HAL
open science

EGG: hatching a mock Universe from empirical prescriptions

C. Schreiber, D. Elbaz, M. Pannella, E. Merlin, M. Castellano, A. Fontana, N. Bourne, K. Boutsia, F. Cullen, J. Dunlop, et al.

To cite this version:

C. Schreiber, D. Elbaz, M. Pannella, E. Merlin, M. Castellano, et al.. EGG: hatching a mock Universe from empirical prescriptions. Astronomy and Astrophysics - A&A, 2017, 602, pp.A96. 10.1051/0004-6361/201629123 . hal-03112459

HAL Id: hal-03112459

https://hal.science/hal-03112459v1

Submitted on 16 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A&A 602, A96 (2017)
DOI: 10.1051/0004-6361/201629123
© ESO 2017

Astronomy & Astrophysics

EGG: hatching a mock Universe from empirical prescriptions[⋆,⋆⋆]

C. Schreiber[1,2], D. Elbaz[2], M. Pannella[3,2], E. Merlin[4], M. Castellano[4], A. Fontana[4], N. Bourne[5], K. Boutsia[4], F. Cullen[5], J. Dunlop[5], H. C. Ferguson[6], M. J. Michałowski[5], K. Okumura[2], P. Santini[4], X. W. Shu (舒新文)[7,2], T. Wang (王涛)[2,8], and C. White[9]

[1] Leiden Observatory, Leiden University, 2300 RA Leiden, The Netherlands
e-mail: cschreib@strw.leidenuniv.nl
[2] Laboratoire AIM-Paris-Saclay, CEA/DSM/Irfu – CNRS – Université Paris Diderot, CEA-Saclay, pt courrier 131, 91191 Gif-sur-Yvette, France
[3] Faculty of Physics, Ludwig-Maximilians Universität, Scheinerstr. 1, 81679 Munich, Germany
[4] INAF–Osservatorio Astronomico di Roma, via Frascati 33, 00040 Monte Porzio Catone (RM), Italy
[5] SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, UK
[6] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[7] Department of Physics, Anhui Normal University, Wuhu, Anhui 241000, PR China
[8] School of Astronomy and Astrophysics, Nanjing University, Nanjing 210093, PR China
[9] Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

Received 15 June 2016 / Accepted 10 January 2017

ABSTRACT

This paper introduces EGG, the Empirical Galaxy Generator, a tool designed within the ASTRODEEP collaboration to generate mock galaxy catalogs for deep fields with realistic fluxes and simple morphologies. The simulation procedure is based exclusively on empirical prescriptions – rather than first principles – to provide the most accurate match with current observations at $0 < z < 7$. We considered that galaxies can be either quiescent or star-forming, and used their stellar mass (M_*) and redshift (z) as the fundamental properties from which all the other observables can be statistically derived. Drawing z and M_* from the observed galaxy stellar mass functions, a star-formation rate (SFR) is attributed to each galaxy from the tight SFR–M_* main sequence, while dust attenuation, optical colors and simple disk plus bulge morphologies are obtained from empirical relations that we established from the high quality *Hubble* and *Herschel* observations from the CANDELS fields. Random scatter was introduced in each step to reproduce the observed distributions of each parameter. Based on these observables, an adequate panchromatic spectral energy distribution (SED) is selected for each galaxy and synthetic photometry is produced by integrating the redshifted SED in common broad-band filters. Finally, the mock galaxies are placed on the sky at random positions with a fixed angular two-point correlation function to implement basic clustering. The resulting flux catalogs reproduce accurately the observed number counts in all broad bands from the ultraviolet up to the sub-millimeter, and can be directly fed to image simulators such as SkyMaker. The images can then be used to test source extraction softwares and image-based techniques such as stacking. EGG is open-source, and is made available to the community on behalf of the ASTRODEEP collaboration, together with a set of pre-generated catalogs and images.

Key words. galaxies: photometry – galaxies: evolution – galaxies: statistics – galaxies: structure

1. Introduction

To a large extent, most of our knowledge of astronomy and astrophysics is derived from two- (or three-) dimensional images of the sky, acquired by observatories in space or on the ground. Different instruments will generally produce images of strongly varying properties, including (but not limited to) zero point calibration, point spread function (PSF), noise, or sky-to-pixel projection. Therefore, extracting observables of astronomical interest – namely fluxes, shapes, positions and the respective uncertainties – requires a good knowledge of the instrument and the image reduction pipeline. For this reason, these observables are typically compiled into catalogs, which can be used with minimal knowledge of the instrument or the image itself and allow a more immediate scientific analysis. Building these catalogs is not straightforward, and various techniques and tools have been introduced during the history of astronomy, ranging in complexity from aperture photometry to multi-component profile fitting (e.g., Bertin & Arnouts 1996; Labbé et al. 2006; De Santis et al. 2006; Laidler et al. 2007; Wuyts et al. 2007; Merlin et al. 2015, to only mention most recent efforts).

In this context, the goal of the ASTRODEEP collaboration[1] is to provide the astronomy community with optimally extracted flux catalogs in the most data-rich cosmological deep fields (i.e., the GOODS and CANDELS fields). To achieve this goal, new photometric codes and techniques are being developed (see, e.g., Merlin et al. 2015; Cappelluti et al. 2016; Shu et al. 2016; Merlin et al. 2016; Castellano et al. 2016; Wang et al., in prep.) to improve on existing practices both in efficiency and accuracy.

A key step in the conception process of such codes is to characterize their performance and accuracy before applying them to real images, checking not only the robustness of the flux

⋆ http://cschreib.github.io/egg/
⋆⋆ Full Table C.1 is only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/602/A96

[1] http://www.astrodeep.eu/

measurements, but also the quality of the error estimates. In addition, the challenge of extracting photometry in far-IR and sub-millimeter images is tied to the large size of the PSF, which generates confusion noise (e.g., Dole et al. 2004, and references therein). In this situation the main issue is the choice of an optimal strategy to select the prior positions at which the sources will be extracted, since using the positions of all known galaxies would result either in over-fitting or a degenerate fit. For this reason, a secondary goal of ASTRODEEP is to provide the astrophysics community with realistic simulations of the sky at different wavelengths and with different angular resolutions, so that astronomers can test their procedures and tools and quantify their respective efficiency. These simulations must provide two essential components: first, simulated images including realistic noise properties, and second, the corresponding mock galaxy catalogs containing the true flux of each object. The resulting mock observations should be as close as possible to the real ones, including in particular the correct flux and color distributions.

Various approaches can be used to generate such mock galaxy catalogs. We separate them in two main classes: physically-motivated approaches on the one hand, and empirical approaches on the other hand. Procedures that belong to the first class are typically based on the output of large-scale cosmological simulations, such as Illustris (Genel et al. 2014), or semi-analytic models (SAMs). These simulations provide dark matter halo populations with realistic redshift-dependent spatial and mass distribution, since the physics governing the growth of these halos is well understood. However, there are two main drawbacks linked to these approaches. First, the physical properties of the galaxies created by such simulations do not always match the observations (e.g., the systematic factor of two underestimation of the star-formation activity in $z = 2$ galaxies; Daddi et al. 2007; Gruppioni et al. 2015); and second, such simulations are computationally expensive and can render some tests impractical if a large number of random realizations is required.

The second class of approach aims at reproducing the observations by construction, sacrificing the consistency of the physics to reach higher fidelity of the mock data. The easiest way to achieve this is to construct a flux distribution from the observed number counts and use it to draw fluxes randomly for each galaxy (see, e.g., the *Stuff* program introduced in Bertin 2009). The process is extremely fast and can be used to generate data sets potentially larger than the observable Universe. But this first order approach also has limitations: first, linking the different photometric bands together and ensuring that colors and their scatter are properly reproduced is non-trivial; and second, it is impractical to extend this model to photometric bands with poorly constrained (or inexistent) number counts. Therefore, a more generic and successful approach would instead generate a realistic spectrum for each galaxy, from which broad band photometry can be immediately derived (e.g., Franceschini et al. 2001; Béthermin et al. 2011; Gruppioni et al. 2011; Béthermin et al. 2012). Choosing the spectrum and luminosity for a particular galaxy then becomes the central question. Galaxy spectra are known to be the sum of multiple components, including mainly stellar, nebular and dust emission. Active galactic nuclei (AGNs) in their various phases can also become the dominant light source in some specific wavelength domains (e.g., Hao et al. 2005; Richards et al. 2006; Hatziminaoglou et al. 2010). Constructing the spectrum of a galaxy hence requires a description of these various components.

To this end, in the present paper we propose a set of empirical relations and recipes that we calibrate on the deepest observations available to date in the *Hubble* CANDELS fields. Crucially, these fields are also covered by the deepest *Herschel* observations, which allow us to derive precise constraints on the dust-obscured properties of galaxies up to $z \sim 3$. We implement these recipes in a new tool named EGG (the Empirical Galaxy Generator) to build realistic mock galaxy catalogs from the UV to the submillimeter.

Briefly, the process we employ considers that galaxies can first be segregated into two broad populations of "star-forming" (SFGs) and "quiescent" galaxies (QGs). We generate galaxies from these two populations based on their observed stellar mass functions. Then, all the other physical properties are statistically inferred from the SFG versus QG classification, the stellar mass, and the redshift of these galaxies. We use these three parameters as the driving factors for a number of observables such as morphology (e.g., size, or bulge vs. disk mass ratio), optical colors, dust attenuation and star-formation rate. To fine tune the fidelity of the final data set, we introduce second order variations by adding a controlled amount of random scatter to most of the observables.

The outline of this paper is the following. In Sect. 2 we describe the observational data set that we use to calibrate our recipes. This includes a brief description of the fields and the methods used to derive the physical parameters of each observed galaxy. In Sect. 3 we describe the recipes to derive the stellar emission, starting from the stellar mass functions (Sect. 3.1), the morphology of the stellar profile (Sect. 3.2), and the optical colors (Sect. 3.3). In Sect. 4 we present our description of the dust emission, including the parametrization of star-formation rate (Sect. 4.1) and obscuration (Sect. 4.2), and the properties of dust, such as its temperature and chemical composition (Sect. 4.4). Section 5 describes the approach we use to generate realistic sky position distributions, including clustering. Section 6 gives an brief overview of how the simulation is assembled and implemented in the code. The resulting catalogs and images are compared to observations in Sects. 7.1 and 7.2, and to recent SAMs in Sect. 7.3. Examples of simulated images are given in Sect. 7.4. Finally, Sect. 8 discusses the limitations of the simulation, and directions to improve it.

In the following, we assume a ΛCDM cosmology with $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\mathrm{M} = 0.3$, $\Omega_\Lambda = 0.7$ and a Salpeter (1955) initial mass function (IMF), to derive both star-formation rates and stellar masses. All magnitudes are quoted in the AB system, such that $M_\mathrm{AB} = 23.9 - 2.5\log_{10}(S_\nu\ [\mu\mathrm{Jy}])$.

2. Sample and observations

We based this analysis on the sample and data described in Schreiber et al. (2015, hereafter S15). In this section, we make a brief summary of these observations.

2.1. Multi-wavelength photometry

The catalogs we used in this work are the official catalogs produced by CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) from the *Hubble* Space Telescope (HST) WFC3 *H*-band imaging of the four fields also covered by deep *Herschel* PACS and SPIRE observations, namely GOODS-North (Barro et al., in prep.), GOODS-South (Guo et al. 2013), UDS (Galametz et al. 2013) and COSMOS (Nayyeri et al., in prep.). Each of these fields is about 150 arcmin^2 and they are evenly distributed on the sky to mitigate cosmic variance.

The ancillary photometry varies from one field to another, being a combination of both space- and ground-based imaging

from various facilities. It is described in detail in the catalog papers cited above, as well as in S15. Briefly, the UV to near-IR wavelength coverage typically goes from the U band up the *Spitzer* IRAC 8 μm, including at least the HST bands *F606W*, *F814W*, and *F160W* and a deep K (or K_s) band, and all these images are among the deepest available views of the sky. These catalogs therefore cover most of the important galaxy spectral features across a wide range of redshifts, even for intrinsically faint objects.

We complemented these catalogs with mid-IR photometry from *Spitzer* MIPS and far-IR photometry from *Herschel* PACS and SPIRE taken as part of the GOODS–*Herschel* (Elbaz et al. 2011) and CANDELS–*Herschel* programs (Inami et al., in prep.).

2.2. Redshifts, stellar masses and rest-frame optical colors

From this observed photometry, photometric redshifts and stellar masses were computed following Pannella et al. (2015). The details of the fitting procedure can be found there and in S15, and we only provide a brief overview in the following.

We used EAzY (Brammer et al. 2008) to derive the photometric redshifts from the CANDELS catalogs, allowing slight adjustments of the photometric zero points by iteratively comparing our photo-zs against the available spec-zs. The stellar masses were then computed using FAST (Kriek et al. 2009) by fixing the redshift to the best-fit photo-z and fitting the observed photometry up to the IRAC 4.5 μm band using the Bruzual & Charlot (2003) stellar population synthesis model, assuming a Salpeter (1955) IMF and a Calzetti et al. (2000) extinction law, allowing a range of attenuation with $A_V = 0$ up to 4. All galaxies were described with a delayed exponentially declining star formation history (SFH), with $\mathrm{SFR}(t) \sim t \exp(-t/\tau)$, and we allow both the age and the exponential time-scale τ to vary. This parametrization allows for both rising, declining and bursty SFHs, depending on the value of the age and age/τ, and therefore covers a large parameter space.

As shown in Pannella et al. (2015, their Appendix C) or Santini et al. (2015, their Fig. 2), the precise choice of the SFH does not significantly impact the stellar mass (unless single bursts are used, see Michałowski et al. 2014). Furthermore, the wavelength coverage in the CANDELS fields is such that the rest-frame UV-to-optical is properly sampled up to $z \sim 7$, thanks in particular to the *Spitzer* IRAC bands, and therefore our mass estimates should not suffer from significant redshift-dependent systematics (see, e.g., Santini et al. 2015; Grazian et al. 2015).

Regardless, the present work is largely insensitive to the accuracy of these redshift and mass determinations. One way to see this is to consider that, throughout this work, we essentially used z and M_* to generate a simpler 2-dimensional space on which any galaxy can be projected. The recipes we describe in the next sections were designed to map a point from this "projected" space back into a much larger space of observables (i.e., all the fluxes and morphological parameters) to reproduce the observed distributions. In this sense, z and M_* can be seen as arbitrary intermediate variables, and their physical meaning (or correctness) is not important. It should therefore be immediately apparent that any systematic error, e.g., in our stellar masses, will simply cancel out by construction and still produce the right distributions for the observables. This is only true because we derived our recipes and stellar mass functions from the same data set (except for the relation between B/T and M_* which we took from Lang et al. 2014, see Sect. 3.2, however their data set and methodology are very similar to ours).

In this work (and as in S15), we only considered galaxies with $H < 26$ to ensure a high quality SED and photometric redshifts. When appropriate, we took into account the selection effects resulting from this magnitude cut (e.g., on the stellar mass functions and mass completeness limits). Sources with an uncertain photometric redshift (redshift `odds` less than 0.8, as given by EAzY) or bad SED fitting (reduced χ^2 larger than ten) were excluded from the present analysis. These represent from 3 to 6% of our sample, depending on the stellar mass and redshift range, and their impact on our results are therefore marginal. We also explicitly removed foreground stars from the catalogs, using a combination of morphology and BzK colors, as in Pannella et al. (2015).

Lastly, the rest-frame U, V and J magnitudes were computed for each galaxy using EAzY, by integrating the best-fit galaxy template from the photo-z estimation. These colors were used, following Williams et al. (2009), to separate those galaxies that are "quiescent" (QGs) from the "star-forming" ones (SFGs). We used the same selection criteria as in S15, that is, a galaxy is deemed star-forming if its colors satisfy

$$UVJ_{\mathrm{SF}} = \begin{cases} U - V & < & 1.3\,, \text{ or} \\ V - J & > & 1.6\,, \text{ or} \\ U - V & < & 0.88 \times (V - J) + 0.49, \end{cases} \tag{1}$$

otherwise the galaxy is considered as quiescent.

3. Stellar properties

3.1. Conditional stellar mass functions

The initial purpose of EGG is to simulate a deep field similar to the GOODS-South field. Therefore, we computed the stellar mass function in this field only, in order to most closely mimic is properties (including, in particular, cosmic variance). To do so, we used the procedure described in S15, which we now briefly recall.

As stated in Sect. 2.2, galaxies from the GOODS-South catalog were selected with $H < 26$ to ensure high quality photometry for all galaxies. This mostly eliminates the effect of the Eddington bias. As in S15, we estimated the evolution of the stellar mass completeness (90%) corresponding to this selection, as inferred from the observed scatter in the M_*-to-$L_{H/(1+z)}$ ratio for SFGs and QGs separately. We found, for example, that at $z = 1$ the completeness is as low as $5 \times 10^8\ M_\odot$ for SFGs. We used this information to correct the counts for incompleteness, and did not attempt to measure the stellar mass function when this correction became larger than a factor of 2. We then defined multiple redshift bins from $z = 0.3$ to $z = 4.5$, and computed within each of these bins the mass distribution for both SFGs and QGs separately, according to the UVJ color–color selection (see Sect. 2.2). Then, we fit a double Schechter law to each distribution:

$$\frac{\mathrm{d}^2 N(z)}{\mathrm{d}\log_{10} M_* \,\mathrm{d}V} = S\left(M_*, \phi_1^\star, M_1^\star, \alpha_1\right) + S\left(M_*, \phi_2^\star, M_2^\star, \alpha_2\right),$$

$$S\left(M_*, \phi^\star, M^\star, \alpha\right) \equiv \log(10)\, \phi^\star \left(\frac{M_*}{M^\star}\right)^{\alpha+1} \exp\left(-\frac{M_*}{M^\star}\right). \tag{2}$$

The results are shown in Fig. 1, and the best-fit parameters are summarized in Tables A.1 and A.2. Our goal here was only to find a functional form that describes well the observed data. We thus attribute no physical origin to each component of the double Schechter law, and because the fit is prone to degeneracy, we allowed ourselves to arbitrarily fix some of the fit parameters. These are surrounded by brackets in the tables.

