SCUBA-2 Ultra Deep Imaging EAO Survey (STUDIES). IV. Spatial Clustering and Halo Masses of Submillimeter Galaxies

Chen-Fatt Lim[1,2], Chian-Chou Chen[1,3], Ian Smail[4], Wei-Hao Wang[1], Wei-Leong Tee[1,5], Yen-Ting Lin[1], Douglas Scott[6], Yoshiki Toba[1,7,8], Yu-Yen Chang[1], YiPing Ao[9], Arif Babul[10], Andy Bunker[11,28], Scott C. Chapman[12,13,14], David L. Clements[15], Christopher J. Conselice[16], Yu Gao[9,17], Thomas R. Greve[18,19], Luis C. Ho[20,21], Sungwook E. Hong[22], Ho Seong Hwang[23], Maciej Koprowski[24], Michał J. Michałowski[25], Hyunjin Shim[26], Xinwen Shu[27], and James M. Simpson[1]

[1] Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan; chenfatt.lim@gmail.com
[2] Graduate Institute of Astrophysics, National Taiwan University, Taipei 10617, Taiwan
[3] European Southern Observatory, Karl Schwarzschild Strasse 2, Garching, Germany
[4] Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK
[5] Department of Physics, University of Arizona, Tucson, AZ 85721, USA
[6] Department of Physics & Astronomy, University of British Columbia, BC, Canada
[7] Department of Astronomy, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
[8] Research Center for Space and Cosmic Evolution, Ehime University, 2-5 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan
[9] Purple Mountain Observatory and Key Laboratory for Radio Astronomy, Chinese Academy of Sciences, Nanjing 210033, People's Republic of China
[10] Department of Physics and Astronomy, University of Victoria, Elliott Building, 3800 Finnerty Road, Victoria, BC V8P 5C2, Canada
[11] Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX13RH, UK
[12] Department of Physics and Astronomy, University of British Columbia, 6225 Agricultural Road, Vancouver, BC, V6T 1Z1, Canada
[13] National Research Council, Herzberg Astronomy and Astrophysics, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada
[14] Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS B3H 4R2, Canada
[15] Department of Physics, Blackett Lab, Imperial College, Prince Consort Road, London, SW7 2AZ, UK
[16] University of Nottingham, School of Physics & Astronomy, Nottingham, NG7 2RD, UK
[17] Department of Astronomy, Xiamen University, Xiamen, Fujian 361005, People's Republic of China
[18] Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[19] Cosmic Dawn Center
[20] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China
[21] Department of Astronomy, School of Physics, Peking University, Beijing 100871, People's Republic of China
[22] Natural Science Research Institute, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Republic Of Korea
[23] Korea Astronomy and Space Science Institute, 776 Daedeokdae-ro, Yuseong-gu, Daejeon 34055, Republic Of Korea
[24] Institute of Astronomy, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Torun, Poland
[25] Astronomical Observatory Institute, Faculty of Physics, Adam Mickiewicz University, 60-286 Poznań, Poland
[26] Department of Earth Science Education, Kyungpook National University, Deagu 41566, Republic Of Korea
[27] Department of Physics, Anhui Normal University, Wuhu, Anhui, 241000, People's Republic of China

Abstract

We analyze an extremely deep 450 μm image ($1\sigma = 0.56$ mJy beam^{-1}) of a $\simeq$300 arcmin^2 area in the CANDELS/COSMOS field as part of the Sub-millimeter Common User Bolometric Array-2 Ultra Deep Imaging EAO Survey. We select a robust (signal-to-noise ratio $\geqslant$4) and flux-limited ($\geqslant$4 mJy) sample of 164 submillimeter galaxies (SMGs) at 450 μm that have K-band counterparts in the COSMOS2015 catalog identified from radio or mid-infrared imaging. Utilizing this SMG sample and the 4705 K-band-selected non-SMGs that reside within the noise level $\leqslant$1 mJy beam^{-1} region of the 450 μm image as a training set, we develop a machine-learning classifier using K-band magnitude and color–color pairs based on the 13-band photometry available in this field. We apply the trained machine-learning classifier to the wider COSMOS field (1.6 deg^2) using the same COSMOS2015 catalog and identify a sample of 6182 SMG candidates with similar colors. The number density, radio and/or mid-infrared detection rates, redshift and stellar-mass distributions, and the stacked 450 μm fluxes of these SMG candidates, from the S2COSMOS observations of the wide field, agree with the measurements made in the much smaller CANDELS field, supporting the effectiveness of the classifier. Using this SMG candidate sample, we measure the two-point autocorrelation functions from $z = 3$ down to $z = 0.5$. We find that the SMG candidates reside in halos with masses of $\simeq(2.0 \pm 0.5) \times 10^{13}\, h^{-1}\, M_\odot$ across this redshift range. We do not find evidence of downsizing that has been suggested by other recent observational studies.

Unified Astronomy Thesaurus concepts: Clustering (1908); Submillimeter astronomy (1647); High-redshift galaxies (734); Galaxy evolution (594); Large-scale structure of the universe (902); Galaxy formation (595)

1. Introduction

Over the past two decades, a class of far-infrared luminous galaxies has been discovered at submillimeter wavelengths.

The extreme infrared luminosities observed in these submillimeter galaxies (SMGs) suggest that they are dusty and considered to be among the most intensively star-forming sources in the universe (Smail et al. 1997; Barger et al. 1998, 1999; Hughes et al. 1998; Eales et al. 1999). SMGs appear to have a redshift distribution peaking at $z \simeq 2.5$ with the majority of them at $z = 1.5$–3.5 (Barger et al. 2000;

---

[28] Visitor, Kavli Institute for the Physics and Mathematics of the Universe (WPI), the University of Tokyo Institutes for Advanced Study, Japan.

Chapman et al. 2003, 2005; Pope et al. 2006; Aretxaga et al. 2007; Wardlow et al. 2011; Michałowski et al. 2012b, 2017; Yun et al. 2012; Simpson et al. 2014, 2017; Chen et al. 2016; Koprowski et al. 2016; Danielson et al. 2017; Dunlop et al. 2017; Stach et al. 2019; Dudzevičiūtė et al. 2020), occupying the same putative peak epoch of unobscured star formation (Madau & Dickinson 2014) and active galactic nucleus (AGN) activity (Schmidt et al. 1995; Hasinger et al. 2005; Wall et al. 2008; Assef et al. 2011). The total infrared luminosities (L_{IR}; 8–1000 μm) of SMGs are similar to local ultra-luminous infrared galaxies (Sanders et al. 1988; Sanders & Mirabel 1996; Farrah et al. 2001; Armus et al. 2009), reaching values greater than a few times $10^{12} L_\odot$ or even higher than $10^{13} L_\odot$ for some of the brightest sources. This corresponds to star formation rates (SFRs) ranging from around $100\, M_\odot\, \mathrm{yr}^{-1}$ to more than $1000\, M_\odot\, \mathrm{yr}^{-1}$ (Hainline et al. 2011; Barger et al. 2012; Swinbank et al. 2014; da Cunha et al. 2015; Simpson et al. 2015; Dudzevičiūtė et al. 2020).

Violent gas accretion, potentially induced by mergers, is the most likely explanation to date for the intensive star formation of SMGs (Frayer et al. 1998; Conselice et al. 2003; Greve et al. 2005; Tacconi et al. 2008; Engel et al. 2010; Swinbank et al. 2010; Alaghband-Zadeh et al. 2012; Menéndez-Delmestre et al. 2013; Chen et al. 2015; Koprowski et al. 2016; Michałowski et al. 2017; Chang et al. 2018). Large amounts of gas accretion can result in a short-lived starburst possibly followed by a quasar phase. Feedback mechanisms from star formation or black hole accretion could have injected sufficient energy to heat the remaining gas or expel it from the galaxy to prevent further star formation (Silk & Rees 1998; Fabian 1999; Trayford et al. 2016). This scenario may be responsible for the formation of the most massive ($M_* > 10^{11} M_\odot$) elliptical galaxies in the local universe (Lilly et al. 1999; Hopkins et al. 2005; Simpson et al. 2014; Toft et al. 2014; Dudzevičiūtė et al. 2020; Rennehan et al. 2020). Therefore, the cosmological evolution of SMGs is crucial for our understanding of the formation of massive galaxies in the universe.

Comparison of clustering measurements with dark matter simulations can provide constraints on the masses of dark matter halos that a given galaxy population resides in (Peebles 1980) and further trace the evolution of the given galaxy population. Previous clustering analyses of SMGs identified in shorter (250–500 μm; Cooray et al. 2010; Maddox et al. 2010; Mitchell-Wynne et al. 2012; van Kampen et al. 2012; Amvrosiadis et al. 2019) and longer (850–1100 μm; Scott et al. 2002; Webb et al. 2003; Weiß et al. 2009; Lindner et al. 2011; Williams et al. 2011; Hickox et al. 2012; Wilkinson et al. 2017; An et al. 2019) submillimeter wave bands have revealed that SMGs reside in high-mass (10^{12}–$10^{13}\, h^{-1} M_\odot$) dark matter halos. These values are also consistent with previous estimates from a sample of obscured starburst galaxies reported by Béthermin et al. (2014), in which they used a combined BzK color criterion and Herschel/PACS data to study the clustering signal of obscured starburst galaxies at $1.5 < z < 2.5$ as a function of their physical parameters. These results suggest that SMGs may be the progenitors of massive elliptical galaxies in the local universe (Hughes et al. 1998; Eales et al. 1999; Swinbank et al. 2006; Targett et al. 2011; Miller et al. 2018; Wang et al. 2019). However, many of these previous studies were limited by either the modest samples of SMGs ($\simeq100$ sources) or a lack of reliable identifications and redshift measurements, which makes their estimated clustering signals highly uncertain.

More precise determinations of the clustering properties with sizable SMG samples have been performed by Chen et al. (2016), Wilkinson et al. (2017), Amvrosiadis et al. (2019), and An et al. (2019). Chen et al. (2016) identified a sample of $\simeq3000$ faint SMGs ($S_{850\,\mu m} < 2$ mJy) using a color selection technique and compared their clustering properties with other galaxy populations in the redshift range $1 < z < 5$. Wilkinson et al. (2017) performed a clustering analysis using a sample of $\simeq600$ 850 μm–selected SMGs in the UKIDSS Ultra Deep Survey field in the redshift interval $1 < z < 3$. Amvrosiadis et al. (2019) studied the clustering properties for a sample of $\simeq120,000$ Herschel-selected SMGs with flux densities $S_{250\,\mu m} > 30$ mJy in low ($z < 0.3$) and high ($1 < z < 5$) redshift intervals. An et al. (2019) identified $\simeq7000$ potential 850 μm–selected SMGs in the COSMOS field based on a radio +machine-learning method trained on the Atacama Large Millimeter/submillimeter Array (ALMA)–identified sample (An et al. 2018) and studied their clustering properties.

These aforementioned recent studies were able to measure the clustering signals from SMGs as a function of redshift. Wilkinson et al. (2017) found that the clustering appears to exhibit tentative evolution with redshift, such that the SMG activity seems to be shifting to less massive halos at lower redshifts $z = 1$–2 and consistent with the downsizing scenario (Cowie et al. 1996; Magliocchetti et al. 2014; Rennehan et al. 2020) that the contribution of luminous sources dominates in the early universe, whereas the growth of the less luminous ones continues at lower redshifts. In contrast, Chen et al. (2016), Amvrosiadis et al. (2019), and An et al. (2019) did not find such a trend, suggesting that SMGs reside in a typical halo mass of about $10^{13}\, h^{-1} M_\odot$ across the redshift range $1 < z < 5$. The discrepancies in the lower-redshift bins could be simply caused by the measurement uncertainties (uncertain identifications and/or poor redshift measurements) or by the different methodologies that are adopted in the clustering analyses, where Wilkinson et al. (2017) relied on the cross-correlation technique with an abundant K-band selected sample, while Chen et al. (2016), Amvrosiadis et al. (2019), and An et al. (2019) adopted an autocorrelation technique.

However, these studies did not probe the clustering signals in a key redshift range ($0.3 < z < 1.0$; cosmic time ranges from 6.0 to 10.4 Gyr) in which the downsizing effect, if it exists, is expected to increase. This is likely caused by the longer-wavelength observations being more sensitive to high-redshift sources. The traditional 850 μm selection allows us to measure clustering at $z > 1$ (Chen et al. 2016; Wilkinson et al. 2017; An et al. 2019), whereas the studies based on Herschel at shorter wavelengths (e.g., 250 μm; van Kampen et al. 2012; Amvrosiadis et al. 2019) are mainly sensitive to the brightest low-redshift sources ($S_{250\,\mu m} > 30$ mJy) due to their positive K-correction. Using spectral energy distribution (SED) template fitting on the far-infrared photometry to estimate redshifts for sources without optical counterparts, Amvrosiadis et al. (2019) extended the Herschel-based clustering studies to $z > 1$, finding results consistent with those obtained from the 850 μm selection. However, the large redshift uncertainties ($\simeq0.3$ for $z = 1$; Amvrosiadis et al. 2019) have the consequence that they cannot meaningfully measure the clustering signals in the low-redshift regime.

Observations at mid-infrared (e.g., 24 μm) can be used to overcome the aforementioned selection bias and probe the clustering signals in the key redshift range of $0.3 < z < 1.0$ (Gilli et al. 2007; Magliocchetti et al. 2008; Starikova et al. 2012; Dolley et al. 2014; Solarz et al. 2015). These studies found relatively lower clustering strengths at $z < 1$ with clustering length $r_0 = 3$–$6\,h^{-1}$ Mpc compared to the high-redshift measurements. This finding is similar to the work of Magliocchetti et al. (2013) based on Herschel/PACS 100 μm–selected sources, in which they found clear evidence for the downsizing effect at redshifts limited to $z \lesssim 1$. However, the clustering lengths also correlate with the infrared luminosities, where the galaxies with higher L_{IR} (higher SFRs) tend to have stronger clustering signals (Dolley et al. 2014; Toba et al. 2017). The majority of the sources in the above studies at 24 and 100 μm are biased toward a fainter population with $L_{IR} \simeq 10^{11}\,L_\odot$ at $z < 1$, which prevents us from making a fair comparison with the SMGs at $z > 1$ that have $L_{IR} > 10^{12}\,L_\odot$.

