



# **Multi-wavelength de-blended Herschel view of the statistical properties of dusty star-forming galaxies across cosmic time**

L. Wang, W. J. Pearson, W. Cowley, J. W. Trayford, M. Béthermin, C. Gruppioni, P. Hurley, M. J. Michalowski

## **► To cite this version:**

L. Wang, W. J. Pearson, W. Cowley, J. W. Trayford, M. Béthermin, et al.. Multi-wavelength de-blended Herschel view of the statistical properties of dusty star-forming galaxies across cosmic time. *Astronomy & Astrophysics* - A&A, 2019, 624, <10.1051/0004-6361/201834093>. <insu-03666761>

**HAL Id: insu-03666761**

**<https://insu.hal.science/insu-03666761v1>**

Submitted on 13 May 2022

**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.



HAL Authorization

# Multi-wavelength de-blended *Herschel* view of the statistical properties of dusty star-forming galaxies across cosmic time

L. Wang<sup>1,2</sup>, W. J. Pearson<sup>1,2</sup>, W. Cowley<sup>2</sup>, J. W. Trayford<sup>3</sup>, M. Béthermin<sup>4</sup>, C. Gruppioni<sup>5</sup>,  
P. Hurley<sup>6</sup>, and M. J. Michałowski<sup>7</sup>

<sup>1</sup> SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, The Netherlands  
e-mail: l.wang@sron.nl

<sup>2</sup> Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands

<sup>3</sup> Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

<sup>4</sup> Aix Marseille Univ., CNRS, LAM, Laboratoire d'Astrophysique de Marseille, Marseille, France

<sup>5</sup> Istituto Nazionale di Astrofisica (INAF) – Osservatorio Astronomico di Bologna, via Gobetti 93/3, 40129 Bologna, Italy

<sup>6</sup> Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Falmer, Brighton BN1 9QH, UK

<sup>7</sup> Astronomical Observatory Institute, Faculty of Physics, Adam Mickiewicz University, ul. Słoneczna 36, 60-286 Poznań, Poland

Received 15 August 2018 / Accepted 25 February 2019

## ABSTRACT

**Aims.** We study the statistical properties of dusty star-forming galaxies across cosmic time, such as their number counts, luminosity functions (LF), and the dust-obscured star formation rate density (SFRD).

**Methods.** We used the most recent de-blended *Herschel* catalogue in the COSMOS field to measure the number counts and LFs at far-infrared (FIR) and sub-millimetre (sub-mm) wavelengths. The de-blended catalogue was generated by combining the Bayesian source extraction tool XID+ and an informative prior derived from the associated deep multi-wavelength photometric data.

**Results.** Through our de-confusion technique and based on the deep multi-wavelength photometric information, we are able to achieve more accurate measurements while at the same time probing roughly ten times below the *Herschel* confusion limit. Our number counts at 250  $\mu\text{m}$  agree well with previous *Herschel* studies. However, our counts at 350 and 500  $\mu\text{m}$  are below previous *Herschel* results because previous *Herschel* studies suffered from source confusion and blending issues. Our number counts at 450 and 870  $\mu\text{m}$  show excellent agreement with previous determinations derived from single-dish and interferometric observations. Our measurements of the LF at 250  $\mu\text{m}$  and the total IR LF agree well with previous results in the overlapping redshift and luminosity range. The increased dynamic range of our measurements allows us to better measure the faint-end of the LF and measure the dust-obscured SFRD out to  $z \sim 6$ . We find that the fraction of obscured star formation activity is at its highest (>80%) around  $z \sim 1$ . We do not find a shift of balance between  $z \sim 3$  and  $z \sim 4$  in the SFRD from being dominated by unobscured star formation at higher redshift to obscured star formation at lower redshift. However, we do find  $3 < z < 4$  to be an interesting transition period as the portion of the total SFRD that is obscured by dust is significantly lower at higher redshifts.

**Key words.** galaxies: abundances – galaxies: evolution – submillimeter: galaxies – galaxies: luminosity function, mass function – galaxies: statistics – infrared: galaxies

## 1. Introduction

About half of all the luminous power from stars and active galactic nuclei (AGN) that makes up the extra-galactic background was emitted in the far-infrared (FIR) and submillimetre (sub-mm), as a result of re-radiation of dust heated by ultraviolet (UV)/optical photons (Puget et al. 1996; Fixsen et al. 1998; Hauser et al. 1998; Lagache et al. 1999; Hauser & Dwek 2001; Dole et al. 2006). Therefore, a complete understanding of the cosmic star formation history (CSFH) depends critically on taking into account the dust-obscured star formation activity from the local Universe out to the highest redshifts (e.g. Madau & Dickinson 2014). For this purpose, it is of fundamental importance to accurately measure the statistical properties of FIR and sub-mm galaxies and their evolution with cosmic time. Number counts (also known as source counts), which is the number density of galaxies as a function of their intrinsic flux, and the luminosity function (LF), which is the volume density of galaxies as a function of their intrinsic luminosity, are statistical descriptions of the galaxy populations at the most basic level

and can provide strong constraints on models of galaxy formation and evolution (e.g. Granato et al. 2004; Baugh et al. 2005; Fontanot et al. 2007; Hayward et al. 2013; Cowley et al. 2015; Lacey et al. 2016).

Conducting observations at the IR and sub-mm wavelengths is challenging because of high background and the limited angular resolution of the single-dish instruments. In the past three decades, tremendous progress has been made in our understanding of the properties of the IR and sub-mm galaxy population through a succession of breakthrough space-based and ground-based telescopes, starting from the Infrared Astronomical Satellite (IRAS; Neugebauer et al. 1984), the Infrared Space Observatory (ISO; Kessler et al. 1996), the Submillimetre Common-User Bolometer Array (SCUBA; Holland et al. 1999) and the Submillimetre Common-User Bolometer Array-2 (SCUBA-2; Holland et al. 2013) camera on the James Clerk Maxwell Telescope (JCMT), the Spitzer Space Telescope (Werner et al. 2004), the AKARI mission (Murakami et al. 2007), the Balloon-borne Large Aperture Submillimetre Telescope (BLAST; Pascale et al. 2008), the Large

APEX Bolometer Camera (LABOCA; [Siringo et al. 2009](#)) on the Atacama Pathfinder Experiment telescope (APEX), and more recently, the *Herschel* Space Observatory ([Pilbratt et al. 2010](#)). Together, these facilities have enabled the IR LF and its evolution to be successfully traced out to  $z \sim 4$ . In particular, *Herschel* surveys allowed us for the first time to select statistically large samples of galaxies at or close to the rest-frame peak of the FIR emission and gave us a direct measure of the bolometric dust emission across a wide redshift range (for a review, see [Lutz 2014](#)).

Despite the impressive progress owing to great advances in the sensitivity of the instruments through the relatively large beam of single-dish instruments, the deepest FIR and sub-mm observations are severely limited by confusion that results from the blending of multiple sources within the same telescope beam (e.g. [Dole et al. 2003](#)). Confusion presents us with several significant challenges, such as contamination of flux density by neighbouring sources, lack of survey dynamic range, and ambiguity in multi-wavelength association (i.e. counterpart identification) and redshift determination. Follow-up high angular resolution observations of the bright SCUBA-2 850  $\mu\text{m}$  sources and *Herschel*-selected sources with interferometric facilities such as the Atacama Large Millimeter/submillimeter Array (ALMA; [Wootten & Thompson 2009](#)) and the Submillimeter Array (SMA; [Ho et al. 2004](#)) have indeed shown that a significant fraction are made up of multiple sources (e.g. [Karim et al. 2013](#); [Hodge et al. 2013](#); [Simpson et al. 2015](#); [Bussmann et al. 2015](#); [Michałowski et al. 2017](#); [Hill et al. 2018](#); [Stach et al. 2018](#)). The multiplicity rate varies from 15–20% to  $\sim 70\%$  in the literature, depending on factors such as sample selection and the exact definition of multiplicity. As a result, the number counts measured with single-dish instruments and interferometers (e.g. [Karim et al. 2013](#); [Simpson et al. 2015](#)) can differ strongly. However, interferometric follow-up observations from the ground are not possible or extremely difficult at the high frequencies of the *Herschel* surveys because of absorption by water vapour in the atmosphere. In order to probe galaxy populations below the confusion limit, various advanced statistical techniques were developed, such as the stacking method (e.g. [Dole et al. 2006](#); [Marsden et al. 2009](#); [Béthermin et al. 2010, 2012a](#); [Viero et al. 2013](#); [Wang et al. 2016](#)) and the map statistics via the pixel intensity distribution, the so-called P(D) measurements (e.g. [Condon 1974](#); [Patanchon et al. 2009](#); [Glenn et al. 2010](#)). However, a common limitation of these statistical methods is that the properties of individual galaxies cannot be determined, which results in the loss of information about the detailed properties of the underlying galaxy population.

A different approach to overcome confusion noise without giving up measuring properties of individual sources is source extraction using the position prior information from galaxy catalogues that are extracted from imaging surveys with higher angular resolution that are conducted at other wavelengths (e.g. at *Spitzer*/MIPS 24  $\mu\text{m}$  and VLA 1.4 GHz) to distinguish the contribution to the total flux density from various sources within the telescope beam (e.g. [Magnelli et al. 2009](#); [Béthermin et al. 2010](#); [Roseboom et al. 2010, 2012](#); [Wang et al. 2013, 2014](#); [Liu et al. 2018](#)). However, most of these prior-based techniques use a maximum-likelihood optimisation approach, which has two main problems. The first problem is that neither variance nor covariance of source fluxes can be properly estimated. The second problem is overfitting when many of the input sources are intrinsically faint. Another potential problem with prior-based source extraction methods (but not limited to methods that use a

maximum-likelihood optimisation approach) is that sources with significant flux densities at the *Herschel* wavelengths might be missed because they were not listed in the prior catalogues. One way to mitigate this potential problem is to use sufficiently deep prior catalogues that can account for most of the emission in the *Herschel* maps.

In this paper, we build on our success of developing a prior-based Bayesian probabilistic de-blending and source extraction tool called XID+ for confusion-dominated maps ([Hurley et al. 2017](#)) as part of the *Herschel* Extragalactic Legacy Project (HELP; [Vaccari 2016](#), Oliver, in prep.) to study the statistical properties of galaxies over an unprecedented dynamic range in luminosity and redshift. Our method is based on using Bayesian inference techniques such as Markov chain Monte Carlo (MCMC) methods to fully explore the posterior probability distribution and therefore to properly estimate the variance and covariance between sources (i.e. how the source fluxes affect each other). Because XID+ is built upon a Bayesian probabilistic framework, it also provides a natural way in which to introduce additional prior information. Subsequently, we introduced prior information on the flux densities themselves through extensive modelling of the spectral energy distributions (SED) and fitting to multi-wavelength imaging data of the galaxies under study, and we were able to show that this SED prior-enhanced XID+ significantly improves the vanilla XID+, based on validation using high angular resolution data from interferometric observations ([Pearson et al. 2017, 2018](#)).

The structure of the paper is as follows. In Sect. 2 we first briefly describe how the SED prior-enhanced XID+ de-confusion technique works, together with the salient features (such as source density and completeness limit) of our most recent de-blended catalogue in the COSMOS field. We then introduce the various theoretical models and simulations (including empirical models, semi-analytic simulations, and hydrodynamic simulations) that we used later on for comparison purposes. In Sect. 3 we present our measurements of the number counts, the monochromatic and total IR LFs, and the CSFH, using our de-blended source catalogue in COSMOS. We also show detailed comparisons between our results and previous measurements in the literature from single-dish and interferometric observations as well as predictions from a range of theoretical models and simulations of galaxy formation and evolution. Finally, discussions and conclusions are presented in Sect. 4. Throughout the paper, we assume  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ , and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . Flux densities are corrected for Galactic extinction ([Schlegel et al. 1998](#)). Unless otherwise stated, we assume a [Chabrier \(2003\)](#) initial mass function (IMF) in this paper.

## 2. Data

### 2.1. SED prior-enhanced XID+ de-confusion technique

We have invested great effort in developing techniques that can use very deep optical/near-infrared(NIR)/mid-infrared(MIR) prior source catalogues to decompose *Herschel* images that suffer from source confusion because the full power of *Herschel* can only be unleashed when it is combined with detailed knowledge of the physical properties of galaxies. A major breakthrough is our development of the prior-based de-blending and source extraction tool XID+ ([Hurley et al. 2017](#)), which is a Bayesian probabilistic framework in which to include prior information. It uses the Bayesian inference tool Stan ([Stan Development Team 2015a,b](#)) to obtain the full posterior probability distribution

on flux estimates. Compared to the previous de-confusion tool DESPHOT (Wang et al. 2014), we can probe much fainter sources (by a factor of  $\sim 10$ ) at the same flux accuracy of 10%.

In the original version, XID+ uses a flat prior in the flux density parameter space (between zero and the brightest value in a given segment in the map), along with the known source positions on the sky. More recently, we have introduced an informative prior on the flux densities themselves in the vanilla XID+ through extensive modelling of the SED and fitting to multi-wavelength photometric data. Using ALMA continuum data as an independent validation, we have shown that by including an informative but still weak prior on the SED, the performance of XID+ can be improved further (Pearson et al. 2017). The systematic bias in flux accuracy, characterised by the difference between our predicted  $870\mu\text{m}$  flux (based on the de-blended XID+ SPIRE fluxes) and the measured flux by ALMA, is reduced when an informative flux prior is used, at a depth of more than ten times below the SPIRE  $5\sigma$  confusion limit of  $\sim 30\text{ mJy}$  (Nguyen et al. 2010).

Our SED prior-enhanced XID+ is detailed in Hurley et al. (2017) and Pearson et al. (2017, 2018). Here we describe the main steps of our method for completeness:

- First, we use the SED modelling and fitting tool Code Investigating GALaxy Emission (CIGALE; Burgarella et al. 2005; Noll et al. 2009; Serra et al. 2011; Boquien et al. 2019) to generate SEDs and to fit these SEDs to the multi-wavelength imaging data, from UV to IR, of the galaxies under study. This step produces estimates for the flux densities and uncertainties in the *Herschel*-SPIRE wavebands at 250, 350, and  $500\mu\text{m}$ . CIGALE uses an energy-balance approach between the attenuated UV/optical emission and the IR/sub-mm emission, allowing the estimation of the IR/sub-mm flux densities. The choices for the SED model components and parameters for the SPIRE band priors follow Pearson et al. (2018) and are briefly repeated here. We use a delayed exponentially declining star formation history (SFH) with an exponentially declining burst, Bruzual & Charlot (2003) stellar emission, Chabrier (2003) IMF, Charlot & Fall (2000) dust attenuation, the updated Draine et al. (2014) version of the Draine & Li (2007) IR dust emission, and Fritz et al. (2006) AGN models.
- The second step is to incorporate the predicted SPIRE flux densities and uncertainties from CIGALE as informative but still weak flux density priors in the probabilistic de-blending and source extraction tool XID+ to estimate the flux densities in the SPIRE bands. Combined with positional information from the prior galaxy catalogue, XID+ then uses the Bayesian inference tool Stan (Stan Development Team 2015a,b) to obtain the full posterior probability distribution on flux estimates by modelling the confusion-limited SPIRE maps.
- When the de-blended XID+ SPIRE flux densities are extracted, these SPIRE data are added to the multi-wavelength photometric dataset and CIGALE is rerun to derive estimates for physical properties such as the monochromatic luminosity, stellar mass ( $M_*$ ), and star formation rate (SFR) for each galaxy. During this step, we also ask CIGALE to give the flux densities estimates and uncertainties at the desired IR and sub-mm wavelengths (e.g. the ALMA  $870\mu\text{m}$  band) for each object. The same CIGALE SED models are used for the flux estimation in the first run and to obtain the physical parameters in the second run.

