

ALMA Science Pipeline Quickstart Guide



www.almascience.org

ALMA, an international astronomy facility, is a partnership of Europe, North America and East Asia
in cooperation with the Republic of Chile.

For further information or to comment on this document, please contact your regional Helpdesk through the ALMA User Portal at 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	September 2014	Pipeline Team
1.1	December 2014	Pipeline Team

In publications, please refer to this document as:

ALMA Pipeline Team, 2014, ALMA Science Pipeline Quickstart Guide, Version 1.1, ALMA

Table of contents

1	Overview	4
2	Obtaining the Pipeline	4
3	Pipeline-related Documentation	4
4	Producing Calibrated Measurement Sets from ALMA Pipeline Deliveries: scriptForPI.py	4
5	The Pipeline Interferometry Calibration Script: casa_pipescript.py	5
5.1	Running casa_pipescript.py from ALMA Delivery Packages	5
5.2	Imaging after using casa_pipescript.py	6
5.3	Modifying the Pipeline Run	7
5.3.1	Note on the Pipeline Mode.....	7
5.3.2	Removing ASDMs from the processing	7
5.3.3	Introducing additional flagging	7
5.3.4	Editing calibrator fluxes used by the Pipeline.....	7
6	The Pipeline WebLog	7
6.1	Navigation	8
6.2	Home Page	8
6.2.1	Observation Summary.....	8
6.3	By Topic	9
6.4	By Task	9
6.4.1	Task sub-pages	10
6.5	Selecting and filtering plot displays	10
6.6	Weblog Quality Assessment (QA) Scoring	11

1 Overview

The ALMA Science Pipeline is being developed for the automated processing of ALMA interferometry and single-dish data, and for the combination of data taken using multiple ALMA arrays. At present, the Pipeline is commissioned for the calibration of the majority of ALMA interferometry datasets and for the provision of diagnostic calibrator images. However, data combination from multiple arrays, science target imaging and single-dish processing are not yet commissioned. This document describes how to obtain the ALMA Pipeline, how to use it to calibrate ALMA data and use of the Pipeline WebLog, the set of plots output by the Pipeline.

The Pipeline calibrates and images ALMA data automatically using Pipeline tasks. Pipeline tasks use CASA tasks wherever possible to perform the data reduction, and Pipeline tasks can be viewed and executed within CASA in exactly the same way as CASA tasks. For example, in a CASA version also containing the Pipeline, to view the inputs possible for Pipeline task `hifa_importdata` type: `inp hifa_importdata`.

The Pipeline is data-driven i.e. the characteristics of each dataset drives the calibration and imaging strategy (the **Pipeline Heuristics**). During the Pipeline run, information on which calibration tables should be used etc are stored in the **Pipeline Context**.

2 Obtaining the Pipeline

The Pipeline can be obtained by downloading a version of CASA that also contains the Pipeline. This is available, along with installation instructions, from the **Documents & Tools** Section of the **ALMA Science Portal**: (<http://www.almascience.org/>). If any issues are encountered with Pipeline installation, please contact the ALMA Helpdesk via the link on the ALMA Science Portal.

3 Pipeline-related Documentation

The User documentation currently relating to Pipeline, available from the ALMA Science Portal, is:

- **ALMA Science Pipeline Quickstart Guide**: Overview of the ALMA Science Pipeline (this document)
- **ALMA Science Pipeline Reference Manual**: Description of individual Pipeline tasks
- **ALMA QA2 Data Products for Cycle 2**: Description of files included in ALMA deliveries

4 Producing Calibrated Measurement Sets from ALMA Pipeline Deliveries: `scriptForPI.py`

If the Pipeline was used during the ALMA Quality Assurance Process, this will be noted in the **README** file of an ALMA delivery. In order to convert ALMA raw data into calibrated Measurement Sets, for example before re-imaging the data, then the script `scriptForPI.py` should be used. `scriptForPI.py` directly applies calibration and flagging tables determined during the ALMA Quality Assurance process to raw data to create calibrated measurement sets. The instructions for using this script will be in the **README** file of the ALMA delivery.

Using `scriptForPI.py` is the recommended and fastest method of obtaining calibrated ALMA data from the delivery. For deliveries that used the ALMA Pipeline, `scriptForPI.py` must be run from within a version of CASA that contains the Pipeline.

