Multi-wavelength Properties of Radio- and Machine-learning-identified Counterparts to Submillimeter Sources in S2COSMOS

Fang Xia An[1,2,3], J. M. Simpson[1,4], Ian Smail[1], A. M. Swinbank[1], Cong Ma[5,6,7], Daizhong Liu[8], P. Lang[8], E. Schinnerer[8], A. Karim[9], B. Magnelli[9], S. Leslie[8], F. Bertoldi[9], Chian-Chou Chen[10], J. E. Geach[11], Y. Matsuda[12,13], S. M. Stach[1], J. L. Wardlow[14], B. Gullberg[1], R. J. Ivison[10,15], Y. Ao[3], R. T. Coogan[16], A. P. Thomson[17], S. C. Chapman[18], R. Wang[19], Wei-Hao Wang[4], Y. Yang[20], R. Asquith[21], N. Bourne[15], K. Coppin[22], N. K. Hine[22], L. C. Ho[19,23], H. S. Hwang[20], Y. Kato[12], K. Lacaille[24], A. J. R. Lewis[15], I. Oteo[10,15], J. Scholtz[1], M. Sawicki[25], and D. Smith[22]

[1] Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham, DH1 3LE, UK; fangxia@idia.ac.za, fangxiaan@gmail.com
[2] Inter-University Institute for Data Intensive Astronomy, and Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, Bellville 7535, Cape Town, South Africa
[3] Purple Mountain Observatory, Chinese Academy of Sciences, 10 Yuanhua Road, Nanjing 210034, People's Republic of China
[4] Academia Sinica Institute of Astronomy and Astrophysics, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
[5] South African Radio Astronomy Observatory, 2 Fir Street, Observatory 7925, Western Cape, South Africa
[6] Department of Mathematics and Applied Mathematics, University of Cape Town, Cross Campus Road, Rondebosch 7700, Western Cape, South Africa
[7] African Institute for Mathematical Sciences, 6–8 Melrose Road, Muizenberg 7945, Western Cape, South Africa
[8] Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany
[9] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany
[10] European Southern Observatory, Karl Schwarzschild Strasse 2, Garching, Germany
[11] Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK
[12] National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo, 181-8588, Japan
[13] The Graduate University for Advanced Studies (SOKENDAI), Osawa, Mitaka, Tokyo, 181-8588, Japan
[14] Physics Department, Lancaster University, Lancaster LA1 4YB, UK
[15] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[16] Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK
[17] The University of Manchester, Oxford Road, Manchester M13 9PL, UK
[18] Department of Physics and Atmospheric Science, Dalhousie University Halifax, NS B3H 3J5, Canada
[19] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China
[20] Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic of Korea
[21] School of Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD, UK
[22] Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, Hatfield AL10 9AB, UK
[23] Department of Astronomy, Peking University, Beijing, 100087, People's Republic of China
[24] Department of Physics and Astronomy, McMaster University, Hamilton, ON L8S 4M1 Canada
[25] Department of Astronomy and Physics, Saint Marys University, Halifax, NS B3H 3C3, Canada

Abstract

We identify multi-wavelength counterparts to 1147 submillimeter sources from the S2COSMOS SCUBA-2 survey of the COSMOS field by employing a recently developed radio+machine-learning method trained on a large sample of Atacama Large Millimeter/submillimeter Array (ALMA)–identified submillimeter galaxies (SMGs), including 260 SMGs identified in the AS2COSMOS pilot survey. In total, we identify 1222 optical/near-infrared (NIR)/radio counterparts to the 897 S2COSMOS submillimeter sources with $S_{850} > 1.6$ mJy, yielding an overall identification rate of $(78 \pm 9)\%$. We find that $(22 \pm 5)\%$ of S2COSMOS sources have multiple identified counterparts. We estimate that roughly 27% of these multiple counterparts within the same SCUBA-2 error circles very likely arise from physically associated galaxies rather than line-of-sight projections by chance. The photometric redshift of our radio+machine-learning-identified SMGs ranges from $z = 0.2$ to 5.7 and peaks at $z = 2.3 \pm 0.1$. The AGN fraction of our sample is $(19 \pm 4)\%$, which is consistent with that of ALMA SMGs in the literature. Comparing with radio/NIR-detected field galaxy population in the COSMOS field, our radio+machine-learning-identified counterparts of SMGs have the highest star formation rates and stellar masses. These characteristics suggest that our identified counterparts of S2COSMOS sources are a representative sample of SMGs at $z \lesssim 3$. We employ our machine-learning technique to the whole COSMOS field and identified 6877 potential SMGs, most of which are expected to have submillimeter emission fainter than the confusion limit of our S2COSMOS surveys ($S_{850\mu m} \lesssim 1.5$ mJy). We study the clustering properties of SMGs based on this statistically large sample, finding that they reside in high-mass dark matter halos ($(1.2 \pm 0.3) \times 10^{13} h^{-1} M_{\odot}$), which suggests that SMGs may be the progenitors of massive ellipticals we see in the local universe.

1. Introduction

Understanding how galaxies form in the early universe and their subsequence evolution through cosmic time is a fundamental goal of modern astrophysics. The discovery of a population of dusty galaxies at high redshifts at far-infrared (FIR) and millimeter/submillimeter wavelengths has a profound impact on the study of galaxy formation and evolution (e.g., Smail et al. 1997; Barger et al. 1998; Hughes et al. 1998; Scott et al. 2002, 2012; Coppin et al. 2006; Weiß et al. 2009; Yamamura et al. 2010; Clements et al. 2011; Ikarashi et al. 2011; Geach et al. 2017; Simpson et al. 2019; see Casey et al. 2014 for a review). The brighter examples of these FIR/submillimeter sources have infrared luminosities of $L_{IR} \geqslant 10^{12} L_{\odot}$, which is comparable to local ultra-luminous infrared galaxies (ULIRGs). Although such infrared luminous galaxies are rare in the local universe, their spatial density increases rapidly with look-back time and appears to peak at $z \sim 2–3$ (e.g., Barger et al. 1999; Chapman et al. 2005; Smolčić et al. 2012; Yun et al. 2012; Simpson et al. 2014; Swinbank et al. 2014; Chen et al. 2016a; Dudzevîĉiūtė et al. 2019). Therefore, these FIR/submillimeter luminous sources host the most intense star formation in the early universe with star formation rates (SFRs) of $\geqslant 10^2–10^3 M_{\odot} \, yr^{-1}$, which would enable them to form the stellar mass of massive galaxies ($M \geqslant 10^{11} M_{\odot}$) within ~ 100 Myr (e.g., Chapman et al. 2005; Bothwell et al. 2013; Casey et al. 2014). These characteristics make these FIR/submillimeter bright sources a key element to constrain models of galaxy formation and evolution.

Thanks to the strong negative k-correction in the submillimeter/millimeter observational wavebands, we can detect these ultra-luminous infrared galaxies at high redshift ($z \sim 1–6$) with a nearly constant sensitivity in terms of dust mass or far-infrared luminosity (although the latter is sensitive to dust temperature). The majority of bright submillimeter sources have been detected in panoramic, ground-based single-dish submillimeter surveys in the past two decades (e.g., Scott et al. 2002, 2012; Coppin et al. 2006; Weiß et al. 2009; Ikarashi et al. 2011; Geach et al. 2017; Wang et al. 2017; Simpson et al. 2019). However, the typical angular resolution of these ground-based single-dish submillimeter surveys is $\sim 10''–20''$ at 450–1100 μm. This coarse resolution made it very difficult to identify the multi-wavelength counterparts of these submillimeter sources and thus is the major challenge for exploiting these panoramic single-dish submillimeter surveys. The indirect tracers of FIR/submillimeter emission, such as radio, 24 μm, or mid-infrared (MIR) properties are traditionally used to identify counterparts to single-dish submillimeter sources (e.g., Ivison et al. 1998, 2007; Smail et al. 2002; Pope et al. 2006; Barger et al. 2012; Michałowski et al. 2012; Cowie et al. 2017). Unfortunately, the completeness of these identifications is typically $\lesssim 50\%$ because of the lack of negative k-correction and the limited observational depth in radio and MIR bands (e.g., Hodge et al. 2013; Chen et al. 2016a, although see Lindner et al. 2011). The completeness is defined as the number of recovered candidate submillimeter galaxies (SMGs) versus the total number of SMGs in the submillimeter survey.

Recently, the field has advanced considerably as a result of interferometric observations at submillimeter/millimeter wavelengths undertaken with the Submillimeter Array (SMA), IRAM's Plateau de Bure Interferometer (PdBI) and Northern Extended Millimetre Array (NOEMA), and especially the Atacama Large Millimeter/submillimeter Array (ALMA).

These facilities can reach arcsecond/subarcsecond positional precision of SMGs, which significantly improved our understanding of these high-redshift, dusty starburst galaxies (e.g., Frayer et al. 1998; Gear et al. 2000; Tacconi et al. 2006; Genzel et al. 2010; Smolčić et al. 2012; Hodge et al. 2013; Swinbank et al. 2014, 2015; Thomson et al. 2014; Aravena et al. 2016; Walter et al. 2016; Danielson et al. 2017; Dunlop et al. 2017; Simpson et al. 2017; Wardlow et al. 2017; Cooke et al. 2018; Gullberg et al. 2018; Stach et al. 2018, 2019). However, for the large single-dish submillimeter surveys, e.g., the SCUBA-2 Cosmology Legacy Survey (S2CLS; Geach et al. 2017) and the S2COSMOS survey (Simpson et al. 2019), the high-resolution interferometric follow-up is still challenging to complete.

Previous work has tried to take advantages of both single-dish (efficient large area surveys) and interferometric (high angular resolution) submillimeter observations to provide a large sample of SMGs with precisely identified multi-wavelength counterparts, which is necessary for investigating the statistical properties, such as spatial clustering, of SMGs (e.g., Hickox et al. 2012; Chen et al. 2016a; Wilkinson et al. 2017). Galaxy clustering is a key measurement that constrains theoretical models of galaxy formation and evolution, as it provides information of the mass of the halos in which the galaxies reside (e.g., Mo & White 1996, 2002; Cooray & Sheth 2002).

Galaxies that follow similar evolutionary tracks are expected to reside in halos with similar masses across cosmic time. Because of their intensively star-forming, massive, and high-redshift nature, SMGs have been suggested to be the progenitors of compact quiescent galaxies at $z \sim 1–2$ and subsequently local massive ellipticals (e.g., Hughes et al. 1998; Eales et al. 1999; Swinbank et al. 2006; Targett et al. 2011; Simpson et al. 2014; Toft et al. 2014; Wang et al. 2019). This scenario can be tested by comparing the spatial clustering of SMGs and other massive galaxy populations at low redshift or in the local universe.

However, because of the coarse angular resolution of single-dish submillimeter surveys and the small survey area of interferometric observations, measurements of SMG clustering have suffered from large uncertainties (e.g., Weiß et al. 2009; Williams et al. 2011; Hickox et al. 2012; Wilkinson et al. 2017). In addition, the previous studies only include brighter SMGs ($S_{850 \, \mu m} \gtrsim 2–3$ mJy), despite the fact that faint SMGs are necessary for a more complete picture of SMG spatial distribution in general. Indeed, it has been suggested that the mass of fainter SMGs' host halos may be comparable with the hosts of the brighter SMGs and that the fainter SMGs may contribute $\sim 80\%$ of the $S_{850 \, \mu m}$ extragalactic background light (e.g., Cowie et al. 2002; Chen et al. 2016b). However, it is impossible to detect faint SMGs through blank-field single-dish submillimeter surveys if the submillimeter emission falls below the corresponding confusion limit (Jauncey 1968). Although ALMA observations can detect the faintest SMGs with $S_{850 \mu m} \lesssim 1$ mJy, their survey area is very limited (Franco et al. 2018; Umehata et al. 2018). By utilising ALMA survey, Chen et al. (2016a, 2016b) developed an optical-infrared triple color–color (OIRTC) technique to select faint SMGs in the UKIDSS-UDS field and measure the clustering strength of SMGs. The main limitation in Chen et al. (2016b) was the small sample size of training SMGs and the moderate survey area of the UDS field (~ 1 degree^2), which in combination

caused large uncertainties in clustering measurements for both SMGs and comparison samples, especially at high redshift.

To exploit deep, wide-field single-dish submillimeter surveys (Geach et al. 2017; Simpson et al. 2019) and obtain a statistically larger and more robust sample of counterparts to SMGs, more advanced techniques for counterpart identification are required. By utilizing a large sample of ALMA-identified SMGs from the ALMA follow-up of the S2CLS submillimeter sources in the UDS field (AS2UDS; Stach et al. 2018, 2019) as a training set, we developed a machine-learning method to identify multi-wavelength counterparts of single-dish submillimeter sources in An et al. (2018) (hereafter An18), and it was supplemented by the use of radio emission as an indirect tracer of submillimeter emission (e.g., Ivison et al. 2002, 2007; Biggs & Ivison 2008; Thomson et al. 2014). The robustness of our method is confirmed by a series of self-tests and independent tests as shown in An18.

In this work, we employ the same radio+machine-learning method developed in An18 to our new SCUBA-2 submillimeter survey in the COSMOS field (S2COSMOS; Simpson et al. 2019) to obtain a large sample of SMGs across a wide field with reliably identified counterparts and investigate their physical and evolutionary properties. The observations of our test sample and training sets, including ALMA-identified SMGs from the AS2COSMOS pilot survey (J. M. Simpson et al., in preparation), as well as the ancillary data in the COSMOS field, are introduced in Section 2. We present our analyses of radio and machine-learning identification of multi-wavelength counterparts to S2COSMOS sources in Section 3. We give our results and discussions of the multi-wavelength and clustering properties of SMGs in Section 4. The main conclusions of this work are given in Section 5. Throughout this paper, we adopt the AB magnitude system (Oke 1974) and assume a flat ΛCDM cosmological model with parameters fixed at the Planck 2015 best-fit values, namely, the Hubble constant $H_0 = 67.27 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, matter density parameter $\Omega_m = 0.32$, and cosmological constant $\Omega_\Lambda = 0.68$ (Planck Collaboration et al. 2016).

2. Observations

2.1. S2COSMOS

The 850 μm SCUBA-2 COSMOS survey (S2COSMOS) was carried out with the East Asian Observatory's James Clerk Maxwell Telescope (JCMT) between 2016 January and 2017 June. We provide a brief overview here, and the full details of observations, data reduction, and catalog are described in Simpson et al. (2019). For S2COSMOS, we adopted a similar observing strategy to S2CLS (Geach et al. 2017) because the partially completed S2CLS map of COSMOS is incorporated into our S2COSMOS survey. Specifically, we first use four PONG-2700 scans, which provide uniform coverage over a circular region with a diameter of 45′ located equidistant from the center of the field to map the full 2 degree^2 COSMOS field. To reduce the inhomogeneous sensitivity caused by the scan overlap, we adopt a smaller scan pattern, PONG-1800, with scan diameter of 30′ to obtain observations in the four corners of the COSMOS field (see Figure 1 in Simpson et al. 2019). The total exposure time is 223 hr with the PONG-2700 and PONG-1800 scans in a ratio of five to one. In total, combining with the SCUBA-2 archival imaging data at 850 μm, which are mostly from S2CLS (Geach et al. 2017), we consider a 640 hr wide-field 850 μm map of the COSMOS field.

