUS README files from Cycle 4 and earlier contain a brief description of project (Project ID, PI, title, sensitivity requested etc.), a summary of the QA2 results and an explanation of the tarball structure as well as file naming conventions. Starting with Cycle 5, the Member OUS README file was reduced to a reference to the location of the QA2 report (See Section 5.1). At the time of this writing, the Member OUS README content is being revised to contain a broader description of the data content.
- **product:** This directory contains the images and image cubes created during QA2.
- **qa:** This contains the output from the quality assurance process. This could include QA0 and QA2 reports and the pipeline WebLog for pipeline calibrated data.
- **raw:** This directory contains the raw data in ASDM format. For Cycle 0 projects, this directory does not contain the raw ASDM files but instead contains measurement sets with system temperatures and water vapour radiometer corrections applied. For these projects, the raw ASDM files are available through a Helpdesk request.
- **script:** This contains the data reduction scripts.

Additionally, the following conventions are used for directories that are not present in most standard datasets from the ASA but that are instead created after the data are downloaded.

- **calibrated:** This directory is generated when the the raw data is calibrated by executing the scripts in the `script` directory (see Section 7.2). For Cycle 0 projects, this directory is present in the tarball and contains the calibrated measurement set.

### 4.3 Archival product files

The `product` directory contains the images and image-cubes of the target fields, the phase calibrator, the bandpass calibrator, the check source (if present), and the polarisation calibrator (for polarimetric observations only)<sup>3</sup>. The naming conventions vary slightly between manually imaged and pipeline imaged data as well as among the manually imaged data from early cycles, but most image products, especially those from Cycle 5 and later, adhere to a strict naming convention. The target intent is coded in the image name. Images for science targets are labelled with `*_sci.*`, though older manual images may contain the (PI given) source name. The names for images of bandpass calibrators use `*_bp.*`, those for phase calibrators have `*_ph.*` and the images for check sources have `*_ch.*`. Polarimetric observations will have also a polarisation calibrator to estimate the polarimetric leakage or  $D$ -terms, and these images have `*_polleak.*`, `*_Pol_Cal*` or just `*_pol*` in their names.

The following naming conventions are used for ALMA images (where `##` is used to indicate a number in the filename). These should be largely identical for manual and pipeline imaging.

The file extensions listed below correspond to the following file types.

- **\*mask.fits:** The mask used by `tclean` or `clean`.

---

<sup>2</sup>For some Cycle 0 Member OUSs, the ASA only contains product and raw data packages. The directory structure and their contents may be different for these Member OUSs.

<sup>3</sup>The archived data for projects from the early ALMA cycles may not include all of the described image products.

- `*pb.fits` or `*.flux.fits`: The primary beam response in `tclean` or `clean` produced images, respectively.
- `*pbcor.fits`: The primary-beam corrected image or data cube.
- `*sd.im.fits`: The single-dish images.

Additional information about the image type as included in the filename is listed below.

- `*mfs.A.*` or `*mfs.POLA*`: The polarisation angle map. This has been the standard used since 2019, but it may have been used earlier.
- `*.mfs.P.*` or `mfs.POLI*`: The linear polarisation intensity map, which is defined as  $\sqrt{U^2 + Q^2}$ . This has been the standard used since 2019, but it may have been used earlier.
- `*spw##.cube.I.*`: The spectral image cube of a single spectral window.
- `*spw##.mfs.I.*`: The multi-frequency synthesis (mfs) continuum stokes I image a single spectral window.
- `*spw##_##_##_#.cont.I.*`: The aggregate bandwidth or continuum Stokes I image based on data from the listed spectral windows.
- `*spw##_##_##_#.cont.I.alpha.*`: The spectral index image, which is defined as  $tt1/tt0$  (see Section 8.5).
- `*spw##_##_##_#.cont.I.tt0.*`: The stokes I continuum image from the mtmfs deconvolver containing the zeroth Taylor term (see Section 8.5).
- `*spw##_##_##_#.cont.I.tt1.*`: The stokes I continuum image from the mtmfs deconvolver containing the first Taylor term (see Section 8.5).
- `*spw##_##_##_#.cont.IQUV.*`: The aggregate bandwidth or continuum full Stokes cube<sup>4</sup> based on data from the listed spectral windows. Individual Q, U, and V images may also be provided.

More information can be found in the ALMA document [ALMA QA2 Data Products for Cycle 7<sup>5</sup>](#).

---

<sup>4</sup>Before Cycle 6, the Stokes V images were not part of the image product package, so the full Stokes cube contained only `*.IQU.*`.

<sup>5</sup>Documents for earlier cycles can be found on <https://almascience.org/processing/qa2-data-products>.



# Chapter 5

## Inspecting archival data

ALMA data undergo a series of quality assessment steps before being delivered to the archive (Petry et al. 2020; see also Sections 2.6.2 and 5.1). This means that, at several stages, expert ALMA data analysts have verified that the data and its calibration are of good quality and meet the PIs' requirements within the ALMA quality assessment standards<sup>1</sup>.

However, it may still be appropriate to review the data quality to understand the specific observations better and to identify any issues in the data, including issues dealt with during the pipeline calibration and previously-unidentified issues (which should be rare). The quality assurance information can also be used to characterise the reliability of the final data in any publications or to understand potential discrepancies between observations. Additionally, archive users may want to create new images or make alterations to the calibration for various reasons, and the quality assessment information could provide guidance on this. Hence, it is important to review the quality of calibrated measurement sets before using it to create final images. Again, remember that ALMA staff have assessed the calibration, so any major problems will likely have been dealt with when creating the archival data.

The steps required to address possible issues spotted with the calibration may vary depending on the situation. Most problems, such as a single bad antenna or a scan with bad amplitudes or phases, will be relatively minor. The solutions to these types of minor problems would simply be to flag the problematic data in the calibrated dataset before re-imaging the data. However, rarer but more complex issues, such as adjusting the flux calibration to use the latest models, may require modification of the calibration scripts downloaded from the ASA. When such complex modifications are needed, users are encouraged to ask their ARC or ARC node for assistance (see Chapter 14).

Results from the quality assurance process can be found in the `qa` directory of each Member OUS. The types of quality assurance data that are present will differ between manually processed and pipeline processed data. The quality assurance data for manually calibrated data consists of a series of image files with diagnostic plots as well as a text file that contains numerical data. The quality assurance files for pipeline processed data consists of a WebLog, which is a series of webpages with figures, and additional files containing QA0 and QA2 report information. More information on the QA2 procedure and the products associated with it can be found in the ALMA QA2 section<sup>2</sup> in the Science Portal.

It is very useful to check several aspects of the calibration before using any manually or pipeline calibrated data, such as the following. More detailed explanations are given in the following sections of this chapter for manually calibrated data (see Section 5.2) and pipeline calibrated data (see Section 5.3).

- **Bandpass calibration:** Are the amplitudes plotted versus frequency devoid of artifacts related to either atmospheric spectral features or spurious lines? Spikes, jumps or regular patterns should not be present

---

<sup>1</sup>In the early ALMA cycles, most data calibration and imaging were performed manually, but since Cycle 5 onwards, pipeline calibration and imaging have been become standard. Note that the ALMA Pipeline is being continuously developed and improved. The ALMA Science Pipeline website (for various CASA versions) is available at <https://almascience.org/processing/science-pipeline>.

<sup>2</sup>ALMA QA2 documentation can be found at <https://almascience.org/processing/qa2-data-products>.

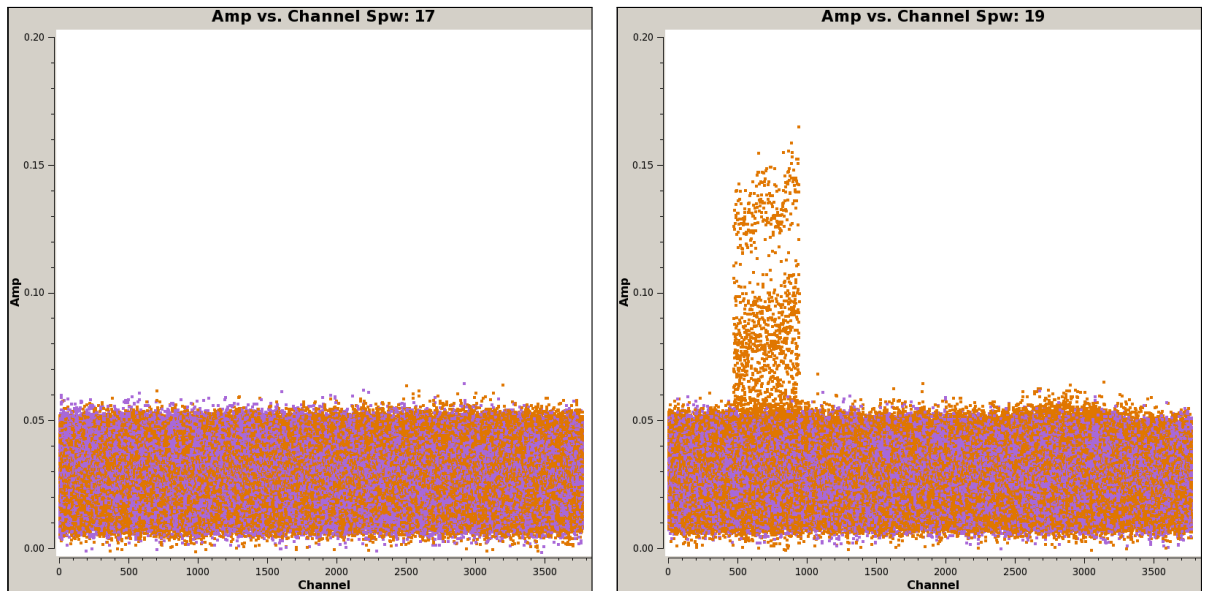


Figure 5.1: Examples of good and bad bandpass calibration coloured by correlation (XX in purple, YY in orange). In the left panel, the amplitudes vary smoothly versus channel (and versus frequency) for all antennas and for both correlations. In the right panel, the YY data for a few baselines jump in amplitude for channels 480-909 and should be flagged. (These plots were produced from manually calibrated ALMA data and are not from the ASA.)

(see for example Figure 5.1). Are the phases versus frequency smooth and continuous, or do they exhibit any sharp jumps or severe scatter that require to be flagged?

- **Flux calibration:** Is the absolute flux scale accurate, and is it compatible with the flux reported in the ALMA Calibrator Source Catalogue for observations near in time (see also Section 11.2)? For Solar System objects, are the latest models (Butler-JPL-Horizons 2012) used for the calibration?
- **Phase calibration:** Do the phase calibrator data plotted versus time exhibit a large amount of scatter in amplitude or phase (see for example Figure 5.2)? Do the phases or amplitudes change suddenly between phase calibrator scans (see for example Figure 5.3)? Does the phase calibrator appear compact in final images? If a check source (an additional calibration source used in observations with configurations with maximum baselines  $> 5$  km or observing frequency  $> 400$  GHz) is present, do the phases and amplitudes vary smoothly versus time, or do they exhibit sharp jumps or high levels of scatter?

ALMA data analysts have assessed these aspects of the data during the QA2 process, and these things should be good for data that are labelled as QA2 Pass. Nevertheless, archive users are strongly encouraged to always investigate the quality of the calibrated data, as this can have a strong impact on the scientific analysis done afterwards. Users should contact their relevant ARC or ARC node via the Helpdesk (see Chapter 14) if they have any concerns about the quality of the calibration of any ALMA dataset.

## 5.1 QA reports for archival data

Since Cycle 5, all manually and pipeline calibrated data comes with a QA0 and a QA2 report in the form of pdf documents<sup>3</sup>. These are part of the auxiliary ASA package, and after that package is unpacked, the reports are placed in the `qa` directory of the corresponding Member OUS.

The version of the QA2 report that is produced for manually calibrated data, which includes all data from before Cycle 5, is discussed further in Section 5.2.

<sup>3</sup>Some Cycle 4 Member OUS may also have QA0 and QA2 reports in pdf format.

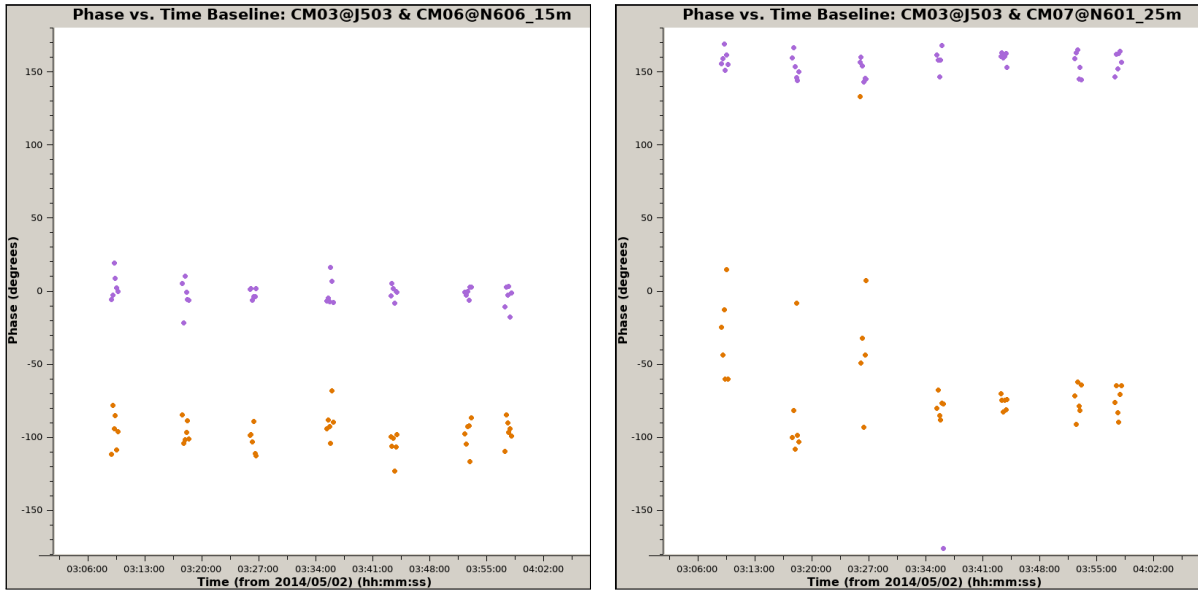


Figure 5.2: Examples of good and bad phase calibration shown for two separate baselines used during the same EB. The XX correlation is shown in purple and the YY correlation is shown in orange. In the left panel, the phases for the CM03 & CM06 baseline show little scatter, which is normal. In the right panel, the phases for the CM03 & CM07 baseline exhibit a large amount of scatter in the first three scans, which is going to cause problems when calibrating the data, and these noisy data should be flagged. (These plots were produced from manually calibrated ALMA data and are not from the ASA.)

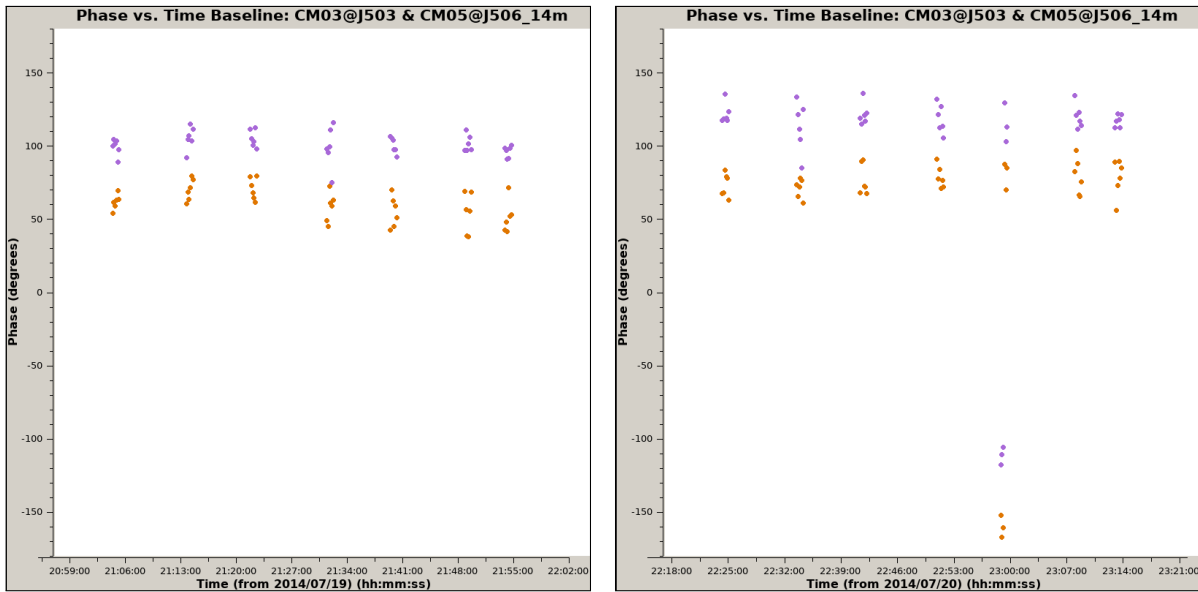


Figure 5.3: Examples of good and bad phase calibration shown for the same baseline but measured during two different EBs in the same program. The XX correlation is shown in purple and the YY correlation is shown in orange. In the left panel, the phases vary smoothly from scan to scan, which is normal. In the right panel, the phases jump in the fifth phase calibrator scan and should be flagged. (These plots were produced from manually calibrated ALMA data and are not from the ASA.)

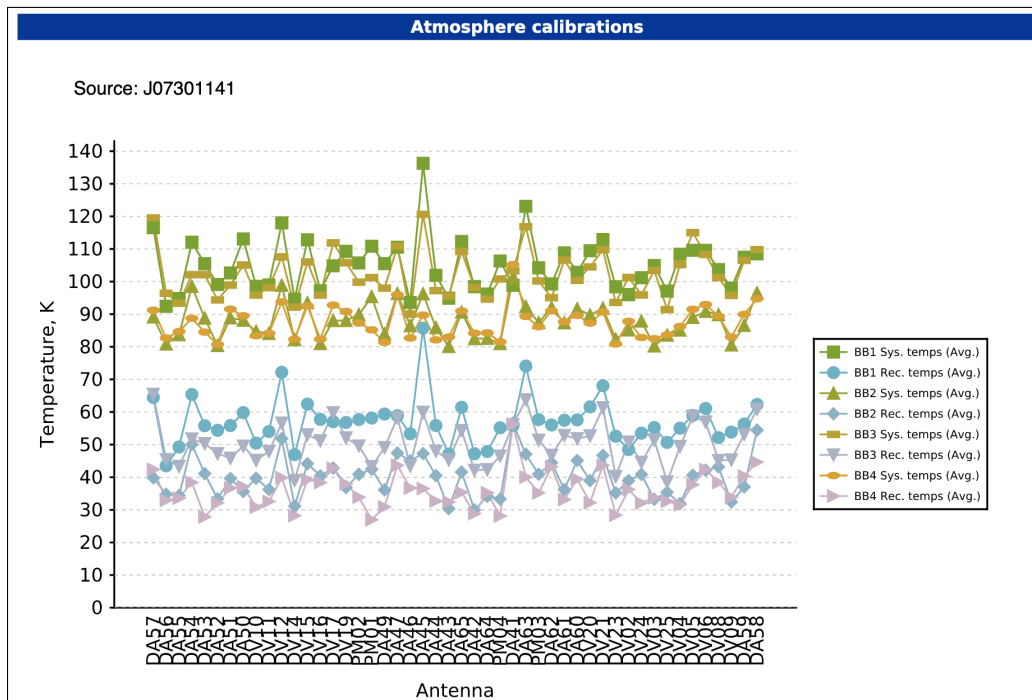


Figure 5.4: An example from SB Z\_CMa\_a\_06\_TM2 (2018.1.01131.S) of the system temperature plotted per antenna in the QA0 report, which can be used to assess the antenna performance.

### 5.1.1 QA0 report

The QA0 report is available in PDF format. The QA0 analysis is performed at the telescope immediately after the observations are completed but before the data are calibrated. Information at the top of the QA0 report includes some basic identifying information about the Member OUS, the array configuration, the weather conditions and the science and calibration sources that were observed, and it will also include any comments from the astronomer on duty during the observations.

The rest of the contents of the QA0 report include various plots used to assess the quality of the observations. The plots showing system temperature plotted versus antenna, as shown in Figure 5.4, can be used to identify any antennas that may not have functioned correctly during the observations. All antennas should have similar system temperature values; any outliers may need to be flagged before imaging if they were not already flagged during the calibration. The plots of phase RMS versus baseline length should be sloped for baselines  $<1000$  m with the phase RMS approaching 0 for very short baselines, but for baselines  $>1000$  m, this slope will become much shallower (see the examples in Figure 5.5 and Matsushita et al. 2017). Outliers in the phase RMS could indicate problems with a noisy antenna, which should be assessed in the WebLogs and dealt with during the data calibration. The pointing offsets show the accuracy of the antenna pointings and are generally not a problem. The baseline distribution simply shows the range of baselines in the interferometric data, which can be good for estimating the beam sizes in the final data.

Note that single dish observations do not produce phase data and that the  $uv$ -coverage does not apply to single dish observations, so plots related to these things are not included in the QA0 reports for total power observations.

### 5.1.2 QA2 report

The version of this report produced from Cycle 5 onwards is available in both PDF and html formats. (QA2 reports from earlier cycles are discussed in Section 5.2.) Unlike the QA2 report for manually calibrated data, the QA2 report for pipeline processed data primarily contains text and tabular information about the data. A



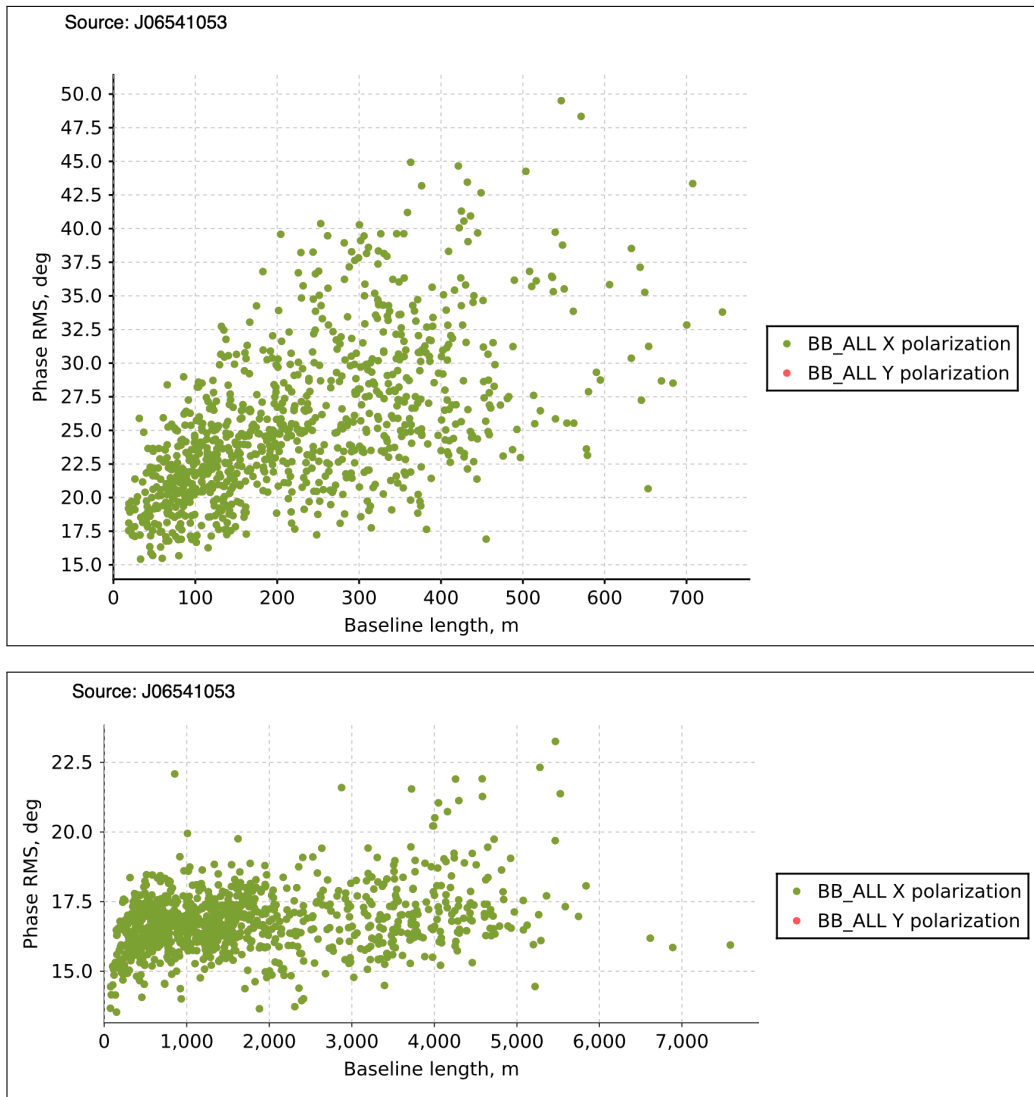


Figure 5.5: An example of the phase RMS behaviour for various baseline lengths, showing calibrator J0654-1053 in Band 6 from SB Z\_CMa\_a\_06\_TM2 (2018.1.01131.S) in the top panel and SB Z\_CMa\_a\_06\_TM1 (2016.1.00110.S) in the bottom panel.

RMS and beam size at representative frequency													
<b>Sensitivity goal</b>	0.01373 [Jy] over bandwidth 2.57812 [GHz]												
<b>Angular resolution goal</b>	0.85017 - 1.27526 [arcsec]												
<b>Achieved RMS for desired bandwidth</b>	0.17000 [mJy]	<b>for continuum</b>							N/A				
<b>Achieved synthesized Major axis (arcsec)</b>	1.260	<b>Minor axis (arcsec)</b>				0.858	<b>Position angle (deg)</b>			N/A			
Execution blocks summary													
EB	N Ant.	Start Time	End Time	ToS (sec)	Avg. Elev. (deg)	Trans. Elev.	Mean PWV (mm)	Phase RMS (deg)	Min BL (m)	Max BL (m)	AR (")	MRS (")	EF
uid://A002/Xd98580/X354	48	2019-03-14 00:58:	2019-03-14 01:19:	1270	70.9	78.5	2.0	0.207	15.0	360.5	1.0	9.6	1.00
Pipeline executions													
CASA version	Pipeline version						Report date			Processing time			
5.4.0-70	42254M (Pipeline-CASA54-P1-B)						2019-04-03 04:36:56			8:22:29			

Figure 5.6: An example of the RMS and beam size, EB summary and pipeline execution sections from the QA2 report of SB Z\_CMa\_a\_06\_TM2 (2018.1.01131.S).

few subsections of this report are particularly useful for assessing the quality of the data and for imaging.

The section with the final QA2 comments contain information on issues that appeared in the pipeline calibration and imaging including specific issues with problematic data that are particularly useful to understanding the calibrated measurement sets. Figure 5.6 shows an example of the sections showing the RMS and beam size section and the EB summary. These sections contain useful indications of the quality of the data in the final images, including the expected sensitivities, beam size and maximum recoverable scale. The table with the calibrator source fluxes provides useful information on the flux densities measured for the calibration sources (other than the flux calibrator itself), and this can be used to assess the stability of the flux calibrations.

The last section of the QA2 report provides instructions for working with the data. This is discussed in more detail in Section 7.2.

## 5.2 Inspecting manually calibrated archival data

As a reminder, the QA2 results from manually calibrated data, which is located in the `qa` directory of every Member OUS, consists of multiple image files and a text file. All files begin with the name of the ASDM. The png files, which have names ending with `_qa2_part*.png`, contain diagnostic plots, and the text file, which ends in `_textfile.txt`, contains calibration data. Figures 5.7, 5.8 and 5.9 show an example of such plots for an ALMA Cycle 3 12-m Array project.