5 The Pipeline Interferometry Calibration Script: `casa_pipescript.py`

To make any alterations to the calibration and flagging performed during the ALMA Quality Assurance Process, data have to be fully re-calibrated using the script `casa_pipescript.py` (instead of using `scriptForPI.py`). `casa_pipescript.py` is a python script for performing full data re-calibration using the Pipeline (typical script shown in Figure 1). Note that the Pipeline works on completed Observing Unit Sets (defined in the ALMA Technical Handbook), such that multiple executions of the same Scheduling Block (execution blocks, ASDMs) can be calibrated using a single script to produce a calibrated Measurement Set per ASDM. Information on individual Pipeline tasks is in the **ALMA Science Pipeline Reference Manual**.

5.1 Running `casa_pipescript.py` from ALMA Delivery Packages

To reproduce the Pipeline calibration performed as part of the ALMA Quality Assurance Process:

- Copy `casa_pipescript.py` from the `script` directory to the `raw` directory
- Copy `flux.csv` and `*flagtemplate.txt` from the `calibration` directory to the `raw` directory

In the `raw` directory:

- Make sure the naming of the raw ALMA data is consistent with those provided in the script (e.g. if the data ends in `.asdm.sdm` then move to names which do not have this suffix)
- Start the version of CASA containing Pipeline using `casapy --pipeline`, then type `execfile('casa_pipescript.py')`

Running the script will create:

- a calibrated MS for each ASDM in the same directory
- test images of the bandpass and phase calibrator (1 per spectral window in `*.image` format. To view a `*.image` file e.g. use `casaviewer image_file_name`)
- a `pipeline*/html` directory containing
 - the Pipeline WebLog - the html pages of plots from the calibration process. Access using e.g. `firefox index.html`
 - `casa_commands.log` –a list of the CASA tasks used by Pipeline tasks during calibration

Note that to re-run the Pipeline multiple times, it is recommended to start each time from a clean directory containing only the raw data, `flux.csv` and `*flagtemplate.txt` files and the `casa_pipescript.py` script. If the `flux.csv` and `*flagtemplate.txt` files are not present in the directory, Pipeline will create new default versions of these files, which will not contain any edits made to them by ALMA staff during the Quality Assurance Process. Also note that it is advisable to have 8-10 GB RAM and 50-75 GB disk space per ALMA raw data file (ASDM) available to perform the Pipeline calibration. Please contact ALMA via the Helpdesk if assistance is needed with data reprocessing.

Figure 1: Example Pipeline Calibration Script

```
# Example Pipeline Data Reduction Script (casa_pipescript.py)
__rethrow_casa_exceptions = True

h_init()

try:
    hifa_importdata(vis=['uid__A002_X72e960_X53d',
'uid__A002_X72e960_X53f'], session=['session_1', 'session_1'])

    hifa_flagdata(pipelinemode="automatic")

    hifa_fluxcalflag(pipelinemode="automatic")

    hif_refant(pipelinemode="automatic")

    hifa_tsyscal(pipelinemode="automatic")

    hifa_tsysflag(pipelinemode="automatic")

    hifa_wvrgcalflag(pipelinemode="automatic")

    hif_lowgainflag(pipelinemode="automatic")

    hif_setjy(pipelinemode="automatic")

    hif_bandpass(pipelinemode="automatic")

    hif_bpflagchans(pipelinemode="automatic")

    hifa_gfluxscale(pipelinemode="automatic")

    hifa_timegaincal(pipelinemode="automatic")

    hif_applycal(pipelinemode="automatic")

    hif_makecleanlist(intent='PHASE,BANDPASS,CHECK')

    hif_cleanlist(pipelinemode="automatic")

finally:
    h_save()
```

5.2 Imaging after using casa_pipescript.py

In order to use the imaging script provided by the ALMA Observatory after using **casa_pipescript.py**, then the science spectral windows must be “split” out from the calibrated measurement sets and the measurement sets output with a .split.cal suffix.

The science spectral windows are specified in the Pipeline Weblog (*Home > Observation Summary > Measurement Set Name > Spectral Setup*, in the ID column) or can be determined using the CASA task **listobs** e.g. `listobs('uid__A002_X89252c_X852.ms')` and results will be in the CASA logger.

