Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **470,** 2253–2261 (2017)
Advance Access publication 2017 June 3

doi:10.1093/mnras/stx1367

On the statistics of proto-cluster candidates detected in the *Planck* all-sky survey

M. Negrello,[1⋆] J. Gonzalez-Nuevo,[2] G. De Zotti,[3] M. Bonato,[3,4] Z.-Y. Cai,[5,6] D. Clements,[7] L. Danese,[4] H. Dole,[8,9] J. Greenslade,[5] A. Lapi[4] and L. Montier[10,11]

[1]*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK*
[2]*Departamento de Física, Universidad de Oviedo, Avda. Calvo Sotelo s/n, Oviedo, Spain*
[3]*INAF, Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122 Padova, Italy*
[4]*SISSA, Via Bonomea 265, I-34136 Trieste, Italy*
[5]*CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei 230026, China*
[6]*School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China*
[7]*Imperial College London, Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
[8]*Institut d'Astrophysique Spatiale, CNRS (UMR 8617) Université Paris-Sud 11, Bâtiment 121, F-91405 Orsay, France*
[9]*Institut Universitaire de France, 103 bd Saint-Michel, F-75005 Paris, France*
[10]*UPS-OMP, IRAP, Université de Toulouse, F-31028 Toulouse Cedex 4, France*
[11]*CNRS, IRAP, 9 Av. colonel Roche, BP 44346, F-31028 Toulouse Cedex 4, France*

Accepted 2017 May 31. Received 2017 May 31; in original form 2017 April 13

ABSTRACT

Observational investigations of the abundance of massive precursors of local galaxy clusters ('proto-clusters') allow us to test the growth of density perturbations, to constrain cosmological parameters that control it, to test the theory of non-linear collapse and how the galaxy formation takes place in dense environments. The *Planck* collaboration has recently published a catalogue of $\gtrsim$2000 *cold* extragalactic sub-millimeter sources, i.e. with colours indicative of $z \gtrsim 2$, almost all of which appear to be overdensities of star-forming galaxies. They are thus considered as proto-cluster candidates. Their number densities (or their flux densities) are far in excess of expectations from the standard scenario for the evolution of large-scale structure. Simulations based on a physically motivated galaxy evolution model show that essentially all *cold* peaks brighter than $S_{545\mathrm{GHz}} = 500$ mJy found in *Planck* maps after having removed the Galactic dust emission can be interpreted as positive Poisson fluctuations of the number of high-z dusty proto-clusters within the same *Planck* beam, rather then being individual clumps of physically bound galaxies. This conclusion does not change if an empirical fit to the luminosity function of dusty galaxies is used instead of the physical model. The simulations accurately reproduce the statistic of the *Planck* detections and yield distributions of sizes and ellipticities in qualitative agreement with observations. The redshift distribution of the brightest proto-clusters contributing to the *cold* peaks has a broad maximum at $1.5 \leq z \leq 3$. Therefore follow-up of *Planck* proto-cluster candidates will provide key information on the high-z evolution of large scale structure.

Key words: galaxies: clusters: general – galaxies: evolution – large-scale structure of Universe – submillimetre: galaxies.

1 INTRODUCTION

Observations of the early evolution of large scale structure are important tools for testing galaxy formation theories, investigating galaxy evolution in dense environments and getting independent constraints on fundamental cosmological parameters (e.g. Harrison & Coles 2012; Casey 2016). This has motivated several observational efforts to identify high-redshift proto-clusters of galaxies (for a review see Overzier 2016) defined as high-z galaxy overdensities that will evolve into the local galaxy clusters.

The detection of these objects, however, proved to be challenging. Above $z \simeq 1.5$ the searches via the X-ray emission or the Sunyaev–Zeldovich (SZ) effect as well as the optical detection of the galaxy

⋆ E-mail: NegrelloM@cardiff.ac.uk

red sequence are inefficient, not only because of the difficulty of high-z observations but even more because these methods are appropriate to detect evolved clusters with mature galaxy populations or the signature of a hot intercluster medium (ICM). But in high-z proto-clusters the ICM may not have yet reached the high temperatures necessary for X-ray or SZ detections, and member galaxies may be still in the active star-formation phase.

Indeed, (sub-)millimeter surveys with *Herschel*, Submillimetre Common-User Bolometer Array-2 (SCUBA-2) or APEX LABOCA proved to be capable of detecting candidate high-z proto-clusters of star-bursting galaxies (Valtchanov et al. 2013; Rigby et al. 2014; Dannerbauer et al. 2014; Casey et al. 2015; Ma et al. 2015; Clements et al. 2016). These objects are however rare and therefore very large area surveys are necessary to get a good statistics. As first pointed out by Negrello et al. (2005) low-resolution sub-mm surveys, such as those carried out by the *Planck* satellite, are very well suited for detecting proto-clusters because they integrate over the emission of *all* star-forming member galaxies, not only of those above some detection limit.

The Planck Collaboration XXXIX (2016, PCXXXIX hereafter) provided a catalogue of 2151 cold sub-mm sources with spectral energy distributions peaking between 353 and 857 GHz at 5arcmin resolution, detected by *Planck* in the cleanest 26 per cent of the sky, with flux density at 545 GHz above 500 mJy. This is referred to as the *Planck* high-z, or PHz, sample. The vast majority of the sources in the sample appear as overdensities of dusty star-forming galaxies, having colours consistent with $z \geq 2$ and were considered as proto-cluster candidates.

Planck Collaboration XXVII (2015) used the Spectral and Photometric Imaging Receiver (SPIRE) on the *Herschel* satellite to perform follow-up observations of 234 *cold Planck* sources, satisfying the criterion of having their rest-frame far-infrared (FIR) peak redshifted to the frequency range $353 - 857$ GHz, as expected for galaxies in the redshift range $z = 2$–4. Usable results were obtained for 228 of them, including 83 of the 203 sources within the area covered by the PHz sample. About 94 per cent of the SPIRE sources associated with the 228 *Planck* objects were found to be consistent with being galaxy overdensities; their (photometric) redshift distribution peaks at $z \simeq 2$ for a typical dust temperature of 35 K. Seven sources (about 3 per cent) are candidate lensed systems; all of them have spectroscopic redshifts and are at $z > 2.2$ (Cañameras et al. 2015).

As pointed out by PCXXXIX the nature of the high-z galaxy overdensities is still uncertain. Given the large uncertainties in the photometric redshift estimates, it is impossible, at present, to establish whether galaxies making up the overdensities are physically related or are just part of random fluctuations in the galaxy number density within the *Planck* resolution element.

PCXXXIX found that the distribution of the total (8–1000 μm) infrared (IR) luminosities of overdensities peaks around $2 \times 10^{14}\,\mathrm{L}_{\odot}$ for their reference dust temperature of 35 K. Using the calibration by Kennicutt & Evans (2012), the corresponding star formation rate (SFR) is $\simeq 3 \times 10^{4}\,\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$. The associated halo mass, $M_{\rm h}$, can be estimated using the SFR–$M_{\rm h}$ relation at $z = 2$ derived by Aversa et al. (2015) exploiting the abundance matching technique. For mean SFRs in the range 30–1000 $\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$ (the typical SFR of $z \simeq 2$ star-forming galaxies is $\simeq 300\,\mathrm{M}_{\odot}\,\mathrm{yr}^{-1}$, see fig. 9 of Cai et al. 2014) the sum of halo masses of star-forming galaxies is $M_{\rm h,sf} \simeq 5 \times 10^{14}\,\mathrm{M}_{\odot}$, independently of the SFR. If the lifetime of the star-forming phase is $t_{\rm sf}$, the *total* halo mass is $M_{\rm h} \simeq M_{\rm h,sf}(t_{\rm u}/t_{\rm sf})$, where $t_{\rm u} \simeq 3.3$ Gyr is the age of the universe at $z \simeq 2$. Since $t_{\rm sf}$ is likely <1 Gyr, $M_{\rm h}$ is probably several times larger than $M_{\rm h,sf}$. The redshift-dependent halo mass function by Sheth & Tormen (1999) gives a surface density of haloes with mass larger than $M_{\rm h,sf}$ at $z > 2$ of $\sim 5 \times 10^{-5}$ deg^{-2} (and much lower than that for $M_{\rm h}$, since the density sinks down exponentially in this mass range) to be compared with the surface density of *Planck* overdensities of 0.21 deg^{-2}. This already highlights a problem with the interpretation of overdensities as proto-clusters.

