# Properties of submillimetre galaxies in a semi-analytic model using the 'Count Matching' approach: application to the ECDF-S

Alejandra M. Muñoz Arancibia,[1][*] Felipe P. Navarrete,[2,3] Nelson D. Padilla,[1,4] Sofía A. Cora,[5,6] Eric Gawiser,[7] Peter Kurczynski[7] and Andrés N. Ruiz[6,8,9]

[1]*Instituto de Astrofísica, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860 Santiago, Chile*
[2]*Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany*
[3]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*
[4]*Centro de Astro-Ingeniería, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860 Santiago, Chile*
[5]*Instituto de Astrofísica de La Plata (CCT La Plata, CONICET, UNLP), Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque s/n, B1900FWA La Plata, Argentina*
[6]*Consejo Nacional de Investigaciones Científicas y Técnicas, Rivadavia 1917, C1033AAJ Buenos Aires, Argentina*
[7]*Department of Physics and Astronomy, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA*
[8]*Instituto de Astronomía Teórica y Experimental (IATE, CCT Córdoba, CONICET, UNC), Laprida 854, X5000BGR Córdoba, Argentina*
[9]*Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000GBR Córdoba, Argentina*

## ABSTRACT

We present a new technique for modelling submillimetre galaxies (SMGs): the 'Count Matching' approach. Using light cones drawn from a semi-analytic model of galaxy formation, we choose physical galaxy properties given by the model as proxies for their submillimetre luminosities, assuming a monotonic relationship. As recent interferometric observations of the *Extended Chandra Deep Field*-South show that the brightest sources detected by single-dish telescopes are comprised by emission from multiple fainter sources, we assign the submillimetre fluxes so that the combined Large APEX BOlometer CAmera (LABOCA) plus bright-end Atacama Large Millimeter/submillimeter Array observed number counts for this field are reproduced. After turning the model catalogues given by the proxies into submillimetre maps, we perform a source extraction to include the effects of the observational process on the recovered counts and galaxy properties. We find that for all proxies, there are lines of sight giving counts consistent with those derived from LABOCA observations, even for input sources with randomized positions in the simulated map. Comparing the recovered redshift, stellar mass and host halo mass distributions for model SMGs with observational data, we find that the best among the proposed proxies is that in which the submillimetre luminosity increases monotonically with the product between dust mass and star formation rate (SFR). This proxy naturally reproduces a positive trend between SFR and bolometric IR luminosity. The majority of components of blended sources are spatially unassociated.

**Key words:** galaxies: evolution – galaxies: formation – early Universe – submillimetre: galaxies.

## 1 INTRODUCTION

Fitting submillimetre galaxies (SMGs) into the current theory of galaxy formation has been a challenge since their discovery. They are the most luminous star-forming sources at the epoch where star formation peaks, being detected by their redshifted far-infrared (FIR) emission from cold dust in the submillimetre (submm) wavebands. SMGs were detected for the very first time in the *Hubble Deep Field* at 850 µm (Barger et al. 1998; Hughes et al. 1998) using the Submillimetre Common-User Bolometer Array on the James Clerk Maxwell Telescope (JCMT/SCUBA; Holland et al. 1999). The advantage of observing SMGs is their negative *k*-correction: the observed flux coming from star-forming regions remains nearly constant with redshift at these wavelengths, so they can be detected out to $z \sim 10$ (Blain et al. 2002).

Many observational efforts have been done with the attempt of mapping the global cosmic star formation history and the build-up of the overall stellar mass content of the Universe through SMGs, as they are useful tracers of the obscured star formation at high

[*] E-mail: amma.19@gmail.com

redshifts, complementing the cosmic census carried on at shorter wavelengths (Blain et al. 1999). These include surveys using single-dish telescopes (Coppin et al. 2006; Weiss et al. 2009; Chen et al. 2013, just to name a few) and interferometric observations, both in the continuum and targeting emission lines (Tacconi et al. 2006; Smolčić et al. 2012; Hatsukade et al. 2013; Hodge et al. 2013; Weiss et al. 2013; Chen et al. 2014, among others).

Identification of counterparts at other wavelengths for different samples of SMGs has been carried out, either through deep optical/near-infrared (NIR) spectroscopy or deep intermediate/high-resolution radio imaging. This allows the spectroscopic follow-up of SMGs, as well as multiwavelength studies. The largest sample of SMGs with spectroscopic follow-up to date was obtained by Chapman et al. (2005), who measured a median spectroscopic redshift of 2.2 for 73 SMGs belonging to several fields, having a median 850 μm flux density of 5.7 mJy. Moreover, using multiwavelength data Wardlow et al. (2011) derived a median photometric redshift of $2.2 \pm 0.1$ for 78 single-dish detected SMGs in the *Extended Chandra Deep Field*-South (ECDF-S) over ∼4 mJy at 870 μm. More recent studies have been successful identifying counterparts of single-dish detected SMGs using interferometers at millimetre (mm) wavebands. For instance, Smolčić et al. (2012) performed interferometric observations at 1.3 mm of SMGs in the central region of the Cosmic Evolution Survey (COSMOS) field, which had been discovered through single-dish observations at 870 μm down to a signal-to-noise ratio (S/N) of 3.8; complementing this with additional multiwavelength catalogues, they derived a mean photometric redshift of $2.6 \pm 0.4$ for 16 of these SMGs.

Having redshift information and multiwavelength photometry allows the performance of spectral energy distribution (SED) fitting, and then deriving other SMG properties as stellar mass and star formation rate (SFR). From several studies exploring these properties (Smail et al. 2004; Borys et al. 2005; Michałowski et al. 2010; Wardlow et al. 2011 and others), there is now some consensus regarding the quite high SFRs of single-dish detected SMGs, having several hundreds and even thousands of $M_{\odot}$ yr$^{-1}$ and hosting significant stellar populations. These findings place them among the most powerful starburst galaxies in the Universe. Additionally, making these sources evolve through models, it has been found that SMGs are likely the progenitors of local luminous early-type galaxies (Smail et al. 2004; Swinbank et al. 2006; Wardlow et al. 2011).

Samples of SMGs are commonly modelled through hydrodynamical simulations (Dekel et al. 2009; Davé et al. 2010) and semi-analytic models (Baugh et al. 2005; Swinbank et al. 2008; Somerville et al. 2012) within a Λ cold dark matter (ΛCDM) cosmology. Since the emission at FIR wavelengths for these sources comes from ultraviolet (UV) light dust-reprocessing, a key process that needs to be addressed by these models is how the dust present in the galaxy absorbs and re-radiates this light. While some of the models relate the parameters controlling the dust attenuation with other galaxy properties and constrain their values with local observations, others make use of radiative transfer codes (Silva et al. 1998; Jonsson 2006; Noll et al. 2009, among others) to compute the FIR luminosity in a self-consistent way. This last method can be very time-consuming when the sample to be modelled comprises thousands or even millions of galaxies, requiring supercomputing facilities.

There is still a controversy between the modelled and observed abundance of SMGs. The lack of enough sources recovered by adopting a Kennicutt (1983) initial mass function (IMF) reported by Granato et al. (2000) was possibly resolved by Baugh et al. (2005), claiming the need of a flat IMF in starbursts, while still reproducing basic properties of the local galaxy population. This was motivated by the increase both in the total UV light radiated per unit mass of stars and in the yield of metals from core-collapse supernovae, which in turn produces more dust to absorb it. However, the assumption subsequently led to underpredict the *K*-band magnitudes for model SMGs (Swinbank et al. 2008). Like these works, other studies were able to predict the abundance of SMGs via the change in a given assumption or process, but failing in the agreement with other observational statistics like their redshift distribution (Fontanot et al. 2007).

Another possible explanation for the discrepancy between observed and modelled counts is that the low resolution achieved with single-dish telescopes masks multiple sources within a beam. This influence on the number of detected sources can be solved through a follow-up of single-dish detected SMGs with (sub)mm interferometry. Hayward et al. (2013b) and Cowley et al. (2014) have discussed the effects of the beamsize on the counts, using models that give the submm emission of galaxies in a self-consistent way.

This work is motivated by recent 870 μm continuum Atacama Large Millimeter/submillimeter Array (ALMA) observations of the ECDF-S, which was previously surveyed using the Large APEX BOlometer CAmera (LABOCA) on the Atacama Pathfinder EXperiment (APEX) telescope (Siringo et al. 2009). Karim et al. (2013) show that the brightest sources detected in the LABOCA ECDF-S Submillimeter Survey (LESS; Weiss et al. 2009) are comprised by emission from multiple fainter sources, namely the ALMA LESS sources (ALESS; Hodge et al. 2013). With the aim of exploring the properties of SMGs in this field, and inspired by the now-standard abundance matching technique (Conroy et al. 2006), we perform a 'Count Matching' approach using light cones drawn from a semi-analytic model of galaxy formation and evolution. We choose various physical galaxy properties given by the model as proxies for their submm luminosities, assuming a monotonic relationship so that the combined LABOCA plus bright-end ALMA observed number counts are reproduced. After turning the catalogues of galaxy positions and fluxes given by the different proxies into submm maps, we perform a source extraction. With this we study the effects of the observational process on the recovered counts, as well as the galaxy properties derived from the detected sources. For finding the best proxy, we explore the redshift, stellar mass and host halo mass distributions. Once the best proxy is determined, several properties for SMGs (as well as for their descendants) can be predicted.

This paper is organized as follows. Section 2 gives a brief explanation of the semi-analytic model used. In Section 3, we detail the procedure for constructing light cones of galaxies, while in Section 4 the Count Matching technique is described. Section 5 outlines the observational process for constructing the submm maps of simulated galaxies. In Section 6 results are discussed, ending with a summary in Section 7.

## 2 GALAXY FORMATION MODEL

We use a combination of a cosmological *N*-body simulation of the concordance ΛCDM universe and the semi-analytic model of galaxy formation SAG, acronym for Semi-Analytic Galaxies. Details about the semi-analytic model are given in Cora (2006), Lagos, Cora & Padilla (2008) and Tecce et al. (2010).

The cosmological simulation was run using the GADGET-2 code (Springel 2005). It gives the dark matter haloes and their embedded substructures, in a periodic box of $150 \, h^{-1}$ Mpc a side, with $640^3$ particles having a mass resolution of $1 \times 10^9 \, h^{-1} \, M_{\odot}$; 100 snapshots giving information for haloes at each epoch are collected,

being equally spaced in logarithm of the scale factor between $z = 20$ and 0. From those outputs, merger trees are constructed and are then used by SAG to generate the galaxy population. The cosmological parameter values assumed are $\Omega_m = 0.28$, $\Omega_\Lambda = 0.72$, $\Omega_b = 0.046$, $H_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with $h = 0.7$ and $\sigma_8 = 0.81$. These parameters are consistent with WMAP7 data (Jarosik et al. 2011). Below we present a big picture of the SAG model, referring the reader to the papers cited above for a deeper insight on the physical processes included.

Hot gas in dark matter haloes is isothermally distributed, being drawn from the assumed baryon fraction. It cools according to a cooling rate, forming a galaxy disc that follows an exponential distribution. Stars form at a rate that depends on the amount of cold gas that gives origin to them, besides dark matter halo properties (virial radii and velocities). Cold gas inflows (Dekel et al. 2009) are not included in the current version of the SAG model.

Stars can form quiescently as well as in starbursts. Starbursts can be triggered by mergers (both major and minor, defined according to the relative masses of the galaxies involved) and disc instabilities, exhausting all the available cold gas instantaneously and contributing to the bulge growth. A merger of two dark matter haloes leads to the merging of the galaxies within them. When a merger occurs, satellite galaxies are accreted by a central one after keeping a circular orbit around it, decaying via dynamical friction. Furthermore, a disc becomes unstable when it is massive enough to be dominated by self-gravity, and therefore sensitive to small external perturbations. Both mergers and disc instabilities, besides cold gas accretion, contribute to the growth of a central black hole.

The stellar mass created in each star formation event gives a distinctive number of core-collapse supernovae according to the adopted IMF, which in this model is Salpeter (1955). These supernovae in turn will reheat the cold gas through galactic outflows, transferring it to the hot gas phase. Apart from this kind of feedback, the model considers the heating produced by the active galactic nucleus (AGN) through black hole accretion as responsible of the gas cooling suppression.

Star formation events lead to metal pollution in galaxies, yielded via mass-losses through stellar winds in stars having low or intermediate mass, as well as supernovae explosions. The chemical enrichment of the gas affects cooling rates and so star formation.

The degree of efficiency, as well as the fulfillment of several criteria involved in the different physical processes, are regulated by free parameters. These are tuned in order to reproduce a number of $z = 0$ observables (see Appendix A, where also some of the galaxy properties predicted by the model and not used as constraints are presented). After assuring this, those parameters remain fixed throughout this work, such that none of the values used is dependent on SMG properties (neither from the model itself nor from observational ones).

## 3 CONSTRUCTING LIGHT CONES

We use the SAG semi-analytic model within the $\Lambda$CDM framework to follow the history of galaxies, obtaining information about them at different epochs which are given by the chosen output redshifts for the model (i.e. the snapshots mentioned in the previous section). We assume that the galaxy sample inside the periodic comoving box of $150\,h^{-1}$ Mpc a side is statistically representative of the overall galaxy population in the Universe, such that we can build a simulated universe repeating (if necessary), in our simulated catalogues, model galaxies at different epochs of their evolution.
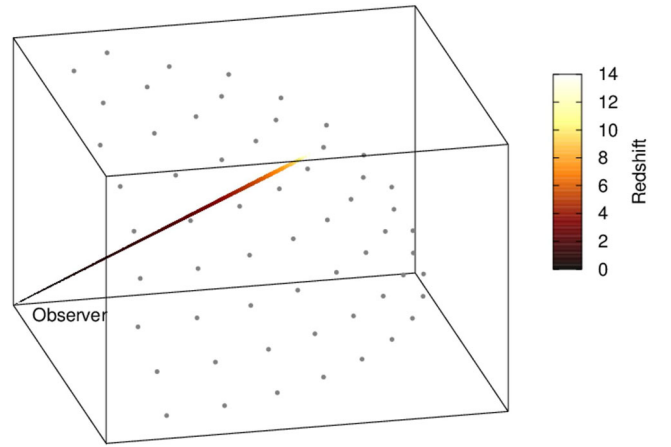


**Figure 1.** Light cone construction scheme. Consider the first octant of a sphere filled with galaxies from the SAG model and centred on the observer. A given interval in RA and Dec (sky plane) and redshift $z$ (line of sight) is chosen, collecting all the model galaxies that fall within that region (dots coloured by $z$). This section of the sphere is mapped taking different orientations (centres indicated by grey points).

For constructing the simulated universe, the observer is placed at the centre of a sphere whose radius extends to the highest redshift where a semi-analytic galaxy exists. The whole spherical volume is populated with galaxies taken from the simulated boxes. Since the redshift separation between two consecutive snapshots is smaller than that corresponding to the side of the box, only one box per epoch is needed in the radial direction.

A light cone consists of all the galaxies that belong to a particular region in the sky and redshift range (see Fig. 1), so the choice of a line of sight and limits in right ascension (RA) and declination (Dec.) around it are needed. These are chosen as follows.[1] Since our aim is to study the counts in the ECDF-S field, we select the same angular area surveyed by LESS, i.e. ~30 arcmin × 30 arcmin. At the redshifts of interest, this area is smaller than the area subtended by the simulation box, so at a given epoch galaxies appear only once in the area covered by the light cone.

