

# Herschel-ATLAS : the spatial clustering of low- and high-redshift submillimetre galaxies

A. Amvrosiadis,<sup>1\*</sup> E. Valiante,<sup>1</sup> J. Gonzalez-Nuevo,<sup>2</sup> S. J. Maddox,<sup>1,3</sup> M. Negrello,<sup>1</sup> S. A. Eales,<sup>1</sup> L. Dunne,<sup>1</sup> L. Wang,<sup>4,5</sup> E. van Kampen,<sup>6</sup> G. De Zotti,<sup>7</sup> M. W. L. Smith,<sup>1</sup> P. Andreani,<sup>6</sup> J. Greenslade,<sup>8</sup> C. Tai-An<sup>8</sup> and M. J. Michałowski<sup>3,9</sup>

<sup>1</sup>*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK*

<sup>2</sup>*Departamento de Física, Universidad de Oviedo, C. Federico Garcia Lorca 18, E-33007 Oviedo, Spain*

<sup>3</sup>*Institute for Astronomy, The University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*

<sup>4</sup>*SRON Netherlands Institute for Space Research, Landleven 12, NL-9747 AD, Groningen, the Netherlands*

<sup>5</sup>*Kapteyn Astronomical Institute, University of Groningen, Postbus 800, NL-9700 AV, Groningen, the Netherlands*

<sup>6</sup>*ESO, Karl-Schwarzschild-Str 2, D-85748 Garching bei Muenchen, Germany*

<sup>7</sup>*INAF-Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122 Padova, Italy*

<sup>8</sup>*Astrophysics Group, Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*

<sup>9</sup>*Astronomical Observatory Institute, Faculty of Physics, Adam Mickiewicz University, ul. Stoleczna 36, 60-286 Poznań, Poland*

Accepted 2018 November 2. Received 2018 November 2; in original form 2018 August 9

## ABSTRACT

We present measurements of the angular correlation function of sub-millimetre (sub-mm) galaxies (SMGs) identified in four out of the five fields of the *Herschel Astrophysical Terahertz Large Area Survey (H-ATLAS) – GAMA-9h, GAMA-12h, GAMA-15h, and NGP* – with flux densities  $S_{250\ \mu\text{m}} > 30\ \text{mJy}$  at  $250\ \mu\text{m}$ . We show that galaxies selected at this wavelength trace the underlying matter distribution differently at low and high redshifts. We study the evolution of the clustering finding that at low redshifts sub-mm galaxies exhibit clustering strengths of  $r_0 \sim 2\text{--}3\ h^{-1}\ \text{Mpc}$ , below  $z < 0.3$ . At high redshifts, on the other hand, we find that sub-mm galaxies are more strongly clustered with correlation lengths  $r_0 = 8.1 \pm 0.5$ ,  $8.8 \pm 0.8$ , and  $13.9 \pm 3.9\ h^{-1}\ \text{Mpc}$  at  $z = 1\text{--}2$ ,  $2\text{--}3$ , and  $3\text{--}5$ , respectively. We show that sub-mm galaxies across the redshift range  $1 < z < 5$ , typically reside in dark-matter halos of mass of the order of  $\sim 10^{12.5}\text{--}10^{13.0}\ h^{-1}\ M_\odot$  and are consistent with being the progenitors of local massive elliptical galaxies that we see in the local Universe.

**Key words:** galaxies: high-redshift – galaxies: evolution – large-scale structure of Universe – submillimetre: galaxies.

## 1 INTRODUCTION

Sub-millimetre galaxies (SMGs, e.g. Smail, Ivison & Blain 1997; Blain et al. 2002, and references therein) are considered to be among the most intensively star-forming objects in the Universe. The ultraviolet (UV) radiation from their newly born stars is absorbed by the dust and then re-emitted at far-infrared (FIR)/sub-millimetre (sub-mm) wavelengths. The total infrared luminosities ( $L_{\text{IR}}$ ) of some of the brightest SMGs can reach values higher than a few times  $10^{12}\ L_\odot$  and sometimes higher than  $10^{13}\ L_\odot$  (comparable to Ultra-Luminous InfraRed Galaxies (ULIRGs; Lonsdale et al. 2006) we see in the local Universe) with inferred star formation rates of  $\sim 1000\ M_\odot\ \text{yr}^{-1}$  (e.g. Swinbank et al. 2014). The shape of their spectral energy distribution (SED) at these wavelengths (Rayleigh–Jeans limit)

approximates a power-law that decreases with increasing wavelength, meaning that it is subject to strong negative K-corrections (e.g. Casey, Narayanan & Cooray 2014). As a result, these objects are predominantly found at high redshift, in the range  $z \sim 1\text{--}3$  (Chapman et al. 2005; Coppin et al. 2006; Simpson et al. 2014; Chen et al. 2016a), although a substantial evolution in the co-moving number density of ULIRGs between  $z = 0$  and  $z = 2\text{--}3$  has also been reported (Daddi et al. 2005).

Despite the great success in characterizing their properties (see Casey et al. 2014, and references therein), the evolutionary stages and the nature of these high-redshift SMGs still remain largely unknown. Various galaxy evolution models have been proposed to explain the morphological transformation and quenching of these objects. The most prevailing scenarios are merger-driven galaxy evolution models, which follow the evolution of both the disc and the spheroidal components of galaxies (Almeida, Baugh & Lacey 2011), models where the star formation is fuelled by steady accre-

\* E-mail: amvrosiadisa@cardiff.ac.uk

tion of large amounts of cold gas (Davé et al. 2010) and a self-regulated galaxy evolution model (Granato et al. 2004; Lapi 2011; Cai et al. 2013).

The evolution of a galaxy population can be constrained from the measurement of its clustering strength, which provides information on the masses of dark matter halos that these galaxies reside in. There have been numerous clustering studies of SMG's identified in the short (250–500  $\mu\text{m}$ ; Cooray et al. 2010; Maddox et al. 2010; Mitchell-Wynne et al. 2012; van Kampen et al. 2012) and long (850–1100  $\mu\text{m}$ ; Webb et al. 2003; Blain et al. 2004; Scott, Dunlop & Serjeant 2006; Weiß et al. 2009; Williams et al. 2011; Hickox et al. 2012; Chen et al. 2016a,b; Wilkinson et al. 2017) submillimetre bands. Similar information can be extracted from the clustering of the unresolved FIR/sub-mm galaxies, through the measurement of the angular power spectrum of CIB anisotropies (Viero et al. 2009; Amblard et al. 2011; Viero et al. 2013; Planck Collaboration XXX 2014).

The most accurate determination of the clustering properties of SMG's up to date has been performed by Chen et al. (2016b). The authors used a sample of  $\sim 3000$  SMGs with redshifts in the range  $z \sim 1-5$ , which were selected using a colour selection technique (Chen et al. 2016a), Optical-Infrared Triple Colour (OIRTC), to preferentially select faint SMG's ( $S_{850} < 2$  mJy) in the K-band from the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007) Ultra Deep Survey (UDS). In their study, they concluded that SMG's, selected with the OIRTC technique, are strongly clustered residing in halos with typical halo masses of the order of  $M_h \sim 10^{13} h^{-1} M_\odot$  across the probed redshift range. However, these sources were not individually detected in the sub-mm wavebands and the evidence that these galaxies are SMGs was based on observations with Atacama Large Millimeter/submillimeter Array (ALMA) training set (a subset of the objects predicted to be brighter sub-mm sources), which implied that the OIRTC method is 87 per cent efficient.

More recently, Wilkinson et al. (2017) studied the clustering properties of SMG's which were identified using the 850  $\mu\text{m}$  maps of the UDS field from the SCUBA-2 Cosmology Legacy Survey (S2CLS; Geach et al. 2013). The authors used a sample of 610 SMGs for which they found counterparts using a combination of radio imaging and the optical/infrared selection technique of Chen et al. (2016a). Using ALMA observations of the brightest sources, they estimate an 80 per cent successful SMG counterpart identification. However, due to the sparse number density of SMGs, the authors relied on a cross-correlation technique, with a more abundant K-band selected sample, to measure their clustering properties. Their analysis yield similar results to Chen et al. (2016b) for  $z > 2$  SMGs, in terms of the halo masses that these galaxies reside to, but for SMGs found in the redshift range  $1 < z < 2$  they reported a downsizing effect where the SMG activity is shifted to halos with typical halo masses of the order of  $M_h \sim 10^{12} h^{-1} M_\odot$ .

In addition, both Chen et al. (2016b) and Wilkinson et al. (2017) performed the clustering analysis for typical star-forming and passive galaxies, identified in the same field using their colours. This is important in order to place the clustering results of SMGs in the broader context of galaxy evolution. However, both these studies were unable to significantly differentiate SMG clustering properties from more typical star-forming galaxies identified in the same redshift range. In addition, Hickox et al. (2012) using a sample of 126 SMGs selected at 870  $\mu\text{m}$  from the *Large APEX Bolometer Camera (LABOCA)* sub-mm survey of the *Extended Chandra Deep Field-South (ECDFS)* concluded that the clustering properties of high-redshift SMGs are consistent with measurements for optically

selected quasi-stellar objects (QSOs). Their findings support evolutionary scenarios in which powerful starburst and QSOs occur in the same systems. In all these studies, high-redshift SMGs reside in dark matter halos of the order of  $\sim 10^{13} h^{-1} M_\odot$  and seem to be consistent with being the progenitors of massive elliptical galaxies that we see in the local universe.

In order to improve the already existing measurements of the angular clustering signal of SMG, we need much larger survey areas to increase the number of detected sources and to obtain accurate redshift information. Concerning the first requirement, the *Herschel Astrophysical Terahertz Large Area Survey (H-ATLAS)* (which covers an area of more than  $\sim 600 \text{ deg}^2$ ) provides almost three orders of magnitude improvement in covered area compared to surveys conducted at 850  $\mu\text{m}$  (Chen et al. 2016a,b; Wilkinson et al. 2017). As for the later requirement, the challenge is to identify optical/near-infrared (NIR) counterparts to the sub-mm sources in order to obtain relatively well-constrained photometric or spectroscopic redshifts. This is especially challenging due to the low angular resolution at sub-mm wavelengths, which results in large positional uncertainties for the sub-mm sources. We thoroughly discuss in Section 2 how we approach this issue.

Nevertheless, these aforementioned studies provide a unique contribution to the field, enabling for the first time the characterization of the clustering properties of SMGs as a function of redshift and their role in galaxy evolution scenarios. However, they do not provide a complete picture as they fold in biases linked to the selection of SMGs at these particular wavelengths, rendering it essential to conduct similar studies at different sub-mm wavebands. In this paper, we will study the clustering properties of SMG's identified at 250  $\mu\text{m}$  in the H-ATLAS survey, with flux densities  $S_{250 \mu\text{m}} > 30$  mJy. Throughout this paper, we assume a flat  $\Lambda$ CDM cosmological model with the best-fit parameters derived from the Planck Observatory (Planck Collaboration XIII 2016), which are  $\Omega_m = 0.307$ ,  $H_0 = 69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , and  $\sigma_8 = 0.816$ .

## 2 H-ATLAS DATA

The *Herschel Astrophysical Terahertz Large Area Survey (H-ATLAS)*; Eales et al. 2010) is a survey of  $\sim 660 \text{ deg}^2$  in five far-infrared (far-IR) to submillimeter (sub-mm) photometric bands – 100, 160, 250, 350, and 500  $\mu\text{m}$  – with the Photoconductor Array Camera and Spectrometer (PACS; Poglitsch et al. 2010) and Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) cameras, which was carried out with the *Herschel Space Observatory* (Pilbratt et al. 2010). The survey comprises five different fields, three of which are located on the celestial equator (GAMA fields; Valiante et al. 2016) covering in total an area of 161.6  $\text{deg}^2$ . The other two fields are centred on the North and South Galactic Poles (NGP and SGP fields; Smith et al. 2017) covering areas of 180.1  $\text{deg}^2$  and 317.6  $\text{deg}^2$ , respectively.