In this paper, we base our analysis on the 450 μm data obtained from the Sub-millimeter Common User Bolometric Array-2 (SCUBA-2; Holland et al. 2013) camera on the 15 m James Clerk Maxwell Telescope (JCMT). The 450 μm observations allow us to obtain photometric measurements closer to the redshifted SED peak of typical SMGs ($\lambda_{rest} \simeq 100\,\mu$m) so they provide a closer match to far-infrared luminosity selection compared to longer-wavelength observations. The 450 μm observations also allow us to probe the SMGs at lower redshifts ($z \simeq 1.5$), with the majority of them at $z = 0.5$–3.0 (Casey et al. 2013; Simpson et al. 2014; Bourne et al. 2017; Zavala et al. 2018; Lim et al. 2020). We have pushed the frontier of the 450 μm imaging by initiating a new SCUBA-2 imaging survey in the CANDELS/Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) field, called the SCUBA-2 Ultra Deep Imaging EAO Survey (STUDIES; Wang et al. 2017). By including all of the archival data, we have constructed an extremely deep single-dish image at 450 μm ($\sigma_{450\,\mu m} = 0.56$ mJy beam^{-1}), which is the deepest image yet obtained at 450 μm. A series of papers has been published based on STUDIES, including number counts (Wang et al. 2017), stellar morphology (Chang et al. 2018), and multi-wavelength properties of the sample (Lim et al. 2020). In this work, we develop a machine-learning classifier based on 164 sources that have 450 μm flux density $\geqslant 4$ mJy, signal-to-noise ratio (S/N) $\geqslant 4$, and K-band counterparts from their radio or mid-infrared identifications. Our machine-learning classifier labels a sample of 6182 SMG candidates in the wider COSMOS field. We employ an autocorrelation technique on the SMG candidates to statistically estimate their clustering signal, which allows us to infer the dark matter halo mass and to constrain the clustering evolution of SMGs from $z = 3$ down to $z = 0.5$.

This paper is structured as follows. In Section 2, we introduce the ancillary data in the COSMOS field, as well as the observations, data reduction techniques, source extraction procedure, and training data set. In Section 3, we describe the machine-learning methodology we use for SMG candidate identification. In Section 4, we verify our machine-learning technique in selecting the SMG candidates. We present the comparison samples in Section 5 and the clustering properties of SMG candidates in Section 6. We summarize our findings in Section 7. Throughout this work, the standard errors of our sample distribution medians are estimated from bootstrap analysis. The term "SMG candidates" in this work represents the machine-learned candidates of 450 μm-selected SMGs, unless otherwise stated. We adopt cosmological parameters $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.70$, $\Omega_m = 0.30$, and $\sigma_8 = 0.83$ (Planck Collaboration et al. 2014).

2. Data

2.1. Main Data

The SCUBA-2 camera contains 5000 pixels (field of view $\simeq 45$ arcmin^2) in each of the 450 and 850 μm detector arrays. The SCUBA-2 camera operates at 450 and 850 μm simultaneously and provides an unprecedented mapping speed, meaning that it can efficiently survey large areas of sky at 450 and 850 μm. The beam size of SCUBA-2 is 7.″9 at 450 μm, which is an order of magnitude smaller in area compared to far-infrared observations from Herschel at similar wavelengths (24″–35″). The 450 μm data presented in this paper come from three sources: STUDIES (Wang et al. 2017), the data taken by Casey et al. (2013, hereafter C13), and the SCUBA-2 Cosmology Legacy Survey 450 μm campaign in the COSMOS field (S2CLS-COSMOS; Geach et al. 2013, 2017). We briefly describe these programs in turn.

STUDIES is a JCMT Large Program that aims to reach the confusion limit at 450 μm within the CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) footprint in the COSMOS field. The mapping center of STUDIES is R.A. $= 10^h00^m30\overset{s}{.}7$ and decl. $= +02°26'40''$. The CV DAISY mapping pattern was used, resulting in an 3′ diameter area of approximately uniform coverage of high sensitivity, with noise level increasing outside of this area as a function of radius. In this work, we only adopt the STUDIES data that were collected until 2019 April, amounting to 252 hr of exposures, about 84% of the total allocated integration (330 hr) for the whole project. We note that data collection for STUDIES is now 99% complete (by 2020 April), and the final version of the deeper image and source catalog will be published in the future. In addition, as part of the S2CLS project, the 450 μm S2CLS-COSMOS was observed with two CV DAISY maps offset by 2′ in decl. from the central pointing of R.A. $= 10^h00^m30\overset{s}{.}7$ and decl. $= +02°22'40''$, with some overlap resulting in an area of $\simeq 13$ arcmin^2 with noise level < 5 mJy beam^{-1}. The total on-sky integration of S2CLS-COSMOS is 150 hr. The survey of C13 used the PONG-900 scan pattern, mapping with a center of R.A. $= 10^h00^m28\overset{s}{.}0$ and decl. $= +02°24'00''$, resulting in a wider circular map that reaches a uniform depth over an area of approximately 700 arcmin^2. The total on-target time of C13 is 38 hr, and these data are much shallower compared to the STUDIES and S2CLS data sets. The majority of the aforementioned observations were conducted under the best submillimeter weather on Maunakea ("Band 1", $\tau_{225\,GHz} < 0.05$, where $\tau_{225\,GHz}$ is the sky opacity at 225 GHz).

The procedure used for data reduction are similar to those in Wang et al. (2017) and Lim et al. (2020). In brief, we apply the following steps.

1. The raw time-stream data from SCUBA-2 were flat-fielded using the flat scans that bracket each science observation, and the data were scaled to units of pW.
2. The time streams were then assumed to be a linear combination of noise and signal from the background (atmospheric water and ambient thermal emission), as well as astronomical objects. The procedure then entered
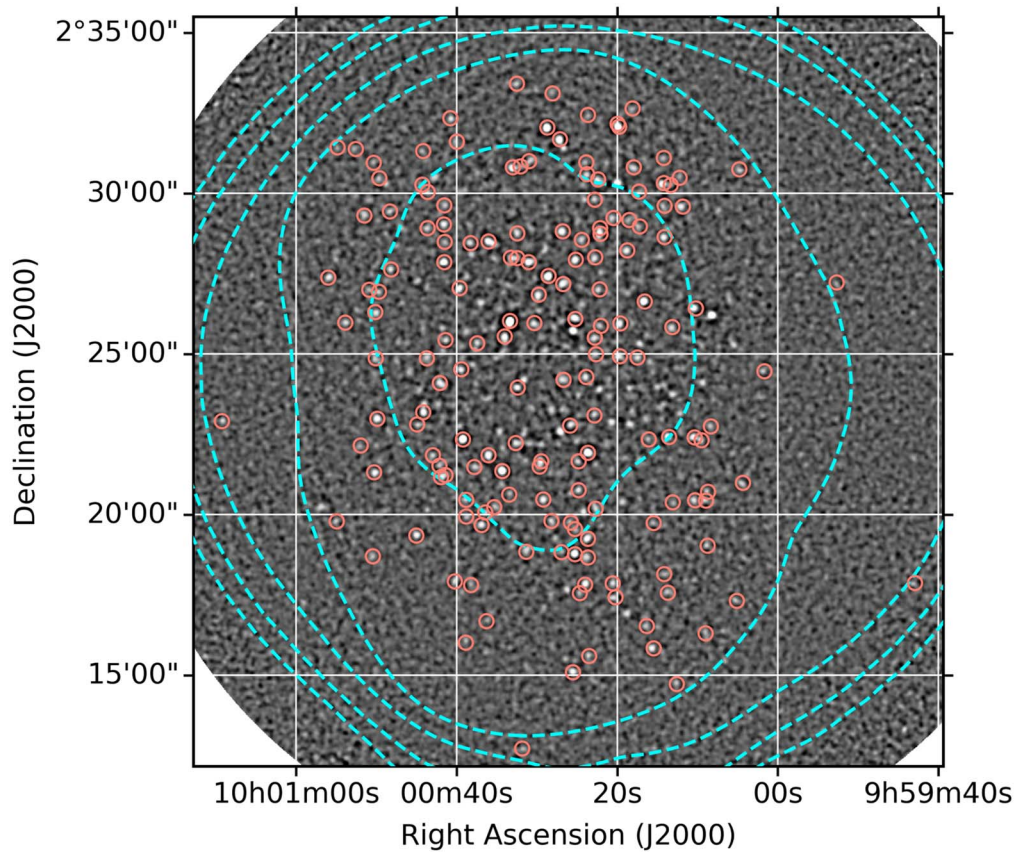
Figure 1. JCMT SCUBA-2 450 μm S/N image, showing the positions of the 164 S/N $\geqslant$ 4 and $S_{450\,\mu m} \geqslant$ 4 mJy sources with K-band counterparts based on the VLA and MIPS identifications (red circles). The cyan contours show the instrumental noise levels with contours at 1, 5, 9, 13, and 17 mJy.

an iterative stage that attempts to fit these components with a model by using the `Dynamic Iterative Map-Maker` routine of Sub-Millimeter Common User Reduction Facility (SMURF; Chapin et al. 2013).

3. We then calibrated the individual reduced scans into units of flux density by using the weighted mean flux conversion factor (FCF) of $476 \pm 95\,\mathrm{Jy\,beam^{-1}\,pW^{-1}}$. The adopted FCF was estimated from a subset of submillimeter calibrators observed under Band-1 weather during the survey campaigns and was consistent with the canonical value estimated from a wider base of calibrators (Dempsey et al. 2013).

4. Individual scans were then co-added in an optimal, noise-weighted manner, by using the `MOSAIC_JCMT_IMAGES` recipe from the PIpeline for Combining and Analyzing Reduced Data (PICARD; Jenness et al. 2008).

5. To improve the detectability of faint point sources, we convolved the map with a broad Gaussian kernel of FWHM $= 20''$ and subtracted the convolved map from the original maps to remove any large-scale structure in the sky background. We then convolved the subtracted map with a Gaussian kernel that is matched to the instrumental point-spread function (FWHM $= 7''\!.9$, Dempsey et al. 2013). We used the `SCUBA2_MATCH-ED_FILTER` recipe in the PICARD environment for this procedure.

Finally, we constructed an extremely deep 450 μm image, with the STUDIES, C13, and S2CLS-COSMOS data combined (Figure 1). The final image covers a sensitive region of $\simeq$300 arcmin^2. The instrumental noise level in the deepest regions is $\simeq$0.56 mJy beam^{-1}, which is roughly 16% deeper than those in previous works (Lim et al. 2020).

2.2. Ancillary Data

We employed radio and near-/mid-infrared identifications to construct a sample of SMG counterparts used for this work (Section 2.3). In the radio, we used the catalog from the Jansky Very Large Array (VLA) Large Project survey at 3 GHz (VLA-COSMOS; Smolčić et al. 2017). In the near-infrared, we employed the archival IRAC catalog (Sanders et al. 2007) obtained from the Spitzer Space Telescope. In the mid-infrared, we generated our own 24 μm catalog using `SExtractor` (Bertin & Arnouts 1996), since the archival MIPS 24 μm catalog (Sanders et al. 2007) only contains sources with $S_{24\,\mu m} > 150\,\mu$Jy. We run `SExtractor` in the S-COSMOS 24 μm image (Sanders et al. 2007) and recalibrated our extracted 24 μm fluxes to their Spitzer General Observer Cycle 3 total flux. Our generated catalog has a 3.5σ detection limit of 57 μJy without using positional priors from other wavelengths. We verify that our generated catalog is in good agreement with the deep Spitzer catalog ($S_{24\,\mu m} > 80\,\mu$Jy) provided by Le Floc'h et al. (2009) with a median value of $|\Delta S_{24\,\mu m}|/S_{24\,\mu m} = 6^{+10}_{-4}\%$ and with median flux peak offsets of $\Delta\alpha = 0''\!.0 \pm 0''\!.3$ and $\Delta\delta = 0''\!.3 \pm 0''\!.3$.

We utilized the multiwavelength band-merged COSMOS2015 catalog compiled by Laigle et al. (2016), which includes 30+ bands of photometric data, spanning from X-ray through the near-ultraviolet and optical to the far-infrared.

We used the combined "FLAG_HJMCC = 0 (or 2)" and "FLAG_COSMOS = 1", which is the region covered by UltraVISTA-DR2 occupying an area of $1.6 \deg^2$ in the COSMOS field. The near-infrared data (e.g., K-band) is essential for accurate photometric redshift and stellar-mass estimates, and the observed K-band magnitude correlates well with stellar mass up to $z \sim 4$ (Laigle et al. 2016). To ensure a uniform selection, we further limited ourselves to galaxies with K-band magnitude of $m_K < 24.5$ mag$_{AB}$ (limiting magnitude at 3σ in a $2''$ diameter aperture from Laigle et al. 2016). The COSMOS2015 catalog also provides stellar-mass and redshift estimations, which are used in our sample selection (Section 5) and clustering analysis (Section 6). The stellar-mass and redshift measurements in the COSMOS2015 catalog were fitted using the LE PHARE code (Arnouts et al. 1999; Ilbert et al. 2006). In this work, we do not exclude AGNs from our sample, and we refer readers to Laigle et al. (2016) for further details.

We employed a stacking technique to assess the averaged 450 and $850\,\mu$m flux density of our SMG candidates (Section 4) using the wide-field SCUBA-2 image from S2COSMOS (Simpson et al. 2019). S2COSMOS is an EAO Large Program that has mapped the entire COSMOS field to a uniform depth at $850\,\mu$m of $\sigma = 1.2$ mJy beam^{-1}. Thanks to the dual-band observing capability of SCUBA-2, $450\,\mu$m data were simultaneously obtained. While S2COSMOS is designed for the $850\,\mu$m imaging (it has been carried out under weather conditions less suitable for $450\,\mu$m observations), the large area and the depth reached $1\sigma \simeq 12$ mJy beam^{-1} at $450\,\mu$m allow us to obtain strong constraints on the stacked flux of the SMG candidates.

2.3. Training Sample

The methodologies employed here for source extraction and counterpart identification are similar to those in Lim et al. (2020). We briefly summarize the method here, referring readers to Lim et al. (2020) for further details.

We employed a source extraction method similar to the "CLEAN" deconvolution in radio interferometric imaging. We identified the most significant peak in the S/N map and subtracted 5% of a peak-scaled model point-spread function from the image at its position. The subtracted flux and coordinates were then cataloged, and the subtraction was iterated until a significance threshold floor ($=3.5\sigma$) was reached. We summed up the subtracted fluxes and the remaining threshold flux density and took this to be the final flux density for each source. In this work, we limited our $450\,\mu$m–selected sample to the sources with total integrated S/N over the "CLEAN"ed area $\geqslant 4$ due to the relatively high fraction of spurious sources ($\geqslant 20\%$) at S/N < 4. We further limited our sources to those with $S_{450\,\mu m} \geqslant 4$ mJy beam^{-1} to achieve a more uniform selection and to address the nonuniform sensitivity coverage of our map. In total, we obtained 221 such sources from a region of $\simeq 300$ arcmin^2.