As described in Simpson et al. (2019), the S2COSMOS data were reduced by using the process described in Chapin et al. (2013) with the Dynamical Iterative Map Maker (DIMM) within the Sub-Millimeter Common User Facility (SMURF), which is provided as part of the STARLINK software suite. We refer the reader to Simpson et al. (2019) for the details of the data reduction procedures.

The instrumental sensitivity varies across the final map. In the center of the image, where the four PONG-2700 scan patterns overlap, the lowest noise reaches $\sigma_{850\mu m} = 0.5 \, \mathrm{mJy \, beam^{-1}}$, while in the outer regions the instrumental noise increases to $\sigma_{850\mu m} \leqslant 5 \, \mathrm{mJy}$. Therefore, Simpson et al. (2019) defined a 1.6 degree^2 region, matching the *Hubble Space Telescope* Advanced Camera for Surveys footprint as the S2COSMOS MAIN survey region with a median noise level of $\sigma_{850\mu m} = 1.2 \, \mathrm{mJy \, beam^{-1}}$ and an additional surrounding 1 degree^2 supplementary (SUPP) survey region with a median 1-σ instrumental sensitivity of 1.7 mJy beam^{-1}. Simpson et al. (2019) present catalogs of the sources detection within these MAIN and SUPP regions. The empirical point-spread function of the S2COSMOS survey is obtained by stacking bright, isolated sources and has an FWHM of 14″8. In total, 1020 and 127 submillimeter sources are detected at a significance level of $>4\sigma$ and $>4.3\sigma$ in the MAIN and SUPP regions, respectively, corresponding to a uniform false detection rate of 2% (Simpson et al. 2019). In this work, we use the whole sample of 1147 sources in our analysis.

2.2. Training Set: ALMA Observations in the COSMOS Field

The use of a larger training sample ensures better performance of machine-learning methods. Hence we prefer to use the largest available sample of ALMA-identified SMGs in the COSMOS field, supplemented by large ALMA samples in other fields, e.g., AS2UDS (Stach et al. 2019), as a training set for identifying multi-wavelength counterparts of our S2COSMOS sources.

2.2.1. AS2COSMOS

We have completed an ALMA Cycle 4 pilot study of the brightest 160 single-dish submillimeter sources from S2COS-MOS (Project ID: 2016.1.00463.S). The ALMA follow-up observations were taken in Band 7 (870 μm) in 2018 April and May. The ALMA primary beam diameter at this frequency is 17″3, which encompasses the area of the SCUBA-2 beam. In addition, there are 24 archival ALMA maps at Band 7 corresponding to our S2COSMOS sources. Therefore, our full AS2COSMOS sample includes 184 ALMA maps with median sensitivity $\sigma_{870\,\mu m} = 0.2 \, \mathrm{mJy \, beam^{-1}}$. The median synthesised beams of the 184 ALMA maps is 0″8 × 0″79. Full details of the observations are presented in J. M. Simpson et al. (in preparation).

The ALMA data were reduced using the the Common Astronomy Software Application (CASA, McMullin et al. 2007) v4.2.2-5.1.1. The data reduction procedures and source detection method are similar to that of the ALMA-SCUBA-2 Ultra Deep Survey (AS2UDS; Stach et al. 2018, 2019). In the 184 ALMA maps, 260 ALMA SMGs were detected at a peak signal-to-noise ratio (S/N) > 4.8 or a 1″2-diameter aperture S/N > 4.9.

2.2.2. A³COSMOS

We use additional ALMA archival data in the COSMOS field to construct our training set. All the publicly available ALMA archive data in the COSMOS field were processed, imaged, and analyzed in the ongoing ALMA archive mining project A³COSMOS (Liu et al. 2019).[26] The latest version of the A³COSMOS data set contains 1909 ALMA pointings, leading to 1134 robust ALMA detections with $S/N \gtrsim 5.4$, corresponding to a spurious fraction of 8% and a completeness >90% (Liu et al. 2019). To limit the complications that arise from different observational frequencies, we only use those continuum observations with frequencies close to the central frequency of the SCUBA-2 filter transmission ($800 < \lambda_c < 1200\ \mu m$). We convert the flux density of all A³COSMOS SMGs to $S_{870\,\mu m}$ by adopting the ratios given in Fujimoto et al. (2016) and limited our training set to $S_{870\,\mu m} > 1$ mJy. We also remove very shallow observations by limiting the rms noise of the map to $\sigma_{870\,\mu m} \leqslant 0.25$ mJy. In total, 984 A³COSMOS SMGs meet these requirements and are used in constructing the training set for identifying counterparts to S2COSMOS submillimeter sources.

2.3. Additional Multi-wavelength Observations/Catalogs

The COSMOS field is one of the largest extragalactic fields with a rich ancillary data set. Here we give a short summary of the data that we use in our analysis.

2.3.1. VLA 3GHz Radio Maps

The radio data we use in this work are from the VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017b), which were taken with the Karl G. Jansky Very Large Array (VLA) at 3 GHz. The reduced continuum data and source catalog have been released by Smolčić et al. (2017b). In summary, the median rms of the final reduced data reaches 2.3 μJy (equivalent to $\simeq 4\ \mu$Jy rms at 1.4 GHz) over the 2 degree² COSMOS field with an angular resolution of $0\rlap{.}''75$. Smolčić et al. (2017b) detected 10,830 3 GHz radio sources at $\geqslant 5\sigma$, which are used to identify radio counterparts of the S2COSMOS submillimeter sources in our work.

2.3.2. Optical/NIR/FIR Catalogs

The COSMOS2015 catalog (Laigle et al. 2016) from the UltraVISTA-DR2 surveys are used to identify optical/NIR counterparts for both ALMA SMGs and S2COSMOS submillimeter sources and are then used to construct a training set and a test sample for the machine-learning analyses. In An18, we found that photometric redshift, absolute H-band magnitude and NIR colors have the greatest diagnostic power to differentiate SMGs from non-SMGs. Since the COSMOS field has deep z^{++} data, which are also used to detect sources in the region that lie outside of the coverage of the K_s band (Laigle et al. 2016), we use the z^{++}, J, K_s, and IRAC 3.6 μm and 4.5 μm band photometries from the COSMOS2015 catalog in our machine-learning analyses.

The 3σ depth of z^{++} is 25.9 mag in a 3″ diameter aperture. The UltraVISTA-DR2 has J- and K_s-band observations reaching a 3σ depth of $J = 24.7$ mag and $K_s = 24.0$ mag in a region of 1.7 degree² and a $J = 24.9$ mag and $K_s = 24.7$ mag in the

four ultra-deep stripes, which cover an area of 0.62 degree². The IRAC 3.6 and 4.5 μm observations have 3σ depth of 25.5 mag within a 3″ diameter aperture. Laigle et al. (2016) used the total flux, which is estimated from the corrected 3″ aperture flux, to fit the SEDs from near-ultraviolet to NIR for their NIR-detected sources. They produced the probability distribution function (PDF) of photometric redshift for each galaxy by matching the observed SED to a set of galaxy templates at a redshift grid with a step of 0.01 and range of $0 < z < 6$ through minimizing the χ^2. The median of this distribution was determined as the photometric redshift of galaxy in the COSMOS2015 catalog. The absolute H-band magnitudes given in the catalog was estimated from the best-fitting SEDs. We refer the reader to Laigle et al. (2016) for the details.

The other photometric catalog used in this work is a "super-deblended" FIR-to (sub)millimeter photometric catalog from Jin et al. (2018). Using the position of K_s-band or radio (in the case of K_s-band non-detection) sources, Jin et al. (2018) adopt a "super-deblended" method developed by Liu et al. (2018) to "deblend" the FIR-to (sub)millimeter photometry of sources from the *Spitzer* (Le Floc'h et al. 2009), *Herschel* (Oliver et al. 2010; Lutz et al. 2011; Béthermin et al. 2012), SCUBA-2 (Cowie et al. 2017; Geach et al. 2017), AzTEC (Aretxaga et al. 2011), MAMBO (Bertoldi et al. 2007), and VLA (1.4 GHz and 3 GHz) surveys (Schinnerer et al. 2010; Smolčić et al. 2017b) in the COSMOS field. Jin et al. (2018) estimated the SFR of NIR or radio-selected galaxies by integrating 8–1000 μm infrared luminosities derived from the FIR+millimeter SED fitting. In our work, we use these estimates of SFR to investigate the star formation efficiency of our identified counterparts of SMGs in the COSMOS field.

3. Analysis

In a similar manner to An18, in order to maximize the completeness, we combine the radio and machine-learning methods to identify the multi-wavelength counterparts of single-dish submillimeter sources.

3.1. Radio Identification

Radio synchrotron emission has been proven to be a useful tracer of obscured star formation, since it is powered by supernova remnants of massive stars. Therefore, radio identification is a traditional method to search for counterparts of submillimeter sources (e.g., Ivison et al. 2002). Instead of considering all radio sources within the SCUBA-2 error circles as potential counterparts, we use the corrected-Poissonian probability, p-value (Downes et al. 1986; Dunlop et al. 1989) to calculate the probabilistic association of radio sources to single-dish submillimeter sources. In An18, we confirmed that the adoption of $p \leqslant 0.065$ increases the precision of radio identification from 64% to 70%. In this work, we adopt this limit in the radio identification.

The VLA 3 GHz radio observations cover all of the S2COSMOS sources, except for the two northernmost ones. In this work, we define the error circle of a SCUBA-2 source as $r = 8\rlap{.}''7$, which is the FWHM of the primary beam of ALMA Band 7, as the precision and recall of radio and machine-learning method are all estimated based on the ALMA SMGs in AS2UDS. Therefore, the performance of the radio and machine-learning identification can only be tested within the

---

[26] https://sites.google.com/view/a3cosmos
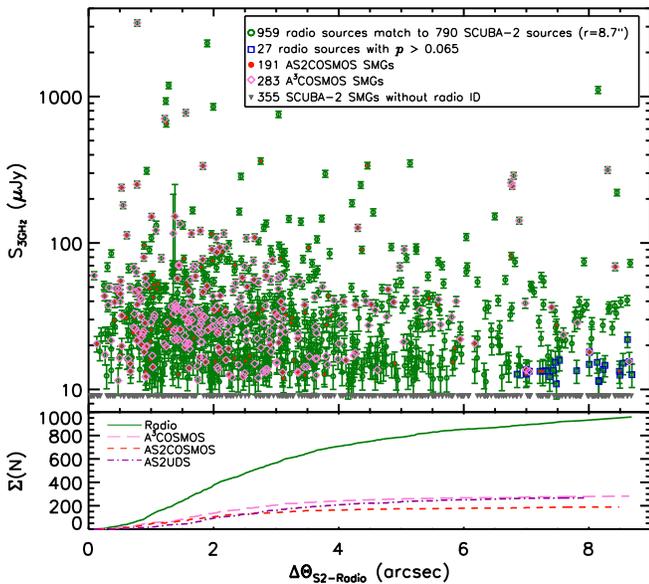
Figure 1. Top: the radio flux densities for all radio sources within the error circles ($r = 8\overset{''}{.}7$) of the SCUBA-2-detected submillimeter sources as a function of the offset of these radio sources from the SCUBA-2 single-dish source. For the 1147 S2COSMOS submillimeter sources, 1145 of them are covered by the VLA 3 GHz radio map. In addition, 959 radio sources lie within 790 SCUBA-2 error circles. We also mark the radio counterparts of ALMA SMGs from our pilot AS2COSMOS survey and from the ALMA archive (A^3COSMOS). We use $p \leqslant 0.065$ as a cut of "robust" radio identification according to our previous work (An et al. 2018). From the test of AS2UDS ALMA SMGs in the UDS field, we expect that the precision of radio identification is ∼70%. Among the 959 radio sources, 27 have $p > 0.065$ and are removed from our identification catalog. Therefore, 932 counterparts of SCUBA-2 sources in the COSMOS field are identified by the radio data alone. There are also 355 SCUBA-2 sources that do not have radio counterparts within $8\overset{''}{.}7$. Bottom: the cumulative number of radio sources and ALMA SMGs as a function of their offset from the SCUBA-2 sources. We also plot the results of AS2UDS SMGs for comparison. In both fields, the number of ALMA SMGs converges at $\Delta\theta > 3''$ while the number of radio sources increases gradually, suggesting that these include some associated companions to the submillimeter sources.

ALMA primary beam. *Precision* is defined as the ratio between the number of correctly identified SMGs and the total number of predicted SMGs using the radio or machine-learning classification. *Recall* is the number of corrected classification versus the total number of ALMA SMGs. In Figure 1, we show that there are $959 \geqslant 5\sigma$ VLA 3 GHz radio sources within 790 SCUBA-2 error circles. Among them, 932 have $p \leqslant 0.065$ and thus are identified as the likely counterparts of S2COSMOS sources. According to the results for AS2UDS in An18, the precision of radio identification is around 70%. The radio imaging in the COSMOS field is ∼2× deeper than that in the UDS field; therefore, the completeness of radio identification in this work should be higher than that of AS2UDS (39%). However, as shown in Figure 1, there are 355 S2COSMOS sources without radio identifications. This may be caused by the fact that the radio observation do not benefit from a negative k-correction, which means even deep radio observation will miss higher-redshift SMGs. Therefore, other methods of identifying counterparts of S2COSMOS sources are necessary to improve the completeness.