To perform the split, in CASA e.g.:

```
split('uid__A002_X89252c_X852.ms',outputvis='uid__A002_X89252c_X852.ms.split.cal',spw='17,19,21,23')
```

If the script named **scriptForFluxCalibration.py** is present in the script directory, this must also then be executed prior to running the imaging script.

5.3 Modifying the Pipeline Run

5.3.1 Note on the Pipeline Mode

In the Pipeline standard calibration script, the Pipeline Mode is defined as “automatic” for each task. In this mode, the task takes the default settings of Pipeline and only a limited number of parameters are exposed for editing by a user. Setting the pipeline mode to “interactive” will usually enable the values of a larger number of parameters to be changed. See the **ALMA Science Pipeline Reference Manual** for more details.

5.3.2 Removing ASDMs from the processing

ASDMs can be removed from the processing by editing the **vis=** and **session=** lists in **hifa_importdata** in **casa_pipescript.py**.

5.3.3 Introducing additional flagging

Additional manual flagging can be introduced to any Pipeline reduction by editing the ***flagtemplate.txt** files. Examples of the syntax to use in editing these files are given at the top of the files. The flags are applied when **hifa_flagdata** is run.

5.3.4 Editing calibrator fluxes used by the Pipeline

Pipeline currently reads quasar calibrator fluxes from the **flux.csv** file. If a regularly-monitored quasar has been used as the flux calibrator, the flux of this provided from the ALMA Quality Assurance process can be over-ridden by editing it in this file. Only values for the flux calibrator can be over-ridden, since values for the bandpass and phase calibrators are derived during data calibration

6 The Pipeline WebLog

The WebLog is a set of html pages that give a summary of how the calibration of ALMA data proceeded. The WebLog produced during the ALMA Quality Assurance Process will be in the **qa** directory of an ALMA delivery. To view the WebLog, untar and unzip the file using e.g. **tar zxvf *weblog.tar.gz**. This will provide a **pipeline*/html** directory containing the WebLog, which can be viewed using a web browser e.g. **firefox index.html**. The WebLog aims to provide both a quick overview of datasets and also give methods for exploring the calibration and flagging in detail. Therefore most calibration pages of the WebLog will first give

a single “representative” view of that calibration (or flagging) step. Users can then choose whether to further select a detailed view of all the plots associated with that calibration step. These plots can easily be filtered by a combination of outlier, antenna and spectral window criteria. Calibration and flagging steps also undergo an automated scoring to give an “at a glance” indication of any trouble points. Throughout the WebLog, links are denoted by text written in blue and it is usually possible to click on thumbnail plots to enlarge them. Where histograms are displayed, in modern web browsers it is possible to draw boxes on multiple histograms to select the plots associated with those data points.

Observation Overview

Project	uid://A001/Xa0/X17e5
Principal Investigator	
OUS Status Entity id	uid://A002/X5d7935/X1d7
Observation Start	2013-12-29 10:24:03 UTC
Observation End	2013-12-29 10:48:40 UTC

Pipeline Summary

Pipeline Version	30425 (Pipeline-Cycle2-R1-B)
CASA Version	4.2.2 r30135
Pipeline Start	2014-07-14 13:34:56 UTC
Execution Duration	1:50:04

Observation Summary

Measurement Set	Receivers	Num Antennas	Time (UTC)			Baseline Length			Size
			Start	End	On Source	Min	Max	RMS	
Observing Unit Set Status: uid://A002/X5d7935/X1d7 Scheduling Block ID: uid://A002/X5d7935/X1c6									
Session: session_2									
uid__A002_X7776f1_X1ddf.ms	ALMA Band 3	27	29/12/2013 10:24:03	29/12/2013 10:48:40	0:08:24	17.3 m	1.3 km	375.0 m	2.7 GB

Figure 2: WebLog Home Page. The navigation bar is circled in red.

6.1 Navigation

To navigate the main pages of the WebLog, click on items given in the bar at the top of the WebLog home page. Also use the **Back** button provided at the upper right on some of the WebLog sub-pages. Avoid using “back/previous page” on your web browser.

6.2 Home Page

The first page in the WebLog gives an overview of the observations (proposal code, data codes, PI, observation start and end time), a pipeline execution summary, and an observation summary. Clicking on the bar at the top of the home page enables navigation to **By Topic** or **By Task**.