To construct the K-band source catalog for our $450\,\mu$m–selected sample, we first cross-matched the positions of $450\,\mu$m sources with positions from the VLA-COSMOS 3 GHz radio catalog (Smolčić et al. 2017) using a $4''$ search radius. In total, we found 131 VLA-identified sources, and this procedure is expected to produce $\simeq 3$ false matches out of the 131. The expected false matches are derived from the number density of the matched catalog at a certain search radius and then multiplied by the total number of matched sources. We then

Table 1
Confusion Matrix for Binary Classification

	Positive prediction	Negative prediction
Positive class	True Positive (TP)	False Negative (FN)
Negative class	False Positive (FP)	True Negative (TN)

cross-matched the radio positions with coordinates from the Spitzer IRAC near-infrared catalog (Sanders et al. 2007) using a $1''$ search radius ($\simeq 4$ expected false matches out of 123). For the remaining 90 $450\,\mu$m sources without radio counterparts, we cross-matched them to the Spitzer MIPS $24\,\mu$m catalog using a search radius of $4''$ ($\simeq 4$ expected false matches out of 65). Based on the $24\,\mu$m positions, we then made a positional matching in the IRAC near-infrared catalog by using a $2''$ search radius ($\simeq 2$ expected false matches out of 52). Once we obtained the IRAC positions from the mid-infrared or radio counterparts, we associated these sources with the band-merged COSMOS2015 photometric catalog (Laigle et al. 2016) using a search radius of $1''$.

Among the 221 sources, 164 ($74\% \pm 6\%$) of the SMGs have K-band counterparts, of which, 117 sources are VLA-identified and 47 are MIPS-identified. All of them are detected significantly according to a corrected-Poissonian probability identification technique (p-values <0.05; see Downes et al. 1986). We employed these 164 sources as our SMG data set for the machine-learning algorithm (Section 3). It is worth noting that 57 sources ($26\% \pm 3\%$ of total 221 sources) do not have MIPS/VLA counterparts and so they are not in the training set. This missing population does not exhibit significant variation in $450\,\mu$m flux density compared to the sources having MIPS/VLA identifications with p-value of 0.16 in a Kolmogorov–Smirnov test (KS test). These 57 sources most likely lie at higher redshifts ($z \gtrsim 3$; see also Figure 4 in Lim et al. 2020) due to the fact that the mid-infrared and radio wave bands do not benefit from a strong negative K-correction, so they are biased against identifying the high-redshift SMGs. Therefore, we do not expect a significant impact from this missing population on our final results (Section 6), which mainly focus on $z < 3$.

To construct a non-SMG sample for the training that is undetected by SCUBA-2 but within the SCUBA-2 footprint, we select 4705 K-band-selected sources that reside within the noise level $\leqslant 1$ mJy beam^{-1} region of the $450\,\mu$m image. We adopt their K-band magnitudes and color–color pairs (i.e., flux ratios) to be the feature vectors in the machine-learning algorithm. Given the faint optical magnitudes of most of our SMG sample, we only adopt the broadband photometries from COSMOS2015 catalog in this work. In total, we have 79 features, of which, 78 features are derived by the interlacing color quantities from 13-band photometry ($uBVri^+z^{++}JHK[3.6][4.5][5.8][8.0]$).

3. Machine-learning Methodology

3.1. Performance Measures of Classification

Before introducing our machine-learning methodology, we first describe how we verify the performance of the classification. In the field of machine learning, a confusion matrix is often used to describe the performance of a classification model (Table 1). The confusion matrix has four terms: true positive (TP) refers to an actual positive sample correctly labeled as

positive; false positive (FP) is a sample incorrectly flagged as positive while in reality it is negative; false negative (FN) corresponds to real positive cases incorrectly flagged as negative; and true negative (TN) represents an actual negative sample correctly labeled as negative. From these categories, one can compute the precision, recall (in other words, the sensitivity or true positive rate, TPR), and false positive rate (FPR) as

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}},$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \text{ and}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}.$$

The precision represents the proportion of all correct positive predictions, while recall is the recovery of all real positive cases that are predicted to be positive. Conversely, the FPR is the ratio between the number of actual negative cases wrongly categorized as positive and the total negative cases.

Several meaningful indicators are often used to verify the performance of a classifier. The f1-measure (Rijsbergen 1979) considers both precision and recall to compute a score. This f1-score can be interpreted as a harmonic average of precision and recall, which can be measured as

$$\mathrm{f1} - \mathrm{score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}. \quad (1)$$

The f1-score reaches its best value at 1 and worst at 0.

Another standard test for evaluating a binary decision problem is using the Receiver Operating Characteristic (ROC) Curve that is plotting TPR against FPR (Provost et al. 1997). A perfect classifier, which has no FN and FP, will have a high value of TPR and low value of FPR. Therefore, a higher value of the area under the ROC curve (AUROCC) corresponds to a better classifier.

3.2. Methodology

In this work, we adopt the extreme gradient boosting (XGBoost; Chen & Guestrin 2016) method that is designed based on a scalable gradient tree boosting learning, since the XGBoost performs the best for identifying the SMG candidates in our sample (Section 3.5; see also other similar works done with XGBoost, An et al. 2018, 2019; Liu et al. 2019). The idea of gradient tree boosting is to build an ensemble of simpler estimators (usually they are decision trees) and convert them sequentially into a complex predictor. During the iterations of building the trees, each tree will correct the error between the predictions and the actual output from the existing trees and minimize the training error of the ensemble. The contribution of each tree can be scaled to reduce the influence of each tree that will leave more space for future trees to improve the ensemble. This process will lead to a better model and prevent the behavior of over-fitting (Friedman 2000, 2002). Typically, smaller values of this weighting (i.e., learning rate or shrinkage) seem to produce a better performance (Friedman 2000).

XGBoost is designed to push the limit of computational resources and improve the model performance for the gradient tree boosting algorithm. XGBoost performs a split finding algorithm that enumerated over all possible splits on all features and finds the best split in tree learning. The advantage of XGBoost over other techniques is that it can handle missing features. When the algorithm encounters a missing value, it is labeled as going into the default direction, which was already learned from the data.

In this work, we define 70% of our data to be the training set and the rest as the test set, where the training set builds the model, and the testing set evaluates the model's performance. In principle, the predictor performs better with a larger fraction of training data. On the other hand, the performance statistic will have greater variance if there is less testing data (Kohavi 1995). To strike a balance, we have tried changing the ratios of our training and testing sets from 50:50 to 90:10, and we verify that the XGBoost performs the best in both AUROCC and f1-score with a 70:30 split.

To avoid over-fitting, we adopt an early stopping in XGBoost after five training iterations that do not yield any improvements. To control the balance of positive and negative weights, we set the *scale_pos_weight* = 28.7, which is given by the ratio between the number of negative and positive instances (4705/164; see Section 2.3).

3.3. Feature Selection

Feature selection is an important process in machine learning that strongly influences the performance of the model. Feature selection is a procedure of selecting a subset of relevant features for model construction. In general, proper feature selection can increase the efficiency by reducing the training duration, enhancing generalization by reducing over-fitting, and improving the prediction performance (see Chen & Guestrin 2016; An et al. 2019).

A trained XGBoost model will automatically calculate feature importance and provide the feature importance scores. In this work, we first trained the XGBoost model based on the training data set and selected features by sorting feature importance scores calculated from the trained model. We then iteratively trained the model based on the selected subset of features until the point of best performance. We verified that both f1-score and AUROCC increase with the number of selected features until we use up all 79 features. Therefore, we did not reduce the number of feature vectors in this work. Considering the limitations of our computational resources, we repeated this procedure 10 times by using a different combination of training and test data sets in each iteration and estimated the average feature importance score from these 10 realizations. The top five important features in our sample are: K, ([3.6]−[4.5]), (K − [4.5]), ([3.6]−[5.8]), and (H − K), which are similar to those photometric wave bands used by Chen et al. (2016; Optical-Infrared Triple Color: (z − K), (K − [3.6]), and ([3.6]–[4.5])) and An et al. (2019; (z − K), (J − K), (K − [3.6]), and ([3.6]–[4.5])). The top five important features in our sample can be associated to fundamental physical properties. The K-band flux roughly maps to the stellar mass, while the SMGs appear to be red and occupy a relatively well-defined region in near-infrared color–color space (see also Chen et al. 2016; An et al. 2019).

3.4. Tuning the Hyper-parameters

The hyper-parameters are a set of parameters that define the machine-learning algorithm as a mathematical formula. The hyper-parameters act as tuning functions that are set during the training of the model. We optimized the hyper-parameters

Table 2
Performance of Test Samples Using Several Machine-learning Methods

Methodology	AUROCC	f1-score	Precision	Recall
Adaptive boosting	0.97 ± 0.01	0.55 ± 0.07	0.80 ± 0.07	0.43 ± 0.09
Decision tree	0.92 ± 0.02	0.36 ± 0.03	0.23 ± 0.03	0.87 ± 0.07
Logistic regression	0.97 ± 0.01	0.56 ± 0.07	0.77 ± 0.08	0.46 ± 0.10
Random forest	0.97 ± 0.01	0.54 ± 0.07	0.81 ± 0.08	0.41 ± 0.07
Stochastic gradient boosting	0.96 ± 0.01	0.61 ± 0.03	0.74 ± 0.06	0.53 ± 0.05
Support vector machines	0.91 ± 0.04	0.51 ± 0.04	0.78 ± 0.06	0.38 ± 0.05
XGBoost	0.97 ± 0.01	0.57 ± 0.07	0.51 ± 0.14	0.73 ± 0.12
XGBoost (with missing values)	**0.97 ± 0.01**	**0.63 ± 0.02**	**0.59 ± 0.04**	**0.68 ± 0.07**

Note. We replace the missing data with the detection limits of each feature vector, since most of the other algorithms do not handle missing values. The XGBoost method with missing values (marked in bold) performs the best for both AUROCC and f1-score.

in XGBoost using k-fold cross-validation (An et al. 2019). This is a resampling procedure used to iteratively evaluate the performance of machine-learning models on a limited data sample. In each iteration, this procedure randomly divides the training data set into k groups/folds (approximately equal size). Each unique fold is treated as a validation set, and the model is fitted on the remaining $k - 1$ folds. The validation set is replaced k times, and the average performance measure of the k sets is then reported. In this work, we used $k = 5$ and adopted the AUROCC as the scoring function to optimize the hyper-parameters of the XGBoost classifier. We used the `RandomizedSearchCV` recipe from the `python`-based `scikit-learn` package (Pedregosa et al. 2011). `RandomizedSearchCV` randomly searches over a combination of hyper-parameter space and finds the best solution for the constructed model. In the same computational budget, `RandomizedSearchCV` can find models that are as good or better compared to a pure grid search (Bergstra & Bengio 2012). In this work, we set the number of iterations to 200. We verify that both f1-score and AUROCC converge after 200 iterations (confirming up to 300 iterations).

3.5. Algorithms Comparison

We repeated the procedure in Section 3.4 10 times by using exactly the same combinations of training and test data sets used in Section 3.3 for each iteration. Each optimizer gives different results in labeling the SMG candidates in accordance with the expectation. We verify that the labeled SMG candidates change within 10% from each optimizer. The mean performance from these 10 optimizers is summarized in Table 2.

We also test the performance of XGBoost against other machine-learning algorithms. To make a fair comparison, we replaced the missing data with the detection limits of each feature vector since most of the other algorithms do not handle missing values. We optimized each of the algorithms (similar to what was done in Section 3.4) and repeated the procedure 10 times by using the exactly same combinations of training and test data sets used in Section 3.3 in each iteration. As shown in Table 2, the XGBoost method with missing values performs the best in terms of both AUROCC and f1-score. Furthermore, we verify that the values of AUROCC, f1-score, precision, and recall do not change significantly with redshift when we review our test data set in specific redshift bins.

3.6. Result from the XGBoost Algorithm

To accommodate the different results from each optimizer due to different combinations of training and test data sets, we adopt the following approach so that the training results are close to the mean performance from the 10 optimizers mentioned in Section 3.4. We combined the predicted class probabilities (output from `predict_proba` algorithm in `scikit-learn` package) from each optimizer by using a combined probability formula $(P_1 P_2 ... P_n)^{\frac{1}{n}}$, where n is 10 in this work. We labeled the sources that have a final class probability $\geqslant 0.5$ (the default threshold for two-classes classification in XGBoost), as SMG candidates. This procedure is similar to the bootstrap aggregating, so called bagging (Breiman 1996). Bagging is a two-step process: bootstrapping and aggregating. Bootstrapping is a sampling method that randomly selects several subsets of samples from the entire data set. The individual subset of samples is then taken as the training data set for the machine-learning models. The aggregation then combines the model predictions from those subsets into a final prediction considering all possible outcomes. In short, the bagging procedure can generate an aggregated predictor from multiple predictors, which can improve the stability and accuracy of machine-learning algorithms and reduce variance to avoid over-fitting (Breiman 1996).

To validate the performance of the aggregated predictor, we isolated 30% of our sample as the independent test sample and split the remaining 70% into 50%:20% for training. We trained the 50%+20% data set 10 times in XGBoost, in which the 50% and 20% subsamples were randomly drawn in each iteration. We used the trained classifiers to estimate the predicted class probabilities of the isolated 30% test set in each iteration and performed the bagging procedure in their probabilities. We confirmed that the performance of the aggregated predictor (precision and recall) is almost the same as the mean performance (precision and recall) from those 10 iterations, indicating that the aggregated predictor is a representative classifier for a number of classifiers. Therefore, we conclude that the performance of the aggregated classifier (after the bagging procedure) should be similar to the mean performance shown in Table 2.

Based on the adopted training results, there is a non-negligible degree of misidentification (precision $= 0.59 \pm 0.04$) and incompleteness (recall $= 0.68 \pm 0.07$) in our classifications. It is therefore worth investigating which population is labeled incorrectly as SMGs while, in reality, it

is not (i.e., the FP) and what fraction of SMGs are labeled incorrectly as field galaxies (i.e., the FN) by our trained XGBoost algorithm. To do this, we investigate the properties of FP in each of the 10 iterations. The median stellar mass and median redshift of our training SMGs are $\log(M_*/M_\odot) = 10.76^{+0.04}_{-0.02}$ and $z = 1.64^{+0.16}_{-0.07}$, respectively, and their median stacked flux is (7.1 ± 0.5) mJy at $450\,\mu$m based on our STUDIES image. We find that the overall properties of FP have a median redshift of $z = 2.0 \pm 0.3$, median stellar mass of $\log(M_*/M_\odot) = 10.7 \pm 0.1$, and median stacked flux of (2.5 ± 0.5) mJy at $450\,\mu$m, based on our STUDIES image. These findings show that the FP has a similar stellar mass to the training SMG sample but appears to have slightly higher redshifts compared to the training SMG sample and is less active in obscured star formation. Meanwhile, we carry out the same investigation for the FN. The overall properties of FN have median redshift $z = 1.4 \pm 0.2$, median stellar mass $\log(M_*/M_\odot) = 10.5 \pm 0.1$, and median stacked flux (6.0 ± 0.5) mJy at $450\,\mu$m. These findings show that our classifier may miss the population that is slightly less massive and lower in redshift compared to our training SMG sample. We conclude that the effectiveness of our classifier could be hindered by the fuzzy boundary between the SMGs and the less-active star-forming galaxies having similar stellar masses. We do not expect that this finding will impact our final clustering measurements significantly in Section 6, where we show that the clustering signals of star-forming galaxies are very similar to the SMG candidates once they are matched in redshift and stellar mass.

