



# The CHILES Continuum and Polarization Survey. I. Survey Design and Noise Characterization

Nicholas M. Luber<sup>1</sup>, Min S. Yun<sup>2</sup>, Hansung B. Gim<sup>3</sup>, Daniel Krista-Kelsey<sup>2</sup>, D.J. Pisano<sup>4</sup>, Emmanuel Momjian<sup>5</sup>, and Chris Hales<sup>5</sup>

<sup>1</sup> Department of Astronomy, Columbia University, Mail Code 5247, 538 West 120th Street, New York, NY 10027, USA; [nicholas.m.luber@gmail.com](mailto:nicholas.m.luber@gmail.com)

<sup>2</sup> Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA

<sup>3</sup> Department of Physics, Montana State University, P.O. Box 173840, Bozeman, MT 59717, USA

<sup>4</sup> Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

<sup>5</sup> National Radio Astronomy Observatory, P.O. Box O, Socorro, NM 87801, USA

Received 2024 October 3; revised 2025 March 31; accepted 2025 April 16; published 2025 June 30

## Abstract

We introduce and describe the CHILES Continuum and Polarisation (CHILES Con Pol) Survey, a 1000 hr 1.4 GHz wideband full polarization radio continuum deepfield with the Very Large Array, commensurate with the CHILES H I deepfield. We describe the observational configuration, outline the calibration of the data, and discuss the effect of radio frequency interference on different observing epochs. In addition, we present a novel radio continuum imaging strategy, using well-known baseline subtraction techniques in radio spectral data, and discuss the applications to the removal of artifacts from sources far from the field center. Additionally, we discuss the nature of a low-level image-wide offset, the so-called “negative bowl,” and simulate our observations in order that we may both properly understand and correct for this artifact. Using these techniques, we present the first total intensity image of the data which achieves an rms noise of  $1.3 \mu\text{Jy beam}^{-1}$  with a synthesized beam of  $4''.5 \times 4''.0$ , the most sensitive *L*-band image ever taken at this resolution. We then place this image into the broader context of 1.4 GHz radio continuum surveys in the literature, in terms of image sensitivity and fidelity, and  $\mu\text{Jy}$  level source counts and  $P(D)$  analysis.

*Unified Astronomy Thesaurus concepts:* [Radio interferometry \(1346\)](#); [Radio continuum emission \(1340\)](#); [Extragalactic radio sources \(508\)](#)

## 1. Introduction

The observation of radio continuum emission can provide crucial measurements for understanding the evolution of galaxies across cosmic time. Radio emission can be used to estimate star formation rates (SFRs) by means of the far-infrared (FIR)-radio correlation (J. J. Condon 1992; M. S. Yun et al. 2001). The use of this relationship is motivated by the scenario in which UV radiation from young bright OB stars is absorbed by dust that produce FIR emission, while supernovae from these same stars produces nonthermal relativistic electrons moving within a galaxy’s magnetic field producing radio synchrotron emission (E. F. Bell 2003). Radio emission can also arise from active galactic nuclei (AGNs) forming radio jets and large radio lobes. These lobes can interact with the interstellar medium within the host galaxy, resulting in negative or positive feedback (G. Cresci & R. Maiolino 2018), and the medium surrounding the galaxy, providing insight into the density and magnetism of intra-galactic matter (B. L. Fanaroff & J. M. Riley 1974; M. J. Hardcastle & J. H. Croston 2020; A. Müller et al. 2021). Additionally, the measurement of polarized radio emission can be used to characterize magnetic fields, in environments spanning from the jets of AGN to cosmological scales (R. A. Laing 1980; T. Kolatt 1998).

These physical phenomenon have been studied with several different large-field 1.4 GHz surveys of varying sensitivity and

resolution, e.g., in the northern hemisphere, the National Radio Astronomical Observatory (NRAO) Very Large Array (VLA) Sky Survey (J. J. Condon 1992) which achieved sensitivity of approximately  $0.5 \text{ mJy beam}^{-1}$  with  $45''$  resolution, and the Faint Images of the Radio Sky at Twenty Centimeters Survey (R. H. Becker et al. 1995), which achieved a sensitivity of approximately  $0.15 \text{ mJy beam}^{-1}$  with a resolution of  $5''$ , and in the southern hemisphere the Sydney University Molonglo Sky Survey (SUMSS; D. C. J. Bock et al. 1999), which achieved a sensitivity of approximately  $1 \text{ mJy beam}^{-1}$  with a resolution of  $45''$ , at the lower frequency of 0.84 GHz. Currently, the VLA is undertaking the Very Large Array Sky Survey (VLASS; M. Lacy et al. 2020) which is expected to achieve  $70.4 \mu\text{Jy beam}^{-1}$  with  $2''.5$  resolution, over a bandwidth spanning 2–4 GHz. These large-field surveys have been able to detect many thousands of different continuum sources at differing levels of intensity, resolutions, and spatial extents. These large-field surveys are ideally complemented by a radio continuum survey that has both excellent resolution ( $<5''$ ) and sensitivity ( $<2 \mu\text{Jy beam}^{-1}$ ), which would be able to probe the radio continuum physics of individual sources out to higher redshifts, as well as low-level diffuse structures from low redshift sources that could be missed by less sensitive all-sky surveys.

To achieve this, such a survey requires a significantly longer integration per field, and thus would likely be limited to a single field, or a few fields. It is also critical that there exist significant multiwavelength coverage to properly characterize the spectral energy distribution (SED), and subsequently properly characterize the nature of the radio emission. With these in mind, it is a natural decision to conduct a deep radio



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

continuum survey that overlaps with the COSMOS field, a  $1.4 \times 1.4$  equatorial field that has extensive multiwavelength coverage (N. Scoville et al. 2007). The COSMOS field has observations that include low-frequency radio coverage (V. Smolčić et al. 2014; E. Schinnerer et al. 2007; V. Smolčić et al. 2017), the submillimeter and infrared (O. Le Fèvre et al. 2020; S. J. Oliver et al. 2012), the near-infrared and optical spectroscopy (P. Capak et al. 2007; L. J. M. Davies et al. 2015; O. Le Fèvre et al. 2015) and imaging (N. A. Grogin et al. 2011; A. M. Koekemoer et al. 2007, 2011; H. J. McCracken et al. 2012), and the UV and X-ray (M. A. Zamojski et al. 2007; F. Civano et al. 2016; S. Marchesi et al. 2016). Additionally, using this extensive multiwavelength coverage, the COSMOS team has catalogs with hybrid measurements, such as the color defined by the near-ultraviolet (NUV) and *R*-band magnitudes (NUV-*R*), and derived properties such as redshifts, both spectroscopic and photometric, SFRs, and stellar masses (L. J. M. Davies et al. 2015; C. Laigle et al. 2016; K. Ito et al. 2022).

The COSMOS field has been previously observed at 1.4 GHz regime by E. Schinnerer et al. (2007) with the VLA in its most extended (A) and second most compact (C) configurations. They achieved an rms noise of approximately  $10 \mu\text{Jy}$  per  $1.5$  beam. This project was a mosaic across the entire COSMOS region with a total of 240 observing hours. More recently, the COSMOS field was observed by the MeerKAT International GHz Tiered Extragalactic Exploration Survey (MIGHTEE; M. Jarvis et al. 2016) using the new South African MeerKAT telescope. The MIGHTEE survey is a multipointing observation that encompasses most of the COSMOS field, as well as mosaics over the XMM-Newton Large Scale Structure, E-CDFS, and ELAIS-S1 fields. They produce images that have a factor  $\approx 5$  worse resolution than in E. Schinnerer et al. (2007), but a factor  $\approx 5$  better sensitivity (I. Heywood et al. 2022). Additionally, the COSMOS field has been observed using very long baseline interferometry, specifically with the Very Long Baseline Array, which achieves milliarcsecond resolution and allows for in-depth studies into AGN fraction of radio source counts (N. Herrera Ruiz et al. 2017, 2018).

In this work, we introduce the CHILES Continuum and Polarization (CHILES Con Pol) Survey, a full polarization 1.4 GHz radio continuum survey taken by the VLA in its second most extended (B) configuration with 1027 hr of total observing time. The CHILES Con Pol survey is taken simultaneously with the CHILES survey and is a single pointing ( $\approx 0.25$  square degrees) coincident with a section of the COSMOS field. The observations were expected to and successfully yielded images with an rms noise of  $\approx 1\text{--}2 \mu\text{Jy}$  per beam and resolution of  $4.5''$ , making them the most sensitive observations at this resolution in *L*-band (1–2 GHz). In Section 2, we discuss the survey design and instrumental setup. In Section 3, we describe the calibration, imaging, and radio frequency interference (RFI) environment of the observations. In Section 4, we show the total intensity image of the field and discuss the applicability of the data to different scientific pursuits. In Section 5, we discuss our results in the context of other radio continuum surveys of COSMOS. In this work, we assume a standard flat  $\Lambda$ CDM cosmology with  $H_0 = 70$ ,  $\Omega_\Lambda = 0.7$  and  $\Omega_M = 0.3$ .

## 2. Survey Design

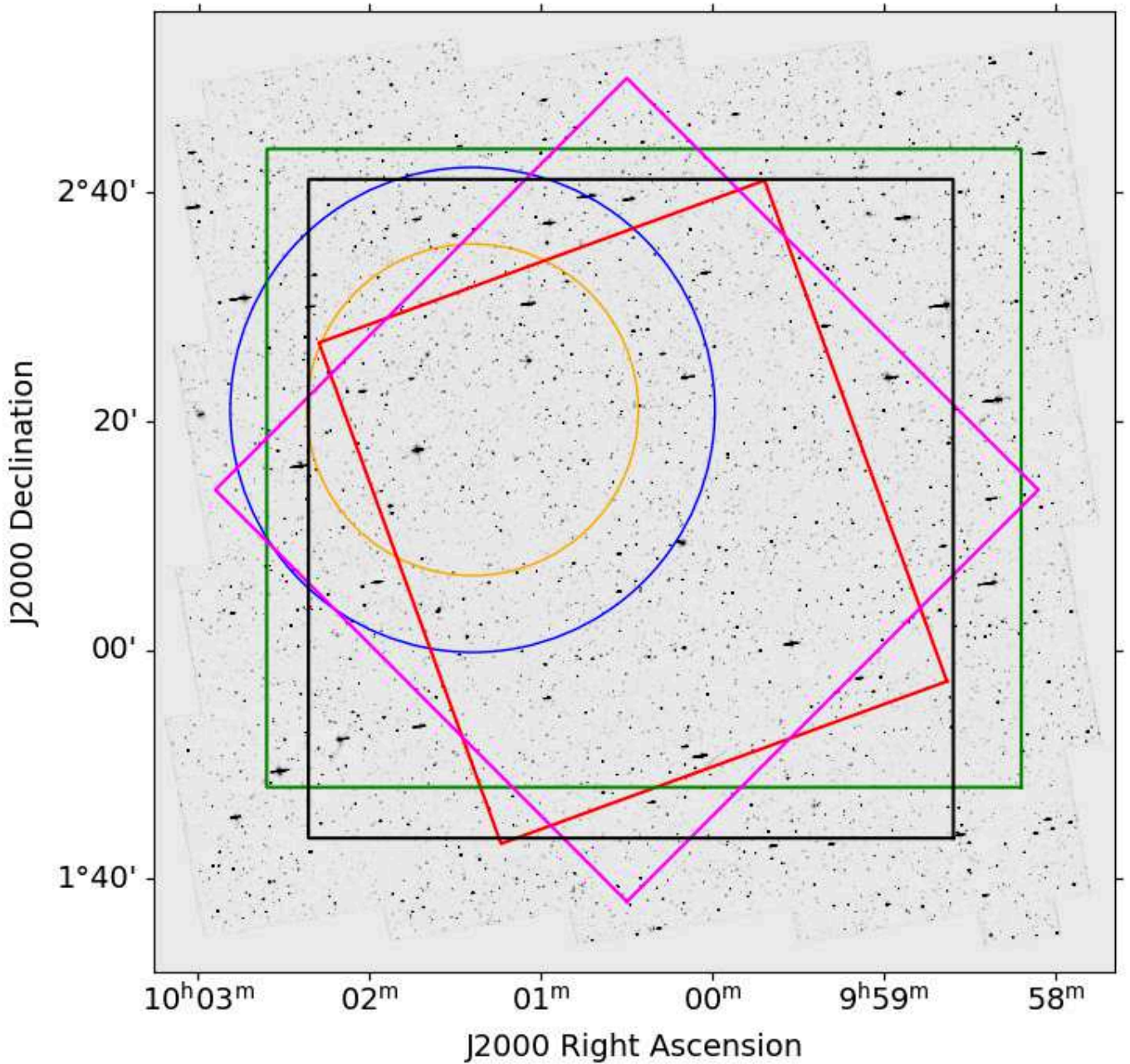
The CHILES Con Pol Survey is a commensurate survey to the COSMOS HI Large Extragalactic Survey (CHILES), which aims to study neutral hydrogen emission out to redshifts  $z \approx 0.45$  (X. Fernández et al. 2013, 2016), and the CHILES Variable and Explosive Radio Dynamic Evolution (CHILES VERDES), a time domain project making use of the CHILES Con Pol data products to explore variations on the timescales of days at an unprecedented sensitivity (S. K. Sarbadhicary et al. 2021). The CHILES pointing was chosen to be coincident with the COSMOS field, thus able to utilize the extensive multiwavelength coverage of the COSMOS survey (N. Scoville et al. 2007). The pointing is centered on J2000  $10^{\text{h}}01^{\text{m}}24^{\text{s}} + 02^{\text{d}}21^{\text{m}}00^{\text{s}}$ , and in Figure 1 we show the CHILES Con Pol field of view overlaid on a Hubble Space Telescope (HST) map along with the footprint of several of the deep multiwavelength surveys that have been undergone in the COSMOS field. The CHILES field is off center from the COSMOS field to avoid strong continuum sources, our strongest in-field continuum source has a flux density of 15 mJy, to reduce the impact of any residual calibration and imaging artifacts.

The CHILES surveys were taken by the Karl G. Jansky VLA operated by the NRAO<sup>6</sup>. In 2011, the VLA was equipped with new wide-bandwidth receivers, including a new updated *L*-band (1–2 GHz) receiver with modern electronics, as part of the Expanded VLA project (R. A. Perley et al. 2011). Additionally, a new correlator, the Wideband Interferometric Digital Architecture, was a part of the upgrade, which now allows for 64 independent sub-bands to be simultaneously observed and subsequently processed (B. R. Carlson & P. E. Dewdney 2000).