To look for a plausible explanation of the results by PCXXXIX we have updated the study by Negrello et al. (2005) who worked out a formalism to derive the IR luminosity functions and number counts of proto-clusters of dusty galaxies. In the last decade the amount of data on the cosmological evolution of the IR luminosity function and on the clustering properties of dusty galaxies has grown enormously, allowing us to put our estimates on a much more solid basis than what was possible in 2005. We will adopt as our reference the latest Cai et al. (2013) version of the self-regulated galaxy evolution model developed by Granato et al. (2004) and Lapi et al. (2006, 2011). However we checked that our results and conclusions do not significantly change if we adopt an empirical fit of the luminosity function of dusty galaxies (e.g. Mancuso et al. 2016).

The paper is organized as follows. In Section 2, we review and update the formalism introduced by Negrello et al. (2005), while in Section 3 we briefly describe the model adopted for the luminosity function and the clustering properties of dusty galaxies. Predictions for the number counts of proto-clusters and comparison with data are presented in Section 4. In Section 5, we investigate the effect on the estimated number counts of Poisson fluctuations of the number of proto-clusters within the *Planck* beam. Section 6 summarizes and discusses our main conclusions.

Throughout the paper we assume a flat cold dark matter cosmology with $\Omega_{0,\,m} = 0.315$, $\Omega_{0,\,b} = 0.044$, $h = 0.67$ and $\sigma_8 = 0.81$, with an index $n = 1$ for the power spectrum of primordial density fluctuations, consistent with the *Planck* results (Planck Collaboration XIII 2016).

2 FORMALISM

The mean luminosity of a clump of *clustered* galaxies or of a *proto-cluster*, at redshift z is[1] (Peebles 1980)

$$\bar{L}_{\rm cl}(z) = \int_{\mathcal{L}} \frac{\mathrm{d}N(L',z)}{\mathrm{d}L'\mathrm{d}V} L'\mathrm{d}L' \int_{V_{\rm cl}} \xi(r,z)\mathrm{d}V, \qquad (1)$$

where $\mathrm{d}N(L,z)/\mathrm{d}L\mathrm{d}V$ is the comoving luminosity function of galaxies and $\xi(r,z)$ is their spatial correlation function; $V_{\rm cl}$ is the volume of the proto-cluster (the choice of the proto-cluster radius will be discussed in Section 3), while r is the comoving radial distance to the volume element $\mathrm{d}V$. The variance of the clump luminosity is

$$\sigma^2_{L_{\rm cl}}(z) = \int_{\mathcal{L}} \frac{\mathrm{d}N(L',z)}{\mathrm{d}L'\mathrm{d}V} L'^2\mathrm{d}L' \int_{V_{\rm cl}} \xi(r,z)\,\mathrm{d}V + \left[\int_{\mathcal{L}} \frac{\mathrm{d}N(L',z)}{\mathrm{d}L'\mathrm{d}V} L'\mathrm{d}L'\right]^2 \int\int_{V_{\rm cl}} [\zeta(r_1,r_2,z) + \xi(r_{12},z) - \xi(r_1,z)\xi(r_2,z)]\mathrm{d}V_1\mathrm{d}V_2, \qquad (2)$$

where r_{12} is the comoving distance between the volume elements $\mathrm{d}V_1$ and $\mathrm{d}V_2$, and ζ is the reduced part of the three-point spatial

[1] At variance with Negrello et al. (2005) we have now dropped from the equations all the Poisson terms, as we are specifically interested in *physically related* galaxies.

correlation function for which we assume the standard hierarchical formula:

$$\zeta(r_1, r_2) = Q(z)\left[\xi(r_1)\xi(r_2) + \xi(r_1)\xi(r_{12}) + \xi(r_{12})\xi(r_2)\right]. \quad (3)$$

The variance of clump luminosities [equation (2)] is the sum of the contributions from fluctuations in the total luminosity and in the number of member sources.

Using equation (3) we rewrite equation (2) as

$$\begin{aligned}\sigma^2_{\rm cl}(z) = &\int_{\mathcal{L}} \frac{{\rm d}N(L', z)}{{\rm d}L'{\rm d}V} L'^2 {\rm d}L' \int_V \xi(r, z){\rm d}V \\ &+ \left[\int_{\mathcal{L}} \frac{{\rm d}N(L', z)}{{\rm d}L'{\rm d}V} L'{\rm d}L'\right]^2 \\ &[Q(z) - 1]\left(\int_V \xi(r, z){\rm d}V\right)^2 \\ &+ \left[\int_{\mathcal{L}} \frac{{\rm d}N(L', z)}{{\rm d}L'{\rm d}V} L'{\rm d}L'\right]^2 \\ &2Q(z)\int\int_V \xi(r_{12}, z)\xi(r_1){\rm d}V_1{\rm d}V_2 \\ &+ \left[\int_{\mathcal{L}} \frac{{\rm d}N(L', z)}{{\rm d}L'{\rm d}V} L'{\rm d}L'\right]^2 \\ &\int\int_V \xi(r_{12}, z){\rm d}V_1{\rm d}V_2. \end{aligned} \quad (4)$$

The statistics of the matter-density distribution found in N-body simulations are successfully reproduced by a log-normal function (Coles & Jones 1991; Kofman et al. 1994; Taylor & Watts 2000; Kayo, Taruya & Suto 2001; Taruya, Hamana & Kayo 2003), not only in weakly non-linear regimes, but also up to density contrasts $\delta \approx 100$. If light is a (biased) tracer of mass, fluctuations in the luminosity density should reflect the statistics of the matter-density field. Therefore we adopt, following N05, a log-normal shape for the distribution of $L_{\rm cl}$. This function is completely determined by its first (mean) and second (variance) moments:

$$\frac{{\rm d}N(L_{\rm cl}, z)}{{\rm d}L_{\rm cl}{\rm d}V} = \frac{\exp[-\frac{1}{2}[\ln L_{\rm cl} - \mu(z)]^2/\sigma^2(z)]}{\sqrt{2\pi\sigma^2(z)}L_{\rm cl}}, \quad (5)$$

where

$$\mu(z) = \ln\left[\frac{\bar{L}^2_{\rm cl}(z)}{\sqrt{\sigma^2_{L_{\rm cl}}(z) + \bar{L}^2_{\rm cl}(z)}}\right], \quad (6)$$

$$\sigma^2(z) = \ln\left[\frac{\sigma^2_{L_{\rm cl}}(z)}{\bar{L}^2_{\rm cl}(z)} + 1\right]. \quad (7)$$

The function is normalized by requiring the conservation of the luminosity density

$$\int_{\mathcal{L}_{\rm cl}} \frac{{\rm d}N(L'_{\rm cl}, z)}{{\rm d}L'_{\rm cl}{\rm d}V} L'_{\rm cl}{\rm d}L'_{\rm cl} = \int_{\mathcal{L}} \frac{{\rm d}N(L', z)}{{\rm d}L'{\rm d}V} L'{\rm d}L'. \quad (8)$$