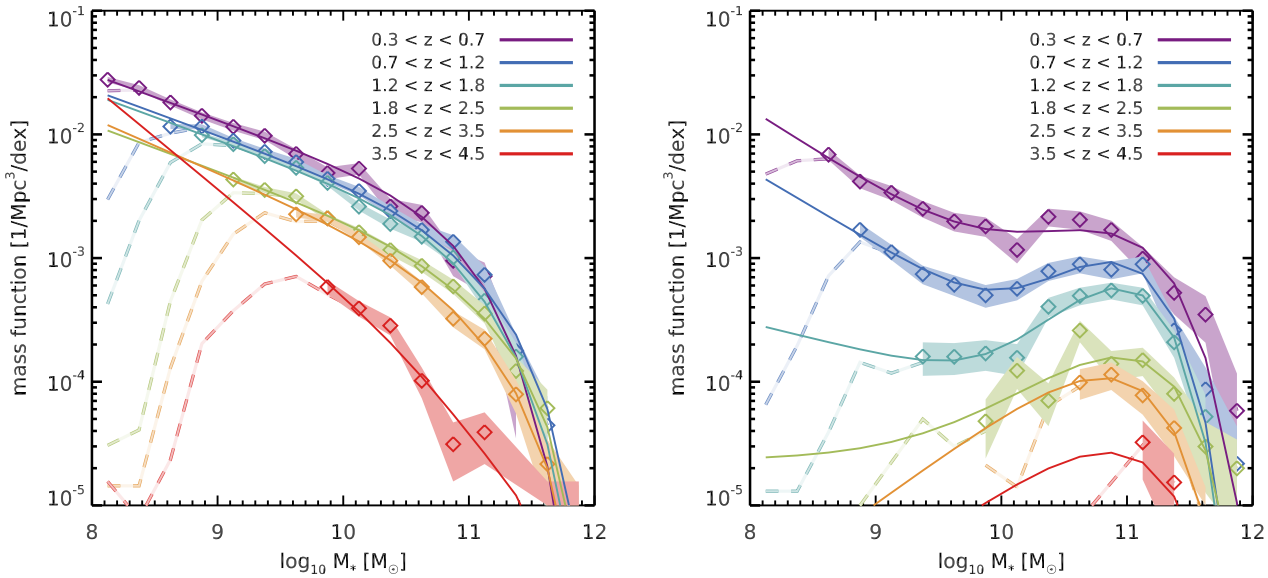

Fig. 1. Conditional stellar mass function at different redshifts for SFGs (*left*) and QGs (*right*), selected with $H < 26$. The dashed lines in the background indicate the raw mass functions, before completeness corrections are applied. The solid colored regions show the completeness-corrected estimate of the mass function, and the width of the region indicates the statistical uncertainty on the measurement (i.e., Poisson noise).

We estimated that our catalog is not complete to assess the mass function of $z = 4$ QGs, and therefore did not attempt to fit it. Instead, we used the same parameters as that obtained at lower redshifts and only adjusted $\phi^\star$ (the normalization) to have a fraction of quiescent galaxies equal to 15% (for $M_* > 4 \times 10^{10}\ M_\odot$), which is the extrapolation of the trend we observed at lower redshifts. This is consistent with what was previously reported by, e.g., Muzzin et al. (2013). However, it was recently suggested that this fraction could be substantially higher, e.g., up to 34% of quiescent galaxies at $z = 3.7$ (Straatman et al. 2014). In any case, this will not change dramatically the quality of our simulated catalogs because these objects are faint and their number density very low.

To reach higher redshifts, we used the results of Grazian et al. (2015) covering $4.5 < z < 7.5$. Their stellar mass functions do not distinguish SFGs from QGs, but since QGs are found in negligible numbers at these redshifts, we simply considered that the Grazian et al. mass functions describe the SFG population. When then assumed that the fraction of QGs remains 15% at $z > 4$ and just rescaled their $z = 4$ mass functions to match this constraint at all $z > 4$. We extrapolated these trends up to $z = 11$ to reach down to the detection limit of today's deep surveys, although we note that this extrapolation is highly uncertain and that galaxies at $z > 7$ in the simulation have essentially unconstrained properties.

As for very low redshifts, we took the $z = 0$ mass function from Baldry et al. (2012), assuming their SFG versus QG separation from a color–magnitude diagram isolates the same populations as the *UVJ* selection. This will have little consequences since we are aiming for pencil-beam surveys which contain very few local galaxies. In fact, to ensure this we imposed a minimum redshift of $z = 0.05$ to avoid having large and bright nearby galaxies in the field of view. This will limit our ability to reproduce the bright end of the counts though, and this limit can be pushed down in EGG if one wants to simulate larger fields accurately (we do so in Appendix C).

Extrapolating these combined mass functions toward the low-mass end (assuming that the low-mass slope is not varying) we can generate a population of galaxies on a wide range of stellar masses in an arbitrary volume between $z = 0$ and $z = 11$.

3.2. Stellar morphology

Following the approach of *Stuff*, we considered that galaxies are made of two components: a bulge (Sérsic index $n = 4$) and a disk (Sérsic index $n = 1$). The fraction of the stellar mass that goes into one or the other component is defined by the bulge-to-total ratio (B/T), and each of these components is described by several morphological parameters, including the projected axis ratio b/a, the half-light radius R_{50}, and the position angle θ. In the following, we present how we calibrated the distributions of these parameters.

The bulge-to-total ratio was estimated following the results of Lang et al. (2014) who conveniently measured the average B/T as a function of stellar mass for both SFGs and QGs in the CANDELS fields at different redshifts. While they found the bulge fraction to increase with stellar mass for both populations, they did not observe any significant difference with redshift between $z = 1$ and $z = 2$, so we chose to make the B/T simply depend on mass following

$$\log_{10}(B/T)_{\rm SF} = -0.7 + 0.27 \times (\log_m - 10) \text{ and}$$
$$\log_{10}(B/T)_{\rm Q} = -0.3 + 0.1 \times (\log_m - 10)\,, \quad (3)$$

with $\log_m \equiv \log_{10}(M_*/M_\odot)$, to which we added a log-normal scatter of 0.2 dex in order to reproduce the width of the distribution reported in Lang et al. (2014). This is a mass-weighted bulge-to-total ratio, therefore we can directly use it to compute the stellar mass inside the disk and the bulge. Estimating the contribution of each component to the luminosity of the galaxy is done in Sect. 3.3.

To calibrate the morphological parameters for bulges and disks, we used the morphological catalogs of van der Wel et al. (2012) who fit single Sérsic profiles of varying index n to all galaxies in the CANDELS fields using the GALFIT software (Peng et al. 2002) on the HST H-band images. In the following, we considered two sub-samples: first, galaxies with $n < 1.5$ and $M_* > 10^9\ M_\odot$, and second, galaxies with $n > 2.5$ and $M_* > 3 \times 10^{10}\ M_\odot$. The cut in stellar mass was used to select galaxies bright enough that the Sérsic fits are reliable. We used these sub-samples to calibrate the morphology of the disk

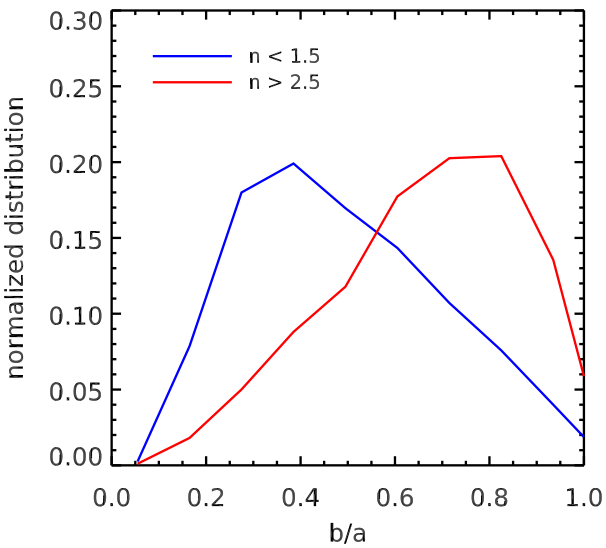

Fig. 2. Observed axis-ratio distribution of disk-dominated ($n < 1.5$, blue) and bulge-dominated ($n > 2.5$, red) samples, combining all redshifts.

and bulges, respectively. Indeed, for galaxies with $n < 1.5$ the presence of a bulge can be neglected so that the measured properties can be attributed to the disk alone (see, e.g., the appendix of Lang et al. 2014), and conversely for $n > 2.5$. This latter sample of $n > 2.5$ galaxies is probably less pure though, since high Sérsic indices can be produced either by a dominant bulge, or by a minor bulge that has a much smaller half-light radius than the disk, as shown in the appendix of Lang et al. (2014). However such cases are relatively rare, and the majority of $n > 2.5$ galaxies can indeed be considered as bulge-dominated.

For each sub-sample, we started by measuring the projected axis ratio distribution (Fig. 2). We found that disk-dominated galaxies have on average lower b/a, which is expected from to their quasi two-dimentional intrinsic geometry: their b/a distribution peaks at 0.3, compared to 0.8 for bulge-dominated galaxies. We considered that these distributions hold for all masses and all redshifts. Van der Wel et al. (2014) reported that the b/a distribution of SFGs at $z = 1.7$ shows a clear mass evolution from 10^9 to 10^{11} $M_\odot$: while their low-mass distribution is very similar to our disk-dominated distribution, they found their high-mass distribution to be bimodal. Without attempting to demonstrate it, we argue here that this trend is probably the result of the increase of the B/T with stellar mass among SFGs (Lang et al. 2014). On the one hand, low-mass galaxies are preferentially bulgeless, and should therefore follow the trend of pure-disks of Fig. 2. On the other hand, high-mass galaxies are more complex systems with a varying mixture of bulges and disks; among these, we expect to find both bulge- and disk-dominated systems, and this would explain the bimodal distribution observed by van der Wel et al. (2014). Interestingly, by comparing their $z = 1.7$ result to a similar analysis in the Sloan Digital Sky Survey (SDSS, $z = 0$), van der Wel et al. (2014) showed that these distributions are also redshift-dependent, so that the bimodality extends down to lower stellar masses at lower redshifts. One possible explanation for this would be that the redshift invariance of the B/T–M_* relation found by Lang et al. (2014) may not hold at $z < 1$, indicating that SFGs in the local Universe have more prominent bulges at fixed stellar mass.

The next step is the calibration of the half-light radius. The size of a galaxy correlates with its stellar mass (i.e., the mass–size relation, see, e.g., Shen et al. 2003), and sizes at fixed mass were also globally smaller in the past (e.g., Ferguson et al. 2004; Daddi et al. 2005). For this reason, we binned our two sub-samples in stellar mass and observed the evolution of the median half-light radius with redshift. The observed trends are reported in Fig. 3. Defining $\log_z \equiv \log_{10}(1+z)$, and we parametrized these relations with the following equations, for disks:

$$\log_{10}\left(R_{50,\mathrm{disk}}[\mathrm{kpc}]\right) = 0.2 \times (\log_m - 9.35) + F_{z,\mathrm{disk}}, \quad (4)$$

$$\text{with } F_{z,\mathrm{disk}} = \begin{cases} 0.41 - 0.22 \times \log_z & \text{for} \quad z \leq 1.7 \\ 0.62 - 0.70 \times \log_z & \text{for} \quad z > 1.7, \end{cases}$$

and for bulges:

$$\log_{10}\left(R_{50,\mathrm{bulge}}[\mathrm{kpc}]\right) = 0.2 \times (\log_m - 11.25) + F_{z,\mathrm{bulge}}, \quad (5)$$

$$\text{with } F_{z,\mathrm{bulge}} = \begin{cases} 0.78 - 0.6 \times \log_z & \text{for} \quad z \leq 0.5 \\ 0.90 - 1.3 \times \log_z & \text{for} \quad z > 0.5, \end{cases}$$

to which we added a log-normal scatter of 0.17 and 0.2 dex respectively. Although these latter values are smaller than the scatter reported, for example, by Shen et al. (2003) or Dutton et al. (2011), we stress that they apply to the disk and bulge components separately. When considering the size of the galaxy as a whole (i.e., the sum of the disk and the bulge, see Appendix B), we found that the additional scatter in the bulge-to-disk ratio is sufficient to reproduce the observed width of the mass–size relation. However, to preserve the normalization of the mass–size relation in composite systems, we used the total mass M_* to derive each component's respective size.

Lastly, we attributed a position angle to each galaxy by randomly drawing from a uniform distribution, and assigned the same angle to both the bulge and disk components.

3.3. Stellar spectral energy distribution

Once the main physical properties were generated, we associated a stellar SED to both the disk and bulge component of each galaxy. Instead of basing our approach solely on physical arguments, for example, stellar age and dust content, we chose a simpler effective prescription where the SED is determined by the position of the galaxy on the *UVJ* diagram. This color–color diagram (already introduced in Sect. 2.2) provides a simple way to describe a wide range of spectral types, in particular "blue and star-forming", "red and dead" and "red and dust-obscured", from low to high redshifts (e.g., Williams et al. 2009; Patel et al. 2012; Straatman et al. 2014). Our first goal was therefore to define a recipe to generate realistic colors distributions in the *UVJ* plane. Since SFGs and QGs are well segregated into two groups (or "clouds"), we treated each population separately, starting with the SFGs.

To this end, we considered all the *UVJ* star-forming galaxies in the CANDELS catalogs and divided this sample in multiple stellar mass bins. For each of these bins, we computed the median $U - V$ and $V - J$ colors at varying redshifts, and display the resulting tracks on the *UVJ* diagram in Fig. 4 (left). At fixed stellar mass, both colors go from blue to red as redshift decreases. A similar trend could already be identified in S15 (Fig. 1), and is expected given the decrease of specific SFR (Noeske et al. 2007; Arnouts et al. 2013) and the increase in metallicity (Mannucci et al. 2010). Interestingly, we found here that all the tracks seem to follow a straight line that we call the "*UVJ* sequence":

$$(U - V)_{\mathrm{SF}} = 0.65 \times (V - J)_{\mathrm{SF}} + 0.45. \quad (6)$$

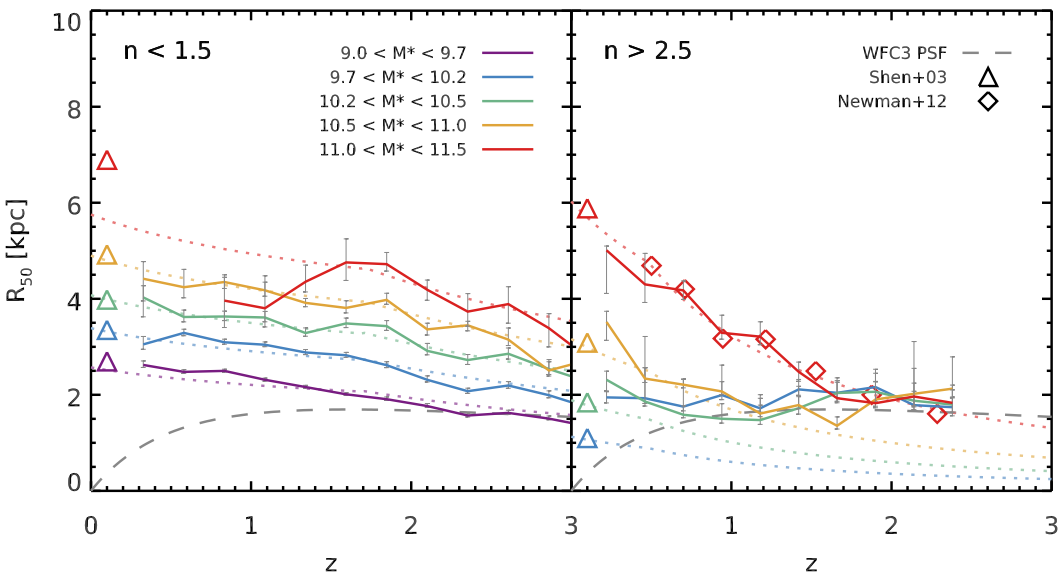

Fig. 3. Observed relation between the half-light radius R_{50} (along the major axis) and redshift of disk-dominated galaxies (*left*) and bulge-dominated galaxies (*right*). Different stellar mass bins are shown with different colors as indicated in the legend. The median values over all CANDELS fields are shown with solid colored lines, error bars indicating the uncertainty on the median from bootstrapping, and the prescription adopted in this work is displayed with a dotted line in the background. We also show how the size of the HST H-band PSF (0.2″) translates into proper distance with a dashed line. Empty triangles at $z = 0.1$ are the values obtained by Shen et al. (2003) in the SDSS, converted from a Kroupa to a Salpeter IMF. Based on the median axis ratios we measure (Fig. 2) and following Dutton et al. (2011), we multiply the Shen et al. values by a factor 1.4 and 1.1 for disk- and bulge-dominated galaxies, respectively, to correct for the fact that their radii were measured in circularized apertures. Finally, for bulge-dominated galaxies, we also display the size measurements of Newman et al. (2012), which were obtained by selecting quiescent galaxies based on their sSFR. Their values are reported as $R_{50}/M_{11}^{0.57}$, which we renormalize to the stellar mass of our highest mass bin.