The H-ATLAS fields were selected to minimize bright continuum emission from dust in the Galaxy, as seen by the *Infrared Astronomical Satellite (IRAS)* at 100  $\mu\text{m}$ . Complementary multi-wavelength data for these fields are provided by surveys spanning ultraviolet (UV) to mid-infrared (mid-IR) regimes. In particular for the GAMA fields – *GAMA-9h*, *GAMA-12h*, *GAMA-15h* – optical spectroscopic data are provided by the Galaxy and Mass Assembly survey (GAMA; Driver et al. 2009), the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009), and the 2-Degree-Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001), while optical photometric data are provided by the Galaxy Evolution Explorer (GALEX; Martin et al. 2005) and the Kilo-Degree Survey (KiDS; de Jong

**Table 1.** Herschel-ATLAS sources with measured spectroscopic redshift from CO observations.

H-ATLAS ID	$z_{\text{spec}}$	$z_{\text{phot}}$	Ref.
J134429.5+303034	2.30	2.31	H12
J114637.9-001132	3.26	2.81	H12
J132630.1+334408	2.95	3.89	H-p
J083051.0+013225	3.63	3.19	R-p
J125632.5+233627	3.57	3.56	R-p
J132427.0+284450	1.68	2.32	G13
J132859.2+292327	2.78	2.81	K-p
J084933.4+021442	2.41	2.91	L-p
J125135.3+261458	3.68	3.63	K-p
J113526.2-014606	3.13	2.28	H12
J133008.6+245900	3.11	2.36	R-p
J142413.9+022303	4.28	4.24	C11
J141351.9-000026	2.48	285	H12
J090311.6+003907	3.04	3.54	F11
J132504.4+311534	1.84	2.12	R-p
J133846.5+255055	2.34	2.69	R-p
J132301.7+341649	2.19	2.58	R-p
J091840.8+023048	2.58	3.06	H12
J133543.0+300402	2.68	2.76	H-p
J091304.9-005344	2.63	2.73	N10
J115820.1-013752	2.19	3.21	H-p
J113243.0-005108	2.58	3.92	R-p
J142935.3-002836	1.03	0.56	P13
J090740.0-004200	1.58	1.05	L12
J085358.9+015537	2.09	1.84	P13
J090302.9-014127	2.31	1.97	L12

*Notes.* The last column corresponds to references for the CO spectroscopic redshifts: N10 = Negrello et al. (2010), C11 = Cox et al. (2011), F11 = Frayer et al. (2011), H12 = Harris et al. (2012), L12 = Lupu et al. (2012), B13 = Bussmann et al. (2013), G13 = George et al. (2013), P13 = Pearson et al. (2013), H-p = Harris et al. (prep), R-p = Riechers et al. (prep), K-p = Krips et al. (prep), L-p = Lupu et al. (prep).

et al. 2015). Besides optical imaging and spectroscopy, imaging data at near-infrared (near-IR) wavelengths are available from the UKIRT Infrared Deep Sky Survey Large Area Survey (UKIDSS-LAS; Lawrence et al. 2007), the Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010), and the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING; Edge et al. 2013). In addition, radio-imaging data in the fields are provided by the Faint Images of the Radio Sky at Twenty-cm survey and the NRAO Very Large Array Sky Survey. The multi-wavelength coverage of the NGP and SGP fields is less extensive. The NGP field is covered in the optical by the SDSS and in near-IR by the UKIDSS-LAS while the SGP field is covered in the optical by KiDS and in the near-IR by VIKING.

The source catalogues of the H-ATLAS fields, which are presented in Valiante et al. (2016) and Maddox et al. (2018) for the GAMA and NGP/SGP fields, respectively, are created in three stages. First, the emission from dust in our Galaxy, which is contained in all *Herschel* images, needs to be removed before the source extraction process. We used the *Nebuliser*<sup>1</sup> algorithm, in order to remove this emission from the SPIRE images in all the three wave-

bands (more details can be found in Valiante et al. (2016) for how the algorithm works). Secondly, the Multiband Algorithm for source Detection and eXtraction (MADX; Maddox et al., in prep.) was used to identify 2.5 $\sigma$  peaks in the 250  $\mu\text{m}$  maps and to measure the flux densities at the position of those peaks in all the SPIRE bands. Before the source extraction, however, the maps were filtered with a matched-filter technique (Chapin et al. 2011) in order to reduce instrumental and confusion noise. Finally, only sources with a signal-to-noise ratio  $\geq 4$  in at least one of the three SPIRE bands were kept in the final catalogue. The 4 $\sigma$  detection limit at 250  $\mu\text{m}$  for a point source ranges from 20 mJy in the deepest regions of the maps (where tiles overlap) to 36 mJy in the non-overlapping regions.

Having extracted our sub-mm sources from our *Herschel* maps, ideally we would like to find the counterparts of these sources in other wavelengths. Identifying counterparts to these sub-mm sources, however, is a challenging task. Using likelihood ratio techniques, Bourne et al. (2016) identified SDSS optical counterparts to the sub-mm sources in the GAMA fields at  $r < 22.4$  with a 4 $\sigma$  detection at 250  $\mu\text{m}$ . The quantity  $R$  (*reliability*) corresponds to the probability that a potential counterpart is associated with a *Herschel* source. They find reliable counterparts ( $R \geq 0.8$ ) for 44,835 sources (39 per cent). In addition, Furlanetto et al. (2018) performed the same analysis for the NGP field and obtained optical counterparts for 42,429 sources (37.8 per cent). One potential caveat of this methodology however, is that it gives an artificially higher likelihood of association for high- $z$  sub-mm sources that are gravitationally lensed by local galaxies or large-scale structure (Bourne et al. 2014).

Finally, we removed local extended sources from our final extracted source catalogue. These sources were selected from the *Herschel* catalogues for having a non-zero aperture radius in either of the three SPIRE wavebands. Custom aperture photometry was carried out by Valiante et al. (2016) for the GAMA fields and Maddox et al. (2018) for the NGP and SGP fields. The number of local extended sources are 231, 226, 284, 889, and 1452 in the GAMA-9h, GAMA-12h, GAMA-15h, NGP, and SGP fields, respectively.

## 2.1 Redshift distribution of sub-mm sources

The redshift distribution of our sources is an essential ingredient in our clustering analysis. It is used to project the angular correlation function,  $w(\theta)$ , in order to recover the spatial correlation function,  $\xi(r)$ , from which the clustering properties of our galaxy population are determined.

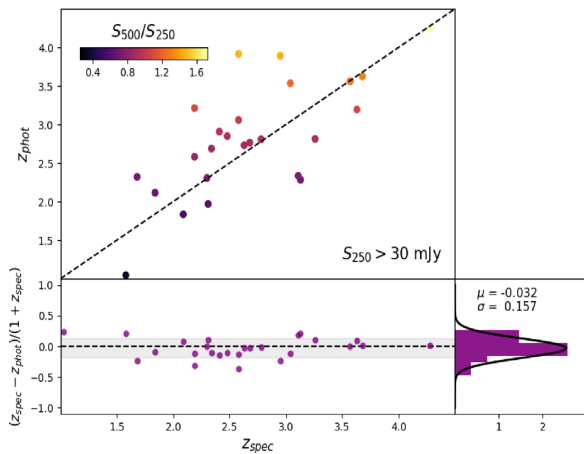
The standard approach used to estimate photometric redshifts, when only IR to sub-mm photometric data are available (in this case the SPIRE 250, 350, & 500  $\mu\text{m}$  flux densities), is to fit a *calibrated* SED template to each source in our sample. This approach has been adopted in many previous studies (Pearson et al. 2013; Ivison et al. 2016; Bakx et al. 2018).

The first step to this procedure is to determine the best-fit values of the parameters of the SED template. We adopt as our SED template a modified blackbody spectral energy distribution, consisting of two dust components with different temperatures. This is given by,

$$S_\nu = A_{\text{off}} [B_\nu(T_h)\nu^\beta + \alpha B_\nu(T_c)\nu^\beta], \quad (1)$$

where  $S_\nu$  is the flux at the rest-frame frequency  $\nu$ ,  $A_{\text{off}}$  is the normalization factor,  $B_\nu$  is the Planck blackbody function,  $\beta$  is the dust emissivity index,  $T_h$  and  $T_c$  are the temperatures of the hot and cold dust components, and  $\alpha$  is the ratio of the mass of the cold to hot dust. In order to compute our sub-mm photometric redshifts,  $z_{\text{phot}}$ ,

<sup>1</sup>The *Nebuliser* algorithm was developed by the Cambridge Astronomical Survey Unit, which can be found at <http://casu.ast.cam.ac.uk/surveys-projects/software-release/background-filtering>.



**Figure 1.** Scatter plot of  $(z_{\text{spec}} - z_{\text{phot}})/(1 + z_{\text{spec}})$  against the spectroscopic redshift,  $z_{\text{spec}}$ , for sources with CO spectroscopic redshifts in the redshift range,  $1 < z < 5$ , which are listed in Table 1. For this comparison, we exclude identified QSOs, as it has been shown that the photometric redshift estimation methodology is only reliable for starburst galaxies (Pearson et al. 2013). In the lower right corner, we show the histogram of  $\Delta z/(1 + z)$  values, as well as the mean and standard deviation from fitting a Gaussian distribution (black curve) to this histogram. Finally, in the upper panel, we show a scatter plot of  $z_{\text{spec}}$  versus  $z_{\text{phot}}$ , where the points are colour-coded based on the flux density ratio  $S_{500}/S_{250}$  at 500 and 250  $\mu\text{m}$ , respectively. The black dashed line shows the 1:1 relation.

we use the parameters found by Pearson et al. (2013):  $T_h = 46.9$ ,  $T_c = 23.9$ ,  $\alpha = 30.1$ , and  $\beta = 2$ .

We evaluate the accuracy of our sub-mm photometric redshift estimates using the same sample of source. In Fig. 1, we show  $(z_{\text{spec}} - z_{\text{phot}})/(1 + z_{\text{spec}})$  vs  $z_{\text{spec}}$ , finding that the template performs reasonably well and does not introduce any systematic offset. Fitting a Gaussian distribution to the histogram of  $\Delta z/(1 + z)$ , shown in the lower right corner of Fig. 1, we find a mean of  $-0.03$  with a standard deviation of  $\sigma_{\Delta z/(1+z)} = 0.157$  and no outliers. Similar conclusions were drawn by Ivison et al. (2016), where the authors used different templates to evaluate their performance. In the top panel of Fig. 1, we see that higher redshift sub-mm sources have preferentially redder colours as expected, where the points are colour-coded based on the flux density ratio  $S_{500}/S_{250}$  at 500 and 250  $\mu\text{m}$ , respectively.

Finally, in order to construct the redshift distributions in Fig. 2 we adopted the following procedure.: (i) if  $R < 0.8$ , we used the sub-mm photometric redshifts that were determined from our SED fitting methodology. (ii) if  $R \geq 0.8$ , we further applied an additional cut in redshift quality parameter  $Q$  (see Driver et al. 2011 for a detailed definition of the redshift quality parameter  $Q$ ). If  $Q \geq 3$ , we used the optical spectroscopic redshift, otherwise we used the optical photometric redshift. In some few cases where  $R \geq 0.8$  but none of the above information was available we used sub-mm photometric redshifts. We need to note here that this selection only concerns the clustering analysis of our low-redshift sample since the completeness of our counterpart identification method drops significantly above  $z > 0.3$  to the point where our high-redshift sample ( $z > 1$ ) is completely dominated by sub-mm sources with no counterparts.

Fig. 2 shows the redshift distribution of our sub-mm sources for all H-ATLAS fields, following the procedure outlined above. The inset plot in each panel shows a zoom into the low-redshift range of the redshift distribution of our sub-mm sources with identified counterparts. The grey histogram corresponds to sources with either

an optical photometric or spectroscopic redshift while the black histogram corresponds to sources with only optical spectroscopic redshift of quality  $Q \geq 3$ . The counterpart identification analysis has not been performed as yet for the SGP field and so in this case we only show the redshift distribution of sub-mm photometric redshifts. One thing to note here is the lack of spectroscopic redshifts in the NGP field compared to the GAMA fields, which are complemented by the GAMA survey (Driver et al. 2009).

We can clearly see that our sample of 250- $\mu\text{m}$ -selected sub-mm galaxies contain different galaxy populations at low and high redshifts (see Pearson et al. (2013) where the authors performed simulations to show that these are, in fact, two different galaxy populations rather than being a bias of the sub-mm photometric redshift estimation methodology). On the one hand, the low-redshift peak around  $z \sim 0.2 - 0.3$  is mostly associated with typical star-forming galaxies (see Bourne et al. (2016) and Furlanetto et al. (2018) for more details on the multiwavelength properties of H-ATLAS galaxies with identified counterpart), while 15–30 per cent would be classified as passive galaxies based on their optical colours (Eales et al. 2018). On the other hand, the broader part of the distribution in the redshift range  $z > 1$  is associated with sub-mm galaxies (Chapman et al. 2005).

## 2.2 Effects of sub-mm photometric redshifts

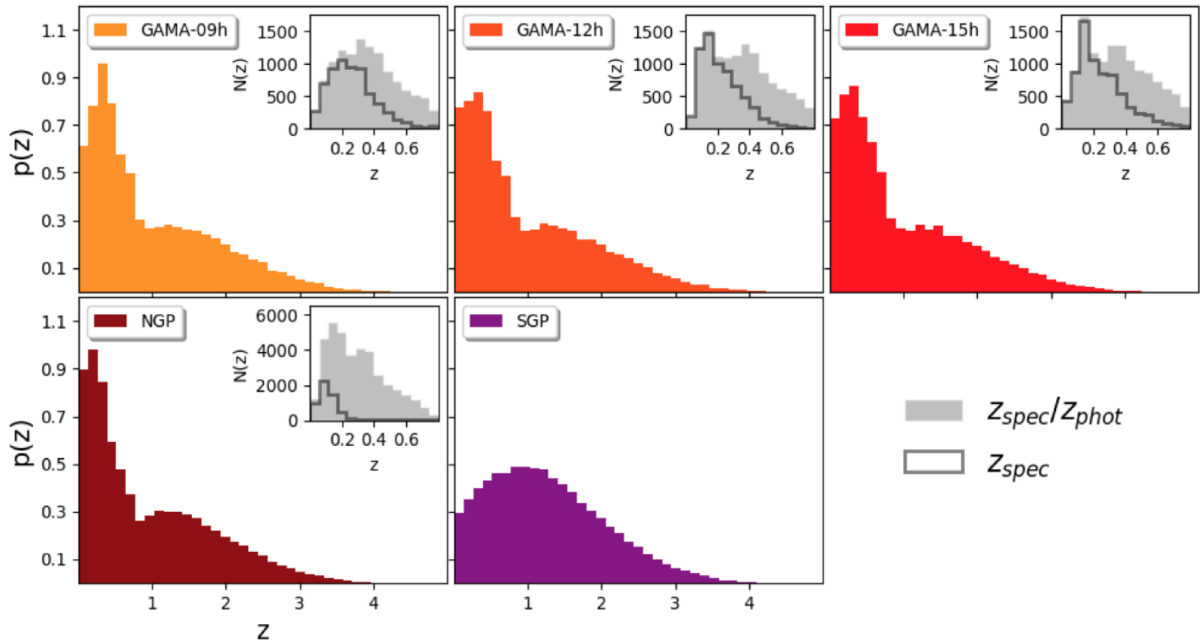
One caveat of using a FIR/sub-mm SED template fitting approach to estimate photometric redshifts for our sub-mm sources, is that the redshifts of low-redshift sources are significantly overestimated. This is clearly demonstrated in Fig. 3, where we show the comparison of  $z_{\text{spec}}$  versus  $z_{\text{phot}}$  for source in the three GAMA fields in the redshift range  $z < 0.3$ , with optically identified counterparts. This comparison highlights the importance of identifying the optical counterparts of low-redshift sub-mm galaxies, when one wishes to measure the clustering properties of high-redshift ( $z > 1$ ) sub-mm galaxies. This is the main reason why we choose not to include the SGP field in the analysis that follows:

In addition, the errors of our sub-mm photometric redshifts,  $z_{\text{phot}}$ , which are derived from the SED fitting methodology, are in most cases quite large. This means that when a tomographic analysis of the clustering is performed, a single source can be found in more than one redshift bins. If this effect is not accounted for properly, it can lead to severe biases. Cowley et al. (2017) demonstrated that seemingly similar correlation functions (from observations and simulations of SMGs) result in significantly different clustering properties due to the incorrect normalization of the correlation function of the underlying dark-matter distribution that these galaxies are tracing.