The exact organization of the plots within the png files as well as which plots are present depends on when the data were originally processed through QA2. Additionally, except for two specific plots, all plots were labelled in Cycles 0-5 but have not been labelled from Cycle 6 onwards. However, the plots always appear in the same sequence through the png files. The list below provides the names or descriptions of the plots and explanations of how they can be used to evaluate the data.

- **Observing Schedule:** This plot, which appears at the top right of Figure 5.7, shows the how the observations were organized. The x-axis shows time, while the y-axis shows the purpose of the observation (or the `SCAN_INTENT`). This is mainly for reference.
- **Mosaic Pointing Configuration:** The next plot in Figure 5.7 shows how the pointings were organized for the on-target observations. Again, this is mainly for reference.
- **Antenna Configuration:** Appearing on the left in the middle row in Figure 5.7, this shows the layout

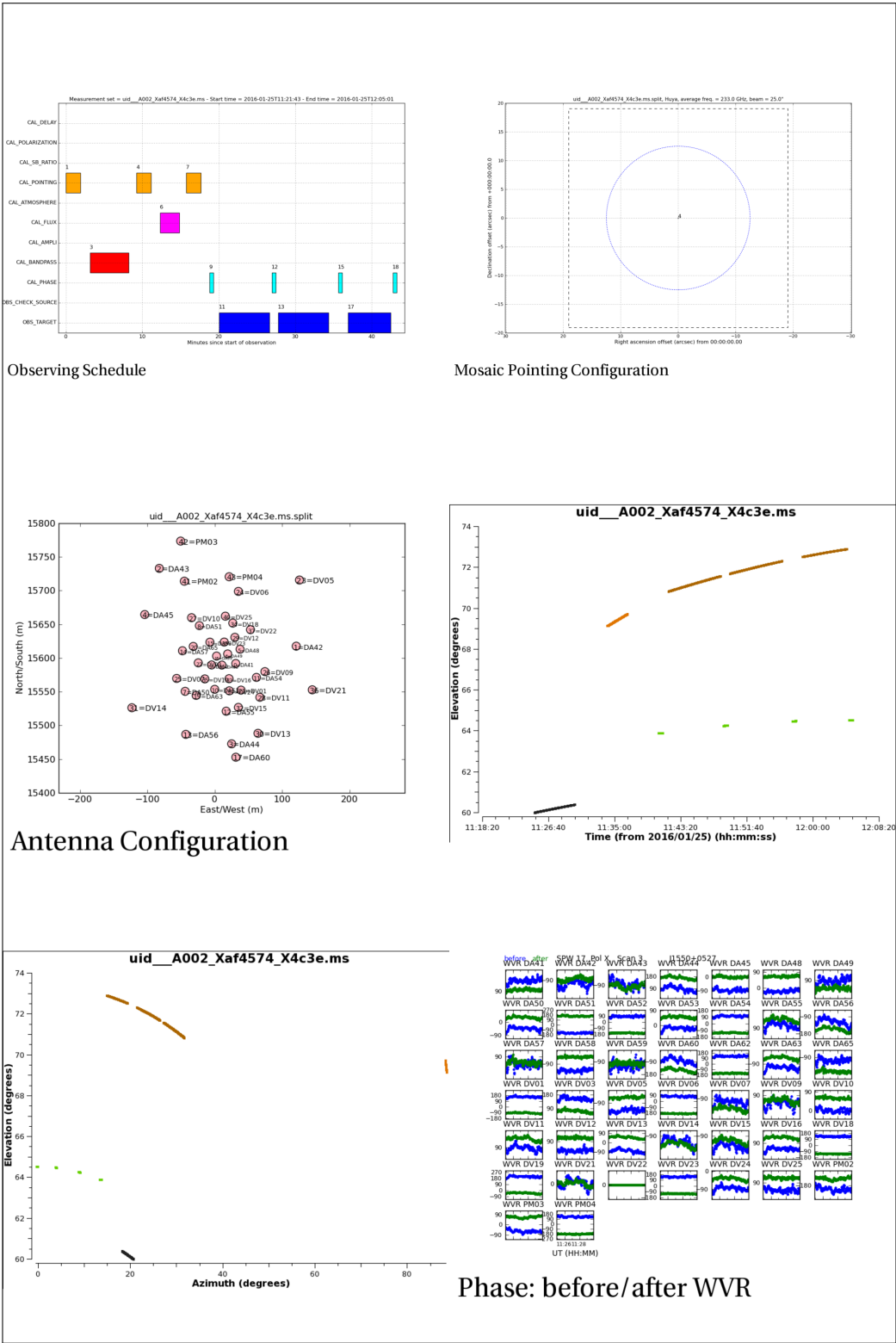


Figure 5.7: Part 1 of the manual calibration QA2 plots from SB Ephemeris\_c\_06\_TE (2015.1.01084.S).

of the antennas during the observations, which can provide some insights into the  $uv$ -coverage of the data and the angular resolution of the observation. Phase and amplitude data from antennas in a compact configuration should generally exhibit less scatter than data from more extended configurations.

- **Elevation versus Time:** This plot in the middle row on the right in Figure 5.7 indicates how much of the atmosphere the telescope was observing through. Moreover, because the  $uv$ -coverage changes between high and low elevations, the elevation data can assist in understanding unusual elongations in the plots of the  $uv$ -coverage.
- **Elevation versus Azimuth:** Located at the bottom left of Figure 5.7, this shows the relative positions of the calibrator and science fields. As with the elevation versus time plot, the elevations can be used to assess how much of the atmosphere the telescope was looking through and how the elevation could affect the  $uv$ -coverage.
- **Phase before/after WVR:** Data from the water vapour radiometers (WVRs) are used to apply a correction to phase versus time for each antenna. These plots, which appear in the bottom right of Figure 5.7, should be checked to ensure that the corrected data, which are shown by the green lines, exhibit generally less scatter than the uncorrected data, which are shown by the blue lines.
- **Temporal gain calibration:** These plots at the top left of Figure 5.8 show the amplitude gains plotted for every antenna. Colour-coded data points are plotted for each spectral window, although the data points may all look like they are one colour if the data are plotted on top of each other. The amplitudes should vary smoothly over time. Any sudden jumps in the measurements that look irregular by eye (other than phases changing between  $180^\circ$  and  $-180^\circ$ ) would be indicative of problems with the data calibration within a specific scan, and the affected calibrated data may need to be flagged.
- **Temporal phase calibration:** This next set of plots are similar to the temporal gain plots except that they show phase instead of amplitudes. Again, the data should vary smoothly over time, and if any sudden jumps are seen in the data, then the data in the affected scans may need to be flagged. The phase data may be more noisy than average for a couple of the antennas, but if the data are very noisy for multiple antennas, then it may be indicative of an issue with the reference antenna used to derive the phase calibration. In this case, it may be useful to attempt to re-calibrate the data with a different reference antenna.
- **Average Bandpass for each spw:** These plots in the middle row of Figure 5.8 show the calibration of the amplitudes in each spectral window versus channel and hence versus frequency. The bandpass data should appear relatively flat and should not vary by more than a few percent. It may drop at the edges of one or more of the spectral windows if these are near the baseband edge where the receivers become insensitive, and it may drop in locations where an atmospheric spectral feature is present, in which case it is worthwhile noting the affected channels. If atmospheric spectral lines are not corrected properly by the bandpass calibration, artefacts could appear in the final spectra of the science targets at the frequencies of those lines.
- **Composite  $T_{\text{sys}}$  for each spw:** The  $T_{\text{sys}}$  measurements are used in an amplitude-based correction versus frequency. The measurements in Figure 5.8 are displayed for each spectral window. The curves for each antenna are normally plotted using different colours, and the different polarisations are plotted using different line styles. Each plot for each spectral window also shows the atmospheric transmission. The  $T_{\text{sys}}$  should appear relatively smooth, although the lines may peak in locations with atmospheric spectral features (as seen in the atmospheric transmission curves); these peaks may also correspond to artefacts in the bandpass data. Also, the curves should all have close to the same magnitudes. If any antenna has a curve that is notably higher than the others, it may not be possible to calibrate those data properly, and it would be important to check whether these data are flagged before the data are used for imaging. The expected  $T_{\text{sys}}$  spectra for each Band are shown in Chapter 5 of the [ALMA Technical Handbook](#). Alternatively, the ALMA Sensitivity Calculator<sup>4</sup> can be used to determine the expected  $T_{\text{sys}}$  at any frequency.
- **Target Spectrum for each spw:** These spectra at the bottom of Figure 5.8 show an example of what the science target data look like after they have been fully calibrated. These spectra are primarily meant to

---

<sup>4</sup>The ALMA Sensitivity Calculator is located at <https://almascience.org/proposing/sensitivity-calculator>.

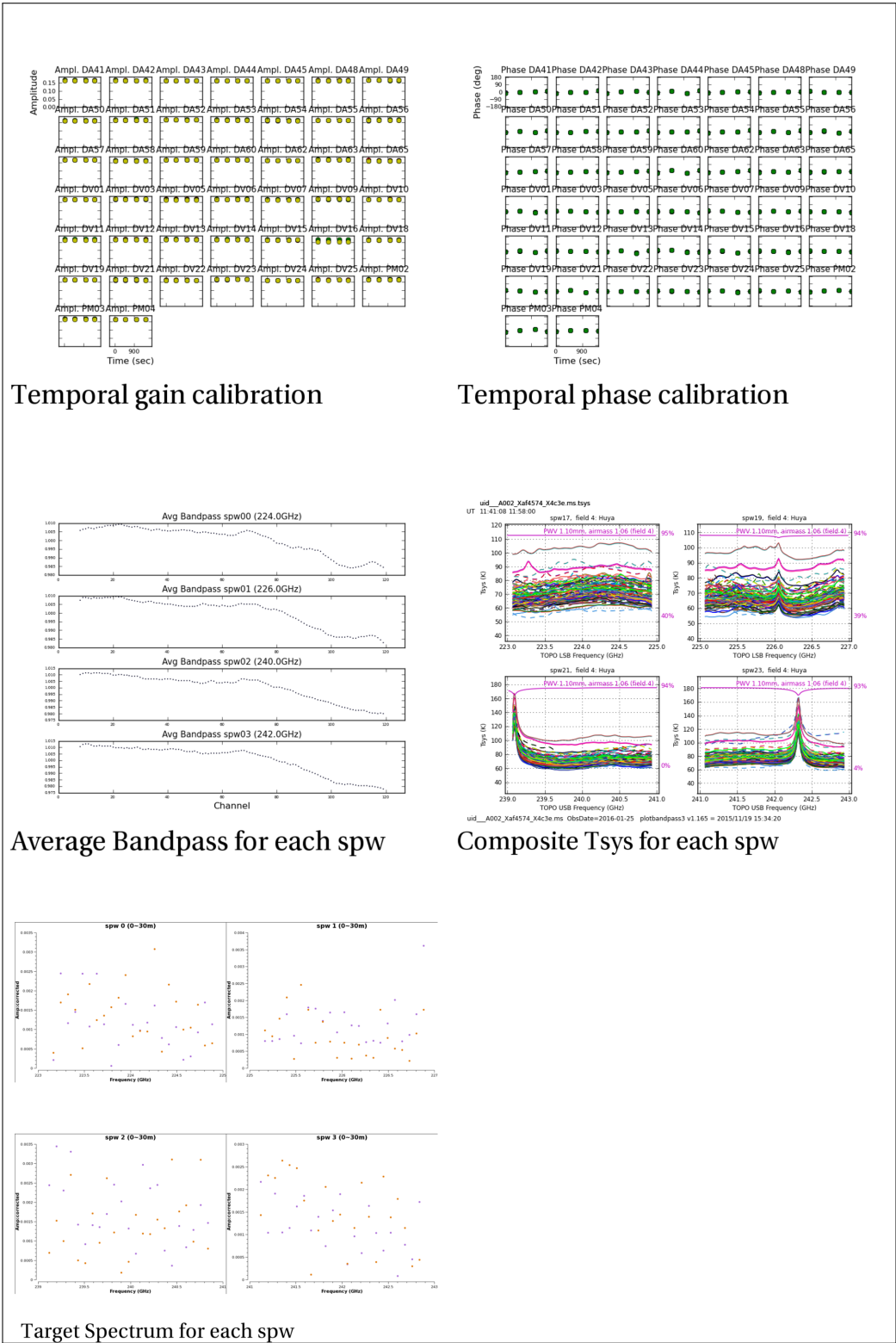


Figure 5.8: Part 2 of the manual calibration QA2 plots from SB Ephemeris\_c\_06\_TE (2015.1.01084.S).

be used for quick-look purposes. The spectra are created using data with  $uv$ -spacings between 0 and 30 m, and the data are averaged over 5 channels. While these settings may not be optimal for line detection in all datasets, the spectra can be still useful for identifying the presence of very strong line emission.

- **Flux calibration model and data:** These plots appearing at the top left of Figure 5.9 show the flux calibration model and the fully calibrated data for the flux calibrator. These data are colour-coded by spectral window. The data for resolved Solar System objects will follow a  $\text{sinc}^2$  function, while the data for quasars (which have coordinate-based names beginning with “J”) will look like flat lines. Two sets of offset colour-coded lines may be present in the plots of corrected data, particularly the quasars; these represent the data from the XX and YY polarisations. The corrected data may exhibit more scatter at longer baselines ( $>1$  km), in high frequency observations (Bands 8, 9, and 10) and in spectral windows with narrow bandwidths, but all of these effects are related to the lower signal-to-noise of the data. Severe outliers in the corrected data may be from antennas that were incorrectly calibrated, and these data may need to be flagged before the data are re-imaged.
- **Phase Calibrator amp/phase vs  $uv$ dist:** These phase calibrated data in the next set of plots in Figure 5.9 are sorted by spw and are colour coded by polarisation (XX and YY). These objects are quasars, which are point-like sources, so the amplitudes should be constant, although the amplitudes of the two polarisations may be slightly offset from each other. The phases should lie within a few degrees of  $0^\circ$ . The data for high frequency observations (Bands 8, 9, and 10) and for observations using spectral windows with narrow bandwidths will generally exhibit more scatter than normal, but this is a consequence of the typically lower signal-to-noise of the data, and these data are still usable for science. Problematic antennas will appear as outliers in these plots, and the data may need flagging before images are made.
- **Phase Calibrator amp/phase vs frequency:** The data in these plots on the left of the middle row of Figure 5.9 are also sorted by spectral window and colour coded by polarisation. These data can be used as an additional assessment of the bandpass calibration. The amplitudes should vary relatively smoothly versus frequency, although the curves for the two polarisations may be offset from each other. The phases should lie within a few degrees of  $0^\circ$ .
- **Target  $u$ - $v$  coverage:** The next plot in Figure 5.9 is useful for understanding the expected beam size and shape and the maximum recoverable scale in the data as well as other aspects of the sampling of the  $uv$ -plane.
- **Target image:** This is an aggregate bandwidth image created of one of the target fields. It can be useful for understanding the expected spatial extent of the emission in the imaged field, which may be useful when setting some of the parameters used by TCLEAN for imaging the data. Additionally, if the image contains any artefacts in the forms of ripple patterns across the image, this would indicate that some of the data were not correctly calibrated and may need to be flagged. Examples of such ripple patterns resulting from a problematic antenna are shown in Figures 6.3 and 6.4.
- **Target psf:** This image can be used to examine the size and shape of the beam directly as well as to understand the structure of the sidelobes.
- **Calibrator images:** From Cycle 6 onwards, images of the calibration sources were also added to these plots. Like the image of the target, the images of these calibration sources can be used to search for artefacts related to calibration issues.

The QA2 text file contains an overview of the observations, the source coordinates, spectral window information, antenna positions, antenna gain and bandpass tables and the flux densities calculated for the bandpass and phase calibrator sources. Figure 5.10 shows part of an example text file. The flux densities in this file can be compared with information from the ALMA calibrator Source Catalogue (Chapter 11) to check the quality of the amplitude calibration. The last part of the text file labelled “check of a target image and sensitivity” gives a good estimate for primary beam size. The aggregate bandwidth sensitivity estimation gives the expected sensitivity for an image based on a single image that includes all channels in all spectral windows; it can potentially be rescaled to estimate the sensitivities in images or individual channels within image cubes with smaller bandwidths, although the sensitivity could also be frequency-dependent.

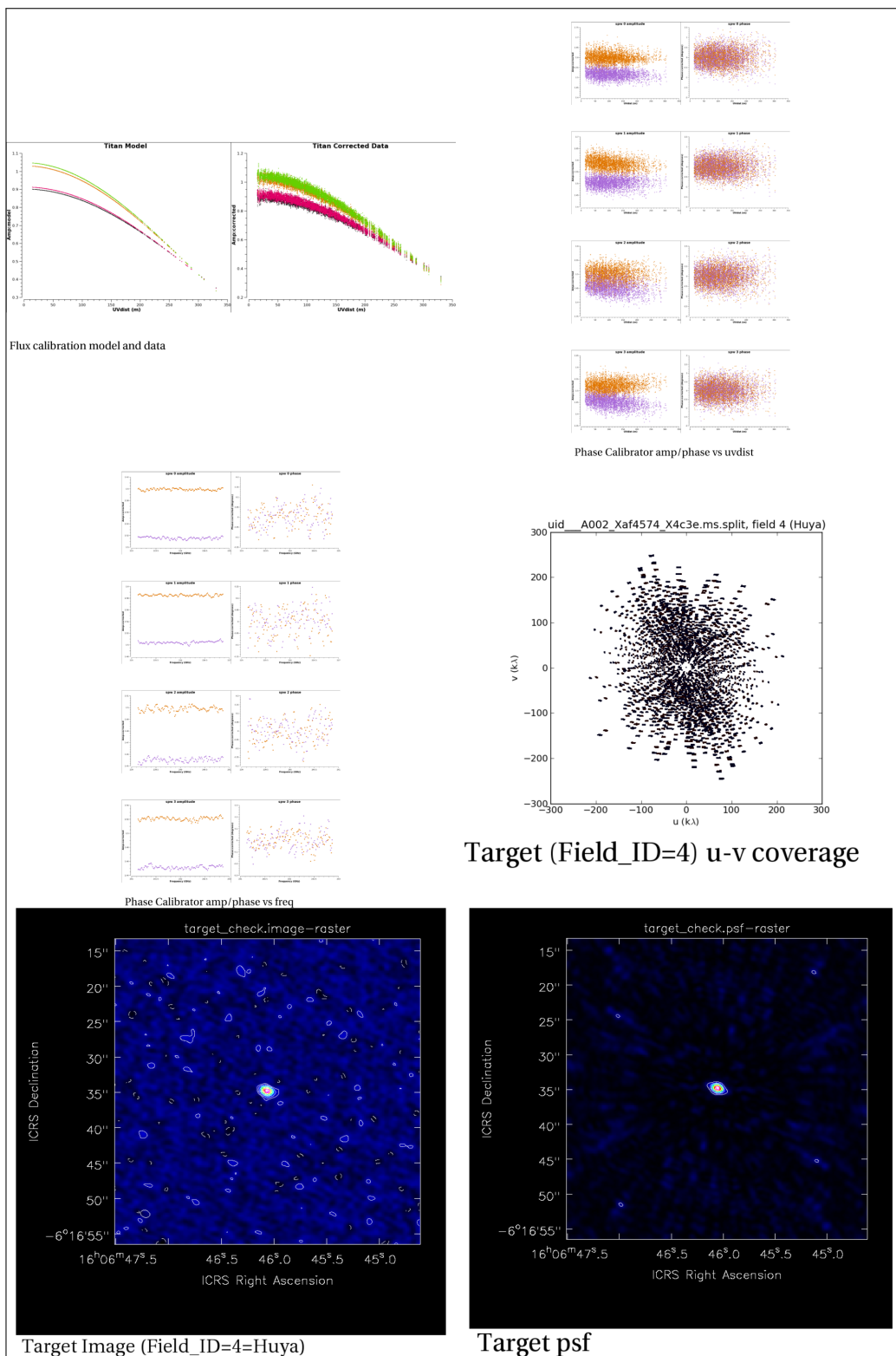


Figure 5.9: Part 3 of the manual calibration QA2 plots from SB Ephemer\_i\_c\_06\_TE (2015.1.01084.S).

```

*****
SUMMARY INFORMATION FOR uid__A002_Xaf4574_X4c3e.ms.split
Experiment Duration: 2016/01/25/11:24:24 to
                    2016/01/25/12:05:00

Processed from ms: uid__A002_Xaf4574_X4c3e.ms.split
Written to file:  NewListobs.txt

SCAN LISTING

Scan FdId srcId FieldName      StartTime   StopTime    Int(s) Elev ScanIntent
  2   0   0   J1550+0527      11:24:24.3 - 11:24:40.4  0.58  59.9 Cal atmos=Tsyst
  3   0   0   J1550+0527      11:24:54.0 - 11:29:56.4  2.02  60.1 Cal Bandpass
  5   2   2   Titan           11:33:36.1 - 11:33:52.2  0.58  69.1 Cal atmos=Tsyst
  6   2   2   Titan           11:34:03.1 - 11:36:34.3  2.02  69.4
  8   3   3   J1549+0237      11:40:01.2 - 11:40:17.3  0.58  63.7 Cal atmos=Tsyst
  9   3   3   J1549+0237      11:40:31.6 - 11:41:01.9  2.02  63.8 Cal Phase
 10   4   4   Huya            11:41:20.4 - 11:41:36.6  0.58  70.7 Cal atmos=Tsyst
 11   4   4   Huya            11:41:46.8 - 11:48:21.5  2.02  71.1 Obs Target
 12   3   3   J1549+0237      11:48:41.2 - 11:49:11.5  2.02  64.1 Cal Phase
 13   4   4   Huya            11:49:31.0 - 11:56:05.6  2.02  71.9 Obs Target
 14   3   3   J1549+0237      11:56:26.2 - 11:56:42.3  0.58  64.3 Cal atmos=Tsyst
 15   3   3   J1549+0237      11:57:22.5 - 11:57:52.7  2.02  64.3 Cal Phase
 16   4   4   Huya            11:58:12.3 - 11:58:28.4  0.58  72.4 Cal atmos=Tsyst
 17   4   4   Huya            11:58:39.2 - 12:04:13.3  2.02  72.6 Obs Target
 18   3   3   J1549+0237      12:04:31.0 - 12:05:01.2  2.02  64.4 Cal Phase

FIELD INFORMATION

Fid  Srd  Field              RA (J2000)    DEC          Fld Time  #Scans
      (min)
  0   0   J1550+0527         15:50:35.26900 +05:27.10.4480  5.31      2
  1   1   J1733-1304         17:33:02.70579 -13:04.49.5482  0.00      0
  2   2   Titan              00:00:00.00000 +00:00.00.0000  2.79      2
  3   3   J1549+0237         15:49:29.43684 +02:37.01.1633  2.55      6
  4   4   Huya                00:00:00.00000 +00:00.00.0000 19.26      5

FREQUENCY INFORMATION

spw  nchan  -----Frequencies (GHz)-----  --Channel Width--
      First      Last      Bandwidth      MHz      km/s      POLN
  0   128   224.992188  223.007812  2.000          -15.625  -20.81  ['XX', 'YY']
  1   128   226.992188  225.007812  2.000          -15.625  -20.63  ['XX', 'YY']
  2   128   239.007813  240.992188  2.000           15.625   19.59  ['XX', 'YY']
  3   128   241.007813  242.992188  2.000           15.625   19.43  ['XX', 'YY']

```

Figure 5.10: An example of the QA2 text file created during the manual calibration of the SB Ephemeris\_c\_06\_TE (2015.1.01084.S). This example shows the observation summary, source coordinates and spectral window information.



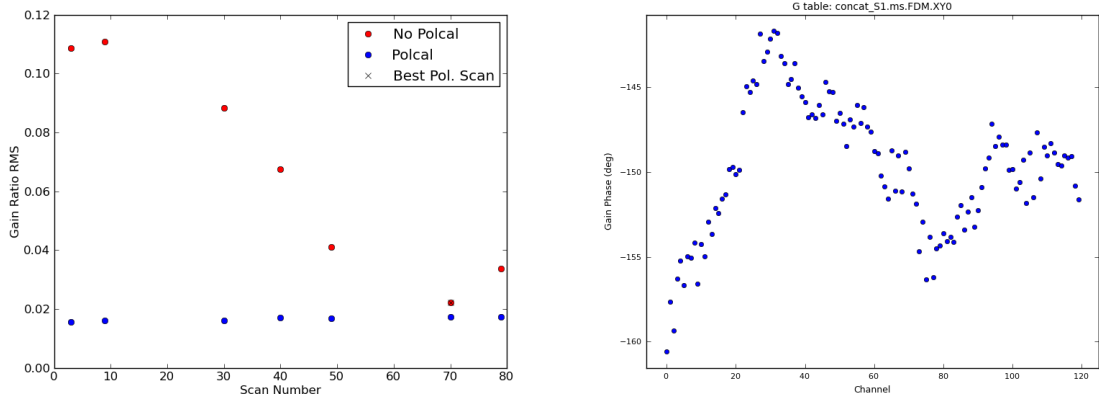


Figure 5.11: Examples of two polarisation calibration plots for SB BRI\_0952\_a\_07\_TM1 (2018.1.01536.S). The left panel shows the gain ratio RMS for the polarisation calibrator before and after applying the polarisation calibration, and the right panel shows the cross-phase solutions per channel.

If any problematic data are present, CASA task `plotms` can be used to inspect the datasets to identify the bad data (see also Section 5.4), and the CASA task `flagdata`<sup>5</sup> can be used to flag these data. If calibrator data are found to require additional flagging, it is recommended to insert the flagging commands in the manual calibration script `*scriptForCalibration.py` at the “Initial flagging” step and rerun the calibration script. In case it is only the science target that requires additional flagging, the the calibration script does not need to be re-run, and the `flagdata` command can be applied to the calibrated measurement set before imaging. Users are encouraged to contact their ARC or ARC node regarding a potential QA3 investigation (see Section 13.8) if they believe they have encounter a dataset with serious problems.

### 5.2.1 Polarisation data

Polarisation data are taken in sessions, with several EBs executed consecutively so as to cover a sufficient range of parallactic angles for accurate polarisation measurements. Most polarisation observations require two or three EBs. The total intensity calibration is performed for each EB. The EBs of each session are then concatenated, and the polarisation calibration is performed. Details about this are reported in the [CASA Guide for Polarisation](#). In the `qa` directory, plots for the polarisation calibration of each session are added to the QA2 reports for each EB. The polarisation calibration is identified by the addition of `*concat_*` to the filename followed by a number indicating the different sessions.

The following plots are delivered per each session.

- `*.Df0.plots`: A directory containing the real and imaginary parts of the  $D$ -terms versus channel for each antenna.
- `*.GainRatiosPol.png`: The gain ratio RMS for each scan of the polarisation calibrator before and after applying the polarisation calibration (see the left panel of Figure 5.11). The RMS after the polarisation calibration should generally be lower.
- `*.Gxyamp.png`: Plots showing the amplitudes of the gains  $G_{xy}$  of the different antennas for each spectral window. These gains should be  $\sim 1$ . Any data exhibiting a deviation  $>10\%$  from 1 should be checked.
- `*.GxyampRatio.png`: Plots showing, for each spectral window, the polarisation ratio  $X/Y$  of the gains  $G_{xy}$  of the different antennas. These gains should be  $\sim 1$  for each antenna. Any data exhibiting a deviation  $>10\%$  from 1 should be checked.

<sup>5</sup>The description for the `flagdata` task is available at <https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.flagging.flagdata.html>.

