Blending bias impacts the host halo masses derived from a cross-correlation analysis of bright submillimetre galaxies

Cowley, William I.; Lacey, Cedric G.; Baugh, Carlton M.; Cole, Shaun; Wilkinson, Aaron

Blending bias impacts the host halo masses derived from a cross-correlation analysis of bright submillimetre galaxies

William I. Cowley,[1,2]* Cedric G. Lacey,[1] Carlton M. Baugh,[1] Shaun Cole[1] and Aaron Wilkinson[3]

[1]*Department of Physics, Institute for Computational Cosmology, University of Durham, South Road, Durham DH1 3LE, UK*
[2]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[3]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

ABSTRACT

Placing bright submillimetre galaxies (SMGs) within the broader context of galaxy formation and evolution requires accurate measurements of their clustering, which can constrain the masses of their host dark matter haloes. Recent work has shown that the clustering measurements of these galaxies may be affected by a 'blending bias', which results in the angular correlation function of the sources extracted from single-dish imaging surveys being boosted relative to that of the underlying galaxies. This is due to confusion introduced by the coarse angular resolution of the single-dish telescope and could lead to the inferred halo masses being significantly overestimated. We investigate the extent to which this bias affects the measurement of the correlation function of SMGs when it is derived via a cross-correlation with a more abundant galaxy population. We find that the blending bias is essentially the same as in the autocorrelation case and conclude that the best way to reduce its effects is to calculate the angular correlation function using SMGs in narrow redshift bins. Blending bias causes the inferred host halo masses of the SMGs to be overestimated by a factor of $\sim$6 when a redshift interval of $\delta z = 3$ is used. However, this reduces to a factor of $\sim$2 for $\delta z = 0.5$. The broadening of photometric redshift probability distributions with increasing redshift can therefore impart a mild halo 'downsizing' effect on to the inferred host halo masses, though this trend is not as strong as seen in recent observational studies.

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

1 INTRODUCTION

Submillimetre galaxies (SMGs; e.g. Blain et al. 2002; Casey, Narayanan & Cooray 2014) are thought to be amongst the most rapidly star-forming objects in the Universe. They are detected at wavelengths that probe the re-emission of radiation by cold interstellar dust. Assuming that the initial radiation field is due to star formation,[1] the extreme luminosity of this dust leads to prodigious inferred star formation rates of $\gtrsim 100\ \mathrm{M_\odot\ yr^{-1}}$ (e.g. Smail et al. 2002; Swinbank et al. 2014). The shape of a galaxy's spectral energy distribution (SED) at these wavelengths (the Rayleigh–Jeans tail of the dust emission) approximates a power law that decreases

with increasing wavelength, meaning that it is subject to a negative *k*-correction (e.g. Blain et al. 2002). For a fixed bolometric luminosity and observer-frame wavelength, shifting the galaxy to higher redshifts means that the SED is sampled at a shorter rest-frame wavelength, where it is intrinsically brighter. This largely cancels out the effect of dimming due to the increasing luminosity distance, meaning that the observed flux of an SMG is roughly constant over $z \sim 1$–10. Thus, SMGs provide a window into (dust-obscured) star formation at high redshift, commonly being found at $z \sim 1$–3 (e.g. Chapman et al. 2005; Simpson et al. 2014).

Placing these extreme galaxies into a consistent evolutionary picture remains challenging. It is not clear what physical mechanisms trigger and quench the star formation rates inferred from observations, and their subsequent evolution to the present day is poorly understood. Simple arguments that make assumptions about the duration of the extreme star formation event and the subsequent evolution of the stellar populations of the SMGs have been used to suggest that they could evolve into massive local elliptical galaxies

---

* E-mail: cowley@astro.rug.nl
[1] Studies that have investigated the X-ray properties of SMGs suggest that their bolometric luminosity is dominated by emission from star formation rather than an active galactic nucleus (AGN; e.g. Alexander et al. 2005).

with most of their stellar mass being assembled during the 'SMG phase' (e.g. Swinbank et al. 2006; Simpson et al. 2014), though see González et al. (2011) for a contrasting view in which this phase accounts for little of the present-day stellar mass in their descendants.

A strong constraint on the evolution of a galaxy population can come from observational measurements of its clustering, which provides information regarding the masses of the dark matter haloes the galaxies inhabit. Growth of structure arguments based on results from *N*-body simulations can then be used to infer the distribution of present-day host halo mass of the galaxies' descendants (e.g. Fakhouri, Ma & Boylan-Kolchin 2010), which can then be compared to the halo masses inferred from the observed clustering of local galaxy populations. However, the spread in the host halo masses of SMG descendants could be significant ($\sim$2 dex; Cowley et al. 2016) due to the hierarchical growth of structure.

Measuring the clustering of far infrared (FIR) or submillimetre-selected galaxies has proven challenging. Some studies have failed to produce significant detections of clustering (e.g. Scott et al. 2002; Webb et al. 2003; Coppin et al. 2006; Williams et al. 2011), or the results derived from similar data have proven contradictory (e.g. Cooray et al. 2010; Maddox et al. 2010). For bright SMGs, a significant difficulty is their sparse number density, meaning large area surveys are required to yield sufficient galaxy pairs for the correlation function to be estimated robustly. An observational study of the clustering of SMGs was performed by Hickox et al. (2012), who ameliorated the problem of small numbers of SMGs by using a cross-correlation (Blake et al. 2006) with a more abundant *Spitzer* InfraRed Array Camera selected galaxy population to find that $z = 1$–3 SMGs in the LESS[2] source catalogue (Weiß et al. 2009) have a correlation length of $r_0 = 7.7^{+1.8}_{-2.3}$ h^{-1} Mpc, corresponding to an inferred halo mass of $M_{\mathrm{halo}} = 10^{12.8^{+0.3}_{-0.5}}$ h^{-1} M$_\odot$. This result is consistent with an earlier study by Blain et al. (2004), who used measured redshift separations of pairs of SMGs in a number of small fields to estimate a correlation length of 6.9 ± 2.1 h^{-1} Mpc. Hickox et al. (2012) used the median growth rate of haloes from Fakhouri et al. (2010) to suggest descendent halo masses consistent with those of local $\sim$2–3 $L_\star$ galaxies.

More recently, Wilkinson et al. (2017) performed a similar analysis. However, these authors were able to improve upon earlier work by making the first measurements of the clustering of SMGs as a function of redshift, owing to the greater number of SMGs detected as part of the SCUBA-2 (Super Common User Bolometer Array 2; Holland et al. 2013) Cosmology Legacy Survey (Geach et al. 2013, 2017) in the UKIDSS–UDS[3] field. Cross-correlating their SMG sample with a more numerous *K*-band-selected galaxy population, Wilkinson et al. (2017) estimated that the halo masses of SMGs ranged from $M_{\mathrm{halo}} \sim 10^{13}$ h^{-1} M$_\odot$ at $z \gtrsim$ 2 to $M_{\mathrm{halo}} \sim 10^{11}$ h^{-1} M$_\odot$ for $1 < z < 2$. Wilkinson et al. (2017) concluded that the $z \gtrsim 2$ SMG population could evolve into local $\sim$2–3 $L_\star$ galaxies.

