Statistics from Cycle 5 ALMA Observations

NAASC Memo #119

Author: Andrew Lipnicky

Date: 9th October 2020

ABSTRACT

This study provides quantitative statistics regarding the types of projects carried out in Cycle 5, investigates the success of completed observations, and makes suggestions based on these findings for improvements to observatory procedures, data reduction workflows, and data archiving and accessibility. In addition to substantial aggregation and distillation of metadata, more in-depth analyses were performed regarding to the success of the Observatory in carrying out the goals of Principal Investigators. This memo is broken into several sections each addressing a different aspect of Observatory operations, and recommendations based on those findings are suggested.

Statistics from Cycle 5 ALMA Observations

Andrew Lipnicky
National Radio Astronomy Observatory

Submitted: October 14, 2019, Published: October 9, 2020, Revised: May 14, 2026*

Contents

*Please note that the bulk of this work was carried out in early 2019, but because of the Covid-19 pandemic the publication of this work was significantly delayed. Therefore, several of the conclusions of this work are made obsolete by newly implemented practices and other concurrent studies, with references to such work noted throughout.

Executive Summary

This memo documents an internal study that was initiated in 2019 soon after the completion of Cycle 5. In late 2025, a minor revision of this document was undertaken with the intention to release the memo as a historical reference. Since many changes and improvements have been made to ALMA systems and observing strategies in the intervening cycles since Cycle 5, some of this memo's conclusions may no longer be current.

The purpose of this study was to provide quantitative statistics about the different types of projects that were carried out in Cycle 5, investigate the success of the observations that were carried out, and make suggestions based on these findings. While much of this report is simple fact finding, more in-depth analyses were performed relating to the success of the Observatory in carrying out the goals of Principal Investigators (PI). This memo is broken into several sections, each addressing a different concept of the Observatory and recommendations based on those findings are suggested.

In the first two sections, an explanation of how ALMA projects are setup and how the data for this study were obtained is provided. First, a global view of the data is provided. Relative numbers of projects based on their scientific category and the number of observations based on the type of observation are presented. Since the Observatory has both standard and non-standard modes, these observation types represent different levels of effort when it comes to carrying out the observation, data reduction, and quality assurance. I found that about 92% of all datasets were able to be pipeline processed with the remaining 8% of projects mostly consisting of high-frequency and polarization data (Table 2). High frequency data are now being pipeline processed, but until Cycle 8, observing practices did not allow for polarization data to go through the data reduction pipeline. Therefore, it was recommended to update the observing strategy to include a bandpass calibrator scan in every Execution Block so that polarization data can be pipeline calibrated.

The next section features information about the different observation setups that were carried out including sensitivity requirements, spectral setups, number of sources observed, and the number of pointings performed per source. Sensitivity requirements define how many repeated observations must be carried out to reach the desired sensitivity. Here it is shown that most observations only require a single visit with the main 12m array, while the 7m and TP arrays require roughly 5 and 6 times (respectively) more time on source, which closely follows the time-multipliers enforced by the ALMA Observing Tool (Figure 2, Table 3).

PIs have a lot of freedom to set up their observations in many different ways to capture specific spectral lines of interest or simply as much bandwidth as possible. This allows PIs to have anywhere from 1 to 16 spectral windows in observing bands 3 through 10. The ALMA correlator provides 4 spectral windows automatically without penalty to bandwidth or resolution due to data rate concerns or technical limitations, yet I found that a small number of observations (about 4%) were carried out with less than 4 spectral windows (Figure 4). As these additional spectral windows come at no additional "cost" and can help with calibration and benefit later archival studies, it is recommended that either the ALMA Observing Tool enforce a minimum of four spectral windows per observation or that the Observatory adds additional spectral windows to PI observations that include less than four spectral windows.

I also investigated the number of channels (Figure 5), total bandwidth (Figure 6), the number of sources per observation setup (Figure 7), and the number of pointings per source (Figure 10). All of these settings, along with the number of spectral windows (Figure 4), have a direct impact on the total data rate coming from the telescope, the imaging load on the data reduction pipeline, and the final data volume that must be stored in the archive. Without data volume mitigation enforced by the pipeline, projects with a large number of images associated with them (large number of spectral windows and large number of sources) could impact quality assurance efforts due to large processing times (Figures 8 & 9). A full investigation is necessary to fully grasp the requested computing resources of the pipeline that incorporates the number of channels, spectral windows, and pointings.

The next section analyzes the observations that were carried out in comparison with what the PI requested. Here, I investigated the interesting interplay of how the array configuration and the total number

of antennas used impacted how well the Observatory natively (i.e. without imaging weighting techniques) met the resolution request of the PI. I found that 41% of observations appear to have been observed in a configuration that is one or more configurations away from the PI requested configuration (Figure 15). Accounting for the fact that actual configurations rarely represent the ideal configurations and directly computing the theoretical achieved resolution instead still yields that 37% of observations were carried out with a configuration that did not meet the PI requested resolution (Figure 16). Interestingly, a large majority of those observations that were observed in a different configuration than requested appear to be grade A and B projects indicating that scheduling algorithms that weight by project grade may not be appropriate. These results indicate that the metrics used by the Observing Tool, scheduling algorithms, and quality assurance could be optimized. It is recommended to consider whether the scheduling algorithm could be changed to weight schedulable projects less on grade and more on achievable angular resolution.

I also investigated how well the Largest Angular Scale (LAS) criteria is met by our observations which is a parameter indicated by the PI but is not a formal quality assurance criterion. By calculating the achieved Maximum Recoverable Scale, as defined by the ALMA Technical Handbook, of each observation and comparing it with the PI defined LAS, I found that for observation sets that require multiple configurations, >35% of them appear to fail to meet the PI goal. Observations that only require a single array fail to meet the PI LAS goal for about 3% of observations (Table 4). Looking in depth at these 3% of projects shows that a combination of configuration and having a large number of antennas may lead to the deviations seen (Figures 11 & 20). This was likely caused by the scheduling algorithm previously ignoring extra antennas above the nominal 43 when considering possible observing blocks, and scheduling algorithms have since improved. With respect to projects requiring multiple configurations, continued investigation is recommended in order to characterize flux recovery across scales probed by different configurations in preparation for Group-level imaging.

Finally, the metadata database where all the data for this project was taken contains a wealth of information from which we can learn and improve the science capable with ALMA. However, because each database table is independently managed and documented internally, it is challenging for independent investigators within ALMA to use the database to its full potential. Suggestions are provided for documentation and use cases of the metadata database to improve its utility.

1 Introduction

The ALMA telescope is incredibly diverse in its range of available observing modes, frequency range, frequency and spatial resolutions, and sensitivity. Likewise, the data reduction pipeline that has been developed to process the large quantity of data that ALMA creates must be extremely versatile. While the pipeline has been created to handle most PI science cases ("standard" modes), there are still some observing modes that are not pipeline supported ("non-standard" modes). As the observatory expands and offers more non-standard modes as standard[1] the pipeline must be further developed to handle these cases. Therefore, it is important to know where significant time is being spent manually calibrating and imaging these currently non-standard modes. Additionally, knowing quantitatively the number of projects that are carried out in non-standard modes informs both the pipeline working group and the Observatory where to best direct future efforts of pipeline development.

Another facet of having such a diverse range of capabilities is being able to carry out successful observations for all accepted PI science. As an observatory, we perform multiple levels of Quality Assurance (QA) to ensure that the data recorded meet the science goals set by the PI. These goals are the sensitivity and resolution of the final delivered image to which we guaranteed in Cycle 5 a sensitivity measurement of 10% or better in Bands 3 – 6, 15% in Bands 7 – 8, or 20% in Bands 9 – 10 and the final 2-D beam size to be within 20% of the requested resolution on both axes. There are three main stages of QA: QA0 where the data integrity and usability is checked to ensure that the data were successfully recorded and no major problems occurred during the observation, QA1 where the ALMA system performance is tracked and calibrator measurements are kept up-to-date, and QA2 where all the QA0 passed data for a set of observations are calibrated and combined to make the final image. If the sensitivity and/or resolution of the final image do not meet the PI's goals, the project is made available in the queue again. Due to the wide range of resolutions that ALMA offers, the array configuration is modified frequently over the course of a Cycle, often only staying in a configuration for one to two weeks before being moved again. Also, due to technical issues or weather, the array is commonly in "hybrid" configurations that involve most antennas in a nominal configuration with a few outlier antennas. Combined with changing weather conditions, these factors make scheduling observations that successfully meet the 20% beam requirement extremely difficult.

The purpose of this study is to provide quantitative statistics about the different types of projects that were carried out in Cycle 5, investigate the success of the observations that were carried out, and make suggestions based on the findings. I chose Cycle 5 since it is the most recent Cycle (at the time of writing) for which almost all the metadata currently exist. In the following section, data acquisition will be covered. Following that, I go through the results and the relative numbers of projects in various forms, how PI's set up their observations, and how those observations were carried out. I will also go into detail about how the data relate to PI goals, and I discuss ongoing and future work related to scheduling and data processing.

1.1 Project Setup

In order to facilitate the following discussions, here is a brief description of the ALMA project setup including the various acronyms that will be used throughout this text. During the Call for Proposals, users (PIs) use the Observing Tool (OT) to set up an Observing Unit Set (OUS) composed of individual "Science Goals" (SG, but also referred to as SGOUS in this text) that they wish to observe. Each SGOUS has a science target(s), a spectral window setup, and performance criteria. The performance criteria is defined by three values: the desired RMS sensitivity, the Largest Angular Scale (LAS) of the science target, and the desired angular resolution. Once a project is submitted, it is given a project identification code (PID).

Each SGOUS then has at least one Group OUS (GOUS) below it. If a SGOUS has specified many targets, these targets can be split into different GOUSs in order to improve observation efficiency and ensure that a standard set of calibrators can work for all science targets. Additionally, time variability studies are often split into many GOUSs that all have the same spectral setup and performance parameters. In all cases, all GOUSs inherit the performance criteria and spectral setup of the parent SGOUS.

[1] In fact, the terminology has evolved and the use of "standard"/"non-standard" is no longer employed as it was before in the ALMA project.

Below the GOUS level are the Member OUSs (MOUS). These are the individual array setups that are required to meet the PI's specified LAS, resolution, and sensitivity. In the absence of an LAS criteria (or an LAS criteria only marginally larger than the requested resolution), only a single MOUS may be required to meet the PI's goals. However, if the PI specifies a large LAS, then up to four MOUSs may be needed to meet this goal, an extended 12M array configuration (TM1), a compact 12M array configuration (TM2), a 7M ACA configuration (7M), and a total power (TP) configuration. Furthermore, additional MOUSs may be contained below the GOUS for different sources in special cases.

Scheduling Blocks (SBs) are then created for each MOUS in order to carry out the observations. SBs are carried out through the use of Execution Blocks (EBs) since it can take many sets of observations to meet the required sensitivity criteria. The data from an EB are called the ASDM (ALMA Science Data Model) which is the actual raw data obtained.

Finally, in order to track all parts of a project, each part from the PID to the ASDM is given a unique identifier (uid).

2 Data Acquisition

All data for this study were obtained via SQL queries of the internal ALMA database which allowed direct access to the ALMA metadata. Eight tables within the ALMA database were used. The data obtained only includes QA0 Pass data, which ensured that only data which went through either pipeline or manual processing were captured. Some information about non-standard projects (mostly related to Table 2) was gathered from the internal ALMA Project Tracker.

Much of the data gathered in the analysis here can be gathered via queries of the public facing ALMA Science Archive using the online user interface or standard tools such as astroquery[2]. Internal queries of the ALMA database were used for consistency and for gathering information related to specific PI requests of resolution and LAS, as well as project grades.

Data cleaning ensured that duplicate entries were removed. The SQL query returned about 120,000 rows which is roughly equivalent to the number of individual pointings performed by the ALMA array during the acquisition of Cycle 5 data. Data in all sections were spot checked against other internal ALMA tracking systems and the Alma Science Archive to ensure that the correct values were being obtained. When possible or applicable, values were also checked against the ALMA Pipeline weblog results. When data were found to be missing or inaccurate the correct values were calculated as explained in the relevant text.

3 Global Properties

3.1 Project Category

When PIs submit projects, they are required to choose a scientific category. Table 1 displays the number of projects and their respective breakdowns by science category type. We see that solar projects are the least common with only 7 projects in all, while ISM/Star Formation are the most common. In Figure 1, we see how the observations of those projects are spread across the sky.

3.2 Project Types

Table 2 shows project breakdowns by type in categories of "standard" and "non-standard" where a project is considered standard if it could be reduced by the ALMA Pipeline at the time of the report. This is perhaps the most relevant breakdown in terms of informing ALMA Pipeline development. In Cycle 5, the following were considered non-standard modes:

- Bands 8, 9, and 10 observations

[2]Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al., AJ, 2019, 157, 98

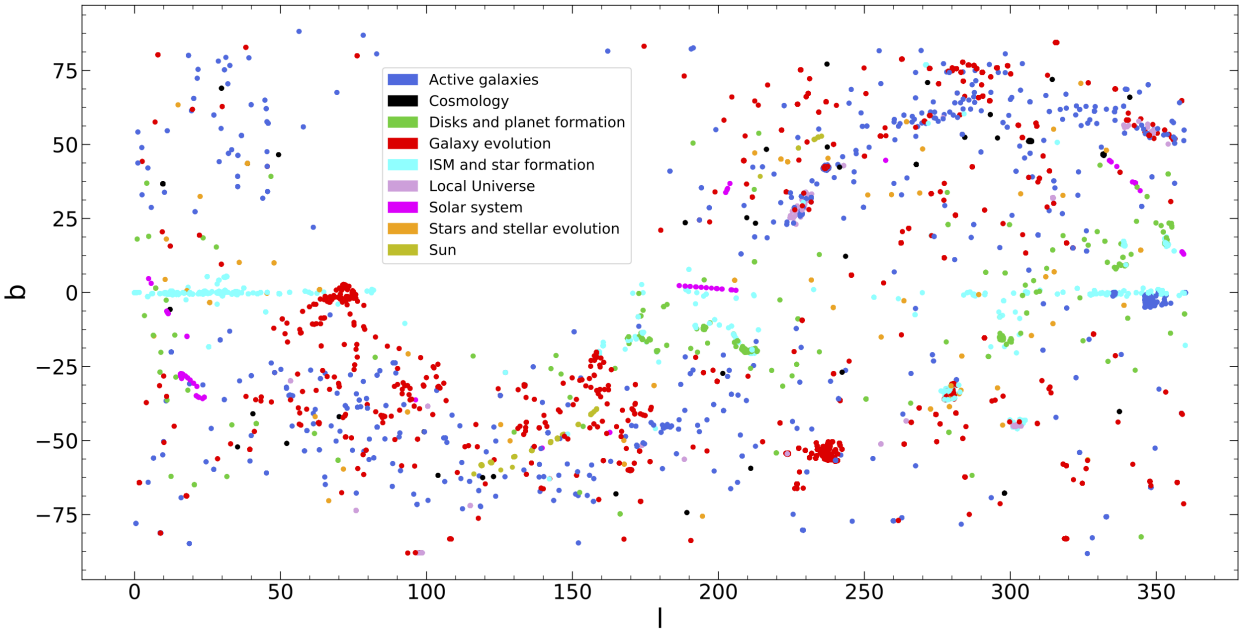

Figure 1: Every ALMA science target pointing from QA0 Passed data in galactic coordinates colored by each project's scientific category.

- Band 7 observations with maximum baselines longer than 5 km (i.e. configurations C-8, C-9, and C-10)
- All full polarization observations
- Spectral scans
- Bandwidth switching projects (having less than 1 GHz aggregate bandwidths over all spectral windows)
- Solar observations
- VLBI observations
- Non-standard calibrations (user-defined calibrations selected in the OT)

Table 1: Total number of projects (PID), science goals (SGOUS), groups (GOUS), members (MOUS), and observations (ASDMs) carried out for Cycle 5 in each scientific category.

Scientific Category	PID	SGOUS	GOUS	MOUS	ASDM
Active galaxies	123	657	663	751	1411
Cosmology	22	61	75	76	190
Disks and planet formation	88	184	216	236	529
Galaxy evolution	86	479	563	823	2323
ISM and star formation	151	333	428	705	2538
Local Universe	11	113	113	138	441
Solar system	16	45	60	65	146
Stars and stellar evolution	39	89	137	154	228
Sun	7	17	17	30	194
Total	543	1978	2272	2978	8000

- Astrometric observations

Table 2: Standard and non-standard modes broken down by project type. Numbers given are the total number of unique PID, SGOUS, GOUS, MOUS, and ASDMs of that type. Since data sets are processed and QA'ed at the MOUS level, the percent of each type versus the total (from Table 1) is given in the MOUS column. Note that because MOUSs can occupy several categories at once, the percentages do not add to 100. The data for spectral scan, ephemeris, user-defined calibrations, bandwidth switching, astrometric, and VLBI projects was found via flags used in the Project Tracker; therefore, it is possible that the numbers are higher if flags were missed.