- `*.XY-CrossPhase.png` or `*.Xfparang.png`: (Note that these files were renamed from `*.XY-CrossPhase.png` to `*.Xfparang.png` late in Cycle 7.) These images show the XY phase solutions versus channel for each spectral window (see the right panel of Figure 5.11). These gains should vary smoothly with frequency. No phase jumps larger than  $5^\circ$  should be present from channel to channel.
- `*.Dterm_sky_imag_spw0-1.png`: (These plots are only produced up to late Cycle 7.) Plots of the imaginary parts of the  $D$ -terms versus frequency for 2 spectral windows. The values should be on the order of a few percent.
- `*.Dterm_sky_real_spw0-1.png`: (These plots are only produced up to late Cycle 7.) Plots of the real parts of the  $D$ -terms versus frequency. The values should be on the order of a few percent.
- `*.XY0_Amb-NoAmb.png`: (These plots are only produced up to late Cycle 7.) The XY phase solution versus frequency before and after solving for the ambiguity. These gains should vary smoothly with frequency. No phase jumps larger than  $5^\circ$  should be present from channel to channel.

## 5.2.2 Total power data

Total power data are delivered with very few diagnostic plots in the `qa` directory, although data from Cycle 5 onwards has been delivered with QA0 and QA2 reports (see Section 5.1). The list below describes the plots that may be present.

- `*.flagcmd.png`: This plot shows the results of a series of flagging steps at the beginning of the data calibrations that remove data known a priori to be unusable for science (such as data with pointing issues). The flagged data are plotted as a function of time. This is useful to check how much data was flagged and also to identify any antennas that were completely flagged at this stage and are therefore not usable for science.
- `*.sampling.png`: These files contain two sets of plots showing the locations observed in the sky during the EB (see Figure 5.12). The top plot in each file shows the angular offset from the reference coordinates as a function of time, while the bottom plot shows the offset in right ascension and declination. The bottom plot is particularly useful for identifying the off-target location used to measure the signal from the sky, which could be important when troubleshooting issues with removing this signal.

## 5.3 Inspecting pipeline calibrated data

For pipeline calibrated data, the WebLog provides the best overview of the data. The WebLog is a set of html pages that includes summaries of the observations, a list of the calibration steps, and sets of diagnostic plots. One WebLog is created for each Member OUS.

The WebLog for a Member OUS is contained within a `.tgz` file located within the `qa` directory. The file can be unpacked by using the `tar xzvf` command. In MacOS, the file can also be unpacked by double-clicking on it in Finder. The extracted files will appear in a directory beginning with the word `pipeline` followed by a time stamp. This directory has one subdirectory named `html` that contains the WebLog.

The report can be viewed by using a web browser<sup>6</sup> to open the file `index.html` within the `html` directory. The first page that is displayed is the Home page (see Figure 5.13), which provides an overview of the observation, a summary of the pipeline used to process the data, and a summary listing the measurement sets for each EB. If the pipeline performed imaging steps, then each measurement set is listed twice, and the second listing ending in “\_target” will contain only the target fields.

Links in the WebLog are light blue in colour. Clicking on these links will open new pages or plots. Clicking on one of the measurement sets in the list on the Home page will open a page with an overview of the observations

<sup>6</sup>Firefox is the officially supported browser to view the ALMA WebLog. In case of problems opening the WebLog in Firefox, consult the Knowledgebase article [What is the best way to view the weblog?](#).

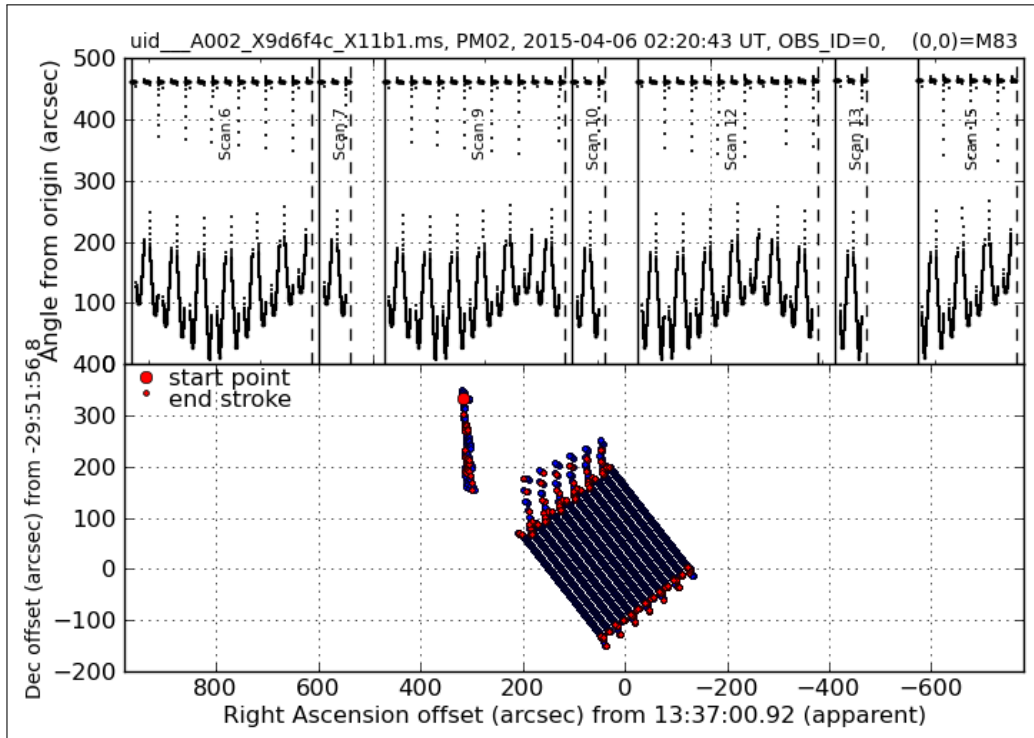


Figure 5.12: An example of a SD sampling plot from SB M83\_b\_03\_TP (2013.1.01312.S) showing the observed footprints on the sky including the off-target positions used for measuring the sky signal for calibration.

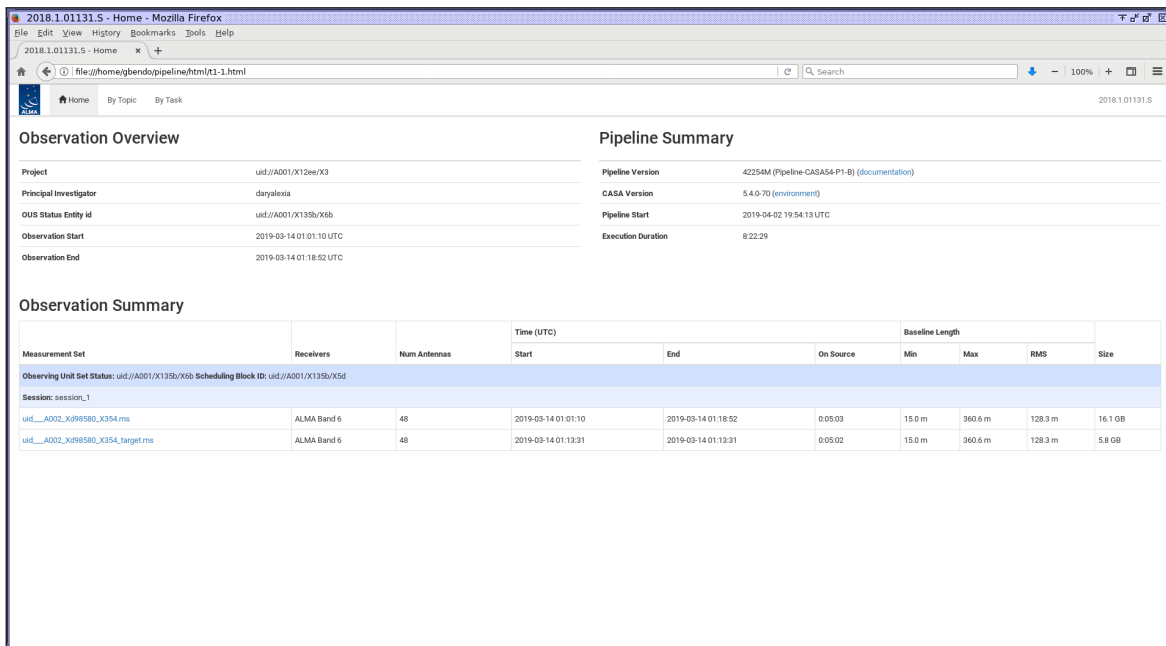


Figure 5.13: An example of a WebLog Home page for SB Z\_CMa\_a\_06\_TM1 (2018.1.01131.S).

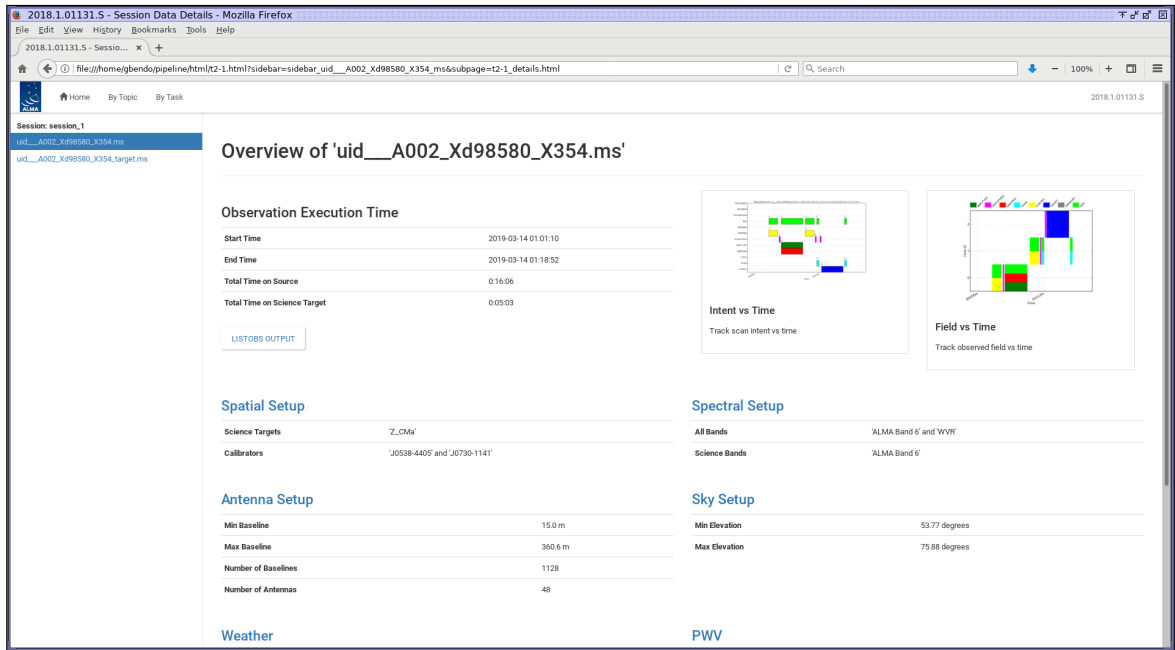


Figure 5.14: An example of a WebLog measurement set overview webpage for SB Z\_CMa\_a\_06\_TM1 (2018.1.01131.S).

QA Score	Colour	Comment
0.90-1.00	Green	No issues identified
0.66-0.90	Blue	No serious issues identified, but a note has been added
0.33-0.66	Yellow	QA warning triggered; carefully inspect the results for this stage
0.00-0.33	Red	Serious issue; may not meet quality standards

Table 5.1: WebLog QA scores.

(see Figure 5.14). This page has many sections labelled with light blue font that, when clicked on, lead to separate webpages displaying further useful tables and figures. Most of the information on this page provides useful summaries of the scheduling of the observations, the observed fields and the spectral window settings. The antenna setup and, to some degree, the sky setup are useful for understanding the *uv*-coverage, which will affect the angular resolution and maximum recoverable scale of the final images. The weather and precipitable water vapour (PWV) plots illustrate the observing conditions; if the weather changes notably during the observations, this could affect the quality of the phase and amplitude data.

Clicking on “By Topic” in the navigation bar at the top of the page leads to a summary of the warnings and other issues encountered during the pipeline calibration and imaging (see Figure 5.15). This is followed by a table showing the WebLog QA scores assigned to the results of a subset of the pipeline tasks. The QA scores range between 0 and 1 and are also coloured (in red, yellow, blue, green) following the definitions in Table 5.1.

More information on the scores can be found in the ALMA Pipeline User Guides in the Documentation Section of the ALMA Science Pipeline<sup>7</sup> for various CASA versions.

The “By Topic” page also contains Flagging Summaries in the form of tables showing the percentage of data flagged for each spectral window, each antenna and each target (see the lower part of Figure 5.15). One set of tables are listed for each EB. It is typical for a fraction of any data to be flagged, and data for some targets may be flagged more than others if that specific target was used for pointing or atmospheric measurements. However,

<sup>7</sup>The ALMA Science Pipeline website (for various CASA versions) is available at <https://almascience.org/processing/science-pipeline>.

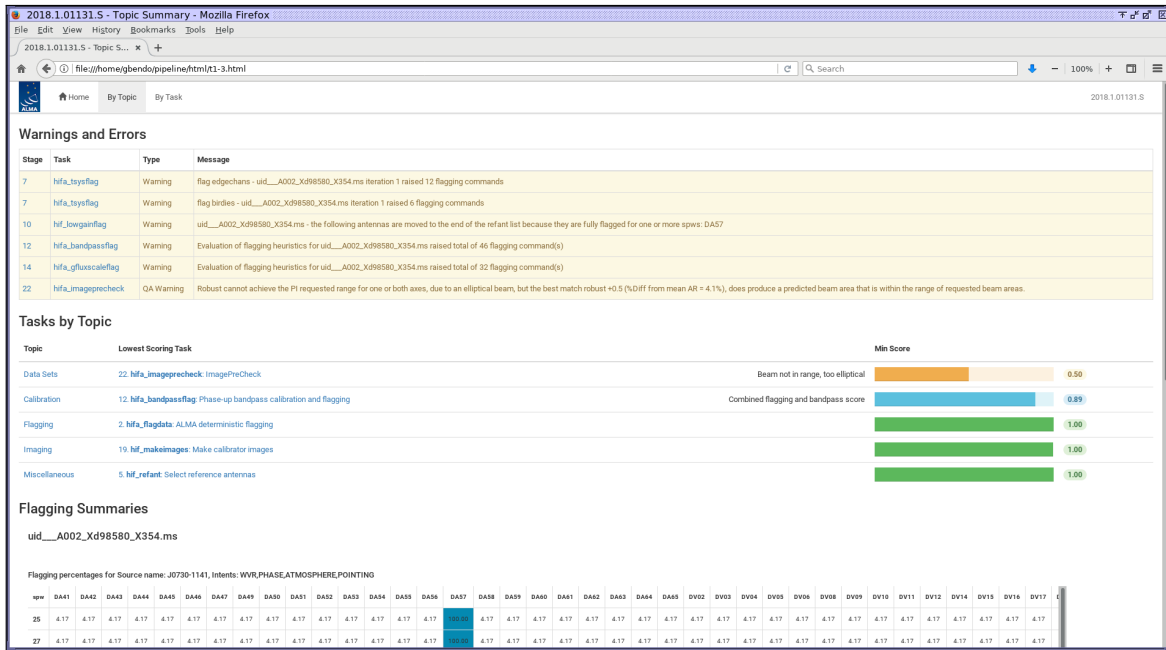


Figure 5.15: An example of a WebLog “By Topic” page for SB Z\_CMa\_a\_06\_TM1 (2018.1.01131.S).

if any antenna is not functioning properly in any spectral window, 100% of the data may be listed as flagged in these tables.

The “By Task” tab contains detailed diagnostic information from each calibration and imaging step, as shown in Figure 5.16. Clicking on the tab itself leads to a page with summaries of all of the task results. Each task is listed with a QA score, and tasks with warnings or errors will have small symbols (a question mark, an exclamation mark or an X) next to them that indicate that queries, warnings or errors were produced when these tasks were executed. Results from tasks with low QA scores (<0.66) or with warning or error messages should be investigated further, although these issues may not necessarily mean that the calibration failed or that the data are unusable for science. Log files as well as scripts for restoring or recalibrating the data are accessible at the bottom of the “By Task” summary page.

Clicking on the names of the individual tasks will lead to individual pages with diagnostic information for each task. The QA scores, input parameters and log files for each task are accessible at the bottom of each individual task’s page. When inspecting these data, be aware of the following issues.

- The two different polarisations will frequently yield slightly different amplitudes before and after calibration. This is normal.
- In plots of amplitude versus  $uv$ -distance, the data for Solar System objects should look like  $\text{sinc}^2$  functions, while the data for quasars will look flat. Data that deviate from these trends may need to be flagged. However, more scatter may be seen for baselines >1 km.

The tasks in the following list provide the most useful information for assessing the data. Note that this list is focused primarily on describing how to use the diagnostic output from a subset of the tasks performed by the pipeline; the pipeline also performs additional tasks that produce minimal or no diagnostic information. As the ALMA Science Pipeline is continuously being improved and refined, not all versions of the pipeline will produce the same plots. New tasks have been introduced over time, and some tasks have been altered. The ALMA Science Pipeline website provides User Guides and Reference Manuals with more detailed descriptions of all of these tasks for various pipeline versions.

- **hifa\_flagdata:** This task performs a series of flagging steps at the beginning of the pipeline (that remove data with pointing calibration failures, data with unwanted intents, shadowed antennas, autocorrelated

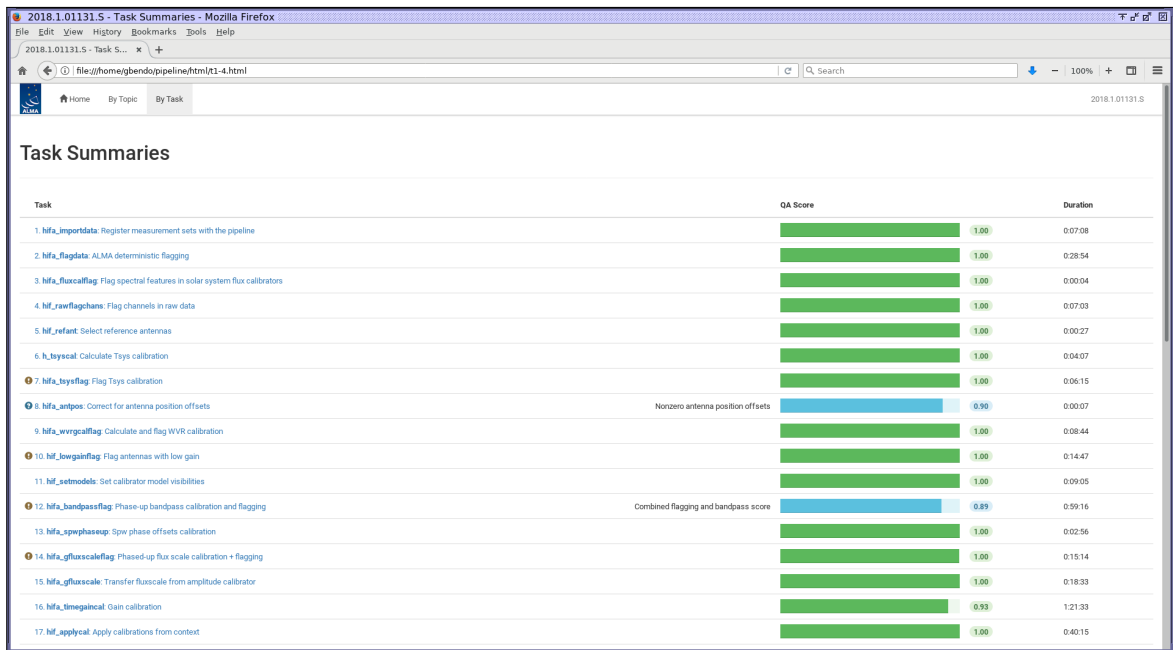


Figure 5.16: An example of a WebLog “By Task” page for SB Z\_CMa\_a\_06\_TM1 (2018.1.01131.S).

data, and edge channels in TDM data) and provides a summary of the data that were flagged during this step. Typical numbers in the total column should be  $\lesssim 20\%$ , although higher numbers may often be seen.

- **hif\_rawflagchans**: The diagnostic table from this task is similar to the information from **hifa\_flagdata**, but this task also provides useful graphs showing the data flagged by channel and baseline (accessible through a link labelled Display in the Flags table at the top of the page).
- **h\_tsyscal** or **hifa\_tsyscal** and **hifa\_tsysflag**: This task produces diagnostic plots showing the system temperature versus frequency for each antenna and each spectral window (as illustrated in Figure 5.17). Each plot also shows the atmospheric transmission. The system temperature will peak in locations with atmospheric features, and the frequencies of these peaks should be checked when imaging data to ensure that no artefacts related to these atmospheric features appear in the final images. Any other peaks in the system temperature data may be problematic data that need to be flagged before imaging, and any antennas with system temperatures significantly offset from the others may also need to be flagged.
- **hifa\_wvrflag**: The plots from this task show how the WVR corrections improved the RMS of the phase data. The first set of plots show the ratio of the RMS of the phases with and without the corrections and will indicate whether any WVR corrections were flagged when correcting the data. If that is so, then the corrections may be interpolated from other antennas as indicated in the table at the bottom of the output from the task. The second set of plots show how the deviation from the median phases of the scans changes when the WVR corrections are applied (see Figure 5.18). The third set of plots show both the ratio of the phase RMS without and with the WVR correction and the deviation from the median phases of the scans with and without the corrections, both of which are plotted as a function of the distance to the reference antenna. Both the second and third sets of plots should show general improvements in the deviation from the median phases and the phase RMS.
- **hif\_setmodels**: The output from this task (which was named **hif\_setjy** in Cycles 3 and 4) shows the flux densities used for the flux calibration source as well as plots of the flux density versus  $uv$ -distance. When Solar System objects are used as flux calibrators, the output from this task should be checked to ensure that it uses the most up-to-date version of the models for the amplitudes. The model used by the task is written in the **standard** parameter listed in the Input Parameters tab on the bottom of the page.
- **hifa\_bandpassflag** and **hifa\_bandpass**: Bandpass-related calibration plots may be produced by one or two tasks with these names depending on the cycle in which the data were acquired. The plots of the

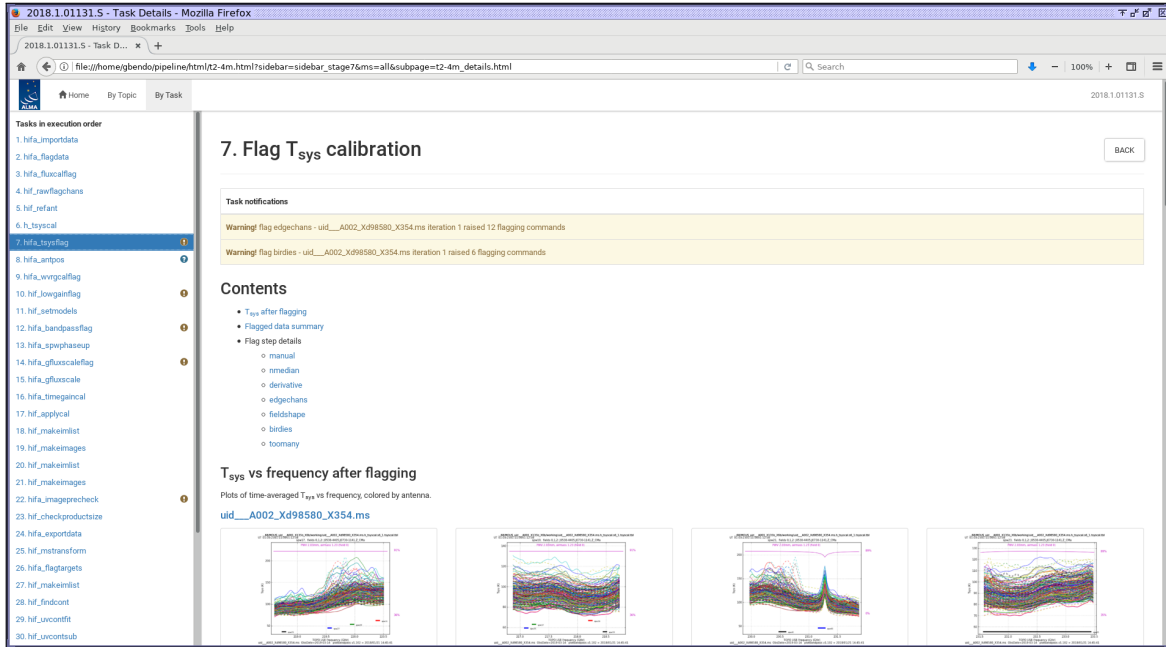


Figure 5.17: An example of a WebLog `hifa_tsysflag` task page for SB Z\_CMa\_a\_06\_TM1 (2018.1.01131.S) showing the diagnostic information of this calibration step.

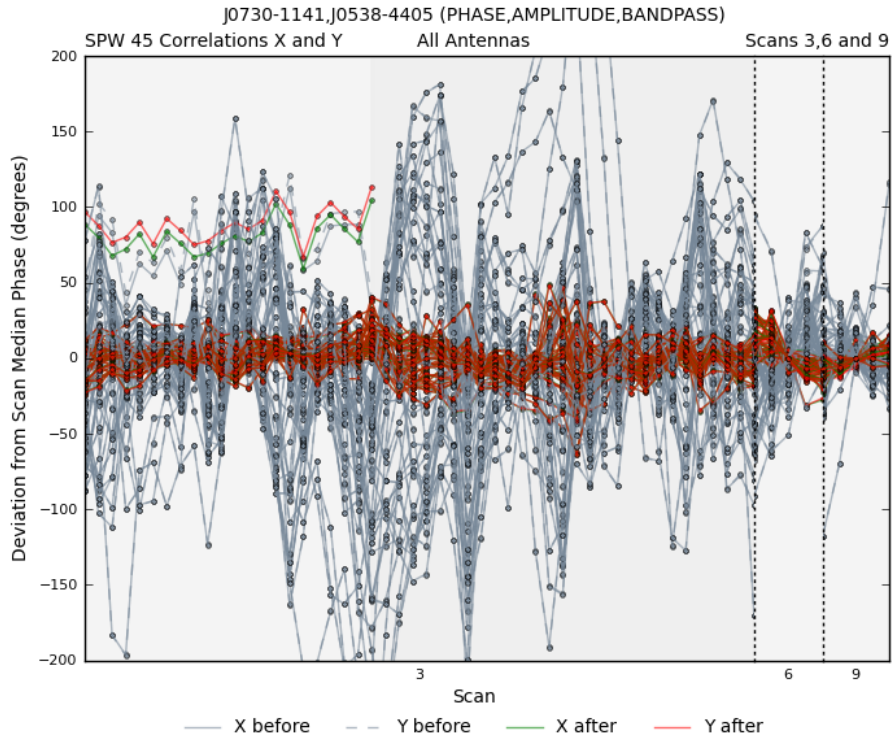


Figure 5.18: An example of the plot from the WebLog task `hifa_wvrgcalflag` showing the deviation from median phases versus time for the calibrator targets in SB Z\_CMa\_a\_06\_TM1 (2018.1.01131.S).

amplitudes and phases versus frequency should be checked to ensure that the corrections are relatively smooth, although atmospheric features may appear in some 12-m data. Any data with sharp dips (>10%) or periodic spikes may need to be flagged before imaging. The plots of the amplitudes versus time after flagging should be checked to ensure that the signal is stable over time, and the plots of amplitudes versus ( $u, v$ ) distance after flagging should be checked to ensure that the data vary relatively smoothly.