However, the work of Hickox et al. (2012) and Wilkinson et al. (2017) is based on source catalogues derived from single-dish imaging surveys with a typical angular resolution of $\sim$20 arcsec at full width half-maximum (FWHM). Interferometers such as the Atacama Large Millimetre Array (ALMA) have an order of magnitude

better resolution, and targeted observations have revealed that many submm sources identified from single-dish imaging are in fact composed of multiple fainter galaxies that could not be distinguished from each other in the original single-dish survey due to its low angular resolution (e.g. Wang et al. 2011; Hodge et al. 2013). The effect this has on the observed number counts has been investigated (Karim et al. 2013; Simpson et al. 2015), but until recently, it has been unclear exactly what impact this has on measurements of the clustering of SMGs.

The first predictions for this were made by Cowley et al. (2016, hereafter C16). There we showed, using the clustering of SMGs predicted by the galaxy formation model of Lacey et al. (2016), that confusion due to the single-dish beam could boost the observed angular correlation function of submm sources by a factor of $\sim$4 relative to the correlation function of the underlying galaxies over all angular scales. This effect was termed 'blending bias'. Many of the blended sources detected in the simulated imaging of C16 comprise physically unassociated galaxies with a typical redshift separation of $\Delta z \sim 1$–2 (Cowley et al. 2015). These galaxies are often fainter than the flux limit of the survey and are boosted above this by the blending, together, of their flux by the beam. Their positions would not be included in the source catalogue otherwise. Though the galaxies that have their flux blended into a single source are generally chance projections along the line of sight, their positions are correlated with the positions of other galaxies at the same redshift. Some of these will also be included in the source catalogue, which leads to 'beam-induced' correlated pairs of sources resulting in a boost in the correlation function (blending bias) on all angular scales. Furthermore, C16 showed that the redshift interval considered has an impact on the blending bias, with narrower redshift intervals including fewer of these beam-induced pairs and so resulting in much smaller blending bias factors.

Wilkinson et al. (2017) performed an analysis similar to C16 for the redshift intervals considered in their work ($\Delta z = 0.5$ for $1.0 < z < 3.5$) and found a blending bias factor of $\sim$1.2 independent of redshift. However, this was for the autocorrelation of submm sources, and also did not consider the effect that the broadening with increasing redshift of the photometric redshift probability distributions of their galaxies would have on the blending bias. Here, we present predictions for the blending bias when the correlation function of submm sources is determined via a cross-correlation with a more abundant galaxy population. Also, in order to provide the best possible comparison of the observations of Wilkinson et al. (2017) and the galaxy formation model used by C16, we choose a *K*-band sample of similar depth and use the same redshift intervals considered in Wilkinson et al. (2017). We also mimic, to first order, the effect of broader photometric redshift probability distributions with increasing redshift on the clustering measurement. In addition, the nature of our simulations allows us to make predictions for the field-to-field variation expected for such observations. In this paper, we focus on observations and predictions made at 850 μm; however, we also expect that blending bias will affect clustering analyses made at shorter FIR wavelengths with *Herschel*–SPIRE[4] (e.g. Cooray et al. 2010; Maddox et al. 2010), where confusion is also significant (e.g. Nguyen et al. 2010).

This paper is structured as follows. In Section 2, we briefly describe the galaxy formation model, the model for computing the dust emission of the simulated galaxies at submm wavelengths and the method for creating the simulated imaging. In Section 3, we present

---

[2] Large APEX (Atacama Pathfinder EXperiment) Bolometer Camera Array Extended *Chandra* Deep Field South Submillimetre Survey.

[3] United Kingdom Infrared Telescope (UKIRT) Infrared Deep Sky Survey–Ultra Deep Survey (UDS).

[4] Spectral and Photometric Imaging Receiver.

our main results[5] and we conclude in Section 4. Throughout we assume a flat Λ cold dark matter cosmology with cosmological parameters consistent with the 7 yr *Wilkinson Microwave Anisotropy Probe* (WMAP7) results (Komatsu et al. 2011) i.e. $\Omega_m = 0.272$, $\Omega_b = 0.0455$, $\Omega_\Lambda = 0.728$, $h = 0.704$, $\sigma_8 = 0.81$ and $n_s = 0.967$.

2   THE THEORETICAL MODEL

Here we introduce our model that combines a dark-matter-only *N*-body simulation, a state-of-the-art semi-analytic model of galaxy formation and a simple model for the reprocessing of stellar radiation by dust in which the dust temperature is calculated self-consistently based on radiative transfer and global energy balance arguments. For further details, we refer the reader to Lacey et al. (2016). We also briefly describe our method for creating the simulated imaging used throughout.

2.1   GALFORM

The Durham semi-analytic model of hierarchical galaxy formation, GALFORM, was introduced by Cole et al. (2000), building on ideas outlined by White & Rees (1978), White & Frenk (1991) and Cole et al. (1994). Galaxy formation is modelled *ab initio*, beginning with a specified cosmology and a linear power spectrum of density fluctuations and ending with predicted galaxy properties at different redshifts.

Galaxies are assumed to form from baryonic condensation within the potential wells of dark matter haloes with their subsequent evolution being controlled in part by the merging history of the haloes. Here, the halo merger trees are extracted directly from a dark-matter-only *N*-body simulation (e.g. Helly et al. 2003; Jiang et al. 2014). We use a $(500\ h^{-1}\ \text{Mpc})^3$ Millennium-style Simulation (Springel et al. 2005; Guo et al. 2013) with cosmological parameters consistent with WMAP7 results (Komatsu et al. 2011), hereafter referred to as MR7. Halo masses are defined using the DHalo algorithm (Jiang et al. 2014).

The baryonic processes thought to be important for galaxy formation are included as a set of coupled differential equations that essentially track the exchange of mass and metals between stellar, cold disc gas and hot halo gas components in each galaxy. Stellar luminosities are computed through coupling the resulting star formation and metal enrichment histories of the simulated galaxies with evolutionary population synthesis models (e.g. Maraston 2005). The values of the parameters in these simplified equations, which describe the complex physical processes involved, are then calibrated against a predetermined set of data from observations and simulations, which provides a strong constraint on the available parameter space (e.g. Lacey et al. 2016). In particular, the Lacey et al. (2016) model is calibrated to reproduce the observed optical and near-infrared luminosity functions for $z \lesssim 3$ and, importantly for this work, the SMG number counts at 850 μm.

In this model, SMGs occupy haloes in the mass range $M_{\text{halo}} \sim 10^{11.5}$–$10^{12}\ h^{-1}\ \text{M}_\odot$ over a large range of redshifts ($0.2 \lesssim z \lesssim 4$). This is because the interplay of physical processes such as gas cooling, supernova feedback and radio-mode AGN feedback means this represents the halo mass range most conducive to star formation in the model (C16; Lacey et al. 2016).

---

[5] Some of the model data presented here will be made available at http://icc.dur.ac.uk/data/. For other requests, please contact the first author.

2.2   The dust model

To determine a simulated galaxy's FIR flux, a model is required to calculate the absorption and re-emission of stellar radiation by interstellar dust.