Mode	Type	PID	SGOUS	GOUS	MOUS (%)	ASDM
Standard	Standard	471	1806	2067	2747 (92.2)	7269
	Long-baseline[1]	110	208	220	220 (7.4)	428
	Ephemeris[2]	7	16	16	16 (0.5)	52
	Mosaic	97	419	429	663 (22.3)	2219
	Target-of-Opportunity	14	39	72	72 (2.4)	111
Non-standard	High Frequency[3]	58	100	101	110 (3.7)	291
	Polarization[4]	24	33	42	45 (1.5)	159
	Solar	7	17	17	30 (1.0)	194
	VLBI	9	10	10	16 (0.5)	102
	Spectral Scan	4	10	32	32 (1.1)	41
	User-defined Calibrations	0	-	-	- (-)	-
	Bandwidth Switching	0	-	-	- (-)	-
	Astrometric	0	-	-	- (-)	-

(1) Long-baseline refers to configurations C-7 and above.
(2) Ephemeris refers to non-solar projects that require an external ephemeris.
(3) High-Frequency projects are considered Band 7 in configuration C-8 or longer and Bands 8, 9, and 10.
(4) Note that while VLBI measurements are full polarization, they are not included in these values.

As can be seen in Table 2, the vast majority of project data sets (92.2%) were able to go through the main ALMA data reduction pipeline. The most obvious candidate for pipeline development and validation would be high frequency observations which consist of nearly 4% of all data sets. As of Cycle 6, Band 8 projects were considered standard, which is a significant fraction of all high frequency observations (see Figure 3). Higher frequency data sets have also gone through pipeline calibration but often need manual intervention due to excessive flagging performed by the pipeline. The remaining non-standard project types appear to consist of about 4% of all projects.

After making high frequency projects standard, the next candidate for pipeline development would likely be adding the ability to process polarization data sets which consist of about 1.5% of all data sets. At the time of writing, polarization MOUSs were carried out in observing sessions that required back-to-back executions of the SB. Between two and four EBs are necessary in order for the polarization calibrator to cover enough range in parallactic angle for proper polarization calibration. Because the observations are carried out back-to-back it is inefficient to observe the bandpass calibrator in neighboring EBs since the solutions from the first EB can be applied to later EBs. Therefore, some of the EBs are purposely missing calibration scans which makes the calibration pipeline (which calibrates each execution individually) unable to process those EBs which are missing calibration scans. Given the current practices of polarization observations and the unique calibration and imaging processes needed to make pipeline calibration possible, it would take considerable development effort. Instead, it is recommended to update the observing strategy to always include bandpass scans in each Execution Block at the expense of some observing efficiency (bandpass scans are typically 5-10 minutes), as this would allow initial pipeline calibration to be possible.[3] With little effort,

[3]As of Cycle 8, this will indeed be the case in order to facilitate data reduction with the pipeline.

this would allow pipeline calibration for another 1-2% of projects and speed up the QA2 process as well. It must be noted that this will only allow initial calibration of the data, rather than full polarization calibration, but it would improve the calibration process overall and decrease the QA2 time.

4 Observation Setups

4.1 Sensitivity Requirements

Table 3 shows the number of MOUSs and ASDMs per array "family" (i.e. the 12M interferometric array, the 7M ACA array, and the TP antenna). It is obvious that the 12M array is the most "popular"; however, the 7M and TP arrays have significantly more ASDMs per MOUS. This mainly occurs due to the difference in collecting area between the different arrays and tracks closely with the overall time-multipliers enforced by the OT and outlined in the ALMA Proposer's Guide. Figure 2 shows a histogram of the number of ASDMs per MOUS. As expected, the majority of projects have only 1 or a few ASDMs. About 10% of MOUSs have more than 5 ASDMs, 3% have more than 10 ASDMs, and less than 1% have more than 20 ASDMs with the maximum value being 105 ASDMs in a single MOUS.

Table 3: Number of MOUSs and ASDMs observed in each array family.

Array Type	Number MOUSs	Number ASDMs	ASDMS per MOUS
12M	1963	3571	1.82
7M	697	2497	5.02
TP	318	1932	6.08

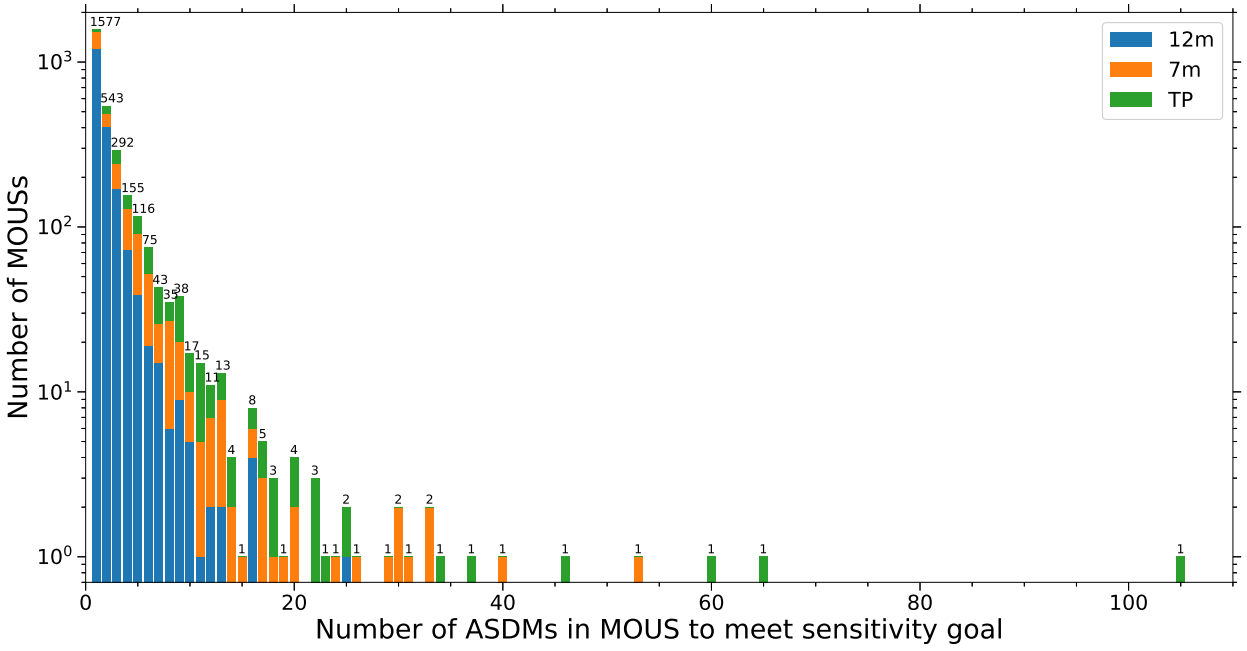

Figure 2: The number of ASDMs that have been observed in a MOUS in order to meet the sensitivity goal. Each array family is given it's own color with 12M, 7M, and TP colored as blue, orange, and green respectively. The total number of MOUSs in a column is displayed in text at the top of each column. Note that the y-axis is in log scale.

4.2 Spectral Setup

Figure 3 shows that Bands 3 and 6 are the most selected. High frequency observations (Bands 8, 9, and 10) occupy only 3.2% of all MOUSs observed,[4] with Band 10 observations being very rare due to both weather constraints and a lower number of approved projects overall.

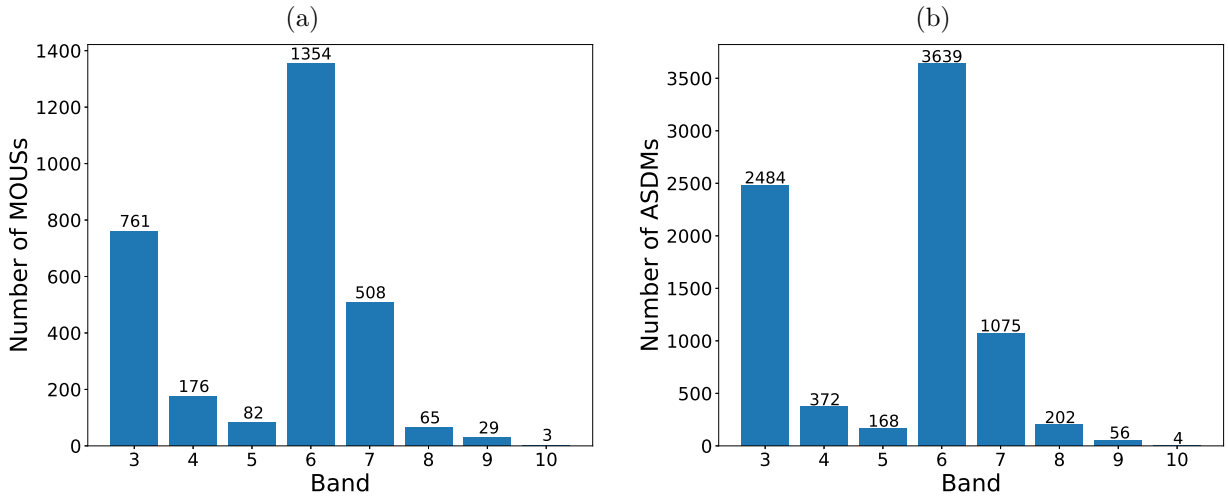

Figure 3: The number of MOUSs (a) and ASDMs (b) observed in each ALMA Band. The value for each column is given at the top of each bar.

As for how PIs setup their projects, Figure 4 shows that the majority (78.7%) of PIs use 4 spectral windows to set up their observations. Additionally, we see that 4.3% of PIs set up their MOUS observations with between 1 and 3 spectral windows which corresponds to about the same fraction of ASDMs (3.6%). MOUSs that contain only one spectral window may hinder calibration, which sometimes requires combining several spectral windows to obtain sufficient signal-to-noise for phase and bandpass solutions; this is especially the case for high frequency observations where strong calibrators are more scarce. Additionally, having more than one spectral window allows for better understanding of flux measurements and their relative errors within the data set. Finally, obtaining data in additional spectral windows does not require additional time on source. Although less than 1% of MOUSs contain only a single spectral window, employing additional spectral windows in such cases would improve calibration and enhance archive products.

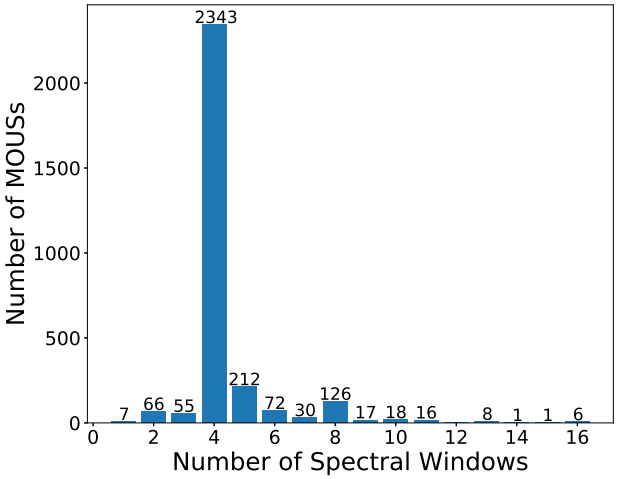

Figure 4: Histogram of the number of MOUSs that use a number of spectral windows.

Another consideration is the number of channels used during the observation of a MOUS. Increasing the number of channels impacts the data volume and memory requirements during pipeline imaging as well as the data download rate while the observations are being performed. Figure 5 shows a histogram of the total number of channels requested. There is a large peak at 512 channels which mostly corresponds

[4]Note that this is consistent with the value of 3.7% reported in Table 2 since that value also includes Band 7 observations taken in configuration C-8 and larger, but is not considered here.

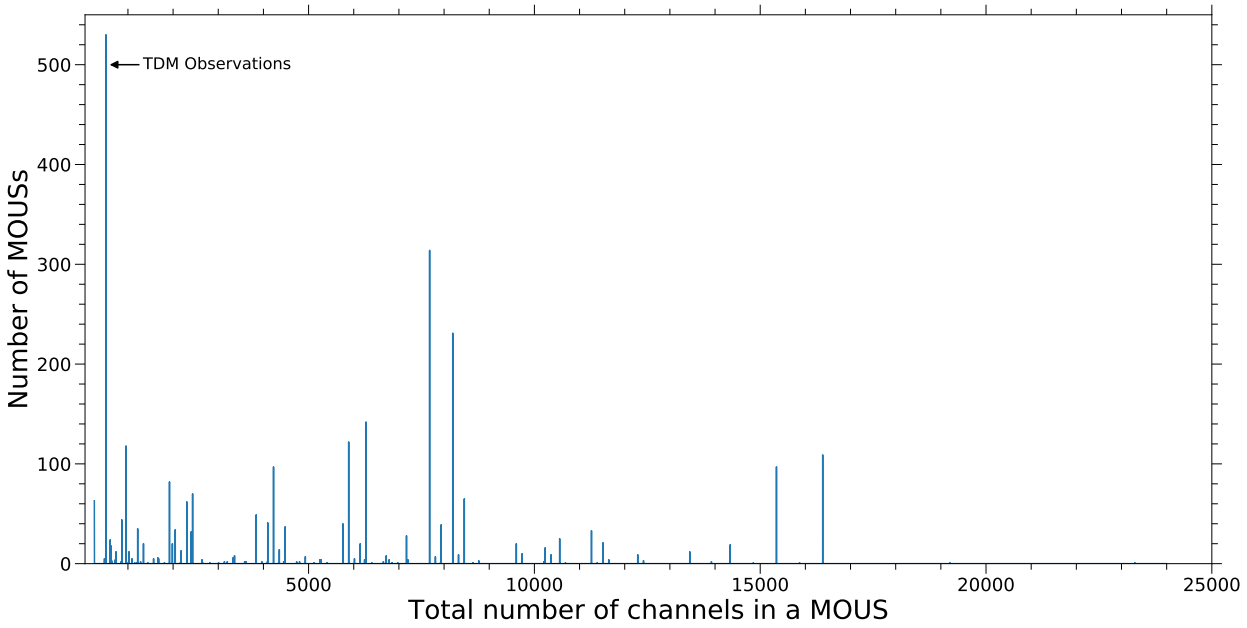

Figure 5: Histogram of the total number of channels in a MOUS in bins of 1. Time-domain measurements (TDM) correspond to projects that requested 128 channels in 4 spectral windows which is pointed out explicitly as the spike at 512. The largest request is a single MOUS at 23,296 channels.

to TDM (time-division measurement) or continuum studies. In total, 25% of all MOUSs are continuum studies (identified as having a representative bandwidth >1.5 GHz). This indicates that most observations contain at least 1 or more FDM (frequency-division measurement) spectral windows. The largest channel request is 23,296 channels and corresponds to a MOUS that requested 3 spectral windows with a single linear polarization (XX) which increases the number of available channels to 7680.

Increasing the number of channels in a spectral window increases the frequency/velocity resolution of the observation; however, this is often at the cost of overall bandwidth coverage. Figure 6 shows a histogram of total bandwidth for each MOUS. Most MOUSs use the full 8 GHz that is available through either FDM coarse resolution or TDM observations. The second most common bandwidth is 7.5 GHz which corresponds to 3 full bandwidth windows and likely 1 higher resolution FDM window. There are additional similar distributions of setups around 2 and 4 GHz which appear to indicate either 1 or 2 TDM windows with a few high resolution FDM windows. We also see that, in agreement with Table 2, there are no projects with an aggregate bandwidth of <1 GHz. Finally, for Bands 9 and 10 observations it is possible to take advantage of the double sideband spectral coverage to double the total bandwidth of the observation. We see that 20 (out of 32) MOUSs take advantage of this and obtain a total of 15+ GHz of spectral coverage.

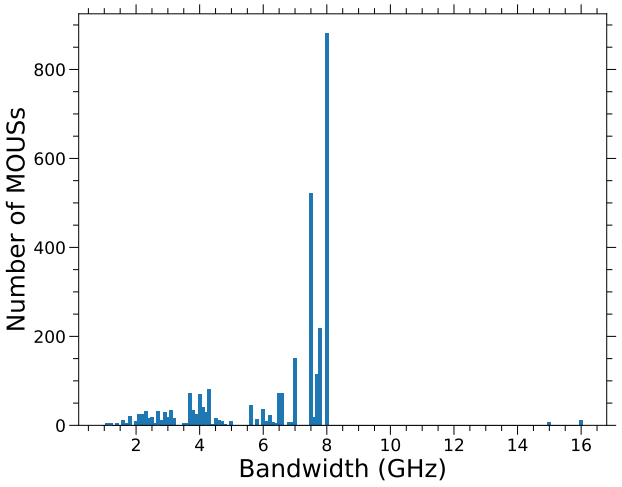

Figure 6: A histogram of the total bandwidth in a MOUS in bins of 100 MHz.