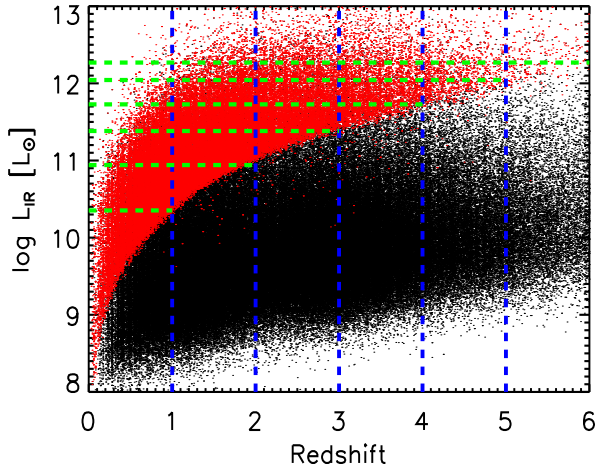
## 2.2. De-blended catalogue in the COSMOS field

In the COSMOS field (Scoville et al. 2007), we used the COSMOS2015 catalogue (Laigle et al. 2016), which contains photometric data in over 30 bands for around 1.2 million objects, as our prior catalogue. We note that the COSMOS2015 catalogue does contain flux densities in the *Herschel*-SPIRE bands for  $\sim 18\,000$  sources. These SPIRE flux densities were extracted with a maximum-likelihood optimisation approach using *Spitzer*/MIPS  $24\mu\text{m}$  sources as priors (Roseboom et al. 2010). In this paper, we do not use these SPIRE flux densities from the COSMOS2015 catalogue. We ran CIGALE to model the SEDs of all galaxies in the COSMOS2015 catalogue and generate flux density priors in the *Herschel*-SPIRE bands at 250, 350, and  $500\mu\text{m}$ . To account for as much dust emission as possible and at the same time keep the level of degeneracy as low as possible, we applied a cut of  $0.7\text{ mJy}$  on the predicted flux density at  $250\mu\text{m}$ , which left us with 205 958 objects over  $2.15$  square degrees (see Pearson et al. 2018 for more details). The 205 958 galaxies with a predicted  $250\mu\text{m}$  flux density above our flux cut were then used in XID+ to model the confusion-limited SPIRE maps and generate de-blended flux densities at 250, 350, and  $500\mu\text{m}$ . Finally, CIGALE was ran again combining the multi-wavelength photometric information and the de-blended SPIRE flux densities to generate estimates and uncertainties on quantities such as the flux density at  $870\mu\text{m}$  (observed-frame), rest-frame monochromatic luminosity at various FIR and sub-mm wavelengths, dust luminosity or the total IR luminosity, stellar mass, and SFR for each galaxy. In CIGALE, dust luminosity is defined as the integrated IR luminosity between 8 and  $1000\mu\text{m}$ , including contributions from both AGN activity and star formation.

Because of our flux cut of  $0.7\text{ mJy}$  at  $250\mu\text{m}$ , we also need to apply an equivalent flux cut at other wavelengths when we compare our number count results with previous measurements in Sect. 3.1. To achieve this, we use the simulated infrared dusty extragalactic sky (SIDES) empirical model, which best matches existing FIR and sub-mm number count measurements (see Sect. 2.3.1). We employ two methods to derive the equivalent flux cuts at other wavelengths. The first method uses the mean flux ratio from the SIDES model to convert the flux cut of  $0.7\text{ mJy}$  at  $250\mu\text{m}$  into an equivalent flux cut of 0.8, 0.7, 0.6, and  $0.3\text{ mJy}$  at 350, 450, 500, and  $870\mu\text{m}$ , respectively. However, because scatter is present in the flux ratios, we also consider a second method, which takes into account the broad correlation between the  $250\mu\text{m}$  flux density ( $S_{250}$ ) and flux densities at other wavelengths. For example, to derive the flux cut at  $350\mu\text{m}$ , we compute the ratio of the number of objects with  $S_{250} > 0.7\text{ mJy}$  and  $S_{350} > x\text{ mJy}$  to the total number of objects with  $S_{350} > x\text{ mJy}$  and then define the flux cut level at  $350\mu\text{m}$  as the value of  $x$  when the ratio is equal to 95%. Using this method, we can derive an equivalent flux cut of 0.7, 0.7, 0.6, and  $0.4\text{ mJy}$  at 350, 450, 500, and  $870\mu\text{m}$ , respectively. In this paper, we use the flux cut values derived from the second method.

Also using the SIDES simulation, we can work out the corresponding limit on the total IR luminosity ( $L_{\text{IR}}$ ) as a function of redshift due to the flux cut of  $0.7\text{ mJy}$  at  $250\mu\text{m}$ . For a given redshift bin with a lower limit ( $z_{\text{lower limit}}$ ) and an upper limit ( $z_{\text{upper limit}}$ ), we compute the ratio of the number of galaxies with  $L_{\text{IR}} > x L_{\odot}$  and  $S_{250} > 0.7\text{ mJy}$  to the total number of galaxies with  $L_{\text{IR}} > x L_{\odot}$  in that redshift bin and then define the IR luminosity limit as the value of  $x$  at which the ratio is equal to 95%. Figure 1 shows the IR luminosity  $L_{\text{IR}}$  versus redshift for mock galaxies from the SIDES simulation.





**Fig. 1.** IR luminosity ( $L_{\text{IR}}$ ) vs. redshift from the SIDES simulation. The black dots are mock galaxies from the SIDES simulation. For clarity, only a random 20% of the simulation is plotted. The red dots are mock galaxies whose  $250\,\mu\text{m}$  flux density lies above  $0.7\,\text{mJy}$ . The vertical dashed blue lines indicate redshift bins that we show for illustration purposes only. We use a variety of redshift binning later on to facilitate the comparisons between our results and previous measurements or predictions from theoretical models. The horizontal dashed green lines indicate the IR luminosity limit above which the sample is 90% complete.

It is worth pointing out that our prior catalogue, that is, the COSMOS2015 catalogue, is limited on stellar mass because it has a magnitude limit in the  $K_s$  band. Section 3.3 in Pearson et al. (2018) details how the stellar mass completeness limit as a function of redshift is derived, which basically follows the procedure in Pozzetti et al. (2010). Using the SIDES simulation, we checked that for our adopted limits on flux (or luminosity), most of our selected galaxies have stellar masses above the stellar mass limit inherent in the COSMOS2015 catalogue. We therefore ignore the effect of the stellar mass limit in our results in Sect. 3.

### 2.3. Empirical models and simulations of galaxy formation physics

Broadly speaking, two different types of models exist: empirical models that are designed to reproduce observations but contain minimum information on galaxy formation physics, and physically motivated models that can be tuned by observations to varying degrees. The latter include the Durham semi-analytic model (SAM), which uses simplified flow equations for bulk components, and the EAGLE numerical hydrodynamic simulation, which solves the equations of gravity, hydrodynamics, and thermodynamics at the same time (see Somerville & Davé 2015 for a review).

#### 2.3.1. Empirical models

We used two different empirical models. The publicly available SIDES simulation<sup>1</sup> is a simulation of the extragalactic sky from the FIR to the sub-mm that includes clustering based on empirical prescriptions. The method used to build this simulated catalogue is described in detail in Béthermin et al. (2017). Briefly, a light cone covering  $2\,\text{deg}^2$  was produced from the Bolshoi-Planck simulation (Klypin et al. 2016; Rodríguez-Puebla et al.

2016). To populate dark matter haloes with galaxies, an abundance-matching technique was used (e.g. Vale & Ostriker 2004). The luminous properties of the galaxies were generated by using an updated version of the two star-formation modes (2SFM) model (Sargent et al. 2012; Béthermin et al. 2012b, 2013), which is based on the observed evolution of the galaxy star-formation main sequence and the observed evolution of the SEDs with redshift. The galaxy star formation main sequence refers to the observed tight correlation (with an intrinsic scatter of  $\sim 0.2\text{--}0.3$  dex) between SFR and stellar mass of star-forming galaxies that exists both in the local Universe and at high redshifts (e.g. Noeske et al. 2007; Elbaz et al. 2007; Daddi et al. 2007; Whitaker et al. 2012; Wang et al. 2013; Speagle et al. 2014; Lee et al. 2015; Schreiber et al. 2015; Tomczak et al. 2016; Pearson et al. 2018). The SIDES simulation reproduces a large set of observables, such as number counts and their evolution with redshift, and cosmic IR background power spectra. The SIDES-simulated light cone contains information such as redshift, halo mass, stellar mass, SFR, SED shape, and flux densities at various wavelengths between  $24$  and  $2000\,\mu\text{m}$  (observed frame).

The empirical galaxy generator (EGG; Schreiber et al. 2017) is a tool for generating mock galaxy catalogues with realistic fluxes and simple morphological types, developed by the ASTRODEEP collaboration. By construction, EGG is designed to match current observations from the UV to the sub-mm at  $0 < z < 7$ . EGG generates mock galaxies that are composed of two broad populations of star-forming galaxies (SFGs) and quiescent galaxies (QGs), based on the observed stellar mass functions of each population. SFRs are assigned (with random scatter) to mock galaxies based on the galaxy star formation main sequence. Other properties such as optical colours, morphologies, and SEDs are assigned using empirical relations derived from *Hubble* and *Herschel* observations from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) fields (Grogin et al. 2011; Koekemoer et al. 2011).

#### 2.3.2. Durham SAMs

In the Durham SAM of galaxy formation, GALFORM, galaxies populate dark matter halo merger trees according to simplified prescriptions for the baryonic physics involved (gas cooling, star formation, feedback, etc.), which result in a set of coupled differential equations that track the exchange of mass and metals between the different baryonic components (stars, cold gas, etc.) of a galaxy. Here we used the version of GALFORM presented in Lacey et al. (2016), with a minor recalibration because the model is implemented in an updated *Planck* cosmology (Baugh et al. 2019, see also Cowley et al. 2018a). This model is calibrated to reproduce a large set of observational data at ( $z \lesssim 6$ ), including sub-mm galaxy number counts such as those presented in this work. To predict sub-mm fluxes, the star formation histories and galaxy properties predicted by GALFORM are coupled with the spectrophotometric code GRASIL (Silva et al. 1998), which computes the absorption and re-emission of stellar radiation by interstellar dust, resulting in self-consistent UV-to-mm SEDs for each simulated galaxy.

One of the main features of the model relevant for the predictions shown here is that it assumes a top-heavy IMF for bursts of star formation triggered by a dynamical process, either disc instabilities, major mergers, or some gas-rich minor mergers, although sub-mm bright galaxies in this model are mainly triggered by disc instabilities. This feature was first introduced into GALFORM by Baugh et al. (2005) so that the model could

<sup>1</sup> <http://cesam.lam.fr/sides/>

simultaneously reproduce observational constraints such as the  $850\mu\text{m}$  galaxy number counts and optical/NIR LFs at  $z = 0$ . A top-heavy IMF is extremely efficient at boosting sub-mm flux due to (i) the increased UV radiation through having more massive stars per unit star formation and (ii) an increased dust mass available to absorb and re-emit this UV radiation through faster recycling of material into the interstellar medium as these massive stars become supernovae. The combination of these two effects allows the model to reproduce the galaxy number counts shown here whilst simultaneously reproducing many other observational datasets (see e.g. Lacey et al. 2016).

### 2.3.3. EAGLE hydrodynamic simulations

The EAGLE simulation suite (described fully in Schaye et al. 2015; Crain et al. 2015) comprises cosmological hydrodynamical simulations of periodic cubic volumes with a range of sizes and numerical resolutions, using a modified version of the Gadget-3 TreeSPH code (an update to Gadget-2, Springel et al. 2005) and a  $\Lambda$ CDM cosmology; the cosmological parameters are those derived in the initial *Planck* release (Planck Collaboration XVI 2014). We used the Ref-100 run, which simulates the evolution of a volume with mean cosmic and a side length of 100 Mpc using the fiducial EAGLE model at a standard resolution. Models for unresolved physical process were implemented to treat star formation and stellar evolution, photoheating and radiative cooling of gas, energetic feedback by supernovae and AGN, and the chemical enrichment of the interstellar medium (ISM) by stars.

Virtual observables are generated for EAGLE galaxies in post-processing using the Monte Carlo dust radiative transfer code SKIRT (Baes et al. 2011; Camps & Baes 2015). The SKIRT modelling approach is briefly summarised here; a full description is given in Camps et al. (2016) and Trayford et al. (2017). Emission from stellar populations older than 10 Myr is represented by the Bruzual & Charlot (2003) spectral libraries. Because dust is not included explicitly in the EAGLE simulations, the diffuse dust distribution is taken to trace that of the sufficiently cool, enriched ISM gas that emerges in EAGLE galaxies, assuming 30% the metal mass is in dust grains and a Zubko et al. (2004) model for grain properties. To include dust associated with the unresolved birth clouds of stars, emission from populations younger than 10 Myr is represented using the HII region spectral libraries of Groves et al. (2008). FIR emission is then produced by the iterative absorption and re-emission of UV-FIR radiation by dust as well as directly from the HII region SEDs, and measured in multiple broad bands in both rest- and observer-frames. The fraction of metals in dust grains, photo-dissociation region covering fraction in HII regions, and temperature threshold for the ISM is set in order to reproduce FIR properties of the local galaxies in the *Herschel* Reference Survey (Boselli et al. 2010; Cortese et al. 2012) for a *K*-band matched sample (Camps et al. 2016). The virtual observations measured at a number of discrete redshifts for galaxies with more than 250 are all publicly accessible at the EAGLE database (McAlpine et al. 2016; Camps et al. 2018), and  $L_{\text{IR}}$  values are estimated by integrating the available broad bands over the 8–1000  $\mu\text{m}$  range.

In contrast to GALFORM, EAGLE assumes a universal Chabrier (2003) IMF. Moreover, while the simulated volume of EAGLE is large for a hydrodynamic simulation of its resolution, the volume is small relative to that achievable for SAMs and empirical models. The Lacey et al. (2016) GALFORM model has a  $\sim 125$  times larger volume, i.e.  $500\text{ h}^{-1}\text{ Mpc}$  on a side. As

a result, more massive systems (i.e. high-mass groups and clusters) are not captured by the simulation, and thus the brightest sources in the FIR are potentially absent in projected counts. It is unclear whether the EAGLE model would produce the correct density of extreme starbursts in a larger volume simulations, or whether it would be necessary to appeal to something like IMF variations to reproduce the observed counts.

## 3. Results

In this section, we present our results of the number counts at various FIR and sub-mm wavelengths (250, 350, 450, 500, and  $870\mu\text{m}$ ), the monochromatic LFs at  $250\mu\text{m}$  (rest-frame) and the total IR LFs as a function of redshift, and the CSFH out to  $z \sim 6$ . We also compare our results with previous measurements in the literature and with predictions from empirical models and physically motivated simulations. We provide our number counts and LFs in a tabular format in the Appendix.

### 3.1. Number counts

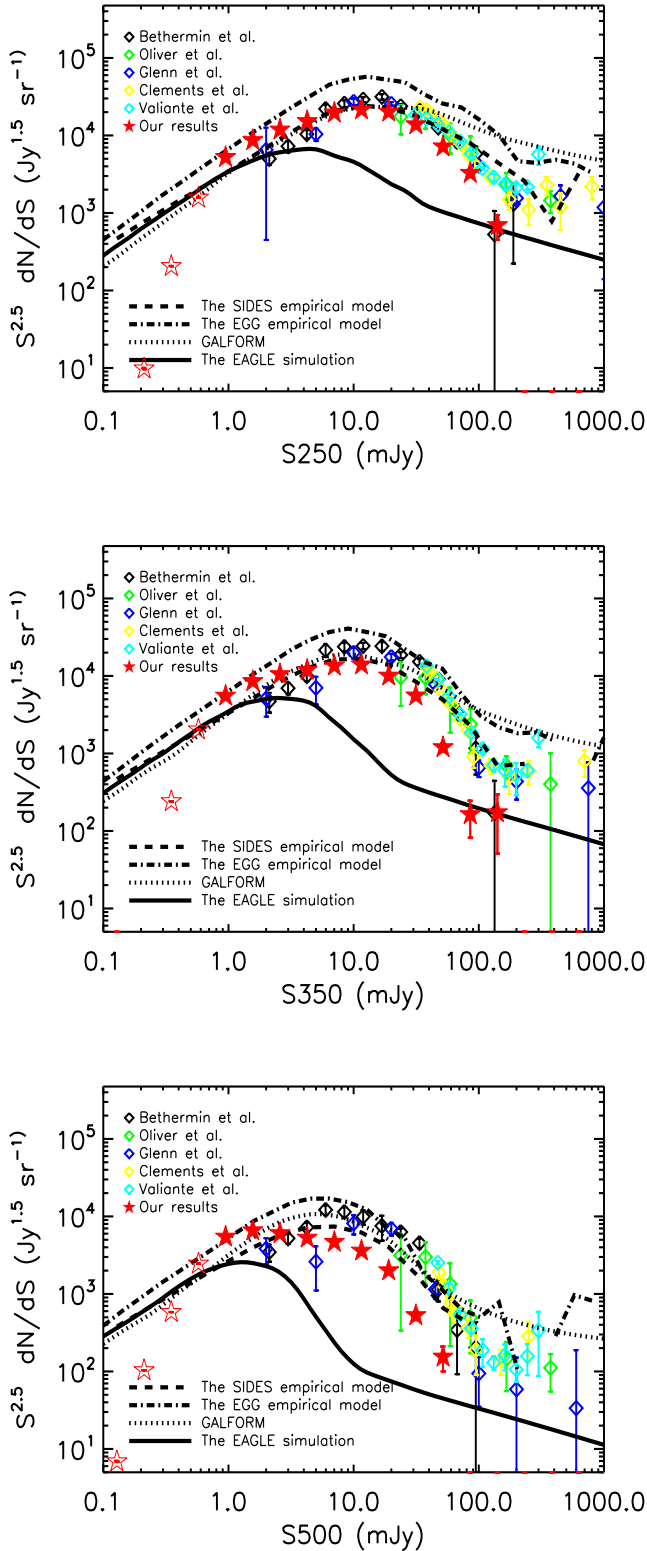
#### 3.1.1. 250, 350, and $500\mu\text{m}$ number counts

In this subsection, we present our measurements of the *Herschel*-SPIRE 250, 350, and  $500\mu\text{m}$  differential number counts using the SED prior-enhanced XID+ de-blended catalogue in the COSMOS field.