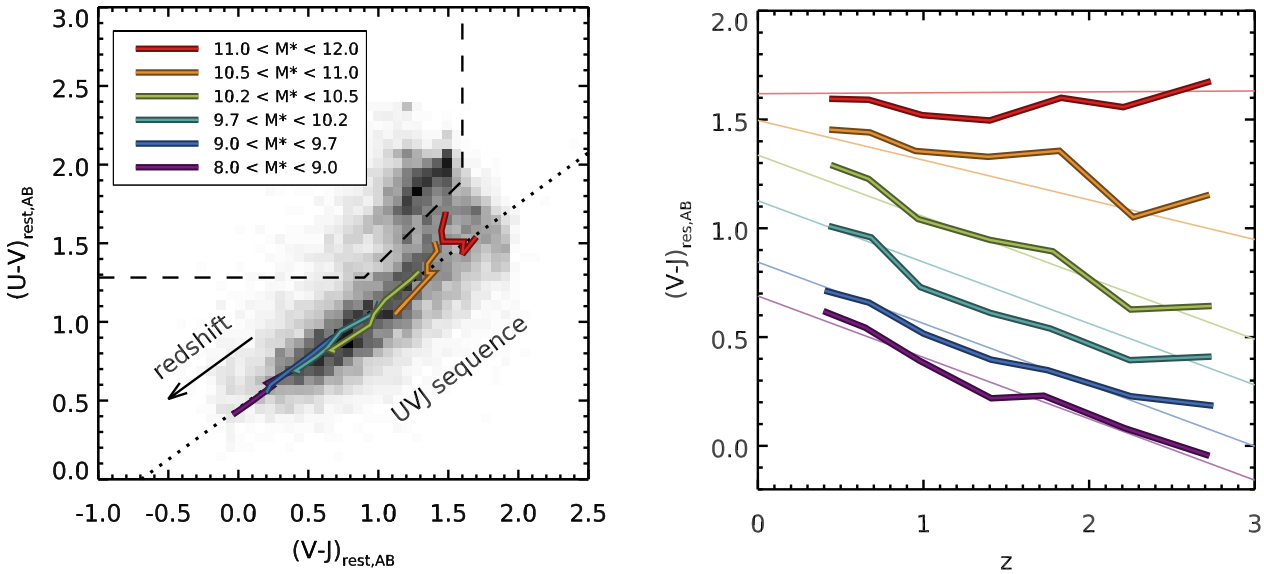

Fig. 4. *Left*: *UVJ* diagram of all galaxies in the CANDELS fields more massive than 10^{10} $M_\odot$ (background gray scale). The redshift evolution of the median $U - V$ and $V - J$ colors of SFGs in different stellar mass bins is overlaid with colored lines. They all fall along a single line we dub the "*UVJ* sequence", which is illustrated by a dotted line. Finally, the adopted dividing line between SFGs and QGs is shown with a long dashed line. *Right*: evolution of the $V - J$ color with redshift for each bin of mass. We show in the background the prescription adopted in this work.

See also Labbé et al. (2007) where such a sequence is found among blue galaxies in a color–magnitude diagram. This sequence also happens to run parallel to the dust attenuation vector (Williams et al. 2009), meaning that the position of a galaxy on this sequence could also be interpreted as a sign of varying dust content.

To better illustrate where galaxies of different mass are located on this sequence and how they evolve with time, we show in Fig. 4 (right) the evolution of the average $V - J$ color with redshift for each mass bin. Massive galaxies appear to always have the same very red colors (consistent with the fact that these are typically the most dusty, e.g., Pannella et al. 2009), while less massive galaxies were substantially bluer in the past. We parametrized the resulting color tracks using the following equations:

$$(V - J)_{\rm SF} = a_0 + a_1 \times \min(z, 3.3), \quad (7)$$

$$\text{with } a_0 = 0.48 \times \mathrm{erf}(\log_m - 10) + 1.15,$$

$$\text{and } a_1 = -0.28 + 0.25 \times \max(0, \log_m - 10.35).$$

We added a Gaussian scatter of 0.1 mag to this color, and limited its value to be at most 1.7 to prevent extreme red colors. We then used Eq. (6) to obtain the $U-V$ color, and added an extra layer of Gaussian scatter of 0.12 mag to both colors to reproduce a wider range than what is allowed by this simple prescription. This scatter is most likely caused by variations of star formation histories, specific star-formation rates (e.g., Arnouts et al. 2013) and inclination (e.g., Patel et al. 2012), but we do not attempt to study its origin here. We emphasize that the redshift evolution of SFGs on the *UVJ* sequence (first line in the above equation) is stopped at $z > 3.3$: not only does this redshift domain go outside of the range in which the recipe was calibrated, but we also found that this step is necessary to reproduce the UV luminosity functions at $4 < z < 7$.

The prescription for QGs is relatively simpler, since these are located within a smaller region of the *UVJ* diagram, the so-called red cloud. We used the same approach as for SFGs, this time selecting the *UVJ* quiescent galaxies in the CANDELS catalog, and computed the median $U-V$ and $V-J$ colors in bins of redshift and mass. Here as well galaxies tend to reside on a sequence:

$$(U-V)_{\rm Q} = 0.88 \times (V-J)_{\rm Q} + 0.75, \tag{8}$$

although the dynamic range is much smaller than for SFGs. In particular we found no significant redshift trend, and a moderate trend with stellar mass such that more massive galaxies are redder, probably because they are older. We parametrized this trend as

$$(V-J)_{\rm Q} = 0.1 \times (\log_m - 11) + 1.25, \tag{9}$$

to which we added 0.1 mag of Gaussian scatter. The resulting color is clamped within 1.15 and 1.45 to prevent galaxies from exiting the red cloud. As for SFGs, the $U-V$ color is obtained by applying Eq. (8), and both colors are perturbed independently with a Gaussian scatter of 0.1 mag.

We used the above relations to derive the colors of the disk and bulge components of each galaxy. To do so, and following the observations of Schreiber et al. (2016) at $z = 1$ (see in particular Fig. 5 from that paper), we considered that all disks are "star-forming" and obtain their colors from Eqs. (6) and 7. Similarly, we considered that all bulges of bulge-dominated galaxies ($B/T > 0.6$) are "quiescent" and described by Eqs. (8) and 9. Determining the colors of disks and bulges in composite systems is challenging, and can only be attempted for the brightest galaxies (see, e.g., Bruce et al. 2014). Here, in the absence of observational constraints, bulges in composite galaxies were randomly chosen to be "star-forming" or "quiescent" with uniform probability to simulate both bulges and pseudo-bulges. This prescription allows us to reproduce the observed total color distributions of galaxies as a function of B/T at $z = 1$ from Schreiber et al. (2016).

We stress here that, with the above prescriptions, "star-forming" colors are not necessarily "blue". As can be seen in Fig. 4, massive SFGs do have red colors up to $(U-V) = 1.5$ owing to dust attenuation. In addition, the random scatter we introduce in the various steps will allow a few SFGs (~1%) to have colors that fall inside the quiescent region (dashed line). As a consequence, 15% of our disk-dominated galaxies ($B/T < 0.3$) at $z < 0.5$ and $M_* > 3 \times 10^{10}\ M_\odot$ have red $g-r$ colors, which is comparable to the value observed by Masters et al. (2010) at $z \sim 0$ (see their Fig. 2). Therefore, these prescriptions allow us to produce many different combinations of colors and morphologies which are representative of the real galaxy population.

The last step is to associate a complete stellar SED to each pair of colors we just generated. From the observed catalogs, we binned uniformly the *UVJ* plane into small buckets of 0.1 mag and gathered all the galaxies that fall inside each bucket, regardless of their redshift and stellar mass. We then computed the average of their rest-frame SED per unit mass, which was obtained from the best-fit template produced by FAST and the Bruzual & Charlot (2003) stellar library when fitting for the stellar mass (Sect. 2.2). We discarded the buckets containing less than ten galaxies, and ended up with an empirical library of 345 SEDs, each corresponding to a given position in the *UVJ* diagram. This library does not cover the whole *UVJ* plane though, and therefore if a simulated galaxy had colors that fall outside of the covered region (which is rare by construction), it was attributed the SED of the closest non-empty bucket. Finally, the stellar SED of both disks and bulges was obtained by rescaling the chosen SED by their respective stellar mass.

The way we built our stellar SED library implicitly assumes that the mass-to-light ratio (M/L) is uniquely determined by the $U-V$ and $V-J$ colors. In practice though, we found that this is not a valid assumption for the lowest redshift galaxies ($z < 1$), for which the procedure produces too low M/L by about 40% on average, and conversely for high redshift galaxies ($z > 6$) where the M/L ratio is overestimated by up to a factor 4. To compensate for this effect, we added a correction term to the stellar mass before using it to rescale the SED (−0.15 dex for $z < 0.45$ and linearly coming back to zero at $z = 1.3$, then +0.6 dex for $z = 8$ and linearly coming back to zero at $z = 6$).

4. Star formation and dust properties

4.1. Star-formation rate

Given the redshift and the stellar mass, we attributed a star-formation rate (SFR) to each galaxy by following the Two Star Formation Mode model (2SFM, Sargent et al. 2012). This model relies on the existence of the SFR–M_* main sequence and has been shown to successfully reproduce the observed flux and redshift distributions from the MIR-to-the submm and even the radio (Béthermin et al. 2012, 2015). Taking advantage of these results, we used here a similar prescription where the model parameters were updated with our latest *Herschel* measurements.

Using the SFR(z, M_*) equation derived from *Herschel* stacking in S15, we associated a "main sequence" (MS) star-formation rate to each SFG with

$$\log_{10}\left({\rm SFR_{MS}}[M_\odot/{\rm yr}]\right) = \log_m - 9.5 + 1.5\ \log_z - 0.3\left[\max\left(0, \log - 9.36 - 2.5\ \log_z\right)\right]^2, \tag{10}$$

where $\log_m \equiv \log_{10}(M_*[M_\odot])$, and $\log_z \equiv \log_{10}(1+z)$.

We then obtained the SFR of each galaxy by applying a log-normal scatter of 0.3 dex to reproduce the observed width of the main sequence, which was found in S15 to be constant both as a function of stellar mass and redshift. In addition, 3% of the galaxies were randomly chosen and placed in a "starburst" mode, where their SFR is enhanced by a factor of 5.24, following S15. Sargent et al. (2012) showed that this last step is necessary to correctly capture the bright-end of the IR luminosity functions. As described in the next section, we used the offset of each galaxy with respect to the main sequence to fine-tune their dust spectrum. Following Elbaz et al. (2011) we quantified this offset with the "starburstiness"

$$R_{\rm SB} \equiv \frac{\rm SFR}{\rm SFR_{MS}}, \tag{11}$$

which is equal to one for a purely main sequence galaxy.

For QGs, we used the IR stacks presented in the appendix of S15, where it was reported that QGs do show some IR emission, typically a factor of ten fainter than SFGs of the same mass. This light may be caused either by residual star formation, AGN torus emission, dust heated by old stars, or by incorrect classification of some SFGs. Although this is an interesting question, its answer is irrelevant for our purposes, and we chose to model this faint emission by interpreting it as residual star-formation. Therefore, QGs were attributed an SFR following

$$\log_{10}(\mathrm{SFR_{QS}}[M_\odot/\mathrm{yr}]) = 0.5\ \log_m + \log_z - 6.1, \tag{12}$$

to which we added a log-normal scatter of 0.45 dex. This latter value was chosen to roughly reproduce the number of 24 μm detected QGs at $z = 1$.

4.2. Obscuration

To estimate the dust luminosity, we decomposed the SFR into a dust-obscured component, which re-emerges in the FIR, and dust-free component, which emerges in the UV. To do so, we used the observed relation between stellar mass and dust obscuration (e.g., Pannella et al. 2009; Buat et al. 2012; Heinis et al. 2014), which we calibrated here in terms of the infrared excess $\mathrm{IRX} \equiv \log_{10}(L_{\mathrm{IR}}/L_{\mathrm{UV}})$ (Meurer et al. 1999). Using the stacked L_{IR} from S15, we found that the relation between IRX and M_* can be described by

$$\mathrm{IRX} = (0.45\min(z, 3.0) + 0.35) \times (\log_m - 10.5) + 1.2. \tag{13}$$

This formula is very similar to that reported by Heinis et al. (2014), except that our relation was found to be redshift dependent: at a fixed mass above $M_* \sim 3 \times 10^{10}\ M_\odot$, attenuation becomes more important at higher redshifts, while it becomes less pronounced for lower mass galaxies. This behavior, at least at the high mass end, is consistent with the results of Pannella et al. (2015) who report that the typical MS galaxy at $z = 2$ is sensibly different from its analog at $z \leq 1$, which they argue is because of modifications in the geometry of the star-forming regions. We also added a scatter of 0.4 dex to this relation: although it has a negligible impact on the generated IR luminosities, Bernhard et al. (2014) showed that this is a necessary ingredient to properly reproduce the bright-end of the UV luminosity function.

4.3. Consistency between the stellar and dust emission

One of the main physical inconsistencies of our model is found here: the stellar UV emission and reddening were already determined indirectly in Sect. 3.3, based on the position of the galaxies in the *UVJ* diagram, and we did not use these values to predict the SFR or the L_{IR}. We made this choice to avoid forcing galaxies to match the "energy balance" between the UV and IR (e.g., da Cunha et al. 2008; Noll et al. 2009); we used independent empirical prescriptions to generate the UV and IR emission separately, so that the actual L_{UV} of our simulated galaxies can differ from the value one would otherwise obtain from Eqs. (10) and (13). In practice the two are still well correlated with no systematic bias over six orders of magnitude, albeit with a substantial scatter of 0.42 dex. Similarly, the SFR one would derive by fitting the stellar SED from Sect. 3.3 is not strictly identical to the SFR of Eq. (10); they correlate with a scatter of 0.36 dex.

We argue that, in spite of being physically inconsistent, this scatter is a feature that allows us to more accurately match the observations. Indeed, different SFR indicators tend to agree in the average sense and for large populations, but often fail to provide consistent results when applied to one single galaxy (e.g., Goldader et al. 2002; Buat et al. 2005; Elbaz et al. 2007; Penner et al. 2012; Oteo et al. 2013). To illustrate this, we selected from the observed CANDELS catalogs all the galaxies at $0.7 < z < 0.8$ with a detection in the *Spitzer* MIPS or *Herschel* bands, and used their L_{IR} and L_{UV} as computed in S15 to derive a "UV+FIR" SFR. We then compared this value against the SFR obtained by FAST when fitting the UV-to-NIR photometry, and found a scatter of 0.42 dex (measurement uncertainties were not subtracted from this value) which is comparable to the 0.36 dex of the simulated catalog.

4.4. Dust temperature and chemical composition

In the previous sections we have attributed an SFR to each galaxy, and estimated what fraction of the associated light comes out in the FIR. From there, the last missing ingredient to predict dust fluxes is a suitable dust SED library, with enough adjustable parameters to reproduce accurately the observed counts. A number of such SED libraries have already been published, calibrated either in 8-to-1000 μm luminosity (L_{IR}) from local galaxies (Chary & Elbaz 2001, hereafer CE01), far-infrared (FIR) colors (Dale & Helou 2002), or intensity of the interstellar radiation field ($\langle U \rangle$) from distant galaxies (Magdis et al. 2012; Béthermin et al. 2015).

We used here a new SED library (introduced in Schreiber et al. 2016) in which both the dust temperature (T_{dust}) and the $\mathrm{IR8} = L_{\mathrm{IR}}/L_8$ (where L_8 is the k-corrected luminosity in the IRAC 8 μm filter) are free parameters. There are two reasons for this choice. First, these are the two parameters that are the easiest to measure without FIR spectroscopy (which is only available for a few bright objects), and are the ones affecting most the shape of the SED. Second, 8 μm is the domain where polycyclic aromatic hydrocarbon molecules (PAHs) emit the bulk of their light through strong emission lines. This is a domain that will be routinely accessed by the *James Webb* Space Telescope in the near future, and there will be a need for a properly calibrated library to exploit these data together with ancillary *Herschel* and *Spitzer* observations.

We calibrated the redshift and mass evolution of both T_{dust} and IR8 using the MIR-to-FIR stacks of S15, to which we added stacks of the *Spitzer* IRS 16 μm imaging (Teplitz et al. 2011) to better constrain the rest-frame 8 μm and the PAH features (available in GOODS-North and South only). We then further refined this calibration using individual *Herschel* detections to constrain the scatter on these parameters, and also to calibrate how they are modified for those galaxies that are offset from the main sequence. These results will be described in more detail in another work (Schreiber et al., in prep.).