- **hifa\_gfluxscaleflag**: This task produces plots of amplitude versus time and  $uv$ -distance. No statistical outliers should appear in either plot. It may be appropriate to flag any data that appear as are outliers in these plots before making final images. The amplitudes should also remain constant versus time. If the amplitudes change significantly in one or more time intervals, those data should be flagged.
- **hifa\_gfluxscale**: This task produces a table and a plot showing the flux densities calculated for the phase calibrator, the check source (if present), and the bandpass calibrator (if it was not used as the phase calibrator). Also shown are the flux densities from the catalogue and the ratio of the observed to measured flux densities. If the catalogued flux densities are based on measurements within a couple of weeks of the observations (which can be determined by looking at the information from the `hifa_importdata` task), then these ratios would be expected to be close to 1. However, phase calibration sources and check sources are rarely observed, and because of the long-term variability of the sources, the ratios often differ significantly from 1.
- **hifa\_timegaincal**: Multiple plots are produced by this task. Some plots are based on integrated measurements for entire scans, while others show data for each integration within each scan. The plots of phase versus time as well as the phase offsets versus time should contain relatively smooth curves close to 0. The data in the plots of amplitude versus time should be close to constant, although the values from some antennas may differ slightly (especially for resolved sources like Solar System objects), and the data may show more scatter for longer baselines or fainter sources. Any statistical outliers in any of these plots may need to be flagged.
- **hif\_applycal**: The output from this task show the calibrated amplitudes plotted versus frequency, time,  $uv$ -distance, and antenna as well as the calibrated phases plotted versus frequency and time. The amplitudes should be relatively constant versus frequency for the calibration sources, although they will show more scatter at the locations of atmospheric spectral features. Astronomical spectral features may also be visible in plots for science targets. The amplitudes should be constant versus time for calibration sources. In plots of the amplitudes versus  $uv$ -distance for calibration sources, the data should either be constant or look like  $\text{sinc}^2$  functions depending on the source, while the data in the plots for science targets may differ from source to source and will depend on the spatial extent of the emission and the signal-to-noise of the source. The plots of the ratios of the observed to model flux densities versus antenna should be close to one, although the data may appear divided into two sets of data points if the values for each polarisation are slightly different. The phases for the calibration sources should be close to 0 as a function of time and frequency. Any deviating data may need to be flagged.
- **hifa\_renorm**<sup>8</sup>: This task was applied from Cycle 7 onwards. The output from this task indicates whether it was necessary to perform a correction for the normalization of the amplitudes caused by bright astronomical lines in the science target. When this correction is applied for any spectral window, the maximum renorm scale factors listed in the output from this task will be higher than 1.02, and the row for the spectral window where this is applicable will be highlighted in red. If the corrections are applied, the diagnostic plots of the autocorrelated data should be checked to ensure that the task was indeed triggered by spectral lines from the astronomical objects and not atmospheric features (which will match up with dips in the atmospheric transmission functions also shown in these plots). If atmospheric features triggered the renormalization corrections, then these corrections by this step should be disabled by setting the `apply` parameter for this task to `False` in the `pprequest.xml`. However, if the renormalization is needed, then this parameter should be set to `True`.
- **hif\_makeimages (cals)**: The page for the first instance of `hif_makeimages` shows the reconstructed beams for the calibration sources as well as the beam dimensions and position angle. The QA scores have the same meaning (and range) as in the WebLog calibration tasks, but for imaging these are assigned

---

<sup>8</sup>More information about the renormalization task can be found in the Knowledgebase article [What are the amplitude calibration issues caused by ALMA's normalization strategy?](#)



based on the ratio of the sensitivity measured in the final image and the theoretical RMS noise times a dynamic range correction factor that takes into account the dynamic range limitation. For more details please consult Chapters 9 and 10 of the ALMA Pipeline User's Guide of the relevant CASA version in Documentation Section of the ALMA Science Pipeline website.

- `hif_makeimages (checksrc)`: Long baseline observations will include observations of an additional point-like object called a check source that is calibrated in exactly the same way as the science targets and that is used to assess the quality of the calibration. When the observations include a check source, images of it will be created in a separate instance of `hif_makeimages`. The diagnostic output are the same as for the first instance of `hif_makeimages`, but it is particularly important here to check the images to ensure that the check source appears point-like. If the source does not appear point-like or if the images contain significant artefacts, then the corresponding data may also not be usable for creating final images of the science targets. (If the images of the check sources look problematic, then the data should be reviewed using the QA3 process as described in Section 13.8.)
- `hifa_imageprecheck`: This task will produce a table listing the predicted beam sizes and effective sensitivities that can be obtained by using different robust values in `tclean`, which can be useful when deciding on how to image data.
- `hif_findcont`: The diagnostic plots from this task show the channels and frequencies (in the LSRK reference frame) in each spectral window used for continuum subtraction by the pipeline. These channels could be used when first manually subtracting the continuum or creating continuum images<sup>9</sup>. This task generally performs very well in identifying the line-free continuum channels for many types of spectral line emission. However, very faint or broad spectral lines may be missed by this task and may only be identifiable after re-imaging the data using parameters that are optimised for detecting these lines. Consequently, the selection of continuum channels may need to be altered when manually imaging the data. Additionally, in cases where the spectra contain many lines and identifying the continuum is very difficult, it may be appropriate to use specialized continuum-identification software designed for such science targets.
- `hif_makeimages (mfs)`, `hif_makeimages (cont)`, `hif_makeimages (cube)`, and `hif_makeimages (cube_repBW)`: The output from the last instances of this task shows continuum images of the science targets based on data from each of the spectral windows, an image of the the aggregate bandwidth, and images of each image cube (which includes a moment 8 map (a maximum total intensity map) for each spectral window). The output also lists the achieved angular resolution and RMS noise for each image, which can be used as references before cleaning the data. Clicking on the link to view the other QA images will show the other outputs from the pipeline imaging. Note that the output from `hif_makeimages (cube_repBW)` will only be present if the PI specified a velocity resolution that involves averaging  $> 4$  spectral channels (for example, if the PI requested  $1 \text{ km s}^{-1}$  resolution for a dataset with  $< 0.25 \text{ km s}^{-1}$  channels).

The `hif_findcont` task and the last `hif_mageimages` tasks will be only present if the imaging pipeline was run on the dataset. In some cases, the results from the imaging pipeline tasks may be in a separated WebLog.

### 5.3.1 Polarisation data

Polarisation data for some Cycle 7 data and all polarisation data from subsequent cycles undergo pipeline assisted calibration. The total intensity part of the calibration of ALMA polarisation projects is done by the ALMA pipeline and will have an associated WebLog. The subsequent polarisation calibration is done manually by ALMA staff and will thus produce the same polarisation-specific diagnostic plots as described for the manually reduced polarisation projects in Section 5.2.1.

The WebLog of the total intensity calibration of a polarisation project is very similar to the WebLogs of other ALMA projects (as described in Section 5.3). The most notable difference is the inclusion of a polarisation calibrator. The task below is the only one that is unique to polarisation projects and that produce useful diagnostic information for assessing the data.

<sup>9</sup>A text file with the identified continuum frequency ranges, named `*cont.dat`, is available in the `calibration/` directory either as a single file or as a part of the `*.auxproducts.tgz` tarball.

- **hifa\_polcalflag**: This task produces plots of outliers in the polarisation calibrator visibility data. The amplitudes versus time and *uv*-distance should ideally have a constant profile or should vary smoothly. Any remaining outliers may need to be flagged after the pipeline calibration steps. Note that the XX and YY correlations, which are shown in different colours in the plots, may have a slight offset in amplitude.

### 5.3.2 Total power data

The pipeline calibrated total power (or single dish) data have a WebLog with the same structure as that of the interferometric ALMA data (discussed in Section 5.3), containing the Home page, the “By Topic” and “By Task” tabs.

Clicking on the name of the measurement set<sup>10</sup> in the Home page, displays the overview of those observations. Most of the information present here has the same structure as the interferometric data (see Section 5.3). Total power data have one unique section called Telescope Pointing details. For every target and antenna pair, two images are present: the telescope raster pattern showing the mapped sky region and the location of the OFF position with respect to the raster. In the second figure, flagged data (typically around the turning points of the raster) are indicated in grey. The coordinates of the OFF position can also be obtained from the Spatial Setup details or in the listobs output.

The “By Task” output contains the detailed diagnostic information for every calibration and imaging step, although the list of tasks differs significantly from those applied to the interferometric data. The following tasks provide the most useful information for assessing the data.

- **hsd\_flagdata**: This task shows a tabular overview of pipeline flags. Users should check that less than about 20% of the data for every EB (or measurement set) are flagged in the “Total” column. An EB with a high amount of flagging may suffer from non-flat baselines, most likely because the data were obtained under non-optimal weather conditions.
- **h\_tsyscal** and **hsd\_tsysflag**: These tasks produce diagnostic plots showing the system temperature versus frequency for each antenna and each spectral window of an EB as well as the atmospheric transmission. The system temperature will peak in locations with atmospheric features, and the frequencies of these peaks should be noted when imaging the data to ensure that no artefacts related to these atmospheric features are mistaken for target signal. From late Cycle 7 onwards, these features should be corrected by the **hsd\_atmcor** task (see below). Any other peaks in the system temperature data may need to be flagged before imaging, and any antennas with system temperatures significantly offset from the others may also need to be flagged.
- **hsd\_skycal**: Three sets of diagnostic plots are produced by this task. The first set show both the sky levels measured in OFF positions and the atmospheric transmission as functions of frequency. The spectra of the sky levels should be devoid of spectral lines. Spectral features coinciding with dips in the atmospheric transmission are likely to be artefacts. The second set of plots show the time variability of the target emission, which should slowly increase or decrease as a function of the OFF position elevation. The third set of plots show the elevation of both the OFF and ON positions and, more importantly, their separation are shown versus time. Elevation differences between the ON and OFF positions of  $\geq 3$  degrees may cause baseline calibration issues because of instabilities in the atmospheric calibration.
- **hsd\_k2jycal**: In the output from this task, all the EBs (or measurement sets) should have similar conversion factors. A conversion factor of unity means that no conversion was performed and that the map will be in units of Kelvin.
- **hsd\_applycal**: This task shows the science target spectra per spectral window *after* the calibration of the system temperature, the sky, and the conversion of Kelvin to Jansky as well as the atmospheric transmission. The spectra should be relatively smooth; ripple-like structures in the spectra could indicate issues with emission in the OFF position. Glitches may appear around atmospheric features; from late

<sup>10</sup>A measurement set may have multiple instances. Measurement sets with `*atmcor.atmtype1*` have the “atmtype1” atmospheric model applied, and measurements sets with `*_b*` have the baseline correction applied. For exploring the measurement set information, the `*.ms` is sufficient.

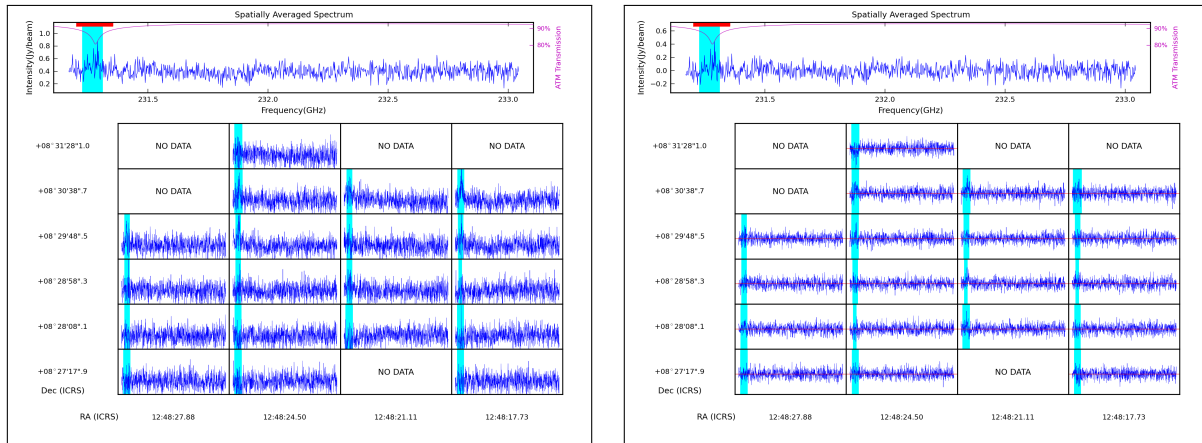


Figure 5.19: Examples of the `hsd_baseline` plots for SB NGC\_4698\_a\_06\_TP (2019.1.00763.L) showing spectral window 19 before (left panel) and after (right panel) the baseline subtraction.

Cycle 7 onwards, these will be corrected by `hsd_atmcor`. Note that the baseline calibration has not yet been applied here.

- `hsd_atmcor`: From late Cycle 7 onwards, this task applies an atmospheric correction to the target data. The plots of data with the atmospheric correction applied are shown here and can be compared with plots prior to the correction in `hsd_applycal`. Any positive or negative features near atmospheric lines should have been at least partially corrected.
- `hsd_baseline`<sup>11</sup>: This task is performed twice; users should focus on the output from the second instance, which follows line masking. The task produces plots gridded by right ascension and declination that show the spectral data before and after baseline subtraction (see Figure 5.19) and also provides maps that provide a first look at the spectral line detections in the target data. If the baseline subtraction has been performed well, the line-free data (shown in blue) in the spectra after baseline subtraction should have values near 0 Jy/beam, and the fitting function (shown in red) should be similar to the line-free data. Identified spectral line will be marked with red bars at the top of these plots, and masked data will be highlighted in cyan; it is important to ensure that all such features have been identified in the second instance of this step. From late Cycle 7 onwards, an additional set of plots will show the baseline flatness after baseline subtraction. Spectral lines should also be identified (with red bars) and flagged (as shown by regions highlighted in cyan) in these plots as well.
- `hsd_blflag`: Additional flagging commands should only comprise a few percent of the data in the EB (as indicated by the “Additional” column). Contact the your ARC or ARC node if an EB has additional flagging higher than this, as this may point to problematic data.
- `hsd_imaging`: This task creates spectral profile maps, channel maps, baseline RMS maps, maximum intensity maps and contamination plots. The spectral profile map should be inspected for baseline flatness, and atmospheric lines should not be misidentified as spectral features. If ripples are found in the baselines (possibly resulting from too large a distance between ON and OFF positions), then the baseline fitting may need to be repeated manually or on a narrower frequency interval around the line of interest. The RMS maps should be smooth without any bright peaks or holes. If this is not the case, the data may need additional baseline flagging. For very bright astronomical line emission, the target emission may be flagged in `hsd_blflag`; to restore this emission, `hsd_blflag` needs to be re-run with different parameters. The contamination plots (displayed from late Cycle 7 onwards), which come in sets of three, are important to assess whether the spectra are contaminated by emission in the OFF position. In each set, the left panel shows the peak signal-to-noise in the spectra at each location, the central panel shows the pixels with low signal-to-noise used to extract the spectra for blank parts of the sky, and the right panel shows in red a

<sup>11</sup>The QA score from this task may appear low. Separate scores are calculated for every EB, antenna, spectral window and correlation, but the lowest value is used as the QA score for the entire task. The list of individual scores, which can be found by clicking on Pipeline QA near the bottom of the page for this task, may be more useful for assessing the task’s performance.

line-free spectrum with, if present, atmospheric features. Any dips that do not coincide with atmospheric lines may come from spectral lines in the OFF position. Users are recommended to contact their ARC or ARC node if any of the above mentions problems occur.

More information on the tasks of the single dish pipeline is available in the Pipeline User Guides and Reference Manuals on the ALMA Science Pipeline website.

## 5.4 Inspecting calibrated data using CASA plotms

Calibrated measurement sets can be inspected using the CASA task `plotms`<sup>12</sup> regardless of whether they were manually or pipeline calibrated. This task can be used to visualize multiple quantities, including the amplitudes and phases of the visibility data, and the data can be colourized in multiple ways, including by antenna and by polarisation, to aid in identifying data that may need to be flagged. The plots discussed here are the the most important ones for assessing the quality of the calibration. The examples below are based on a dataset where J1517-2422 is the bandpass and flux calibrator and J1610-3958 is the phase calibrator.

Plotting the amplitudes versus *uv*-distances and colourizing the data by antenna allows for quickly identifying antennas with outlying amplitudes, particularly in data for the flux and bandpass calibrators. Solar System objects should have amplitudes that look like sinc<sup>2</sup> functions in these plots, while quasars, which are point-like, should have constant amplitudes. The example below uses the parameter `iteraxis='spw'` to iterate through plots for the data in each of the spectral windows specified by the `spw` parameter. The result for spectral window 31 is shown in Figure 5.20.

```
casa% plotms(vis = 'my_calibrated_data.ms',
             xaxis='uvdist',
             yaxis='amp',
             ydatacolumn='corrected',
             field='J1517-2422',
             spw='25,27,29,31',
             avgchannel='1e4',
             avgtime='1e6',
             iteraxis='spw',
             coloraxis='antenna1')
```

Plots of the calibrated amplitudes versus channel or frequency can be used to inspect the spectral profile of a calibrator, particularly the bandpass calibrator, for spectral lines. Atmospheric lines should not need to be flagged because the bandpass calibration will remove these feature, although the channels covering these atmospheric lines may exhibit a higher amount of scatter in amplitudes and phases because of the lower atmospheric transmission. Setting `showatm=True` will also display the atmospheric transmission curve in the plot as shown in Figure 5.21.

```
casa% plotms(vis = 'my_calibrated_data.ms',
             xaxis='channel',
             yaxis='amp',
             ydatacolumn='corrected',
             field='J1517-2422',
             spw='25,27,29,31',
             avgtime='1e6',
             iteraxis='spw',
             coloraxis='antenna1',
             showatm=True)
```

---

<sup>12</sup>The `plotms` task description is available at <https://casadocs.readthedocs.io/en/stable/api/tt/casaplotms.plotms.html>.

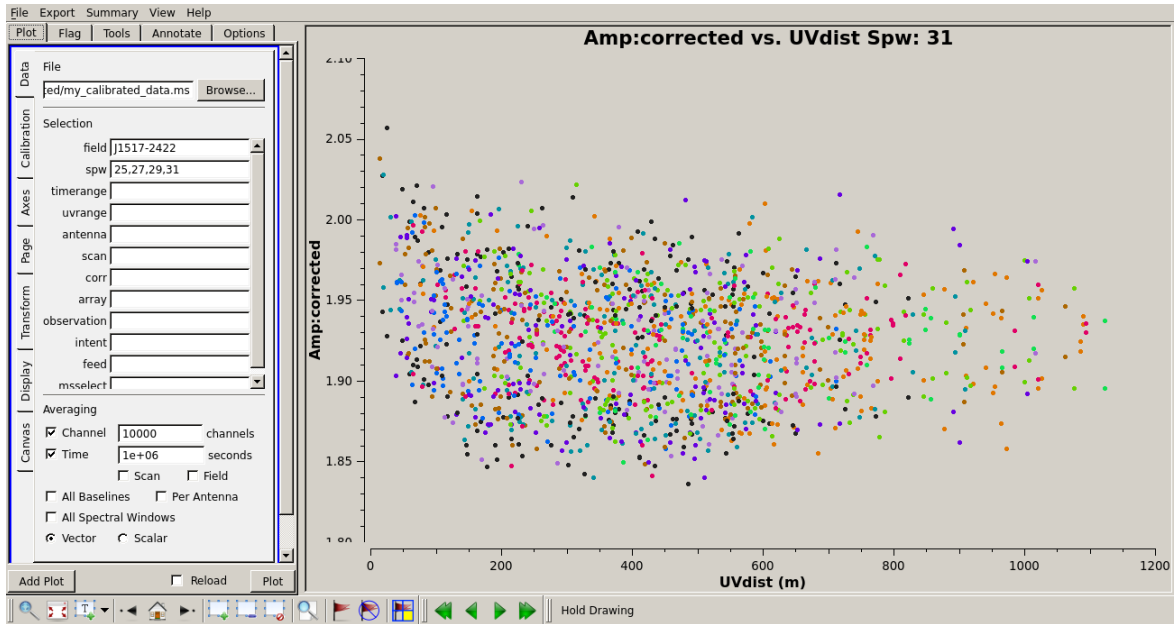


Figure 5.20: An example of the plotms GUI showing amplitude versus  $uv$ -distance within spectral window 31 for the bandpass calibrator J1517-2422.

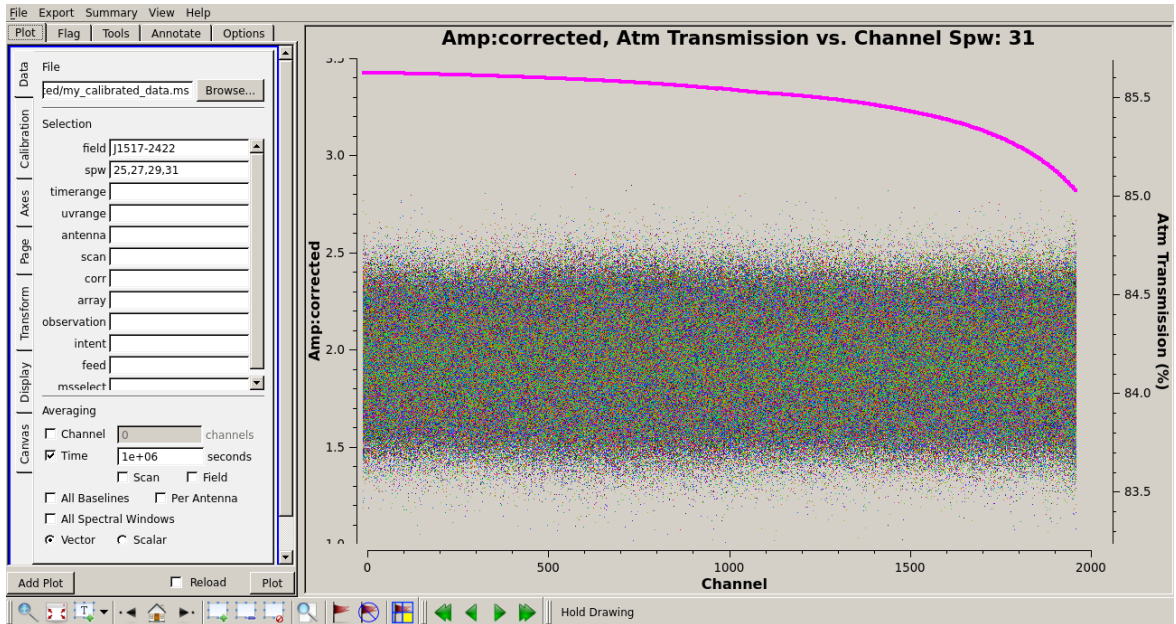


Figure 5.21: An example of the plotms GUI showing amplitude versus channel for the bandpass calibrator J1517-2422 for spectral window 31. The atmospheric transmission curve is shown in violet.

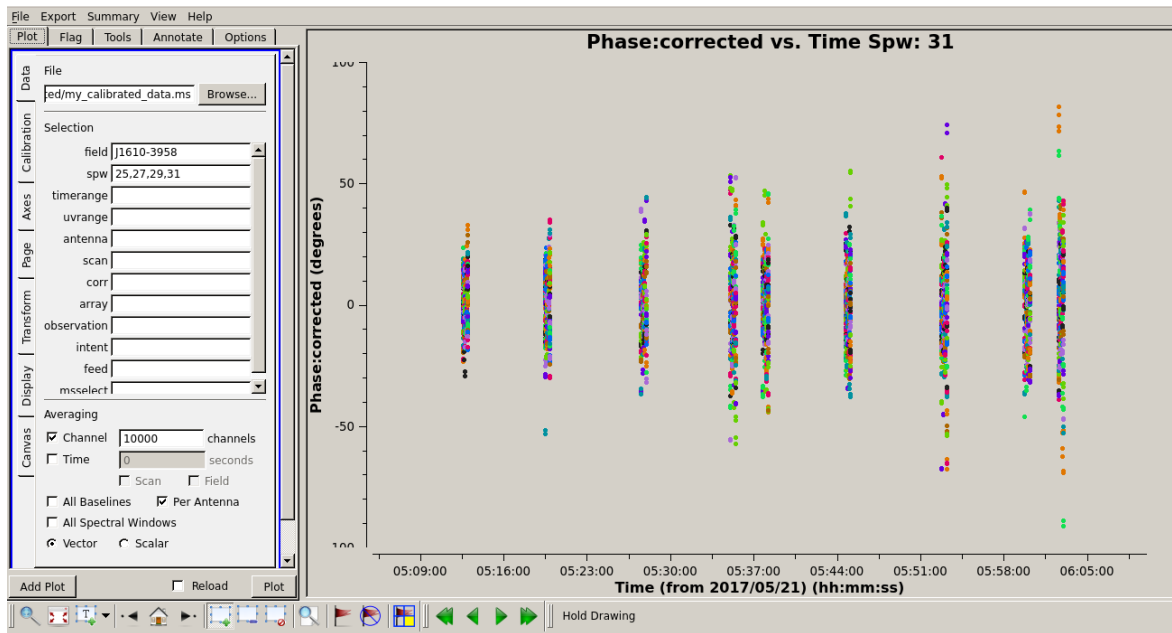


Figure 5.22: An example of the `plotms` GUI showing phase versus time for the phase calibrator J1610-3958 for spectral window 31.

Plotting the calibrated amplitudes or phases versus time or scan for the phase calibrator is important for investigating the phase and amplitude stability of the observations. Particular attention should be paid to whether the phases jump suddenly between scans or if any scan exhibits a large amount of scatter. The example below shows how to plot phase versus time, and Figure 5.22 shows one of the resulting plots.

```

casa% plotms(vis = 'my_calibrated_data.ms',
             xaxis='time',
             yaxis='phase',
             ydatacolumn='corrected',
             field='J1610-3958',
             spw='25,27,29,31',
             avgchannel='1e4',
             avgchannel=True,
             iteraxis='spw',
             coloraxis='antenna1')

```

Additional examples of inspecting ALMA data can be found in the [CASA guides for reducing ALMA Science Verification data](#).

## Chapter 6

# Checking the quality of the archived images and image cubes

### 6.1 Images available in the ASA

Each archived data set has a number of images and image cubes downloadable from the ASA in FITS format. For each science target, a pipeline imaged Member OUS will include three different types of images: continuum-subtracted spectral cubes for each spectral window covering the full channel range of that spectral window at the native spectral resolution; continuum images for each spectral window (based on the continuum channel selection); and continuum images combining all continuum channels from all spectral windows. The calibrators (phase calibrator, bandpass calibrator, and check source) will have only continuum images for each spectral window.

A manually imaged Member OUS will likely include a subset of these continuum images or spectral cubes depending on the science goals of the project. Hence, it is possible, for many manually imaged datasets, that some continuum images or spectral cubes will not be available in the ASA or that the spectral cubes will include only a subset of the channels covering the spectral line of interest. Also, for manually imaged Member OUSs, not all science targets nor all of the calibrators may have been imaged during QA2. ARI-L continuum images and spectral cubes (of the science targets and the calibrators) have been created for a significant number of Cycle 2-4 manually imaged Member OUSs and are available from the ASA as external product packages (see also Section 3.1.1).

The ALMA QA2 procedure checks that the calibration quality of the visibility data is good for science. The imaging products are considered to be reference images. They are particularly useful as quick-look images for performing initial rough analyses to determine whether spectral line or continuum emission is present or for assessing the flux density distribution in the map so as to understand the spatial scales of the observation. However, the images are not necessarily optimised for all science cases. When necessary, users can create ALMA images that are better tailored for their scientific purposes by, for example, changing the spectral resolution, applying a better source model or modifying the weighting scheme.

The following sections describe how to measure some of the most important properties of ALMA images.

### 6.2 Tools for inspecting ALMA imaging products

ALMA images and especially large spectral cubes can be up to tens of GB. CARTA<sup>1</sup> and the CASA Viewer are the two recommended viewers and analysis tools for inspecting ALMA images. In the future, CARTA is expected to fully replace the CASA Viewer for image visualisation and analysis. At the time of this writing, though, both viewers are still in use, and each one may have some advantages over the other. For more information, see the

---

<sup>1</sup>The CARTA website is accessible at <https://cartavis.github.io/>.

section on the remaining capability gaps on the CASA website.

Other FITS viewers can also be used to inspect and analyse ALMA image cubes that have been exported to FITS format.