The spectral line CHILES observations require 60 of the 64 available baseline board pairs to deliver 15 32 MHz sub-bands with a total of 30,720 channels across 480 MHz of bandwidth with dual polarization (RR, LL) products. The remaining four baseline board pairs were configured to deliver four 128 MHz wide sub-bands, each with 64 channels and full polarization products (RR, LL, RL, LR). Therefore, the resulting total bandwidth of the CHILES Con Pol data is 512 MHz. These four spectral windows are spread across *L*-band with the central channels in each spectral window being 1064, 1448, 1704, and 1832 MHz, respectively, and are chosen to properly sample *L*-band frequencies while also avoiding regions that are known to be heavily impacted by RFI. The correlator integration time used for the observations is eight seconds to ensure that time smearing is minimal, while also minimizing the total data volume.

The CHILES Con Pol observations were taken with the VLA in its second most extended array form, the B-configuration. The VLA-B configuration has a maximum baseline length of  $\approx 11$  km, corresponding to an angular resolution of  $4.8''$  at 1.4 GHz. In Figure 2, we show the percentage of the total CHILES Con Pol observations as a function of baseline length (left) and the percentages with the corresponding angular sky resolution (right). These percentages were taken from the idealized case that no data were flagged due to RFI, and that all 27 antennas were functional for

<sup>6</sup> The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.



**Figure 1.** The CHILES Con Pol field, defined as the 20% (blue circle) and 50% (orange circle) of the VLA primary beam at 1.4 GHz, overlaid on the COSMOS HST coverage (A. M. Koekemoer et al. 2007). In addition, to provide context for the extensive multiwavelength coverage in the COSMOS field, we present the footprints of just a few of the many surveys the JWST COSMOS-Web field-of-view is presented in red (C. M. Casey et al. 2023), the AxTEC submillimeter survey is presented in magenta (I. Aretxaga et al. 2011), the CFHT NUV survey is presented in black (P. Capak et al. 2007), and the GAMA G10 region of reprocessed COSMOS photometry and spectroscopic is shown in green (S. K. Andrews et al. 2017).

every observation. This is an approximation because it is known that RFI will preferentially affect some baselines, and that some antennas are out of the array during different observing sessions. However, on average, Figure 2 is an accurate representation of the distribution of baselines for the full survey. The VLA B-configuration was chosen because it has excellent resolution and can resolve the HI kinematics in higher redshift sources, while also being less affected by RFI than the more compact VLA configurations. Additionally, for the CHILES Con Pol survey that will approach unprecedented sensitivity, if our resolution were any worse, the data would have been fully confusion limited. As a result, the

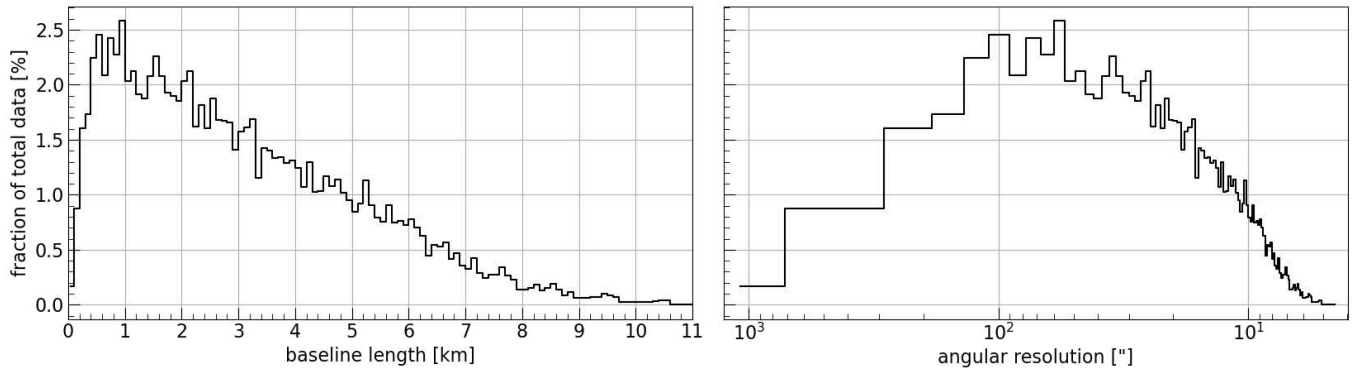
B-configuration allows for increased source detection and characterization for the CHILES Con Pol survey.

### 3. Data Processing

#### 3.1. Calibration

The calibration of the data is done entirely using the Common Astronomical Software Application (CASA; J. P. McMullin et al. 2007), version 4.7, and employs standard procedures and diagnostics. Before we begin the bandpass, complex gain, and polarization calibration, we perform a number of processing steps that provide an overview of the





**Figure 2.** Left: The percentages of total visibilities as a function of baseline length, with the data put into bins of width 100 m. Right: The percentages of total visibilities, using the same bins as the plot to the left, as a function of the angular scale to which they are sensitive to emission.

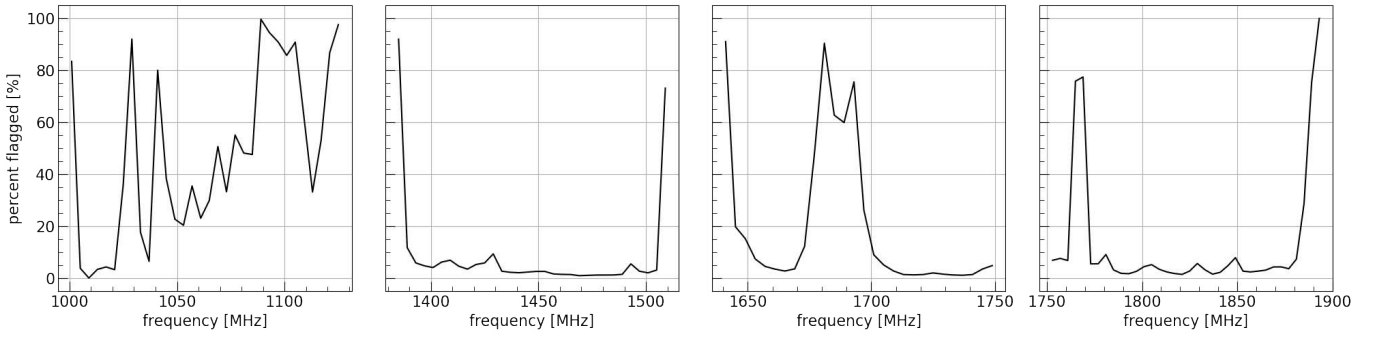
data and proper calibration formatting. These are summarized and undertaken in the following steps:

1. We create a per-session working directory, download the VLA observer's log, and the data in its SDM-BDF format. The data is then imported into the standard CASA measurement set format while applying the online flags. The online flags and weather information are then plotted, and initial flagging statistics saved. This first measurement set includes the 15 32 MHz sub-bands of the CHILES data and the four 128 MHz CHILES Con Pol data. We then split off the four CHILES Con Pol 128 MHz sub-bands into their own measurement set and delete the full measurement set, as well as the SDM-BDF directory.
2. We retrieve the specifics of the observations from the measurement set and write out the channel number-frequency relationship for each spectral window. We calculate the antenna position corrections, and plot the antenna positions and elevation versus time. We apply the pure zero, shadow, and autocorrelation flags, and Hanning smooth the data. We then flag two channels at the beginning and end of each spectral window, to mitigate the effects from the edge of the bandpass, and flag the first scan after each field change, to ensure the antennae are on source for the entirety of the scans used for calibration.
3. We create some initial plots for data inspection, specifically the amplitude versus frequency for the two calibrators and target field. We set the model for 3C286 using the *L*-band standard from R. A. Perley & B. J. Butler (2013) and the position angle of 3C286 to 33° across the entire bandwidth. We then create weblogs for the flagging, and the data overview plots made subsequent to this step. Finally, we create text files for manual flags and for the reference antenna, spectral window, and channel set-up to be used in subsequent steps.

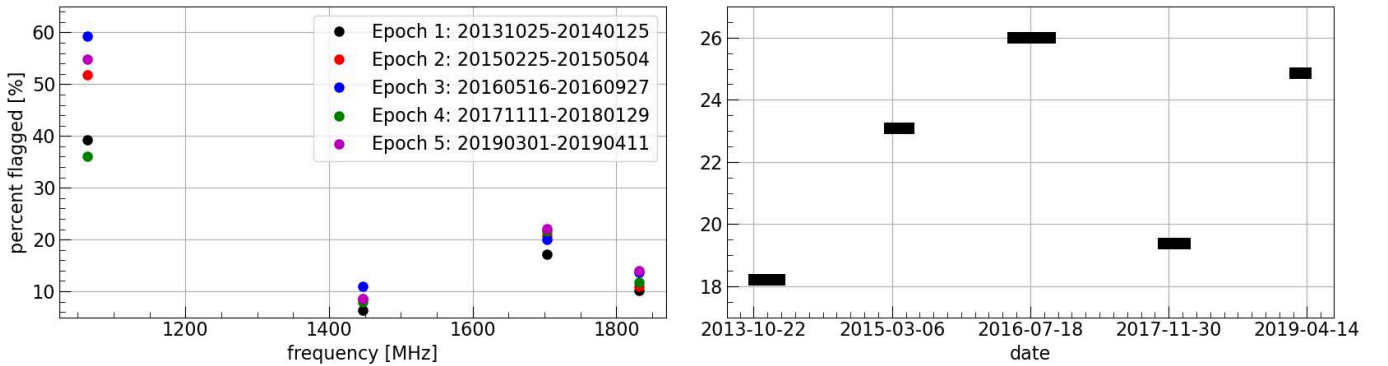
After these initial steps, we perform the gain calibrations. These calibrations make use of several tasks custom made for the reduction of this data, specifically *interpgain* (C. A. Hales 2016a) and *antintflag* (C. A. Hales 2016b). Additionally, the gain calibration routine uses the custom task *pieflag* (C. A. Hales & E. Middelberg 2014), which we will describe in more detail in Section 3.2, in addition to the final

flagging statistics for the target data. The gain calibration steps are summarized and undertaken in the following steps:

1. Import the ionospheric corrections and generate the corresponding calibration table. Create three-dimensional plots of time and frequency as the *x* and *y*-axis, and amplitude as the color scale for all targets and spectral windows for the RL correlation. Using these plots, select the reference channels, and document and apply any channel or baseline flagging for all fields, and any time flagging for 3C286 and the reference channels, and report the flagging statistics after this step.
2. Perform the initial phase calibration for 3C286, three channels wide around the reference channel, plot the primary calibrator initial phase solutions, and refine flagging, if required. Perform and plot the delay calibration, and refine flagging, if needed. Do the bandpass calibration, plot the solutions, and perform any necessary additional flagging. We then do the final temporal gain calibration for 3C286, plot the solutions, and flag as required.
3. We do an initial amplitude and phase calibration for J0943-0819, again three channels wide around the reference channel, plot the initial solutions, perform any necessary flagging, and interpolate the gain when necessary to account for any flagging. Next, we apply the antenna position corrections, bandpass, delay, and preliminary gain calibration to J0943-0819 and perform additional flagging using *pieflag*.
4. We now do the final gain calibration for J0943-0819. We then plot the final solutions and perform any additional flagging, if need be, and finally, interpolate the gain solutions on account of any of the additional flagging.
5. We perform and plot the per-spectral window cross-hand delay calibration solutions, and do any additional flagging. Next, assuming zero source polarization, we perform, and plot the solutions of, the leakage calibration, and perform any additional flagging. We then do the RL phase calibration, and subsequently plot the solutions and flag as necessary.
6. We now scale the amplitude gains for J0943-0819 and plot the rescaled amplitude data. We then apply the full calibration to the two calibrators, and plot and inspect the data. If the calibration is satisfactory, we apply the calibration to the target field, and use *pieflag* on the calibrated target data. Finally, we create plots of



**Figure 3.** The percent of the total visibilities flagged for the entire CHILES Con Pol database. The four panels of the plot represent the split for the four spectral windows used in the observations.



**Figure 4.** Left: The percentages of total visibilities flagged per each 128 MHz spectral window, and per the consecutive VLA-B configurations. Right: The percent of total visibilities flagged per the consecutive VLA-B configurations.

amplitude versus frequency and UV distance, and then report the final flagging statistics.

### 3.2. Radio Frequency Interference

The data were flagged on a per-session basis using the flagging agent, *pieflag*, which is run in the CASA environment (C. A. Hales & E. Middelberg 2014). *pieflag* runs by comparing several different statistics on visibility amplitudes for each channel, to a user-determined RFI-free channel, on a baseline-by-baseline basis. For a full description of the algorithm, see E. Middelberg (2006). This results in 21.6% of the total target data being flagged. In Figure 3, we show the percent of total visibilities flagged per channel, where each of the panels corresponds to a spectral window. The lowest frequency spectral window is clearly the most affected by RFI due to aircraft navigation radars. Similarly, the third spectral window is impacted by RFI from weather satellites.

The CHILES Con Pol observations were taken in a variety of times of the day and seasons of the year. As a result, the different observing epochs have differing exposures to RFI sources and subsequent flagging statistics. The first and fourth observing epochs were observed, generally, in the early morning near sunrise in late fall/early winter, the second and fifth observing epochs were observed in the middle of the night in late winter/early spring, and the third epoch was observed in the late afternoon/early evening in the summer time. In Figure 4, we can see that the flagging statistics for the observing epochs that were observed in similar seasons/times are remarkably similar, as to be expected. Interestingly, the best RFI conditions seem to be the early winter and early

morning observations. The fact that the third observing epoch has the most data flagged is to be expected given that it was the most susceptible to daytime RFI, and the latter half of the observing epoch has additional flags as a result of the field becoming closer to the Sun, and shorter baselines being highly susceptible to solar RFI. Finally, each spectral window varies similarly in time, indicating, that there is no preferential season/time for observations across the entire *L*-band, but rather the effects of worsening RFI are felt across all low frequencies.

### 3.3. Imaging

The production of accurate high dynamic range radio continuum total intensity images is a well-studied problem. As surveys achieve higher and higher sensitivity, the largest source of errors in the images come from direction-dependent effects caused by time, frequency, and direction-dependent effects in the telescope's primary beam pattern, as well as antenna pointing errors (I. Heywood et al. 2022). The effects of these errors appear as deconvolutional errors where a source is unable to be CLEANed properly, and the total intensity image has significant artifacts from the point source function (PSF) response to these sources. These errors become more pronounced the farther away from field center a source is, since these are the places in which the telescope's primary beam response varies the most. Additionally, the intrinsically stronger a source is, the stronger the residual sidelobes and subsequent artifacts.

