Article:

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Astronomy & Astrophysics

GOODS-ALMA 2.0: Source catalog, number counts, and prevailing compact sizes in 1.1 mm galaxies

C. Gómez-Guijarro[1], D. Elbaz[1], M. Xiao[1,2], M. Béthermin[3], M. Franco[4], B. Magnelli[1], E. Daddi[1],
M. Dickinson[5], R. Demarco[6], H. Inami[7], W. Rujopakarn[8,9,10], G. E. Magdis[11,12,13], X. Shu[14], R. Chary[15],
L. Zhou[2,16], D. M. Alexander[17], F. Bournaud[1], L. Ciesla[3], H. C. Ferguson[18], S. L. Finkelstein[19],
M. Giavalisco[20], D. Iono[21,22], S. Juneau[5], J. S. Kartaltepe[23], G. Lagache[3], E. Le Floc'h[1], R. Leiton[6], L. Lin[24],
K. Motohara[25], J. Mullaney[26], K. Okumura[1], M. Pannella[27,28], C. Papovich[29,30], A. Pope[20],
M. T. Sargent[31,32], J. D. Silverman[10], E. Treister[33], and T. Wang[2]

(Affiliations can be found after the references)

ABSTRACT

Submillimeter/millimeter observations of dusty star-forming galaxies with the Atacama Large Millimeter/submillimeter Array (ALMA) have shown that dust continuum emission generally occurs in compact regions smaller than the stellar distribution. However, it remains to be understood how systematic these findings are. Studies often lack homogeneity in the sample selection, target discontinuous areas with inhomogeneous sensitivities, and suffer from modest uv coverage coming from single array configurations. GOODS-ALMA is a 1.1 mm galaxy survey over a continuous area of 72.42 arcmin^2 at a homogeneous sensitivity. In this version 2.0, we present a new low resolution dataset and its combination with the previous high resolution dataset from the survey, improving the uv coverage and sensitivity reaching an average of $\sigma = 68.4\,\mu$Jy beam^{-1}. A total of 88 galaxies are detected in a blind search (compared to 35 in the high resolution dataset alone), 50% at $S/N^{\mathrm{peak}} \geq 5$ and 50% at $3.5 \leq S/N^{\mathrm{peak}} \leq 5$ aided by priors. Among them, 13 out of the 88 are optically dark or faint sources (H- or K-band dropouts). The sample dust continuum sizes at 1.1 mm are generally compact, with a median effective radius of $R_{\mathrm{e}} = 0\farcs10 \pm 0\farcs05$ (a physical size of $R_{\mathrm{e}} = 0.73 \pm 0.29$ kpc at the redshift of each source). Dust continuum sizes evolve with redshift and stellar mass resembling the trends of the stellar sizes measured at optical wavelengths, albeit a lower normalization compared to those of late-type galaxies. We conclude that for sources with flux densities $S_{1.1\,\mathrm{mm}} > 1$ mJy, compact dust continuum emission at 1.1 mm prevails, and sizes as extended as typical star-forming stellar disks are rare. The $S_{1.1\,\mathrm{mm}} < 1$ mJy sources appear slightly more extended at 1.1 mm, although they are still generally compact below the sizes of typical star-forming stellar disks.

Key words. galaxies: evolution – galaxies: high-redshift – galaxies: photometry – galaxies: star formation – galaxies: structure – submillimeter: galaxies

1. Introduction

Galaxies luminous in the infrared (IR) and submillimeter/millimeter (submm/mm) wavelengths are intense star-forming systems, some of which constitute the most powerful starbursts in the universe. They are the so-called dusty star-forming galaxies (DSFGs; see Casey et al. 2014, for a review). Characterized by star formation rates (SFRs) of hundreds and up to thousands of solar masses per year (e.g., Magnelli et al. 2012; Swinbank et al. 2014; Simpson et al. 2015a), their high dust content absorbs the intense ultraviolet (UV) emission from the burst of star formation and radiates it at far-IR and mm wavelengths. Their redshift distribution peaks at $z \sim 2-3$ (e.g., Chapman et al. 2005; Yun et al. 2012; Smolčić et al. 2012; Dudzevičiūtė et al. 2020) and they constitute a key galaxy population emitting half of the IR luminosity of the universe at $z \sim 2$ (e.g., Magnelli et al. 2011, see also Pérez-González et al. 2005; Magnelli et al. 2009). Furthermore, being already massive galaxies ($M_* > 10^{10.5}\,M_\odot$, e.g., Wardlow et al. 2011; Hainline et al. 2011; Michałowski et al. 2012; Pannella et al. 2015) and capable of assembling large amounts of stellar mass very quickly owed to their high SFRs, and they have been proposed as progenitors of quiescent galaxies at high redshift and eventually the most massive elliptical galaxies in the local universe (e.g., Cimatti et al. 2008; Ricciardelli

et al. 2010; Fu et al. 2013; Ivison et al. 2013; Toft et al. 2014; Gómez-Guijarro et al. 2018; Valentino et al. 2020).

From the first samples of mid-IR and far-IR bright galaxies uncovered by the IRAS satellite (e.g., Neugebauer et al. 1984; Rowan-Robinson et al. 1984; Elbaz et al. 1992) and the SCUBA bolometer at submm wavelengths (e.g., Smail et al. 1997; Hughes et al. 1998; Barger et al. 1998), the Atacama Large Millimeter/submillimeter array (ALMA) has recently opened a new era in the studies of DSFGs (see Hodge & da Cunha 2020, for a review). ALMA is capable of performing high-resolution and high-sensitivity observations. Its improved angular resolution enables secure galaxy identification of those otherwise affected by source confusion and blending single-dish observations. Its sensitivity reaching submilliJansky (sub-mJy) levels (e.g., Carniani et al. 2015; Fujimoto et al. 2016; González-López et al. 2017) allows for one to access a less extreme DSFG population. Even more normal star-forming galaxies, such as those located within the scatter of the correlation between the SFR and the stellar mass of SFGs (the so-called main sequence (MS) of star-forming galaxies (SFGs), e.g., Noeske et al. 2007; Elbaz et al. 2007; Daddi et al. 2007), are detected at submm/mm wavelengths with ALMA (e.g., Papovich et al. 2016; Schreiber et al. 2017a; Dunlop et al. 2017). Therefore, ALMA observations are bringing together the

previously existing gap between massive extreme starbursts and more normal MS-like SFGs, newly incorporated into the overall population of DSFGs as those galaxies luminous at IR and submm/mm wavelengths.

In the last years, thanks to the ALMA capabilities, a number of studies have uncovered that the submm/mm dust continuum emission occurs in compact areas smaller than the stellar sizes (e.g., Simpson et al. 2015a; Ikarashi et al. 2015; Hodge et al. 2016; Fujimoto et al. 2017; Gómez-Guijarro et al. 2018; Elbaz et al. 2018; Lang et al. 2019; Rujopakarn et al. 2019; Gullberg et al. 2019; Franco et al. 2020a). These findings are not only associated with the dust continuum, as other studies including CO lines (e.g., Puglisi et al. 2019) or radio emission (e.g., Jiménez-Andrade et al. 2019) have found more compact emission compared to the stellar sizes in these tracers as well. However, there are also examples of observations of more extended galaxy-wide dust continuum emission (e.g., Rujopakarn et al. 2016; Cheng et al. 2020; Sun et al. 2021; Cochrane et al. 2021) and simulations indicating that differential attenuation could play an important role in how observations compare the extent of the dust continuum emission to that of the stars (Popping et al. 2022). Therefore, it remains to be understood how systematic compactness is in DSFGs.

Many ALMA studies have been biased to follow-ups of galaxy samples, often targeting discontinuous areas and with inhomogeneous sensitivities. Recent ALMA blind surveys started to tackle these issues going from a deep pencil beam approach to larger areas at shallower depths (e.g., Walter et al. 2016; Dunlop et al. 2017; Franco et al. 2018; Hatsukade et al. 2018; Decarli et al. 2019; Zavala et al. 2021). However, another major roadblock in ALMA studies to date is the detection and measurement of accurate fluxes and sizes of sources spanning a wide range of intrinsic properties. While ALMA provides a tremendous advantage by resolving dust continuum sizes, single array configurations providing angular resolution sufficient to measure sizes of intrinsically compact sources could be missing more extended sources for which a coarser angular resolution would be better suited. Therefore, it has yet to be understood whether the current literature results have been affected by observational biases in order to answer to the question of how systematic compactness is in DSFGs.

In this work, we present GOODS-ALMA 2.0, an ALMA blind survey at 1.1 mm in the Great Observatories Origins Deep Survey South field (GOODS-South; Dickinson et al. 2003; Giavalisco et al. 2004). In this version 2.0, we present a new low resolution dataset and its combination with the previous high resolution dataset, GOODS-ALMA 1.0 (see Franco et al. 2018), improving the *uv* coverage and sensitivity. We aim at addressing the matters outlined above with the particularity of covering a large contiguous area using two array configurations at a similar and homogeneous depth over the whole field. The layout of the paper is as follows. An overview of the GOODS-ALMA survey, the observations and data processing involved in this work is in Sect. 2. In Sect. 3 we present the catalog of sources, including their flux and size measurements. Section 4 is dedicated to the study of the number counts, containing the relevant completeness and flux tests through simulations. We characterize source redshifts and stellar masses in Sect. 5, where we also report some new optically dark/faint sources. In Sect. 6 we discuss how systematic compactness is in DSFGs. We summarize the main findings and conclusions in Sect. 7.

Throughout this work we adopt a concordance cosmology $[\Omega_\Lambda, \Omega_M, h] = [0.7, 0.3, 0.7]$ and a Salpeter initial mass function

(IMF; Salpeter 1955). When magnitudes are quoted they are in the AB system (Oke 1974).

2. ALMA 1.1 mm galaxy survey in GOODS-South

2.1. Survey overview

GOODS-ALMA is a 1.1 mm galaxy survey carried out with ALMA Band 6 in GOODS-South. GOODS-ALMA 2.0 covers a continuous area of 72.42 arcmin² (primary beam response level ≥20%) centered at $\alpha = 3^h32^m30^s$, $\delta = -27°48'00$ at a homogeneous sensitivity with two different array configurations, to include both small and large spatial scales (see Fig. 1). Cycle 3 observations (program 2015.1.00543.S; PI: D. Elbaz) correspond to a more extended array configuration providing the high-resolution small spatial scales (hereafter, high resolution dataset and abbreviated as high-res in figures). Cycle 5 observations (program 2017.1.00755.S; PI: D. Elbaz) concern a more compact array configuration supplying the low-resolution large spatial scales (hereafter, low resolution dataset and abbreviated as low-res in figures). The high resolution dataset was presented in Franco et al. (2018). In this 2.0 version, we present the low resolution dataset and its combination with the high resolution dataset (hereafter, combined dataset). In Table 1 we summarize the angular resolution and sensitivity of the high resolution, low resolution, and combined datasets.

The survey area corresponds to a field of $\sim 10' \times 7'$ (15.1 Mpc × 10.5 Mpc comoving scale at $z = 2$). In order to cover this extension both the high and low resolution observations were designed as a 846-pointing mosaic separated by 0.8 times the antenna primary beam (half power beam width, HPBW $\sim 23''$), with each high and low resolution pointing centered at the exact same position. The pointings were grouped into six parallel and slightly overlapping $\sim 6.8' \times 1.5'$ slice submosaics, with a position angle of 70 deg and composed of 141 pointings each.

2.2. Observations and data processing

Low resolution dataset observations were carried out between 2018 July 17 and 2019 March 22. Each slice submosaic was completed in three execution blocks (EBs), 18 EBs in total. The number of antennas ranged 42−47, in configuration C43-2. The array configuration was not the most compact of the ALMA configurations due to a scheduling problem. The shortest and longest baselines were 15.0 m and 360.5 m. The mean precipitable water vapor (PWV) ranged 1.16−2.90 mm. The resulting total on source exposure time was 14.39 h (~ 2.40 h per slice submosaic). The correlator was set up in a single tuning centered at 265.0 GHz ($\lambda = 1.13$ mm) containing four spectral windows 1.875 GHz wide with 128 channels of 15.625 MHz each (17 km s⁻¹ at 265.0 GHz) in dual polarization, centered at 256.0, 258.0, 272.0, and 274.0 GHz covering a bandwidth of 7.5 GHz in the frequency range 255.0625−274.9375 GHz. These frequencies correspond to the highest frequencies of the band 6, optimal for a continuum survey. This tuning is the same as that of the high resolution dataset observations. The radio quasar J0519−4546 was observed as a bandpass and flux calibrator in 10 EBs, J0522−3627 in three, J0423−0120 in three, and J2357−5311 in two EBs. The radio quasar J0342−3007 was observed as a phase calibrator in 10 EBs, while J0348−2749 was used in seven, and J0336−2644 in one EB. High resolution dataset observations were presented in Franco et al. (2018). The shortest and longest baselines were 16.7 m and 1808.0 m and the
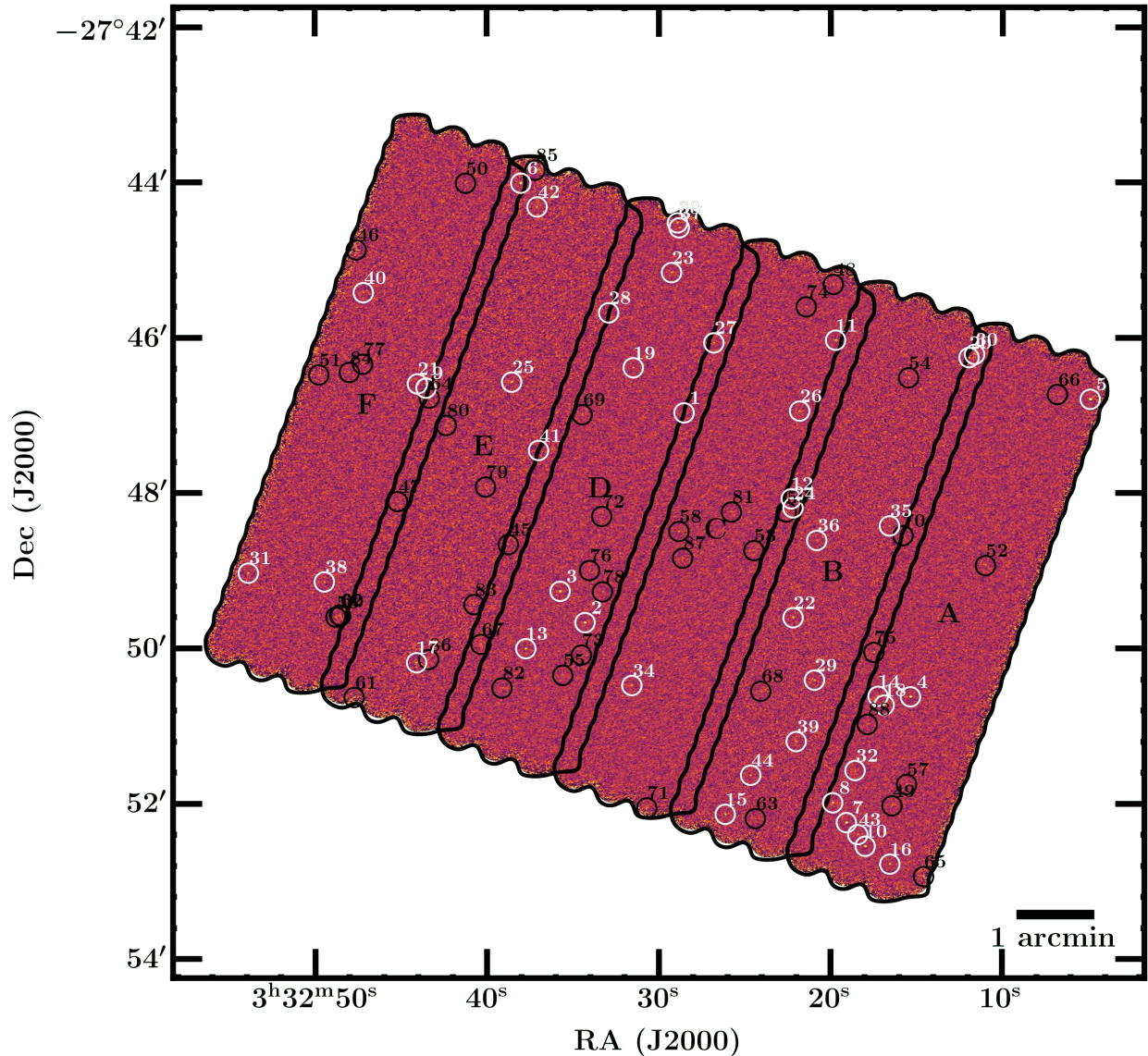
Fig. 1. GOODS-ALMA 2.0 map at 1.1 mm constructed by the combination of the high resolution and low resolution datasets (combined dataset). The sources detected with a purity of 100% as explained in Sect. 3.1.1 are marked with white circles and the sources detected using priors as described in Sect. 3.1.2 are marked with black circles. North is up, east is to the left. Cutouts of each source are shown in Appendix A.

resulting total on source exposure time was 14.06 h (~2.34 h per slice submosaic). For further details about the high resolution dataset observations, we refer the reader to Franco et al. (2018).

The Common Astronomy Software Applications (CASA; McMullin et al. 2007) version 5.6.1-8 was employed for data reduction using the scripts provided by the observatory. We processed both the high resolution and low resolution datasets in a similar manner to avoid introducing systematics. In the case of the high resolution dataset, it implies an independent data processing to the one already presented in Franco et al. (2018), leading to a 20% improved sensitivity when compared at the same angular resolution achieved by using the same natural weighting scheme. We inspected the visibilities and added additional flagging required besides those already included in the original scripts. We checked the flux calibration by verifying the calibrator flux density estimations. In order to reduce the computational time required for subsequent imaging we also averaged the calibrated visibilities in time over 120 s and in frequency over 8 channels.

Imaging was performed using the multifrequency synthesis algorithm collapsing all the frequency channels for continuum imaging. This is implemented within the CASA task TCLEAN, allowing for one to generate a dirty map. We decided to work with the dirty map instead of the clean map as: (1) the coverage in the uv plane is well sampled (see Fig. 2), which results in the absence of strong sidelobes; (2) the absence of very bright sources or a large dynamic range in flux densities; (3) the absence of very extended emission as the sources are marginally resolved, since the overall purpose is to measure global sizes. Imaging and deconvolution techniques suffer from specific issues related to the combination of datasets coming from multiple array configurations due the differences in the shapes of the synthesized dirty and clean beams, which requires to implement corrections (Czekala et al. 2021). Therefore, working directly on the dirty map is the best choice. This choice is also supported by the negligible difference (<1%) in the noise comparison between the dirty and clean maps. A natural weighting scheme was also chosen to get the best

Table 1. Summary of the data.

Slice	High resolution Date	Ant.	t target (s)	t total (s)	Beam (arcsec × arcsec)	σ (μJy beam⁻¹)	Low resolution Date	Ant.	t target (s)	t total (s)	Beam (arcsec × arcsec)	σ (μJy beam⁻¹)	Combined Beam (arcsec × arcsec)	σ (μJy beam⁻¹)
A	2016 August 17	42	46.52	72.12	0.″251 × 0.″199	89.6	2018 July 17	44	46.30	65.50	1.″38 × 1.″07	93.2	0.″503 × 0.″453	68.1
	2016 August 31	39	50.36	86.76			2019 January 19	47	48.03	68.18				
	2016 August 31	39	46.61	72.54			2019 January 19	46	48.03	67.45				
B	2016 September 01	38	46.87	72.08	0.″207 × 0.″185	88.8	2019 March 05	43	48.07	71.08	1.″22 × 1.″05	93.2	0.″366 × 0.″321	67.7
	2016 September 01	38	48.16	72.48			2019 March 05	43	48.05	68.15				
	2016 September 02	39	46.66	75.06			2019 March 12	42	48.08	70.10				
C	2016 August 16	37	46.54	73.94	0.″248 × 0.″234	89.1	2019 March 17	44	48.05	69.48	1.″48 × 0.″86	96.0	0.″419 × 0.″377	68.6
	2016 August 16	37	46.54	71.58			2019 March 17	47	48.05	70.25				
	2016 August 27	42	46.52	74.19			2019 March 19	47	48.03	67.05				
D	2016 August 16	37	46.54	71.69	0.″257 × 0.″233	88.8	2019 March 21	47	48.05	69.17	1.″29 × 0.″94	96.4	0.″460 × 0.″429	68.7
	2016 August 27	44	46.52	72.00			2019 March 21	45	48.07	69.98				
	2016 August 27	44	46.52	72.08			2019 March 22	44	48.03	69.28				
E	2016 August 01	39	46.54	71.84	0.″284 × 0.″260	88.7	2019 March 05	46	48.07	71.25	1.″41 × 0.″97	95.7	0.″452 × 0.″411	68.4
	2016 August 01	39	46.53	72.20			2019 March 07	46	48.05	69.62				
	2016 August 02	40	46.53	74.46			2019 March 14	42	48.03	70.95				
F	2016 August 02	40	46.53	72.04	0.″294 × 0.″256	88.8	2019 March 21	45	48.05	69.37	1.″28 × 0.″95	96.4	0.″510 × 0.″455	68.8
	2016 August 02	41	46.53	71.61			2019 March 22	44	48.02	69.57				
	2016 August 02	39	46.53	71.55			2019 March 22	47	48.07	67.62				
Mean					0.″251 × 0.″232	89.0					1.″33 × 0.″935	95.2	0.″447 × 0.″418	68.4

Notes. The columns show the slice ID, date, number of antennae, time on target, total time (time on target + calibration time), angular resolution, and sensitivity for the high resolution, low resolution, and combined datasets.
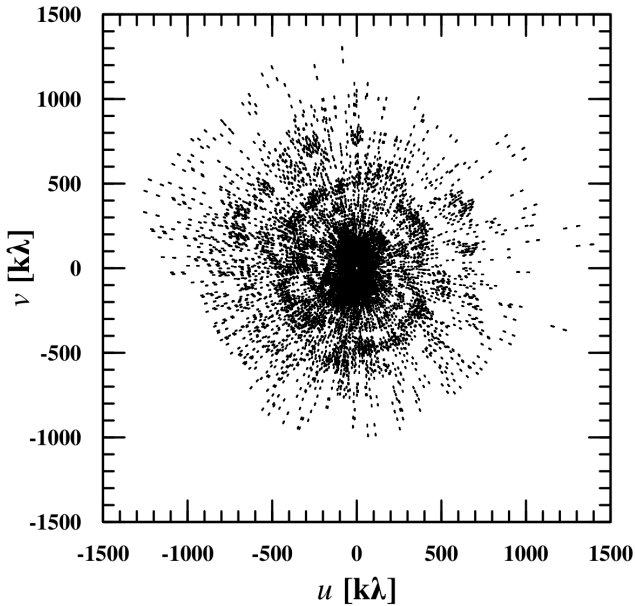

Fig. 2. Example of *uv* coverage in the combined dataset for one of the sources in the catalog.

point-source sensitivity, optimal for source detection. We combined each slice submosaic of the high resolution and low resolution datasets separately to produce both a high resolution and a low resolution dataset mosaic. Each of the high resolution and low resolution slice submosaics were also combined together to produce combined slice submosaics, which are also concatenated to form the combined dataset mosaic. All these combinations are always done following the original weight of each visibility.

In Table 1 we summarize the angular resolution and sensitivity of the high resolution, low resolution, and combined mosaics and slice submosaics. Since the observing conditions were slightly different during the observations of the different

slices, these submosaics have small differences in angular resolution and sensitivity. The point spread function (PSF) associated to each slice submosaic are shown in Fig. 3. There are slight differences in the PSF profile, but these differences do not introduce systematics in neither the source detection nor the flux density measurements, the two pieces of the data analysis for which imaging was used and the resulting synthesized beam plays a role (see Sects. 3.1.1 and 3.2, for details). In particular, for flux density measurements, the beam used was that corresponding to the specific location of the source for which the measurement was carried out and, thus, there are no systematics associated to the location of a source in the map. Therefore, we did not apply any specific tapering to homogenize the beam, since that would be at the expense of the survey depth compromising the optimal strategy for source detection, while measurements in the untapered map do not lead to systematics. On average, the high resolution and low resolution mosaics have similar sensitivities of 89.0 and 95.2 μJy beam⁻¹ at an angular resolution of 0.″251 × 0.″232 and 1.″33 × 0.″935, respectively (synthesized beam full width at half maximum (FWHM) along the major × minor axis). The combined mosaic reaches an average sensitivity of 68.4 μJy beam⁻¹ at an average angular resolution of 0.″447 × 0.″418.