## 6.3 Properties of ALMA imaging products

### 6.3.1 Spatial and spectral coverage

An interferometer is only sensitive to a subset of spatial scales. These angular scales can be derived from the  $uv$ -distribution using the formulas in Chapter 7 of the [ALMA Technical Handbook](#). The spatial scales can be, to some degree, assessed by looking at the images. Weather conditions may alter the expected outcome; for example, during sub-optimal weather conditions, long baselines are more strongly affected by phase stability issues than short baselines.

The maximum sky coverage of ALMA images is set by the primary beam, which is typically on the order of tens of arcseconds depending on the observed frequency (see Chapter 3 of the [ALMA Technical Handbook](#)). Hence, ALMA images may also include serendipitous targets. The sensitivity of an ALMA image, as shown by the primary beam response, is best near the centre of image and decreases towards the edges.

Archival spectral line cubes can be searched for specific spectral lines of interest. While weak spectral lines may not be easily visible in the  $uv$ -data, they might be more easily seen in the image cube. Spectral cubes may also be smoothed to increase the signal-to-noise per channel (but obviously at the cost of spectral resolution). Edge channels or channels with atmospheric lines may contain higher noise levels.

Generally, the continuum emission will have been subtracted from spectral cubes in the ASA. For pipeline imaged data, a `*.continuum.dat` or `*.cont.dat` file is present in the `calibration` directory listing the line-free frequency ranges in the LSRK reference frame. For manually imaged data, the line-free channels can be identified from the input of the CASA `uvcontsub` task in the `*scriptForImaging.py`.

### 6.3.2 Sensitivity

A comparison of the measured RMS noise with the expected theoretical sensitivity can be used to assess whether the image can be improved by either deeper cleaning or self-calibration.

An estimate of the RMS noise can be obtained from the primary beam corrected image (`*pbcor.fits`) by measuring the RMS noise in an empty part of sky near the center of the image. (This can be done using the CASA Viewer or CARTA.) In continuum-subtracted spectral cubes, these measurements should be done in line free channels. In images where the primary beam correction has been applied, the sensitivity at the edges of the images will always be worse because of the lower primary beam response, so the RMS noise near the edges of the images are not useful except for assessing the signal-to-noise of science targets in these regions.

The theoretically-expected sensitivity for pipeline imaged data can be obtained from the CASA WebLog. The theoretically-expected sensitivities and the measured sensitivities are given in each `hif_makeimages` step. Alternatively, the theoretical sensitivity per pointing can be calculated using the ALMA Sensitivity Calculator<sup>2</sup>. This requires inputting the appropriate bandwidth, frequency, time on source and polarisation. To make the estimate more realistic, values of the system temperature, PWV and target declination should be provided as well. For mosaics, which are observations done with multiple pointings sampling the sky area at the Nyquist spatial frequency, the RMS noise of the central pointings will be better by a factor 1.58 than a single pointing (as discussed in Chapter 7 of the [ALMA Technical Handbook](#)). See Chapter 5 for instructions on how to find this information in either the WebLog or the manual QA2 products. More information on the ALMA Sensitivity Calculator can be found in Chapter 9 of the [ALMA Technical Handbook](#).

If the measured image noise is significantly higher than the theoretically expected noise, this could indicate one of several problems. It is possible that the target was not cleaned deeply enough and the image residuals still contain target emission. Another possibility is that the parameters used for cleaning (such as the robust

---

<sup>2</sup>The ALMA Sensitivity Calculator is located at <https://almascience.org/proposing/sensitivity-calculator>.



factor) were not set to optimize the RMS noise levels. In these situations, carefully re-imaging the data may yield better results.

For images with a very bright source, the RMS will be limited by the dynamic range (the ratio of the peak intensity over the RMS noise) of the image, and self-calibration will be necessary to decrease the RMS noise (see Section 8.6). Typically, ALMA images are expected to reach a dynamic range of about 100 in Bands 3–8 and about 50 in Bands 9 and 10. See the Knowledgebase article [What is meant by imaging dynamic range?](#) for more information.

The theoretical sensitivity may not always be reached for multiple reasons, including but not limited to sparse  $(u,v)$  coverage, a bright source just outside of the field of view, or bad weather conditions such as high wind speeds during the observation.

### 6.3.3 Beam sizes and $uv$ -coverage

The synthesized beam represents the smallest angular scale of the image and hence the angular resolution of the image. Beam size and beam orientation information for ALMA images are stored in the image FITS headers as BMAJ, BMIN, and BPA keywords. Images of point-like sources, such as most calibrators, have the shape of the image beam.

The size and shape of the beam are determined by the  $uv$ -coverage of the data, which is a function of the antenna positions as projected onto the sky during the observations. The  $uv$ -coverage is therefore affected by the spacing of the antennas, the elevation of the targets observed and the period of time over which the field is observed. The spatial extent of the antennas is inversely proportional to the size of the beam, and if the  $uv$ -coverage is elongated in one dimension (because of either a non-axisymmetric antenna layout or projection effects), then the beam will appear elongated in the perpendicular dimension.

The maximum recoverable scale (MRS) is determined by the shortest baselines in the  $uv$ -coverage. Visibilities at locations on the  $uv$ -plane at or near its origin are not sampled. Images from interferometric data will only contain structures smaller than the MRS; larger structures are not reproduced in the final images. Negative artefacts around source emission may be present if structures larger than the MRS are present. For more information, consult Chapters 3 and 7 of the [ALMA Technical Handbook](#).

The  $uv$ -coverage of a target source determines how well `tclean` can reconstruct the true flux distribution. Snap-shot observations can be functional for a point-like target, but for a target with a complex morphology a well-sampled  $uv$ -plane is preferable. Figure 6.1, taken from the [ALMA Technical Handbook](#), shows ALMA simulations of the observations of galaxy M51 with different 12-m Array configurations ranging from compact (C43-1) to extended (C43-6). The differences in the  $uv$ -coverage cause changes in the level of detail, the angular resolution and the recovery of extended structures in these images.

For pipeline calibrated data, the  $uv$ -coverage can be examined in the WebLog in the “Observation Summary” for each measurement set by clicking on the “Antenna Setup” link. Additionally, plots of the amplitude versus  $uv$ -distance can be viewed in the output from `hif_applycal` under the “By Task” tab. The robust factor applied by the pipeline when imaging the data as well as the synthesized beam size can be found under “By Task” in the output from `hifa_imageprecheck`.

For manually calibrated data, plots of the  $uv$ -coverage can be found in the the QA2 results, and the weighting parameters for `tclean` commands can be found in the `*scriptForImaging.py` within the `script` directory. Alternatively, a plot of the antenna distribution can be generated with CASA task `plotants`, or the  $uv$ -coverage can be plotted using `plotms` as shown below. Figure 6.2 shows an example of antenna positions plot generated by the `plotants` task.

```
casa% plotants(vis='my_calibrated_data.ms',
              figfile='antenna_plot.png')
casa% plotms(vis='my_calibrated_data.ms',
            xaxis='U',
            yaxis='V',
            coloraxis='field')
```

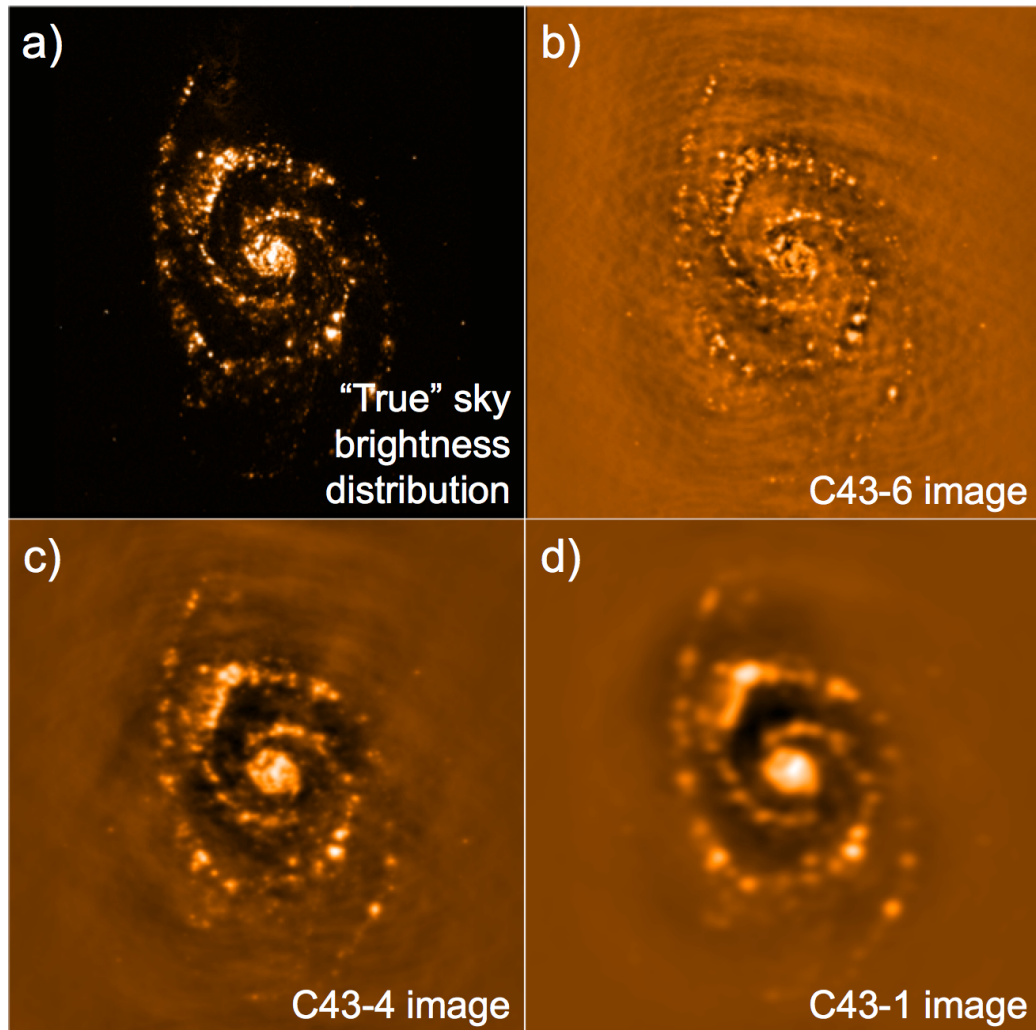


Figure 6.1: ALMA `simobserve` simulations showing how  $uv$ -coverage affects the angular resolution and recovery of extended structure. Panel a shows the “true” sky brightness, while panels b, c and d show images from the extended 12-m Array configuration C43-6, the intermediate configuration C46-3, and the compact configuration C43-1. (Image credit: [ALMA Technical Handbook](#).)

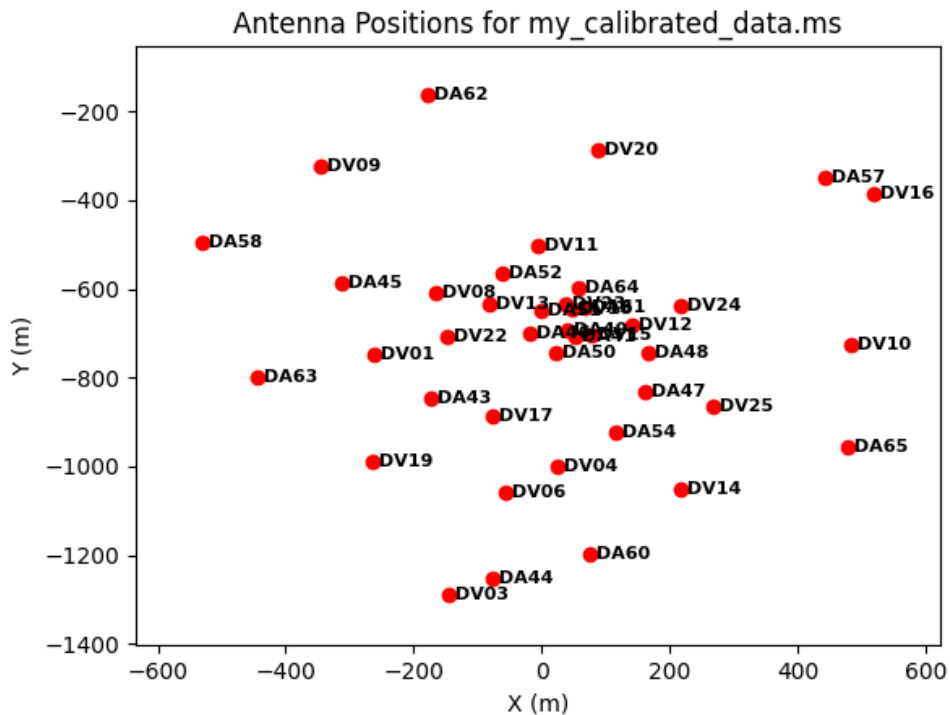


Figure 6.2: An example of the antenna position plot made with the `plotants` task

Different beam sizes may be obtained when re-imaging the data by varying the robust factor or using natural or uniform weighting schemes, which will apply different weights to data in  $uv$ -space. In natural weighting, visibility weights are not changed, but because interferometric observations generally have more shorter baselines than longer baselines, images created using natural weighting tend to produce large scale structures more accurately and also tend to produce images with larger beams. In uniform weighting, the weights are adjusted to account for the relative sparseness of long baseline data, which results in images that tend to reproduce emission from compact structures, and the beams also tend to be smaller. Naturally weighted images will generally have a better RMS noise than uniformly weighted images.

To accurately sample the beam shape, it is important to set pixel sizes that are smaller than the beam size in the final images. Ideally, the full width at half maximum (FWHM) of the beam along both its major and its minor axes should be sampled by five to eight pixels. For manually calibrated data from Cycle 3 and earlier, the beam sizes are listed in the `README` file, while for manually calibrated data from Cycle 4 and later, the beams are listed in the QA2 Report PDF file. For pipeline processed data, beam sizes are listed in the output from any `hif_makeimages` task, which can be found under the “By Task” tab. Alternately, the expected beam sizes can be calculated using the  $uv$ -distances along with the appropriate equations in Chapter 7 of the [ALMA Technical Handbook](#).

For manually calibrated data from Cycle 4 and later, the MRS is listed in the QA2 Report PDF file, but this is not listed for manually calibrated data from Cycle 3 or earlier or for pipeline calibrated data. In these situations, the MRS can be derived using the shortest  $uv$ -distance values and equations from Chapter 7 of the [ALMA Technical Handbook](#).

More details on the angular resolution and visibility weighting schemes are given in Section 8.4.

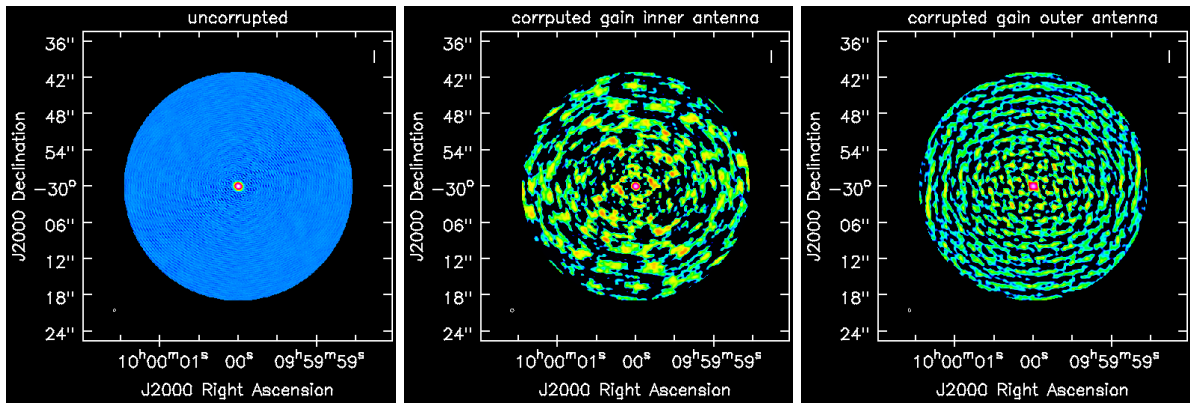


Figure 6.3: Three images of a Gaussian source model produced with the same `tclean` command and shown with the same flux scaling. The images are created from the same simulated measurement set, but the amplitude gain has been corrupted for one of the inner antennas in the centre image, and the amplitude gain has been corrupted for one of the outer antennas in the right image. No primary beam correction has been applied.

### 6.3.4 Residual features and artefacts

Although archive images are made from science-ready calibrated data, it is still possible that the image may contain artefacts from minor calibration issues such as problems with the phase or amplitude calibration of one problematic antenna. Such issues may be resolved by flagging the problematic data without the need to re-calibrate the data. A number of examples of possible calibration artefacts based on simulated data are presented below.

Ideally, the noise in the image should appear random, although it will increase towards the edges of primary beam-corrected images. A recurring ripple or grid-like pattern could be indicative of issues with an individual antenna, which may mean that additional flagging needs to be applied before images are made. Figure 6.3 shows three simulated images (created using `simobserve` in the CASA toolkit; see Section 12.2) of a  $0.5 \times 0.4$  arcsecond Gaussian source as it would appear in observations with the ALMA 12-m Array in configuration-4. The left image is an observation based on the array performing nominally. The centre image was made of the same measurement set after corrupting the gain amplitude of an inner antenna, while the right image was created after corrupting the gain amplitude of an outer antenna. Note that the spatial frequency of the noise is lower when an inner antenna is corrupted and higher when an outer antenna is corrupted.

Additional substructures around bright sources, particularly symmetrical patterns, may be indicative of calibration errors that can be corrected using self-calibration (see Section 8.6). This substructure will otherwise limit the dynamical range (the ratio between the peak of the emission and the RMS noise). Figure 6.3 shows again three simulated and imaged datasets of a relatively compact Gaussian source of  $0.06 \times 0.03$  arcsec. The left image is again based on the array performing ideally, while the center image has one antenna with corrupted amplitude gains, and the right image has one antenna with corrupted phase gains. Bad amplitude gains give raise to symmetrical substructures, while badly calibrated phases are visible as asymmetric structures.

Images with emission that is larger than the maximum recoverable scale may contain negative artefacts around the emission, as can be seen, for example, the extended 12-m Array configuration (C43-6) simulation of spiral galaxy M51 in Figure 6.1, where negative bowls have appeared around regions with bright emission, and a wave pattern is seen across the map. The only way to correct this is to combine these data with either interferometric data with shorter baselines or single dish data, as discussed in Chapter 9.

Non-optimal masks applied during the cleaning process may cause artefacts. If the target mask used in `tclean` has no target emission, `tclean` will interpret the background emission as source emission and will start “cleaning” it. This will produce negative bowl-like artefacts in the images. Careful target masking is therefore important. For spectral line masking, where the spectral cube may have many channels with and without line emission, it is important to mask the emission region only in the channels with line emission and not to apply one mask automatically to all channels. Automasking in `tclean` is possible for complicated line profiles or profiles

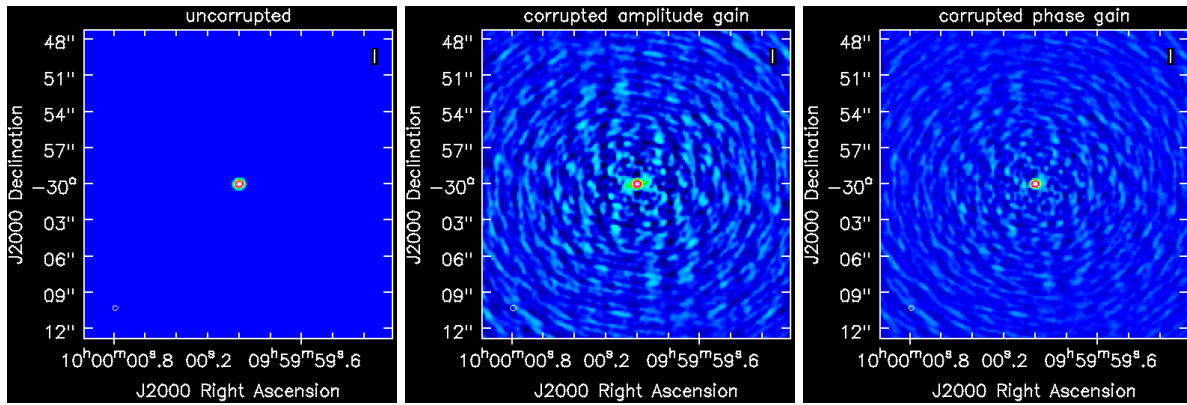


Figure 6.4: Three images of a Gaussian source model produced with the same `tclean` command and shown with the same flux scaling. The images are created from the same simulated measurement set, but antenna A25 has corrupted amplitude gains in the centre image and corrupted phase gains in the right image. No primary beam correction has been applied.

present over many channels where it may be difficult to set the mask manually (see the [CASA Automasking Guide](#)).



# Chapter 7

## Requesting and recreating calibrated visibilities for archival data

For users to produce their own images or perform analyses on the visibility data, they need to obtain the calibrated visibilities. Currently, the ALMA Science Archive stores the uncalibrated raw visibilities with a calibration script and calibration tables that can be used to create the calibrated visibility data. However, the calibration of ALMA data may require significant computational resources, so any ALMA users who encounter such problems can contact their ARCs through the Helpdesk to either request calibrated visibility data (for either publicly-available data or their proprietary data (for PIs and delegated Co-Is) or to otherwise get help with the data. Both of these options are discussed in this chapter.

### 7.1 Requesting calibrated visibilities for archival data

For public ALMA datasets, archive miners may request directly calibrated visibility data by writing a Helpdesk ticket to their ARCs. For proprietary ALMA datasets, PIs and their delegees may request their calibrated proprietary visibility data (and the data access rights will be checked by the ARC) . See Chapter 14 for information about how to open a Helpdesk ticket. In the EA and NA ARCs, the PI has the option of downloading the calibrated data for their project when their data are delivered. More information about these policies is available in the Knowledgebase article [How do I obtain a file of calibrated visibilities \(“measurement set”\) for ALMA data?](#).

ARI-L, which provides pipeline images of manually calibrated datasets from Cycles 2-4, can also release the calibrated datasets on request. These visibility data will be available for at least 3 years after the project’s end in June 2022. The procedure for obtaining calibrated data can be found on the ARI-L website<sup>1</sup>.

ALMA Large Programs may publicly release their calibrated data in addition to the recommended primary beam corrected images.

### 7.2 Recreating calibrated visibilities from/for archival data

#### 7.2.1 Initial inspection of the data quality and calibration procedure

For all ALMA Member OUSs in the ASA, the raw data are available in ASDM format alongside the calibration scripts for creating the calibrated measurement sets for the Member OUS<sup>2</sup>.

Before starting the calibration procedure, it is good practice to inspect the contents of the `qa` directory, the WebLog or the manual QA2 report as discussed in Chapter 5. The extensive quality assurance (QA2) process

---

<sup>1</sup>The ARI-L website is located at <https://almascience.org/alma-data/aril>.

<sup>2</sup>Cycle 0 ALMA Member OUSs are an exception. For that Cycle, fully calibrated or a priori calibrated data will be available instead of raw data.

ensures that the ALMA calibration will be of good quality for the large majority of archival data sets. Once the calibration is deemed satisfactory and appropriate for the intended scientific use, the calibration scripts in the `script` directory can be used to recreate the calibrated measurement sets.

The calibration of both pipeline calibrated and manually calibrated ALMA data is performed by running `*.scriptForPI.py` in the `script` directory. To do this, the correct version of CASA must be used, and the pipeline mode should be specified if appropriate.

The first step is to identify whether the Member OUS was calibrated manually or by the calibration pipeline. If the data were pipeline calibrated, the `script` directory will contain files named `casa_pipescript.py`, `casa_piperestorescript.py` and `PPR*.xml`. These files are not present for manually calibrated datasets. Instead, the `script` directory for manually calibrated data will contain a `*.ms.scriptForCalibration.py` file for every EB. The `*.ms.scriptForCalibration.py` scripts are called by `*.scriptForPI.py` and should not be run separately by hand<sup>3</sup>

The second step is to identify the CASA version that was used the calibration, as the calibration scripts are optimised for specific versions of CASA and may not work well or may produce errors when well executed using other versions of CASA. For both pipeline and manually calibrated Member OUSs from Cycle 5 onwards, the CASA version can be found in the QA2 pdf report in the `qa` directory (see Section 5.1). For pipeline calibrated data, the CASA version is also stated on the start page of the WebLog (see Section 5.3), while for most manually calibrated data, the Member OUS README files contain the CASA version information. See Section 13.2 for more details on finding the CASA version.

## 7.2.2 Applying the ALMA calibration tables

### Applying the pipeline calibration

To apply the pipeline calibration, go to the `script` of the ALMA Member OUS. Launch the appropriate version of CASA with the pipeline extension by the following command.

```
> casa --pipeline
```

Then, in CASA execute the `*.scriptForPI.py` file.

```
casa% execfile('member.uid__A001_X1296_Xe30.scriptForPI.py')
```

or alternatively

```
casa% execfile('member.uid__A001_X1296_Xe30.casa_pipescript.py')
```

### Applying the manual calibration

To apply the manual calibration, go to the `script` directory and start CASA without calling any extensions.

```
> casa
```

---

<sup>3</sup>The `script` directories for Member OUSs with multiple EBs from Cycles 0-2 also contain `scriptForFluxCalibration.py` files. These files performed flux equalisation steps that were considered useful during the earlier cycles. If these files are present, they will also be executed by `*.scriptForPI.py`. The `scriptForFluxCalibration.py` files may also contain special calibration or flagging steps that may be useful to refer to for more information about the calibration.



Then, in CASA execute the `*.scriptForPI.py` file

```
casa% execfile('member.uid__A001_X885_X1f6.scriptForPI.py')
```

### 7.2.3 Files created by the calibration script

When the calibration script is launched, it creates a new directory named `calibrated` in which the results are placed, including a `working` directory containing the calibration tables and one or more `pipeline*` directories containing regenerated WebLogs<sup>4</sup>.

For every manually calibrated EB, the script creates a calibrated measurement set called `*.ms.split.cal` containing the calibrated data<sup>5</sup>. The visibility data in the measurement set is organised into data columns as explained in Section 8.1. The `*.ms.split.cal` datasets contain columns labelled `DATA` that correspond to the calibrated data.

For every pipeline calibrated EB, the scripts create a calibrated measurement set that ends in `*.ms` and that contains the calibrated data in a column labelled `CORRECTED_DATA`. For imaging these data, it may be convenient to split the `CORRECTED_DATA` column into a new measurement set using either `mstransform` or `split` in CASA. These tasks can also be used to select a subset of the data (such as only the science spws or specific targets of interest), as shown in the example below for `mstransform`, or perform a number of other measurement set manipulation operations<sup>6</sup>.

```
casa% mstransform(vis='all_my_data.ms',
                 outputvis='my_target.ms.cal.split',
                 field='4U_1728-34',
                 spw='9,11,13,15',
                 datacolumn='corrected')
```

In measurement sets created by `mstransform`, the spws and field ids are renumbered. The example above will create a measurement set labelled `my_target.ms.cal.split` with spws numbered 0, 1, 2 and 3 and a field with an id 0. This can be checked by using the CASA task `listobs`.

For a Member OUS with multiple EBs, the calibrated measurement sets may be concatenated into one measurement set. The information about the different EBs are retained within concatenated measurement sets, and this information can be seen when using `listobs` to create summaries of the measurement sets. The example below, which is meant to be used in the `calibrated` directory of a Member OUS, uses `concat` to create a concatenated dataset of all `*.ms` files in the directory (which are identified using the Python module `glob`) and then lists the contents of that dataset using `listobs`. The `concat` task can also perform time and spectral averaging if desired (see Chapter 8).

```
casa% import glob
casa% msList=glob.glob('*ms')
casa% concat(vis=msList,
             outputvis='calibrated.ms')
casa% listobs(vis='calibrated.ms',
             listfile='lisobs.txt')
```

<sup>4</sup>Before recalibrating the data, the calibration tables and WebLogs will already be present in the `calibration` and `qa` directories, respectively.