3 ADOPTED MODEL FOR DUSTY GALAXIES

We compute the clump total-infrared ($8-1000\,\mu$m) luminosity function using the redshift-dependent IR luminosity function of dusty galaxies provided by the Cai et al. (2013, C13 hereafter) model which accurately reproduces the observational determinations, available up to $z \simeq 4$ (see Fig. 1 of Bonato et al. 2014),

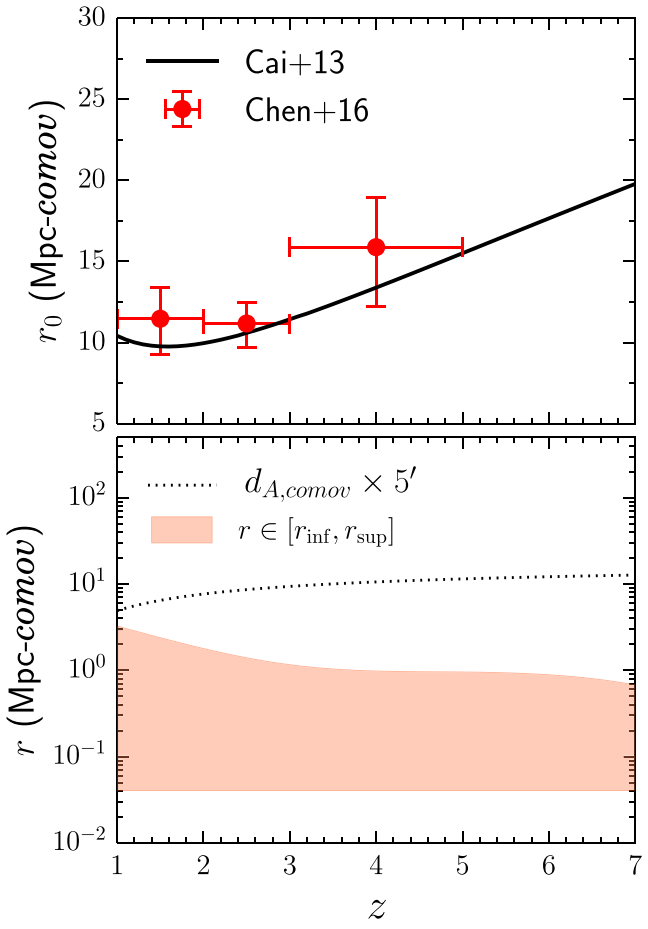

Figure 1. Upper panel: spatial correlation length of proto-spheroids estimated by Cai et al. (2013; solid line) compared to the observational determination by Chen et al. (2016) from a sample of near-infrared selected sub-millimeter galaxies (red dots). Lower panel: lower and upper limits on the value of r adopted for integration in equations (1) and (4). The dotted curve shows, for comparison, the comoving scale corresponding to an angular separation of 5 arcmin, approximately the FWHM of the *Planck* beam at 545 GHz.

as well as a broad variety of other relevant data.[2] In order to assess the robustness of our results we also work out predictions based on a parametric fit to the infrared luminosity function of dusty galaxies. Specifically, we use the results by Mancuso et al. (2016) who preformed a fit to the global, redshift-dependent, SFR function of galaxies measured at UV and FIR wavelengths by the *Hubble* and the *Herschel* space telescopes. As we are interested in the dust-obscured phase of the formation of galaxy clusters at $z \gtrsim 1$, we just need the part of the global SFR function above SFR $\gtrsim 50\,{\rm M}_\odot\,{\rm yr}^{-1}$ [corresponding to luminosities $L_{\rm SFR} \gtrsim few \times 10^{11}\,{\rm L}_\odot$ (Kennicutt & Evans 2012)], where, generally, the star formation is heavily obscured by dust. In practise we identify the value of the SFR at which the global SFR function is a factor of 10 higher than the SFR function derived by Mancuso et al. (2016) from UV (dust-corrected) data alone and apply an exponential cut-off

[2] A tabulation of the IR luminosity function at several redshifts as well as figures showing fits to other relevant data sets are available at http://people.sissa.it/zcai/galaxy_agn/ or http://staff.ustc.edu.cn/zcai/galaxy_agn/index.html.

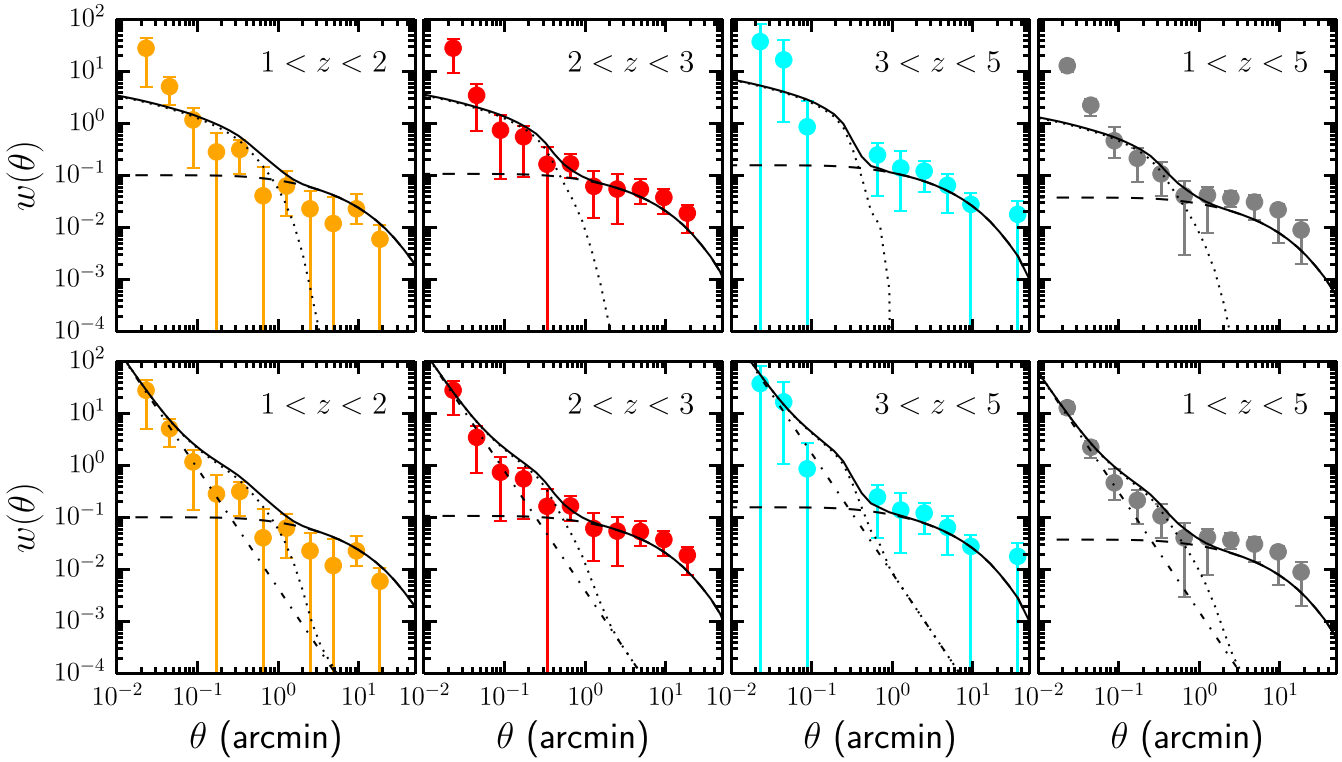

Figure 2. Angular correlation function of faint sub-millimeter galaxies measured by Chen et al. (2016; data points) in different photometric redshift bins. Top panels: comparison with the theoretical angular correlation function derived from the model of Cai et al. (2013; solid black curve) without any fine tuning of the clustering parameters. The 1- and 2-halo terms are shown by the dotted and the dashed curves, respectively. Bottom panels: same as above but with an additional contribution to the 1-halo term, modelled as a power law [$\xi(r) \propto r^{-\gamma}$, with $\gamma = 3.3$, dot–dashed curve; see text].

below that value. The cut-off correponds to SFR $\sim 60\,\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ to $120\,\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ from $z = 2$ to $z = 5$, or to infrared luminosities $L_{\rm IR} \sim 4 \times 10^{11}\,\mathrm{L}_\odot$ to $\sim 8 \times 10^{11}\,\mathrm{L}_\odot$, respectively.