4.3 Number of Sources

Figure 7 plots the number of sources per MOUS (a) and ASDM (b). A significant majority of projects have a single source (82% of all MOUSs and ASDMs); however, the number of sources observed extends out to 129. In an ASDM sense (Figure 7b), this impacts scheduling constraints since projects with more than 1 source must be scheduled in a way that allows all of the sources to be observed while using the same calibrators for pipeline calibration purposes.

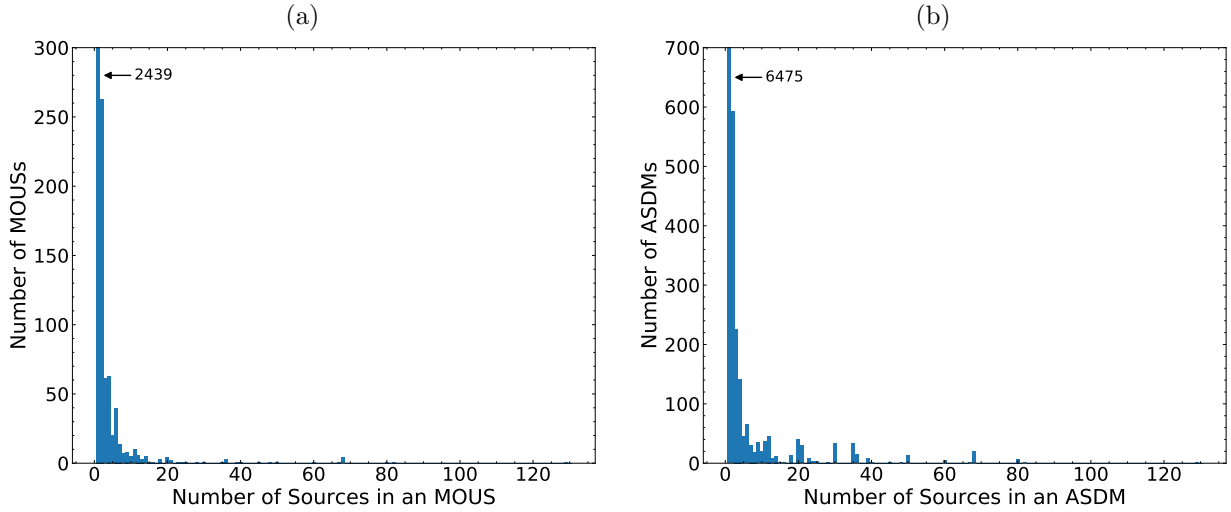

Figure 7: Histograms of the number of (a) MOUSs and (b) ASDMs that have a certain number of sources. Since the first bar of the graph extends far above the y-axis maximum, its value has been printed explicitly.

Figure 7a hints at another issue, which is the fact that the pipeline creates many images for each science target source in a MOUS and for MOUSs with more than one source, this number can quickly grow. Furthermore, if the MOUS has many spectral windows, the number of pipeline images can grow dramatically. The number of images the pipeline creates for every MOUS (ignoring any data mitigation) is shown in equation 1.

$$N_{images} = (1 + 2 * N_{spws}) * N_{sources} \quad (1)$$

Where N_{spws} is the number of spectral windows and $N_{sources}$ is the number of science target sources. This is because the pipeline creates a continuum image and a line cube for every spectral window of every source plus an aggregate continuum image for every source. Note that this formula slightly underestimates the total number of images since it does not include the additional pipeline imaging of the representative spectral window if the PI has specified a bandwidth that requires averaging of more than 4 channels but is not continuum.

Figure 8 shows a 2-D histogram of the number of sources versus the number of spectral windows in a MOUS. Again, we see that most projects have a single source and 4 spectral windows; however, there are also many MOUSs that request many spectral windows and/or many sources. In all, 65% of MOUSs request 4 spectral windows of a single source and 31% of MOUSs request more than 4 spectral windows or more than one source. Given the multiplicative nature of Equation 1, the total number of pipeline generated images quickly adds up. Fortunately, it appears that large numbers of sources with large numbers of spectral windows (i.e. the top right quadrant of Figure 8) is not a preferred mode of observation for users. As it is, using the values shown in Figure 8, at least 56,743 images were requested from the pipeline from the data obtained in Cycle 5.

While 65% of MOUSs request 4 spectral windows of a single source, this only accounts for 31% of the total pipeline imaging load. To visually demonstrate the number of images requested of the pipeline (i.e. ignoring any data mitigation), Figure 9 shows a histogram of the number of MOUSs that require a specific

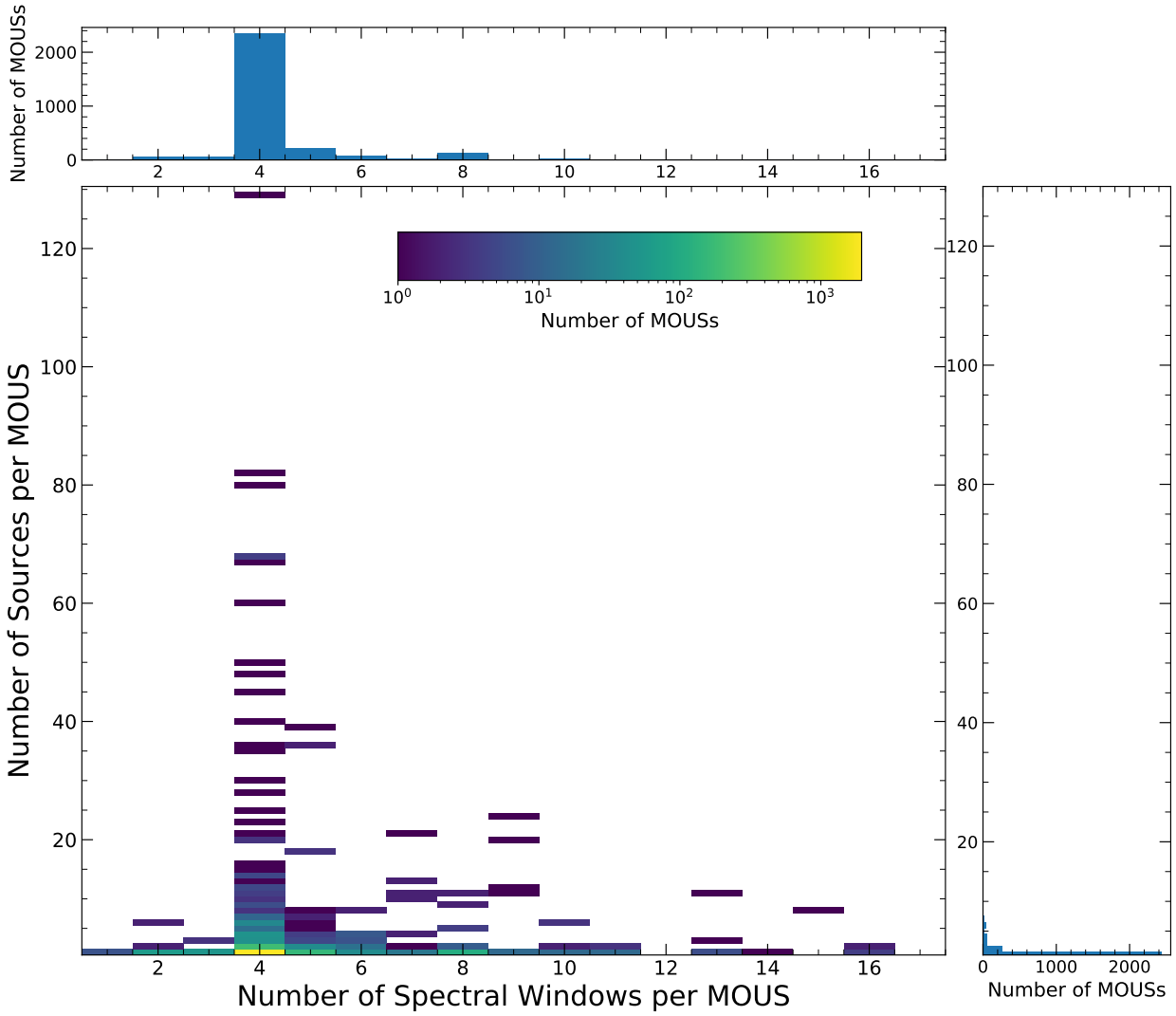

Figure 8: The number of sources versus the number of spectral windows in a MOUS. Histograms are also displayed showing the relative number of items in each axis; these are the same plots as shown in Figures 4 and 7a but are displayed here for a more direct comparison.

numbers of images. The largest number of pipeline images is 1161 images from the MOUS that observed 129 sources with 4 spectral windows.

Obviously, actually carrying out all the imaging requested would result in large data volumes and would most likely slow down the QA2 process. Therefore, the pipeline performs a data mitigation stage where the final data volume is calculated before imaging takes place and various steps are performed to reduce the final data volume to a reasonable size. This mitigation includes reducing the number of imaged sources and/or spectral windows as well as decreasing the image size and increasing the pixel size. Sometimes, especially for long baseline data, the final data volume cannot be mitigated further without significantly compromising the image products. In those cases, the data reducer manually generates the image of the representative target spectral window and source.

The data presented here still only provide a simplified picture of the requested pipeline resources when it comes to imaging. As an avenue for future research into the resources needed to fulfill PI's requested science goals, the total number of channels, selected frequency, configuration, number of executions, and pointings

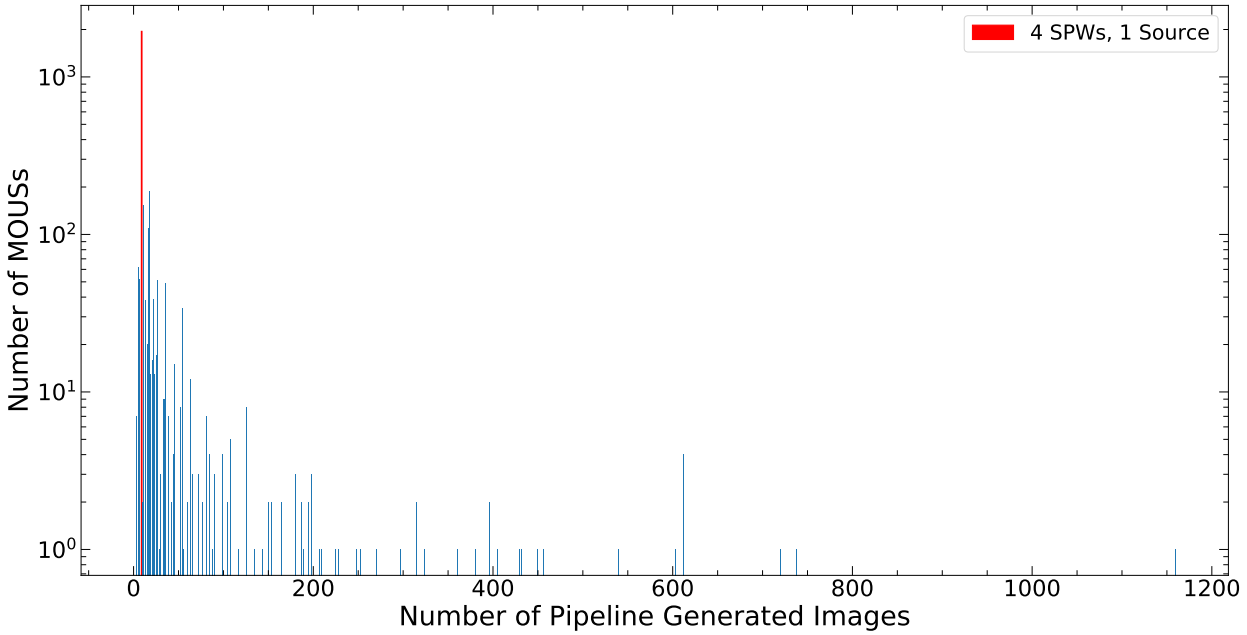

Figure 9: Histogram of the number of MOUSs that request a specific number of images. The number of images requested by the pipeline for a MOUS was calculated using Equation 1. The most common request is displayed in red and shows the number of MOUSs that request 9 images, the result of observing a single source with 4 spectral windows. This accounts for 65% of all MOUSs but only 31% of all images. Note that the y-axis is in log scale.

would need to be considered, as all these factors impact the memory requirements and final data size of the imaging products.

4.4 Pointings and Mosaic Setups

As shown in Table 2, 22.3% of all MOUSs contain a mosaic setup. Currently, the OT restricts the total number of pointings to 150 or less. In total, the ALMA array performed about 115,000 pointings carrying out Cycle 5 projects. Not all ASDMs of a MOUS may contain all pointings due to aborted/incomplete executions, and an observation can be set up to contain multiple small mosaics or combinations of mosaics and single pointings, Figure 10 shows a histogram of the number of pointings per source of an ASDM and ignoring single pointings. Because multiple ASDMs in a MOUS would repeat observations, Figure 10 is not a direct indication of the total number of sources but is an indication of which types or mosaic setups are most popular. From the obvious spike, it appears that small mosaics, peaking at 3 pointings, are the most common, and mosaics above 20 pointings are fairly rare.

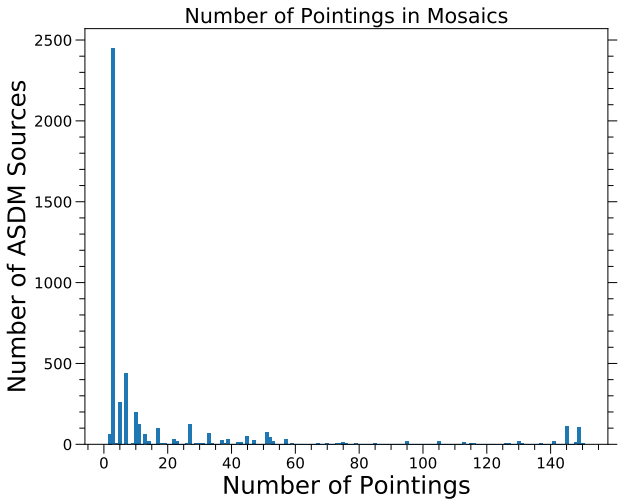

Figure 10: Histogram of the number of pointings performed for a mosaic source in an ASDM (single pointings are ignored). Since an ASDM may contain many sources, the number of pointings for each source within every ASDM was evaluated and plotted here.

5 Observations Carried Out

5.1 Number of Antennas Used

According to the Cycle 5 Proposer's Guide, ALMA guarantees a minimum of 43 antennas in the 12M array, 10 antennas in the 7M ACA, and 3 TP antennas. For QA2 purposes, this number can shrink down to 80% of this value due to flagging; however, observations should not be carried out if the minimum number of antennas is not met. Figure 11 shows a histogram of the number of antennas used in an ASDM by array type. For each array type, between 8-10% of all ASDMs are taken with less than the minimum number of antennas guaranteed; however, only QA0 pass data are shown here which indicates that some error margin appears acceptable. On the other side of the figure, there appears to be a large scatter with the total number of 12M antennas used for each observation with more than 50% of ASDMs being observed with more than 45 antennas. While this leads to excellent sensitivity measurements, in the current system, it can appear to be detrimental to achieving the requested resolution and LAS since having a larger number of antennas complicates the scheduling of projects (see Section 5.3 below).