2.3. Noise map

In addition to the high resolution, low resolution, and combined image maps, we built a noise map for each of these datasets by using a sliding box sigma clipping methodology. Every 2 × 2 pixels in the image map, we calculated the standard deviation in a 10″ × 10″ box (200 × 200 pixels for a 0.″05/pix scale). This box is large enough to converge to the noise level, but small enough to reflect the local noise variations. The pixel step is below the scale where the noise varies significantly (i.e., FWHM) and saves computation time. The pixels with values above 3σ from the median value are masked (clipped). This procedure is repeated three times. After these clipping iterations the standard deviation is computed one last time and assigned as the noise level of the 2 × 2 pixels. The box slides across the mosaic
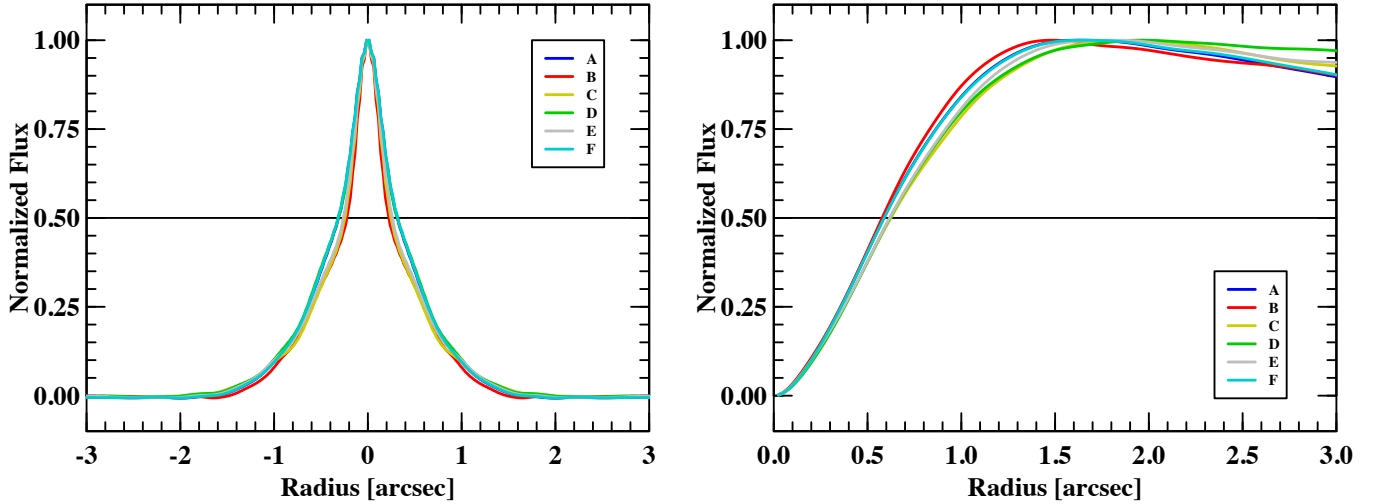
Fig. 3. *Left panel*: Azimuthal-average profile of the combined dataset PSFs associated to the six slice submosaics (normalized to the value of the central pixel). The solid black line marks the definition of the synthesized beam FWHM. *Right panel*: cumulative flux density enclosed in the PSFs (normalized to the maximum value with the solid black representing 50% of this value as reference.

until the entire map is covered. We note that these noise maps are built using the nonprimary beam corrected image maps. In Table 1 we summarize the noise levels of the high resolution, low resolution, and combined mosaics and slice submosaics.

3. Catalog

3.1. Source detection

In order to detect the sources we employed the Python Blob Detector and Source Finder (PyBDSF[1]), a tool designed to decompose radio interferometry images into sources. PyBDSF reads in the input image, calculates background rms and mean images, finds islands of emission, fits Gaussians to the islands, and groups the Gaussians into sources. A threshold for separating sources from noise is set, either using a constant hard threshold or a false detection rate algorithm. In the constant hard threshold method, the user defines a pixel threshold (σ_p), which corresponds to the signal-to-noise ratio (S/N) to identify an island of emission, and an island threshold (σ_f), which corresponds to the S/N that defines the boundaries of an island of emission. Islands of contiguous emission are identified by finding all the pixels greater that the pixel threshold. Starting from each of these pixels, all contiguous pixels (the surrounding eight pixels) higher than the island boundary threshold are identified as belonging to one island. Next, it fits multiple Gaussians to each island. If multiple Gaussians are fit and one of them is a bad solution then the number of Gaussians is decreased by one and the fit is done again, until all solutions in the island are good. After that, it groups nearby Gaussians within an island into sources. Once the Gaussians that belong to a source are identified, fluxes for the grouped Gaussians are summed to obtain the total flux of the source. The source position is set to be its centroid which, along with the source size, is determined from moment analysis.

3.1.1. Blind detection

In this work, we only used PyBDSF for source detection, but not for flux density or size measurements. We performed a blind

detection by running the `process_image` task with our own rms noise map built as explained in Sect. 2.3, overriding the internal background rms calculated by PyBDSF to have better control of the detection process. We used the hard threshold method with a grid of σ_p = 3.0−6.0 and σ_f = 2.0−4.0 in steps of 0.05, as opposed to the false detection rate algorithm to also have control of the false detection probability. Since the sources are only marginally extended and substructure is beyond the scope of the survey we activated the option of grouping by island (`group_by_isl` = TRUE), which assigns a single source per detected island. Besides, we included the sources for which the Gaussian fit for flux measurements failed (`incl_empty` = TRUE). These sources do not have a valid PyBDSF flux density measurement, but we do not use PyBDSF for flux density measurements, only for source detection. As shown in Sect. 2.2 the synthesized beam varies slightly over the image map. We tested whether the observed beam variation influences the detection method by running PyBDSF modifying the beam FWHM in the header of the image map within the FWHM range across the different slice submosaics (see Table 1). We found that the detection method does not depend on this beam variation. Therefore, we proceeded with a single blind detection carried out with a common average beam across the field.

In the left panel of Fig. 4 we show the pixel distribution of the combined dataset map. The right-hand side tail reflects the excess created by real sources. The sources detected in the image map are positive sources. However, the noise is Gaussian and negative detections also exist. The difference between positive and negative sources is a proxy for the number of real sources, being the purity defined as:

$$p = \frac{N_{\text{pos}} - N_{\text{neg}}}{N_{\text{pos}}}. \tag{1}$$

We inverted the image map by multiplying it by −1 and repeated the detection procedure for the inverse image map. In the right panel of Fig. 4 we plot the number of positive and negative detections in the combined dataset map as a function of σ_p for a fixed σ_f = 2.7. Given the chosen detection options in the `process_image` task explained above, the number of detections does not depend on σ_f. For consistency within the GOODS-ALMA survey we chose a value σ_f = 2.7, equivalent to the

---

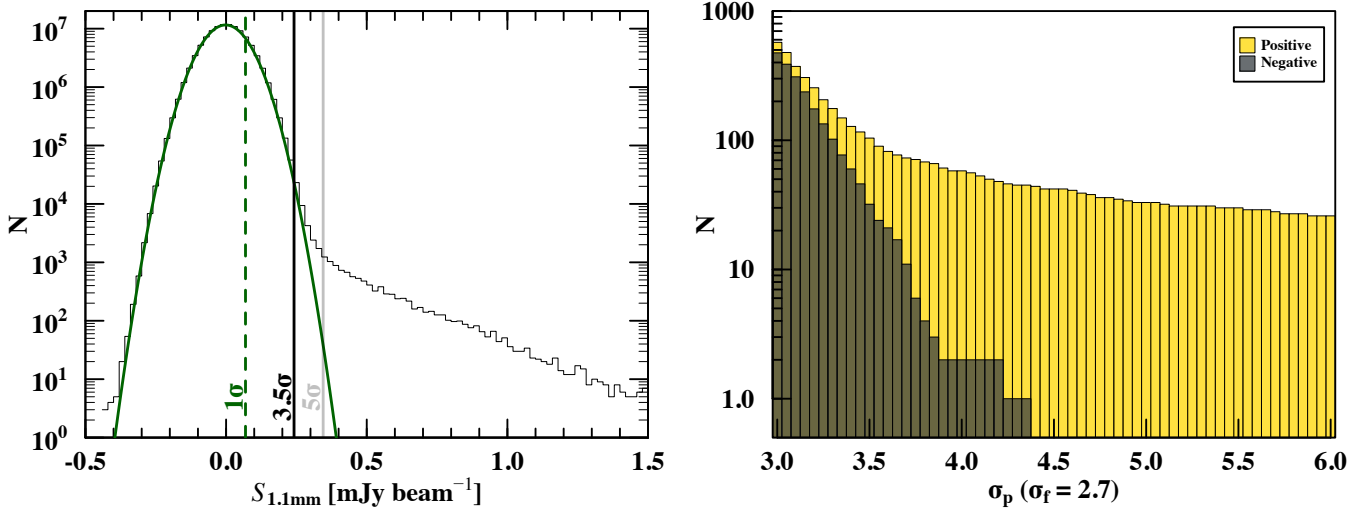[1] https://www.astron.nl/citt/pybdsf/

Fig. 4. *Left panel*: pixel distribution in the combined dataset map. The solid green curve shows the result of a Gaussian fit and indicates the noise level ($\sigma = 68.4\,\mu\text{Jy beam}^{-1}$). *Right panel*: number of positive (yellow histogram) and negative (black histogram) detections as a function of the pixel threshold σ_p for a fixed island threshold $\sigma_\text{f} = 2.7$ in the combined dataset map. We note that the number of detections is a differential value and not cumulative with decreasing σ_p.

floodclip threshold $\sigma_\text{f} = 2.7$ used with the tool BLOBCAT (Hales et al. 2012) in Franco et al. (2018).

A purity of $p = 1$ (sources detected with a purity of 100% associated to the absence of negative detections) is found for $\sigma_\text{p} \geq 4.4$ in the combined dataset map. This value is $\sigma_\text{p} \geq 5.2$ in the high resolution map and $\sigma_\text{p} \geq 4.2$ in the low resolution map. The number of sources with a purity of 100% is 44 in the combined map. Independently detected, this number is 8 in the high resolution map and 38 in the low resolution map. In Table 2 we present the main catalog of 100% pure sources detected in the combined dataset, labeling also which ones appear independently detected in the high resolution or low resolution datasets as 100% pure sources. We note that in Table 2 we include the detection S/N^peak for each source, defined as the PyBDSF peak flux density divided by the average background rms noise of the island. This detection S/N^peak is not strictly the same as the pixel threshold σ_p. The former is usually higher by construction of the PyBDSF detection methodology. While the PyBDSF peak flux density is calculated by the code performing Gaussian fitting and then divided by the average background rms noise of the island to obtain the detection S/N^peak, the pixel threshold σ_p is the number of sigma above the mean for an individual pixel and, thus, it involves the read of the image and noise maps at a given pixel.

3.1.2. Prior-based selection

At $\sigma_\text{p} = 3.0$, the blind detection results in 573 positive and 475 negative detections in the combined dataset map. A proxy for the expected number of real sources is $N_\text{real} = N_\text{pos} - N_\text{neg} = 98 \pm 32$ (where the uncertainty is calculated from Poisson statistics: $\Delta N_\text{real} = \sqrt{(\Delta N_\text{pos})^2 + (\Delta N_\text{neg})^2} = \sqrt{(\sqrt{N_\text{pos}})^2 + (\sqrt{N_\text{neg}})^2}$. This number is much larger than those in the 100% pure main catalog. Therefore, we created a prior-based supplementary catalog following the methodology presented in Franco et al. (2020b) using *Spitzer*/IRAC and the *Karl G. Jansky* Very Large Array (VLA) prior positions.

Spitzer/IRAC 3.6 and 4.5 μm observations in the GOODS-South field come from the IRAC Ultradeep Field (IUDF; Labbé

et al. 2015), which combines all ultradeep data ever taken with IRAC at 3.6 and 4.5 μm over GOODS-South and the HUDF. The deepest observations come from the IRAC Ultra Deep Field (IUDF; PI: I. Labbé) and IRAC Legacy over GOODS (IGOODS; PI: P. Oesch) programs, combined with archival data from GOODS (PI: M. Dickinson), SEDS (Ashby et al. 2013, PI: G. Fazio), S-CANDELS (Ashby et al. 2015, PI: G. Fazio), ERS (PI: G. Fazio), and UDF2 (PI: R. Bouwens). The combined IRAC images reach a 5σ point source sensitivity of 26.7 and 26.5 AB mag at 3.6 and 4.5 μm, respectively. VLA observations in GOODS-South (PI: W. Rujopakarn) were taken at 3 GHz (10 cm), covering the entire GOODS-ALMA field reaching a rms noise of $\sim 2.1\,\mu$Jy at an angular resolution $\sim 0\rlap{.}''3$ (Rujopakarn et al., in prep.), and at 6 GHz (5 cm), around the HUDF with partial coverage of GOODS-ALMA reaching a rms noise of $\sim 0.32\,\mu$Jy and an angular resolution of $0\rlap{.}''31 \times 0\rlap{.}''61$ (Rujopakarn et al. 2016).

First, the purpose is to get the distance within which a given ALMA source has a secure IRAC/VLA counterpart. We searched for IRAC/VLA counterparts in the 100% pure main catalog. In the case of IRAC we used the last publicly available catalog in GOODS-South by Ashby et al. (2015), which includes all the IRAC datasets listed above except for IGOODS, enough for the purpose of getting the distance within which a given ALMA source has a secure IRAC counterpart. For VLA, we employed the 3 GHz (10 cm) image catalog (Rujopakarn et al., in prep.). We note that in order to make the counterpart assignation we corrected the coordinates in the ancillary catalogs and images for the GOODS-South astrometry offsets reported in the literature when comparing with ALMA coordinates, except for the case of the VLA catalogs and images which do not suffer from such offsets. Franco et al. (2020b) reported a global astrometry offset of ΔRA $= -96$ mas, ΔDec $= 252$ mas, but also a non negligible local offset caused by distortions in the original HST/ACS and WFC3 GOODS-South mosaics that can reach $\sim 0\rlap{.}''15$ in the edges of the GOODS-ALMA field. We applied both in the present analysis.

Using the 100% pure main catalog we find that 44/44 have an IRAC counterpart. Among them, 35/44 have an IRAC counterpart located at $\leq 0\rlap{.}''4$. However, we visually inspected the images

Table 2. Source catalog: 100% pure.

ID	α (J2000) (deg)	δ (J2000) (deg)	ID_{ZF}	z	$\log(M_*/M_\odot)$	S/N^{peak}	$S_{1.1\,mm}$ (mJy)	θ (arcsec)	High-res	Low-res	Other ID
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A2GS1	53.118813	−27.782884	17856	2.309 [a]	11.06	18.78	2.15 ± 0.11	0.18 ± 0.02	Yes/100pur	Yes/100pur	AGS1,AS2,SG5,GS6
A2GS2	53.142840	−27.827895	12333	3.556 [b]	11.09	18.65	2.25 ± 0.12	0.20 ± 0.02	Yes/100pur	Yes/100pur	AGS4,SG12
A2GS3	53.148870	−27.821185	13086	2.582 [c]	11.44	18.11	2.11 ± 0.11	0.17 ± 0.02	Yes/100pur	Yes/100pur	AGS3,AS3,SG9,GS5
A2GS4	53.063887	−27.843804	10316	2.918 [d]	10.76	17.19	2.06 ± 0.10	0.17 ± 0.01	Yes/100pur	Yes/100pur	AGS2,SG3
A2GS5	53.020352	−27.779931	18282	1.797 [d]	11.09	15.80	2.75 ± 0.09	0.35 ± 0.03	Yes/100pur	Yes/100pur	AGS8,SG4,LESS18
A2GS6	53.158397	−27.733588	23898	3.46	11.14	14.55	2.15 ± 0.12	0.19 ± 0.02	Yes/100pur	Yes/100pur	AGS5,SG7
A2GS7	53.079416	−27.870820	6964	3.467 [b]	10.52	14.21	2.05 ± 0.12	0.48 ± 0.04	Yes/100pur	Yes/100pur	AGS17,SG11,LESS10
A2GS8	53.082752	−27.866560	7867	3.29	11.44	12.07	1.40 ± 0.12	0.13 ± 0.02	Yes/Prior	Yes/100pur	AGS7,SG16
A2GS9	53.181383	−27.777572	18645	2.696 [e]	10.90	11.12	1.41 ± 0.13	0.29 ± 0.05	Yes/Prior	Yes/100pur	AGS18,AS4,UDF2,C06,SG25
A2GS10	53.074868	−27.875889	6755	3.47	10.56	10.91	1.67 ± 0.10	0.11 ± 0.02	Yes/Prior	Yes/100pur	AGS15,SG13,LESS34
A2GS11	53.082069	−27.767279	19833	2.41	11.33	10.39	1.34 ± 0.11	0.21 ± 0.03	Yes/Prior	Yes/100pur	AGS10,SG10
A2GS12	53.092818	−27.801328	15639	3.847 [f]	10.87	9.66	1.39 ± 0.10	0.21 ± 0.03	Yes/Prior	Yes/100pur	AGS9,SG18
A2GS13	53.157199	−27.833492	11442	1.619 [g]	10.96	9.30	1.36 ± 0.12	0.44 ± 0.08	Yes/Prior	Yes/100pur	AGS26,SG23,GS7
A2GS14	53.071752	−27.843698	10241	1.956 [c]	11.19	8.30	0.83 ± 0.10	0.31 ± 0.06	Yes/Prior	Yes/100pur	AGS37,SG30,GS4
A2GS15	53.108810	−27.869037	7589	3.47	10.24	8.08	1.24 ± 0.10	<0.13	Yes/Prior	Yes/100pur	AGS11,SG19
A2GS16	53.068881	−27.879724	6153	2.45	11.39	8.03	1.03 ± 0.11	0.16 ± 0.03	Yes/Prior	Yes/100pur	AGS31,SG43
A2GS17	53.183697	−27.836500	11353	4.64	10.39	7.96	1.23 ± 0.12	0.25 ± 0.05	Yes/Prior	Yes/100pur	AGS25,SG58
A2GS18	53.070260	−27.845595	10152	3.689 [h]	10.56	7.73	0.83 ± 0.09	0.14 ± 0.03	Yes/Prior	Yes/100pur	AGS21,SG26
A2GS19	53.131138	−27.773202	19033	2.225 [i]	11.43	7.56	0.72 ± 0.11	<0.13	Yes/Prior	Yes/100pur	AGS13,AS8,SG40
A2GS20	53.049739	−27.770990	19463	2.68	10.72	7.49	0.93 ± 0.10	<0.13	No	Yes/100pur	AGS33,SG32
A2GS21	53.183469	−27.776661	18658	2.698 [e]	10.30	7.27	1.49 ± 0.13	<0.14	Yes/100pur	Yes/100pur	AGS6,AS1,UDF1,SG22
A2GS22	53.092372	−27.826850	12416	2.72	10.77	6.87	1.07 ± 0.09	0.17 ± 0.03	Yes/Prior	Yes/100pur	AGS20,SG14
A2GS23	53.121858	−27.752778	21730	2.00	10.76	6.68	0.73 ± 0.14	<0.14	No	Yes/100pur	AS11
A2GS24	53.092391	−27.803269	15342	2.32	11.38	6.66	0.85 ± 0.11	0.28 ± 0.06	No	Yes/100pur	SG38
A2GS25	53.160620	−27.776287	18701	2.543 [j]	10.13	6.65	0.99 ± 0.12	0.17 ± 0.05	Yes/Prior	Yes/100pur	AGS12,AS5,UDF3,C01,SG48
A2GS26	53.090782	−27.782492	17976	1.927 [d]	10.64	6.55	0.76 ± 0.09	<0.14	No	Yes/100pur	SG34
A2GS27	53.111595	−27.767864	19964	4.72	11.01	6.49	0.64 ± 0.12	0.22 ± 0.06	No	Yes/Prior	AGS32
A2GS28	53.137093	−27.761411	(…)	1.967 [d]	10.71	6.49	0.92 ± 0.10	0.17 ± 0.05	No	Yes/100pur	AS7,SG29
A2GS29	53.087184	−27.840242	(…)	3.47	11.32	6.41	0.85 ± 0.11	<0.15	Yes/Prior	Yes/100pur	AGS24,SG44
A2GS30	53.048384	−27.770312	(…)	3.80	11.46	6.31	0.71 ± 0.11	1.66 ± 0.31	No	Yes/100pur	SG55
A2GS31	53.224499	−27.817250	13388	2.15	11.18	6.21	0.79 ± 0.11	0.28 ± 0.08	Yes/Prior	Yes/Prior	AGS28
A2GS32	53.077331	−27.859632	8449	2.251 [d]	11.55	6.02	0.51 ± 0.10	<0.15	No	Yes/Prior	SG31
A2GS33	53.120402	−27.742111	(…)	(…)	(…)	5.91	0.95 ± 0.12	0.95 ± 0.30	No	No	AS20,SG68
A2GS34	53.131474	−27.841396	10345	1.613 [g]	11.43	5.86	0.74 ± 0.10	0.41 ± 0.11	No	Yes/100pur	SG65
A2GS35	53.069006	−27.807141	14926	4.73	10.94	5.72	0.72 ± 0.11	0.60 ± 0.16	Yes/Prior	Yes/100pur	AGS27
A2GS36	53.086635	−27.810257	14543	2.36	11.27	5.69	0.49 ± 0.10	0.17 ± 0.05	Yes/Prior	Yes/Prior	AGS23,SG36
A2GS37	53.119994	−27.743167	22905	3.85	11.16	5.66	0.71 ± 0.12	0.66 ± 0.20	No	Yes/100pur	SG72
A2GS38	53.206064	−27.819142	(…)	(…)	(…)	5.55	0.70 ± 0.13	(…)	No	Yes/100pur	AS17
A2GS39	53.091617	−27.853421	9248	2.36	10.61	5.44	0.81 ± 0.10	0.33 ± 0.10	No	Yes/Prior	AGS39,SG42
A2GS40	53.196569	−27.757065	21234	2.46	10.21	5.44	0.62 ± 0.12	<0.16	No	Yes/Prior	AS9
A2GS41	53.154105	−27.790947	16952	1.759 [d]	10.54	5.30	0.79 ± 0.12	0.30 ± 0.10	No	Yes/100pur	AS12,UDF5,C02
A2GS42	53.154440	−27.738686	23441	2.29	11.12	5.25	0.80 ± 0.11	0.42 ± 0.13	No	Yes/100pur	
A2GS43	53.076655	−27.873394	6921	3.54	10.54	5.12	0.70 ± 0.10	0.87 ± 0.21	No	Yes/100pur	
A2GS44	53.102654	−27.860660	8455	4.19	11.05	5.00	0.76 ± 0.10	0.42 ± 0.13	No	Yes/100pur	

Notes. (1) ALMA source ID; (2) right ascension (J2000) in degrees of the ALMA source as detected by PyBDSF; (3) declination (J2000) in degrees of the ALMA source as detected by PyBDSF; (4) ID associated to the stellar counterpart in the ZFOURGE catalog (5) redshift with spectroscopic redshifts shown with three decimal digits; (6) stellar mass; (7) detection signal to noise ratio from PyBDSF, measured as the peak flux density over the local rms noise; (8) 1.1 mm flux density measured using a 1″6 diameter aperture corrected from aperture losses and flux boosting; (9) source size given by the FWHM of a circular Gaussian model fit in the uv plane; (10) and (11) source presence in the high/low resolution datasets (Yes or No) and whether it was detected as a 100% pure source or using priors (100pur or Prior) in those datasets; (12) source ID in other surveys: AGS (GOODS-ALMA; Franco et al. 2018, 2020b); AS (ASAGAO; Hatsukade et al. 2018), UDF (Dunlop et al. 2017); C (ASPECS; Aravena et al. 2020); SG (SUPER GOODS; Cowie et al. 2018); LESS (Hodge et al. 2013); GS (Elbaz et al. 2018). Spectroscopic redshift references: [a]Kurk et al. (2013); [b]Zhou et al. (2020); [c]Wuyts et al. (2009); [d]Momcheva et al. (2016); [e]Decarli et al. (2019); [f]B. Mobasher, priv. comm.; [g]Vanzella et al. (2008); [h]Garilli et al. (2021); [i]Kriek et al. (2008); [j]Inami et al. (2017).

and the remaining 9/44 correspond to blends of multiple sources. Once we corrected their coordinates accounting for blending by fitting a point source model using GALFIT (Peng et al. 2002) (where the number of priors is set to the number of sources in the $F160W$-band image), all the 44 ALMA sources in the 100% pure main catalog have an IRAC counterpart located at ≤0″4 (see Fig. 5, panel A). Similarly, 35 ALMA sources in the 100% pure main catalog have a VLA counterpart. Among them, all 35/35 have a VLA counterpart located at ≤0″4 (see Fig. 5, panel B). Therefore, ≤0″4 is the robust counterpart search radius to look for further prior-based ALMA sources with either IRAC or VLA counterpart in the σ_p = 3.0 blind detection. We calculated the probability of random association between a potential ALMA source and an IRAC/VLA counterpart. In order to do this, we selected 100 000 random positions in the GOODS-ALMA field and checked for counterparts in the IRAC/VLA catalogs at a distance ≤0″4. The probability of randomly finding an IRAC counterpart located at ≤0″4 of a given ALMA source is 0.5% and 0.05% in the case of VLA. The number of negative ALMA sources with an IRAC counterpart is consistent with this estimation as there is one negative ALMA source with an IRAC counterpart (the probability calculation yields 475 · 0.5% = 2.4). The number of negative ALMA sources with a VLA counterpart is zero (the probability calculation yields 475 · 0.05% = 0.2).