<sup>5</sup>For Member OUS data from Cycles 0-2 where flux equalisation was applied using `scriptForFluxCalibration.py` as called by `*.scriptForPI.py`, the `calibrated` directory will contain a concatenated, flux-equalised dataset called `calibrated.ms` with the flux-equalised visibilities in the `CORRECTED_DATA` column.

<sup>6</sup>Descriptions of the `mstransform` and `split` tasks are available at <https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.manipulation.mstransform.html> and <https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.manipulation.split.html>.

For reference, ALMA staff may have included concatenation task in the `*scriptForImaging.py` or `*scriptForImagingPrep.py` files in the `script` directories for manually calibrated data.

## 7.2.4 Possible issues and caveats

Before launching the calibration script, no `calibrated` directory should be present in the Member OUS directory. The calibration script will raise an error when it finds that this directory already exists. A list of the most common errors and problems with running the `*.scriptForPI.py` calibration script are described in Section 13.3.

Prior to the middle of Cycle 5, the delivery package of manually calibrated Total Power data did not include a `*.scriptForPI.py`. For these projects, the `README` gives special instructions on how to obtain the calibrated measurement set.

If computational resources, such as disk space, RAM or the availability to install the desired CASA version, are a problem, please contact your ARC or ARC node via the Helpdesk (see Chapter 14).

## 7.2.5 Modifying the calibration script

For pipeline calibrated projects, the ALMA Pipeline User Guides, available on the Science Pipeline website<sup>7</sup> for various CASA versions, give instructions on how to manipulate inputs such as, for example, introducing additional flags or editing calibrator fluxes.

For manually calibrated projects the `*.scriptForCalibration.py` can be edited as needed. Care should be taken to make these edits in the most appropriate places. For example, additional flagging commands of calibrator targets are best inserted at step 10 “Initial flagging”.

The ARCs are available to assist with the modifications of the calibration scripts (see Chapter 14).

## 7.3 Solar System flux calibrators

In early ALMA cycles, Solar System objects were often used as flux calibrators for individual SBs. The expected flux calibration accuracy per ALMA band can be found in the [ALMA Proposer’s Guide](#) for the cycle from which the data have been taken.

Early ALMA data from Cycle 0, including the early Science Verification data, scaled the amplitudes of the Solar System objects using fluxes from the Butler–JPL–Horizons 2010 models, which are now deprecated. For these datasets, the calibration should be repeated in CASA version 4.0.1 or higher using the Butler–JPL–Horizons 2012 model, which requires modifying the CASA `setjy` task in the calibration script as described in the [CASA Guide for Solar System models in CASA 4.0](#). (For data calibrated using CASA versions earlier than 3.4, additional changes need to be made to the calibration scripts as described in this CASA Guide.) Alternatively, the science target images can be scaled by the appropriate correction factor reported in this CASA Guide.

For the asteroids Pallas and Ceres, the Butler–JPL–Horizons 2012 models were updated in 2015. These newer models were applied to all data taken after Jan 1, 2015, which corresponds to CASA Versions 4.7 and higher. However, the models for Ceres may still be unreliable. For data taken before 2015 that observed Pallas or Ceres as flux calibrators, the following steps should be taken.

- If Ceres was previously used as a flux calibrator, use the bandpass calibrator as a flux calibrator instead if possible.
- If the data were taken in 2013 or earlier, use the bandpass calibrator as a flux calibrator if possible.
- If the data were taken in 2014 with Pallas as the flux calibrator, then it may be possible to calibrate the data. The current plans are to extend the latest flux models backwards in time.

---

<sup>7</sup>The ALMA Science Pipeline website is available at <https://almascience.org/processing/science-pipeline>.

To ascertain whether the quality of the flux calibration is within the ALMA limits, the phase, bandpass and, if present, check source calibrator fluxes can be compared to ALMA flux monitoring values using the ALMA Source Calibrator Catalogue See Section 11.2 for an example of how to do this.

More information on updating Cycle 0 data to the Butler-JPL-Horizons 2012 can be found in the Knowledgebase article [How can I update Cycle 0 absolute fluxes to the Butler-JPL-Horizons 2012 model standards?](#).

Assistance with updated Solar System Object models can be requested from the ARCs via the ALMA Helpdesk (Chapter 14).

## 7.4 Data averaging

Data averaging can be used to reduce the overall data volume of measurement sets. CASA tasks such as `tclean` will run faster on smaller measurement sets. The CASA tasks `split` and `mstransform` can average the visibility data in time using the `timebin` parameter or average the data in frequency using the `width` (in `split`) or `chanbin` (in `mstransform`) parameters.

Frequency averaging is limited by bandwidth smearing, which causes a reduction of the intensity of point-like sources at the edge of the primary beam. Bandwidth smearing is relatively unimportant for short baselines (compact ALMA configurations), but for the more extended configurations and especially at the lower frequencies, it becomes very important. For compact configurations with baselines  $\leq 500$  m, the CASA Guide rule-of-thumb values for frequency averaging are `chanbin=125MHz/(Channel Width [MHz])` for Bands 3, 4, 5 and 6 and `chanbin=250MHz/(Channel Width [MHz])` for Bands 7, 8, 9 and 10. For more extended configurations, the spectral averaging should be reduced if it is applied at all. See the beam size and  $(u, v)$  coverage heading in Section 6.3 for how to obtain the baseline distribution of a measurement set. For more information, consult the [CASA Guide on Continuum Imaging](#) and the [Bandwidth Smearing Tutorial](#).

Time averaging can be performed by specifying the desired time interval using the `timebin` parameter (e.g., `timebin='30s'`). By default, the time span of the time averaging will be limited by the scan boundaries (i.e. the time average will not be extended to average data from multiple scans together). Similarly, time averaging will by default not extend across state boundaries (which, for ALMA, corresponds to a sub-scan). These default limits on time averaging can be modified by setting `combine`, a sub-parameter of `timebin`, to `'scan'`, `'state'` or `'state,scan'` to ignore one or both types of boundaries and average across them.

Time averaging will decrease the  $uv$ -plane sampling and may thus reduce the image fidelity. Time averaging must be kept short enough that the Earth's rotation will not move the source position in the  $uv$ -frame by as much as the synthesized beam. Large time-averaging values may thus tangentially broaden the synthesized beam. Hence, the time averaging  $\Delta t$  should be set so that

$$\Delta t \ll \frac{\theta_{\text{res}}}{\omega_e * \theta_{\text{FOV}}}$$

where  $\theta_{\text{res}}$  is the size of the synthesized beam,  $\theta_{\text{FOV}}$  is the field of view, and  $\omega_e$  is the Earth's angular rotation speed of  $7.3 \times 10^{-5} \text{ rad s}^{-1}$ . For example, a Band 6 observation using the most extended array will produce images with fields of view of 25 arcsecond and synthesized beams of 45 milliarcsecond. For these observations, the averaging time must be  $\ll 0.045/25/7.3 \times 10^{-5} = 25$  s. For more information on time averaging, consult Chapter 6.4.1 in [Interferometry and Synthesis in Radio Astronomy by Thompson, Moran & Swenson \(2017\)](#).

More information on data averaging with `mstransform` can be found in the [CASA documentation on  \$uv\$ -manipulation](#).



# Chapter 8

## Re-imaging archival data

While all archive images are intended to be useful as reference images, they are not necessarily optimized for all scientific applications. While PIs and archive data miners may want to use these images as indicators of the data quality and coverage of the observations, it may be desirable, especially for data from the older ALMA cycles, to re-image the data using the latest version of the imaging task `tclean`<sup>1</sup> with parameter settings that are optimized for the specific science project. Note that the archival data will first need to be calibrated before they can be re-imaged, as explained in Sect 7.2. Re-imaging can be performed starting with the existing imaging scripts or pipeline task parameters available in the auxiliary package from the ASA (see Section 3.1.1).

- For pipeline imaged datasets, the image pipeline inputs can be altered to make new images. Consult the [CASA Guide for the ALMA imaging pipeline reprocessing](#).
- The CASA `tclean` task can be used to manually make new images of any ALMA data. In this case, it is still very useful to start with the existing imaging script parameters available in the ASA auxiliary files. Manually imaged data have a `*.scriptForImaging.py` in the `script` directory. Pipeline imaged data have `tclean` parameters listed within the WebLog. To see these parameters, go to any page for the `hif_makeimages` task and either click on “Input Parameters” near the bottom of the page or either view or download the CASA logs using the links at the very bottom of the page.

While the current CASA imaging task is `tclean`, older archival data would have been imaged with the older version of this task, which is called `clean` and which will be deprecated. It is recommended that `tclean` is used to re-image these older data.

The parameters of `tclean` can be set to customize the final images, including changing the spectral resolution, the spectral range and the angular resolution. In the following subsections, we briefly review some of the most commonly-used `tclean` parameters. For detailed information on the `tclean` task, consult the [First Look at Imaging, Image Continuum](#) or [First Look at Line Imaging](#) CASA Guides for introductory information, or look at the more specific examples for multiple ALMA cycles listed at [ALMA Imaging Pipeline Reprocessing](#). Several [ALMA-specific CASA Guides](#) show examples of how to image various kinds of ALMA data.

### 8.1 The `datacolumn` parameter

A measurement set may contain both the visibilities prior to calibration as well as the calibrated visibility data. These data are placed in different data columns, where `datacolumn='data'` is the version of the data created prior to the application of calibration tables to the measurement set and `datacolumn='corrected'` is the calibrated version of the data.

It is important to specify which data column should be used when imaging with `tclean`. As a rule of thumb,

---

<sup>1</sup>The description of the `tclean` task is available at <https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.imaging.tclean.html>.

after having run the standard calibration script as provided in the ASA, the following options should be used based on the output.

- `datacolumn = 'data'` should be used if the calibrated measurement set created in the `calibrated` directory has a name that ends in `*.split.cal.ms`. Files with these names were probably created using manual calibration scripts, in which case the corrected data will be in the `DATA` column. (Subsequent calibration steps, such as self-calibration, will fill in the `CORRECTED_DATA` column of the measurement set.)
- `datacolumn = 'corrected'` should be used if the `calibrated` directory contains only files ending in `*.ms` and not any files ending in `*.ms.split.cal`. These data would have probably been calibrated with the pipeline. The pipeline creates measurement sets with separate `DATA` and `CORRECTED_DATA` columns, with the calibrated data in the `CORRECTED_DATA` column.

Note that, if `datacolumn` is set to `CORRECTED_DATA` but that column does not exist, `tclean` and `plotms` will use the `DATA` column instead.

One way to check whether the `CORRECTED_DATA` column exists is to try to extract it with the CASA `split` or `mstransform` tasks as in the following example.

```
casa% split(vis='my_calibrated_data.ms',
           outputvis='spw31_data.ms',
           datacolumn='corrected',
           spw='31')
```

If the `CORRECTED_DATA` column is not present, this will produce an error.

## 8.2 Spectral line images and continuum subtraction

The spectral information of ALMA visibility data is natively stored in the TOPO (topocentric) reference frame<sup>2</sup>. This is a time-dependent reference frame. Using the CASA `tclean` task, the `outframe` parameter can be set to a time-independent reference frame such as `LSRK` or `BARY`. When combining multiple spectral line EBs, it can be useful to transform each EB measurement set to a time-independent reference frame by using the CASA task `mstransform`. See also Section 13.6.

When creating image cubes for science, it may be practical to use `tclean` to create first-look image cubes for each spectral window (or at least for each one that includes any spectral lines of interest) covering all the channels but using the `width` parameter in the `tclean` task to optimize the spectral resolution. These first-look cubes are useful both for identifying the locations of specific spectral lines and continuum emission, especially when imaging concatenated data from different EBs (in which case `outframe` should be set to a time-independent reference frame), and for testing how the channel width affects the spectral line detection.

For pipeline imaged datasets, most if not all target data should be fully imaged<sup>3</sup>. It is advisable to first look at these image cubes, although keep in mind that the spectral resolution in the image cubes is not necessarily optimized for line detection.

If continuum-subtracted data are needed, then one of a few different methods could be used to identify the continuum channels.

- The first-look image cubes can be used to identify the continuum channels. However, be aware that the channels in the image cubes may not have a one-to-one correspondence with the channels in the measurement sets if the `start` or `width` parameters were set when creating the cubes, so appropriate

---

<sup>2</sup>Consult the Knowledgebase article [What are the frequency reference frames in CASA?](#) for more information about reference frames.

<sup>3</sup>For very large datasets only a subset of the images cubes may have been generated.

conversions will need to be applied to create a list of continuum channels in the visibility data. On the other hand, the frequencies in the spectral cube can be used in place of the channels when performing the continuum subtraction.

- Line-free channels can potentially be identified by plotting the amplitudes versus channel for the visibility data as described by the CASA Guide [First Look at Line Imaging](#) and shown in the example below.

```
casa% plotms(vis='my_data.ms',
            xaxis='channel',
            yaxis='amp',
            field='my_field',
            avgspw=False,
            avgtime='1e9',
            avgscan=True,
            avgbaseline=True,
            showgui = True)
```

However, fainter spectral lines may be difficult to see using `plotms`, in which case the first-look image cubes would be more useful for identifying the line-free channels.

- If the data were pipeline imaged, then the pipeline identified line-free channels will be listed in `*.continuum.dat` or `*.cont.dat` files in the `calibrated` directory and can be inspected in the WebLog by checking out task `hif_findcont`.
- If the data were manually imaged, the `*scriptForImaging.py` file in the `script` directory may contain the `uvcontsub` task with the line-free channels identified by the ALMA staff member.

Even if the continuum emission is not detectable in the individual image cubes, it is still good practice to perform the continuum-subtraction steps if spectral line data are needed for the analysis.

Once the continuum channels have been identified, `uvcontsub` can be used to fit and subtract the continuum from the visibility data as shown in the example below.

```
casa% uvcontsub(vis = 'my_data.ms',
               field = 'M100',
               fitspw = '0:5~50;170~240',
               fitorder = 1,
               solint='int')
```

The continuum channels or frequencies are specified using the `fitspw` parameter. The output is a continuum-subtracted *uv*-dataset that has the file extension `*.contsub`, and the data in the output should contain only the spectral line emission. The `plotms` task in the example above can be used with the `*.contsub` file as the input measurement set to verify that the continuum was subtracted; the amplitudes in the line-free channels should be approximately 0. The `*.contsub.ms` file can be used as an input in `tclean` for spectral line imaging.

When creating image cubes, the following parameters should be set.

- `specmode = 'cube'` should be set to create image cubes.
- `restoringbeam = 'common'` should be set to ensure a constant beam size for each plane in the cube.
- `start` should specify what to use as the first channel in the image cube. This could be either a frequency, velocity or a channel in the input visibility data.
- `width` should specify the channel width in the image cube.

- `nchan` should specify the number of channels (or should be set to -1 to image all data available from the start position to the end of the cube).
- `restfreq` optionally should be set to the rest frequency of a spectral line of interest. If this is set, then `start` and `width` could be specified in velocity.
- `outframe` should set the reference frame of the image. Typically, time-independent reference frames such as “LSRK” or “BARY” are used.

The CASA Guide [First Look at Line Imaging](#) gives examples of spectral line imaging with CASA.

## 8.3 Continuum imaging

Continuum imaging should be done following steps to identify the continuum channels in the visibility data. This could include using the continuum channels identified by the `hifa_findcont` step in the WebLogs, using the images from the ALMA archive or using new first-look images created from the archival data. When combining data from multiple SBs or multiple programs, new first-look images would provide the best opportunities to find spectral lines and therefore to identify the continuum channels.

When creating continuum images, the following parameters should be set.

- `specmode = 'mfs'` should be set to create continuum images.

Note that if the fractional bandwidth  $\Delta\nu/\nu$  is greater than 10%, then `deconvolver='mtmfs'` should be set to account for the spectral slope (see Section 8.5 for more details).

## 8.4 Weighting and angular resolution

The restoring beam is a 2D Gaussian fit to the central peak of the synthesised beam or point spread function (PSF) of the data. The angular resolution (i.e. the restoring beam) of an interferometric radio image can be adjusted by modifying the `weighting`, `uvtaper` or `uvrange` parameters of the `tclean` task.

The `weighting` parameter controls the weighting scheme applied to the visibility data (e.g. the weight that each visibility measurement is multiplied by during the image generation). Four weighting schemes are most commonly used.

- `weighting = 'natural'` will weight data based on the sample density pattern on the  $uv$ -plane. In practice, this gives more weight to shorter baselines and thus decreases the angular resolution. It provides the maximum imaging sensitivity as well as the best sensitivity to large-scale emission, but the images will have a broader beam with high side lobes. It is most appropriate for detection experiments where sensitivity is more important than angular resolution or when the detection of extended emission is important.
- `weighting = 'uniform'` will weight data so that the  $uv$ -plane is uniformly weighted. This is done by down-weighting short baselines, which are more numerous, and giving more weight to long baselines, which are less common. The resulting images have higher angular resolutions, and the beam will have narrower main lobes and suppressed sidelobes. However, the imaging sensitivity will not be as high as with natural weighting, and large-scale emission will not be reproduced as effectively. For imaging interferometric data in general, this is most appropriate when narrower beam sizes are needed. In practice, though, images created from ALMA data (as well as images created from other data with poorly-sampled long baselines) using this weighting scheme will generally be excessively noisy.
- `weighting = 'briggs'` applies a weighting scheme that varies smoothly between natural and uniform weightings based on the `robust` parameter, which should also be specified when using this weighting scheme. The `robust` parameter can be varied between 2, which is equivalent to natural weighting, and -2, which



is equivalent to uniform weighting. This allows for customizing the trade-off between the narrower beam size but noisier images provided by uniform weighting and the better imaging sensitivity and sensitivity to large-scale emission but broader beam sizes provided by natural weighting. The standard value used for ALMA data is typically `robust=0.5`.

- `weighting = 'briggsbw taper'` applies, just like the `briggs` scheme, a weighting scheme between natural and uniform depending on the `robust` parameter setting. In addition, `briggsbw taper` applies an inverse *uv*-taper per frequency channel so that the cube imaging weights have a similar density as continuum (multi-frequency synthesis) weights. This is essentially a frequency-related correction for changes in the angular scales sampled by the individual baselines. This weighting scheme should be used in combination with the `tclean` parameter `perchanweightdensity=True`.

Two different `tclean` parameters, `uvtaper` and `uvrange`, can be used to enhance the emission from low surface brightness components. When using these parameters, it is important to use natural visibility weighting scheme.

- `uvtaper` applies a taper in addition to the weighting scheme specified via the `weighting` parameter. This taper is a Gaussian function expressed in terms of distance in *uv*-space. Longer baselines (corresponding to higher spatial frequencies or smaller angular scales) are weighted down relative to shorter baselines (corresponding to lower spatial frequencies or larger angular scales) to suppress artifacts arising from poorly sampled areas of the *uv*-plane. This is equivalent to smoothing the beam obtained by other weighting schemes. It decreases the angular resolution to reveal large-scale structures in the image. However, as more data are down-weighted, the image noise will increase.
- `uvrange` is more extreme, as it limits the *uv*-range of the data used for imaging. This is recommended for determining the length scale (angular resolution/baseline) at which sources become resolved or to exclude point-like objects, which dominate the emission at long baselines. Again, image noise will increase as more data are excluded.

The final synthesised beam, which ideally should be Gaussian, can be inspected after running `tclean` by displaying the resulting `*.psf` image in the CASA viewer or CARTA. This image can also be useful for understanding whether any artefacts are caused by a prominent non-Gaussian beam shape.

## 8.5 Other commonly-used parameters

Users can set additional `tclean` parameters to customize the images for their science goals.

- `vis`: This parameter specifies the input visibility data. It is possible to image combined datasets by setting the `vis` parameter to a list of measurement sets. This list could include multiple measurement sets from an SB that contains multiple EBs, measurement sets from different 12-m configurations, or measurement sets from 7-m and 12-m observations. For more information on data combination, see Chapter 9. When combining measurement sets from antennas with different diameters, the parameter `gridding='mosaic'` must be set to derive the appropriate primary beam correction.
- `imsize`: This sets the image size in pixels. The image size should be slightly larger than the area covered by the primary beam (or, more precisely, the area of the primary beam that is above the value set by `pblimit`, which has a default value of 0.2). The FWHM primary beam for the ALMA 12-m and 7-m is determined by the antenna diameter  $D$  and the observing wavelength  $\lambda$  as given by the equation  $FWHM = 1.13 * \lambda / D$ . However, smaller image sizes are more practical to use when imaging sources that are small relative to the primary beam, such as individual point-like sources that are observed using long baseline configurations.
- `cell`: The cell size, which is specified by this parameter, should be at least between 1/5 to 1/8 of the width of both axes of the synthesized beam to allow for proper sampling of the beam shape. See the CASA Guide on [Continuum Imaging Parameters](#) for more details on `cell` and `imsize` parameters.

- **phasecenter**: This sets the centre of the image produced by `tclean`. This parameter should be set when creating mosaics based on data with multiple pointings.
- **stokes**: The default setting for this parameter is 'I', which will image only intensity. Polarisation imaging is possible if the observations were conducted in full polarisation mode, in which case the parameters `stokes= 'IQUV'` (or a single Stokes parameter) and `psfmode='clarkstokes'` should be set<sup>4</sup>.
- **gridding**: The default setting of `gridding='standard'` works for images based on single pointings. However, `gridding='mosaic'`, which uses a joint deconvolution algorithm that works in the  $uv$ -plane, must be set when imaging mosaic data and when combining 12-m Array and 7-m Array data. The mosaic response function, which is a convolution of the primary beam patterns for each pointing in the mosaic, will be created as a file named `*.pb`. See the [CASA Guide on Mosaics Imaging](#) for examples. Data combination is discussed in Chapter 9.
- **deconvolver**: A number of deconvolvers are available. The standard is `deconvolver='hogbom'`, which is recommended for data with poor  $uv$ -coverage. Use `deconvolver='multiscale'` if complicated, large-scale structures are expected in the image. When creating continuum images (specifically when the aggregate fractional bandwidth  $\Delta\nu/\nu$  is greater than 10%) and the target has a high signal-to-noise, then `deconvolver='mtmfs'` (multi-term multi frequency synthesis) is recommended. When 'mtmfs' is set, the `nterm` parameter can be set to fit the spectral profile with multiple terms (which is useful for accounting for non-flat profiles). Hence, this will also directly derive spectral index maps (named `*.alpha`)<sup>5</sup>.
- **pbcor**: When set to `True` the primary beam correction is applied to the final restored image. The default setting is `False`. The primary beam response is typically a Gaussian with a value 1 at the center of the field (the `*.pb` file). To form an image of the sky where the flux densities of off-centre sources are scaled correctly, the output of `tclean` needs to be divided by this primary beam (or, in the case of mosaics, the combination of primary beam patterns used to make the mosaic)<sup>6</sup>.
- **threshold**: This is the global stopping threshold that the peak residual (as measured within the region of the clean mask across all image channels) is compared to. It is recommended to specify this parameter, especially when using `interactive=False`, to avoid overcleaning, which may add false sources to the deconvolved images by treating random noise spikes as sources. A suitable number for this parameter is 3 times the expected noise of the observation.
- **interactive**: When this is set to `True`, a GUI will appear during the clean process that will allow users to manually draw clean masks, which will make the cleaning process more precise. When this is set to `False`, the cleaning is performed without additional user interaction.
- **usemask**: This specifies an input mask for the cleaning process. When this is set to `auto-multitresh`, the `tclean` task will mimic what an experienced user would do when manually masking images, which may be particularly useful for maps with a large number of sources or maps with very complex emission structure (see the [CASA Automasking Guide](#)).

## 8.6 Self calibration

Self calibration is an advanced data calibration technique that aims to improve the calibration and enhance the signal to noise ratios of the images by calibrating the visibility data on the science target itself. The process can be applied to both continuum images and spectral cubes. For self-calibration to be effective, the target image has to contain a bright source. A discussion of this topic is beyond the scope of this document, but a number of CASA Guides describe the self-calibration process, including the [First Look at Self Calibration Guide](#), the

<sup>4</sup>Once a Stokes image cube has been made, images of the linear polarisation and polarisation angle can be created from the Q and U images with the CASA task `immath` using either `mode='poli'` for a linear polarisation image or `mode='pola'` for a polarisation angle image. It is also possible to create the fractional linear polarisation image using `immath` with `mode='evalexpr'`. See the [CASA Guide for Polarisation Imaging](#) for examples.

<sup>5</sup>Spectral index images can also be obtained with the CASA task `immath` using the `mode='spix'` parameter (see the [CASA Guide for Spectral Index Imaging](#)). In this case, it is recommended to use `deconvolver='mtmfs'` and `nterms=2`.

<sup>6</sup>When the primary beam correction is not implemented in `tclean`, the correction can be applied to the uncorrected image (`*.image`) using the CASA task `impbcor`.

[Self-Calibration Template](#) and the [VLA Self Calibration Tutorial](#). Additionally, more theoretical information on gain calibration techniques can be found in [Brogan, Hunter & Fomalont \(2018\)](#).



# Chapter 9

## Data combination

Multiple ALMA array configurations may be needed to cover high spatial dynamic ranges in fields with both diffuse, extended emission and compact, localised emission. As discussed in Section 1.2, ALMA consists of fifty antennas in the 12-m Array (which can be moved into multiple configurations), twelve 7-m antennas and four 12-m antennas in the Total Power (TP) Array. Together, the 7-m and TP Arrays form the ACA. Data from these arrays are acquired and calibrated as separate Member OUSs, but after calibration, the data may be combined to cover a range of spatial scales from tens of milliarcseconds to tens of arcseconds. At present, the ASA and the imaging pipeline do not provide images based on data from multiple arrays, even in the case where multiple SBs were executed using multiple arrays with the purpose of being combined to produce final images. Combining and imaging the data is left to users, who are encouraged to seek help from the ARCs, as data combination may be technically challenging and require accessing advanced computing resources.

This chapter provides a broad overview of the array combination process. Array combination, and particularly the combination of single dish and interferometric data, is currently the subject of ongoing development, so other references should be used to look up detailed information on the subject. The technical aspects of data combination are given in Chapter 7 of the [ALMA Technical Handbook](#). The [CASA Guide for the ALMA Science Verification observations of M100](#) provides an example showing how to combine and image data from the 12-m, 7-m and the ALMA Total Power Array. The [CASA Documentation on combining measurement sets](#)<sup>1</sup> provides details on combining *uv*-data using the CASA `concat` task.

### 9.1 Preparation steps before array combination

ALMA visibility weights are formally set to  $1/\sigma^2$ , where  $\sigma$  is the RMS noise expected per integration time and per bandwidth<sup>2</sup>. The weights column in the measurement set is normally updated during the calibration process<sup>3</sup>. These weights are used by `tclean` in CASA when producing images. When concatenating multiple measurement sets, the weights need to be consistent. Any ALMA data that were pipeline calibrated using CASA version 4.2.2 or later or that were manually calibrated using CASA 4.3 or later have consistent weight calculations; see Section 13.2 for information on finding the CASA version. When combining older ALMA data, it is imperative to verify the visibility weights. The [CASA Guide on Data weights and combination](#) provides more information about weights in ALMA data.

A few issues regarding fields need to be taken care of when concatenating or imaging multiple measurement sets. If the target positions are not the same in all of the measurement sets, they may be handled by CASA as different fields when the data are concatenated and may be renumbered. As a consequence, the `field` and `phasecenter` parameters for `tclean` that are used for imaging a concatenated dataset may differ from the parameters used for the individual measurement sets in that concatenated dataset. In observations of Solar

---

<sup>1</sup>Information on how to combine measurements sets with CASA is available at [https://casadocs.readthedocs.io/en/stable/notebooks/uv\\_manipulation.html#Combine-MeasurementSets](https://casadocs.readthedocs.io/en/stable/notebooks/uv_manipulation.html#Combine-MeasurementSets).

<sup>2</sup>For CASA versions earlier than 4.2.2 (which were used in early ALMA cycles), the weights were defined differently. More information is provided by the [CASA memo on weights](#) and the [CASA description of visibility weighting](#).