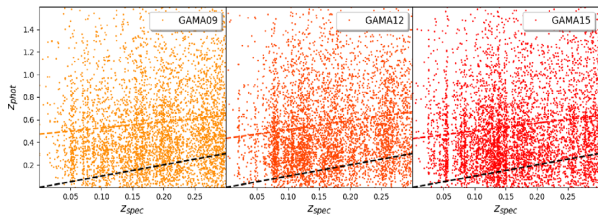
In order to account for the effect of random errors in photometric redshift estimates on  $dN/dz$ , following the analysis by Budavári et al. (2003), we estimate the redshift distribution  $p(z|W)$  of galaxies selected by our window function  $W(z_{\text{ph}})$ , as

$$p(z|W) = p(z) \int dz_{\text{ph}} W(z_{\text{ph}}) p(z_{\text{ph}}|z), \quad (2)$$

where  $p(z)$  is the initial redshift distribution,  $W(z_{\text{ph}})$  is a top-hat window function where  $W = 1$  for  $z_{\text{ph}}$  in the selected redshift interval  $z_{\text{min}} < z < z_{\text{max}}$  and  $W = 0$  otherwise, and  $p(z_{\text{ph}}|z)$  is the probability that a source with true redshift  $z$  has a photometric redshift  $z_{\text{ph}}$ . The function  $p(z_{\text{ph}}|z)$  is parametrized as a Gaussian distribution with a



**Figure 2.** The redshift distribution of sub-mm sources detected in the five fields of the *H-ATLAS* survey: *GAMA-09h* (top-left), *GAMA-12h* (top-middle), *GAMA-15h* (top-right), *NGP* (bottom-left), and *SGP* (bottom-middle). The histograms are normalized so that the area is equal to unity. The inset plot in each panel shows a zoom into the low-redshift range of the redshift distribution of our sub-mm sources with identified counterparts. The grey histogram corresponds to sources with either an optical photometric or spectroscopic redshift (see the main text for more details) while the black histogram corresponds to sources with only optical spectroscopic redshift of quality  $Q \geq 3$ . For the case of the *SGP* field, only sub-mm photometric redshifts are available.



**Figure 3.** Comparison of spectroscopic,  $z_{\text{spec}}$ , and photometric,  $z_{\text{phot}}$ , redshifts in the redshift range  $z < 0.3$ , for submillimeter sources with identified counterparts. The black dashed lines show the 1:1 relation, while the coloured dashed lines show the best-fit line that goes through the data points.

zero mean and variance  $(1+z)\sigma_{\Delta z/(1+z)}$ ,

$$p(z_{\text{ph}}|z) = \frac{1}{\sqrt{2\pi(1+z)^2\sigma_{\Delta z/(1+z)}^2}} \exp\left\{-\frac{(z - z_{\text{ph}})^2}{2(1+z)^2\sigma_{\Delta z/(1+z)}^2}\right\}, \quad (3)$$

where the dispersion is taken to be  $\sigma_{\Delta z/(1+z)} = 0.15$  as determined from the comparison of our sub-mm photometric redshifts and a sample of 26 sources with reliable CO spectroscopic redshifts.

This correction is only relevant for the clustering analysis of our high-redshift sample ( $z > 1$ ). This is because our sample is completely dominated by sources with only sub-mm photometric redshift information. Therefore, in this case, the initial redshift distribution,  $p(z)$ , is estimated by excluding sources with identified counterparts (i.e. those with reliability  $R \geq 0.8$ ).

### 3 CLUSTERING ANALYSIS

In this section, we describe the methodology we followed in order to measure the angular clustering signal.

#### 3.1 The angular two-point correlation function

The angular two-point auto-correlation function (ACF),  $w(\theta)$ , is a measure of the excess probability, compared with a random distribution, of finding a galaxy at an angular separation  $\theta$  from another,  $P(\theta) = N[1 + w(\theta)]$ , where  $N$  is the surface density of galaxies. To calculate the angular two-point autocorrelation function we use the Landy & Szalay (1993) estimator,

$$w(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (4)$$

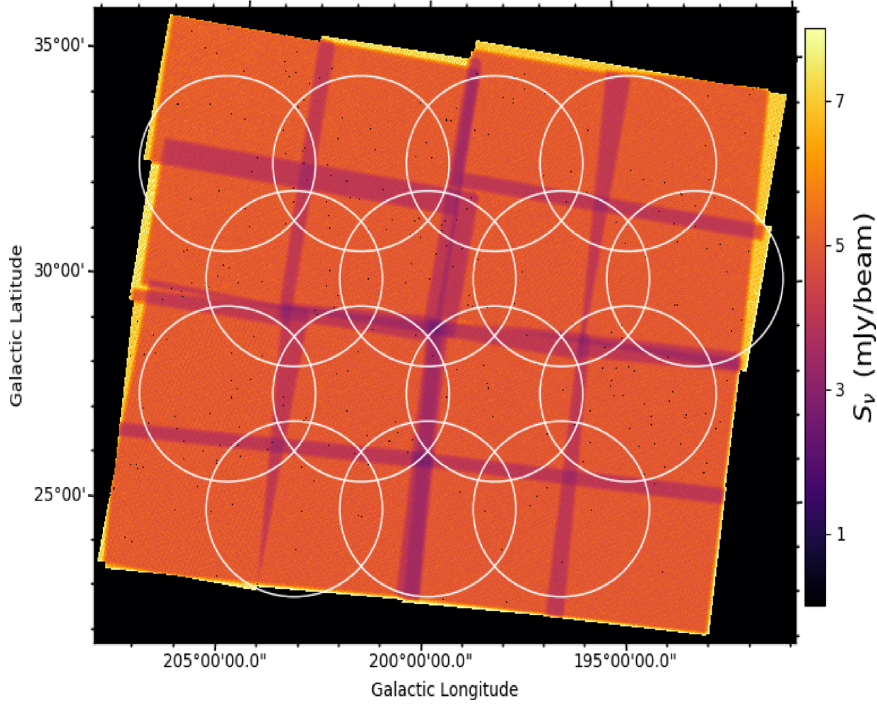
where  $DD(\theta)$  is the number of data–data pairs,  $DR(\theta)$  is the number of data–random pairs, and  $RR(\theta)$  is the number of random–random pairs, each at separation  $\theta$ . The  $DR(\theta)$  and  $RR(\theta)$  are normalized to have the same number of total pairs as  $DD(\theta)$ , so that given  $N_D$  sample sources and  $N_R$  random points, then  $DR(\theta) = [(N_D - 1)/2N_R]N_{DR}(\theta)$  and  $RR(\theta) = [N_D(N_D - 1)/N_R(N_R - 1)]N_{RR}(\theta)$ , where  $N_{DR}(\theta)$  and  $N_{RR}(\theta)$  are the original counts.

The error on  $w(\theta)$  at each angular separation, which is associated with the Landy & Szalay (1993) estimator, is defined as

$$\sigma_w^2 = \frac{(1+w)^2}{DD}. \quad (5)$$

However, these errors are considerably underestimated as the variance only accounts for the shot noise from the sample of the random points (which is folded in the measurement of  $w$ ) and the Poisson uncertainties of the  $DD$  counts. For a more accurate representation of the errors, we consider a ‘delete one jackknife’ resampling method (Norberg et al. 2009), which also accounts for systematic uncertainties due to the field-to-field variations.

In order to implement this approach, the area of each field was divided into  $N_{\text{sub}}$  circular sub-regions (as seen in Fig. 4), each with a radius of  $\sim 120$  arcmin. Similarly to González-Nuevo et al. (2017), we allowed for a 30 per cent overlap between sub-regions and about



**Figure 4.** The filtered variance map of the *NGP* field. The circular areas correspond to the 15 individual sub-regions that the field is divided, in order to perform the ‘delete one jackknife’ resampling method. The black holes in the map indicate the regions covered by extended sources that were masked out.

less than 10 per cent of each sub-region did not contain any sources (essentially falling outside of the image). These constraints were introduced in order to maximize the usable area and resulted in four independent sub-regions in each of the GAMA fields and 15 for the *NGP* field (as shown in Fig. 4).

Each jackknife sample is defined by discarding, in turn, each of the  $N_{\text{sub}}$  sub-regions into which each field has been split. The covariance matrix for the  $N_{\text{sub}}$  jackknife resamplings is then estimated using,

$$C_{i,j} = \frac{N_{\text{sub}} - 1}{N_{\text{sub}}} \sum_{k=1}^{N_{\text{sub}}} (w(\theta_i)^k - \bar{w}(\theta_i)) (w(\theta_j)^k - \bar{w}(\theta_j)), \quad (6)$$

where  $w(\theta_{i,j})^k$  are the auto-correlation functions measured in each jackknife realization and  $\bar{w}(\theta_{i,j})$  is the average auto-correlation function from all jackknife realizations.

We also corrected the measured correlation function for the integral constraint (IC; Roche & Eales 1999). Assuming that the true correlation function  $w(\theta)$  can be described as a power-law model,  $w_{\text{model}}(\theta) = A\theta^{-\gamma}$ , the observed one will be given by

$$w(\theta) = w_{\text{model}}(\theta) - IC, \quad (7)$$

where the IC can be numerically evaluated (Adelberger et al. 2005) using the RR counts from,

$$IC = \frac{\sum_i RR(\theta_i) w_{\text{model}}(\theta_i)}{\sum_i RR(\theta_i)}. \quad (8)$$

The best-fit values for the power-law model,  $w_{\text{model}}$ , from which the IC correction was evaluated, were determined by restricting the angular distance range to  $\theta > 4$  arcmin.

### 3.2 Construction of the random catalogues

We mentioned in Section 2 that local extended sources were removed from our *Herschel* catalogues, prior to calculating  $w(\theta)$ . Consequently, we need to account for the removal of these sources when constructing our random catalogues. This is accomplished by masking out the regions covered by extended source in order to avoid placing random sources in those regions. The masked regions were elliptical in the case where a custom aperture was created (using the minor semi-axis as well as the position angle; see Section 5.2 in Valiante et al. (2016) for details), otherwise they were circular (see Fig. 4).

The random catalogues were then created by drawing ten times more points, than in our real catalogue of sources, from a uniform distribution.

In practice, however, our noise maps are not completely uniform (as seen in Fig. 4), due to overlapping scanned regions. It is important that these non-uniformities not be imprinted on the measured clustering signal. We consider a similar approach to that adopted by Maddox et al. (2010), where we incorporated the noise (instrumental + confusion) information, while making sure we conserve the number counts of our real catalogues. This was achieved as follows: (a) a flux was chosen randomly using the cumulative probability distribution of fluxes of our real sources, (b) a random position was generated on the image, (c) the local noise was estimated as the quadratic sum of the instrumental noise in that pixel and the confusion noise (see Table 3; Valiante et al. 2016, for the GAMA fields), (d) we kept the source if its flux, perturbed by a Gaussian deviate equal to the total local noise estimate, was greater than  $4\sigma$ ; otherwise, the process was repeated starting from (a). The measurement of the angular correlation function using random catalogues generated this way, however, shows no significant difference compared to the simple uniform random catalogues. This is due to the fact that we apply a cut in flux-density at  $250 \mu\text{m}$ ,

which ensures that the fluxes of these sources are not significantly boosted.

### 3.3 The real-space correlation length

The simplest way to interpret the clustering strength of a galaxy population is to estimate its correlation length,  $r_0$ . We will determine this value for our SMG population at different redshift slices. We assume that the spatial correlation function,  $\xi(r)$ , is described by a power-law,

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \quad (9)$$

where  $r$  is the comoving distance between two points,  $r_0$  is the correlation length and  $\gamma$  is the power-law index.