6.2.1 Observation Summary

The Observation Summary table lists all the measurement sets included in the pipeline processing. Each measurement set is calibrated independently by the pipeline. The table provides a quick overview of the ALMA receiver band used, the number of antennas, the start/end date and time, the time spent on source, the array minimum and maximum baseline length, the rms baseline length and the size of that measurement set. To view the observational setup of each measurement set in more detail, click on the name of it to go to its overview page.

6.2.1.1 Measurement set overview pages

Each measurement set overview page has a number of tables: **Observation Execution Time**, **Spatial Setup**, **Antenna Setup**, **Spectral Setup** and **Sky Setup**. For more information on the tables titled in blue text, click on these links. There are additionally links to **Weather** and **Scans** information. Two thumbnail plots, which can be enlarged by clicking on them, show the observation structure either as **Field Source Intent vs Time** or **Field Source ID vs Time**. To view the CASA listobs output from the observation click on **listobs output**.

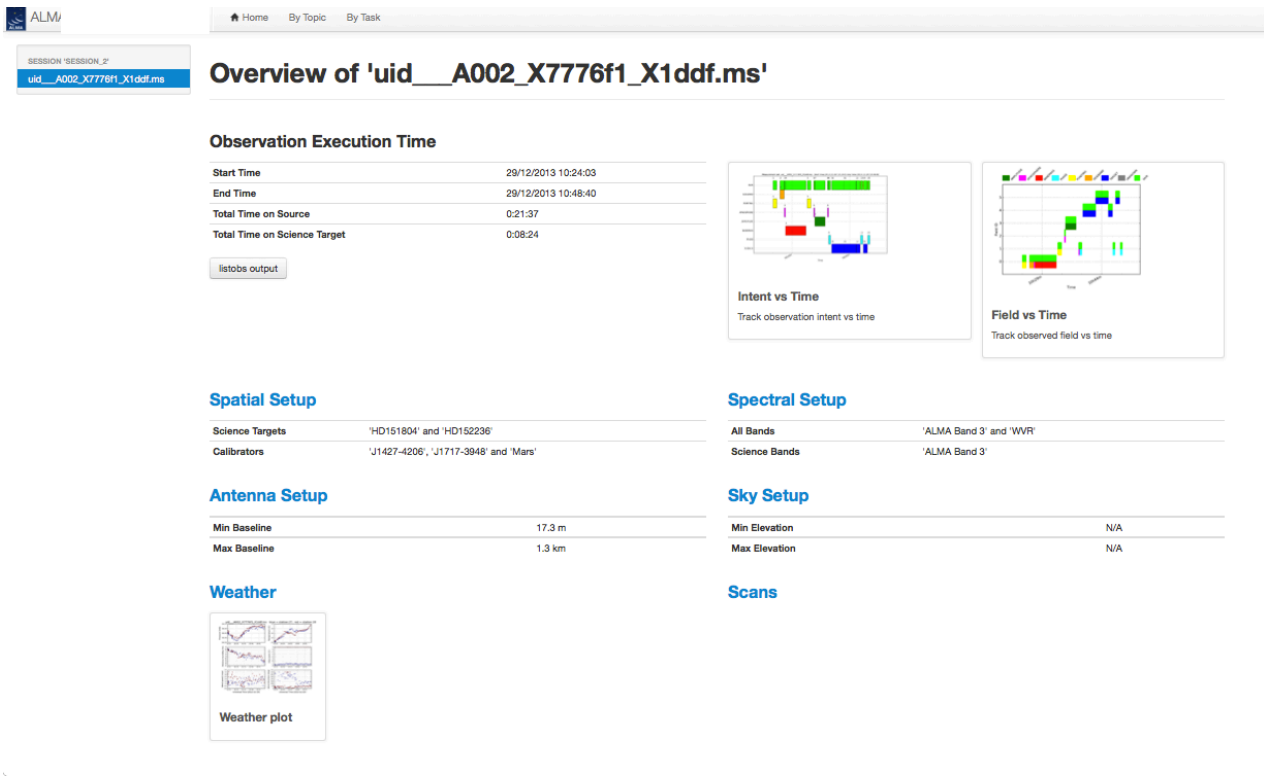


Figure 3: A Measurement Set Overview Page. Click on the table headings in blue to see more information about each.