Several different methods exist to try and counter this effect. One such example is that of peeling, where the traditional approach is to subtract all sources from the field, except for the

problematic source, phase rotate the data to be centered on this source, derive a direction-dependent calibration, apply the calibration and subtract the source from the data, and then return the data to its initial pointing and calibration with all other sources added back to the field (P. K. G. Williams et al. 2019). However, a technique such as this requires that sources be of a significant enough flux density to solve for the gain solutions. The CHILES Con Pol problematic sources have flux densities that do not meet this criterion for a single observing session, which is the scale on which this must be done, but are still strong enough to produce image artifacts in the final combined image.

In order for CHILES Con Pol to produce accurate total intensity image reconstructions, we pursue a novel approach for dealing with these problematic sources in radio continuum data. The approach we introduce below is done entirely with CASA version 5.6, was done using the West Virginia University High Performance Computing Cluster, Thorny Flat,<sup>7</sup> and takes approximately one week of computation using 96 GB of memory.

For each individual session, we create a square image with a side length of  $1.7^\circ$  with a Briggs robust value of  $R = +0.5$ , 1024 wplanes, four Taylor terms, and CLEAN down to the  $1\sigma$  level in the areas of known continuum sources. These sources are then subtracted from the field. We then phase shift to a source far from the field center (outside of our  $1.7^\circ$  image) and perform a  $uv$ -plane linear fit, per each spectral window, and subtraction of the problematic source. We repeat this process for two other problematic sources far from the field center. The use of a  $uv$ -plane linear fit is preferred because it is known to be capable of subtracting sources subject to imperfect bandpass calibration, such as our distant sources suffering from direction-dependent effects. As a result, subtracting them from the field produces higher quality images of the science field in a manner similar to peeling, without requiring the high signal-to-noise required by peeling to produce appropriate gain solutions. After this is done, we return the field to the proper phase center and restore the CLEANed sources.

Following the per-session problematic source subtraction, we create a square image per observing epoch (see Table 1 for the breakdown of sessions per epoch) with a side length of  $1.7^\circ$  with a Briggs robust value of  $R = -1$ , 1024 wplanes, four Taylor terms, and CLEAN down to the  $1\sigma$  level in the areas of known continuum sources. We then combine these five per-epoch images in the image plane using an inverse variance weighted mean.

#### 4. Results

In Figures 5 and 6, we present the total intensity image using the CHILES Con Pol data, made from the full 1000 hr of total observing time.<sup>8</sup> Figure 5 shows the full field of view of the observations (the 20% point of the VLA primary beam response at 1.4 GHz), and a zoom in on the inner quarter of field of view is shown in Figure 6. In this field, we detect approximately 1636 sources that will be characterized and matched with COSMOS multiwavelength data in H. B. Gim et al. (2025). A visual inspection of the total intensity map

**Table 1**  
Summary of Observations

a	b
Target Pointing	10h01m24s +02°21′00″ <sup>a</sup>
Observation Dates	2013/10/25–2019/04/11 <sup>b</sup>
Total Integration Time	1027 hr <sup>c,d</sup>
Bandpass Calibrator	3C286
Complex Gain Calibrator	J0943-0819
Dump Time	8 s
Spectral Window Setup	$4 \times 128$ MHz <sup>e</sup>
Channel Width	2 MHz

**Notes.**

<sup>a</sup> In the J2000 coordinate standard.

<sup>b</sup> The observations were taken over five different epochs, corresponding to five consecutive VLA-B array configurations. This results in the observations being taken over different parts of the year and having a mix of both day and night time observations.

<sup>c</sup> The total integration time is split among the five observing epochs with 188, 220, 190, 237, and 192 hr, respectively, in a total of 210 scheduling blocks.

<sup>d</sup> Individual sessions have total observation times ranging from 1 to 8 hr, with a mean time of 4.75 hr.

<sup>e</sup> The frequencies for the first channel in the four spectral windows are 1000, 1384, 1640, and 1768 MHz, respectively.

without applying the primary beam cutoff makes it obvious that there is significant extra radio continuum emission within the VLA primary beam area, presumably arising from the high density of  $\mu\text{Jy}$  radio sources (see Section 5.1 for further discussions), indicating that these ultra-deep radio continuum data are confusion limited.

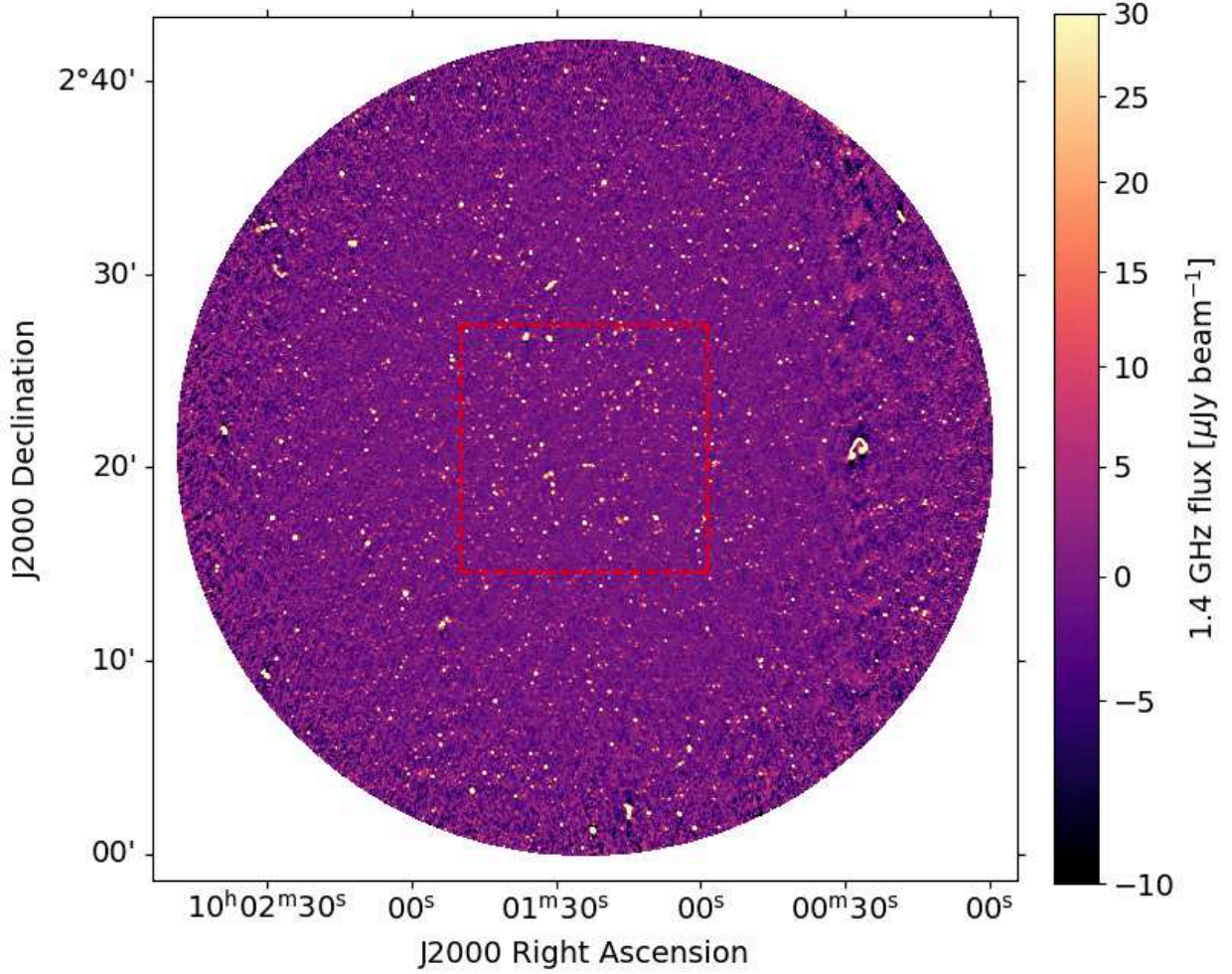
We estimate the intrinsic map noise in five featureless regions far outside the primary beam. The five featureless regions were at  $39'$ ,  $44'$ ,  $48'$ ,  $49'$ , and  $57'$  away from the phase center and have sizes of  $3' \times 3'$ . For the total intensity image, the distributions for the five regions are shown in Figure 7. We drew the histogram of flux densities after accumulating all signals in five regions to remove the stochastic effect. The bin width is robustly measured by applying the Freedman–Diaconis rule. Then, we applied the Gaussian fit to the histogram using the general-purpose optimization function *optim* in *R* to estimate the rms noise, which yielded the rms of 1.09, 0.79, 0.76, and  $0.80 \mu\text{Jy beam}^{-1}$  for *I*, *Q*, *U*, and *V*-stoke, respectively. The rms noise measured in the center of the image is significantly higher,  $2.03 \mu\text{Jy beam}^{-1}$  (see Table 2), and we attribute this to additional “noise” to the source confusion and imaging artifacts (see Section 5.3). We estimate the expected theoretical thermal noise by calculating the average system equivalent flux density (SEFD) of the observations by using the known frequency-dependent SEFD values from E. Momjian (2017), and the frequency-dependent number of visibilities as the weighting. Using this method, the theoretical thermal noise expected from the total number of visibilities is  $0.63 \mu\text{Jy beam}^{-1}$ , which is much closer to the measured image noise far outside the primary beam, especially in the *Q*, *U*, and *V*-Stokes images where both confusion and deconvolution noises are expected to be much lower.

The images presented in Figures 5 and 6 are of unprecedented sensitivity and resolution, but are affected by several sources of systematic errors. The most noticeable artifact in the image plane is a deconvolution failure around the extended source at 10h 25 m and  $+2^\circ 22'$ , colloquially referred to as the “Earmuffs” galaxy. This complex source is at

<sup>7</sup> Computational resources were provided by the WVU Research Computing Thorny Flat HPC cluster, which is funded in part by NSF OAC-1726534.

<sup>8</sup> These data are publicly available and can be found: [10.5281/zenodo.13877267](https://zenodo.org/record/13877267)





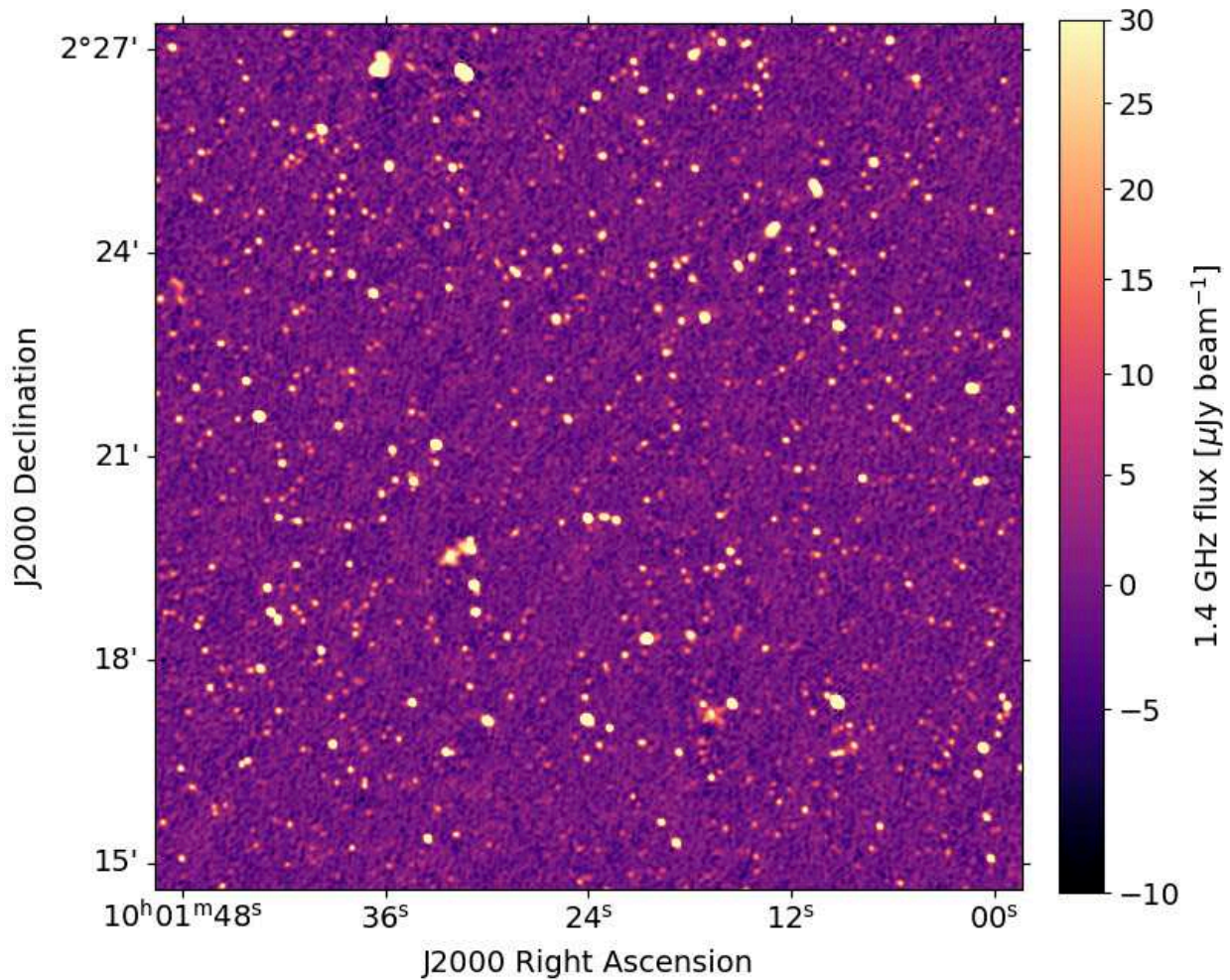
**Figure 5.** The primary beam corrected CHILES Continuum Polarization total intensity map at the resolution of  $4''.5 \times 4''.0$ , out to the 20% point of the VLA at 1.4 GHz. The central red-dashed square, of side length  $8''.6$ , corresponds to the region shown in the following Figure 6.

the approximate 50% point of the VLA primary beam, resulting in an imperfect calibration. Additionally, there are some low-level spoke-like features emanating across the image (from SE to NW) that are a result of sources beyond the first null of the primary beam. Although these image-plane errors are noticeable, we still achieve close to Gaussian noise and within a factor of 4 of the theoretical expectation. Future data releases will aim to address these issues using more advanced imaging and deconvolution techniques, while also producing high fidelity images of other three Stokes planes.