The angular correlation function, parametrized as power-law model,  $w(\theta) = A_w \theta^{-\delta}$ , can be deprojected using the Limber approximation (Limber 1954) to yield a measurement on the correlation length over different redshift bins. This conversion is performed as follows:

$$r_0^\gamma(z) = A_w \left\{ \frac{H_0 H_\gamma \int_{z_i}^{z_j} N^2(z) (1+z)^{\gamma-(3+\epsilon)} \chi^{1-\gamma}(z) E(z) dz}{c \left( \int_{z_i}^{z_j} N(z) dz \right)^2} \right\}^{-1}, \quad (10)$$

where the value  $\epsilon = \gamma - 3$  is assumed, which corresponds to a constant clustering in comoving coordinates. In addition,

$$H_\gamma = \Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{\gamma-1}{2}\right) / \Gamma\left(\frac{\gamma}{2}\right) \quad (11)$$

with  $\Gamma(x)$  being the gamma function and  $\chi(z)$  is the radial comoving distance which can be computed from,

$$\chi(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}. \quad (12)$$

where  $H_0$  is the Hubble constant and  $E^2(z) = \Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}(1+z)^{3(1+w)}$ . Finally,  $N(z)$  is the number of sources per unit of redshift interval within a solid angle. The redshift distributions are determined differently for the analysis of our low- and high-redshift samples.

## 4 RESULTS

In this section, we present our results of the angular auto-correlation function,  $w(\theta)$ , for source samples selected with at least a  $4\sigma$  detection at  $250\text{-}\mu\text{m}$  ( $\sim 30\text{ mJy}$ ). The measurements were performed for evenly spaced logarithmic bins of angular separation in the range  $0.5' < \theta < 50'$ , where the lower limit comes from the FWHM of the SPIRE instrument's PSF at  $250\text{ }\mu\text{m}$  ( $0.3'$ ; Griffin et al. 2010). In addition, as discussed in Section 2, our sample of sources comprises different galaxy populations at low and high redshifts. Therefore, we will examine these two cases individually in the sections that follow.

### 4.1 Evolution of clustering with redshift for $z < 0.3$ SMGs

The clustering evolution of sub-mm sources selected at  $250\text{ }\mu\text{m}$ , in the low-redshift regime ( $z < 0.3$ ), has previously been studied by van Kampen et al. (2012). In their study, the authors used a sample of sources selected from the H-ATLAS Science Demonstration Phase

(SDP) field at a  $5\sigma$  significance level accounting for both instrumental and confusion noise. This resulted in a flux-density cut of  $S > 33\text{ mJy/beam}$  at  $250\text{ }\mu\text{m}$ . Additional selection criteria that were introduced in their study, specifically concerning the reliability of counterpart identification and the quality of optical spectroscopic redshifts, were identical to the ones introduced here.

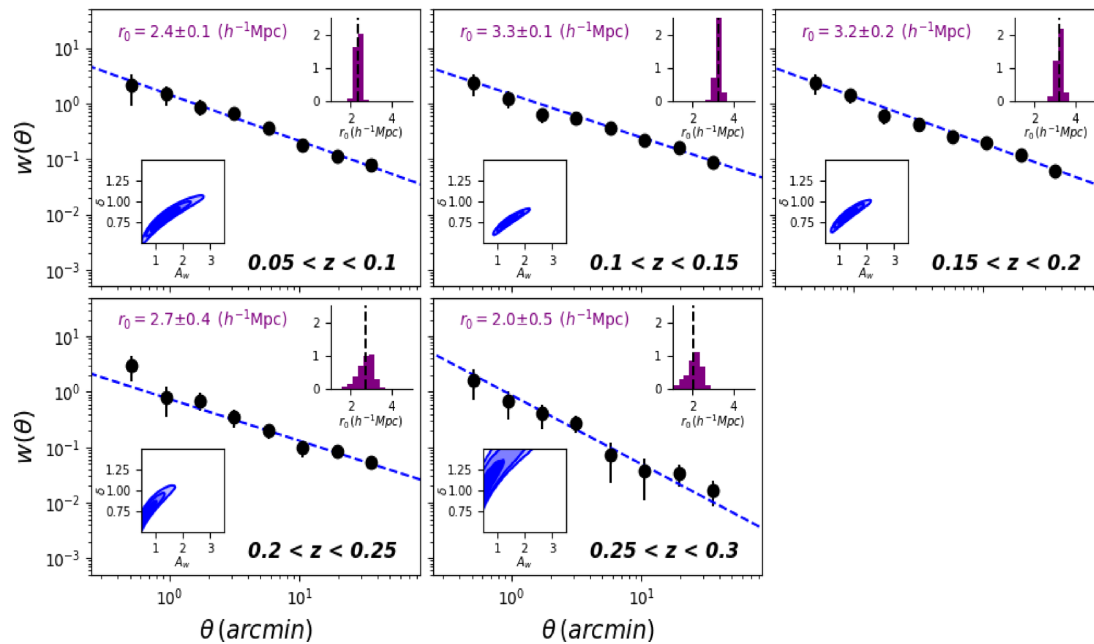
In this section, we repeat the analysis of van Kampen et al. (2012) for a sample of sources selected at  $250\text{ }\mu\text{m}$ , from the *GAMA + NGP* fields of the *H-ATLAS* survey. The *SGP* field was not used in this analysis since the optical counterpart identification analysis has not been performed as yet for this field. Similarly to van Kampen et al. (2012), we start our analysis at redshift  $z \sim 0.05$  where the redshift distribution starts to pick up (see Fig. 2) and end at  $z \sim 0.3$  where the completeness starts to drop sharply (see Bourne et al. 2016). We use a width size of  $\Delta z \sim 0.05$  which results in five individual redshift bins.

Our clustering measurements are shown in Fig. 5, where each panel corresponds to a different redshift bin indicated at the bottom right corner. The redshift distribution of sources for which this measurement corresponds to, are shown as the grey histograms in each panel of Fig. 2. One thing to note is that the clustering signal in the *NGP* field is slightly weaker compared to the *GAMA* fields, which is probably due to the lack of spectroscopic redshifts coverage.

In order to model the clustering signal, we used a two-parameter power-law model,  $w(\theta) = A_w \theta^{-\delta}$ , and performed an MCMC fitting method using the *emcee* package (Foreman-Mackey et al. 2013). The  $1$ ,  $1.5$ , and  $2\sigma$  contours of the fitted  $(A_w, \delta)$  parameter space are shown in the bottom left inset plot in each panel of Fig. 5. The resulting best-fit values for the parameters of our model in each redshift bin are presented in Table 2. These correspond to predictions that are shown as blue dashed lines in each panel of Fig. 5. The power-law slopes in all redshift bins are broadly consistent with that of normal star-forming galaxies,  $\delta \sim 0.8$  (Zehavi et al. 2011). Although 20–30 per cent of H-ATLAS galaxies have the red optical colours typical of traditional passive galaxies, Eales et al. (2018) show that these are still star-forming galaxies, although with a significant old stellar population. Therefore, it is not surprising that we find a clustering signal typical of star-forming galaxies.

The clustering length,  $r_0$ , in each redshift slice was calculated following a bootstrap method. We performed  $N \sim 1000$  realizations where in each realization we randomly draw a parameter value pair  $(A_w, \delta)$  without replacement from the output MCMC chain of our fitting method. In this way, we also account for the degeneracies in the parameters of our model. The resulting normalized histograms of  $r_0$  values from our bootstrap method are shown in the upper right corner inset plot of each panel in Fig. 5. The black vertical dashed line indicates the mean of the distribution, which was derived by fitting a Gaussian distribution to the histogram. This value corresponds to our measurement of the clustering length,  $r_0$  which is shown in the upper left corner of each panel, where the  $1\sigma$  uncertainty is taken as the standard deviation of the fitted Gaussian distribution. Our results are shown in the last column of Table 2 and seem to agree fairly well with van Kampen et al. (2012) measurements, even though their uncertainties were considerable.

We need to note here that we find a significant difference in the measurement of the correlation length,  $r_0$ , in the redshift bin  $0.15 < z < 0.2$  compared to van Kampen et al. (2012). The authors report in their study the existence of a structure around  $z \sim 0.164$ , which might be responsible for the excess clustering strength. Due to the small area used in their analysis, this structure dominates the clustering signal in this redshift bin. However, we are using a much



**Figure 5.** The angular correlation function of sub-mm galaxies for each redshift slice in the redshift range  $z < 0.3$ . The dashed lines show the best-fit two-parameter model,  $w(\theta) = A_w \theta^{-\delta}$ , where the best-fit values can be found in Table 2. The inset plot in the lower left corner in each panel corresponds to the 1, 1.5, and  $2\sigma$  contours in the fitted  $(A_w, \delta)$  parameter space. The inset plot in each panel shows the histogram of correlation length values which were derived from our bootstrap method. The black dashed vertical line in the inset plot of each panel, indicates the mean of the distribution which is also shown in the upper left corner in each panel.

**Table 2.** Results of Clustering Analysis for  $z < 0.3$  SMGs.

	$0.05 < z < 0.10$	$0.10 < z < 0.15$	$0.15 < z < 0.20$	$0.20 < z < 0.25$	$0.25 < z < 0.30$
$N_{\text{gal}}$	6225	8284	8385	7914	6744
$A_w$	$1.44^{+0.67}_{-0.50}$	$1.45^{+0.37}_{-0.30}$	$1.34^{+0.41}_{-0.36}$	$0.74^{+0.48}_{-0.32}$	$0.88^{+1.25}_{-0.57}$
$\delta$	$0.83^{+0.14}_{-0.14}$	$0.77^{+0.08}_{-0.08}$	$0.85^{+0.09}_{-0.10}$	$0.75^{+0.18}_{-0.19}$	$1.24^{+0.43}_{-0.40}$
$r_0$ ( $h^{-1}$ Mpc)	$2.4 \pm 0.1$	$3.3 \pm 0.1$	$3.2 \pm 0.2$	$2.7 \pm 0.3$	$2.0 \pm 0.5$

larger area in our study and this signal gets diluted, which is what probably causes this difference in the measurement of the clustering length.

## 4.2 Clustering of $z > 1$ SMGs

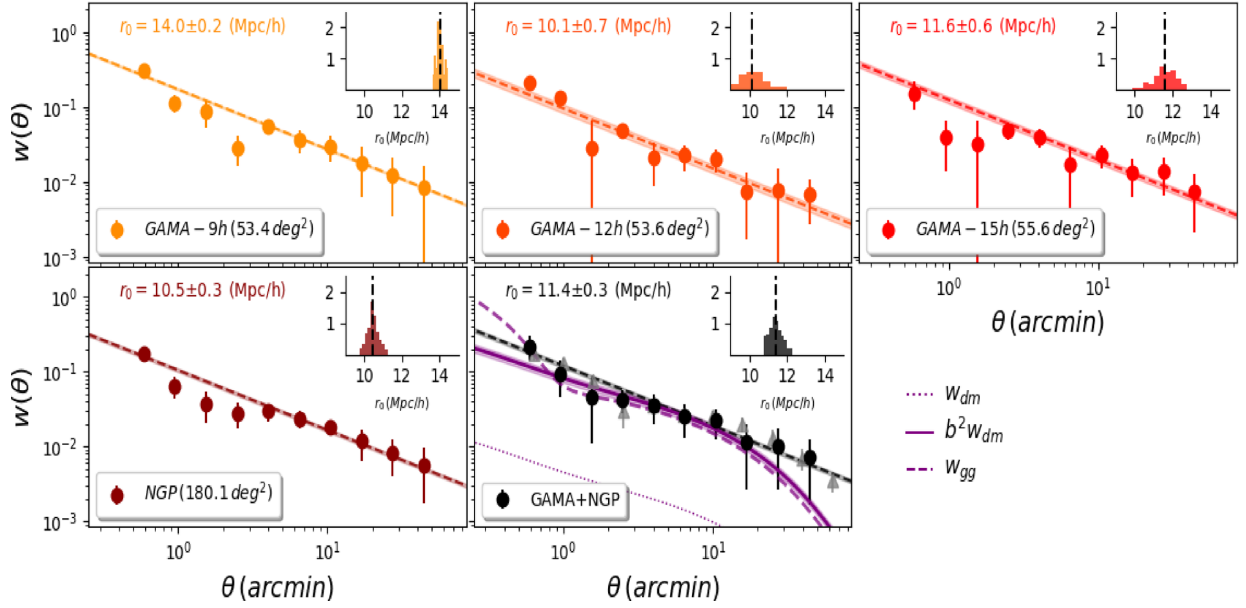
Previous studies on the clustering of SMGs have focused on the broad redshift range  $1 < z < 5$  (e.g. Webb et al. 2003; Blain et al. 2004; Scott et al. 2006; Weiß et al. 2009; Cooray et al. 2010; Maddox et al. 2010; Williams et al. 2011; Hickox et al. 2012; Mitchell-Wynne et al. 2012; Chen et al. 2016a,b; Wilkinson et al. 2017). As seen from the redshift distributions in Fig. 2, the majority of our sources lie in that redshift range with the peak of the redshift distribution occurring around  $z \sim 1.25$  (excluding the optically identified counterparts which typically reside at  $z < 0.5$ ). Therefore, in order to make a direct comparison with previous clustering measurements, we first perform our clustering analysis for sources within this redshift range.

The measured angular correlation functions of sub-mm sources, for each of the *H-ATLAS* fields under investigation, are shown in Fig. 6: top panels for the three equatorial *GAMA* fields and bottom-left for the *NGP* field. Our measurements were corrected by a factor of 1.25, as determined from our simulations in Appendix, for the

effect of filtering with a matched-filter to remove the background cirrus emission. The error bars were determined as  $\sigma_i \sim \sqrt{C_{ii}}$ .