Figure 2 compares our counts with previous measurements based on *Herschel* observations and predictions from various models and simulations (including the empirical models SIDES and EGG, GALFORM, and the EAGLE hydrodynamic simulation). These observations include the *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2012) and the *Herschel* Astrophysical Terahertz Large Area Survey (H-ATLAS; Eales et al. 2010). The number counts are multiplied by a factor of  $S^{2.5}$  to reduce the dynamic range of the plot and to highlight the plateau at high flux densities where the Euclidian approximation is valid. Note that as discussed in Sect. 2.2, our flux cut is based on the predicted  $250\mu\text{m}$  flux and therefore should be treated as a guide because it is not likely to be exact. However, our number counts below the flux cut begin to drop rapidly, which indicates that our flux cut is a reasonable representation of the completeness limit. Our error bars only include Poisson errors (and so are likely to be underestimated), while other studies have included uncertainties due to field-to-field variations.

The top panel in Fig. 2 shows that at  $250\mu\text{m}$ , our measurements agree very well with previous observational results that were derived using various techniques such as blind source extraction, stacking, and a P(D) analysis using pixel flux distribution. The lack of very bright sources with  $S_{250} > \sim 100\text{ mJy}$  in our counts is due to the limited size of the COSMOS field. Our counts are more than ten times deeper than the counts derived from blind-source extraction, which probes the counts above the confusion limit (Oliver et al. 2010; Clements et al. 2010; Valiante et al. 2016) and roughly twice deeper than counts derived from P(D) or stacking (Glenn et al. 2010; Béthermin et al. 2012a), which can also probe the counts below the confusion limit.

At the longer wavelengths of 350 and  $500\mu\text{m}$ , the agreement between our number counts and previous measurements decreases, which is understandable because the effects of confusion and source blending become progressively stronger. In general, our counts are lower than previous *Herschel* measurements



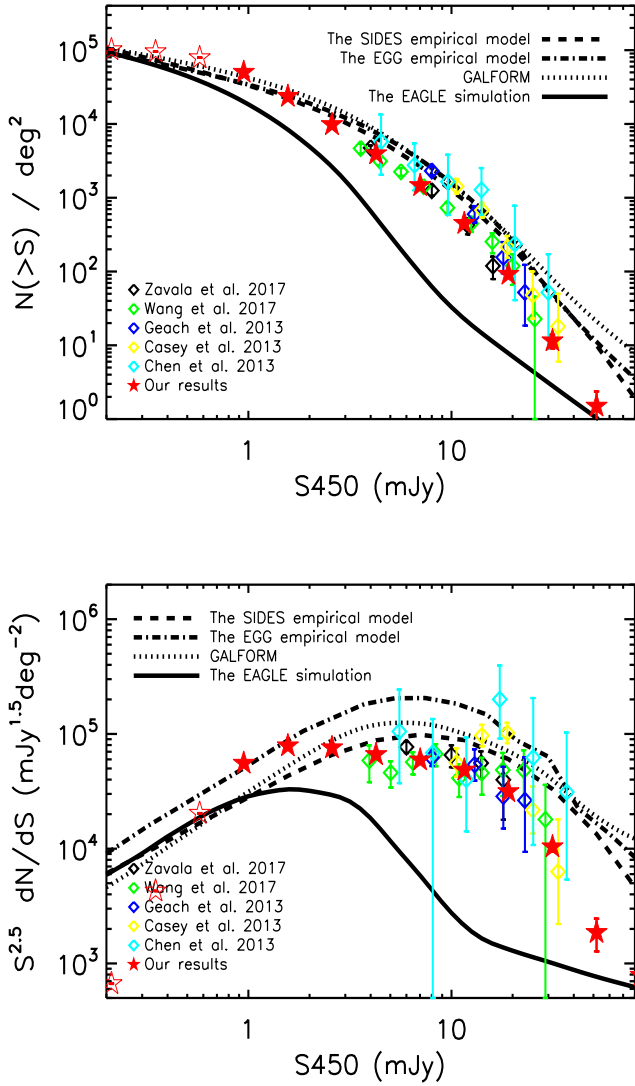
**Fig. 2.** Differential number counts at 250, 350, and 500  $\mu\text{m}$ . Our number counts derived from the SED prior-enhanced XID+ de-blended catalogue in COSMOS are shown as red stars (filled red stars: our number counts above the flux cut; empty red stars: our number counts below the flux cut). Error bars on the red stars only represent Poisson errors. The lines show the predicted number counts from various models and simulations (including SIDES, EGG, GALFORM, and the EAGLE hydrodynamic simulation). The empty diamonds are previous measurements based on *Herschel* observations.

at  $S_{350} > \sim 8$  and  $S_{500} > \sim 5$  mJy. At 500  $\mu\text{m}$ , our counts are lower by as much as  $\sim 0.5$  dex. In addition, our counts indicate that the turnover in the number counts occurs at fainter flux levels and the shift is more pronounced at 500  $\mu\text{m}$  than at 350  $\mu\text{m}$ . This could suggest that either previous *Herschel* count measurements still suffer confusion and source blending problems, which is progressively worse at longer wavelength because of the larger beam size, or our results could be too strongly de-blended. As we discuss in Sects. 3.1.2 and 3.1.3, the excellent agreement between our number counts at 450  $\mu\text{m}$  and the SCUBA-2 450  $\mu\text{m}$  counts, and between our number counts at 870  $\mu\text{m}$  and the SCUBA-2 850  $\mu\text{m}$ , SMA 860  $\mu\text{m}$ , and ALMA 870  $\mu\text{m}$  number counts (derived from single-dish observations with much higher angular resolution and interferometric observations with arcsec or even sub-arcsec resolution) strongly supports the former interpretation.

Except for the EAGLE hydrodynamic simulation, all other models give roughly similar predictions of the number counts at the SPIRE wavelengths and agree well with measurements from previous *Herschel* studies. The SIDES empirical model best reproduces the number counts from previous *Herschel*-based studies. The EGG empirical model tends to over-predict (by a factor of  $\sim 2$ ) the number density of galaxies at all fluxes compared to the SIDES empirical model and GALFORM. GALFORM, although very closely reproducing the predictions from SIDES, does over-predict the number of bright galaxies, and this effect is increasingly pronounced towards shorter wavelength. Again, our measurements are lower (increasingly so towards longer wavelength) than predictions from the three models (SIDES, EGG, and GALFORM). This is not surprising because these models are more or less designed to reproduce the previous *Herschel* measurements. SIDES and EGG are purely empirical models, and therefore by construction, they will agree well with the input observational constraints. However, the downside of these empirical models is that they provide little physical insight about the galaxy formation physics (e.g. physical processes driving the formation and evolution of dusty star-forming galaxies).

In comparison, the predicted number counts from the EAGLE hydrodynamic simulation are much lower (increasingly so towards longer wavelength) than all observational measurements and predictions from the empirical models and GALFORM at all wavelengths. The reason is that unlike GALFORM, the statistical properties of the dusty star formation galaxies are not used to tune the EAGLE simulation. In general, physically-motivated models of galaxy formation struggle to reconcile the statistical properties (such as the number counts and LFs) of sub-mm galaxies without appealing to something like a top-heavy IMF in starburst galaxies (e.g. [Baugh et al. 2005](#); [Lacey et al. 2016](#)). However, alternative solutions to solve the problem of matching the observed sub-mm number counts, such as changes to prescriptions for star formation and feedback, have also been suggested in the literature (e.g. [Hayward et al. 2013](#); [Safarzadeh et al. 2017](#)). In addition, as described in Sect. 2.3.3, the limited volume of EAGLE (which is 125 times smaller than that of GALFORM) means that it misses rarer objects such as luminous starbursts, which significantly contribute to the number counts at the bright end. In order to assess how much the volume problem can affect the number counts, we checked a smaller simulation box with a side length of 50 Mpc. We conclude that the relatively small volume of EAGLE does affect the counts at the bright end, but it is unlikely to account for the large discrepancy with the observations.





**Fig. 3.** *Upper panel:* cumulative (or the integral) number counts at  $450\mu\text{m}$ . Our results derived from the de-blended catalogue in the COSMOS field are plotted as red stars (filled red stars: our number counts above the flux cut; empty red stars: our number counts below the flux cut). Error bars on the red stars only represent Poisson errors. The lines show the predicted number counts from various models and simulations (including SIDES, EGG, GALFORM, and the EAGLE hydrodynamic simulation). The empty diamonds are previous results based on JCMT SCUBA-2 observations. *Lower panel:* differential number counts at  $450\mu\text{m}$ .

### 3.1.2. $450\mu\text{m}$ number counts

In this subsection, we present our measurements of the  $450\mu\text{m}$  number counts using the SED prior-enhanced XID+ de-blended catalogue in the COSMOS field. Our  $450\mu\text{m}$  number counts were generated using the de-blended  $500\mu\text{m}$  flux densities after applying a scaling factor of the flux ratio  $S_{450}/S_{500} = 0.86$ , derived from the SIDES simulation. The standard deviation of the flux ratio  $S_{450}/S_{500}$  is very small ( $\sim 0.079$ ) and therefore is ignored here.

Figure 3 compares our counts with previous measurements using observations carried out with the SCUBA-2 camera on the 15 m JCMT. In comparison, the primary mirror of the *Herschel* satellite is 3.5 m in diameter. Chen et al. (2013) presented SCUBA-2  $450\mu\text{m}$  observations in the field of the massive

lensing cluster A370 (total survey area  $>100\text{arcmin}^2$ ) and 20 detected sources with a signal-to-noise ratio  $S/N > 4$ . The intrinsic number counts derived in Chen et al. (2013), which probe flux levels below the *Herschel* confusion limit, are plotted as light blue diamonds in Fig. 3. The first deep blank-field cosmological  $450\mu\text{m}$  imaging covering an area of  $140\text{arcmin}^2$  of the COSMOS field was conducted as part of the SCUBA-2 Cosmology Legacy Survey (S2CLS) and was presented in Geach et al. (2013). Consequently, Geach et al. (2013) made the first number counts at  $450\mu\text{m}$  from an unbiased blank-field survey at a flux density limit  $S_{450} > 5\text{mJy}$ ; they are plotted as dark blue diamonds in Fig. 3. Later on, Casey et al. (2013) studied the number counts at  $450\mu\text{m}$  using 78 sources detected from a wider and shallower  $394\text{arcmin}^2$  area in COSMOS observed with SCUBA-2 with more uniform coverage. Their counts are plotted as yellow diamonds. More recently, Zavala et al. (2017), using deep observations in the Extended Groth Strip (EGS) field taken as part of the S2CLS, detected 57 sources at  $450\mu\text{m}$ . They presented one of the deepest number counts available so far, derived using directly extracted sources from only blank-field observations, which are plotted as black diamonds. Wang et al. (2017) used a new program on the JCMT, the SCUBA-2 Ultra Deep Imaging EAO (East Asian Observatory) Survey (STUDIES), and detected  $\sim 100$   $450\mu\text{m}$  sources in the COSMOS field. They presented a determination of the number counts down to a flux limit of  $3.5\text{mJy}$  (green diamonds in Fig. 3).

Figure 3 shows that previous observational measurements of the  $450\mu\text{m}$  counts from blank-field and lensing-cluster surveys carried out with SCUBA-2 on the JCMT are more or less consistent with each other within the errors, especially when the counts derived from the lensing cluster observations by Chen et al. (2013) with very large uncertainties are excluded. These SCUBA-2 number counts results are all significantly lower than the *Herschel* counts at  $350$  and  $500\mu\text{m}$ . As pointed out in previous studies such as Wang et al. (2017), the much higher *Herschel* counts are mostly due to confusion and source blending, which is more severe when sources below the beam are strongly clustered. At  $450\mu\text{m}$ , the angular resolution achievable with JCMT is roughly a factor of 5 better than that of *Herschel* at  $500\mu\text{m}$ . Because of its much higher angular resolution and much fainter confusion limit (about seven times fainter), SCUBA-2 counts at  $450\mu\text{m}$  do not suffer as strongly from confusion and source blending issues.

Our measurements of the  $450\mu\text{m}$  number counts, which are lower than previous *Herschel* results (as shown in Sect. 3.1.1), are now in excellent agreement with the number counts derived from the higher resolution SCUBA-2 observations. We also extend the number count measurements by a factor of  $\sim 4$  compared to the deepest SCUBA-2 study by Wang et al. (2017). Previous SCUBA-2 measurements did not find any evidence of a turnover in the number counts at the faint end, while *Herschel* studies suggest a turnover at around  $5\text{mJy}$  at  $500\mu\text{m}$  (or around  $4\text{mJy}$  at  $450\mu\text{m}$ , using the colour ratio of  $S_{450}/S_{500} = 0.86$ ). Based on the increased dynamic range, our number counts derived using the SED prior-enhanced XID+ de-blended catalogue in COSMOS suggest a turnover in the counts at around  $2\text{mJy}$ .

The empirical models SIDES, EGG, and GALFORM generally over-predict the number counts compared to the SCUBA-2 measurements and our measurement. This is not surprising because the models are more or less tuned to reproduce the *Herschel*/SPIRE number counts results, as discussed in Sect. 3.1.1. Again, the predicted number counts from the EAGLE hydrodynamic simulation (which have not been tuned



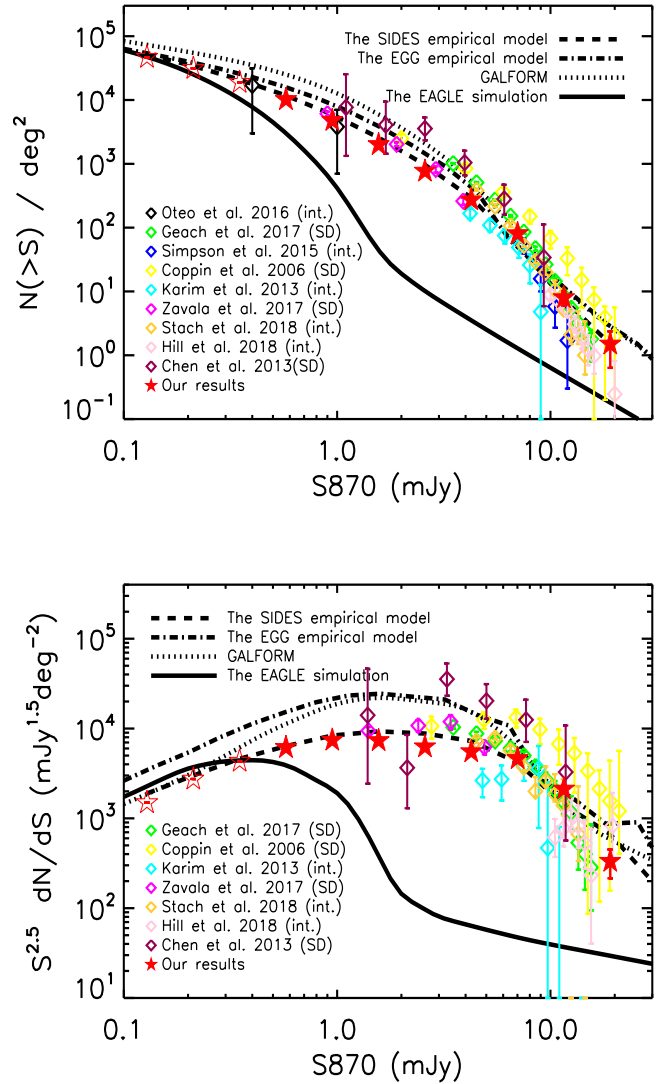
to match the statistical properties of dusty star-forming galaxies) are much lower than the observations and predictions from the empirical models and GALFORM.

### 3.1.3. $870\mu\text{m}$ number counts

In this subsection, we present our measurements of the  $870\mu\text{m}$  number counts using the SED prior-enhanced XID+ de-blended catalogue in COSMOS. As described in Sects. 2.1 and 2.2, in the second run of CIGALE, which combined the multi-wavelength photometric information with the de-blended XID+ SPIRE flux densities, we also generated the predicted flux densities and uncertainties at  $870\mu\text{m}$ . Therefore, we can compare our predicted  $870\mu\text{m}$  number counts with previous measurements using either single-dish (e.g. SCUBA-2 on the JCMT) or interferometric observations (e.g. ALMA and SMA). In general, single-dish observations cover a much larger area than interferometric observations, but angular resolution and depth are limited. In this paper, we ignore the small differences that are due to the slightly different effective wavelengths of the different instruments.