The two-point correlation function of dusty galaxies was computed using the halo occupation distribution (HOD) formalism. The power spectrum of the galaxy distribution is parametrized as the sum of two terms: a 1-halo term, dominating on small scales, that accounts for the distribution of galaxies within the same halo; and a 2-halo term, that describes the correlations amongst galaxies residing on different haloes and, therefore dominates on larger scales. The HOD provides a statistical description of how galaxies populate dark matter haloes and we model it using the central-satellite formalism (see e.g. Zheng et al. 2005). This assumes that the first galaxy to be hosted by a halo lies at its centre, while the remaining galaxies are distributed in proportion to the halo mass profile and are classified as satellites. The mean halo occupation function of satellite galaxies is parametrized as: $\langle N_{\rm sat}\rangle \propto (M_{\rm vir}/M_{\rm sat})^{\alpha_{\rm sat}}$, where $M_{\rm vir}$ is the halo mass and the power-law index $\alpha_{\rm sat}$ is a free parameter. As for the 2-halo term, the key parameter, determining the amplitude of the effective bias function $b_{\rm eff}(z)$, is the minimum halo mass, $M_{\rm vir,\,min}$.

We have adopted the HOD parameter values derived by C13 by fitting the power spectra of the cosmic infrared background (CIB) measured by *Planck* and *Herschel*. The spatial correlation function, $\xi(r, z)$, was then computed as the Fourier antitransform of the 3D power spectrum. We refer the reader to Xia et al. (2012) for all the relevant details on the formalism used. The clustering radius, or correlation length, r_0, defined by $\xi(r_0, z) = 1$, is found to be, in comoving units, $r_0 \simeq 10 - 20$ Mpc in the redshift range $z \sim 1 - 7$. It is shown by the solid line in the upper panel of Fig. 1.

Recently, Chen et al. (2016b) have measured the correlation length of a sample of ∼3000 sub-millimetre galaxies, with redshifts $z \sim 1{-}5$ and star formation rates $\gtrsim 60{-}100\,\mathrm{M}_\odot\ \mathrm{yr}^{-1}$, identified using a new colour selection technique, which combines three optical-near-infrared colours (Chen et al. 2016a). Their estimates of r_0 in 3 redshift intervals ($1 < z < 2$, $2 < z < 3$ and $3 < z < 5$) are shown in the same figure and are in good agreement with the results of C13.

In Fig. 2, we show the angular correlation function measured by Chen et al. (2016b) in the 3 redshift slices and for the full sample ($1 < z < 5$). The prediction of the C13 model [computed using the redshift distribution of the Chen et al. (2016b) sample], without any tuning of the HOD parameters, is also shown for comparison (dashed curve for the 2-halo term; dotted curve for the 1-halo term). The agreement is remarkably good on angular scales $\theta \gtrsim 0.1$arcmin; however, on the smallest scales the model significantly underestimate the clustering of sub-mm galaxies. This is not a surprise considering that the constraints derived by C13 were obtained from measurements performed at larger angular scales than those probed by Chen et al. It is beyond the scope of this paper to perform a detailed analysis of the signal measured by Chen et al. within the HOD formalism. Therefore, in order to achieve a good agreement with the data we add an extra contribution to the 1-halo term below $\theta = 0.2$ arcmin, that we model as a power law, $\xi(r) \propto r^{-\gamma}$. The fit to the data gives $\gamma = 3.3 \pm 0.2$. The extra term is shown by the dot–dashed curve in the lower panels of Fig. 2.

We adopt the transition scale between the 2- and the 1-halo terms as the radius of the proto-cluster, $r_{\rm cl}$. As illustrated by Fig. 2, the corresponding angular scale ranges from $\simeq$1 to $\simeq$0.4 arcmin for z varying from $\simeq$1 to $\simeq$5. Hence, the *Planck* resolution is not ideal

for detecting individual proto-clusters. Higher resolution surveys, such as those planned with the cosmic origins explorer (CORE), will be far more efficient for this purpose (De Zotti et al. 2016).

The additional power-law contribution to $w(\theta)$ dominates on scales $\lesssim 0.1 - 0.2$ arcmin corresponding to *physical* linear scales of 50–100 kpc. Interestingly, this is the scale of the proto-cluster cores at $z \simeq 2.4$–2.5 detected by Ivison et al. (2013) and Wang et al. (2016).

The integrals over r in equations (1) and (4) are carried out up to $r_{\rm sup} = r_{\rm cl}$, while the lowest value of r is set by the minimum angular scale probed by the Chen et al. measurements, i.e. $\sim$0.025 arcmin (or 1.5 arcsec), which corresponds to a *comoving* length of $\sim 25-64$ kpc for $z = 1 - 7$. In our calculations we set $r_{\rm inf} =$ 40 kpc (comoving), corresponding to $\simeq$13 kpc (physical) at $z \simeq 2$ where the distribution of *Planck* proto-cluster candidates peaks. The values of $r_{\rm inf}$ and $r_{\rm sup}$, as a function of redshift, are shown by the upper and lower limits of the shaded region in lower panel of Fig. 1. In the same figure, the dotted curve represents the scale sampled by the beam of *Planck* at sub-millimeter wavelengths (i.e. 5arcmin). It is clear that such proto-clusters would appear as point-like sources in the *Planck* maps.

For the amplitude Q of the three-point correlation function ζ we assumed $Q(z) = 1/b^2(M_{\rm eff}, z)$. This formula was derived by Szapudi et al. (2001) from N-body simulations and holds on scales $\lesssim$10 Mpc h^{-1}, i.e. on the scales of interest here. We neglect any dependence of Q on the linear scale. At variance with N05 we do not consider a model with $Q = 1/b$ suggested by perturbation theory (Fry & Gaztanaga 1993) as this holds on larger scales, i.e. in the limit of small density fluctuations. We also discard the unrealistic model corresponding to $Q(z) = 1$, as this holds for a pure dark-matter distribution, i.e. not for visible galaxies (e.g. Szapudi, Meiksin & Nichol 1996).

We note that since $Q(z)$ decreases with increasing redshift, the contribution to $\sigma^2_{\rm cl}$ from the second term in the right-hand side of equation (2) may become negative. When this happens we set such contribution to zero.

4 RESULTS AND COMPARISON WITH OBSERVATIONS

The search of proto clusters at high redshift have been carried out using many different approaches and several of these structures have already been found. We provide below a summary of some of the most significant findings.

By exploiting the wealth of data covering the entire accessible electromagnetic spectrum in the 2 deg^2 Cosmological Evolution Survey (COSMOS; Scoville et al. 2007) field, Capak et al. (2011) searched for $z > 4$ proto-clusters finding one at $z = 5.30$. This primordial cluster is characterized by an overdensity of galaxies within a region of $\sim$2 Mpc (comoving) radius around the extreme starburst galaxy COSMOS AzTEC-3 and has an estimated total (baryonic+dark) mass of $\gtrsim 4 \times 10^{14}$ M$_\odot$.

Ivison et al. (2013) identified at least four intrinsically luminous galaxies at $z = 2.41$ across an $\sim$100 kpc region by studying the CO line properties of the sub-mm brightest sources in the *Herschel* Astrophysical Terahertz Large Area Survey (*H*-ATLAS; Eales et al. 2010). Specifically, they focused on a sample of candidate lensed galaxies with $2.1 \lesssim z \lesssim 3.5$, extracted from $\sim$330 deg^2 of the *H*-ATLAS (i.e. the Equatorial + North Galactic pole fields), for which measurements of the CO(1-0) emission line were previously obtained with the Green Bank Telescope (Harris et al. 2012). The clump of luminous galaxies discovered by Ivison et al. (2013) was interpreted as the core of an overdensity that will evolve into a massive ($\sim 10^{14.6}$ M$_\odot$) cluster.

