MNRAS 000, 1–25 (2016) Preprint 18 July 2016 Compiled using MNRAS LATEX style file v3.0

Evolution of cosmic star formation in the SCUBA-2 Cosmology Legacy Survey

N. Bourne,[1⋆] J. S. Dunlop,[1] E. Merlin,[2] S. Parsa,[1] C. Schreiber,[3] M. Castellano,[2] C. J. Conselice,[4] K. E. K. Coppin,[5] D. Farrah,[6] A. Fontana,[2] J. E. Geach,[5] M. Halpern,[7] K. K. Knudsen,[8] M. J. Michałowski,[1] A. Mortlock,[1] P. Santini,[2] D. Scott,[7] X. W. Shu,[9] C. Simpson,[10] J. M. Simpson,[1] D. J. B. Smith,[5] P. P. van der Werf[3]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[2]*INAF-Osservatorio Astronomico di Roma, via di Frascati 33, I-00078 Monte Porzio Catone, Roma, Italy*
[3]*Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands*
[4]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[5]*Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, Hatfield AL10 9AB, UK*
[6]*Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA*
[7]*Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada*
[8]*Department of Earth and Space Sciences, Chalmers University of Technology, Onsala Space Observatory, SE-43992 Onsala, Sweden*
[9]*Department of Physics, Anhui Normal University, Wuhu, Anhui, 241000, China*
[10]*Gemini Observatory, Northern Operations Center, 670 N. A'ohōkū Place, Hilo, HI 96720-2700, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We present a new exploration of the cosmic star-formation history and dust obscuration in massive galaxies at redshifts $0.5 < z < 6$. We utilize the deepest 450 and 850μm imaging from SCUBA-2 CLS, covering 230 arcmin^2 in the AEGIS, COSMOS and UDS fields, together with 100–250μm imaging from *Herschel*. We demonstrate the capability of the T-PHOT deconfusion code to reach below the confusion limit, using multi-wavelength prior catalogues from CANDELS/3D-HST. By combining IR and UV data, we measure the relationship between total star-formation rate (SFR) and stellar mass up to $z \sim 5$, indicating that UV-derived dust corrections underestimate the SFR in massive galaxies. We investigate the relationship between obscuration and the UV slope (the IRX–β relation) in our sample, which is similar to that of low-redshift starburst galaxies, although it deviates at high stellar masses. Our data provide new measurements of the total SFR density (SFRD) in $M_* > 10^{10}$M$_\odot$ galaxies at $0.5 < z < 6$. This is dominated by obscured star formation by a factor of > 10. One third of this is accounted for by 450μm-detected sources, while one fifth is attributed to UV-luminous sources (brighter than $L^\star_{\rm UV}$), although even these are largely obscured. By extrapolating our results to include all stellar masses, we estimate a total SFRD that is in good agreement with previous results from IR and UV data at $z \lesssim 3$, and from UV-only data at $z \sim 5$. The cosmic star-formation history undergoes a transition at $z \sim 3-4$, as predominantly unobscured growth in the early Universe is overtaken by obscured star formation, driven by the build-up of the most massive galaxies during the peak of cosmic assembly.

Key words: methods: statistical – galaxies: high-redshift – submillimeter: galaxies – submillimeter: diffuse background

1 INTRODUCTION

A key element in understanding the evolution of galaxies and the build-up of the present-day population is the cosmic star-formation history, i.e. the overall comoving volume-density

⋆ nbourne22@gmail.com

of the star-formation rate (SFR) within galaxies throughout the Universe, measured as a function of look-back time. This has been observationally determined from ultraviolet (UV) emission from star-forming galaxies up to $z \sim 9$, within the first few hundred Myr of the Universe (Bouwens et al. 2007; Reddy et al. 2008; Cucciati et al. 2012; McLure et al. 2013; Duncan et al. 2014; Bouwens et al. 2015; Mashian et al. 2015; McLeod et al. 2015, 2016; Parsa et al. 2016). However, it is well known that, over most of cosmic history, the great majority of the UV radiation from young stars is absorbed by dust within galaxies and is thermally re-radiated in the far-infrared (FIR; Desert et al. 1990). Unobscured UV observations can be corrected for this deficit using empirical calibrations based on the slope of the spectral energy distribution (SED) in the UV, but the effectiveness of these assumptions at high redshifts is unclear due to potential variations in the intrinsic UV slope and the dust attenuation curve (Gonzalez-Perez et al. 2013; Castellano et al. 2014; Heinis et al. 2014; Price et al. 2014; Yajima et al. 2014; Capak et al. 2015; Coppin et al. 2015; Pannella et al. 2015; Zeimann et al. 2015; Talia et al. 2015; Watson et al. 2015). Discrepancies between observations and the most up-to-date hydrodynamical simulations and models suggest that either there must be systematic errors in the observational results, or the models do not adequately describe properties such as the dust attenuation curve (Somerville et al. 2012; Kobayashi et al. 2013; Genel et al. 2014; Furlong et al. 2015). At high redshifts we lack direct measurements of the obscuration fraction in typical star-forming galaxies, and therefore generally assume a theoretical relationship linking it to the UV slope, which is calibrated on low-redshift samples (Meurer et al. 1999; Calzetti et al. 2000; Kong et al. 2004; Boissier et al. 2007; Muñoz-Mateos et al. 2009; Overzier et al. 2011; Hao et al. 2011; Takeuchi et al. 2012). This relationship depends on the UV attenuation curve and metallicity, and remains poorly constrained at high-redshifts, in spite of extensive efforts using *Herschel* and *Spitzer* (Buat et al. 2012; Reddy et al. 2012; Burgarella et al. 2013; Heinis et al. 2014; Price et al. 2014).

The chief problem is the difficulty of measuring accurate rest-frame FIR emission in representative samples of star-forming galaxies at $z > 3$, as a result of the high confusion noise in sub-millimetre (submm) surveys with moderate-sized, single-dish telescopes such as *Herschel* (Reddy et al. 2012; Burgarella et al. 2013; Gruppioni et al. 2013; Heinis et al. 2014). Interferometric imaging surveys of deep fields (e.g. with ALMA) offer a high-resolution alternative to single-dish surveys (Bouwens et al. 2016; Kohno et al. 2016; Hatsukade et al. 2016; Dunlop et al. 2016), but these only probe relatively small volumes which have limited statistical power.