In the bottom-middle panel of the same figure we show the measured angular correlation function by combining the three equatorial *GAMA* fields with the *NGP* field. In the same panel, we overlay the measurement from González-Nuevo et al. (2017; see also González-Nuevo et al. 2014) which was obtained using 250- $\mu\text{m}$ -selected sources in the redshift range  $z > 1.2$  from the *GAMA* fields as well as a small part of the *SGP* field. In this study, the authors used sources selected with at least a  $4\sigma$  detection at 250  $\mu\text{m}$ , which results in an  $S > 29$  mJy cut in flux density, and a  $3\sigma$  detection at 350  $\mu\text{m}$  in order to preferentially select high-redshift sources. The two measurements seem to agree fairly well across all angular scales. We can also compare our results with Cooray et al. (2010), who used the two widest fields from the Herschel Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2010), Lockman-SWIRE, and Spitzer First Look Survey (FLS). For this comparison, there seems to be a large disagreement, with the authors of this paper reporting a stronger clustering signal across all angular scales. This disagreement, which was first realized by comparing the results from Maddox et al. (2010), is alarming and it is not fully understood. We will discuss this further in Appendix A3, where we suggest that the removal of the background cirrus emission being one possibility for this difference.





**Figure 6.** The angular correlation function of sub-mm sources identified in the four *H-ATLAS* fields: *GAMA-9h* (top-left), *GAMA-12h* (top-middle), *GAMA-15h* (top-right), and *NGP* (bottom-left). The error bars are derived using a ‘delete one jackknife’ resampling method. The bottom-middle panel shows the measured angular correlation function of the combined *GAMA + NGP* fields. The measurements were corrected by a factor of 1.25, as determined from our simulations in Appendix, for the effect of filtering with a matched-filter. The dashed line shows the best-fit one-parameter power-law model with fixed slope,  $w(\theta) = A\theta^{-0.8}$ , where the  $1\sigma$  uncertainty is shown as the shaded region. The inset plot in each panel shows the histogram of correlation length values which were derived from our bootstrap method. The black dashed vertical line in the inset plot of each panel, indicates the mean of the distribution which is also shown in the upper left corner in each panel. The purple dotted curve shows the dark matter angular correlation function,  $w_{\text{dm}}$ . This has been scaled by the best-fit value of the linear bias factor,  $b$ , which is shown as the solid purple curve, with the  $1\sigma$  uncertainty shown as the shaded region. Finally, we show as the purple dashed curve the galaxy–galaxy angular correlation function,  $w_{\text{gg}}$ , that corresponds to the best-fit HOD model. In addition, we show the results from González-Nuevo et al. (2017) as grey triangles.

As a first step towards modelling the clustering signal, we use a one-parameter power-law model,  $w(\theta) = A_w\theta^{-0.8}$ , with a fixed slope. We perform the fitting for each individual field, as well as for the combined *GAMA + NGP*. The resulting best-fit values for the parameter of our model are summarized in Table 3. These correspond to the dashed coloured lines in each panel of Fig. 6, where the  $1\sigma$  uncertainty is shown as the shaded region.

The correlation length,  $r_0$ , was calculated following a bootstrap method in order to consider the uncertainty in the best-fit value of the power-law model. In each realization, we randomly sample the parameter  $A_w$  from a Gaussian distribution, centred at the best-fit value with a standard deviation equal to its error, and use equation (10) to calculate the correlation length. The resulting normalized histograms of  $r_0$  values, from our bootstrap method, are shown in the upper right corner of each panel in Fig. 6. The black vertical dashed line indicates the mean of the distribution, which was derived by fitting a Gaussian distribution to the histogram. This value corresponds to our measurement of the clustering length,  $r_0$ , which is shown in the upper left corner of each panel, where the  $1\sigma$  uncertainty is taken as the standard deviation of the fitted Gaussian distribution. The results are summarized in Table 3. We need to note here that the redshift distribution that enters the calculation of the correlation length,  $r_0$ , has been corrected for the effect of random photometric redshift errors, as described in Section 2.2.

We estimate the correlation length to be  $r_0 = 11.4 \pm 0.4 h^{-1} \text{Mpc}$ . The error in the measurement is relatively small, which is due to the assumption of a power-law model with a fixed slope thus reducing the uncertainties from introducing additional parameters. The measurement of the correlation length is in gen-

eral agreement with previous studies (Webb et al. 2003; Blain et al. 2004; Weiß et al. 2009; Williams et al. 2011; Hickox et al. 2012; Chen et al. 2016a,b). The measurement is also in agreement with Maddox et al. (2010), who used 250- $\mu\text{m}$ -selected sources from the *SDP* field of *H-ATLAS*, reporting a clustering length in the range  $r_0 \sim 7 - 11 h^{-1} \text{Mpc}$  when considering additional colour cuts to preferentially select high-redshift sub-mm sources. However, comparing our results with Wilkinson et al. (2017) we seem to find a larger clustering strength, even when compared with their sample of SMGs with radio-identified counterparts which typically comprise of more luminous SMGs.

#### 4.2.1 Halo Bias model

In order to convert the clustering strength to the inferred dark-matter halo mass,  $M_{\text{halo}}$ , we need to compute the galaxy bias,  $b$ . This quantity can be inferred by scaling the dark-matter angular correlation function,  $w_{\text{dm}}(\theta)$ , according to the following relation:

$$w(\theta) = b^2 w_{\text{dm}}(\theta). \quad (13)$$

In the above expression the dark matter angular correlation function,  $w_{\text{dm}}(\theta)$ , can be computed using the Limber’s equation which is used in order to convert a 3D power spectrum,  $P(k)$  into a projected angular correlation function from,

$$w(\theta) = \frac{1}{c} \int \left( \frac{dN}{dz} \right)^2 H(z) \int \frac{k}{2\pi} P(k, z) J_0 \left( \frac{k\theta}{\chi^{-1}(z)} \right) dk dz, \quad (14)$$

where  $J_0$  is the zero-th order Bessel function and  $dN/dz$  is the corrected redshift distribution as described in Section 2.2. In this case,

**Table 3.** Results of clustering analysis for  $z > 1$  SMGs.

Sample	Field	$N_{\text{gal}}$	$\langle z \rangle$	$A_w$	$r_0$ ( $h^{-1}$ Mpc)	$b$	$\log\left(\frac{M_{\text{halo}}}{h^{-1} M_{\odot}}\right)$	$\log\left(\frac{M_{\text{cen}}}{h^{-1} M_{\odot}}\right)$	$\log\left(\frac{M_{\text{sat}}}{h^{-1} M_{\odot}}\right)$
$1 < z < 2$	<i>GAMA + NGP</i>	85319	$1.75^{+0.55}_{-0.70}$	$0.13 \pm 0.01$	$11.4 \pm 0.4$	$4.3 \pm 0.3$	$13.2 \pm 0.1$	$13.00^{+0.14}_{-0.22}$	$14.17^{+0.65}_{-0.75}$
$2 < z < 3$	<i>GAMA + NGP</i>	55749	$1.47^{+0.38}_{-0.46}$	$0.09 \pm 0.01$	$8.1 \pm 0.5$	$3.2 \pm 0.1$	$13.1 \pm 0.1$	$12.91^{+0.11}_{-0.14}$	$14.78^{+0.50}_{-0.64}$
$3 < z < 5$	<i>GAMA + NGP</i>	25108	$2.21^{+0.42}_{-0.49}$	$0.11 \pm 0.02$	$8.5 \pm 0.8$	$4.5 \pm 0.4$	$12.9 \pm 0.1$	$12.50^{+0.28}_{-0.21}$	$12.32^{+1.71}_{-1.29}$
$1 < z < 5$	<i>GAMA + NGP</i>	4462	$2.93^{+0.45}_{-0.48}$	$0.31 \pm 0.16$	$13.9 \pm 3.9$	$9.0 \pm 2.5$	$13.2 \pm 0.2$	$12.88^{+0.35}_{-0.41}$	$12.93^{+1.71}_{-1.64}$

$P(k, z)$  is the non-linear dark matter power spectrum,  $P_{\text{NL}}(k, z)$  which was computed using the HALOMOD package (Murray et al. 2013; Murray et al. in prep). This package implements the HaloFit code (Smith et al. 2003) with improved parametrization provided by Takahashi et al. (2012).

Fitting our modelled angular correlation function, which is given by equation (13), we determined the galaxy bias. Our theoretical prediction is shown as the purple curves in bottom-middle panel of Fig. 6, where the  $1\sigma$  uncertainty is shown as the shaded region. The best-fit value of the galaxy bias for the combined *GAMA + NGP* is  $b = 4.26 \pm 0.27$  (see Table 3).

Finally, in order to infer the dark matter halo mass that corresponds to a specific value of the galaxy bias we need to assume a bias function,  $b(M, z)$ . The value of the halo mass,  $M_{\text{halo}}$ , will strongly depend on the assumed parametrization of the bias function. We opted to use the function introduced by Tinker et al. (2010),

$$b(v) = 1 - \frac{v^{\alpha}}{v^{\alpha} + \delta_c^{\alpha}} + Bv^b + Cv^c, \quad (15)$$

where  $B = 0.183$ ,  $b = 1.5$ ,  $c = 2.4$ ,  $\delta_c$  is the critical density for collapse, and  $v = \delta_c \sigma(M, z)$  is the ‘peak height’ in the linear density field, with  $\sigma(M, z)$  being the linear matter variance.

The inferred dark matter halo mass using the bias function, which was detailed above, is  $\log(M_{\text{halo}}) = 13.2 \pm 0.1$  (see Table 3) and was calculated at the median redshift  $\langle z \rangle \sim 1.75$ .

#### 4.2.2 Halo occupation distribution (HOD) model

We can see for Fig. 6 that our model adopting the halo bias formalism does not provide an accurate fit to the small angular scales. In an attempt to model this clustering signal, we make use of the halo model power spectrum,  $P(k)$ , which is written as the sum of two terms. The 1-halo term arises from interactions between galaxies within a single dark matter halo and dominates on small scales, while the 2-halo term arises from interactions of galaxies that belong to different halos and dominates on large scales (see Cooray & Sheth 2002). These terms are computed from,

$$P_{\text{gg}}^{1h}(k, z) = \int n(M, z) \frac{\langle N(N-1) | M \rangle}{\bar{N}_{\text{gal}}^2} y^2(k|M, z) dM \quad (16)$$

$$P_{\text{gg}}^{2h}(k, z) = P_{\text{lin}}(k, z) \times \left( \int n(M, z) b(M, z) \frac{\langle N | M \rangle}{\bar{N}_{\text{gal}}} y(k|M, z) dM \right)^2, \quad (17)$$

where  $n(M, z)$  is the halo mass function (?),  $y(k|M, z)$  is the normalized Fourier transform of the halo density profile,  $b(M, z)$  is the linear large-scale bias, and  $P_{\text{lin}}(k, z)$  is the linear matter power spectrum which is computed using the *CAMB* code (Lewis, Challinor & Lasenby 2002).

This formalism introduces the Halo Occupation Distribution (HOD) parametrization to the clustering signal arising from galaxy

populations. In this parameterization, the mean numbers of central and satellite galaxies in a halo of mass  $M$  are given by,

$$\langle N_{\text{cen}} | M \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log M - \log M_{\text{cen}}}{\sigma_{\log M}} \right) \right], \quad (18)$$

$$\langle N_{\text{sat}} | M \rangle = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\log M - \log M_{\text{cen}}}{\sigma_{\log M}} \right) \right] \left( \frac{M}{M_{\text{sat}}} \right)^{\alpha_{\text{sat}}}, \quad (19)$$

where  $\text{erf}(x)$  is the error function,  $M_{\text{cen}}$  is the minimum halo mass above which all halos host a central galaxy,  $\sigma_{\log M}$  is the width of the central galaxy mean occupation,  $M_{\text{sat}}$  is the mass scale at which one satellite galaxy per halo is found, in addition to the central galaxy, and  $\alpha_{\text{sat}}$  is the power-law slope of the satellite occupation number with halo mass.

The best-fit values of the parameters of our HOD model, which resulted from our MCMC analysis, are summarized in the first row of Table 3 for which we used flat priors for the parameters of our model within the range:  $12 < \log(M_{\text{cen}}/h^{-1} M_{\odot}) < 14$ ,  $10 < \log(M_{\text{sat}}/h^{-1} M_{\odot}) < 15$  with a fixed power-law slope for the satellite occupation number,  $\alpha_{\text{sat}} = 1.0$ , and width of the central galaxy mean occupation,  $\sigma_{\log M} = 0.3$ . Our theoretical prediction is shown as the purple dashed curve in bottom-middle panel of Fig. 8 and seems to provide a more accurate fit to the data. Using the Bayesian information criterion (BIC), we find:

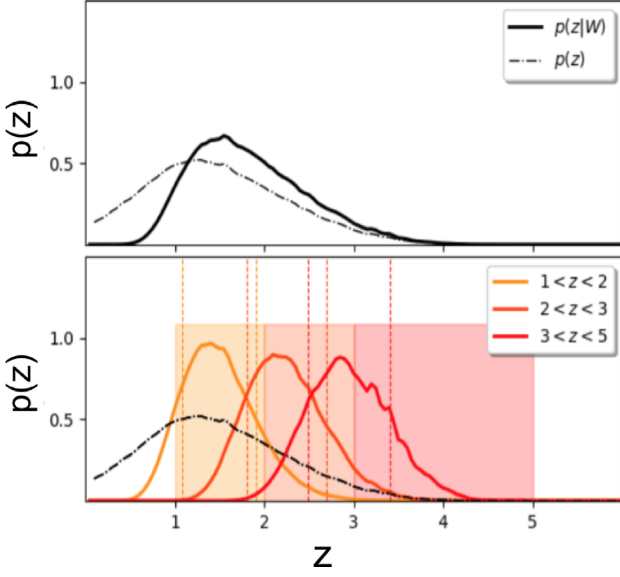
- (i)  $w(\theta) = A\theta^{-0.8}$  : 6.00 (1 parameter)
- (ii)  $w(\theta) = b^2 w_{\text{dm}}$  : 6.32 (1 parameter)
- (iii)  $w(\theta) = w_{\text{gg}}$  : 7.71 (2 parameter)

The BIC seem to prefer a power-law model,  $w(\theta) = A\theta^{-0.8}$ , which provides a better fit to the largest scales. If we were to exclude the last two data points, the BIC seem to prefer the  $w(\theta) = w_{\text{gg}}$  model.