3.2. Machine-learning Classification

We apply the machine-learning classification in An18 to this work to identify the optical/NIR counterparts of S2COSMOS

sources. In An18, we use two machine-learning algorithms, Support Vector Machine (SVM; Vapnik 1995) and XGBoost (Chen & Guestrin 2016), to classify SMGs from field galaxies. The performance of these two algorithms is similar as shown in An18. Liu et al. (2019) tested a number of machine-learning algorithms and found most of them, including the deep learning, have very similar performance in differentiating SMGs from field galaxies. We should point out that the performances shown in Liu et al. (2019) is slighter better than that in An18 and this work mainly because they used a smaller search radius ($r = 7''$), which affects the final completeness of identification, as there are ALMA SMGs with offset $>7''$ from SCUBA-2 sources as shown in Figure 1. Therefore, in this work, we prefer to retain the two machine-learning algorithms we validated in An18 to identify multi-wavelength counterparts to SCUBA-2 sources in the COSMOS field.
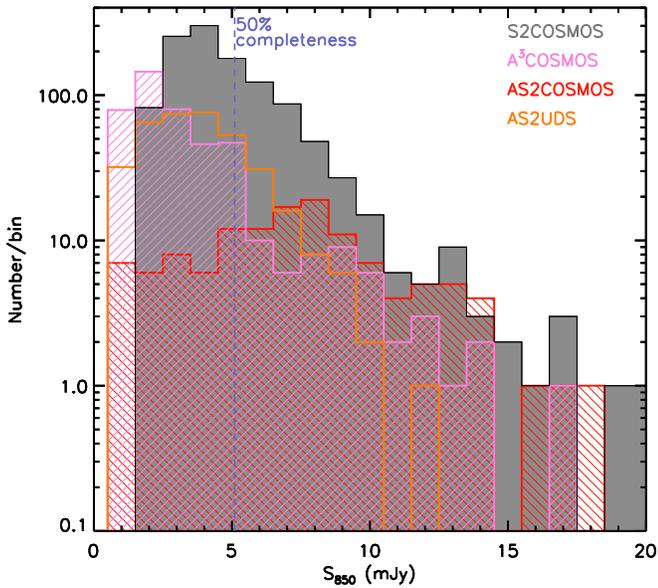
3.2.1. Training Set

The effectiveness of machine-learning algorithms depends sensitively on the completeness and precision of the training set. On the one hand, a larger sample size provides better performance of the machine-learning classification. On the other hand, as we demonstrated in An18, differences of photometric systems between the training set and test samples will affect the performance of the machine-learning algorithms. Therefore, in this section, we compare the performance of both SVM and XGBoost classifiers by using training sets based on three different ALMA surveys. We described two of them, AS2COSMOS pilot and A^3COSMOS, in Section 2.2. The other ALMA survey we use is the 870 μm ALMA survey of 716 SCUBA-2 sources in the UDS field (AS2UDS; Stach et al. 2019), which was used to build the training set in An18.

We show the flux density distributions of ALMA SMGs in these three ALMA surveys in Figure 2. The flux densities of A^3COSMOS SMGs are converted to $S_{870\,\mu m}$ by adopting the ratios estimated by Fujimoto et al. (2016).

Following An18, a non-SMG is defined as any optical/NIR source that lies within the ALMA primary beams but does not have a secure detection from ALMA (e.g., does not have $\geqslant 4.8\sigma$ detection in the AS2COSMOS pilot). In An18, we identified that the photometric redshift (z_{phot}), absolute H-band magnitude (M_H) and NIR colors are the most efficient properties for differentiating SMGs from field galaxies. The COSMOS field has deep IRAC 3.6 and 4.5 μm data along with the z^{++}-band data, which is used to detect sources outside the regions of UltraVISTA-DR2 by Laigle et al. (2016). Hence the NIR colors we choose in this work are ($z - [3.6]$) and ([3.6]–[4.5]). While the SVM classifier requires detection in all the selected properties, the XGBoost classification can be performed with missing features. Therefore, when constructing the training sets for XGBoost, we do not require secure detection in z^{++}-band. In addition, we include ($J - K_s$) and ($K_s - [3.6]$) colors if the source has secure detections in the corresponding band(s).

We list the number of SMGs and non-SMGs with the secure measurements of selected features in each training set in Table 1. Without the limitation of detection in z^{++}-band, the sample sizes of XGBoost increase ∼10%–15% compared to that of SVM. We also combine the training sets based on AS2UDS, AS2COSMOS, and A^3COSMOS surveys to enlarge the training set. The duplicates between AS2COSMOS pilot and A^3COSMOS have been removed in the combined training

Figure 2. Flux density distributions of S2COSMOS submillimeter sources and ALMA SMGs in the training set used in this work. We also mark the flux densities that correspond to the 50% completeness of S2COSMOS sources (Simpson et al. 2019). The flux density of all ALMA archive SMGs (A³COSMOS) are converted to $S_{870\mu m}$ if necessary using the flux ratios given in Fujimoto et al. (2016). We also present the flux distribution of ALMA SMGs in the UDS field (AS2UDS) for comparison. The flux density of our training sets covers that of single-dish S2COSMOS submillimeter sources.

set to guarantee a uniform weight for all sources when training the machine-learning classifiers.

The parameters of both SVM and XGBoost classifiers are optimized by five-fold cross-validation (Kohavi 1995), which means we first divide the training set into five subsets and train the machine-learning classifier on four folds and validate on the remaining one. We repeat the five-fold cross-validation 100 times to estimate the scatter in evaluation metrics. The evaluation metrics we use in this work are *Recall*, *Precision*, False-Positive Rate (FPR) ,and the F_1 score. We have defined *Precision* and *Recall* in Section 3.1. FPR is defined as the number of sources that are incorrectly classified as SMGs over the total number of non-SMGs in the data set. We also use the F_1 score, which is defined as $2\times$ (Precision $\times$ Recall)/ (Precision+Recall), as one of the metric to evaluate the performance of machine-learning classifier trained by different training sets. We show the *Recall* and *Precision* of five-fold cross-validation with the optimized SVM classifiers trained on different training sets in the left panel of Figure 3 and list the values of all evaluation metrics in Table 1.

The SVM classifier trained by the sample based on AS2UDS has the lowest F_1 score, 76%. Possible reasons for this are that the training set based on AS2UDS comprises the *K*-band-selected sources, and that the 3.6 and 4.5 μm observations in the UDS field are shallower than those in the COSMOS field. Therefore, the selected input features based on the multi-wavelength photometry in the COSMOS field are not the best properties for the photometric systems in the UDS field to train the machine-learning classifier. For instance, the F_1 score of five-fold cross-validation of the AS2UDS training set with the different input features in An18 is ~80%. The SVM classifier trained by the A³COSMOS training sets has a performance with $F_1 = 81\%$ as shown in Table 1. The performance of SVM is slightly increased to $F_1 = 82\%$ if we combine the training sets

of AS2COSMOS pilot and A³COSMOS (A³+AS2COSMOS). Although the SVM classifier trained by AS2COSMOS pilot shows the best performance with $F_1 = 83\%$, the scatter is larger because of the small sample size. In addition, the AS2COSMOS pilot corresponds to the brighter single-dish submillimeter sources ($S_{850\mu m} > 6$ mJy). Therefore, we also evaluate the SVM classifiers that are trained on different training sets by using the other training set as the test sample and show the results in the left panel of Figure 3. We find that if we test the machine-learning method on the test sample that corresponds to the brighter single-dish submillimeter sources, i.e., AS2COS-MOS pilot and the brighter AS2UDS, the results will be overestimated. As shown in Figure 3, if we use the training set that is constructed by the brighter SCUBA-2 sources, then we will fail to recover ~5% SMGs with moderate/fainter submillimeter emission. We also notice that if we use the training set from UDS field to recover SMGs in the COSMOS field and vice versa, the differences between observation depth, source selection, and the estimates of the photometric redshift will cause a ~4%–6% decrease in the success of machine-learning classification. Overall, the SVM classifier trained on A³+AS2COSMOS has better performance in all cases as shown in Figure 3.

We do the same evaluation for the XGBoost classifier. We show the Receiver Operating Characteristic (ROC) curves from the optimized XGBoost classifier trained on different training sets in the right panel of Figure 3. The ROC curves are constructed by comparing the *Recall* against the FPR, as the probability threshold is varied (Fawcett 2004). We then use the area under the curve (AUC) of an ROC curve (Fawcett 2004) and the evaluation metrics shown in Table 1 to evaluate the performance of XGBoost classifiers. The classifier trained on the A³+AS2COSMOS training set has a better performance with the maximal AUC (Figure 3) and $F_1 = 81\%$. Therefore, we choose the training set based on the A³+AS2COSMOS surveys to train the machine-learning classifier in this work.

For the two machine-learning algorithms, SVM has a slightly better performance compared to XGBoost as shown in Table 1. However, the SVM cannot deal with missing features unless they are artificially filled by imputed data. In this work, the primary missing features are NIR colors of sources. Unfortunately, the causes of the lack of measurement of NIR colors are various. They could be due to dust reddening, geometry, star formation history, redshift, and so on. For classifying SMGs and non-SMGs, the statistical imputation techniques do not improve the performance of machine-learning classifiers and even make the results worse in some cases, as demonstrated in Liu et al. (2019). In contrast, the XGBoost classifier can perform classification with missing features by classifying the instance into the optimal default direction that is learned from the data (Chen & Guestrin 2016). If we adopt the XGBoost classifier, the sample size of both the training set and the test sample increase without the limitation of requiring secure detection in the z^{++}-band, as shown in Table 1. And these two classifiers have a very similar precision while the recall of SVM is ~2%–4% higher than that of XGBoost. However, the sample size of XGBoost is ~10% larger than that of SVM. Therefore, the final completeness of the machine-learning method is still higher if we adopt the XGBoost classifier. We therefore prefer to use the XGBoost classifier in this work to increase the sample size of both the training set and the test sample and hence the completeness of
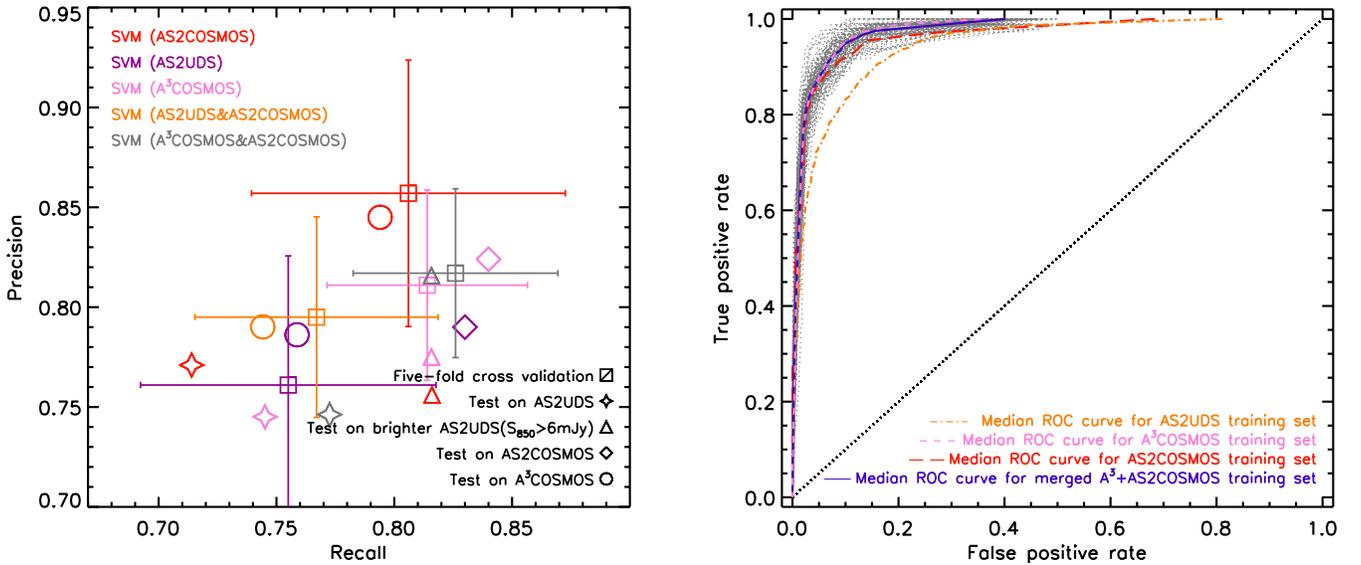
Figure 3. Left: comparison of the SVM classification performances based on different training sets. We use different colors to show the results based on different training sets and different symbols represent the results of different tests. For each training set, we use five-fold cross-validation to optimize the parameters of the SVM classifier and show the optimized results as squares in the plot, where the errors are standard deviation of precision and recovery rate based on 100 bootstrap-simulations. Then we test the performance of the SVM classifier based on a different training set by using the other training sets as the test sample. Through these comparisons, we find the SVM based on the combined A^{3}COSMOS and AS2COSMOS pilot training sets has the best performance in differentiating SMGs from field galaxies. Right: comparison of the Receiver Operating Characteristic (ROC) curves for optimized XGBoost classifier trained on different training sets. The XGBoost classifier is optimized through the five-fold cross-validation. We repeat the five-fold cross-validation 100 times, calculate the median true and false-positive rates and use them to represent the curve. As shown in the plot, the ROC curve based on merged A^{3}COSMOS and AS2COSMOS pilot training set has the maximal area-under-the-curve (AUC). Therefore, this merged sample is adopted as the training set in this work. The background gray dotted lines represent the individual ROC curves of five-fold cross-validation based on this merged training set.

Table 1
Machine-learning Performance Based on Different Training Sets

Training Set/SVM	SMG[a]	non-SMG[a]	Recall (%)	Precision (%)	FPR (%)	F_1 Score (%)
AS2UDS	255	1224	76 ± 6	76 ± 7	5.2 ± 1.6	76 ± 6
AS2COSMOS pilot	100	850	81 ± 7	86 ± 7	1.4 ± 0.9	83 ± 7
A^{3}COSMOS	357	4186	81 ± 4	81 ± 5	1.6 ± 0.5	81 ± 5
AS2COSMOS+UDS	355	2074	77 ± 5	80 ± 5	3.4 ± 1.0	78 ± 5
A^3+AS2COSMOS[b]	394	4509	83 ± 4	82 ± 4	2.0 ± 0.6	82 ± 4
Training set/XGB						
AS2UDS	364	1279	74 ± 5	81 ± 4	5.2 ± 1.4	77 ± 5
AS2COSMOS pilot	126	928	78 ± 9	82 ± 8	2.2 ± 1.0	80 ± 8
A^{3}COSMOS	445	4528	80 ± 5	81 ± 4	2.4 ± 0.6	81 ± 5
AS2COSMOS+UDS	458	1999	74 ± 4	79 ± 4	4.4 ± 1.1	77 ± 4
A^3+AS2COSMOS[b]	490	4904	79 ± 4	82 ± 4	2.4 ± 0.5	81 ± 4

Notes.
[a] The number of SMG and non-SMGs in the training set.
[b] Duplicates in these two training sets have been removed.

the identified counterparts of the S2COSMOS submillimeter sources.