Apart from biased samples of bright Lyman-alpha emitters and submm-selected starbursts (which are rare and therefore unrepresentative of the overall star-forming-galaxy population), the only significant source of information on the cosmic SFR at early times comes from samples of photometrically selected Lyman-break galaxies (LBGs). These samples can be used to measure the rest-frame UV luminosity function (LF), from which the total SFR density (SFRD) can be extrapolated by integrating the LF model and applying (potentially large) corrections for dust obscuration. The obscuration of star formation in $z > 3$ LBGs has been studied in the submm via stacking of SCUBA-2 and *Herschel* data by Coppin et al. (2015), and individually using ALMA and the Plateau de Bure Interferometer (e.g. Schaerer et al. 2015; Capak et al. 2015; Bouwens et al. 2016; and see also Chapman et al. 2000; Peacock et al. 2000; Stanway et al. 2010; Davies et al. 2012, 2013), but the question of whether low-redshift calibrations hold true at high redshifts remains open.

The cosmic star-formation history can also be explored through measurements of the average specific SFR (SSFR = SFR/stellar mass) of galaxy samples. This is known to increase with look-back time, but may plateau or rise more slowly at $z > 3$ (Madau & Dickinson 2014). Meanwhile, stellar mass density is constrained out to high redshifts with smaller systematic errors due to the reduced dust extinction effects in the rest-frame near-IR (Ilbert et al. 2013; Muzzin et al. 2013; Grazian et al. 2015). This must approximately trace the integral of total SFRD over all masses and times up to a given look-back time, so we can expect to see the rate of growth of stellar mass trace the cosmic SFR evolution (Wilkins et al. 2008; Madau & Dickinson 2014).

It is clear that the vast majority of the SFR in the Universe is obscured by dust, and this fraction appears to increase from $z = 0$ to $z = 1$. However, the behaviour at higher redshifts is uncertain (Burgarella et al. 2013). There are systematic uncertainties in the true behaviour of the dust-obscured (hence total) SFRD at high redshifts due to uncertainties about the nature of star-forming galaxies at high redshifts. Typical star-forming galaxies at high redshifts have higher SSFRs than their counterparts at low redshifts, so it is unclear whether they resemble their low-redshift counterparts (which might be implied by the existence of a common mass-SFR relation known as the "main sequence"; Noeske et al. 2007) or whether they are more similar to high-SSFR galaxies at low redshifts (because they are similarly rich in dense gas and have a clumpy mass distribution; Price et al. 2014).

In summary we need to improve our knowledge of the obscured SFRD from FIR observations at $z > 3$. Currently we are limited by issues including sample bias towards unrepresentative bright objects (either from UV selection or FIR selection), and uncertainties in the stacking of fainter objects. In the FIR, the major obstacle is the low resolution of single-dish FIR/submm surveys (typically 15–35 arcsec), which limits our ability to detect and identify individual sources above the confusion limit, and also hampers stacking below the confusion limit due to the difficulty of separating the emission from heavily blended positions.

In this paper we attempt to improve this situation with a combination of three key ingredients: (i) deep, high resolution submm imaging of blank fields with JCMT/SCUBA-2 (Holland et al. 2013); (ii) rich multi-wavelength catalogues containing positions, redshift information and UV-to-mid-IR SEDs to support the submm data; (iii) the latest deconfusion techniques developed within the ASTRODEEP consortium, which allow us to maximize the useful information available from these combined data sets. In Section 2 we briefly describe the submm imaging data and the sample used in this work. In Section 3 we explain how T-PHOT is applied to measure deconfused submm photometry for the sample, and in Section 4 we describe and validate the stacking technique. Section 5 presents the results and discussion

in the context of previous literature. Throughout this paper we assume a flat ΛCDM cosmology with $\Omega_M = 0.3$, $h = H_0/100\ \mathrm{km\,s^{-1}\,Mpc^{-1}} = 0.7$. All magnitudes are in the AB system (Oke 1974; Oke & Gunn 1983) and we assume the Kroupa & Weidner (2003) initial mass function (IMF) throughout, unless otherwise stated.

2 DATA

We use data from the SCUBA-2 Cosmology Legacy Survey (S2CLS; Geach et al. 2013, 2016; Roseboom et al. 2013), which provides imaging over ∼ 5 deg^2 at 850 μm in several wide/deep survey fields, as well as ∼ 0.25 deg^2 at 450 and 850 μm in several ultra-deep fields, which benefit from multi-wavelength coverage from CANDELS (Grogin et al. 2011; Koekemoer et al. 2011).

In order to probe the SFRs of the faint and highly-confused source population at the highest redshifts we require the best possible submm resolution (from a single-dish blind survey) to minimise confusion noise, and the deepest available imaging to minimise instrumental noise. This is provided by the 450-μm imaging of the CANDELS COSMOS, UDS and AEGIS fields, covering a total of 230 arcmin^2 to depths of ∼ 1.0 mJy beam^{-1} rms (average instrumental noise in match-filtered maps) with resolution FWHM = 7.5 arcsec. Due to its high angular resolution and long wavelength, the 450-μm imaging is central to our analysis. However, we additionally benefit from SCUBA-2 850-μm imaging with rms ∼ 0.2mJy beam^{-1} (match-filtered), FWHM = 14 arcsec; *Herschel*/PACS imaging with rms ∼ 1 − 2 mJy beam^{-1} and FWHM = 9, 11 arcsec at 100 and 160 μm respectively (from PEP and HerMES; Lutz et al. 2011; Oliver et al. 2012); and *Herschel*/SPIRE imaging with rms ∼ 2 − 4 mJy beam^{-1} and FWHM = 18 arcsec at 250 μm (from HerMES). The SCUBA-2 maps have been reduced using the SMURF pipeline described by Chapin et al. (2013), which includes a band-pass filter in the time series equivalent to angular scales of 2 to 120 arcsec (for further details see Geach et al. 2013, 2016). However, the maps that we use have not been processed with a matched filter since our methodology effectively performs this filtering independently. The flux-conversion factors applied to the 450 and 850-μm data are 540 and 591 Jy beam^{-1} pW^{-1}; these are the canonical values with an additional 10 per cent correction to account for losses in the filtering procedure (Chapin et al. 2013; Dempsey et al. 2013).