We assume that interstellar dust exists in two components: spherical molecular clouds of a fixed gas surface density in which stars form and a diffuse interstellar medium distributed smoothly throughout an exponential disc. The energy from stellar radiation absorbed by each component can be calculated using the star formation and metal enrichment histories for a galaxy predicted by GALFORM, and then solving the equations of radiative transfer in this assumed geometry. The dust emission is then calculated using global energy balance arguments, assuming that it emits as a modified blackbody. Crucially, this means that dust temperature is not a free parameter in the model but is calculated self-consistently for each dust component in each galaxy.

Despite its simplicity, the model is able to reproduce the predictions of the more sophisticated spectrophotometric code GRASIL (Silva et al. 1998), offset only by minor factors of $\lesssim 2$ with a very tight scatter, for $\lambda_{\text{rest}} \gtrsim 70\ \mu$m (Cowley et al. 2017). The sheer computational expense of codes such as GRASIL ($\sim$3–5 CPU min per galaxy), however, makes them unsuitable for the large number of galaxies contained in the cosmological volumes used in this work.

2.3   Creating simulated imaging of SMGs

The lightcone code presented in Merson et al. (2013) is used to create mock surveys of SMGs using 50 randomly orientated lines of sight through the simulation volume. For the purposes of this study, we use an AB apparent magnitude of $m_K < 25$ to select our *K*-band population, similar to that used by Wilkinson et al. (2017), and a limit of $S_{850\,\mu\text{m}} > 0.35$ mJy. This 850 μm limit is chosen as it is the flux above which 90 per cent of the total predicted background light is included in a typical image. This prediction is in good agreement with the observations of Puget et al. (1996) and Fixsen et al. (1998), and so means that our submm maps have a realistic extragalactic background light (Cowley et al. 2015). We choose a maximum area of 4 deg^2 for our fields, which is larger than that currently surveyed in submm observations, to reduce the effect of a finite survey area on our results. An example of the *K*-band absolute magnitude versus redshift for the resulting input catalogue is shown in Fig. 1.

Following the method presented in Cowley et al. (2015), galaxies in each mock catalogue are mapped on to a grid of pixels according to their position, such that the value in a pixel is equal to the sum of the 850-μm flux of all the galaxies that fall within it. The resulting image is then smoothed with a Gaussian beam, with the pixel scale chosen such that the beam is well sampled. Instrumental (white) noise is then added, and matched-filtering is performed prior to source extraction. In order to mimic the SCUBA-2 observational data, we choose a Gaussian with a 15 arcsec FWHM and $\sim$1 mJy beam^{-1} of instrument noise. Source counterparts are identified as the galaxy that makes the dominant contribution to the overall submm flux of the source. The statistics of the resulting source catalogue can then be compared to those of the underlying galaxies. The resulting redshift distributions are compared in Fig. 2. We can see that the source counterparts are more numerous than galaxies at the same flux limit, and that their distribution has a more prominent high-redshift tail. The surface number densities of the flux-limited *K*-band sample, the $S_{850\,\mu\text{m}} > 4$ mJy galaxies and the
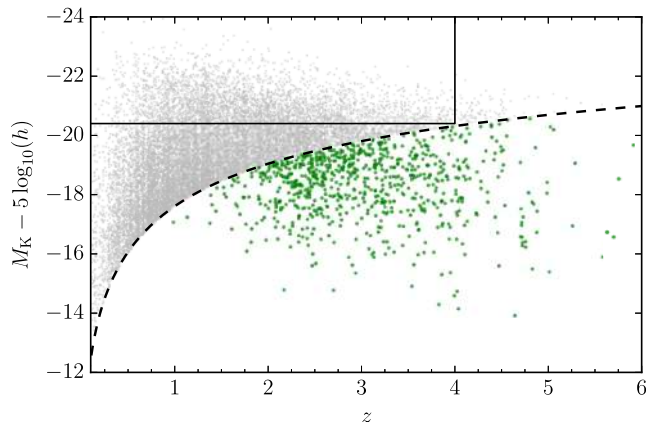
Figure 1. The K-band absolute magnitude versus redshift for lightcone galaxies flux-limited in the K band at an apparent magnitude of 25 (grey dots) for one of the 50×4 deg^2 fields. For clarity, only 1 per cent of this sample is shown. The K-band flux limit is indicated by the dashed black line. The green points indicate galaxies with $S_{850\,\mu m} > 4$ mJy, which are not selected in the K band. The black solid lines indicate a volume-limited K-band sample for $z < 4$. All magnitudes are in the AB system.
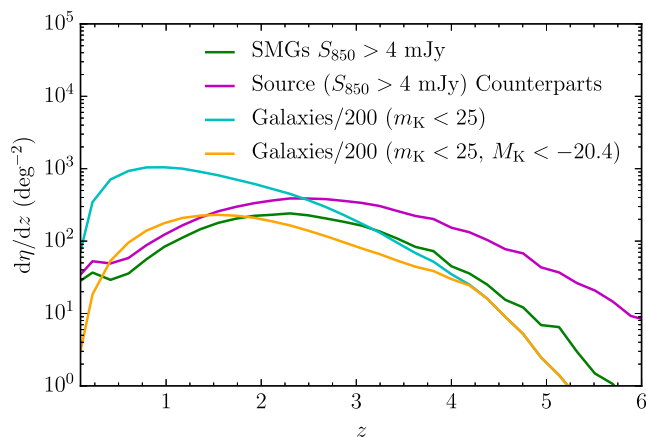
Figure 2. Predicted average redshift distributions from our 50×4 deg^2 fields for $S_{850\,\mu m} > 4$ mJy galaxies (green line), the counterparts (see the text) of sources with $S_{850\,\mu m} > 4$ mJy extracted from the simulated submm imaging (magenta line), flux-limited sample of K-band-selected galaxies (cyan line) and a volume-limited sample (for $z < 4$) of K-band-selected galaxies (orange line). The latter two lines have both been divided by a factor of 200 for presentation purposes. All magnitudes are in the AB system.

counterparts to $S_{850\,\mu m} > 4$ mJy sources are 4.02×10^5, 5.54×10^2 and 1.05×10^3 deg^{-2}, respectively.

3 RESULTS

The simplest measure of clustering from a galaxy imaging survey is the two-point angular correlation function $w(\theta)$. The probability of finding two objects separated by an angle $\theta > 0$ is defined as (e.g. Peebles 1980)

$$\delta P_{12}(\theta) = \eta^2 \left[1 + w(\theta)\right] \delta\Omega_1 \delta\Omega_2, \tag{1}$$

where η is the mean surface density of objects per unit solid angle and $\delta\Omega_i$ is a solid angle element, such that $w(\theta)$ represents the excess probability of finding objects at angular separation, θ, relative to a random (Poisson) distribution. The measured angular correlation function for a given galaxy population, w_g, can then be compared

to that expected for the dark matter, w_{DM}, to yield the large-scale bias of the galaxy population, calculated as

$$b_g(\theta) = \left[\frac{w_g(\theta)}{w_{DM}(\theta)}\right]^{1/2}. \tag{2}$$

Although the bias is scale dependent (e.g. Angulo et al. 2008), it is usually approximated as a constant on large (linear) scales, where it can be compared to an expected bias computed by a weighted average of the bias values over the haloes that are occupied (e.g. Cooray & Sheth 2002):

$$b_{eff}(z) = \frac{\int b(M_h, z)\, n(M_h, z)\, \langle N_{gal}|M_h\rangle\, \mathrm{d}M_h}{\int n(M_h, z)\, \langle N_{gal}|M_h\rangle\, \mathrm{d}M_h}. \tag{3}$$

Here, $b(M_h, z)$ is the large-scale bias of halos with mass M_h at redshift z, $n(M_h, z)$ is the halo mass function at redshift z such that $n(M_h, z)\, \mathrm{d}M_h$ is the comoving number density of halos in the mass range $[M_h, M_h + \mathrm{d}M_h]$, and $\langle N_{gal}|M_h\rangle$ is the mean of the halo occupation distribution (HOD, the expected mean number of galaxies within a halo of mass M_h). Thus, measuring the large-scale bias can yield information regarding the halo masses that the galaxy population occupy.