3.2.2. Machine-learning Results

Having selected a training set and machine-learning algorithm, we then use the optimized XGBoost classifier to identify optical/NIR counterparts to S2COSMOS submillimeter sources. For the 1147 S2COSMOS sources, 1066 of them are within optical/NIR coverages. Within their error circles, there are 5655 NIR-selected sources that have secure detection in IRAC 3.6 and 4.5 μm band and also have estimated photometric redshift and absolute H-band magnitude in the COSMOS2015 catalog. These are used to construct the test sample for classifying counterparts to

S2COSMOS sources. We show the results of classification in Figure 4. The XGBoost classifier identifies 658 counterparts of S2COSMOS sources from the 5655 NIR-selected sources. Among them, 368 also have radio-identified counterparts. According to the self-test of the AS2UDS data set in An18, the *Precision* of radio and machine-learning classification could reach 90%. Combining with radio identification, in total, we identify 815 optical/NIR/radio counterparts to S2COSMOS submillimeter sources from the 5655 NIR-selected sources. The expected *Recall* of the radio combined with the machine-learning method in this work is $\geqslant 85\%$, according to the self-test in An18 and the relatively deeper radio observation in the COSMOS field.
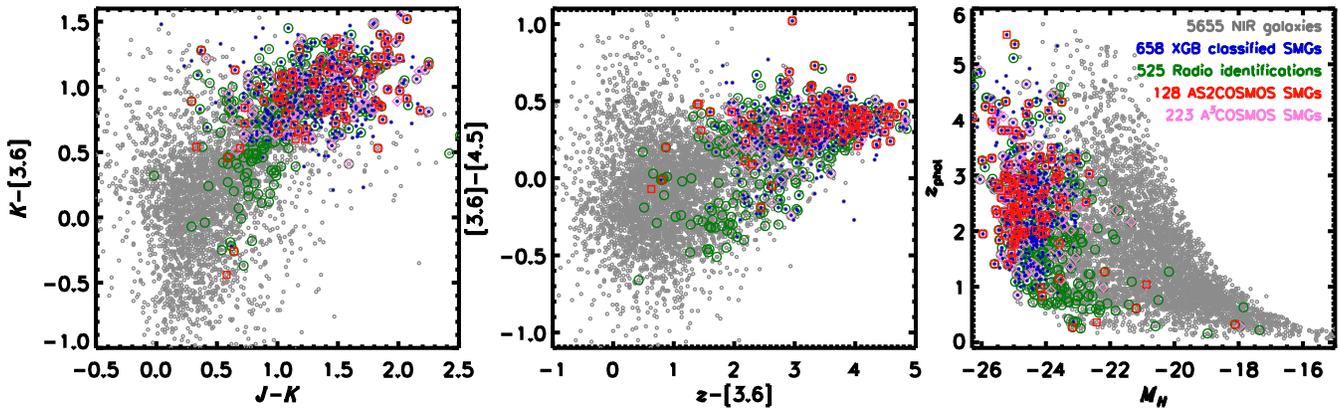
Figure 4. Results of applying the XGBoost classifier to identify counterparts to SCUBA-2-detected submillimeter sources in NIR observations of the COSMOS field. The gray circles are 5655 NIR-selected galaxies, which are located within 1066 SCUBA-2 error circles and have secure detection in 3.6 and 4.5 μm band and sufficient photometry for estimating the redshift and absolute H-band magnitude. The blue points represent the 658 counterparts of SCUBA-2 submillimeter sources classified by machine learning. The red squares show the 128 ALMA SMGs from our pilot ALMA follow-up observations of brighter ($S_{850\mu m} > 6$ mJy) S2COSMOS submillimeter sources. XGBoost recalls 88% ALMA SMGs that corresponded to brighter single-dish submillimeter sources. There are also 223 ALMA SMGs from the ALMA archive located within the SCUBA-2 error circles that meet the requirement of the machine-learning classification. XGBoost recovers 90% of these ALMA SMGs. However, we note that all of these ALMA SMGs (AS2COSMOS pilot and A³COSMOS) are part of the training set. Therefore, the *Recall* is higher than that of the five-fold cross-validation of the training set, which is ~80%. We also mark the 525 NIR-selected galaxies with $>5\sigma$ 3 GHz radio detection and $p \leqslant 0.065$, i.e., radio-identified counterparts to SCUBA-2 sources. There are 368 NIR galaxies classified as the SMGs by both radio and machine learning. Combining with the radio identification, in total, we identify 815 counterparts of SMGs from 5655 NIR-selected galaxies in the COSMOS field. According to the "self-test" of the AS2UDS sample and the deeper radio observation in COSMOS, we expect that our radio+machine-learning method recovers $\geqslant$85% ALMA SMGs from these 5655 NIR galaxies.

In Section 3.1, we identified 932 radio counterparts to S2COSMOS sources by using the p-value to calculate the probabilistic association of radio sources to S2COSMOS sources. Of these 368 are also identified by machine learning, while radio identification alone finds 564 counterparts to S2COSMOS sources. Therefore, in total, our radio+machine-learning method identifies 1222 optical/NIR/radio counterparts of single-dish-detected submillimeter sources in the COSMOS field, which are listed in Table 2.

3.2.3. Identification Rate

The completeness of the radio+machine-learning method is expected to be 64% based on the test of AS2UDS in An18. This fraction should be higher for brighter submillimeter sources because of the increased identification rate as shown in Figure 5. The identification rate is defined as the number of single-dish submillimeter sources having at least one identified counterpart versus the total number of single-dish submillimeter sources in the survey. For the 1145 S2COSMOS sources with radio or NIR coverage, our method identified multi-wavelength counterparts for 897 of them. Therefore, the average identification rate is 78%. From Figure 5, we can see that the identification rate increases with the flux density of S2COSMOS sources. We also show the identification rate of AS2COSMOS pilot SMGs as a function of $S_{870\,\mu m}$ in Figure 5. The identification rate of ALMA SMGs also increases with submillimeter flux density. The identification rate of AS2UDS SMGs in An18 is shown for comparison. On average, the identification rate of S2COSMOS sources is higher than that of AS2UDS SMGs, which is most likely caused by the false-positive identifications of the radio+machine-learning method. The expected identification rate of SMGs within the SCUBA-2 error circles in the COSMOS field is slighter higher than that in the UDS field because of the slightly deeper radio data.

4. Results and Discussion

Having identified the multi-wavelength counterparts of S2COSMOS sources, we now analyze the physical properties of these single-dish-detected submillimeter sources.

4.1. Redshift Distribution

For the 1222 radio+machine-learning-identified counterparts, 819 have estimated photometric redshifts from the COSMOS2015 catalog with $z_{median} = 2.3 \pm 0.1$. We show the redshift distribution of our identified optical/NIR/radio counterparts of S2COSMOS sources in Figure 6.

By comparing the redshift distribution of machine-learning- or radio-identified counterparts with that of ALMA SMGs in the UDS (AS2UDS) and the ECDFS (ALESS) fields, we find that machine learning fails to recover some low-redshift counterparts while radio identification includes some low-redshift contamination. For the machine-learning classification, photometric redshift is one of the input features in the analyses. Therefore, the trained classifier tends to recover SMGs at high redshift. Radio observations do not benefit from a negative k-correction, hence they will miss SMGs at high redshift and preferentially include some contaminations at low redshift. Nevertheless, the overall redshift distribution of radio+machine-learning-identified counterparts of S2COSMOS sources is broadly consistent with that of ALMA SMGs (Simpson et al. 2014; Dudzevičiūtė et al. 2019) as shown in Figure 6.

4.2. Multiplicity

Recent studies using interferometric observations in submillimeter/millimeter suggest $\gtrsim$20% of single-dish-detected submillimeter sources actually correspond to blends of multiple SMGs (e.g., Wang et al. 2011; Karim et al. 2013; Simpson et al. 2015a, 2015b; Stach et al. 2018). In this work, multiplicity is defined as single-dish-detected submillimeter sources having

Table 2
Radio+Machine-learning-identified Counterparts to S2COSMOS Sources

ID_{S2}	$R.A._{S2}$	$Decl._{S2}$	$R.A._{NIR}$	$Decl._{NIR}$	$Flag_{xgb}$[a]	$R.A._{radio}$	$Decl._{radio}$	$Flag_{radio}$[b]	z_{phot}
S2COS850.0001	150.033530	2.436552	...	...	...	150.033508	2.436735	1	...
S2COS850.0002	150.064676	2.263777	150.065170	2.263673	1	150.065063	2.263606	1	3.31
S2COS850.0003	150.238150	2.337106	150.238708	2.336827	1	150.238647	2.336836	1	2.29
S2COS850.0003	150.238150	2.337106	150.239273	2.336381	1	...	...	...	2.36
S2COS850.0006	149.989043	2.458214	149.988724	2.458332	0	149.988693	2.458483	1	3.30
...	...	...	...	...	...	...	...	...	...

Notes.
[a] $Flag_{xgb} = 1$: XGBoost classified SMGs; 0: classified non-SMGs; ...: not qualified for machine-learning classification.
[b] $Flag_{radio} = 1$: radio-identified SMGs; ...: do not have radio detection.
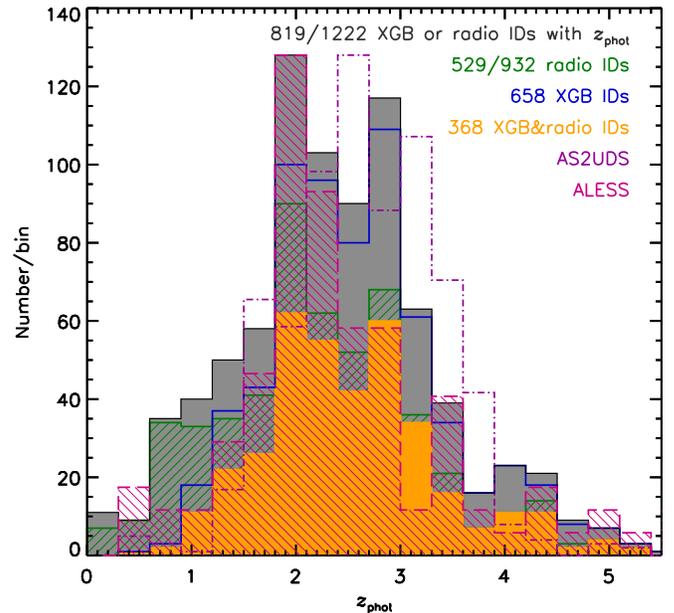
(This table is available in its entirety in machine-readable form.)
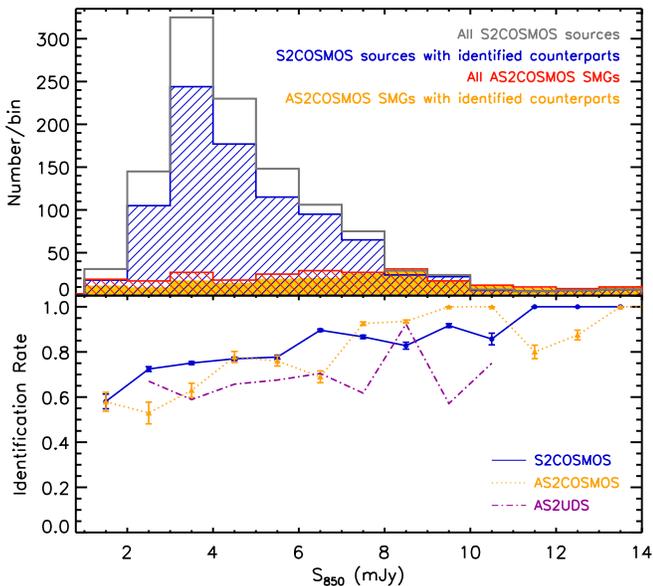
Figure 5. Identification rate as a function of submillimeter flux density. Top: the flux density distributions of all SCUBA-2 sources, SCUBA-2 sources having at least one identified counterpart, and AS2COSMOS pilot SMGs having identified radio or NIR counterparts. Bottom: the fraction of SCUBA-2 sources with at least one identified counterpart and the fraction of ALMA SMGs with an identified counterpart as a function of $S_{850\,\mu m}$. It can be seen that the identification rate increases for both brighter single-dish-detected submillimeter sources and ALMA-detected SMGs. We also show the completeness of the radio+machine-learning method of AS2UDS sample as a function of $S_{850\,\mu m}$ for comparison. We expect a slightly higher completeness of S2COMSOS SMGs because of deeper radio data compared to that in the UDS field.

Figure 6. Redshift distributions of radio+machine-learning-identified counterparts to SCUBA-2 detected submillimeter sources in the COSMOS field. We plot the distribution of all 819/1122 XGBoost or radio identifications with photometric redshift (z_{phot}). We also separately present the redshift distribution of radio+machine-learning-identified counterparts to submillimeter sources. The distribution of galaxies that are identified as counterparts of SMGs by both radio and machine learning is shown as the orange area. For comparison, we also show the redshift distribution of ALMA SMGs in the UDS field (AS2UDS; Dudzevičiūtė et al. 2019) and the ECDFS field (ALESS; Simpson et al. 2014). The distributions of these two ALMA samples are scaled to compare with our results. In general, our radio+machine-learning-identified counterparts of SMGs have a similar redshift distribution to ALMA-detected SMGs.
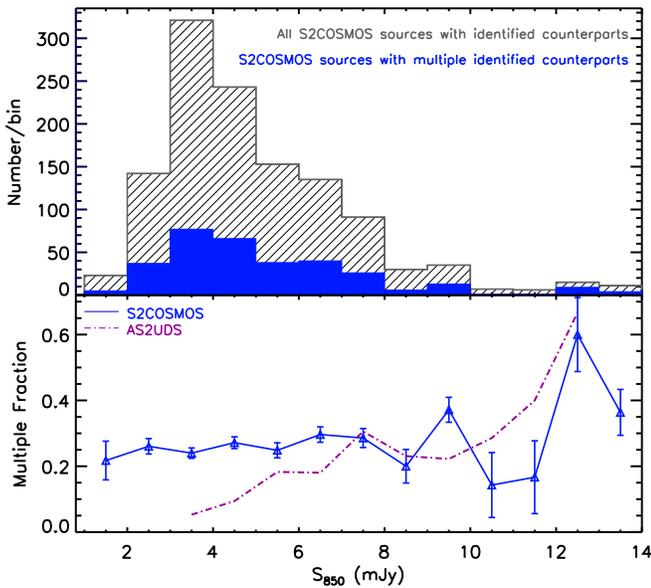
more than one identified counterpart from our radio+machine-learning method. We show the multiple fraction as a function of the flux density of single-dish-detected submillimeter sources in Figure 7. The average fraction of S2COSMOS sources with multiple identified counterparts is $(26 \pm 5)\%$. This fraction is fairly constant, and only slightly increases for brighter single-dish submillimeter sources as shown in Figure 7.