We use multi-wavelength catalogues compiled by the 3D-HST team (Brammer et al. 2012; Skelton et al. 2014), which include photometry spanning the u–8μm bands from CFHTLS, Subaru, CANDELS, NMBS, WIRDS, UKIDSS-UDS, UltraVISTA, SEDS, S-COSMOS and EGS. These catalogues are derived from *HST*/WFC3 imaging and are effectively H-band selected from imaging with a median 5-σ depth of H(F160W) = 26.4 in a 1-arcsec aperture; the catalogues have 50 and 95 per cent completeness limits of H = 26.5 and 25.1 respectively (Skelton et al. 2014). We use the combined "Z_BEST" redshift data from 3D-HST (Momcheva et al. 2015) which comprise collected spectroscopic redshifts from the literature, *HST* grism redshifts where reliable, and photometric redshifts from EAZY (Brammer et al. 2008) otherwise. The spectroscopic and photometric redshifts are described by Skelton et al. (2014); EAZY photometric redshifts are based on SED fitting with a linear combination of seven templates, which consist of the five default PÉGASE stellar population synthesis (SPS) models from Fioc & Rocca-Volmerange (1997), in addition to a young, dusty template and an old, red template from Whitaker et al. (2011). Grism redshifts are described by Momcheva et al. (2015), and are based on observations with the WFC3 G141 grism covering 60 per cent of the CANDELS imaging in AEGIS, COSMOS and UDS, for photometric objects selected down to a depth of $JH = 26$ in the co-added F125W+F140W+F160W images. Redshifts were determined from fitting the 2D spectra and multi-band photometry simultaneously, using a modified version of the EAZY templates with additional emission-line templates from Dobos et al. (2012). Redshift fits to all grism spectra were visually inspected for quality, and typical redshift uncertainties are $\Delta z/(1+z) \approx 0.003$ (Momcheva et al. 2015). Photometric redshift uncertainties are typically $\Delta z/(1+z) \leq 0.03$ (and ≤ 0.01 in COSMOS due to good medium-band coverage), with fewer than 5 per cent significant outliers in all fields (Skelton et al. 2014). Our final sample (see Section 3.1) consists of 6 per cent spectroscopic redshifts, 49 per cent grism redshifts and 45 per cent photometric redshifts. Most of the redshifts at $z > 3$ are photometric, but this subset also has typical uncertainties ≤ 0.03.

Skelton et al. (2014) also provide stellar-population parameters derived from SED fitting with FAST (Kriek et al. 2009). These are based on Bruzual & Charlot (2003) SPS models with a Chabrier (2003) IMF;[1] solar metallicity; exponentially-declining star-formation histories with a minimum e-folding time of 10^7 yr; minimum age of 40 Myr; $0 < A_V < 4$ and a Calzetti et al. (2000) dust attenuation law. We use the stellar masses from these SED fits (which are well-constrained by the available optical-near-IR photometry; Skelton et al. 2014), and we use $M_{\rm UV} = M_{1600\text{Å}}$ (i.e. *FUV*-band) estimated from the rest-frame SED fit from EAZY (Skelton et al. 2014). At this wavelength, $M_{\rm UV}$ traces emission from stellar populations with a mean age of 10^7yr (Kennicutt & Evans 2012). More detailed sample selection and binning is described in Section 3.1.

3 METHODS

Many different algorithms have been developed to solve the problem of deconfusing low-resolution imaging using prior information from high-resolution surveys (Béthermin et al. 2010; Kurczynski & Gawiser 2010; Roseboom et al. 2010; Bourne et al. 2012; Viero et al. 2013; Wang et al. 2014; MacKenzie et al. 2015; Safarzadeh et al. 2015; Hurley et al. 2016; Wright et al. 2016). In this study we use T-PHOT (Merlin et al. 2015, 2016)[2] to fit submm low-resolution images

[1] We make no adjustment for the difference between Chabrier (2003) and Kroupa & Weidner (2003) IMFs, which are essentially identical for our purposes (Chomiuk & Povich 2011).

[2] T-PHOT is a multi-purpose deconfusion code applicable to multi-wavelength data sets with a range of angular resolutions; it was developed within the ASTRODEEP consortium and is available from `http://astrodeep.eu`

(LRI) with positional priors from higher-resolution data-sets. T-PHOT is a versatile tool which can be applied to a wide range of problems as described by Merlin et al. (2015). In general it provides a fast and efficient algorithm for measuring deblended photometry in low-resolution imaging based on prior information from higher-resolution data-sets, which may consist of high-resolution images, parametric light-profile models, and/or simply positional priors, depending on the particular data-set in question. The specific application of T-PHOT to confusion-limited, unresolved submm images such as those from SCUBA-2 involves a number of elements which differ significantly from applications in less confused optical-to-mid-IR images. These can be broken down into the following four-stage procedure:

(i) **Selection:** select a suitable source catalogue to use for the positional priors.
(ii) **Input:** provide as inputs the list of prior positions, the LRI and associated rms noise map, and the PSF.
(iii) **Optimization:** construct a model of the entire LRI consisting of scaled point sources at every prior position, and obtain a set of flux measurements (i.e. the best-fitting normalisation for each point source) that minimizes χ^2 between this model and the LRI.
(iv) **Background:** measure the overall background level of the LRI, and account for this in the flux measurements of sources. The latest version of T-PHOT (Merlin et al. 2016) fits the background as a free parameter in the model. We investigate the accuracy of the background measurement and whether any further refinement is required.

To fully explain the application of T-PHOT to confused submm images with dense prior lists, we expand on of each of these steps in the following subsections (3.1–3.4).

3.1 Selection

The definition of the prior catalogue is of great importance when attempting to model a confused map by χ^2 minimisation. However, obtaining a formally good fit does not guarantee accurate estimates for the fluxes or their associated errors or covariances: this relies on an appropriate set of priors being used. For example, if an object which is bright in the LRI is not represented in the priors then the fluxes of any nearby priors will be over-estimated due to blending with the additional bright object not being accounted for, and this systematic error will not be included in the covariance matrix (see Merlin et al. 2015).