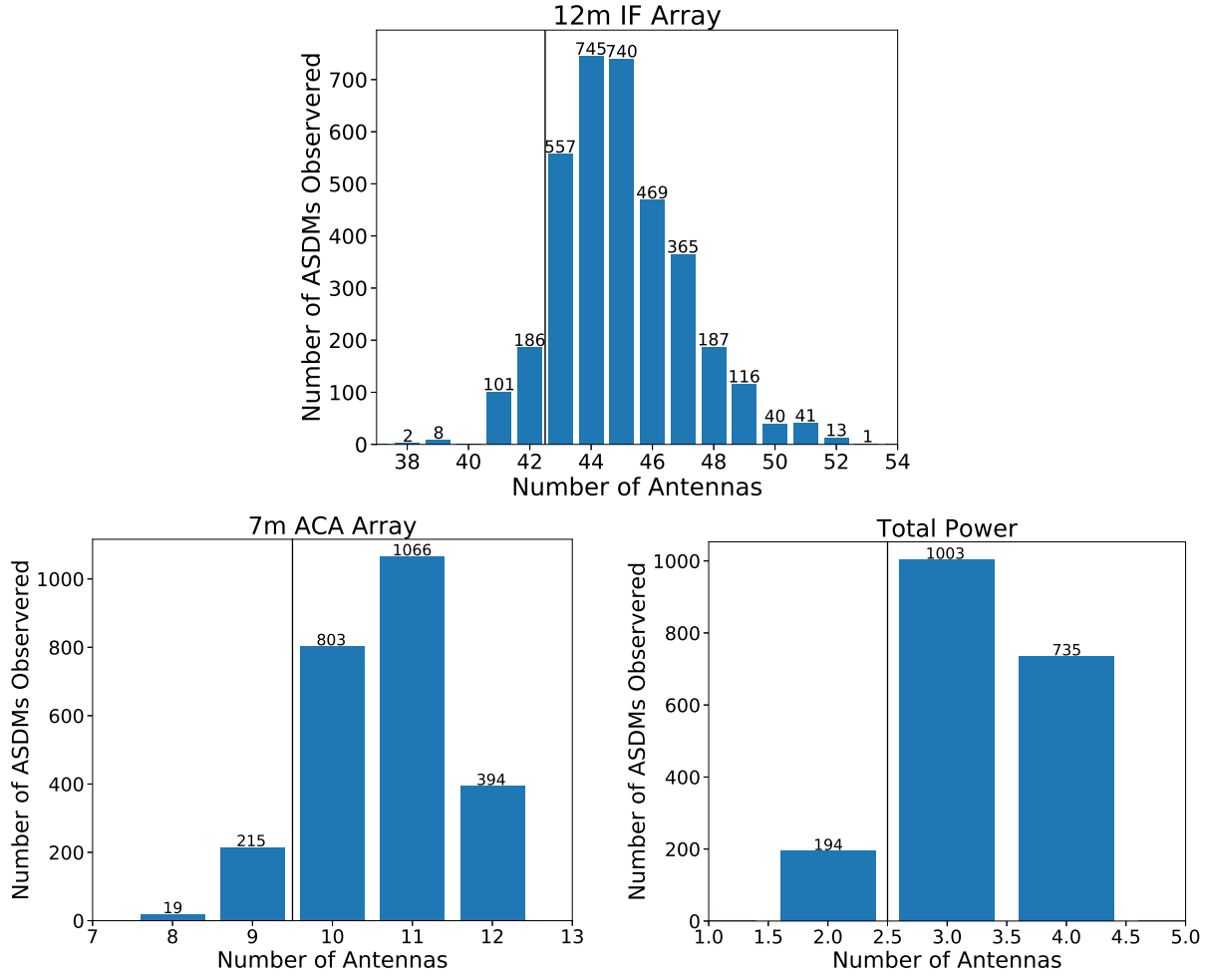

Figure 11: The number of antennas used to observe an ASDM by array type. The vertical black line in each panel represents the minimum number of antennas that we guarantee each observation will have. The total number of ASDMs in a column is displayed in text at the top of each column.

5.2 Array Configurations

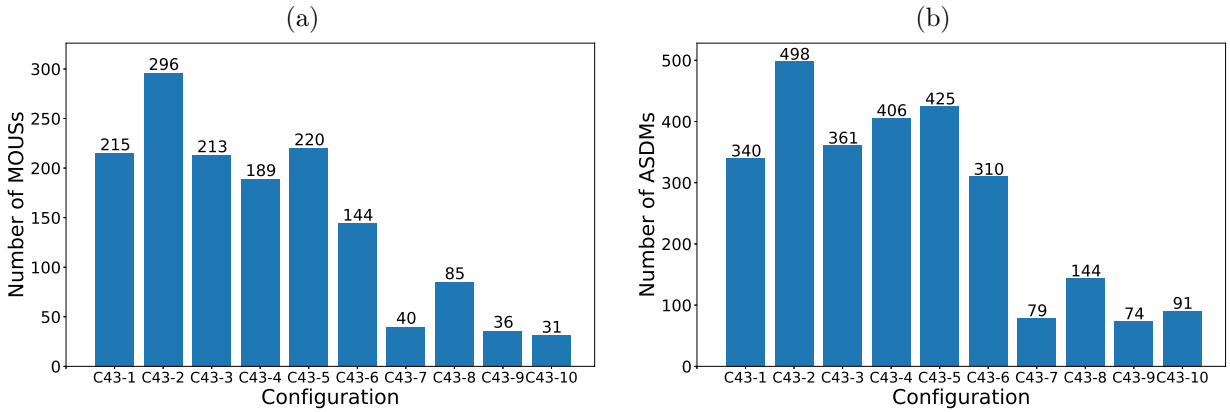

Figure 12: Number of (a) MOUSs and (b) ASDMs carried out in each PI requested ALMA configuration.

In Figure 12 we see the number of MOUS and ASDMs that were carried out for each nominal configuration. To be clear, Figure 12 shows the PI's requested resolution mapped to an ALMA configuration as determined by the OT. To make these plots (and other similar plots in this section), any SGOUS (and therefore subsequent MOUS/ASDMs) that requested a resolution range that included more than one nominal configuration (such as Target-of-Opportunity projects) were removed. Therefore, all the data plotted include only those data sets that requested a single configuration.

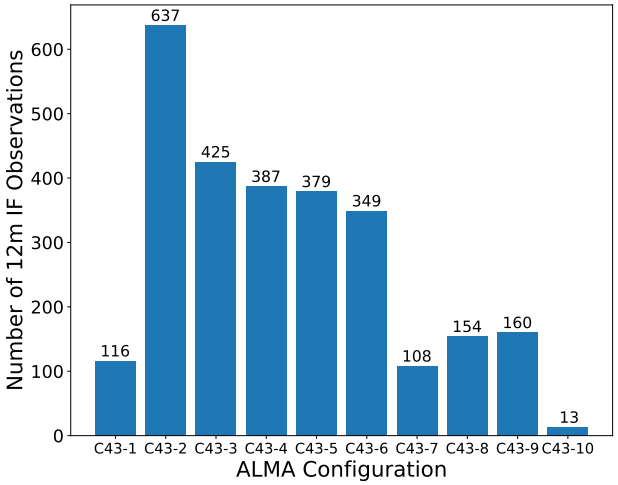

Figure 13: The number of ASDMs observed in each ALMA configuration as determined by the 80th percentile of the baselines, L_{80}, from the actual observation to assign a nominal configuration.

Since the ALMA array is highly dynamic, with antennas are being more-or-less continually relocated, and the total number of antennas included in the array changes rapidly, the configuration that is used to carry out a measurement may not always match an ALMA nominal array corresponding to what a PI has requested. Therefore, in Figure 13 I show the number of observations carried out in each configuration by using the actual antenna locations from each observation to assign a nominal ALMA configuration. To assign a configuration to each ASDM, I take the antennas that were actually used for the observation, calculate all the projected baselines, and use the 80th percentile of the baselines to assign a nominal configuration. Theoretically, resolution is determined by

$$\theta_{\rm res} = \frac{k\lambda}{\rm L_{max}} \quad [\rm radians] \tag{2}$$

where k is a uv weighting factor, λ is the central wavelength in meters, and $\rm L_{max}$ is the longest baseline. However, during the course of observing, ALMA array configurations often include outlier antennas such that the most extended antenna is not representative of the array or of the obtained resolution. Using the longest baseline therefore would skew the theoretical resolution while not contributing heavily to the final data product. Therefore, the resolution that was obtained with each observed array is better represented using

$$\theta_{\rm res} = \frac{0.574\ \lambda}{\rm L_{80}}\ \ [\rm radians] \tag{3}$$

where instead $\rm L_{80}$ is the 80th percentile of the baselines and has been found to more representative of the achieved resolution (ALMA Cycle 5 Technical Handbook, equation 7.4). Furthermore, this equation represents the resolution obtained from an ALMA array using a Briggs robust weighting value of 0.5 which is the ALMA imaging standard. Since there can be variations in the setups as antennas are relocated, I use the mean $\rm L_{80}$ baseline between configurations as a cut-off value for the assignment of configuration. For example, using the nominal configurations[5], C-1 has an $\rm L_{80,C\text{-}1} = 108$ m and C-2 has an $\rm L_{80,C\text{-}2} = 144$ m; therefore, I use the mean value of 126 m as the cut-off value to determine whether an ASDM was observed in C-1 ($\rm L_{80,obs} < 126$ m) or C-2 ($\rm L_{80,obs} > 126$ m).

It appears that only about 100 observations were carried out in configuration C-1 despite over 300 ASDMs corresponding to that configuration. A similar situation applies for the other configurations shown in Figure 12b. To help visualize the difference, I plot in Figure 14 a histogram of the difference between the observed array and the requested nominal array for each project, such that an observation carried out in C-6 that requested C-5 would result in an answer of 1, and in Figure 15 I show the observed configuration versus the requested configuration.

In Figure 15 we see that projects were mostly observed in their requested configuration (along the diagonal), and when they were not observed in the corresponding configuration, there was a preference towards more extended configurations. The most common example is C-1 requests being carried out in C-2 which we also see in Figure 13. The one exception appears to be configuration C-10 where it appears that most observations were carried out in C-9. To further quantify the extent of this effect, Figure 14 shows that ∼41% of observations were carried out in configurations that did not match the nominal configuration indicated by the PI. Positive numbers in Figure 14 indicate that the observations were carried out in a configuration that was more extended than the request which appears to be a more favored sub-optimal mode (in agreement with other internal studies that have been performed related to scheduling). There also appears to be a small number of observations that were carried out more than one configuration away from the request.

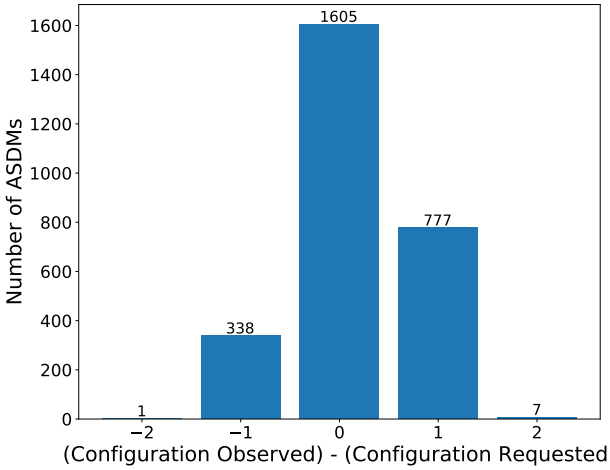

Figure 14: A histogram of the difference between the observed configuration and the requested observation as determined by using the 80th percentile of the baselines, $\rm L_{80}$, to assign an observed configuration. Positive numbers represent more extended configurations observed than requested.

Although a configuration does not match a PI request, it may not exactly translate into whether or not the observations met the resolution goals of the PI. Therefore, using equation 3, I calculate the theoretical resolution obtained by the observed array and show in Figure 16 the resolution requested and obtained for every ASDM that failed to meet the PI's request. Note that here I use the term "failure" to simply mean that the theoretical resolution fails to meet the QA2 criteria on it's own. This does **not** indicate that the data actually failed QA2 or have no value. During QA2 imaging one has the ability to use image weighting schemes (e.g. Briggs robust, UV tapering) to adjust the beam size. Whereas an internal investigation did find that 6% of datasets between October 2017 and February 2019 (i.e. Cycle 5, beginning Cycle 6) failed to meet QA2 standards based on beam size, that is not discussed further here. The total number of "failures" is 998 out of the 2728 ASDMs that meet the criteria (i.e. only a single array configuration requested), or 37%. In Figure 16 the different configurations can be clearly seen as plateaus in the data; furthermore, the requested configurations also can be clearly

[5]https://almascience.org/tools/casa-simulator

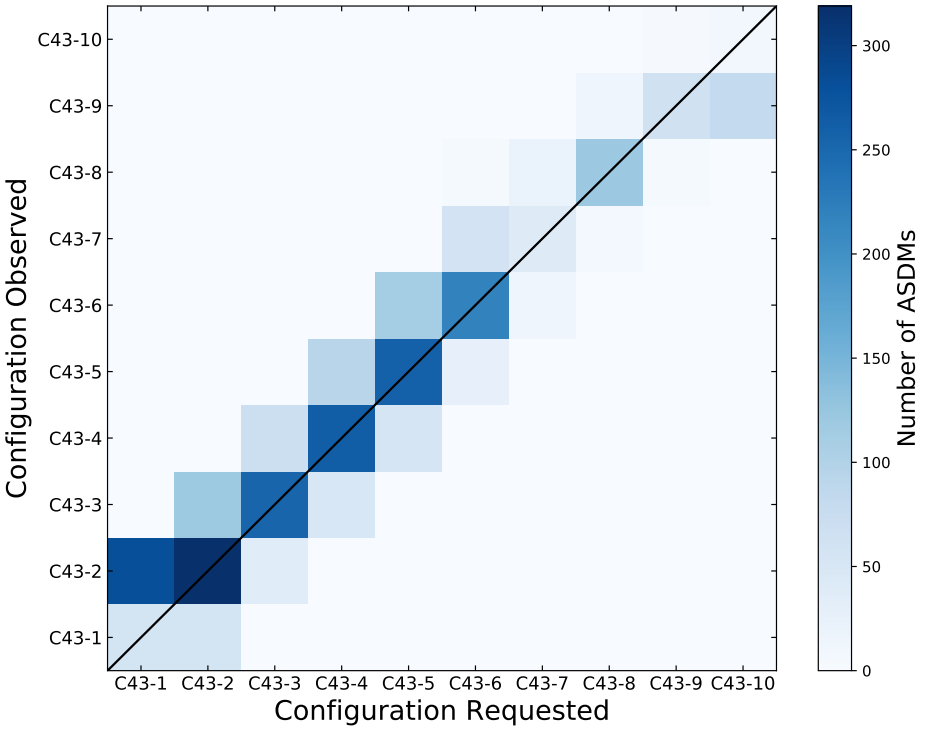

Figure 15: A 2-D histogram of the observed array configuration versus the requested array configuration. The observed array configuration was determined using the 80th percentile of the baselines, L_{80}, to assign a nominal configuration. The colorbar shown on the right represents the relative number of ASDMs in each bin.

seen as the request for similar resolutions are grouped in blocks. For example, the outlier resolution requests around ASDM index of 500 clearly show that those ASDMs requested a resolution of C-5 but were observed in C-4. Another interesting group of outliers are the blocks of requests near ASDM index 700 and 950 that requested a resolution of about 2.5 arcseconds. It appears that this resolution was only briefly available while the array transitioned between configurations C-2 and C-3 which resulted in many of those ASDMs failing to meet the beam size with the given array.

Another notable feature of Figure 16 is that the relative number of failures appears to favor grade A and B ASDMs which both have nearly four times as many failures as grade C ASDMs. This feature may indicate an issue with how projects are scheduled. For a given configuration, grade A and B projects are queued preferentially over grade C. As antennas are relocated toward a new configuration, higher priority projects are more likely to be observed than lower-ranking (but perhaps better suited for the array) observations. In other words, many grade A and B projects get scheduled during the antenna relocation process and near the beginning of an offered configuration, perhaps before the full configuration is constructed.