Stach et al. (2018) in their ALMA AS2UDS survey studied the multiplicity of S2CLS-UDS single-dish submillimeter sources and found a multiplicity of $(28 \pm 2)\%$ for submillimeter sources with $S_{850\,\mu m} \geqslant 5$ mJy and of $(44^{+16}_{-14})\%$ at $S_{850\,\mu m} \geqslant 9$ mJy. We plot the multiplicity based on AS2UDS in Figure 7 for comparison. The multiple fraction of S2COSMOS sources increases with the flux density of the SCUBA-2 source and is higher than that of AS2UDS at the

faint end, which most likely is a result of incompleteness in identifying faint companions in those ALMA maps. In An18, we stacked the submillimeter emission in ALMA maps at the position of the machine-learning-classified but individually ALMA-undetected NIR-selected galaxies and found that the machine learning can recover the faint submillimeter emissions even if they are below the detection threshold of ALMA observations ($S_{870\,\mu m} \sim 1$–2 mJy). This is further confirmed by deeper ALMA observations in Cycle 5 for the 10 brightest SCUBA-2 sources in AS2UDS, which did not have any secure ALMA detections in Cycles 3 and 4 (Stach et al. 2019). These deeper observations recovered counterparts for 9 of the 10 target fields. This may explain the decline in multiplicity at the faint end in AS2UDS shown in Figure 7. However, the

Figure 7. Top: the flux density distribution of all SCUBA-2 sources that have radio or machine-learning-identified counterparts in the COSMOS field and the distribution of SCUBA-2 sources with multiple identified counterparts. Bottom: the fraction of SCUBA-2 sources with multiple identified counterparts as a function of $S_{850\mu m}$. On average, the multiple fraction is $(26 \pm 5)\%$ and this fraction increases for brighter single-dish-detected sources. We also show the multiple fraction of the AS2UDS sample for comparison. The higher multiple fraction at the faint end of this work might be caused by the false-positive detections of our radio+machine-learning method. However, our stacking analyses shown in An et al. (2018) confirm that the machine-learning method can identify the faint/diffuse submillimeter emissions even below the ALMA detection limit. The accuracy of the multiplicity of S2COSMOS sources is affected by both incompleteness and false-identifications of the radio+machine-learning method.
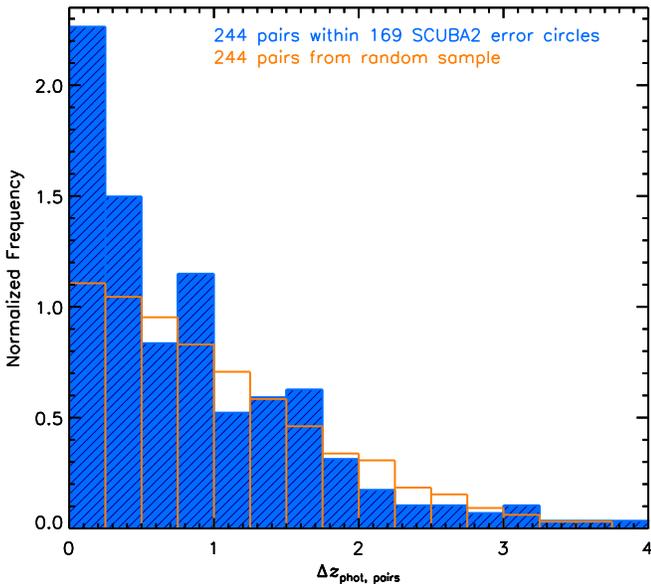
Figure 8. Normalized distribution of Δz for the pairs of radio+machine-learning-identified counterparts within the same SCUBA-2 error circle, compared to the pairs randomly selected from the distribution of all isolated counterparts with photometric redshift. The distribution is normalized by assuming that each SCUBA-2 source only has a single counterpart. The strong peak at $\Delta z < 0.25$ for the identified counterpart pairs compared with a random sample indicates that a moderate fraction of multiple counterparts in the same SCUBA-2 error circles arises from physically associated galaxies, rather than chance line-of-sight projections.
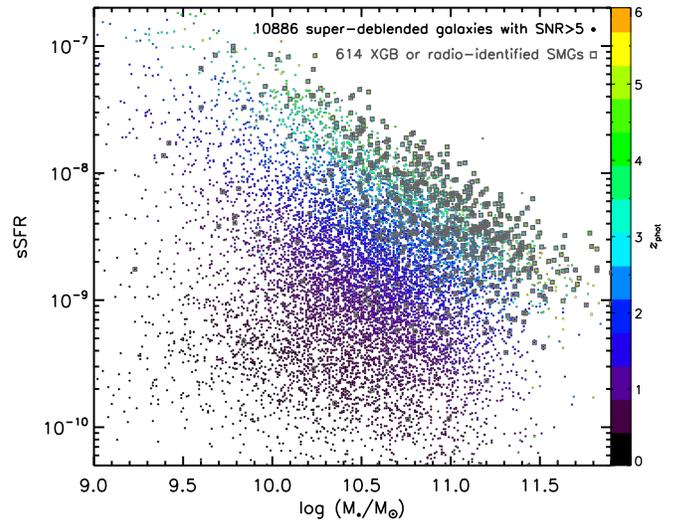

Figure 9. Specific star formation rate (sSFR) vs. stellar mass for "super-deblended" K_s or radio-detected galaxies in the COSMOS field. The SFRs are computed from the integrated 8–1000 μm infrared luminosities derived from FIR+millimeter SED fitting (Jin et al. 2018). Stellar masses are from Laigle et al. (2016) and Muzzin et al. (2013) with a Chabrier IMF (Chabrier 2003). We mark our radio+machine-learning-identified counterparts of SMGs with gray squares. The identified counterparts are the most strongly star-forming and most massive galaxies, compared with the remaining K_s or radio-detected field galaxies.

false-positive detection of the radio+machine-learning method may also cause a modest increase of multiplicity of S2COSMOS sources, although this is not expected to depend on $S_{850\,\mu m}$.

As described in Section 4.1, 819/1222 radio+machine-learning-identified counterparts have estimated photometric redshift from the COSMOS2015 catalog (Laigle et al. 2016). Among their corresponding single-dish SCUBA-2 sources, our study of multiplicity shows that 133 (15%) of them have two, 35 (4%) have three and one has four radio+machine-learning-identified counterparts. For the SCUBA-2 sources that have multiple SMGs, Simpson et al. (2015a) and Stach et al. (2018) suggested that ~30% of them arise from physically associated galaxies based on their photometric redshift (Δz_{phot}), while Wardlow et al. (2018), using an ALMA CO survey of six single-dish submillimeter sources, found that $(36 \pm 18)\%$ of multiple SMG components in blended single-dish submillimeter sources are closely physically associated.

In this work, we show the normalized distribution of Δz_{phot} for the 244 pairs that are identified as counterparts to the same S2COSMOS sources in Figure 8. We find that 65 of 244 pairs (27%) have $\Delta z_{phot} < 0.25$. The choice of $\Delta z_{phot} < 0.25$ is to compare with the results of ALMA SMGs in Stach et al. (2018). We also note that the choice of $\Delta z_{phot} < 0.25$ corresponds to ~2.5σ of the uncertainty of the photometric redshift of radio+machine-learning-identified counterparts of SMGs as shown in Section 4.1. To test the significance of this result, we randomly select 244 pairs from the 1222 identified counterparts of S2COSMOS sources. We repeat this random selection 100 times and show the median value of the distribution of Δz for these random pairs in Figure 8 for comparison. On average, 15% of the random pairs have $\Delta z_{phot} < 0.25$, which is half of that for identified counterpart pairs. This suggests that a moderate fraction (~27%) of multiple counterparts to the same SCUBA-2 sources arise from physically associated galaxies, rather than line-of-sight
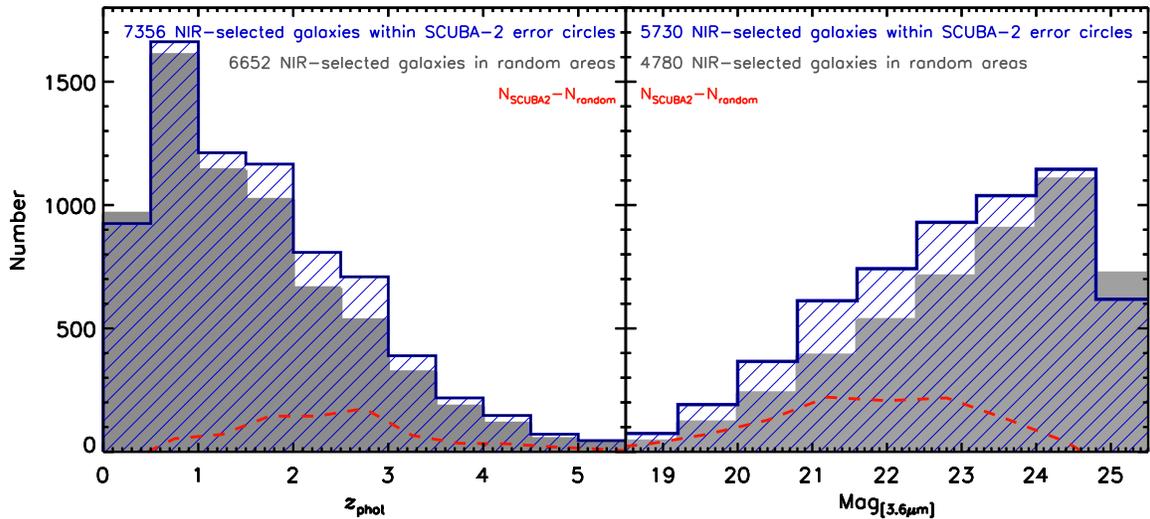
Figure 10. Redshift distribution (left) and 3.6 μm magnitude (right) distribution of NIR-selected galaxies within SCUBA-2 error circles (blue shade area), compared to those galaxies within sky (gray area). The galaxies in the right panel also have secure detection at 4.5 μm band and have estimated photometric redshift and absolute H-band magnitude from COSMOS2015 catalog, i.e., are qualified for our machine-learning method. We can see an excess in number density within SCUBA-2 error circles, especially at higher redshift and the brighter end of 3.6 μm magnitude.

projections by chance, although our result is affected by the incompleteness and false-positive identification of the radio +machine-learning method.

4.3. AGN Fraction

Both theoretical simulations and observational studies suggest a link between the growth of galaxies and their supermassive black holes (SMBHs, $M_{BH} = 10^6$–$10^9 M_\odot$, e.g., Hopkins et al. 2008; Ishibashi & Fabian 2016). Under this paradigm, starburst-dominated galaxies and the active galactic nuclei (AGNs) dominated QSOs are essentially the same systems observed at different evolutionary stages (e.g., Sanders et al. 1988; Perna et al. 2018). Therefore, surveying the AGN activity in the SMG population provides insights into not only SMBH growth, but also potentially the evolutionary cycle of massive galaxies.

We start by looking at the AGN population in radio-identified SMGs. We take advantage of the fact that Smolčić et al. (2017a) has identified AGNs from the VLA-COSMOS 3 GHz radio data using the X-ray, MIR color–color, and multi-wavelength SEDs selection methods. For the 932 radio-identified counterparts of S2COSMOS sources, 80 have X-ray detections in the *Chandra* COSMOS-Legacy Survey (Civano et al. 2016; Marchesi et al. 2016). Among these, 74 are classified as X-ray AGNs according to their X-ray luminosity ($L_{[0.5–8] Kev} > 10^{42}$ erg s^{-1}, although we note that this is conservative for strongly star-forming galaxies; Smolčić et al. 2017a). Since the X-ray selection of AGNs is progressively missing faint AGNs at high redshift because of the limitation of survey depth, Smolčić et al. (2017a) also adopt the MIR color–color selection method of Donley et al. (2012) and the SED selection to complement the X-ray selection criterion. Of the radio counterparts of S2COSMOS sources with $z_{phot} \lesssim 3.0$, 125 meet the MIR AGN criterion, although 41 of them are also X-ray AGNs. Smolčić et al. (2017a) also classified radio sources that show AGN signatures in their optical to millimeter SED as SED-selected AGNs. Among these, 163 are counterparts of S2COSMOS sources. In total, 225/932 (24 ± 4)%

radio-identified counterparts of S2COSMOS sources are classified as AGNs in Smolčić et al. (2017a).

For the remaining 290 optical/NIR counterparts of S2COSMOS sources, which lack radio counterparts and are classified by machine learning, none of these have X-ray detections in *XMM*-COSMOS (Cappelluti et al. 2007; Hasinger et al. 2007; Brusa et al. 2010) or *Chandra* surveys (Civano et al. 2016; Marchesi et al. 2016). Therefore, the fraction of X-ray-detected AGNs in our radio+machine-learning-identified counterparts of S2COSMOS sources is (6 ± 2)%. We also adopt the MIR color–color selection criteria in Donley et al. (2012) for the 170/290 optical/NIR counterparts of SCUBA-2 sources, which have $z_{phot} \leqslant 3.0$ and $>3\sigma$ detections in all four IRAC bands (see Stach et al. 2019). Three of these meet the selection criteria of AGNs from Donley et al. (2012). Therefore, combining the MIR color–color and SED selection, the AGN fraction in our radio+machine-learning-identified counterparts of S2COSMOS sources is (19 ± 4)%, which is consistent with the AGN fraction in ALMA SMGs of the ALESS (Wang et al. 2013) and the AS2UDS samples ($\lesssim$28%, Stach et al. 2019) but lower than a sample of 1.1 mm selected ALMA SMGs in the GOODS-South field (Franco et al. 2018).

4.4. Star Formation Rate and Stellar Mass

SMGs are believed to be massive starburst galaxies with total infrared luminosity of $L_{IR} \gtrsim 10^{12} L_\odot$. In this section, we investigate the star formation rate and stellar mass of our identified counterparts of SCUBA-2 sources in the COS-MOS field.

We first match our sample to the FIR-to-(sub)millimeter photometric catalog from Jin et al. (2018), which is described in Section 2.3.2. This catalog is created by using the position of K_s-band or radio-detected sources and a "super-deblended" method developed by Liu et al. (2018) to estimate the FIR-to-(sub)millimeter photometries. The SFR of these "super-deblended" sources is estimated by integrating the 8–1000 μm infrared luminosity from the best-fit FIR+mm SED (Jin et al. 2018). We show the specific SFR, which is defined as the ratio of SFR and stellar mass, as a function of stellar mass for the

10,886 "super-deblended" sources with $S/N_{FIR+mm} > 5$ in Figure 9. Among these, 614 are our radio+machine-learning-identified counterparts of S2COSMOS sources. The stellar mass estimates are from the catalogs of Laigle et al. (2016) and Muzzin et al. (2013) with a Chabrier IMF (Chabrier 2003). The photometric redshift of sources is shown by the color scale in Figure 9.

As shown in Figure 9, our identified counterparts of single-dish submillimeter sources tend to be at higher redshift, while having higher star formation activity and higher stellar mass, compared with the NIR or radio-selected galaxies with fainter submillimeter emission. These differences support the reliability of our classification. We thus conclude that statistical analyses of SMGs can be undertaken based on this large sample of identified counterparts of submillimeter sources from the panoramic single-dish survey.