Second, we validated the ≤0″4 counterpart radius by checking the stellar masses of the counterparts. We checked whether
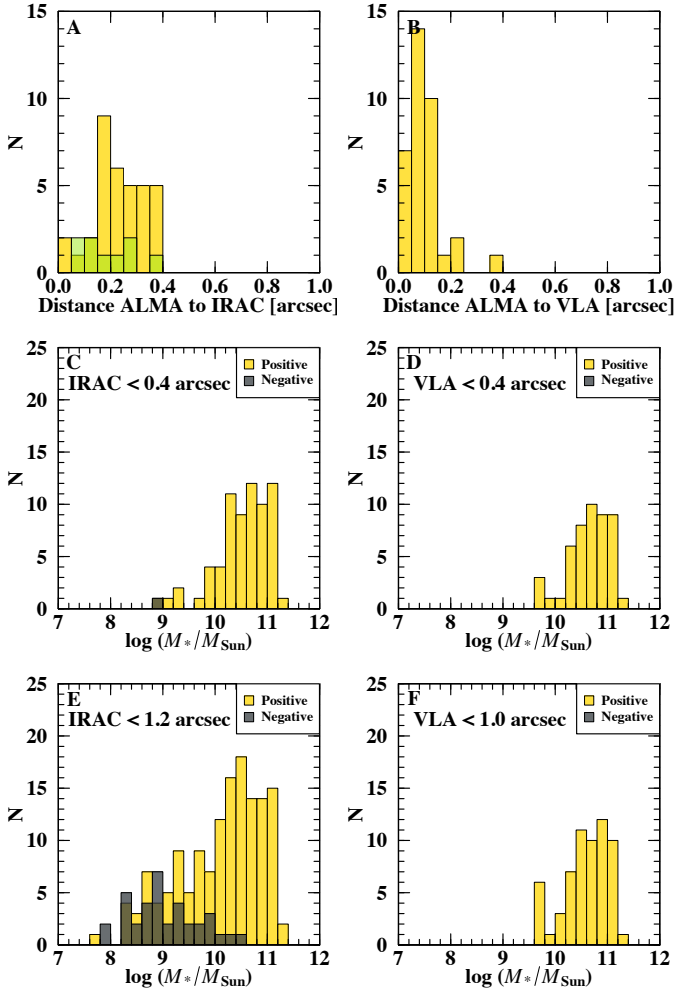
Fig. 5. *Panels A and B*: distance from the ALMA sources in the 100% pure main catalog to the closest IRAC counterpart in Ashby et al. (2015) (*panel A*) and VLA counterpart at 3 GHz (Rujopakarn et al., in prep.) (*panel B*). *Panel A*: sources whose coordinates were corrected accounting for blending in IRAC are highlighted in green (see main text). *Panels C, D, E, and F*: number of positive (yellow histogram) and negative (black histogram) detections in the $\sigma_p = 3.0$ blind detection in the combined dataset map with an IRAC counterpart at $\leq 0.''4$ (*panel C*) and $\leq 1.''2$ (*panel E*) or with a VLA counterpart at $\leq 0.''4$ (*panel D*) and $\leq 1.''0$ (*panel F*) as a function of stellar mass.

there are negative detections with a massive counterpart at $\leq 0.''4$, providing the number of expected spurious detections. The FourStar galaxy evolution survey (ZFOURGE; PI. I. Labbé) provides the deepest K_s-band image with a total 5σ sensitivity of up to 26.5 AB mag, after combining their own survey image with all the other K_s-band images available in GOODS-South. We looked for counterparts in the K_s-band selected ZFOURGE catalog by Straatman et al. (2016). There are no negative detections at $\log(M_*/M_\odot) \geq 9.0$ (see Fig. 5, panels C and D) for IRAC/VLA $\leq 0.''4$ counterparts in the $\sigma_p = 3.0$ blind detection. Therefore, we selected all the $\sigma_p = 3.0$ detections with an IRAC or VLA counterpart at $\leq 0.''4$ with stellar masses $\log(M_*/M_\odot) \geq 9.0$. In Table 3 we present the supplementary catalog of prior-based selected sources detected in the combined dataset, adding another 44 sources. We include the detection S/N^{peak} for each source as defined in Sect. 3.1.1. We also label which ones appear independently in the high resolution or low resolution datasets at $\sigma_p = 3.0$.

Finally, we checked the limits of the prior methodology. For IRAC, at $\leq 1.''2$ and $\log(M_*/M_\odot) \geq 10.0$ there is still an excess of positive sources with a massive counterpart associated compared to negative detections (see Fig. 5, panel E), expecting around three to be spurious. For VLA, there are still no negative detections found at $\log(M_*/M_\odot) \geq 9.0$ for counterparts at $\leq 1.''0$ (see Fig. 5, panel F). We report these extra 16 sources as uncertain sources (i.e., sources with either an IRAC counterpart at $\leq 1.''2$ with stellar masses $\log(M_*/M_\odot) \geq 10.0$ or a VLA counterpart at $\leq 1.''0$ with stellar masses $\log(M_*/M_\odot) \geq 9.0$, see Table B.1), but we did not employ them for further analysis.

As a sanity check, we visually inspected the IRAC 3.6 and 4.5 μm ultradeep images and VLA 3 GHz maps at the positions of ALMA source candidates and tagged "real" counterparts by our own personal judgement. We created an alternative prior-based catalog leading to the exact same result compared to the analysis using the catalog counterparts and the stellar mass validation criteria.

In comparison with the high resolution dataset analysis presented in Franco et al. (2018, 2020b) there are three sources (namely AGS32, AGS33, and AGS39) reported in Franco et al. (2020b) using the prior methodology that do not appear in the high resolution dataset in this work, although they are found in the low resolution dataset (A2GS20, A2GS27, and A2GS39, respectively). We note that Franco et al. (2018, 2020b) detection in the high resolution dataset is slightly different compared to the high resolution dataset detection here, since the former was carried out in a tapered map with a homogeneous and circular synthesized beam of $0.''6$ FWHM and, besides, with a different detection tool (BLOBCAT). Data tapering is beneficial to optimize the sensitivity to sources that are larger than the angular resolution. Therefore, some differences are expected between the sources detected in the tapered high resolution dataset used in Franco et al. (2018, 2020b) and the ones detected in the untapered high resolution dataset employed here. In addition, Franco et al. (2020b) reported three sources that are not part of the 100% pure main catalog or the prior-based supplementary catalog presented in this work (namely AGS34, AGS35, and AGS36), being among those with the lowest S/N ($S/N = 3.72$, 3.71, and 3.66) in Franco et al. (2020b). Similarly two sources appear in the high resolution dataset here, but do not in the high resolution dataset analysis in Franco et al. (2018, 2020b) (namely A2GS58 and A2GS71). This points to the differences in the detection tool PyBDSF versus BLOBCAT introducing some differences in the sources detected at the lowest S/N regimes.

3.2. Flux and size measurements

GOODS-ALMA 2.0 was designed to retrieve the spatial information for both the small and large spatial scales. Compact array configurations, sensitive to large spatial scales with large maximum recoverable scales, are suitable to get total flux measurements with minimum flux losses. Extended array configurations yield information on the small spatial scales. Together, the information on multiple spatial scales makes it possible to measure a large range of intrinsic source sizes, while retrieving accurately the total fluxes.

A variety of techniques are used for flux density measurements in the literature, including aperture photometry or 2D functional fitting in the image plane, peak flux measurements in the case of unresolved sources also in the image plane, or functional fitting in the *uv* plane. In this work, we explored flux densities measurements using these techniques. Eventually, we report the values obtained from aperture photometry. This is the

Table 3. Source catalog: Prior-based.

ID	α (J2000) (deg)	δ (J2000) (deg)	ID_{ZF}	z	$\log(M_*/M_\odot)$	S/N^{peak}	$S_{1.1\,mm}$ (mJy)	High-res	Low-res	Other name
A2GS45	53.161435	−27.811158	14419	2.77	11.06	5.27	0.53 ± 0.13	No	Yes/Prior	AS29
A2GS46	53.198290	−27.747905	22177	1.910 [a]	10.96	4.92	0.65 ± 0.12	No	No	AS6,SG20
A2GS47	53.188278	−27.801928	15703	3.83	10.03	4.71	0.37 ± 0.11	No	No	
A2GS48	53.082624	−27.755313	21397	2.926 [b]	10.07	4.71	0.48 ± 0.09	No	No	
A2GS49	53.068362	−27.867197	7676	1.973 [c]	10.31	4.65	0.46 ± 0.10	No	No	
A2GS50	53.171785	−27.733608	24110	2.89	10.31	4.64	0.42 ± 0.13	No	Yes/Prior	
A2GS51	53.207365	−27.774726	18890	2.36	10.83	4.49	0.59 ± 0.11	No	Yes/Prior	
A2GS52	53.045714	−27.815630	13780	0.88	9.54	4.48	0.43 ± 0.12	No	No	
A2GS53	53.101870	−27.812437	14215	2.024 [d]	10.53	4.44	0.40 ± 0.11	No	Yes/Prior	
A2GS54	53.064376	−27.775348	18738	0.735 [e]	10.06	4.44	0.55 ± 0.10	No	Yes/Prior	
A2GS55	53.148238	−27.839232	10844	1.545 [f]	10.76	4.44	0.41 ± 0.11	No	No	
A2GS56	53.203243	−27.826719	12427	2.78	11.04	4.41	0.62 ± 0.12	No	Yes/Prior	
A2GS57	53.064807	−27.862613	8323	4.64	11.15	4.40	0.54 ± 0.10	No	Yes/Prior	SG52
A2GS58	53.120073	−27.808327	14700	1.83	11.03	4.33	0.55 ± 0.12	Yes/Prior	Yes/Prior	AS44,SG63
A2GS59	53.094033	−27.804167	15251	2.475 [d]	11.40	4.30	0.42 ± 0.11	No	No	SG59
A2GS60	53.202104	−27.826442	(...)	1.120 [g]	10.53	4.28	0.45 ± 0.13	No	No	
A2GS61	53.198834	−27.843955	10096	1.615 [g]	11.40	4.28	1.01 ± 0.12	No	No	SG53
A2GS62	53.202342	−27.826284	12438	1.120 [a]	10.82	4.25	0.48 ± 0.12	Yes/Prior	No	AGS29,AS18
A2GS63	53.101527	−27.869956	7453	3.19	10.54	4.21	0.34 ± 0.11	No	No	
A2GS64	53.180574	−27.779729	18336	2.67	10.67	4.19	0.53 ± 0.12	No	No	AS14,UDF7,C11
A2GS65	53.060693	−27.882386	5860	2.22	11.15	4.16	0.98 ± 0.10	No	Yes/Prior	GS1
A2GS66	53.028301	−27.778904	18460	1.686 [f]	10.81	4.16	0.35 ± 0.11	No	No	
A2GS67	53.168078	−27.832547	11581	0.650 [h]	10.30	4.15	0.60 ± 0.11	No	Yes/Prior	AGS30
A2GS68	53.100219	−27.842636	10418	1.413 [a]	10.10	4.10	0.43 ± 0.09	No	Yes/Prior	
A2GS69	53.143518	−27.783274	17733	1.414 [i]	10.89	4.09	0.55 ± 0.12	No	Yes/Prior	AS13,UDF6,C13
A2GS70	53.065710	−27.809225	14760	2.61	9.61	4.06	0.46 ± 0.12	No	No	
A2GS71	53.127853	−27.867664	7653	3.026 [c]	10.26	4.05	0.67 ± 0.10	Yes/Prior	No	
A2GS72	53.138771	−27.805110	15233	2.28	9.37	4.04	0.25 ± 0.12	No	No	
A2GS73	53.143706	−27.834860	11381	1.987 [d]	10.15	4.03	0.40 ± 0.12	No	Yes/Prior	
A2GS74	53.089200	−27.760137	20735	1.61	11.31	3.97	0.54 ± 0.11	No	Yes/Prior	
A2GS75	53.072754	−27.834274	11449	1.618 [g]	11.25	3.96	0.69 ± 0.11	No	Yes/100pur	SG33
A2GS76	53.141751	−27.816725	13714	2.53	11.24	3.94	0.39 ± 0.12	No	No	AS31
A2GS77	53.196745	−27.772431	19313	2.805 [j]	10.55	3.90	0.40 ± 0.11	No	No	
A2GS78	53.138604	−27.821299	13269	3.65	10.15	3.89	0.38 ± 0.11	No	No	
A2GS79	53.166960	−27.798800	15702	1.998 [a]	10.94	3.84	0.44 ± 0.12	No	No	AS15,UDF11,C10
A2GS80	53.176495	−27.785575	17465	1.314 [a]	11.06	3.82	0.52 ± 0.12	No	Yes/Prior	AGS38,UDF16,C15
A2GS81	53.107330	−27.804163	15305	2.66	10.63	3.82	0.47 ± 0.11	No	Yes/Prior	
A2GS82	53.162978	−27.841940	10656	4.38	10.61	3.79	0.40 ± 0.12	No	No	
A2GS83	53.169775	−27.824041	12763	2.130 [c]	11.00	3.78	0.50 ± 0.11	No	Yes/Prior	
A2GS84	53.199969	−27.774185	19133	4.36	10.74	3.71	0.47 ± 0.11	No	No	
A2GS85	53.154933	−27.730797	24422	2.72	10.77	3.68	0.65 ± 0.11	No	Yes/Prior	
A2GS86	53.180784	−27.835827	11385	1.95	10.12	3.65	0.41 ± 0.12	No	No	
A2GS87	53.119150	−27.814066	14122	3.32	10.56	3.63	0.43 ± 0.13	No	Yes/Prior	AS33
A2GS88	53.074368	−27.849730	8874	0.123 [k]	10.21	3.59	0.41 ± 0.11	No	No	

Notes. Spectroscopic redshift references: [a]Vanzella et al. (2008); [b]Garilli et al. (2021); [c]Balestra et al. (2010); [d]Morris et al. (2015); [e]Mignoli et al. (2005); [f]Momcheva et al. (2016); [g]Wuyts et al. (2009); [h]Ferreras et al. (2009); [i]Decarli et al. (2019); [j]Kurk et al. (2013); [k]Cooper et al. (2012).

best choice for the dataset in this work because this technique does not assume an a priori functional form and it can be applied to both the 100% pure main catalog and the prior-based supplementary catalog, as the latter sources are not detected at a S/N^{peak} high enough for reliable uv plane estimates.

We measured total flux densities of the sources in both the 100% pure main catalog and the prior-based supplementary catalog using aperture photometry in the primary beam corrected combined dataset dirty map (e.g., Simpson et al. 2015a; Scoville et al. 2016; Liu et al. 2019). Through growth curves, we tested a range of apertures from 0.2 to 4″ diameter in steps of 0″.2. We chose an aperture of 1″.6 diameter, which gave the optimal trade-off between total flux retrieval and S/N^{total}, defined as the total flux density divided by its uncertainty. The fluxes were corrected by the appropriate aperture corrections to account for the flux losses outside the aperture. This aperture correction was calculated by dividing the flux within the aperture of 1″.6 diameter by the flux enclosed

in the synthesized dirty beam within the same aperture (normalized to its maximum value). As shown in Sect. 2.2, while there are slight variations in the beam profile over the map, these differences do not introduce systematics in the flux measurements. The reason is that the beam used to perform the correction was that associated to the specific location of a given source. Besides, the nature of the aperture correction is independent of the specific shape of the synthesized dirty beam or its deviation from a Gaussian shape compared to a clean beam. This also supports the aperture technique as the best choice for the dataset in this work. In any case, we checked whether a common aperture correction significantly influences the flux density measurements. This is shown in the right panel of Fig. 3, where the aperture correction is by definition the inverse of the y-axis. The correction range is 1.38−1.55 for the aperture radius 0″.8 across the beams associated to the different slice submosaics, yielding a flux density variation of <10% if a common average beam correction across the field is used.

As a sanity check, we measured total flux densities using 2D functional fitting in the image plane via the CASA task `imfit`, yielding consistent results with the aperture photometry methodology with a relative difference between them given by the median $(S_{\text{imfit}} - S_{\text{ap}})/S_{\text{ap}} = 0.01 \pm 0.14$ (where the uncertainty is the median absolute deviation). In Tables 2, 3, and B.1 we present the flux density measurements obtained from the aperture photometry methodology.

In comparison with the flux measurements in Franco et al. (2018, 2020b) for the sources in common with this work, the relative difference between them given by the median $(S_{\text{AGS}} - S_{\text{A2GS}})/S_{\text{A2GS}} = -0.11 \pm 0.28$ (where the uncertainty is the median absolute deviation). Therefore, on average the flux measurements in Franco et al. (2018, 2020b) are systematically slightly lower due to limited *uv* coverage.

Sizes can be measured in the image plane or directly in the *uv* plane using the information from the complex visibilities, leading to more reliable results than image plane techniques. Therefore, we measured the sizes directly in the *uv* plane using the combined dataset to include the information on multiple spatial scales and access the largest possible range of intrinsic source sizes with minimum biases. We employed the CASA task `UVMODELFIT` that allows us to fit single component models to single sources. Since the scope of this work is to get global size measurements, we fit a Gaussian model with fixed circular axis ratio. In order to isolate each source we split the combined dataset mosaic into single source measurement sets as follows: for each source pair of coordinates we searched for all the pointings that contained data on that source (each source is covered by six pointings typically). Next, each pointing was phase shifted to set the phase center at the source coordinates using the CASA task `fixvis`. Data and weights are modified to apply the appropriate primary beam correction that correspond to the phase shift, by using the CASA toolkit `MeasurementSet` module. After that, by using the CASA task `fixplanets` the phase center was set to the source coordinates for all the pointings that contained data on the source and the visibility weights recomputed with the task `statwt`. Finally, the pointings were concatenated into a single measurement set. In Table 2, we present the size measurements obtained from the *uv* plane fitting methodology in the combined dataset. We only report the values for the sources in the 100% pure main catalog, which have a detection $S/N^{\text{peak}} \geq 5$ (defined in Sect. 3.1.1) as the PyBDSF peak flux density divided by the average background rms noise of the island). Below this S/N, size measurements are unreliable. For this subset of sources we also determined the minimum possible size that can be reliably measured from the formula by Martí-Vidal et al. (2012):

$$\theta_{\min} = \beta \left(\frac{\lambda_c}{2S/N^2} \right)^{1/4} \theta_{\text{beam}} \simeq 0.88 \frac{\theta_{\text{beam}}}{\sqrt{S/N}}, \quad (2)$$

with $\lambda_c = 3.84$ and $\beta = 0.75$ as in Franco et al. (2018, 2020b). The minimum size ($\theta_{\min}$) is given in units of the synthesized beam FWHM (θ_{beam}) depending on the source S/N. Values below this minimum are assumed to be unresolved and we report them as upper limits in Table 2.

3.2.1. Size distribution

In Fig. 6 we show the size distribution of the sources in the 100% pure main catalog (gray). Sizes are displayed as the effective (half-light) radius of the circular Gaussian model fit in the *uv* plane. We distinguish between sources present in the high resolution dataset (blue) and sources detected in the low resolution but
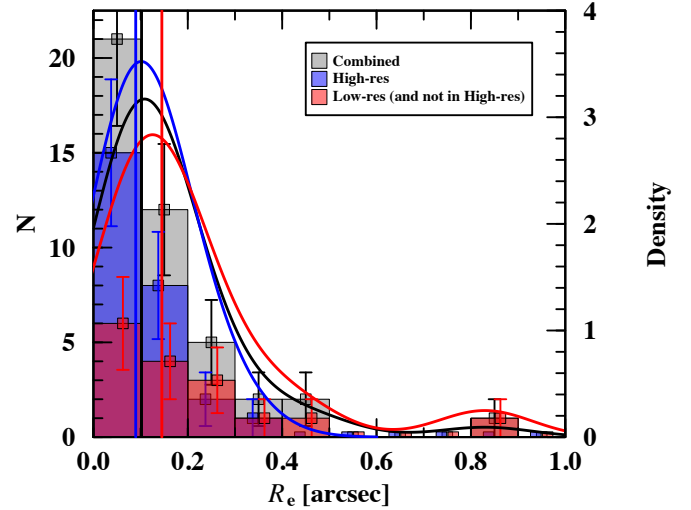

Fig. 6. Size distribution of the 100% pure main catalog detected in the combined dataset (gray). We show histograms with Poisson error bars and probability density curves (kernel density estimates, by definition normalized to an area under the curve equal to one). Medians are displayed as a solid vertical line. The sizes were measured as the effective radius of the circular Gaussian model fit in the *uv* plane. Sources detected in the high resolution dataset are shown in blue, while sources also present in the low resolution dataset but not in the high resolution dataset, are shown in red. We note that the black, blue, and red histograms are overlaid, not stacked. The histogram bins are such that all the upper limits fall in the first bin to keep a correct shape.

not in the high resolution dataset (red). There is only one source that appears in the 100% pure main catalog from the combined dataset but was not in the high/low resolution datasets, namely A2GS33.

First, we notice that the distribution appears skewed toward small sizes. Sources are compact with a median effective (half-light) radius of $R_e = 0\farcs10 \pm 0\farcs05$ and a median physical size of $R_e = 0.73 \pm 0.29\,\text{kpc}$ calculated at the redshift of each source (where the uncertainties are given by the median absolute deviation). Among them $R_e = 0\farcs09 \pm 0\farcs03$ ($R_e = 0.70 \pm 0.23\,\text{kpc}$ calculated at the redshift of each source) corresponds to those also present in the high resolution dataset and $R_e = 0\farcs15 \pm 0\farcs10$ ($R_e = 1.21 \pm 0.82\,\text{kpc}$ calculated at the redshift of each source) corresponds to those detected in the low resolution but not in the high resolution dataset (upper limits are taken at face value, with three and five sources respectively on each group).

Franco et al. (2018) extraction in the high resolution dataset was carried out in a tapered map with a homogeneous and circular synthesized beam of $0\farcs6$ FWHM under the assumption that the sources were point-like at that angular resolution. For these sources detected in the tapered map, Franco et al. (2018) measured sizes in another high resolution dataset map constructed by employing a natural weighting scheme leading to the same resolution we achieve in this work with the same dataset in our independent analysis. Franco et al. (2018) reported a median size of $R_e = 0\farcs09$ ($0\farcs18$ FWHM also fit with a circular Gaussian model using `UVMODELFIT`), with 85% of the sources with $R_e < 0\farcs125$. Therefore, Franco et al. (2018) results are perfectly consistent with our median for sources detected in the high resolution dataset, with 73% of them indeed below $R_e < 0\farcs125$. In conclusion, even after including the information on multiple spatial scales making possible to access a wider range of intrinsic source sizes, the results in Franco et al. (2018) hold and were not biased to small values due to limited *uv* coverage.

A variety of literature studies in the ALMA era have concluded that the dust continuum emission appears located in compact regions with median sizes among the different samples within a circularized $R_e = 0.''10-0.''15$ (e.g., Simpson et al. 2015a; Ikarashi et al. 2015; Hodge et al. 2016; Fujimoto et al. 2017; Gómez-Guijarro et al. 2018). In principle, some of the results could be biased toward small sizes if lacking *uv* coverage and/or sensitivity. However, Elbaz et al. (2018) also reported compact dust continuum emission ($R_e = 0.''10-0.''15$, circularized) in a sample of DSFGs at $z \sim 2$ with long integration times reaching a typical $S/N \sim 35$. Our results with improved *uv* coverage are consistent with the conclusion that dust continuum emission occurs in compact regions. Of course, GOODS-ALMA 2.0 is a flux limited survey and we discuss this conclusion more extensively in this regard in Sect. 6.