As an additional test, shown in the Appendix, we have applied two distinct machine-learning algorithms with different levels of recall and precision and show that the clustering signals recovered for the SMGs are robust. The SMG candidates that are identified by decision tree (better recall) and random forest (better precision) yield clustering signals that are not significantly different from the results of XGBoost-identified SMG candidates, indicating that our final results are insensitive to the chosen classifiers (see the Appendix).

The spurious sources in the SMG training sample, which are wrongly labeled as SMGs, may impact the training sample and bias the performance estimators. We check that the cumulative spurious fraction of our SMG training sample is $\simeq 10\%$. To quantify how sensitive our training is to this contamination, we randomly choose 10% of the SMG training sample and swap them with the MIPS- or VLA-detected non-SMGs in the STUDIES field. We then train the XGBoost by using this training sample and repeat the procedure 10 times. The mean AUROCC and f1-scores from these procedures are reduced by 3% and 10%, respectively, compared to the measurements shown in Table 2. These offsets are expected, since the contamination in the training sample will reduce the precision of the predictor. Considering the uncertainties in the measured AUROCC and f1-score, we conclude that our final clustering measurements (Section 6) will not be impacted significantly by the small fraction of spurious sources.

In total, our trained XGBoost algorithm labels 6182 SMG candidates from the COSMOS2015 catalog containing 307,374 sources that have an $m_K < 24.5$ mag$_{AB}$ across an effective area of $1.6\,deg^2$.

4. Verifications of the SMG Candidates

While machine learning is a powerful and promising technique for data analysis, it inevitably appears to be a black box to typical users. Checks need to be performed to ensure that the selected SMG candidates have properties (especially those not used for the training) similar to those of the parent SMG sample.

The total number of SMG candidates (6182 sources) is consistent with the predicted number counts of 6100^{+1800}_{-1400} of $450\,\mu$m sources in the field. The predicted number count is derived by integrating the best-fitted function of Wang et al. (2017) over a flux range of 3.4–36 mJy. The predicted median flux in this flux density range is (6.8 ± 0.1) mJy. The lower end of the range is determined from the boosting corrected $450\,\mu$m flux density range of our training SMG data set. We further apply the fraction of 0.74 ± 0.06 to take into account the selection bias in our training data set (see Section 2.3), since our training SMG data set only includes SMGs having K-band counterparts.

To probe the submillimeter emissions of our 6182 SMG candidates, we directly measure their $450\,\mu$m flux density from the S2COSMOS image. The $450\,\mu$m flux distribution of SMG candidates is consistent with the measurement from the SMG training sample, based on the same image (p-value = 0.36 in KS test). The median stacked $450\,\mu$m flux density from the S2COSMOS image of the SMG candidates is (5.3 ± 0.2) mJy (Figure 3). This value is tentatively offset ($\simeq 2.3\sigma$) from the median stacked flux of the SMG training sample (7.8 ± 1.1) mJy derived from the same S2COSMOS image but significantly offset from the aforementioned predicted median flux of (6.8 ± 0.1) mJy and the median stacked flux of the SMG training sample, based on our deeper STUDIES image (7.1 ± 0.5) mJy. The stacked value of the SMG training sample could be biased high, since sources close to the detection threshold will be included if they are on a peak of the noise. Indeed, the deboosted $450\,\mu$m stacked flux of the SMG training sample is 6.4 ± 0.5 mJy, only marginally higher than the stacked flux of the SMG candidates. To have an independent test, we apply the stacking analyses in the $850\,\mu$m images from the S2COSMOS and STUDIES surveys by assuming that the 450 and $850\,\mu$m noise maps are reasonably independent. The median stacked $850\,\mu$m flux density from the S2COSMOS image of our SMG candidates is (1.23 ± 0.03) mJy. This value is significantly offset from the median stacked $850\,\mu$m flux of the SMG training sample (1.9 ± 0.2) mJy based on the same image, but tentatively offset ($\simeq 1.9\sigma$) from similar measurements based on our deeper STUDIES image (1.7 ± 0.2) mJy. The fainter stacked flux of the SMG candidates compared to the stacked flux of SMG training sample is expected, since our machine-learning algorithm misidentifies a fraction of SMGs ($\simeq 30\%$; see the recall value in Table 2). However, the stacked flux only provides the overall emission properties of our sample. We do not expect that our results in the following analyses will be affected significantly due to contamination, since the contamination arises from less dusty star-forming galaxies with similar stellar masses to the training SMG sample (see Section 3.6), and the clustering signals of star-forming galaxies are very similar to the SMG candidates once they are matched in redshift and stellar mass (Section 6).
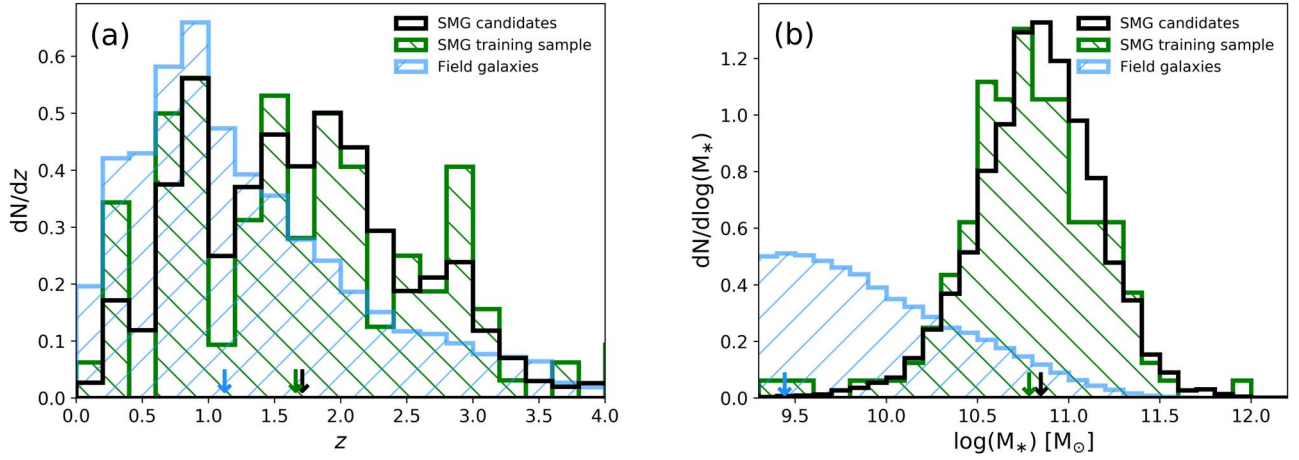
Figure 2. Panel (a): normalized histograms of photometric redshift; and panel (b): normalized histograms of stellar mass for the SMG training sample, SMG candidates, and field galaxies. The median values are marked as downward arrows for the corresponding sample. The median redshifts and stellar masses of SMG candidates are consistent with the estimations from the parent SMG training data set.
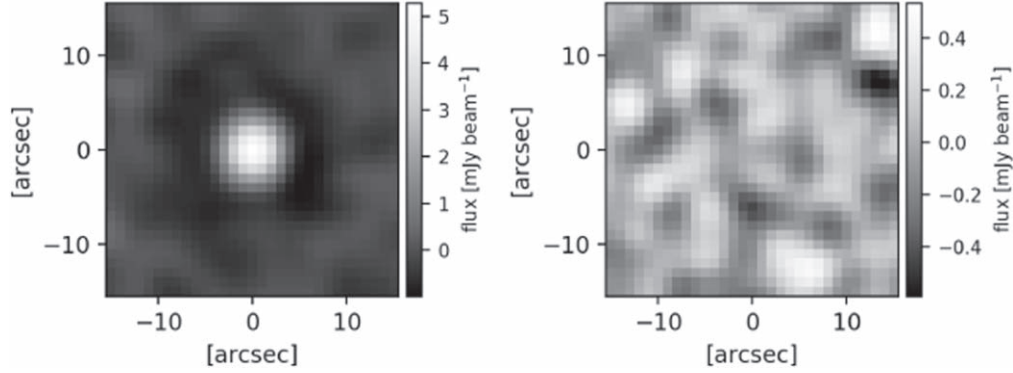

Figure 3. Left panel: stacked SCUBA-2 450 μm image from the S2COSMOS survey (Simpson et al. 2019) at the positions of 6182 SMG candidates. The negative flux trough surrounding the source is caused by the matched-filter procedure from the `SCUBA2_MATCHED_FILTER` recipe (see Section 2). The stacked 450 μm flux density is (5.3 ± 0.2) mJy. Right panel: stacked 450 μm image at 6182 random positions.

All of the SMG training sample in the machine-learning algorithm has matched detections from the MIPS at 24 μm or VLA at 3 GHz, or both (see Section 2.3). The fraction of our SMG candidates that have MIPS and/or VLA detections is $\simeq 88\% \pm 1\%$ (=5442/6182), indicating that our SMG candidates behave similarly at mid-infrared and/or radio wavelengths compared with the SMG training sample. In addition, the MIPS- and VLA-flux distributions of our SMG candidates are consistent with the SMG training data set, having p-values of 0.77 and 0.31 in the KS test, respectively. Based on the S2COSMOS image, the stacked flux of our SMG candidates with MIPS or VLA detections is (5.6 ± 0.2) mJy, while the stacked flux of the remaining SMG candidates is (3.0 ± 0.5) mJy. This indicates that $\simeq 12\%$ (=740/6182) of our SMG candidates are misidentified and/or are biased toward fainter SMG population. Again, we do not expect that our results in the following analyses will be affected significantly due to this finding. It is worth noting that the MIPS 24 μm and VLA 3 GHz photometry are not part of the training features (Section 3.3). The high MIPS- or VLA-detection rates strongly support the reliability of our machine-learning algorithm.

By comparing other populations extracted from the COSMOS2015 catalog, we can test whether the distribution of our SMG candidates is consistent with the SMG training sample. Figure 2 shows the normalized histograms of photometric

redshift and stellar mass for the 450 μm-selected SMG training sample, 450 μm-selected SMG candidates, and field galaxies. The median redshift of our SMG candidates is $z = 1.69 \pm 0.02$ (Figure 2(a)), which is in excellent agreement with that of the SMG training sample ($z = 1.64^{+0.16}_{-0.07}$). The redshift distribution of our SMG candidates is consistent with the SMG training data set, having a p-value = 0.63 in the KS test, indicating that we cannot reject the null hypothesis of no difference between our SMG candidates and parent SMG training sample. On the other hand, the p-value is essentially zero in the KS test between the redshift distribution of our SMG candidates and that of the field galaxies in the COSMOS2015 catalog.

Figure 2(b) shows the stellar-mass distributions of the SMG candidates, SMG training sample, and the field galaxies over a redshift range of $z = 0$–6. The median stellar mass of our SMG candidates ($\log(M_*/M_\odot) = 10.83 \pm 0.01$) is consistent with the median stellar mass of the SMG training sample ($\log(M_*/M_\odot) = 10.76^{+0.04}_{-0.02}$). A KS test performed on the stellar-mass distributions of SMG candidates and parent SMG training data set shows that the test cannot distinguish between these two populations at the 95% significance level (p-value = 0.15). The KS test rejects the null hypothesis that the stellar-mass distributions of our SMG candidates and field galaxies are drawn from the same distribution (p-value $\simeq 0$). The similarities of physical properties (e.g., redshift and stellar mass) between our SMG candidates and the parent
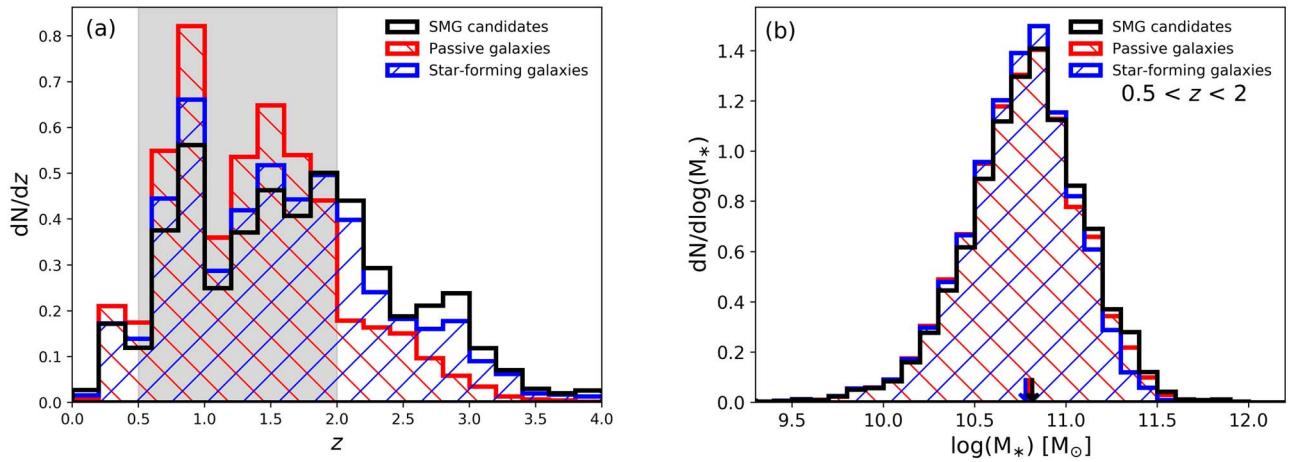
Figure 4. Panel (a): normalized redshift distributions for SMG candidates, and comparison samples of star-forming galaxies and passive galaxies. All of these populations are matched as closely as possible in redshift, stellar mass, and sample size. In this work, we restrict the comparison samples to the galaxies at $0.5 < z < 2$ (gray shaded region), since it is difficult to find sufficient numbers of passive and massive star-forming galaxies at $z > 2$. Panel (b): normalized stellar-mass distributions for SMG candidates and comparison samples of star-forming galaxies and passive galaxies at $0.5 < z < 2$.

SMG training data set are expected. This is because the features for estimating the physical properties in the COSMOS2015 catalog, which are based on the photometric data, are similar to the adopted features in our machine-learning algorithm.

We also test our SMG-identification technique on ALMA observations. ALMA follow-up observations resolved 260 $850\,\mu$m–selected SMGs in the S2CLS-COSMOS program (AS2COSMOS; J.M. Simpson et al. 2020, in preparation) from 183 $850\,\mu$m submillimeter sources with $S/N > 4.3\sigma$ (rms $\simeq 0.2$ mJy). There are 165 ALMA-detected sources with K-band counterparts in the COSMOS2015 that further satisfy our selection criterion of $m_K < 24.5$ mag$_{AB}$ (see Section 2.2). Among these, our machine-learning algorithm successfully identifies 126 ALMA-detected sources as SMG candidates, indicating that the completeness of our identification is $76\% \pm 7\%$. On the other hand, 148 SMG candidates are located within the 183 ALMA primary-beam areas (FWHM = 17″.3). Among them, 121 sources are detected by ALMA, suggesting that the precision of our machine-learning classifier is $82\% \pm 7\%$. However, the ALMA observations should not necessarily have high identification rates in our sample, since ALMA and JCMT were observing in different wavelengths, and therefore, the sensitivity limits are different. To coarsely estimate the $450\,\mu$m sensitivity in the AS2COSMOS survey, we employ the typical $S_{450\,\mu m}/S_{850\,\mu m}$ ratio of 2.5–4.5 at the faint end (Hsu et al. 2016). The sensitivity of AS2COSMOS ($\simeq 0.86$ mJy) is equivalent to a $450\,\mu$m sensitivity of $\simeq 2$–4 mJy, which is roughly our selection limit. Therefore, we conclude that our SMG candidates should be detected with AS2COSMOS, and the high identification rate of AS2COSMOS in our sample is expected, which again supports the reliability of our machine-learning algorithm.