Any radio interferometric map, at a given resolution, with suitably long integration will reach a level where the predicted thermal noise becomes of similar order to the confusion noise, the noise caused by low flux-density sources that are not directly detected, but sufficiently dense as to bias flux-density measurements. Once this level is reached, further integration will not result in the usual decrease of noise by the inverse-square-root of the integration time. In addition to this confusion noise, there is a secondary confusion term which arises from residual sidelobes from sources that are unCLEANed because they lie below the CLEANing threshold, which in the case of CHILES Con Pol that is  $5 \mu\text{Jy}$ . Additionally, any source above  $5 \mu\text{Jy}$  is CLEANed down to this level, but will remain unCLEANed below this level, and thus contribute some confusing sidelobes. These sidelobes

spread across the field and introduce a noise term that further increases the measured noise from thermal noise (J. J. Condon et al. 2012).

Interestingly, these residuals are not uniform across the field of view, but rather their prevalence decreases with increasing radial distance from the phase center. This is a natural consequence of the fact that there are more unCLEANed sources below the threshold in the center of the field, thus more unCLEANed sidelobes, and an increase in this confusion style-noise. The net effect of these confusions from interferometric response is to produce a “negative bowl,” due to the strong negative features in the point-spread function of the telescope. To illustrate this, in Figure 9, we show the PSF response for each of the four sub-bands. Directly north and south of the peak response are regions of significant negative response, ranging from  $-0.09$  and  $-0.11$  for the four sub-bands, that create this negative bowl feature. This effect would vary with different values of robust weighting, with the worst case being for a naturally weighted image. Given the long integration of CHILES Con Pol and subsequent low thermal noise, these unCLEANed sources produce a non-negligible effect on the fidelity of our maps. Recently, T. Mauch et al. (2020) discuss this issue in their deep MeerKAT observations 1.28 GHz observations of the DEEP2 field. To correct for this negative bowl, they multiply their derived empirical primary



**Figure 6.** The primary beam corrected CHILES Continuum Polarization total intensity map at the resolution of  $4''.5 \times 4''.0$ , for the red-dashed square, of side length  $8''.6$ , indicated in the previous Figure 5.

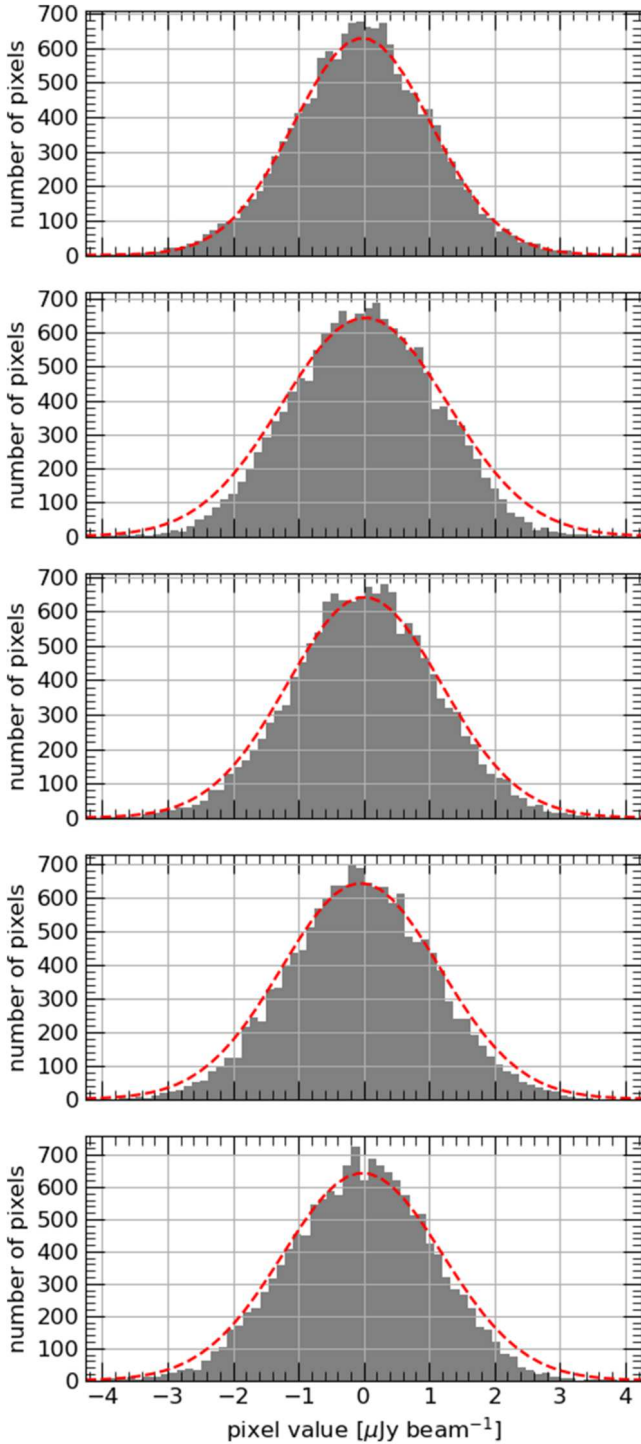
beam pattern by the central offset which effectively flattens the image, and produces accurate flux densities. The approach we take is similar, although below we describe simulated observations to probe specifically the role of faint unCLEANed sources.

Imperative to the measurement of accurate flux densities for the CHILES Con Pol sources is the characterization and correction of this “negative bowl” feature in our maps, which leads to a maximum systematic offset of approximately  $-0.4 \mu\text{Jy beam}^{-1}$  at field center. In order to understand this effect, and to study its trend as a function of radial distance from the center of the field, we simulate the low flux-density regime, flux densities less than  $5 \mu\text{Jy}$  and greater than  $0.1 \mu\text{Jy}$ , and characterize the noise caused by their unCLEANed sidelobes. To do this, we first construct an image of  $51.2 \times 51.2$  and populate it uniformly with point sources that have a distribution of flux densities given by the source count formalism in J. J. Condon et al. (2012). This image is then primary beam attenuated, to mimic the actual response of the telescope, and all flux densities above  $5 \mu\text{Jy}$ , the CLEAN threshold of CHILES Con Pol, are set to  $5 \mu\text{Jy}$ . The map from the previous step is then used as an input model that we use to simulate the visibilities CHILES Con Pol observations using the CASA task *simobserve*. We then take these visibilities and produce an image using the same parameters as with our

CHILES Con Pol map, but with no CLEANing applied. This procedure provides us with a simulated image of the contribution of the confusion due to low flux-density sources and sidelobes from unCLEANed sources.

Using this simulated image, we can now explore how the peak of the pixel-value distribution for the simulated map varies as a function of radial distance. To do this, we take annuli at increasing radii, taken on the center of the image, and fit a Gaussian to the pixel-value distribution and save the mean of this fit as the mean pixel-value offset for these annuli. For the CHILES Con Pol image, we follow the same procedure with the same annuli in order to ensure the closest comparison. This comparison is shown in Figure 10, where we show the simulated measurements (black markers) and observed measurements (red markers) as a function of radial distance. From Figure 10, we see that the overall radial trend between the two is consistent, with the effect lessening at increasing radial distance. Interestingly, although the slope is consistent between the two, the values for the observed map are systematically lower. We hypothesize that this is due to the fact that the CHILES Con Pol data suffer from imperfect calibration, and thus imperfect deconvolution. Specifically, when an image with some remaining calibration errors is being CLEANed, flux will be properly added to the CLEAN model, but the PSF structure in the image attributed to the CLEAN





**Figure 7.** The distribution of pixel values for the total intensity image over five blank sky regions 39', 44', 48', 49', and 57' away from the phase center (shown here in descending plots) with sizes of  $3' \times 3'$ . The distributions are shown for the range  $\pm 4 \mu\text{Jy beam}^{-1}$  range with the bin size dictated by the Freedman–Diaconis rule. The red-dashed lines indicate the Gaussian fits to the distributions used in estimating the rms noise.

model flux will not be perfectly subtracted from the image, resulting in low-level artifacts. As a net result, there are more residual artifacts from these sources that are not CLEANed properly producing a larger systematic offset. Additionally, by measuring the noise in our simulated data, we estimate that the residual sidelobes are contributing approximately 50% of the measured noise in the image.

**Table 2**  
Summary of Image Noise Characterization

Method	rms Noise ( $\mu\text{Jy beam}^{-1}$ )
Histogram of $R = -1$ I-image center <sup>a</sup>	2.04
Histogram of $R = +0.5$ I-image center <sup>a</sup>	2.03
Histogram of $R = +0.5$ I-image outside 40' radius	1.09
Histogram of $R = +0.5$ Q- and U-image	0.79, 0.76
Residual rms noise from the $dN/dS$ analysis	1.67

**Notes.**

<sup>a</sup> In this measurement we define image center as a square region of  $8.5''$  over the central pointing of the field, corresponding to the region shown in Figure 6. We calculate this value on the primary beam corrected image; however, the primary beam response for our observations does not fall below 85%, allowing us to make an estimate of image noise without significant primary beam effects.

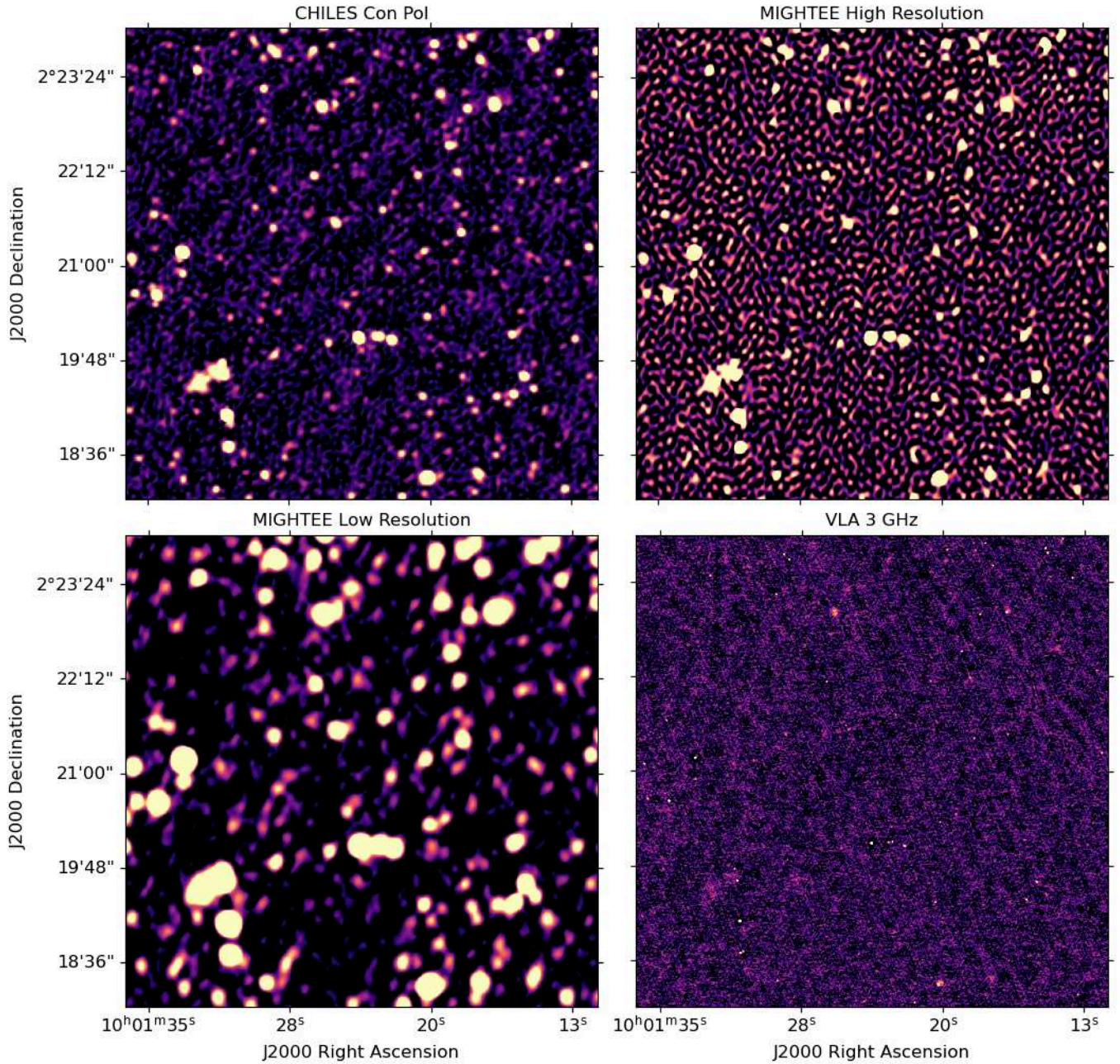
From the above investigation, we conclude that any sufficiently deep radio interferometric observation must correct for this systematic effect in a nature consistent with the observational set up of each individual observation in order to recover accurate flux densities, especially those observations that are plagued by negative sidelobes above the 5% level. For the CHILES Con Pol image, we correct for this offset by empirically finding the mean offset in a range of radial annuli and adding this value to all pixels in each annulus. As a result, we produce a zero-meant image corrected for this issue.

## 5. Discussion

### 5.1. Studying Individual Micro-Jansky Radio Sources

A zoom-in view of the CHILES Con Pol data shown in the top left panel of Figure 8 highlights the capability of the CHILES Con Pol data to identify individual radio continuum sources down to  $5 \mu\text{Jy}$  level. In comparison, the early science MIGHTEE 1.4 GHz survey image shown in the lower left panel has a lower resolution ( $\theta = 8.6''$ ) and a slightly lower thermal noise ( $\sigma = 1.7 \mu\text{Jy beam}^{-1}$ ), while the higher resolution MIGHTEE data shown on the upper right panel has a comparable resolution ( $\theta = 5''$ ) and a significantly higher noise ( $\sigma = 5.5 \mu\text{Jy beam}^{-1}$ , I. Heywood et al. 2022). Also shown for comparison is the VLA-COSMOS 3 GHz image (bottom right panel, V. Smolčić et al. 2017), which has a significantly higher angular resolution ( $\theta = 0.75''$ ) but is slightly shallower ( $\sigma \approx 4 \mu\text{Jy}$  at 2.3 GHz), and is at higher frequencies.