We select priors from the 3D-HST parent catalogue (which is effectively H-band selected – see Section 2) by first imposing limits in AB magnitude of $K_s < 24$ or IRAC [3.6]< 24, which are chosen to maximise the completeness of massive galaxies at $z < 6$ (we analyse the completeness in Section 3.5). We include only galaxies with the "USE" photometric flag from 3D-HST, which ensures sufficient photometry for a photometric redshift fit. The USE flag excludes stars; objects whose photometry may be affected by nearby bright stars; objects which don't have at least two individual exposures in F125W or F160W; objects with F160W signal-to-noise ratio ≤ 3; objects with catastrophic photometric-redshift fits ($\chi^2 \geq 1000$) or catastrophic stellar-population fits ($\log M_* < 0$). The fraction of objects in the 3D-HST catalogues of AEGIS, COSMOS and UDS which have the USE flag is 87 per cent, and this fraction is essentially independent of magnitude at F160W< 26.5.

To avoid excessive crowding of the priors at lower redshifts we also impose a cut on stellar mass of $\log(M_*/\mathrm{M}_\odot) > 9$. In this way we deblend only massive galaxies in the model; galaxies of lower mass become part of the background to be subtracted (see Section 3.4). By restricting the priors in this way, we implicitly assume that any population of objects that contributes to the map, but is missing from the priors, is not spatially correlated with the priors. Clearly this assumption is flawed when considering that low-mass galaxies will be clustered around high-mass galaxies, but we can justify this decision as follows. The contribution of $\log(M_*/\mathrm{M}_\odot) < 9$ galaxies to the 450-μm sky (i.e. the cosmic infrared background, CIB) will be relatively small compared to higher-mass galaxies, assuming (i) that 450-μm flux is roughly proportional to SFR at $z < 6$ (Blain et al. 2002),[3] (ii) that SFR, in particular obscured SFR, is roughly proportional to stellar mass (e.g. Noeske et al. 2007; Elbaz et al. 2011), and (iii) that most of the stellar mass resides in galaxies more massive than $10^9\mathrm{M}_\odot$ (Kauffmann et al. 2003; Kajisawa et al. 2009; Marchesini et al. 2009; Mortlock et al. 2011). The stellar mass threshold ensures that any bias incurred from this assumption will be roughly invariant with redshift, whereas if we used only a magnitude cut the bias would increase as the mass cut rises with increasing redshift. In order to test for potential bias as a result of the mass limit, we tested the effect of using a higher limit of $\log(M_*/\mathrm{M}_\odot) = 9.5$. If there is a significant bias due to flux from lower-mass galaxies (outside of the sample) being attributed to galaxies in the sample, that bias would be larger when we increase the log mass limit from 9.0 to 9.5. In fact, we find that the results obtained with the two different mass limits are fully consistent with each other, and all of the measured trends that we discuss are robust.

Our final sample consists of 8809 positional priors within an area of 230 arcmin^2 over the AEGIS, COSMOS and UDS fields.

3.2 Input

The inputs for T-PHOT are specified within the parameter file as described by Merlin et al. (2015). In this case we provide the 450-μm image of each field (the LRI), along with the rms instrumental noise map from the SCUBA-2 pipeline (Chapin et al. 2013), and the PSF. The data have been band-pass-filtered as part of the map-making procedure in order to remove large-scale background variations, and the maps have zero mean (Geach et al. 2013; Chapin et al. 2013). We use images that are not PSF-filtered or match-filtered because the T-PHOT algorithm itself effectively filters the image and deconfuses all sources, so there is no benefit to using additional filtering. The PSF of these pre-matched-filter images is therefore assumed to be a symmetrical Gaussian function with FWHM=7.5 arcsec (Chapin et al. 2013). The prior cat-

[3] In fact there is a redshift-dependence, but it is weak at $1 < z < 4$, and the only lower-mass galaxies with significant 450-μm emission would have to be either at low redshift, or have extremely high SFR/M_* ratios.

alogue is also provided as a list of x, y positions relative to the image.

The 450-μm data provide the backbone of our analysis, but in order to characterise FIR/submm SEDs we also apply T-PHOT to images at 100, 160, 250 and 850μm. We assume PSFS with FWHM of 9, 11, 18 and 14″ in each of these bands respectively. The larger beam sizes in these lower-resolution images means that confusion-related uncertainties are much larger, as a result of the far greater degeneracies between highly blended galaxies. However, the uncertainties output by T-PHOT account for the covariances between blended galaxies, and therefore the output measurements can be successfully combined to constrain the average properties of samples at these wavelengths, as described in Section 4.5.

When fitting confused maps in this way, it is of vital importance that the map is appropriately masked, since any sources lying outside the region covered by the priors will not be included in the model and can easily dominate the residual, leading to degenerate χ^2 values. We therefore ensure that the prior catalogues and images are matched, by masking the outer parts of the image that are at least 10 pixels (approximately 3×FWHM of the PSF) outside the area covered by the prior catalogue.

3.3 Optimization

We run T-PHOT using the recommended options for unresolved priors: we use a single pass, since the precise astrometric positions of the priors cannot be improved by allowing them to shift in the "dance" stage;[4] we fit the entire image with a single model rather than dividing it into cells; and we allow T-PHOT to fit a single-valued sky background as a free parameter (see Section 3.4). An example of a T-PHOT parameter file used in this analysis is provided in Appendix B. Once the best-fitting model has been obtained, T-PHOT outputs a model image comprising a "collage" of all the priors with their best-fitting normalizations (flux densities), a residual given by the difference LRI − model, a catalogue of best-fitting parameters for each object, and the covariance matrix of the best fit. The fluxes in the catalogue are background-subtracted, while the model and residual can be background-subtracted using the output background value in the T-PHOT log file.

3.4 Background

The sky background (i.e. any signal that remains in the image after the individual sources under consideration have been subtracted) is an important property of the image, especially when we want to measure very faint sources close to and below the confusion noise level. The current version of T-PHOT includes the background as a free parameter in the model, and in this section we explore the accuracy and uncertainty of this fit.

Background subtraction in general is a problem for confusion-limited maps, which have no empty regions of sky in which to measure the background. For this reason, such maps are usually set to have zero mean, such that positive bright sources and over-dense regions are balanced by negative surface brightness in regions where the source density is low. In order to obtain accurate flux densities for both bright and faint sources it is necessary to ensure that the background "behind" all the sources of interest is zero. After removal of large-scale background variations such as foreground cirrus, one can assume that a confusion-limited map consists of two separate populations of point sources: the ones we wish to fit, which are part of the model; and those we do not fit, which constitute the background. It is therefore generally true that the background is a function of the map itself (particularly the source density and beam size), and also a function of the source population that one is measuring.