5. Comparison Samples

To put the SMG candidates into the context of general galaxy populations at the same redshifts, we construct comparison samples of star-forming galaxies and passive galaxies.

5.1. Rest-frame NUV–r–J Color

Our comparison star-forming and passive galaxies are selected based on the rest-frame $M_{NUV} - M_r$ and $M_r - M_J$ color cuts (Ilbert et al. 2013) for galaxies not flagged as SMG candidates by our machine-learning algorithm. To avoid incompleteness, in the following analyses, we only consider sources that have stellar-mass estimates above the 95% mass completeness. To empirically estimate the stellar-mass completeness as a function of redshift, we follow a procedure similar to that in Pozzetti et al. (2010). We take the faintest 20% of galaxies in K-band magnitude in several redshift bins to be a representative observational limit for our whole sample. We then find the 95th upper percentile from the stellar-mass distribution of this subsample in each of the redshift bins and take the values to be the stellar-mass limit for the corresponding redshift bins. The 95% mass completeness as a function of redshift can be described with a polynomial function $\log(M_{lim}/M_\odot) = 8.14 + 0.95z - 0.09z^2$.

Since it is known that galaxy clustering evolves with redshift and is a strong function of stellar mass (McCracken et al. 2015), to make a fair comparison, we need to construct a sample of comparison galaxies that are matched as closely as possible to our SMG candidates in redshift, stellar mass, and sample size. We first adopt the `binned_statistic_2d` algorithm from `scipy` package (Jones et al. 2001) to generate the two-dimensional histograms of our SMG candidates and comparison galaxies by using specific redshift and stellar-mass bins. We then randomly select a number of comparison galaxies in each bin, which is matched with the SMG candidates. The normalized redshift distributions for our SMG candidates and comparison samples are shown in Figure 4(a). As we can see, the number of passive galaxies drops at $z > 2$. Similarly, it is also hard to find sufficient numbers of massive star-forming galaxies that are not SMG candidates at $z > 2$. Therefore, in this work, we restrict the comparison samples to the galaxies at $0.5 < z < 2$. In total, we randomly select 3021 and 3083 passive galaxies and star-forming galaxies, respectively, at $0.5 < z < 2$ (summarized in Table 3). The redshift estimations for SMG candidates may be less reliable compared to the
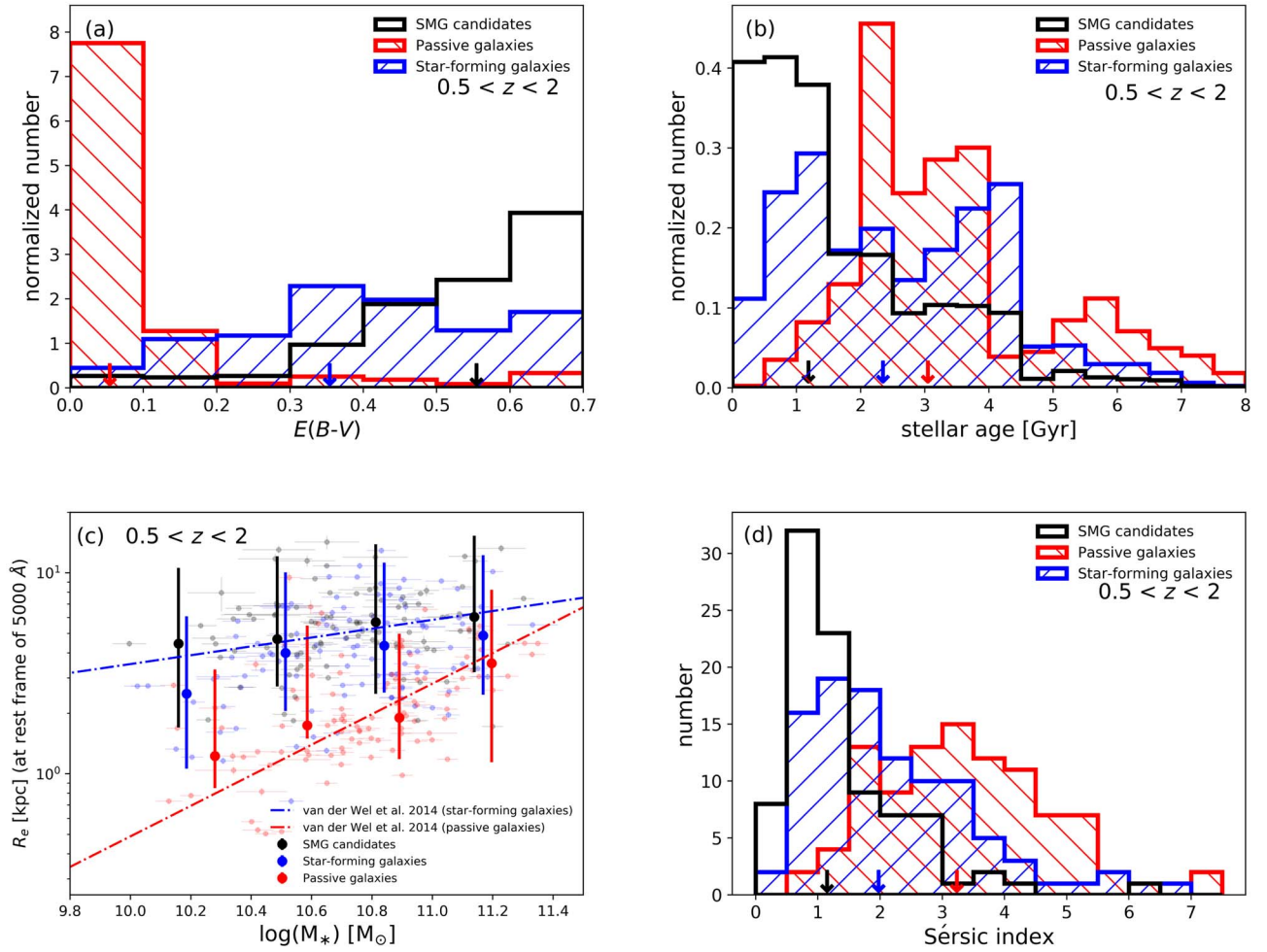
Figure 5. Physical properties of SMG candidates and comparison galaxy samples. Panel (a): normalized histograms of extinction; panel (b): normalized histograms of stellar population age; panel (c): effective radius and stellar-mass relations; and panel (d): histograms of Sérsic index. In panels (a), (b), and (d), the median values are marked as downward arrows for the corresponding sample. In panel (c), the larger points show the running median of our sample and the $\pm1\sigma$ scatter, while the smaller circles show the effective radius and stellar-mass relation of individual sources for the corresponding sample. In panel (c), we also show the best-fit size–mass relations of star-forming and passive galaxies at $z = 1.25$ from van der Wel et al. (2014), which correspond to the median redshifts of our sample. In summary, we find that the SMG candidates are dustier, younger, larger, and more disk-like than the comparison samples that are matched in redshift and stellar mass.

Table 3
Results of Clustering Analyses

Sample	Redshift	$N_{\rm s}$[a]	b	r_0 $(h^{-1}\,{\rm Mpc})$	$\log(M_{\rm halo})$ $(h^{-1}\,M_\odot)$
SMG candidates	$0.5 < z < 1.0$	1112	$2.4^{+0.3}_{-0.3}$	$8.1^{+1.0}_{-1.1}$	$13.4^{+0.2}_{-0.2}$
	$1.0 < z < 2.0$	2205	$3.9^{+0.4}_{-0.4}$	$9.8^{+1.1}_{-1.1}$	$13.4^{+0.1}_{-0.2}$
	$2.0 < z < 3.0$	1518	$5.9^{+0.7}_{-0.9}$	$10.9^{+1.5}_{-1.7}$	$13.2^{+0.2}_{-0.2}$
Passive galaxies	$0.5 < z < 1.0$	1112	$3.3^{+0.2}_{-0.2}$	$11.5^{+0.8}_{-0.8}$	$13.9^{+0.1}_{-0.1}$
	$1.0 < z < 2.0$	1909	$4.3^{+0.4}_{-0.5}$	$10.8^{+1.2}_{-1.3}$	$13.5^{+0.1}_{-0.2}$
Star-forming galaxies	$0.5 < z < 1.0$	1090	$2.5^{+0.3}_{-0.3}$	$8.4^{+1.0}_{-1.0}$	$13.5^{+1.0}_{-0.2}$
	$1.0 < z < 2.0$	1993	$4.1^{+0.4}_{-0.5}$	$10.4^{+1.2}_{-1.3}$	$13.5^{+0.1}_{-0.2}$

Note.
[a] Sample sizes of our samples in the corresponding redshift bins.

comparison passive and star-forming galaxies, since SMG candidates are expected to be dusty and consequently fainter in the optical. Interestingly, at $0.5 < z < 2$, the median redshift errors for SMG candidates ($\sigma_z = 0.02 \pm 0.06$) is statistically

consistent with the comparison passive ($\sigma_z = 0.02 \pm 0.06$) and star-forming galaxies ($\sigma_z = 0.03 \pm 0.06$). Therefore, we conclude that the redshift estimations for all of our samples are equally reliable. The stellar-mass distribution for sources at

$0.5 < z < 2$ is plotted in Figure 4(b). The samples all show similar stellar-mass distributions according to the KS test (p-value >0.05).

5.2. Physical Properties of SMG Candidates and Comparison Samples

In this section, we compare some physical properties between our SMG candidates and comparison samples. The stellar mass, dust extinction, and age of the stellar population assembled in the model can be recovered from SED template fitting. In this work, we adopted the COSMOS2015 catalog values, estimated with the LE PHARE code. In short, they used a library of synthetic spectra generated using stellar population synthesis from Bruzual & Charlot (2003) and assuming a Chabrier (2003) initial mass function. Both declining and delayed star formation histories were used. The SEDs were generated for a grid of 42 ages in the range 0.05–13 Gyr, and the attenuation curve of Calzetti et al. (2000) was applied to the templates with color excess in range of $E(B - V) = 0$–0.7.

Our results show that the SMG candidates, which are expected to be dusty, tend to have higher values of $E(B - V)$, compared to star-forming and passive comparison samples (Figure 5(a)). A significant fraction of our SMG candidates have $E(B - V) = 0.7$, which is a cap on the maximum $E(B - V)$ value introduced in the LE PHARE SED fitting procedure. This implies that the extinction corrections, which are applied to the sources in the COSMOS2015 catalog, may be insufficient, and consequently, their dust-corrected SFR estimations may be underestimated, particularly for the dusty population (see also Casey et al. 2013; Simpson et al. 2014; Elbaz et al. 2018; Dudzevičiūtė et al. 2020).

Figure 5(b) shows histograms of model stellar population age for our SMG candidates and comparison samples. It is worth noting that the age measurements derived from SED fitting alone should perhaps be treated with skepticism (see Dudzevičiūtė et al. 2020), since the galaxy colors become redder with increasing extinction or age (the age-dust degeneracy; e.g., Calzetti 2001; Pforr et al. 2012). This degeneracy might be more severe in the starburst systems such as SMGs (Hainline et al. 2011; Michałowski et al. 2012a; Simpson et al. 2014). On the other hand, the age measurements are highly dependent on the assumed star formation history, in which the bursty star formation models tend to produce the youngest ages, while the continuous star formation models tend to produce the oldest ages (Maraston et al. 2010; Hainline et al. 2011). Nevertheless, the distribution of stellar population age in SMG candidates peaks at younger ages (Figure 5(b)), even though the overall distribution of stellar masses is similar to that in comparing star-forming and passive galaxies. This suggests that the SMG candidates may be galaxies with more recent star formation (i.e., a higher proportion of young stars), since the amount of attenuation in the rest-frame ultraviolet or optical is so large that the SED fitting code will have to de-redden the optical SED to the youngest available stellar populations.

To investigate the spatial structure of our SMG candidates and comparison samples, we adopt the morphological properties of Hubble Space Telescope (HST) H_{F160W}-selected galaxies in the CANDELS/COSMOS field from van der Wel et al. (2012). We only use the sources with "flag = 0" in this catalog, which consists of the objects with a good Sérsic model fits (Sérsic 1963, 1968) measured by the GALFIT code (Peng et al. 2010). According to van der Wel et al. (2012), the

structural parameters can be measured with a precision and accuracy better than 10% down to $H_{F160W} \simeq 24.5$ mag$_{AB}$. It is worth noting that our work here is similar to that in Chang et al. (2018). Chang et al. (2018) used a sample of SCUBA-2 SMGs with $S_{450\,\mu m} > 2$ mJy, while we adopt machine-learned SMG candidates with $S_{450\,\mu m} > 4$ mJy (see Section 4). Our sample size is about 50% larger than that in Chang et al. (2018), since we can use the entire catalog from van der Wel et al. (2012) for our SMG candidates. In total, we have 91 SMG candidates, 102 star-forming galaxies, and 99 passive galaxies with a robust Sérsic model fit.

Figure 5(c) shows the effective radius (R_e) at a rest-frame wavelength of $\simeq 5000$ Å and stellar-mass relations for our SMG candidates and the comparison samples. To determine the R_e value at a rest-frame wavelength of $\simeq 5000$ Å, we follow Equations (1) and (2) in van der Wel et al. (2014), which consider the wavelength dependence of R_e as a function of redshift and stellar mass. In Figure 5(c), we also show the best-fit size–mass relations of star-forming and passive galaxies at $z = 1.25$ from van der Wel et al. (2014), which correspond to the median redshifts of these subsamples. The comparison passive galaxies are, on average, smaller than star-forming galaxies, which is consistent with earlier studies on both local and high-redshift samples (Shen et al. 2003; Ichikawa et al. 2012; Newman et al. 2012; Fernández Lorenzo et al. 2013; van der Wel et al. 2014). The median R_e of our SMG candidates ($5.5^{+0.3}_{-0.4}$ kpc) is consistent with previous studies of ALMA follow-up at LABOCA 870 μm–selected SMGs at $z \simeq 2$ ($R_e = 4.4^{+1.1}_{-0.5}$ kpc; Chen et al. 2015), SCUBA-2 SMGs at $z = 0.5$–1.5 ($R_e = 4.9 \pm 0.3$ kpc; Chang et al. 2018), and ALMA follow-up at SCUBA-2 SMGs ($R_e = 4.8 \pm 0.3$ kpc; Lang et al. 2019), within the errors, but somewhat higher than the measurement of radio-identified SMGs at $z \simeq 2$ ($R_e = 2.8 \pm 0.4$ kpc; Swinbank et al. 2010). Chen et al. (2015) attributed this to the fact that Swinbank et al. (2010) used shallower HST-NICMOS images, which tend to show smaller sizes. Our result shows that the SMG candidates are significantly ($\simeq 3\sigma$) more extended than the comparison star-forming galaxies ($R_e = 4.0^{+0.3}_{-0.2}$ kpc). Even though we split our samples by stellar mass, we still find the trend that the SMG candidates are slightly larger than comparison star-forming galaxies in all mass bins (Figure 5(c)). The two-dimensional KS test (python-package ks2d2s; original estimation references from Peacock 1983) in the size–mass plane shows that the SMG candidates are significantly different from comparison star-forming galaxies (p-value = 0.006). We therefore confirm and strengthen the tentative findings in Chang et al. (2018), in which they showed that the 450 and 850 μm–selected SMGs are marginally different from a stellar-mass- and SFR-matched star-forming sample.