The parametrizations for T_{dust} and IR8 for a galaxy lying exactly on the main sequence are:

$$T_{\mathrm{dust}}^{\mathrm{MS}}[\mathrm{K}] = 4.65 \times (z-2) + 31 + 1.5 \times \min(z-2, 0) \times \mathrm{clamp}(\log_m - 10.7, 0, 1), \tag{14}$$

$$\log_{10}(\mathrm{IR8_{MS}}) = \log_{10}(7.73 + 1.95 \times \min(z-2, 0)) - 1.8 \times \mathrm{clamp}(\log_m - 10, -1, 0). \tag{15}$$

The second line in both equations is a second order dependence on the stellar mass motivated by the stacked fluxes. We made massive galaxies at $z < 2$ slightly colder (owing to their reduced

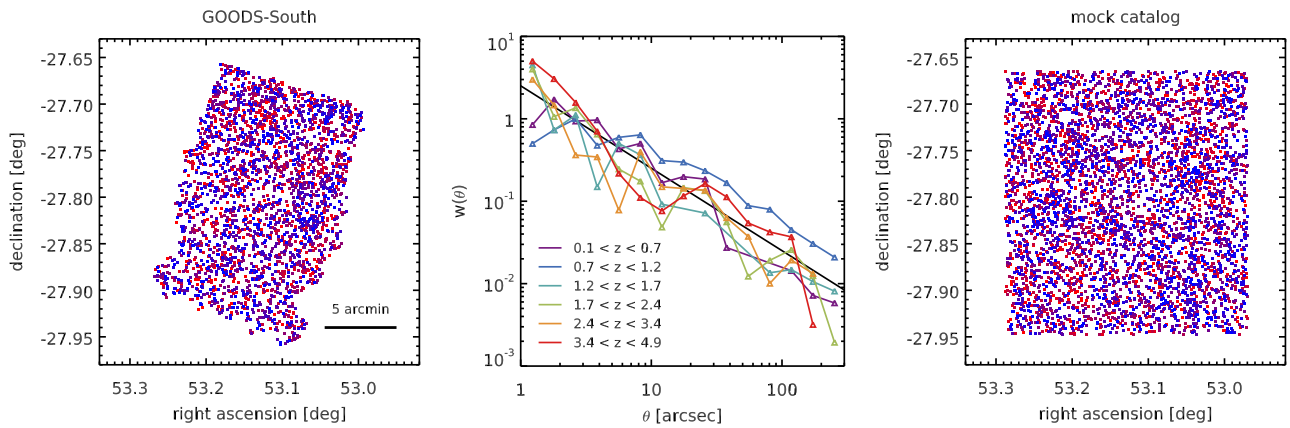

Fig. 5. *Left*: sky positions of the galaxies in the GOODS-South field with $H < 23$. The colors indicates the redshift: blue points are $z < 0.3$, red points are $z > 1.5$, and galaxies in between are shown in shades of purple. *Middle*: angular two-point correlation function for GOODS-South galaxies with $9 < \log_{10}(M_*/M_\odot) < 10.3$, in various redshift slices as indicated in the legend. The straight black line is a power law of index -1. Error bars are not shown for clarity. *Right*: same as left, but for the mock catalog produced by EGG.

star formation efficiency, see Schreiber et al. 2016), while low-mass galaxies were given a lower IR8 essentially consistent with no PAH emission. The latter is known to happen for sub-solar metallicity objects in the local Universe (e.g., Madden et al. 2006; Wu et al. 2006; O'Halloran et al. 2006; Galliano et al. 2008; Ciesla et al. 2014).

As shown in Elbaz et al. (2011), starbursting galaxies have a depleted 8 μm luminosity compared to the total $L_{\rm IR}$, and an increased dust temperature (see also Nordon et al. 2012; Magdis et al. 2012; Magnelli et al. 2014; Béthermin et al. 2015). This is a sign that these galaxies are experiencing an episode of star formation in a more compact interstellar medium. To take this effect into account, we included a dependence of both $T_{\rm dust}$ and IR8 on the starburtiness (see previous section):

$$T_{\rm dust}[{\rm K}] = T_{\rm dust}^{\rm MS} + 6.6 \times \log_{10}(R_{\rm SB}), \quad (16)$$

$$\log_{10}({\rm IR8}) = \log_{10}({\rm IR8_{MS}}) + 0.43 \times \log_{10}(R_{\rm SB}), \quad (17)$$

to which we added a Gaussian scatter of 4.1 K for $T_{\rm dust}$ and a log-normal scatter of 0.1 dex for IR8.

Using these two quantities, we picked a FIR SED from our library and rescaled it with the $L_{\rm IR}$ computed in the previous section to build the dust SED of the galaxy.

5. Sky position

The simplest approach to generate the position on the sky of each galaxy would be to draw these positions uniformly on the sphere, within the region of the sky that is covered by the simulated survey. The stellar mass functions we used in Sect. 3.1 ensure that we get a correct sky density of object over the whole simulated area. However, within the ΛCDM cosmology, we expect galaxies to form large-scale structures by following the merging history of their dark matter halos (Peebles 1982). In other words, galaxies tend to cluster on the sky, and we need to simulate this effect to generate realistic sky positions. In S15 (see also Béthermin et al. 2010), we showed that clustering can have a significant impact on the statistical properties of confused, long-wavelength images from *Spitzer* and *Herschel*: it will tend to increase the contrast compared to a uniform position distribution, that is, creating overdense and underdense regions within the survey area. On the other hand, we expect clustering to be no more than a cosmetic change for the high-resolution HST images, which do not suffer from confusion.

The procedure we used here is to aim at reproducing the observed angular two-point correlation function $w(\theta)$, that is, the excess probability of finding a galaxy at a given angular distance θ from another, as compared to a uniform position distribution. The first step was therefore to measure this two-point correlation function in the real GOODS-South field. To do so, we binned the whole catalog in redshift slices of width $\Delta = 0.25 \times (1 + z)$, and only two mass bins because the statistics is limited ($M_* = 10^9$ to 3×10^{10} $M_\odot$, and $M_* = 3 \times 10^{10}$ to 10^{12} $M_\odot$), and we did not attempt to further refine the sample by separating different galaxy types. We then used the Landy & Szalay (1993) estimator to compute the two-point correlation function of each sample, and observed a significant clustering signal between 1″ and 5′ at all redshifts (see Fig. 5, middle panel). This signal is well described by a single power law

$$w(\theta) \sim \theta^{-\gamma}, \quad (18)$$

with $\gamma = 1$, and where θ is the angular distance between two sources. As in S15, we found no significant change in angular clustering amplitude with redshift between $z = 0.3$ and $z = 4$ (which is also consistent with the results of Béthermin et al. 2015), but we did find that the amplitude of the two-point correlation function for massive galaxies is on average about three times larger. The fact that massive galaxies are more clustered should not come as a surprise knowing that we selected all massive galaxies, including red quiescent galaxies which are known to be the best tracers of the large scale structures, both locally and in the distant Universe (e.g., Cooper et al. 2006; Cucciati et al. 2006; Elbaz et al. 2007).

However, it is important to note that these measured two-point correlation functions are affected by the uncertainties on the photometric redshifts (photo-zs). Indeed, within each adopted redshift bin, there is a chance that we miss some galaxies that scattered out of the bin, and another chance that we are contaminated by some galaxies scattering into the bin. The net result is that we observe a clustering amplitude that is lower than the intrinsic one. This effect can be simulated (and we do so in the following) once the uncertainty on the observed redshifts is known. Redshift uncertainties also contribute to some extent to the mass-dependent clustering that we observed, since massive bright galaxies have more robust photo-zs and are therefore expected to show a cleaner clustering signal. To measure

this uncertainty, we cross-matched our GOODS-South catalog against the 3DHST catalog (DR1, Skelton et al. 2014). While the two catalogs are based on the same raw observations, the data reduction and photometry were performed independently with different tools. On the other hand, the photo-zs were estimated with the same code, so we likely underestimated the real redshift uncertainty. We measured the distribution of redshift differences between the two catalogs, and took into account that what we observed was the combination of uncertainties coming from both catalogs (i.e., assuming they are independent, $\sqrt{2}$ higher than that of a single catalog). We found that the redshift uncertainty in $\Delta z/(1+z)$ is well described by the combination of two zero-mean Gaussians at high (low) masses: a first distribution of width 2.2% (2.5%) that describes 84% (76%) of galaxies, and a second distribution of width 6.6% (9.3%) that describes the remaining 16% (24%).

To produce sky positions that resemble these observations, we used the Soneira & Peebles (1978) algorithm, which is a simple method to produce a two-point correlation function with an adjustable power-law slope. Briefly, the algorithm starts by drawing η random positions within a circle of radius R on the sky. This is the first level. At each of these positions, a new set of η positions is randomly generated within a smaller circle of radius R/λ, and the ensemble of all these new positions make up the second level. This procedure is repeated up to a given level L, reducing the size the circles by a factor λ for each step. The final level contains a set of η^L positions on which we randomly place our simulated galaxies drawn from narrow redshift bins ($\Delta z/(1+z) \sim 0.1$) to roughly mimic the redshift-space clustering.

We tweaked the parameters of the Soneira & Pebbles algorithm to recover the right power law slope $\gamma = 1$ ($\eta = 5$ and $\lambda = 6$) and to generate a large enough number of positions (i.e., by varying L, which does not affect the shape of the power law). The last parameter, R, was arbitrarily fixed[2] to $R = 3'$, which truncates large-scale clustering beyond this angular scale (i.e., beyond ~1 Mpc at all $z > 0.5$). To fill the whole mock survey area, we randomly placed several $3'$ circles within the field with a uniform probability distribution and used the Soneira & Peebles algorithm only within each of these circles.

Using this method, we produced a catalog of clustered positions with the right power-law slope, which we checked by measuring the clustering of the generated positions with the Landy & Szalay estimator. However, we still had to tune the amplitude of this clustering. We chose here a simple approach where we used the Soneira & Peebles algorithm only for a given fraction f of the simulated galaxies, and used uniformly distributed positions for the remaining fraction. We chose this fraction by first generating a set of positions with $f = 100\%$ (i.e., maximum clustering), apply the above procedure to measure the correlation function, and compared it to the observed one. The difference of amplitude then told us by how much we needed to reduce the simulated clustering. We stress that it is important here to take into account the redshift uncertainties that affect the observed relation. To do so, we measured the two-point correlation function in the simulation using "wrong" redshifts, which were obtained from the "true" redshifts of the simulation and then perturbed within the uncertainty described above. After taking this into account, we found that $f = 25\%$ for $M_* < 3 \times 10^{10}\ M_\odot$, and $f = 60\%$ for more massive galaxies[3]. An example of the resulting sky distribution for massive galaxies is shown in Fig. 5 (right).

To double check, we also computed the angular correlation function of the whole catalog above $M_* > 10^{10}\ M_\odot$, mixing all redshifts. Doing so, we got rid of the issue of the redshift uncertainty, and found also a good agreement with the observations.

6. Generating a light cone

6.1. Standard procedure

With all these recipes, we generated a complete catalog of galaxies, each with its own UV-to-submm SED. In this section, we summarize the procedure that is implemented in EGG to produce a final flux catalog. For a quick overview, one can refer to the flow chart presented in Fig. 6.

Given the area of the mock survey, the first step is to choose the number of galaxies that will be generated. Since EGG uses the stellar mass function as a starting point, this amounts to choosing the lowest stellar mass that will be generated. This threshold can be chosen to be constant, for example down to $M_* = 10^8\ M_\odot$, but this is in fact quite inefficient. Observations in the real GOODS-South field are flux-limited; it is possible to detect galaxies that are less massive than $10^8\ M_\odot$ at low redshifts, but the smallest measured stellar mass at $z > 2$ is closer to $10^9\ M_\odot$. Therefore, this approach can result either in a catalog that is incomplete (if the mass threshold is too high and one misses detectable galaxies at low redshift), or bloated (if the threshold is too low and one generates galaxies that will never be observed).

A more efficient approach is to use a redshift dependent threshold, so that galaxies are generated down to low stellar mass at low redshifts, and then increase this threshold to generate only the most massive galaxies at higher redshifts. To do so, the user first chooses a "selection band", for example, the HST *F160W* or the VISTA K_s band, and a magnitude limit, for example $H < 29$, above which the catalog will be at least 90% complete. EGG then builds a redshift grid, and for each redshift in that grid it computes the distribution of mass-to-light ratios in the selection band for all the optical SEDs in the library. It picks the tenth percentile of this distribution, and uses it to compute the minimum stellar mass at this redshift given the magnitude limit.

Once the stellar mass and the redshift are generated from the mass functions, the program uses the method described in Sect. 5 to place these galaxies on the sky, and applies all the above recipes to generate the SFR, the $L_{\rm IR}$ and other dust related parameters ($T_{\rm dust}$ and IR8), the *UVJ* colors, and the morphological parameters (B/T, R_{50}, b/a). Then, the optical SED is chosen based on the generated *UVJ* colors (Sect. 3.3) and scaled by the stellar mass, while the FIR SED is chosen from the $T_{\rm dust}$ and IR8 (Sect. 4.4) and scaled by the $L_{\rm IR}$. The two SEDs are redshifted and co-added to form a single, panchromatic SED that ranges from the FUV up to the submm, as shown in Fig. 7 (the radio and X-ray domains are not yet implemented). Lastly, the SED is multiplied by the response curve of each broadband filter and integrated to generate a flux in each band. This is all done with a single call to the `egg-gencat` program, and takes 20 s on a single-core of a regular desktop computer to generate a CANDELS-like field down to $H = 28$.

[2] Choosing larger values for R would lead to situations where most galaxies in a given redshift bin could fall out of the field of view, generating additional cosmic variance. We did not implement this in the code.

[3] These values depend largely on the details of the clustering modeling (i.e., the choice of R, η and λ) as well as the width of the redshift windows within which the angular correlation function is measured, and have therefore little meaning in the absolute sense. However the trend for massive galaxies to be relatively more clustered should reflect a real phenomenon.

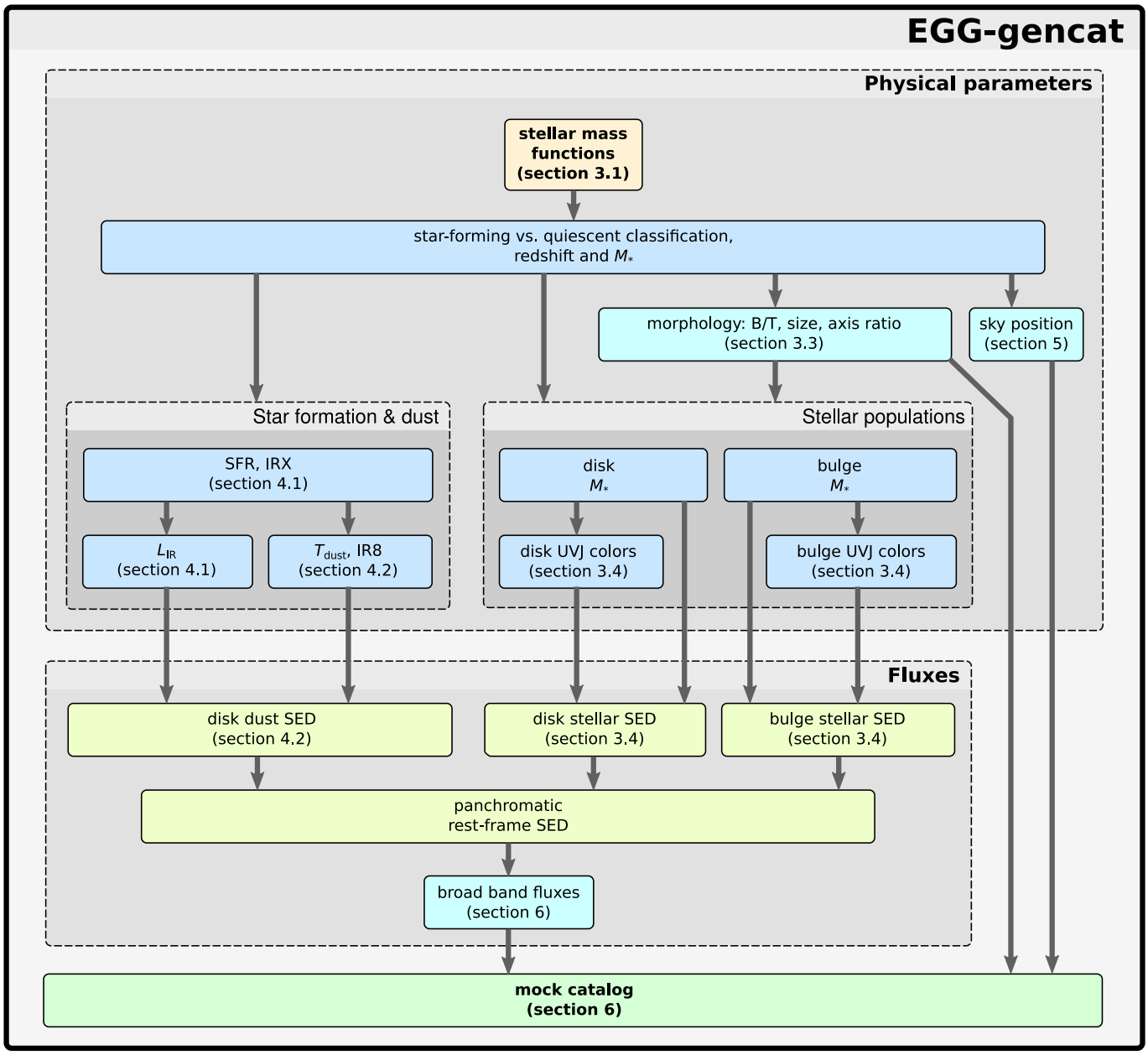

Fig. 6. Flow chart of the catalog simulation procedure, as implemented in `egg-gencat`.