To translate this detection limit to galaxy properties at cosmological distances, we plot the  $5\sigma$  detection limit for SFR and the 1.4 GHz power as a function of redshift in Figure 11. To calculate SFR, we use the calibration in E. J. Murphy et al. (2011) that relates SFR and 1.4 GHz luminosity using the FIR–radio correlation. Our ability to identify and study *individual* star-forming galaxies (SFGs) with  $\text{SFR} \geq 10 M_{\odot} \text{ yr}^{-1}$  to  $z = 3$  is particularly important for characterizing star formation activities among the field galaxy populations in a wide range of cosmic epochs, including during the “Cosmic Noon.” These detections, and the cross matching of them with the existing COSMOS multiwavelength data, will allow for explorations into radio continuum science for fainter luminosities and more distant redshifts (H. B. Gim et al. 2025).



**Figure 8.** The total intensity radio continuum images of a  $3' \times 3'$  region for a detailed comparison among the CHILES Con Pol (top left), MIGHTEE high ( $\theta = 5''$ ; upper right) and low resolution ( $\theta = 8''$ ; 6; lower left I. Heywood et al. 2022), and VLA-COSMOS 3 GHz ( $\theta = 0''.75$ ; lower right V. Smolčić et al. 2017) data. The intensity ranges shown are between  $-1$  and  $+20 \mu\text{Jy}$  for the first three panels and  $-1$  and  $+10 \mu\text{Jy}$  for the VLA-COSMOS 3 GHz image. The VLA-COSMOS 1.4 GHz survey data (E. Schinnerer et al. 2007) is too shallow ( $\sigma \approx 10 \mu\text{Jy}$ ) to offer a meaningful comparison and is not shown.

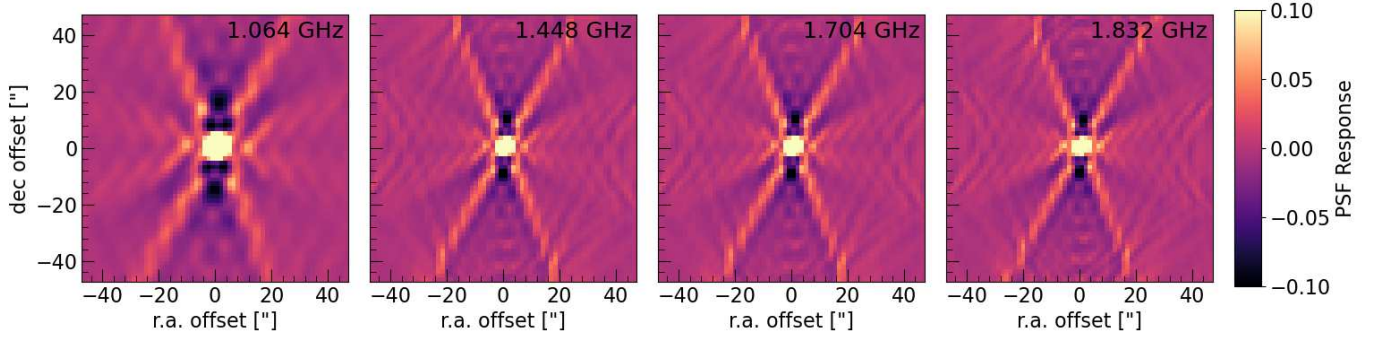
### 5.2. Statistical Inferences on the Sub- $\mu\text{Jy}$ Radio Population

Studies of source counts are a simple but powerful analysis method that can yield important insights into the cosmic evolution of an astronomical source population. The “ $P(D)$ ” analysis (P. A. G. Scheuer 1957), which is statistical modeling of a pixel distribution of a confusion-limited image, can provide further quantitative constraints on the source counts at flux-density ranges well below the individual detection limits (see a review by J. J. Condon et al. 2012, and references therein). Following the initial characterization of the sub-mJy radio population by K. J. Mitchell & J. J. Condon (1985) using

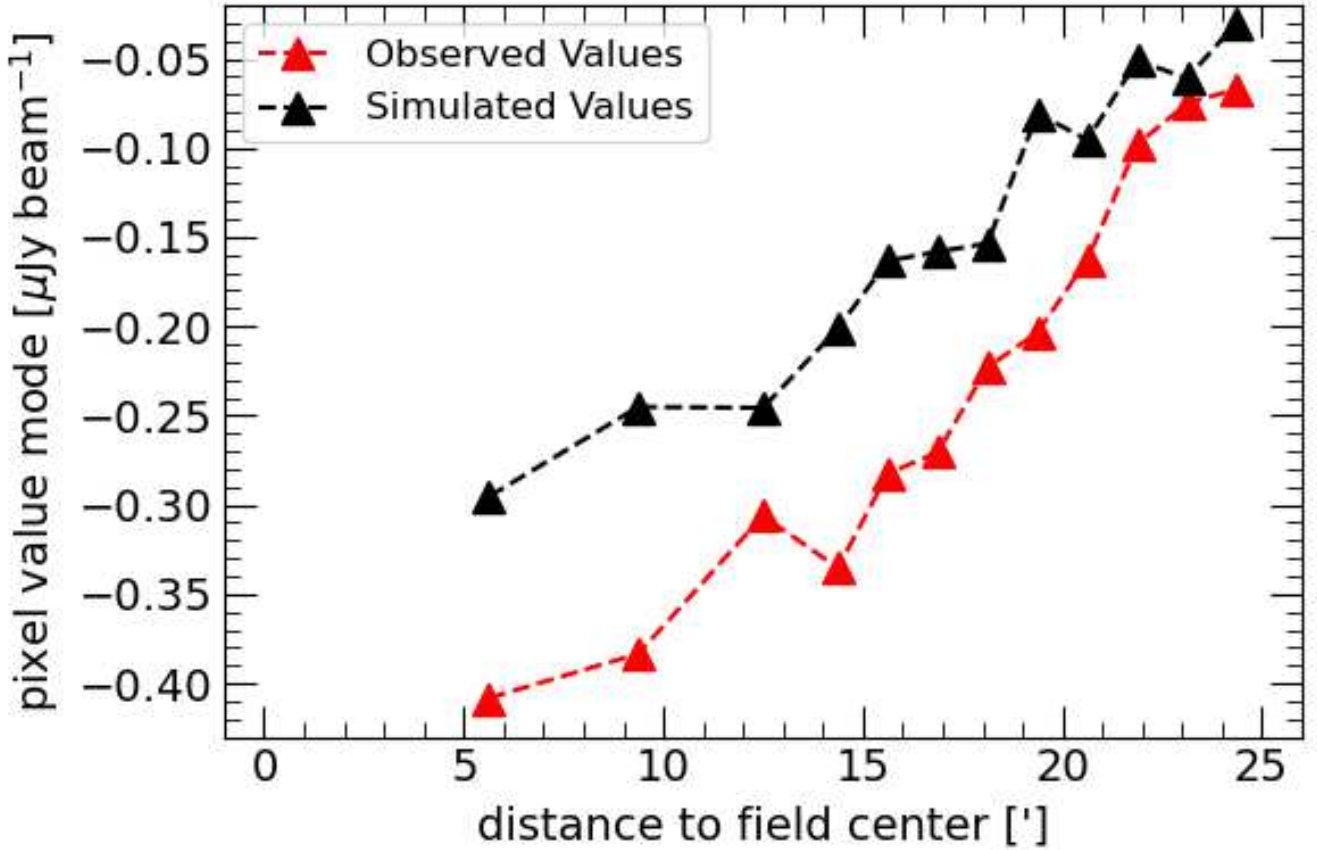
some of the early deep radio surveys, subsequent generations of ever deeper and wider surveys have extended the radio source counts to increasingly fainter flux-density levels and have characterized the nature of the radio sources as rapidly evolving populations of SFGs and AGNs (see A. M. Matthews et al. 2021, and references therein).

Apparent discrepancy among different surveys around  $S_{1.4\text{GHz}} \lesssim 100 \mu\text{Jy}$  have raised possible systematic problems with observations and modeling assumptions, such as angular resolution and flux sensitivity, source size and clustering, sample variance, etc. (see M. Bondi et al. 2008; F. N. Owen & G. E. Morrison 2008; G. de Zotti et al. 2010; I. Heywood et al.





**Figure 9.** The PSF response in each of the four sub-bands, whose center frequencies are indicated in the upper right corner of each panel, with a common color scale across each panel of  $-0.1$  to  $0.1$ .



**Figure 10.** The mean of the pixel-value distribution as a function of distance from field center for the simulated CHILES Con Pol observations of flux densities below  $5 \mu\text{Jy}$  (black markers), and for the CHILES Con Pol maps presented in this work (red markers).

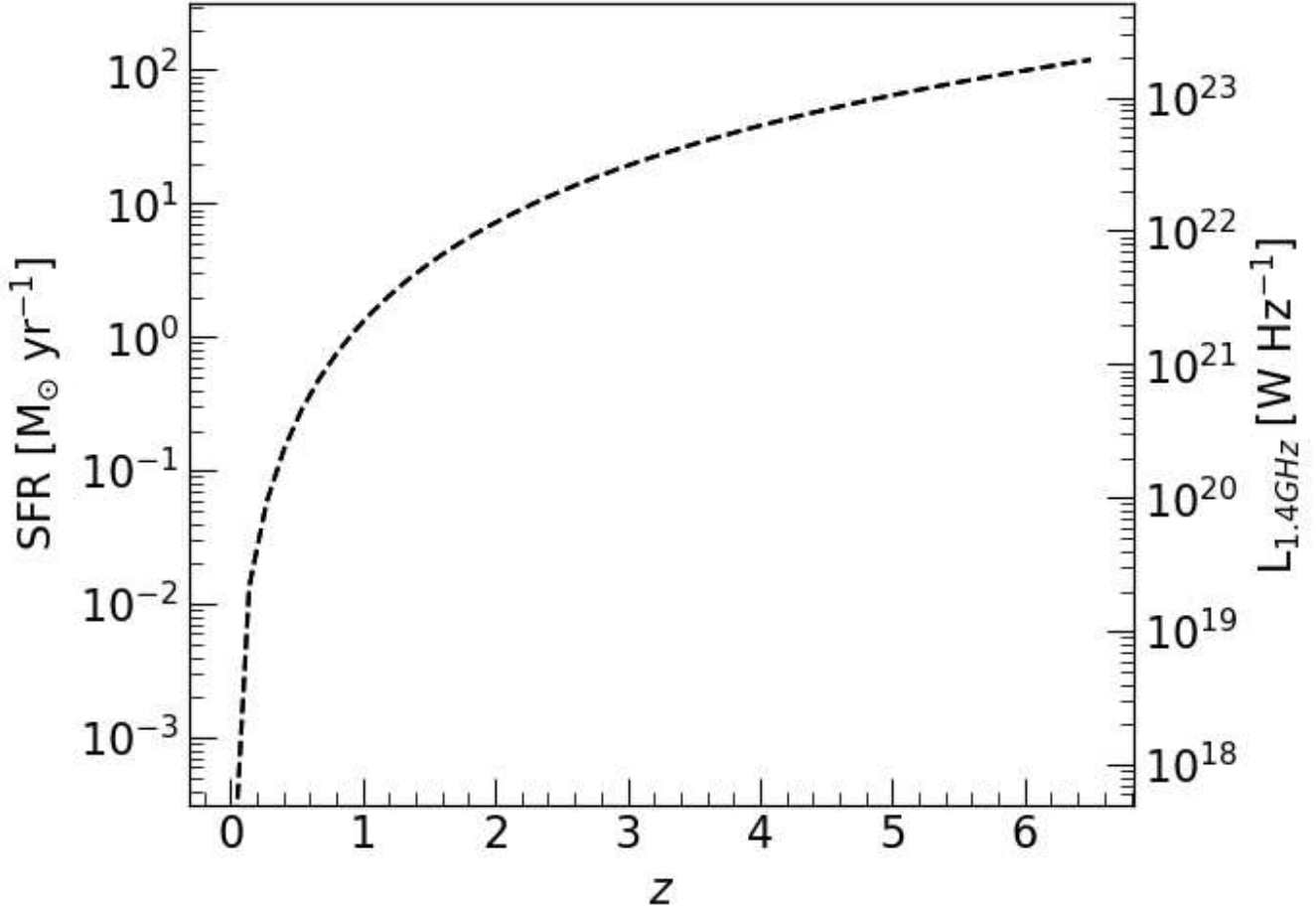
2013). Possible extra sources of extragalactic radio background by ARCADE 2 experiments (D. J. Fixsen et al. 2011) also raised new interests in the faint radio source counts and their contribution to the cosmic radio background (M. Seiffert et al. 2011; T. Vernstrom et al. 2011).

The Euclidean-normalized source counts derived from our  $P(D)$  analysis of the CHILES Con Pol data are shown in Figure 12, and they confirm the steepening of the number counts below  $S_{1.4\text{GHz}} = 100 \mu\text{Jy}$  as reported by recent studies such as by J. J. Condon et al. (2012) and T. Mauch et al. (2020). Our source counts derived for all pixels fainter than  $10 \mu\text{Jy}$  shown on the left panel nearly exactly match the source

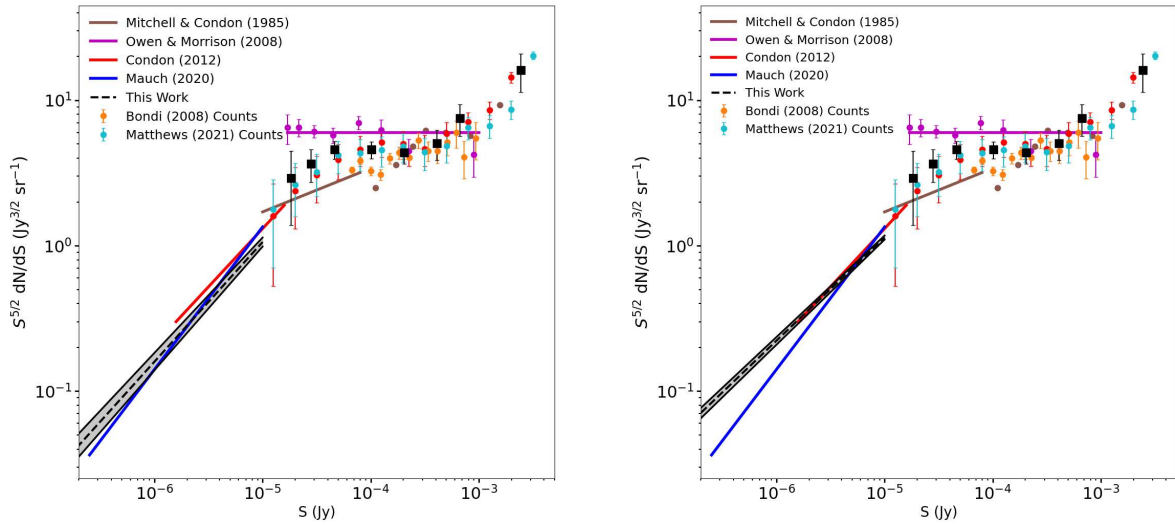
counts derived by J. J. Condon et al. (2012) using a VLA 3 GHz survey specifically designed for this purpose. Our source counts ( $dN/dS = 13,746S^{-1.677} \text{Jy}^{-1} \text{sr}^{-1}$ ) are slightly steeper than the recent source counts reported by T. Mauch et al. (2020) from their 1.28 GHz MeerKAT survey ( $dN/dS = 107,000S^{-1.52} \text{Jy}^{-1} \text{sr}^{-1}$ ) of the DEEP2 field.<sup>9</sup> T. Mauch et al. (2020) analyzed the same flux-density range using a least-squares power-law fit with a fixed image noise of  $\sigma_n = 0.55 \mu\text{Jy}$ . These authors did not offer a detailed

<sup>9</sup> Our Euclidean-normalized source counts  $S^{5/2}n(>S)$  shown in Figure 12 are integral counts and are flatter in comparison.