The Sérsic indices (n_s) for our SMG candidates and comparison samples, which are from van der Wel et al. (2012), are shown as histograms in Figure 5(d). The median n_s of our SMG candidates (1.1 ± 0.1) is consistent with previous studies, including $n_s = 1.4 \pm 0.8$ (Swinbank et al. 2010), $n_s = 1.2 \pm 0.3$ (Chen et al. 2015), $n_s = 1.1 \pm 0.1$ (Chang et al. 2018), and $n_s = 1.0 \pm 0.2$ (Lang et al. 2019). Our SMG candidates are statistically distinguishable, with a lower median of n_s, compared to comparison star-forming ($n_s = 1.9 \pm 0.1$) and passive galaxies ($n_s = 3.2^{+0.2}_{-0.1}$). However, our result does not necessarily show that the SMG candidates are dominated by disk-like structures ($n_s \simeq 1$), since most of the SMGs with
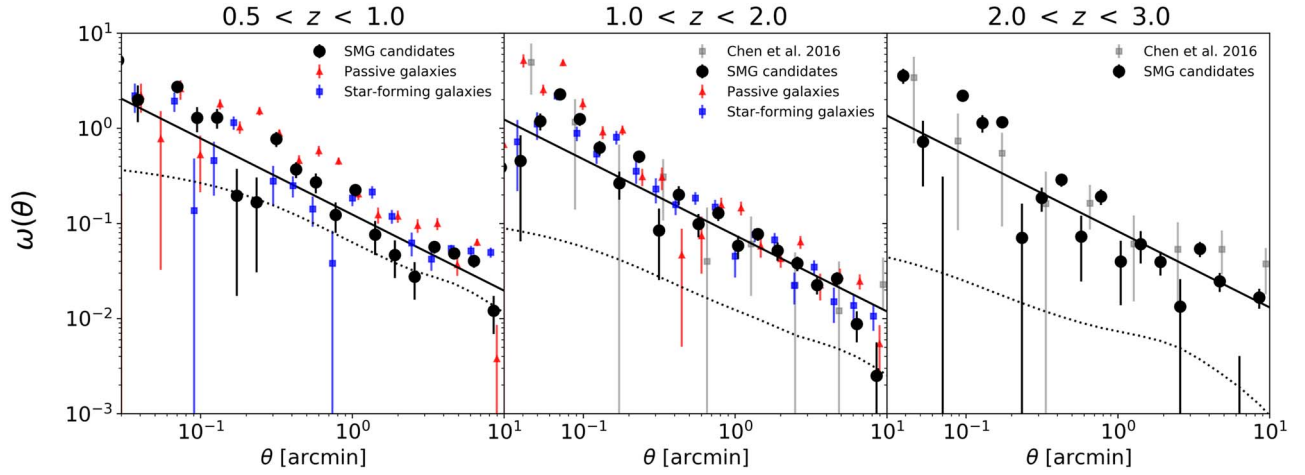
Figure 6. Two-point autocorrelation function of our SMG candidates and the comparison samples at $0.5 < z < 3.0$. The data points are offset horizontally, to avoid confusion. The dotted curves show the autocorrelation functions of the dark matter in the corresponding redshift bins.

low n_s ($\simeq 1$) are in fact visually classified as either irregulars or interacting systems (Chen et al. 2015).

In summary, we find that the SMG candidates are dustier, younger, larger, and more disk-like than the comparison samples that are matched in redshift and stellar mass. With these differences in mind, in the next section, we discuss their clustering properties.

6. Clustering and Halo Mass

We now investigate the standard two-point clustering statistics that can quantitatively measure the large-scale structure of the universe and trace the amplitude of galaxy clustering as a function of scale. The clustering measurements can further allow us to infer the dark matter halo mass and to constrain the evolution of clustering.

6.1. Two-point Autocorrelation Function

We measure the two-point galaxy autocorrelation function by using the Landy & Szalay (1993) estimator,

$$\omega(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (2)$$

where $DD(\theta)$, $DR(\theta)$, and $RR(\theta)$ are the normalized number of galaxy–galaxy, galaxy–random, and random–random pairs, respectively, counted within a given angular separation bin of $\theta \pm \delta\theta/2$. The $DR(\theta)$ and $RR(\theta)$ are normalized to the same total number of pairs as $DD(\theta)$, with $DR(\theta) = \frac{(N_D - 1)}{2N_R} N_{DR}(\theta)$, and $RR(\theta) = \frac{N_D(N_D - 1)}{N_R(N_R - 1)} N_{RR}(\theta)$, where N_D and N_R are the number of sources in the galaxy and random samples, respectively, while $N_{DR}(\theta)$ and $N_{RR}(\theta)$ are the original counts of galaxy–random and random–random pairs, respectively. We adopted 10 times as many random points as the number of our galaxies.

The projected angular two-point autocorrelation function, $\omega(\theta)$, can generally be described as a power law,

$$\omega(\theta)_{mod} = A\theta^{-\delta}, \quad (3)$$

where A is the clustering amplitude, and δ is the slope of the correlation function. The value $\delta = 0.8$ has been found to be appropriate for both observations and theories at the physical

separation of $\simeq 0.1$–$10\,h^{-1}$ Mpc (e.g., Peebles 1980; Davis & Peebles 1983; Zehavi et al. 2005; Coil et al. 2007, 2008). This statement still holds for the galaxy samples at redshift up to $\simeq 5$ ($\delta = 0.8$–1.1; Ouchi et al. 2005; Coil et al. 2006; Kashikawa et al. 2006; Lee et al. 2006; Hildebrandt et al. 2009; Durkalec et al. 2015; Harikane et al. 2016; Jose et al. 2017).

Measurement of $\omega(\theta)$ could be biased, since our sample is located in a single, area-limited region that may not be representative of the true underlying mean density over the whole sky. The observed $\omega(\theta)$ is usually biased lower than the true value $\omega(\theta)_{mod}$,

$$\omega(\theta) = \omega(\theta)_{mod} - IC, \quad (4)$$

by an additive factor of IC known as the integral constraint. In practice, IC can be numerically estimated (e.g., Infante 1994; Roche & Eales 1999; Adelberger et al. 2005) over the survey geometry using the random–random pairs of the form

$$IC = \frac{\sum_i N_{RR}(\theta_i)\omega(\theta_i)_{true}}{\sum_i N_{RR}(\theta_i)}. \quad (5)$$

We repeated the estimate of Equation (2) 25 times. We calculated the variance, mean $\omega(\theta)$, and the mean N_{rr} using these 25 estimates in Equation (5). We then employed a χ^2 minimization in Equation (4) to derive the best-fit $\omega(\theta)_{true}$ over scales of $0'.1$–$6'$ ($\simeq 0.1$–$6\,h^{-1}$ Mpc at $z = 2$ for co-moving distance). The IC-corrected $\omega(\theta)$ is shown as black symbols in Figure 6.

The errors in the clustering amplitude of $\omega(\theta)_{true}$ are expected to be underestimated, since the variance only accounts for the shot noise from the sample of the random points and the Poisson uncertainties of the DD counts, and it does not include the field-to-field variance (often called cosmic variance) of a field-limited survey. To quantify the uncertainty more realistically, we employed the "delete one jackknife" resampling method (see also Scranton et al. 2002; Zehavi et al. 2002, 2005; Norberg et al. 2009) to estimate the covariance matrix for the autocorrelation function measurements. The principle of the jackknife method is to first divide the data set into N independent subsamples, and then, one copy of the subsamples is omitted systematically at a time, similar to the k-fold cross-validation used in machine learning. Therefore, the resampling data set consists of $N_{sub} - 1$ remaining subarea, with area

$(N_{\rm sub} - 1)/N_{\rm sub}$ times the full area of the original data set. The covariance matrix of N independent realizations can be obtained as

$$C_{ij} = \frac{N_{\rm sub} - 1}{N_{\rm sub}} \sum_{k=1}^{N_{\rm sub}} (\omega(\theta_i)^k - \overline{\omega(\theta_i)})(\omega(\theta_j)^k - \overline{\omega(\theta_j)}),$$

where $\omega(\theta_{i,j})^k$ is the autocorrelation function measured with the kth area removed, and $\overline{\omega(\theta_i)}$ is the average autocorrelation function of the jackknife realizations. In practice, we divide our sample into $N_{\rm sub} = 13$ nearly equally sized stripe-shaped subareas. We verify that the shapes of the subarea do not affect the jackknife results significantly. To determine the best-fit power-law model (Equation (3)) for each correlation function, we perform a χ^2 minimization, where

$$\chi^2 = \sum_i \sum_j (\omega(\theta_i)^k - \omega(\theta_i)_{\rm mod}) C_{ij}^{-1} (\omega(\theta_j)^k - \omega(\theta_j)_{\rm mod}).$$

The 1σ error is estimated based on finding where $\Delta\chi^2 = 1$ (Avni 1976).

6.2. Dark Matter Halo Mass

To quantify the underlying dark matter halo mass ($M_{\rm halo}$) of our sample, we first need to compute the galaxy bias (b), which can be defined as the square root of the ratio of the two-point correlation function of the galaxies relative to the dark matter,

$$b = \left(\frac{\omega(\theta)}{\omega(\theta)_{\rm DM}}\right)^{\frac{1}{2}},$$

where the $\omega(\theta)_{\rm DM}$ is the projected angular two-point correlation function of the dark matter. To reproduce the clustering model in dark matter, we use the HALOFIT code from Smith et al. (2003) with improved fitting formulae provided by Takahashi et al. (2012), which can predict the nonlinear and dimensionless power spectrum of dark matter $\Delta^2(k) = k^3 P(k)/(2\pi)^2$ for a wide range of cold dark matter cosmologies. The Fourier transform of the two-point correlation function is the power spectrum. We then project the power spectrum into the angular correlation function by using Limber's equation (Limber 1953; Peebles 1980; Peacock 1991; Baugh & Efstathiou 1993), specifically via the Equation (A6) in Myers et al. (2007). The $\omega(\theta)_{\rm DM}$ profiles at $z = 0.5$–1, $z = 1$–2, and $z = 2$–3 are shown as dotted curves in Figure 6. We fit a single b parameter from the observed galaxy correlation function and the dark matter correlation function by minimizing χ^2 on the scales of $0.'1$–$6'$. Finally, we convert the b to $M_{\rm halo}$ using the ellipsoidal collapse model of Sheth et al. (2001).

In the case of the small-angle approximation ($\theta \ll 1$ rad) and assuming no clustering evolution over the redshift bin, we can de-project the angular autocorrelation function to the power spectrum by inverting Limber's equation (e.g., Myers et al. 2006; Hickox et al. 2012; see Peebles 1980 for full detivation) and further estimate the clustering scale length (r_0) as follows:

$$A = H_\gamma \frac{\int_0^\infty (dN/dz)^2 E_z \chi^{1-\gamma} dz}{[(dN/dz)dz]^2} r_0^\gamma, \tag{6}$$

where $H_\gamma = \Gamma(1/2)\Gamma([\gamma - 1]/2)/\Gamma(\gamma/2)$, Γ is the gamma function, $\gamma = \delta + 1$ (=1.8 in this work), χ is the radial

co-moving distance, dN/dz is the redshift selection function, and $E_z = H_z/H_0 = dz/d\chi$ ($H_z^2 = H_0^2[\Omega_{\rm m}(1 + z)^3 + \Omega_\Lambda]$).

According to the formalism of Peebles (1980), r_0 is related to b via

$$r_0 = 8\left(\frac{\Delta_8^2}{C_\gamma}\right)^{1/\gamma} = 8\left(\frac{b^2\sigma_8^2 D^2}{C_\gamma}\right)^{1/\gamma}, \tag{7}$$

where $C_\gamma = 72/(3 - \gamma)(4 - \gamma)(6 - \gamma)2^\gamma$, σ_8 is the amplitude of matter clustering (= 0.83; Planck Collaboration et al. 2014), and Δ_8 is the clustering strength of dark matter haloes more massive than stellar mass M at redshift z, which can be defined as $\Delta_8 = b(M, z)\sigma_8 D(z)$. The function $D(z)$ is the growth factor of linear fluctuations in the dark matter distribution, which can be computed from

$$D(z) = \frac{5\Omega_{\rm m} E_z}{2} \int_z^\infty \frac{1 + y}{E_z^3(y)} dy.$$

In our subsequent analysis, the redshift value is assumed to be the median of the distribution of the sources. The results are summarized in Table 3.