4.5. Excess of Number Density within SCUBA-2 Error Circles

To study the environment properties of single-dish-detected submillimeter sources, we first compare the number density of NIR-selected galaxies within the SCUBA-2 beam with that in the random areas. These are 1066/1147 SCUBA-2 sources lying within the NIR coverage in the COSMOS field. We then randomly offset the 1066 SCUBA-2 position in R.A. or decl. by $9''$, which is slighter larger than the SCUBA-2 beam, and use the average number of NIR-selected galaxies within these regions as the number density of NIR galaxies in the random areas. We show the distribution of redshift and $3.6\,\mu m$ magnitude of NIR galaxies within the SCUBA-2 error circle and within these random areas in Figure 10. The number density of NIR galaxies within the SCUBA-2 beam exhibits an excess at high redshift and bright $3.6\,\mu m$ magnitudes. The NIR galaxies shown in the right panel of Figure 10 are qualified for our machine-learning classification, i.e., they have secure detection at 3.6 and $4.5\,\mu m$ band while having estimated photometric redshift and absolute H-band magnitude. The total number of excess NIR galaxies within the SCUBA-2 error circles is 950. In Section 3, we identified 815 counterparts of SMGs from 5655 NIR-detected galaxies within these 1066 SCUBA-2 error circles. According to An18, the recall of the radio+machine-learning method is ~85% for the AS2UDS SMGs that qualified for machine-learning analyses. Therefore, the excess of NIR galaxies within SCUBA-2 regions can be roughly explained by the contribution from SMGs if we consider the incompleteness of radio+machine-learning identification.

4.6. Properties of "Blank"-SCUBA-2 Sources

From the 1145 SCUBA-2 sources within the VLA 3 GHz radio survey region or NIR coverage, we find no radio or optical/NIR counterparts for 248 SCUBA-2 sources ($22\% \pm 5\%$). We call these "blank"-SCUBA-2 sources. We find only seven radio sources within the error circles of these 248 "blank"-SCUBA-2 sources and all of them have p-value >0.065, thus are only classified as "possible" counterparts. There are 1960 NIR galaxies within these 248 "blank"-SCUBA-2 error circles, and 778/1960 (40%) of them meet the requirement for the machine-learning analyses, but are classified as non-SMGs. Comparing to machine-learning-identified counterparts of SMGs (Figure 4), these 778 NIR galaxies within the "blank"-SCUBA-2 regions are either fainter in absolute H-band or at lower redshift or have a
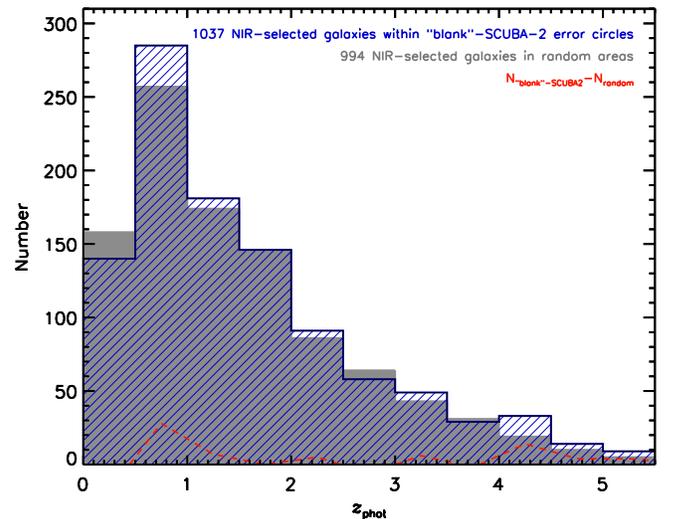
Figure 11. Redshift distribution of NIR-selected galaxies within 248 "blank"-SCUBA-2 error circles, compared with those galaxies within random areas. We can also see an excess of number density within "blank"-SCUBA-2 error circles, but with a weaker significance compared with that in Figure 10.

blue NIR color and therefore are classified as "non-SMGs" by the XGBoost classifier.

We also study the environment properties of these "blank"-SCUBA-2 sources by using the method described in Section 4.5. There is also an excess of 43 in the number of NIR galaxies that have estimated photometric redshift within the "blank"-SCUBA-2 sources compared with that in random areas as shown in Figure 11. Therefore, for "blank"-SCUBA-2 sources, the number density of this excess is $0.17\,beam^{-1}$, which is weaker than that of all 1066 SCUBA-2 sources with an average density of $0.66\,beam^{-1}$ as shown in the left panel of Figure 10.

We stack the *Herschel*/SPIRE maps at the positions of "blank"-SCUBA-2 sources and detect significant emission with the averaged peak flux densities of 8.3 ± 0.4, 10.5 ± 0.6, and 9.5 ± 0.7 mJy at $250\,\mu m$, $350\,\mu m$, and $500\,\mu m$, respectively. This suggests that the main explanation for "blank"-SCUBA-2 sources is the incompleteness of radio+machine-learning method. The major causes of the incompleteness is that some SMGs lack radio or NIR counterparts or do not have secure measurements of all the input features required for the machine-learning analyses.

4.7. SMG Clustering

Spatial clustering is a powerful way to study galaxy evolution and the evolutionary connections between different galaxy populations, since it provides a direct method to constrain the mass of a halo in which galaxies reside (e.g., Cooray & Sheth 2002).

For the bright SMGs selected at $850\,\mu m$, the measurements of clustering have suffered from small number statistics because of their low spatial density and small survey area. Previous works resorted to a cross-correlation technique by using other galaxy populations with higher source surface densities (Hickox et al. 2012; Wilkinson et al. 2017). Chen et al. (2016b) studied the clustering of SMGs by using a sample of ~3000 OIRTC-identified counterparts of SMGs in the UDS field, which includes faint SMGs below the single-dish confusion limit. However, the clustering measurements in
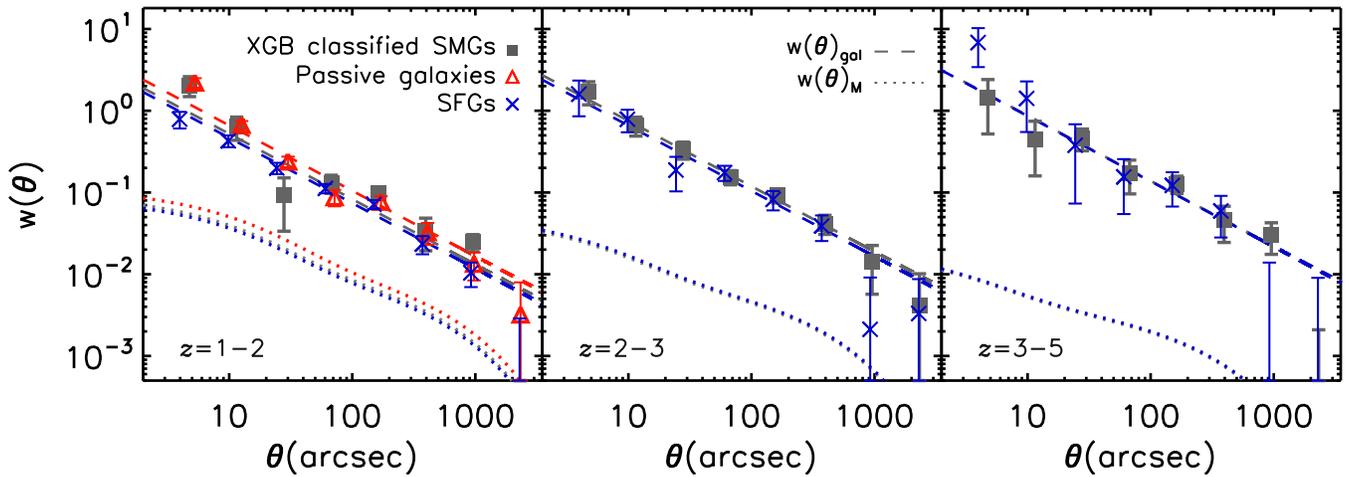
Figure 12. Two-point autocorrelation function of machine-learning-identified counterparts at $z = 1$–2, 2–3, and 3–5 in the full COSMOS field, compared with (NUV $-r$)/($r - J$) color–color selected quiescent and star-forming galaxies from the COSMOS2015 catalog (Laigle et al. 2016). The stellar masses of galaxies in all three samples are limited with $\log_{10}(M_*/M_\odot) \geqslant 10.5$. The dashed lines represent the best-fit power-law models, $w(\theta) = A\,\theta^{-0.8}$. The dotted curves represent the autocorrelation function of the total matter. As shown in Appendix, the redshift distributions of machine-learning-identified SMGs, passive galaxies and SFGs are slightly different in each redshift bin. We therefore evaluate the autocorrelation functions of the total matter for the three different samples separately and show them in different colors. In the lowest redshift bin, the passive galaxies show a relatively stronger clustering, compared with machine-learning-identified counterparts of SMGs.

Chen et al. (2016b) were limited by the moderate survey area of the UDS field. We similarly apply our machine-learning classification to the NIR-selected galaxies in the whole COSMOS field to identify faint SMGs, whose submillimeter emission is fainter than the confusion limit of S2COSMOS surveys. Although recent interferometric instruments can detect faint SMGs ($S_{850\,\mu m} \lesssim 1.0$ mJy, e.g., Franco et al. 2018; Umehata et al. 2018), it is observationally prohibitive to map the full 2 degree^2 COSMOS field with ALMA. By applying our machine-learning method to the whole COSMOS field, we obtain a sample of faint SMG candidates for future interferometric follow-up observations. In total, 356,673 NIR-selected galaxies meet the requirement of the machine-learning classification, i.e., having secure detection at 3.6 and 4.5 μm and having estimated photometric redshift and absolute H-band magnitude in the COSMOS2015 catalog (Laigle et al. 2016). Among them, 6877 (2%) are classified as the likely counterparts of SMGs.

With this statistically large sample, we investigate the clustering properties of SMGs. We first divide our sample into three redshift bins (Figure 12) to study the evolution of SMGs and compare to clustering results in the literature (Hickox et al. 2012; Chen et al. 2016b; Wilkinson et al. 2017). We then calculate the two-point autocorrelation function (ACF) $w(\theta)$ for each subsample by using the Landy & Szalay (1993) estimator:

$$w(\theta) = \frac{(\mathrm{DD} - 2\mathrm{DR} + \mathrm{RR})}{\mathrm{RR}}, \qquad (1)$$

where DD, DR, and RR are the number of data-data, data-random, and random–random galaxy pairs in each θ bin respectively. The bright stars and bad pixels in the source-detection images of Laigle et al. (2016) have been masked out before the clustering analysis. We then generate a random sample within this masked region and with a sample size of ~1000 times larger than that of the machine-learning-identified counterparts in each redshift bin.

A power-law model $w(\theta) = A\,\theta^{-0.8}$ is assumed for the ACF of galaxies, which is suggested both observationally and

theoretically, at the physical scale of ~0.1–10 h^{-1} Mpc (Postman et al. 1998; Zehavi et al. 2002; Springel et al. 2005). However, because our sample is in a field comparable in size to the expected clustering length, the observed ACF needs to be corrected for the integral constraint (IC; Groth et al. 1977):

$$w(\theta) = A\theta^{-0.8} - \mathrm{IC}. \qquad (2)$$

The integral constraint can be numerically estimated (e.g., Infante 1994; Roche & Eales 1999; Adelberger et al. 2005) using the random–random pairs with the following form:

$$\mathrm{IC} \approx \int w(\theta)dF_r, \qquad (3)$$

where $F_r(\theta)$ is the cumulative distribution function of pair angular-separation (θ) estimated from RR counts.

The systematic uncertainties of ACF due to field-to-field variation is estimated by using the Jackknife resampling method (e.g., Norberg et al. 2009; Coupon et al. 2012; Chen et al. 2016a). In practice, we first divide our sample into $N_{\mathrm{sub}} = 32$ subsamples and remove one subsample at a time for each Jackknife realization. We then estimate the $w(\theta)_{jk}$ based on each Jackknife sample and repeat this process N_{sub} times. The covariance matrix is derived through the variance of these $w(\theta)_{jk}$. We then fit the observed ACF (Equation (2)) by performing a multivariate Gaussian fit on the scales of $8''$–$500''$ (~0.1–7 h^{-1} Mpc at $z = 2$) where the aforementioned covariance-matrix estimation is used to characterize the correlated uncertainties of the ACF estimator values in each angular-separation bin. The total uncertainties include Poisson noises of number counts and resampling variances, which are propagated into the uncertainty of the normalizing factor A of the ACF model in Equation (2). We show the best-fit results in Figure 12.

To investigate the relation of SMGs to other galaxy populations, we build two comparison samples of star-forming galaxies and passive galaxies using the COSMOS2015 catalog. The star-forming and quiescent galaxies are classified using the location of galaxies in the (NUV $-r$) versus ($r - J$) color–color plane (Laigle et al. 2016). Since the clustering properties of

galaxies depend on their star formation rate, stellar mass, and redshift (e.g., Wake et al. 2011; Mostek et al. 2013; Coupon et al. 2015; Wilkinson et al. 2017; Cochrane et al. 2018; Lin et al. 2019), we match the two comparison samples to our machine-learning-classified SMGs in redshift and stellar mass. About 85% of machine-learning-classified SMGs have stellar mass $\log_{10}(M_*/M_\odot) \geqslant 10.5$. Therefore, we limit all three samples with $\log_{10}(M_*/M_\odot) \geqslant 10.5$ in our clustering analyses.

For the selection of quiescent galaxies, the UVJ/UrJ color–color selection technique has proved to be more reliable at $z < 2$ (Williams et al. 2009; Ilbert et al. 2013). In addition, the sample size of quiescent galaxies decreases steeply beyond $z > 2$. Therefore, in our analyses, we only include quiescent galaxies at $1 < z < 2$. We find that 6.4% of these red UrJ sources comprise machine-learning-identified counterparts of submillimeter sources, which are very likely obscured dusty galaxies. However, the precision of our machine-learning classification being (83 ± 5)%, we cannot rule out that these sources might be real red passive galaxies. Hence we first keep these overlapped IDs and estimate the ACF for passive galaxies as shown in Figure 12. Then we remove machine-learning IDs from the quiescent sample and estimate their ACF again. We also estimate the effect on ACF of machine-learning-classified counterparts of SMGs by removing the IDs that are classified as quiescent galaxies. We consider the variances of $w(\theta)$ between these tests as the uncertainties when we calculate the amplitude of $w(\theta)$ for the corresponding sample.

To measure how well galaxies trace the underlying dark matter distribution, we compute the galaxy bias, which is quantified by the relationship

$$b_{\rm gal}^2 = \frac{w(\theta)}{w(\theta)_{\rm M}}, \qquad (4)$$

where $w(\theta)_{\rm M}$ is the two-point ACF of the total matter, including contributions from both cold dark matter (CDM) and baryons (Desjacques et al. 2018).