3.1 A cross-correlation analysis of submm sources

As we are comparing the predictions of our model to the analysis of Wilkinson et al. (2017), we begin with the source catalogue ($S_{850\,\mu m} > 4$ mJy) derived from source extraction from the simulated images of C16 as described previously. The SMG sample used by Wilkinson et al. (2017) has a slightly fainter flux limit ($\sim$3.5 mJy; Chen et al. 2016); however, we do not expect this to have a significant impact on our science results. In C16, we showed that the angular autocorrelation of the submm sources, w_s, was boosted by a 'blending bias' factor, b_b, relative to that of the underlying galaxy population, w_g, such that $w_s = b_b^2 w_g$. In this paper, we calculate w_s via a cross-correlation with a volume-limited K-band-selected galaxy population ($m_K < 25$). Assuming linear theory, the large-scale bias of the submm sources, b_s, can be determined using
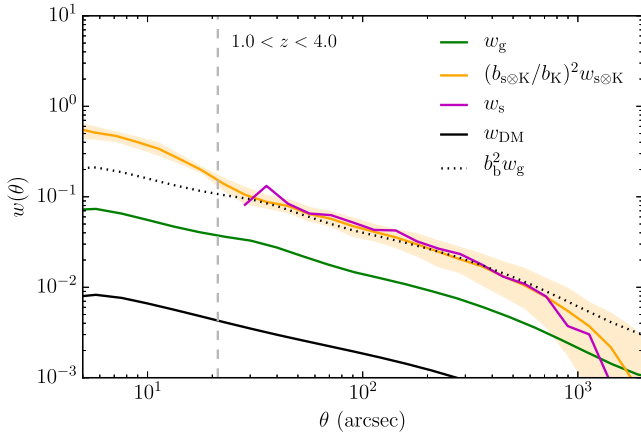
$$b_s = b_{s\otimes K}^2 / b_K, \tag{4}$$

where b_K represents the bias of the K-band-selected galaxy population as measured from its autocorrelation function and $b_{s\otimes K}$ is the bias of the cross-correlation of the two populations. This means that $(b_{s\otimes K}/b_K)^2 w_s$ is equivalent to w_s, provided that the blending bias effects both measurements in the same way.

To calculate the angular cross-correlation of the submm sources and the K-band galaxy sample, $w_{s\otimes K}$, we use the Landy & Szalay (1993) estimator adapted for cross-correlations:

$$w_{s\otimes K}(\theta) = \frac{DD_{sK} - DR_{sK} - DR_{Ks} + RR_{sK}}{RR_{sK}}, \tag{5}$$

where DD, DR and RR represent data–data, data–random and random–random pairs, respectively, and the subscripts s and K represent the submm sources and K-band-selected galaxies, respectively. In calculating $w_{s\otimes K}$, we use the actual number of sources in each field to estimate the mean surface density, rather than the true surface density. This causes the angular correlation function to be underestimated by an average amount, σ^2, often referred to as the integral constraint (Groth & Peebles 1977). For the cross-correlation functions, this quantity is related to the field-to-field variation in the number counts through

$$\sigma_{s\otimes K}^2 = \frac{\langle (\eta_s - \langle\eta_s\rangle)(\eta_K - \langle\eta_K\rangle)\rangle}{\langle\eta_s\rangle\langle\eta_K\rangle} - \frac{\langle\eta_{sK}\rangle}{\langle\eta_s\rangle\langle\eta_K\rangle}, \tag{6}$$
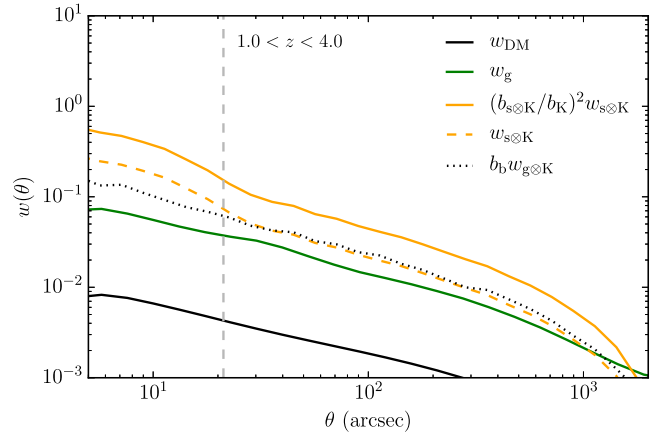
Figure 3. Predicted angular correlation functions in the redshift range $1.0 < z < 4.0$. The angular correlation function of galaxies selected by $S_{850\,\mu m} > 4$ mJy is shown by the green line. The cross-correlation of counterparts to sources with $S_{850\,\mu m} > 4$ mJy with a volume-limited K-band-selected sample, averaged over our 50×4 deg^2 fields and scaled to remove the bias of the K-band sample, is shown by the orange line. The shaded orange region corresponds to the predicted 1σ (16–84 percentile) field-to-field variation for the 4 deg^2 field area used. The autocorrelation of the source counterparts (averaged over 50×4 deg^2) is shown by the magenta line; the correlation function of dark matter in the MR7 simulation is shown by the black line; and the correlation function of the galaxies scaled by the blending bias squared (here $b_b = 1.7$) is shown by the black dotted line. The vertical dashed line indicates the FWHM of the match-filtered point spread function used to create the simulated imaging $\sim$21.2 arcsec.

where η_{sK} represents the surface density of objects that are in both populations. We evaluate this quantity for our 50 lightcone fields and add it on to our computed cross-correlation functions. We also make the corresponding correction to our autocorrelation functions. These corrections are typically of the order of $\sim 10^{-3}$. We note that equation (6) is not how this correction is usually calculated in observational studies, where the expression $\sigma^2 = \sum RR(\theta)w(\theta)/\sum RR(\theta)$ is more commonly used to evaluate the integral constraint Roche & Eales (1999), in the absence of multiple fields. However, we have checked that this expression gives essentially identical results to equation (6).