For a given line of sight, the repetition of a model galaxy at various redshifts is limited as much as possible, since the repetition of simulated structure might lead to a wrong interpretation of how model galaxies are spatially distributed once we turn them into a map. In order to test the amount of sources appearing in the light cone at more than one epoch, we map the first octant of the sphere in the redshift range $0 < z < 5$ with 58 light cones having 0.3 deg² and orientations with roughly uniform coverage in RA and Dec. Instead of using model galaxies, the periodic boxes are filled with a Cartesian grid of dots; they emulate galaxies, so these dots are labelled for keeping control of their frequency of appearance at different epochs. A catalogue is then created for each light cone containing all the 'grid galaxies' that fall within it, and a quality factor is computed for evaluating the repetition of structures. This factor is defined as

$$Q = \frac{\#\,\text{points appearing more than once}}{\#\,\text{points appearing just once}}, \tag{1}$$

---

[1] In the following, we do not include any characteristic of the real sky in the simulated celestial sphere, so none of the selected orientations need to be corrected for the presence of the Milky Way, closeness to bright stars, etc.
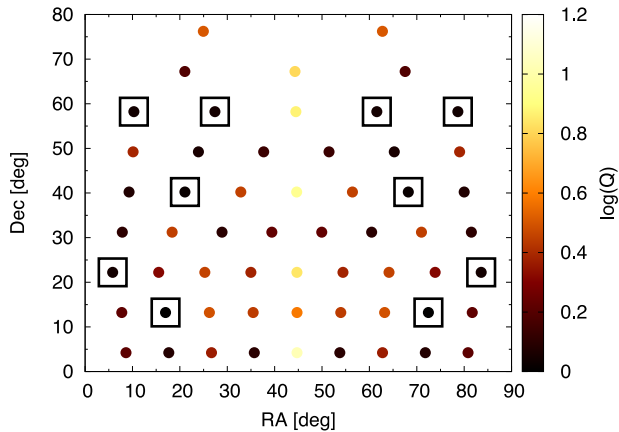
**Figure 2.** Study of the best orientations for the light cones in RA and Dec, according to the $Q$ quality factor (see text). Orientations indicated in Fig. 1 are filled with dots representing galaxies, drawn from a cubic grid in the range $z = 0$–$5$, taking angular areas of 0.3 deg$^2$. For each orientation, $Q$ is computed (colour code gives its logarithm). The $n$ best orientations are such that $Q$ is lowest (black squares). For this paper $n = 10$ light cones were selected.

i.e. the ratio between the amount of sources that appear at more than one epoch and the amount of sources that are not repeated within the light cone. With this definition, a low repetition of structure is expressed as low $Q$. We choose the $n$ best orientations as those with the lowest $Q$. For this paper, $n = 10$ light cones were selected (see Fig. 2). The choice of this number of orientations has the aim of testing the effects of cosmic variance over the results. The 58 orientations tested span a range in $Q$ from 1.14 to 12.66, with the 10 chosen light cones in the range $Q = 1.14$–$1.24$. Through the analysis of galaxy properties presented in this work, we have checked that there are no trends between $Q$ values and the recovered distributions for these 10 light cones.

This technique for choosing the best orientations for light cones was tested only up to $z = 5$. Since we are interested in galaxy redshift distributions of observed SMGs, which peak at $z \sim 2$, this upper limit is adequate for our purposes.

Once a light cone is constructed, the redshifts of galaxies are computed from their coordinates within it. Identifying the particular snapshots of the simulation to which they belong, several galaxy properties given by SAG can be recovered, e.g. stellar masses, SFRs, etc.

## 4 THE COUNT MATCHING TECHNIQUE

Motivated by abundance matching techniques (e.g. Conroy et al. 2006), we propose a new approach for exploring the properties of SMGs, namely the Count Matching technique. Here, a given physical galaxy property (or a combination of several ones) is chosen as a proxy for another property whose numerical value is unknown, assuming a monotonic relationship. In our case, the unknowns are the 870 μm luminosities of a galaxy sample. We use this monotonic relationship to assign a third property, namely, the 870 μm flux density, to the simulated galaxies in such a way that an observational statistics for the latter is reproduced. The chosen observational statistics are the observed galaxy number counts at this wavelength. This allows us to explore several other statistics for simulated galaxies selected using the matched fluxes, like redshifts, masses, luminosities, etc. By comparing them with distributions

derived from observations, the quality of the proposed proxies can be analysed.

The selection of properties used in the proxies is motivated by the theoretical understanding of the mechanism that triggers FIR and submm emission in luminous and ultraluminous infrared galaxies (LIRGs and ULIRGs, respectively; Sanders & Mirabel 1996), with the SMGs being thought of as the high-$z$ cousins of these sources. These galaxies emit the bulk of their bolometric emission in the FIR, corresponding to absorbed UV light by dust that is thermally re-radiated at FIR wavelengths; this UV light comes mostly from young stars, with a minor contribution from AGN.[2] Within this picture, it makes sense to propose that the bright submm fluxes measured for these sources are correlated in some way with the stellar mass, for instance, in a very simplistic model motivated by the strong correlation of many galaxy properties with stellar mass and the observation of large stellar masses for SMGs (e.g. Michałowski et al. 2010). Or with the dust mass, that partially reprocesses the short wavelength light, or with the SFR, as high values give rise to more stars and then more UV emission. Given its low computational cost, the technique offers us the advantage to explore the relationship between several properties (or combinations of them) available from the simulation, and the FIR emission in a galaxy.

We are interested in applying this technique when simulating sources in the ECDF-S, given the well-studied SMGs detected using the LABOCA bolometer (Weiss et al. 2009) and the recently detected sources using the ALMA interferometer at high submm flux densities (Karim et al. 2013).

The steps for this approach are the following:

(i) Select a galaxy sample. We take all the galaxies within each light cone with stellar mass over $10^8 \, \mathrm{M_\odot}$. This lower limit is imposed by the mass resolution of the underlying cosmological simulation; the luminosity functions (LFs) obtained for our model galaxies can be considered complete till that limit (see Ruiz et al. 2014).

(ii) Choose a physical galaxy property as a proxy for the 870 μm rest-frame luminosity, assuming that sources having higher values of the property have higher luminosities. With this we are imposing a monotonic relationship between proxy and luminosity, without placing any restriction about the exact shape of this relation (which can vary depending on the chosen proxy). We test several properties given by SAG, as well as combinations of two or more. The chosen properties are: stellar mass ($M_{\mathrm{stellar}}$), star formation rate (SFR), cold gas mass ($M_{\mathrm{coldgas}}$), disc scalelength ($r_d$), dust mass ($M_{\mathrm{dust}}$), stellar age ($t_{\mathrm{stellar}}$) and cold gas phase metallicity ($\log (\mathrm{O/H}) + 12$). The proxies constructed with these properties are defined in Table 1. The H13 proxy corresponds to equation 15 of Hayward et al. (2013a), i.e. $M_{\mathrm{dust}}^{0.54} \times \mathrm{SFR}^{0.43}$, being the best fit over their simulated galaxies. For the MDA, MDD, MDS and H13 proxies, the dust mass is computed assuming $M_{\mathrm{dust}} \propto M_{\mathrm{coldgas}} \times (\log (\mathrm{O/H}) + 12)$. This is inspired by the assumption of a dust-to-gas mass ratio proportional to the metallicity, adopted by the radiative transfer code GRASIL (Silva et al. 1998).

---

[2] Wang et al. (2013) found a fraction of AGN of $17^{+16}_{-6}$ per cent for ALESS SMGs having a rest-frame 0.5–8 keV absorption-corrected luminosity greater than $7.8 \times 10^{42} \, \mathrm{erg \, s^{-1}}$. Moreover, Laird et al. (2010) found that in at least ∼85 per cent of GOODS-N SMGs the star formation dominates the FIR emission, via the study of their X-ray spectra.

**Table 1.** Quantities used as proxies for the rest-frame submm luminosity.

| Proxy | Definition |
|---|---|
| Stellar mass | $M_{stellar}$ |
| SFR | Star formation rate |
| SFR surface density | $SFR/r_d^2$ (for $r_d > 0$) |
| sSFR | $SFR/M_{stellar}$ |
| Dust mass | $M_{dust}$ |
| MDA | $M_{dust}/t_{stellar}$ |
| MDD | $M_{dust} \times SFR/r_d^2$ (for $r_d > 0$) |
| MDS | $M_{dust} \times SFR$ |
| H13 | $M_{dust}^{0.54} \times SFR^{0.43}$ |

(iii) For a given property or a combination of properties, let the proxy value, adjusted for cosmological distance, be known as

$$y = \frac{\text{Proxy}}{f_k \, D_L^2}, \tag{2}$$

and sort the values for all galaxies in increasing order. Here $D_L$ is the luminosity distance of each galaxy and $f_k$ is a factor giving the $k$-correction corresponding to its redshift (Hogg et al. 2002) at rest-frame 870 µm, defined in terms of redshift and monochromatic luminosities (in the $L_\nu$ formalism) as

$$f_k = \frac{L_{870\,\mu m}}{(1+z)\,L_{870\,\mu m/(1+z)}}. \tag{3}$$

Note that the definition of $y$ introduced above turns it into an analogue to a flux density (despite its arbitrary units, which vary according to the proxy), since the rest-frame 870 µm luminosity can be related to the observed 870 µm flux density through

$$S_{870\,\mu m} = \frac{L_{870\,\mu m}}{4\pi\, f_k\, D_L^2}. \tag{4}$$

The $k$-correction is recovered using a template for Arp 220 (Blain 1999), which is a typical ULIRG. Additionally, we explore the effect of assuming that the negative $k$-correction eliminates completely the diminution of the submm flux due to distance; in this approach, the amount of submm flux assigned to each galaxy only depends on the value of the property or combination of properties selected as proxy.

(iv) Assign a submm flux to each galaxy according to its value of $y$, such that sources with higher values will have higher fluxes. This is the key step in the Count Matching process. Submm fluxes are drawn from a Monte Carlo simulation following the observed cumulative number counts, where we have combined the LABOCA counts at low fluxes (Weiss et al. 2009) and ALMA counts (Karim et al. 2013) over 8 mJy. The choice of ALMA counts only for the bright end is because we want to test whether we are able to recover the LABOCA counts after simulating the observational process (including blending), while avoiding biases in the ALMA counts arising from targeting only bright LESS sources and not other regions in the ECDF-S with S/N slightly lower than the LESS threshold (Karim et al. 2013); we tested the effect of switching counts between ALMA and LABOCA data at different fluxes in the range 7.5–8.5 mJy, which does not lead to significantly different results. We translate this combination of cumulative counts into differential counts, and use one realization of the Monte Carlo simulation as the random fluxes that follow these counts.
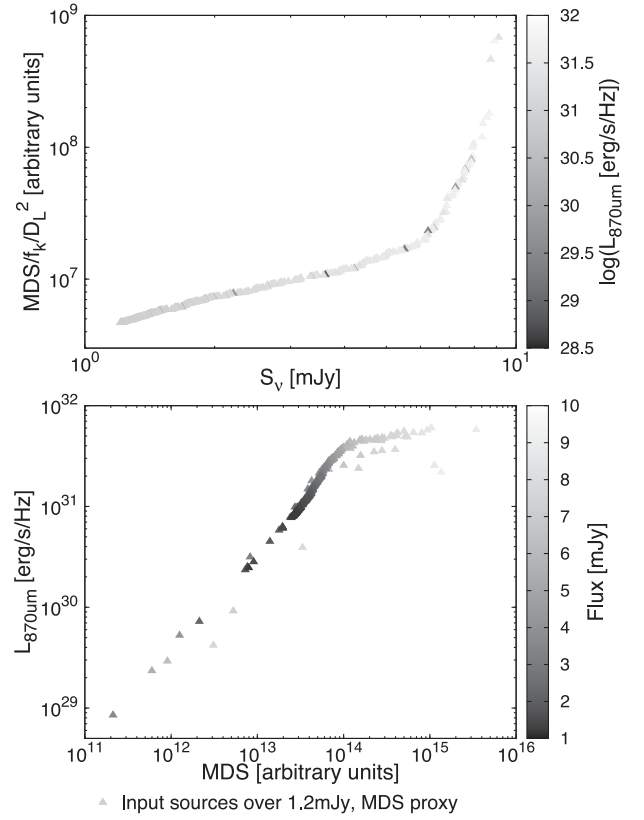


**Figure 3.** The Count Matching process. Top panel: relation between the 870 µm flux and the flux-analogue quantity $y = \text{Proxy}/f_k/D_L^2$ for the MDS proxy. Bottom panel: recovered relation between rest-frame 870 µm luminosity and proxy value. For clarity, this is shown only for sources brighter than 1.2 mJy for a given light cone.

The assignment of submm fluxes for galaxies in the light cone can be summarized as

$$\int_{S_\nu'}^{\infty} n(S_\nu)\mathrm{d}S_\nu = \int_Y^{\infty} n(y)\mathrm{d}y \tag{5}$$

where $S_\nu$ is the flux density at 870 µm and $n(S_\nu)$ is the amount of galaxies having flux densities between $S_\nu$ and $S_\nu + dS_\nu$ (and similarly for $n(y)$). $Y$ and $S_\nu'$ stand for particular values of $y$ and $S_\nu$. Then $Y(S_\nu')$ gives the transformation from flux to proxy, which in turn can be used to recover the numerical value of the submm luminosity for each galaxy.

To illustrate the procedure, in Fig. 3 top panel we show the relation between the 870 µm flux, which comes from the Monte Carlo simulation, and the flux-analogue quantity $y$ for the MDS proxy, which comes from the model (for clarity, this is only shown for sources over 1.2 mJy for a given light cone). Once the 870 µm flux is assigned to a given galaxy, the SED template can be used to recover the numerical value of its rest-frame 870 µm luminosity. The shape of the monotonic relation between this luminosity and the proxy value for MDS is shown in the bottom panel; the trend is robust, despite some outliers having combinations of extreme proxy and redshift values.

Proxies proposed in this work then have simple, one-to-one dependences with rest-frame 870 µm luminosity, for a stellar mass limited sample of model galaxies. No additional selections are included at this stage, except for sources where a proxy has an un-

defined value (e.g. removing galaxies without disc in SFR surface density and MDD proxies). Definitely, these proxies can be refined selecting only a subset of the stellar mass limited sample, for instance removing passive galaxies because of their low SFR; we leave these improvements for a future work.

The advantage of using this simple recipe instead of a full radiative transfer is the considerably lower computational cost, when applied to big samples of galaxies. However, it also has limitations. In the abundance matching formalism, the halo mass function must be correctly predicted by the *N*-body simulation, as well as the stellar mass function should be well derived from the observations. In our approach, we are assuming that the bright-end ALMA and faint-end LABOCA counts are correct, an observable that does not provide redshift information. Because submm luminosities are difficult to model directly, we choose a proxy for the rest-frame submm luminosity for each model galaxy, and use its redshift and an appropriate SED template to predict its submm flux analogue. We then perform an abundance-type matching in the space of observed versus predicted submm fluxes following equation (5). This determines the redshift distribution of the brightest sources (i.e. SMGs) as a function of flux, which in turn influences their clustering and therefore the amount of blending that occurs before bright-end LABOCA counts are measured.

## 5 SIMULATING THE OBSERVATIONAL PROCESS

We turn the catalogues of galaxy positions and fluxes given by the different proxies into submm maps that include a modelling of the observational process.

First, simulated sources are injected in a noise map having a spatial resolution of 19.5 arcsec (as the LABOCA beamwidth). The resulting map is beamsmoothed using a 90 arcsec Gaussian kernel, for removing low spatial frequency structures as is commonly done for observational data (e.g. Weiss et al. 2009). The latter map is then substracted from the former, and the result is convolved with a 19.5 arcsec Gaussian kernel, giving rise to a map resolution of 27.6 arcsec. The beamsmoothing process produces a decrease in the source fluxes by ~8.6 per cent, which are hence rescaled.

One map is constructed for each orientation and proxy. Injected sources span a region of ~30 arcmin × 30 arcmin. They are taken from the simulated light cones, being the *N* brightest sources of each catalogue. We choose $N = 5 \times 10^3$, as it gives source fluxes well below the values reported by observations, while the computing time is reasonable (we checked that a change in one order of magnitude in *N* does not have a significant effect on the statistics recovered from the maps). An example of these simulated maps is given in Fig. 4, indicating the region where model sources are injected as a black square.

We perform a source extraction as done for maps obtained through observations. With this we study the effects of the observational process on the recovered counts, as well as the galaxy properties derived from the counterparts of detected sources for each proxy.