6.3 By Topic

The **By Topic** page provides an overview of all **Warnings and Errors** triggered, a Quality Assessment overview in **Tasks by Topic** and **Flagging Summaries** for the processing.

6.4 By Task

The **By Task Summary** page gives a list of the calibration and flagging steps performed on the dataset. It is not displayed per measurement set as the Pipeline performs each step on every measurement set sequentially before proceeding to the next step e.g. it will import and register all measurement sets with the Pipeline before proceeding to perform the ALMA deterministic flagging step on each measurement set. The name of each step on the By Task page is a link to more information. On the right hand side of the page, the

bars and scores indicate how well the Pipeline processing of that stage went. All green bars should indicate a fairly problem-free dataset.

6.4.1 Task sub-pages

The task sub-pages (accessed by clicking on e.g. hifa_importdata) provide the outcome, or the representative outcome, of each Pipeline task executed. Additionally by clicking on the appropriate blue links on each sub-page, the **CASA Log**, **Pipeline QA**, **Input Parameters** and **Task Execution Times** can be accessed. Most sub-pages have further links in order to access a more detailed view of the outcome of each task. These links are often labelled by the measurement set name. For a fast assessment of results, go straight to the **By Task > hif_applycal** page.

6.5 Selecting and filtering plot displays

Using the **By Task > hifa_tsysflag: Flag Tsys calibration pages** as an example.

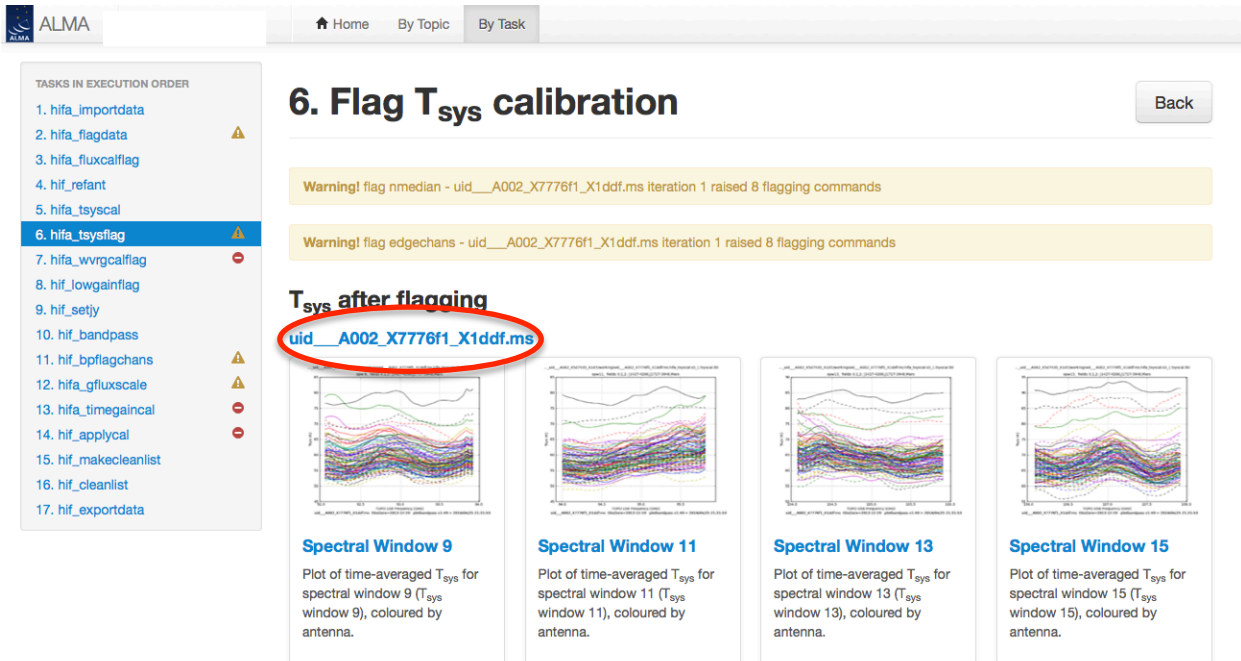
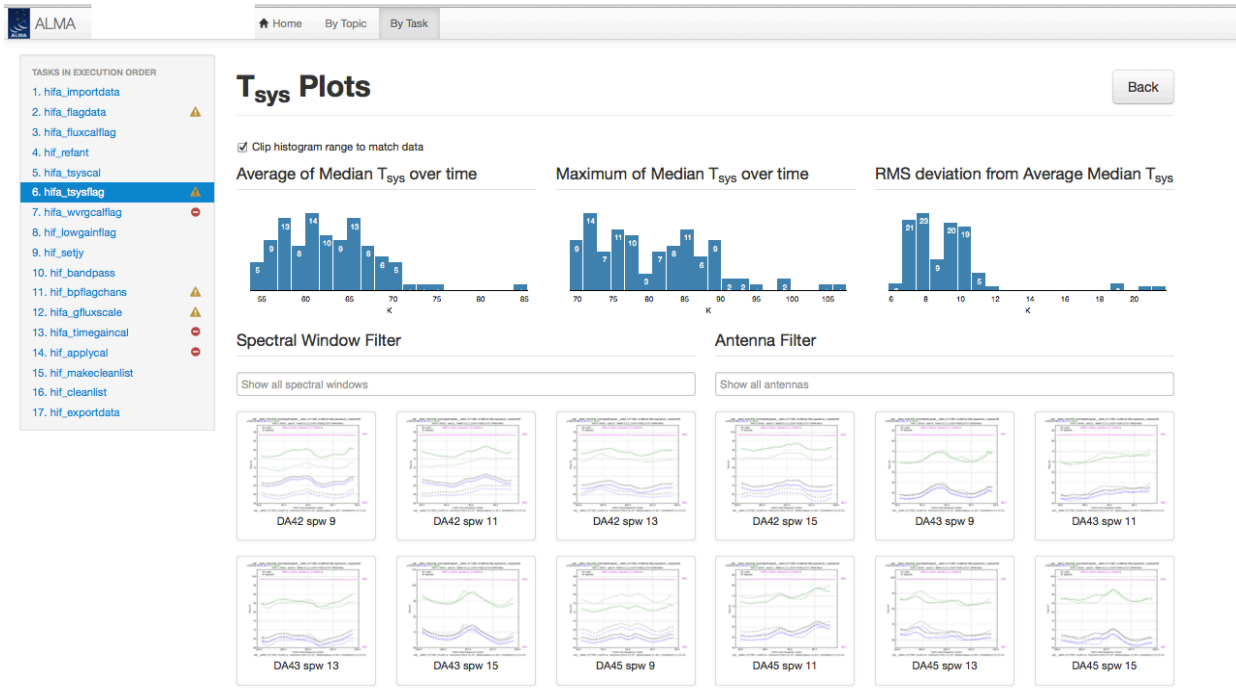