We notice in Fig. 6 a small difference between the size distribution of sources present in the high resolution dataset compared to those that are in the low resolution but not in the high resolution dataset, with the latter skewed toward larger sizes and with a larger scatter. There is also one outlier to the smooth size distribution, namely A2GS30 with $R_e = 0.''86 \pm 0.''16$. A2GS30 is located at a distance $<5''$ with respect to another source, namely A2GS20 with a size of $R_e < 0.''13$. The latter could be an indication that our size measurements are biased to larger sizes in the vicinity of close neighbors. Therefore, we inspected all $<5''$ pairs. A2GS33/37 are slightly larger than the average and have a similar $z_{\rm phot}$ (see Table 2). It could be an indication that size measurements of close pairs are systematically affected or that they are physically larger due to interactions. However, there are another three $<5''$ pairs (A2GS12/24, A2GS9/21, A2GS14/18) and none of them have anomalously large sizes even when located at similar $z_{\rm phot}$ (A2GS9/21). As a sanity check, we measured sizes for these sources using the GILDAS task `uvfit` as an alternative *uv* plane fitting tool. We fit two Gaussian models with fixed circular axis ratio simultaneously, yielding consistent results. A common characteristic that the three galaxies with the larger sizes (A2GS30/33/37) share is their ID ≥ 30, pointing to the S/N as a potential reason since our IDs are ordered with decreasing detection $S/N^{\rm peak}$. Either some of the lower $S/N^{\rm peak}$ sources are systematically larger due to an artificial bias in the size measurements for lower $S/N^{\rm peak}$ or, as a low $S/N^{\rm peak}$ is also related with a generally lower flux density, these sources are physically fainter and larger. Franco et al. (2020b) indeed argued that the prior-based methodology allowed for one to access a population of fainter and slightly larger sources. While a detail analysis about the dust continuum emission profiles is out of the scope of this paper, we discuss potential size variations due to differences in flux densities in Sect. 6.

It is important to consider that the dust continuum sizes at 1.1 mm in this work are associated to sources that span a wide redshift range ($0 < z < 5$; see Sect. 5). The dust continuum emission in the Rayleigh–Jeans (RJ) side ($\lambda_{\rm rf} \gtrsim 250\,\mu{\rm m}$) of the IR spectral energy distribution (SED) is more sensitive to the dust mass, while the dust continuum emission around the peak of the IR SED is sensitive to variations in the dust temperature. Most of the sources are located in the redshift range $1 < z < 4$, spanning rest-frame wavelengths $0.22-0.55\,\mu{\rm m}$ and, therefore, the dust continuum emission traced is that of the RJ side. For sources with increasingly high redshifts, specially those at $z > 4$, the dust continuum emission gets closer to the peak of the IR SED. Popping et al. (2022) has recently studied the extent of the dust continuum emission for thousands of MS galaxies drawn from the TNG50 simulation (Nelson et al. 2019; Pillepich et al. 2019) between $1 < z < 5$. The authors concluded that the half-light

radii of galaxies at observed-frame wavelengths from 700 μm to 2 mm are similar to those at rest-frame 850 μm at the $5-10\%$ level, for galaxies at redshifts $1 < z < 5$ and stellar masses $9.0 < \log(M_*/M_\odot) < 11.0$. Therefore, the dust continuum sizes for the sources in this work which rest-frame dust continuum emission gets closer to the peak of the IR SED are at most 10% smaller than those associated to the RJ side of the IR SED.

3.2.2. Test: Flux growth curves from tapering

Although the *uv* coverage is sensitive to both small and large spatial scales and flux losses are expected to be negligible, we tested the flux density measurements using another methodology: a growth curve built after tapering the data (Xiao et al., in prep.). Data tapering adds an additional weight function that reduces the weights of the outer visibilities at the expense of also reducing the collecting area and, thus, the sensitivity. Nevertheless, the tapering also reduces the angular resolution, which is beneficial to optimize the sensitivity to sources that are larger than the angular resolution. We created tapered mosaics of the combined dataset, starting from the original resolution and stopping when the resulting tapered PSF $R_e = 1.''5$, much larger than any reasonable source size. Then, for a given source, we measured the peak flux density on every tapered mosaic and built growth curves (peak flux as a function of the tapered PSF R_e). When the tapering reaches the point where the intrinsic source size is below the angular resolution, the entire flux of the source is retrieved in a single beam and the total flux can be read as the peak flux. In order to decide what tapering length is the one for which the entire source flux is measured in a single beam (the position in the *x*-axis at which we read the source flux in the source growth curve) we set the criteria: (1) measure always at least when the maximum $S/N^{\rm total}$ is reached; (2) measure either when the first derivative of the $S/N^{\rm total}$ (signal increase with respect to noise increase) is below one (more noise than signal enters the beam respect to the previous step) or the first derivative of the $S/N^{\rm total}$ has a local minimum (to avoid including nearby noise peaks in the flux measurements). In the left panel of Fig. 7 we compare the flux densities obtained using the aperture and the tapering methodologies. The relative difference between them is given by the median $(S_{\rm ap} - S_{\rm tap})/S_{\rm tap} = -0.03 \pm 0.18$ (where the uncertainty is the median absolute deviation), with -0.04 ± 0.15 for the 100% pure main catalog and -0.01 ± 0.21 for the prior-based supplementary catalog. In addition, we compare the flux densities associated to the size measurements in the *uv* plane for the 100% pure main catalog in the right panel of Fig. 7, being the relative difference in this case 0.04 ± 0.08. Therefore, the different methodologies provide consistent flux density measurements and in particular the agreement with the fluxes obtained in the *uv* plane also contributes to the robustness of the size measurements.

4. Number counts

Calculation of the number counts requires to assess different aspects of the survey that influence the ability to retrieve sources of a given flux and size or the flux measurements themselves. For this reason, before jumping into the calculation of the number counts, we dealt with these aspects.

4.1. Completeness and boosting

The completeness is the fraction of real sources that are actually detected for a given flux and size. In order to compute the completeness, we performed simulations by injecting artificial
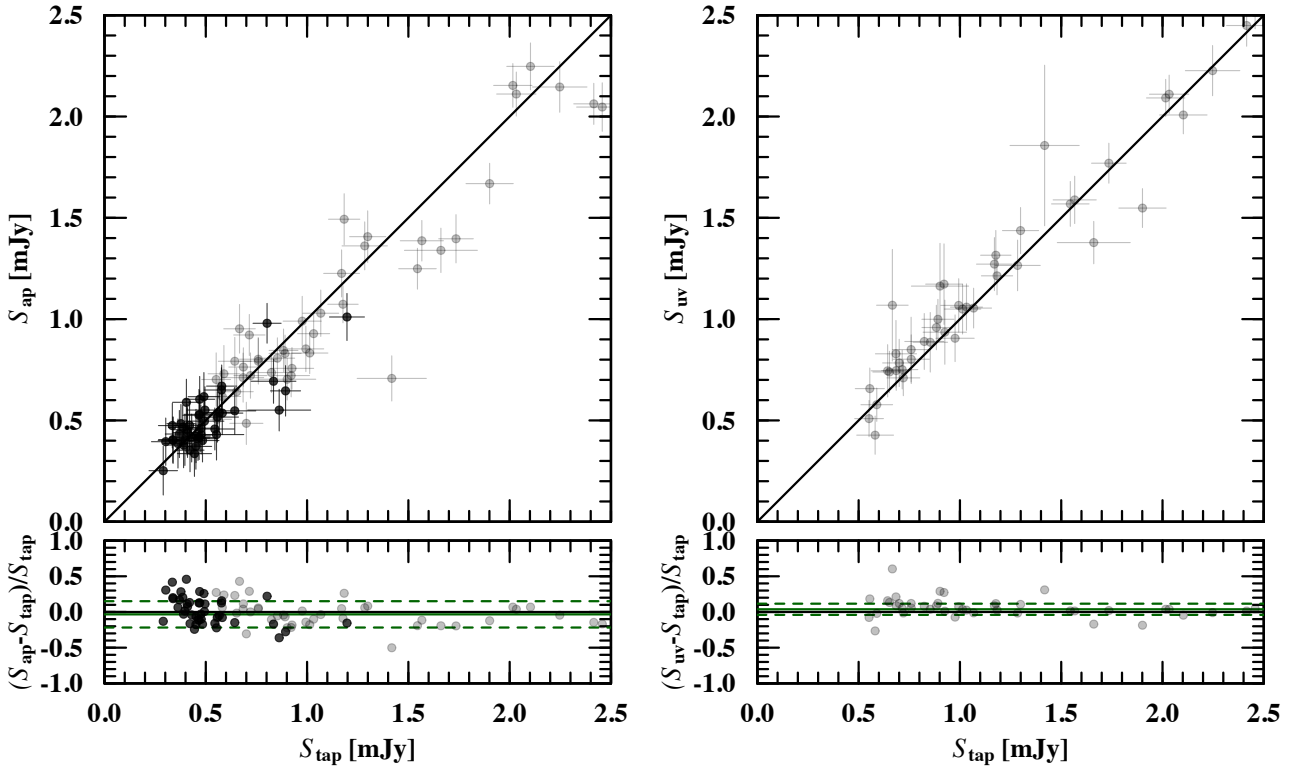
Fig. 7. *Left panel*: comparison between the flux densities obtained using the aperture and the tapering methodologies for both the 100% pure main catalog (gray symbols) and the prior-based supplementary catalog (black symbols). The median relative difference is $(S_{ap} - S_{tap})/S_{tap} = -0.03 \pm 0.18$ (shown with green lines in the *bottom panel*). *Right panel*: comparison between the flux densities associated to the size measurements in the *uv* plane and the flux densities from the tapering methodology for the 100% pure main catalog (grey symbols). The median relative difference in this case is $(S_{uv} - S_{tap})/S_{tap} = 0.04 \pm 0.08$.

model sources in the combined dataset map. We modeled Gaussian sources, convolved them with the combined dataset synthesized dirty beam, and injected them in the combined dataset map at random locations. In total we injected 450 sources for a given input flux and size. This number is such that, considering the number of independent beams, the probability of two sources overlapping is negligible (∼1%). Besides, the scarcity of real sources in the map allowed us to work directly with the dirty map. Even if the overlapping probability is very low, model sources could be close enough to other model or real sources and affect their flux measurements. To be sure that the latter is not the case, we eliminated any model source within 5″ diameter of another model or real source. The simulations were carried out for flux densities ranging 0.1−3 mJy in steps of 0.1 mJy and sizes from pure point sources to 1″ FWHM in steps of 0″.1. In total, a grid of 30 fluxes and 11 sizes composed of 450 sources each.

After the injection of artificial model sources in the combined dataset map, we performed the same blind source detection procedure described in Sect. 3. In the left panel of Fig. 8 we show the completeness as a function of the input flux density (S_{in}) for the different simulated source sizes as detected for $\sigma_p = 3.0$. The survey reaches a ∼100% completeness for all simulated sizes for flux densities $S_{in} > 1$ mJy. The completeness is also lower for increasing source sizes at fixed flux densities.

For the purpose of knowing the behavior of the survey and the detection procedure in retrieving certain types of sources, the completeness analysis in S_{in} is relevant. However, input fluxes in the real data are unknown by nature and, thus, for the practical purpose of applying a completeness correction to the number counts we need to know its behavior as a function of the flux that

we are actually able to measure, the output flux density (S_{out}). S_{out} measurements were performed following the same aperture photometry methodology described in Sect. 3.2. In the right panel of Fig. 8 we show the completeness as a function of S_{out} for the different simulated source sizes as detected for $\sigma_p = 3.0$. This plot provides the completeness correction to be applied to the number counts. Qualitatively, the behavior in terms of S_{in} or S_{out} is similar (i.e., the completeness reaches ∼100% for sources with $S_{1.1\,mm} > 1$ mJy, it progressively decays below this value, and it is also lower for increasing source sizes at fixed flux densities), although quantitatively the behavior changes.

Another aspect to characterize is flux boosting, which consists in the artificial increase of S_{out} compared to S_{in} typically observed for sources detected at very low S/N (e.g., Murdoch et al. 1973; Hogg & Turner 1998; Coppin et al. 2005). Flux boosting is connected to the detection threshold, as it reflects the fact that detectable sources at very low S/N are those located in a noise peak and, thus, their flux measurements are systematically boosted, leading to the observed increase of S_{out} compared to S_{in} at low S/N. We used the same set of simulations to study the S_{out}/S_{in} ratio as a function of the output detection S/N^{peak} (S/N_{out}^{peak}) as shown in the left panel of Fig. 9, where we represent the range of sizes (0″.1−0″.4 FWHM), S/N^{peak} (3.5−20), and flux densities (0.25−2.75 mJy) measured in the real sources. The S_{out}/S_{in} ratio is stable over the whole studied range of detection S/N^{peak} and fluxes, as traced by the sliding median. There is evidence for a small level of flux boosting at $3.5 < S/N_{out}^{peak} < 5.0$, reaching 4% at $S/N_{out}^{peak} = 3.5$. We applied flux boosting corrections as a function of the source detection S/N^{peak}

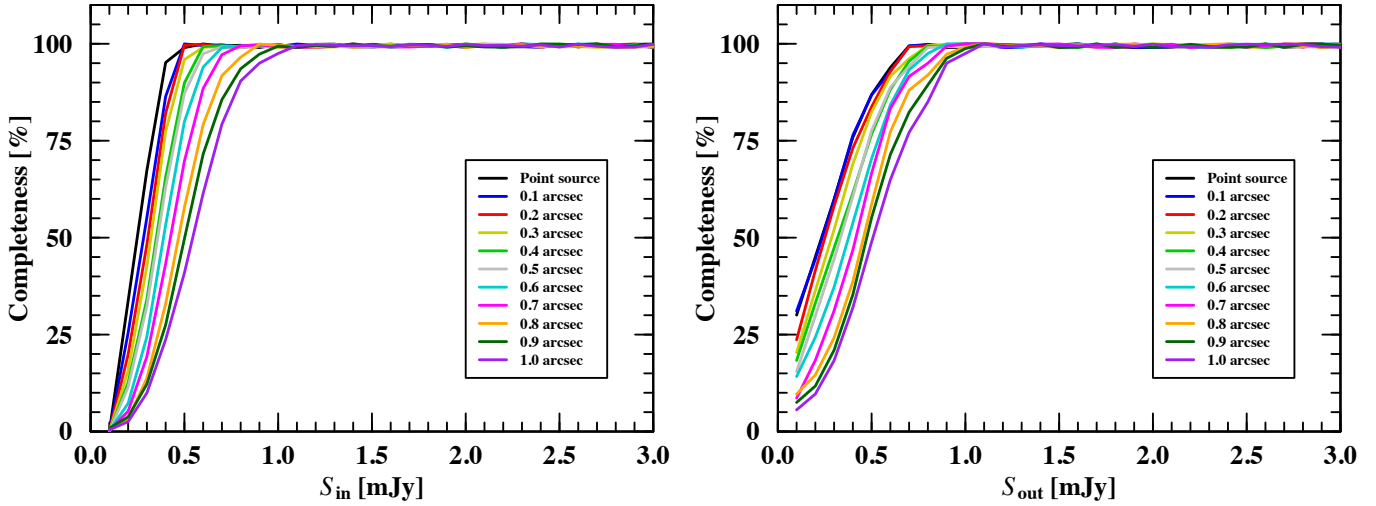**Fig. 8.** Completeness as a function of the input flux density (S_{in}; *left panel*) and as a function of the output flux density (S_{out}; *right panel*) for different model source sizes ranging from pure point sources to $1''$ FWHM.
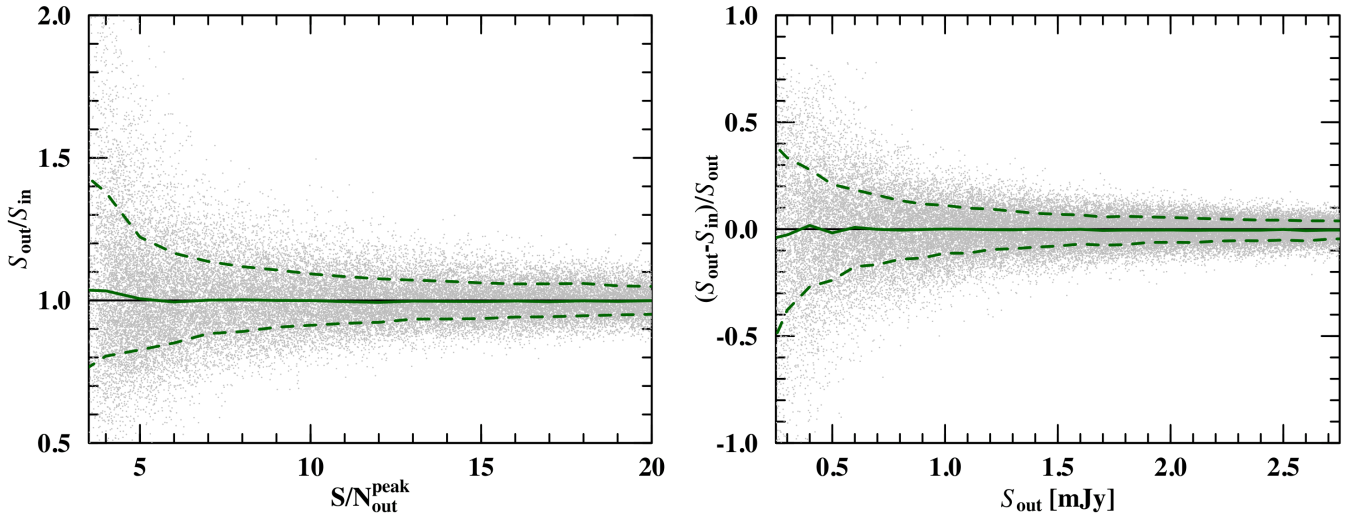

Fig. 9. Ratio of the output over input flux densities (S_{out}/S_{in}) as a function of the output detection S/N^{peak} (S/N_{out}^{peak}) from PyBDSF, measured as the peak flux density over the local rms noise (*left panel*), and flux density accuracy (($S_{out} - S_{in})/S_{out}$) as a function of the output flux density (S_{out}) measured with the aperture photometry methodology (*right panel*) for simulated sources with sizes ranging $0''.1-0''.4$ FWHM. The distribution of the whole set of simulations is shown as gray symbols with their sliding median and standard deviation in solid and dashed green, respectively.

accordingly in the flux density measurements presented in Tables 2, 3, and B.1.

The set of simulations was also used to assess the accuracy of the flux density measurements and whether they are affected by systematics to be corrected. In the right panel of Fig. 9 we show the flux density measurements accuracy as given by ($S_{out} - S_{in})/S_{out}$ as a function of S_{out}. The sliding median of ($S_{out} - S_{in})/S_{out}$ is ~1 over the entire S_{out} range studied. Therefore, we did not add any further correction to the measured flux densities based on our simulations.

We also notice the fact that the $1''.6$ diameter aperture where we measured the fluxes provides on average a similar S/N^{total} to that of the detection S/N^{peak}. This is seen in the left panel of Fig. 9 as how the detection S/N_{out}^{peak} (*x*-axis) coincides with the S_{out}/S_{in} ratio (*y*-axis), a proxy for S/N^{total}. For example, $S/N_{out}^{peak} = 10$ is associated with $S_{out}/S_{in} \sim 0.9{-}1.1$ and, thus, a ~10σ detection has typically a $S/N^{total} \sim 10$. The latter does not necessarily mean that a source that has a flux density 10 times

over the rms noise has a flux accuracy of ~10%, since this is only true for pure point sources. This is seen in the right panel of Fig. 9 as, for example, a $S/N_{out}^{peak} = 10$ detection, considering the average sensitivity of $\sigma = 68.4 \, \mu$Jy beam^{-1}, has a total flux density of 0.68 μJy in the case of a pure point source associated with $S_{out}/S_{in} \sim 0.9{-}1.1$. However, ($S_{out} - S_{in})/S_{out}$ is on the level of ~15%, which is expected since what we represent are extended sources with a range of sizes $0''.1-0''.4$ FWHM and, thus, there are sources with high total flux densities but with low detection S/N^{peak} widening the distribution of the ($S_{out} - S_{in})/S_{out}$ ratio as a result.

4.2. Effective area

The survey sensitivity is not perfectly homogeneous and there are small differences between regions seen between the different slice submosaics (see Table 1). We calculated the effective area for a given sensitivity as shown in the curve in
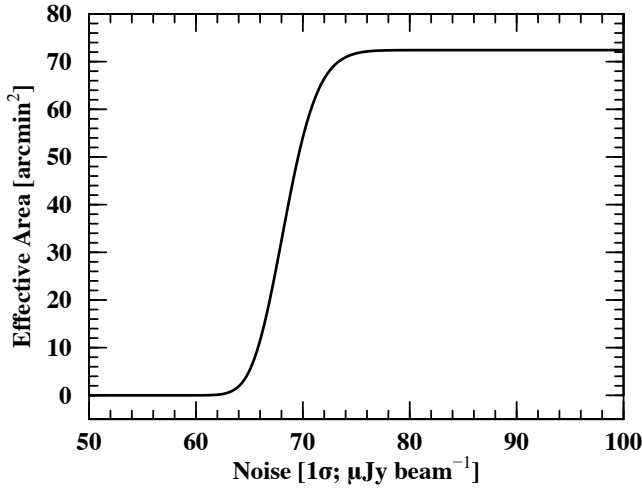
Fig. 10. Effective area versus noise. The curve is built by counting the area above a given noise (1σ) threshold.

Fig. 10. The curve is built by counting the area in the combined dataset noise map above a given noise (1σ) threshold. The total survey area is 72.42 arcmin^2, with 100% of the area reaching a sensitivity of at least 83.5 µJy and 90% of at least 71.7 µJy.

4.3. Differential and cumulative number counts

The contribution of a source i to the number counts at a frequency ν is:

$$\xi(S_{\nu,i}) = \frac{p(S_{\nu,i}^{\mathrm{peak}})}{A_{\mathrm{eff}}(S_{\nu,i}^{\mathrm{peak}})\, C(S_{\nu,i}, \theta)}, \tag{3}$$

where $p(S_{\nu,i}^{\mathrm{peak}})$ is the purity as defined in Eq. (1), $A_{\mathrm{eff}}(S_{\nu,i}^{\mathrm{peak}})$ is the effective area as given by the curve in Fig. 10, which are both associated to the detection S/N^{peak} and, thus, described in terms of the peak flux density $S_{\nu,i}^{\mathrm{peak}}$. $C(S_{\nu,i}, \theta)$ is the completeness for a pair flux density, size ($S_{\nu,i}, \theta$), as explained in Sect. 4.1.

The differential number counts are obtained adding the contribution of sources within a flux density interval ΔS_ν:

$$\frac{\mathrm{d}N(S_\nu)}{\mathrm{d}S_\nu} = \frac{1}{\Delta S_\nu} \sum_i \xi(S_{\nu,i}). \tag{4}$$

Cumulative number counts are calculated summing over all the sources with a flux density higher than S_ν:

$$N(>S_\nu) = \sum_{\forall S_{\nu,i} > S_\nu} \xi(S_{\nu,i}). \tag{5}$$

We calculated the contribution of each source to the number counts. In the case of the purity correction we applied $p = 1$ for all sources. This is the case for the 100% pure main catalog by definition. For the prior-based supplementary catalog the purity correction to be applied is, in principle, that studied in Sect. 3.1.1 as a function of the detection S/N^{peak}. However, the purity correction applied in this way is valid only for sources from the blind detection procedure before the prior-based selection. Once the sources are validated by the priors the purity correction has to be adjusted to a smaller value, reflecting the better knowledge of the actual real sources aided by priors. In order to assess

what the adjusted purity correction is, we compared the number of sources in the 100% pure plus prior-based catalogs (88 sources) with the expected number of real sources (98 ± 32, see Sect. 3.1.2). Both are consistent and, therefore, likely possible that we are capturing all the real sources down to a detection $S/N^{\mathrm{peak}} = 3.5$. Therefore, we assumed $p \sim 1$ for the sources in the prior-based supplementary catalog, knowing that if a ($\sim$10%) fraction of the sources with at $S/N^{\mathrm{peak}} = 3.5$–5.0 were missed it does not significantly affect the number counts.

For the size dependency we used the sizes from the uv plane fitting as derived in Sect. 3.2 for the 100% pure main catalog. If a source lacked of size estimation we used the median size of the 100% pure sources. In the case of the prior-based supplementary catalog sources, whose sizes are not reliable through uv plane fitting, we employed the median size of the 100% pure sources at $S_{1.1\,\mathrm{mm}} < 1$ mJy, since the prior-based sources are in this flux density regime.