As another way of looking at the resolutions obtained by the various configurations throughout the course of Cycle 5, Figure 17 shows a density plot of baseline percentiles for every observed ASDM that had requested ALMA configuration C-6 compared with the nominal C-6 configuration. In agreement with Figure 15, we see that many observations were carried out around the nominal configuration C-6; however, many observations were outside of the QA2 allowed resolution, and some observations were observed in configurations extending all the way to nominal C-8. In Appendix A, I have included the same plot for every nominal configuration. All plots agree with the data shown in Figure 15, and in almost all cases there appears to be a bi-modal distribution of L80 values that lie on both sides of the nominal L80. This may be caused by observations that use more than the nominal number of antennas and include antenna positions that are very extended. As we saw in Figure 11, more than half of all 12M observations are carried out using 45 or more antennas. Adding more antennas than the nominal array creates about $43 * (N_{ants} - 43)$ additional baselines in the

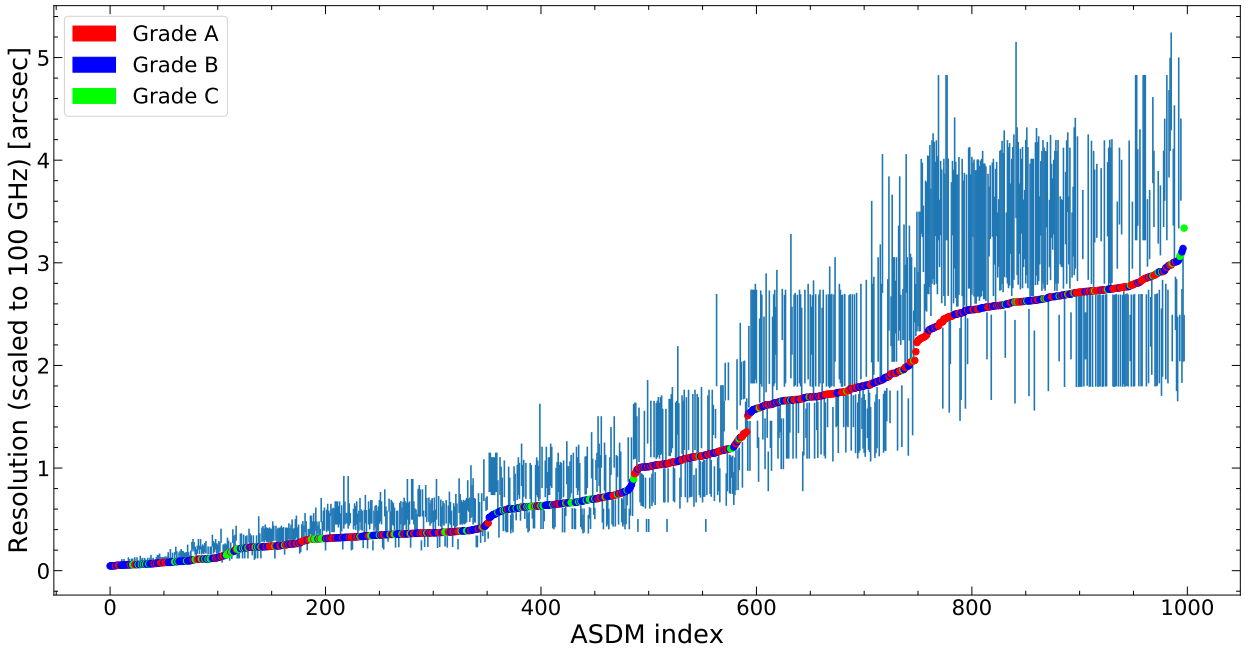

Figure 16: The resolution requested by a PI is shown as a range (blue bars) and the resolution obtained in the observation is shown as a point. The color of the obtained resolution reflects the relative project grade; there are 398 grade A, 474 grade B, and 126 grade C ASDMs shown here. The x-axis is simply an index, each ASDM that failed to meet the request is shown. Note that the data has been sorted from lowest to highest resolution and all resolutions have been scaled to 100 GHz.

array. If the additional antenna positions are extended, then this may push the value of L_{80} upward in length and therefore drop the achieved theoretical resolution. However, because the outer baselines are sparsely populated (as seen by the long tail after the 90th percentile in Figure 17), they may not in fact adversely affect the final achieved resolution. Therefore, this may be as much an indication that the value L_{80} is not a good proxy for the resolution that will be obtained as it is an indication of sub-optimal scheduling of observations. Additional work is needed to investigate which is the case.

It must be noted that the values obtained and shown in Figure 16 and 17 are the theoretical resolutions obtained from the array configuration that was used to take the data for every ASDM. As such, it does not include information from combining multiple ASDMs in a MOUS (of which 38.4% of MOUSs contain more than a single ASDM, Figure 2), any flagging that occurred during calibration, or the actual beam size that could be achieved using Briggs robust or *uv*-tapering (which is an avenue for future investigation). Therefore this is not an indication of QA2 failures but of a misalignment between scheduling and QA2 that appears to have resulted in more than a third of the ASDMs studied here being observed in a suboptimal configuration. This possibly resulted in a large number of manual imaging intervention cases during the QA2 process. We have also seen from Figure 11 that often (>50%) more than the nominal number of 43 antennas are included in the array. This means that manual imaging using increased robust values and/or *uv*-tapering are likely successful at meeting the beam size and RMS requirements set by the PI but, in fact, may be the root cause of the missed beam size in the first place. Additionally, as we will see later in Section 5.3, improperly scheduling observations and requiring manual imaging has negative consequences for obtaining the PI specified LAS. More work must be done to investigate how many of the MOUSs pointed out in the above Figures resulted in QA2 failures and/or manual imaging intervention as this would indicate the true extent of the issue.

In summary, the results presented in this section indicate that the metrics used by the OT, scheduling algorithms, and QA2 are in some cases suboptimal (like L_{80}) and/or are not aligned well with each other.

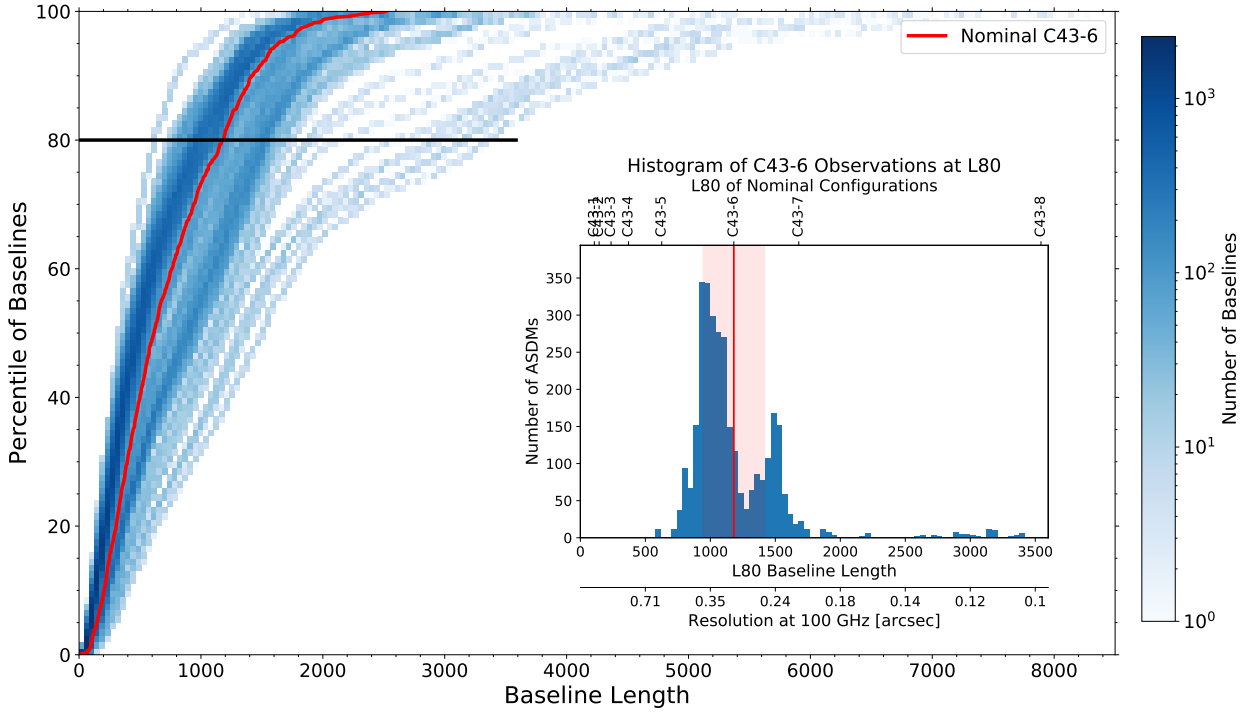

Figure 17: A 2-D histogram of baselines versus the baseline percentile (i.e. the distance within which X percent of baselines are contained). The data in blue are the binned baselines from all observations carried out for every ASDM that requested only configuration C-6 and the relative density of values is displayed in the colorbar (note log scale). The ALMA nominal C-6 configuration is shown as a red line in both the main and inset panels. In the inset panel, the red shaded region corresponds to the QA2 allowed range in resolution values. The inset panel is a histogram of values extracted from the black line in the main panel. The values in the inset panel represent the number of ASDMs that have the 80th percentile of their baselines (L_{80}) at a certain distance which directly translates to the resolution obtained using a given array (equation 3) and is shown on a secondary x-axis. The L_{80} of each nominal array within range is shown on the upper x-axis.

In particular, grade A and B projects seem to be most affected by scheduling issues which may indicate that scheduling algorithms should more heavily weight achievable resolution rather than grade.

5.3 Largest Angular Scale Analysis

As stated in Section 1.1, PIs are required to define the resolution, LAS, and sensitivity that they need in order to carry out their Science Goals. If the PI specifies a LAS that is much larger than the resolution, it is highly likely that multiple configurations are needed in order to meet the LAS criteria. Note that the LAS is a property of the science target, while the largest angular features to which an observation is sensitive is a property of an array configuration known as the Maximum Recoverable Scale (MRS). Ideally, the MRS can be found by

$$\mathrm{MRS} \approx \frac{0.6\,\lambda}{\mathrm{L_{min}}} \ \text{[radians]} \tag{4}$$

where λ is the central wavelength in meters and $\mathrm{L_{min}}$ is the shortest baseline. However, because the array configuration can sometimes have outlier antennas and/or an inconsistently spaced array, the MRS is better determined via:

$$\mathrm{MRS} \approx \frac{0.983\,\lambda}{\mathrm{L}_5} \ [\mathrm{radians}] \tag{5}$$

where L_5 is the fifth percentile of the baselines in meters and the constant 0.983 has been empirically determined from simulations (ALMA Technical Handbook, Chapter 7).

In particular, the most compact configuration for a project must ensure that the MRS is larger than or equal to the LAS of the science target. The necessity of more compact configurations and their duration are calculated using various time multipliers and image combination theory as described in Chapter 7 of the ALMA Cycle 5 Technical Handbook.

All the observations in a GOUS should meet the user specified LAS under the assumption that the data from all configurations will be combined later. However, through the QA2 process, only the resolution and sensitivity parameters of each individual MOUS are compared to the user request. The actual MRS obtained in a "Group" sense is not considered. Therefore, we are not currently aware of how well the LAS parameter is met by the observations that we deliver.

To answer the question of how well we meet the PI's specified LAS goal, we need to calculate the MRS from the aggregate sum of the observations. This is complicated by several factors:

1. The above equation 5 has only been used to describe single configurations, not combinations of arrays. Therefore, it is not currently known if the equation properly represents combined arrays. Here, I will assume that it does for this analysis but this needs to be explored further in the future.

2. The ACA and TP arrays have significantly less collecting area compared to the 12M array. This is compensated by using various time multipliers so that the smaller arrays observe the science targets for longer total integration times in order to match the sensitivity of the 12M array. However, these time multipliers are mostly empirically derived and may end up skewing the final data product. It is also worth noting that these time multipliers assume an exact ALMA configuration which is unrealistic as we saw in the previous section the array is often shifting between configurations.

3. In order to meet the sensitivity required, MOUSs may have many ASDMs each of which were observed in various weather conditions and slightly different antenna configurations as the array is continuously shifting. Therefore, some baselines may have more or less weight in the final data product. However, the median system temperature obtained in the representative spectral window is not easily obtained from the database since the representative spectral window is not in the database. It should also be noted that since the final data product is not considered here, I cannot also consider flagging that occurred during the calibration process.

4. Finally, the baselines are not kept within the database, only the antennas used and their pad locations. Likewise, while the MRS obtained per ASDM is reported in the database, this value uses the non-projected baselines which misrepresents science targets at low elevation.

I first make several cuts to the data to ensure a consistent and meaningful sample. First, I only consider a GOUS if all its dependent MOUSs have been fully observed. This ensures that the data required to meet the LAS goal actually exists. Next, I eliminate any GOUS that has a TP component. This is done because the existence of a TP SB in a GOUS means that the chosen LAS was large enough to demand it and thus the TP observations meet the goal or the LAS was larger than the TP array can provide, in which case ALMA cannot meet the goal anyway. Finally, I remove any GOUSs that have an LAS goal that is smaller than the requested resolution of the most extended array. This final cut eliminates projects which have specified that any resolution is okay (i.e. point sources or target-of-opportunity projects) and projects for which the LAS is not an important factor (such as the native array meeting the LAS goal anyway). After these cuts, there are 1347 out of 1701 grade A/B and 2272 overall GOUSs that meet the criteria.

Then, for every ASDM within a GOUS, I calculate the projected baselines using the actual array configuration used and the average target source elevation and azimuth via the CASA Analysis Utils task *getBaselineLengths*. These projected baselines are then concatenated and the 5th percentile baseline (L_5) is

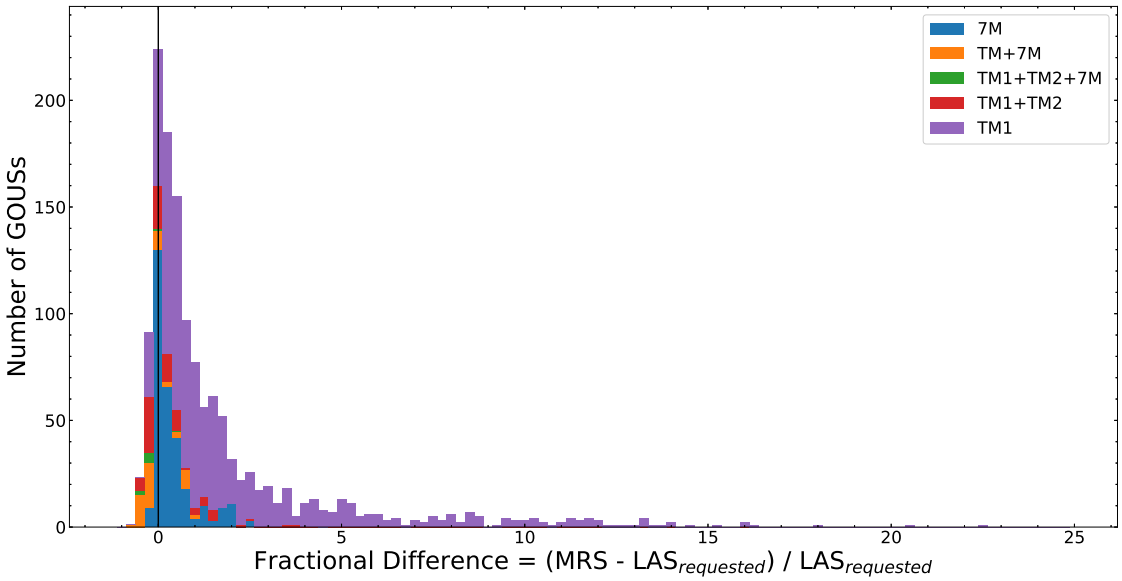

Figure 18: Histogram of the fractional difference between the MRS obtained and the LAS requested by the PI in bins of 0.25. The color coding represents the array combinations that were used to reach the LAS goal. Note that the values have **not** been scaled (i.e. they are not represented as percents).

found for this concatenated array. In essence, this approximately represents the L_5 obtained of a "snapshot" observation with the same array configurations and combinations since I don't consider the total integration time. Using that value for L_5 and the representative frequency for the GOUS, I calculate the observed MRS using equation 5. The results are shown Table 4 and in Figures 18 and 19, the latter shows the same results but zoomed in around 0.0.

In Figures 18 and 19, positive fractional differences indicate an observed MRS that is larger than the LAS request while negative values indicate an MRS that does not meet the LAS requirement. At a glance, it appears that most projects exactly meet the LAS requirement and a large majority meet the goal overall. The long tail off in the positive direction comes mainly from projects that have requested an LAS that is only slightly larger than the requested resolution. Thus, with the observed configuration the MRS obtained is often much larger than the request is these cases.

Table 4 gives the total number of failures by array combination type. We see from the table that individual arrays (TM1 and 7M) are robust against MRS failures with percentages below 4%. However, a significant percentage of GOUSs fail to meet the LAS requested when multiple configurations are needed which may be due to several factors. As noted above in point (1), equation 5 has only been used to describe single

Table 4: The number of GOUSs per array, as well as respective counts and percentages of GOUSs with MRS that fail to meet the corresponding LAS requirement.