Figure 4 compares our results with previous measurements using single-dish observations from SCUBA-2 and interferometric observations from SMA and ALMA. The first estimates of the  $850\mu\text{m}$  number counts were presented in Coppin et al. (2006) using  $>100$  detected sources from the SCUBA HALF Degree Extragalactic Survey (SHADES; Mortier et al. 2005; van Kampen et al. 2005) over an area of  $720\text{ arcmin}^2$ . Using SCUBA-2 observations of a field ( $>100\text{ arcmin}^2$ ) lensed by the massive cluster A370, Chen et al. (2013) detected 26 sources at  $850\mu\text{m}$  with an  $S/N > 4$ . Through the effect of gravitational lensing, Chen et al. (2013) were able to probe fainter galaxies than the sources detected in Coppin et al. (2006). More recent single-dish observations have been reported by Geach et al. (2017) and Zavala et al. (2017). Geach et al. (2017) detected  $\sim 3000$  sub-mm sources at  $S/N > 3.5$  at  $850\mu\text{m}$  over  $\sim 5\text{ deg}^2$  that were surveyed as part of the S2CLS. This is the largest survey of its kind at this wavelength, which increases the sample size selected at  $850\mu\text{m}$  by an order of magnitude. As a result, Geach et al. (2017) were able to measure the number counts at  $850\mu\text{m}$  with unprecedented accuracy. In particular, the large area of the survey enabled a better determination of the counts at the bright end. Zavala et al. (2017) used deep observations with SCUBA-2 in the EGS as part of the S2CLS to detect 90 sources at  $850\mu\text{m}$  with an  $S/N > 3.5$  over  $70\text{ arcmin}^2$  and derived the deepest number counts from blank-field single-dish observations at  $S_{850} > 0.9\text{ mJy}$ .

Karim et al. (2013) reported the first determination of the number counts at  $870\mu\text{m}$  based on arcsecond-resolution observations with ALMA for a sample of 122 sub-mm sources selected from the LABOCA Extended Chandra Deep Field South Submillimetre Survey (LESS; Weiß et al. 2009). They found that the ALMA-derived number counts broadly agree with previous determinations from single-dish observations. Following Karim et al. (2013), Simpson et al. (2015) presented high-resolution  $870\mu\text{m}$  ALMA observations of a representative sample of the brightest 30 sub-mm sources in the UKIDSS UDS field, which are selected from the S2CLS. Fifty-two sub-mm galaxies were at an  $S/N > 4$ . They found that the level of multiplicity present in their observations boosts the number counts from single-dish observations by 20% at  $S_{870} > 7.5\text{ mJy}$  and by 60% at  $S_{870} > 12\text{ mJy}$ . Oteo et al. (2016) exploited sub-arcsec resolution ALMA calibration observations in a variety of frequency bands and array configurations and were able to reach lower noise levels ( $25\mu\text{Jy beam}^{-1}$ ) to detect faint dusty star-forming galaxies. Oteo et al. (2016)



**Fig. 4.** *Upper panel:* cumulative number counts at  $870\mu\text{m}$ . Our results derived from the de-blended catalogue in COSMOS are plotted as red stars (filled red stars: our number counts above the flux cut; empty red stars: our number counts below the flux cut). Error bars on the red stars only represent Poisson errors. The lines show the predicted number counts from various models and simulations (including SIDES, EGG, GALFORM, and the EAGLE hydrodynamic simulation). The empty diamonds are previous results using either single-dish (SD) observations or interferometric (int.) observations. The SCUBA-2 number counts are shown as a function of  $850\mu\text{m}$  flux density and the SMA number counts are plotted as a function of  $860\mu\text{m}$  flux density, but we ignored the slightly different effective wavelengths of the different instruments. *Lower panel:* differential number counts at  $870\mu\text{m}$ .

presented cumulative number counts at  $870\mu\text{m}$  based on 11 sub-mm sources detected in ALMA band 7 at an  $S/N > 5$ . Following Simpson et al. (2015), Stach et al. (2018) reported the first results of the recently completed ALMA  $870\mu\text{m}$  continuum survey of a complete sample of over 700 sources from the UKIDSS/UDS field ( $50\text{ arcmin}^2$ ). They were able to derive the number counts at  $S_{870} > 4\text{ mJy}$  and confirmed that the number counts derived from single-dish SCUBA-2 observations are about 28% too high in comparison. Hill et al. (2018) observed the brightest sources (down to  $S_{850} \sim 8\text{ mJy}$ ) in the S2CLS with SMA at  $860\mu\text{m}$  at an average synthesised beam of  $2.4\text{ arcsec}$ . Their number counts are consistent with previous single-dish results, but the cumulative counts are systematically lower by  $\sim 14\%$ .

Except for the earliest measurement from Coppin et al. (2006) and the measurement from the lensing cluster field (Chen et al. 2013), all other estimates agree more or less well with each other within the errors. Our predicted 870  $\mu\text{m}$  number counts and previous measurements based on SCUBA-2 850  $\mu\text{m}$ , SMA 860  $\mu\text{m}$ , and ALMA 870  $\mu\text{m}$  observations also agree excellently well. Based on our de-confusion technique and on the wealth of deep multi-wavelength photometric information in COSMOS, we were able to extend the 870  $\mu\text{m}$  count measurements down to fainter flux levels (by a factor of  $\sim 2$ ) than in the deepest observations carried out by SCUBA-2.

The SIDES empirical model agrees best with the observational measurements. In addition, there is an excellent agreement between our 870  $\mu\text{m}$  number counts and the predicted counts from SIDES throughout the entire dynamic range where our measurements are available. Both the EGG empirical model and GALFORM over-predicted the number counts, especially in the flux range between  $\sim 0.3$  and  $\sim 6$  mJy. Again, the predicted number counts from the EAGLE hydrodynamic simulation are much lower than the observations and predictions from the empirical models and GALFORM. It is also interesting to see that the under-prediction of EAGLE number counts compared to the observed counts increases towards longer wavelengths.

### 3.2. Luminosity functions and their evolution

In this subsection, we first present our results on the monochromatic rest-frame 250  $\mu\text{m}$  LF and then the total IR LF (integrated from 8 to 1000  $\mu\text{m}$ ) in various redshift bins. We also compare our results with previous measurements and with predictions from the Durham SAM and the EAGLE hydrodynamic simulations.

#### 3.2.1. Monochromatic rest-frame 250 $\mu\text{m}$ LF

In Fig. 5 we compare our monochromatic rest-frame 250  $\mu\text{m}$  LF in four redshift bins from  $z \sim 0.5$  to  $z \sim 4.5$  with those in Koprowski et al. (2017). Koprowski et al. (2017) used SCUBA-2 850  $\mu\text{m}$  observations in the COSMOS and UKIDSS-UDS fields from the S2CLS together with ALMA 1.3 mm imaging data of the HUDF (Dunlop et al. 2017) to determine the rest-frame 250  $\mu\text{m}$  LFs out to redshift  $z \sim 5$ . Because the mean redshift of the population of their 850  $\mu\text{m}$  detected sources is around 2.5 (probing a rest-frame around 250  $\mu\text{m}$  at the mean redshift), the average sub-mm galaxy template from Michałowski et al. (2010) was adopted to convert the observed-frame 850  $\mu\text{m}$  into the rest-frame 250  $\mu\text{m}$  flux density. Koprowski et al. (2017) also presented the best-fitting Schechter functions, which are parametrised as

$$\phi(L, z) = \phi_* \left( \frac{L}{L_*} \right)^\alpha \exp \left( -\frac{L}{L_*} \right), \quad (1)$$

where  $\phi_*$  is the normalisation parameter,  $\alpha$  is the faint-end slope, and  $L_*$  is the characteristic luminosity. Figure 5 shows that our rest-frame 250  $\mu\text{m}$  LF in the four redshift bins agrees well with the measurements from Koprowski et al. (2017) in the overlapping luminosity range, but our measurements also extend to much fainter luminosities (roughly ten times fainter). In the redshift bin  $1.5 < z < 2.5$ , the dynamic range in luminosity probed by our study is the same as that probed by Koprowski et al. (2017) because the two faintest points in Koprowski et al. (2017) in the  $1.5 < z < 2.5$  bin were derived using the ALMA 1.3 mm data. Koprowski et al. (2017) derived the faint-end slope  $\alpha$  to be  $\alpha = -0.4$  in the redshift bin  $1.5 < z < 2.5$  and kept it fixed at

this value for the remaining three redshift bins. While our rest-frame 250  $\mu\text{m}$  LF measurements agree well with the best-fitting Schechter functions presented in Koprowski et al. (2017) in the two lowest redshift bins, our measurements indicate higher volume densities towards the faint end at higher redshifts.

Koprowski et al. (2017) found that their total IR LF measurements based on SCUBA-2 observations have a much smaller number of bright sources at all redshifts than the *Herschel*-based studies of Magnelli et al. (2013) and Gruppioni et al. (2013). However, the Koprowski et al. (2017) study used a single SED (i.e. the average sub-mm galaxy SED template from Michałowski et al. 2010) in order to convert the observed 850  $\mu\text{m}$  flux density into a total IR flux (integrated between 8 and 1000  $\mu\text{m}$ ). Because we agree reasonably well with the monochromatic rest-frame 250  $\mu\text{m}$  LF from Koprowski et al. (2017) and also with the total IR LF from Magnelli et al. (2013) and Gruppioni et al. (2013), as we show in Sect. 3.2.2, we conclude that the likely cause for the disagreement between the SCUBA-2 based study and the *Herschel*-based studies is the use of a single SED shape instead of the full wide range of SEDs present in the dusty star-forming galaxy population. Gruppioni & Pozzi (2019) conducted a thorough investigation into the large discrepancies seen in the total IR LF from Koprowski et al. (2017) and the *Herschel*-based measurements of Magnelli et al. (2013) and Gruppioni et al. (2013). They concluded that the discrepancy is mainly caused by the use of a single template in Koprowski et al. (2017) and sample incompleteness because SCUBA-2 surveys are biased against galaxies with “warm” SED shapes.

#### 3.2.2. Total IR LF

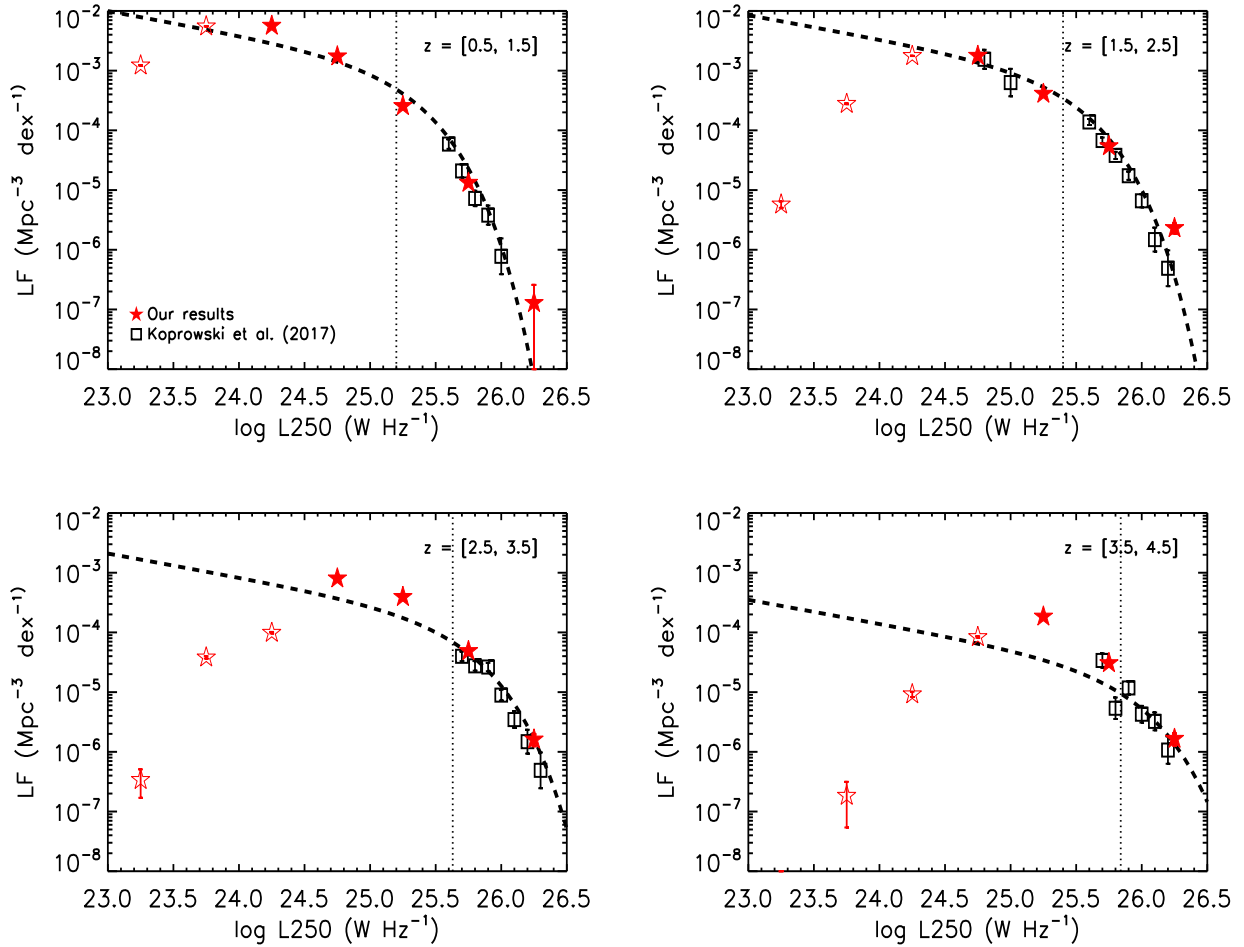
In Fig. 6 we compare our total IR LFs with those in Magnelli et al. (2013) derived from the deepest *Herschel* PACS surveys at 70, 100, and 160  $\mu\text{m}$  in the GOODS fields obtained by the PACS Evolutionary Probe (PEP; Lutz et al. 2011) and GOODS-*Herschel* (GOODS-H; Elbaz et al. 2011). Magnelli et al. (2013) used the positional information of *Spitzer*/IRAC 3.6  $\mu\text{m}$  sources to extract sources from the *Spitzer*/MIPS 24  $\mu\text{m}$  map, which were in turn used as positional priors for source extraction in the PACS maps. To obtain the required photometric redshift information, they cross-matched the IRAC-MIPS-PACS source catalogues with the shorter wavelength GOODS catalogues (optical + NIR). LFs were then constructed using the  $1/V_{\text{max}}$  method, but limited to redshifts  $z < 2.3$  because the PACS data do not extend beyond 160  $\mu\text{m}$ . Magnelli et al. (2013) fit the IR LFs with a double power-law function that was parameterised as follows:

$$\phi = \phi_{\text{knee}} L^{-0.6}, \text{ when } \log(L/L_\odot) < L_{\text{knee}}, \quad (2)$$

and

$$\phi = \phi_{\text{knee}} L^{-2.2}, \text{ when } \log(L/L_\odot) > L_{\text{knee}}. \quad (3)$$

The free parameters are the normalisation  $\phi_{\text{knee}}$  and the transition luminosity  $L_{\text{knee}}$ . The fixed power-law slopes are taken from Sanders et al. (2003), who studied the IR LF of IR-bright galaxies selected from the IRAS all-sky survey in the local Universe. Our measurements of the total IR LFs agree quite well with the measurements from Magnelli et al. (2013) and their best-fitting double power-law functions out to redshifts  $z < 2.3$ . In the highest redshift bin  $1.8 < z < 2.3$ , our measurements extend to fainter luminosities by almost one dex than in Magnelli et al. (2013).



**Fig. 5.** Rest-frame 250  $\mu\text{m}$  LF. The red stars are derived from our de-blended catalogue in COSMOS (filled red stars: our LF above the completeness limit; empty red stars: our LF below the completeness limit). Error bars on the red stars only represent Poisson errors. The black empty squares are taken from [Koprowski et al. \(2017\)](#), based on SCUBA-2 850  $\mu\text{m}$  observations of the COSMOS and UKIDSS-UDS fields as part of the S2CLS. The two faintest points in the [Koprowski et al. \(2017\)](#) measurements in the  $1.5 < z < 2.5$  redshift bin (which has the largest dynamic range) were derived using the ALMA 1.3 mm data. The dashed line is the best-fit Schechter function adopted in [Koprowski et al. \(2017\)](#). The faint-end slope of the Schechter function was found to be  $\alpha = -0.4$  in the  $1.5 < z < 2.5$  redshift bin in [Koprowski et al. \(2017\)](#) and was kept fixed in the remaining three redshift bins. The vertical dotted line indicates the location of the characteristic luminosity, i.e.  $L_*$  in Eq. (1), as derived in [Koprowski et al. \(2017\)](#).