In addition, we would like to note that we considered using the HOD parametrization of Geach et al. (2012), which is more appropriate for star formation rate selected samples. However, this HOD model has a lot more free parameters which were impossible to constrain given the errors in our measurements and the fact that we do not probe scales far in to the non-linear regime. The only parameter that is well constrained using this alternative parametrization is the minimum halo mass above which all halos host a central galaxy  $M_{\text{cen}}$ . This is the main parameter of interest for this work and its value was consistent between the two different parameterizations.

#### 4.3 Evolution of clustering with redshift for $z > 1$ SMGs

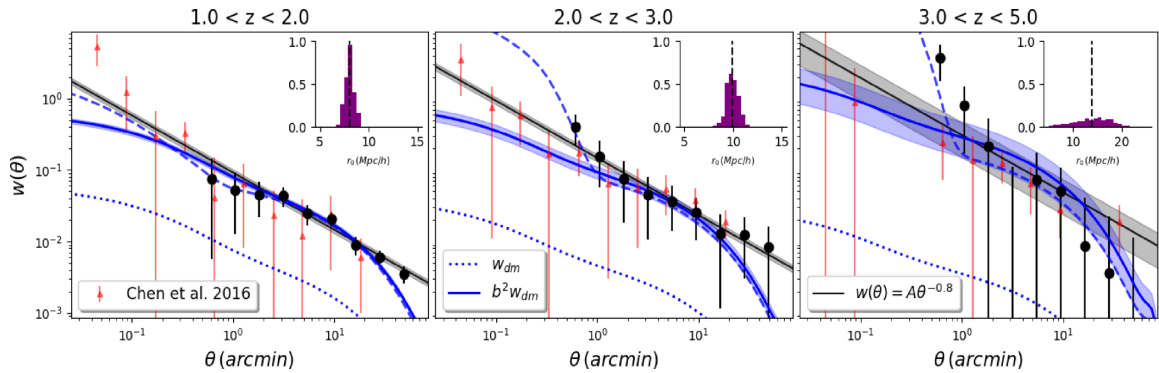
The large sample of high- $z$  sub-mm sources ( $z > 1$ ) in the combined *GAMA + NGP* fields allow us to investigate the redshift evolution of the clustering signal. To do that, we split our sample into three redshift bins,  $1 < z < 2$ ,  $2 < z < 3$ , and  $3 < z < 5$  similarly to Chen et al. (2016b). The redshift distributions,  $p(z|W)$ , after accounting for the effect of random photometric redshifts are shown in last three panels of Fig. 7 for the different redshift bins. We restricted



**Figure 7.** The estimated redshift distributions  $p(z|W)$  taking into account the window functions,  $W(z_{\text{ph}})$ , and the photometric redshift error function,  $p(z_{\text{ph}}|z)$ . The black dot-dashed line shows the initial redshift distribution,  $p(z)$  of our sources. The top panel shows the ‘corrected’ redshift distribution for sources in the redshift range  $1 < z < 5$ , while similarly in the bottom panel for the different redshift bins indicated at the right upper corner. The vertical solid lines correspond to the 50th percentile of the distribution, while the vertical dashed lines left and right of it correspond to the 16th and 84th percentiles, respectively. The shaded regions show the width of our window functions.

our analysis to three redshift bins in order to avoid excessive overlap between the corrected redshift distribution.

The resulting clustering measurements are shown in Fig. 8 for each redshift bin. The measurements were corrected by a factor of 1.25 as determined from our simulations in Appendix. In each panel, we also include the measurement from Chen et al. (2016b) as red triangles, which probe angular scales down to  $\sim 1'$ . The two measurements agree fairly well in the angular scales probed by *Herschel*. However, in the highest redshift bin, we find an excess



**Figure 8.** The angular correlation function of sub-mm galaxies for each redshift slice in the redshift range  $1 < z < 5$  (black circles). The measurements were corrected by a factor of 1.25, as determined from our simulations in Appendix. The black solid lines corresponds to our fitted power-law model with a fixed slope,  $w(\theta) = A\theta^{-0.8}$ , where the  $1\sigma$  uncertainty is shown as the black shaded region. The inset plot in each panel show the histogram of correlation length values which were derived from our bootstrap method. The black dashed vertical line in the inset plot of each panel, indicates the mean of the distribution. The blue dotted curve shows the dark matter angular correlation function,  $w_{\text{dm}}$ . This has been scaled by the best-fit value of the linear bias factor,  $b$ , which is shown as the solid blue curve, with the  $1\sigma$  uncertainty shown as the blue shaded region. Finally, we show as the blue dashed curve the galaxy–galaxy angular correlation function,  $w_{\text{gg}}$ , that corresponds to the best-fit HOD model. In addition, we also include the measurements from Chen et al. (2016) shown as red triangles.

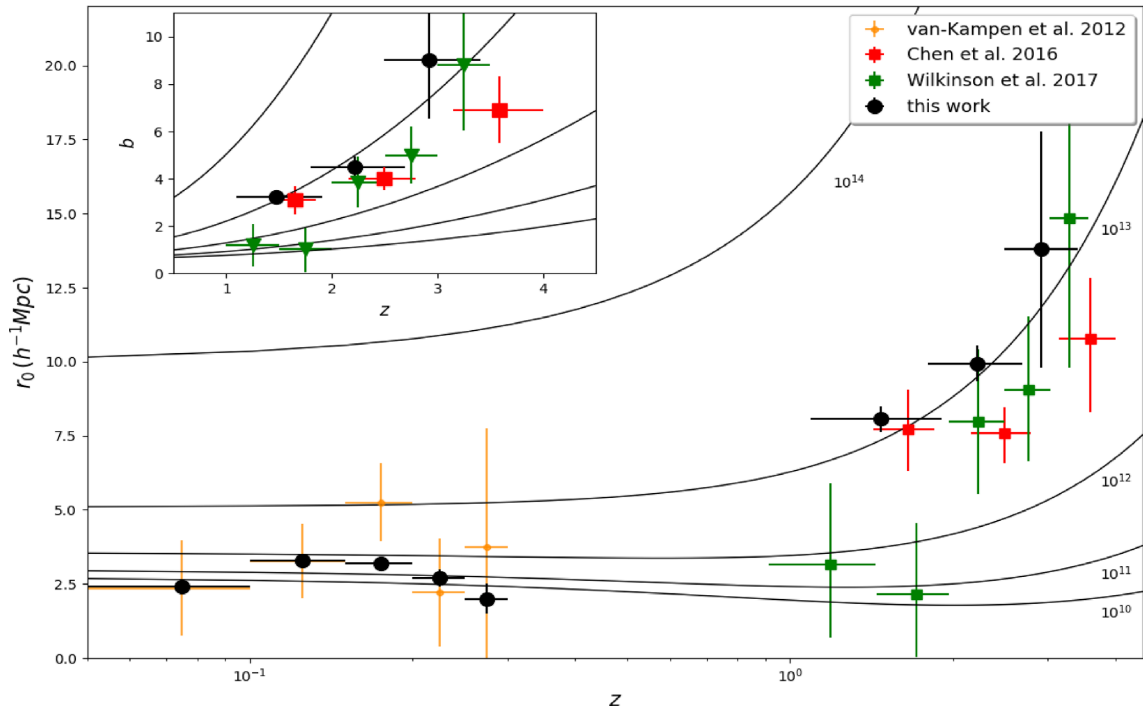
signal in the lowest probed angular bin compared to Chen et al. (2016b).

We fit a one-parameter power-law model with a fixed slope,  $w(\theta) = A_w\theta^{-0.8}$ , in order to model the angular correlation functions in each redshift bin. The resulting best-fit values for the parameter of our model, in each redshift bin, are shown in Table 3. These correspond to the black lines in each panel of Fig. 8, where the  $1\sigma$  uncertainty is shown as the grey-shaded region.

The correlation length,  $r_0$ , in each redshift slice was calculated following a bootstrap method, in order to consider the uncertainty in the best-fit value of the power-law model. In each realization, we randomly sample the parameter  $A_w$  from a Gaussian distribution, centred at the best-fit value with a standard deviation equal to its error, and use equation (10) to calculate the correlation length. The resulting normalized histograms of  $r_0$  values, from our bootstrap method, are shown in the upper right corner of each panel in Fig. 8. The black vertical dashed line indicates the mean of the distribution, which was derived by fitting a Gaussian distribution to the histogram. This value corresponds to our measurement of the clustering length,  $r_0$ , where the  $1\sigma$  uncertainty is taken as the standard deviation of the fitted Gaussian distribution. Our results are shown in Table 3 for each redshift slice.

Finally, we compute the bias parameters,  $b$ , for each redshift slice following the same methodology outlined in Section 4.2, using the corrected redshift distributions shown in Fig. 7 in order to compute the projected the dark matter angular correlation functions,  $w_{\text{dm}}(\theta)$ . In Fig. 8, we show  $w_{\text{dm}}(\theta)$  in each panel as the blue dashed lines. In the same Figure, the blue solid lines in each panel show the projected the dark matter angular correlation functions scaled by the best-fit value of the linear bias parameters, where the  $1\sigma$  uncertainty is shown as the blue-shaded region. Our results are shown in Table 3 for each redshift slice along with the halo masses,  $M_{\text{halo}}$ , that correspond to these bias measurements according to equation (15).

In the last two panels of Fig. 8 we see that the scaled dark matter angular correlation functions does not provide a satisfactory fit to the data, indicating the need of using an HOD model, similar to the analysis in the previous section. The results from our MCMC analysis are shown in Table 3 for which we used flat priors for the parameters of our model within the range:  $12 < \log(M_{\text{cen}}/h^{-1} M_{\odot}) < 14$



**Figure 9.** The evolution of the correlation length  $r_0$  with redshift for our sample of 250- $\mu\text{m}$ -selected sources with flux densities  $S > 30$  mJy (black points). We also show the clustering results from previous studies: Herschel-ATLAS science demonstration phase (SDP) field 250- $\mu\text{m}$ -selected sources at  $0.05 < z < 0.3$  (van-Kampen et al. 2012; yellow points), UKIRT Infrared Deep Sky Survey (UKIDSS) 850- $\mu\text{m}$ -selected SMGs at  $1 < z < 5$  (Chen et al. 2016b; red points), SCUBA-2 Cosmology Legacy Survey 850- $\mu\text{m}$ -selected SMG's at  $1 < z < 3.5$  (Wilkinson et al. 2017; green points). The black solid lines show the evolution of  $r_0$  with redshift for dark matter halos of different masses (in units of  $h^{-1} M_\odot$ ) using equation (20). The inset plot of the top left corner shows the evolution of the galaxy bias as a function of redshift, where the black solid lines show the theoretical predictions using equation (15) of the bias function from Tinker et al. (2010).

and  $10 < \log(M_{\text{sat}}/h^{-1} M_\odot) < 15$  with a fixed power-law slope for the satellite occupation number,  $\alpha_{\text{sat}} = 1.0$ , and width of the central galaxy mean occupation,  $\sigma_{\log M} = 0.3$ . We were not able to set good constraints on  $M_{\text{sat}}$ . The resulting errors depend strongly on the range of prior, we adopted for this parameter.

## 5 DISCUSSION

Our findings are summarized in Fig. 9 where we plot the evolution of the correlation length,  $r_0$ , as a function of redshift for our sample of 250- $\mu\text{m}$ -selected sources with flux densities  $S > 30$  mJy. The green points correspond to measurements from Wilkinson et al. (2017), while the red points correspond to measurements from Chen et al. (2016b). The black lines are the theoretical predictions for the evolution of the correlation length with redshift for different halo masses, which were estimated using the formalism of Peebles (1980). According to that formalism, the correlation length is related to the bias parameter, as

$$r_0 = 8 \left( \frac{\Delta_8^2}{C_\gamma} \right)^{1/\gamma} = 8 \left( \frac{b^2 \sigma_8^2 D^2}{C_\gamma} \right)^{1/\gamma}, \quad (20)$$

where  $\Delta_8$  is the clustering strength of haloes, more massive than the mass  $M$  at redshift  $z$  and is defined as  $\Delta_8(M, z) = b(M, z) \sigma_8 D(z)$ , with  $D(z)$  being the growth factor of linear fluctuations in the dark matter distribution which is computed from,

$$D(z) = \frac{5\Omega_m E(z)}{2} \int_z^\infty \frac{1+y}{E^3(y)} dy. \quad (21)$$

The factor  $C_\gamma$  is computed from,  $C_\gamma = 72/(3-\gamma)(4-\gamma)(6-\gamma)2^\gamma$ , where  $\gamma$  is the slope of power-law model which parametrizes the spatial correlation function and is taken to be  $\gamma = 1.8$  (since we assume the same power-law slope when computing the correlation length, see Section 3.3). The inset plot in Fig. 9 shows the evolution of the bias parameter as a function of redshift, where the green and red points correspond to the values found in the aforementioned studies. The black solid lines correspond to theoretical predictions using equation (15) of the bias function from Tinker et al. (2010).