Array Combination	Number of GOUSs	GOUS with suboptimal MRS	Percentage
TM1	880	30	3.4
TM1+TM2	92	32	34.8
TM1+TM2+7M	8	7	87.5
TM1+7M	62	37	59.7
7M	305	9	3.0
All	1347	115	8.5

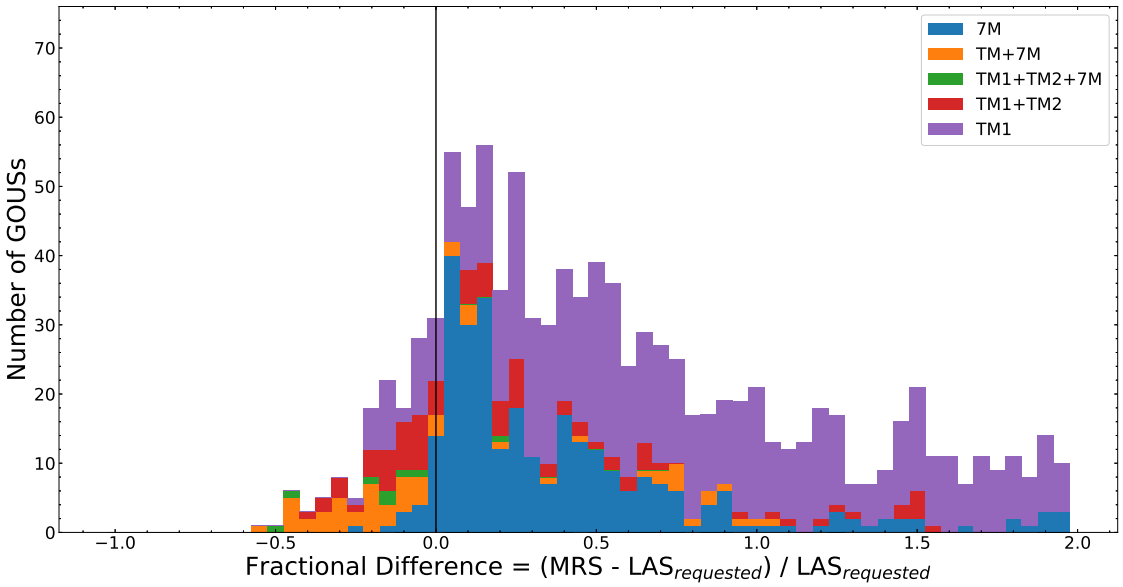

Figure 19: Same as Figure 18 except zoomed in around 0.0 and in bins of 0.05.

configurations and not array combinations. It may be the case that the MRS is being artificially lowered due to a suboptimal factor when multiple configurations are involved or that L_5 is not the best proxy for the shortest baseline when combining data because it is moved to more extended baselines when the total number of baselines is increased.

The values calculated for the MRS recovered per array that are given in the ALMA Cycle 5 Technical Handbook (and used in the OT) are calculated assuming an imaging robust value of 0.5. Currently, I consider the aggregate number of baselines used from all observations; however, the actual observed *uv* data would better represent the MRS because the relative gridding and weighting of the *uv* data can significantly affect the achieved MRS. The MRS presented here is comparable to what would be expected when using Robust = 2.0 (natural weighting) since all baselines are considered (i.e. no *uv* gridding). Alternatively, using Robust = 0.5 effectively down-weights short baselines by gridding them together, and thus decreases the MRS observed, shifting the trends in Figures 18-19 slightly in the negative direction.

An in-depth analysis of the MRS recovered by ALMA data from distinct arrays would be needed to characterize the GOUSs with suboptimal MRS indicated in Table 4. Current studies only consider single arrays (point 1 above); however, combining data from multiple arrays is common practice and embedded in the process of how projects are set up, observed, and delivered to PIs.

It currently appears that observations requesting multiple array configurations have a significant risk of not meeting the LAS requirements set by the PI. This points to the need for a thorough investigation into GOUS imaging and data combination through the use of simulations to properly determine the correct array combinations and corresponding integration times for GOUS-based imaging.

Since equation 5 should only be used to describe single configurations, we look more in depth at failures that occur during 12M only observations (purple bars in Figure 18) since these should be the most robust against failure. Note that I don't investigate the 7M array since the array is static and when used on its own either meets the PI's goal or does not, in which case TP data should have been requested. Here I investigate a few cases in detail.

For MOUS uid://A001/X1284/X151c, the user specified Band 8 measurements with a beam size in the range between 0.3 and 0.9 arcsec and an LAS of 5 arcsec. With an LAS requirement up to 4.9 arcsec, the OT returns the possible configurations of C-1 and C-2. However, with the LAS requirement of 5 arcsec, the OT drops the C-2 configuration and returns a configuration and time estimate considering only C-1.

Unfortunately, this consideration does not make it to scheduling or influence the angular resolution to enforce a C-1 only configuration. Because of this, scheduling only considered the beam size requirement and observed the MOUS in a configuration that very closely resembled the nominal C-2 configuration. Since the beam size requirement was met, the MOUS passed QA2 without any issues.

A more complicated example is MOUS uid://A001/X1284/Xbed. Here, the user specified a range of resolutions between 1.0 and 1.35 arcsec and an LAS of 10 arcsec, thus the OT returned the C-1 configuration as necessary. However, since the specified range is smaller than 20% (the QA2 beam size criteria), scheduling expanded the allowed resolution to be between 0.94 and 1.41 arcsec. This expansion of the beam size allowed for the MOUS to be picked up in the C-2 configuration where it met the beam size criteria with a Robust = 2.0 during imaging; as a result, the MRS recovered was 8.8 arcsec.

The scenarios described above do not appear to be unique, as other MRS failures also reveal that the user's choice of LAS was not fully considered during scheduling. Looking in depth at these failures from 12M-only observations, we see that there was a mismatch between the OT, QA2 criteria, and scheduling which ultimately ends in the LAS requirement being unmet.

To investigate whether or not these results are systematic of specific configurations, Figure 20a shows histograms of the fractional difference between the obtained MRS and the requested LAS, dependent on the requested array configuration. We see that observations using C-1, C-2, and C-8 were more likely to result in suboptimal MRS; however, no single array configuration appears to stand out as being significantly prone to failures in this respect. Configurations C-1/2 were already discussed in Section 5.2. For another example, MOUS uid://A001/X1296/Xe99 was observed in C-9 even though C-8 corresponded to the requested angular resolution. During the QA2 process the data reducer noted that the MRS was not met but it was passed anyway since MRS achieved is not a formal QA2 criteria.

I also plot the fractional difference as a function of the number of antennas used in Figure 20b. We see that the percent of failures more-or-less increases steadily as we include more antennas in the array. Although we begin to suffer from small numbers of datasets from which to draw conclusions based on individual array setups, the majority of the contribution to LAS failures comes from arrays that include more that 45 antennas (53% of all array setups studied here, see Figure 11 as well). This is most likely a symptom of scheduling practices when more antennas are included in the array than the nominal 43. When more than the nominal number of antennas are in the array, the scheduling algorithm modifies the array by subtracting antennas from the calculation of a theoretical beam in order to see what beam sizes the array is currently capable of obtaining. It will continue to remove antennas down to 43 until it finds projects that are currently schedulable. This is done in an attempt to balance the beam requirement and schedulability as well as acting as a rough robust weighting factor. In other words, it is assumed that performing this process of removing antennas from the array is similar to what can be obtained during imaging by either removing antennas through flagging or down-weighting them via the robust factor or uv-tapering.

While this process may increase the number of projects scheduled, it appears to adversely effect the MRS as we have calculated it here, as well as possibly the achieved beam size (Section 5.2, Figures 16 and 17). Furthermore, this practice likely hurts projects which require combinations of arrays since this may be performed on some arrays but not other arrays in a combined dataset. Thus, different robust values may be needed per dataset in order to obtain the requested beam size and MRS which is not possible to do during image combination. Based on this analysis, scheduling algorithms should incorporate all currently available antennas when determining whether beam sizes and MRS values are expected to meet the requirements set forth by the PI.

5.4 Sky Pointings

In Figure 21 we see the difference between the transit elevation and the average observed elevation (a proxy for Hour Angle) versus azimuth for all pointings performed in Cycle 5 and the relative density of pointings in each axis. We see that projects are evenly spread across the sky by science category and the vast majority of observations are carried out when the target is near transit elevation. Only a small fraction of pointings are carried out at an hour angle greater than 2 (>30°) which is as expected. This is especially true in the

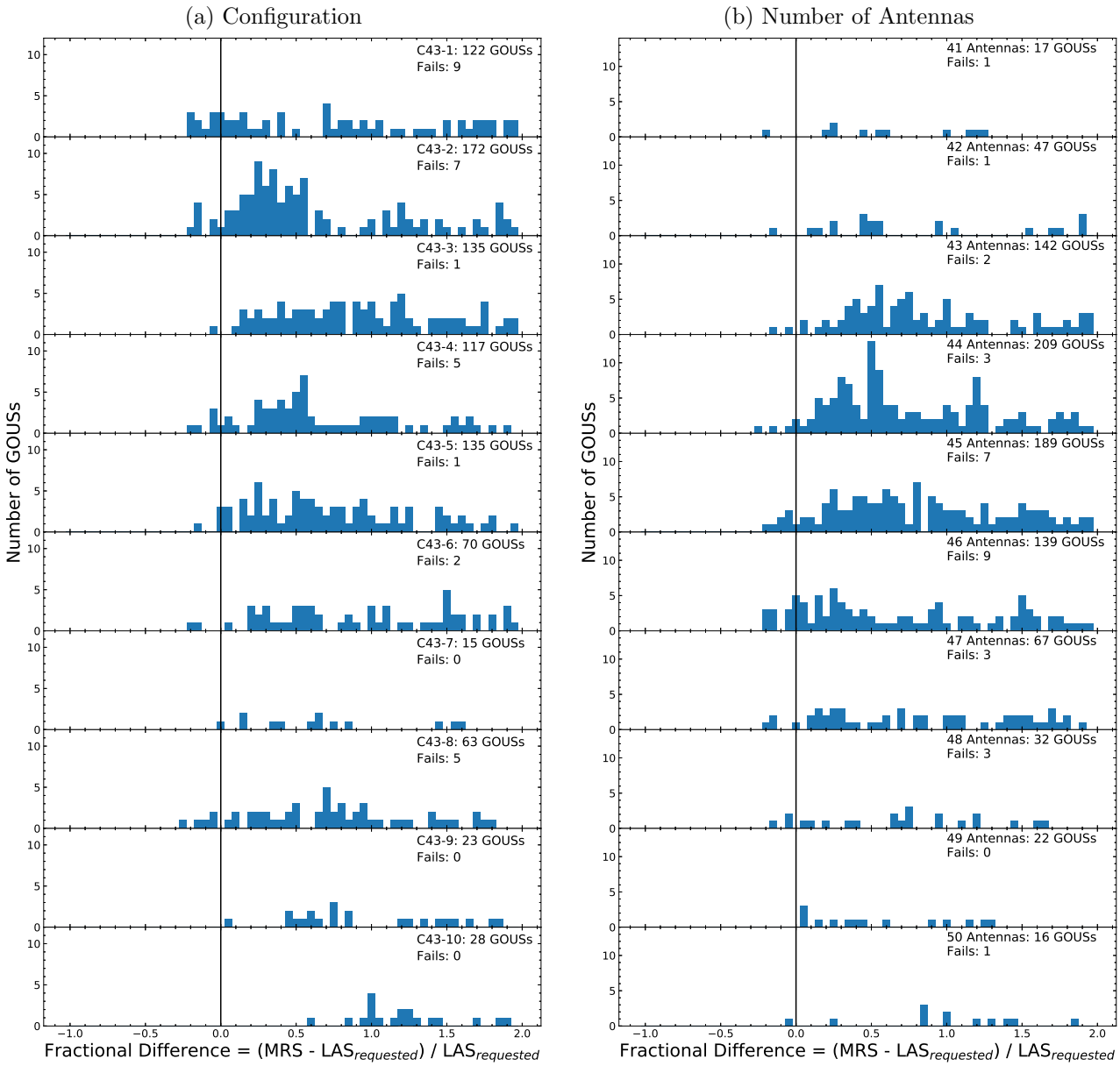

Figure 20: Histograms of the fractional difference between the MRS obtained and the LAS requested by the PI in bins of 0.05 for observations that required only TM1 observations. Note that these plots only show data in the range [-1,2] so that the long tail of data shown in Figure 18 is ignored here. In the left column (a), each panel represents one of the nominal configurations that was requested. In the right column (b), each panel represents the mean number of antennas that were used in all the observations of each GOUS. Note that the top (bottom) panel contains all the values for GOUSs with less (more) than 41 (50) antennas. The configuration or number of antennas and the total number of GOUSs in that category are displayed in the upper right corner of each panel as well as the total number of GOUSs with a fractional difference less than 0.0 (i.e. where the observations failed to meet the MRS requirement.). Note that the values have **not** been scaled (i.e. they are not represented as percentages.)

Northern and Southern directions. This indicates that scheduling is appropriately timing observations to capture targets at their zenith and have the greatest chance of producing a circular beam shape.

Mosaics clearly stand out as horizontal lines in Figure 21 since the average elevation of each ASDM is constant but the Azimuth changes throughout the observation. Another clear feature are the vertical lines around +/-100° which occur for projects which have a large number of sources but also from any sources nearly directly East or West. From the upper histogram in Figure 21, it appears that projects near +/-90° are not preferred and the most preferred direction is towards the Southern pole (+/-180°).

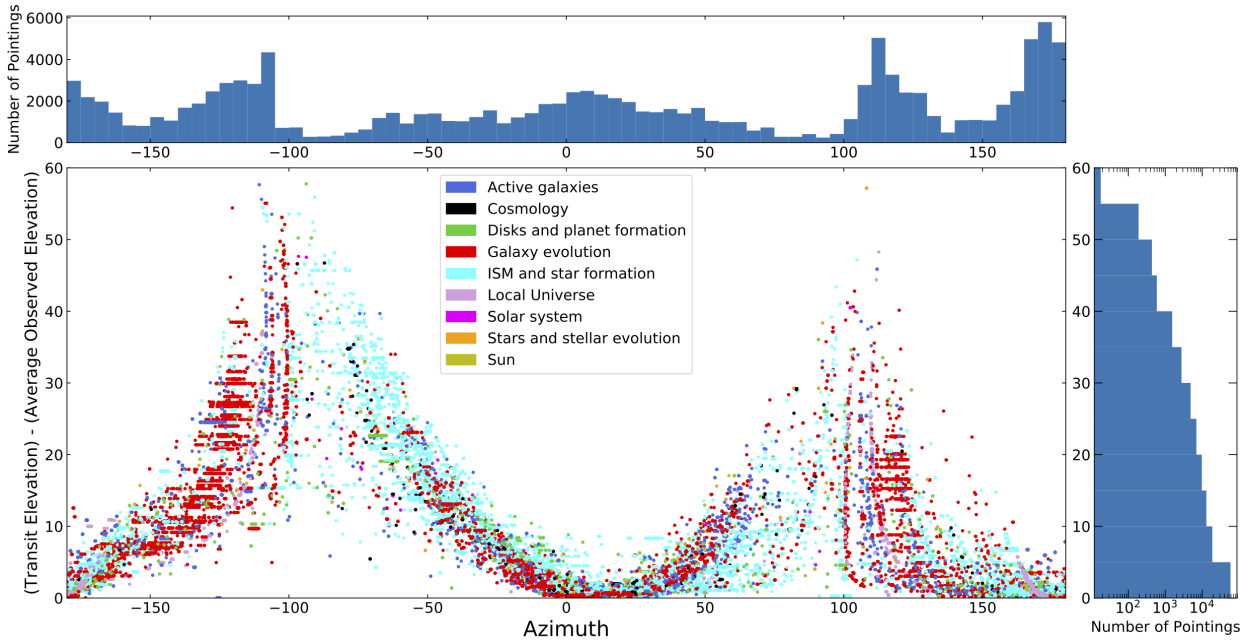

Figure 21: The difference between the transit elevation of a source and the average observed elevation (a proxy for Hour Angle) versus Azimuth for every pointing colored by scientific category of the project. Additionally there are histograms in Azimuth and "Hour Angle" in 5° bins showing the relative density of pointings shown as well. Note that the right-hand histogram is in log scale. Here, 0° indicates North.

6 Summary of Recommendations

I have investigated many different aspects of PI projects, from how they are set up to how they are carried out. Here we summarize the primary findings and recommendations:

- High frequency and polarization projects are the most highly requested non-standard observing modes (Table 2). Incorporating these projects into the pipeline will make nearly 97% of MOUSs standard-mode. This will require a combination of scheduling and pipeline changes. If scheduling included a bandpass and flux scan intent in each EB, pipeline calibration of polarization MOUSs should (in principle) become a standard mode.[6] For high frequency MOUSs, the pipeline will likely need improved heuristics to deal with weak calibrators and reduce over-flagging.

- A small number of projects utilize only between 1 and 3 spectral windows (Figure 4). Because calibration often needs to combine spectral windows to improve signal-to-noise and provide checks on flux

[6]Note that this is already a Cycle 8 goal, and reinforced by the data presented here.

measurement and increasing the number of spectral windows does not require any additional time-on-source, a minimum of 4 spectral windows should be enforced by the OT. This would have the benefit of both improving calibration and also archive products.

- Pipeline imaging can result in a large number of science images, and some requests differ substantially from the nominal case of 1 source and 4 spectral windows per MOUS (Figure 9). The total memory and hardware requirements necessary for providing this service should be more fully investigated.

- It was found that more than a third of all ASDMs appear to have been observed outside of the configuration that corresponds to their requested angular resolution (Figure 16). A more robust method of scheduling that is better aligned with QA2 requirements and utilizes the corresponding configuration would ensure a better match to PI requests.