The $w(\theta)_{\rm M}$ is evaluated based on the small-angle approximation to the projection of isotropic density-fluctuation power spectrum $P(k, z)$ onto a transverse, two-dimensional surface (see Kaiser 1992; Baugh & Efstathiou 1993; Dodelson 2003; LoVerde & Afshordi 2008). Formally, the two-point correlation function is a Hankel transform of the power spectrum,

$$w(\theta)_{\rm M} = \frac{1}{2\pi} \int_0^\infty \int_0^\infty \frac{H(z)}{c} V^2(z) k J_0[\chi(z)k\theta] P(k, z) \, dz dk, \qquad (5)$$

where $H(z)$ is the Hubble parameter, $V(z)$ is the selection function in redshift z such that $\int_0^\infty V(z) \, dz = 1$, $\chi(z) = \int_0^z c/H(z') \, dz'$ is the radial comoving distance, and J_0 denotes the zeroth-order Bessel function of the first kind. We assume a flat ΛCDM cosmological model with parameters fixed at the *Planck* 2015 best-fit values (Planck Collaboration et al. 2016). The power spectrum $P(k, z)$ in this fiducial cosmology is computed using the software CAMB (Lewis et al. 2000, 2019), with the nonlinear evolution of CDM clustering described by the halo model (see Smith et al. 2003) as re-calibrated with cosmological N-body simulations by Takahashi et al. (2012). In each redshift bin, the selection function $V(z)$ is computed from the sources' photometric redshifts ($z_{\rm phot}$), the redshift uncertainties, and their SMG-classification probabilities (the last one only
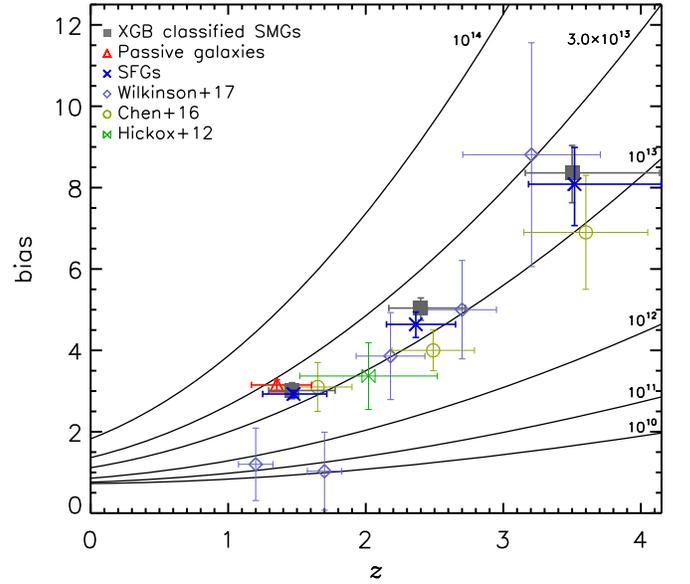
Figure 13. Galaxies bias as a function of redshift for machine-learning-classified counterparts of SMGs and comparison samples of passive and star-forming galaxies from COSMOS2015 catalog (Laigle et al. 2016). We also present the estimated bias of SMGs in the literature for comparison. The solid lines show the modeled bias as a function of redshift for various halo masses (labeled by the halo mass at $z = 0$ in solar mass, Mo & White 2002). Our measurements show that the typical host halo mass of SMGs with $\log_{10}(M_*/M_\odot) \geqslant 10.5$ is $(1.2 \pm 0.3) \times 10^{13} \, h^{-1} \, M_\odot$ at $z > 1$.

Table 3
Results of Clustering Measurements

Sample[a]	$z_{\rm median}$	$N_{\rm s}$[b]	$b_{\rm gal}$	$M_{\rm halo}$ $(10^{13} \, h^{-1} \, M_\odot)$
$z = 1–2$				
XGB IDs	$1.5^{+0.3}_{-0.2}$	2226	$3.0^{+0.2}_{-0.2}$	1.2 ± 0.3
Passive galaxies	$1.4^{+0.3}_{-0.2}$	3478	$3.2^{+0.1}_{-0.1}$	1.7 ± 0.2
SFGs	$1.5^{+0.3}_{-0.2}$	5337	$2.9^{+0.1}_{-0.1}$	1.0 ± 0.1
$z = 2–3$				
XGB IDs	$2.4^{+0.3}_{-0.2}$	2111	$5.0^{+0.2}_{-0.3}$	1.2 ± 0.3
SFGs	$2.4^{+0.3}_{-0.2}$	1885	$4.6^{+0.3}_{-0.3}$	1.0 ± 0.3
$z = 3–5$				
XGB IDs	$3.5^{+0.6}_{-0.3}$	1072	$8.4^{+0.7}_{-0.7}$	1.3 ± 0.4
SFGs	$3.5^{+0.6}_{-0.3}$	656	$8.1^{+0.9}_{-1.0}$	1.2 ± 0.5

Notes.
[a] The stellar mass limit of all subsamples is $\log_{10}(M_*/M_\odot) \geqslant 10.5$.
[b] The number of sources used in the clustering analyses.

for the machine-learning IDs). The technical details about computing the selection function $V(z)$ are presented in the Appendix.

We show our measured galaxies bias for machine-learning-classified counterparts of SMGs and two comparison samples at three redshift bins and compare them with the model predictions (Sheth et al. 2001; Mo & White 2002) in Figure 13 and Table 3. We also compare our clustering measurements of these three galaxies populations with the results in the literature (Hickox et al. 2012; Hartley et al. 2013; Chen et al. 2016b; Wilkinson et al. 2017; Amvrosiadis et al. 2018; Lin et al. 2019). Our results are consistent with the main results in the

literature, which suggest that the SMGs resided in high-mass ($(1.2 \pm 0.3) \times 10^{13} \, h^{-1} \, M_\odot$) halos. At $z \sim 1\text{--}2$, the passive galaxies show a slightly stronger clustering compared to machine-learning-classified counterparts of SMGs. Star-forming galaxies are less clustered compared with the other two populations. This is consistent with the evolutionary scenario that SMGs may be the progenitors of the most massive quiescent galaxies at low redshift.

5. Conclusion

By utilizing the high angular resolution ALMA data and rich ancillary data available in the COSMOS field, we employ our previously developed radio+machine-learning method to identify multi-wavelength counterparts of S2COSMOS single-dish-detected submillimeter sources. We then provide a large sample of SMGs with robustly identified radio/optical/NIR counterparts and study the physical and statistical properties of SMGs. Our main conclusions are as follows.

1. Using the deep VLA 3 GHz radio data in the COSMOS field, we identify 932 radio counterparts to the S2COSMOS submillimeter sources by adopting a p-value cut of $p \leqslant 0.065$. The expected precision of radio identification is 70% from the self-test of AS2UDS in An18.

2. We use three ALMA data sets in both COSMOS and UDS fields to build training sets for machine-learning algorithms and compare the performance of machine-learning classifiers trained on these different training sets. The sample constructed from the combined AS2COSMOS and A^{3}COSMOS data sets is chosen as the training set for this work because the machine-learning classifiers trained on this sample can best classify SMGs and non-SMGs in the COSMOS field. There are 5655 NIR galaxies located within SCUBA-2 error circles that meet the requirements of machine-learning classification. Among them, 658 are classified as optical/NIR counterparts of S2COSMOS submillimeter sources by the machine-learning classifier. Combining with the radio identification, we identify 1222 radio/optical/NIR counterparts to 897 of the 1145 single-dish-detected submillimeter sources in the COSMOS field. The identification rate is $(78 \pm 9)\%$ and it increases for bright SCUBA-2 sources.

3. For the 897 S2COSMOS sources that have at least one radio- or machine-learning-identified counterpart, $(26 \pm 5)\%$ of them have multiple counterparts. The multiple fraction increases with the flux densities of single-dish submillimeter sources. We estimate that $\sim 27\%$ of the multiple counterparts within the same SCUBA-2 error circles arise from physically associated galaxies by comparing the difference of their photometric redshift.

4. We study the physical properties of the 1222 radio +machine-learning-identified counterparts to the S2COSMOS submillimeter sources. The redshift distribution of these counterparts peaks at $z = 2.3 \pm 0.1$ and has a redshift range of $z = 0.2\text{--}5.7$, which is consistent with that of ALMA SMGs from AS2UDS and ALESS surveys. The AGN fraction of our radio+machine-learning-identified counterparts to S2COSMOS sources is $(19 \pm 4)\%$, which is similar to the AGN fraction in the ALESS and AS2UDS samples. Compared with NIR or

radio-selected galaxies in the COSMOS field, our radio +machine-learning-identified counterparts of S2COSMOS sources have higher star formation rates and higher stellar masses. These results mean that our radio +machine-learning-identified counterparts constitute a comprehensive and representative sample of SMGs indicated by their physical properties.

5. We investigate the environment properties of bright SCUBA-2 sources and find a significant excess of NIR galaxies at higher redshifts and brighter NIR magnitudes within SCUBA-2 error circles compared to those within the random fields. We find that the excess of NIR galaxies can be roughly explained by the contribution from SMGs within these regions.

6. Among the 1145 S2COSMOS submillimeter sources that lie within the coverage of radio or NIR observations in the COMSOS field, 248 of them do not have any radio- or machine-learning-identified counterparts. We study the properties of these 248 "blank"-SCUBA-2 sources and confirm that the main cause of the lack of an identified counterpart is the incompleteness of our radio+machine-learning method.

7. We employ our machine-learning technique to the whole COSMOS field and identify 6877 optical/NIR counterparts of faint SMGs, whose submillimeter emission lies below the confusion limit of our S2COSMOS submillimeter surveys ($S_{850 \, \mu m} \lesssim 1.5$ mJy). By using this statistically large sample of SMGs with precisely identified multi-wavelength counterparts, we investigate the clustering properties of this galaxies population. The clustering measurements show that SMGs reside in CDM halos with mass of $\sim (1.2 \pm 0.3) \times 10^{13} \, h^{-1} \, M_\odot$, which is relatively unchanged across cosmic time. We compare the clustering strength and galaxies bias of SMGs to those of SFGs and passive galaxies at the similar redshift ranges and with the same stellar mass limit of $\log_{10}(M_*/M_\odot) \geqslant 10.5$. We find that at $z = 1\text{--}2$, passive galaxies show a slightly stronger clustering compared with SMGs. Star-forming galaxies are less clustered compared with SMGs at $z = 1\text{--}5$. These results are consistent with the suggested scenario that SMGs may be the progenitors of most massive quiescent galaxies in the low-redshift universe.

Appendix
Redshift Selection Function

Following the discussions of Chen et al. (2016b), in the absence of accurate information about intrinsic redshift and luminosity distributions for the heterogeneous sample (Laigle et al. 2016), we may approximate the selection function $V(z)$ by the PDF of the sample's estimated redshifts in each slice (see also Baugh & Efstathiou 1993). To do this, it is necessary to account for the redshift uncertainties. In this work, we use a mixture model as our underlying probabilistic assumption, to be detailed as follows.

We denote the redshift (cumulative) distribution function (CDF) of the ith source, as indicated by its z_{phot} estimation and uncertainty, by $F_i(z)$, and let p_i be the classification probability of the same source. We consider the mutually exclusive, equal-probability collection of random events S_i, defined as $S_i = \{$the ith source being selected$\}$, such that $\Pr(\bigcup_{i=0}^{N} S_i) = 1$, where N

is the total number of sources in a catalog. Our assumption amounts to $\Pr(S_i) = 1/N$. Furthermore, we identify each p_i as the conditional probability $\Pr(C|S_i)$, where C is the event of inclusion by the classifier. The conditional CDF,

$$F_{\mathrm{Z}}(z) = \Pr(\{Z < z\} \mid C), \qquad (6)$$

is the antiderivative of the selection function $V(z)$ for the whole catalog. To express the selection function based on only a subset (redshift slice) of the catalog, we denote the index set of the sources in the slice by $\mathcal{J}$ and let $S_{\mathcal{J}} = \bigcup_{i \in \mathcal{J}} S_i$. The CDF specific to (and conditional upon) the slice identified by the indices in $\mathcal{J}$ is

$$
\begin{aligned}
F_{\mathrm{Z}}(z; \mathcal{J}) &= \Pr(\{Z < z\} \mid C \cap S_{\mathcal{J}}) \\
&= \frac{1}{\Pr(C \cap S_{\mathcal{J}})} \sum_{i \in \mathcal{J}} \Pr(\{Z < z\} \mid C \cap S_i) \\
&\quad \times \Pr(C \mid S_i)\Pr(S_i) \\
&= \frac{1}{N\Pr(C \cap S_{\mathcal{J}})} \sum_{i \in \mathcal{J}} F_i(z)p_i, \qquad (7)
\end{aligned}
$$

following the definition of conditional probability and the theorem of total probability. The numerical value of $\Pr(C \cap S_{\mathcal{J}})$ is determined by the normalization to unity, $F_{\mathrm{Z}}(+\infty; \mathcal{J}) = 1$, i.e.,

$$\Pr(C \cap S_{\mathcal{J}}) = \frac{1}{N} \sum_{i \in \mathcal{J}} p_i. \qquad (8)$$

It follows that the conditional CDF to be found is a weighted sum of individual components (the mixture model), as in

$$F_{\mathrm{Z}}(z; \mathcal{J}) = \frac{\sum_{i \in \mathcal{J}} p_i F_i(z)}{\sum_{i \in \mathcal{J}} p_i} = \sum_{i \in \mathcal{J}} w_i F_i(z), \qquad (9)$$

where $w_i = p_i/(\sum_{i \in \mathcal{J}} p_i)$ is the weight. Thus, the selection function for the redshift slice is

$$V(z; \mathcal{J}) = \frac{dF_{\mathrm{Z}}(z; \mathcal{J})}{dz} = \sum_{i \in \mathcal{J}} w_i f_i(z), \qquad (10)$$

where f_i is the PDF characterizing the z_{phot} uncertainty of the ith source. In our work, the individual f_is (see Equation (10)) are
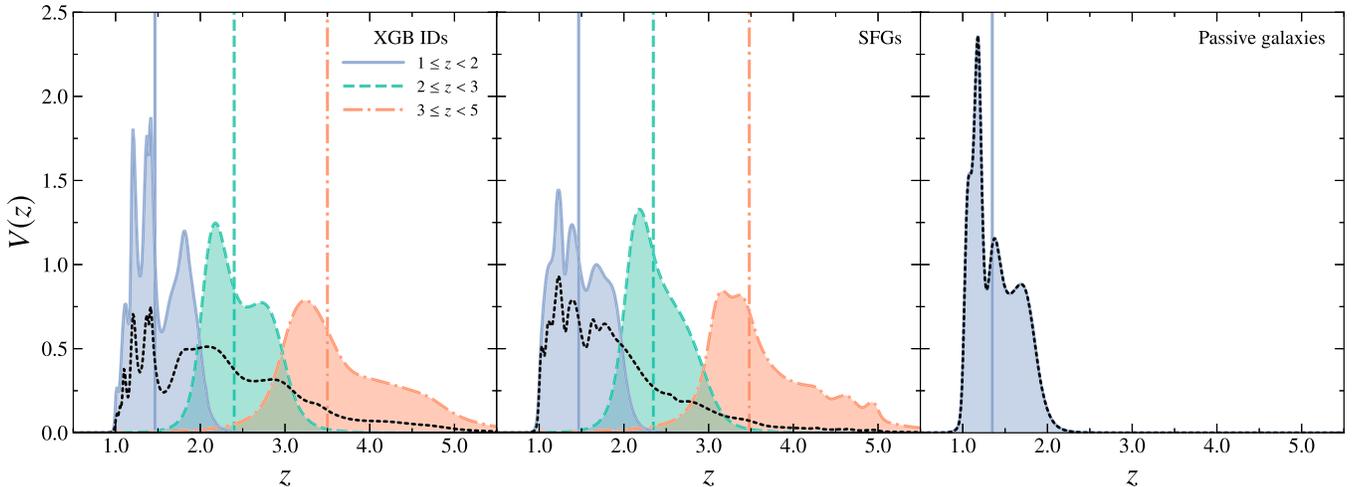
Figure 14. Selection functions for subsamples from each redshift bin, with the redshift PDF of the combined sample shown for comparison (dotted lines). Each selection function reflects the aggregated probability density of source redshift from the respective subsample, and is normalized to unity individually. Vertical lines mark the median redshifts of the corresponding subsample (see Table 3).

modeled by Gaussian distributions indicated by z_{phot} and the 68.3% uncertainty bounds. As a result of such uncertainties, the selection function based on a slice can spread outside its redshift cut and overlap with neighboring ones, as can be observed in Figure 14. Also shown therein is the selection function built from the full catalog, which is comparable to the Figure 3 of Chen et al. (2016b).