In Fig. 3, we show the angular cross-correlation function of submm sources with the K-band galaxy population and (for comparison) the autocorrelation of submm sources, over the redshift range $1 < z < 4$. For our submm sources, we use the position and redshift of the galaxy that makes the largest contribution to the flux of the source. The angular correlation functions for the galaxies and dark matter are calculated from their spatial correlation functions using the Limber (1953) equation (computed using a method similar to that described in Gonzalez-Perez et al. 2011), appropriately changing the redshift limits, as is done in C16.[6] We derive a blending bias factor of $b_b \sim 1.7$ comparing the clustering of submm sources and galaxies. For reference, we also show the galaxy correlation function scaled by b_b^2. For calculating the biases, we restrict ourselves to the angular range over which the dark matter correlation function is approximately linear. We do this by excluding scales for which $w_{DM,non-linear} > 1.2 \times w_{DM,linear}$ from our computation of the bias. We also exclude angular scales larger than 10^3 arcsec to

ensure that the bias measurements are not affected by the finite area of our mock surveys. We can see that the autocorrelation and the scaled cross-correlation functions are essentially the same. It therefore appears that blending bias behaves in a similar manner to a linear scale-independent bias. In this regime, the ratio of the cross-correlation of the K-band sample with the submm sources to the cross-correlation of the K-band sample with SMGs should simply be equal to the blending bias i.e. $w_{s\otimes K} = b_b w_{g\otimes K}$. We show that this is the case in Fig. 4.
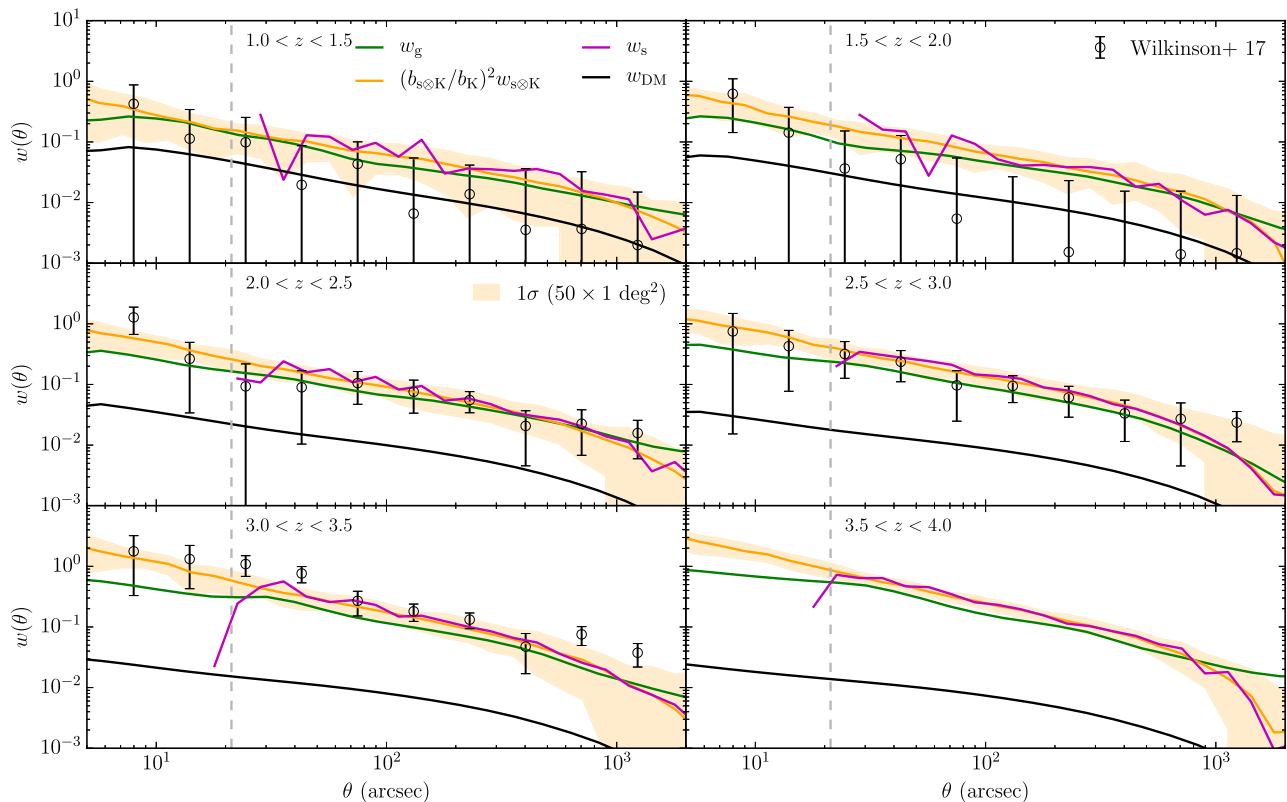
Thus, whilst the cross-correlation technique can provide smaller statistical errors than the autocorrelation due to the larger number of objects considered, it is still affected by blending bias in the same way.

In order to compare the predictions of our model to the observations of Wilkinson et al. (2017), we repeat this analysis using their quoted redshift intervals with $\Delta z = 0.5$. This is shown in Fig. 5. We also show the predicted 16–84 percentile field-to-field variance, estimated from 50 lightcone fields. For calculating the predicted field-to-field variance, we assume an area of 1 deg^2 comparable to that used in Wilkinson et al. (2017).

The agreement between the model and the observations appears to be generally favourable, with the majority of observed data points in each redshift bin (apart from the $1.5 < z < 2$ bin) lying within the predicted 1σ region, indicating that the model is broadly consistent with the observed data.

We can also see from Fig. 5 that the blending bias factors have been reduced (to $b_b \sim 1.1$–1.2) due to the narrower redshift interval than considered previously. Again, they are essentially the same as those that would be derived from the autocorrelation of the submm sources and are very similar to those derived in Wilkinson et al. (2017) for the autocorrelation case (see their table 2).

In Fig. 6, we show the large-scale bias calculated from the cross-correlation derived function compared to that of the actual underlying galaxies. We can see that blending bias still affects the inferred halo mass of the SMGs, although to a much lesser extent than it would for the broader $1 < z < 4$ redshift interval, where $b_b \sim 1.7$. Using the large-scale bias–halo mass relations of Sheth, Mo & Tormen (2001), we find that the blending bias ($b_b \sim 1.1$–1.2) results

---

[6] In principle, these could be derived from lightcone catalogues giving essentially identical results; however, we prefer using Limber's equation as it utilizes all of the clustering information in our simulation volume.

Figure 5. Predicted angular correlation functions for different redshift intervals indicated in the panels for galaxies selected with $S_{850\,\mu m} > 4$ mJy (green lines), the cross-correlation of counterparts to sources with $S_{850\,\mu m} > 4$ mJy with a volume-limited K-band-selected sample, averaged over 50×4 deg^2 fields and scaled so as to remove the bias of the K-band sample (orange line), and the autocorrelation of the source counterparts (averaged over the 50×4 deg^2 fields; magenta line). We also show observational data from Wilkinson et al. (2017), which are derived from a cross-correlation of sources with a K-band-selected galaxy sample, and so should be compared with our orange line. The shaded orange region corresponds to the 1σ (16–84 percentile) scatter derived from field-to-field variations, calculated from the central 1 deg^2 region in each of our fields in order to match the area used in the observations of Wilkinson et al. (2017). The vertical dashed line indicates the FWHM of the match-filtered point spread function used to create the simulated imaging ~ 21.2 arcsec.

in the halo masses of SMGs being overestimated by a factor of ~ 2. For the broader $1 < z < 4$ redshift interval, this overestimate is a factor of ~ 6. For this, we have assumed that all galaxies occupy host dark matter haloes of the same mass (i.e. the $\langle N_{gal}|M_h \rangle$ term in equation 3 is described by a Dirac delta function) and used the median redshift of the relevant population (SMGs or submm source counterparts) in the redshift interval considered. We also show in Fig. 6, for comparison, the large-scale bias values derived by Wilkinson et al. (2017), though recomputed assuming the same WMAP7 cosmological parameters as assumed in this work.