**Figure 11.** The  $5\sigma$  detection limit, assuming a  $1 \mu\text{Jy}$  noise level, for the CHILES Con Pol Survey for radio continuum derived SFR and the 1.4 GHz luminosity as a function of redshift.



**Figure 12.** Euclidean-normalized source counts. The 1.4 GHz  $P(D)$  distribution derived from the CHILES Con Pol data for the pixel ranges of  $0.1\text{--}10 \mu\text{Jy beam}^{-1}$  (left panel) and  $0.1\text{--}60 \mu\text{Jy beam}^{-1}$  (right panel) are shown by black-dashed lines. The source count derived from the CHILES Con Pol component catalog is shown by large black dots (H. B. Gim et al. 2025). They are compared with the published  $P(D)$  (K. J. Mitchell & J. J. Condon 1985; F. N. Owen & G. E. Morrison 2008; J. J. Condon et al. 2012; T. Mauch et al. 2020) and compilations of number counts by M. Bondi et al. (2008) and A. M. Matthews et al. (2021).

descriptions of uncertainties in their model parameters, but their uncertainties shown in their Figure 14 are large enough to be consistent with the J. J. Condon et al. (2012) results, and thus in turn also with our results.

This comparison is interesting because the CHILES Con Pol data can, in principle, probe deeper into the fainter flux levels ( $S_{1.4\text{GHz}} \leq 5 \mu\text{Jy}$ ) by lowering the confusion limit with a higher angular resolution, as discussed in Section 5.1. We have also taken a different mathematical formulation for the  $P(D)$  analysis, using a Lévy  $\alpha$ -stable distribution (D. Herranz et al. 2004) and a Markov Chain Monte Carlo (MCMC) model fitting, rather than the more commonly used method by J. J. Condon (1974), in order to obtain more robust constraints on the modeling parameters (see Appendix B for a full description). Therefore, this steepening of the power-law slope below  $S_{1.4\text{GHz}} = 100 \mu\text{Jy}$  seems to be a robust result.

Source confusion depends on the actual source number counts and observed angular resolution, and the classic source count based on individual detections is presented in the catalog paper (H. B. Gim et al. 2025). We repeated the same  $P(D)$  analysis to a high flux-density range, up to  $60 \mu\text{Jy}$  (right panel of Figure 12) and  $100 \mu\text{Jy}$  (not shown but see Appendix C) in order to test the robustness of the standard assumption of a single power law, which may be hampered by an inflection point between 20 and  $100 \mu\text{Jy}$ , and a confusion limit. Both the derived power-law index and the normalization increased when the pixel ranges are expanded to higher flux-density levels. In hindsight, this is not surprising since the source counts model by K. J. Mitchell & J. J. Condon (1985) found an even steeper slope with a higher normalization when they analyzed a higher flux-density range (between 10 and  $86 \mu\text{Jy}$ ), indicating a significant change in the shape of the source counts. As discussed by T. Mauch et al. (2020) in their choice of a power-law model, this is probably not a robust mathematical form. We also note that the right flux-density range to model was not a priori obvious when we started this work, and this changing shape of the source counts is something future studies should keep in mind.

### 5.3. Imaging Extended Sources

To highlight the significance of the sensitivity and resolution of CHILES Con Pol, we compare our total intensity image to those presented in other surveys in the COSMOS field, specifically the Deep VLA-COSMOS Survey (E. Schinnerer et al. 2007) and the MIGHTEE Survey (M. Jarvis et al. 2016). In brief, the VLA-COSMOS survey is a mosaic over the entire COSMOS field at 1.4 GHz taken with the VLA in A-configuration that achieves an rms noise of approximately  $12 \mu\text{Jy beam}^{-1}$  at  $1''.5$  resolution (see E. Schinnerer et al. (2007) for further details). The MIGHTEE survey is a series of four mosaics, including a field approximately the entire size of the COSMOS field, with the MeerKAT telescope at 1.4 GHz. In the COSMOS field, the MIGHTEE survey achieves an rms noise of  $2.2 \mu\text{Jy beam}^{-1}$  at  $5''.2$  resolution and  $1.6 \mu\text{Jy beam}^{-1}$  at  $8''.6$  resolution (see C. L. Hale et al. 2025 for more details).

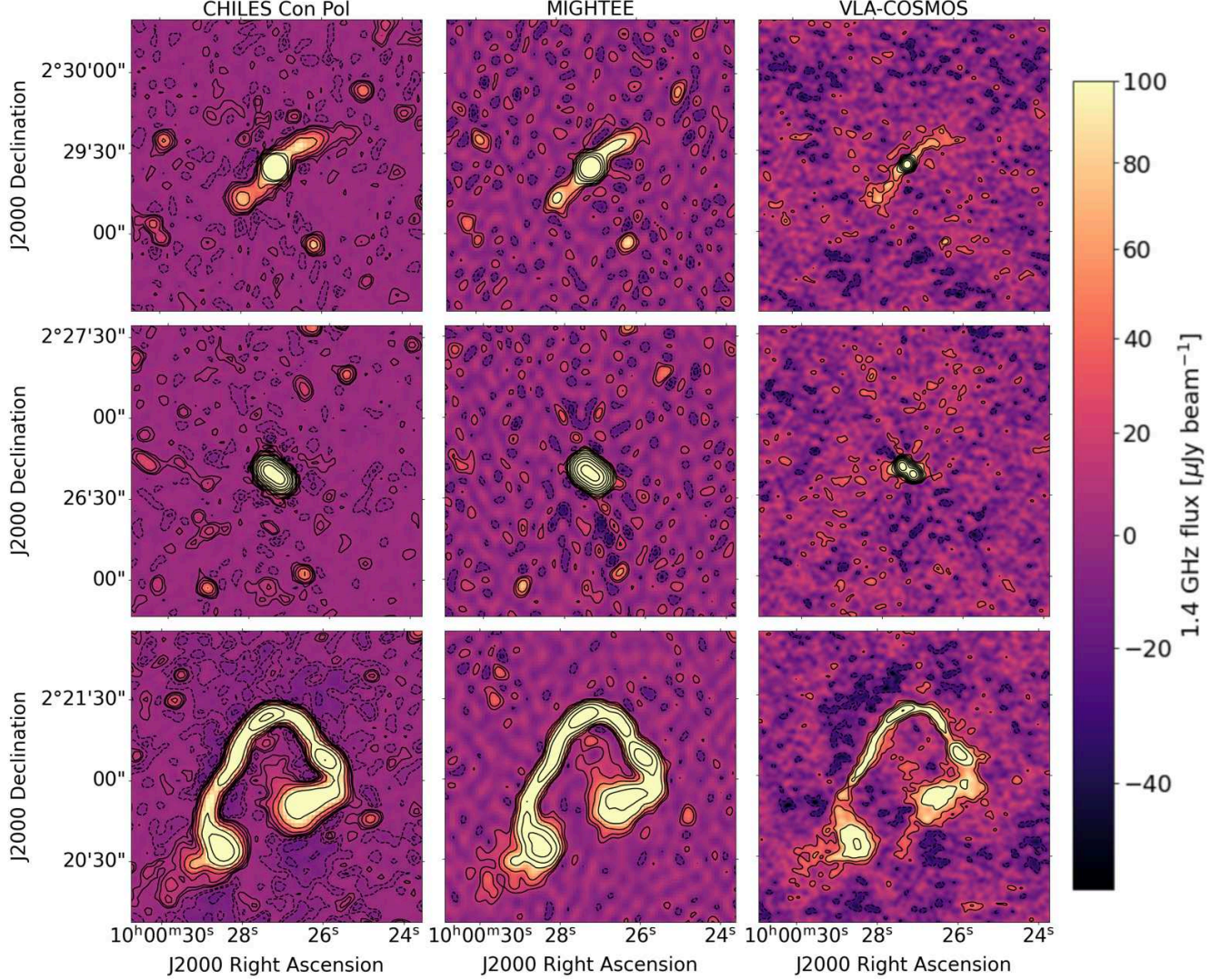
We measured our rms noise to be  $1.3 \mu\text{Jy beam}^{-1}$  at  $4''.5$  resolution, when measured via a pixel histogram far from field center. This translates to a factor of approximately two better sensitivity than MIGHTEE at similar resolutions, and a factor of nine better sensitivity than VLA-COSMOS, but at a factor three worse resolution. These differences in sensitivity, survey size, and resolution result in these three surveys being

complementary data sets for different scientific pursuits. CHILES Con Pol has better point source sensitivity than MIGHTEE, and will therefore be able to better detect individual SFGs and AGN at higher redshifts and will be less affected by confusion noise. However, the primary beam of MeerKAT is significantly larger than the VLA’s, they are covered by the 50% point at 1.4 GHz is approximately 3.5 times larger, which allowed for MIGHTEE to probe a larger area and observe interesting objects not seen by CHILES Con Pol, such as those in J. Delhaize et al. (2021). Both MIGHTEE and CHILES Con Pol have better sensitivity than VLA-COSMOS, but the excellent resolution of VLA-COSMOS allows for the data to be used to show smaller features, such as the complex structures in AGN jets, such as the exquisite detail in jets of the “Earmuffs” galaxy, an example of a wide-angle head tail galaxy (V. Smolčić et al. 2007).

In order to further compare our results to that of MIGHTEE and VLA-COSMOS, we compare the three surveys for two sources in Figure 13. In the top row of Figure 13, we show the CHILES Con Pol, MIGHTEE, and VLA-COSMOS, respectively, from left to right, images centered on an interesting extended source present in all three images. In these, we see that all three surveys detect the extended source, but that CHILES Con Pol also has more detections of point sources around it than VLA-COSMOS, and of higher signal-to-noise than MIGHTEE. In the middle row of Figure 13, we show images over the strongest in-field CHILES Con Pol source (approximately 15 mJy at 1.4 GHz) for the three surveys. In the bottom row, we see an example in which CHILES Con Pol has the strongest artifacts, demonstrated by the negative residuals surrounding the positive flux. However, we are still able to detect extended features at sufficient signal-to-noise compared to the other surveys, as well as recover the point sources around the source, evident in both CHILES Con Pol and MIGHTEE. The three surveys each show some residual imaging artifacts, seen as stronger positive and negative contours surrounding the source, but these artifacts are of a similar level, indicating that each survey reaches similar acceptable levels of accuracy in calibration.

The imaging approach we used in imaging our field is nontraditional, as opposed to the methods used for MIGHTEE and VLA-COSMOS. More specifically, for MIGHTEE the team used second and third generation calibration, self-calibration, and direction-dependent mitigation techniques in order to correct for sources of error in their field (I. Heywood et al. 2022). Our utilization of uv-baseline fitting and subtraction, a technique common for spectral line data (T. J. Cornwell et al. 1992), of sources out of the field is significantly different. However, a comparison of images show that our technique is able to just as successfully produce images of high dynamic range, with artifacts from these sources being no higher than that of those in the MIGHTEE data. One of our problematic sources, the aforementioned “Earmuffs” galaxy, does continue to cause problems in our image—since it lies in our field, we are unable to use this subtraction technique as we would risk subtracting flux from in-field sources. The data is, fortunately, only affected at the R. A. of the source due to the sidelobes being limited to the north–south direction, due to the fact that COSMOS is an equatorial field.





**Figure 13.** The 1.4 GHz total intensity radio continuum images for an interesting likely AGN within the COSMOS (top) a 15 mJy source, the brightest source within CHILES Con Pol, (middle), and the “Earmuffs” galaxy (bottom), for CHILES Con Pol, MIGHTEE (I. Heywood et al. 2022), and VLA-COSMOS (E. Schinnerer et al. 2007), from left to right. The color bar on the far right applies to all images and each map has contours at the  $\pm 2\sigma$  for  $n$  values 1–11, with  $\sigma$  corresponding to their respective reported rms sensitivities.

## 6. Summary

In this paper, we introduced and described the CHILES Con Pol Survey. We discussed the survey design, specifically the ability of the survey to utilize unused resources from the CHILES survey, allowing for CHILES Con Pol to obtain 1027 hr of integration time from the VLA in B-configuration. We described our calibration and imaging routines, and discussed the RFI environment and mitigation techniques for our observations. Using these routines, we present the first total intensity image of the CHILES Con Pol data, which has achieved an rms noise of  $1.3 \mu\text{Jy}$  per synthesized beam of size  $4''.5 \times 4''.0$ , the best sensitivity, to date, at this resolution. Our image is then discussed in the context of other 1.4 GHz surveys of the COSMOS field, where we compare and contrast the image properties of these surveys. Additionally, we compare and contrast source counts from the CHILES Con Pol data and place these results in the context of the previous literature. Finally, we discussed the scientific capabilities of

this data set to probe galaxy evolution across redshift space using radio continuum data at an unprecedented level. Future data releases will include images of all four stokes planes, and in subsequent papers we will discuss the source identification and COSMOS catalog matching (H. B. Gim et al. 2025), as well as probe the properties of these sources.

## Acknowledgments

We thankfully acknowledge the thorough review and helpful comments from the anonymous referee that aided in improving the demonstration and presentation of results, scientific discussion, and clarity. We are very grateful for helpful discussion with, and insights from, J. H. van Gorkom.