Wang et al. (2016) looked instead for concentrations of K_s-band selected red galaxies at $z > 2$ over a 1.62 deg^2 region of the COSMOS field, finding a significant overdensity of massive galaxies at $z = 2.506$ associated with extended X-ray emission. The high star formation rate ($\sim$3400 M$_\odot$ yr^{-1}) measured in the central 80 kpc region of the overdensity suggests that also in this case we are witnessing the rapid build-up of a cluster core.

Clements et al. (2014) examined the *Herschel*/SPIRE images of the 16 Planck Early Release Compact Source Catalog sources lying within the 90 deg^2 of the Herschel Multitiered Extragalactic Survey (HerMES; Oliver et al. 2012), finding four overdensities of sub-mm sources. Using existing multiwavelength photometric data they estimated redshifts in the range $z \sim 0.8-2.3$ and typical star formation rates SFR $>$ 1000 M$_\odot$ yr^{-1} for the proto -clusters.

In Fig. 3 we compare these observations with the predictions based on our model. The measured infrared luminosities of the proto-clusters, listed in Table 1, are presented as lower limits. In fact they only account for cluster members detected above the flux limit of the observations, while a significant, or even dominant, contribution is expected from fainter members. For example, using the Cai et al. (2013) model we find that sources brighter than 50 mJy at 500 μm, approximately the 90 per cent completeness level of *Herschel* observations used by Planck Collaboration XXVII (2015), comprise 29, 43, 65 and 66 per cent of the luminosity obtained integrating the whole luminosity function at $z = 1.5$, 2, 2.5 and 3, respectively. In other words, proto-clusters stand out more clearly in sub-mm maps, such as those that have been provided by *Planck* and will be hopefully provided by CORE, than in much higher resolution point source surveys at the same wavelengths. This is in keeping with the findings by Clements et al. (2014) and Planck Collaboration XXXIX (2016) who noted that the sub-mm flux densities of proto-cluster candidates measured by *Planck* are about two to three times larger than the summed luminosities of member galaxies detected with *Herschel* within the *Planck* beam, although part of the difference is to be attributed to the 'flux boosting' affecting the low signal-to-noise *Planck* measurements (sources are preferentially detected if they are on top of positive confusion plus instrumental noise fluctuations).

Not only the proto-cluster luminosities but also the number densities inferred from the Ivison et al. (2013) and Wang et al. (2016) detections are lower limits. Ivison et al. (2013) only focused on *Herschel* sources with existing CO line measurements, while Wang et al. (2016) based their selection on a K_s-band catalogue, that may miss proto-clusters of heavily dust-enshrouded galaxies. Information on the redshift and infrared luminosity of the proto-clusters in Fig. 3 are summarized in Table 1.

We also show, in the same figure, the predictions based on the infrared luminosity function of Mancuso et al. (2016). For $2 \lesssim z \lesssim 3.5$ the latter predicts $\times 1.5-2$ more proto-clusters than the Cai model. This is because the Mancuso et al. (2016) infrared luminosity function is higher than the Cai et al. (2013) one for $L_{\rm IR} \lesssim 10^{12}$ L$_\odot$ and $z \gtrsim 2$, which leads to higher values of the mean and variance of the clump luminosity. However, the difference between the two models is still consistent with the uncertainties in the measured luminosity function of dusty galaxies at those luminosities and redshifts (see fig. 1 of Mancuso et al. 2016).

Having checked that our model yields results consistent with observations, that refer to the redshift range probed by the sample

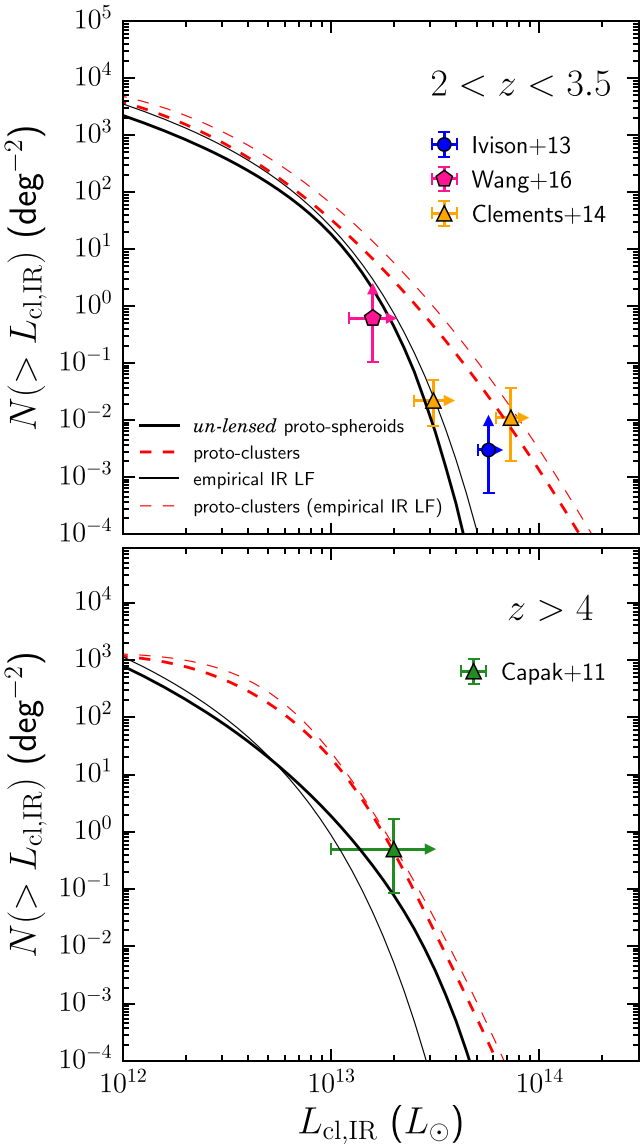

Figure 3. Cumulative infrared luminosity function of proto-clusters with redshifts in the range $2 < z < 3.5$ (upper panel) and with $z > 4$ (lower panel). The predictions, based on the Cai et al. model are shown by the thick dashed red curve while, for comparison, the abundance of un-lensed proto-spheroids is shown by the thick solid black curve. The excess density of proto-clusters compared to unlensed proto-spheroids is compensated by a deficit at faint luminosities (not shown): the luminosity density is conserved (equation 8). The lighter curves are predictions based on the infrared luminosity function of Mancuso et al. (2016) for dusty galaxies (solid back curve) and proto-clusters (dashed red curve). The data points mark the constraints on the abundance of dusty proto-clusters set by the discoveries of such objects reported in Capak et al. (2011), Ivison et al. (2013), Clements et al. (2014) and Wang et al. (2016). The luminosities of detected proto-clusters are the sum of those of member galaxies above the detection limits. Hence, they do not include the contribution of fainter member galaxies, whose summed emission may be dominant (see text).

of PCXXXIX, we compare, in Fig. 4, our predictions (dashed red curve; thicker for the Cai et al. model and lighter for the Mancuso et al. (2016) empirical luminosity function) with the counts at 545 GHz (550 μm) of the PH*z* sources, reported in the latter paper (filled blue circles). The number counts at a given wavelength are derived from the IR luminosity function of the clumps [equation (5)] by adopting the spectral energy distribution (SED) of the $z = 2.3$ star forming galaxy SMM J2135-0102 (Swinbank et al. 2010) for the proto-spheroids, as done by Cai et al. (2013).