6.3. Clustering Signals

Figure 7 shows the values of r_0 as a function of redshift for our SMG candidates and for comparison samples. We also plot the measurements from the literature for 24 μm–selected galaxies (Dolley et al. 2014; Solarz et al. 2015), 100 μm–selected Herschel SMGs (Magliocchetti et al. 2013), 250 μm–selected Herschel sources (Amvrosiadis et al. 2019), 850 μm–selected SMGs (Webb et al. 2003; Blain et al. 2004; Weiß et al. 2009; Williams et al. 2011; Hickox et al. 2012; Chen et al. 2016; Wilkinson et al. 2017; An et al. 2019), and quasars (Myers et al. 2006; Porciani & Norberg 2006; Shen et al. 2007; Eftekharzadeh et al. 2015). The measured r_0 (or b) of our SMG candidates and the comparison samples decline with decreasing redshift (Table 3). This trend is expected, since dark matter clustering evolves rapidly as the universe evolves with time (dotted curves in Figure 6), resulting in a decrease of the measured value of r_0 (or b) of a biased population. The preceding analysis only illustrates the clustering signals of our sample, so combining the knowledge of clustering signals and halo masses of our samples is more meaningful. Our SMG candidates reside in a halo with a typical mass of $\simeq(2.0 \pm 0.5) \times 10^{13} h^{-1} M_\odot$ across the redshift range $0.5 < z < 3$. In general, passive galaxies have stronger clustering signals than the comparison star-forming galaxies and SMG candidates, indicating that passive galaxies preferentially reside in more massive halos compared to star-forming galaxies and SMG candidates at fixed stellar mass and redshift (see also Hartley et al. 2008, 2010; McCracken et al. 2010; Lin et al. 2012, 2016; Sato et al. 2014; Ji et al. 2018). On the other hand, there is no significant difference between the SMG candidates and the comparison star-forming galaxies for the same stellar-mass cut, suggesting that these two populations reside in similar mass halos. Similar trends can also be found in earlier studies of galaxy samples at $1.5 < z < 2.5$ (Béthermin et al. 2014) and at $1 < z < 5$ (Chen et al. 2016; An et al. 2019). This result implies that merging events may not be the only triggering mechanism for SMGs, since we expect that more biased regions will result in higher merging and/or interaction rates (Lin et al. 2010; de Ravel et al. 2011; Sobral et al. 2011). However, further studies
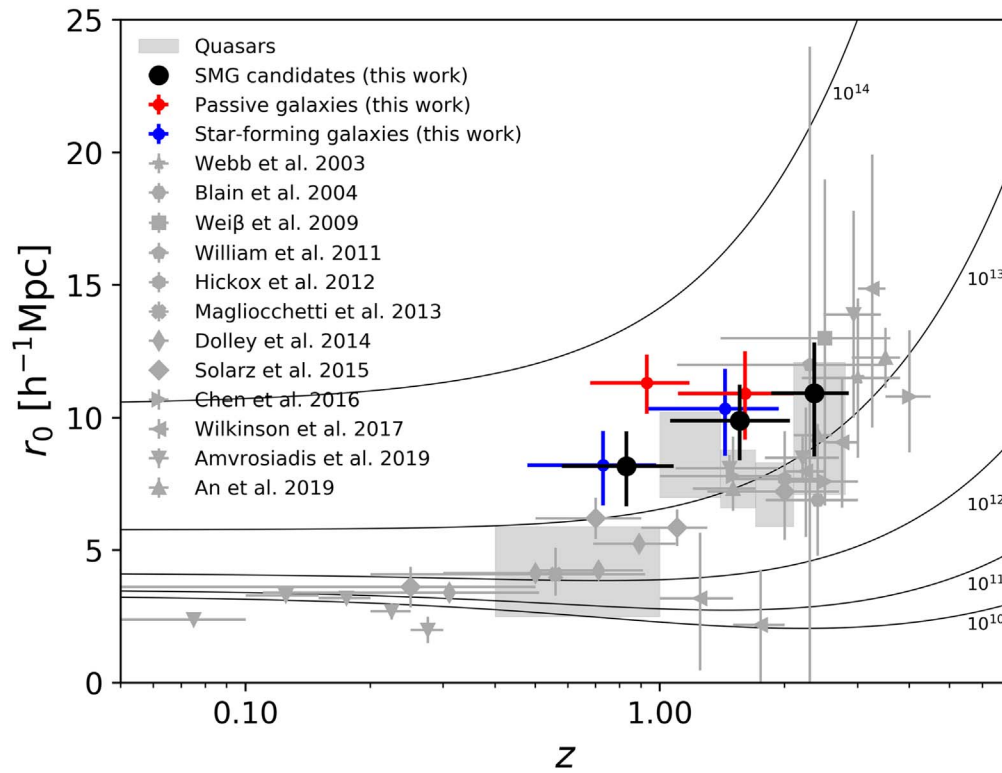
Figure 7. Redshift evolution of the clustering length r_0 for our SMG candidates and the comparison samples. The data points are slightly offset horizontally, to avoid overlap. We find no evidence that the clustering signal of SMG candidates exhibits an evolution with redshift. SMG candidates reside in a typical halo mass of $\simeq (2.0 \pm 0.5) \times 10^{13} \, h^{-1} \, M_\odot$ across the redshift range of $0.5 < z < 3$. We also show the estimated r_0 of 24 μm–selected galaxies (Dolley et al. 2014; Solarz et al. 2015), SMGs (Webb et al. 2003; Blain et al. 2004; Weiß et al. 2009; Williams et al. 2011; Hickox et al. 2012; Magliocchetti et al. 2013; Chen et al. 2016; Wilkinson et al. 2017; Amvrosiadis et al. 2019; An et al. 2019), and quasars (Myers et al. 2006; Porciani & Norberg 2006; Shen et al. 2007; Eftekharzadeh et al. 2015) in literature for comparison.

are still needed to address the question of which mechanisms increase the dust obscuration within galaxies.

At $z = 1$–3, the clustering measured for our SMG candidates is in broad agreement with that of previous studies, except that Wilkinson et al. (2017) found a weaker clustering strength at $z \simeq 1$–2, although, with large uncertainties (so not significantly different from our results). We find no strong evidence that the halo masses of SMG candidates exhibit any evolution with redshift. This is in agreement with what was found in Chen et al. (2016), Amvrosiadis et al. (2019), and An et al. (2019) but is contrary to what was suggested in Wilkinson et al. (2017), in which they found that SMG activity seems to be shifting to less massive halos, consistent with an early downsizing scenario. We attribute this to their large measurement uncertainties or the different methodologies that are adopted in the clustering analyses. We note, however, that the term "downsizing" is a relative notion. While we do not observe a significant decrease of halo masses of SMGs with decreasing redshifts, which is why we say we do not observe downsizing, we are ultimately comparing halo masses at different redshifts. Given the same mass, halos at higher redshifts are expected to grow into more massive halos at the present day. From this perspective, our results support the downsizing scenario such that the SMGs at lower redshifts are formed within halos that are, on average, smaller in mass at the present day.

Our result provides a meaningful constraint on the clustering amplitude of SMGs at $z \simeq 0.5$–1, a key redshift range where downsizing effects are expected to take place. At $z \simeq 0.5$–1, the clustering signal of our SMG candidates appears to be a little higher than in the previous studies of Magliocchetti et al. (2013), Dolley et al. (2014), and Solarz et al. (2015); although, there are large uncertainties. It is worth noting that the majority of the sources in the aforementioned studies represent a fainter population with $L_{\rm IR} \simeq 10^{11} \, L_\odot$. On the other hand, our SMG candidates are bright at 450 μm ($\simeq$5 mJy; see Section 4), which corresponds to $L_{\rm IR} \simeq 10^{12} \, L_\odot$ at $z = 0.5$–1 if we convert the 450 μm flux density into the total $L_{\rm IR}$ by using the average ALESS 870 μm SEDs (da Cunha et al. 2015). Therefore, a stronger clustering signal is expected in our sample, since the galaxies with higher $L_{\rm IR}$ tend to have stronger clustering (Dolley et al. 2014; Toba et al. 2017).

By comparing the $z < 1$ measurements and those at $z > 1$, along with the quasars, there is a clear trend for clustering length to increase with redshift at earlier epochs, which is consistent with downsizing behavior. The question now arises: are the SMGs at $z < 1$ the same dusty population as those at $z > 1$? The dusty star-forming galaxy population seems to have lower $L_{\rm IR}$ at low redshifts (e.g., Magliocchetti et al. 2013; Dolley et al. 2014; Solarz et al. 2015), which prevents us from making a direct comparison with the SMGs at high redshifts. Nevertheless, our understanding of the clustering properties in SMGs is still far from complete. The study of the redshift

evolution of the clustering properties in SMGs with comparable L_{IR} at a similar wavelength is required to solve this problem. We expect this situation to be improved with next-generation submillimeter telescopes, such as Atacama Large Aperture Submm/mm Telescope (AtLAST; Klaassen et al. 2019). A future AtLAST survey with increased sensitivity at $350\,\mu m$, and a larger field of view ($\simeq 1°$) will detect fainter populations ($L_{IR} \simeq 10^{11}\,L_{\odot}$) at larger scales.

7. Summary

By combining SCUBA-2 data from the ongoing JCMT Large Program STUDIES and the archive in the CANDELS/ COSMOS field, we have obtained an extremely deep $450\,\mu m$ image ($1\sigma = 0.56$ mJy beam^{-1}) covering $\simeq 300$ arcmin^2: by far the deepest image ever observed at $450\,\mu m$. We obtain a sample of 221 SMGs from this image; however, the sample size is too small to meaningfully study the redshift evolution of the clustering of the population.

We select a robust (S/N $\geqslant$ 4) and flux-limited ($\geqslant 4$ mJy) sample of 164 SMGs that have K-band counterparts in the COSMOS2015 catalog identified based on radio or mid-infrared imaging, which allows us to employ their optical and near-infrared colors. Ultilizing this SMG sample and the 4705 K-band-selected non-SMGs lying within the $\leqslant 1$ mJy beam^{-1} noise level region of the $450\,\mu m$ image as a training set, we develop a machine-learning classifier to identify SMG candidates in the full COSMOS field. We employ the K-band magnitudes and color–color pairs based on the 13-broadband photometry measurements ($uBVri^{+}z^{++}JHK[3.6][4.5][5.8][8.0]$) available in this field for the machine-learning algorithm. Our main findings are as follows:

1. Our trained classifier labels 6182 SMG candidates in the wider COSMOS field from the COSMOS2015 catalog with $m_K < 24.5$ mag$_{AB}$ across an effective area of 1.6 deg^2.
2. The number density, VLA 3 GHz and/or MIPS $24\,\mu m$ detection rates, redshift and stellar-mass distributions, and the stacked $450\,\mu m$ flux densities of the SMG candidates across the COSMOS field agree with the measurements made in the much smaller CANDELS field, all supporting the effectiveness of the classifier. The high completeness (76% $\pm$ 7%) and precision (82% $\pm$ 7%) of our SMG candidates as judged from their detection in longer-wavelength ALMA observations further supports our machine-learning algorithm.
3. We found that the SMG candidates tend to have higher reddening compared to comparison star-forming and passive galaxies that are matched in redshift and stellar mass. The SMG candidates also have younger stellar population ages than the comparison star-forming and passive galaxies, even though their overall distribution of stellar masses is similar. This suggests that the SMG candidates may have more recent star formation and, consequently, have a higher proportion of young stars.
4. The SMG candidates have a median effective radius of $5.5^{+0.3}_{-0.4}$ kpc and a median Sérsic index of 1.1 ± 0.1. These measurements are consistent with previous studies within the uncertainties. Our results show that SMG candidates are significantly more extended and more disk-like than the comparison star-forming and passive galaxies.
5. We measured the two-point autocorrelation function of the SMG candidates from $z = 3$ down to $z = 0.5$

and found that they reside in halos with masses of $\simeq (2.0 \pm 0.5) \times 10^{13} h^{-1} M_{\odot}$ across this redshift range. However, we do not find evidence of downsizing that has been suggested by other recent observational studies.

Appendix
Clustering Results of SMG Candidates Identified by Other Algorithms

We adopt the SMG candidates that are identified by a decision tree (better recall) and random forest (better precision), and we estimate their clustering signals by using the same procedures as in Section 6. The results are shown in Table A1.

Table A1
Clustering Results of SMG Candidates Identified by Decision Tree and Random Forest Algorithms

Sample	Redshift	N_s[a]	b	r_0 (h^{-1} Mpc)	$\log(M_{\mathrm{halo}})$ ($h^{-1} M_\odot$)
SMG candidates (Decision Tree)	$0.5 < z < 1.0$	2406	$2.4^{+0.2}_{-0.2}$	$8.1^{+0.7}_{-0.8}$	$13.4^{+0.1}_{-0.1}$
	$1.0 < z < 2.0$	8919	$3.2^{+0.2}_{-0.2}$	$8.0^{+0.5}_{-0.5}$	$13.1^{+0.1}_{-0.1}$
	$2.0 < z < 3.0$	3115	$6.2^{+0.5}_{-0.5}$	$11.5^{+1.0}_{-1.0}$	$13.3^{+0.1}_{-0.1}$
SMG candidates (Random Forest)	$0.5 < z < 1.0$	143	...	...	...
	$1.0 < z < 2.0$	730	$5.3^{+1.0}_{-1.3}$	$13.8^{+3.0}_{-3.6}$	$13.8^{+0.2}_{-0.4}$
	$2.0 < z < 3.0$	993	$6.0^{+1.3}_{-1.7}$	$11.0^{+2.8}_{-3.5}$	$13.2^{+0.3}_{-0.5}$

Note.
[a] Sample sizes of our samples in the corresponding redshift bins.

The decision tree and random forest label 14440 and 1866 SMG candidates, respectively, from the COSMOS2015 catalog at redshift $0.5 < z < 3.0$. The larger number of SMG candidates found using the decision tree algorithm is expected, since the higher recall will select more sources, but as a tradeoff, the precision decreases. In contrast, the situation is reserved for the random forest algorithm. The stacked fluxes of the SMG candidates identified by the decision tree and random forest methods are (3.0 ± 0.1) mJy and (6.4 ± 0.3) mJy, respectively. As expected, we find a lower stacked flux with the decision tree and a higher one with the random forest, essentially reflecting their precision. It is worth noting that the random forest algorithm only labels $\sim$100 SMG candidates in the redshift bin of $0.5 < z < 1.0$, and consequently, we do not have sufficient data points for the clustering analyses in this redshift bin. Nevertheless, the clustering signals do not show significant differences from the results of XGBoost-identified SMG candidates (Table 3), indicating that our final results do not strongly depend upon the method we choose.