We decided the optimal bin width and first bin to calculate the number counts as an optimal trade-off between resolution and sufficient number of sources per bin. The chosen bin width was $\Delta \log(S_\nu) = 0.20$. The uncertainties for each bin were calculated from 10 000 Monte-Carlo simulations varying the source fluxes randomly within their uncertainties added in quadrature to the Poisson uncertainties.

Another correction applied here was the Eddington bias, as it is sometimes called. According to this effect, because the sources with lower luminosities are more numerous than brighter sources, the noise leads to an overestimation of the number counts in the fainter flux bins. To take into account this effect we simulated a physically informed number of sources using the slope of the number counts in Franco et al. (2018) and added Gaussian noise to each simulated source. We applied a correction factor to each flux density bin as the ratio between the flux distribution before and after adding the noise.

Cosmic variance was not taken into account when calculating the uncertainties in the number counts. As also discussed in Franco et al. (2018), while cosmic variance it is expected to be significant for massive galaxies in small solid angles, the strong negative K-correction for redshifts above $z > 1.8$ at 1.1 mm and up to the highest redshift in our catalog ($z = 4.73$) counterbalances it. The comoving volume over $\Delta z \sim 3$ of 1400 Gpc^3 is relatively large. Using Moster et al. (2011) to estimate the effect of cosmic variance, it results in $\sim$15% and, thus, it does not significantly affect the number counts.

The resulting number counts derived using the whole 100% pure main plus prior-based supplementary catalogs are shown in Table 4 and in Fig. 11 (black symbols; combined). Additionally, we studied the contribution to these combined number counts from the 100% pure main catalog, the prior-based supplementary catalog, and the high and low resolution datasets independently to have an idea of which type of sources contribute to the number counts as a function of flux density. In order to do this we derived alternative versions of the number counts as if the only sources detected were those associated to the desired type of sources to be studied. The completeness correction applied was that associated to the combined number counts from the whole 100% pure plus prior-based catalogs and, thus, by definition they are insufficient to reach the combined number counts level, but in turn they reflect the contribution from that particular type of sources to the number counts. The results from the 100% pure main catalog only (gray symbols) manifest the effect of the prior methodology in the number counts when compared to the combined number counts from the 100% pure plus prior-based catalogs (black symbols). Adding, the prior-based supplementary catalog it is

Table 4. Differential and cumulative number counts derived using the 100% pure plus prior-based catalogs.

$S_{1.1\,\rm mm}$ (mJy)	$dN/dS_{1.1\,\rm mm}$ (mJy^{-1} deg^{-2})	$N_{\rm diff}$	$N(>S_{1.1\,\rm mm})$ (deg^{-2})	$N_{\rm cum}$
(1)	(2)	(3)	(4)	(5)
(0.30)	(2900^{+5200}_{-1500})	(4)	(5110^{+100}_{-240})	(88)
0.48	9900^{+1500}_{-3500}	34	4700^{+380}_{-730}	84
0.75	4030^{+450}_{-470}	28	2510^{+150}_{-130}	50
1.19	1260^{+97}_{-270}	14	1100^{+100}_{-100}	22
1.89	400^{+110}_{-20}	7	400^{+100}_{-20}	8
3.00	36^{+36}_{-6}	1	49^{+50}_{-7}	1

Notes. (1) Central flux density in the bin; (2) differential number counts; (3) number of sources per bin in the differential number counts; (4) cumulative number counts; (5) number of sources per bin in the cumulative number counts. Bin width is $\Delta \log(S_\nu) = 0.20$. The uncertainties are calculated from Monte-Carlo simulations and Poisson added in quadrature. In parentheses, the first bin not used in our analysis.

possible to access a population of fainter 1.1 mm sources as indicated by Franco et al. (2020b). The results from the high resolution (blue symbols) and low resolution (red symbols) datasets manifest that both are able to capture the bright end of the combined number counts, but the high resolution dataset is very inefficient in retrieving the faint end as it rapidly decays at $S_{1.1\,\rm mm} < 1$ mJy. The low resolution dataset is much more complete at lower flux densities and similar to the 100% pure main catalog from the combined dataset. The difference between the high resolution and low resolution can be explicitly seen as the dashed red line that represents the sources detected in the low resolution dataset but not present in the high resolution dataset. In conclusion, the low resolution dataset is very efficient at retrieving sources for a wide range of flux densities, but the high resolution dataset is biased to the brighter ones. The combination of the high resolution and low resolution datasets (combined dataset), along with the prior methodology, allow us to be more complete at the faint end.

In comparison with the number counts analysis presented in Franco et al. (2018) from the high resolution dataset, the cumulative number counts in our analysis are slightly higher for a fixed flux density bin (see right panel of Fig. 11, dark gray circles from Franco et al. 2018 compared to blue squares from this work). We note that Franco et al. (2018) extraction in the high resolution dataset is slightly different compared to the high resolution dataset detection here, since the former was carried out in a tapered map with a homogeneous and circular synthesized beam of 0″6 FWHM. Besides Franco et al. (2018) used a 80% pure catalog, while in this work we employed the 100% pure plus prior-based catalogs. This work is more complete and less dependent on both the purity and completeness corrections compared to Franco et al. (2018) analysis and, thus, the latter is potentially more affected by systematics that lead to slightly different number counts.

In Fig. 11 we also compare our number counts with the literature that studied number counts at similar wavelengths from both ALMA (Hatsukade et al. 2013, 2016, 2018; Karim et al. 2013; Ono et al. 2014; Carniani et al. 2015; Simpson et al. 2015b, 2020; Aravena et al. 2016; Fujimoto et al. 2016; Oteo et al. 2016; Dunlop et al. 2017; Umehata et al. 2017; Franco et al. 2018; Muñoz Arancibia et al. 2018; Stach et al. 2018; González-López et al. 2020; Béthermin et al. 2020) and

single-dish (e.g., Lindner et al. 2011; Scott et al. 2012; Geach et al. 2017; Simpson et al. 2019; Shim et al. 2020) facilities. When the wavelength is different than 1.1 mm we applied a correction factor to the flux densities so the number counts are comparable. These factors were calculated using a modified black body (MBB) model, assuming a dust emissivity index $\beta = 1.5$ and dust temperature $T_{\rm dust} = 35$ K. The corrections are: $S_{1.1\,\rm mm}/S_{1.3\,\rm mm} = 1.79$, $S_{1.1\,\rm mm}/S_{1.2\,\rm mm} = 1.36$, $S_{1.1\,\rm mm}/S_{870\,\mu\rm m} = 0.44$, $S_{1.1\,\rm mm}/S_{850\,\mu\rm m} = 0.41$. Our results are in agreement with the general trends of literature studies covering the flux densities around the knee of the number counts accurately with dozens of galaxies in most of the flux density bins.

A discrepancy between the ALMA results from Karim et al. (2013) and some of the single-dish studies (e.g., Lindner et al. 2011; Scott et al. 2012) exists at the bright end (e.g., Franco et al. 2018). The results presented here reduce the tension compared to the results in Franco et al. (2018), although the discrepancy still exist. The origin of the discrepancy was suggested to come from the challenging boosting and blending effects that affect the single-dish measurements leading to an overestimation of the bright end of the number counts (e.g., Karim et al. 2013; Ono et al. 2014; Simpson et al. 2019). Recent ALMA studies at the bright end (Stach et al. 2018; Simpson et al. 2020) have shown indeed agreement with the first ALMA results by Karim et al. (2013). The latter tendency has also been strengthened by more recent single-dish studies (Geach et al. 2017; Simpson et al. 2019; Shim et al. 2020).

The wealth of ALMA studies in the literature along with the results presented in this work cover a wide range of flux densities spanning from the faint to the bright end of the number counts, making it possible to accurately perform a fit. We fit a Schechter (Schechter 1976) and a double power law (DPL) function to both the differential and cumulative number counts (see Table 5):

$$\phi(S_\nu) = \frac{N_0}{S_0} \left(\frac{S_\nu}{S_0}\right)^\alpha \exp\left(-\frac{S_\nu}{S_0}\right) \tag{6}$$

is a modified Schechter function conventionally used in similar studies at these wavelengths (e.g., Coppin et al. 2006; Knudsen et al. 2008; Austermann et al. 2010), where N_0 is the normalization, S_0 the characteristic flux density, and α is the faint-end slope.

$$\phi(S_\nu) = \frac{N_0}{S_0} \left[\left(\frac{S_\nu}{S_0}\right)^\alpha + \left(\frac{S_\nu}{S_0}\right)^\beta\right]^{-1} \tag{7}$$

is the DPL (e.g., Scott et al. 2002; Coppin et al. 2006; Knudsen et al. 2008). In this fit we did not consider our first flux density bin since it appears clearly incomplete. We did not take into account upper limits presented in some of the literature studies and neither did we consider the results from single-dish surveys to keep the fit as clean as possible restricted to ALMA. The Schechter fit performs better than the double power law, specially at the bright end.

In Fig. 11 we also display the predictions of an updated version of the modeled Simulated Infrared Dusty Extragalactic Sky (SIDES) by Béthermin et al. (2017) at 1.1 mm for both the differential and cumulative number counts. This model shows an overall good agreement with our derived best-fit to the ALMA-based observed 1.1 mm number counts, albeit a prediction of a slightly more pronounced flattening at the faint end and a mild excess of sources at the bright end. In the right panel of Fig. 11 we also include the cumulative number counts predictions by different galaxy evolution models in the literature
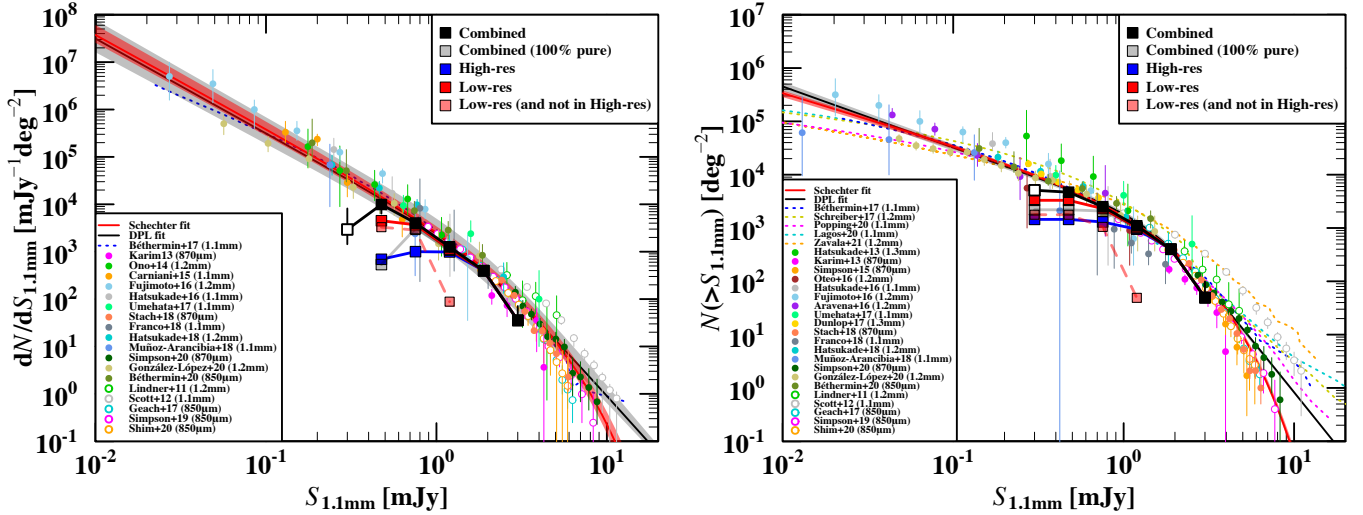
Fig. 11. Differential (*left panel*) and cumulative (*right panel*) number counts from the combined dataset using the 100% pure plus prior-based catalogs (black symbols). The first bin (white symbol) is not used in our analysis. Also displayed the contribution of the 100% pure catalog only (gray symbols), the sources extracted in the high resolution (blue symbols) and low resolution (red symbols) datasets. In both panels we display literature studies of ALMA number counts at similar wavelengths of 1.3, 1.2, 1.1 mm, 870 μm, and 850 μm (filled circles) converted to 1.1 mm as explained in the main text. We also add some other single-dish literature studies (open circles). Best-fit Schechter and DPL functions using the entire set of number counts from all the ALMA studies are shown in red and black solid lines, respectively, with their uncertainties corresponding to the 16% and 84% percentiles as a shaded area. Predicted number counts from several models in the literature are shown with dotted lines.

Table 5. Best-fit parameters to the differential and cumulative number counts for a Schechter and double power law function.

Function	N_0 (deg^{-2})	S_0 (mJy)	α	β
Differential number counts				
Schechter	2290^{+370}_{-330}	$1.89^{+0.17}_{-0.10}$	$-1.97^{+0.05}_{-0.03}$	
DPL	1800^{+1200}_{-200}	$1.75^{+0.24}_{-0.32}$	$4.02^{+0.22}_{-0.19}$	$2.00^{+0.05}_{-0.07}$
Cumulative number counts				
Schechter	4010^{+170}_{-160}	$1.12^{+0.03}_{-0.02}$	$-0.96^{+0.02}_{-0.03}$	
DPL	2170^{+170}_{-120}	$1.56^{+0.05}_{-0.09}$	$4.00^{+0.09}_{-0.08}$	$1.14^{+0.02}_{-0.03}$

Notes. The uncertainties correspond to the 16% and 84% percentiles.

(Schreiber et al. 2017b; Popping et al. 2020; Lagos et al. 2020; Zavala et al. 2021). These models also predict a more pronounced flattening at the faint end with different degrees of normalization compared to our derived best-fit to the ALMA-based observed 1.1 mm cumulative number counts. The latter is steeper since it includes both the flatter tendency observed by Muñoz Arancibia et al. (2018) and González-López et al. (2019), and the steeper results by Fujimoto et al. (2016). Around the knee the models are overall consistent with our best-fit result, except for the model in Schreiber et al. (2017b) which exhibits a higher normalization. The model by Zavala et al. (2021) also departs from the ALMA-based observations toward higher values as it approaches the brighter flux density regimes. In the bright end the models predict an excess in the cumulative number counts not seeing so far in the observations.

4.4. Cosmic infrared background

The extragalactic background light (EBL) is the integrated intensity of all of the light emitted throughout the history of the universe across the whole of the electromagnetic spectrum. The

EBL constitutes the second most energetic source of background after the cosmic microwave background (CMB). The EBL SED is composed of two main components: the cosmic optical background (COB) and the cosmic infrared background (CIB). While the COB is due to radiation from stars, the CIB comes from the absorption of UV/optical emission that is re-emitted at IR wavelengths by dust. COB and CIB have a similar contribution to the total EBL (e.g., Dole et al. 2006), which indicates that half of the stellar emission in galaxies is absorbed and re-emitted by dust. Millimeter ALMA number counts can resolve 50–100% of the CIB (e.g., Carniani et al. 2015; Fujimoto et al. 2016; González-López et al. 2020).

The surface brightness of the CIB down to a given flux limit $S_\nu^{\lim}$ is given by integrating the differential number counts:

$$I_\nu(S_\nu > S_\nu^{\lim}) = \int_{S_\nu^{\lim}}^{\infty} S_\nu \frac{dN(S_\nu)}{dS_\nu} dS_\nu. \tag{8}$$

We calculated the amount of CIB resolved in this work by integrating the Schechter fit to the differential number counts down to the faintest flux density bin probed in the survey, which corresponds to 0.3 mJy. The result is $0.0289^{+0.0011}_{-0.0006}$ MJy sr^{-1} (where uncertainties correspond to the 16% and 84% percentiles). The Cosmic Background Explorer (COBE) measured the absolute surface brightness of the CIB (e.g., Fixsen et al. 1998; Lagache et al. 1999; Odegard et al. 2019). Using the analytical fit from Fixsen et al. (1998), we calculated a reference absolute value of 0.076 MJy sr^{-1} at 1.13 mm. Therefore, the amount of CIB resolved by the survey is $37.9^{+1.4}_{-0.8}$% down to 0.3 mJy.

5. Source properties

The wealth of ancillary data in the GOODS-South field allows us to study some basic properties of our ALMA sources. Particularly, we are interested in characterizing the redshift and stellar mass distributions.

We looked for stellar counterparts of the 100% pure main and the prior-based supplementary catalogs in the K_s-band selected ZFOURGE catalog by Straatman et al. (2016), which provides photometry and other products including photometric redshift and stellar mass estimations. ZFOURGE (PI: I. Labbé) is a program carried out with the FourStar instrument (Persson et al. 2013) on the 6.5 m *Magellan Baade* Telescope using five near-IR medium bands (J_1, J_2, J_3, H_s, and H_l), covering the same range as classical J and H broadband filters, and a K_s-band. It includes the CDFS field (encompasing GOODS-South), among other fields. ZFOURGE combines dedicated FourStar/K_s-band observations with pre-existing K-band imaging to create super-deep detection images. In the CDFS it incorporates VLT/HAWK-I/K from HUGS (Fontana et al. 2014), VLT/ISAAC/K from GOODS, with ultra deep data in the HUDF region (Retzlaff et al. 2010), CFHST/WIRCAM/K from TENIS (Hsieh et al. 2012), and *Magellan*/PANIC/K in HUDF (PI: I. Labbé). In addition to the dedicated observations, the ancillary CDFS UV to near-IR filters in ZFOURGE include VLT/VIMOS/U, R-imaging (Nonino et al. 2009), HST/ACS/B, V, I, Z-imaging (Giavalisco et al. 2004; Wuyts et al. 2008), ESO/MPG/WFI/U_{38}, V, R_c-imaging (Erben et al. 2005; Hildebrandt et al. 2006), HST/WFC3/F098M, F105W, F125W, F140W, F160W and HST/ACSF606W, F814W-imaging (Grogin et al. 2011; Koekemoer et al. 2011; Windhorst et al. 2011; Brammer et al. 2012), and 11 Subaru/Suprime-Cam optical medium bands (Cardamone et al. 2010). *Spitzer*/IRAC/3.6 and 4.5 μm images are the ultradeep mosaics from the IUDF (Labbé et al. 2015), using data from the IUDF (PI: I. Labbé) and IGOODS (PI: P. Oesch) programs, combined with GOODS (PI: M. Dickinson), ERS (PI: G. Fazio), S-CANDELS (PI: G. Fazio), SEDS (PI: G. Fazio), and UDF2 (PI: R. Bouwens). Mid-IR *Spitzer*/IRAC/5.8 and 8.0 μm images are from GOODS (PI: M. Dickinson).

We also searched for updated spectroscopic redshifts in a recent compilation in GOODS-South (N. Hathi, priv. comm.) and additional recent surveys that have supplied new spectroscopic information in the field: VANDELS (Garilli et al. 2021), the MUSE-Wide survey (Herenz et al. 2017; Urrutia et al. 2019), and ASPECS LP (Decarli et al. 2019; González-López et al. 2019; Boogaard et al. 2019). In Tables 2 and 3 we specify the reference for each spectroscopic redshift. We note that when more than one redshift was available we chose the one with the highest reported quality.

Stellar masses and photometric redshifts were taken from ZFOURGE, except when there were updated spectrocopic redshifts. In that case we calculated the stellar masses using the same methodology as ZFOURGE to keep a consistent analysis. Photometry was fit using FAST++[2], an updated version of the spectral energy distribution (SED) fitting code FAST (Kriek et al. 2009) employed in ZFOURGE. The stellar population models are from Bruzual & Charlot (2003, BC03), with exponentially declining star formation histories (SFHs), a Calzetti et al. (2000) dust attenuation law, and fixed solar metalicity. FAST++ input files had the same parameters and grid of models as in ZFOURGE. The stellar masses were multiplied by a factor of 1.7 to scale them from a Chabrier (Chabrier 2003) to a Salpeter (Salpeter 1955) IMF (e.g., Reddy et al. 2006; Santini et al. 2012; Elbaz et al. 2018). We note that the stellar masses are dependent on the methodology and the assumptions made in the SED fitting. In the literature some studies have tested the impact of different codes and SFHs on the stellar masses of DSFGs

(e.g., Hainline et al. 2011; Michałowski et al. 2014; Simpson et al. 2020). Michałowski et al. (2014) reported that exponentially declining models with a code which does not assume energy balance, as employed here, were able to recover the stellar masses of their simulated submillimeter galaxies (SMGs), albeit a slight underestimation (∼0.05 dex) and significant scatter. Conversely, alternative approaches using a code assuming an energy balance between the UV emission absorbed and radiated at far-IR and mm wavelengths resulted in a systematic over-estimation (∼0.1 dex) of the stellar masses of their simulated SMGs.

Additionally, we substituted the stellar masses and redshifts of the six optically dark sources studied in detail in Zhou et al. (2020), namely AGS4, AGS11, AGS15, AGS17, AGS24, and AGS25 (A2GS2, A2GS15, A2GS10, A2GS7, A2GS29, and A2GS17, respectively). Zhou et al. (2020) reported spectroscopic confirmation for AGS4 (A2GS2) and AGS17 (A2GS7) and argued for AGS11, AGS15, and AGS24 to be at the median redshift of a $z \sim 3.5$ overdensity. Although independent spectroscopic confirmation is still needed for these sources, we used the assumed redshifts in Zhou et al. (2020) for them. For these sources, Zhou et al. (2020) used as well the same methodology as ZFOURGE to keep a consistent analysis (i.e., photometry was fit using FAST++, BC03 stellar population modes, exponentially declining SFHs, and Calzetti et al. 2000 dust attenuation law).

After visual inspection of all the ALMA sources in comparison with the K_s-band image, there are five of them with blending issues in the ZFOURGE catalog. It leads to a smaller number of catalog entries than actual sources and, thus, the photometry is affected and requires an improved tailored analysis. One of them is AGS24 in Zhou et al. (2020, A2GS29), who already solved the blending issue and we used their stellar masses and redshifts. The other four are A2GS28, A2SGS30, A2GS33, and A2GS60. An extra source, A2GS38 does not have a counterpart in the ZFOURGE catalog and also required photometry, in this case without blending from close neighbors.