Immediately apparent from Fig. 6 is that despite the general agreement between the predicted and observed correlation functions shown in Fig. 5, the inferred large-scale bias values do not agree. We attribute this to photometric redshift probability distributions used for the observed galaxies, and discuss this in more detail in the following section. We list our results from this section, the predicted large-scale SMG bias (b_g), blending bias (b_b) and large-scale submm source bias (b_s) for each $\Delta z = 0.5$ redshift interval [listed in column (a)] in Table 1. The table also lists the results from Section 3.2, where we use the redshift intervals described in column (b), as discussed below. For reference, we also list the large-scale bias values derived by Wilkinson et al. (2017).

3.2 The effect of photometric redshifts

Given the apparent good agreement between the predicted and observed correlation functions in Fig. 5, the cause of the extreme

differences in the derived bias values (and subsequent conclusions about the host halo masses), shown in Fig. 6, is worthy of further investigation. As mentioned earlier, we attribute this to the width of the photometric redshift probability distributions used for each galaxy by Wilkinson et al. (2017), a necessary consequence of the available photometry. The redshifts in Wilkinson et al. (2017) were mostly obtained from the UDSz ESO Large Programme (ID:180. A-0776; PI: O. Almaini). The EAZY template-fitting pack (Brammer, van Dokkum & Coppi 2008) was used to derive a photometric redshift probability distribution for each galaxy through a maximum likelihood analysis. SMG counterparts were assigned using the Optical InfraRed Triple Colour method (Chen et al. 2016). A galaxy in the Wilkinson et al. (2017) analysis is able to appear in multiple redshift intervals, weighted by the integral of its probability distribution between the limits of the redshift interval.

A consequence of this is that the effective redshift distributions used for each bin are typically broader than the quoted limits of the bin would suggest, and become broader with increasing redshift as the quality of the photometric redshifts generally degrade i.e. the probability distributions become broader. In Fig. 7, we show the average SMG counterpart redshift distributions for each redshift interval from Wilkinson et al. (2017), these being calculated as the sum of all the individual galaxy photometric redshift probability distributions weighted by their integral between the quoted limits.
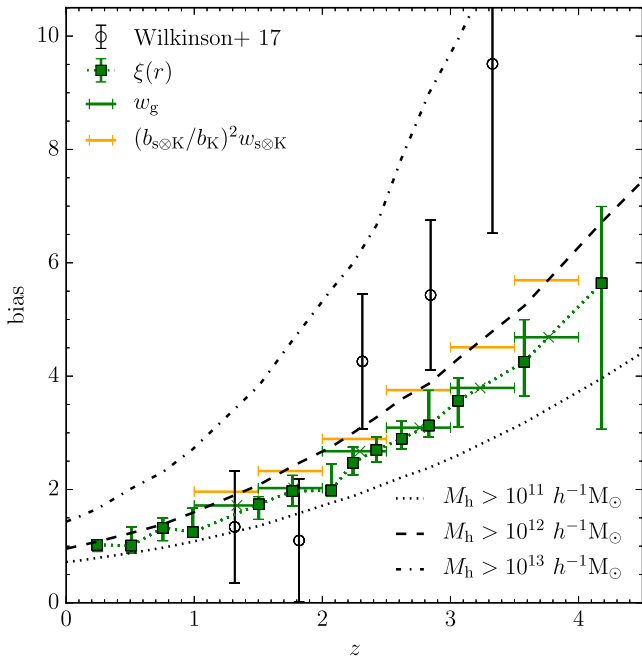
Figure 6. Predicted evolution of large-scale bias with redshift. Green squares with error bars represent the bias measured directly from the 3D spatial correlation function of SMGs with $S_{850\,\mu m} > 4$ mJy, as is done in C16. The 1σ errors are calculated using the volume bootstrap method advocated in Norberg et al. (2009). The horizontal green bars show the large-scale bias of the SMGs with $S_{850\,\mu m} > 4$ mJy derived from the angular correlation function over the redshift range indicated by the width of the bar. The horizontal orange bars show the same but for the angular correlation function of submm sources calculated via a cross-correlation with a volume-limited K-band-selected sample. The dotted, dashed and dash–dotted black lines show the evolution of the large-scale bias of haloes with $M_{halo} > 10^{11}$, 10^{12} and $10^{13}\ h^{-1}$ M$_{\odot}$, respectively, measured directly from the MR7 simulation. Observational data (black circles with errors) are from Wilkinson et al. (2017).

Thus, the angular correlation functions for dark matter used by Wilkinson et al. (2017) would typically have a lower normalization than that shown in Fig. 5 (where we used the true redshifts of the galaxies in the simulation and a top-hat redshift window of $\Delta z = 0.5$) as the spatial correlation function of the dark matter, $\xi_{DM}(r, z)$, has effectively been projected over a larger volume. This explains how the agreement between the angular correlation functions shown in Fig. 5 is consistent with the disagreement in the inferred large-scale bias shown in Fig. 6.

To mimic the effect of the width of photometric redshift distribution to first order, we increase the width of the redshift intervals we
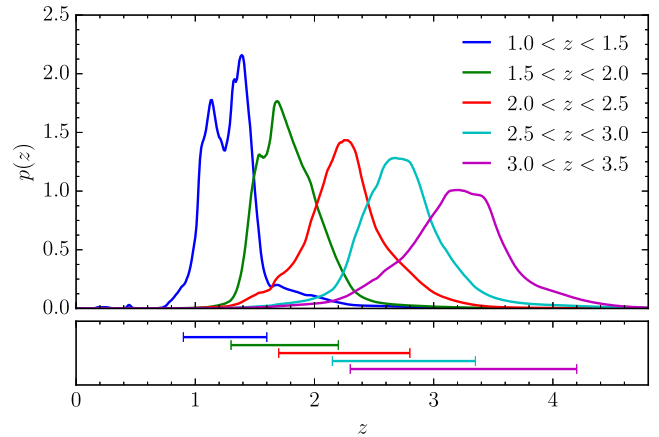
Figure 7. Top panel: SMG photometric redshift distributions from Wilkinson et al. (2017). The distributions are shown for the redshift intervals indicated in the legend and are normalized to have unit area. Bottom panel: the width of the top-hat redshift interval required (with the same central redshift) so that the angular dark matter correlation functions computed using the predicted redshift distributions in Fig. 2 have the same normalization as those computed using the redshift distributions in the top panel.

consider (symmetrically in redshift, maintaining the same central redshift) until our dark matter correlation functions have a similar normalization to those calculated using the redshift distributions of Wilkinson et al. (2017) for each bin. These new redshift interval widths are shown in the bottom panel of Fig. 7 and listed in column (b) of Table 1.