In Fig. 7 we compare our LFs with the results from [Gruppioni et al. \(2013\)](#). [Gruppioni et al. \(2013\)](#) used the datasets (at 70, 100, and 160  $\mu\text{m}$ ) from the *Herschel* PEP Survey, in combination with the HerMES imaging data at 250, 350, and 500  $\mu\text{m}$ , to derive the evolution of the rest-frame 35, 60, and 90  $\mu\text{m}$  and the total IR LFs up to  $z \sim 4$ . The inclusion of the SPIRE imaging data allowed [Gruppioni et al. \(2013\)](#) to determine IR luminosities without large uncertainties due to extrapolations. [Gruppioni et al. \(2013\)](#) used a modified-Schechter function (e.g. [Saunders et al. 2009](#); [Wang & Rowan-Robinson 2010](#); [Marchetti et al. 2016](#); [Wang et al. 2016](#)) to fit the total IR LF,

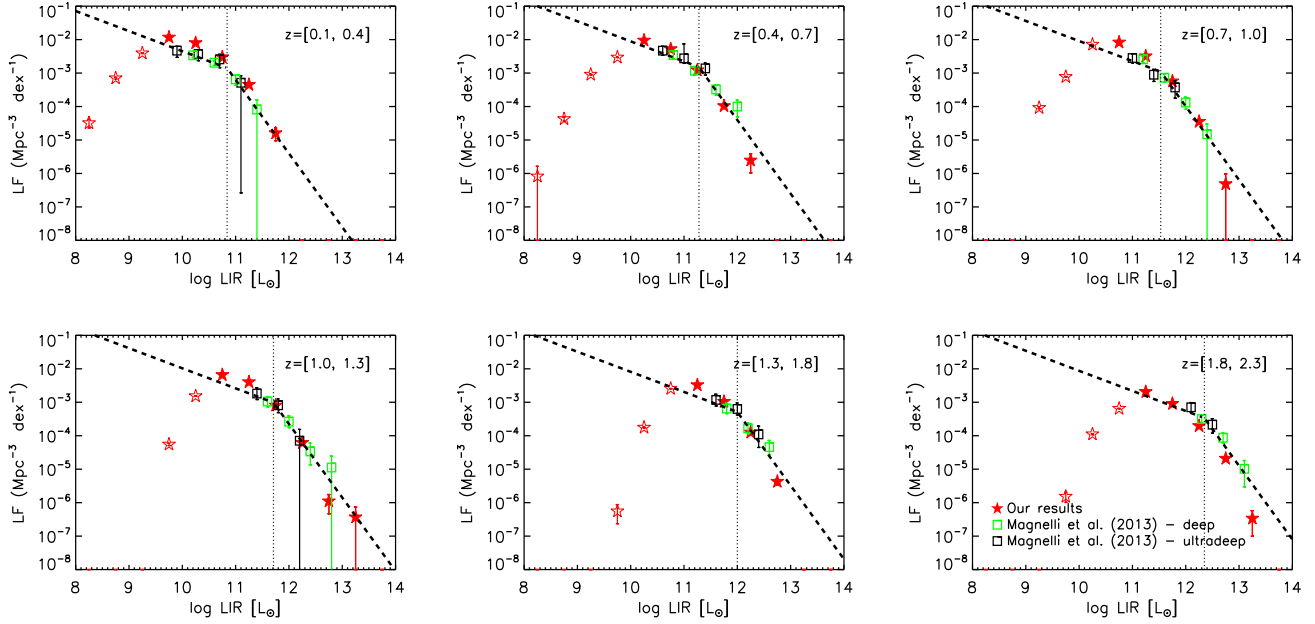
$$\phi(L) = \phi_* \left( \frac{L}{L_*} \right)^{1-\alpha} \exp \left[ -\frac{1}{2\sigma^2} \log_{10}^2 \left( 1 + \frac{L}{L_*} \right) \right], \quad (4)$$

which behaves as a power law for  $L < L_*$  and as a Gaussian in  $\log L$  for  $L > L_*$ . In principle, there are four free parameters in the modified Schechter function, that is,  $\alpha$ , which describes the faint-end slope,  $\sigma$ , which controls the shape of the cut-off at the bright end,  $L_*$ , which is the characteristic luminosity, and  $\phi_*$ , which is the characteristic density. However, due to a lack of dynamic range, [Gruppioni et al. \(2013\)](#) adopted the faint-end

slope value  $\alpha = 1.2$  and the Gaussian width parameter  $\sigma = 0.5$  derived in the first redshift bin ( $0.0 < z < 0.3$ ) for all higher redshift bins. In other words, only  $L_*$  and  $\phi_*$  are allowed to vary freely in the higher redshift bins.

In general, our measurements based on the de-blended catalogue in the COSMOS field and measurements from [Gruppioni et al. \(2013\)](#) in the overlapping luminosity range agree well. Additionally, we are able to extend the LF measurements down to much fainter luminosities and out to higher redshifts. Our measurements of the LF also seem to suggest that there are fewer sources at the bright end, which could be partially caused by the limited size of the COSMOS field. In Fig. 7 we also show the best-fit modified Schechter function to our measurements alone (the red lines) and the best-fit function to both our total IR LFs and the measurements in [Gruppioni et al. \(2013\)](#) (the blue lines). During the fitting process using the MCMC sampler emcee ([Foreman-Mackey et al. 2013](#)), we assumed that the characteristic luminosity  $L_*$  and characteristic density  $\phi_*$  evolve with redshift, but the bright-end Gaussian width  $\sigma$  and the faint-end slope  $\alpha$  do not depend on redshift. There are 13 redshift bins in Fig. 7, which means that in total, we have 28 free parameters. Table 1 lists the best-fit values and marginalised





**Fig. 6.** Total IR LF out to  $z \sim 2.3$ . The red stars are derived from our de-blended catalogue in COSMOS (filled red stars: our LF above the completeness limit; empty red stars: our LF below the completeness limit). Error bars on the red stars only represent Poisson errors. The black squares are from Magnelli et al. (2013) based on observations of the GOODS-S ultra-deep field. The green squares are also from Magnelli et al. (2013), but based on observations of the GOODS-N/S deep fields. The dashed line in each panel is the best-fit double power-law from Magnelli et al. (2013). The vertical dotted line indicates the location of the transition luminosity, i.e.  $L_{\text{knee}}$  in Eq. (2) and (3), as derived in Magnelli et al. (2013).

errors of the parameters in the modified Schechter functions derived from fitting to our measurements only (based on the de-blended catalogue in COSMOS). Table 2 lists the best-fit values and marginalised errors of the parameters in the modified Schechter functions derived from fitting to our measurements based on the de-blended catalogue and to the measurements from Gruppioni et al. (2013) presented in Fig. 7.

In Fig. 8 we plot the evolution of the characteristic luminosity  $L_*$  and normalisation  $\phi_*$  in the best-fitting modified Schechter function as a function of redshift or lookback time. Regarding the evolution of the characteristic density  $\phi_*$ , the two sets of measurements are consistent with each other within errors although the red stars (derived based on the de-blended catalogue in COSMOS only) are consistently below the blue stars. Similar to the conclusions reached in Gruppioni et al. (2013), we also find that the characteristic density evolves very mildly for the 8 Gyr or so ( $z \sim 1$ ) and then decreases rapidly (by about two orders of magnitude) from  $z \sim 1$  to  $\sim 6$ . Regarding the evolution of the characteristic luminosity  $L_*$ , the two measurement sets are again consistent with each other within errors. The red stars are consistently above the blue stars as a result of anti-correlation between  $L_*$  and  $\phi_*$ . As a function of redshift,  $L_*$  increases quickly with redshift out to  $z \sim 2$  and then seems to more or less flatten out to  $z \sim 6$ . As a function of lookback time,  $L_*$  seems to evolve in a simple linear fashion with time. The evolution of  $L_*$  is qualitatively similar to the evolution of the normalisation of the galaxy star-forming main sequence (e.g. Koprowski et al. 2016; Pearson et al. 2018).

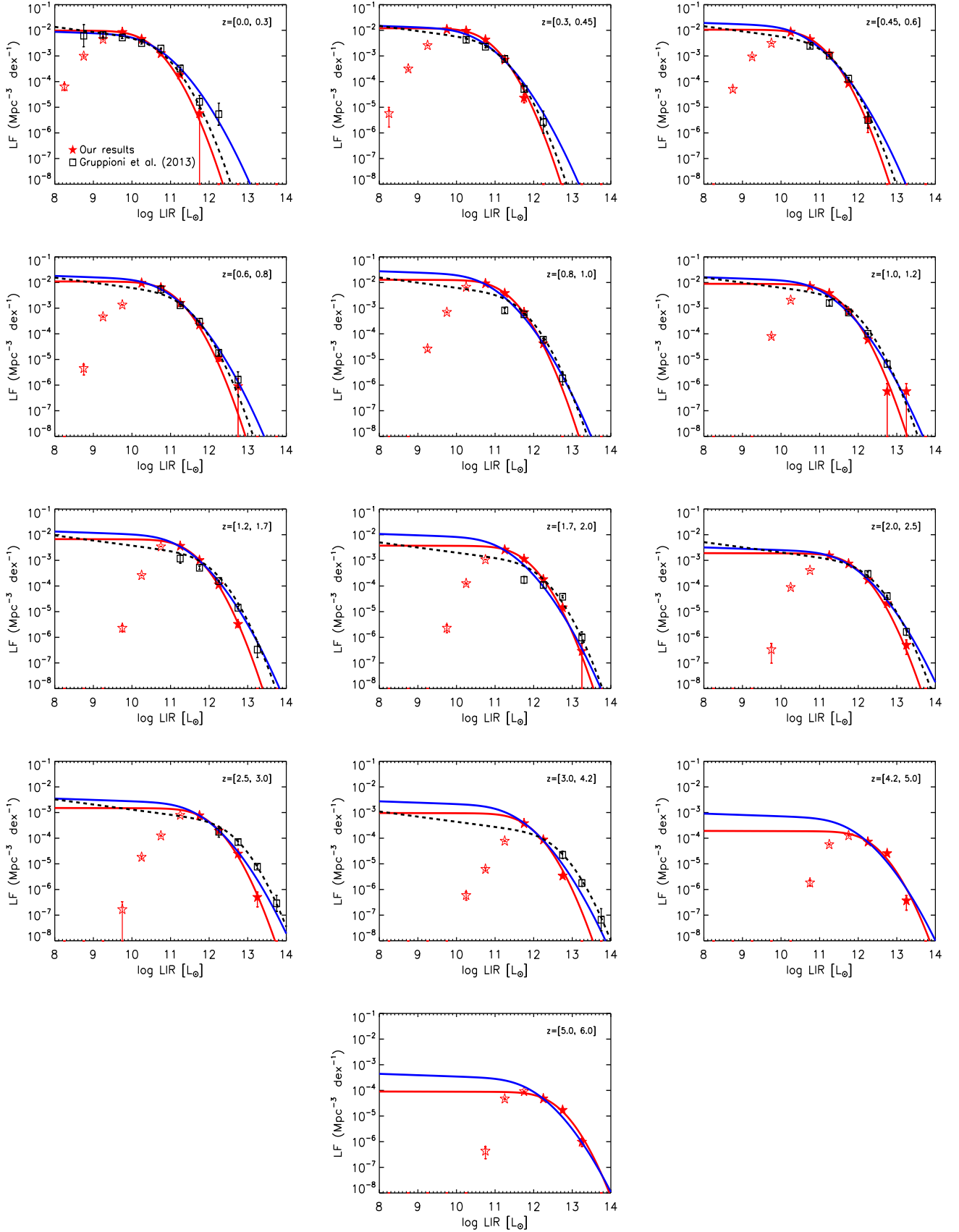
In Fig. 9 we compare our total IR LF with predictions from GALFORM out to  $z = 6$ . There is a lack of very bright sources with  $L_{\text{IR}} > 10^{13} L_{\odot}$  that is caused by the limited volume of the simulation (500 Mpc on a side). We further decomposed the predicted IR LF from GALFORM into starburst and quiescent populations. The transition between starburst and quiescence occurs at around  $10^{12} L_{\odot}$  at low redshifts and decreases to around  $10^{11} L_{\odot}$  towards high redshift. The overall agreement

**Table 1.** Best-fit values and marginalised errors of the parameters in the modified Schechter functions derived from fitting to our measurements only (based on the de-blended catalogue in COSMOS).

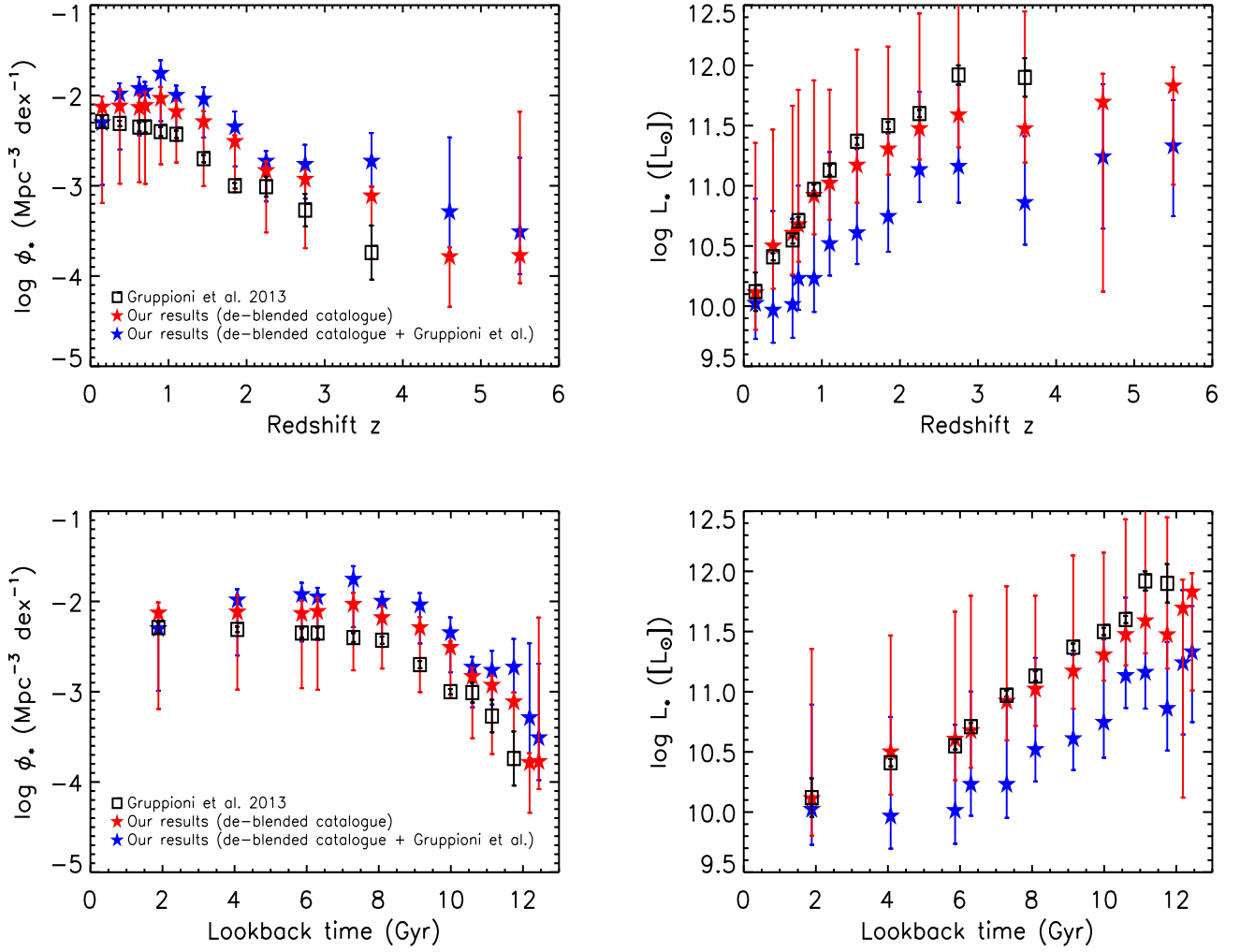
Redshift range	$\phi_*$	$L_*$	$N$
$0.0 < z < 0.3$	$-2.13^{+0.12}_{-1.06}$	$10.11^{+1.24}_{-0.31}$	2932
$0.3 < z < 0.45$	$-2.12^{+0.18}_{-0.86}$	$10.50^{+0.97}_{-0.36}$	7990
$0.45 < z < 0.6$	$-2.13^{+0.14}_{-0.83}$	$10.61^{+1.06}_{-0.34}$	9113
$0.6 < z < 0.8$	$-2.11^{+0.15}_{-0.87}$	$10.68^{+1.12}_{-0.31}$	18660
$0.8 < z < 1.0$	$-2.03^{+0.13}_{-0.73}$	$10.92^{+0.95}_{-0.33}$	26469
$1.0 < z < 1.2$	$-2.18^{+0.12}_{-0.56}$	$11.02^{+0.76}_{-0.30}$	21127
$1.2 < z < 1.7$	$-2.29^{+0.12}_{-0.71}$	$11.17^{+0.96}_{-0.31}$	36380
$1.7 < z < 2.0$	$-2.51^{+0.10}_{-0.56}$	$11.31^{+0.85}_{-0.21}$	15491
$2.0 < z < 2.5$	$-2.83^{+0.10}_{-0.69}$	$11.48^{+0.96}_{-0.26}$	13429
$2.5 < z < 3.0$	$-2.93^{+0.11}_{-0.76}$	$11.59^{+1.02}_{-0.27}$	9323
$3.0 < z < 4.2$	$-3.11^{+0.10}_{-0.70}$	$11.47^{+0.97}_{-0.28}$	5037
$4.2 < z < 5.0$	$-3.79^{+0.10}_{-0.56}$	$11.70^{+0.24}_{-1.57}$	1369
$5.0 < z < 6.0$	$-3.77^{+1.59}_{-0.31}$	$11.83^{+0.16}_{-0.82}$	599
$\alpha = 1.26^{+0.36}_{-0.25}$			
$\sigma = 0.44^{+0.06}_{-0.27}$			

**Notes.** The faint-end slope parameter  $\alpha$  and the bright-end Gaussian width parameter  $\sigma$  are assumed to be independent of redshift. The last column  $N$  shows the number of galaxies above the completeness limit in the corresponding redshift bin.

between our measurements and the GALFORM predictions at the bright end ( $L_{\text{IR}} > 10^{11} L_{\odot}$ ) is reasonably good, especially at  $z < 2.5$ . It is clear that in order to match the observations at the bright end, the population of starburst galaxies in the simulation is of great importance. On the other hand, the population



**Fig. 7.** Total IR LF out to  $z \sim 6$ . The red stars are derived from our de-blended catalogue in COSMOS (filled red stars: our LF above the completeness limit; empty red stars: our LF below the completeness limit). Error bars on the red stars only represent Poisson errors. The black squares are from Gruppioni et al. (2013). The dashed line is the best-fit modified-Schechter function from Gruppioni et al. (2013). The red line is the best-fit function derived from fitting to our measurements of the IR LF only (i.e., the filled red stars). The blue line is the best-fit function derived from fitting to both our IR LFs and the measurement in Gruppioni et al. (2013) (i.e. the filled red stars and the black squares).