For the sources above that required an improved tailored analysis, we carried out the photometry following the methodology described in Gómez-Guijarro et al. (2018) for crowded and blended sources. Since we used ZFOURGE there are three types of datasets: HST bands that provide the best spatial resolution, ground-based bands homogenized to a common Moffat PSF profile (0.″9 FWHM; see Straatman et al. 2016), and *Spitzer*/IRAC bands which are the ones with a coarser PSF and more affected by blending. The *Hubble* Legacy Fields (HLF) v2.0 for the GOODS-South region (HLF-GOODS-S) includes all ultraviolet, optical, and IR data taken to date by HST over 14 years across the field (Illingworth et al. 2016; Whitaker et al. 2019). Therefore, instead of the ZFOURGE HST data, we used the updated HLF-GOODS-S v2.0 mosaics homogenized to the WFC3/F160W-band PSF, along with the ground-based and *Spitzer*/IRAC bands. Briefly, the photometry methodology is as follows: all the bands affected by blending were fit with a model using GALFIT (Peng et al. 2002), where the number of priors is set to the number of sources in the F160W-band image. While for those bands unaffected by blending, we performed aperture photometry with aperture diameters 0.″7 for HST (as in Whitaker et al. 2019) and 1.″2 for the ground-based (as in Straatman et al. 2016), along with aperture corrections derived by tracing the PSF growth curves to account for the flux losses outside the aperture, plus also PSF photometry with GALFIT for *Spitzer*/IRAC bands. Uncertainties were derived from empty aperture measurements.
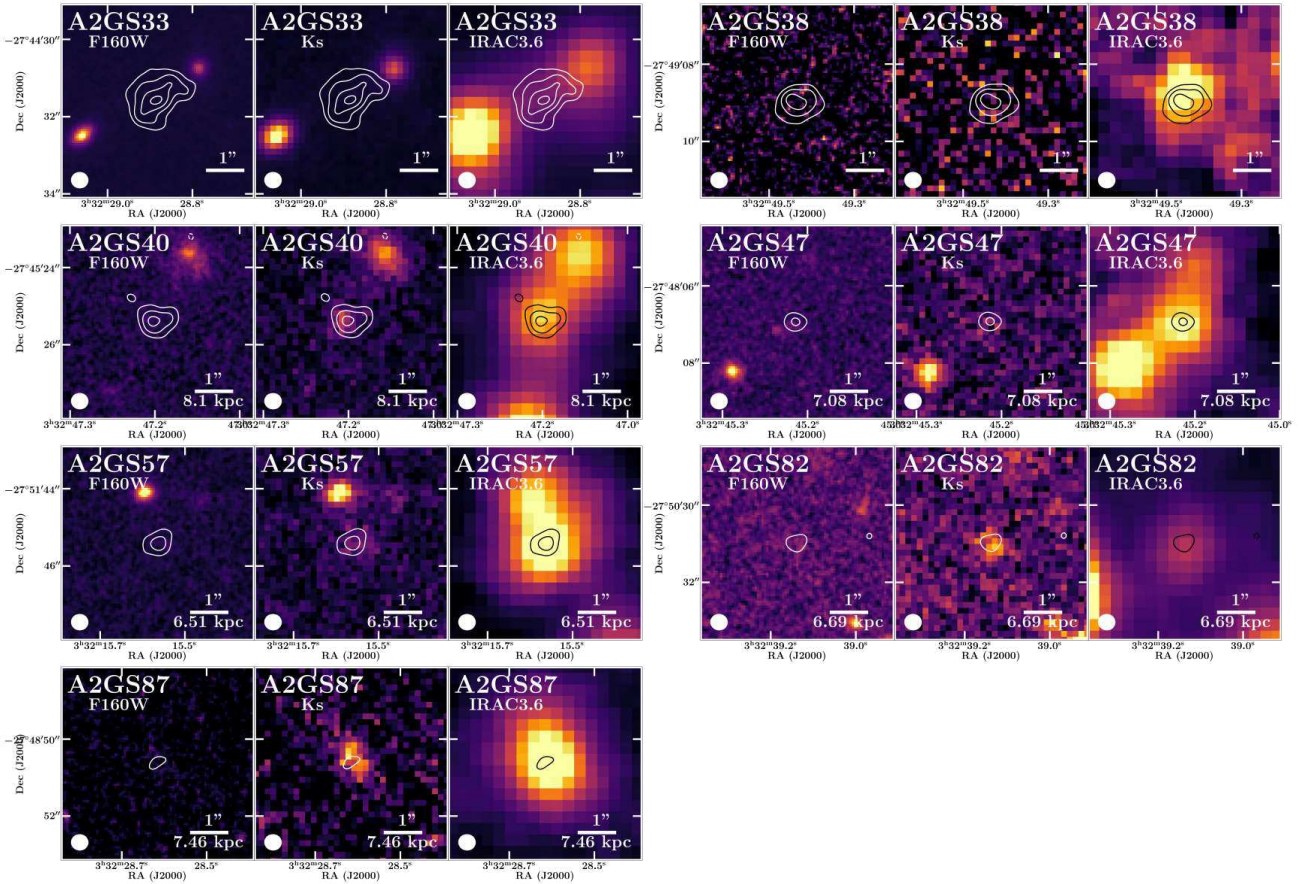
---

[2] https://github.com/cschreib/fastpp

Fig. 12. Optically dark/faint galaxies. $5'' \times 5''$ HST/WFC3 $F160W$, ZFOURGE K_s, and *Spitzer*/IRAC 3.6 μm images with ALMA 1.1 mm contours overlaid in white (starting at $\pm 3\sigma$ and growing in steps of $\pm 1\sigma$, where positive contours are solid and negative contours dotted). North is up, east is to the left.

5.1. Optically dark sources

In the last years a new population of galaxies missed in optical surveys but bright at far-IR/mm wavelengths has been discovered. This type of DSFGs differs from previously known intense starbursts (e.g., Walter et al. 2012; Riechers et al. 2013; Marrone et al. 2018) as they have lower SFRs characteristic of MS SFGs, rather than starbursts. Their space density is two orders of magnitude greater than equally massive $z \sim 3$−4 SFGs (Wang et al. 2019). They are of great interest as they are seen as a key population of galaxies that dominate the contribution of massive ($M_* > 10^{10.3}\,M_\odot$) galaxies to the SFR density of the universe at $z > 3$ (Wang et al. 2019). These optically-invisible massive galaxies (also known as HST-dark galaxies) are currently undetected or very faint in all optical and near-IR bands up to and including the H-band (H-band dropouts) in the deepest cosmological fields ($H > 27$ mag; 5σ point source), but bright at longer near-IR bands ([4.5] < 24 mag) (Wang et al. 2016, 2019; Alcalde Pampliega et al. 2019). It should be noted, as is not always the case, that the selection and characterization of this population of galaxies depends on the depth of the observations. For a fixed [4.5] magnitude, it is particularly important to know the depth at which they remain undetected or very faint in the H-band, as their $z > 3$ and massive nature relies in their red H−[4.5] color (e.g., Wang et al. 2019). DSFGs surveys with ALMA combined with shallower H-band observations hinted for this optically dark galaxies (e.g., Simpson et al. 2014, $H > 23$ mag; 3σ), although their red H−[4.5] color remained

uncertain in the absence of deeper H-band observations. More examples of this galaxy population continue to be discovered in new surveys (e.g., Franco et al. 2018, 2020b; Yamaguchi et al. 2019; Williams et al. 2019; Romano et al. 2020; Toba et al. 2020; Umehata et al. 2020; Gruppioni et al. 2020; Smail et al. 2021; Fudamoto et al. 2021).

In GOODS-ALMA, Franco et al. (2018, 2020b) already reported six of these galaxies (AGS4, AGS11, AGS15, AGS17, AGS24, and AGS25, which correspond to A2GS2, A2GS15, A2GS10, A2GS7, A2GS29, and A2GS17, respectively). In this work, there exist some additional sources without or with very faint emission at bands up to and including the H-band (H-band dropouts), namely: A2GS40, A2GS47, A2GS57, A2GS82, and A2GS87 (see Fig. 12). Furthermore, A2GS33, after subtraction of the $F160W$-band neighbors, shows no emission in the K_s-band image at the position of the ALMA source. We performed aperture photometry in the residual image, confirming no significant ($<3\sigma$) emission in all bands up to and including the K_s-band (K_s-band dropouts). A2GS33 is also a candidate *Spitzer*/IRAC 3.6 μm dropout (see Fig. 12), although its emission is highly contaminated by blending with close neighbors in this band. In addition, A2GS38, which did not have a ZFOURGE counterpart at the start, is another K_s-band dropout, but detected in *Spitzer*/IRAC 3.6 μm (see Fig. 12). A2GS33 and A2GS38 coincide respectively with ID 20 and 17 in Yamaguchi et al. (2019), who also reported them as K-band dropouts.

Therefore, the total number of optically dark/faint sources uncovered so far in GOODS-ALMA is 13 (ALMA detected
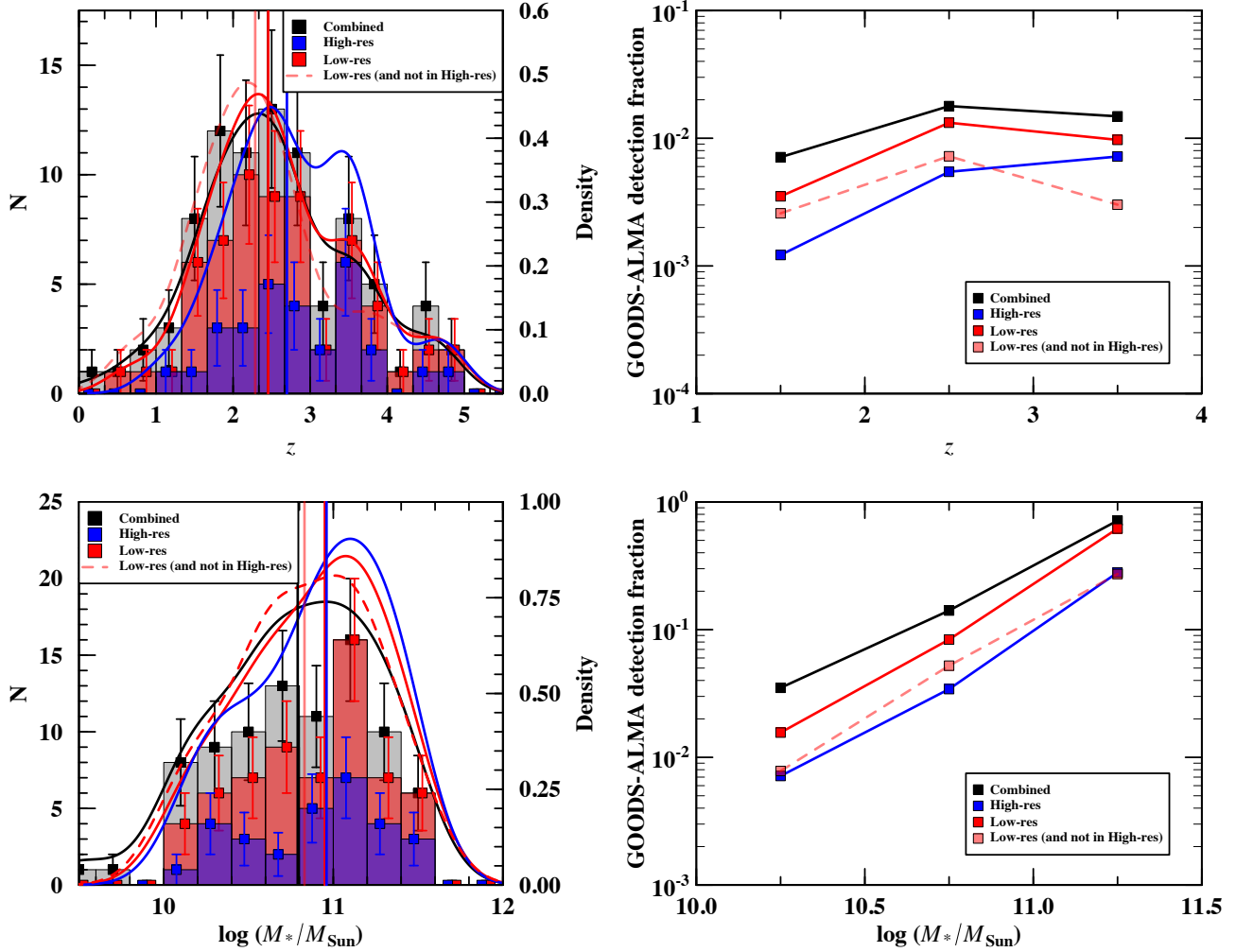
Fig. 13. *Top left panel*: redshift distribution. We show histograms with Poisson error bars and probability density curves (kernel density estimates, by definition normalized to an area under the curve equal to one). We note that the histograms are overlaid, not stacked. Medians are displayed as a solid vertical line. The distribution was constructed using the values in Tables 2 and 3, which represent the best redshift estimate for each source from a spectroscopic or photometric origin. *Top right panel*: detection fraction of sources in GOODS-ALMA 2.0 compared to all the galaxies in ZFOURGE located within the same area as a function of redshift for galaxies with stellar masses $10.0 < \log(M_*/M_\odot) < 11.5$. *Bottom left panel*: similar to the top left panel but in terms of the stellar mass. *Bottom right panel*: similar to the top right panel but in terms of the stellar mass for galaxies with redshift $1 < z < 4$. In all panels, we represent the whole 100% pure plus prior-based catalogs from the combined dataset (black), the sources detected in the high resolution (blue) and low resolution (red) datasets, along with the sources that appear in the low resolution but do not in the high resolution dataset (dashed red). We note that for simplicity, in the case of the sources that appear in the low resolution but do not in the high resolution dataset the histograms are omitted and only the probability density curves are displayed.

H- or *K*-band dropouts). In particular, Zhou et al. (2020) analyzed in detail the six optically dark sources in Franco et al. (2018, 2020b) and found that almost all of them (4/6) are associated to the same $z \sim 3.5$ structure. In fact, along with the latter sources we detected around a dozen more ALMA sources potentially associated with the same structure (see the southwest region of the GOODS-ALMA 2.0 map in Fig. 1), with some of them located along two streams that connect at the center of the structure (pinpointed by AGS24/A2GS29). Spectroscopic confirmation is still needed to confirm this hypothesis. A detailed analysis of the optically dark/faint sources in this work and their potential link to overdense structures is beyond the scope of this paper.

5.2. Redshift and stellar mass distributions

In Fig. 13 we present the redshift distribution along with the detection fraction of sources in GOODS-ALMA 2.0 compared

to all the galaxies in ZFOURGE located within the same area as a function of redshift. The redshift distribution was constructed using the values in Tables 2 and 3, which represent the best redshift estimate for each source from a spectroscopic or photometric origin. The different datasets studied in this work are represented: combined, high resolution, and low resolution datasets, and also those sources that appear in the low resolution but do not in the high resolution dataset.

First, we see that the high resolution dataset redshift distribution is skewed toward higher redshifts compared to that of the low resolution dataset, which has a similar distribution compared to the combined dataset. In fact, the dashed red line that represents the sources that appear in the low resolution but do not in the high resolution dataset reflects the difference between the high resolution and low resolution datasets clearly as it is skewed toward lower redshifts. The difference between the two datasets is also clear in the detection fraction, with the high resolution

dataset being efficient in picking up sources at higher redshifts, while the low resolution dataset achieves a higher detection fraction and exhibits a similar shape as the combined dataset across redshift. The combined dataset reaches naturally a higher detection fraction than the low resolution dataset since it is a deeper map.

The median redshift of the GOODS-ALMA 2.0 survey from the combined dataset using the 100% pure plus prior-based catalogs is $z_{med} = 2.46$. This value is in line with literature studies of DSFGs peaking at $z = 2-3$ (e.g., Chapman et al. 2005; Yun et al. 2012; Smolčić et al. 2012; Dudzevičiūtė et al. 2020). Among other literature studies at ~1 mm in GOODS-South, Dunlop et al. (2017) reported a lower mean redshift of $z = 2.15$ in a ×2.0 deeper 1.3 mm survey (average sensitivity of 35 μJy beam^{-1} at an average angular resolution of 0.″7), while covering a ×16 smaller area (4.5 arcmin^2). Yamaguchi et al. (2020) reported a slightly lower median redshift of $z = 2.38$ in a ×1.0−2.3 deeper 1.2 mm survey (average sensitivity of 30−70 μJy beam^{-1} at an average angular resolution of 0.″59×0.″53), while covering a ×2.8 smaller area (26 arcmin^2). Aravena et al. (2020) reported a lower redshift, yielding a median value of $z = 1.8$ in a ×7.4 deeper 1.2 mm survey (average sensitivity of 9.3 μJy beam^{-1} at an average angular resolution of 1.″53 × 1.″08), while covering a ×15 smaller area of 5 arcmin^2. These results are in line with the idea that shallower and larger-area surveys are better at detecting brighter sources at higher redshifts, while deeper and smaller-area surveys access fainter sources at lower redshifts and, therefore, the redshift distribution of DSFGs is dependent on the survey depth (e.g., Ivison et al. 2007; Béthermin et al. 2015; Aravena et al. 2020).

A second peak appears in the redshift distribution for all datasets at $3 < z < 4$ due to the $z \sim 3.5$ overdensity of sources reported in Zhou et al. (2020). This second peak is more prominent in the high resolution dataset where these source are detected. We note that although the histogram of the high resolution dataset shows a higher peak at the location of the overdense structure (all sources are located in a single histogram bin), the probability density curve shows that the main peak is that at $z = 2-3$ (sources across various histogram bins add up and account for a higher percentage of the total number of sources), consistent with the other datasets involved in this work.

Similarly, in Fig. 13 we present the stellar mass distribution along with the detection fraction for the different datasets studied. The differences here are more subtle and specially show up in the dashed red line of sources that appear in the low resolution but do not in the high resolution dataset. Sources missed by the high resolution are skewed to lower stellar masses. In terms of the detection fraction we see again that the low resolution dataset achieves a higher detection fraction than the high resolution dataset and the deeper combined dataset reaches a higher detection fraction than the low resolution dataset. In particular, the fraction of detected sources at $\log(M_*/M_\odot) > 11.0$ in the low resolution dataset is very similar to that of the combined dataset. In other words, the newly incorporated sources in the combined dataset are more in the lower stellar mass regime.

Complementing the redshift and stellar mass characterization, Fig. 14 shows the stellar mass as a function of redshift for the sources in the 100% pure plus prior-based catalogs from the combined dataset. We distinguish between sources extracted in the high resolution dataset (median redshift and stellar masses: $z = 2.7\pm1.1$, $\log(M_*/M_\odot) = 10.96\pm0.46$, where the uncertainties are given by the median absolute deviation), sources that appear in the low resolution but do not in the high resolution dataset ($z = 2.29 \pm 0.73$, $\log(M_*/M_\odot) = 10.83 \pm 0.43$), and also sources that
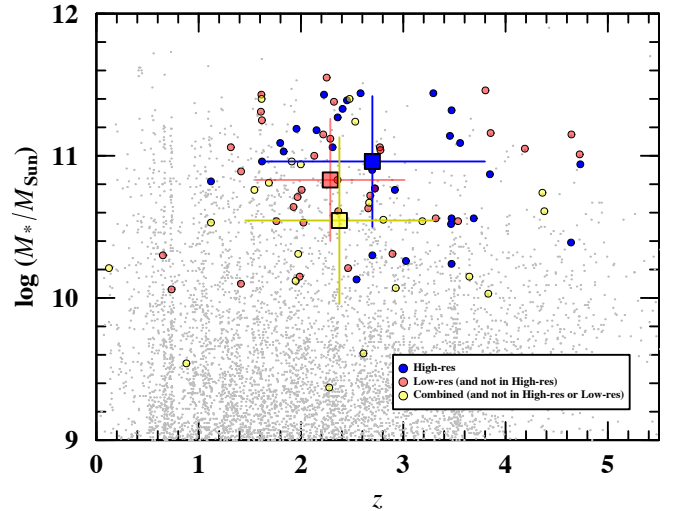

Fig. 14. Stellar mass versus redshift for the sources in the 100% pure plus prior-based catalogs from the combined dataset, distinguishing between sources that appear in the high resolution dataset (blue), in the low resolution but not in the high resolution dataset (red), and in the combined but not in the high resolution or low resolution datasets (yellow). Medians are shown as big squares, where the uncertainties are given by the median absolute deviation.

appear in the combined but do not in the high resolution or low resolution datasets ($z = 2.38\pm0.92$, $\log(M_*/M_\odot) = 10.55\pm0.59$). Therefore, we see that the sources missed by the high resolution that are in the low resolution dataset are skewed to lower redshifts and stellar masses, and the sources missed by these two datasets individually and retrieved in the combined dataset are skewed to lower stellar masses. We note that the median redshift of the high resolution dataset becomes slightly lower if the sources in the $z \sim 3.5$ overdensity are not taken into account ($z = 2.56 \pm 0.65$), but even in this case the sources missed by the high resolution that are in the low resolution dataset are still skewed to lower redshifts.

6. Discussion

6.1. ALMA array configuration impact on source detection

ALMA has opened the possibility of resolving dust continuum emission providing size estimates of the emitting regions. However, a very important concern in studies to date is the detection and measurement of accurate fluxes and sizes of sources spanning a wide range of intrinsic properties. Single array configurations providing angular resolution sufficient to measure sizes of intrinsically compact sources could be missing more extended sources for which a coarser angular resolution would be better suited.

A given array configuration yielding an angular resolution smaller than the intrinsic source size limits the survey depth to avoid false detections resulting from the excessively large number of independent beams. This is directly related with the concept of purity as explained in Sect. 3.1.1. A survey with a large number of independent beams leads to a high detection S/N^{peak} required for a source to be regarded as real and, thus, effectively limits the survey depth. In addition, completeness decreases as a function of the source size as described in Fig. 8 and, thus, such array configuration providing high angular resolution is worse suited for increasingly larger source sizes. Besides, this type of high resolution array configuration could not properly

account for the flux coming from larger spatial scales leading to potential flux losses. Tappering techniques can be used to mitigate the aforementioned purity and completeness issues, but they come at the expense of the survey depth. Combining high resolution and low resolution array configurations improves the purity by reducing the number of independent beams, which avoids compromising the survey depth to mitigate false detections and improves completeness with minimum flux biases in a wider range of intrinsic source properties.

Our analysis regarding source sizes in Sect. 3.2.1 concludes that dust continuum emission occurs in compact regions, in line with a variety of literature studies in the ALMA era (e.g., Simpson et al. 2015a; Ikarashi et al. 2015; Hodge et al. 2016; Fujimoto et al. 2017, 2018; Gómez-Guijarro et al. 2018; Elbaz et al. 2018; Franco et al. 2018; Rujopakarn et al. 2019; Gullberg et al. 2019). One remaining question to address is the reason why the sources that appear in the low resolution but do not in the high resolution, although skewed to slightly larger sizes and with larger scatter as shown in Fig. 6, did not appear originally in the high resolution dataset being overall compact sources. The most relevant property that distinguish them are the flux densities. The number counts analysis in Sect. 4.3 reflects that the sources detected in the high resolution dataset are brighter than those that appear in the low resolution but do not in the high resolution dataset. The answer to this question comes jointly from purity and completeness as outlined above. In terms of purity, a compact source that appears in the low resolution but do not in the high resolution dataset, being fainter has a lower detection S/N^{peak} compared to another similarly compact source with a higher flux density. These sources have a higher purity in the low resolution compared to that in the high resolution dataset. This effect is seen in the resulting σ_{p} (directly related to S/N^{peak}) for a purity of $p = 1$ found to be $\sigma_{\text{p}} \geq 4.2$ in the low resolution map and $\sigma_{\text{p}} \geq 5.2$ in the high resolution map. Besides, the Tables 2 and 3 are ordered with decreasing detection S/N^{peak}. As we move down the table the sources are no longer detected in the high resolution dataset (see Table 2, Col. 10). In terms of completeness, a compact source that appear in the low resolution but do not in the high resolution dataset, being fainter has a lower completeness compared to another similarly compact source with a higher flux density. These sources have a higher completeness in the low resolution compared to that in the high resolution dataset. Both to mitigate purity and completeness issues, this is the reason why in Franco et al. (2018) a tapering technique was applied at the expense of lowering the survey sensitivity to an average of 182 μJy beam^{-1} at an angular resolution of $0\farcs614 \times 0\farcs587$.

6.2. Conversion of angular into physical sizes: redshift and stellar mass dependency

It is well known that the stellar sizes measured at optical wavelengths vary with redshift and stellar mass for both early and late-type galaxies (e.g., van der Wel et al. 2014). Galaxies are smaller with increasing redshift at fixed stellar mass and larger with increasing stellar mass at fixed redshift. Therefore, in order to fairly compare galaxy sizes they need to be expressed in the same terms of redshift and stellar mass. We could correct the dust continuum sizes of each source to a common redshift and stellar mass by using the $R_{\text{e}}(z, M_*)$ dependency of late-type galaxies of van der Wel et al. (2014). However, we first need to verify the hypothesis of whether dust continuum sizes also vary in terms of redshift and stellar mass resembling the behavior known for the stellar sizes measured at optical wavelengths.

In order to verify the aforementioned assumption, we split the source catalog in four bins according to whether the redshifts and stellar masses are above or below the median values of the whole 100% pure main plus prior-based supplementary catalogs ($z_{\text{med}} = 2.46$ and $\log(M_{*\text{med}}/M_\odot) = 10.79$). We stacked the sources in each bin, measuring the dust continuum size of the stack. This approach has the advantage of higher number statistics, although the ALMA centroids become less constrained for increasingly lower detection S/N^{peak} and, thus, potentially leading to an artificial increase of the stacked size (e.g., Fujimoto et al. 2017). Alternatively, we also performed the stacks using the *Spitzer*/IRAC 3.6 counterparts, detected at much higher S/N and, thus, with better constrained centroids. The stacking was performed in the uv plane as follows: for each source pair of coordinates to be stacked we searched for all the pointings that contained data on that source (each source is covered by six pointings typically). Next, each pointing was phase shifted to set the phase center at the source coordinates using the CASA task `fixvis`. Data and weights are modified to apply the appropriate primary beam correction that correspond to the phase shift with the CASA toolkit `MeasurementSet` module. After that, by using the CASA task `fixplanets` the phase center was set to a common pair of coordinates $\alpha = 00\,\text{h}\,00\,\text{m}\,00\,\text{s}$; $\delta = 00\,\text{d}\,00\,\text{m}\,00\,\text{s}$ (J2000) for all the pointings to be stacked and the visibility weights recomputed with the task `statwt`. Finally, the pointings were concatenated into a single measurement set, where the weight of each source was normalized by its flux density. We measured the sizes of the stacks as explained in Sect. 3.2 employing the CASA task `UVMODELFIT` fitting a Gaussian model with fixed circular axis ratio. There are no significant differences in the stacked dust continuum sizes when ALMA or *Spitzer*/IRAC 3.6 centroids are used. Finally, we quantified the $R_{\text{e}}(z, M_*)$ dependency to be compared with that for the stellar sizes measured at optical wavelengths by fitting the dust continuum sizes measured for the stacks in each bin using the expressions:

$$R_{\text{e}} = A\left(M_*/5 \times 10^{10}\,M_\odot\right)^\alpha, \tag{9}$$

where A is a normalization constant given in kpc and α expresses the $R_{\text{e}} - M_*$ dependency, and

$$R_{\text{e}} = B(1 + z)^\beta, \tag{10}$$

where B is a normalization constant given in kpc and β expresses the $R_{\text{e}} - z$ dependency.