One cannot fit all of the sources in the map because this would mean a greater number of degrees of freedom in the model than the number of independent data points (beams) in the data. However, one can treat the net contribution of

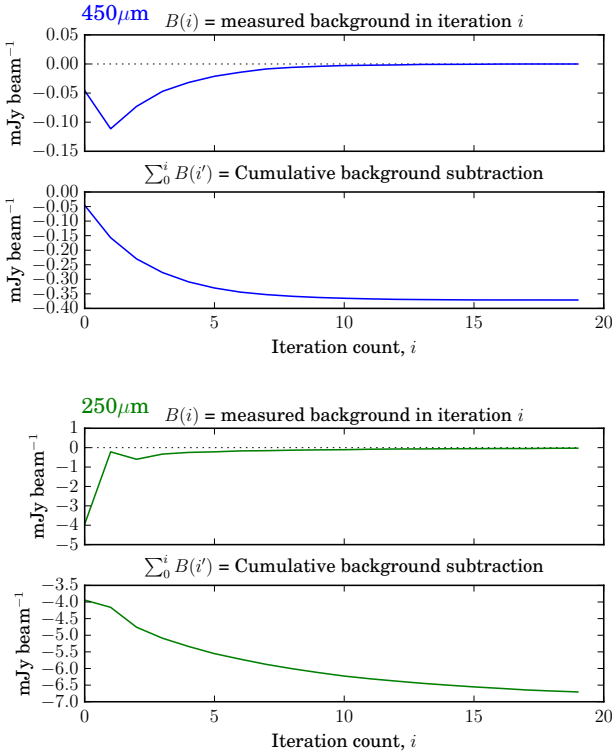

Figure 1. Demonstration of convergence on the background subtraction by iterating over T-PHOT. In each iteration, T-PHOT is run with the background fixed at zero; instead of fitting for the background, it is measured from the modal pixel value in the residual from the previous iteration, and is subtracted from the LRI before running T-PHOT. Each pair of panels show the background value B measured after each iteration, and the cumulative background value subtracted from the LRI (the sum of B over all previous iterations). Results are shown for the 450-μm and the 250-μm LRIs. This technique gives results consistent with a single run of T-PHOT in which the background is treated as a free parameter.

[4] The "dance" stage allows for precise re-registration of image priors to account for astrometric shifts between bands, and would involve a second pass of the optimization routine to improve the photometry (Merlin et al. 2015).

the background sources (which are not part of the priors) as a constant background level, assuming that the variations in their surface density are uncorrelated with the sources that are in the model. This is effectively what is assumed in the T-PHOT background-fitting model.

As an alternative to fitting for the background as a free parameter, we can run T-PHOT with the background fixed to be zero (i.e. making the assumption that the LRI is already background-subtracted). If the background is in fact not zero, then this should result in a residual with zero mean but some skewness, and the model fluxes will be biased to balance the non-zero background. We can then iterate the following steps to converge on the best estimate for the background:

(i) estimate the background level from the mode of the residual of the previous run (for the first run, the mode of the LRI);
(ii) subtract this background from the LRI used in the previous run;
(iii) run T-PHOT on the new background-subtracted LRI and output a new residual.

Rather than using the mean, we estimate the mode from the centre of a Gaussian function fitted to the 2.5-σ-clipped histogram of pixel values in the residual map, because this is less susceptible to bias from the few bright sources that may not have been included in the model. Convergence on the background value is generally achieved within five iterations with the 450-μm LRI (and 20 iterations with 250-μm LRI, which has much lower resolution), as shown in Fig. 1. If this did not converge it would indicate that too many priors have been included in the model and the fit is unstable. The independent results of this iterative method generally agree reasonably well with the background obtained from a single run of T-PHOT with the background as a free parameter; the difference is $\lesssim 0.1$ mJy beam^{-1} using the prior catalogues described above. This improves significantly when using less dense prior lists; for example using a *Spitzer* 24-μm-selected prior list results in agreement within 0.01 mJy beam^{-1}.

We also ran source-injection simulations in order to estimate any remaining background not captured by T-PHOT's internal background fitting. This method provides an estimate of the average difference between output and input fluxes of simulated sources added to the LRI. We simulated 1000 realizations, in each case adding a single point source at a random position in the LRI, and appending its coordinates to the prior catalogue of one of our fields. The fluxes of the simulated sources were randomly drawn from a distribution that was uniform in log(flux) between 0.01 and 10 mJy. We then ran T-PHOT and allowed it to fit the background as a free parameter. The statistics of output – input flux (shown in Fig. 2) are approximately Gaussian with some skewness, because most injected sources fall in faint regions of the map but a few fall on top of bright sources. There is no dependence on the input flux. The median and the 2.5-σ-clipped mean are both −0.27 mJy, but the overall mean is +0.15 mJy. The mean is positively biased by objects with large covariance because they were injected close to existing bright sources in the map. However, the inverse-variance-weighted mean of output–input flux is very close to zero (within 0.05 mJy), indicating that the T-PHOT background subtraction is adequate if we measure the fluxes of our samples using the weighted average.