Sources are extracted using SEXTRACTOR (Bertin & Arnouts 1996), for a limit in S/N of 3.8. This cut is chosen following Weiss et al. (2009), according to the expected number of false detections. The source extraction comprises two iterations, where in the second one the sources detected in the first iteration are removed from the map and the noise level is recalculated. We consider even sources extracted outside the limits given by the injected sources.
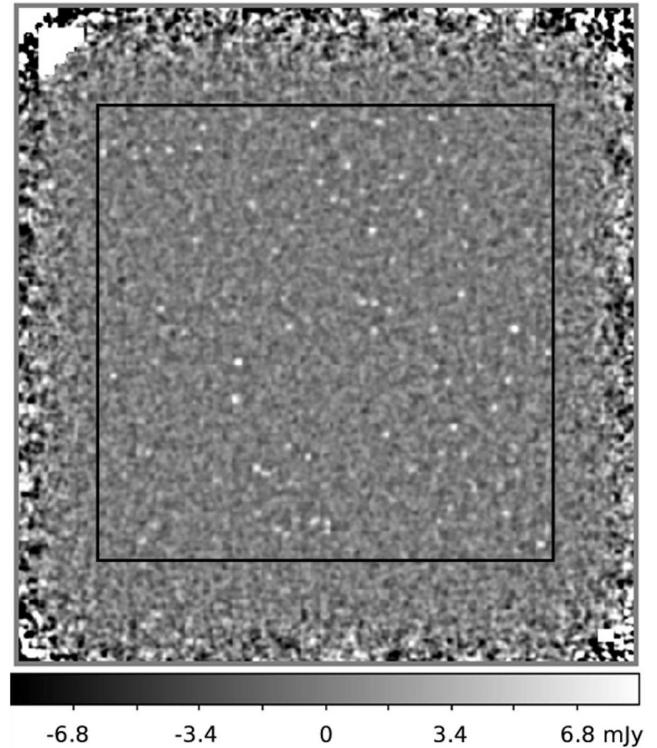


**Figure 4.** Example of a simulated map, corresponding to a given light cone in the MDS proxy. The black square encloses the region where simulated sources are injected (~30 arcmin × 30 arcmin). Grey-scale gives submm flux per pixel. The map resolution is 6 arcsec pix$^{-1}$.

As our resulting source catalogue is limited by S/N, we need to correct the extracted fluxes for boosting effects: the systematic enhancement on the measured fluxes coming from noise introduces a bias in the number of sources exceeding the chosen limit (Scott et al. 2008). For simulated maps, the deboosting correction is found comparing the extracted flux of a given source with the flux of its brightest counterpart in the input catalogue (see Section 6.1) and then using the fitting function $S_{\nu, \text{deboosted}} = (S_{\nu, \text{extracted}} + A)^{1/2}$, with *A* the free parameter. This function is chosen such that it has a cutoff at fluxes near the S/N limit, and tends to the equality at high fluxes. Taking these deboosted fluxes, cumulative submm number counts are obtained for each simulated map, including a proper correction for completeness taken from Weiss et al. (2009).

In order to validate these methods of source extraction and flux deboosting, we apply them to the actual LESS field (Weiss et al. 2009): first, we extract sources from the actual LESS map down to S/N = 3.8 with SEXTRACTOR, and secondly we apply the flux deboosting correction to the extracted sources. Since in this case the map comes from real observations, we do not have an input catalogue for it, so the *A* parameter is found computing the median value among all the simulated maps. The resulting deboosted fluxes and counts are shown in Fig. 5. Compared to Weiss et al. (2009) data, we recover a lower amount of sources in the flux range ~4–8 mJy by a factor of ~1.5. However, this may come from their different deboosting flux technique, which makes use of a *P*(*D*) analysis.

We have tried different parameters for SEXTRACTOR. In a first extraction with conservative parameters (DETECT_MINAREA=5 plus defaults) and only one iteration, we obtained ~20 per cent less sources than reported by Weiss et al. (2009). Trying with a grid of parameters in DETECT_MINAREA, and also modifying the parameters related to the deblending of sources (DEBLEND_NTHRESH
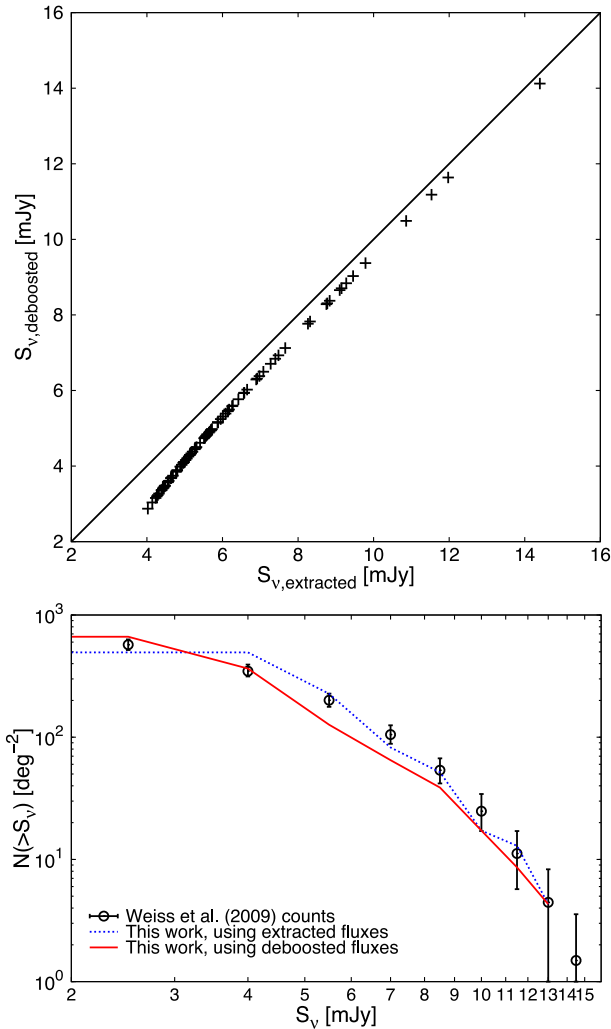
**Figure 5.** Source extraction and flux deboosting methods used in this work, applied to the actual LESS map (Weiss et al. 2009). Top panel: comparison between extracted and deboosted fluxes for our detected sources. Bottom panel: cumulative number counts recovered in each case, compared to actual data for this field (Weiss et al. 2009).

and DEBLEND_MINCONT), we increased the matching with the original catalogue. Finally, after extracting the sources from the first iteration, we re-run SEXTRACTOR on the residual map. This procedure assures us good agreement with the catalogue from Weiss et al. (2009), thus minimizing differences in the final counts arising from different detection methods.

# 6 RESULTS AND DISCUSSION

We are interested in those proxies where the simulated sources that follow the observed LABOCA plus bright-end ALMA counts, after going through the observational process, give (1) counts comparable to LABOCA data, (2) redshift distributions consistent with observational values, (3) other properties in agreement with observations (including clustering, stellar mass, host halo mass, etc). For the last two requirements, we compare model distributions with different surveys from the literature.

Note that it is possible that the properties are affected by the environment of ECDF-S, as this field appears to be underdense when compared to other deep fields probed with submm wavebands (see

Weiss et al. 2009 discussion, but also Chen et al. 2013, who find no difference with other fields but for a smaller area). However, we find that there are only minor changes in the proxy versus luminosity at 870 μm relation (see Fig. 3) between light cones of different density. In particular, this relation changes only by ∼10 per cent between the different light cones for the MDS proxy.

## 6.1 Number counts

Fig. 6 shows the cumulative submm number counts obtained for the different proxies, compared to the input number counts (i.e. before passing through the observational process) and to those extracted from LABOCA and ALMA observations. Compared to the observed counts, sources extracted from the simulated maps give counts closer to LABOCA data.

In order to quantify the goodness of each proxy at this step, we compare their $\chi^2$ values given by

$$\chi^2 = \sum_{S_\nu} \frac{[(\mathrm{d}N/\mathrm{d}S)_{\mathrm{LABOCA}} - \overline{(\mathrm{d}N/\mathrm{d}S)}]^2}{[\delta(\mathrm{d}N/\mathrm{d}S)_{\mathrm{LABOCA}}]^2 + [\delta\overline{(\mathrm{d}N/\mathrm{d}S)}]^2}, \tag{6}$$

where $(\mathrm{d}N/\mathrm{d}S)_{\mathrm{LABOCA}}$ corresponds to the observed LABOCA differential counts at a given flux density, $\overline{(\mathrm{d}N/\mathrm{d}S)}$ gives our mean differential counts over the ten light cones, and $\delta(\mathrm{d}N/\mathrm{d}S)$, $\delta\overline{(\mathrm{d}N/\mathrm{d}S)}$ values are the reported uncertainties in LABOCA counts and scatter in our simulated counts, respectively. Since sources injected in the maps follow the combined bright-ALMA plus faint-LABOCA counts by construction, this $\chi^2$ definition allow us to quantify the influence coming from (a) the observational process (blending, noise, etc) and (b) clustering given by the model (as different proxies give different spatial distribution for bright sources).

The reduced $\chi^2$ is presented for all proxies in Table 2, second column. There, as well as in Fig. 6, we also show the results for the case where the spatial distribution of input sources along each of the 10 maps is random, i.e. does not come from any proxy and so no clustering is provided (note that the number of sources injected in each map is the same as used for the proxies, as well as their submm flux distribution). All proxies are quite similar, so this criterion is not enough to choose a proxy as the best. A further analysis of the predicted distribution of other galaxy properties may discriminate better between proxies. Even the counts obtained for input sources with randomized positions in the simulated map are consistent with those where the Count Matching process was applied (which include realistic galaxy clustering provided by the semi-analytic model). This finding indicates that the clustering has a minor influence when determining the cumulative submm number counts.

For recovering the properties of model galaxies, we perform a cross-match between all the extracted sources and the input galaxies having submm fluxes down to 1.2 mJy (hereafter the input catalogue). A search radius of 13.8 arcsec is used, corresponding to the beam radius after the map processing (see Section 5). When more than one input galaxy falls within the search radius around a given extracted source, we refer to it as a multiple source composed by blended galaxies; otherwise, we call it a single source.

Fig. 7 illustrates this with a small region within the map shown in Fig. 4. Note that among the extracted sources, besides singles and multiples there can be also sources that do not have a counterpart in the input catalogue. Although most of these non-matched sources have S/N ratios around the chosen limit for extraction and can be considered as spurious sources whose flux is boosted by the map noise, some of them have S/N over 5.
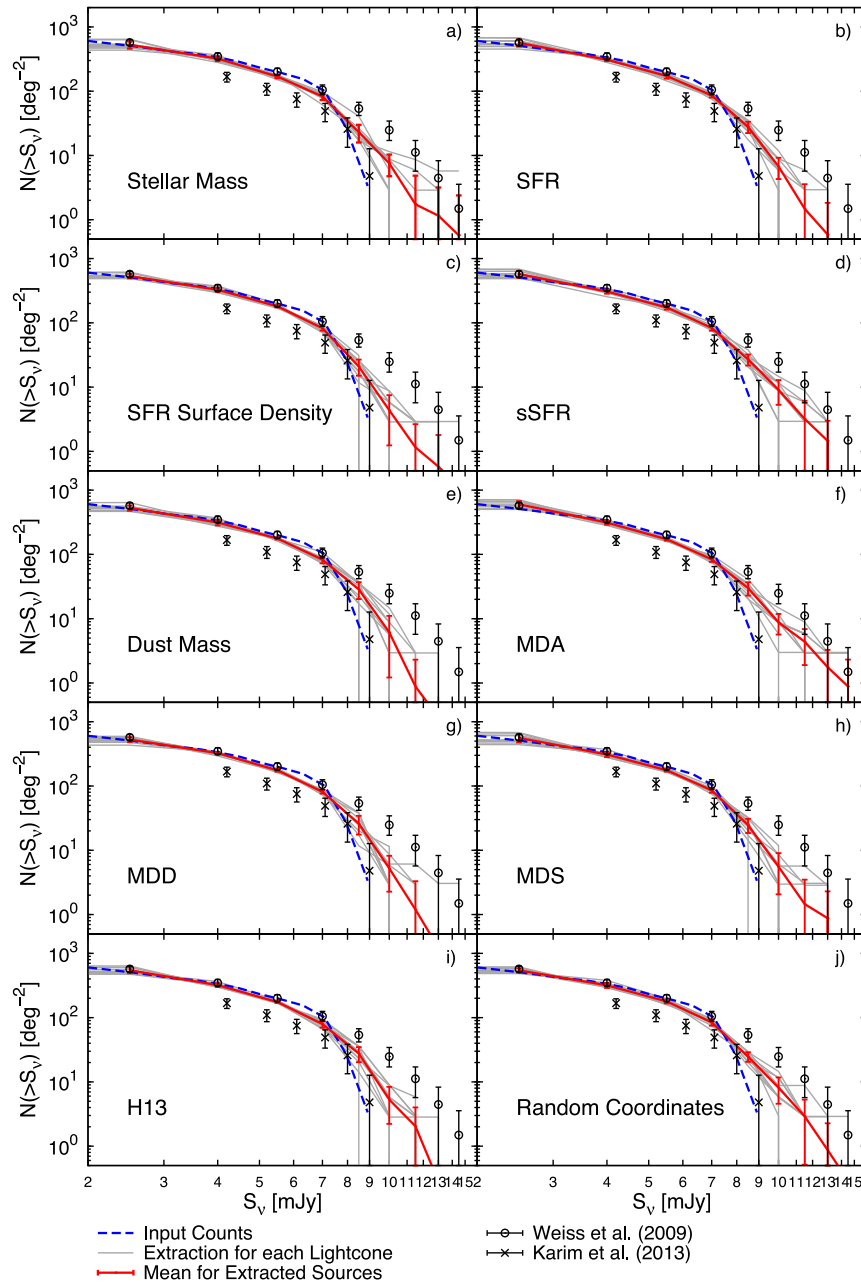
**Figure 6.** Cumulative number counts at 870 μm obtained for the different proxies: (a) stellar mass, (b) SFR, (c) SFR surface density, (d) sSFR, (e) dust mass, (f) dust mass divided by stellar age (MDA), (g) dust mass times SFR divided by the square of the disc scalelength (MDD), (h) dust mass times SFR (MDS), and (i) equation 15 of Hayward et al. (2013a), which we refer to as the H13 proxy. Blue thick dashed lines give the input counts for each proxy, using all the input galaxy fluxes given by the Count Matching. Grey thin solid lines correspond to the counts recovered for the 10 light cones after passing through the observational process, i.e. obtained from sources extracted from the simulated maps. Red thick solid lines connect mean values for simulated sources at each flux level, with standard deviations as error bars. As a comparison, observational data are shown from LABOCA (Weiss et al. 2009) and ALMA studies (Karim et al. 2013). Additionally, counts obtained for input sources with randomized positions in the simulated map are given in panel (j). In all panels, there are lines of sight giving counts consistent with those derived from LABOCA observations.

A fraction of ∼22–25 per cent of the extracted sources do not have counterparts (see Table 2, third column); in the following, we remove these sources from our analyses. The fraction of multiple sources across the proxies, in the flux range given by the recovered counts, goes from ∼9 to ∼14 per cent (see Table 2, fourth column). We have checked that the coordinates of counterparts do not follow a preferential direction in the map when compared to the coordinates of extracted sources.

### 6.2 Recovered redshift distributions for detected sources and *k*-correction effects

As shown above, all the proposed proxies (even taking random positions for galaxies in the sky) can successfully reproduce the observed LABOCA counts. We then explore other predicted properties of bright model galaxies, to find out which (if any) proxies for submm luminosity are plausible.