In Table 6 we present the dust continuum sizes measured for the stacks in each bin. The results of the fits are shown in Table 7. In Fig. 15 we show the $R_{\text{e}} - M_*$ and $R_{\text{e}} - z$ planes displaying the sources in the 100% pure main catalog, the dust continuum sizes measured for the stacks, and the fits. Along with these, we display the stellar size measured at optical wavelengths evolution with redshift and stellar mass for both early and late-type galaxies from van der Wel et al. (2014) at the median values of the source catalog ($z_{\text{med}} = 2.46$ and $\log(M_{*\text{med}}/M_\odot) = 10.79$) for comparison.

Dust continuum sizes appear to be smaller with increasing redshift at fixed stellar mass and larger with increasing stellar mass at fixed redshift. Overall, the $R_{\text{e}} - M_*$ dependency given by α resembles that of the stellar sizes measured at optical wavelengths of late-type galaxies in (e.g., van der Wel et al. 2014), albeit a lower normalization constant. The $R_{\text{e}} - z$ dependency given by β resembles as well that of the stellar sizes measured at optical wavelengths of late-type galaxies in (e.g., van der Wel et al. 2014), with a lower normalization constant. Therefore, we

Table 6. Dust continuum sizes at 1.1 mm as measured for stacks of sources in bins of redshift and stellar mass.

	Number	R_e
$z < z_{med}, M_* < M_{*med}$	17	$0\rlap{.}''103 \pm 0\rlap{.}''008$ (0.87 $\pm$ 0.07 kpc)
$z > z_{med}, M_* < M_{*med}$	26	$0\rlap{.}''095 \pm 0\rlap{.}''006$ (0.73 $\pm$ 0.05 kpc)
$z < z_{med}, M_* > M_{*med}$	26	$0\rlap{.}''127 \pm 0\rlap{.}''006$ (1.06 $\pm$ 0.05 kpc)
$z > z_{med}, M_* > M_{*med}$	17	$0\rlap{.}''117 \pm 0\rlap{.}''007$ (0.86 $\pm$ 0.05 kpc)

Notes. The columns show the size measurements for the stacks of sources in four redshift and stellar mass bins according to their location relative to the median values of the source catalog (z_{med} = 2.46 and $\log(M_{*med}/M_\odot)$ = 10.79). The number of sources in each bin is also indicated. The physical dust continuum sizes in kpc are calculated at the median redshift and stellar mass of each bin.

Table 7. Redshift and stellar mass dependency of dust continuum sizes at 1.1 mm.

	$\log A$	α
$R_e - M_*$ ($z < z_{med}$)	-0.020 ± 0.028	0.106 ± 0.049
$R_e - M_*$ ($z > z_{med}$)	-0.117 ± 0.026	0.113 ± 0.062
	$\log B$	β
$R_e - z$ ($M_* < M_{*med}$)	0.194 ± 0.030	-0.55 ± 0.33
$R_e - z$ ($M_* > M_{*med}$)	0.274 ± 0.024	-0.52 ± 0.19

Notes. $R_e - M_*$ at fixed redshift and $R_e - z_*$ at fixed stellar mass fit to Eqs. (9) and (10), respectively. Results are given by fitting the sizes measured for stacks of sources in bins of redshift and stellar mass shown in Table 6.

verified the hypothesis of whether dust continuum sizes vary in terms of redshift and stellar mass resembling the behavior known for the stellar sizes measured at optical wavelengths. At the moment the relatively small number of sources in this work for this type of study does not allow us to better constrain α and β or to group the sources within narrower redshift and stellar mass bins to explore the $R_e(z, M_*)$ dependency for dust continuum size measured at 1.1 mm more accurately. Consequently, to correct the dust continuum sizes of each source to a common redshift and stellar mass, we employed the $R_e(z, M_*)$ dependency of late-type galaxies of van der Wel et al. (2014).

The lower normalization values of dust continuum sizes compared to those of the stellar sizes of late-type galaxies better resemble those of early-type galaxies. This fact reinforces the idea of the stellar mass-size place locus as a proof of DSFGs being progenitors of massive elliptical galaxies (e.g., Toft et al. 2014; Barro et al. 2016; Gómez-Guijarro et al. 2018; Tadaki et al. 2020; Franco et al. 2020a) and, reflected in the compact sizes of various far-IR to radio tracers, the build-up of central stellar cores prior to the quenching of star formation (e.g., Barro et al. 2017; Jiménez-Andrade et al. 2019; Gómez-Guijarro et al. 2019; Puglisi et al. 2021; Suess et al. 2021).

In the literature Fujimoto et al. (2017) studied dust continuum sizes at 1.1 mm for a large compilation of 1627 massive SFGs observed with ALMA, finding that dust continuum sizes are more compact than those at UV/optical wavelengths, although by a factor somewhat smaller than the factor 3−4 more compact than the typical sizes of stellar disks for late-type galaxies at UV/optical wavelengths we find. Fujimoto et al. (2017) also argued that dust continuum sizes at 1.1 mm follow a similar evolutionary trend with redshift than the stellar sizes of late-type galaxies. Jiménez-Andrade et al. (2019) studied radio sizes at 3 GHz for a mass-complete sample of 3184

radio-selected SFGs and found a flatter evolutionary trend with redshift $\beta = -0.26 \pm 0.08$ (0.12 ± 0.14) for galaxies on (above) the MS of SFGs, but no clear variation of radio sizes with stellar mass. Using 3 GHz and 6 GHz radio emission Jiménez-Andrade et al. (2021) found radio emission more compact by a factor 2−3 than UV/optical sizes and also no variation of radio sizes with stellar mass.

6.3. The systematicy of compactness in DSFGs

Our results conclude that dust continuum emission occurs in compact regions. However, GOODS-ALMA 2.0 is a flux limited survey and we need to understand the extend of this conclusion in terms of the flux density completeness limits. In addition, the sources that appear in the low resolution but do not in the high resolution, although compact, are skewed to slightly larger sizes and with larger scatter as shown in Fig. 6. These sources are mainly characterized by lower flux densities, but are also located at slightly lower redshifts and stellar masses.

In order to be able to fairly compare galaxy sizes across flux densities, they need to be expressed in the same terms of redshift and stellar mass as explained above, where we verified that dust continuum sizes evolve with redshift and stellar mass resembling the trends of the stellar sizes measured at optical wavelengths, albeit a lower normalization compared to those of late-type galaxies. Therefore, we corrected the dust continuum sizes of each source to a common redshift and stellar mass as given by the median values of the source catalog (z_{med} = 2.46 and $\log(M_{*med}/M_\odot)$ = 10.79) by using the $R_e(z, M_*)$ dependency of late-type galaxies of van der Wel et al. (2014). In Fig. 16 we show the corrected dust continuum sizes at 1.1 mm as a function of the 1.1 mm flux densities for the sources in the 100% pure plus prior-based catalogs from the combined dataset. We also display the sources in the prior-based supplementary catalog even if their sizes are unreliable owing to their low S/N as we do know their fluxes and, thus, while their position in the y-axis is uncertain, their location in the x-axis is well constrained. We clearly see that the upper-right quadrant of the plane is empty.

In Fig. 16 we also show the typical size of the stellar distribution measured at optical wavelengths for both early and late-type galaxies from van der Wel et al. (2014) evaluated at the common redshift and stellar mass (z_{med} = 2.46 and $\log(M_{*med}/M_\odot)$ = 10.79). Sources with $S_{1.1\,mm} > 1$ mJy are always below the typical size of star-forming stellar disks with a median corrected size of $R_e^{cor} = 0.72 \pm 0.03$ kpc ($R_e = 0\rlap{.}''094 \pm 0\rlap{.}''004$), meaning that there is no dust continuum emission as extended as typical star-forming stellar disks in the $S_{1.1\,mm} > 1$ mJy regime. At most only one source is consistent within the scatter of the typical size of star-forming stellar disks, namely AGS17 (A2GS7). However, this galaxy exhibits signs of being a merger with a double peak dust continuum emission (see Fig. A.1). It was also reported as an extended source in the analysis of Franco et al. (2018), where the extraction was carried out in the high resolution dataset using a tapered map with a homogeneous and circular synthesized beam of $0\rlap{.}''6$ FWHM under the assumption that the sources were point-like at that angular resolution. AGS17 was reported as one of the outliers that did not meet the point-like criteria.

In terms of completeness, in Fig. 16 we also show the ~100% completeness region (blank area without the line grid) drawn from the simulations for sources up to $1''$ FWHM presented in Sect. 4.1. This angular size corresponds to a physical size of R_e = 4.05 kpc at the common redshift and stellar mass
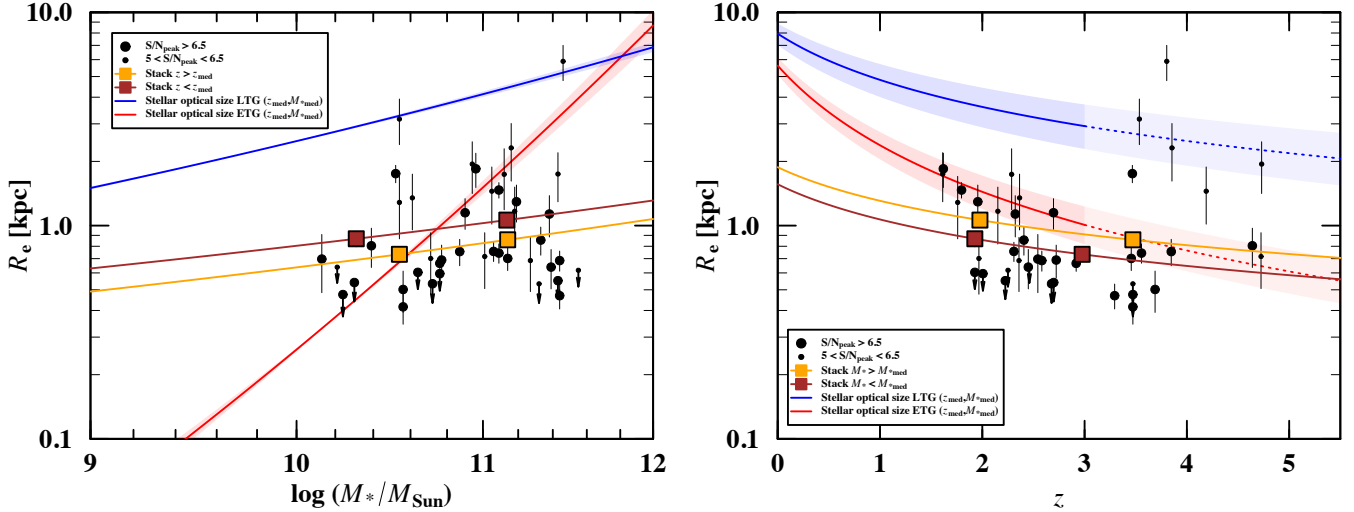
Fig. 15. $R_e - M_*$ (*left panel*) and $R_e - z$ (*right panel*) planes. Dust continuum sizes at 1.1 mm as measured for stacks of sources in bins of redshift and stellar mass and the associated fits shown in Tables 6 and 7 are displayed. The dust continuum sizes measured for individual sources in the 100% pure main catalog are also shown distinguishing sources with a detection $S/N^{peak} > 6.5$ (big black circles, representing approximately the top third of the source catalog) and sources with a detection $5 < S/N^{peak} < 6.5$ (small black circles). For comparison, the stellar size measured at optical wavelengths evolution with redshift and stellar mass for both early and late-type galaxies from van der Wel et al. (2014) at the median values of the source catalog ($z_{med} = 2.46$ and $\log(M_{*med}/M_{\odot}) = 10.79$) are also shown with their scatter displayed as a shaded areas (at $z > 3$ extrapolations of the van der Wel et al. 2014 evolutionary trends are shown as dotted lines).

($z_{med} = 2.46$ and $\log(M_{*med}/M_{\odot}) = 10.79$). We are ~100% complete for sources with $S_{1.1mm} > 1$ mJy covering the scatter of the typical size of star-forming stellar disks. Therefore, we conclude that compact dust continuum emission at $S_{1.1mm} > 1$ mJy prevails and that dust continuum emission as extended as typical star-forming stellar disks at these flux densities is rare (at most we only detected one source in the entire 72.42 arcmin² area of the survey, if we consider AGS17 as such type of source).

At lower flux densities in the $S_{1.1mm} < 1$ mJy regime the picture is less clear. In this flux regime we have a mix of sources in the 100% pure main catalog with reliable size measurements, plus all sources in the prior-based supplementary catalog with unreliable size measurements. The median corrected size of the $S_{1.1mm} < 1$ mJy sources in the 100% pure main catalog is $R_e^{cor} = 0.92 \pm 0.06$ kpc ($R_e = 0''.138 \pm 0''.009$). This value appears slightly more extended compared to that of the $S_{1.1mm} > 1$ mJy sources. In order to confirm the slight size difference between both flux density regimes we stacked the sources at $S_{1.1mm} > 1$ mJy and at $S_{1.1mm} < 1$ mJy and measured the dust continuum size of the stacks as explain in Sect. 6.2. $S_{1.1mm} > 1$ mJy sources result in $R_e^{cor} = 0.80 \pm 0.05$ kpc and $S_{1.1mm} < 1$ mJy sources result in $R_e^{cor} = 0.97 \pm 0.05$ kpc, confirming the slight size difference in these flux density regimes.

It is clear is that, beyond any slight difference between flux density regimes, even in the $S_{1.1mm} < 1$ mJy regime most of the sources appear compact below the sizes of typical star-forming stellar disks as in the $S_{1.1mm} > 1$ mJy regime. The most distinct characteristic of the $S_{1.1mm} < 1$ mJy regime is the larger scatter with some of the sources possibly entering the scatter of the typical size of star-forming stellar disks. However, the lower S/N could lead to systematically larger sizes due to an artificial bias at this S/N and larger samples with higher S/N would be required for confirmation. We note also that in the regime of $S_{1.1mm} < 1$ mJy, and specially for larger sources, the completeness drops (see Fig. 8) and further data would be also required to know the abundance of these sources. Naturally, there should be a regime where the millimeter observations get deep enough

to start detecting secular star formation in regular star forming disks, with their associated extended dust continuum disks. In the literature some studies have suggested that fainter millimeter sources could be physically larger. Franco et al. (2020b) argued that the prior-based methodology allowed for one to access a population of fainter and slightly larger sources. The ALMA Frontier Fields work by González-López et al. (2017) at 1.1 mm probed the sub-mJy regime and found similar results to our work. The lensing-corrected sizes were in the same range as those measured in brighter samples, with possibly larger dispersion. 2/3 of the sources had sizes a factor ×1.6 larger that the brighter sources and suggested that a substantial portion of the sub-mJy sources may be mildly more extended than brighter ones. Rujopakarn et al. (2016) reported more extended galaxy-wide dust continuum emission at 1.1 mm in a sample of 11 normal MS SFGs, Cheng et al. (2020) also found extended dust continuum emission at 870 μm in four DSFGs, mostly with mild IR luminosities $\log(L_{IR}/L_{\odot}) < 12.0$, and Sun et al. (2021) also found extended dust continuum emission at 1.3 μm in two DSFGs with mild IR surface brightness associated to less vigorous star formation.

7. Summary and conclusions

The GOODS-ALMA survey is a 1.1 mm galaxy survey carried out with ALMA in the GOODS-South field. GOODS-ALMA 2.0 covers a continuous area of 72.42 arcmin² at a homogeneous sensitivity with two different array configurations: a more extended configuration providing the high-resolution small spatial scales (high resolution dataset) and a more compact configuration supplying the low resolution large spatial scales (low resolution dataset). The results based on the first high resolution dataset alone were already published in Franco et al. (2018, 2020a,b) and Zhou et al. (2020). In this 2.0 version we present the low resolution dataset and its combination with the high resolution dataset (combined dataset), reaching an average sensitivity of $\sigma = 68.4$ μJy beam^{-1} at an average angular resolution
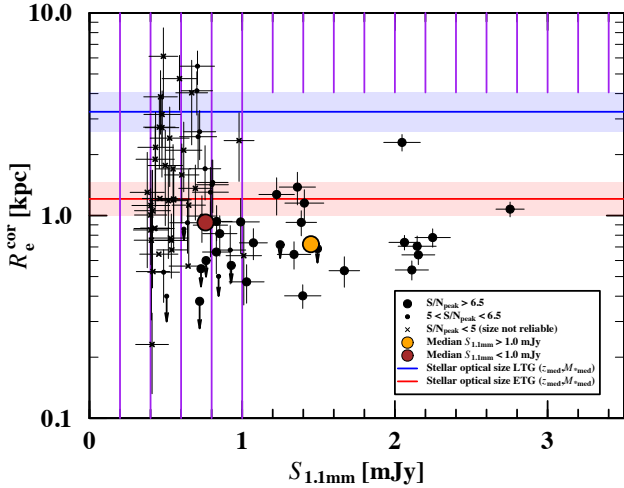
Fig. 16. Physical dust continuum size at 1.1 mm corrected to a common redshift and stellar mass ($z_{med} = 2.46$ and $\log(M_{*med}/M_\odot) = 10.79$) versus 1.1 mm flux density for the sources in the 100% pure plus prior-based catalogs from the combined dataset. For sources in the 100% pure main catalog we distinguish detection $S/N^{peak} > 6.5$ (big black circles, representing approximately the top third of the source catalog) and detection $5 < S/N^{peak} < 6.5$ (small black circles). The sizes of sources in the prior-based supplementary catalog (which have a detection $S/N^{peak} < 5$) are unreliable (shown as crosses), but their 1.1 mm flux densities are well constrained and, thus, we know their position in the x-axis. Median corrected dust continuum sizes for $S_{1.1\,mm} > 1$ mJy sources (orange circle) and $S_{1.1\,mm} < 1$ mJy sources (brown circle) are also displayed. The typical size of the stellar distribution measured at optical wavelengths for both early and late-type galaxies from van der Wel et al. (2014) at z_{med} and M_{*med} are also shown with their scatter as a shaded areas. The grid of purple lines shows the region where we are no longer ~100% complete. Compact dust continuum emission at $S_{1.1\,mm} > 1$ mJy prevails and sizes as extended as typical star-forming stellar disks in this flux density regime are rare.

of $0\farcs447 \times 0\farcs418$. In particular we construct a source catalog, deriving number counts, and dust continuum sizes at 1.1 mm. In summary we find:

– A total of 88 galaxies are detected in a blind search, compared to 35 in the high resolution dataset alone. We find 44 sources with a detection $S/N^{peak} \geq 5$ associated to a purity $p = 1$ (100% pure main catalog). Using a prior-based methodology we find another 44 sources with a detection $3.5 \leq S/N^{peak} \leq 5$ (prior-based supplementary catalog).

– We find a total of 13 optically dark/faint sources (ALMA detected H- or K-band dropouts). This adds seven sources to those already reported in the GOODS-ALMA survey.

– Number counts are fully consistent with the wealth of literature studies. We derived best-fit parameters to the differential and cumulative number counts from a compilation of ALMA-based studies at 850 μm–1.3 mm (see Table 5). In addition, we dissected the contribution to the number counts from our different datasets: while the high resolution dataset is efficient at picking up the bright end of the number counts at $S_{1.1\,mm} > 1$ mJy, it missed sources at $S_{1.1\,mm} < 1$ mJy that appear in the low resolution dataset, which is efficient at retrieving sources for a wide range of flux densities. The combination of the high resolution and low resolution datasets (combined dataset), along with the prior methodology, allow us to be more complete at the faint end $S_{1.1\,mm} < 1$ mJy regime.

– Dust continuum sizes at 1.1 mm are generally compact with a median effective radius of $R_e = 0\farcs10 \pm 0\farcs05$, corresponding to a physical size of $R_e = 0.73 \pm 0.29$ kpc (at the redshift of each source). This result takes advantage of the improved uv coverage and sensitivity in a wider range of spatial scales given the combination of high resolution and low resolution datasets from two array configurations.

– Dust continuum sizes at 1.1 mm evolve with redshift and stellar mass resembling the trends of the stellar sizes measured at optical wavelengths, albeit a lower normalization compared to those of late-type galaxies.

– Compact dust continuum emission at 1.1 mm prevails for sources with flux densities $S_{1.1\,mm} > 1$ mJy. Sizes as extended as typical star-forming stellar disks are rare.

– At $S_{1.1\,mm} < 1$ mJy, dust continuum emission at 1.1 mm appears slightly more extended, although they are still generally compact below the sizes of typical star-forming stellar disks. A larger scatter in the sizes in this flux regime is also seen, with some of the sources possibly entering the regime of the typical size of star-forming stellar disk, but the lower S/N and completeness associated to this flux regime would require further data to confirm and evaluate the abundance of such sources.

After covering a large contiguous area using two array configurations at a similar and homogeneous depth providing both small and large spatial scales, our findings indicate that dust continuum emission occurring in compact regions smaller than the stellar distribution appears to be a norm in DSFGs.