**Fig. 8.** Evolution of the characteristic luminosity  $L_*$  and normalisation  $\phi_*$  in the best-fitting modified Schechter function, i.e. Eq. (4). The red stars are derived from fitting to the measurements based on the de-blended catalogue in COSMOS only. The blue stars are derived from fitting to both the measurements from the de-blended catalogue in COSMOS and the measurements from Gruppioni et al. (2013). The black squares show the measurements taken from Gruppioni et al. (2013). *Top panels:* evolution of  $L_*$  and  $\phi_*$  as a function of redshift. *Bottom panels:* evolution of  $L_*$  and  $\phi_*$  as a function of lookback time.

of quiescent galaxies in the simulation is important to match the observations at the faint end. However, at  $z < 1$ , GALFORM predictions at the faint end, where the quiescent population dominates the starburst population, are much lower than our measurements. At higher redshifts  $z > 2.5$ , GALFORM seems to over-predict the number of bright dusty star-forming galaxies compared to our measurements, and this over-prediction generally becomes worse towards higher redshift. Further studies are needed to understand the cause of this prediction; it might be an over-production of starburst galaxies or may have to do with the top-heavy IMF. Figure 9 demonstrates that it is also informative to compare predictions of the LF as a function of redshift, and not just the number counts.

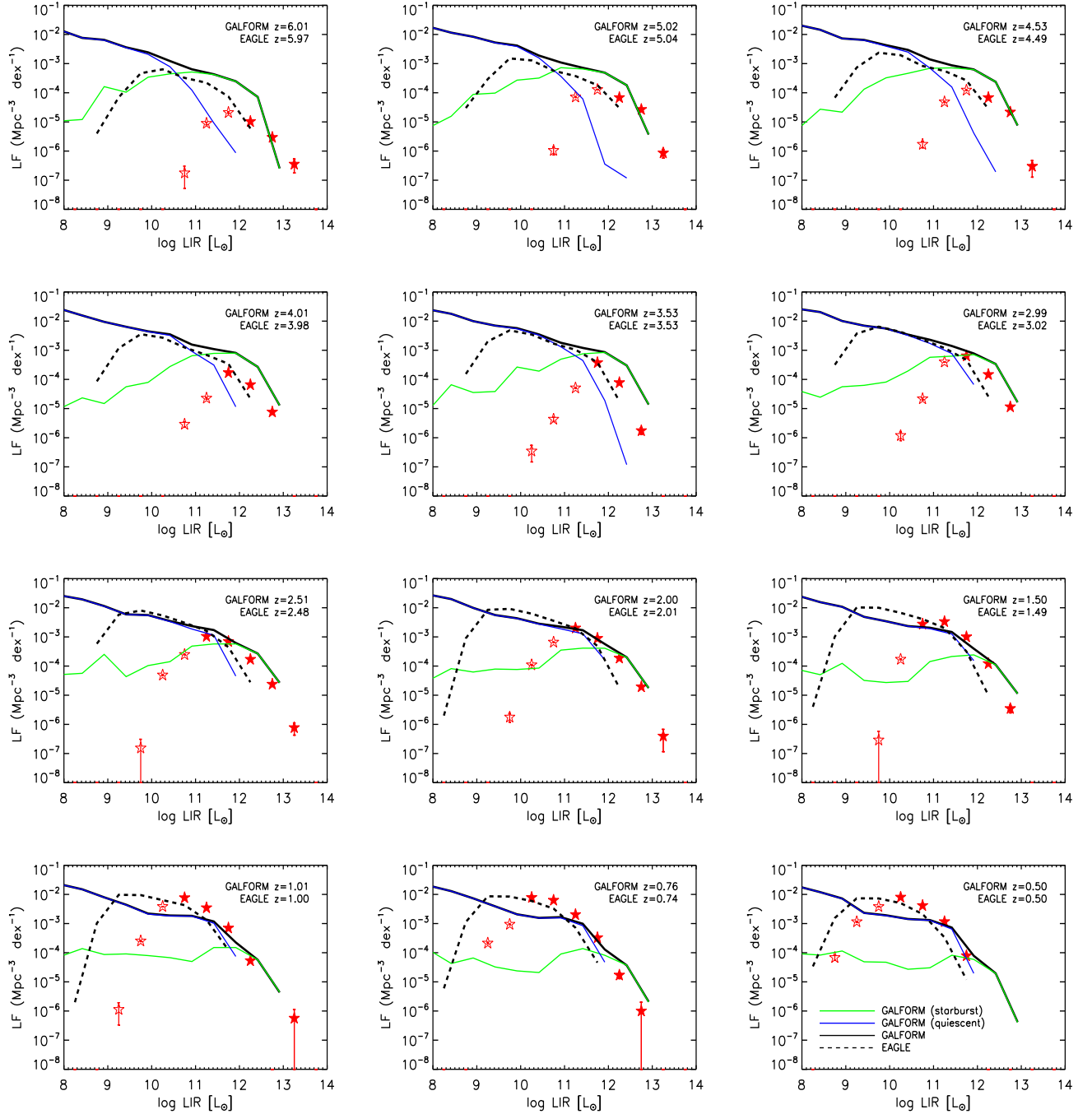
In Fig. 9 we also compare our total IR LF with predictions from the EAGLE hydrodynamic simulation out to  $z = 6$ . The lack of sources with  $L_{\text{IR}} > 10^{12} L_{\odot}$  is caused by the limited volume of the EAGLE simulation (100 Mpc on a side). The drop at the faint end is due to the criteria that only galaxies with stellar masses in excess of  $10^{8.5} M_{\odot}$  and with dust distributions resolved by  $\geq 250$  particles are included in the EAGLE catalogue. In general, the total IR LF predicted from the EAGLE hydrodynamic simulation is lower at the bright end than our measurements, which reflects the under-prediction seen in the

number count plots (from Fig. 2 to Fig. 4). This could be caused by a lack of starburst galaxies, the poor sampling of the largest haloes in the 100 Mpc EAGLE box, the need of adopting a different IMF (e.g. a top-heavy IMF) for the starburst population, or the need of changing the subgrid physical prescriptions of the simulation related to star formation, feedback, and so on. Interestingly, the total IR LF predicted from EAGLE agrees better with the observations than with the predictions from GALFORM in the intermediate-luminosity range (between  $L_{\text{IR}} \sim 10^{10} L_{\odot}$  and  $L_{\text{IR}} \sim 10^{11} L_{\odot}$ ) at  $z < 1$ . At  $z > 2.5$ , where GALFORM over-predicts the IR LF, EAGLE predictions agree reasonably well with the observations in the overlapping dynamic range.

### 3.3. Cosmic star-formation history

In the past two decades, impressive progress has been made in charting star formation from the local Universe to the epoch of re-ionisation (e.g. Hopkins & Beacom 2006; Madau & Dickinson 2014; Gruppioni et al. 2013; Bouwens et al. 2015), for which a multitude of SFR tracers has been used. It is a remarkable achievement that there is a reasonable consensus regarding the recent history below  $z \sim 2$ . However, above  $z \sim 3$ , great differences of more



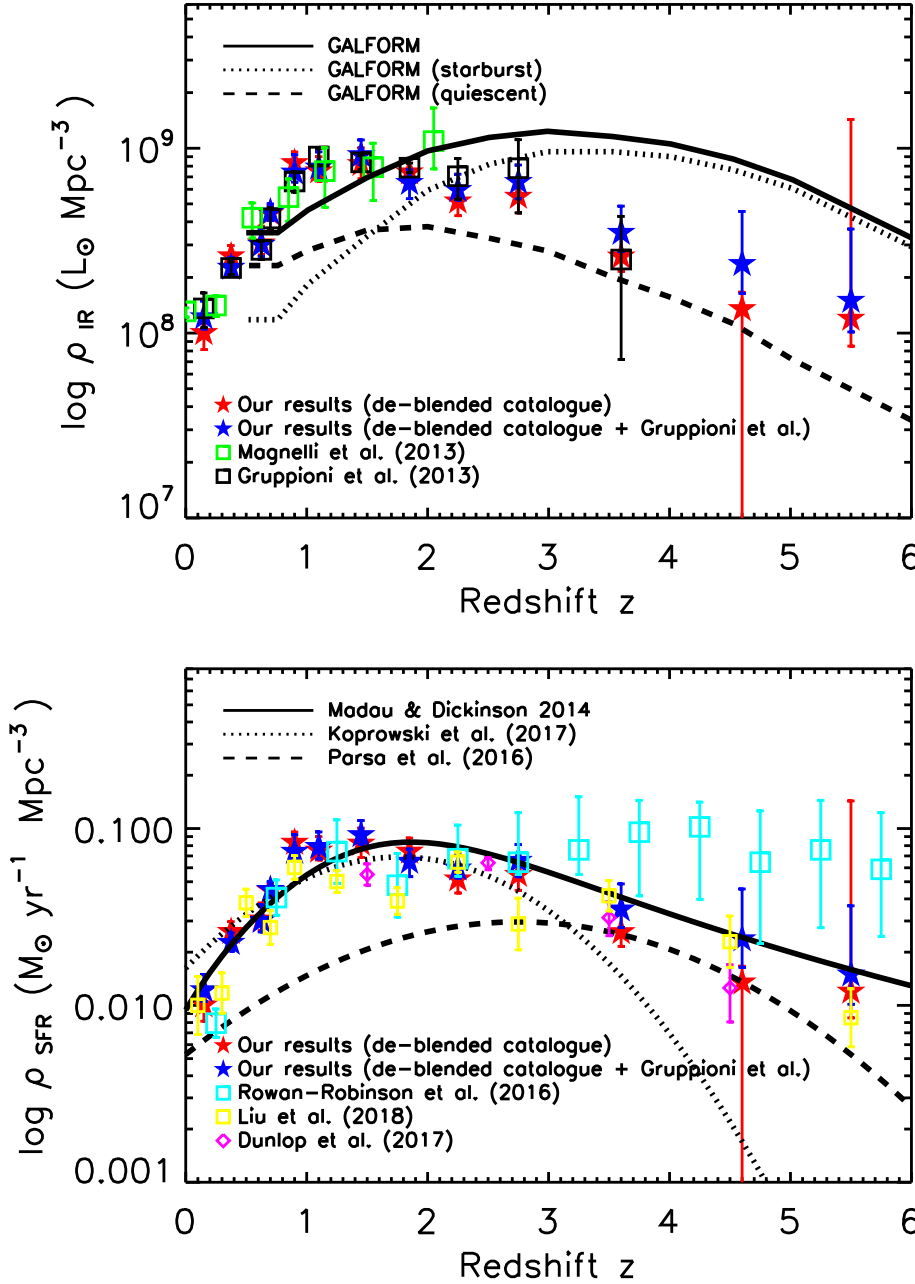


**Fig. 9.** Total IR LF. The red stars are derived from our de-blended catalogue in COSMOS (filled red stars: our LF above the completeness limit; empty red stars: our LF below the completeness limit). Error bars on the red stars only represent Poisson errors. The thick black solid lines are from the Durham SAM. The thin blue solid lines correspond to the LFs of the quiescent galaxies from the Durham SAM. The thin green solid lines correspond to the LFs of the starburst galaxies from the Durham SAM. The black dashed lines are from the EAGLE hydrodynamic simulation.

than one order of magnitude still exist. This order-of-magnitude difference encompasses multiple very different predictions from competing galaxy evolution models (e.g. [Grupponi et al. 2015](#); [Henriques et al. 2015](#); [Lacey et al. 2016](#)), therefore it is vital to measure the CSFH with much greater precision and accuracy. The cosmic epoch over  $3 < z < 4$  is also very interesting; some studies suggest that the balance of power may shift from unobscured star formation to dusty star formation ([Koprowski et al. 2017](#); [Dunlop et al. 2017](#); [Bourne et al. 2017](#)). The most direct SFR tracer measures UV light, which is redshifted to optical and IR for distant galaxies. As a result, very sensitive instruments (such

as the *Hubble Space Telescope*) can be used to probe the SFR density (SFRD) in these early cosmic epochs ([McLure et al. 2013](#); [Bouwens et al. 2015, 2016](#); [Finkelstein et al. 2015](#); [Parsa et al. 2016](#)). However, large and uncertain dust extinction correction needs to be applied to these UV-only observations. To directly probe the dust-obscured star formation, we need FIR and sub-mm SFR tracers.

The deep SPIRE maps contain most of the emission in the cosmic IR background, which arises from the integrated dust emission over the entire cosmic history ([Puget et al. 1996](#)). However, because of the large beam, SPIRE observations



**Fig. 10.** *Top panel:* redshift evolution of the total IR luminosity density  $\rho_{\text{IR}}$  out to  $z \sim 6$ . The results of integrating the best-fit modified Schechter function for our observed total IR LF only (based on the de-blended catalogue in COSMOS) are shown as red stars. The results of integrating the best-fit function for both our observed total IR LF and the Gruppioni et al. (2013) IR LF are shown as blue stars. Error bars on the stars represent the  $1\sigma$  uncertainty. The measurements from Magnelli et al. (2013) are shown as green squares. The measurements from Gruppioni et al. (2013) are shown as black squares. The black solid line shows the predicted  $\rho_{\text{IR}}$  as a function of redshift from GALFORM. The black dotted line shows the predicted  $\rho_{\text{IR}}$  of the starburst galaxy population from GALFORM. The black dashed line shows the predicted  $\rho_{\text{IR}}$  of the quiescent galaxy population from GALFORM. *Bottom panel:* co-moving SFRD  $\rho_{\text{SFR}}$  as a function of redshift out to  $z \sim 6$ . Estimates of the dust-obscured SFRD  $\rho_{\text{SFR}}^{\text{obscured}}$  based on the best-fit function for our total IR LF only are shown as red stars. Estimates of  $\rho_{\text{SFR}}^{\text{obscured}}$  based on the joint constraints from our total IR LF and the Gruppioni et al. (2013) IR LF are shown as blue stars. We also compare with other *Herschel*-based studies such as Rowan-Robinson et al. (2016) and Liu et al. (2018) and the ALMA-based study of Dunlop et al. (2017). The black dotted line shows the best-fit function of the evolution of  $\rho_{\text{SFR}}^{\text{obscured}}$  from Koprowski et al. (2017). The black dashed line corresponds to the UV-based unobscured SFRD estimates  $\rho_{\text{SFR}}^{\text{unobscured}}$  from Parsa et al. (2016). The black solid line shows the parametric description of the evolution of the total SFRD  $\rho_{\text{SFR}}$  provided by Madau & Dickinson (2014).

suffer from source confusion and blending, which limits our ability to detect faint objects and de-blend neighbouring sources. At  $z \sim 3$ , the SPIRE  $5\sigma$  confusion limit corresponds to an IR luminosity  $10^{13} L_{\odot}$ , which is many times brighter than the expected turnover in the LF. The advent of ALMA finally closed the gap with UV/optical observations in sensitivity and angular resolution. Dunlop et al. (2017) presented the first deep ALMA image of the 4.5 arcmin<sup>2</sup> *Hubble* Ultra Deep Field and measurement of the SFRD using a total sample of 16 sources over  $1 < z < 5$ . Even with its extraordinary sensitivity, the small field of view means that it is still impractical to use ALMA to carry out large blank-field surveys to address the CSFH controversy, which requires statistically large samples of galaxies. To help resolve the SFRD controversy at  $z > 3$ , we can exploit our SED prior-enhanced XID+ de-blended *Herschel* photometry (which significantly extends the dynamic range probed by previous studies) to probe the knee of the IR LF and derive much tighter constraints of the SFRD.