- For GOUSs that require multiple configurations to recover flux on large scales, it was found that about half of all GOUSs fail to achieve an MRS comparable to or larger than the LAS indicated by the PI (Table 4). Further investigation into data combination and the configurations needed to properly recover flux information on many scales is recommended to optimize this aspect. A valuable first step is to create simulations of observations using nominal ALMA configurations and combinations of arrays that the OT suggests. Further, inconsistencies seem to arise from the use of more antennas than the nominal 43 (Figure 20b) when scheduling. Finally, scheduling, QA2, and the OT standards should be aligned when it comes to projects corresponding to large LAS, as was outlined in Section 5.3.

- Scheduling can utilize results from this report in order to optimize outcomes with respect to angular resolution and the LAS criteria set by the PI. Previously, scheduling considers subsets of the current array by "flagging" outer or inner antenna to achieve a certain beam size.[7] This was done in an effort to mimic robust weighting or *uv*-tapering while imaging; however, this practice may lead to the suboptimal cases that have been presented here. We provide some recommendations for improved scheduling outcomes.

- The ALMA metadata database is rich with potential to significantly improve the observations we perform and the data we deliver to PIs.

7 Conclusion/Future Work

By far, the work presented here just scratches the surface of questions that can be answered with the large amount of metadata that the ALMA database contains. This work has sought to answer some of the larger outstanding questions such as simply classifying the types of projects that are requested and how well we meet some of the goals set by the PIs. However, this raises even more questions that should be answered. With the work carried out here, the machinery and expertise has been developed to both investigate deeper issues and carry a similar analysis into future Cycles as the data become available. Already there are a few avenues which I was not able to fully investigate here but need to be investigated in order to properly direct future work.

For example, the number of spectral windows versus number of sources indicates the number of pipeline images that would be made (Figure 9); however, it does not indicate the memory resources necessary to carry out those demands. That would require compounding information related to the number of executions, channels, spectral windows, pointings, sources, band used, and configuration, all factors which would impact the number of pixels, the image size, and thus the data size of the final products. This type of analysis is needed in order to both direct pipeline development efforts but also archive storage demands and future hardware requirements for both the ARCs and our users.

The LAS study performed here has shown that, in general, we meet the LAS requirement of our PIs but only when a single configuration is requested. If more than one configuration is needed there is clearly a

[7] Note that this is no longer common practice as scheduling algorithms have improved for Cycle 7+.

need to better understand how well we meet that goal. Using the tools available to us now (i.e. Equation 5) it appears that most projects that require multiple configurations fail to meet the LAS requirement set by the PI (Table 4). This highlights the need for a more detailed investigation into LAS recovery through more detailed simulations using combinations of arrays to figure out how best to both analyze how well we have met the PI's LAS goals and perform better as an Observatory in the future. Also, using the projected baselines, the time on source, elevation, and integration times, one can re-create the obtained *uv* data and find the PSF of every GOUS. This work is currently on-going and can offer more insight into how well we actually meet LAS goals of PIs.

Furthermore, obtaining the *uv* data will also improve the investigation into the achieved beam sizes as it can then more clearly show what robust imaging or *uv*-tapering can do to help meet the PI resolution goals. It will also help show how much observing data outside of the nominal configurations hinders data combination or resolution goals (Figures 16 and 17).

Finally, I would simply like to mention that further questions, feedback, or investigations sparked by the above are welcome.

8 Acknowledgements

I would like to thank Ryan Loomis for the many in-depth discussions and ideas related to the various aspects of this study. I would also like to thank Ignacio Toledo for all the help related to gathering data and information from the ALMA database. Thanks also to Catarina Ubach, John Hibbard, and Tony Remijan for their careful reading and feedback for this work.

A Additional Plots

Here I include the same plot as shown in Figure 17 of Section 5.2 but now for each configuration. Each figure displays a 2-D histogram of baselines versus the baseline percentile (i.e. the distance within which X percent of baselines are contained). The data in blue are the binned baselines from all observations carried out for every ASDM that requested only the relevant configuration shown and the relative density of values is displayed in the colorbar (note log scale). The ALMA nominal configuration is shown as a red line in both the main and inset panels. In the inset panel, the red shaded region corresponds to the QA2 allowed range in resolution values. The inset panel is a histogram of values extracted from the black line in the main panel. The values in the inset panel represent the number of ASDMs that have the 80th percentile of their baselines (L_{80}) at a certain distance which directly translates to the resolution obtained using a given array (equation 3) and is shown on a secondary x-axis. The L_{80} of each nominal array within range is shown on the upper x-axis.

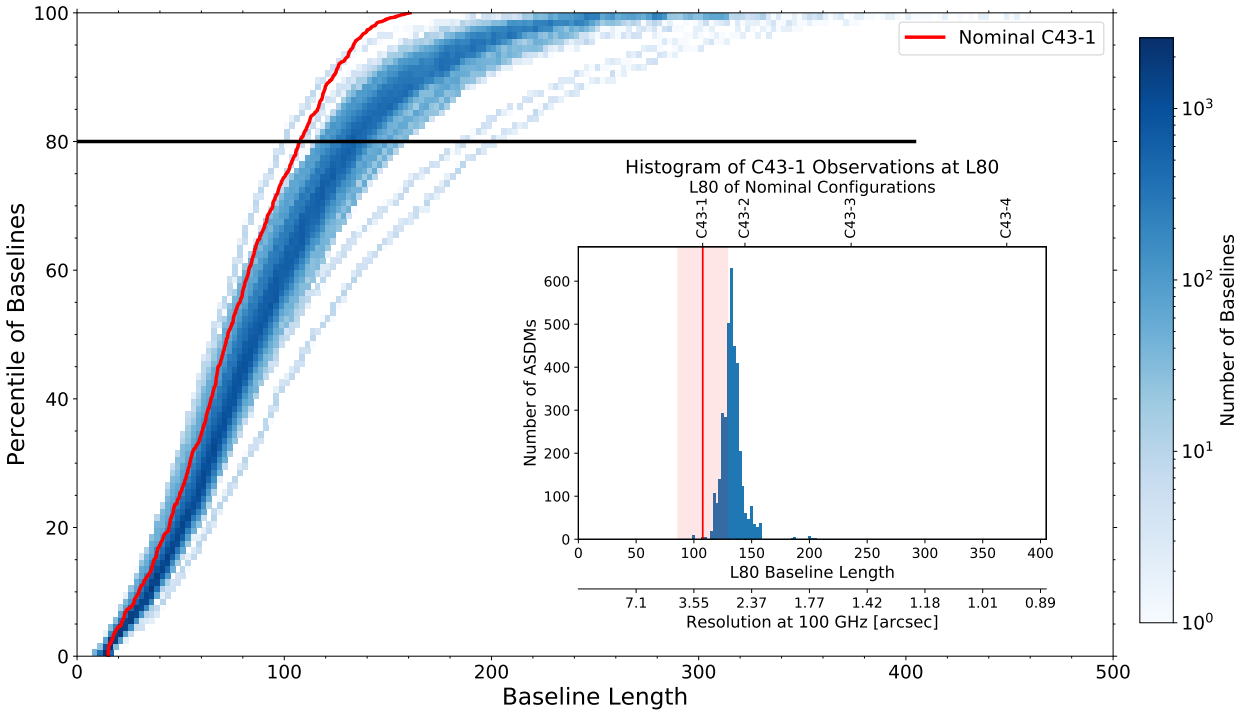

Figure 22: See Appendix A text. Same as Figure 17 but for C-1.

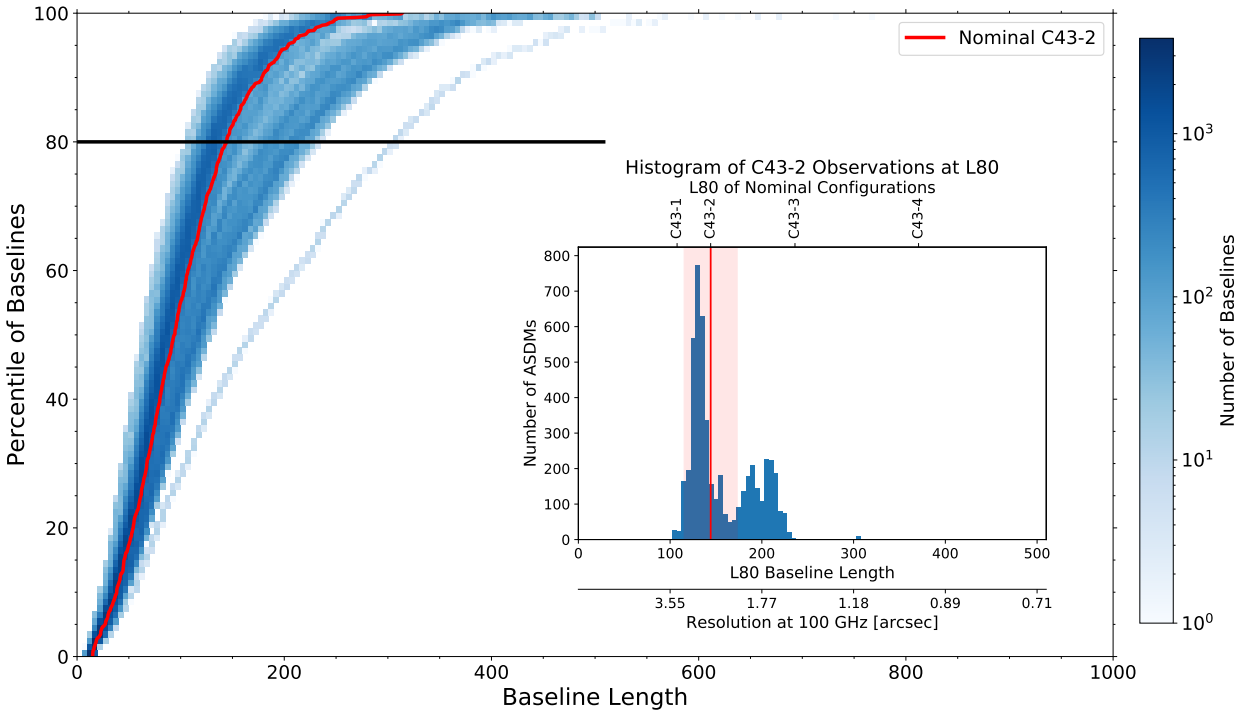

Figure 23: Same as Figure 17 but for C-2.

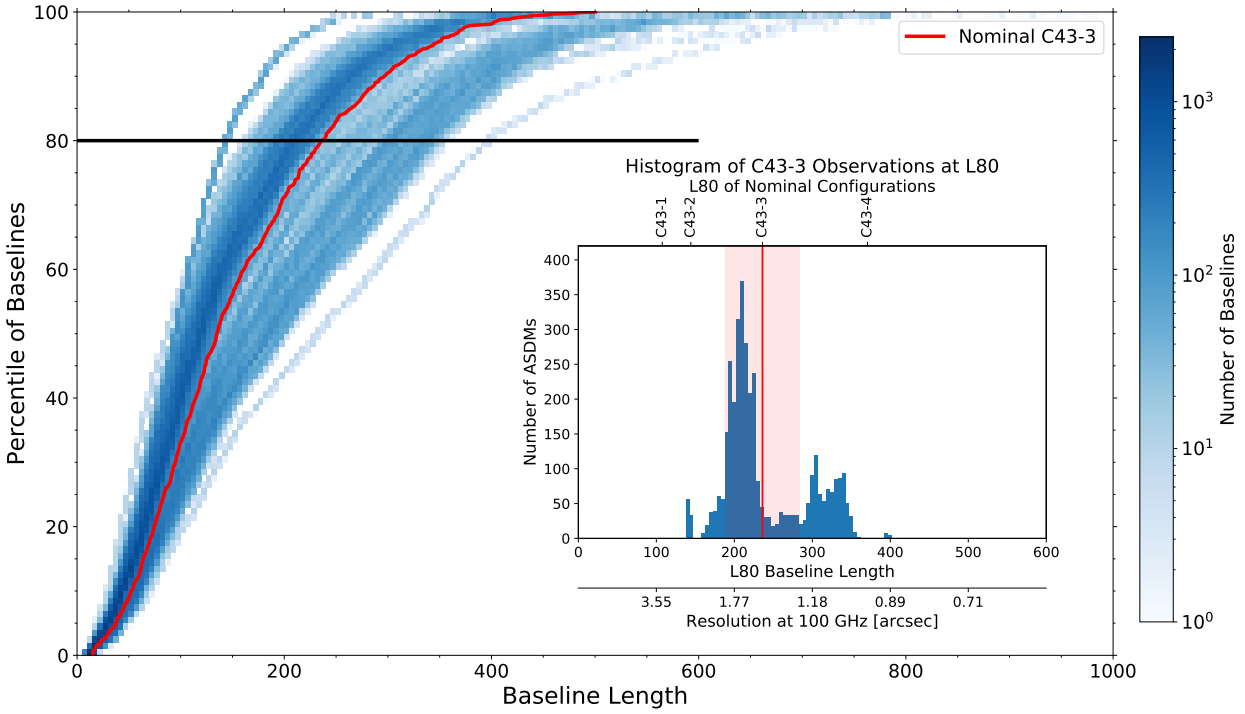

Figure 24: Same as Figure 17 but for C-3.

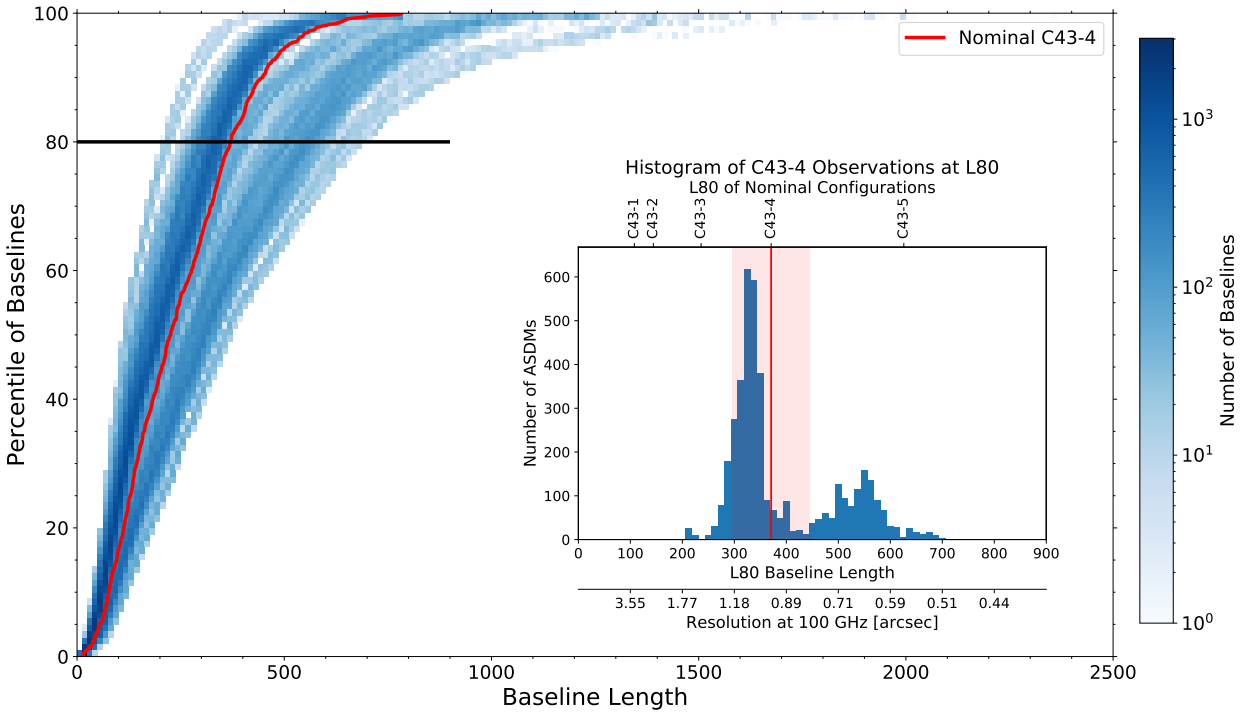

Figure 25: Same as Figure 17 but for C-4.