**Table 2.** Properties from the counts derived for each proxy, using sources extracted from the simulated maps (extraction down to the $3.8\sigma$ detection level). Fractions are mean values over the 10 light cones, reporting also standard deviations.

| Proxy | $\chi^2/\text{dof}^a$ | Mean fraction of non-matches[b] | Mean fraction of multiples[c] |
|---|---|---|---|
| Stellar mass | 3.598 | $0.223 \pm 0.027$ | $0.113 \pm 0.044$ |
| SFR | 3.120 | $0.250 \pm 0.032$ | $0.105 \pm 0.028$ |
| SFR surface density | 3.689 | $0.232 \pm 0.029$ | $0.090 \pm 0.033$ |
| sSFR | 3.617 | $0.237 \pm 0.026$ | $0.123 \pm 0.028$ |
| Dust mass | 3.536 | $0.243 \pm 0.028$ | $0.091 \pm 0.021$ |
| MDA | 3.400 | $0.228 \pm 0.021$ | $0.141 \pm 0.034$ |
| MDD | 3.642 | $0.232 \pm 0.029$ | $0.102 \pm 0.033$ |
| MDS | 3.558 | $0.244 \pm 0.025$ | $0.096 \pm 0.028$ |
| H13 | 3.599 | $0.245 \pm 0.041$ | $0.089 \pm 0.034$ |
| Random coordinates | 3.470 | $0.248 \pm 0.021$ | $0.093 \pm 0.027$ |

[a]Reduced $\chi^2$ for each proxy, with $\chi^2$ calculated according to equation (6) and dof the degrees of freedom (i.e. number of flux bins).
[b]Mean fraction of sources without having a counterpart in the input catalogue (over all the extracted sources).
[c]Mean fraction of sources having more than one counterpart in the input catalogue (over the extracted sources that have at least one counterpart).
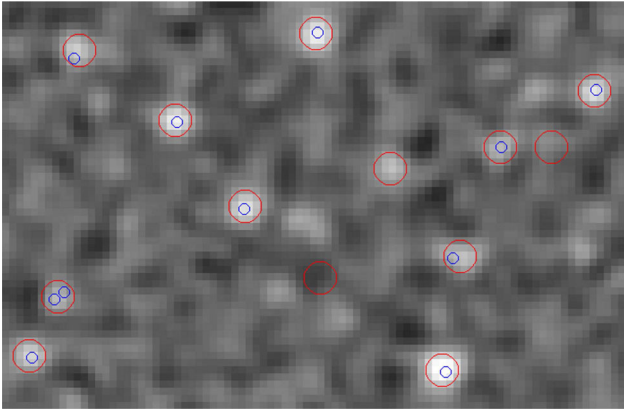


**Figure 7.** Small section of the simulated map shown in Fig. 4, illustrating the cross-match between extracted and input sources. Sources extracted from the map lie in the centres of the red big circles, which have a radius of 13.8 arcsec; this is the search radius used for finding the counterparts of these sources in the input catalogue down to 1.2 mJy, which are shown as blue small circles. They allow us to explore the predicted properties for extracted sources in each light cone and proxy.

The properties of extracted sources are recovered assigning to them the galaxy properties of their counterpart(s) in the input catalogue. Including in our analysis the properties of all the input sources within the search radius down to 1.2 mJy allows us to compare the distribution of recovered properties for each proxy with statistics derived from observations in the literature, even from follow-ups of single-dish detected SMGs with interferometry. If we restrict the comparison to single-dish detected SMGs, and so assign to a extracted multiple source the properties of the brightest model galaxy within the search radius, there are no significant differences in the recovered distributions compared to the inclusion of all counterparts. This shows that the observational process of source extraction does not impose systematic biases in this respect. In addition, we can compare the observations in the literature with the distributions obtained for input sources brighter than 3.8 mJy, i.e. the properties of the brightest model galaxies in submm before going through the observational process.

Fig. 8 shows the mean redshift distributions for the extracted sources in each proxy. For all proxies, there is no significant difference between them and the distributions for the input brightest sources. The proxy that gives the closest match with the observed distribution (i.e. Wardlow et al. 2011 for the ECDF-S) is MDS, giving mean values around $z = 2$. Although the SFR proxy gives a distribution with similar shape to the observational one, it has the problem of including a significant fraction of very low-$z$ sources. Hence, even though the MDS proxy prediction is much broader compared to the observed distribution, it provides an acceptable agreement for such a simple model where the rest-frame 870 µm luminosity is assumed to be related to galaxy properties in a direct dependence (see Fig. 3), suggesting that both dust mass and SFR might play an important role in the process of FIR emission. This dependence can surely be improved.

For the brightest sources in the submm, however, redshift distributions also depend on the $k$-correction factor assumed for the model galaxies, as it affects the assignment of submm fluxes to each galaxy in the sample and thus the flux of sources detected in the simulated maps; similarly, the distributions may change if very low-$z$ sources are excluded in the Count Matching process. Assuming that the negative $k$-correction eliminates completely the dilution due to distance, the amount of submm flux assigned to each galaxy only depends on the value of the property selected as proxy; we call this case 'no $k$-correction', and explore it particularly for the MDS and H13 proxies. The predicted distributions are shown as green histograms in panels (h) and (i) of Fig. 8, respectively, for comparison with the fiducial case. As it is shown, H13 becomes a good proxy only if model sources having $z < 0.5$ are excluded in the Count Matching process. In this case, there are no significant modifications to the shape of the distribution when applying the $k$-correction to model spectra.

The dependence of the submm flux density with dust mass and SFR was addressed in detail by Hayward et al. (2011), using the results of hydrodynamical simulations of isolated disc and merging galaxies connected to a dust radiative transfer code. Hayward et al. (2011) results motivated the fitting function used by Hayward et al. (2013a) in the assignment of submm fluxes. They found that this relation is accurate to within 0.3 dex in the redshift range $\sim$1–6, but severely underpredicts the galaxy fluxes for $z \lesssim 0.5$. Our results are in line with those findings, as the H13 proxy reproduces the observed redshift distribution only when discarding low-redshift sources during the flux assignment. Over $z > 0.5$, this proxy works applying or not the $k$-correction factor. It is remarkable that the MDS proxy gives good agreement with observations even when sources at $z < 0.5$ are excluded in the Count Matching process, and independent of the assumption for the shape of the galaxy spectrum.

Besides the proposed proxies, we explore the redshift distribution obtained when a random assignment of submm luminosities is done for model sources, which are injected in the simulated map with a random spatial distribution. As is shown in panel (j) of Fig. 8, the recovered distribution is wider than the observed for SMGs and reaches its maximum around $z = 2$, resembling the statistics for the whole galaxy population in the simulated light cones.

We apply some robustness tests to our technique. First, we test the effects of adding scatter to the proxy, in light of the findings by Moster et al. (2010) and Behroozi et al. (2010) about how scatter can bias the results of the abundance matching. We explore this using a Gaussian distribution centred in the value given by the model and taking a standard deviation of 30 per cent (this choice is close to typical errors in galaxy properties derived from observations,
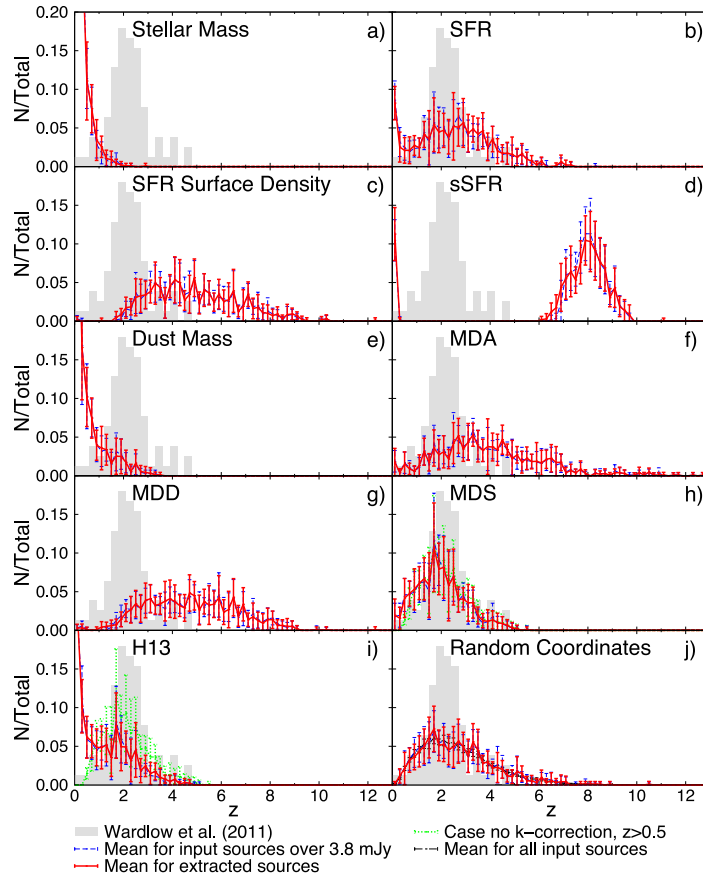
**Figure 8.** Mean redshift distributions over the 10 light cones obtained for the different proxies, taking all the sources extracted from simulated maps having counterparts in the catalogue of input sources down to 1.2 mJy, considering all the input sources that lie within a search radius of 13.8 arcsec (red thick solid lines). This is compared to the mean histogram obtained taking all the input sources brighter than 3.8 mJy (blue thin dashed lines). For each proxy the distribution of photometric redshifts obtained by Wardlow et al. (2011) for 78 LESS sources is overplotted (grey histograms). The submm flux for each galaxy in the input catalogues, which give origin to the simulated maps, is obtained according to equation (5) (see also equation 2), where the $k$-correction factor is given by an Arp 220 template spectrum. The MDS proxy is the only one that is in good qualitative agreement with the observed distribution for the ECDF-S. The H13 proxy only gives good agreement when model sources having $z < 0.5$ are excluded in the Count Matching process (green thick dotted lines); in this case, $k$-correcting the spectra does not introduce a significant change in the distribution, and MDS remains as a good proxy. Additionally, the last panel shows the distribution recovered when a random assignment of submm luminosities is done for model sources, injected in a map following a random spatial distribution. In this panel, we also include the distribution for the whole galaxy population in the simulated light cones (black thin dot–dashed line).

as stellar mass or SFR). After this, the trend between proxy and rest-frame 870 μm luminosity is still robust. The recovered redshift distributions are consistent with the ones obtained without considering the scatter, with their medians varying only 0.15 units at most (except for the sSFR proxy, where the median redshift decreases by 0.9 units).

In addition, we test the effects that come from using models that give poorer fits to the observed galaxy population. We achieve this modifying two key features in SAG: (a) excluding AGNs in the whole model, and (b) using the De Lucia et al. (2004) prescription for star formation (fiducial is Croton et al. 2006, which unlike the former involves a cold gas mass density threshold below which there is no star formation). Both tests were done keeping the same model parameters as in the fiducial case.

Excluding AGNs changes dramatically the SMG redshift distribution for the MDS proxy, which now peaks at $z = 0.3$. For the stellar mass, dust mass and H13 proxies it gives no sources over $z = 1$, while for the SFR proxy it gives ∼3 times more low-$z$ sources in proportion. For this model, none of our proposed proxies gives an acceptable redshift distribution when compared to observational data.

The model with the alternative prescription for star formation introduces changes in the SMG redshift distributions for some of the proxies. For the MDS and H13 proxies it gives 2 and 1.6 times more low-$z$ sources in proportion, respectively. For the dust mass proxy it gives 1.5 times more $z = 0$ sources and essentially zero sources over $z = 1$. For the SFR surface density proxy it gives no sources below $z = 3$, while for the MDD proxy there are no SMGs below $z = 4$.

Finally, we tested the effects on the recovered distributions of using the median SED for observed SMGs found by Michałowski et al. (2010) to compute $k$-corrections, finding no significant differences with respect to the fiducial model except for the specific SFR (sSFR) proxy, where the median redshift decreases by 2 units. For the H13 proxy, an extra peak emerges in the predicted redshift distribution at $z = 0.8$. We also tried Magdis et al. (2012) SED templates as they allowed Béthermin et al. (2012) to reproduce the counts from the mid-IR to the mm domain using an empirical model. Using these templates, we find no significant differences with respect to the fiducial model. For instance, in the SFR, MDS and H13 proxies the median redshift moves to a lower value by ∼0.5 units. To sum up, the changes from modifications to the model produce
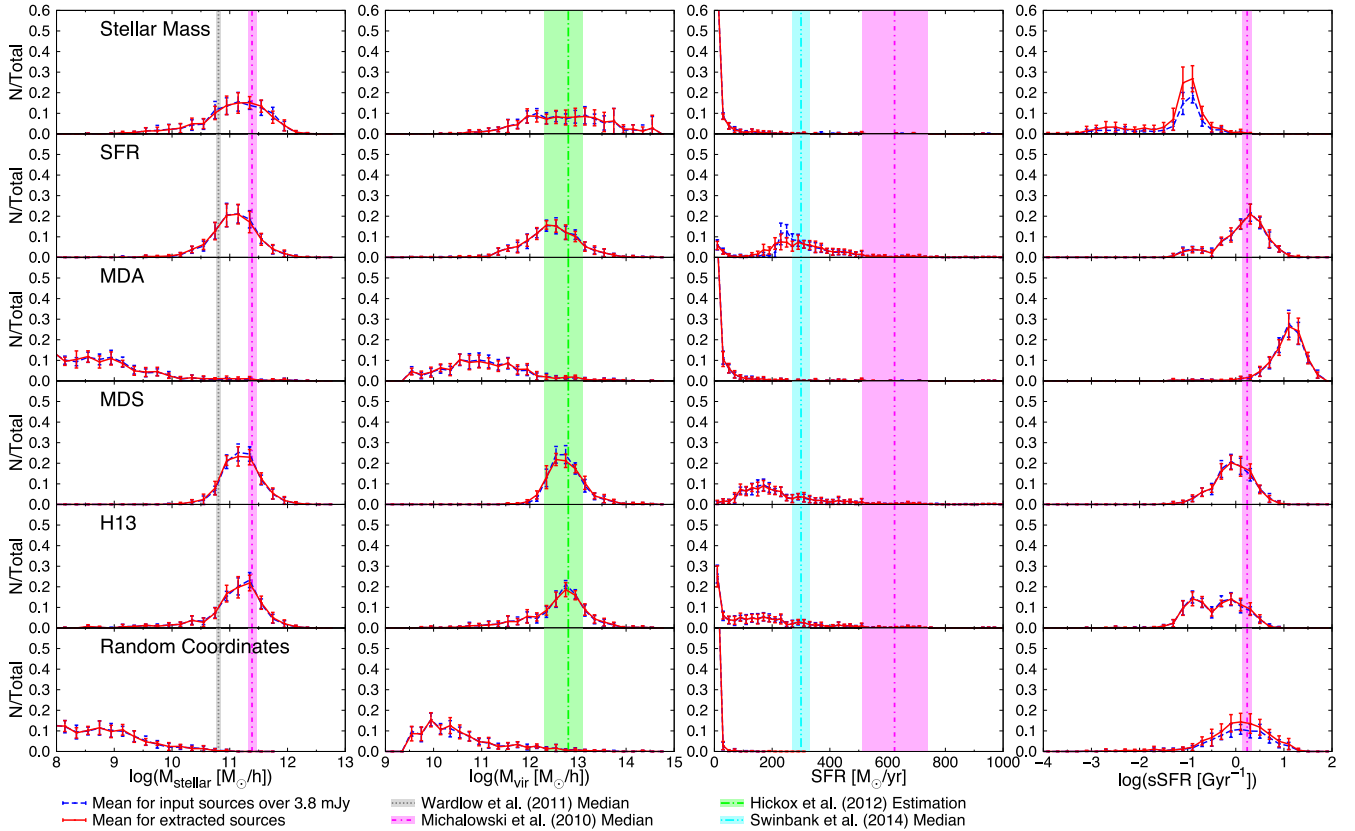
**Figure 9.** Mean distributions over the 10 light cones of several properties obtained for five selected proxies: stellar mass, SFR, dust mass divided by stellar age (MDA), dust mass times SFR (MDS), and equation 15 of Hayward et al. (2013a), which we refer to as the H13 proxy. Sources extracted from simulated maps having counterparts in the catalogue of input sources down to 1.2 mJy are taken, considering all the input sources that lie within a search radius of 13.8 arcsec (red solid lines). Properties are compared to the mean histograms obtained taking all the input sources brighter than 3.8 mJy (blue dashed lines). Standard deviations for model distributions are shown as error bars. In columns from left to right, the predicted properties are: stellar mass, compared to the median values obtained by Wardlow et al. (2011) and Michałowski et al. (2010), in vertical grey dotted and magenta dot–dashed lines, respectively; virial halo mass, compared to the Hickox et al. (2012) estimation, as a vertical green long dash–dotted line; SFR, compared to the Michałowski et al. (2010) median value as a vertical magenta dot–dashed line and to the Swinbank et al. (2014) median value as a vertical cyan double dot–dashed line; and specific SFR, compared to Michałowski et al. (2010) median value, again in a vertical magenta dot–dashed line. Standard errors for observational values are presented as coloured shaded regions. Bottom panels show the distributions recovered when a random assignment of submm luminosities is done for model sources, injected in a map following a random spatial distribution.

small changes in the results as long as the $z = 0$ model galaxies are consistent with observations.