6.2. Customization

Most parameters of the simulation can be customized easily by using command line arguments, including in particular the dimensions of the mock field, the random seed, the depth, the redshift range, or the set of bands in which fluxes should be produced. New filters can be added to the default list with minimal effort. The program can also generate rest-frame magnitudes in any band if requested, or even output the full medium-resolution spectrum of each galaxy.

It is also possible to feed EGG with a pre-existing catalog of redshifts, stellar masses, star-forming versus quiescent flags, and positions, e.g., coming from a real catalog: the program will then apply the same recipes and predict fluxes in any band for each input galaxy. This can be useful for proposal writing, or to test the systematics introduced when stacking FIR images for specific populations (e.g., flux boosting from clustering) as was done in S15.

The program and its various options are described in full detail in the documentation, which is provided with the source code or can be browsed on line[4].

7. Quality of the mock catalogs

7.1. Number counts

In this section we quantify the accuracy of the mock catalogs produced by EGG by comparing the generated number counts in various bands from the optical to the sub-mm against the observed counts in the GOODS-South or CANDELS fields and other literature data.

All the number counts predicted by the simulation over a large dynamic range of fluxes are provided in Appendix C. Here in Fig. 8, we focus on the HST *F435W* and *F160W* bands.

[4] http://cschreib.github.io/egg/files/EGG.pdf

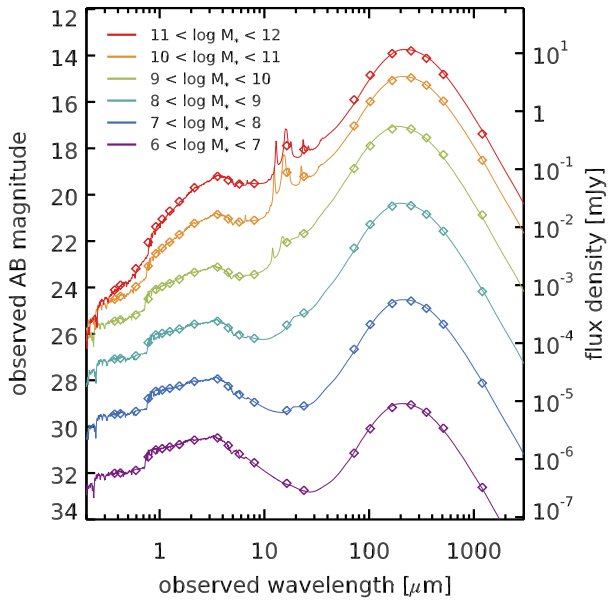

Fig. 7. Average SEDs of $z = 1$ SFGs in the EGG simulation. Each color corresponds to a different bin of stellar mass, as shown in the legend. The open diamonds are the corresponding broad band fluxes integrated in a set of common bands from *Hubble*, *Spitzer*, *Herschel* and ALMA.

We compared these to the observed counts in GOODS-South, splitting the field into two parts: the HUDF, which is deeper, and the rest of the field. We recall that stars were excluded from the observed catalogs.

The agreement is found to be excellent in the NIR. Because these wavelengths are most closely correlated to the stellar mass of the galaxies, and since the mock catalog was built to reproduce exactly the stellar mass function in GOODS-South, this should not come as a surprise. Reproducing the UV-optical (e.g., *F435W*) fluxes is less trivial, because these bands rather trace the emerging UV light coming from star formation, modulated by dust extinction. Nevertheless, the agreement here is also good. For all bands from *F435W* to *Spitzer* IRAC channel 4, we find a reduced $\chi^2 < 2$ between simulated and observed counts in the regimes where observations are complete.

Figure 9 shows instead the MIR, FIR and sub-mm counts in a few selected bands (other bands can be found in Appendix C). Since the number of detected galaxies in this wavelength domain is low, both observed and simulated counts are more strongly affected by statistical fluctuations than in the optical. To mitigate this effect, we used here the four CANDELS fields covered with deep *Herschel* data, rather than just GOODS-South. In addition, to estimate the amplitude of these fluctuations, we produced 100 realizations of the mock catalog, each time using a different seed to initialize the random number generator. The spread in counts among all realizations gave us a first order estimate of the cosmic variance. We then computed the average counts among all realizations, and compared this against the observations. This comparison shows that our simulation is able to capture the right shape and normalization of the counts, including in particular the turnover of the MIPS 24 μm counts around 200 μJy. The observations in the other bands are too shallow to probe this regime, but we confirm at least that the power law slope at the bright end is correctly reproduced.

In the same figure we show a prediction of the number counts at 1.2 mm, a wavelength domain in which our FIR SEDs and recipes are not calibrated. The agreement with published number counts from recent ALMA and single dish AzTEC observations is also satisfactory, reinforcing the validity of our approach. EGG also predicts 25^{+8}_{-7} detections above $S_{1.3\ \mathrm{mm}} > 150\ \mu$Jy in the 4.5 arcmin^2 of the HUDF, which is consistent with the 16 detections found by Dunlop et al. (2016). The error bars at the faint end are very permissive though, owing to the small area covered by ALMA to date. Single dish data covers much larger fields, but are limited both in depth and angular resolution. Poor angular resolution causes blending issues, and affect the measured flux catalogs by merging multiple moderately bright sources – too close to be reliably separated – into a single brighter one (see, e.g., Hodge et al. 2013). This can artificially boost the counts at the bright end, and make the comparison with models difficult. To some extent, a similar issue must be affecting the *Herschel* SPIRE fluxes.

7.2. Pixel statistics

To get rid of the uncertainty caused by blending, we also directly analyzed the pixel value distributions of the simulated maps (Fig. 10), avoiding the problems of having to find the right (number of) counterparts for the FIR sources. This procedure also takes into account the effect of clustering, which will tend to increase the contrast of the map without actually changing the number counts. To build the simulated *Spitzer* and *Herschel* images, we used the `egg-gennoise` tool to produce empty maps only containing Gaussian noise. The amplitude of this noise was adjusted to match the rms of empty regions in the observed maps. We then rescaled the fluxes of our simulated galaxies to match the zero point of the image, and painted them on the map using `egg-genmap`. Since the angular resolution of these images is low, we considered our galaxies to be unresolved and model them all as point sources, convolved with the observed PSF from GOODS-South. The resulting maps were finally median subtracted, as were the observed maps.

To display and analyze the full dynamic ranges of the images, we built histograms of the pixel values divided by the noise RMS, then rescaled by the hyperbolic arcsine function (asinh). This is similar to a logarithmic scale, except that it behaves linearly close to zero, which allows a proper representation of both negative fluctuations from the background and extreme values from bright objects.

We found the agreement with the observed maps to be good over most of the dynamic range, except possibly at the very bright end. In fact, the downside of this approach is that the bright pixel counts are very sensitive to statistical fluctuations, and a single bright (but usually rare) object can drastically impact the measured distributions. This effect can be seen in Fig. 10 for the 24 to 160 μm bands. In practice, the brightest pixels in the observed GOODS-South 24 μm map belong to: a) a $z = 0.3$ strong starburst (factor of ten above the MS) with an AGN; b) a pair of interacting $z = 0.07$ galaxies; and c) another $z = 0.03$ galaxy. Except for the AGN, these objects could be found in the simulation, although they will be rare because the volume probed at these redshifts is small. Furthermore, the minimum redshift we imposed in the simulation ($z = 0.05$) prevents us from properly sampling this low redshift population.

7.3. Comparison to semi-analytic models

We now compare EGG to mock catalogs generated from state-of-the-art semi-analytic models. In particular, the predicted number counts in the UV-NIR are compared in Fig. 11.

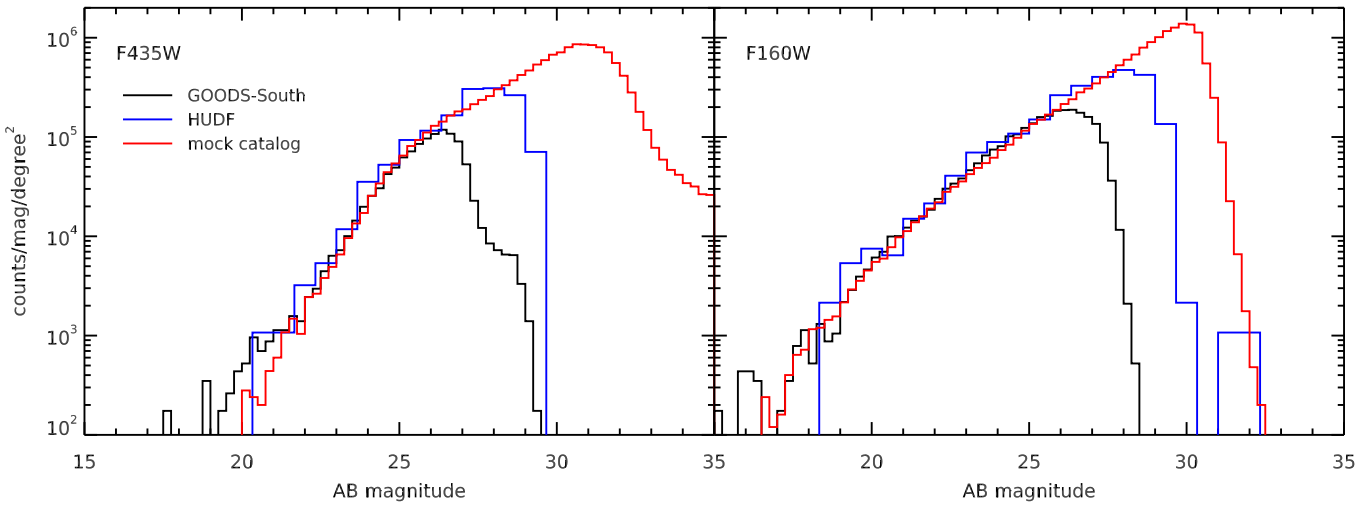

Fig. 8. Observed magnitude distribution from the HST *F435W* and *F160W* bands (other bands can be found in the appendix). The simulated fluxes (red histogram) come from a mock field of $10' \times 10'$ that is 90% complete down to $H < 30$. These are compared to the observed fluxes in the *Hubble* Ultra Deep Field (HUDF, blue) and the rest of the GOODS-South field (shallower, in black). Stars were excluded from the observed counts, since they are absent from the simulation.

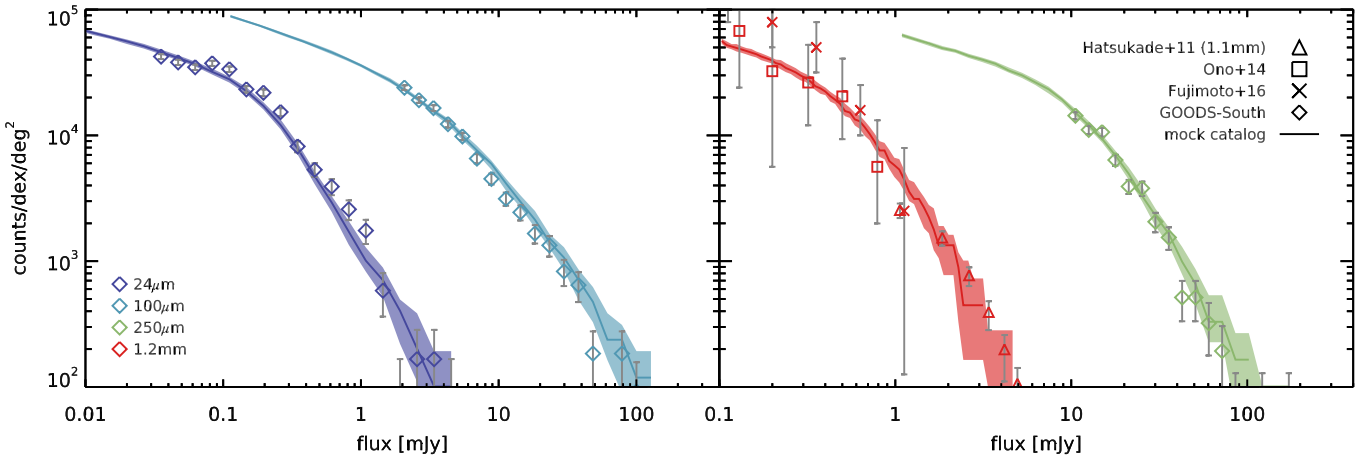

Fig. 9. Number counts in the MIR (24 μm), FIR (100 and 250 μm) and sub-mm (1.2 mm). Additional MIR and FIR bands can be found in the appendix. The observed counts in the CANDELS fields are reported as open diamonds with Poisson error bars, and we compare these to the median counts of 100 simulated catalogs, shown with a solid line of the same color. In the background, the shaded area show the range covered between the 16th and 84th percentile of 100 simulated catalogs, to illustrate how much scatter one should expect simply due to cosmic variance. Observed counts for 24 μm are from GOODS only, since they are substantially deeper. For 1.2 mm, observations are from Ono et al. (2014) and Fujimoto et al. (2016) using ALMA 1.2 mm and Hatsukade et al. (2011) using AzTEC 1.1 mm, reported with open squares, crosses and open triangles, respectively. The 1.1 mm fluxes were scaled down by a factor 0.77 following Fujimoto et al. (2016). These observations were not done in GOODS-South, and are therefore affected by a different cosmic variance.

We investigated first the Theoretical Astronomy Observatory (TAO[5]; Bernyk et al. 2016), which uses the SAGE SAM (Croton et al. 2016) and an additional module to produce galaxy SEDs and broadband fluxes. We used the TAO v3.0 web interface to generate a mock catalogs of about 0.2 deg^2, covering $0.01 < z < 10$, and keeping all galaxies with a K_s-band magnitude brighter than 31. We generated apparent AB magnitudes in all *Hubble*, *Spitzer* and *Herschel* passbands, as well as the U and K_s bands. Producing the catalog took about 27 h (against seven minutes for EGG with a similar setup).

While we requested a generous limiting magnitude of 31 in the K_s band, we found that the TAO mock catalog was incomplete below $K \sim 27$, or equivalently below a stellar mass of $\sim 10^8$ $M_\odot$ at nearly all redshifts, because of the finite mass resolution of the underlying dark matter simulation. This is impractical if one is interested in simulating faint galaxies, however this limit is still below the depth of CANDELS and allows a fair comparison with observations, which we provide in Fig. 11.

We found that the shape of the counts is well reproduced by TAO, however these predicted counts were systematically brighter than the observed counts by about one magnitude in all passbands from the *F435W* to the *F160W*. In the *Spitzer* IRAC passbands, the situation was worse and the shape of the TAO counts deviated significantly from the observations, overproducing galaxies at magnitudes fainter than ~20. In contrast, the counts predicted by EGG followed very closely the observations with no significant deviation.

We also compared EGG against the Henriques et al. (2015) SAM (hereafter H15), which can produce fluxes of galaxies from the U up to the *Spitzer* IRAC bands. Among the available

[5] https://tao.asvo.org.au/tao/

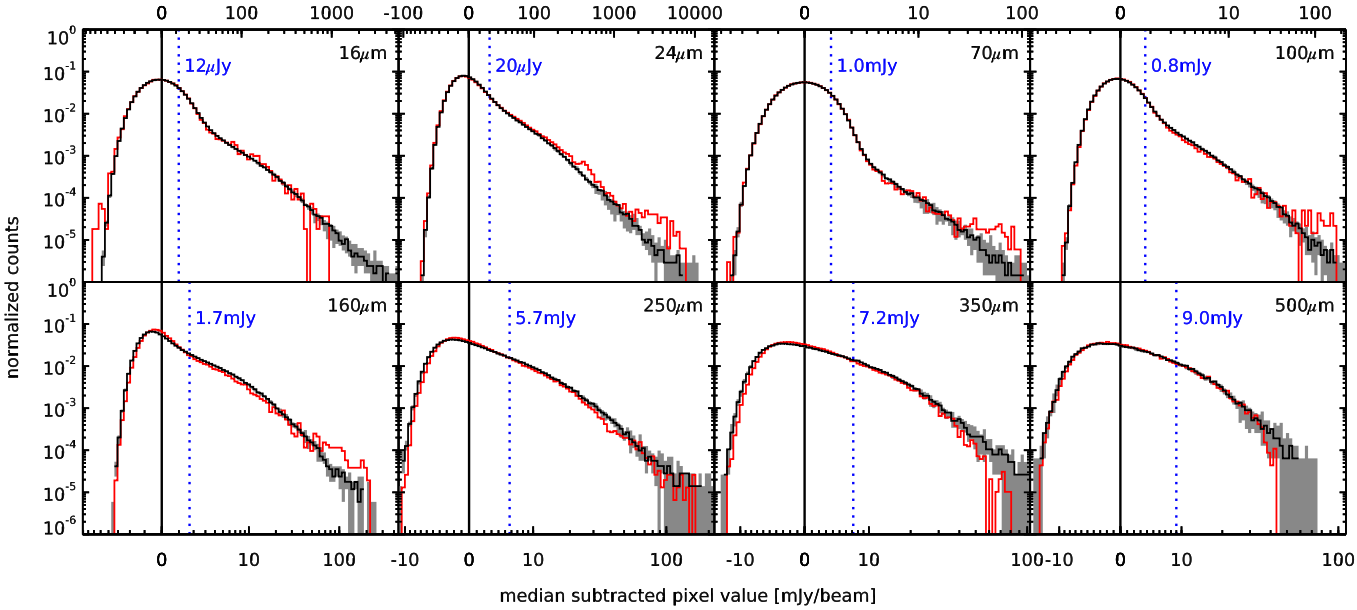

Fig. 10. Pixel value distributions of the 16 μm to 500 μm maps, in μJy/beam for 16 and 24 μm and mJy/beam otherwise. We show the observed distribution in GOODS-South in red, and compare this reference to 100 simulated catalogs generated with different random realizations. The median of these 100 realizations is shown with a solid black line, while the range covered by the 16th and 84th percentiles is shaded in gray in the background. Each map is median-subtracted, and the pixel values displayed here are scaled using the hyperbolic arcsine function (see text). We show the location of the median of the map with a vertical solid black line, and the 3σ point-source detection limit with a vertical blue dotted line.