<sup>3</sup>When manually calibrating data, the weights are updated by `applycal` when the `calwt=True` has been set.

System objects and near-field high-proper motion objects, `phasecenter='TRACKFIELD'` should be used so that `tclean` uses the ephemerides attached to the `FIELD` table in the measurement set.

Before creating an image based on data from multiple measurement sets, it is also important to check that the flux densities (i.e. the amplitudes) from the different datasets are consistent when plotted in  $uv$ -space. This can be done using the example `plotms` command shown below. Inconsistencies in the flux densities can be corrected either by re-scaling the visibilities or by using the image created with data from a more compact configuration as a model for calibration.

## 9.2 Combining different 12-m Array configurations

Measurement sets from two or more 12-m Array observations can be given as a list of inputs to the CASA task `tclean`, which combines the visibility data to make one image. When doing this, it is important to set `beam` to `'common'` and set the `phasecenter` parameter to a coordinate string (with more information on this given by the `tclean` documentation). Alternately, multiple measurement sets can be combined to form one larger measurement set using the CASA task `concat`, and the output from that task can then be used as input for `tclean`.

## 9.3 Combining 7-m and 12-m Arrays

An overview of the 12-m Array configurations that can be combined with the 7-m Array data can be found in Chapter 7 of the [ALMA Technical Handbook](#). In general, the  $uv$ -plane should be sampled as well as possible to improve to restoration of the emission on various spatial scales. If the data from the two arrays overlap in  $uv$ -space, this can be used to check that the flux scales for the two datasets are equivalent on the same spatial scales.

The overlap between baselines can be checked by first combining the data with `concat` and subsequently using the CASA task `plotms` to plot the amplitudes of the visibility data versus  $uv$ -distance. The example command below shows how to launch `plotms` from the CASA command line interface and save the output as a file, but it is also possible to examine the data in the GUI.

```
casa% concat(vis=['7mdata.ms', 12mdata.ms'],
             combinedvis='7and12mdata.ms')
casa% plotms(vis='7and12mdata.ms',
             xaxis='uvdist',
             yaxis='amp',
             field='M100',
             plotfile='M100_combined_uvdistance.png',
             coloraxis='observation',
             avgtime='60',
             avgscan=True,
             showgui=False)
```

Although `plotms` can be used to simultaneously plot data from multiple measurement sets, it is easier to create plots using concatenated data. Setting `coloraxis` to `'observation'` will plot the data from each EB as different colours, so the data from the 12-m and 7-m Arrays will be coloured differently. Averaging by time, channel or scan will speed up the task significantly and is recommended for this comparison.

After this flux scaling check, the data can be imaged in the same way as two 12-m Array datasets, which was described in the previous section. Either a single concatenated dataset or separate 12-m and 7-m measurement sets can be used as inputs in `tclean`. However, the `gridder` parameter needs to be set to `'mosaic'` to properly perform the primary beam correction for the two different antenna sizes.

## 9.4 Combining 7-m/12-m and TP Arrays

Three methods are available for combining interferometric and single dish (or total power) data: feathering, joint deconvolution, and assisted cleaning. The main issues when combining data are related to the file headers, differences in the frequencies or spectral setup of the different arrays and the structure of the source that has been imaged. Currently, a data combination working group is testing different combination methods and preparing a series of documents explaining the main difficulties and possible solutions. This working group maintains a website with all documentation, scripts and presentations<sup>4</sup>.

### 9.4.1 Feathering

With the CASA task `feather`<sup>5</sup>, an image from single-dish observations and an image from an interferometer are combined in the Fourier transform plane (see [Cotton 2017](#)). This is the current recommended method for combining ALMA Total Power and ALMA interferometric observations. The CASA Guide Tutorial on [M100 Data Combination](#) provides instructions on using the feathering technique. The command below shows an example of how feather is used.

```
casa% feather(imagename='M100_Feather_CO.image',
             highres='M100_combine_CO_cube.image.subim',
             lowres='M100_TP_CO_cube.regrid.subim.dep')'
```

### 9.4.2 Joint deconvolution

In this method, both the interferometer and single-dish data are combined in the  $uv$ -plane (or visibility space). The combined visibilities are then imaged as regular ALMA data. It is thus necessary to create visibilities for the single-dish data. While CASA does not yet have any tasks for doing this, two tools have been developed that can be used within CASA: `tp2vis`<sup>6</sup>, which is described by [Koda et al. \(2019\)](#) and has been developed as a North American ALMA Development Study; and `sd2vis`<sup>7</sup>, which has been developed by the Nordic ARC node.

### 9.4.3 Joint reconstruction (sdintimaging)

The CASA task `sdintimaging`<sup>8</sup> uses the interferometric measurement set, the single dish image and their PSF images to create a joint image and PSF cubes. First, the interferometer data are gridded into an image or image cube, and a corresponding PSF is created for the data. Next, the single dish image and PSFs are combined with their interferometric counterparts in a feathering step. The joint image and PSF data are then passed through a deconvolution algorithm. Model images from the deconvolution algorithm are then subtracted from both the single dish image cube as well as the interferometer data to form the new pair of residual images to be feathered in the next iteration. More information `sdintimaging` can be found in [Rau et al. \(2019\)](#).

### 9.4.4 Model-assisted cleaning

The model-assisted cleaning method uses the single dish image as a model when cleaning the interferometric data. Using a model usually helps in the convergence of the imaging algorithm `tclean`, although total flux is not constrained. As a second step, the resulting interferometric image is combined with the single dish image using the feathering technique.

<sup>4</sup>The data combinations working group website is located at <https://github.com/teuben/dc2019/>.

<sup>5</sup>A description of the `feather` task is available at <https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.imaging.feather.html>.

<sup>6</sup>Information about the `tp2vis` task can be found at <https://github.com/tp2vis/distribute>.

<sup>7</sup>Information about the `sd2vis` task is available at <https://www.oso.nordic-alma.se/software-tools.php>.

<sup>8</sup>A description of the `sdintimaging` task is available at <https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.imaging.sdintimaging.html>.

## 9.5 Combining ALMA data with other single dish or interferometer data

The process of combining ALMA data with other single dish or interferometer data should be the same as combining data from different ALMA arrays. However, the header information for data from other telescopes must have the same format as the ALMA data that it will be combined with. An example of how to combine ALMA 12-m Array data with data from the Green Bank Telescope, using the “model-assisted cleaning” method, can be found in [GBT Memo #300](#) by Hoffman & Kepley (2018).



# Chapter 10

## Using the calibrators as science targets

### 10.1 ALMA calibrator targets

ALMA calibrators are typically compact quasars and Solar System objects. The latter are only used for flux calibration, while the quasars are used for bandpass, flux and phase calibration as well as check sources. The ALMA Calibrator Source Catalogue (see Chapter 11) is the database of quasar calibrator observations and can be used to look up measured fluxes, positions and other properties of the objects.

ALMA calibrator data are always publicly available. For projects still within their proprietary periods, the calibrator data cannot be directly downloaded from the ASA but can be requested from the ARCs via the Helpdesk. The ASA can be used to search specifically for calibrator data by ticking the “calibrator-only” box in the search-box. Generally, the ASA contains not only the visibility data for the calibration sources but also the calibrator continuum images that were produced for QA2. For pipeline calibrated data, all calibrators will have one continuum image per spw. However, for manually calibrated data, the availability of calibrator images is variable, and the images may not have even been created.

When visibility data are downloaded from the archive and re-calibrated using the ALMA calibration scripts, the final measurement sets will contain calibrated data for the calibrator sources. Nevertheless, the use of calibrators as science targets may require modifying the calibration script for a number of reasons, including removing or excluding any commands that flag spectral lines (see next Section) or changing the calibration scheme to provide an independent bandpass calibration (if another suitable calibrator target was observed).

Modifications of the calibration scheme may be complicated, and users are encouraged to contact their local ARC or ARC node through the Helpdesk for assistance (see Chapter 14).

### 10.2 Identifying flagged spectral features in calibrators

Spectral lines in calibrator spectra are generally flagged during standard manual or pipeline calibration. To inspect and image the spectral features in calibrator data, it is necessary to first find the flagging commands and then re-run the calibration script without these flags.

For some Solar System objects, the latest flux models also include some strong molecular lines such as CO, which means that these lines may not need to be flagged if these flux models are used. However, spectral lines not included in the models of Solar System objects will be flagged during QA2. The spectra of phase calibrators are checked and flagged for spectral lines during QA2 as well, but weak spectral lines may sometimes remain unseen and therefore unflagged.

Bandpass calibrators, which may also be used as flux calibrator, are selected on being line-free sources and should in general not have any spectral line features. In the rare case that the bandpass calibrator has spectral lines these would have been flagged in QA2 (otherwise these would have caused a systematic offset in amplitude in the spectral window).

The quickest way to check whether any spectral ranges have been flagged in the calibrator data is to check either the `*scriptForCalibration.py` in the `script` directory (for manually calibrated data) or the WebLog in the `qa` directory (for pipeline calibrated data).

- **manual calibration:** The `flagdata` commands are generally listed in either the “A priori flagging” step (near the beginning of the script) or the “Initial flagging” step (in the middle of the script). Many manual calibration scripts from the ALMA archive also include comments describing the reasons that data were flagged during QA2.
- **pipeline calibration:** The output from `hif_fluxcalflag` in the “By Task” tab of the WebLog can be used to check which Solar System flux calibrator data may have been flagged. Additional manually added flags may be found in the `hifa_flagdata` output and flagging template files.

Once the flagging commands have been identified, they can be commented out in the manual calibration script or removed from the flagging template files. Re-running the pipeline calibration without `hif_fluxcalflag` requires commenting out this step in the `casa_pipescript.py` file (see the ALMA Science Pipeline User’s Guide<sup>1</sup> for details). The calibration has then to be re-run to obtain the measurement set with the spectral lines.

---

<sup>1</sup>The ALMA Science Pipeline website available at <https://almascience.org/processing/science-pipeline> and contains User’s Guides for various CASA versions.

# Chapter 11

## ALMA Calibrator Source Catalogue

The ALMA Calibrator Source Catalogue<sup>1</sup> is the database of ALMA calibrator sources, which are mostly bright, compact quasars. It contains a web-based query interface<sup>2</sup> (shown in the top panel of Figure 11.1) that can be used to search for observations of the calibrator sources.

In this chapter, we give an overview of the strategies available to assess the flux calibration of ALMA archival data using the Calibrator Source Catalogue. The astronomical sources that are usually adopted as flux density calibrators are presented in Section 11.1. General strategies to assess flux calibration are provided in Section 11.2.

### 11.1 ALMA calibrator (grid) sources

ALMA data is flux calibrated using astronomical sources with known flux densities in the observations. This could be a bright quasar or a Solar System object (when requested in the Observing Tool at proposal submission). About 40 bright quasar sources, called grid sources, are frequently monitored by the observatory so as to track their flux densities, which may vary over time. These monitoring observations are typically performed in Band 3 and in either Band 6 or Band 7, which allows for calculating a spectral index for these sources. In earlier cycles, grid sources were used as flux calibrators whenever a Solar System object was not available. In later cycles, thanks to the monitoring efforts of grid sources, these are now principally used as secondary flux calibrators in science observations. These sources are also normally used as bandpass calibrators.

The flux accuracy of ALMA data is estimated to be better than 5% for Bands 3-5, 10% for Bands 6-8, and 20% for Bands 9 and 10. More details on ALMA flux calibration can be found in Appendix A of the [ALMA Proposer's Guide](#) and in Chapter 10 in the [ALMA Technical Handbook](#).

### 11.2 Checking the flux calibration of archival data

After downloading an ALMA dataset, it may be useful to examine the accuracy of the flux scale of the data (for example, when estimating the uncertainties of the target source flux). The consistency of the flux calibration of a dataset can be assessed by comparing the flux density values derived for the grid sources with the values listed in the ALMA Calibrator Source Catalogue. The fluxes listed in this catalogue are used by the observatory at the time of data reduction to perform the flux calibration. Therefore, the values should be consistent by default. However, it may be useful to, for example, investigate the impact of alternative spectral index calculations and to set up different flux density scales for the flux calibration sources, especially since these sources are variable. This section provides the tools for such checks and alternative flux extrapolations.

---

<sup>1</sup>The ALMA Source Calibrator Catalogue website is available at <https://almascience.org/alma-data/calibrator-catalogue>.

<sup>2</sup>The ALMA Calibrator Source Catalogue query-interface is available at <https://almascience.org/sc/>.

## 11.2.1 Obtaining calibrator flux densities

Querying the ALMA Calibrator Source Catalogue will show the flux history of any specified source (see Figure 11.1). The calibrator catalogue can be queried by source name, position, energy, observing dates, or polarisation fraction. After pressing “Search”, the catalogue provides the results in the form of a table and a plot, which are shown in separate tabs. The table provides the measurements within the specified time range. If no time range is specified, only the most recent measurements are listed. The plot initially shows the sources in a map of the sky. Clicking on a source will display a second graph showing flux density versus time in various ALMA bands (Figure 11.1). It is possible to magnify part of the plot to see the history of a particular source in a specific time interval and flux range.

The ALMA Calibrator Source Catalogue can provide a good visual overview of the existing measurements and variability of calibrator sources. To obtain numeric values, two auxiliary tasks for CASA exist in the Analysis Utilities package (see Section 12.1) that facilitate the extrapolation of flux measurements to a given frequency and time: `getALMAFlux` and `getALMAFluxForMS`<sup>3</sup>.

The task `getALMAFlux` can be used in CASA to estimate the flux density of a specified grid source at a given time and frequency as seen in the example below.

```
casa% import analysisUtils as aU
casa% aU.getALMAFlux('J1924-292', 230, date='2019-12-12', showplot=True,
plotfile='J1924_230.png')
Closest Band 3 measurement:  5.500 +- 0.130 (age=-2 days) 91.5 GHz (will extrapolate from this datum
using spectral index)
Closest Band 7 measurement:  2.780 +- 0.080 (age=+2 days) 337.5 GHz getALMAFluxCSV(Cycle6):  Fitting
for spectral index with 1 measurement pair of age 6.0 days from 2019-12-12, with age separation of 0
days
2019-12-06:  freqs=[103.5, 91.47, 337.47], fluxes=[5.33, 5.56, 2.75]
Wrote plot:  J1924_230.png
Median Monte-Carlo result for 230.000000 = 3.422817 +- 0.275106 (scaled MAD = 0.273568)
Result using spectral index of -0.538980 for 230.000 GHz from 5.500 Jy at 91.470 GHz = 3.346019
+- 0.275106 Jy
Out[3]:  'ageDifference':  4.0, 'covar':  array([[ 0.00034049,  0.00136793],[ 0.00136793,
0.00887766]]), 'fluxDensity':  3.3460192750107671, 'fluxDensityUncertainty':  0.27510639305798373,
'intercept':  0.63837706251223336, 'interceptUncertainty':  0.0, 'meanAge':  -2.0, 'meanOfLogX':
2.168151400125081, 'monteCarloFluxDensity':  3.422816598521556, 'spectralIndex':
-0.53898041071174752, 'spectralIndexAgeOldest':  6, 'spectralIndexAgeSeparation':  0,
'spectralIndexAgeYoungest':  6, 'spectralIndexDate':  '20191211', 'spectralIndexNPairs':  1,
'spectralIndexUncertainty':  0.012793175399861925
```

The task first prints a list of ALMA flux density measurements near the requested date. The procedure then uses these measurements to fit a power law function through these data, and the power law is then used to determine the flux density at the requested frequency. Several quantities related to the power law function are printed on the screen. The task also returns a python dictionary containing the estimated flux density and its uncertainty, the slope and its uncertainty, and the mean age of the observations. If the `showfile` parameter is set to `True`, then a plot is made showing the ALMA data points, the best-fitting power law, and the flux density at the specified input frequency. If `showfile` is set to `True` and if the `plotfile` parameter is set, the image is saved as a file. Many optional parameters are available for refining the flux extrapolation calculations as discussed in the `getALMAFlux` documentation.

The task `getALMAFluxForMS` can be used to estimate the flux density of sources in a given measurement set. This runs `getALMAFlux` for all calibrators present in a measurement set and returns a python dictionary

<sup>3</sup>Documentation for the Analysis Utilities tasks `getALMAFlux` and `getALMAFluxForMS` can be found at <https://safe.nrao.edu/wiki/bin/view/ALMA/GetALMAFlux> and <https://safe.nrao.edu/wiki/bin/view/ALMA/GetALMAFluxForMS>, respectively.

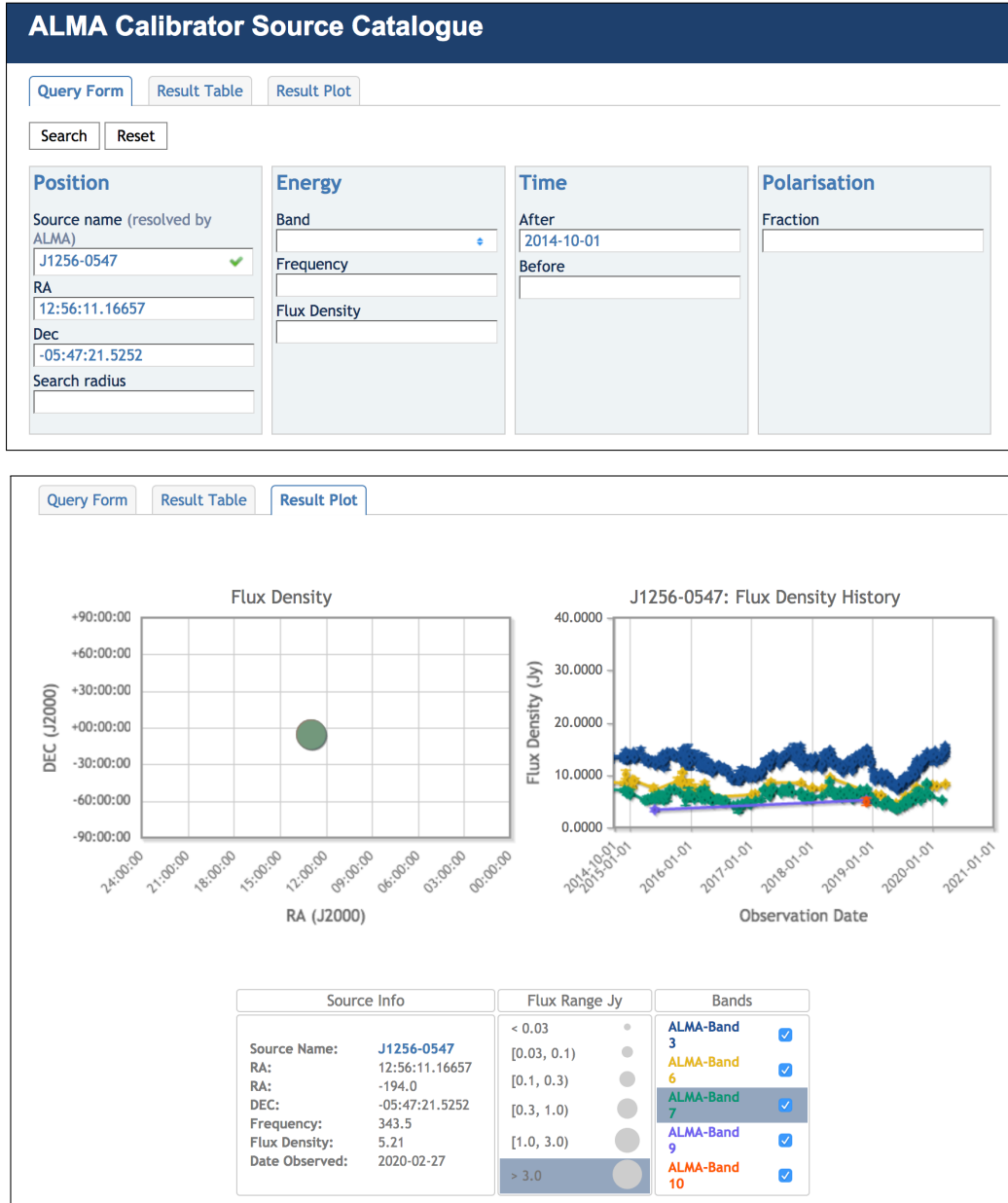


Figure 11.1: Top panel: The query interface to the ALMA Calibrator Source Catalogue. Bottom panel: An example of a plot from the ALMA Calibrator Source Catalogue for J1256-0547.

containing data for each calibrator source. It also generates CASA `setjy` commands that can be used in manual calibration scripts.

In addition, a Flux Service<sup>4</sup> is provided in the Calibrator Source Catalogue via a VO compatible URL query interface. The Flux Service provides an estimate of the flux and spectral index in VO format with uncertainties and quality flags for a given source at a certain date and frequency based on the available flux measurements in the database<sup>5</sup>. The algorithm uses the fit of a linear model for the flux density in time and log frequency space and decides on which of the flux measurements closest to the query date can be included into the fit based on goodness-of-fit criteria. An acceptable result is provided in all cases of well-sampled grid sources. For more sparsely sampled sources, such as phase calibrators, a different approach is required, and the heuristic of the `getALMAFlux` algorithm is employed instead and based on the availability of measurements. Note that the Flux Service is used by the ALMA Science Pipeline as the primary method of estimating the flux densities of flux calibration sources from mid-Cycle 7.

## 11.2.2 Comparing the flux densities of the archival data to the catalogued values

The QA2 products in the `qa` directory of a Member OUS can be used to check the final measured flux densities for all of the observed calibrators in a Member OUS. For manually calibrated data (see Section 5.2), the derived flux densities of the bandpass, phase and check source calibrators can be found in the file `*textfile.txt`. For pipeline calibrated data, the output from the pipeline step `hifa_gfluxscale`, which can be found in the “By Task” section of the WebLog (see Section 5.3), lists the catalogue flux density values for the bandpass calibrator (which may also be the flux calibrator) as well as the catalogue and derived flux densities for the phase calibrator and (if present) check source.

The flux densities for the flux calibration source as listed in the QA2 products should match the flux density value in the ALMA Calibrator Source Catalogue, which can be found either using the web interface or `getALMAFlux` as described above. A reason for disagreement between `getALMAFlux` and the archived value can be further updates to the Calibrator Source Catalogue database since QA2, in which case the later value would be more accurate. If the difference in the flux densities is significantly larger than the typical uncertainties quoted for that frequency band (see Section 11.1), the data may need to be recalibrated. In this case, users are advised to contact their ARC or ARC node for assistance, and it may be appropriate to submit a Helpdesk ticket to initiate a QA3 process (see Section 13.8). For sources other than the flux calibration sources, the discrepancy between the observed flux densities and the values from the catalogues may not be problematic. The flux densities may naturally differ because the other calibration sources are variable and because the catalogue measurements may be from months earlier or later than the observations, especially since phase calibrators and check sources are observed very infrequently.

The catalogued flux density of quasar calibrator sources used in the pipeline reduction for a given ALMA dataset are reported in the file `flux.csv` in the `calibration` directory (see Section 4). For each calibration source, the `flux.csv` file contains one row of information with the measurement set name, field id, spw number, I, Q, U, V flux densities, spectral index and comments. If a grid source was used as the flux calibrator, its reference flux density will be listed.

The ALMA Pipeline User’s Guide provides details on how to reprocess a dataset with different flux density values for the flux calibrator. The different versions of this guide, which correspond to different versions of CASA, are accessible from the ALMA Science Pipeline website<sup>6</sup>.

---

<sup>4</sup>Consult the document [Flux Service of the ALMA Source Catalogue](#) for more information.

<sup>5</sup>An example a Flux Service query is available at <https://almascience.org/sc/flux?DATE=15-Oct-2019&FREQUENCY=233.0E+09&NAME=J0006-0623>

<sup>6</sup>The ALMA Science Pipeline website (for various CASA versions) is available at <https://almascience.org/processing/science-pipeline>.

# Chapter 12

## Useful additional CASA packages

### 12.1 Analysis Utilities

The Analysis Utilities package is a Python module that can be used for further analysis of ALMA data, such as, for example, to produce additional diagnostic plots. Some examples of Analysis Utilities package are the `aU.getALMAflux` command, which can be used to obtain the calibrator flux as described in Chapter 11, or `aU.plotmosaic`, which generates a png image of the mosaic footprint. The Analysis Utilities package together with the installation instructions and documentation of the analysis tasks can be found at the [Analysis Utilities website](#).

To use the Analysis Utilities module, download the package and unpack it into its own directory (labelled 'path-to-analysisUtils' in the example below). The Analysis Utilities module then needs to be called from the CASA command line or in a script. The example below shows the commands to import the module and to execute the `aU.getALMAflux` function in CASA.

```
casa% import sys
casa% import os
casa% sys.path.append('path-to-analysisUtils')
casa% import analysisUtils as aU
casa% aU.getALMAflux('J1924-292', 230, date='2019-12-12', showplot=True,
plotfile='J1924_230.png')
```

### 12.2 CASA Toolkit

The CASA Toolkit is a collection of general modules that can be used for analysing and manipulating measurement set data and images. For example, the MeasurementSet (`ms`) module together with the Table (`tb`) module can be used to manipulate or read parts of measurement sets or calibration tables. The Images module (`ia`) has many functions for accessing, analyzing and creating CASA images, such as computing histograms of image data.

```
casa% ia.open('my_image.image')
casa% res = ia.histograms()
casa% ia.close()
```

Model images can be created using the Images, Coordsys (`cs`) and Componentlist (`cl`) modules. The example below, based on information from the [Simulation Guide Component Lists for CASA 5.1](#), shows how to make an image containing one compact Gaussian source and one more extended Gaussian source. The result is shown in Figure 12.1.

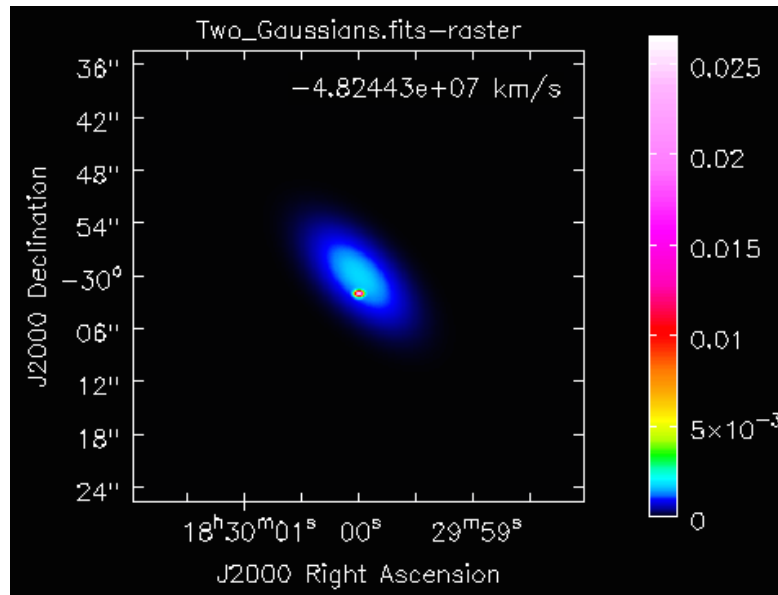


Figure 12.1: The two Gaussian model image created using the CASA Toolkit commands from the Images, Coordsys and Componentlist modules.

```

casa% direction_ext='J2000 18h30m00.0s -30d00m00.0s'
casa% direction_comp='J2000 18h30m00.0s -30d00m02.0s'
casa% direction='J2000 18h30m00.0s -30d00m00.0s'
casa% cl.done()
casa% cl.addcomponent(dir=direction_ext, flux=10.0, fluxunit='Jy', freq='230.0GHz',
    shape='Gaussian', majoraxis='10.0arcsec', minoraxis='5.0arcsec',
    positionangle='45.0deg')
casa% cl.addcomponent(dir=direction_comp, flux=1.0, fluxunit='Jy', freq='230.0GHz',
    shape='Gaussian', majoraxis='0.7arcsec', minoraxis='0.5arcsec',
    positionangle='90.0deg')
casa% ia.fromshape('Two_Gaussians.im', [512,512,1,1], overwrite=True)
casa% cs=ia.coordsys()
casa% cs.setunits(['rad', 'rad', '', 'Hz'])
casa% cell_rad=qa.convert(qa.quantity('0.1arcsec'), 'rad')['value']
casa% cs.setincrement([-cell_rad, cell_rad], 'direction')
casa% cs.setreferencevalue([qa.convert('18h30m', 'rad')['value'],
    qa.convert('-30deg', 'rad')['value']], type='direction')
casa% cs.setreferencevalue('230GHz', 'spectral')
casa% cs.setincrement('1GHz', 'spectral')
casa% ia.setcoordsys(cs.torecord())
casa% ia.setbrightnessunit('Jy/pixel')
casa% ia.modify(cl.torecord(), subtract=False)
casa% exportfits(imagename='Two_Gaussians.im', fitsimage='Two_Gaussians.fits',
    overwrite=True)

```

Model images are very useful as inputs for the CASA task `simobserve`, which produces simulations of ALMA observations. For more information consult the [Guide To Simulating ALMA Data](#).