We then repeat our analysis using these new top-hat redshift intervals. We show two examples of this, for the $2.0 < z < 2.5$ and $3.0 < z < 3.5$ bins (for which we now use redshift intervals of $1.7 < z < 2.8$ and $2.3 < z < 4.2$, respectively), in Fig. 8 and list the results for each interval in Table 1. Considering a broader redshift distribution brings the large-scale bias values we measure for the simulated submm sources into broad agreement with the values quoted by Wilkinson et al. (2017), apart from the $1.0 < z < 1.5$ bin where the large-scale bias is overpredicted, and the $3.0 < z < 3.5$ bin where it is underpredicted.

Our reasoning for the agreement between the observed and predicted large-scale bias values for submm sources found here is as follows. As the width of the redshift interval we consider increases, the blending bias also increases. This is due to the inclusion of more 'beam-induced' correlated pairs in the correlation function calculation as is discussed in C16. However, the intrinsic galaxy large-scale bias remains approximately constant. Therefore, the increase in blending bias means that the inferred large-scale bias for the sources becomes greater.

Table 1. Predicted large-scale bias of SMGs (b_g), blending bias (b_b) and large-scale bias of submm sources (b_s; note that $b_g b_b = b_s$) for the top-red redshift intervals indicated in columns (a) and (b). The large-scale bias observed by Wilkinson et al. (2017) is also shown.

| (a)	b_g	b_b	b_s	(b)	b_g	b_b	b_s	b_s (Wilkinson et al. 2017)
$1.0 < z < 1.5$	1.7	1.1	2.0	$0.9 < z < 1.6$	1.7	1.1	1.9	1.34 ± 0.99
$1.5 < z < 2.0$	2.0	1.1	2.3	$1.3 < z < 2.2$	2.1	1.2	2.5	1.10 ± 1.09
$2.0 < z < 2.5$	2.7	1.1	2.9	$1.7 < z < 2.8$	2.6	1.3	3.3	4.26 ± 1.19
$2.5 < z < 3.0$	3.1	1.2	3.8	$2.1 < z < 3.3$	3.1	1.4	4.2	5.43 ± 1.32
$3.0 < z < 3.5$	3.8	1.2	4.5	$2.3 < z < 4.2$	3.4	1.5	5.0	9.51 ± 2.99

Notes. (a) Top-hat redshift interval used in Section 3.1 and quoted by Wilkinson et al. (2017).
(b) Top-hat redshift interval used in Section 3.2, chosen such that the normalization of the dark matter correlation function is the same as used by Wilkinson et al. (2017).
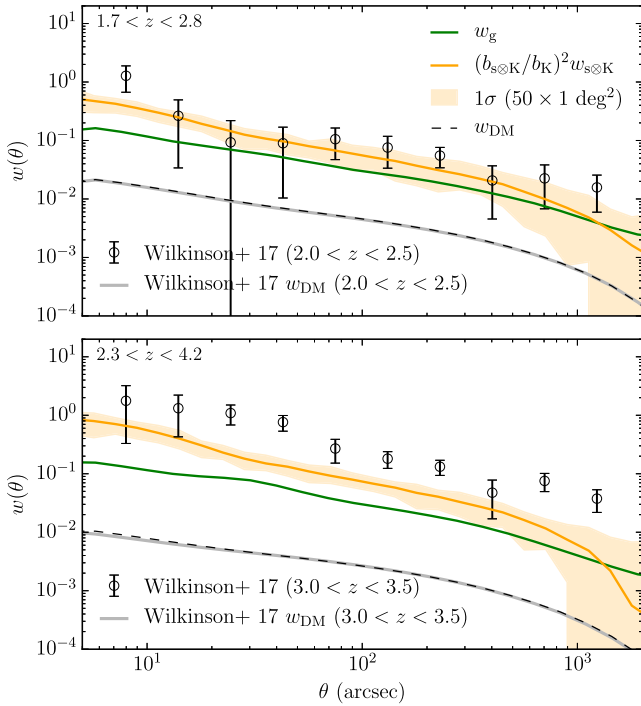
Figure 8. Predicted angular correlation functions for the redshift intervals $1.7 < z < 2.8$ (top panel) and $2.3 < z < 4.2$ (bottom panel) that correspond to the $2.0 < z < 2.5$ and $3.0 < z < 3.5$ intervals in Fig. 5, respectively. These broader intervals are chosen such that the angular correlation function for dark matter (dashed black line) is in agreement with that used by Wilkinson et al. (2017, solid grey line) for the redshift bin. All other lines and symbols have the same meaning as in Fig. 5.
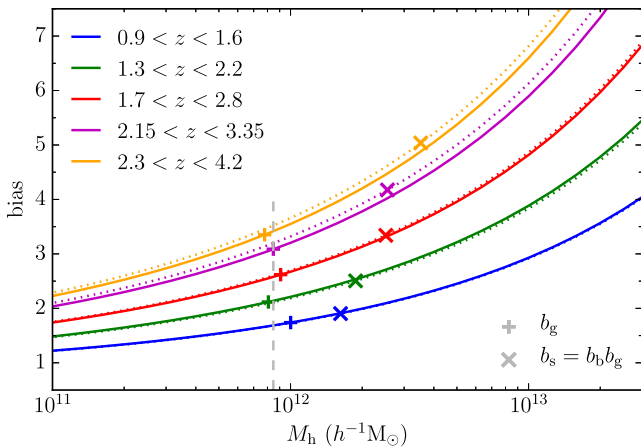
Figure 9. Large-scale bias-to-dark matter halo mass relations of Sheth et al. (2001) calculated at the median redshift of the interval considered for galaxies (solid lines) and source counterparts (dotted lines). The different colours are for the redshift intervals indicated in the legend. Plus signs (crosses) indicate the position of galaxies (source counterparts) on this plane using the biases derived from the corresponding angular correlation functions. The vertical dashed grey line shows the median inferred halo mass for the galaxies.

In Fig. 9, we show the effect this has on the inferred host halo masses as a function of redshift. We use the large-scale bias-to-halo mass relations of Sheth et al. (2001) and assume that the objects occupy haloes of a single mass at the median redshift of the interval considered, which we calculate using the relevant redshift distributions from Fig. 2. For the galaxies, we find this yields inferred halo

masses consistent with those that the galaxies are known to occupy in the model (see fig. 5 of C16) and with no significant redshift evolution over this range. For the sources, however, we observe a mild evolution in halo mass from $\sim 4 \times 10^{12} \ h^{-1} \ M_\odot$ at $z \sim 3$ to $\sim 2 \times 10^{12} \ h^{-1} \ M_\odot$ at $z \sim 1$, due to the blending bias being larger at higher redshift as the redshift interval considered is broader. Whilst it appears unlikely from this analysis that this effect could account for all of the very strong halo mass 'downsizing' found by Wilkinson et al. (2017), it is possible that the apparent downsizing trend was amplified by this effect, as the broadening of the redshift intervals with increasing redshift was not considered by Wilkinson et al. (2017) when deriving their blending bias factors.