Support for this work was partially provided by the NSF through award SOSP 21A-002 from the NRAO. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under a cooperative agreement by Associated Universities, Inc. D.J.P.’s research is partially



supported by the South African Research Chair Initiative of the Department of Science & Innovation and the National Research Foundation (NRF). D.J.P. and N.M.L. thank the WVU Eberly College Dean's office for partial support, and partial support from NSF grant AST 1412578. M.S.Y.'s research is partially supported by an NSF grant AST 1413102.

*Facility:* National Radio Astronomy Observatory (NRAO) Expanded Very Large Array (EVLA).

*Software:* astropy (Astropy Collaboration et al. 2013), CASA (J. P. McMullin et al. 2007).

## Appendix A

### $P(D)$ Analysis, Source Counts, and Source Confusion

For a confusion limited image, the deflection  $D$  of pixel values from the mean  $\mu$  is described as the “probability of deflection” or  $P(D)$  (P. A. G. Scheuer 1957), and this quantity represents a spatial variation in the integrated total flux due to a random fluctuation in source density within an observing beam. Assuming differential source counts of the form:

$$\frac{dN}{dS} = N_0 S^{-\kappa}. \quad (\text{A1})$$

J. J. Condon (1974) have shown that the noise-free confusion probability distribution function (PDF) is an integral and leads to a scaling relation:

$$P[(N_0 \Omega_e)^{1/(\kappa-1)} D] = (N_0 \Omega_e)^{-1/(\kappa-1)} P(D), \quad (\text{A2})$$

where the effective solid angle  $\Omega_e \equiv \frac{\Omega_b}{\kappa-1}$  depends on the Gaussian beam solid angle  $\Omega_b \equiv \frac{\pi \theta_1 \theta_2}{4 \ln 2}$  ( $\theta_1$  and  $\theta_2$  are the full width at half-maximum widths of the elliptical Gaussian beam). The source counts within an effective beam area  $\Omega_e$  is  $N_0 \Omega_e = \eta_1^{-1}$ , where:

$$\eta_1^{-1} = \frac{2\Gamma(\kappa/2)\Gamma[(\kappa+1)/2]\sin[\pi(\kappa-1)/2]}{\pi^{\kappa+1/2}} \quad (\text{A3})$$

( $\Gamma$  is the factorial function). Therefore, the functional shape of  $P(D)$  relies upon only  $\kappa$  for power-law source counts.

The observed  $P(D)$  is the convolution of the source confusion and image noise  $\sigma_n$ . Since  $\Omega_e$  is derived from the normalized PSF with a gain  $G$  (i.e.,  $\Omega_e \equiv \int G^{\kappa-1} d\Omega$ ), the sidelobes of the PSF have a larger effect as  $\kappa$  decreases. The rms confusion variance  $\sigma_c^2$  diverges for all  $\frac{dN}{dS}$ , and  $P(D)$  must be truncated above a cutoff deflection  $D_c$  (J. J. Condon 1974; T. Mauch et al. 2020). The confusion noise  $\sigma_c$  due to sources fainter than a signal-to-noise ratio  $q = \frac{D_c}{\sigma_c}$  is given by:

$$\sigma_c = \left( \frac{q^{3-\kappa}}{3-\kappa} \right)^{1/(\kappa-1)} (N_0 \Omega_e)^{1/(\kappa-1)}, \quad (\text{A4})$$

and the number of beam solid angles per source brighter than  $D_c$  is  $\beta = \frac{q^2}{3-\kappa}$  (J. J. Condon et al. 2012). This leads to the rule-of-thumb for source confusion, “one  $5\sigma$  source per 25 beam areas” for  $\kappa \sim 2$ .

## Appendix B

### Methodology

To accurately extract the noise-free  $P(D)$  (and thus  $\frac{dN}{dS}$  and  $\sigma_c$ ),  $N_0$  and  $\kappa$  have to be optimized as free parameters. From the determined or parameterized  $\sigma_n$ , the observed  $P(D)$  is extracted from the convolution of the image noise and the

noise-free  $P(D)$ . Equation (A2) is ultimately derived from the  $P(D)$  equation (J. J. Condon 1974):

$$P(D) = 2 \int_0^\infty \exp(-N_0 \Omega_e \eta_1 \omega^{\kappa-1}) \cos(N_0 \Omega_e \eta_2 \omega^{\kappa-1} + 2\pi \omega D) d\omega, \quad (\text{B1})$$

where  $\eta_1$  is defined in Equation (A3) and  $\eta_2$  is

$$\eta_2 = \pi^{\kappa-1/2} \Gamma\left[\frac{1}{2}(2-\kappa)\right] / \left\{ 2\Gamma\left[\frac{1}{2}(\kappa+1)\right] \right\}. \quad (\text{B2})$$

The  $D$  term in Equation (B1) now represents the mean deflection  $\bar{D}$  instead of absolute zero, unless  $1 < \kappa < 2$ , where the sky brightness no longer diverges and Olbers' paradox can be avoided (T. Mauch et al. 2020).

### B.1. The Lévy Distribution

D. Herranz et al. (2004) have noted that the pixel distribution of a deep radio survey image resembles an extremely impulsive noise PDF described by a Lévy  $\alpha$ -stable distribution, a statistics model popularly used in economics and engineering. The characteristic function for the Lévy distribution completely describes any PDF, such as Gaussian or Poisson distributions, and has no analytical form. Four parameters govern the behavior of the Lévy distribution:

1.  $\alpha$  is the characteristic exponent that describes the impulsiveness of the function. In the domain  $\alpha = (0, 2]$ ,  $\alpha = 2$  describes a Gaussian, and as  $\alpha \rightarrow 0$ , the function becomes more impulsive;
2.  $\beta$  controls how skewed or symmetric the function is on the domain  $\beta = [-1, 1]$ , where  $\beta = 0$  resembles a symmetric distribution while  $\beta = 1$  completely left-skews the function;
3.  $\gamma$  is the scale parameter, which controls how spread out the distribution is for  $\gamma > 0$ ;
4.  $\mu$  is the shift parameter, which reflects the mean of the distribution.

The Lévy distribution is then described in the Fourier space as (see D. Herranz et al. 2004):

$$\psi(w) = \exp\{i\mu w - \gamma|w|^\alpha B_{w,\alpha}\}, \quad (\text{B3})$$

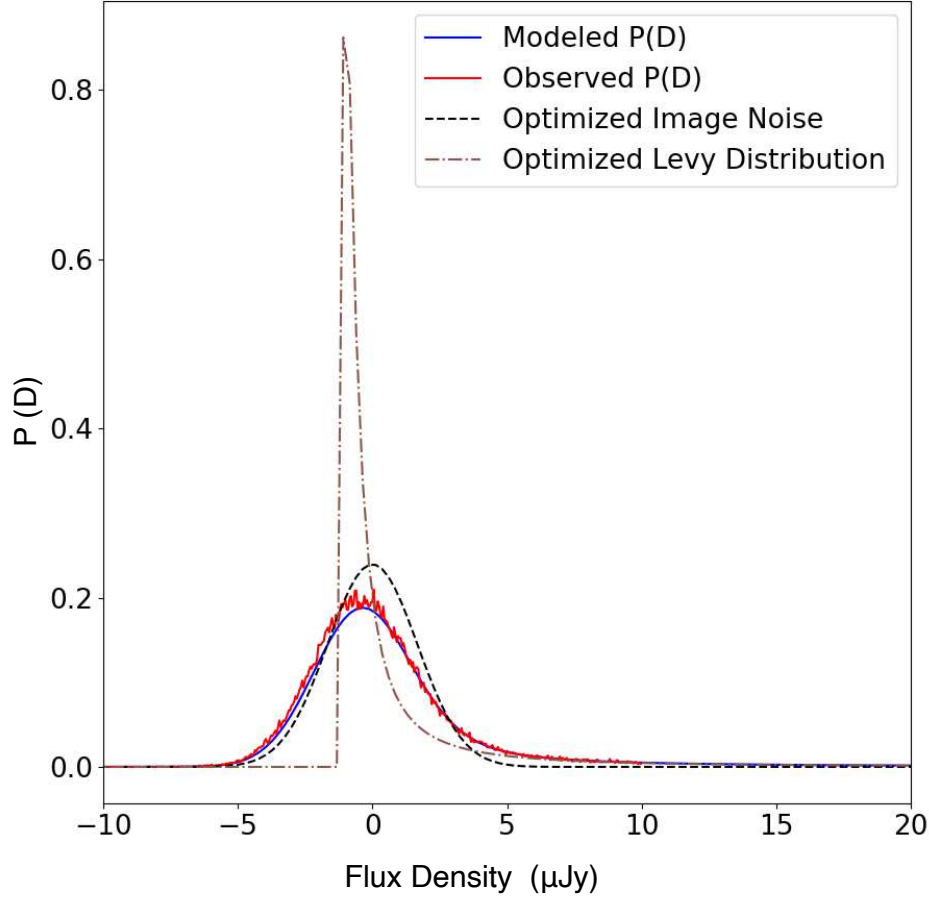
$$B_{w,\alpha} = \begin{cases} \left[ 1 + i\beta \operatorname{sgn}(w) \tan\left(\frac{\alpha\pi}{2}\right) \right] & \text{for } \alpha \neq 0 \\ \left[ 1 + i\beta \operatorname{sgn}(w) \frac{2}{\pi} \log|w| \right] & \text{for } \alpha = 0. \end{cases} \quad (\text{B4})$$

A new model for  $P(D)$  can then be constructed by convolving a parameter-optimized Lévy distribution with a Gaussian distribution, representing the image noise. As a result, the Lévy  $\alpha$ -stable distribution becomes a maximum-skewed ( $\beta = 1$ ) distribution with no localization due to convolution ( $\mu$  is arbitrary).

We have chosen to take this formulation of  $P(D)$  over the more commonly adopted form described by J. J. Condon (1974) because this form is better suited for the MCMC parameter optimization, which can yield a more robust characterization of the source counts model parameters and their uncertainties.

### B.2. $P(D)$ Implementation

$P(D)$  is now described as the convolution of the noise-free distribution due to the underlying confused source population and Gaussian-assumed image noise. Figure 14 illustrates these



**Figure 14.** Distribution functions of observed  $P(D)$ , Gaussian noise, Lévy distribution, and modeled  $P(D)$  are shown with red solid, black-dashed, black-dashed-dotted, and blue solid lines, respectively.

distributions as a function of flux density, depicting the observed  $P(D)$  distribution (red solid line), Gaussian noise distribution (black-dashed line), and Lévy distribution (black-dashed-dotted line). The modeled  $P(D)$  distribution (blue solid line) represents the convolution of the Lévy distribution and image noise, demonstrating excellent agreement with the observed distribution across the measured flux-density range.

The parameters that define the noise-free distribution, shown in Equation (B4), are  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\mu$ . Since  $\beta$  and  $\mu$  are not relevant to describing the confusion source density, only  $\alpha$  and  $\gamma$  are used in the calculations, where

$$\alpha = \kappa - 1, \quad (\text{B5})$$

$$\gamma = \frac{\pi^{3/2} N_0 \Omega_e}{2^{\alpha+1} \Gamma\left(\frac{\alpha+1}{2}\right) \Gamma\left(\frac{\alpha+2}{2}\right) \sin\left(\frac{\alpha\pi}{2}\right)}. \quad (\text{B6})$$

Rearranging these equations to solve for  $N_0$  and  $\kappa$ , the source counts parameters can be found by parameterizing and optimizing  $\alpha$  and  $\gamma$  as:

$$\kappa = \alpha + 1, \quad (\text{B7})$$

$$N_0 = \gamma \frac{2^{\alpha+1} \Gamma\left(\frac{\alpha+1}{2}\right) \Gamma\left(\frac{\alpha+2}{2}\right) \sin\left(\frac{\alpha\pi}{2}\right)}{\pi^{3/2} \Omega_e}. \quad (\text{B8})$$

Therefore, Euclidean-normalized source count is expressed as

$$S^{2.5} \frac{dN}{dS} = N_0 S^{2.5-\kappa} = N_0 S^{1.5-\alpha}, \quad (\text{B9})$$

where  $\kappa$  and  $N_0$  are defined as Equations (B7) and (B8).

### B.3. Data Modeling Using MCMC

The most efficient way to model  $P(D)$  is to utilize the MCMC method due to its effective ability to model complicated functions. The advantage of using MCMC methods over other optimization methods, such as the nonlinear least squares (NLS) regression optimizer, is that MCMC incorporates randomness to best optimize multi-parameter spaces with trial-and-error precision. The NLS regression technique requires precise knowledge of the priors, is more susceptible to outliers, and can also get stuck on local minimums rather than the true minimum.

The downsides of using an MCMC are computational intensity and time. Producing one realization of  $P(D)$  computationally takes a significant fraction of a second on a computer with a powerful CPU, which is much longer than computing the result for a much simpler function. Since optimizing  $P(D)$  sufficiently requires hundreds of iterations, producing a good result can take many hours. Therefore, a

good advanced planning of the resolution of the data and the number of iterations is required. For convenience and reliability, we have utilized the Python package `emcee` (D. Foreman-Mackey et al. 2013). This package is widely used and is well documented. We also found it to work well with our complex function for the optimization.