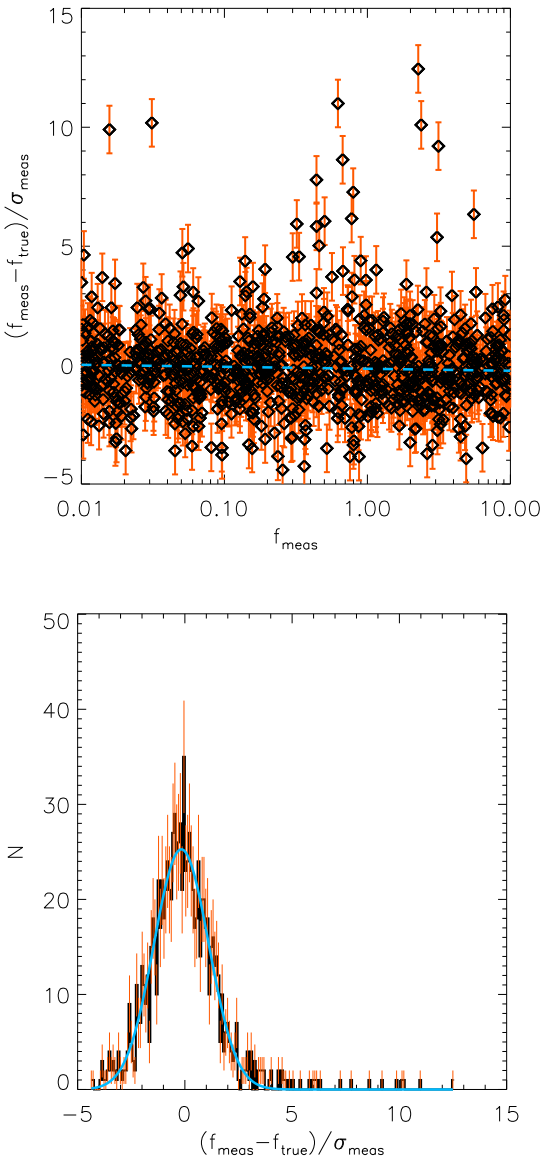

Figure 2. Results of source-injection simulations to test the accuracy of T-PHOT background-subtraction. Top: the output–input flux-density difference (normalized by output error) as a function of input flux density. The linear least-squares fit shown by the dashed line is $y = (0.01\pm0.06)-(0.03\pm0.02)x$. Bottom: the histogram of flux-density offsets, with mean of 0.15 ± 0.18 and weighted mean of -0.02 ± 0.05. The solid line shows a least-squares Gaussian fit with mean of -0.17 ± 0.03 and width of 1.28 ± 0.03. The simulation results indicate that output flux densities are unbiased at all flux levels.

3.5 Completeness of the sample

There are several causes of potential incompleteness in our sample, resulting from the following levels of sampling:

(i) the parent sample used to compile the 3D-HST catalogue;
(ii) the magnitude-limited sample within the 3D-HST catalogue;
(iii) the fact that not all objects in the magnitude-limited sample will have reliable stellar mass and photo-z estimates.

These have the potential to impact on our results by introducing unwanted bias into the sample, as a function of redshift and/or stellar mass. We address the potential for bias in each stage of sampling below.

The 3D-HST parent sample originates from *HST*/WFC3 imaging in AEGIS, COSMOS and UDS, and is essentially H-band limited with 50-per-cent completeness at F160W = 26.5 (Skelton et al. 2014). This is extremely deep, but is potentially biased against objects that are faint in the rest-frame UV at $z > 4$ (Chen et al. 2015). The additional constraints of $K_s < 24$ or $[3.6] < 24$ effectively select on stellar mass over the full redshift range. These combined criteria are complete to a stellar mass limit of $10^{10}\mathrm{M}_{\odot}$ or lower at $z < 4$, but will introduce further incompleteness at higher redshifts. We can estimate this incompleteness using realistic simulated galaxy catalogues from Schreiber et al. (2016), as shown in Fig. 3. These show that roughly 95 per cent of $M_* > 10^{10}\mathrm{M}_{\odot}$ galaxies at $4 < z < 6$ have F160W < 26.5, at which limit the parent catalogue is approximately 50 per cent complete (Skelton et al. 2014). The median magnitude of this sample is F160W $\approx$ 25, at which magnitude the catalogue is approximately 90 per cent complete. Fig. 3 also shows that the fraction of simulated $M_* > 10^{10}\mathrm{M}_{\odot}$ galaxies at $4 < z < 6$ that satisfy the criterion $K_s < 24$ is only 33 per cent, indicating severe incompleteness; however, the addition of the $[3.6] < 24$ criterion increases this to 84 per cent.[5] We therefore conclude that our sample is sufficiently ($\gtrsim 80$ per cent) complete with respect to massive galaxies ($> 10^{10}\mathrm{M}_{\odot}$) in the redshift range of interest ($0 < z < 6$).

We also consider the potential incompleteness of the subset of galaxies in the sample that have reliable photometric redshifts and stellar mass estimates from 3D-HST. Skelton et al. (2014) state that reliable measurements can be ensured by combining a magnitude cut (such as our K_s/IRAC criteria) with the USE flag (a flag indicating that photometry is sufficiently reliable to use for SED-fitting; see Section 3.1). We therefore require the USE flag for our sample, which does not exclude a large fraction of sources even at the magnitude limit of 24. The fraction of sources with $K_s < 24$ or $[3.6] < 24$ that have USE = 1 is 83 per cent. We cannot break this down by redshift or mass since these quantities are unreliable for sources with USE = 0; however, we note that the fraction does not decrease for fainter magnitudes, so we do not expect this incompleteness to be worse at $z > 4$.

Fig. 4 shows the distribution of galaxies in our sample, and the estimated completeness of the sample, as a function of redshift and stellar mass. Above a stellar mass limit of $10^{10}\mathrm{M}_{\odot}$, the sample is greater than 90 per cent at $z < 3$, and roughly 80 per cent at $3 < z < 6$. We therefore conclude that our results are unlikely to be affected significantly by incompleteness.

[5] This is slightly better than the estimated 70-per-cent completeness at $[3.6] < 24$ from S-CANDELS (Ashby et al. 2015), but is broadly consistent.

4 ANALYSIS OF T-PHOT OUTPUTS

4.1 Flux density and error estimates

The outputs from T-PHOT provide estimates for 450-μm flux densities and errors associated with each prior position. The T-PHOT error estimates are derived from the full covariance matrix and therefore account for all confusion between galaxies with $M_* > 10^9\mathrm{M}_{\odot}$ that meet the magnitude limits of the selection. Furthermore, we measure an additional error from the rms of the T-PHOT residual map, which represents the residual confusion noise resulting from objects missing from the prior list. This is added in quadrature to the T-PHOT error estimate to form the full uncertainty on each of our flux measurements, which is used for defining signal-to-noise ratios of detections. A third source of error is the systematic flux calibration uncertainty of the instrument. Since this represents a systematic offset for all flux measurements, it does not contribute to errors between measurements made in the same waveband or instrument, but it does contribute to uncertainties between measurements from different instruments, and we consider this in Section 4.5.