ORCID iDs

Fang Xia An ⓘ https://orcid.org/0000-0001-7943-0166
J. M. Simpson ⓘ https://orcid.org/0000-0002-8521-1995
Ian Smail ⓘ https://orcid.org/0000-0003-3037-257X
A. M. Swinbank ⓘ https://orcid.org/0000-0003-1192-5837
Cong Ma ⓘ https://orcid.org/0000-0001-6109-2365
Daizhong Liu ⓘ https://orcid.org/0000-0001-9773-7479
P. Lang ⓘ https://orcid.org/0000-0002-5681-3575
E. Schinnerer ⓘ https://orcid.org/0000-0002-3933-7677
A. Karim ⓘ https://orcid.org/0000-0002-8414-9579
B. Magnelli ⓘ https://orcid.org/0000-0002-6777-6490
S. Leslie ⓘ https://orcid.org/0000-0002-4826-8642
F. Bertoldi ⓘ https://orcid.org/0000-0002-1707-1775
Chian-Chou Chen ⓘ https://orcid.org/0000-0002-3805-0789
S. M. Stach ⓘ https://orcid.org/0000-0003-1122-6948
J. L. Wardlow ⓘ https://orcid.org/0000-0003-2376-8971
R. J. Ivison ⓘ https://orcid.org/0000-0001-5118-1313
Y. Ao ⓘ https://orcid.org/0000-0003-3139-2724
R. Wang ⓘ https://orcid.org/0000-0003-4956-5742
Wei-Hao Wang ⓘ https://orcid.org/0000-0003-2588-1265
Y. Yang ⓘ https://orcid.org/0000-0003-3078-2763
K. Coppin ⓘ https://orcid.org/0000-0002-0729-2988
N. K. Hine ⓘ https://orcid.org/0000-0002-7975-0474
L. C. Ho ⓘ https://orcid.org/0000-0001-6947-5846
H. S. Hwang ⓘ https://orcid.org/0000-0003-3428-7612
K. Lacaille ⓘ https://orcid.org/0000-0001-9870-5663
I. Oteo ⓘ https://orcid.org/0000-0001-5875-3388

References

Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697
Amvrosiadis, A., Eales, S. A., Negrello, M., et al. 2018, MNRAS, 475, 4939
An, F. X., Stach, S. M., Smail, I., et al. 2018, ApJ, 862, 101
Aravena, M., Decarli, R., Walter, F., et al. 2016, ApJ, 984, 68
Aretxaga, I., Wilson, G. W., Aguilar, E., et al. 2011, MNRAS, 415, 3831
Barger, A. J., Cowie, L. L., Sanders, D. B., et al. 1998, Natur, 394, 248
Barger, A. J., Cowie, L. L., Smail, I., et al. 1999, AJ, 117, 2656
Barger, A. J., Wang, W.-H., Cowie, L. L., et al. 2012, ApJ, 761, 89
Baugh, C. M., & Efstathiou, G. 1993, MNRAS, 265, 145
Bertoldi, F., Carilli, C., Aravena, M., et al. 2007, ApJS, 172, 132
Béthermin, M., Le Floc'h, E., Ilbert, O., et al. 2012, A&A, 542, A58
Biggs, A. D., & Ivison, R. J. 2008, MNRAS, 385, 893
Bothwell, M. S., Smail, I., Chapman, S. C., et al. 2013, MNRAS, 429, 3052
Brusa, M., Civano, F., Comastri, A., et al. 2010, ApJ, 716, 348
Cappelluti, N., Hasinger, G., Brusa, M., et al. 2007, ApJS, 172, 341
Casey, C. M., Narayanan, D., & Cooray, A. 2014, PhR, 541, 45
Chabrier, G. 2003, PASP, 115, 763
Chapin, E. L., Berry, D. S., Gibb, A. G., et al. 2013, MNRAS, 430, 2545
Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, ApJ, 622, 772
Chen, C.-C., Smail, I., Ivison, R. J., et al. 2016a, ApJ, 820, 82
Chen, C.-C., Smail, I., Swinbank, A. M., et al. 2016b, ApJ, 831, 91
Chen, T., & Guestrin, C. 2016, arXiv:1603.02754
Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62
Clements, D. L., Bendo, G., Pearson, C., et al. 2011, MNRAS, 411, 373
Cochrane, R. K., Best, P. N., Sobral, D., et al. 2018, MNRAS, 475, 3730
Cooke, E. A., Smail, I., Swinbank, A. M., et al. 2018, ApJ, 861, 100
Cooray, A., & Sheth, R. 2002, PhR, 372, 1
Coppin, K., Chapin, E. L., Mortier, A. M. J., et al. 2006, MNRAS, 368, 1621
Coupon, J., Arnouts, S., van Waerbeke, L., et al. 2015, MNRAS, 449, 1352
Coupon, J., Kilbinger, M., McCracken, H. J., et al. 2012, A&A, 542, A5
Cowie, L. L., Barger, A. J., Hsu, L.-Y., et al. 2017, ApJ, 837, 139
Cowie, L. L., Barger, A. J., & Kneib, J.-P. 2002, AJ, 123, 2197
Danielson, A. L. R., Swinbank, A. M., Smail, I., et al. 2017, ApJ, 840, 78
Desjacques, V., Jeong, D., & Schmidt, F. 2018, PhR, 733, 1
Dodelson, S. 2003, Modern Cosmology (San Diego, CA: Academic)
Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, ApJ, 748, 142
Downes, A. J. B., Peacock, J. A., Savage, A., & Carrie, D. R. 1986, MNRAS, 218, 31
Dudzevičiūtė, U., Smail, I., & Swinbank, S. M. 2019, arXiv:1910.07524
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Dunlop, J. S., Peacock, J. A., Savage, A., et al. 1989, MNRAS, 238, 1171
Eales, S., Lilly, S., Gear, W., et al. 1999, ApJ, 515, 518
Fawcett, T. 2004, Machine Learning, 31, 1
Franco, M., Elbaz, D., Béthermin, M., et al. 2018, A&A, 620, A152
Frayer, D. T., Seaquist, E. R., Thuan, T. X., et al. 1998, ApJ, 503, 231
Fujimoto, S., Ouchi, M., Ono, Y., et al. 2016, ApJS, 222, 1
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789
Gear, W. K., Lilly, S. J., Stevens, J. A., et al. 2000, MNRAS, 316, L51
Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091
Groth, E. J., Peebles, P. J. E., Seldner, M., et al. 1977, SciAm, 237, 76
Gullberg, B., Swinbank, A. M., Smail, I., et al. 2018, ApJ, 859, 12
Hartley, W. G., Almaini, O., Mortlock, A., et al. 2013, MNRAS, 431, 3045
Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, ApJS, 172, 29
Hickox, R. C., Wardlow, J. L., Smail, I., et al. 2012, MNRAS, 421, 284
Hodge, J. A., Karim, A., Smail, I., et al. 2013, ApJ, 768, 91
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2008, ApJS, 175, 356
Hughes, D. H., Serjeant, S., Dunlop, J., et al. 1998, Natur, 394, 241
Ikarashi, S., Kohno, K., Aguirre, J. E., et al. 2011, MNRAS, 415, 3081
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Infante, L. 1994, A&A, 282, 353
Ishibashi, W., & Fabian, A. C. 2016, MNRAS, 463, 1291
Ivison, R. J., Greve, T. R., Dunlop, J. S., et al. 2007, MNRAS, 380, 199
Ivison, R. J., Greve, T. R., Smail, I., et al. 2002, MNRAS, 334, 1
Ivison, R. J., Smail, I., Le Borgne, J.-F., et al. 1998, MNRAS, 298, 583
Jauncey, D. L. 1968, ApJ, 152, 647
Jin, S., Daddi, E., Liu, D., et al. 2018, ApJ, 864, 56
Kaiser, N. 1992, ApJ, 388, 272
Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, MNRAS, 432, 2
Kohavi, R. 1995, Proc. of the Int. Joint Conf. on Artificial Intelligence 14 (San Francisco, CA: Morgan Kaufmann Publishers), 1137
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Lewis, A., Vehreschild, A., Mead, A., et al. 2019, CAMB: Code for Anisotropies in the Microwave Background, 1.0.7, Zenodo
Lin, X., Fang, G., Cai, Z.-Y., et al. 2019, ApJ, 875, 83
Lindner, R. R., Baker, A. J., Omont, A., et al. 2011, ApJ, 737, 83
Liu, D., Daddi, E., Dickinson, M., et al. 2018, ApJ, 853, 172
Liu, D., Lang, P., Magnelli, B., et al. 2019, ApJS, 244, 40
Liu, R. H., Hill, R., Scott, D., et al. 2019, MNRAS, 489, 1770
LoVerde, M., & Afshordi, N. 2008, PhRvD, 78, 123506
Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, A&A, 532, A90
Marchesi, S., Civano, F., Elvis, M., et al. 2016, ApJ, 817, 34
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, adass XVI, 376, 127
Michałowski, M. J., Dunlop, J. S., Ivison, R. J., et al. 2012, MNRAS, 426, 1845
Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
Mo, H. J., & White, S. D. M. 2002, MNRAS, 336, 112
Mostek, N., Coil, A. L., Cooper, M., et al. 2013, ApJ, 767, 89
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Norberg, P., Baugh, C. M., Gaztañaga, E., et al. 2009, MNRAS, 396, 19
Oke, J. B. 1974, ApJS, 27, 21
Oliver, S. J., Wang, L., Smith, A. J., et al. 2010, A&A, 518, L21
Perna, M., Sargent, M. T., Brusa, M., et al. 2018, A&A, 619, A90
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Pope, A., Scott, D., Dickinson, M., et al. 2006, MNRAS, 369, 1185
Postman, M., Lauer, T. R., Szapudi, I., et al. 1998, ApJ, 506, 33
Roche, N., & Eales, S. A. 1999, MNRAS, 307, 703
Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988, ApJL, 328, L35
Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, ApJS, 188, 384
Scott, K. S., Wilson, G. W., Aretxaga, I., et al. 2012, MNRAS, 423, 575
Scott, S. E., Fox, M. J., Dunlop, J. S., et al. 2002, MNRAS, 331, 817

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015a, ApJ, 807, 128
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015b, ApJ, 815, 81
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2017, ApJ, 839, 58
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2019, ApJ, 880, 43
Simpson, J. M., Swinbank, A. M., Smail, I., et al. 2014, ApJ, 788, 125
Smail, I., Ivison, R. J., & Blain, A. W. 1997, ApJL, 490, L5
Smail, I., Ivison, R. J., Blain, A. W., & Kneib, J.-P. 2002, MNRAS, 331, 495
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Smolčić, V., Aravena, M., Navarrete, F., et al. 2012, A&A, 548, A4
Smolčić, V., Delvecchio, I., Zamorani, G., et al. 2017a, A&A, 602, A2
Smolčić, V., Novak, M., Bondi, M., et al. 2017b, A&A, 602, A1
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Natur, 435, 629
Stach, S. M., Dudzevičiūtė, U., Smail, I., et al. 2019, MNRAS, 487, 4648
Stach, S. M., Smail, I., Swinbank, A. M., et al. 2018, ApJ, 860, 161
Swinbank, A. M., Chapman, S. C., Smail, I., et al. 2006, MNRAS, 371, 465
Swinbank, A. M., Dye, S., Nightingale, J. W., et al. 2015, ApJL, 806, L17
Swinbank, A. M., Simpson, J. M., Smail, I., et al. 2014, MNRAS, 438, 1267
Tacconi, L. J., Neri, R., Chapman, S. C., et al. 2006, ApJ, 640, 228
Takahashi, R., Sato, M., Nishimichi, T., et al. 2012, ApJ, 761, 152
Targett, T. A., Dunlop, J. S., McLure, R. J., et al. 2011, MNRAS, 412, 295
Thomson, A. P., Ivison, R. J., Simpson, J. M., et al. 2014, MNRAS, 442, 577
Toft, S., Smolčić, V., Magnelli, B., et al. 2014, ApJ, 782, 68
Umehata, H., Hatsukade, B., Smail, I., et al. 2018, PASJ, 70, 65
Vapnik, V. N. 1995, The Nature of Statistical Learning Theory (Berlin: Springer)
Wake, D. A., Whitaker, K. E., Labbé, I., et al. 2011, ApJ, 728, 46
Walter, F., Decarli, R., Aravena, M., et al. 2016, ApJ, 984, 67
Wang, S. X., Brandt, W. N., Luo, B., et al. 2013, ApJ, 778, 179
Wang, T., Schreiber, C., Elbaz, D., et al. 2019, Natur, 572, 211
Wang, W.-H., Cowie, L. L., Barger, A. J., & Williams, J. P. 2011, ApJL, 726, L18
Wang, W.-H., Lin, W.-C., Lim, C.-F., et al. 2017, ApJ, 850, 37
Wardlow, J. L., Cooray, A., Osage, W., et al. 2017, ApJ, 837, 12
Wardlow, J. L., Simpson, J. M., Smail, I., et al. 2018, MNRAS, 479, 3879
Weiß, A., Kovács, A., Coppin, K., et al. 2009, ApJ, 707, 1211
Wilkinson, A., Almaini, O., Chen, C.-C., et al. 2017, MNRAS, 464, 1380
Williams, C. C., Giavalisco, M., Porciani, C., et al. 2011, ApJ, 733, 92
Williams, R. J., Quadri, R. F., Franx, M., et al. 2009, ApJ, 691, 1879
Yamamura, I., Makiuti, S., Ikeda, N., et al. 2010, yCat, 298, 0
Yun, M. S., Scott, K. S., Guo, Y., et al. 2012, MNRAS, 420, 957
Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172