At high redshifts, our results are in general agreement with previous studies for the evolution of clustering of SMGs. Above redshift of about  $\sim 2$ , however, we find our population of bright SMGs selected at 250  $\mu\text{m}$  with flux densities  $S_{250} > 30$  mJy exhibit larger clustering strengths ( $2\sigma$  discrepancy) compared to Chen et al. (2016b) where the authors studied a sample of faint SMGs selected at 850  $\mu\text{m}$  with flux densities  $S_{850} < 2$  mJy. This indicates that brighter SMGs cluster more strongly than their faint counterparts even at high redshifts, which is also supported by the fact that our results are more in agreement with Wilkinson et al. (2017) where the authors studied bright SMGs selected at 850  $\mu\text{m}$  with flux densities  $S_{850} > 2$  mJy. On the other hand, we find that SMGs in the redshift range  $1 < z < 2$  follow the same evolutionary track as those at higher redshifts, in contrast to the findings of Wilkinson et al. (2017) where the authors reported a downsizing effect ( $3\sigma$  discrepancy). However, this effect is not present in the analysis of Chen et al. (2016b). It is not straightforward to determine the cause for this difference, as there might be biases folded in the measurements associated with the selection of these SMGs. Finally our measurements of the halo masses are in very good agreement with studies

of the cross-correlation between the cosmic microwave background (CMB) lensing potential derived from Planck mission data and the high-redshift galaxies detected by the H-ATLAS survey (Bianchini et al. 2015, 2016).

The discrepancies of the aforementioned differences are at  $2-3\sigma$ . If we also were to consider that the errors on our measurements are slightly underestimated as determined from our simulations in Appendix, the agreement becomes even better. This suggests that we need to improve the accuracy of our measurements in order to confidently differentiate the clustering properties of faint and bright SMGs, as well as SMGs selected at different wavelengths. This improvement can come by obtaining more realistic photometric redshift measurements and potentially use the Atacama Large Millimeter/submillimeter Array (ALMA) to improve our counterpart identification techniques (Jin et al. 2018).

## 6 CONCLUSIONS

We measured the angular auto-correlation function of low- and high-redshift sub-mm sources identified in the *GAMA + NGP* fields of the *H-ATLAS*, which comprise the largest area extragalactic survey at sub-mm wavelengths. We selected a sample of sources detected at the  $4\sigma$  significance level ( $\sim 30$  mJy) at  $250\text{-}\mu\text{m}$ . Our main results are summarized as follows:

(i) We performed simulations of clustered sources and assessed our methodology of extracting sources from our  $250\text{-}\mu\text{m}$  *H-ATLAS* maps. We estimated the correction factor that needs to be applied to the measured angular correlation function due to the loss of clustering power from our method for removing Galactic cirrus emission. Our simulations and methodology for calculating this correction factor are described in Appendix.

(ii) First, we studied the evolution of clustering with redshift for our low-redshift ( $z < 0.3$ ) sample. We showed that SMGs in this redshift range exhibit clustering lengths of the order of  $\sim 2-3 h^{-1}$  Mpc, similar to normal galaxies selected at optical wavelengths. Our results agree with the findings of van Kampen et al. (2012), albeit with much improved errors on the measurement due to our larger sample.

(iii) We performed an auto-correlation analysis of SMG in the redshift range  $1 < z < 5$ , which is similar to the redshift range of many previous studies. We showed that SMGs are strongly clustered, finding a clustering length  $r_0 = 11.4 \pm 0.4 h^{-1}$  Mpc. We modelled the clustering signal by scaling the dark matter angular correlation function by the linear bias factor, finding a value of  $b = 4.26 \pm 0.27$  that corresponds to a dark matter halo mass of  $\log(M_{\text{halo}}) = 13.2 \pm 0.1 h^{-1} M_{\odot}$ .

(iv) In addition, we studied the evolution of clustering with redshift for our sample of high-redshift ( $z > 1$ ) sub-mm sources. We showed that SMGs occupy dark matter halos with masses of the order of  $10^{12.9}$ ,  $10^{12.5}$ , and  $10^{12.9} h^{-1} M_{\odot}$  at  $z = 1-2$ ,  $2-3$ , and  $3-5$ , respectively. We did not find a downsizing effect for SMG below redshift of about  $\sim 2$ , as reported in Wilkinson et al. (2017).

(v) Finally, we point out that galaxies selected at  $250 \mu\text{m}$  at high and low redshifts are not the same population. The star formation activity seems to be shifting from high mass halos at  $z > 1$  to less massive halos at  $z < 1$ , consistent with the downsizing effect reported in Magliocchetti et al. (2014).

## ACKNOWLEDGEMENTS

We thank the anonymous referee for the insightful comments and suggestions which helped to improve this manuscript. EV and SAE acknowledge funding from the UK Science and Technology Facilities Council consolidated grant ST/K000926/1. JGN acknowledges financial support from the I + D 2015 project AYA2015-65887-P (MINECO/FEDER) and from the Spanish MINECO for a ‘Ramon y Cajal’ fellowship (RYC-2013-13256). MN acknowledges financial support from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 707601. MS and SAE have received funding from the European Union Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under grant agreement No. 607254. LD and SM acknowledge support from the European Research Council (ERC) in the form of Consolidator Grant COSMIC-DUST (ERC-2014-CoG-647939, PI H L Gomez). GDZ acknowledges support from ASI/INAF agreement n. 2014-024-R.1 and from the ASI/Physics Department of the university of Roma–Tor Vergata agreement n.2016-24-H.0.

## REFERENCES

- Abazajian N. K., Adelman-McCarthy J. K., Agueros M. A. et al., 2009, *ApJS*, 182, 543
- Adelberger K. L., Steidel C. C., Pettini M., Shapley A. E. et al., 2005, *ApJ*, 619, 697
- Almeida C., Baugh C. M., Lacey C. G., 2011, *MNRAS*, 417, 2057
- Amblard A., Cooray A., Serra P., Altieri B., Arumugam V. et al., 2011, *Nature*, 470, 510
- Bakx T. J. L. C., Eales S. A., Negrello M., Smith M. W. L., Valiante E. et al., 2018, *MNRAS*, 473, 1751
- Bianchini F. et al., 2015, *ApJ*, 802, 64
- Bianchini F. et al., 2016, *ApJ*, 825, 24
- Blain A. W., Chapman S. C., Smail I., Ivison R., 2004, *ApJ*, 611, 725
- Blain A. W., Smail I., Ivison R. J., Kneib J.-P., Frayer D. T., 2002, *Phys. Rep.*, 369, 111
- Bourne N., Dunne L., Maddox S. J., Dye S., Furlanetto C. et al., 2016, *MNRAS*, 462, 1714
- Bourne N., Maddox S. J., Dunne L., Dye S., Eales S. et al., 2014, *MNRAS*, 444, 1884
- Budavári T., Connolly A. J., Szalay A. S., Szapudi I., Csabai I. et al., 2003, *ApJ*, 595, 59
- Bussmann R. S., Pérez-Fournon I., Amber S., Calanog J., Gurwell M. A. et al., 2013, *ApJ*, 779, 25
- Cai Z.-Y., Lapi A., Xia J.-Q., De Zotti G., Negrello M., Gruppioni C. et al., 2013, *ApJ*, 768, 21
- Casey C. M., Narayanan D., Cooray A., 2002, *Phys. Rev.*, 541, 45
- Chapin E. L., Chapman S. C., Coppin K. E., Devlin M. J. et al., 2011, *MNRAS*, 411, 505
- Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, *ApJ*, 622, 772
- Chen C.-C., Smail I., Ivison R. J., Arumugam V., Almaini O. et al., 2016a, *ApJ*, 820, 82
- Chen C.-C., Smail I., Swinbank A. M., Simpson J. M. et al., 2016b, *ApJ*, 831, 91
- Colless M., Dalton G., Maddox S., Sutherland W., Norberg P., Cole S. et al., 2001, *MNRAS*, 328, 1039
- Cooray A., Amblard A., Wang L., Arumugam V., Auld R. et al., 2010, *A&A*, 518, 22
- Cooray A., Sheth R., 2002, *Phys. Rev.*, 372, 1
- Coppin K., Chapin E. L., Mortier A. M. J., Scott S. E. et al., 2006, *MNRAS*, 372, 1621
- Cowley W. I., Lacey C. G., Baugh C. M., Cole S., Wilkinson A., 2017, *MNRAS*, 469, 3396
- Cox P., Krips M., Neri R., Omont A., Güsten R., Menten K. M. et al., 2011, *ApJ*, 740, 63

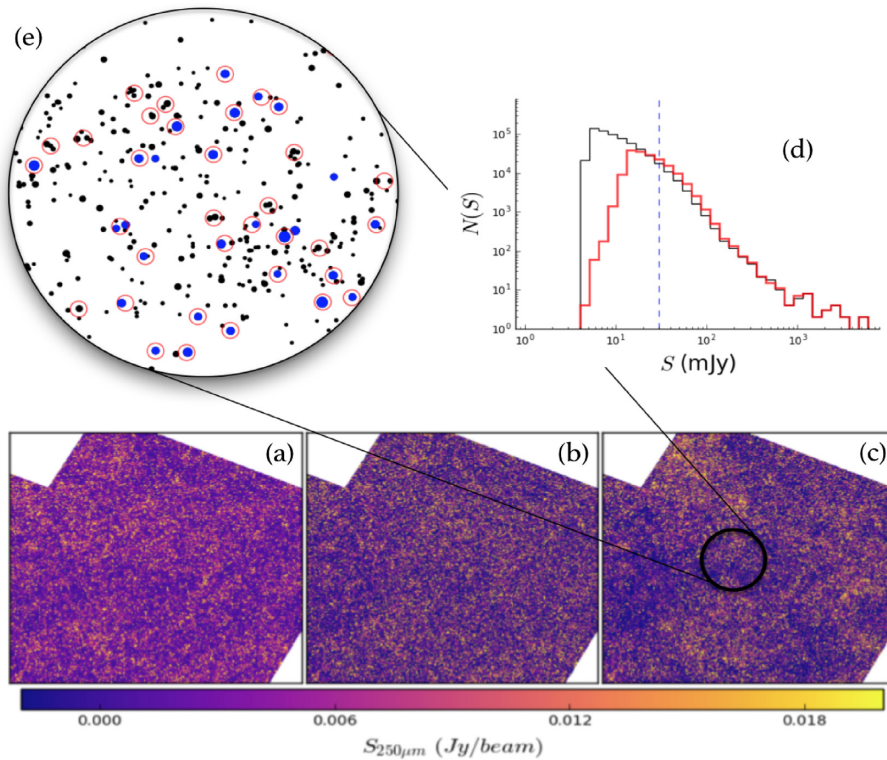
- Daddi E., et al., 2005, *ApJ*, 631, L13
- Davé R., Finlator K., Oppenheimer B. D., Fardal et al., 2010, *MNRAS*, 404, 1355
- de Jong J. T. A., Verdoes K., G. A., Boxhoorn D. R., Buddelmeijer H., Capaccioli M. et al., 2015, *A&A*, 582, A62
- Driver S. P., et al., 2011, *MNRAS*, 413, 971
- Driver S. P., GAMA Team, Baldry I. K., Bamford S., Bland-Hawthorn J., Bridges T. et al., 2009, *Astron. Geophys.*, 50, 12
- Eales S., Dunne L., Clements D., Cooray A., De Zotti G., Dye S. et al., 2010, *PASP*, 122, 499
- Eales S., Smith D., Bourne N., Loveday J., Rowlands K. et al., 2018, *MNRAS*, 473, 3507
- Edge A., Sutherland W., Kuijken K., Driver S., McMahon R., Eales S., Emerson J. P., 2013, *MNRAS*, 409, 109
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Frayser D. T., Harris A. I., Baker A. J., Ivison R. J., Smail I., Negrello M. et al., 2011, *ApJ*, 726, 22
- Furlanetto C., Dye S., Bourne N., Maddox S., Dunne L. et al., 2018, *MNRAS*, 476, 961
- Geach J. E., Chapin E. L., Coppin K. E. K., Dunlop J. S., Halpern M., Smail I. et al., 2013, *MNRAS*, 432, 53
- Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I. et al., 2012, *MNRAS*, 426, 679
- George R. D., Ivison R. J., Hopwood R., Riechers D. A., Bussmann R. S., Cox P. et al., 2013, *MNRAS*, 436, 99
- González-Nuevo J., Lapi A., Bonavera L., Danese L., de Zotti G., Negrello M. et al., 2017, *JCAP*, 10, 024
- González-Nuevo J., Lapi A., Negrello M., Danese L., De Zotti G. et al., 2014, *MNRAS*, 442, 2680
- González-Nuevo J., Toffolatti L., Argueso F., 2005, *ApJ*, 621, 1
- Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, *ApJ*, 600, 580
- Griffin M. J. et al., 2010, *A&A*, 518, L3
- Harris A. I., Baker A. J., Frayer D. T., Smail I., Swinbank A. M., Riechers D. A. et al., 2012, *ApJ*, 752, 152
- Hickox R. C., Wardlow J. L., Smail I., Myers A. D., Alexander D. M. et al., 2012, *MNRAS*, 421, 284
- Ivison R. J., Lewis A. J. R., Weiss A., Arumugam V., Simpson J. M., Holland W. S. et al., 2016, *ApJ*, 832, 781
- Jin S., Daddi E., Liu D., Smolčić V., Schinnerer E. et al., 2018, *ApJ*, 864, 56
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lapi A., González-Nuevo J., Fan L., Bressan A., De Zotti G., Danese L. et al., 2011, *ApJ*, 742, 24
- Lawrence A., Warren S. J., Almaini O., Edge A. C., Hambly N. C. et al., 2007, *MNRAS*, 379, 1599
- Lewis A., Challinor A., Lasenby A., 2000, *Phys. Rev. D*, 538, 473
- Limber D. N., 1954, *ApJ*, 119, 655
- Lonsdale C. J., Farrah D., Smith H. E., 2006, *We would argue it would be better to write*. Praxis Publishing Ltd., Chichester, UK, p. 285
- Lupu R. E., Scott K. S., Aguirre J. E., Aretxaga I., Auld R., Barton E. et al., 2012, *ApJ*, 757, 135
- Maddox S. J., Dunne L., Rigby E., Eales S., Cooray A. et al., 2010, *MNRAS*, 404, 518, 11
- Maddox S. J., Valiante E., Cigan P., Dunne L., Eales S. et al., 2018, *ApJS*, 236, 30
- Magliocchetti M., Lapi A., Negrello M., De Zotti G., Danese L., 2014, *MNRAS*, 437, 2263
- Martin D. C., Fanson J., Schiminovich D., Morrissey P., Friedman P. G. et al., 2005, *ApJ*, 619, 1
- Mitchell-Wynne K., Cooray A., Gong Y., Bethermin M., Bock J. et al., 2012, *ApJ*, 753, 23
- Murray S. G., Power C., Robotham A. S. G., 2013, *Astron. Comput.*, 3, 23
- Negrello M., Hopwood R., De Zotti G., Cooray A., Verma A., Bock J., Frayer D. T. et al., 2010, *Science*, 330, 800
- Norberg P., Baugh C. M., Gaztanaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Oliver S. J., Wang L., Smith A. J., Altieri B., Amblard A. et al., 2010, *A&A*, 518, 21
- Pearson E. A., Eales S., Dunne L., González-Nuevo J., Maddox S. et al., 2013, *MNRAS*, 435, 2753
- Peebles P. J. E., 1980, *The Large Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
- Pilbratt G. L., Riedinger J. R., Passvogel T., Crone G., Doyle D. et al., 2010, *A&A*, 518, L1
- Planck Collaboration XIII et al., 2015, *A&A*, 594, 63
- Planck Collaboration XXX et al., 2014, *A&A*, 571, 30
- Poglitsch A., Waelkens C., Geis N., Feuchtgruber H., Vandenbussche B. et al., 2010, *A&A*, 518, L2
- Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
- Scott S. E., Dunlop J. S., Serjeant S., 2006, *MNRAS*, 370, 1057
- Simpson J. M., Swinbank A. M., Smail I. et al., 2014, *ApJ*, 788, 125
- Smail I., Ivison R. J., Blain A. W., 1997, *ApJ*, 490, L5
- Smith M. W. L., Ibar E., Maddox S. J., Valiante E., Dunne L., Eales S. et al., 2017, *ApJS*, 233, 26
- Smith R. E., et al., 2003, *MNRAS*, 341, 1311
- Swinbank A. M., Simpson J. M., Smail I., Harrison C. M., Hodge J. A. et al., 2014, *MNRAS*, 438, 1267
- Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *ApJ*, 761, 521
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A. et al., 2010, *ApJ*, 724, 878
- Valiante E., Smith M. W. L., Eales S., Maddox S. J. et al., 2016, *MNRAS*, 462, 3146
- van Kampen E., Smith D. J. B., Maddox S., Hopkins A. M. et al., 2012, *MNRAS*, 426, 3455
- Viero M. P., Ade P. A. R., Bock J. J. et al., 2009, *ApJ*, 707, 1766
- Viero M. P., Wang L., Zemcov M., Addison G., Amblard A., Arumugam V. et al., 2013, *ApJ*, 772, 77
- Webb T. M., Eales S., Foucaud S., Lilly S. J., McCracken H., Adelberger K. et al., 2003, *ApJ*, 582, 6
- Weiß A., Kovács A., Coppin K., Greve T. R., Walter F., Smail I. et al., 2009, *ApJ*, 707, 1201
- Wilkinson A., Almaini O., Chen C., Smail I. et al., 2017, *MNRAS*, 464, 1380
- Williams C. C., Giavalisco M., Porciani C., Yun M. S., Pope A. et al., 2011, *ApJ*, 733, 92
- Wright E. L., Eisenhardt P. R. M., Mainzer A. K., Ressler M. E., Cutri R. M. et al., 2010, *AJ*, 140, 1868
- Zehavi I., Zheng Z., Weinberg D. H., Blanton M. R., Bahcall N. A. et al., 2011, *ApJ*, 736, 59