References

Alcalde Pampliega, B., Pérez-González, P. G., Barro, G., et al. 2019, ApJ, 876, 135
Aravena, M., Decarli, R., Walter, F., et al. 2016, ApJ, 833, 68
Aravena, M., Boogaard, L., Gónzalez-López, J., et al. 2020, ApJ, 901, 79
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, ApJS, 218, 33
Austermann, J. E., Dunlop, J. S., Perera, T. A., et al. 2010, MNRAS, 401, 160
Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12
Barger, A. J., Cowie, L. L., Sanders, D. B., et al. 1998, Nature, 394, 248
Barro, G., Kriek, M., Pérez-González, P. G., et al. 2016, ApJ, 827, L32
Barro, G., Faber, S. M., Koo, D. C., et al. 2017, ApJ, 840, 47
Béthermin, M., De Breuck, C., Sargent, M., & Daddi, E. 2015, A&A, 576, L9
Béthermin, M., Wu, H.-Y., Lagache, G., et al. 2017, A&A, 607, A89
Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, A&A, 643, A2
Boogaard, L. A., Decarli, R., González-López, J., et al. 2019, ApJ, 882, 140
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cardamone, C. N., van Dokkum, P. G., Urry, C. M., et al. 2010, ApJS, 189, 270
Carniani, S., Maiolino, R., De Zotti, G., et al. 2015, A&A, 584, A78

Casey, C. M., Narayanan, D., & Cooray, A. 2014, Phys. Rep., 541, 45
Chabrier, G. 2003, PASP, 115, 763
Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, ApJ, 622, 772
Cheng, C., Ibar, E., Smail, I., et al. 2020, MNRAS, 499, 5241
Cimatti, A., Cassata, P., Pozzetti, L., et al. 2008, A&A, 482, 21
Cochrane, R. K., Best, P. N., Smail, I., et al. 2021, MNRAS, 503, 2622
Cooper, M. C., Yan, R., Dickinson, M., et al. 2012, MNRAS, 425, 2116
Coppin, K., Halpern, M., Scott, D., Borys, C., & Chapman, S. 2005, MNRAS, 357, 1022
Coppin, K., Chapin, E. L., Mortier, A. M. J., et al. 2006, MNRAS, 372, 1621
Cowie, L. L., González-López, J., Barger, A. J., et al. 2018, ApJ, 865, 106
Czekala, I., Loomis, R. A., Teague, R., et al. 2021, ApJS, 257, 2
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Decarli, R., Walter, F., Gónzalez-López, J., et al. 2019, ApJ, 882, 138
Dickinson, M., Giavalisco, M., & GOODS Team 2003, in The Mass of Galaxies at Low and High Redshift, eds. R. Bender, & A. Renzini, 324
Dole, H., Lagache, G., Puget, J. L., et al. 2006, A&A, 451, 417
Dudzevičiūtė, U., Smail, I., Swinbank, A. M., et al. 2020, MNRAS, 494, 3828
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Elbaz, D., Arnaud, M., Casse, M., et al. 1992, A&A, 265, L29
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Elbaz, D., Leiton, R., Nagar, N., et al. 2018, A&A, 616, A110
Erben, T., Schirmer, M., Dietrich, J. P., et al. 2005, Astron. Nachr., 326, 432
Ferreras, I., Pasquali, A., Malhotra, S., et al. 2009, ApJ, 706, 158
Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L., & Shafer, R. A. 1998, ApJ, 508, 123
Fontana, A., Dunlop, J. S., Paris, D., et al. 2014, A&A, 570, A11
Franco, M., Elbaz, D., Béthermin, M., et al. 2018, A&A, 620, A152
Franco, M., Elbaz, D., Zhou, L., et al. 2020a, A&A, 643, A30
Franco, M., Elbaz, D., Zhou, L., et al. 2020b, A&A, 643, A53
Fu, H., Cooray, A., Feruglio, C., et al. 2013, Nature, 498, 338
Fudamoto, Y., Oesch, P. A., Schouws, S., et al. 2021, Nature, 597, 489
Fujimoto, S., Ouchi, M., Ono, Y., et al. 2016, ApJS, 222, 1
Fujimoto, S., Ouchi, M., Shibuya, T., & Nagai, H. 2017, ApJ, 850, 83
Fujimoto, S., Ouchi, M., Kohno, K., et al. 2018, ApJ, 861, 7
Garilli, B., McLure, R., Pentericci, L., et al. 2021, A&A, 647, A150
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93
Gómez-Guijarro, C., Toft, S., Karim, A., et al. 2018, ApJ, 856, 121
Gómez-Guijarro, C., Magdis, G. E., Valentino, F., et al. 2019, ApJ, 886, 88
González-López, J., Bauer, F. E., Romero-Cañizales, C., et al. 2017, A&A, 597, A41
González-López, J., Decarli, R., Pavesi, R., et al. 2019, ApJ, 882, 139
González-López, J., Novak, M., Decarli, R., et al. 2020, ApJ, 897, 91
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, A&A, 643, A8
Gullberg, B., Smail, I., Swinbank, A. M., et al. 2019, MNRAS, 490, 4956
Hainline, L. J., Blain, A. W., Smail, I., et al. 2011, ApJ, 740, 96
Hales, C. A., Murphy, T., Curran, J. R., et al. 2012, MNRAS, 425, 979
Hatsukade, B., Ohta, K., Seko, A., Yabe, K., & Akiyama, M. 2013, ApJ, 769, L27
Hatsukade, B., Kohno, K., Umehata, H., et al. 2016, PASJ, 68, 36
Hatsukade, B., Kohno, K., Yamaguchi, Y., et al. 2018, PASJ, 70, 105
Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, A&A, 606, A12
Hildebrandt, H., Erben, T., Dietrich, J. P., et al. 2006, A&A, 452, 1121
Hodge, J. A., & da Cunha, E. 2020, R. Soc. Open Sci., 7, 200556
Hodge, J. A., Karim, A., Smail, I., et al. 2013, ApJ, 768, 91
Hodge, J. A., Swinbank, A. M., Simpson, J. M., et al. 2016, ApJ, 833, 103
Hogg, D. W., & Turner, E. L. 1998, PASP, 110, 727
Hsieh, B.-C., Wang, W.-H., Hsieh, C.-C., et al. 2012, ApJS, 203, 23
Hughes, D. H., Serjeant, S., Dunlop, J., et al. 1998, Nature, 394, 241
Ikarashi, S., Ivison, R. J., Caputi, K. I., et al. 2015, ApJ, 810, 133
Illingworth, G., Magee, D., Bouwens, R., et al. 2016, ArXiv e-prints [arXiv:1606.00841]
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Ivison, R. J., Greve, T. R., Dunlop, J. S., et al. 2007, MNRAS, 380, 199
Ivison, R. J., Swinbank, A. M., Smail, I., et al. 2013, ApJ, 772, 137
Jiménez-Andrade, E. F., Magnelli, B., Karim, A., et al. 2019, A&A, 625, A114
Jiménez-Andrade, E. F., Murphy, E. J., Heywood, I., et al. 2021, ApJ, 910, 106
Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, MNRAS, 432, 2
Knudsen, K. K., van der Werf, P. P., & Kneib, J. P. 2008, MNRAS, 384, 1611
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Franx, M., et al. 2008, ApJ, 677, 219
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kurk, J., Cimatti, A., Daddi, E., et al. 2013, A&A, 549, A63
Labbé, I., Oesch, P. A., Illingworth, G. D., et al. 2015, ApJS, 221, 23
Lagache, G., Abergel, A., Boulanger, F., Désert, F. X., & Puget, J. L. 1999, A&A, 344, 322

Lagos, C. D. P., da Cunha, E., Robotham, A. S. G., et al. 2020, MNRAS, 499, 1948
Lang, P., Schinnerer, E., Smail, I., et al. 2019, ApJ, 879, 54
Lindner, R. R., Baker, A. J., Omont, A., et al. 2011, ApJ, 737, 83
Liu, D., Lang, P., Magnelli, B., et al. 2019, ApJS, 244, 40
Magnelli, B., Elbaz, D., Chary, R. R., et al. 2009, A&A, 496, 57
Magnelli, B., Elbaz, D., Chary, R. R., et al. 2011, A&A, 528, A35
Magnelli, B., Lutz, D., Santini, P., et al. 2012, A&A, 539, A155
Marrone, D. P., Spilker, J. S., Hayward, C. C., et al. 2018, Nature, 553, 51
Martí-Vidal, I., Pérez-Torres, M. A., & Lobanov, A. P. 2012, A&A, 541, A135
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Data Analysis Software and Systems XVI, eds. R. A. Shaw, F. Hill, & D. J. Bell, ASP Conf. Ser., 376, 127
Michałowski, M. J., Dunlop, J. S., Cirasuolo, M., et al. 2012, A&A, 541, A85
Michałowski, M. J., Hayward, C. C., Dunlop, J. S., et al. 2014, A&A, 571, A75
Mignoli, M., Cimatti, A., Zamorani, G., et al. 2005, A&A, 437, 883
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Morris, A. M., Kocevski, D. D., Trump, J. R., et al. 2015, AJ, 149, 178
Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, ApJ, 731, 113
Muñoz Arancibia, A. M., González-López, J., Ibar, E., et al. 2018, A&A, 620, A125
Murdoch, H. S., Crawford, D. F., & Jauncey, D. L. 1973, ApJ, 183, 1
Nelson, D., Pillepich, A., Springel, V., et al. 2019, MNRAS, 490, 3234
Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, ApJ, 278, L1
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43
Nonino, M., Dickinson, M., Rosati, P., et al. 2009, ApJS, 183, 244
Odegard, N., Weiland, J. L., Fixsen, D. J., et al. 2019, ApJ, 877, 40
Oke, J. B. 1974, ApJS, 27, 21
Ono, Y., Ouchi, M., Kurono, Y., & Momose, R. 2014, ApJ, 795, 5
Oteo, I., Zwaan, M. A., Ivison, R. J., Smail, I., & Biggs, A. D. 2016, ApJ, 822, 36
Pannella, M., Elbaz, D., Daddi, E., et al. 2015, ApJ, 807, 141
Papovich, C., Labbé, I., Glazebrook, K., et al. 2016, Nat. Astron., 1, 0003
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Pérez-González, P. G., Rieke, G. H., Egami, E., et al. 2005, ApJ, 630, 82
Persson, S. E., Murphy, D. C., Smee, S., et al. 2013, PASP, 125, 654
Pillepich, A., Nelson, D., Springel, V., et al. 2019, MNRAS, 490, 3196
Popping, G., Walter, F., Behroozi, P., et al. 2020, ApJ, 891, 135
Popping, G., Pillepich, A., Calistro Rivera, G., et al. 2022, MNRAS, 510, 3321
Puglisi, A., Daddi, E., Liu, D., et al. 2019, ApJ, 877, L23
Puglisi, A., Daddi, E., Valentino, F., et al. 2021, MNRAS, 508, 5217
Reddy, N. A., Steidel, C. C., Erb, D. K., Shapley, A. E., & Pettini, M. 2006, ApJ, 653, 1004
Retzlaff, J., Rosati, P., Dickinson, M., et al. 2010, A&A, 511, A50
Ricciardelli, E., Trujillo, I., Buitrago, F., & Conselice, C. J. 2010, MNRAS, 406, 230
Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, Nature, 496, 329
Romano, M., Cassata, P., Morselli, L., et al. 2020, MNRAS, 496, 875
Rowan-Robinson, M., Clegg, P. E., Beichman, C. A., et al. 1984, ApJ, 278, L7
Rujopakarn, W., Dunlop, J. S., Rieke, G. H., et al. 2016, ApJ, 833, 12
Rujopakarn, W., Daddi, E., Rieke, G. H., et al. 2019, ApJ, 882, 107
Salpeter, E. E. 1955, ApJ, 121, 161
Santini, P., Fontana, A., Grazian, A., et al. 2012, A&A, 538, A33
Schechter, P. 1976, ApJ, 203, 297
Schreiber, C., Pannella, M., Leiton, R., et al. 2017a, A&A, 599, A134
Schreiber, C., Elbaz, D., Pannella, M., et al. 2017b, A&A, 602, A96
Scott, S. E., Fox, M. J., Dunlop, J. S., et al. 2002, MNRAS, 331, 817
Scott, K. S., Wilson, G. W., Aretxaga, I., et al. 2012, MNRAS, 423, 575
Scoville, N., Sheth, K., Aussel, H., et al. 2016, ApJ, 820, 83
Shim, H., Kim, Y., Lee, D., et al. 2020, MNRAS, 498, 5065
Simpson, J. M., Swinbank, A. M., Smail, I., et al. 2014, ApJ, 788, 125
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015a, ApJ, 799, 81
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015b, ApJ, 807, 128
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2019, ApJ, 880, 43
Simpson, J. M., Smail, I., Dudzevičiūtė, U., et al. 2020, MNRAS, 495, 3409
Smail, I., Ivison, R. J., & Blain, A. W. 1997, ApJ, 490, L5
Smail, I., Dudzevičiūtė, U., Stach, S. M., et al. 2021, MNRAS, 502, 3426
Smolčić, V., Aravena, M., Navarrete, F., et al. 2012, A&A, 548, A4
Stach, S. M., Smail, I., Swinbank, A. M., et al. 2018, ApJ, 860, 161
Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, ApJ, 830, 51
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2021, ApJ, 915, 87
Sun, F., Egami, E., Rawle, T. D., et al. 2021, ApJ, 908, 192
Swinbank, A. M., Simpson, J. M., Smail, I., et al. 2014, MNRAS, 438, 1267
Tadaki, K.-I., Belli, S., Burkert, A., et al. 2020, ApJ, 901, 74
Toba, Y., Goto, T., Oi, N., et al. 2020, ApJ, 899, 35
Toft, S., Smolčić, V., Magnelli, B., et al. 2014, ApJ, 782, 68

Umehata, H., Tamura, Y., Kohno, K., et al. 2017, ApJ, 835, 98
Umehata, H., Smail, I., Swinbank, A. M., et al. 2020, A&A, 640, L8
Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, A&A, 624, A141
Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, ApJ, 889, 93
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
Vanzella, E., Cristiani, S., Dickinson, M., et al. 2008, A&A, 478, 83
Walter, F., Decarli, R., Carilli, C., et al. 2012, Nature, 486, 233
Walter, F., Decarli, R., Aravena, M., et al. 2016, ApJ, 833, 67
Wang, T., Elbaz, D., Schreiber, C., et al. 2016, ApJ, 816, 84
Wang, T., Schreiber, C., Elbaz, D., et al. 2019, Nature, 572, 211
Wardlow, J. L., Smail, I., Coppin, K. E. K., et al. 2011, MNRAS, 415, 1479
Whitaker, K. E., Ashas, M., Illingworth, G., et al. 2019, ApJS, 244, 16
Williams, C. C., Labbe, I., Spilker, J., et al. 2019, ApJ, 884, 154
Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, ApJ, 682, 985
Wuyts, S., van Dokkum, P. G., Franx, M., et al. 2009, ApJ, 706, 885
Yamaguchi, Y., Kohno, K., Hatsukade, B., et al. 2019, ApJ, 878, 73
Yamaguchi, Y., Kohno, K., Hatsukade, B., et al. 2020, PASJ, 72, 69
Yun, M. S., Scott, K. S., Guo, Y., et al. 2012, MNRAS, 420, 957
Zavala, J. A., Casey, C. M., Manning, S. M., et al. 2021, ApJ, 909, 165
Zhou, L., Elbaz, D., Franco, M., et al. 2020, A&A, 642, A155

[1] AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, 91191 Gif-sur-Yvette, France
e-mail: carlos.gomezguijarro@cea.fr

[2] School of Astronomy and Space Science, Nanjing University, Nanjing 210093, PR China

[3] Aix Marseille Université, CNRS, LAM, Laboratoire d'Astrophysique de Marseille, Marseille, France

[4] Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK

[5] Community Science and Data Center/NSF's NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA

[6] Departamento de Astronomía, Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción, Concepción, Chile

[7] Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan

[8] Department of Physics, Faculty of Science, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand

[9] National Astronomical Research Institute of Thailand (Public Organization), Don Kaeo, Mae Rim, Chiang Mai 50180, Thailand

[10] Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

[11] Cosmic Dawn Center (DAWN), Denmark

[12] DTU-Space, Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark

[13] University of Copenhagen, Jagtvej 128, 2200 Copenhagen N, Denmark

[14] Department of Physics, Anhui Normal University, Wuhu, Anhui 241000, PR China

[15] Infrared Processing and Analysis Center, MS314-6, California Institute of Technology, Pasadena, CA 91125, USA

[16] Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing 210093, PR China

[17] Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham DH1 3LE, UK

[18] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

[19] Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

[20] Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA

[21] National Astronomical Observatory of Japan, National Institutes of Natural Sciences, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

[22] SOKENDAI (The Graduate University for Advanced Studies), 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

[23] School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA

[24] Institute of Astronomy & Astrophysics, Academia Sinica, Taipei 10617, Taiwan

[25] Institute of Astronomy, Graduate School of Science, The University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181-0015, Japan

[26] Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK

[27] Astronomy Unit, Department of Physics, University of Trieste, Via Tiepolo 11, 34131 Trieste, Italy

[28] Fakultät für Physik der Ludwig-Maximilians-Universität, 81679 München, Germany

[29] Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA

[30] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA

[31] Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK

[32] International Space Science Institute (ISSI), Hallerstrasse 6, 3012 Bern, Switzerland

[33] Instituto de Astrofísica, Facultad de Física, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile

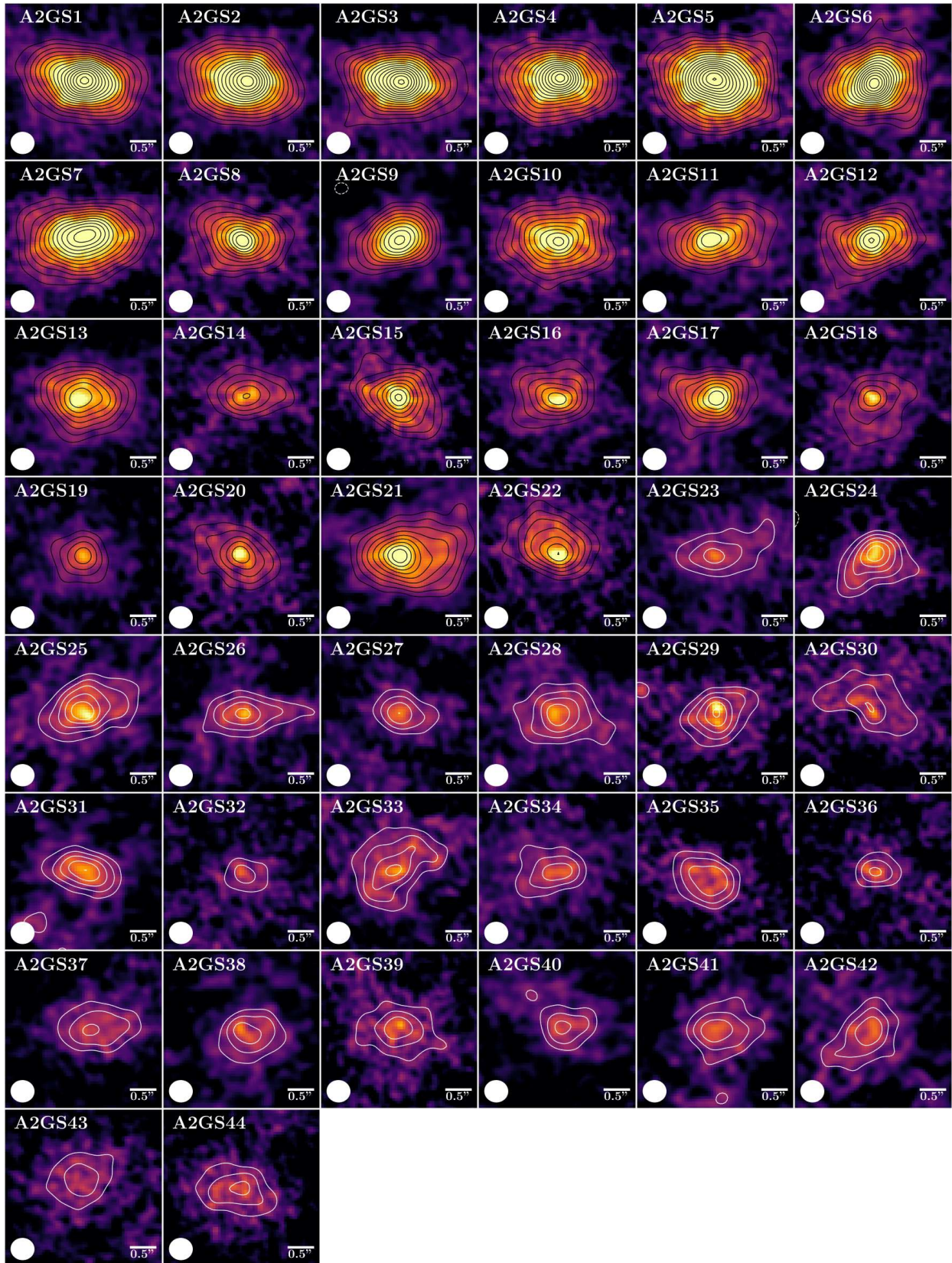## Appendix A: Source catalog ALMA 1.1 mm images

Fig. A.1. Source catalog (100% pure sources) ALMA 1.1 mm images from the combined dataset map with contours overlaid starting at $\pm 3\,\sigma$ and growing in steps of $\pm 1\,\sigma$ ($\sigma = 68.4\,\mu\mathrm{Jy\,beam^{-1}}$). Positive contours are solid and negative contours dotted. ALMA synthesized beam FWHM is shown in the bottom left corner. Images are $3'' \times 3''$ with north up and east to the left.
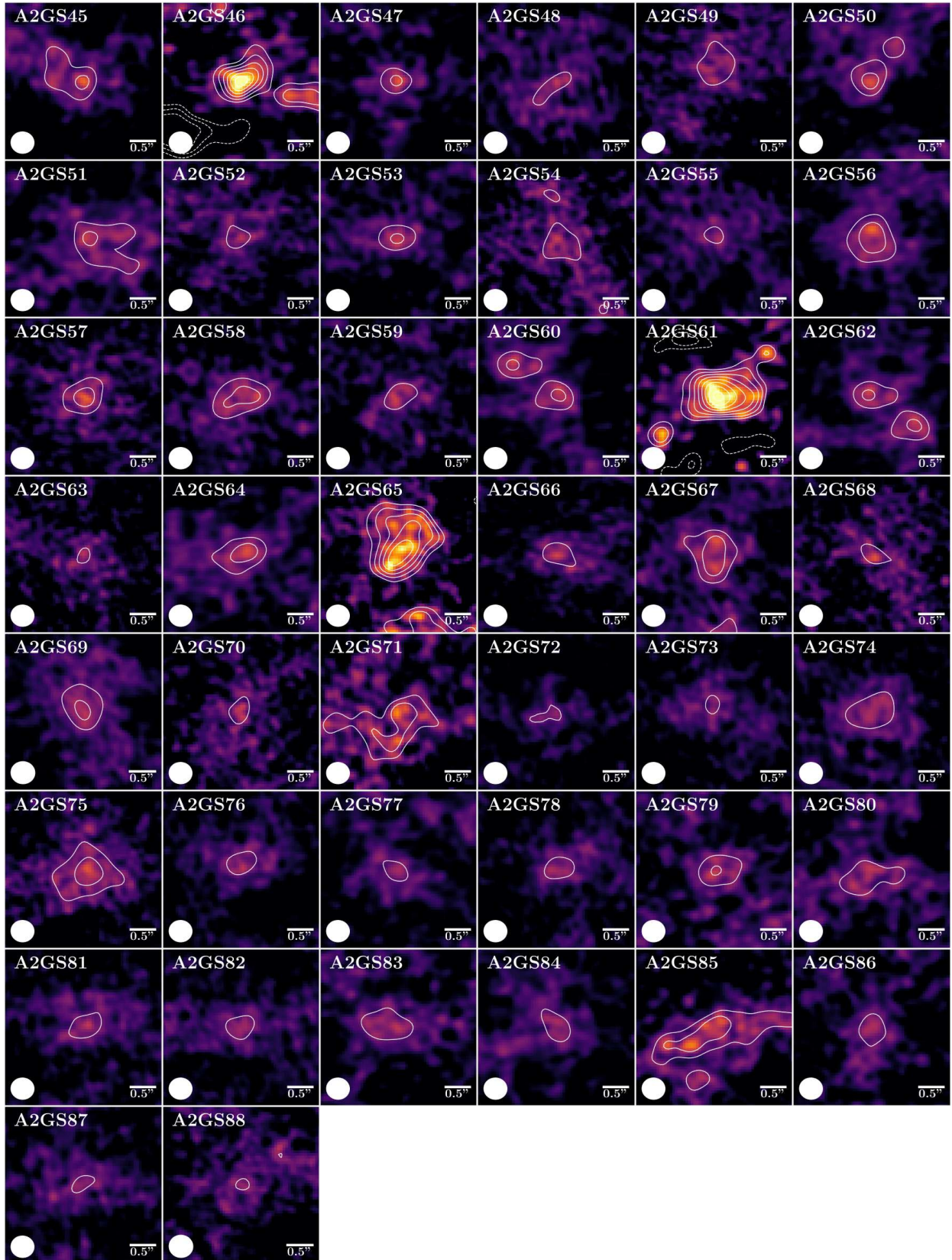
Fig. A.2. Source catalog (prior-based sources) ALMA 1.1 mm images as in Fig. A.1. Images are 3″ × 3″ with north up and east to the left.

Appendix B: Catalog of uncertain sources

Table B.1 contains the extra 16 sources that we labeled as uncertain drawn from the analysis or the prior methodology and stellar mass verification pushed to the limit as described in Sect. 3.1.2. For IRAC priors, at $\leq 1\overset{''}{.}2$ and $\log(M_*/M_\odot) \geq 10.0$ there exists an excess of positive sources with a massive counterpart associated compared to negative detections, expecting around three to be spurious. For VLA, there are no negative detections found at $\log(M_*/M_\odot) \geq 9.0$ for counterparts at $\leq 1\overset{''}{.}0$.

Table B.1. Source catalog: uncertain.

Name	α(J2000) (deg)	δ(J2000) (deg)	S/N	$S_{1.1\,mm}$ (mJy)
A2GS89	53.162554	-27.739038	4.62	0.48 ± 0.12
A2GS90	53.101624	-27.836201	4.41	0.43 ± 0.11
A2GS91	53.126213	-27.756910	4.26	0.38 ± 0.11
A2GS92	53.092587	-27.820566	4.06	0.33 ± 0.10
A2GS93	53.205681	-27.807847	3.99	0.42 ± 0.11
A2GS94	53.041306	-27.793312	3.94	0.19 ± 0.09
A2GS95	53.052008	-27.772592	3.91	0.41 ± 0.09
A2GS96	53.218272	-27.826649	3.83	0.37 ± 0.13
A2GS97	53.171641	-27.733416	3.79	0.45 ± 0.11
A2GS98	53.194773	-27.744600	3.75	0.55 ± 0.11
A2GS99	53.044350	-27.772527	3.59	0.35 ± 0.10
A2GS100	53.173802	-27.772325	3.58	0.35 ± 0.10
A2GS101	53.145150	-27.748866	3.51	0.52 ± 0.11
A2GS102	53.148828	-27.819157	3.49	0.37 ± 0.11
A2GS103	53.136098	-27.784996	3.49	0.39 ± 0.11
A2GS104	53.168012	-27.832939	3.44	0.44 ± 0.11