Table 1. Information about the proto-clusters at redshift $z > 2$ in Fig. 3. The quoted infrared luminosities are those derived by summing the contributions from all the detected star forming galaxy members. Therefore they do not account for the emission due to unresolved/undetected cluster members.

z	$L_{\rm IR}$ ($\times 10^{12}$ L$_\odot$)	Reference
2.05 ± 0.09	31 ± 6	Clements et al. (2014)
2.27 ± 0.12	73 ± 11	Clements et al. (2014)
2.41	57 ± 6	Ivison et al. (2013)
2.506	15.8	Wang et al. (2016)
5.30	17 ± 8	Capak et al. (2011)

Apart from the small fraction of strongly lensed galaxies (around 3 per cent; Planck Collaboration XXVII 2015), these candidate high-*z* sources may be either overdensities of bright star-forming galaxies (i.e. proto-clusters) or high peaks of the cosmic infrared background (CIB) fluctuations which have red colours because they are contributed by physically unrelated high-*z* galaxies.

Figure 4 also shows, for comparison, the differential number counts of strongly lensed galaxies. Such counts were computed using the formalism of Lapi et al. (2012, their SISSA model) with a maximum amplification $\mu_{\rm max} = 15$ that is found by Negrello et al. (2017) to reproduce the counts of strongly lensed galaxies in the *H*-ATLAS. The number densities of the PH*z* sample are well in excess of those expected from our model, also when using the empirical luminosity function of Mancuso et al. (2016), and, more importantly, of the number densities of high-*z* haloes capable of hosting them, as pointed out in Section 1. In the next Section we discuss a plausible explanation of the discrepancy.

5 THE EFFECT OF SOURCE CONFUSION

Thanks to the very-low instrumental noise of the *Planck* High Frequency Instrument (HFI), the fluctuation field measured in the HFI maps is signal-dominated. In regions with low Galactic dust content, clustering of high-*z* sources making up the CIB is the dominant source of intensity fluctuations, at sub-mm wavelengths, on the several arcmin scales of interest here. The probability distribution function (PDF) of such intensity fluctuations is highly skewed, with an extended tail towards high signal-to-noise ratios (cf. fig. 10 of De Zotti et al. 2015).

To investigate the nature of high signal-to-noise intensity peaks we have resorted to Monte Carlo simulations. Proto-clusters have been randomly distributed in infrared luminosity and redshift over an area of 5000 deg^2, according to the modelled luminosity functions. The area was divided into 1.7×1.7 arcmin pixels[3] and the flux densities of proto-clusters within each pixel were summed up. The map was then convolved with the *Planck* beam and background subtracted. The source extraction was performed by looking for connected pixels with signal-to-noise ratio (SNR)[4] SNR ≥ 5 as in PCXXXIX. A couple of examples of the content of simulated intensity peaks with SNR ≥ 5 are shown in Fig. 6. One or two very luminous proto-clusters are generally present; however, most of the flux density within the *Planck* beam is accounted for by much

[3] 1.7 arcmin corresponds to 1/3 of the *Planck* Full Width at Half Maximum (FWHM) at 545 GHz.
[4] A detection requires at least one pixel above the SNR threshold.

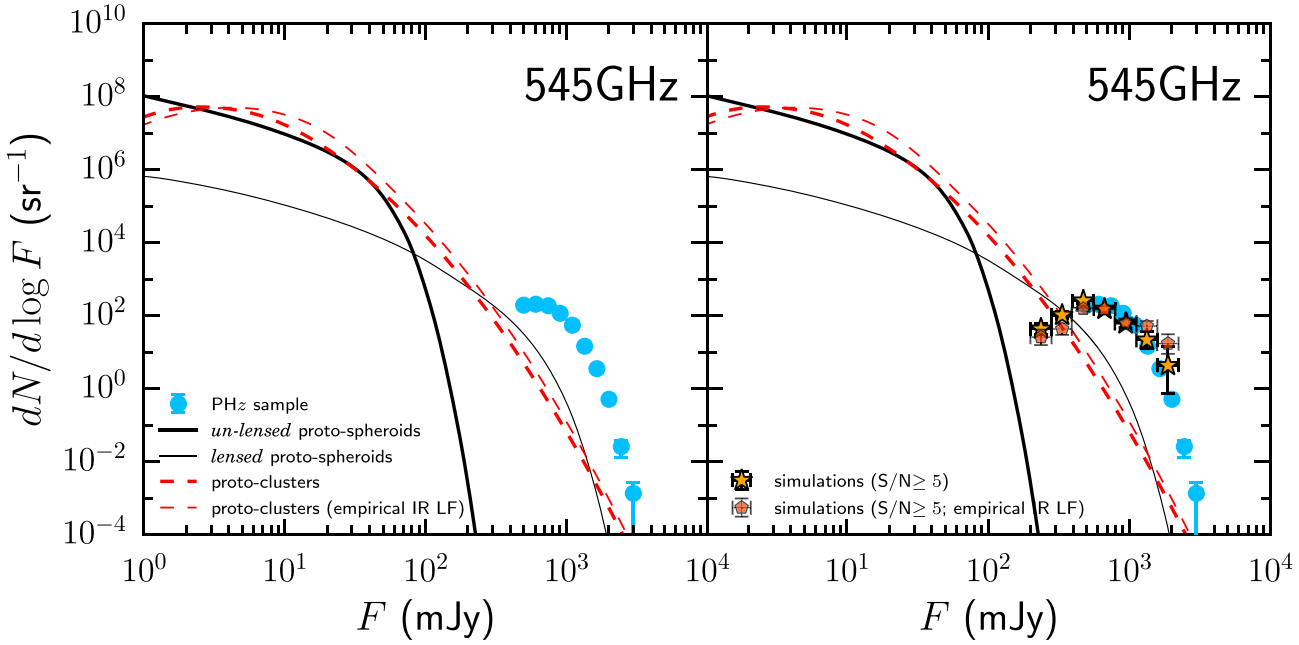

Figure 4. Predicted differential number counts of proto-clusters at 545 GHz (550 μm) derived from the Cai et al. model (2013; thick dashed red curve) and the infrared luminosity function of Mancuso et al. (2016; light dashed red curve). Data points are for the PHz sample (PCXXXIX; blue circles). For comparison, the number counts of both un-lensed and strongly lensed proto-spheroidal galaxies, from the Cai et al. model, are shown by the thick and the thin solid black curves, respectively. On the right-hand panel the *Planck* results are compared with the number counts of sources detected with signal-to-noise ratio SNR ≥ 5 in the simulated map (yellow stars for the Cai model and orange pentagons for the empirical luminosity function of Mancuso et al. 2016).

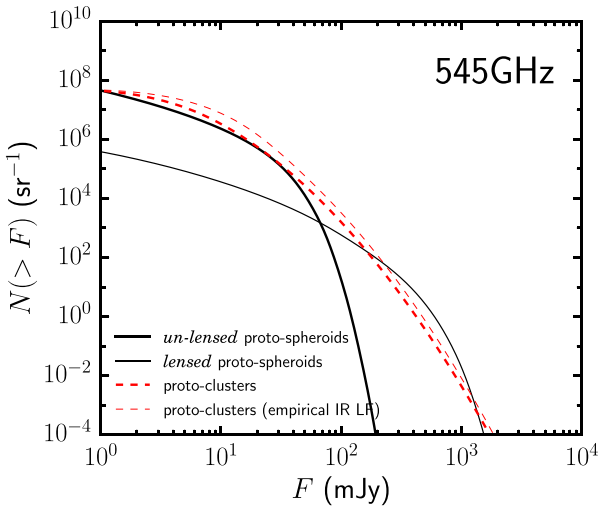

Figure 5. Predicted integral number counts of proto-clusters at 545 GHz (550 μm; dashed red curves). The meaning of the curves is the same as in Fig. 4.

fainter objects, as implied by the steepness of the bright portion of the proto-cluster counts (Figs 4 and 5).