### Appendix C Results

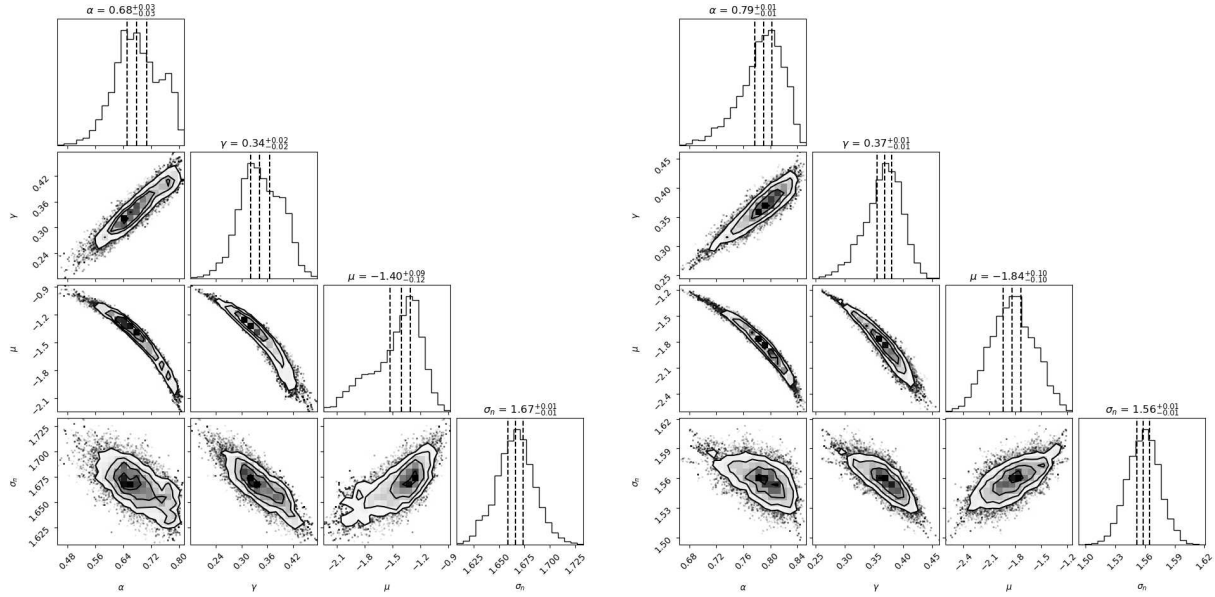
We performed the MCMC modeling with the CHILES Con Pol image which was generated with Briggs weighting at  $R = +0.5$ . The primary beam attenuation was not corrected in this image. We selected pixel values within a radius of 180 pixels (corresponding to  $4.5''$ ), ensuring minimal primary beam attenuation (6%) while maintaining statistical robustness.

Figure 15 presents the posterior probability distributions of the four model parameters ( $\alpha$ ,  $\gamma$ ,  $\mu$ , and  $\sigma_0$ ) for input pixel ranges of  $[-10, +10]$  and  $[-60, +60] \mu\text{Jy beam}^{-1}$ , respectively. The corresponding best-fit number count models are illustrated in Figure 12, along with observational data from the literature and the CHILES Con Pol survey (H. B. Gim et al. 2025). The best-fit model parameters are also summarized in Table 3, including the model fitting results for the input pixel range of  $[-100, +100] \mu\text{Jy beam}^{-1}$  (no plots are shown).

It should be noted that the best-fit model parameters are sensitive to the initial parameters, particularly for the input

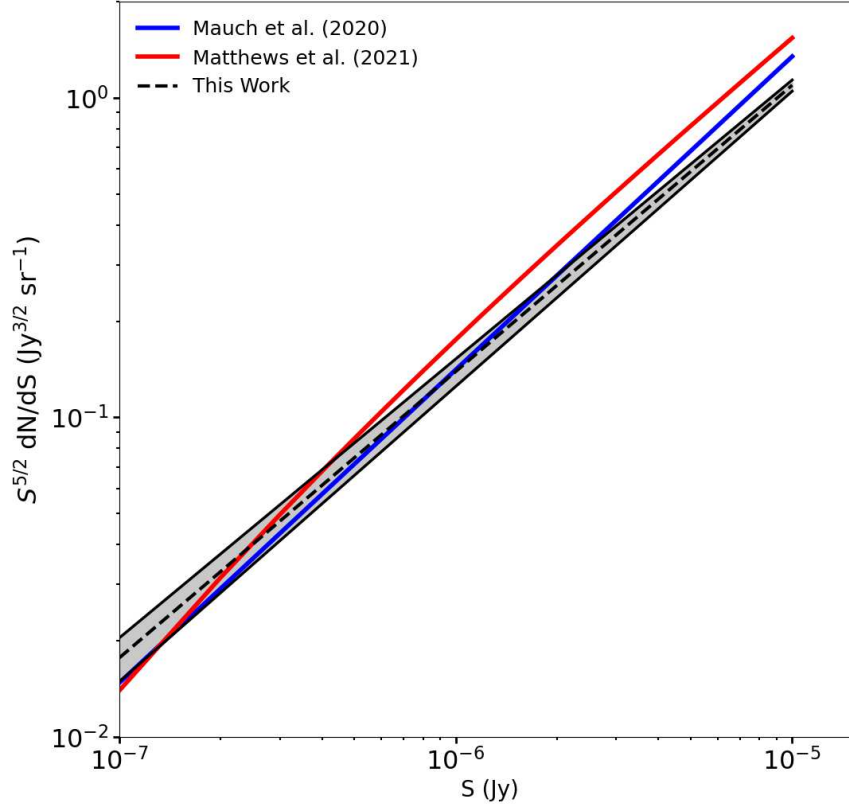
range of  $[-10, +10] \mu\text{Jy beam}^{-1}$ . This sensitivity may be attributed to the presence of multiple local optima. To address this, we tested various initial parameters and identified the best-fit model parameters by quantifying the difference between observed  $P(D)$  and  $P(D)$  constructed from the best-fit model parameters.

Furthermore, we extended our  $P(D)$  analysis to the MeerKAT DEEP2 image. Following the methodology established by T. Mauch et al. (2020), we applied the same code to the MeerKAT DEEP2 image within a 1250-pixel radius and flux-density range of  $-10 < S < 10 \mu\text{Jy}$ . Comparisons with the prior studies using the same observational data demonstrate consistency between models, as illustrated in Figure 16, which presents Euclidean-normalized source counts from T. Mauch et al. (2020, blue solid line), A. M. Matthews et al. (2021, red solid line), and our  $P(D)$  distribution model (black-dashed line). While our  $P(D)$  model exhibits a moderately shallower slope, it remains consistent with T. Mauch et al.'s (2020) model within  $1\sigma$  uncertainty bounds. The model presented by A. M. Matthews et al. (2021) diverges slightly from both our model and that of T. Mauch et al. (2020), particularly above  $1 \mu\text{Jy}$ . This discrepancy likely stems from differing source count formulations; specifically, our approach and that of T. Mauch et al. (2020) implemented power-law models for source counts, whereas A. M. Matthews et al. (2021) employed a cubic polynomial function approximation.



**Figure 15.** The posterior probability distributions of the four model parameters such as  $\alpha$ ,  $\gamma$ ,  $\mu$ , and  $\sigma_0$  are presented in 2D corner plots for models with an input pixel-value range of  $[-10, +10]$  in the left and  $[-60, +60] \mu\text{Jy beam}^{-1}$  in the right panel, respectively.





**Figure 16.** Comparisons of Euclidean-normalized source counts with MeerKAT DEEP2 survey. Blue and red solid lines indicate the source counts obtained from T. Mauch et al. (2020) and A. M. Matthews et al. (2021) for the DEEP2 field observed at the MeerKAT, respectively, while the black-dashed line shows the source count obtained using our  $P(D)$  model to the MeerKAT DEEP2 image.

**Table 3**  
Summary of the Best-fit Model Parameters

Flux-density Range	$\alpha$	$\gamma$	$\mu$	$\sigma_0$
$0.1 \leq S_{1.4\text{GHz}} \leq 10 \mu\text{Jy}$	$0.677^{+0.028}_{-0.030}$	$0.343^{+0.021}_{-0.023}$	$-1.404^{+0.122}_{-0.091}$	$1.666^{+0.008}_{-0.007}$
$0.1 \leq S_{1.4\text{GHz}} \leq 60 \mu\text{Jy}$	$0.790^{+0.013}_{-0.012}$	$0.369^{+0.013}_{-0.012}$	$-1.837^{+0.105}_{-0.099}$	$1.558^{+0.0007}_{-0.006}$
$0.1 \leq S_{1.4\text{GHz}} \leq 100 \mu\text{Jy}$	$0.854^{+0.011}_{-0.012}$	$0.424^{+0.012}_{-0.013}$	$-2.576^{+0.214}_{-0.179}$	$1.538 \pm 0.008$

### ORCID iDs

Min S. Yun <https://orcid.org/0000-0001-7095-7543>  
Hansung B. Gim <https://orcid.org/0000-0003-1436-7658>  
D.J. Pisano <https://orcid.org/0000-0001-7996-7860>  
Emmanuel Momjian <https://orcid.org/0000-0003-3168-5922>  
Chris Hales <https://orcid.org/0000-0002-3733-2565>

### References

- Andrews, S. K., Driver, S. P., Davies, L. J. M., et al. 2017, *MNRAS*, **464**, 1569  
Aretxaga, I., Wilson, G. W., Aguilar, E., et al. 2011, *MNRAS*, **415**, 3831  
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33  
Becker, R. H., White, R. L., & Helfand, D. J. 1995, *ApJ*, **450**, 559  
Bell, E. F. 2003, *ApJ*, **586**, 794  
Bock, D. C. J., Large, M. I., & Sadler, E. M. 1999, *AJ*, **117**, 1578  
Bondi, M., Ciliegi, P., Schinnerer, E., et al. 2008, *ApJ*, **681**, 1129  
Capak, P., Aussel, H., Ajiki, M., et al. 2007, *ApJS*, **172**, 99  
Carlson, B. R., & Dewdney, P. E. 2000, *EIL*, **36**, 64  
Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2023, *ApJ*, **954**, 31  
Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, **819**, 62  
Condon, J. J. 1974, *ApJ*, **188**, 279  
Condon, J. J. 1992, *ARA&A*, **30**, 575  
Condon, J. J., Cotton, W. D., Fomalont, E. B., et al. 2012, *ApJ*, **758**, 14  
Cornwell, T. J., Uson, J. M., & Haddad, N. 1992, *A&A*, **258**, 583  
Cresci, G., & Maiolino, R. 2018, *NatAs*, **2**, 179  
Davies, L. J. M., Driver, S. P., Robotham, A. S. G., et al. 2015, *MNRAS*, **447**, 1014  
de Zotti, G., Massardi, M., Negrello, M., & Wall, J. 2010, *A&ARv*, **18**, 1  
Delhaize, J., Heywood, I., Prescott, M., et al. 2021, *MNRAS*, **501**, 3833  
Fanaroff, B. L., & Riley, J. M. 1974, *MNRAS*, **167**, 31P  
Fernández, X., Gim, H. B., van Gorkom, J. H., et al. 2016, *ApJL*, **824**, L1  
Fernández, X., van Gorkom, J. H., Hess, K. M., et al. 2013, *ApJL*, **770**, L29  
Fixsen, D. J., Kogut, A., Levin, S., et al. 2011, *ApJ*, **734**, 5

- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Gim, H. B., Yun, M. S., Luber, N. M., et al. 2025, *AJ*, **170**, 60
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35
- Hale, C. L., Heywood, I., Jarvis, M. J., et al. 2025, *MNRAS*, **536**, 2187
- Hales, C. A. 2016a, interpgain, v1.1, Zenodo, doi:10.5281/zenodo.163002
- Hales, C. A. 2016b, antintflag, v1.0, Zenodo, doi:10.5281/zenodo.163546
- Hales, C. A., & Middelberg, E., 2014 pieflag: CASA Task to Efficiently Flag Bad Data, Astrophysics Source Code Library, ascl:1408.014
- Hardcastle, M. J., & Croston, J. H. 2020, *NewAR*, **88**, 101539
- Herranz, D., Kuruoglu, E. E., & Toffolatti, L. 2004, *A&A*, **424**, 1081
- Herrera Ruiz, N., Middelberg, E., Deller, A., et al. 2017, *A&A*, **607**, A132
- Herrera Ruiz, N., Middelberg, E., Deller, A., et al. 2018, *A&A*, **616**, A128
- Heywood, I., Jarvis, M. J., & Condon, J. J. 2013, *MNRAS*, **432**, 2625
- Heywood, I., Jarvis, M. J., Hale, C. L., et al. 2022, *MNRAS*, **509**, 2150
- Ito, K., Tanaka, M., Miyaji, T., et al. 2022, *ApJ*, **929**, 53
- Jarvis, M., Taylor, R., Agudo, I., et al. 2016, in *MeerKAT Science: On the Pathway to the SKA 6*, ed. R. Taylor, F. Camilo, L. Leeuw, & K. Moodley (Trieste: Proceedings of Science), **1**
- Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, **172**, 196
- Koekemoer, A. M., Faber, S. M., & Ferguson, H. C. 2011, *ApJS*, **197**, 36
- Kolatt, T. 1998, *ApJ*, **495**, 564
- Lacy, M., Baum, S. A., Chandler, C. J., et al. 2020, *PASP*, **132**, 035001
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24
- Laing, R. A. 1980, *MNRAS*, **193**, 439
- Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2020, *A&A*, **643**, A1
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, **576**, A79
- Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, **817**, 34
- Matthews, A. M., Condon, J. J., Cotton, W. D., & Mauch, T. 2021, *ApJ*, **909**, 193
- Mauch, T., Cotton, W. D., Condon, J. J., et al. 2020, *ApJ*, **888**, 61
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
- McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in *ASP Conf. Ser. 376, CASA Architecture and Applications*, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), **127**
- Middelberg, E. 2006, *PASA*, **23**, 64
- Mitchell, K. J., & Condon, J. J. 1985, *AJ*, **90**, 1957
- Momjian, E. 2017, VLA Expansion Project Memoranda #204
- Müller, A., Pfrommer, C., Ignesti, A., et al. 2021, *MNRAS*, **508**, 5326
- Murphy, E. J., Condon, J. J., Schinnerer, E., et al. 2011, *ApJ*, **737**, 67
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, **424**, 1614
- Owen, F. N., & Morrison, G. E. 2008, *AJ*, **136**, 1889
- Perley, R. A., & Butler, B. J. 2013, *ApJS*, **204**, 19
- Perley, R. A., Chandler, C. J., Butler, B. J., & Wrobel, J. M. 2011, *ApJL*, **739**, L1
- Sarbadhicary, S. K., Tremou, E., Stewart, A. J., et al. 2021, *ApJ*, **923**, 31
- Scheuer, P. A. G. 1957, *PCPS*, **53**, 764
- Schinnerer, E., Smolčić, V., Carilli, C. L., et al. 2007, *ApJS*, **172**, 46
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Seiffert, M., Fixsen, D. J., Kogut, A., et al. 2011, *ApJ*, **734**, 6
- Smolčić, V., Ciliegì, P., Jelić, V., et al. 2014, *MNRAS*, **443**, 2590
- Smolčić, V., Novak, M., Bondi, M., et al. 2017, *A&A*, **602**, A1
- Smolčić, V., Schinnerer, E., Finoguenov, A., et al. 2007, *ApJS*, **172**, 295
- Vernstrom, T., Scott, D., & Wall, J. V. 2011, *MNRAS*, **415**, 3641
- Williams, P. K. G., Allers, K. N., Biller, B. A., & Vos, J. 2019, *RNAAS*, **3**, 110
- Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, *ApJ*, **554**, 803
- Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. 2007, *ApJS*, **172**, 468