Based on the parametric descriptions of the total IR LF, we can carry out luminosity-weighted integration over a sufficient dynamic range in each redshift bin to study the time evolution of the total IR luminosity density  $\rho_{\text{IR}}$ . In the top panel of Fig. 10, we show the evolution of  $\rho_{\text{IR}}$  out to  $z \sim 6$ . Our two sets of measurements (red stars and blue stars) are consistent with each other within errors. It is clear that our measurements are also consistent with these two previous *Herschel*-based studies, Gruppioni et al. (2013) and Magnelli et al. (2013), in the overlapping redshift range. Because of the large fraction of photometric redshifts and because the PEP selection might miss high-redshift sources, the Gruppioni et al. (2013) estimate in the redshift bin  $3.0 < z < 4.2$  is likely to be a lower limit. We also plot the predicted total IR luminosity density as well as the IR luminosity density from the starburst and quiescent galaxy populations separately from GALFORM as a function of redshift. We find that the starburst population dominates at high redshift and the quiescent population dominates at low redshift. The

**Table 2.** Best-fit values and marginalised errors of the parameters in the modified Schechter functions derived from fitting to both our measurements and the measurements from Gruppioni et al. (2013).

Redshift range	$\phi_*$	$L_*$
$0.0 < z < 0.3$	$-2.30^{+0.14}_{-0.69}$	$10.02^{+0.87}_{-0.29}$
$0.3 < z < 0.45$	$-1.98^{+0.11}_{-0.62}$	$9.97^{+0.82}_{-0.27}$
$0.45 < z < 0.6$	$-1.92^{+0.13}_{-0.52}$	$10.01^{+0.71}_{-0.27}$
$0.6 < z < 0.8$	$-1.95^{+0.10}_{-0.48}$	$10.23^{+0.77}_{-0.26}$
$0.8 < z < 1.0$	$-1.75^{+0.14}_{-0.53}$	$10.23^{+0.73}_{-0.28}$
$1.0 < z < 1.2$	$-2.00^{+0.11}_{-0.50}$	$10.52^{+0.76}_{-0.27}$
$1.2 < z < 1.7$	$-2.04^{+0.13}_{-0.43}$	$10.61^{+0.70}_{-0.26}$
$1.7 < z < 2.0$	$-2.35^{+0.17}_{-0.44}$	$10.75^{+0.69}_{-0.30}$
$2.0 < z < 2.5$	$-2.73^{+0.12}_{-0.44}$	$11.13^{+0.65}_{-0.27}$
$2.5 < z < 3.0$	$-2.76^{+0.21}_{-0.38}$	$11.16^{+0.68}_{-0.30}$
$3.0 < z < 4.2$	$-2.73^{+0.31}_{-0.28}$	$10.86^{+0.55}_{-0.35}$
$4.2 < z < 5.0$	$-3.29^{+0.83}_{-0.45}$	$11.24^{+0.60}_{-0.60}$
$5.0 < z < 6.0$	$-3.51^{+0.82}_{-0.47}$	$11.33^{+0.38}_{-0.58}$
$\alpha = 1.28^{+0.39}_{-0.20}$		
$\sigma = 0.65^{+0.04}_{-0.12}$		

**Notes.** The faint-end slope parameter  $\alpha$  and the bright-end Gaussian width parameter  $\sigma$  are assumed to be independent of redshift.

transition occurs at around  $z \sim 1.5$ . The GALFORM-predicted total IR luminosity densities agree reasonably well with the observations at  $z < 3$ . At  $z > 3$ , the GALFORM predictions are much higher than the observed values. Cowley et al. (2018b) discussed the considerable difference between the intrinsic CSFH predicted by GALFORM and the apparent CSFH derived by converting the IR luminosity density into SFR volume density.

To derive the dust-obscured SFR volume density  $\rho_{\text{SFR}}^{\text{obscured}}$ , we multiplied our estimates of  $\rho_{\text{IR}}$  by a constant factor of  $10^{-10} M_{\odot} \text{ yr}^{-1} / L_{\odot}$  (Béthermin et al. 2017), which is derived from the Kennicutt (1998) conversion factor after converting into the Chabrier (2003) IMF. In the bottom panel of Fig. 10, we show our measurements of  $\rho_{\text{SFR}}^{\text{obscured}}$ . To avoid overcrowding, estimates of  $\rho_{\text{SFR}}^{\text{obscured}}$  based on measurements from Magnelli et al. (2013) and Gruppioni et al. (2013) are not shown because they are consistent with our results based on the good agreement seen in the top panel of Fig. 10 and because the same conversion factor is applied to convert  $\rho_{\text{IR}}$  into  $\rho_{\text{SFR}}$ . We also compare with two other *Herschel*-based studies, that is, Rowan-Robinson et al. (2016) with small updates from Rowan-Robinson et al. (2018) in some redshift bins and Liu et al. (2018), and the ALMA-based study by Dunlop et al. (2017). Rowan-Robinson et al. (2016) used a novel approach of selecting  $500 \mu\text{m}$  sources from a combination of several large HerMES fields totalling  $\sim 20 \text{ deg}^2$  in order to extend the measurements of Gruppioni et al. (2013) out to  $z \sim 6$ . We find that the measurements of Rowan-Robinson et al. (2016) are systematically higher than our results at  $z > 3$ , although they are still marginally consistent. The estimates of Liu et al. (2018) were derived by using super-deblended dust emission in galaxies in the GOODS-North field, based on prior catalogues constructed from deep *Spitzer*  $24 \mu\text{m}$  and VLA 20 cm detections and progressive de-blending from less strongly to more strongly confused bands. Our results agree well with the  $\rho_{\text{SFR}}^{\text{obscured}}$  estimates derived by Liu et al. (2018). The ALMA-derived measurements of the dust obscured  $\rho_{\text{SFR}}$  based on a sample of 16 sources

from Dunlop et al. (2017) also agree reasonably well with our estimates.

We also compare with the unobscured SFRD estimates  $\rho_{\text{SFR}}^{\text{unobscured}}$  from Parsa et al. (2016), which were based on converting from the rest-frame UV ( $1500 \text{ \AA}$ ) luminosity into UV-visible SFR. We do not find evidence for a shift of balance between  $z \sim 3$  and  $z \sim 4$  in the CSFH from being dominated by unobscured star formation at high redshift to obscured star formation at low redshift (subject to uncertainties associated with estimates of the unobscured SFRD), as found by previous studies of Koprowski et al. (2017), Dunlop et al. (2017), and Bourne et al. (2017). For example, the black dotted line in the right panel of Fig. 10 shows the best-fitting function of the evolution of  $\rho_{\text{SFR}}^{\text{obscured}}$  from Koprowski et al. (2017), which crosses the evolution of  $\rho_{\text{SFR}}^{\text{unobscured}}$  from Parsa et al. (2016) (i.e. the dashed line) at  $z \sim 3$ . However, we do find the redshift range  $3 < z < 4$  to be an interesting transition period because the fraction of the total SFRD that is obscured by dust is significantly lower at higher redshift than at lower redshifts. The fraction of dust-obscured SF activity is at its highest ( $> 80\%$ ) around  $z \sim 1$ , which then decreases towards both low and high redshift.

## 4. Discussions and conclusions

We used our multi-wavelength de-blended *Herschel*-SPIRE catalogue in the COSMOS field to study the number counts, the monochromatic and total IR (integrated from 8 to  $1000 \mu\text{m}$ ) LFs, and the dust-obscured CSFH. We compared our results with previous determinations from single-dish and interferometric observations and predictions from both empirical models and physically motivated models (including semi-analytic simulations and hydrodynamic simulations). Our main conclusions are the following:

- Our number counts at the SPIRE wavelength  $250 \mu\text{m}$  derived from the multi-wavelength de-blended catalogue in the COSMOS field show good agreement with previous *Herschel* measurements. However, the agreement is increasingly worse towards the longer wavelengths  $350$  and  $500 \mu\text{m}$ . At  $500 \mu\text{m}$ , our number counts can be as much as  $0.5$  dex lower than previous *Herschel* studies because previous *Herschel* studies suffered from confusion and source-blending problems, which are increasingly more severe towards longer wavelengths.
- Our number counts at  $450 \mu\text{m}$  from the de-blended catalogue in COSMOS agree very well with the JCMT SCUBA-2  $450 \mu\text{m}$  measurement (with a factor of  $\sim 5$  improvement in angular resolution), especially with the most recent SCUBA-2 measurements from Zavala et al. (2017) and Wang et al. (2017).
- Excellent agreement is found between our predicted number counts at  $870 \mu\text{m}$  based on the de-blended catalogue in COSMOS and the SCUBA-2  $850 \mu\text{m}$ , SMA  $860 \mu\text{m}$ , and ALMA  $870 \mu\text{m}$  measurements, which are derived either from single-dish observations or interferometric observations achieving arcsec or even sub-arcsec angular resolution.
- Our monochromatic rest-frame  $250 \mu\text{m}$  LFs agree well with SCUBA-2 measurements (Koprowski et al. 2017) in the overlapping luminosity and redshift range. We extended the Koprowski et al. (2017) measurements by around 1 dex at the faint end, except in the redshift bin  $1.5 < z < 2.5$ , where measurements in the two faintest luminosity bins in Koprowski et al. (2017) were derived from ALMA  $1.3 \text{ mm}$  observations.



- Our total IR LF agrees well with previous *Herschel* PACS and SPIRE measurements Magnelli et al. 2013; Gruppioni et al. 2013 in the overlapping luminosity and redshift range. Through our de-blending technique and based on the wealth of multi-wavelength photometric information in the COSMOS field, we also probed much fainter luminosities and out to higher redshifts. We derived the best-fitting modified Schechter function in a number of redshift bins out to  $z \sim 6$ . We find that the characteristic density evolves very mildly for the 8 Gyr and then decreases rapidly (by about two orders of magnitude) from  $z \sim 1$  to  $z \sim 6$ . The characteristic luminosity  $L_*$  increases quickly with redshift out to  $z \sim 2$  and then seems to more or less flatten out to  $z \sim 6$ . As a function of lookback time,  $L_*$  evolves simply in a linear fashion.
- We find a reasonable agreement between our total IR LF and the predictions from GALFORM. The population of starburst galaxies with top-heavy IMF is important to match the observed LF at the bright end. On the other hand, the population of quiescent galaxies in the simulation is important to match the observations at the faint end. However, at the faint end, GALFORM predictions are considerably below our measured total IR LF at  $z < 1$ .
- The predicted total IR LF from the EAGLE hydrodynamic simulation are generally lower at the bright end than our measurement, which could be caused by the limited volume of EAGLE and lack of high mass haloes, a lack of starburst galaxies, the need of adopting a different IMF (e.g. a top-heavy IMF) for the starburst population, or the need of changing the subgrid physical prescriptions of the simulation. Interestingly, towards the faint end, the predicted total IR LF from EAGLE agrees fairly well with our measurements.
- Our measurement of the co-moving IR luminosity density and the dust-obscured SFR volume density as a function of redshift agree well with previous *Herschel* studies but extends to higher redshifts. By comparing our results with the SFRD estimates derived from UV-based studies, we find that the fraction of dust-obscured SF activity is at its highest (>80%) around  $z \sim 1$ , which then decreases towards both low and high redshift. We do not find evidence for a shift of balance between  $z \sim 3$  and  $z \sim 4$  in the CSFH from being dominated by unobscured star formation at high redshift to obscured star formation at low redshift. However, we do find the redshift range  $3 < z < 4$  to be an interesting transition period because the fraction of the total SFRD that is obscured by dust is significantly lower at higher redshift than at lower redshift.

**Acknowledgements.** M.J.M. acknowledges the support of the National Science Centre, Poland, through the POLONEZ grant 2015/19/P/ST9/04010; this project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 665778. We thank an anonymous referee for helpful comments that have improved the paper.

## References

Asboth, V., Conley, A., Sayers, J., et al. 2016, *MNRAS*, **462**, 1989  
 Baes, M., Davies, J. I., Dejonghe, H., et al. 2003, *MNRAS*, **343**, 1081  
 Baes, M., Verstaappen, J., De Looze, I., et al. 2011, *ApJS*, **196**, 22  
 Baugh, C. M., Lacey, C. G., Frenk, C. S., et al. 2005, *MNRAS*, **356**, 1191  
 Baugh, C. M., Gonzalez-Perez, V., del P Lagos, C., et al. 2019, *MNRAS*, **483**, 4922  
 Béthermin, M., Dole, H., Cousin, M., & Bavouzet, N. 2010, *A&A*, **516**, A43  
 Béthermin, M., Le Floch, E., Ilbert, O., et al. 2012a, *A&A*, **542**, A58  
 Béthermin, M., Daddi, E., Magdis, G., et al. 2012b, *ApJ*, **757**, L23  
 Béthermin, M., Wang, L., Doré, O., et al. 2013, *A&A*, **557**, A66

Béthermin, M., Wu, H.-Y., Lagache, G., et al. 2017, *A&A*, **607**, A89  
 Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, *A&A*, **622**, A103  
 Boselli, A., Eales, S., Cortese, L., et al. 2010, *PASP*, **122**, 261  
 Bourne, N., Dunlop, J. S., Merlin, E., et al. 2017, *MNRAS*, **467**, 1360  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, **803**, 34  
 Bouwens, R. J., Oesch, P. A., Labbé, I., et al. 2016, *ApJ*, **830**, 67  
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000  
 Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, *MNRAS*, **360**, 1413  
 Bussmann, R. S., Riechers, D., Fialkov, A., et al. 2015, *ApJ*, **812**, 43  
 Camps, P., & Baes, M. 2015, *Astron. Comput.*, **9**, 20  
 Camps, P., Trayford, J. W., Baes, M., et al. 2016, *MNRAS*, **462**, 1057  
 Camps, P., Trčka, A., Trayford, J., et al. 2018, *ApJS*, **234**, 20  
 Casey, C. M., Chen, C.-C., Cowie, L. L., et al. 2013, *MNRAS*, **436**, 1919  
 Chabrier, G. 2003, *PASP*, **115**, 763  
 Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718  
 Chen, C.-C., Cowie, L. L., Barger, A. J., et al. 2013, *ApJ*, **762**, 81  
 Clements, D. L., Rigby, E., Maddox, S., et al. 2010, *A&A*, **518**, L8  
 Condon, J. J. 1974, *ApJ*, **188**, 279  
 Coppin, K., Chapin, E. L., Mortier, A. M. J., et al. 2006, *MNRAS*, **372**, 1621  
 Cortese, L., Ciesla, L., Boselli, A., et al. 2012, *A&A*, **540**, A52  
 Cowley, W. I., Lacey, C. G., Baugh, C. M., & Cole, S. 2015, *MNRAS*, **446**, 1784  
 Cowley, W. I., Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 2018a, *MNRAS*, **474**, 2352  
 Cowley, W. I., Lacey, C. G., Baugh, C. M., et al. 2018b, *MNRAS*, submitted [arXiv:1808.05208]  
 Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, **450**, 1937  
 Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, **670**, 156  
 Dole, H., Lagache, G., & Puget, J.-L. 2003, *ApJ*, **585**, 617  
 Dole, H., Lagache, G., Puget, J.-L., et al. 2006, *A&A*, **451**, 417  
 Draine, B. T., & Li, A. 2007, *ApJ*, **657**, 810  
 Draine, B. T., Aniano, G., Krause, O., et al. 2014, *ApJ*, **780**, 172  
 Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, *MNRAS*, **466**, 861  
 Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, **468**, 33  
 Eales, S., Dunne, L., Clements, D., et al. 2010, *PASP*, **122**, 499  
 Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, *A&A*, **533**, A119  
 Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., et al. 2015, *ApJ*, **810**, 71  
 Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L., & Shafer, R. A. 1998, *ApJ*, **508**, 123  
 Fontanot, F., Monaco, P., Silva, L., & Grazian, A. 2007, *MNRAS*, **382**, 903  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306  
 Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, *MNRAS*, **366**, 767  
 Geach, J. E., Chapin, E. L., Coppin, K. E. K., et al. 2013, *MNRAS*, **432**, 53  
 Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, *MNRAS*, **465**, 1789  
 Glenn, J., Conley, A., Béthermin, M., et al. 2010, *MNRAS*, **409**, 109  
 Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, *ApJ*, **600**, 580  
 Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, **518**, L3  
 Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35  
 Groves, B., Dopita, M. A., Sutherland, R. S., et al. 2008, *ApJS*, **176**, 438  
 Gruppioni, C., & Pozzi, F. 2019, *MNRAS*, **483**, 1993  
 Gruppioni, C., Pozzi, F., Rodighiero, G., et al. 2013, *MNRAS*, **432**, 23  
 Gruppioni, C., Calura, F., Pozzi, F., et al. 2015, *MNRAS*, **451**, 3419  
 Hauser, M. G., & Dwek, E. 2001, *ARA&A*, **39**, 249  
 Hauser, M. G., Arendt, R. G., Kelsall, T., et al. 1998, *ApJ*, **508**, 25  
 Hayward, C. C., Narayanan, D., Kereš, D., et al. 2013, *MNRAS*, **428**, 2529  
 Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, *MNRAS*, **451**, 2663  
 Hill, R., Chapman, S. C., Scott, D., et al. 2018, *MNRAS*, **477**, 2042  
 Ho, P. T. P., Moran, J. M., & Lo, K. Y. 2004, *ApJ*, **616**, L1  
 Hodge, J. A., Karim, A., Smail, I., et al. 2013, *ApJ*, **768**, 91  
 Holland, W. S., Robson, E. I., Gear, W. K., et al. 1999, *MNRAS*, **303**, 659  
 Holland, W. S., Bintley, D., Chapin, E. L., et al. 2013, *MNRAS*, **430**, 2513  
 Hopkins, A. M., & Beacom, J. F. 2006, *ApJ*, **651**, 142  
 Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, *MNRAS*, **464**, 885  
 Karim, A., Swinbank, A. M., Hodge, J. A., et al. 2013, *MNRAS*, **432**, 2  
 Kennicutt, Jr., R. C. 1998, *ApJ*, **498**, 541  
 Kessler, M. F., Steinz, J. A., Anderregg, M. E., et al. 1996, *A&A*, **315**, L27  
 Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, *MNRAS*, **457**, 4340  
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36  
 Koprowski, M. P., Dunlop, J. S., Michałowski, M. J., et al. 2016, *MNRAS*, **458**, 4321  
 Koprowski, M. P., Dunlop, J. S., Michałowski, M. J., et al. 2017, *MNRAS*, **471**, 4155  
 Lacey, C. G., Baugh, C. M., Frenk, C. S., et al. 2016, *MNRAS*, **462**, 3854  
 Lagache, G., Puget, J.-L., & Gispert, R. 1999, *Ap&SS*, **269**, 263  
 Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24

- Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, *ApJ*, **801**, 80
- Liu, D., Daddi, E., Dickinson, M., et al. 2018, *ApJ*, **853**, 172
- Lo Faro, B., Buat, V., Roehlly, Y., et al. 2017, *MNRAS*, **472**, 1372
- Lutz, D. 2014, *ARA&A*, **52**, 373
- Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, *A&A*, **532**, A90
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Magnelli, B., Popesso, P., Berta, S., et al. 2013, *A&A*, **553**, A132
- Marchetti, L., Vaccari, M., Franceschini, A., et al. 2016, *MNRAS*, **456**, 1999
- Marsden, G., Ade, P. A. R., Bock, J. J., et al. 2009, *ApJ*, **707**, 1729
- McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, *Astron. Comput.*, **15**, 72
- McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, *MNRAS*, **432**, 2696
- Michałowski, M., Hjorth, J., & Watson, D. 2010, *A&A*, **514**, A67
- Michałowski, M. J., Dunlop, J. S., Koprowski, M. P., et al. 2017, *MNRAS*, **469**, 492
- Mortier, A. M. J., Serjeant, S., Dunlop, J. S., et al. 2005, *MNRAS*, **363**, 563
- Murakami, H., Baba, H., Barthel, P., et al. 2007, *PASJ*, **59**, S369
- Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, *ApJ*, **278**, L1
- Nguyen, H. T., Schulz, B., Levenson, L., et al. 2010, *A&A*, **518**, L5
- 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
- Oliver, S. J., Wang, L., Smith, A. J., et al. 2010, *A&A*, **518**, L21
- Oliver, S. J., Bock, J., Altieri, B., et al. 2012, *MNRAS*, **424**, 1614
- Oteo, I., Zwaan, M. A., Ivison, R. J., Smail, I., & Biggs, A. D. 2016, *ApJ*, **822**, 36
- Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, *MNRAS*, **456**, 3194
- Pascale, E., Ade, P. A. R., Bock, J. J., et al. 2008, *ApJ*, **681**, 400
- Patanchon, G., Ade, P. A. R., Bock, J. J., et al. 2009, *ApJ*, **707**, 1750
- Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, *A&A*, **603**, A102
- Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, *A&A*, **615**, A146
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *A&A*, **518**, L1
- Planck Collaboration XVI. 2014, *A&A*, **571**, A16
- Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, *A&A*, **523**, A13
- Puget, J.-L., Abergel, A., Bernard, J.-P., et al. 1996, *A&A*, **308**, L5
- Rodríguez-Puebla, A., Behroozi, P., Primack, J., et al. 2016, *MNRAS*, **462**, 893
- Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, *MNRAS*, **409**, 48
- Roseboom, I. G., Ivison, R. J., Greve, T. R., et al. 2012, *MNRAS*, **419**, 2758
- Rowan-Robinson, M., Oliver, S., Wang, L., et al. 2016, *MNRAS*, **461**, 1100
- Rowan-Robinson, M., Wang, L., Farrah, D., et al. 2018, *A&A*, **619**, A169
- Safarzadeh, M., Lu, Y., & Hayward, C. C. 2017, *MNRAS*, **472**, 2462
- Sanders, D. B., Mazzarella, J. M., Kim, D.-C., Surace, J. A., & Soifer, B. T. 2003, *AJ*, **126**, 1607
- Sargent, M. T., Béthermin, M., Daddi, E., & Elbaz, D. 2012, *ApJ*, **747**, L31
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, **500**, 525
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, **575**, A74
- Schreiber, C., Elbaz, D., Pannella, M., et al. 2017, *A&A*, **602**, A96
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Serra, P., Amblard, A., Temi, P., et al. 2011, *ApJ*, **740**, 22
- Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, *ApJ*, **509**, 103
- Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2015, *ApJ*, **807**, 128
- Siringo, G., Kreysa, E., Kovács, A., et al. 2009, *A&A*, **497**, 945
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, **53**, 51
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, **214**, 15
- Springel, V., Di Matteo, T., & Hernquist, L. 2005, *MNRAS*, **361**, 776
- Stan Development Team 2015a, *PyStan: The Python Interface to Stan*, Version 2.7.0
- Stan Development Team 2015b, *Stan: A C++ Library for Probability and Sampling*, Version 2.7.0
- Stach, S. M., Smail, I., Swinbank, A. M., et al. 2018, *ApJ*, **860**, 161
- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, *ApJ*, **817**, 118
- Trayford, J. W., Camps, P., Theuns, T., et al. 2017, *MNRAS*, **470**, 771
- Vaccari, M. 2016, *The Universe of Digital Sky Surveys*, 42, 71
- Vale, A., & Ostriker, J. P. 2004, *MNRAS*, **353**, 189
- van Kampen, E., Percival, W. J., Crawford, M., et al. 2005, *MNRAS*, **359**, 469
- Valiante, E., Smith, M. W. L., Eales, S., et al. 2016, *MNRAS*, **462**, 3146
- Viero, M. P., Moncelsi, L., Quadri, R. F., et al. 2013, *ApJ*, **779**, 32
- Wang, L., & Rowan-Robinson, M. 2010, *MNRAS*, **401**, 35
- Wang, L., Farrah, D., Oliver, S. J., et al. 2013, *MNRAS*, **431**, 648
- Wang, L., Viero, M., Clarke, C., et al. 2014, *MNRAS*, **444**, 2870
- Wang, L., Norberg, P., Béthermin, M., et al. 2016, *A&A*, **592**, L5
- Wang, W.-H., Lin, W.-C., Lim, C.-F., et al. 2017, *ApJ*, **850**, 37
- Weiβ, A., Kovács, A., Coppin, K., et al. 2009, *ApJ*, **707**, 1201
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, **154**, 1
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, *ApJ*, **754**, L29
- Wootten, A., & Thompson, A. R. 2009, *IEEE Proceedings*, **97**, 1463
- Zavala, J. A., Aretxaga, I., Geach, J. E., et al. 2017, *MNRAS*, **464**, 3369
- Zubko, V., Dwek, E., & Arendt, R. G. 2004, *ApJS*, **152**, 211

## Appendix A: Number counts and LF measurements

We provide our measurements of number counts in COSMOS at the *Herschel*-SPIRE wavelengths (250, 350, and 500  $\mu\text{m}$ ) in Tables A.1–A.3. The uncertainties on the number counts only account for Poisson errors. At the bright end, we expect field-to-field variations to be a larger source of uncertainty. In Tables A.4 and A.5, we provide number counts in COSMOS at 450 and 870  $\mu\text{m}$ . The 450  $\mu\text{m}$  counts are derived from the de-blended 500  $\mu\text{m}$  flux densities after applying a scaling factor of the 450–500  $\mu\text{m}$  flux ratio. The 870  $\mu\text{m}$  counts are derived from the predicted 870  $\mu\text{m}$  flux densities based on the best-fit SED. For more details, we refer to Sects. 3.1.2 and 3.1.3. In Table A.6, we list our measurements of the total IR LF in 13 redshift bins.

**Table A.1.** Our measurement of the number counts at 250  $\mu\text{m}$  in COSMOS.

$S$ (mJy)	$S^{2.5}dN/dS$ ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )	Poisson error ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )
0.95	5247.40	15.8654
1.56	8670.32	29.6727
2.58	11872.1	50.5201
4.25	15274.6	83.3767
7.01	19670.4	137.666
11.56	21567.6	209.740
19.06	20242.8	295.650
31.42	13876.4	356.156
51.80	7198.93	373.247
85.41	3277.45	366.430
140.81	693.836	245.308

**Notes.** The flux density  $S$  is the centre of the bin. Uncertainties on the counts only represent Poisson errors.

**Table A.2.** Our measurement of the number counts at 350  $\mu\text{m}$  in COSMOS.

$S$ (mJy)	$S^{2.5}dN/dS$ ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )	Poisson error ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )
0.95	5547.49	16.3128
1.56	8550.90	29.4676
2.58	10501.0	47.5132
4.25	11935.8	73.7033
7.01	13523.4	114.147
11.56	13996.3	168.962
19.06	10061.0	208.431
31.42	5576.15	225.772
51.80	1199.82	152.378
85.41	163.872	81.9362
140.81	173.459	122.654

**Notes.** The flux density  $S$  is the centre of the bin. Uncertainties on the counts only represent Poisson errors.

**Table A.3.** Our measurement of the number counts at 500  $\mu\text{m}$  in COSMOS.

$S$ (mJy)	$S^{2.5}dN/dS$ ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )	Poisson error ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )
0.95	5464.69	16.1906
1.56	6592.61	25.8743
2.58	6054.29	36.0771
4.25	5303.91	49.1313
7.01	4682.50	67.1676
11.56	3606.16	85.7637
19.06	2012.19	93.2130
31.42	530.191	69.6175
51.80	154.816	54.7356

**Notes.** The flux density  $S$  is the centre of the bin. Uncertainties on the counts only represent Poisson errors.

**Table A.4.** Our measurement of the number counts at 450  $\mu\text{m}$  in COSMOS.

$S$ (mJy)	$S^{2.5}dN/dS$ ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )	Poisson error ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )
0.95	55612.8	160.304
1.56	78844.2	277.717
2.58	75856.4	396.346
4.25	66137.9	538.472
7.01	59380.2	742.368
11.5	48491.2	976.092
19.06	31445.7	1143.67
31.42	10390.6	956.535
51.80	1864.15	589.496
85.41	789.281	558.106

**Notes.** The flux density  $S$  is the centre of the bin. Uncertainties on the counts only represent Poisson errors.

**Table A.5.** Our measurement of the number counts at 870  $\mu\text{m}$  in COSMOS.

$S$ (mJy)	$S^{2.5}dN/dS$ ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )	Poisson error ( $\text{Jy}^{1.5} \text{sr}^{-1}$ )
0.58	6161.7472	36.673168
0.95	7420.0241	58.554406
1.56	7352.2708	84.806301
2.58	6266.4922	113.91740
4.25	5523.9150	155.61868
7.01	4640.5271	207.53068
11.56	2121.9831	204.18791
19.06	332.75839	117.64786

**Notes.** The flux density  $S$  is the centre of the bin. Uncertainties on the counts only represent Poisson errors.



**Table A.6.** Our measurement of the total IR LF in COSMOS.

Redshift bin	$\log L_{\text{IR}}(L_{\odot})$	LF ( $\text{Mpc}^{-3} \text{dex}^{-1}$ )	Poisson error ( $\text{Mpc}^{-3} \text{dex}^{-1}$ )
$0.0 < z < 0.3$	9.75	0.0084	0.0002
$0.0 < z < 0.3$	10.25	0.0045	0.0002
$0.0 < z < 0.3$	10.75	0.0012	8.4947e-05
$0.0 < z < 0.3$	11.25	0.0002	3.2696e-05
$0.0 < z < 0.3$	11.75	5.7799e-06	5.7799e-06
$0.3 < z < 0.45$	9.75	0.0109	0.0002
$0.3 < z < 0.45$	10.25	0.0094	0.0002
$0.3 < z < 0.45$	10.75	0.0043	0.0001
$0.3 < z < 0.45$	11.25	0.0007	4.5427e-05
$0.3 < z < 0.45$	11.75	2.3124e-05	8.1754e-06
$0.45 < z < 0.6$	10.25	0.0086	0.0001
$0.45 < z < 0.6$	10.75	0.0044	8.7627e-05
$0.45 < z < 0.6$	11.25	0.0012	4.5986e-05
$0.45 < z < 0.6$	11.75	8.5720e-05	1.2246e-05
$0.45 < z < 0.6$	12.25	3.4988e-06	2.4740e-06
$0.6 < z < 0.8$	10.25	0.0095	9.2757e-05
$0.6 < z < 0.8$	10.75	0.0064	7.5702e-05
$0.6 < z < 0.8$	11.25	0.0016	3.7752e-05
$0.6 < z < 0.8$	11.75	0.0002	1.4314e-05
$0.6 < z < 0.8$	12.25	1.1722e-05	3.2510e-06
$0.6 < z < 0.8$	12.75	9.0166e-07	9.0166e-07
$0.8 < z < 1.0$	10.75	0.0091	7.9075e-05
$0.8 < z < 1.0$	11.25	0.0038	5.1005e-05
$0.8 < z < 1.0$	11.75	0.0007	2.1651e-05
$0.8 < z < 1.0$	12.25	4.1722e-05	5.3420e-06
$1.0 < z < 1.2$	10.75	0.0070	6.3435e-05
$1.0 < z < 1.2$	11.25	0.0038	4.6495e-05
$1.0 < z < 1.2$	11.75	0.0007	2.0429e-05
$1.0 < z < 1.2$	12.25	5.9408e-05	5.8254e-06
$1.0 < z < 1.2$	12.75	5.7123e-07	5.7123e-07
$1.0 < z < 1.2$	13.25	5.7123e-07	5.7123e-07
$1.2 < z < 1.7$	11.25	0.0036	2.6300e-05
$1.2 < z < 1.7$	11.75	0.0010	1.3871e-05
$1.2 < z < 1.7$	12.25	0.0001	4.6736e-06
$1.2 < z < 1.7$	12.75	3.2517e-06	7.8865e-07
$1.7 < z < 2.0$	11.25	0.0026	2.7386e-05
$1.7 < z < 2.0$	11.75	0.0011	1.7924e-05
$1.7 < z < 2.0$	12.25	0.0002	7.2243e-06
$1.7 < z < 2.0$	12.75	1.4403e-05	2.0368e-06
$1.7 < z < 2.0$	13.25	2.8805e-07	2.8805e-07
$2.0 < z < 2.5$	11.25	0.0015	1.5965e-05
$2.0 < z < 2.5$	11.75	0.0007	1.1118e-05
$2.0 < z < 2.5$	12.25	0.0002	5.4133e-06
$2.0 < z < 2.5$	12.75	2.1059e-05	1.8761e-06
$2.0 < z < 2.5$	13.25	5.0141e-07	2.8949e-07
$2.5 < z < 3.0$	11.75	0.0008	1.1329e-05
$2.5 < z < 3.0$	12.25	0.0002	5.7064e-06
$2.5 < z < 3.0$	12.75	2.4608e-05	2.0296e-06
$2.5 < z < 3.0$	13.25	5.0220e-07	2.8995e-07
$3.0 < z < 4.2$	11.75	0.0004	5.2906e-06
$3.0 < z < 4.2$	12.25	8.6658e-05	2.5324e-06
$3.0 < z < 4.2$	12.75	3.4042e-06	5.0192e-07
$4.2 < z < 5.0$	12.25	7.1666e-05	2.9656e-06
$4.2 < z < 5.0$	12.75	2.5525e-05	1.7698e-06
$4.2 < z < 5.0$	13.25	3.6815e-07	2.1255e-07
$5.0 < z < 6.0$	12.25	4.7607e-05	2.2721e-06
$5.0 < z < 6.0$	12.75	1.7026e-05	1.3588e-06
$5.0 < z < 6.0$	13.25	9.7599e-07	3.2533e-07

**Notes.** The total IR luminosity  $\log L_{\text{IR}}$  is the centre of the bin. Uncertainties on the counts only represent Poisson errors.