### 6.3 Prediction of physical properties of SMGs

Our analysis of the submm extracted sources is also extensible to other galaxy properties given by the semi-analytic model. Figs 9 and 10 show some of the distributions predicted when taking *k*-corrected galaxies at $z > 0$ in the Count Matching process: stellar mass, halo mass, SFR, sSFR, gas fraction, cold gas phase metallicity, effective radius and depletion time. Note that the intrinsic values of these properties are not affected by the Count Matching, since they are computed entirely by the SAG model. Keeping the values of these properties without modifications, assures that the final SAG galaxy population remains unaffected by the Count Matching technique. What may vary across the proposed proxies are the distributions of these properties for bright SMGs, since model sources having assigned the highest 870 μm flux densities for one proxy could have, for instance, intrinsically lower SFRs when compared to model bright sources selected using another proxy.

All the properties listed above can be compared with observations.[3] For brevity, this is presented only for five of the proxies; we select the proxies giving the best agreement with the observed redshift distribution shown in Fig. 8, as well as some of the proxies that do not. This is done in order to check if the former ones predict, in a consistent way, other galaxy properties when compared to observations. Median values for these properties are shown in Tables 3 and 4. The Arp 220 template is used for *k*-correcting the galaxy spectra. Again, there is no significant difference between the recovered distributions and those for the input brightest sources.

We compare the distribution of stellar masses with the median values derived by Wardlow et al. (2011) and Michałowski et al. (2010), who consider single-dish detected sources mapping the submm continuum; the former analyse 78 detected counterparts to 72 SMGs in the ECDF-S, while the latter consider 76 SMGs from the Chapman et al. (2005) sample. Except for the MDA proxy, all of them give reasonable distributions compared to SMG observations;

---

[3] When comparing results for the proxies with observational properties that depend on assumed evolutionary synthesis models, we do not make any rescaling for IMFs different from the one assumed in this work (Salpeter IMF).
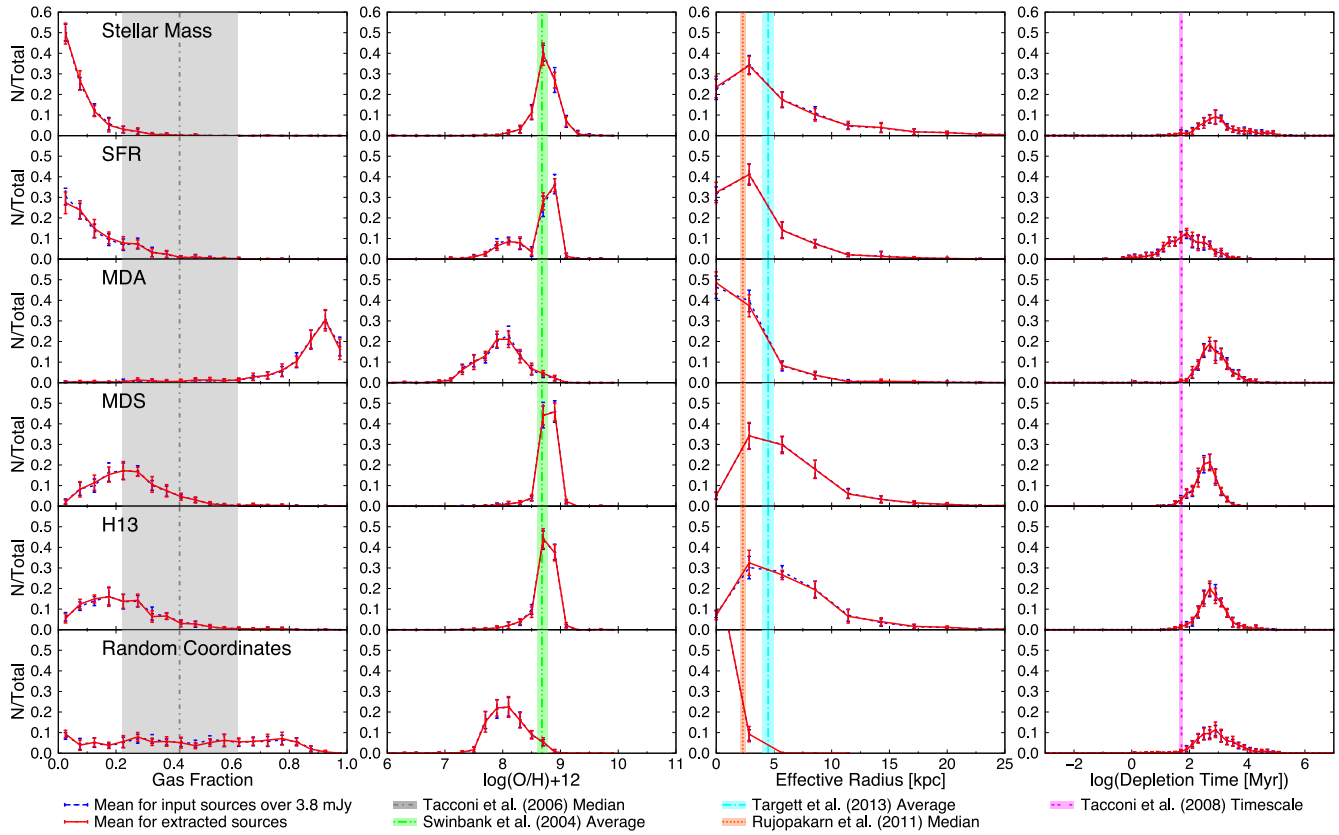
**Figure 10.** Mean distributions over the 10 light cones of several properties obtained for five selected proxies: stellar mass, SFR, dust mass divided by stellar age (MDA), dust mass times SFR (MDS), and equation 15 of Hayward et al. (2013a), which we refer to as the H13 proxy. Red solid and blue dashed lines are as in Fig. 9. In columns from left to right, the predicted properties are: gas fraction, defined as cold gas mass over the sum between cold gas and stellar mass, compared to the median value found by Tacconi et al. (2006) as a vertical grey dot–dashed line; cold gas phase metallicity in terms of oxygen and hydrogen abundances, compared to the average value found by Swinbank et al. (2004) as a vertical green double dot-dashed line; effective radius, computed considering the contribution of both bulge and disc, compared with values found by Targett et al. (2013) and Rujopakarn et al. (2011) in vertical cyan long-dash–dotted and orange dotted lines, respectively; and depletion time, defined as the ratio between the cold gas mass and SFR, compared to the gas exhaustion time-scale found by Tacconi et al. (2008) as a vertical magenta double-dotted line. Bottom panels show the distributions recovered when a random assignment of submm luminosities is done for model sources, injected in a map following a random spatial distribution.

**Table 3.** Properties of sources extracted from the simulated maps. Uncertainties for each proxy are standard deviations of the median, considering all sources belonging to the 10 light cones altogether.

| Proxy | Redshift | $\log$ (Stellar mass [$M_\odot h^{-1}$]) | $\log$ (Virial halo mass [$M_\odot h^{-1}$]) | SFR ($M_\odot$ yr$^{-1}$) |
|---|---|---|---|---|
| Stellar mass | $0.223 \pm 0.261$ | $11.135 \pm 0.764$ | $12.716 \pm 1.353$ | $1.552 \pm 3.406$ |
| SFR | $2.542 \pm 2.333$ | $11.074 \pm 0.552$ | $12.447 \pm 0.785$ | $287.299 \pm 173.629$ |
| SFR surface density | $4.715 \pm 2.922$ | $9.972 \pm 0.979$ | $10.959 \pm 0.809$ | $61.755 \pm 61.321$ |
| sSFR | $7.923 \pm 1.251$ | $7.982 \pm 0.208$ | $9.778 \pm 0.661$ | $4.540 \pm 1.984$ |
| Dust mass | $0.247 \pm 0.437$ | $10.993 \pm 0.977$ | $12.566 \pm 1.237$ | $8.239 \pm 18.077$ |
| MDA | $3.656 \pm 2.759$ | $8.716 \pm 1.081$ | $10.996 \pm 1.238$ | $10.105 \pm 15.559$ |
| MDD | $4.677 \pm 3.114$ | $10.392 \pm 0.645$ | $11.362 \pm 0.581$ | $102.731 \pm 61.160$ |
| MDS | $1.862 \pm 1.448$ | $11.186 \pm 0.463$ | $12.728 \pm 0.498$ | $190.734 \pm 160.044$ |
| H13 | $0.998 \pm 1.785$ | $11.177 \pm 0.528$ | $12.674 \pm 0.696$ | $129.016 \pm 237.874$ |

note that the standard errors reported for observational quantities do not take into account the individual uncertainties in their determination. However, moving to the predicted virial halo masses, we find that only MDS and H13 proxies are able to reproduce the value estimated by Hickox et al. (2012) for SMGs in the ECDF-S field.

Our model underpredicts the SFRs by a factor of ∼3 when compared to observed values for SMGs by Michałowski et al. (2010), being more consistent with SFRs measured for high-$z$ star-forming galaxies (SFGs) by Tacconi et al. (2010), which have mean values of 95 and 135 M$_\odot$ yr$^{-1}$ at $z = 1.2$ and 2.3, respectively (note that

this is not a sample of SMGs). Deblending FIR observations for the ALESS SMG positions with *Herschel*, Swinbank et al. (2014) find a median SFR ∼2 times lower than the one derived for single-dish selected SMGs, although their sample includes 870 μm fluxes down to 2 mJy (including only sources brighter than 4.2 mJy gives an SFR of 530 ± 60 M$_\odot$ yr$^{-1}$). Regarding the sSFR distribution, the stellar mass proxy gives considerably lower values compared to the median value found by Michałowski et al. (2010), the MDA proxy considerably higher values, and the H13 proxy a bimodal distribution; the remaining proxies give a reasonable prediction.

**Table 4.** Properties of sources extracted from the simulated maps (continued).

| Proxy | sSFR (Gyr$^{-1}$) | Gas fraction | log(O/H)+12 | Effective radius (kpc) | Depletion time (Myr) |
|---|---|---|---|---|---|
| Stellar mass | $0.014 \pm 0.030$ | $0.050 \pm 0.077$ | $8.767 \pm 0.220$ | $3.422 \pm 5.238$ | $5035.675 \pm 10884.577$ |
| SFR | $1.652 \pm 2.242$ | $0.097 \pm 0.144$ | $8.769 \pm 0.222$ | $2.199 \pm 3.074$ | $70.547 \pm 130.123$ |
| SFR surface density | $4.787 \pm 5.569$ | $0.007 \pm 0.008$ | $8.288 \pm 0.401$ | $0.021 \pm 0.012$ | $1.617 \pm 1.151$ |
| sSFR | $35.106 \pm 7.081$ | $0.762 \pm 0.158$ | $7.834 \pm 0.141$ | $0.153 \pm 0.164$ | $101.953 \pm 94.529$ |
| Dust mass | $0.115 \pm 0.246$ | $0.215 \pm 0.201$ | $8.671 \pm 0.319$ | $6.940 \pm 8.190$ | $2260.248 \pm 3800.487$ |
| MDA | $12.474 \pm 12.398$ | $0.896 \pm 0.113$ | $7.975 \pm 0.591$ | $1.479 \pm 1.953$ | $628.211 \pm 826.719$ |
| MDD | $3.197 \pm 3.538$ | $0.009 \pm 0.014$ | $8.476 \pm 0.258$ | $0.039 \pm 0.036$ | $2.938 \pm 3.372$ |
| MDS | $0.878 \pm 1.061$ | $0.238 \pm 0.172$ | $8.796 \pm 0.134$ | $5.127 \pm 4.998$ | $377.963 \pm 429.295$ |
| H13 | $0.429 \pm 0.705$ | $0.203 \pm 0.181$ | $8.773 \pm 0.168$ | $5.222 \pm 5.838$ | $576.980 \pm 738.643$ |

Reproducing the observed SFR distribution for SMGs is an outstanding challenge for semi-analytic models. Some aspects in SAG regarding the modelling of the star formation (e.g. the complete removal of hot gas when galaxies become satellites, or not including cold gas inflows) might be leading to lower predicted SFRs not only for SMGs, but also for the whole galaxy population across redshift (see the discussion about Fig. A1 in Appendix A). This translates in globally underpredicting galaxy SFRs at the redshifts where SMGs lie. It affects the predicted distributions for SMGs given by all our proposed proxies, recalling that the intrinsic SFR of each galaxy is not affected by the Count Matching process.

However, for each model SMG we can explore the relation between the SFR, given by the semi-analytic model, and the bolometric IR luminosity, computed from its 870 μm flux (using its redshift and the SED template to integrate its emission in the rest-frame wavelength range 8–1000 μm). The MDS proxy predicts a positive trend between SFR and bolometric IR luminosity, in line with observations: on average, a model source classified as ULIRG has larger SFR than one classified as LIRG. We highlight this prediction because, when performing the Count Matching, we pose no requirements regarding this relationship. Therefore, different proxies can give completely different SFR versus bolometric IR luminosity laws.

We also compute the model gas fraction for the brightest submm sources, as the ratio between the cold gas mass and the sum of cold gas and stellar mass. Compared to the average that Tacconi et al. (2006) found for eight SMGs at $z \sim 2$–3.4 (estimated from both continuum and CO emission lines), the MDS and H13 proxies give less gas-rich sources than the observed SMGs, while the stellar mass and SFR proxies give very gas-poor galaxies, and the MDA proxy very gas-rich sources. In all cases except for MDA proxy (where it is underpredicted), the metallicity of the remaining cold gas in the galaxy is close to the average value that Swinbank et al. (2004) obtained for 15 sources including SMGs and optical faint radio galaxies (Chapman et al. 2004) targeting the Hα line.

We compute the effective (i.e. projected) radius of a given galaxy weighting the contributions of both bulge and disc as

$$r_{\rm eff} = \frac{M_{\rm bulge}\, r_{\rm bulge,eff} + M_{\rm disc}\, r_{\rm disc,eff}}{M_{\rm bulge} + M_{\rm disc}}, \qquad (7)$$

where $M_{\rm bulge}$ is the bulge stellar mass, $M_{\rm disc}$ is the sum of cold gas and stellar mass of the disc, $r_{\rm bulge,\ eff} = r_{\rm bulge}/1.35$ and $r_{\rm disc,\ eff} = r_{\rm disc}/1.68$ are the effective radii for bulge[4] and disc, respectively, given in terms of the half-mass radii in three dimensions; the factors 1.35 and 1.68 correspond to concentration values for elliptical and spiral galaxies, respectively, taken from Graham et al. (2005). Compared to the

[4] A brief description of the adopted model for bulge sizes is given in Appendix B.

average half-light radius found by Targett et al. (2013) for 24 SMGs in the Great Observatories Origins Deep Survey-South (GOODS-S) and to the median effective radius found by Rujopakarn et al. (2011) for 48 sources at high redshift (median of 1) comprising SMGs and ULIRGs, we find that the stellar mass and SFR proxies give distributions peaking at the value for observed high-$z$ ULIRGs and SMGs, while the MDA proxy isolates mostly very compact sources. The MDS and H13 proxy prefer sources with size in agreement with Targett et al. (2013) data, although both distributions are quite broad.

Finally, we compute the depletion time for model galaxies, defined as the ratio between the cold gas mass and SFR. This gives a measure of the gas exhaustion time-scale. Only the SFR proxy gives a distribution consistent with the Tacconi et al. (2008) time-scale for four SMGs, which was derived from mm CO interferometry. Conversely, the MDA, MDS and H13 proxies give distributions consistent with the value found by Tacconi et al. (2010) for high-$z$ SFGs ($0.9 \pm 0.6$ Gyr); the stellar mass proxy gives a similar distribution too, but as is shown in its SFR distribution, most of the brightest galaxies have SFRs close to zero, giving depletion times tending to infinite values, which do not appear in the depletion time plot but are still considered when calculating the median value for the model (see Table 4).