The source flux density is obtained by first fitting a Gaussian to a 20 × 20 arcmin postage stamp centred on the detected object and then integrating the best-fitting model. The differential number counts measured in the simulated map are shown in the right-hand panel of Fig. 4 by the yellow stars. The effect of source confusion is dramatic and brings the number counts in line with the measured ones (blue dots). The same conclusion is reached when we use the Mancuso et al. (2016) empirical luminosity function for dusty galaxies, as illustrated by the orange pentagons in the same figure.

The model integral number counts of *individual* high-z ($z \gtrsim 2$) proto-clusters, displayed in Fig. 5, show that those brighter than the flux limit of the PHz sample (500 mJy at 545 GHz), hence detectable by *Planck*, are expected to be very rare. The CORE project can do much better. Its estimated 4σ detection limits, at the nearby frequency of 520 GHz, range from 141 mJy for the 1 m telescope option to 82.5 mJy for a 1.5 m telescope (De Zotti et al. 2016). From Fig. 5 the corresponding surface densities range from 230 sr^{-1} to 3300 sr^{-1}.

The yellow histograms in Fig. 7 show the distributions of the FWHMs and of the ellipticities, ϵ, of sources detected in simulated maps with SNR≥5 and $F_{545\mathrm{GHz}} \geq 500$ mJy [very similar results are obtained using the empirical luminosity function of Mancuso et al. (2016)]. The distributions are qualitatively consistent with those of the PHz sample (solid histograms, taken from fig. 14 of PCXXXIX), suggesting that the elongated shape of the detected sources is an effect of source confusion rather than being an intrinsic property of the proto-clusters. Note that a close agreement between the results of our simulations and the observed distributions is not to be expected because simulations cannot reproduce exactly the selection criteria adopted by PCXXXIX (simultaneous detection within a 5arcmin radius in the 545 GHz excess map, with SNR > 5, and in the 857, 545, and 353 GHz cleaned maps with SNR > 3; absence, at 100 GHz, of any local maximum with SNR > 3 within a radius of 5arcmin of the 545 GHz position; colour-colour selection) and their procedures for determining sizes and ellipticities.

Finally, in Fig. 8 we show the redshift distribution of the brightest proto-clusters associated with the sources detected in the simulated map with SNR ≥ 5 and $F_{545\,\mathrm{GHz}} \geq 500$ mJy. For comparison, we also show the photometric redshift distribution (obtained assuming a modified blackbody spectrum with dust temperature $T = 35$ K

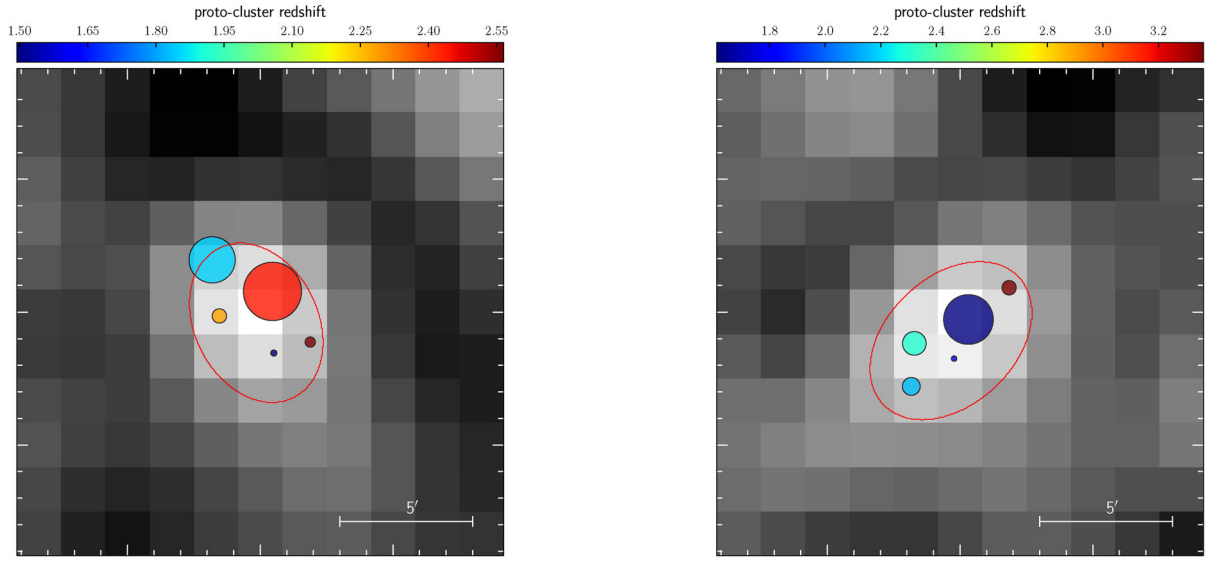

Figure 6. Examples of sources detected in the 545 GHz simulated map (black and white image in the background). The red ellipsis is the Gaussian fit to the detection, while the spots mark the five brightest proto-clusters within the ellipsis, scaled in size according to their flux density (i.e. the brighter the clump, the bigger the spot).

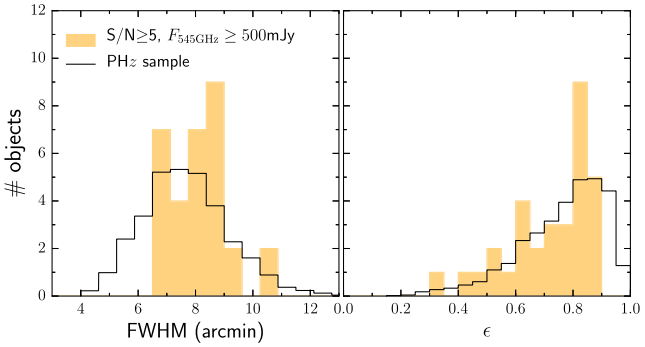

Figure 7. Distribution of FWHMs and ellipticities, ϵ, derived from the simulations for detections with SNR ≥ 5 and $F_{545\mathrm{GHz}} \geq 500$ mJy (yellow histograms), compared to those measured for the PHz sample, normalized to the number of sources in the simulation (solid line). The FWHM is defined as $(\mathrm{FWHM_{major}} \times \mathrm{FWHM_{minor}})^{1/2}$ where $\mathrm{FWHM_{major}}$ and $\mathrm{FWHM_{minor}}$ are the major and the minor FWHM derived from the Gaussin fit, while the ellipticity is defined as $\epsilon = [1 - (\mathrm{FWHM_{minor}}/\mathrm{FWHM_{major}})^2]^{1/2}$. Note that the simulations cannot reproduce accurately the selection criteria of the PHz sample, nor the procedures for measuring the FWHM and the ellipticity. Hence a precise match between the model and the observed distributions cannot be expected (see text).

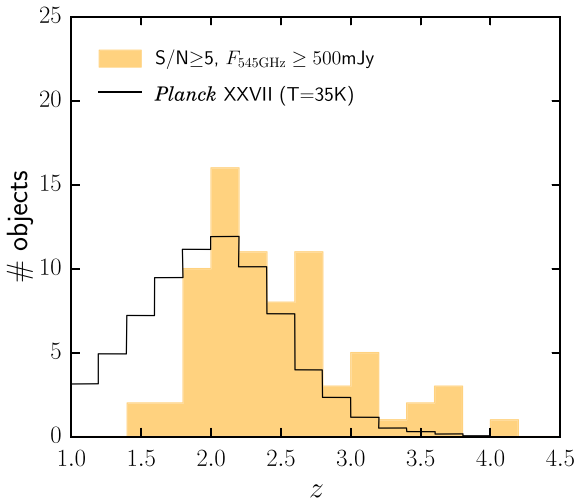

Figure 8. Redshift distribution of the brightest proto-clusters associated with the sources detected in the simulation with SNR ≥ 5 and $F_{545\mathrm{GHz}} \geq 500$ mJy (yellow histogram). For comparison we show the photometric redshift distribution of the *Planck* candidate proto-clusters (Planck Collaboration XXVII 2015) derived from the associated *Herschel*/SPIRE sources (black line) assuming a modified blackbody spectrum with dust temperature $T = 35$ K and dust emissivity index $\beta = 1.5$.

and dust emissivity index $\beta = 1.5$) of the *Herschel*/SPIRE detected sources associated with the *Planck* candidate proto-clusters (Planck Collaboration XXVII 2015, black histogram in the same figure). Both histograms peak in the redshift range $z = 1.5-3$ showing the great potential offered by the *Planck* catalogue for studying the early phases of clusters formation.