## APPENDIX: CORRECTION TO THE ANGULAR CORRELATION FUNCTION

One concern that was pointed out by Maddox et al. (2010) is the effect of removing the Galactic cirrus emission, on the measured clustering signal of submillimeter sources. This emission was removed, using the *Nebuliser* algorithm, before trying to detect extragalactic sources with the MADX (Maddox et al. in preparation) algorithm on the *Herschel* images (as discussed in Section 2). In addition to the background cirrus emission, *Nebuliser* can also remove any large-scale background produced by clustered faint sources that cannot be individually resolved and can ultimately affect the measured clustering signal (Valiante et al. 2016; Maddox et al. 2018).

### A1 Generating realistic maps

In order to quantify this effect, we create catalogues of clustered source positions on the sky region covered by the *GAMA-15h* field



**Figure A1.** The different steps illustrating the procedure followed to produce realistic maps of clustered point sources for the *GAMA-15h* field (see text for more details). Panel (a) shows the convolved map of our input point sources, panel (b) the noise was added, and in panel (c) the background emission from the *Nebulizer* using a pixel scale  $N_{\text{pix},b} = 30$  arcmin was added and the mean of the map was set to zero. Applying our source extraction algorithm (MADX) using a pixel scale  $N_{\text{pix},f} = 30$  arcmin we obtain the number counts as a function of flux density shown as the red histogram in panel (d), with the input number counts shown as the black histogram, respectively. Finally, panel (e) shows a zoomed in region of panel (c) where the black and blue points correspond to the input sources, with the blue being the sources with flux densities  $>30$  mJy. The red circles on the other hand correspond to our extracted sources with flux densities  $>30$  mJy.

according to an input power-law power spectrum  $P_{\text{corr}}(k)$ .<sup>2</sup> For these catalogues, we follow the methodology of González-Nuevo, Tofolatti & Argüeso (2005) (G05). We adopt the model by Cai et al. (2013) for the number count distribution of these sources assuming a minimum flux limit of  $S_{\text{min}} = 5$  mJy, unless otherwise stated (we refer the reader to the G05 paper for more details).

These catalogues were then used to produce realistic maps of the *GAMA-15h* field. We start by creating a high-resolution map ( $1''/\text{pixel}$ ) on top of which our simulated sources are laid down. This map is convolved with a PSF (the measured FWHM of the azimuthally average circular PSF is  $17.8$  at  $250 \mu\text{m}$ ) and consequently rescaled to the real *Herschel* pixel size at  $250$  microns ( $6''/\text{pixel}$ ). In panel (a) of Fig. A1, a small patch of this map is shown. Subsequently, the noise map was added to the image. This map was created by assigning each pixel a value drawn from a Gaussian distribution with zero mean and standard deviation equal to the corresponding pixel value of the raw instrumental noise map of the *GAMA-15h* field (see Valiante et al. 2016). In panel (b), the same patch of the map is shown after the noise was added. Finally, we included the Galactic background cirrus emission that was estimated by *Nebulizer* on our real *Herschel* image of the *GAMA-15h* field, using a filtering scale of  $N_{\text{pix}}$  pixels. We will use the notation  $N_{\text{pix},b}$  when we refer to the background that was added to our sim-

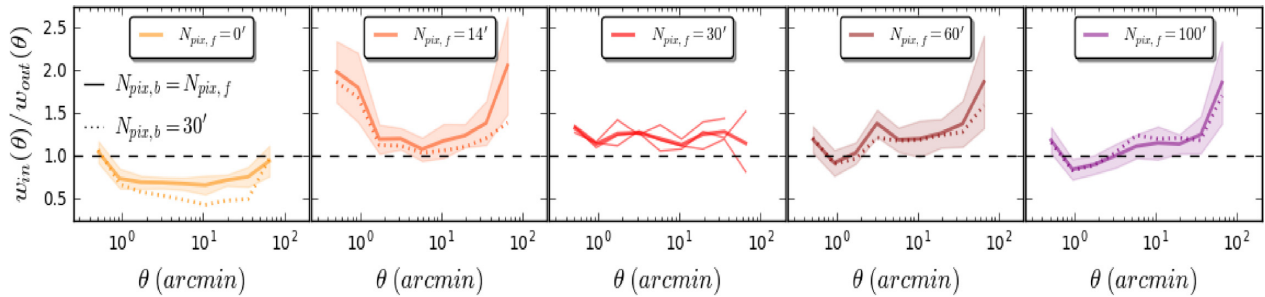
ulated map. The last step of the map-making procedure is to set the mean of the map to zero. In panel (c) of Fig. A1, we show the case with  $N_{\text{pix},b} = 30$ .

Once our simulated map is created, we then execute our source extraction algorithm after we filter our map with *Nebulizer* with a scale in pixels equal to  $N_{\text{pix},f}$ . We will use the notation  $N_{\text{pix},f}$  when we refer to the filtering of our simulated maps prior to source extraction. In Fig. A1, we have used  $N_{\text{pix},b} = N_{\text{pix},f} = 30$ . In the right top corner, we show the number counts of our input catalogue of simulated point sources as the black histogram along with the number counts of sources extracted from our simulated map as the red histogram. For a visual aid in the left top corner, we show the input and output source catalogues in a smaller region of the map. The black and blue points correspond to our input sources (the size of these points is indicative of their flux densities), where blue points are sources with flux densities  $S \geq 30$  mJy. Equivalently, the open red circles are sources in the output catalogue with flux densities  $S \geq 30$  mJy. As shown from this figure, as well as from the histograms, there is a clear flux boosting effect taking place due to the confusion of low flux density sources.

## A2 Determining the correction factor

In order to now quantify the effect in question, we compute the ratio of input to the output angular correlation function for sources with  $S > 30$  mJy, where the results are shown in Fig. A2. The different panels correspond to the different filtering pixel scales used in the

<sup>2</sup>In the case of a power-law power spectrum  $P(k) \propto k^{-1.2}$ , the angular correlation function  $w(\theta) \propto \theta^{-0.8}$



**Figure A2.** The ratio of the measured input to the output (after filtering) angular correlation function. The different panels, going from left to right, correspond to a filtering scale of  $N_{\text{pix},f} = 0, 14, 30, 60, 100$ . In all panels, the dotted lines correspond to a added background emission that was produced by the *Nebuliser* algorithm using a filtering scale  $N_{\text{pix},b} = 30$ . The continuous lines, on the other hand, correspond to an added background emission using  $N_{\text{pix},b} = N_{\text{pix},f}$ , where the shaded regions are the  $1\sigma$  Poisson uncertainties to the measurements. In the middle panel, corresponding to a filtering scale  $N_{\text{pix},b} = 30$ , we performed this procedure for three realizations of our simulations. These are shown as the thin faded continuous lines where the thick line shows their average.

analysis  $N_{\text{pix},f} = 0, 14, 30, 60, 100$  starting from left to right, respectively. The continuous lines correspond to a background with  $N_{\text{pix},b} = N_{\text{pix},f}$  pixel scale, while for the dotted lines we used the same background for all panels  $N_{\text{pix},b} = 30$ . The case of  $N_{\text{pix},b} = 0$  means that no background emission was added.

In the case where no filtering is being applied prior to the source extraction (first panel of Fig. A2), the resulting clustering strength of our extracted sources is enhanced in all angular scales, whether or not a background is added. This results from the combination of two effects: (i) The low flux density sources of our input catalogue, which constitute the unresolved background, are also clustered. (ii) The added background cirrus emission is also contributing to the enhancement of the clustering signal. In the remaining cases, we filter our simulated map prior to the extraction of sources using the different filtering pixel scales. The aim of filtering the map is to remove the cirrus emission only, but as a consequence some real clustering is also removed.

We determine that the case of  $N_{\text{pix},f} = 30$ , which is in fact the one used for extracting sources from the H-ATLAS maps, performs best. From the ratio of input to output angular correlation functions, we work out that a 25 per cent correction needs to be applied, at all angular scales, to the clustering measurements. In fact, for this case only, we performed this procedure for three realizations of our simulation, which are shown as the thin faded red lines in the middle

panel of Fig. A2 where the thicker red line is their average. We use this correction factor in our analysis, only when specifically stated. In addition, we see from the different realizations that there is non-negligible scatter, which will contribute to the final error budget of the measured clustering properties presented in Fig. 9. As a result, the quoted errors in Fig. 9 are slightly underestimated.

### A3 Maddox et al. (2010) versus Cooray et al. (2010)

We pointed out in Section 4.2 that our measurement of the angular correlation function is significantly lower compared to the measurement of Cooray et al. (2010). In their case, no filtering was applied prior to source extraction, even though there should be cirrus contamination to some degree in the HERMES fields. This means that their measurement would fall into the case of  $N_{\text{pix},b} = 30$  and  $N_{\text{pix},f} = 0$ . On the other hand, our measurement would correspond to  $N_{\text{pix},b} = 30$  and  $N_{\text{pix},f} = 30$ . Our simulations in Fig. A2 suggest that there is more than a 50 per cent difference between these two cases. This could be the reason of the disagreement between the two measurements, which was first pointed out by Maddox et al. (2010).

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.