Figure 4: Flag Tsys Calibration Overview Page. A representative view of the Tsys measurements for each spectral window is shown. To see the measurements for individual antenna/spectral windows, click on the measurement set name, circled here in red.

Figure 5: The Tsys plots for all antennas. Above the plots are histograms indicating the distributions of metrics based



on the median Tsys.

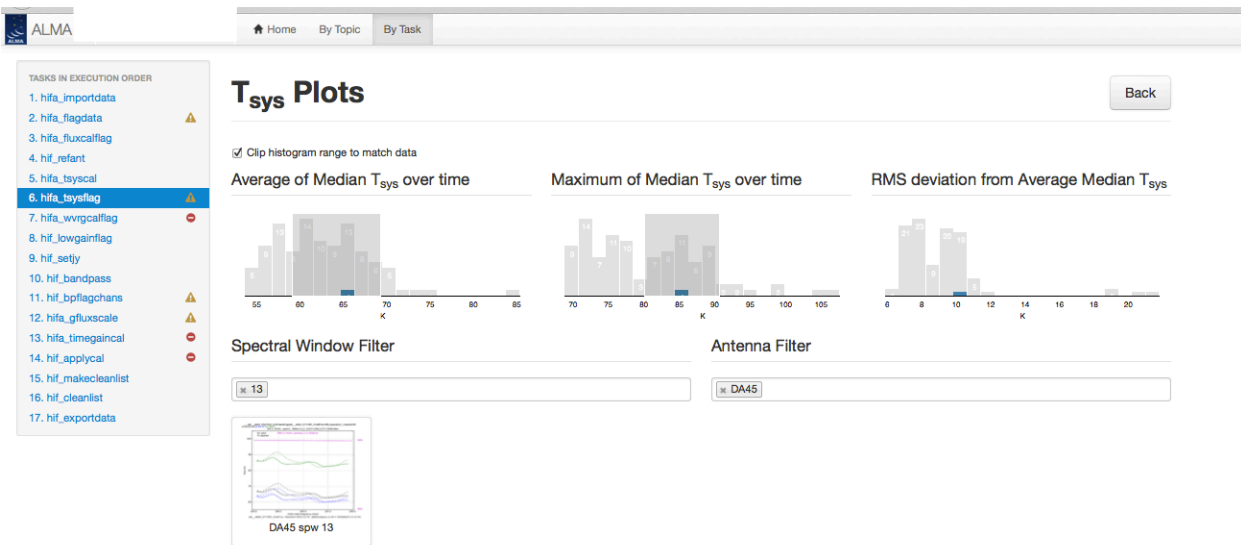


Figure 6: Filtering to the plot of interest by drawing grey boxes on two of the histograms using a mouse button, and adding additional spectral window and antenna filters. To clear the grey box filters on the histograms, click on white space in the histograms.

6.6 Weblog Quality Assessment (QA) Scoring

Pipeline tasks have scores associated with them in order to quantify the quality of the dataset and the

calibration. The scores are between 0.0 and 1.0 and are colourized such that:

Score	Colour	Comment
0.90-1.00	Green	Standard/Good
0.66-0.90	Blue	Below standard
0.33-0.66	Yellow	Warning
0.00-0.33	Red	Error

Pipeline Task	Pipeline QA Scoring Metric	Score
hifa_importdata	Checking that the required calibrators are present	1.0 all present
		0.1 subtracted for missing bandpass or flux calibrator
		1.0 subtracted for missing phase calibrator or Tsys calibration
		0.5 subtracted for existing processing history
hifa_flagdata	Determining percentage of incremental flagging	Shadowing: $0 < \text{score} < 1 = 50\% < \text{fraction flagged} < 20\%$
		Others: 0%-5% -> 1.0, 5%-50% -> 1.0...0.5, >50% -> 0.0
hifa_fluxcalflag	Determining percentage of incremental flagging	0%-5% -> 1.0, 5%-50% -> 1.0...0.5, >50% -> 0.0
hif_refant	Determining if a reference antenna could be found	1.0 if reference antenna could be determined, 0.0 otherwise
hifa_tsysflag	Determining percentage of incremental flagging	0%-5% -> 1.0, 5%-50% -> 1.0...0.5, >50% -> 0.0
hifa_wvrgcalflag	Checking phase RMS improvement	0.0 if $\text{RMS}(\text{before})/\text{RMS}(\text{after}) < 1$, 0.5 ... 1.0 for ratios between 1 and 2, and 1.0 for ratios > 2
hif_lowgainflag	Determining percentage of incremental flagging	0%-5% -> 1.0, 5%-50% -> 1.0...0.5, >50% -> 0.0
hif_bandpass	Judging phase and amplitude solution flatness per antenna, spectral window	two algorithms: Wiener entropy and derivative deviation, and signal-to-noise ratio (scores: Wiener entropy: error function with 1-sigma deviation of 0.001

	and polarization (interim measure, future scoring will be based on data with solutions applied)	from 1.0; derivative deviation: error function with 1-sigma deviation of 0.03 for the outlier fraction; signal-to-noise ratio: error function with 1-sigma deviation of 1.0 for the signal-to-noise ratio)
hifa_gfluxscale	Determining SNR of fitted flux values	fitted flux values with SNR < 5.0 are assigned a score of 0.0, SNR > 20.0 a score of 1.0, and a linearly scaled value in between
hifa_timegaincal	Determining X-Y / X2-X1 phase solution deviations	Standard deviation of X-Y phase difference converted to path length: 1.0 if lower than 4.25e-6 m, 0.0 if higher than 7.955e-2 m, with an exponential decrease in between. Standard deviation of X2-X1 phase differences of subsequent integrations converted to path length: 1.0 if lower than 3.08-e-5 m, 0.0 if higher than 2.24e-2 m, with an exponential decrease in between. NB: The high limits are currently dummies. Determining their realistic values is still under development.
hifa_applycal	Determining percentage of incremental flagging	0%-5% -> 1.0, 5%-50% -> 1.0...0.5, >50% -> 0.0



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.