As was done for galaxy redshifts, we compute the recovered distribution of all these properties when a random assignment of submm luminosities is done for model sources, being injected in the simulated map with a random spatial distribution. This is shown in the bottom panels of Figs 9 and 10. The distributions obtained in this case are quite different from those for observed SMGs, except for the sSFRs, where the apparent agreement with observations comes from the distribution of this property for the whole model galaxy population.

It is worth noting that the validity of the predicted properties is preserved even after taking into account scatter in the proxy, as well as when different SED templates (Michałowski et al. 2010 or Magdis et al. 2012) are assumed to compute $k$-corrections. Nevertheless, when excluding AGNs in the SAG model, SFR distributions peak at lower values for all proxies. In particular, it decreases the median SFR by $\sim$200 M$_\odot$ yr$^{-1}$ for the MDS proxy. Predicted stellar masses change slightly, remaining consistent with observational data. The large changes are expected though, as this model is extremely different from the fiducial one; the bright end of the local optical LF is affected, as well as the slope of the cosmic SFR across redshift, departing considerably from observations. On the other hand, adopting the De Lucia et al. (2004) prescription for star formation gives roughly the same stellar and virial masses compared to the fiducial case, except for the H13 proxy, where a tail with lower masses appears. For the SFR and MDS proxies the SFRs are larger, changing the median value by $\sim$50 M$_\odot$ yr$^{-1}$ (affecting their depletion times accordingly). The largest impact is on the gas

fractions, which now peak at 0 (recall that in this prescription stars can form for any amount of cold gas). For the stellar mass proxy, a tail with large metallicities appears.

Therefore, models with very different star formation laws still broadly agree in the predicted SMG properties, unlike models where AGNs are not taken into account. As long as the model provides a reasonable fit to the general galaxy population, the predictions are robust.

### 6.4 Effect of galaxy blending on the recovered counts

We study the contribution of multiple sources to the cumulative number counts at submm wavebands, separating them from single sources. The recovered fraction of blends in a multiple source depends naturally on the flux depth, as is pointed out by Chen et al. (2013). At the counts level shown in Fig. 11, left-hand column, we obtain a maximum of three blends down to an input flux level of

1.2 mJy. For the stellar mass, SFR and H13 proxies, there are no single sources brighter than 10 mJy, but only multiples. For MDA and MDS proxies, the importance of multiple sources composed by two blends begins at extracted fluxes higher than ~9 and ~11 mJy, respectively.

In order to address if there is a spatial association between the blended sources that compose a multiple one, we compute the redshift separation between components as

$$\Delta z = \sqrt{\sum_{i \neq j} (z_i - z_j)^2}. \tag{8}$$

The mean distributions are shown in Fig. 11, right-hand column. For distinguishing between associated and unassociated sources, we follow the limit of $\Delta z = 0.02$ proposed by Hayward et al. (2013b). For the five proxies, we find that most of the components are unassociated (in agreement with findings by Hayward et al. 2013b and Cowley et al. 2014 for model SMGs), with a significant
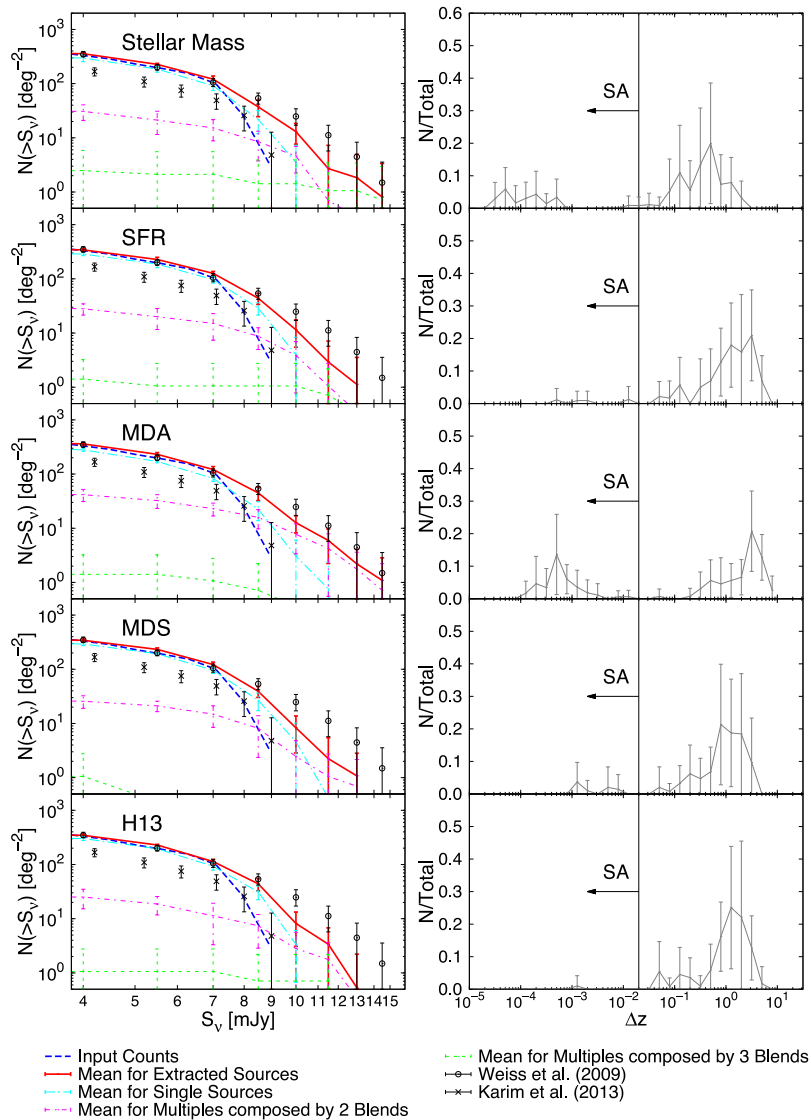


**Figure 11.** Properties of multiple sources for five selected proxies. Left column: recovered cumulative number counts at 870 μm (red thick solid lines) split in the contribution by single sources (cyan thin long-dash–dotted lines), multiples composed by two blended sources (magenta thin dot–dashed lines) and multiples composed by three blended sources (green thin short-dashed lines); these are compared with the input counts (blue thick dashed lines) and to observational data from LABOCA and ALMA studies (symbols as in Fig. 6). Right column: mean distribution of the redshift separation between components of multiple sources (grey solid lines). 'SA' stands for spatially associated sources, according to the cut proposed by Hayward et al. (2013b).

**Table 5.** Mean cumulative fraction over flux density $S_\nu$ at 870 μm for MDS proxy. Standard deviations are reported.

| Type of source | $N/N_{\text{extractions}}(>S_\nu)$ | | | |
| --- | --- | --- | --- | --- |
| | 7 mJy | 8.5 mJy | 10 mJy | 11.5 mJy |
| Single | $0.842 \pm 0.134$ | $0.714 \pm 0.308$ | $0.584 \pm 0.795$ | $0.199 \pm 0.690$ |
| Multiple composed by 2 blends | $0.155 \pm 0.068$ | $0.274 \pm 0.206$ | $0.363 \pm 0.419$ | $0.595 \pm 1.275$ |
| Multiple composed by 3 blends | $0.003 \pm 0.012$ | $0.012 \pm 0.039$ | $0.053 \pm 0.173$ | $0.206 \pm 0.712$ |
| Total | 1 | 1 | 1 | 1 |

amount of associated sources only for the stellar mass and MDA proxies. This result is in line with the lack of a dependence in the counts on the clustering of sources of different proxies, which is also similar to the case of random positions (see Section 6.1); in a given submm image, most sources that lie in a same line of sight are there by chance.

### 6.5 Characteristics of our best proxy: MDS

We end this section presenting some predictions for SMGs in the MDS proxy, which is the only one giving at the same time red-shift, stellar mass and host halo mass distributions consistent with observed ones for SMGs.

#### 6.5.1 Multiplicity of single-dish detected SMGs and redshift distribution of all blends

Table 5 shows the cumulative fraction $N/N_{\text{extractions}}(>S_\nu)$ of single and multiple sources at several submm fluxes, averaged over the 10 light cones. Note that over 8.5 mJy this proxy predicts that ~27 per cent of the detected sources are multiples composed by two blends. Over 10 mJy the statistics is noisier, but still not consistent with the finding by Karim et al. (2013) of 100 per cent multiple sources over ~9 mJy in the ECDF-S. Over our S/N limit of 3.8, the fraction of multiples is in the range ~9–14 per cent (see Table 2). This is lower than the range reported by Hodge et al. (2013) for LESS sources, ~35–45 per cent. Conversely, it is closer to the findings of $12.5^{+12.1}_{-6.8}$ per cent by Chen et al. (2013) for SCUBA-2 SMGs in the CDF-N, and to the $22 \pm 9$ per cent by Smolčić et al. (2012) for LABOCA SMGs in the COSMOS field.

We explore whether the redshift distribution of both single sources and the blends of multiple sources is different from the distribution of all model sources over our flux cut of 1.2 mJy in the whole field. This statistics could be obtained for the actual ECDF-S once interferometric observations over the entire field are carried out. So far, the determination of photometric redshifts for actual sources over ~1.2 mJy in the ECDF-S is restricted to ALMA sources surrounding/comprising bright LABOCA SMGs (Simpson et al. 2014); they consider a subset of the Hodge et al. (2013) catalogue.

We show the comparison between model and observed data in Fig. 12. Compared to ALMA sources, in proportion we recover more sources at low redshift. Restricting the discussion to our model data, there is a slight change in the distribution taking input sources in the whole field compared to the subset of sources that lie around or comprise SMGs extracted from the simulated maps: considering all sources in the field, we recover slightly less low-$z$ and more high-$z$ sources in proportion.
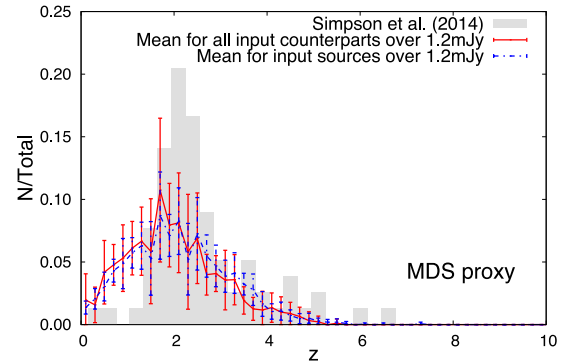


**Figure 12.** Mean redshift distribution over the 10 light cones for the MDS proxy, considering all the counterparts for extracted sources in the input catalogue over 1.2 mJy (red solid line). This is compared with the distribution of photometric redshifts for ALMA sources in the ECDF-S from Simpson et al. (2014) displayed as a grey histogram. The mean distribution for all input sources over 1.2 mJy, i.e. not only around bright extracted sources, is also shown (blue dot–dashed line). Standard deviations are reported for each redshift bin.

#### 6.5.2 Star formation rate versus stellar mass

We also compare our best proxy with the stellar mass and SFR ones, putting together the sources selected by each proxy in all light cones, after going through the observational process. For each source extracted from the simulated maps, we consider the properties of all counterparts within the search radius down to 1.2 mJy.

Fig. 13 shows the SFR as a function of stellar mass for the three proxies, including also the distribution for all galaxies in the 10 input light cones and the median sSFR for SMGs by Michałowski et al. (2010), translated to a function with the shape SFR $\propto M_{\text{stellar}}$. Results are in line with the findings shown in Fig. 9, fourth column; the MDS and SFR proxies give most of the SMGs lying in the high-mass end of the relationship between SFR and stellar mass that can be derived from the observational median sSFR, while the stellar mass proxy gives SMGs lying below the relation.

#### 6.5.3 Luminosity functions

Rest-frame LFs at other wavebands can be recovered using the input 870 μm flux and redshift of all sources in a given light cone and making use of the Arp 220 template spectrum. For instance, Fig. 14 shows rest-frame LFs at 90 μm in six redshift ranges spanning from $z = 0.4$ to 4.5, predicted for model SMGs in each of the three proxies. The MDS and SFR proxies predict similar distributions at all redshifts, giving a reasonable agreement with Gruppioni et al. (2013) *Herschel* data (taken with the Photoconductor Array Camera
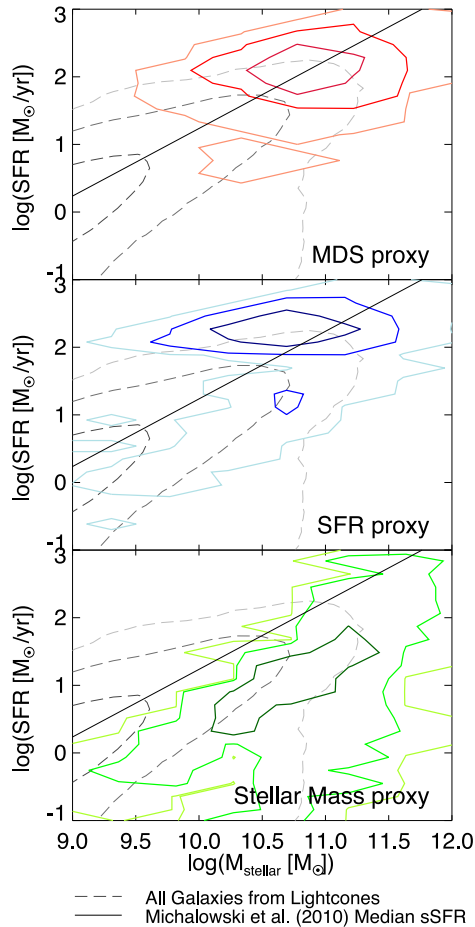
**Figure 13.** SFR versus stellar mass including sources extracted from all simulated maps. Three proxies are compared in separate diagrams for clarity: the results for our best proxy, MDS (red solid contours) are compared to those for stellar mass and SFR proxies (green and blue solid contours, respectively). Also displayed are the distribution for galaxies in the 10 input light cones as grey dashed contours, and the relation between SFR and stellar mass derived from the median specific SFR determined by Michałowski et al. (2010) as a black solid line. From the darkest to the lightest, contours enclose 68, 95 and 99 per cent of the sources.

and Spectrometer, PACS) at $z = 3.5$–$4.5$. For all these proxies, the underprediction of the observed LFs at $z = 1.8$–$3.5$ is related to overpredicting low-$z$ ones; this is particularly notorious for the stellar mass proxy even at the highest redshifts. Trying different SED templates for computing $k$-corrections, as described in Section 6.2, does not improve our results.

In this case, the observational distributions are presented only as a reference, since in principle we do not expect to recover similar LFs. Gruppioni et al. (2013) use samples from the Guaranteed Time Observation PACS Evolutionary Probe Survey, selected through their emission observed at 160 µm, whereas in our case only the 870 µm waveband is used as a selection criterion. In addition, our assumptions of fixed spectra may not represent galaxies at all the redshifts with which we are making the comparison.

### 6.5.4 *Descendants of model SMGs*

Once we determine the input counterparts for sources extracted from simulated maps, we follow their history in the SAG model until the present epoch.

A colour magnitude diagram for these descendants is shown in Fig. 15 for the three proxies, giving also the distribution for all $z = 0$ SAG galaxies with stellar mass over $10^8\,\mathrm{M_\odot}$. While the stellar mass proxy gives a significant population of blue-cloud galaxies, descendants for the MDS and SFR proxies lie in the red-sequence region. Note that in this approach the history of model $z = 0$ early-type galaxies is not followed backwards, so this does not necessarily ensures that the SMG phase is common to all early-type galaxies, nor that passing through an SMG phase is the only way to build up a red-sequence galaxy in our model.

Fig. 16 shows the distribution of stellar mass, virial halo mass, SFR and galaxy type for $z = 0$ descendants given by each proxy. As a reference, we also present observational data for the Milky Way: the stellar mass range computed by Flynn et al. (2006) using the local optical LF and the vertical structure of the Galaxy disc, the total mass within $\sim 200\,\mathrm{kpc}$ obtained by Bhattacharjee et al. (2014) measuring the rotation curve of the Galaxy, and the SFR determined by Murray & Rahman (2010) using the total Galaxy free–free emission.