ORCID iDs

Chen-Fatt Lim ⓘ https://orcid.org/0000-0003-1213-9360
Chian-Chou Chen ⓘ https://orcid.org/0000-0002-3805-0789
Ian Smail ⓘ https://orcid.org/0000-0003-3037-257X
Wei-Hao Wang ⓘ https://orcid.org/0000-0003-2588-1265
Yen-Ting Lin ⓘ https://orcid.org/0000-0001-7146-4687
Yoshiki Toba ⓘ https://orcid.org/0000-0002-3531-7863
Yu-Yen Chang ⓘ https://orcid.org/0000-0002-6720-8047
YiPing Ao ⓘ https://orcid.org/0000-0003-3139-2724
Arif Babul ⓘ https://orcid.org/0000-0003-1746-9529
David L. Clements ⓘ https://orcid.org/0000-0002-9548-5033
Christopher J. Conselice ⓘ https://orcid.org/0000-0003-1949-7638
Yu Gao ⓘ https://orcid.org/0000-0003-0007-2197
Thomas R. Greve ⓘ https://orcid.org/0000-0002-2554-1837
Luis C. Ho ⓘ https://orcid.org/0000-0001-6947-5846
Sungwook E. Hong ⓘ https://orcid.org/0000-0003-4923-8485
Ho Seong Hwang ⓘ https://orcid.org/0000-0003-3428-7612
Maciej Koprowski ⓘ https://orcid.org/0000-0001-5785-1154
Michał J. Michałowski ⓘ https://orcid.org/0000-0001-9033-4140
Xinwen Shu ⓘ https://orcid.org/0000-0002-7020-4290
James M. Simpson ⓘ https://orcid.org/0000-0002-8521-1995

References

Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, ApJ, 619, 697
Alaghband-Zadeh, S., Chapman, S. C., Swinbank, A. M., et al. 2012, MNRAS, 424, 2232
Amvrosiadis, A., Valiante, E., Gonzalez-Nuevo, J., et al. 2019, MNRAS, 483, 4649
An, F. X., Simpson, J. M., Smail, I., et al. 2019, ApJ, 886, 48
An, F. X., Stach, S. M., Smail, I., et al. 2018, ApJ, 862, 101
Aretxaga, I., Hughes, D. H., Coppin, K., et al. 2007, MNRAS, 379, 1571
Armus, L., Mazzarella, J. M., Evans, A. S., et al. 2009, PASP, 121, 559
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Assef, R. J., Kochanek, C. S., Ashby, M. L. N., et al. 2011, ApJ, 728, 56
Avni, Y. 1976, ApJ, 210, 642
Barger, A. J., Cowie, L. L., & Richards, E. A. 2000, AJ, 119, 2092
Barger, A. J., Cowie, L. L., Sanders, D. B., et al. 1998, Natur, 394, 248
Barger, A. J., Cowie, L. L., & Sanders, D. B. 1999, ApJL, 518, L5
Barger, A. J., Wang, W.-H., Cowie, L. L., et al. 2012, ApJ, 761, 89
Baugh, C. M., & Efstathiou, G. 1993, MNRAS, 265, 145
Bergstra, J., & Bengio, Y. 2012, J. Mach. Learn. Res., 13, 281
Bertin, E., & Arnouts, S. 1996, A&As, 117, 393
Béthermin, M., Kilbinger, M., Daddi, E., et al. 2014, A&A, 567, A103
Blain, A. W., Chapman, S. C., Smail, I., & Ivison, R. 2004, ApJ, 611, 725
Bourne, N., Dunlop, J. S., Merlin, E., et al. 2017, MNRAS, 467, 1360
Breiman, L. 1996, Machine Learning, 24, 123
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Calzetti, D. 2001, PASP, 113, 1449
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Casey, C. M., Chen, C.-C., Cowie, L. L., et al. 2013, MNRAS, 436, 1919
Chabrier, G. 2003, PASP, 115, 763
Chang, Y.-Y., Ferraro, N., Wang, W.-H., et al. 2018, ApJ, 865, 103
Chapin, E. L., Berry, D. S., Gibb, A. G., et al. 2013, MNRAS, 430, 2545
Chapman, S. C., Blain, A. W., Ivison, R. J., & Smail, I. R. 2003, Natur, 422, 695
Chapman, S. C., Blain, A. W., Smail, I., & Ivison, R. J. 2005, ApJ, 622, 772
Chen, C.-C., Smail, I., Swinbank, A. M., et al. 2015, ApJ, 799, 194
Chen, C.-C., Smail, I., Swinbank, A. M., et al. 2016, ApJ, 831, 91
Chen, T., & Guestrin, C. 2016, arXiv:1603.02754
Coil, A. L., Hennawi, J. F., Newman, J. A., Cooper, M. C., & Davis, M. 2007, ApJ, 654, 115
Coil, A. L., Newman, J. A., Cooper, M. C., et al. 2006, ApJ, 644, 671
Coil, A. L., Newman, J. A., Croton, D., et al. 2008, ApJ, 672, 153
Conselice, C. J., Chapman, S. C., & Windhorst, R. A. 2003, ApJL, 596, L5
Cooray, A., Amblard, A., Wang, L., et al. 2010, A&A, 518, L22
Cowie, L. L., Songaila, A., Hu, E. M., & Cohen, J. G. 1996, AJ, 112, 839
da Cunha, E., Walter, F., Smail, I. R., et al. 2015, ApJ, 806, 110
Danielson, A. L. R., Swinbank, A. M., Smail, I., et al. 2017, ApJ, 840, 78
Davis, M., & Peebles, P. J. E. 1983, ApJ, 267, 465
de Ravel, L., Kampczyk, P., le Fèvre, O., et al. 2011, arXiv:1104.5470
Dempsey, J. T., Friberg, P., Jenness, T., et al. 2013, MNRAS, 430, 2534
Dolley, T., Brown, M. J. I., Weiner, B. J., et al. 2014, ApJ, 797, 125
Downes, A. J. B., Peacock, J. A., Savage, A., & Carrie, D. R. 1986, MNRAS, 218, 31
Dudzevičiūtė, U., Smail, I., Swinbank, A. M., et al. 2020, MNRAS, 494, 3828
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Durkalec, A., Le Fèvre, O., Pollo, A., et al. 2015, A&A, 583, A128
Eales, S., Lilly, S., Gear, W., et al. 1999, ApJ, 515, 518
Eftekharzadeh, S., Myers, A. D., White, M., et al. 2015, MNRAS, 453, 2779
Elbaz, D., Leiton, R., Nagar, N., et al. 2018, A&A, 616, A110
Engel, H., Tacconi, L. J., Davies, R. I., et al. 2010, ApJ, 724, 233
Fabian, A. C. 1999, MNRAS, 308, L39
Farrah, D., Rowan-Robinson, M., Oliver, S., et al. 2001, MNRAS, 326, 1333

Fernández Lorenzo, M., Sulentic, J., Verdes-Montenegro, L., & Argudo-Fernández, M. 2013, MNRAS, 434, 325
Frayer, D. T., Ivison, R. J., Scoville, N. Z., et al. 1998, ApJL, 506, L7
Friedman, J. H. 2000, AnSta, 29, 1189
Friedman, J. H. 2002, Comput. Stat. Data Anal., 38, 367
Geach, J. E., Chapin, E. L., Coppin, K. E. K., et al. 2013, MNRAS, 432, 53
Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789
Gilli, R., Daddi, E., Chary, R., et al. 2007, A&A, 475, 83
Greve, T. R., Bertoldi, F., Smail, I., et al. 2005, MNRAS, 359, 1165
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Hainline, L. J., Blain, A. W., Smail, I., et al. 2011, ApJ, 740, 96
Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, ApJ, 821, 123
Hartley, W. G., Almaini, O., Cirasuolo, M., et al. 2010, MNRAS, 407, 1212
Hartley, W. G., Lane, K. P., Almaini, O., et al. 2008, MNRAS, 391, 1301
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hickox, R. C., Wardlow, J. L., Smail, I., et al. 2012, MNRAS, 421, 284
Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, A&A, 498, 725
Holland, W. S., Bintley, D., Chapin, E. L., et al. 2013, MNRAS, 430, 2513
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2005, ApJ, 630, 705
Hsu, L.-Y., Cowie, L. L., Chen, C.-C., Barger, A. J., & Wang, W.-H. 2016, ApJ, 829, 25
Hughes, D. H., Serjeant, S., Dunlop, J., et al. 1998, Natur, 394, 241
Ichikawa, T., Kajisawa, M., & Akhlaghi, M. 2012, MNRAS, 422, 1014
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., McCracken, H. J., le Fèvre, O., et al. 2013, A&A, 556, A55
Infante, L. 1994, A&A, 282, 353
Jenness, T., Cavanagh, B., Economou, F., & Berry, D. S. 2008, in ASP Conf. Ser. 394, Astronomical Data Analysis Software and Systems XVII, ed. R. W. Argyle, P. S. Bunclark, & J. R. Lewis (San Francisco, CA: ASP), 565
Ji, Z., Giavalisco, M., Williams, C. C., et al. 2018, ApJ, 862, 135
Jones, E., Oliphant, T., & Peterson, P. 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org
Jose, C., Baugh, C. M., Lacey, C. G., & Subramanian, K. 2017, MNRAS, 469, 4428
Kashikawa, N., Yoshida, M., Shimasaku, K., et al. 2006, ApJ, 637, 631
Klaassen, P., Mroczkowski, T., Bryan, S., et al. 2019, BAAS, 51, 58
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kohavi, R. 1995, in Proc. 14th Int. Joint Conf. on Artificial Intelligence 2, IJCAI'95 (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 1137
Koprowski, M. P., Dunlop, J. S., Michałowski, M. J., et al. 2016, MNRAS, 458, 4321
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Lang, P., Schinnerer, E., Smail, I., et al. 2019, ApJ, 879, 54
Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, ApJ, 703, 222
Lee, K.-S., Giavalisco, M., Gnedin, O. Y., et al. 2006, ApJ, 642, 63
Lilly, S. J., Eales, S. A., Gear, W. K. P., et al. 1999, ApJ, 518, 641
Lim, C.-F., Wang, W.-H., Smail, I., et al. 2020, ApJ, 889, 80
Limber, D. N. 1953, ApJ, 117, 134
Lin, L., Capak, P. L., Laigle, C., et al. 2016, ApJ, 817, 97
Lin, L., Cooper, M. C., Jian, H.-Y., et al. 2010, ApJ, 718, 1158
Lin, L., Dickinson, M., Jian, H.-Y., et al. 2012, ApJ, 756, 71
Lindner, R. R., Baker, A. J., Omont, A., et al. 2011, ApJ, 737, 83
Liu, R. H., Hill, R., Scott, D., et al. 2019, MNRAS, 489, 1770
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Maddox, S. J., Dunne, L., Rigby, E., et al. 2010, A&A, 518, L11
Magliocchetti, M., Cirasuolo, M., McLure, R. J., et al. 2008, MNRAS, 383, 1131
Magliocchetti, M., Lapi, A., Negrello, M., de Zotti, G., & Danese, L. 2014, MNRAS, 437, 2263
Magliocchetti, M., Popesso, P., Rosario, D., et al. 2013, MNRAS, 433, 127
Maraston, C., Pforr, J., Renzini, A., et al. 2010, MNRAS, 407, 830
McCracken, H. J., Capak, P., Salvato, M., et al. 2010, ApJ, 708, 202
McCracken, H. J., Wolk, M., Colombi, S., et al. 2015, MNRAS, 449, 901
Menéndez-Delmestre, K., Blain, A. W., Swinbank, M., et al. 2013, ApJ, 767, 151
Michałowski, M. J., Dunlop, J. S., Cirasuolo, M., et al. 2012a, A&A, 541, A85
Michałowski, M. J., Dunlop, J. S., Ivison, R. J., et al. 2012b, MNRAS, 426, 1845
Michałowski, M. J., Dunlop, J. S., Koprowski, M. P., et al. 2017, MNRAS, 469, 492
Miller, T. B., Chapman, S. C., Aravena, M., et al. 2018, Natur, 556, 469
Mitchell-Wynne, K., Cooray, A., Gong, Y., et al. 2012, ApJ, 753, 23
Myers, A. D., Brunner, R. J., Nichol, R. C., et al. 2007, ApJ, 658, 85
Myers, A. D., Brunner, R. J., Richards, G. T., et al. 2006, ApJ, 638, 622

Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162
Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, MNRAS, 396, 19
Ouchi, M., Hamana, T., Shimasaku, K., et al. 2005, ApJL, 635, L117
Peacock, J. A. 1983, MNRAS, 202, 615
Peacock, J. A. 1991, MNRAS, 253, 1P
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Peebles, P. J. E. 1980, The large-scale structure of the universe (Princeton, NJ: Princeton Univ. Press)
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097
Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16
Pope, A., Scott, D., Dickinson, M., et al. 2006, MNRAS, 370, 1185
Porciani, C., & Norberg, P. 2006, MNRAS, 371, 1824
Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, A&A, 523, A13
Provost, F., Fawcett, T., & Kohavi, R. 1997, in Proc. of the Fifteenth Int. Conf. on Machine Learning (San Francisco, CA : Morgan Kaufmann Publishers Inc.), 445
Rennehan, D., Babul, A., Hayward, C. C., et al. 2020, MNRAS, 493, 4607
Rijsbergen, C. J. V. 1979, Information Retrieval (2nd ed.; Newton, MA: Butterworth-Heinemann)
Roche, N., & Eales, S. A. 1999, MNRAS, 307, 703
Sanders, D. B., & Mirabel, I. F. 1996, ARA&A, 34, 749
Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86
Sanders, D. B., Soifer, B. T., Elias, J. H., Neugebauer, G., & Matthews, K. 1988, ApJL, 328, L35
Sato, T., Sawicki, M., & Arcila-Osejo, L. 2014, MNRAS, 443, 2661
Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, AJ, 110, 68
Scott, S. E., Fox, M. J., Dunlop, J. S., et al. 2002, MNRAS, 331, 817
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Scranton, R., Johnston, D., Dodelson, S., et al. 2002, ApJ, 579, 48
Sérsic, J. L. 1963, BAAA, 6, 41
Sérsic, J. L. 1968, Atlas de Galaxias Australes (Cordoba: Observatorio Astronomico)
Shen, S., Mo, H. J., White, S. D. M., et al. 2003, MNRAS, 343, 978
Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, AJ, 133, 2222
Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1
Silk, J., & Rees, M. J. 1998, A&A, 331, L1
Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015, ApJ, 807, 128
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
Smith, R. E., Peacock, J. A., Jenkins, A., et al. 2003, MNRAS, 341, 1311
Smolčić, V., Delvecchio, I., Zamorani, G., et al. 2017, A&A, 602, A2
Sobral, D., Best, P. N., Smail, I., et al. 2011, MNRAS, 411, 675
Solarz, A., Pollo, A., Takeuchi, T. T., et al. 2015, A&A, 582, A58
Stach, S. M., Dudzevičiūtė, U., Smail, I., et al. 2019, MNRAS, 487, 4648
Starikova, S., Berta, S., Franceschini, A., et al. 2012, ApJ, 751, 126
Swinbank, A. M., Simpson, J. M., Smail, I., et al. 2014, MNRAS, 438, 1267
Swinbank, A. M., Smail, I., Chapman, S. C., et al. 2010, MNRAS, 405, 234
Swinbank, M., Smail, I., Lindner, C., et al. 2006, MNRAS, 371, 465
Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246
Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, ApJ, 761, 152
Targett, T. A., Dunlop, J. S., McLure, R. J., et al. 2011, MNRAS, 412, 295
Toba, Y., Nagao, T., Kajisawa, M., et al. 2017, ApJ, 835, 36
Toft, S., Smolčić, V., Magnelli, B., et al. 2014, ApJ, 782, 68
Trayford, J. W., Theuns, T., Bower, R. G., et al. 2016, MNRAS, 460, 3925
van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van Kampen, E., Smith, D. J. B., Maddox, S., et al. 2012, MNRAS, 426, 3455
Wall, J. V., Pope, A., & Scott, D. 2008, MNRAS, 383, 435
Wang, T., Schreiber, C., Elbaz, D., et al. 2019, Natur, 572, 211
Wang, W.-H., Lin, W.-C., Lim, C.-F., et al. 2017, ApJ, 850, 37
Wardlow, J. L., Smail, I., Coppin, K. E. K., et al. 2011, MNRAS, 415, 1479
Webb, T. M., Eales, S., Foucaud, S., et al. 2003, ApJ, 582, 6
Weiß, A., Kovács, A., Coppin, K., et al. 2009, ApJ, 707, 1201
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
Yun, M. S., Scott, K. S., Guo, Y., et al. 2012, MNRAS, 420, 957
Zavala, J. A., Aretxaga, I., Dunlop, J. S., et al. 2018, MNRAS, 475, 5585
Zehavi, I., Blanton, M. R., Frieman, J. A., et al. 2002, ApJ, 571, 172
Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1