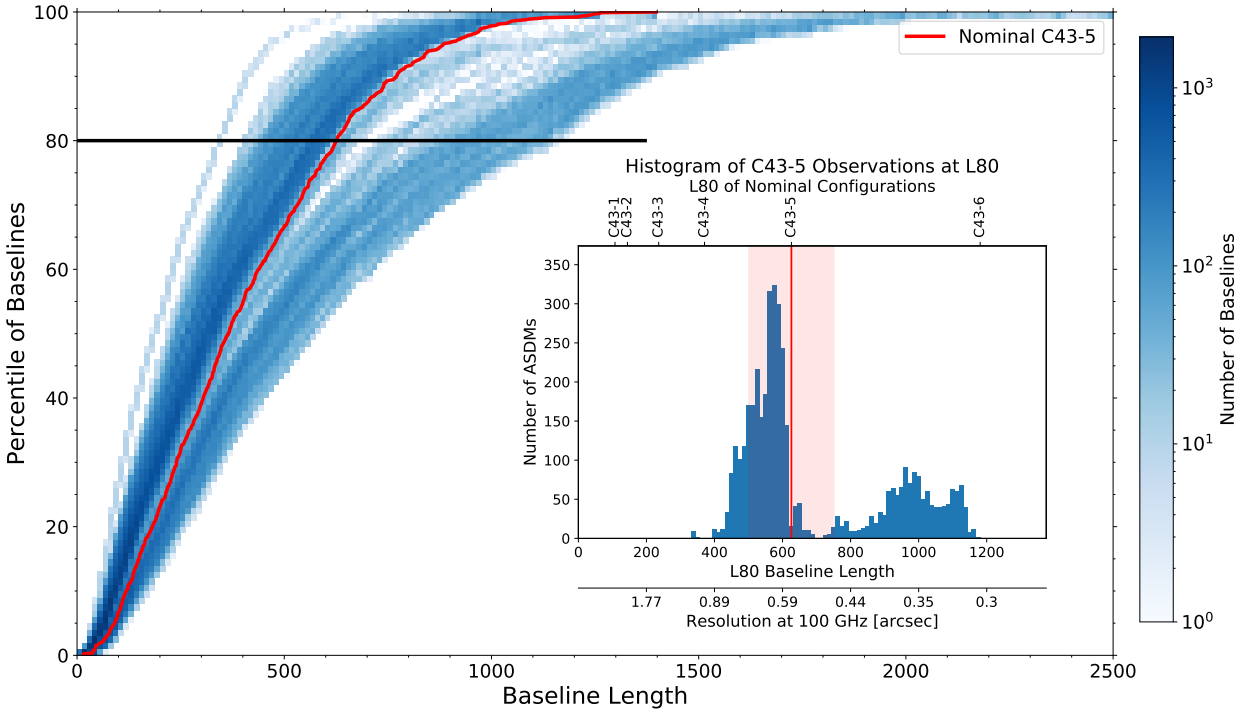

Figure 26: Same as Figure 17 but for C-5.

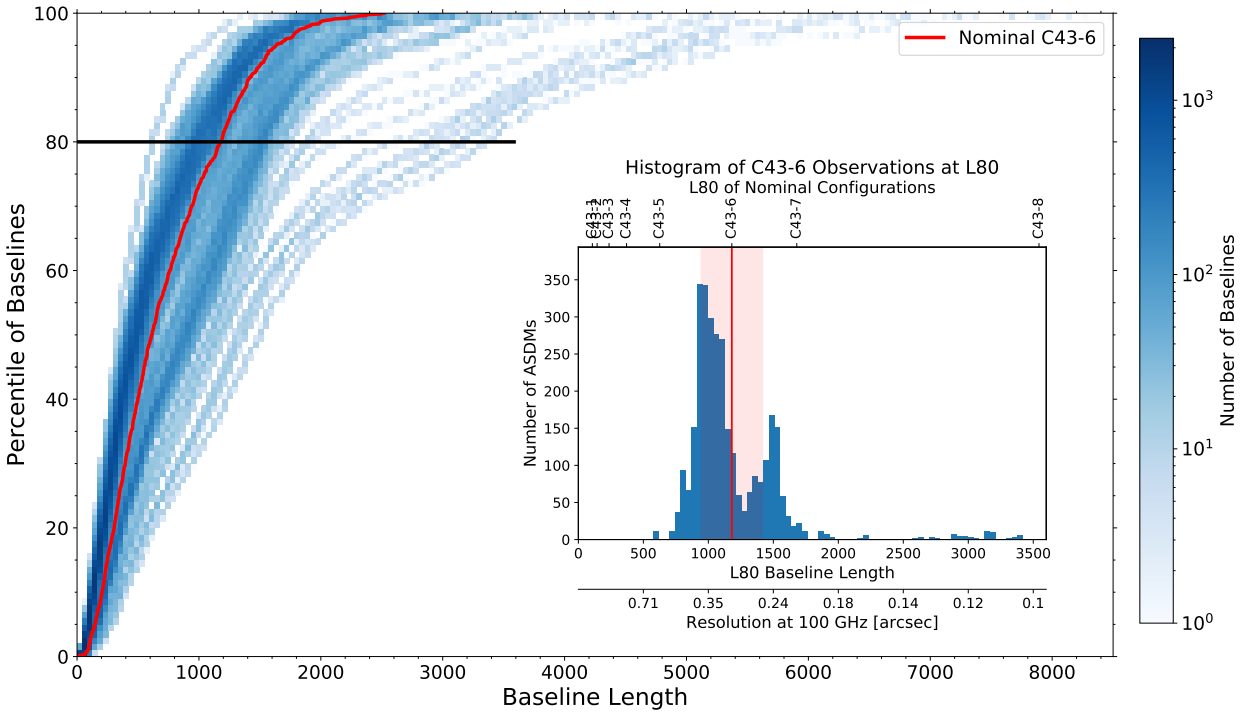

Figure 27: Same as Figure 17.

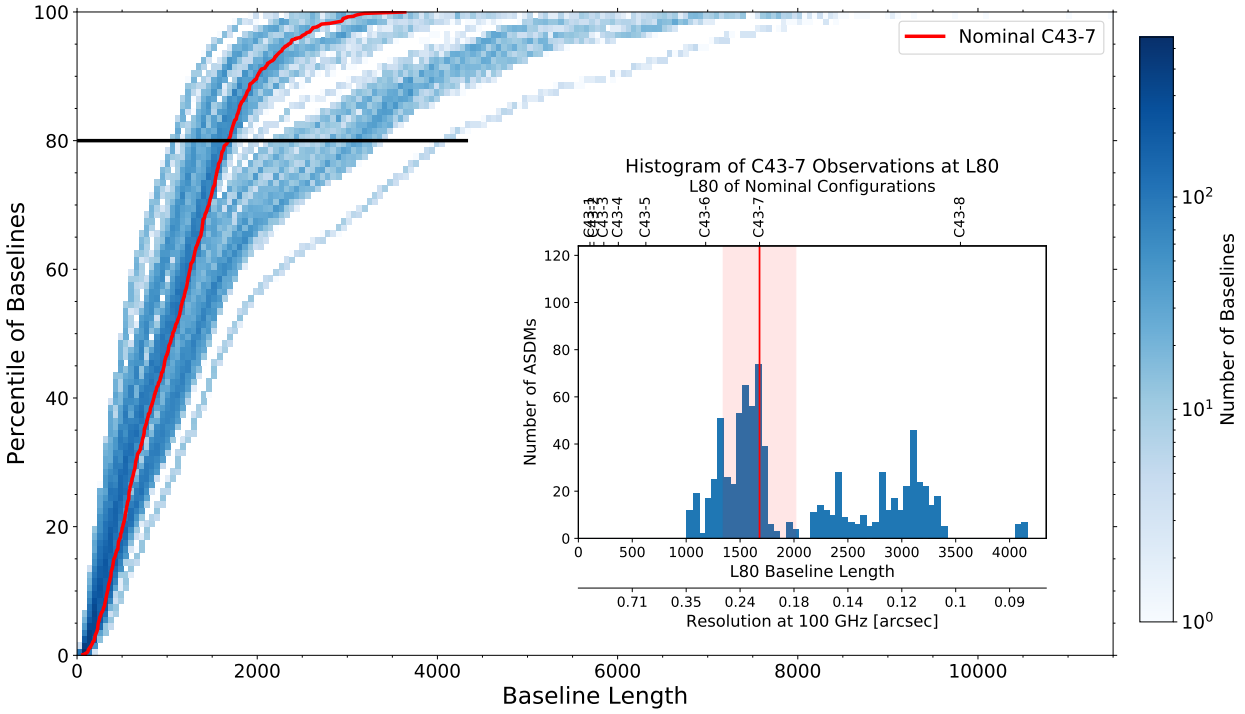

Figure 28: Same as Figure 17 but for C-7.

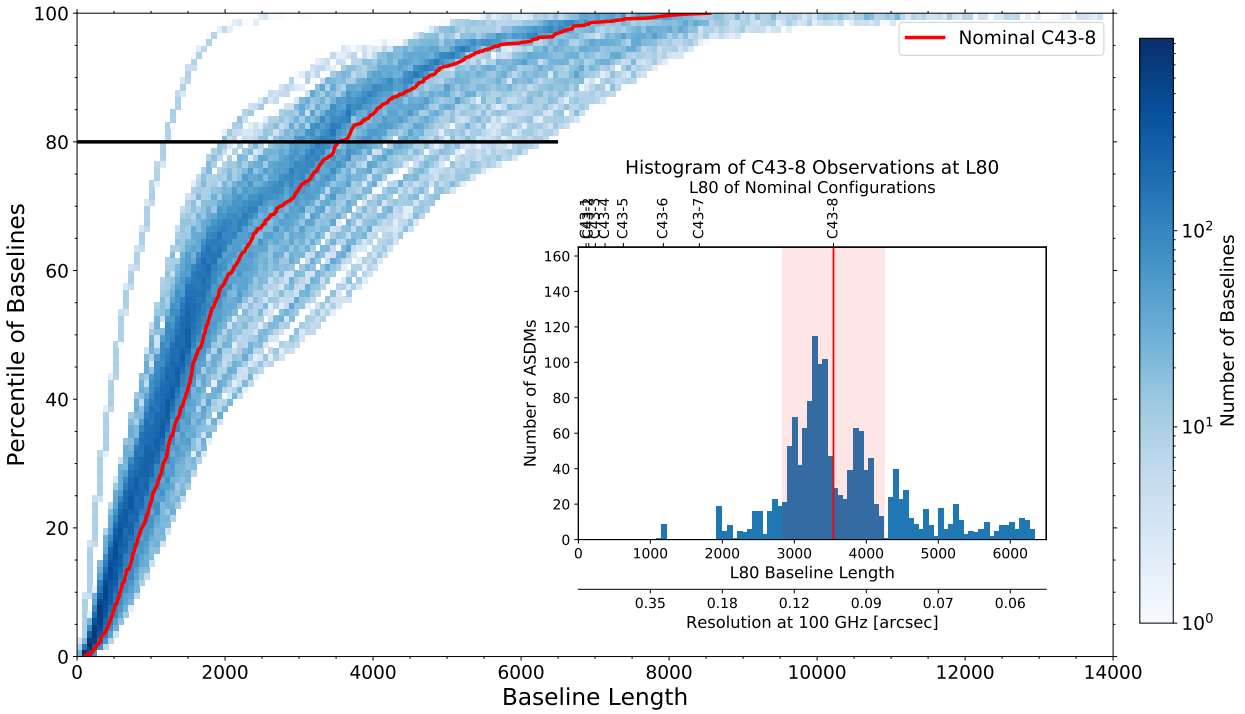

Figure 29: Same as Figure 17 but for C-8.

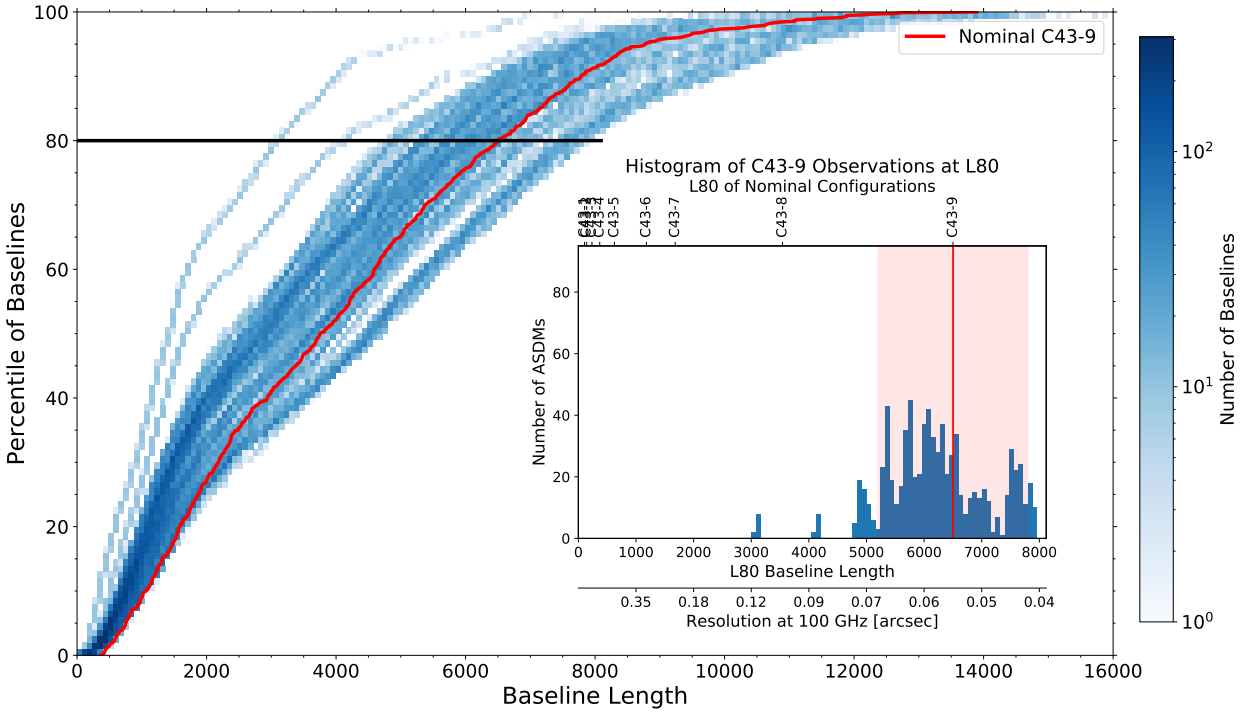

Figure 30: Same as Figure 17 but for C-9.

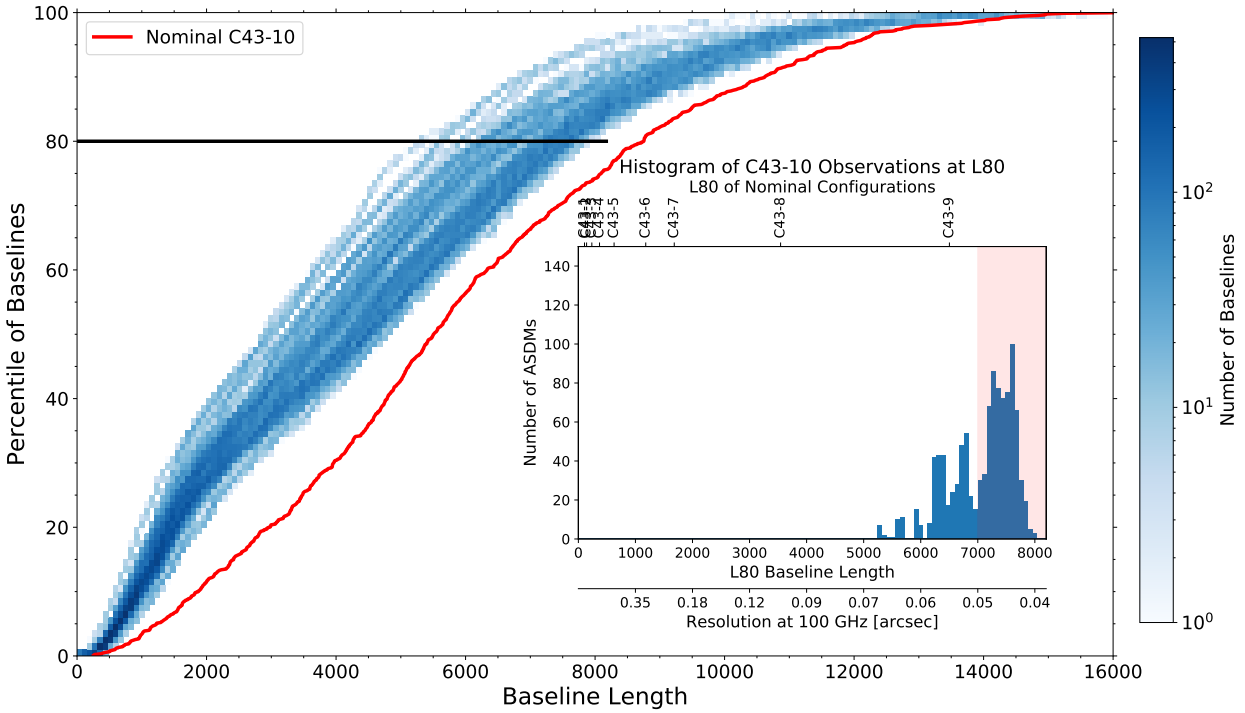

Figure 31: Same as Figure 17 but for C-10.