While in all cases most of the descendants are central galaxies having SFRs in the range $0$–$10\,\mathrm{M_\odot\,yr^{-1}}$, the distribution of stellar masses is different for each proxy, being the lowest for the stellar mass proxy and the highest for the MDS one. Regarding the virial halo masses, the stellar mass proxy gives a very wide distribution, while the SFR and MDS proxies give mainly sources lying at and to higher values than the median $z = 0$ value predicted by Hickox et al. (2012) of $\log(M_{\mathrm{vir}}\,[\mathrm{M_\odot}/h]) = 13.3^{+0.3}_{-0.5}$. Comparing the SFR distribution for descendants with the one for model $z = 0$ ellipticals, or even extending the comparison to all model $z = 0$ galaxies (see the inset in Fig. 16), the average descendant of SMG tends to have a higher tail of high SFR; this is found regardless of the proxy. At this point it is worth noting the presence of outliers in the large sample of SAG model galaxies used to construct the light cones. Among these are some sources with red colours but SFRs over $\sim 10\,\mathrm{M_\odot\,yr^{-1}}$, which will leave the red sequence after a few tens of Myr.

## 7 SUMMARY

In light of the new findings regarding SMGs in the ECDF-S with ALMA, namely, the brightest single-dish detected sources being comprised by emission from multiple fainter sources (Karim et al. 2013), we develop a new technique for modelling SMGs, using light cones of galaxies drawn from the semi-analytic model SAG in a ΛCDM framework. In this paper, we introduce the Count Matching approach, where physical galaxy properties (or their combinations) given by the model are selected as proxies for the submm galaxy emission. Submm luminosities are assigned to a mass-limited sample of model galaxies assuming a monotonic relationship with the proxies, in such a way that a combination of LABOCA (Weiss et al. 2009) plus bright-end ALMA observed number counts (Karim et al. 2013) are reproduced. We make model catalogues given by each proxy pass through the observational process just as single-dish observations at 870 µm, allowing us to recover cumulative number counts as is done for actual data. Going back to the properties of counterparts for each extracted source (i.e. the injected sources whose coordinates are within a given search radius from each detection), we compare the recovered distributions of several galaxy properties with current observations from the literature.

Our main findings are the following.

(i) For the nine proposed proxies, there are lines of sight that give cumulative submm number counts consistent with LABOCA data
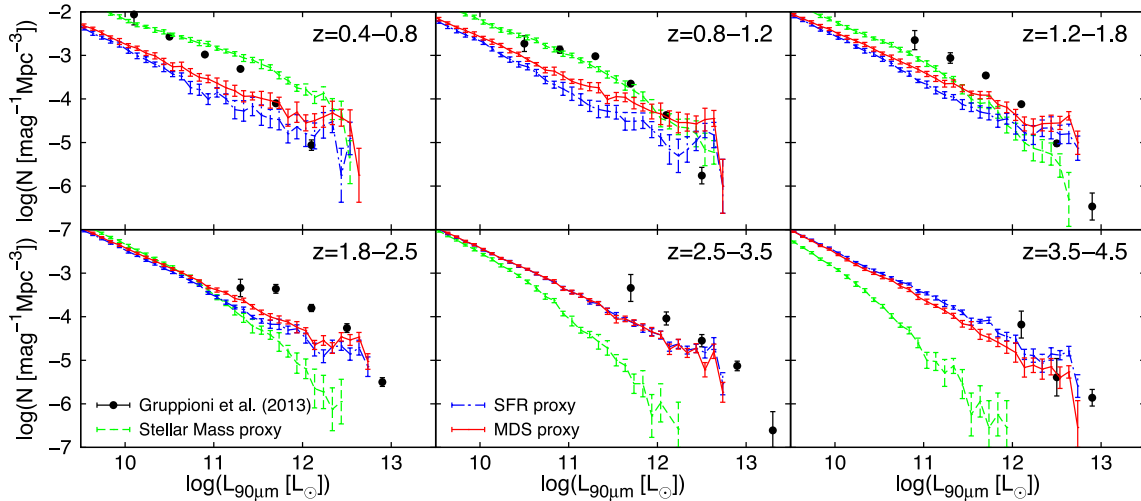
**Figure 14.** Rest-frame LFs at 90 µm predicted for the brightest input sources in each proxy, at six redshift ranges. Results are presented for MDS (red solid lines), stellar mass (green dashed lines) and SFR proxies (blue dot–dashed lines), averaging over the 10 light cones. Standard deviations are reported. Observational data by Gruppioni et al. (2013) of *Herschel* sources in the Guaranteed Time Observation PACS Evolutionary Probe Survey are also shown (black filled circles).

(Weiss et al. 2009). This is found even for maps having the same distribution of injected fluxes but randomized source coordinates, giving a hint that the clustering has a minor influence in the counts (see Fig. 6).

(ii) The majority of components of blended SMGs are spatially unassociated, meaning that most sources lying in a single line of sight are there by chance (see Fig. 11). Across the proxies, the fraction of multiple sources goes from ∼9 to ∼14 per cent, underpredicting the actual fraction for SMGs in the ECDF-S of ∼35–45 per cent (Hodge et al. 2013) but closer to the values found by Chen et al. (2013) in the CDF-N ($12.5^{+12.1}_{-6.8}$ per cent) and Smolčić et al. (2012) in the COSMOS field ($22 \pm 9$ per cent).

(iii) The different proxies retrieve a variety of distributions for a given galaxy property. We explore SMG distributions of redshift, stellar mass, host halo mass, SFR, sSFR, gas fraction, metallicity, effective radius and depletion time (see Figs 9 and 10, and Tables 3 and 4).

(iv) The SAG model underpredicts the individual measured SFRs by a factor of ∼3 when compared to actual SMG data from single-dish observations. This could be affected by some aspects in the modelling of star formation, like the complete removal of hot gas when galaxies become satellites.

(v) While the proxy where the submm luminosity increases monotonically with the ratio between dust mass and stellar age (the MDA proxy) is slightly better than the others when reproducing LABOCA cumulative counts, it fails to reproduce stellar and host halo masses, as well as the gas fraction and other properties.

(vi) The recovered redshift, stellar mass and host halo mass distributions (among others) for model SMGs are consistent with observations, when the submm luminosity increases monotonically with the product between dust mass and SFR (the MDS proxy). In this proxy most of the SMG $z = 0$ descendants are central galaxies with SFRs between 0 and $10 \, \mathrm{M_\odot \, yr^{-1}}$, with $u - r$ rest-frame colours lying in the red sequence (see Figs 15 and 16).

(vii) Using a fixed Arp 220 for all model galaxies, we are able to reproduce the rest-frame LF observed by Gruppioni et al. (2013) at 90 µm in the redshift range $z = 3.5$–4.5, but underpredicting it at $z = 1.8$–3.5. This is related to overpredicting low-$z$ distributions (see Fig. 14). In future work, computing LFs at other FIR wavebands
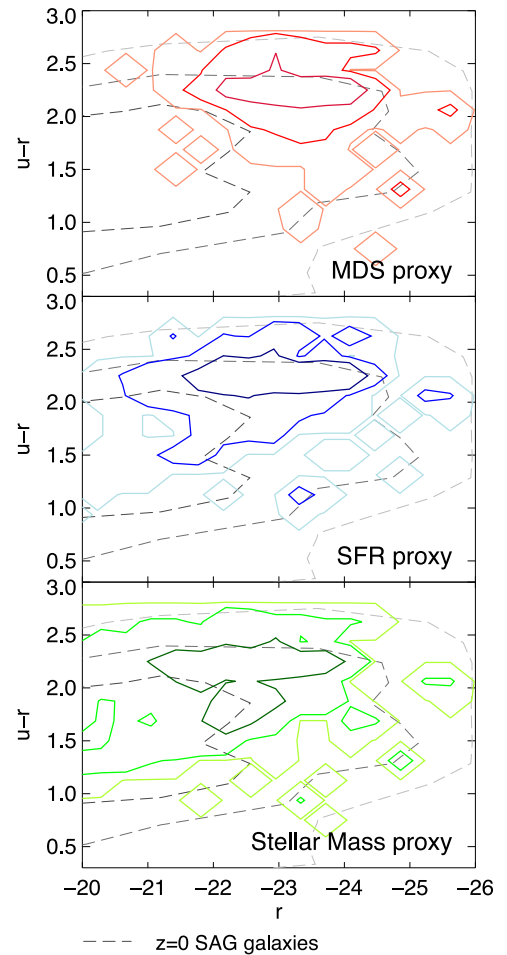


**Figure 15.** Colour magnitude diagram for descendants at $z = 0$ of extracted sources from all simulated maps. Magnitudes are in the rest frame. Same proxies as in Fig. 13 are considered, now compared to the distribution for all $z = 0$ SAG galaxies having stellar masses over $10^8 \, \mathrm{M_\odot}$ (grey dashed contours). From the darkest to the lightest, contours enclose 68, 95 and 99 per cent of the sources.
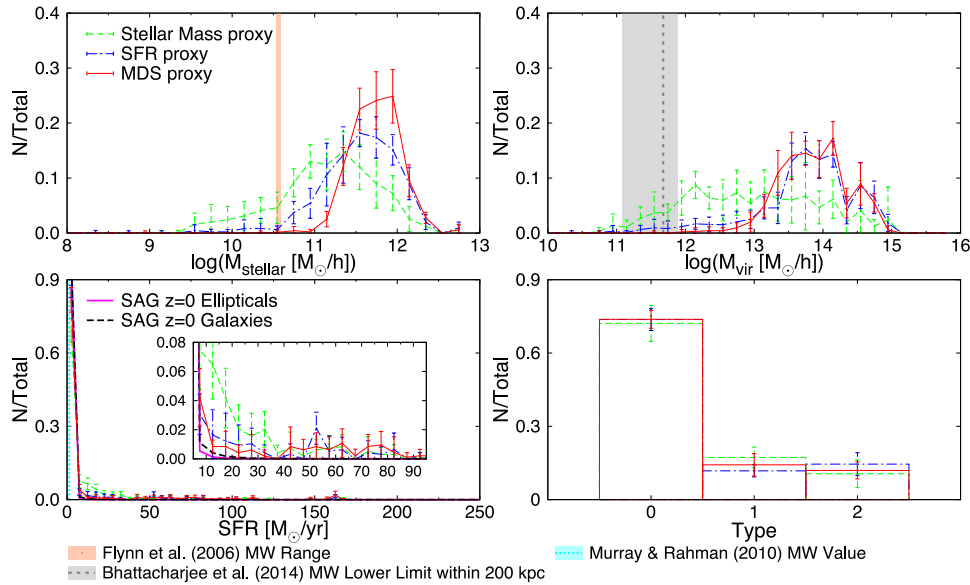
**Figure 16.** Distributions of stellar mass, virial halo mass, SFR and galaxy type for descendants at $z = 0$ of extracted sources. Histograms are averaged over the 10 light cones, taking the same proxies as in Fig. 13. In the bottom-left panel, the SFR distribution for model $z = 0$ ellipticals and for all model $z = 0$ galaxies are shown as magenta thick solid and black thick dashed lines, respectively. Galaxy types shown in the bottom-right panel correspond to: 0 for central galaxies, 1 for substructure satellites and 2 for satellites whose substructure has been lost, but which have not yet merged. As a reference, observational data for the Milky Way (MW) are shown: the stellar mass range computed by Flynn et al. (2006) as a light orange region, the total mass within ∼200 kpc obtained by Bhattacharjee et al. (2014) as a vertical grey short-dashed line, and the SFR determined by Murray & Rahman (2010) as a vertical cyan dotted line. The errors for observational values are presented as coloured shaded regions.

could also be used to determine a more precise proxy, specially if we use counts from larger surveys.

An immediate benefit of using this technique to assign submm luminosities to big galaxy samples is the low computational cost, unlike the use of full radiative transfer including dust over each galaxy, which computes the FIR emission in a self-consistent way. Our phenomenological approach gives results consistent with the Hayward et al. (2013b) prediction regarding the lack of spatial association between blended galaxies, and we see a remarkable similarity between the predictions of the MDS proxy with the fitting function used in their paper to compute the submm flux, which was obtained using results of the SUNRISE dust radiative transfer code (Jonsson 2006) on hydrodynamical simulations (see Hayward et al. 2013a for details). The high proportion of spatially unassociated sources has also qualitative agreement with Cowley et al. (2014) results, where the galaxy properties (including submm flux) were obtained using a semi-analytic model plus a dust model that resembles the treatment performed by the GRASIL spectrophotometric code.

Further improvements in the Count Matching approach could consider the enhancement in mass resolution for the dark matter simulation on which the semi-analytic model is run, the increase in the simulation volume, the incorporation of AGN properties as proposed proxies and the inclusion of a more gradual star formation during starbursts (Gargiulo et al. 2014) or cold gas inflows in the SAG model. When modelling the observational process, it is also possible to quantify the influence of gravitational lensing on the fluxes and redshift distribution of detected sources, including the boosting of higher redshift sources above the survey limit (Weiss et al. 2013); nevertheless, lensing would affect number counts well beyond our covered range for this field and wavelength, according to models (Lima et al. 2010; Er et al. 2013).

The flexibility in the methodology underlying the Count Matching approach can be extended to other galaxy populations and/or wavelength ranges and/or observed fields to be modelled, as well as to other galaxy formation models and/or cosmology.

## REFERENCES

Andrews B. H., Martini P., 2013, ApJ, 765, 140
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945

Barger A. J., Cowie L. L., Sanders D. B., Fulton E., Taniguchi Y., Sato Y., Kawara K., Okuda H., 1998, Nature, 394, 248

Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Béthermin M. et al., 2012, ApJ, 757, L23

Bhattacharjee P., Chaudhury S., Kundu S., 2014, ApJ, 785, 63

Blain A. W., 1999, MNRAS, 304, 669

Blain A. W., Smail I., Ivison R. J., Kneib J.-P., 1999, MNRAS, 302, 632

Blain A. W., Smail I., Ivison R. J., Kneib J.-P., Frayer D. T., 2002, Phys. Rep., 369, 111

Borys C., Smail I., Chapman S. C., Blain A. W., Alexander D. M., Ivison R. J., 2005, ApJ, 635, 853

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Chapman S. C., Smail I., Windhorst R., Muxlow T., Ivison R. J., 2004, ApJ, 611, 732

Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, ApJ, 622, 772

Chen C. C., Cowie L. L., Barger A. J., Casey C. M., Lee N., Sanders D. B., Wang W.-H., Williams J. P., 2013, ApJ, 776, 131

Chen C. C., Cowie L. L., Barger A. J., Wang W.-H., Williams J. P., 2014, ApJ, 789, 12

Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

Coppin K. et al., 2006, MNRAS, 372, 1621

Cora S. A., 2006, MNRAS, 368, 1540

Cowley W. I., Lacey C. G., Baugh C. M., Cole S., 2014, MNRAS, preprint (arXiv:1406.0855)

Croton D. J. et al., 2006, MNRAS, 365, 11

Davé R., Finlator K., Oppenheimer B. D., Fardal M., Katz N., Kereš D., Weinberg D. H., 2010, MNRAS, 404, 1355

De Lucia G., Kauffmann G., White S. D. M., 2004, MNRAS, 349, 1101

Dekel A. et al., 2009, Nature, 457, 451

Er X. et al., 2013, MNRAS, 430, 1423

Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813

Flynn C., Holmberg J., Portinari L., Fuchs B., Jahreiß H., 2006, MNRAS, 372, 1149

Fontanot F., Monaco P., Silva L., Grazian A., 2007, MNRAS, 382, 903

Gargiulo I. D. et al., 2014, MNRAS, preprint (arXiv:1402.3296)

Garnett D. R., 2002, ApJ, 581, 1019

González J. E., Lacey C. G., Baugh C. M., Frenk C. S., Benson A. J., 2009, MNRAS, 397, 1254