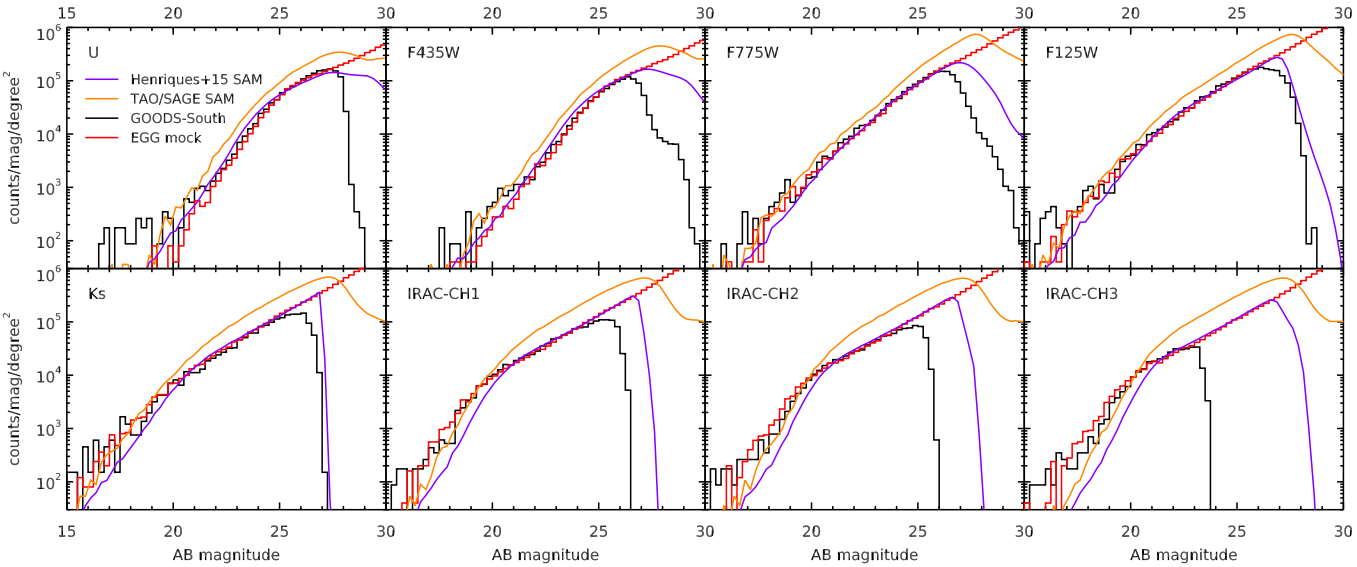

Fig. 11. Same as Fig. 8, but showing more bands and adding the counts predicted by the TAO/SAGE SAM (orange curve) and the H15 SAM (purple curve).

pre-built catalogs[6], we considered a single 3 deg^2 light cone selected with $K < 27$. We found that the counts predicted by the H15 model were in excellent agreement with the observations in all bands, except at the bright end in the NIR where the model was underpredicting the counts. As discussed in Henriques et al. (2012), this is most likely caused by the lack of dust emission in their model, which would increases the NIR-MIR fluxes of the bright local galaxies in the field of view.

Reproducing the far-IR counts is generally more challenging for SAMs. As pointed out by Gruppioni et al. (2015), they typically fail at producing the most actively star-forming galaxies, which are necessarily dusty. This inevitably impedes their ability to reproduce far-IR counts. While TAO is currently advertised to generate fluxes from the UV to the FIR, we found that the current implementation failed at generating any sensible dust emission in the mock galaxies; fluxes were too low by several orders of magnitude and SEDs were clearly unphysical. This was most likely an implementation issue, and we therefore did not attempt to compare these fluxes here. Instead, we compared our results to the IR counts predicted by the Somerville et al. (2012, hereafter S12) and Lacey et al. (2016, hereafter L16) SAMs in Fig. 12.

The S12 model reproduced correctly the bright counts at 24 μm, but clearly overpredicted the faint end, below 0.2 mJy.

[6] http://galformod.mpa-garching.mpg.de/public/LGalaxies/downloads.php

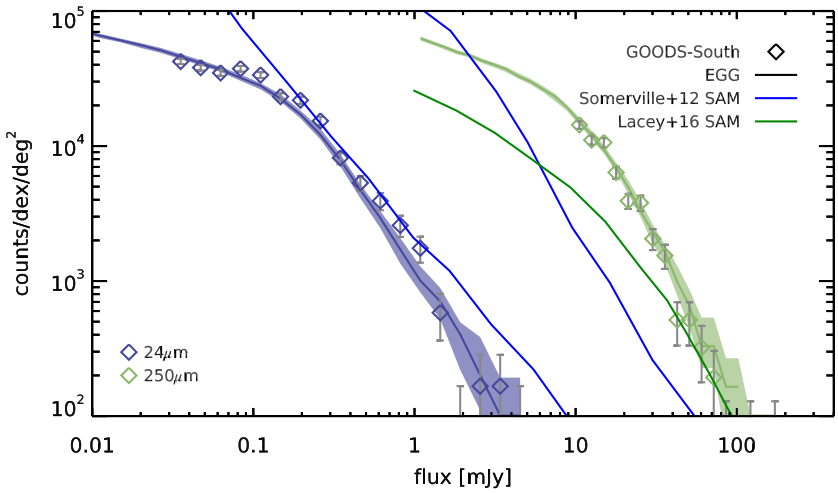

Fig. 12. Same as Fig. 9, but showing only the 24 and 250 μm bands and adding the counts predicted by the S12 SAM (blue curves) and the L16 SAM (green curve).

At longer wavelengths, the SPIRE 250 μm counts were about an order of magnitude lower than observations for 10 mJy $< S_{250} <$ 100 mJy. The L16 model provided a better fit at the bright end, but still globally underpredicted the counts.

In summary, different SAMs have different strengths. Some may give an accurate description of the counts in a given wavelength regime, but will fail (or not provide any prediction) in another. Ultimately, while they have an important predictive power that is lacking to EGG, no single SAM is currently able to reproduce the emission of galaxies from the UV to the FIR at the redshifts considered here.

7.4. Example of simulated images

A more qualitative way to assess the realism of the mock catalogs is to produce mock images of the sky and compare them by eye to observed ones. For this test, we used the FIR images introduced in the previous section, and built synthetic *Hubble* and *Spitzer*-IRAC images by feeding the flux catalogs produced by `egg-gencat` to SkyMaker (Bertin 2009). Contrary to `egg-genmap`, SkyMaker takes into account the detailed morphology we generated for each galaxy in Sect. 3.2. We tuned the background surface brightness and exposure time in SkyMaker to match the noise amplitude of the observed images and used the observed PSFs from the GOODS-South field; the `egg-2skymaker` script is provided with EGG to make the transition to SkyMaker straightforward and to automatically prepare the configuration files.

We display individual bands separately in Fig. 13, and a three-color image in Fig. 14 (top). From afar, no obvious difference can be spotted. Only when zooming in on the high resolution HST image, as in Fig. 14 (bottom), can one identify galaxies with complex morphologies (e.g., clumps or spiral arms) that are by construction absent from the simulation.

These images and the associated catalogs will be made publicly available on the ASTRODEEP website.

8. Limitations and missing ingredients

The main objective of EGG is to simulate deep cosmological fields similar to the CANDELS fields. Because our recipes are generic and easily extensible, EGG can also be used to simulate wider areas up to a few square degrees, like the COSMOS field (Scoville et al. 2007), or fields deeper than what *Hubble* has currently observed. However, this requires some caution:

- As shown in Appendix C, our simulated number counts are consistent with the observations in 48 deg^2 from GAMA (Hill et al. 2011) up to bright magnitudes from the g to z bands, but tend to under-predict the counts in the U band (at magnitudes <20 by a factor two) and over-predict the counts in the K band (at magnitudes <18 by a similar factor). Improving the agreement here would require a finer tuning of our recipes for low redshifts ($z < 0.3$), which we have mostly neglected in the present work.
- The counts in the MIR-to-submm at the extreme bright end can be significantly affected by strongly lensed galaxies (e.g., Béthermin et al. 2011), which we did not consider in the model.
- As noted in Sect. 5, our method to implement clustering will not generate any correlation on scales larger than 3′, which is about an order of magnitude too small for galaxies in the local Universe (e.g., Connolly et al. 2002).
- When going deeper than $H \sim 29$, the catalogs rely on extrapolations of our recipes in domains that still remain to be observationally constrained (very low mass galaxies at low redshift, or very high redshift galaxies). For example, when constructing the stellar mass functions, we assumed for simplicity that the fraction of quiescent galaxies remains constant at all $z > 4$ at the level of 15%. This assumption is totally unconstrained at present. Similarly, we had to explicitly adjust the M/L ratios of high redshift galaxies to be able to reproduce the $z > 5$ UV luminosity functions, and similar correction might be required for $z > 8$ galaxies that will be detected by future facilities.

Therefore, while EGG can technically produce a catalog of any area and depth, it is important to consider the aforementioned limitations before attempting to do so.

Another limitation already mentioned in the previous section, and illustrated in Fig. 14 (bottom), is that the morphologies of our galaxies are idealistic. In particular, in EGG disks are only described as smooth profiles, based only on their projected axis ratio (i.e., inclination) and half-light radius; no detail is simulated on the spiral arm structure, bars, or the presence of clumps and dust lanes. Given a statistical description of their occurrence (e.g., Simmons et al. 2014; Guo et al. 2015; Willett et al. 2015), spiral arms, bars and clumps can be straightforwardly added as a post-processing step outside of EGG by redistributing the flux of the disk, and mock images of these galaxies can be created using GALFIT. Dust lanes on the other hand are more complex to simulate, since they should modify both morphology and colors as function of the disk's inclination (e.g., Tuffs et al. 2004). Reddening is presently independent on inclination in EGG, so the recipe to generate axis ratios would need to be modified to take

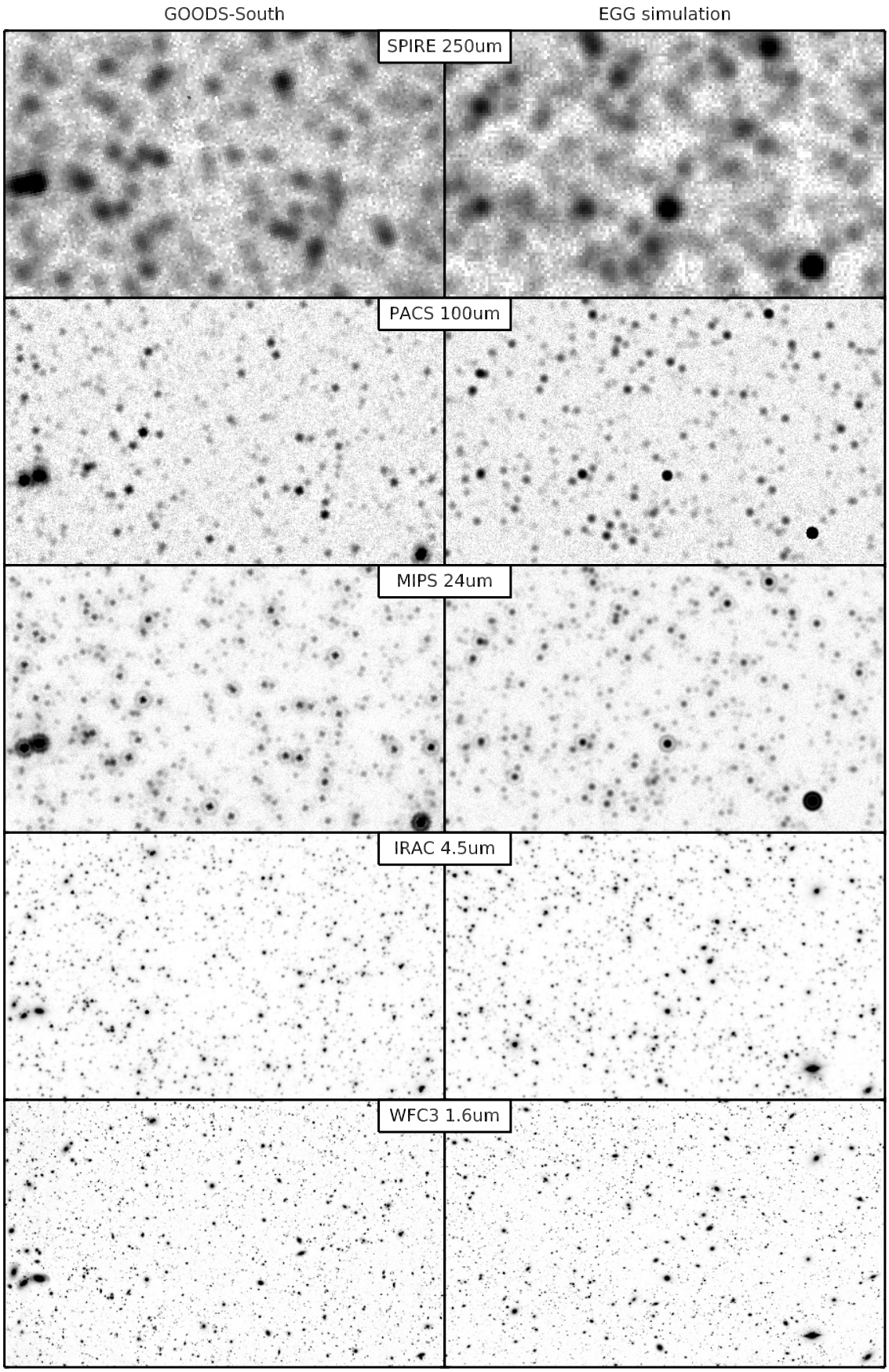

Fig. 13. Comparison of real images from GOODS-South (*left*) and a random excerpt from a simulated field produced by EGG (*right*). Each image is $9' \times 5.4'$. From *top to bottom* (*FWHM* of the PSF): *Herschel* SPIRE 500 μm (36.6″) and PACS 100 μm (6.7″); *Spitzer* MIPS 24 μm (5.7″) and IRAC-ch2 4.5 μm (1.6″); and *Hubble* WFC3-F160W 1.6 μm (0.19″). Images of a given band are shown with the same color bar.

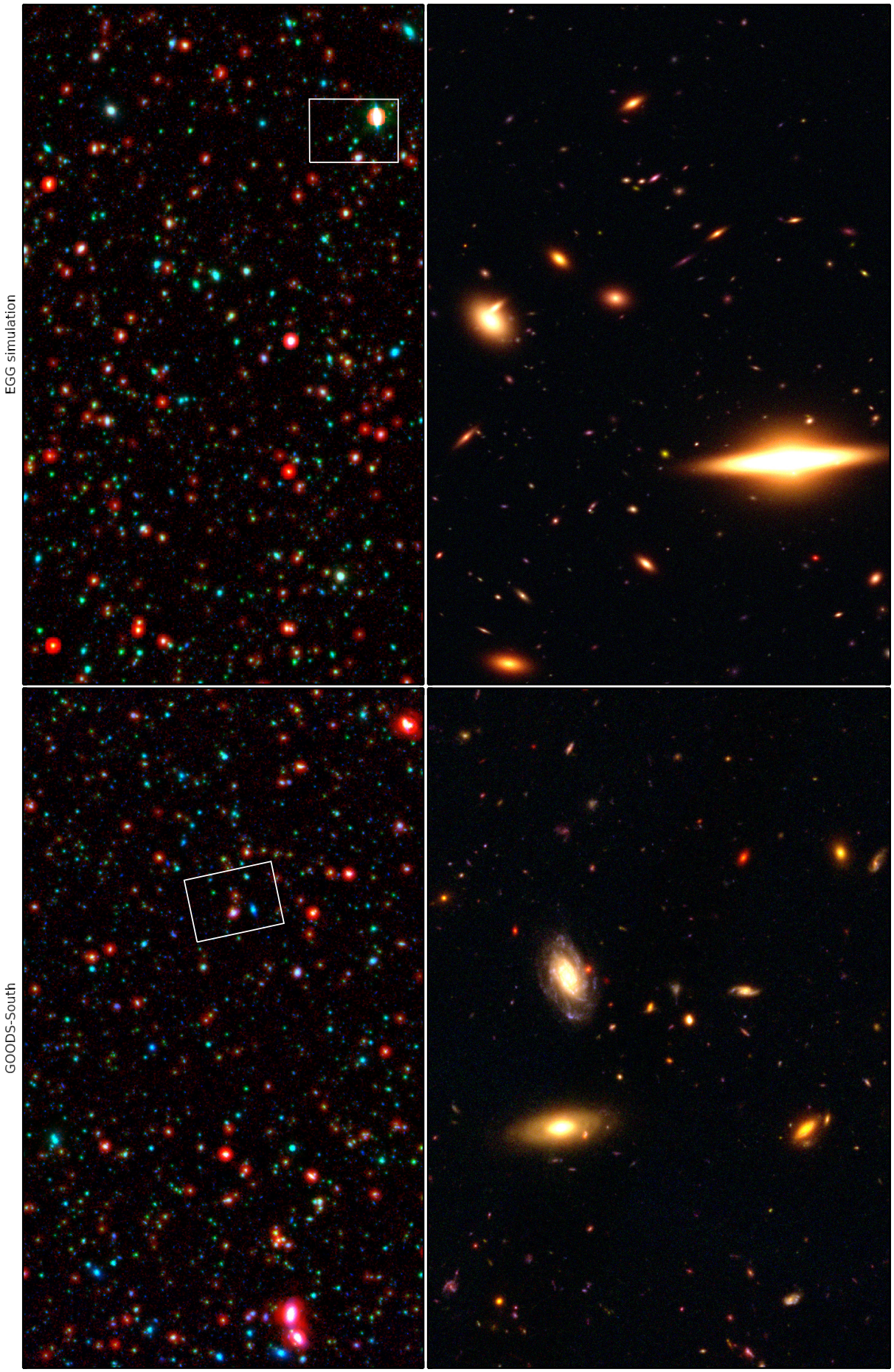

Fig. 14. *Top*: three-color version of Fig. 13, with *Hubble* WFC3-F160W (blue), *Spitzer* IRAC-ch2 (green) and *Herschel* PACS 100 μm (red). The real images from GOODS-South are shown on the *left*, and the simulated images are shown on the *right* with the same color bar. *Bottom*: zoom-in on a small region of the above image, showing the *Hubble* ACS-F435W (blue), ACS-F850LP (green) and WFC3-F160W (red) bands. These regions were chosen to contain large disk galaxies.

into account the chosen SED of the galaxy (or conversely). It is also questionable whether high redshift galaxies can be described by a simple disk+bulge model. High redshift galaxies at $z > 3$ tend to have more irregular morphologies (see, e.g., Kartaltepe et al. 2015; Huertas-Company et al. 2016), although part of this is related to the fact that *Hubble* only observes their rest-frame UV emission, which is affected by variable dust extinction and bright clumps; it is unknown whether disks already existed at these epochs, and only upcoming rest-optical imaging from *JWST* will answer this question.