6 CONCLUSIONS

The PHz catalogue appears to be dominated by overdensities of star-forming galaxies plus a small fraction of strongly gravitationally lensed galaxies (around 3 per cent; Planck Collaboration XXVII 2015). Motivated by these results, we have updated the N05 predictions for the number counts of proto-clusters of star-forming galaxies and compared them with observational estimates.

We stress that our results are largely model-independent. The basic ingredients of our calculations are the sub-mm luminosity functions of high-z galaxies and their spatial correlation function. The luminosity functions have been observationally determined up to $z \simeq 4$, based on *Herschel* survey data (Gruppioni et al. 2013). The model we have used reproduces very well the observational determinations at each redshift (Cai et al. 2013; Bonato et al. 2014). Furthermore, we have shown that the results do not change significantly if the model is replaced by an empirical fit of the observed luminosity function of dusty galaxies (Mancuso et al. 2016). As

shown by Fig. 2, our model also provides an accurate representation of the observed redshift dependent correlation functions.

We find that the counts of sub-mm bright proto-clusters, obtained in the framework of the standard Λ cold dark matter cosmology, are well below the observational estimates by PCXXXIX. A similar conclusion can be inferred from the results of Granato et al. (2015) who also predict, by using a completely different approach, a number of proto-clusters well below the one suggested by the statistics of *Planck* detections.

We have shown that this basically follows from the fact that, at high z, there are not enough sufficiently massive haloes. On the other hand we have shown, by means of simulations, that the *Planck cold* intensity peaks are fully consistent with being mostly random fluctuations in the number of unrelated proto-clusters at $z \geq 2$ within the *Planck* beam. This is highlighted in Fig. 6 where examples of sources detected in the simulation are shown together with the brightest proto-clusters they are comprised of. The counts of these fluctuations match very well with those of *cold* peaks and there is qualitative agreement also with the distributions of their FWHM's and of their ellipticities. Interestingly, the only *Planck* overdensity for which spectroscopic or photometric redshifts of member galaxies have been obtained, was found to consist of two physically unrelated structures at $z \sim 1.7$ and $z \sim 2$ (Flores-Cacho et al. 2016).

The redshift distribution of the brightest proto-clusters contributing to the *cold* peaks, given by the model, has a broad maximum between $z = 1.5$ and $z = 3$. Therefore, follow-up observations of galaxies within the *Planck* overdensities providing redshift estimates would be a powerful tool to investigate the early phases of cluster formation, inaccessible by other means.

ACKNOWLEDGEMENTS

We thank the referee for helpful suggestions. MN acknowledges financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 707601. GDZ acknowledges financial support by ASI/INAF agreement n. 2014-024-R.1 for the *Planck* LFI activity of Phase E2. JGN acknowledges financial support from the Spanish MINECO for a 'Ramon y Cajal' fellowship (RYC-2013-13256) and the I+D 2015 project AYA2015-65887-P (MINECO/FEDER)

REFERENCES

Aversa R., Lapi A., de Zotti G., Shankar F., Danese L., 2015, ApJ, 810, 74
Bonato M. et al., 2014, MNRAS, 438, 2547
Cai Z.-Y. et al., 2013, ApJ, 768, 21
Cai Z.-Y., Lapi A., Bressan A., De Zotti G., Negrello M., Danese L., 2014, ApJ, 785, 65
Cañameras R. et al., 2015, A&A, 581, A105
Capak P. L. et al., 2011, Nature, 470, 233
Casey C. M., 2016, ApJ, 824, 36
Casey C. M. et al., 2015, ApJ, 808, L33
Chen C. C. et al., 2016a, ApJ, 820, 82
Chen C. C. et al., 2016b, ApJ, 831, 91
Clements D. L. et al., 2014, MNRAS, 439, 1193
Clements D. L. et al., 2016, MNRAS, 461, 1719
Coles P., Jones B., 1991, MNRAS, 248, 1
Dannerbauer H. et al., 2014, A&A, 570, A55
De Zotti G. et al., 2015, JCAP, 6, 018
De Zotti G. et al., 2016, JCAP, preprint (arXiv:1609.07263)
Eales S. et al., 2010, PASP, 122, 499
Flores-Cacho I. et al., 2016, A&A, 585, A54
Fry J. N., Gaztanaga E., 1993, ApJ, 413, 447
Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580
Granato G. L., Ragone-Figueroa C., Domínguez-Tenreiro R., Obreja A., Borgani S., De Lucia G., Murante G., 2015, MNRAS, 450, 1320
Gruppioni C. et al., 2013, MNRAS, 432, 23
Harris A. et al., 2012, ApJ, 752, 152
Harrison I., Coles P., 2012, MNRAS, 421, L19
Ivison R. et al., 2013, ApJ, 772, 137
Kayo I., Taruya A., Suto Y., 2001, ApJ, 561, 22
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kofman L., Bertschinger E., Gelb J. M., Nusser A., Dekel A., 1994, ApJ, 420, 44
Lapi A. et al., 2006, ApJ, 650, 42
Lapi A., Shankar F., Mao J., Granato G. L., Silva L., De Zotti G., Danese L., 2011, ApJ, 742, 24
Lapi A., Negrello M., Gonzalez-Nuevo J., Cai Z.-Y., De Zotti G., Danese L., 2012, ApJ, 755, 46
Ma C.-J. et al., 2015, ApJ, 806, 257
Mancuso C., Lapi A., Shi J., Cai Z.-Y., Gonzalez-Nuevo J., Béthermin M., Danese L., 2016, ApJ, 833, 152
Negrello M., Gonzalez-Nuevo J., Magliocchetti M., Moscardini L., De Zotti G., Toffolatti L., Danese L., 2005, MNRAS, 358, 869
Negrello M. et al., 2017, MNRAS, 465, 3558
Oliver et al., 2012, MNRAS, 424, 1614
Overzier R. A., 2016, A&AR, 24, 14
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Planck Collaboration XIII, 2016, A&A, 594, A13
Planck Collaboration XXVII, 2015, A&A, 582, A30
Planck Collaboration XXXIX, 2016, A&A, 596, A100
Rigby E. E. et al., 2014, MNRAS, 437, 1882
Scoville N. Z. et al., 2007, ApJ, 172, 1
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Swinbank A. M. et al., 2010, Nature, 464, 733
Szapudi I., Meiksin A., Nichol R. C., 1996, ApJ, 473, 15
Szapudi I., Postman M., Lauer T. R., Oegerle W., 2001, ApJ, 548, 114
Taruya A., Hamana T., Kayo I., 2003, MNRAS, 339, 495
Taylor A. N., Watts P. I. R., 2000, MNRAS, 314, 92
Valtchanov I. et al., 2013, MNRAS, 436, 2505
Wang T. et al., 2016, ApJ, 828, 56
Xia J.-Q., Negrello M., Lapi A., De Zotti G., Danese L., Viel M., 2012, MNRAS, 422, 1324
Zheng Z. et al., 2005, ApJ, 633, 791

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