The CASA toolkit is already part of CASA and requires no specific download or installation. For the description of the various modules and tasks, go to the CASA Toolkit website<sup>1</sup>.

<sup>1</sup>The CASA Toolkit website is available at <https://casa.nrao.edu/docs/casaref/CasaRef.html>.



# Chapter 13

## Frequently encountered issues

This is a list of the most common issues raised by ALMA users and the solutions.

### 13.1 Errors when running the `download*.sh` script

The most frequent errors with the `download*.sh` script (see Section 3.2) are caused by data transfer problems. The `download*.sh` script first attempts to use `wget` to transfer the data. If `wget` does not work, `curl` is used instead. When `curl` fails, it will stop and report a number corresponding to an error that can be found in a list of exit codes given by the `man curl` command.

```
> ... interrupted with error code curl/28
> man curl
```

Some common `curl` error codes are `curl/18` and `curl/28`, which refer to a partial data transfer and an operation time-out, respectively. After checking that enough disk space is available for the data<sup>1</sup> and that the internet connection is functional, the `download*.sh` script should be run again. If the problem persists, please open a Helpdesk ticket (see Chapter 14).

### 13.2 Finding the CASA version used for the data reduction

The `scriptForPI.py` calibration script should always be run in the CASA version it was written for because CASA tasks may change slightly from version to version (see Section 13.4). In addition, the `scriptForPI.py` will throw an error if not run in the appropriate version (see Section 13.3).

The CASA version of the calibration script can be found in one of the following locations.

- In the `README` file of the Member OUS for manually calibrated datasets and for datasets from Cycle 3 and earlier.
- In the `*.scriptForCalibration.py` scripts used for manual calibration (near the top of the script before the first calibration step).
- In the QA2 report PDF available in the `qa` directory, which is mainly available for data from Cycle 5 and later and which is available for both manually and pipeline calibrated data (see Section 5.1).
- At the top of the CASA log files present in the `log` directory or in the `qa/pipeline*/html` directory that is created when untarring the `/qa/*weblog.tgz` file.

---

<sup>1</sup>In the future, the `download*.sh` script will check whether enough disk space is available before downloading.

- On the home page of the WebLog for pipeline calibrated data (see Section 5.3).

For more information, see the Knowledgebase article on [What CASA version was used for QA2?](#).

## 13.3 Problems with running `*.scriptForPI.py`

Section 7.2 explained how to re-calibrate archival ALMA data using the available calibration scripts. However, several problems may arise during this process. The most common problems are described below with a possible solution.

### Not enough disk space

ALMA data can be up to hundreds of GB in size. The calibration procedure typically requires free disk space of about ten times the raw data size. When a typical `*.scriptForPI.py` is executed, it starts by calculating how much disk space CASA will require. The script will crash when this space is not available and will report the reason for the crash. To avoid this problem, the `*.scriptForPI.py` can be executed in CASA using the space saving mode, as shown in the example below.

```
> cd script
> casa -c 'SPACESAVING -N; execfile('scriptForPi.py')'
```

N refers to four options of increasing disk space saving: 0 - no space saving mode; 1 - all intermediate measurement sets `*ms.split` are deleted; 2 - all intermediate measurement sets `*ms.split` and `*ms` are deleted;  $\geq 3$  - all intermediate measurement sets `*ms.split`, `*ms` and `*ms.split.cal` (if possible) are deleted. See the [ALMA QA2 Data Products for Cycle 7 document](#) for more information.

### TypeError

The reason for this error might be that the `*.scriptForPI.py` was executed with the wrong CASA version. See Section 13.2 for information on identifying the correct CASA version.

### NameError

This is a CASA version error specifically produced by CASA version 5.1.1 and higher when restoring the pipeline calibration in the wrong CASA version. To avoid this error, the `*.scriptForPI.py` should always be run in the same version of CASA as it was generated in. See also Knowledgebase article [Why does scriptForPI.py crash with NameError: name hif\\_restoredata is not defined?](#).

### “No such file or directory”

The error message “No such file or directory” is produced by CASA versions of 4.7.0 and higher when executing scripts created for older versions of CASA. Newer CASA versions expect the calibration files in the `calibration` directory to have the extensions `*.tgz` rather than the `*.tar.gz` extensions that were used for files produced using earlier CASA versions. This problem can be fixed by either using the CASA version that corresponds to the one used to generate the scripts or by changing the file extensions to `*.tgz`.

### The calibrated directory

Each time `*.scriptForPI.py` is executed, it will create a `calibrated` directory in the Member OUS directory in which it stores the calibrated data and calibration tables. If this directory already exists when the

`*.scriptForPI.py` is executed, then a “SystemExit” warning will appear reporting that the `calibrated` directory already exists, and the calibrated data will not be regenerated. Removing or renaming the `calibrated` directory solves this problem.

### Pipeline not working

When attempting to run `*.scriptForPI.py` for pipeline calibrated data, CASA should be started using the same version of CASA that was initially used for calibration with the `-- pipeline` option. Otherwise, the error “name 'h\_init' is not defined” will appear.

## 13.4 Issues related to the CASA version used for the data reduction

CASA has evolved and improved over the years, meaning that older versions may have slightly different standards than new versions. In addition, new CASA versions often require relatively recent operating systems to function. The list below describes some of the issues related to changes between CASA versions.

ALMA calibration scripts are written for a specific CASA version, and start typically with a CASA version check. In manual calibration scripts, the CASA version check may be bypassed by commenting out these lines in the script (typically at the top of the file), although the resulting calibration quality is not guaranteed when doing this.

### OS incompatibility

All CASA versions have some dependencies on the OS. Old CASA versions may only operate on outdated OS versions, while the newest CASA version might only run on the latest OSs. Contact your nearest ARC through the ALMA Helpdesk for help with these problems (see Chapter 14).

### Butler-JPL-Horizons models

The Butler-JPL-Horizons models are used to calculate the flux densities of the planets and asteroids that are used for flux calibration. The Butler-JPL-Horizons 2010 models were used up to CASA versions 3.4, but the newer 2012 models, which have been used in CASA version 4.2.7 and later versions, include more spectral lines in the models and more precise continuum models.

It is strongly recommended that data originally calibrated using the 2010 models are re-processed with the 2012 models. The model flux densities have been found to change by 20% between the two versions. A description of how to implement the 2012 models can be found in Knowledgebase article [How can I update Cycle 0 absolute fluxes to the Butler-JPL-Horizons 2012 model standards?](#), and the ARCs may also be contacted via the Helpdesk for assistance (see Chapter 14).

See also Section 7.3 for more information of Solar System flux calibrators Pallas and Ceres.

### Visibility weights in CASA

When multiple measurement sets are combined to produce an image, the relative values of the visibility weights from the different measurement sets are important (see Chapter 9). When the visibility weights are calculated correctly, more noisy *uv*-data points are down-weighted, thus improving the quality of the final images. The measurement set has a `WEIGHT` column where the corresponding weight for every *uv*-data point is written. Tasks that apply calibration tables, such as `gencal` or `applycal`, update these weights when the parameter `calwt=True` is set. It is recommended to always update the weights during data calibration.

The formulas for calculating the visibility weights were changed between CASA 4.2.1 and CASA 4.3, as discussed by the [ALMA memo on data weights](#). Data created using the earlier visibility weighting scheme

cannot be combined with data created using the current scheme without applying corrections to the data. The CASA Guide on [Data Weights and Combination](#) has more detailed instructions on adjusting the visibility weights.

## 13.5 Unit conversion

ALMA flux densities are commonly expressed in Janskys (Jy). The conversion from Jy to cgs units is given below.

$$1 \text{ Jy} = 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1} \quad (13.1)$$

ALMA interferometric images have surface brightness units of Jy/beam. The beam referred to here is the synthesised beam of the final images, and the area of this beam depends on the  $uv$ -coverage of observations and the  $uv$ -weighting scheme as well as any tapering applied in the imaging. The image header gives the full width at half maximum for the major and minor axes  $\theta_{maj}$  and  $\theta_{min}$  for a 2-D Gaussian function fit to the beam. The area enclosed by this beam, the solid angle  $\Omega$ , can be calculated using the equation below.

$$\Omega_{\text{beam}} = \frac{\pi}{4 \ln 2} \theta_{maj} \theta_{min} \quad (13.2)$$

Flux densities  $S_\nu$  in radio astronomy, particularly single dish measurements, are also frequently expressed in terms of (the Rayleigh-Jeans approximation of) brightness temperatures  $T_B$ <sup>2</sup>. The conversion between these two quantities is given by

$$S_\nu = \frac{2k\nu^2}{c^2} \int T_B d\Omega \quad (13.3)$$

where  $\nu$  is the frequency and  $\Omega$  is the solid angle on the sky.

More details unit conversion can be found in the [Observing with ALMA - A Primer](#) and the [ALMA Technical Handbook](#).

## 13.6 Reference frames

ALMA observes in the topocentric reference frame (TOPO), which is the reference frame of the antennas. The frequencies for the raw visibility data are also in the TOPO frame. Consequently, if several EBs are observed at different times using the same spectral set-up, the frequencies of the channels may be slightly shifted in the TOPO frame as a result of changes in the relative velocity between the observatory and the target. If necessary, it is possible to change the reference frame of the measurement set to LSRK or other reference frames using the CASA task `mstransform`.

When the visibilities are imaged, an output reference frame is selected in `tclean` using the parameter `outframe`, and then `tclean` will align the visibility data into the new reference frame. The default option as well as the option used for most ALMA imaging is `outframe='LSRK'`, but in earlier ALMA cycles, extragalactic targets were imaged using `outframe='BARY'`. For spectral lines, the velocities will be calculated relative to a given a rest frequency specified by the `restfreq` parameter.

If the data have been pipeline imaged, the continuum frequency selection reported in the `*continuum.dat` or `*cont.dat` file is in LSRK because the ranges are determined from images in the LSRK frame.

For more information on reference frames, consult the Knowledgebase article [What are the frequency reference frames in CASA?](#).

---

<sup>2</sup>ALMA is sufficiently sensitive and works at sufficiently high frequencies that using a Planck function to convert between surface brightness in Jy/beam and brightness temperature in K would yield different results from the Rayleigh-Jeans approximation.

## 13.7 Leap second table TAI\_UTC is out-of-date

A number of CASA tasks may throw an error stating that the “Leap second table TAI\_UTC seems out-of-date”. Each CASA release contains the latest data (observatory coordinates, leap seconds, ephemerides, calibrator models, Earth-Orientation-Parameters etc.) available at the time of the release. CASA will give this leap second error when the table is out of date for the data being processed, which occurs when the data were observed after the release of the specific version of CASA in use. Although this is a severe error, it does not interrupt any CASA tasks.

The error can be solved by updating the CASA Data Repository either in CASA using

```
casa% !update-data
```

or at the command line using

```
> casa-config --exec update-data
```

Updating the CASA Data Repository is most important for astrometry programmes, but it is generally recommended that these files are kept up to date.

See the CASA Guide on [Updating the CASA data repository](#), the [CASA documentation on this topic](#) or the Knowledgebase article [How can I update the data repository copy of my CASA installation?](#) for more information.

## 13.8 Problems with delivered ALMA data and QA3

When a Member OUS is discovered to have problems with the calibration, imaging or observational setup, the user may request that the data be re-calibrated. This is called Quality Assurance 3 (QA3). The request should be sent as a Helpdesk ticket to the ARC, and the ARC will then forward the request to JAO, where it will be evaluated and resolved. This may include re-observation, re-calibration or re-imaging of the data. If the data are re-calibrated, the re-calibrated and re-imaged data will be ingested into the ASA, and the old data will be removed. For more information, consult Section 8.4 in the [ALMA Users' Policies](#).



# Chapter 14

## Assistance with Archive mining and science exploitation

Multiple resources are available to assist ALMA users with their data. Some of these have already been mentioned in this guide, but they are listed again here for clarity.

### ALMA Helpdesk

The ALMA Helpdesk is located at <https://help.almascience.org>. The Helpdesk provides users with a line of communication to ask questions that will be reviewed and replied to by ARC staff. Users need to login to create tickets or view previously-submitted tickets. To troubleshoot an issue related to the data or data reduction, as much information as possible should be provided, including but not limited to the project code, the SB name, the CASA version being used, the command used when the error was encountered, the entire error message and an accompanying CASA log file. The activation of Target-of-Opportunity observations is also done through the ALMA Helpdesk.

### Knowledgebase

The ALMA Knowledgebase is a directory of articles and known issues that may be of help to general users. It includes, for example, the formulas for converting flux measurements from Jy/beam to Kelvin and instructions on identifying the version of CASA used to process data. The Knowledgebase can be accessed from the Helpdesk or directly through <https://help.almascience.org/kb>.

### ALMA Regional Centres (ARCs)

The ARCs can provide face-to-face support for proposal preparation, data reduction and archival research.

To arrange a visit, please review this Knowledgebase article [How do I arrange a visit to one of the ARCs?](#). Funding for face-to-face visits is also available via different avenues depending on region.

For more general information about the ARCs, please visit their websites.

- The [East Asian ARC](#) is based at the National Astronomical Observatory of Japan (NAOJ), which has its headquarters in Mitaka, Tokyo, Japan. Taiwan has an [EA ARC station](#) at ASIAA in Taipei, and Korea has an [EA ARC station](#) at KASI in Daejeon.
- The [European ARC](#) is organised as a coordinated network of scientific support nodes distributed across Europe. The central node is located at ESO Headquarters in Garching bei München, Germany. The support nodes offering face-to-face user support are located in Bonn-Cologne (Germany), Bologna (Italy),

Onsala (Sweden), IRAM-Grenoble (France), Leiden (The Netherlands), Manchester (United Kingdom) and Ondřejov (Czech Republic). Users based in Europe can request a face-to-face visit for support with their archival research, regardless of the existence of a local node.

- The [North American ALMA Science Center](#) (NAASC) is located at the NRAO headquarters in Charlottesville, Virginia, USA. The North American ARC is a partnership between NRAO, the National Research Council of Canada (NRC), which is supported by the Millimeter Astronomy Group at the Herzberg Institute of Astrophysics, and the National Science Council of Taiwan (NSC).

## Documentation from the science portal

The ALMA Science Portal contains a documentation section<sup>1</sup>. The documents for the current cycle, the manual for the ALMA Science Archive, and other guides related to ALMA data products, data processing and CASA can be found on the main documentation page. However, documents from previous cycles are also archived in the portal.

## CASA documentation

The latest CASA documentation is available at <https://casadocs.readthedocs.io/en/stable/>, and its content is updated with every CASA release. The website contains descriptions of CASA tasks with a detailed description of the input parameters, information on CARTA, and the CASA Memo Series and Knowledgebase as well as many additional pages discussing various other topics.

## CASA Guides

ALMA CASA Guides are available at <https://casaguides.nrao.edu/index.php?title=ALMAGuides> and contain imaging and analysis examples based on ALMA datasets, including detailed guides to the calibration and imaging of some of the publicly available ALMA Science Verification data. Many of these CASA Guides have been referenced throughout this document.

## ALMA Science Archive Notebooks

A series of Jupyter Notebooks have been created that show how Python (version 3) can be used to query the archive and download data. The notebooks primarily focus on using pyVO, although one of the notebooks includes examples based on astroquery. These notebooks can be accessed from <https://almascience.org/alma-data/archive/archive-notebooks>.

## Video tutorials

The Interactive Training in Reduction and Analysis of INterferometric data (I-TRAIN) series is a series of recorded training sessions covering various topics related to working with ALMA data, including using the ALMA Science Archive itself. The videos in this series are maintained at <https://almascience.org/tools/eu-arc-network/i-train>.

The [ALMA Primer channel](#) on YouTube has a set of videos that discuss many basic interferometry concepts as they apply to ALMA. The topics include discussions related to proposal preparation, such as sensitivity estimation, and discussions about data processing, such as image and cell sizes, as well as topics such as  $(u, v)$  coverage that are pertinent to both proposal preparation and data processing.

---

<sup>1</sup>ALMA Documents and Tools can be found at <https://almascience.org/documents-and-tools>.



# Appendix A

## Acronym List

**ACA:** Atacama Compact Array  
**ADMIT:** ALMA Data Mining Toolkit  
**AGB:** Asymptotic Giant Branch  
**ALMA:** Atacama Large Millimeter/submillimeter Array  
**AOS:** Array Operations Site  
**ARC:** ALMA Regional Centre  
**ARI-L:** Additional Representative Images for Legacy  
**ASA:** ALMA Science Archive  
**ASCII:** American Standard Code for Information Interchange  
**ASDM:** ALMA Science Data Model  
**ASIAA:** Academia Sinica Institute of Astronomy and Astrophysics  
**BARY:** Barycentric reference frame  
**CARTA:** Cube Analysis and Rendering Tool for Astronomy  
**CASA:** Common Astronomy Software Applications  
**CO:** Carbon Monoxide  
**CSV:** Comma Separated Values  
**DDT:** Directors Discretionary Time  
**EA ARC:** East Asian ALMA Regional Centre  
**EB:** Execution Block  
**ESO:** European Southern Observatory  
**FAQ:** Frequently Asked Questions  
**FWHM:** Full Width at Half Maximum  
**Group OUS:** Group Observation Units Set  
**GUI:** Graphical User Interface  
**html:** hypertext markup language  
**I-TRAIN:** Interactive Training in Reduction and Analysis of INterferometric data  
**JAO:** Joint ALMA Observatory  
**KASI:** Korea Astronomy and Space Science Institute  
**LSRK:** Kinematic Local Standard of Rest  
**Member OUS:** Member Observing Unit Set  
**mfs:** multi-frequency synthesis  
**MRS:** Maximum Recoverable Scale  
**mtmfs:** multi-term multi frequency synthesis  
**NA ARC:** North American ALMA Regional Centre  
**NAASC:** North American ALMA Science Center  
**NAOJ:** National Astronomical Observatory of Japan  
**NED:** NASA/IPAC Extragalactic Database  
**NRAO:** National Radio Astronomy Observatory  
**NRC:** National Research Council of Canada

**NSC:** National Science Council of Taiwan  
**OS:** Operating System  
**OSF:** Operations Support Facility  
**OUS:** Observation Units Set  
**pb:** primary beam  
**pbcor:** primary beam corrected  
**PDF:** Portable Document Format  
**PI:** Principal Investigator  
**png:** portable network graphics  
**PSF:** Point Spread Function  
**pyVO:** python Virtual Observatory **PWV:** Precipitable Water Vapour  
**QA:** Quality Assurance  
**RMS:** Root Mean Square  
**SB:** Scheduling Block  
**SG:** Science Goal  
**SIMBAD:** Set of Identifications, Measurements and Bibliography for Astronomical Data  
**spw:** spectral window  
**TOPO:** Topometric reference frame  
**TP:** Total Power  
**TSV:** Tab Separated Values  
**URL:** Uniform Resource Locator  
**VLBI:** Very Long Baseline Interferometry  
**VO:** Virtual Observatory  
**VOTable:** Virtual Observatory Table  
**WVR:** Water Vapour Radiometer  
**XML:** Extensible Markup Language

# Appendix B

## Further reading

### ALMA documents

ALMA Science Archive Manual

<https://almascience.org/documents-and-tools/latest/science-archive-manual>

ALMA Proposer's Guide

<https://almascience.org/documents-and-tools/latest/alma-proposers-guide>

ALMA Technical Handbook

<https://almascience.org/documents-and-tools/latest/alma-technical-handbook>

ALMA QA2 Data Products for Cycle 7

<https://almascience.org/documents-and-tools/cycle7/alma-qa2-data-products-for-cycle-7>

ALMA Users' Policies

<https://almascience.org/documents-and-tools/latest/documents-and-tools/alma-user-policies>

Flux Service of the ALMA Source Catalogue

<https://almascience.org/documents-and-tools/latest/flux-service-of-the-alma-source-catalogue>

Knowledgebase article repository

<https://help.almascience.org/kb>

Observing with ALMA - A Primer

<https://almascience.org/documents-and-tools/latest/alma-science-primer>

Repository of ALMA Science Pipeline User's Guides and Reference Manuals for various CASA versions

<https://almascience.org/processing/science-pipeline#docs>

Repository of Interactive Training in Reduction and Analysis of INterferometric (I-TRAIN) sessions

<https://almascience.org/tools/eu-arc-network/i-train>

### CASA Guides

3C286 Polarization

[https://casaguides.nrao.edu/index.php?title=3C286\\_Polarization](https://casaguides.nrao.edu/index.php?title=3C286_Polarization)

ALMA CASA Guides

<https://casaguides.nrao.edu/index.php?title=ALMAGuides>

ALMA Imaging Pipeline Reprocessing

[https://casaguides.nrao.edu/index.php/ALMA\\_Imaging\\_Pipeline\\_Reprocessing](https://casaguides.nrao.edu/index.php/ALMA_Imaging_Pipeline_Reprocessing)

Analysis Utilities

[https://casaguides.nrao.edu/index.php?title=Analysis\\_Utilities](https://casaguides.nrao.edu/index.php?title=Analysis_Utilities)

Automasking Guide

[https://casaguides.nrao.edu/index.php/Automasking\\_Guide](https://casaguides.nrao.edu/index.php/Automasking_Guide)

Data Weights and Combination

<https://casaguides.nrao.edu/index.php/DataWeightsAndCombination>

First Look at Imaging

[https://casaguides.nrao.edu/index.php/First\\_Look\\_at\\_Imaging](https://casaguides.nrao.edu/index.php/First_Look_at_Imaging)

First Look at Line Imaging

[https://casaguides.nrao.edu/index.php/First\\_Look\\_at\\_Line\\_Imaging](https://casaguides.nrao.edu/index.php/First_Look_at_Line_Imaging)

First Look at Self Calibration

[https://casaguides.nrao.edu/index.php/First\\_Look\\_at\\_Self\\_Calibration](https://casaguides.nrao.edu/index.php/First_Look_at_Self_Calibration)

Guide to Simulating ALMA Data

[https://casaguides.nrao.edu/index.php/Guide\\_To\\_Simulating\\_ALMA\\_Data](https://casaguides.nrao.edu/index.php/Guide_To_Simulating_ALMA_Data)

Image Continuum

[https://casaguides.nrao.edu/index.php/Image\\_Continuum](https://casaguides.nrao.edu/index.php/Image_Continuum)

Imaging Mosaics

[https://casaguides.nrao.edu/index.php/AntennaeBand7\\_Imaging\\_5.4#Imaging\\_Mosaics](https://casaguides.nrao.edu/index.php/AntennaeBand7_Imaging_5.4#Imaging_Mosaics)

M100 Band 3 (ALMA Science Verification observations)

[https://casaguides.nrao.edu/index.php/M100\\_Band3](https://casaguides.nrao.edu/index.php/M100_Band3)

Self-Calibration Template

[https://casaguides.nrao.edu/index.php/Self-Calibration\\_Template](https://casaguides.nrao.edu/index.php/Self-Calibration_Template)

Simulation Guide Component Lists (CASA 5.1)

[https://casaguides.nrao.edu/index.php?title=Simulation\\_Guide\\_Component\\_Lists\\_\(CASA\\_5.1\)](https://casaguides.nrao.edu/index.php?title=Simulation_Guide_Component_Lists_(CASA_5.1))

Solar System Models in CASA 4.0

[https://casaguides.nrao.edu/index.php?title=Solar\\_System\\_Models\\_in\\_CASA\\_4.0](https://casaguides.nrao.edu/index.php?title=Solar_System_Models_in_CASA_4.0)

Spectral Index Imaging

[https://casaguides.nrao.edu/index.php/EVLA\\_Advanced\\_Topics\\_3C391#Spectral\\_Index\\_Imaging](https://casaguides.nrao.edu/index.php/EVLA_Advanced_Topics_3C391#Spectral_Index_Imaging)

Updating the CASA Data Repository

[https://casaguides.nrao.edu/index.php/Updating\\_the\\_CASA\\_Data\\_Repository](https://casaguides.nrao.edu/index.php/Updating_the_CASA_Data_Repository)

VLA Self-calibration Tutorial CASA 5.7.0

[https://casaguides.nrao.edu/index.php?title=VLA\\_Self-calibration\\_Tutorial-CASA5.7.0](https://casaguides.nrao.edu/index.php?title=VLA_Self-calibration_Tutorial-CASA5.7.0)

## References

[Advanced Gain Calibration Techniques in Radio Interferometry](#), Brogan, Hunter & Fomalont, 2018, arXiv:1805.05266

[Fourier Plane Image Combination by Feathering](#), Cotton, 2017, PASP, 129, 094501

[astroquery: An Astronomical Web-querying Package in Python](#) Ginsburg et al., 2019, AJ, 157, 98

[Total Power Map to Visibilities \(TP2VIS\): Joint Deconvolution of ALMA 12m, 7m, and Total Power Array Data](#) Koda, Teuben, Sawada, Plunkett & Fomalont, 2019, PASP, 131, 4505

[The Additional Representative Images for Legacy \(ARI-L\) Project for the ALMA Science Archive](#), Massardi et al., 2021, PASP, 133, 085001

[ALMA Long Baseline Campaigns: Phase Characteristics of Atmosphere at Long Baselines in the Millimeter and Submillimeter Wavelengths](#) Matsushita et al., 2017, PASP, 129, 035004

[CASA Architecture and Applications](#) , McMullin, Waters, Schiebel, Young & Golap, 2007, ASPC, 376, 127

[ALMA Data Quality Assurance and the Products it Delivers - The Contribution of the European ARC](#), Petry et al., 2020, The Messenger, 181, 16

[A Joint Deconvolution Algorithm to Combine Single-dish and Interferometer Data for Wideband Multiterm and Mosaic Imaging](#), Rau, Naik & Braun, 2019, AJ, 158, 3

[Interferometry and Synthesis in Radio Astronomy](#), Thompson, Moran & Swenson, 2017, 3rd Edition, Springer



The Atacama Large Millimeter/submillimeter Array (ALMA), an international astronomy facility, is a partnership of Europe, North America and East Asia in cooperation with the Republic of Chile. ALMA is funded in Europe by the European Organization for Astronomical Research in the Southern Hemisphere (ESO), in North America by the U.S. National Science Foundation (NSF) in cooperation with the National Research Council of Canada (NRC) and the National Science Council of Taiwan (NSC) and in East Asia by the National Institutes of Natural Sciences (NINS) of Japan in cooperation with the Academia Sinica (AS) in Taiwan. ALMA construction and operations are led on behalf of Europe by ESO, on behalf of North America by the National Radio Astronomy Observatory (NRAO), which is managed by Associated Universities, Inc. (AUI) and on behalf of East Asia by the National Astronomical Observatory of Japan (NAOJ). The Joint ALMA Observatory (JAO) provides the unified leadership and management of the construction, commissioning and operation of ALMA.