Future improvements to the tool will aim toward the inclusion of emission lines in the generated spectra, including lines in both the UV-optical (e.g., Lyα, Hα, Hβ, [O II], [O III]) and the FIR (i.e., mainly [C II] and the CO ladder). While being a convenient mean to estimate line fluxes and integration time when designing observing proposals, this will most importantly allow the simulation of grism data, which will be particularly useful for Euclid. Adding emission lines to the simulation will also improve the accuracy of the rest-optical SEDs of high-redshift galaxies ($z > 5$) which are know to have emission lines of particularly large equivalent widths (e.g., Stark et al. 2013). At this stage, the inclusion of AGNs will become an important step to properly reproduce the line ratios and luminosity functions (e.g., Kauffmann et al. 2003; Hao et al. 2005). This would also allow us to refine the counts at the bright end in the mid-IR (see Sect. 7.1). Once AGNs are added to the simulation, the spectrum of each galaxy can be safely extended to cover more extreme wavelength regimes, including in particular the X-ray and the radio (the simulation of the X-ray domain has already been successfully attempted in a companion paper; Cappelluti et al. 2016).

Lastly, another improvement that will come in a future version is the inclusion of foreground stars from the Milky Way, which can be used to evaluate the efficiency of standard algorithms for stars-galaxies separation in deep fields. Stars also dominate the bright end of the counts, and the halos of the brightest ones can severely impact one's ability to recover accurate fluxes in their surroundings; it can therefore be an important component especially in the simulation of wider fields where such stars are common.

9. Conclusions

We presented EGG, a simple tool to generate realistic galaxy mock catalogs. Produced and released to the astrophysics community as part of the ASTRODEEP collaboration, this program was designed to run on personal computers and can be used to create ab initio complete flux catalogs across the whole UV-to-submm spectrum for hundreds of thousands of galaxies, taking into account the emission of both stars and dust. The whole process takes only a few seconds of computation time on a single CPU.

The approach chosen to generate each galaxy is purely empirical, in that it is based on observed distributions and scaling relations, the stellar mass and the redshift of the galaxy being the two main drivers of the generated properties and spectra. The strength of this approach is that, almost by construction, it is able to accurately reproduce not only the flux distributions within the whole mock catalog, but also the relations between the individual fluxes of each galaxies. The catalogs created by this tool can therefore be used to optimize or validate the usual flux extraction techniques on real images, as well as to provide hints regarding the proper choice of priors when extracting highly confused images.

Lastly, since counts simulated by EGG accurately reproduce the observations, they can also be extrapolated to fainter limits to predict the outcome of future surveys, for example, with *JWST* or ALMA. To this end, we provide tabulated number counts (both differential and cumulative) on the EGG website (see Appendix C).

Acknowledgements. The authors want to thank the anonymous referee for their comments that improved the consistency and overall quality of this paper. EGG is written using phy++, a free and open source C++ library for fast and robust numerical astrophysics (http://cschreib.github.io/phypp/). This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. This research was supported by the French Agence Nationale de la Recherche (ANR) project ANR-09-BLAN-0224 as well as from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 312725.

References

Arnouts, S., Le Floc'h, E., Chevallard, J., et al. 2013, A&A, 558, A67
Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, MNRAS, 421, 621
Bernhard, E., Béthermin, M., Sargent, M., et al. 2014, MNRAS, 442, 509
Bernyk, M., Croton, D. J., Tonini, C., et al. 2016, ApJS, 223, 9
Bertin, E. 2009, Mem. Soc. Astron. It., 80, 422
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Béthermin, M., Dole, H., Cousin, M., & Bavouzet, N. 2010, A&A, 516, A43
Béthermin, M., Dole, H., Lagache, G., Le Borgne, D., & Penin, A. 2011, A&A, 529, A4
Béthermin, M., Daddi, E., Magdis, G., et al. 2012, ApJ, 757, L23
Béthermin, M., Daddi, E., Magdis, G., et al. 2015, A&A, 573, A113
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, MNRAS, 440, 2810
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bruce, V. A., Dunlop, J. S., McLure, R. J., et al. 2014, MNRAS, 444, 1001
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Buat, V., Iglesias-Páramo, J., Seibert, M., et al. 2005, ApJ, 619, L51
Buat, V., Noll, S., Burgarella, D., et al. 2012, A&A, 545, A141
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cappelluti, N., Comastri, A., Fontana, A., et al. 2016, ApJ, 823, 95
Castellano, M., Amorín, R., Merlin, E., et al. 2016, A&A, 590, A31
Chary, R., & Elbaz, D. 2001, ApJ, 556, 562
Ciesla, L., Boquien, M., Boselli, A., et al. 2014, A&A, 565, A128
Connolly, A. J., Scranton, R., Johnston, D., et al. 2002, ApJ, 579, 42
Cooper, M. C., Newman, J. A., Croton, D. J., et al. 2006, MNRAS, 370, 198
Croton, D. J., Stevens, A. R. H., Tonini, C., et al. 2016, ApJS, 222, 22
Cucciati, O., Iovino, A., Marinoni, C., et al. 2006, A&A, 458, 39
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
Daddi, E., Renzini, A., Pirzkal, N., et al. 2005, ApJ, 626, 680
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Dale, D. A., & Helou, G. 2002, ApJ, 576, 159
De Santis, C., Grazian, A., & Fontana, A. 2006, Mem. Soc. Astron. It. Suppl., 9, 454
Dole, H., Rieke, G. H., Lagache, G., et al. 2004, ApJS, 154, 93
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, MNRAS, 466, 861
Dutton, A. A., van den Bosch, F. C., Faber, S. M., et al. 2011, MNRAS, 410, 1660
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, ApJ, 600, L107
Finkelstein, S. L., Ryan, Russell E., J., Papovich, C., et al. 2015, ApJ, 810, 71
Franceschini, A., Aussel, H., Cesarsky, C. J., Elbaz, D., & Fadda, D. 2001, A&A, 378, 1
Fujimoto, S., Ouchi, M., Ono, Y., et al. 2016, ApJS, 222, 1
Galametz, A., Grazian, A., Fontana, A., et al. 2013, ApJS, 206, 10
Galliano, F., Dwek, E., & Chanial, P. 2008, ApJ, 672, 214
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, MNRAS, 445, 175
Goldader, J. D., Meurer, G., Heckman, T. M., et al. 2002, ApJ, 568, 651
Grazian, A., Fontana, A., Santini, P., et al. 2015, A&A, 575, A96
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Gruppioni, C., Pozzi, F., Zamorani, G., & Vignali, C. 2011, MNRAS, 416, 70
Gruppioni, C., Pozzi, F., Rodighiero, G., et al. 2013, MNRAS, 432, 23
Gruppioni, C., Calura, F., Pozzi, F., et al. 2015, MNRAS, 451, 3419
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39

Hao, L., Strauss, M. A., Fan, X., et al. 2005, AJ, 129, 1795
Hatsukade, B., Kohno, K., Aretxaga, I., et al. 2011, MNRAS, 411, 102
Hatziminaoglou, E., Omont, A., Stevens, J. A., et al. 2010, A&A, 518, L33
Heinis, S., Buat, V., Béthermin, M., et al. 2014, MNRAS, 437, 1268
Henriques, B. M. B., White, S. D. M., Lemson, G., et al. 2012, MNRAS, 421, 2904
Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, MNRAS, 451, 2663
Hill, D. T., Kelvin, L. S., Driver, S. P., et al. 2011, MNRAS, 412, 765
Hodge, J. A., Karim, A., Smail, I., et al. 2013, ApJ, 768, 91
Huertas-Company, M., Bernardi, M., Pérez-González, P. G., et al. 2016, MNRAS, 462, 4495
Kartaltepe, J. S., Mozena, M., Kocevski, D., et al. 2015, ApJS, 221, 11
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, ApJ, 649, L67
Labbé, I., Franx, M., Rudnick, G., et al. 2007, ApJ, 665, 944
Lacey, C. G., Baugh, C. M., Frenk, C. S., et al. 2016, MNRAS, 462, 3854
Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Lang, P., Wuyts, S., Somerville, R. S., et al. 2014, ApJ, 788, 11
Madden, S. C., Galliano, F., Jones, A. P., & Sauvage, M. 2006, A&A, 446, 877
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6
Magnelli, B., Lutz, D., Saintonge, A., et al. 2014, A&A, 561, A86
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Masters, K. L., Mosleh, M., Romer, A. K., et al. 2010, MNRAS, 405, 783
Merlin, E., Fontana, A., Ferguson, H. C., et al. 2015, A&A, 582, A15
Merlin, E., Amorín, R., Castellano, M., et al. 2016, A&A, 590, A30
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Michałowski, M. J., Hayward, C. C., Dunlop, J. S., et al. 2014, A&A, 571, A75
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Newman, A. B., Ellis, R. S., Bundy, K., & Treu, T. 2012, ApJ, 746, 162
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43
Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793
Nordon, R., Lutz, D., Genzel, R., et al. 2012, ApJ, 745, 182
O'Halloran, B., Satyapal, S., & Dudik, R. P. 2006, ApJ, 641, 795
Ono, Y., Ouchi, M., Kurono, Y., & Momose, R. 2014, ApJ, 795, 5
Oteo, I., Magdis, G., Bongiovanni, A., et al. 2013, MNRAS, 435, 158
Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, ApJ, 698, L116
Pannella, M., Elbaz, D., Daddi, E., et al. 2015, ApJ, 807, 141
Patel, S. G., Holden, B. P., Kelson, D. D., et al. 2012, ApJ, 748, L27
Peebles, P. J. E. 1982, ApJ, 263, L1
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H. 2002, AJ, 124, 266
Penner, K., Dickinson, M., Pope, A., et al. 2012, ApJ, 759, 28
Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, ApJS, 166, 470
Salpeter, E. E. 1955, ApJ, 121, 161
Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, ApJ, 801, 97
Sargent, M. T., Béthermin, M., Daddi, E., & Elbaz, D. 2012, ApJ, 747, L31
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Schreiber, C., Elbaz, D., Pannella, M., et al. 2016, A&A, 589, A35
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Shen, S., Mo, H. J., White, S. D. M., et al. 2003, MNRAS, 343, 978
Shu, X. W., Elbaz, D., Bourne, N., et al. 2016, ApJS, 222, 4
Simmons, B. D., Melvin, T., Lintott, C., et al. 2014, MNRAS, 445, 3466
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Somerville, R. S., Gilmore, R. C., Primack, J. R., & Domínguez, A. 2012, MNRAS, 423, 1992
Soneira, R. M., & Peebles, P. J. E. 1978, AJ, 83, 845
Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJ, 783, L14
Teplitz, H. I., Chary, R., Elbaz, D., et al. 2011, AJ, 141, 1
Tuffs, R. J., Popescu, C. C., Völk, H. J., Kylafis, N. D., & Dopita, M. A. 2004, A&A, 419, 821
van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24
van der Wel, A., Chang, Y., Bell, E. F., et al. 2014, ApJ, 792, L6
Willett, K. W., Schawinski, K., Simmons, B. D., et al. 2015, MNRAS, 449, 820
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Wu, Y., Charmandaris, V., Hao, L., et al. 2006, ApJ, 639, 157
Wuyts, S., Labbé, I., Franx, M., et al. 2007, ApJ, 655, 51

Table A.1. Double Schechter function parameters for the SFG population.

z	$\phi_1^\star$ dex^{-1} Mpc^{-3}	$\log_{10}(M_1^\star)$ $\log_{10}(M_\odot)$	α_1	$\phi_2^\star$ dex^{-1} Mpc^{-3}	$\log_{10}(M_2^\star)$ $\log_{10}(M_\odot)$	α_2
0.3–0.7	8.90×10^{-4}	[11]	[–1.4]	8.31×10^{-5}	10.64	[0.5]
0.7–1.2	7.18×10^{-4}	[11]	[–1.4]	4.04×10^{-4}	10.73	[0.5]
1.2–1.8	4.66×10^{-4}	[11]	[–1.5]	4.18×10^{-4}	10.67	[0.5]
1.8–2.5	2.14×10^{-4}	[11]	[–1.57]	4.06×10^{-4}	10.84	[0.5]
2.5–3.5	2.12×10^{-4}	[11]	[–1.6]	9.07×10^{-5}	10.94	[0.5]
3.5–4.5	4.45×10^{-5}	[11]	[–1.7]	8.60×10^{-6}	11.69	[0.5]

Notes. Parameters that were chosen manually are enclosed in brackets.

Table A.2. Double Schechter function parameters for the QG population.

z	$\phi_1^\star$ dex^{-1} Mpc^{-3}	$\log_{10}(M_1^\star)$ $\log_{10}(M_\odot)$	α_1	$\phi_2^\star$ dex^{-1} Mpc^{-3}	$\log_{10}(M_2^\star)$ $\log_{10}(M_\odot)$	α_2
0.3–0.7	7.77×10^{-5}	[11]	[–1.65]	1.54×10^{-3}	11.04	–0.48
0.7–1.2	3.54×10^{-5}	[11]	[–1.60]	1.04×10^{-3}	10.86	0.06
1.2–1.9	2.30×10^{-5}	[11]	[–1.25]	6.25×10^{-4}	10.83	0.30
1.9–2.5	[10^{-5}]	[11]	[–1]	1.73×10^{-4}	11.05	–0.17
2.5–3.5	[0]	[11]	[–1]	1.22×10^{-4}	10.94	–0.26
3.5–4.5	[0]	[11]	[–1.35]	[3×10^{-5}]	[11]	[–0.30]

Notes. Parameters that were chosen manually are enclosed in brackets.

Table C.1. Differential and cumulative number counts in UV to millimeter bands as produced by EGG.

	VIMOS u			ALMA 2.0 mm	
Flux density μJy	Differential counts dex^{-1} deg^{-2}	Cumulative counts deg^{-2}	...	Differential counts dex^{-1} deg^{-2}	Cumulative counts deg^{-2}
1.189×10^{-5}	2.273×10^{7}	2.158×10^{7}	...	3.066×10^{6}	4.854×10^{6}
1.682×10^{-5}	1.906×10^{7}	1.808×10^{7}	...	2.771×10^{6}	4.393×10^{6}
2.379×10^{-5}	1.599×10^{7}	1.514×10^{7}	...	2.527×10^{6}	3.975×10^{6}
...	...	...	...	...	...
4.642×10^{6}	0	0	...	0	0

Notes. This full table is only available at the CDS. We only show here a sample of columns and rows to help visualize the table format.

Appendix A: Fit parameters for the conditional stellar mass functions

We provide in Tables A.1 and A.2 the values of the parameters of the two Schechter functions we fit to the observed stellar mass functions, as described in Sect. 3.1.

Appendix B: Estimating the effective half-light radius of a double Sérsic profile

Using GALFIT, we created a suite of ideal double Sérsic profiles ($n = 1$ disk + $n = 4$ bulge) of varying relative fluxes and sizes. We then used GALFIT again to fit the generated image with a single Sérsic profile of variable index n and size $R_{50,\mathrm{total}}$. Using a simpler growth curve analysis, we confirmed that the resulting size is an unbiased measure of the effective half-light radius of the galaxy. Assuming the two components have the same projected axis ratio, we found that this value can be approximated as

$$R_{50,\mathrm{total}} = R_{50,\mathrm{disk}} \times (1 - (B/T)^\alpha) + R_{50,\mathrm{bulge}} \times (B/T)^\alpha, \quad \text{(B.1)}$$

where $\alpha = 1 - 0.8 \times \log_{10}(R_{50,\mathrm{disk}}/R_{50,\mathrm{bulge}})$.