We conclude that measuring the correlation function for submm sources via an autocorrelation or cross-correlation is affected by blending bias in the same way. Measuring the cross-correlation using objects within a relatively narrow redshift range is the best way to perform such a measurement, due to the increased statistical significance from the cross-correlation with a more abundant sample and to the reduced blending bias due to the narrower redshift range being investigated. Such an analysis is performed by Wilkinson et al. (2017). However, this comes with the important caveat that accurate redshifts for the correct counterpart to the submm source are required, and there are a sufficient number of objects in each redshift bin for the result to be statistically significant. Alternatively, as is discussed in C16, a significant targeted follow-up campaign with interferometers such as ALMA would allow the blended sources in the single-dish catalogue to be identified and removed from the clustering analysis, providing a result free from blending bias. Investigation of the evolution of the SMG clustering with redshift will still require accurate redshifts (at the level that the typical redshift error is expected to be factors of a few smaller than the width of the redshift bin), but this is an issue separate from the blending bias.

4 CONCLUSIONS

We study the effect of 'blending bias', the square root of the factor by which the angular correlation function of sources identified through source extraction of confused imaging is boosted relative to that of the underlying galaxy population. In particular, we focus on its applications to the measurement of the clustering of SMGs, an important population of galaxies that exhibit some of the highest inferred star formation rates in the Universe, as this can constrain the dark matter halo masses of these galaxies and thus their subsequent evolution.

To do so, we use the galaxy formation model presented in Lacey et al. (2016), which can successfully reproduce the observed number counts of SMGs at 850 μm, and the methodology of Cowley et al. (2015) to create simulated imaging based on the model galaxies.

We compare our model predictions to the recent analysis of Wilkinson et al. (2017), who cross-correlated a sample of SMGs in the UKIDSS–UDS field with a more numerous K-band-selected sample to derive the large-scale bias and halo masses of SMGs in five redshift intervals from $1.0 \lesssim z \lesssim 3.5$.

Importantly, we find that the blending bias factors are essentially the same whether the correlation function is derived through an autocorrelation or cross-correlation technique, though they can be reduced by decreasing the width of the redshift interval considered. This adds weight to the accuracy of the Wilkinson et al. (2017) study, which is the first to measure the evolution in the clustering of SMGs. However, our predictions indicate their results may still be affected systematically by blending biases of at least $b_b \sim 1.1$–1.2, which can lead to the host halo masses being overestimated by a factor of ~ 2.

We investigate the effect of the redshift intervals considered becoming effectively broader with increasing redshift due to the use of photometric redshift probability distributions, a necessary consequence of the available photometry used in observational studies. We find that this can result in inferring a spurious halo mass 'downsizing' trend, where the halo masses inferred from the clustering at $z \sim 3$ are a factor of ~ 2 greater than those inferred at $z \sim 1$. This is due to the blending bias factors being larger at higher redshift as a result of the broader redshift distributions used. However, this trend is not as strong as the one observed by Wilkinson et al. (2017).

Finally, we note that the blending bias values quoted in this work may be somewhat model dependent and caution that further work is required to fully understand the implications of this bias for measurements made from catalogues derived from single-dish imaging surveys at FIR/submm wavelengths. Additionally, we hope that this effect will be confronted directly with future ALMA observations, which would allow the clustering of an SMG sample to be measured free from blending bias.

REFERENCES

Alexander D. M., Bauer F. E., Chapman S. C., Smail I., Blain A. W., Brandt W. N., Ivison R. J., 2005, ApJ, 632, 736
Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, MNRAS, 383, 755
Blain A. W., Smail I., Ivison R. J., Kneib J.-P., Frayer D. T., 2002, Phys. Rep., 369, 111
Blain A. W., Chapman S. C., Smail I., Ivison R., 2004, ApJ, 611, 725
Blake C., Pope A., Scott D., Mobasher B., 2006, MNRAS, 368, 732
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45
Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, ApJ, 622, 772
Chen C.-C. et al., 2016, ApJ, 820, 82
Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
Cooray A. et al., 2010, A&A, 518, L22
Coppin K. et al., 2006, MNRAS, 372, 1621
Cowley W. I., Lacey C. G., Baugh C. M., Cole S., 2015, MNRAS, 446, 1784
Cowley W. I., Lacey C. G., Baugh C. M., Cole S., 2016, MNRAS, 461, 1621
Cowley W. I., Béthermin M., del P. Lagos C., Lacey C. G., Baugh C. M., Cole S., 2017, MNRAS, 467, 1231
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Fixsen D. J., Dwek E., Mather J. C., Bennett C. L., Shafer R. A., 1998, ApJ, 508, 123
Geach J. E. et al., 2013, MNRAS, 432, 53
Geach J. E. et al., 2017, MNRAS, 465, 1789
González J. E., Lacey C. G., Baugh C. M., Frenk C. S., 2011, MNRAS, 413, 749
Gonzalez-Perez V., Baugh C. M., Lacey C. G., Kim J.-W., 2011, MNRAS, 417, 517
Groth E. J., Peebles P. J. E., 1977, ApJ, 217, 385
Guo Q., White S., Angulo R. E., Henriques B., Lemson G., Boylan-Kolchin M., Thomas P., Short C., 2013, MNRAS, 428, 1351
Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., 2003, MNRAS, 338, 903
Hickox R. C. et al., 2012, MNRAS, 421, 284
Hodge J. A. et al., 2013, ApJ, 768, 91
Holland W. S. et al., 2013, MNRAS, 430, 2513
Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, MNRAS, 440, 2115
Karim A. et al., 2013, MNRAS, 432, 2
Komatsu E. et al., 2011, ApJS, 192, 18
Lacey C. G. et al., 2016, MNRAS, 462, 3854
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Limber D. N., 1953, ApJ, 117, 134
Maddox S. J. et al., 2010, A&A, 518, L11
Maraston C., 2005, MNRAS, 362, 799
Merson A. I. et al., 2013, MNRAS, 429, 556
Nguyen H. T. et al., 2010, A&A, 518, L5
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Puget J.-L., Abergel A., Bernard J.-P., Boulanger F., Burton W. B., Desert F.-X., Hartmann D., 1996, A&A, 308, L5
Roche N., Eales S. A., 1999, MNRAS, 307, 703
Scott S. E. et al., 2002, MNRAS, 331, 817
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103
Simpson J. M. et al., 2014, ApJ, 788, 125
Simpson J. M. et al., 2015, ApJ, 799, 81
Smail I., Ivison R. J., Blain A. W., Kneib J.-P., 2002, MNRAS, 331, 495
Springel V. et al., 2005, Nature, 435, 629
Swinbank A. M., Chapman S. C., Smail I., Lindner C., Borys C., Blain A. W., Ivison R. J., Lewis G. F., 2006, MNRAS, 371, 465
Swinbank A. M. et al., 2014, MNRAS, 438, 1267
Wang W.-H., Cowie L. L., Barger A. J., Williams J. P., 2011, ApJ, 726, L18
Webb T. M. et al., 2003, ApJ, 587, 41
Weiß A. et al., 2009, ApJ, 707, 1201
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wilkinson A. et al., 2017, MNRAS, 464, 1380
Williams C. C. et al., 2011, ApJ, 733, 92

This paper has been typeset from a TEX/LATEX file prepared by the author.