Graham A. W., Driver S. P., Petrosian V., Conselice C. J., Bershady M. A., Crawford S. M., Goto T., 2005, AJ, 130, 1535

Granato G. L., Lacey C. G., Silva L., Bressan A., Baugh C. M., Cole S., Frenk C. S., 2000, ApJ, 542, 710

Gruppioni C. et al., 2013, MNRAS, 432, 23

Hatsukade B., Ohta K., Seko A., Yabe K., Akiyama M., 2013, ApJ, 769, L27

Hayward C. C., Kereš D., Jonsson P., Narayanan D., Cox T. J., Hernquist L., 2011, ApJ, 743, 159

Hayward C. C., Narayanan D., Kereš D., Jonsson P., Hopkins P. F., Cox T. J., Hernquist L., 2013a, MNRAS, 428, 2529

Hayward C. C., Behroozi P. S., Somerville R. S., Primack J. R., Moreno J., Wechsler R. H., 2013b, MNRAS, 434, 2572

Hickox R. C. et al., 2012, MNRAS, 421, 284

Hodge J. A. et al., 2013, MNRAS, 768, 91

Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, preprint (astro-ph/0210394)

Holland W. S. et al., 1999, MNRAS, 303, 659

Hughes D. H. et al., 1998, Nature, 394, 241

Jarosik N. et al., 2011, ApJS, 192, 14

Jonsson P., 2006, MNRAS, 372, 2

Karim A. et al., 2013, MNRAS, 432, 2

Kennicutt R. C., 1983, ApJ, 272, 54

Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183

Lagos C. del P., Cora S. A., Padilla N. D., 2008, MNRAS, 388, 587

Laird E. S., Nandra K., Pope A., Scott D., 2010, MNRAS, 401, 2763

Lima M. et al., 2010, ApJ, 717, L31

Magdis G. E. et al., 2012, ApJ, 760, 6

Magnelli B. et al., 2013, A&A, 553, A132

Michałowski M., Hjorth J., Watson D., 2010, A&A, 514, A67

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macci A. V., Naab T., Oser L., 2010, ApJ, 710, 903

Murray N., Rahman M., 2010, ApJ, 709, 424

Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793

Noordermeer E., van der Hulst J. M., Sancisi R., Swaters R. A., van Albada T. S., 2005, A&A, 442, 137

Padilla N. D., Salazar-Albornoz S., Contreras S., Cora S. A., Ruiz A. N., 2014, MNRAS, 443, 2801

Ruiz A. N. et al., 2014, ApJ, preprint (arXiv:1310.7034)

Rujopakarn W., Rieke G. H., Eisenstein D. J., Juneau S., 2011, ApJ, 726, 93

Salpeter E. E., 1955, ApJ, 121, 161

Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749

Scott K. S. et al., 2008, MNRAS, 385, 2225

Shankar F., Marulli F., Bernardi M., Dai X., Hyde J. B., Sheth R. K., 2010, MNRAS, 403, 117

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Simpson J. M. et al., 2014, ApJ, 788, 125

Siringo G. et al., 2009, A&A, 497, 945

Smail I., Chapman S. C., Blain A. W., Ivison R. J., 2004, ApJ, 616, 71

Smolčić V. et al., 2012, A&A, 548, A4

Somerville R. S., Gilmore R. C., Primack J. R., Domnguez A., 2012, MNRAS, 423, 1992

Springel V., 2005, MNRAS, 364, 1105

Swaters R. A., Balcells M., 2002, A&A, 390, 863

Swinbank A. M., Smail I., Chapman S. C., Blain A. W., Ivison R. J., Keel W. C., 2004, ApJ, 617, 64

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., 2008, MNRAS, 391, 420

Swinbank A. M. et al., 2014, MNRAS, 438, 1267

Tacconi L. J. et al., 2006, ApJ, 640, 228

Tacconi L. J. et al., 2008, ApJ, 680, 246

Tacconi L. J. et al., 2010, Nature, 463, 781

Targett T. A. et al., 2013, MNRAS, 432, 2012

Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. D. P., 2010, MNRAS, 408, 2008

Tremonti C. A. et al., 2004, ApJ, 613, 898

Wang S. X. et al., 2013, ApJ, 778, 179

Wardlow J. L. et al., 2011, MNRAS, 415, 1479

Weiss A. et al., 2009, ApJ, 707, 1201

Weiss A. et al., 2013, ApJ, 767, 88

## APPENDIX A: PREDICTIONS FROM THE SAG MODEL

The semi-analytic model involves a set of free parameters that regulate the action of the different physical processes, namely gas cooling, star formation, feedback from core-collapse supernova explosions and AGN, galaxy mergers and chemical enrichment of baryons. The values of these parameters are tuned by using a calibration method based on the Particle Swarm Optimization technique (Ruiz et al. 2014). Observational constraints for this procedure are $z = 0$ LFs at $b_J$ and $r$ bands, and the black hole–bulge mass relation. Rest-frame magnitudes in several filters from the UV to the NIR are computed considering CB07 evolutionary synthesis models (which are an update of Bruzual & Charlot 2003 models) for different cold gas metallicities and assuming a Salpeter IMF. This magnitudes are corrected by dust extinction following De Lucia et al. (2004). To
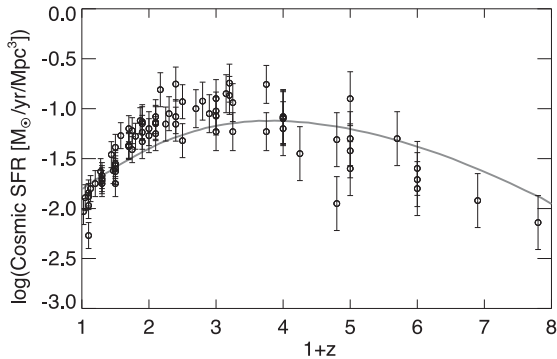
**Figure A1.** Cosmic SFR as a function of redshift: predicted by SAG as a grey line, compared to observational data compiled by Behroozi et al. (2013) shown as open circles.

validate the model used throughout this work, we present some of the predicted galaxy properties.

Fig. A1 shows the evolution of the cosmic SFR density, compared to observational data compiled by Behroozi et al. (2013) scaled to our adopted IMF. Our model predicts consistently the rise and fall of the cosmic SFR going from higher to lower redshifts, although at $z = 0$ we have some excess compared to observations. Compared to Magnelli et al. (2013) data based on *Herschel* observations (corrected to a Salpeter IMF as well), we underpredict the cosmic SFR by a factor of 1.2 at $z = 0$, and by a factor of 3 in the range $z = 1$–2. Note that this does not minimize the potential of the Count Matching technique, since when performing it we are only concerned about reproducing the observed trend for the cosmic SFR across redshift (not the exact values) which is successfully achieved with the model used.

Fig. A2 shows the predicted mass–metallicity relation at $z = 0$–2. Compared to Tremonti et al. (2004) and Andrews & Martini (2013) data, our cold gas phase metallicities at $z = 0$ are systematically shifted downwards.

If we consider the wide spread in the fit to the mass–metallicity relation given by different metallicity calibrations (Kewley & Ellison 2008), the gas metallicities from our model can be considered acceptable. The low gas metallicity values that characterize $z = 0$ model galaxies might be related to the low SFR in the redshift range $z = 2$–3 as is evident from the marginal agreement with observational data depicted in Fig. A1; the number of core-collapse supernovae, which are the main source of oxygen, is directly related with the SFR. Levels of star formation may increase by considering the effects of accretion of material with misaligned angular momenta on the gaseous disc from which stars are formed (Padilla et al. 2014). On the other hand, feedback from supernova explosions plays a crucial role in determining the mass–metallicity relation. Both these processes are far from being well understood and their modelization needs improvements.

Considering galaxies at $z = 2$, we predict slightly higher cold gas metallicities when compared to Erb et al. (2006) data. In any case, when applying the Count Matching technique we are interested only in the trends followed among properties, and not in the exact values; in this sense, the observed trend between mass and metallicity is well reproduced. This also applies to the ratio between cold gas and stellar mass as a function of stellar mass at $z = 0$, shown in Fig. A3. In our model, this ratio follows the observed trend from measurements by Garnett (2002), Swaters & Balcells (2002), Noordermeer et al. (2005) and Baldry et al. (2008), but underpredicts it by a factor of ∼3.
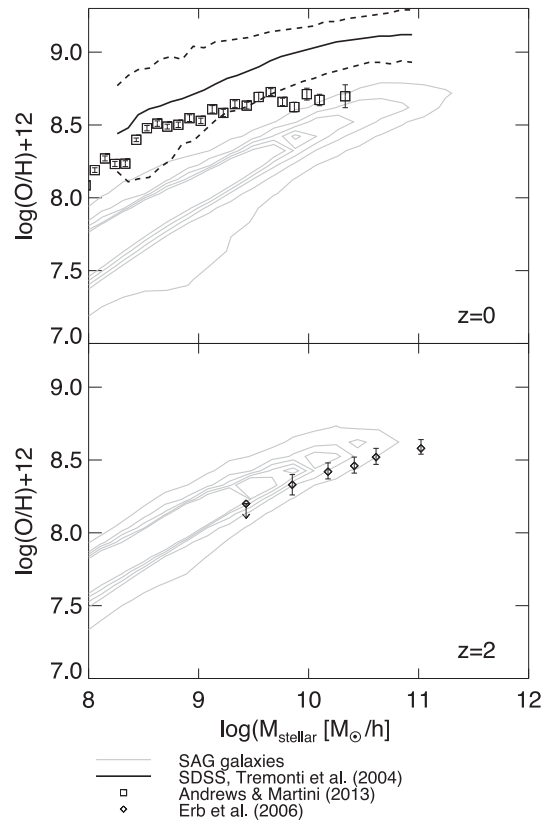
**Figure A2.** Mass–metallicity relation predicted by SAG (contours). At $z = 0$ (top panel), the model distribution is compared to the median value measured by Tremonti et al. (2004) as a black solid line, with dashed lines as the 68 percentile and Andrews & Martini (2013) in black squares using SDSS SFGs, while at $z = 2$ (bottom panel) is compared to data from Erb et al. (2006) using UV-selected SFGs.
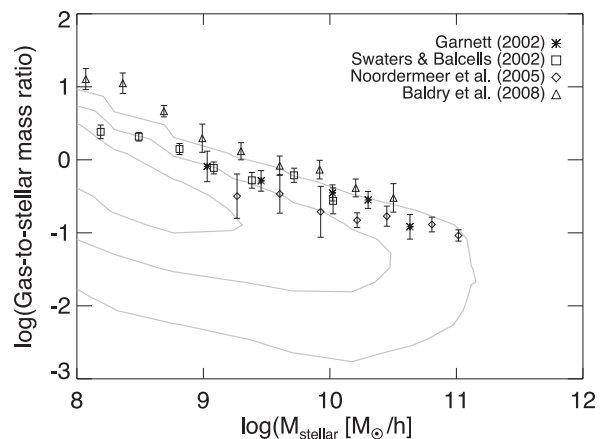


**Figure A3.** Cold gas to stellar mass ratio as a function of stellar mass. Contours: predictions from SAG. Symbols: measured values by Garnett (2002) as asterisks, Swaters & Balcells (2002) as open squares, Noordermeer et al. (2005) as open diamonds and Baldry et al. (2008) as open triangles.

## APPENDIX B: MODEL FOR BULGE SIZES

We implement a model for bulge sizes in SAG, based on the one introduced by Cole et al. (2000) with further improvements by González et al. (2009). In this approach, virial equilibrium and energy conservation are considered for computing the bulge size of a given galaxy.

Following González et al. (2009), the bulge size is computed when galaxy mergers and disc instabilities occur. When there is a merger, the bulge radius of the remnant galaxy $r_{\text{new, merger}}$ is given by

$$\frac{(M_1 + M_2)^2}{r_{\text{new,merger}}} = \frac{M_1^2}{r_1} + \frac{M_2^2}{r_2} + \frac{f_{\text{orbit}}}{\bar{c}} \frac{M_1 M_2}{r_1 + r_2}, \tag{B1}$$

where the subscripts 1 and 2 stand for the central and satellite galaxies, respectively. In the case of a major merger, the masses involved are $M_i = M_{i,\text{stellar}} + M_{i,\text{coldgas}} + f_{\text{dark}} M_{i,\text{DM}}$ (with $i = 1, 2$), while for minor mergers are $M_1 = M_{1,\text{stellar, bulge}} + f_{\text{dark}} M_{1,\text{DM}}$ and $M_2 = M_{2,\text{stellar}} + f_{\text{dark}} M_{2,\text{DM}}$. Note that in all cases there is a contribution from the dark matter halo mass $M_{\text{DM}}$, determined by a factor with fiducial value $f_{\text{dark}} = 2$. The form factor adopted is $\bar{c} = 0.5$. The sizes involved are radii weighted by mass, i.e. $r_i = (M_{i,\text{bulge}} r_{i,\text{bulge}} + M_{i,\text{disc}} r_{i,\text{disc}})/(M_{i,\text{bulge}} + M_{i,\text{disc}})$ (with $i = 1, 2$).

When there is a disc instability, the bulge size $r_{\text{new, instab}}$ is given by

$$\frac{c_B (M_{\text{bulge}} + M_{\text{disc}})^2}{r_{\text{new,instab}}} = \frac{c_B M_{\text{bulge}}^2}{r_{\text{bulge}}} + \frac{c_D M_{\text{disc}}^2}{r_{\text{disc}}} + f_{\text{int}} \frac{M_{\text{bulge}} M_{\text{disc}}}{r_{\text{bulge}} + r_{\text{disc}}}, \tag{B2}$$

where the disc mass includes the contribution of both the stellar and cold gas components. As in Cole et al. (2000), the adopted factors are $c_D = 0.49$, $c_B = 0.45$ and $f_{\text{int}} = 2$.

In this work, we introduce an update of the bulge size when a starburst occurs, inspired by the model for disc instabilities. In this case, the bulge size immediately after the starburst event $r_{\text{new, starb}}$ is given by

$$\frac{c_B (M_{\text{bulge}} + M_{\text{starb}})^2}{\Delta r_{\text{new,starb}}} = \frac{c_B M_{\text{bulge}}^2}{r_{\text{bulge}}} + \frac{c_D M_{\text{starb}}^2}{r_{\text{starb}}} + f_{\text{int}} \frac{M_{\text{bulge}} M_{\text{starb}}}{r_{\text{bulge}} + r_{\text{starb}}}, \tag{B3}$$

$$r_{\text{new,starb}} = r_{\text{bulge}} + \Delta r_{\text{new,starb}}, \tag{B4}$$

where $M_{\text{starb}}$ is the mass of stars formed in the starburst and $r_{\text{starb}}$ is a characteristic scalelength associated with it. For simplicity, we take
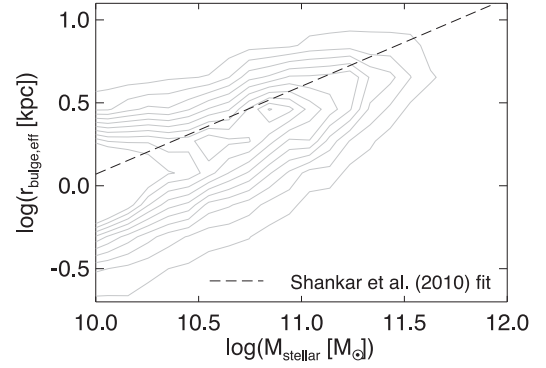


**Figure B1.** Size–stellar mass relation: for elliptical galaxies at $z = 0$ in the SAG model as contours, and the linear fit for a sample of SDSS ellipticals obtained by Shankar et al. (2010) as a dashed black line.

$r_{\text{starb}} = r_{\text{disc}}$. The adopted factors $c_D$, $c_B$ and $f_{\text{int}} = 2$ are the same as for disc instabilities.

For validating this model in SAG, we present our predicted size–stellar mass relation for elliptical galaxies at $z = 0$ in Fig. B1. The size is characterized by the bulge effective radius; the projected density is assumed to be well described by a de Vaucouleurs profile, $\rho(r) \propto r^{1/4}$. Our model is in good agreement with the linear fit obtained by Shankar et al. (2010) for a sample of ellipticals in SDSS.

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