Appendix C: Number counts and map statistics for additional bands

We provide in Table C.1 the differential and cumulative number counts in an exhaustive list of bands from the VIMOS u to the ALMA 2.0 mm bands. For illustration, the differential counts are also displayed in Figs. C.1 and C.2.

To obtain a wide dynamic range, we generated three catalogs with a "wedding cake" strategy. The first catalog covered 20 deg^2 down to [4.5] = 20 (where [4.5] is the AB magnitude in the *Spitzer* IRAC 4.5 μm band); the second catalog covered 1 deg^2 down to [4.5] = 30; and the last catalog covered 0.05 deg^2 down to [4.5] = 36. For all three catalogs, the minimum redshift was lowered to $z = 0.005$. These catalogs contained ~330 000, ~4 330 000 and ~3 400 000 galaxies, respectively. The widest catalog was used for counts below 100 dex^{-1} deg^{-2}, and the deepest catalog was used for counts above 300 000 dex^{-1} deg^{-2}.

The counts produced by EGG are compared to the observations in GOODS-South (open diamonds). For UV-NIR bands, we also show the counts from GAMA (Hill et al. 2011, open triangles), corrected for the different filter response curves.

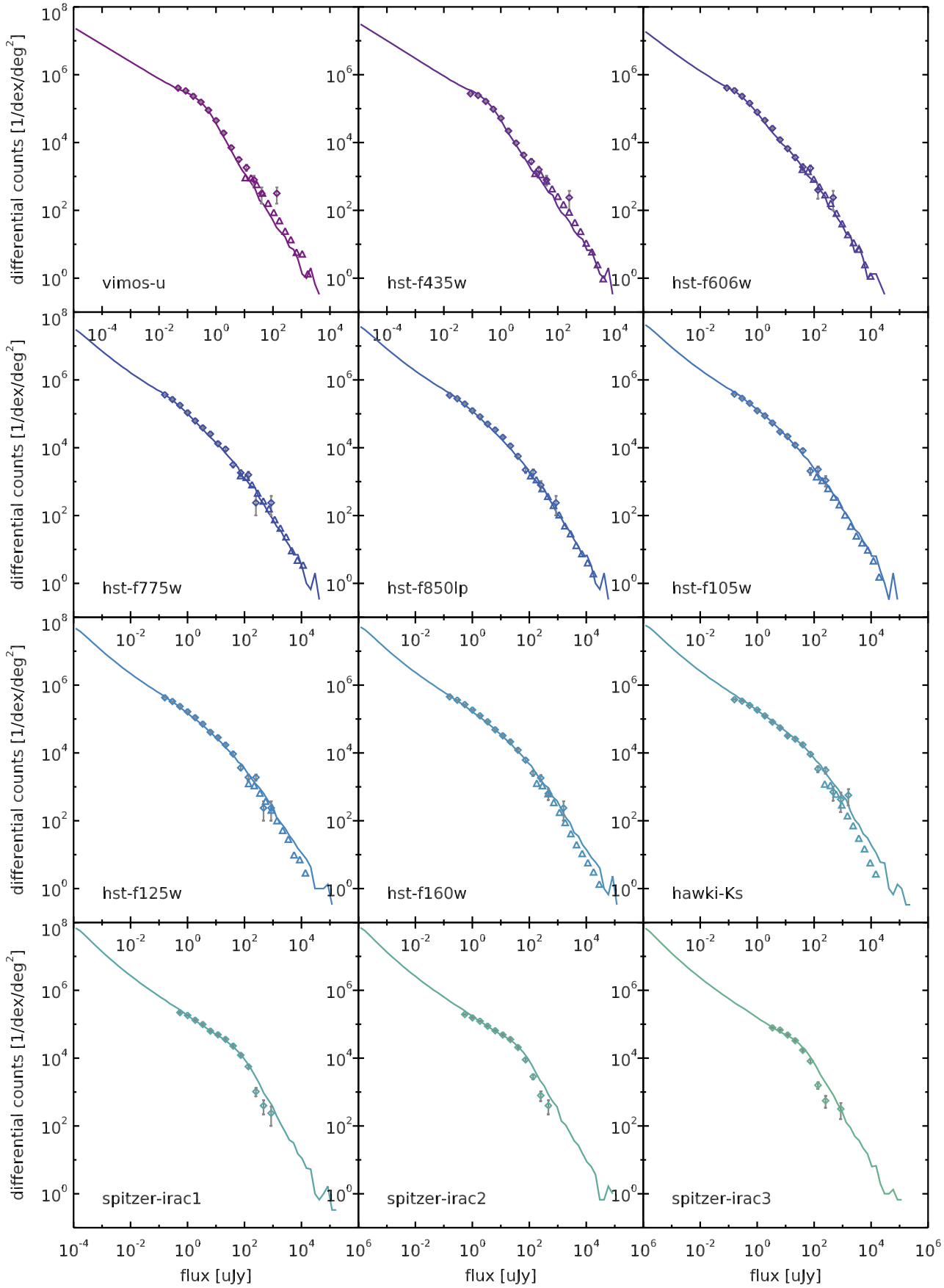

Fig. C.1. Differential counts produced by EGG (solid lines), compared to observations (open symbols). See text for details and references.

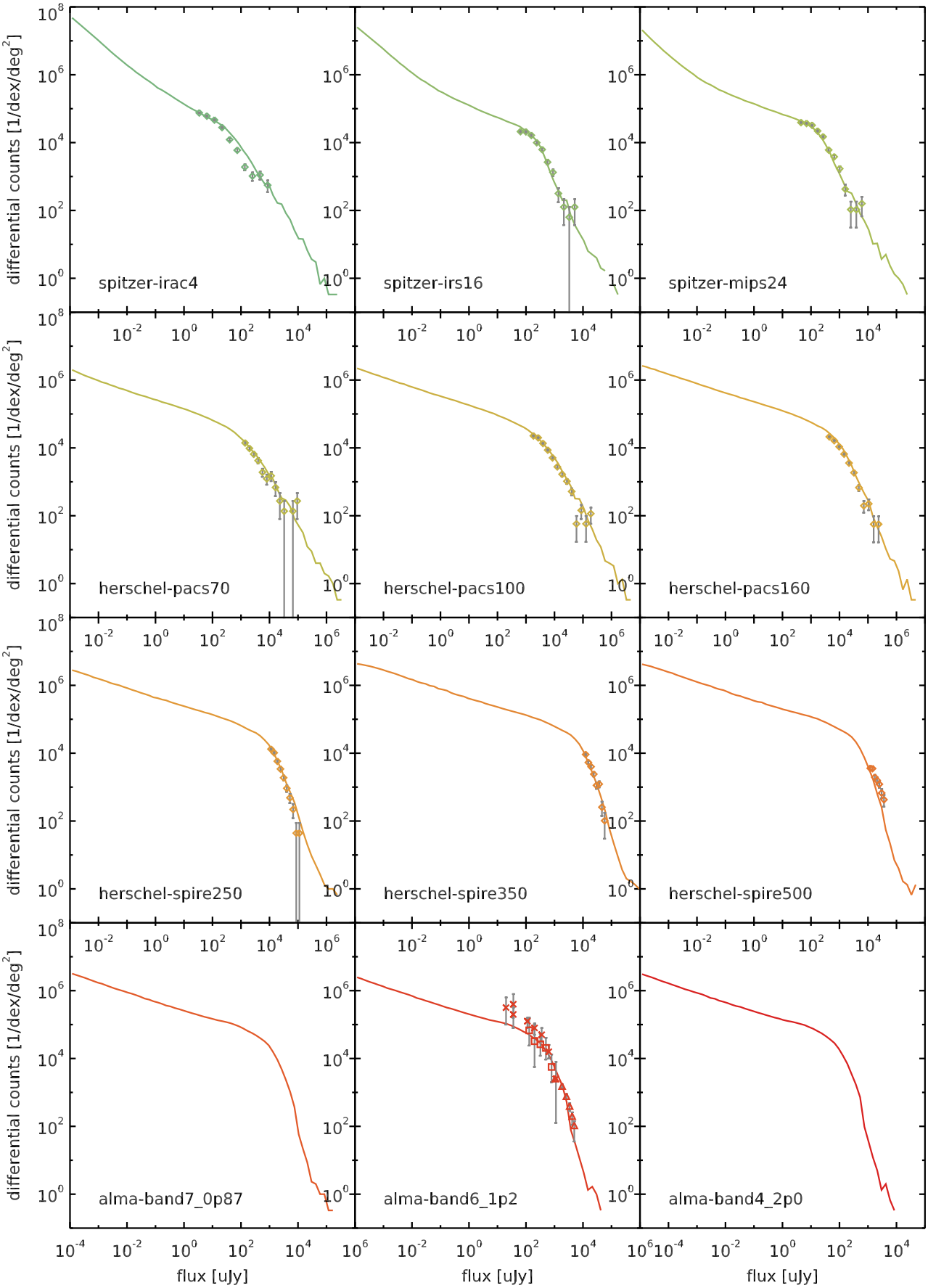

Fig. C.2. Same as Fig. C.1 for the remaining bands.

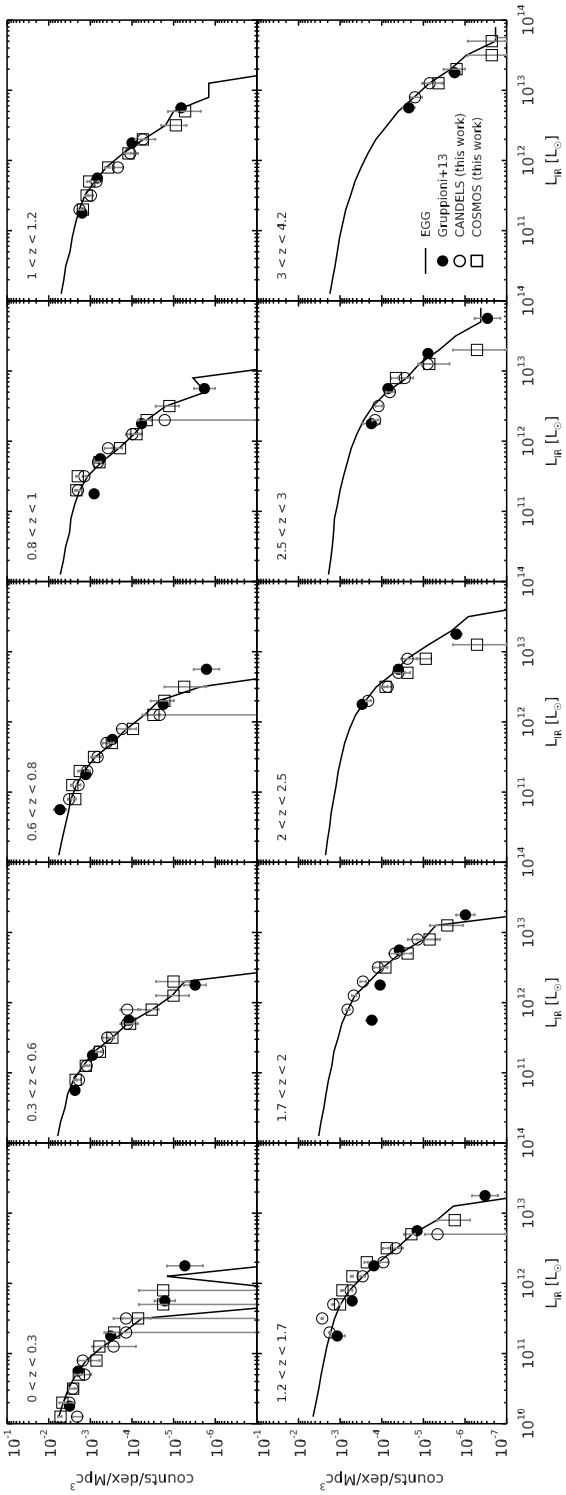

Fig. D.1. Infrared luminosity ($L_{\rm IR}$) function at various redshifts. The solid line is the output of EGG. It is compared against observations from our own catalogs in CANDELS (open circles) and COSMOS 2 deg^2 (open squares), and the results of Gruppioni et al. (2013) in the same fields (filled circles). All observations use either *Spitzer* MIPS or *Herschel* photometry to derive $L_{\rm IR}$.

To perform the correction, we generated fluxes in both the HST bands and the corresponding closest band from SDSS or UKIDSS, computed the median ratio of the two, and then multiplied the fluxes from GAMA by this factor. This correction is 1.07, 0.84, 0.93, 1.08, 1.05, 1.04, 1.0, 0.91 and 0.97 for the u, g, r, i, z, Y, J, H and K bands, respectively. For clarity, we only displayed one point every five from the Hill et al. data. Lastly, for ALMA 1.2 mm, we showed the same literature data as in Fig. 9.

These counts are tabulated and freely accessible on the EGG website[7], as well as on the CDS.

Appendix D: UV and IR luminosity functions

Figures D.1 and D.2 show, respectively, the IR and UV luminosity functions (LFs) produced by EGG in various redshifts bins and compare them against observations from our own catalogs or the literature. The IR LFs are in good agreement with the observations, although this is true almost by construction given that we use the same mass functions and the observed $L_{\rm IR}$–M_* relation. The UV luminosity is less clearly controlled by our simulation procedure, as it was indirectly determined by the *UVJ* colors and stellar masses. Still, the EGG LFs reproduce correctly the shape and normalization of the LFs at $3.5 < z < 7.5$, although our LFs seem to match better the straight power-law shape observed by Bowler et al. (2014) at $z = 7$, as opposed to an actual Schechter function (especially so at $z > 5.5$). At $z = 8$, our LF agrees better with the observations of Finkelstein et al. (2015) rather than that of Bouwens et al. (2015), which are each based on a different combination of *Hubble* deep fields. Given the small volumes we are dealing with at these high redshifts, this could be explained by cosmic variance. On the other hand, our stellar mass functions were not observationally constrained at this redshift, hence we could agree with either of these observational data by tweaking the high redshift extrapolation of the mass function.

[7] http://cschreib.github.io/egg/counts.html

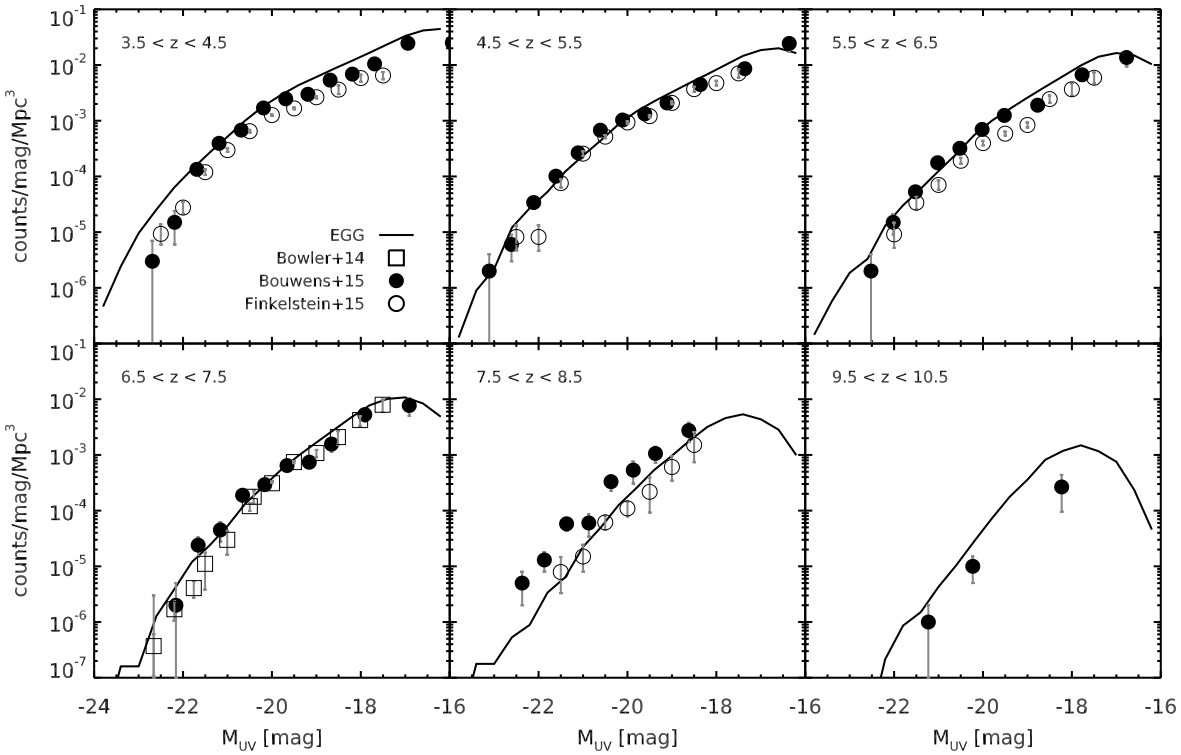

Fig. D.2. Ultraviolet absolute magnitude ($M_{\rm UV}$, 1600 Å) function at various redshifts. The solid line is the output of EGG. It is compared against observations from Bouwens et al. (2015, filled circles) and Finkelstein et al. (2015, open circles), each compiling multiple fields observed with *Hubble* WFC3, and Bowler et al. (2014, open squares, only at $z = 7$) which uses the UltraVISTA imaging of COSMOS.