4.2 450μm-detected sources

The output catalogue contains sources detected down to a limit of approximately 3 mJy at a signal/noise ratio $S/N = 3$, similar to the flux limit of blind catalogues created from the same images (e.g. Roseboom et al. 2013). By using the prior catalogue we can unambiguously associate each 450-μm source with the supporting multi-wavelength photometry and SED-fitting information. We detect 165 objects from our full prior list, 130 of which have $M_* > 10^{10}\mathrm{M}_{\odot}$, and 66 per cent of which have spectroscopic or grism redshifts (the rest have photometric redshifts). The T-PHOT residuals show that there are no remaining significant 450-μm sources missing from our priors. We can therefore make a reliable estimate of the redshift distribution of 450-μm sources, as shown in Fig. 5. The median redshift of our sample is 1.68. It is evident that sources above a fixed flux limit have a broad redshift distribution between $0.5 \lesssim z \lesssim 3$, as a result of the negative k-correction. At higher redshifts, the detection rate drops as the 450-μm band probes rest-frame wavelengths blue-ward of the SED peak of star-forming galaxies, at which point the k-correction is less favourable. Nevertheless, as Fig. 5 shows, our detections do include a small fraction of sources at $z > 3$ which are largely absent from the 450-μm-selected catalogue of Roseboom et al. (2013). This can be explained by the difficulty of obtaining unambiguous identifications for submm sources extracted blindly from the image (see also Casey et al. 2013). A similar high-redshift tail can be seen the 850-μm-selected sample of Koprowski et al. (2016), in which many of the sources at $z > 3$ had their redshifts estimated from the 100–850μm+1.4GHz SED, due to the absence of secure near-IR counterparts.

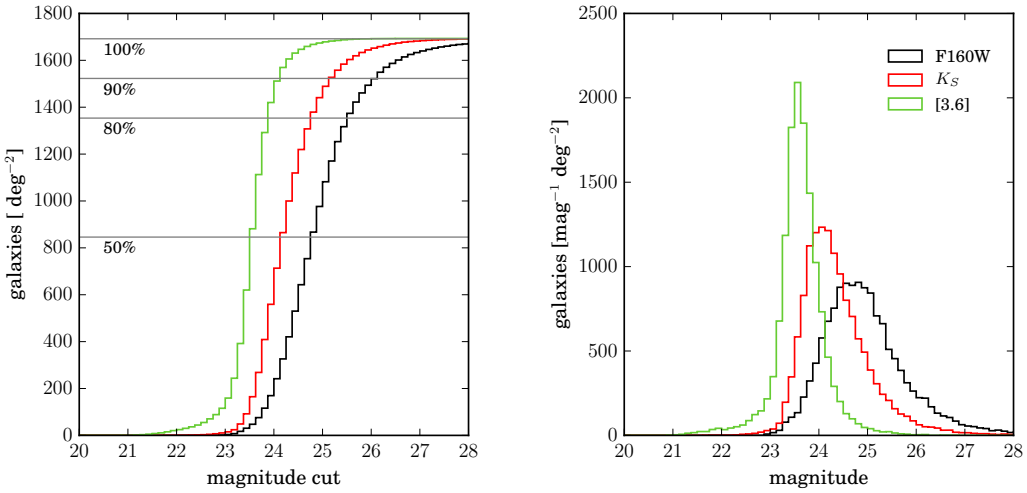

Figure 3. Completeness (left) and magnitude histograms (right) of simulated galaxies in the mass range $M_* > 10^{10}\mathrm{M}_\odot$ and redshift range $4 < z < 6$, drawn from a simulated EGG catalogue covering 10 deg^2 (Schreiber et al. 2016). The three bands relevant to our selection are plotted: F160W (the effective selection band of 3D-HST, in black); K_s (red); and IRAC [3.6] (green).

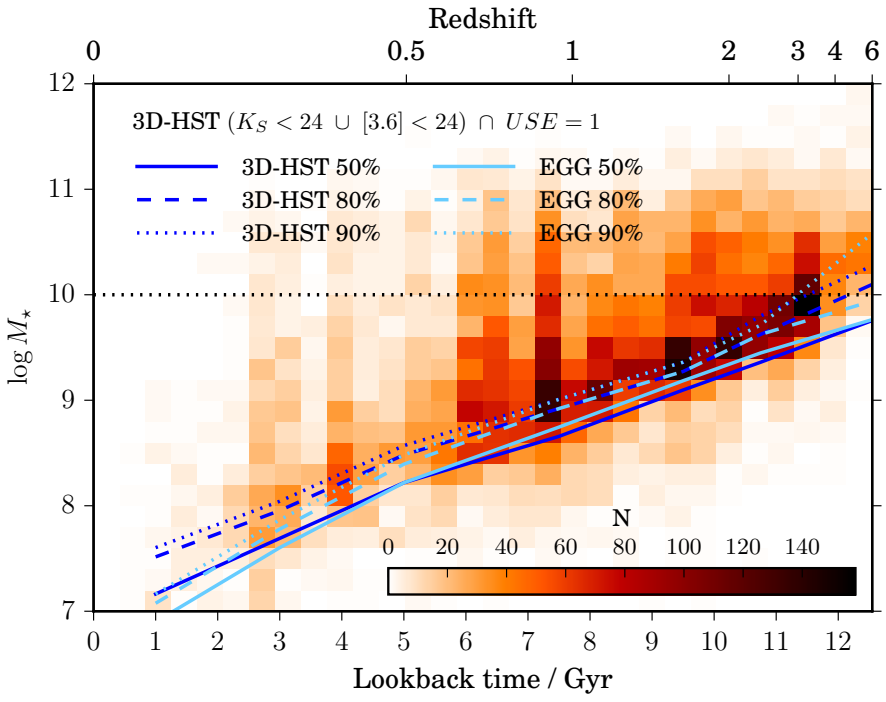

Figure 4. Distribution of stellar masses as a function of redshift or lookback time, for the subset of the 3D-HST galaxy catalogue meeting our selection criteria (shown by the coloured density histogram). The selection criteria are given in the label. The mass limit of $10^{10}\mathrm{M}_\odot$ is shown by the horizontal dotted line. The three *dark blue* lines mark the lowest masses to which the subset is complete at 50, 80 and 90 per cent respectively (relative to the full F160W-limited 3D-HST sample with USE=1). The three *light blue* lines mark the 50, 80 and 90-per-cent completeness limits predicted from the EGG simulated catalogue, where the completeness as a function of z and M_* is given by the fraction of galaxies that meet the same magnitude limits, $K_s < 24$ or $[3.6] < 24$. The agreement between the two sets of lines shows that the F160W limit in the 3D-HST catalogue does not reduce the completeness of our selection, and the overall completeness is at least 80 per cent at all redshifts.

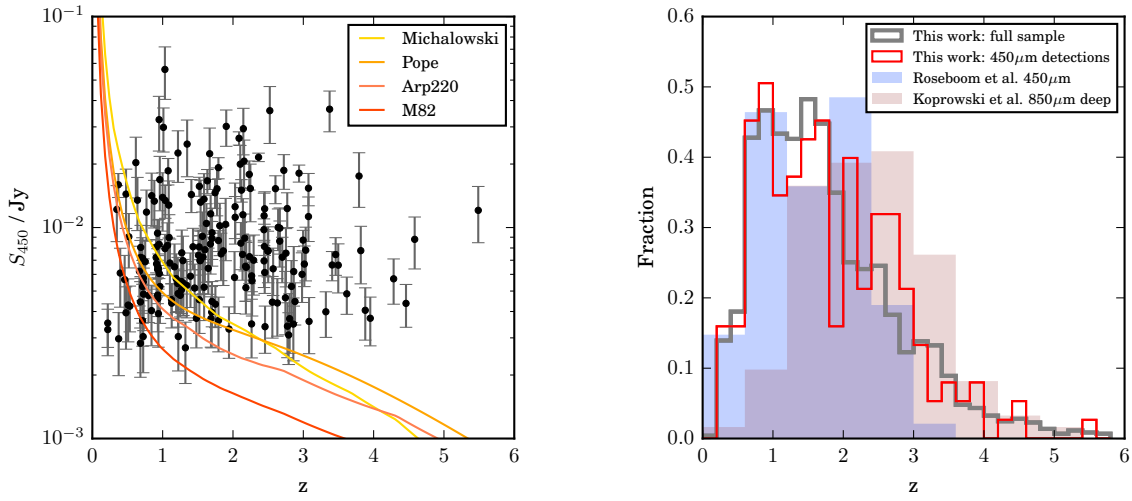

Figure 5. Properties of 450-μm sources detected at $S/N \geq 3$ by the T-PHOT algorithm. Left: 450-μm flux as a function of the redshift of the prior. Solid lines show the tracks of various SED templates scaled to a constant $\mathrm{SFR_{IR}} = 100\,\mathrm{M_\odot yr^{-1}}$ ($L_{\mathrm{IR}} = 6.7 \times 10^{11}\,\mathrm{L_\odot}$); templates are for SMGs from Michałowski et al. (2010) and Pope et al. (2008), and for Arp220 and M82 (Silva et al. 1998). Right: redshift distribution of detections (red line) and of the full prior sample with $M_* > 10^9\mathrm{M_\odot}$ (grey line). We also show, for comparison, the 450-μm-selected sources (with identifications) in the COSMOS and UDS deep fields of S2CLS (Roseboom et al. 2013); and the 850-μm-selected sources (with identifications) in the COSMOS deep field of S2CLS (Koprowski et al. 2016).

4.3 Stacked results

The key advantage of our technique is the ability to estimate unbiased fluxes for sources below the confusion limit, accounting for any clustering of the sources down to scales smaller than the beam. While only a small fraction (2 per cent) of the priors have a "detected" 450-μm flux at $S/N \geq 3$, it is still possible to utilise results from all priors by combining them statistically in order to measure average trends of 450-μm flux as a function of properties derived from the multi-wavelength priors. We therefore investigate "stacked" 450-μm properties from the T-PHOT measurements, in bins of stellar mass ($M_* > 10^{10}\mathrm{M_\odot}$), redshift ($0 < z < 6$) and rest-frame UV absolute magnitude M_{UV}. We use broad bins of redshift and stellar mass to maintain a sufficiently large sample in each bin, with boundaries at $z = 0.5$, 1.5, 2.5, 4.0, 6.0; $\log(M_*/\mathrm{M_\odot}) = 10.0$, 10.5, 11.0, 12.0. Note that our sample is complete to stellar masses $M_* > 10^{10}\mathrm{M_\odot}$ at all redshifts, but by including in our T-PHOT fits all galaxies selected down to a mass limit of $10^9\mathrm{M_\odot}$, we ensure that the covariance matrix accounts for all cross-correlations between galaxies above this limit, in order to minimise potential correlation-induced biases. Stacked fluxes are calculated using an inverse-variance-weighted mean in order to maximise the signal-to-noise ratio of stacked measurements. We include all binned sources in our stacks, regardless of signal-to-noise ratio, since excluding detections would lead to a bias against bright sources. Repeating the analysis using a median instead of a variance-weighted mean does not alter any of our conclusions.

In dividing our sample into bins of M_{UV}, we wish to probe galaxies as a function of their position on the UV luminosity function (LF). In this way we hope to understand how the UV luminosity relates to both SFR and dust obscuration. Since the UV LF evolves with redshift, we select M_{UV} bins as a function of redshift. Parsa et al. (2016) investigated the evolution of the LF via the Schechter parameters M^*_{UV}, ϕ^* and α in data from a large variety of surveys spanning $0 < z < 8$. They showed that the evolving LF could be empirically described by allowing these parameters to evolve smoothly with redshift. We adopt their best-fitting evolution,

$$M^*_{\mathrm{UV}} = (1+z)^{0.206}(-17.793 + z^{0.762}), \quad (1)$$

in order to define our M_{UV} bins as a function of redshift. We divide our sample into three M_{UV} bins within each redshift bin, placing the bin boundaries at +4, +2, 0, and −4 relative to $M^*_{\mathrm{UV}}(\bar{z})$ (where $\bar{z}$ is the mean redshift in the bin). The bin boundaries have been chosen to ensure similar numbers of objects in each bin, while at the same time sampling the UV LF in a consistent way at all redshifts.

4.4 Validating the stacking of T-PHOT measurements in simulations

In order to test whether stacked fluxes are representative, we applied our method to a realistic mock galaxy catalogue produced by the EGG simulation tool (Schreiber et al. 2016).[6] We took a simulated catalogue of a 412 arcmin^2 region, and created a 450-μm image with realistic instrumental noise (Gaussian rms=2.5 mJy beam^{-1}), injecting point

[6] Empirical Galaxy Generator (EGG) available from http://astrodeep.eu

sources with a Gaussian PSF (FWHM=7.5 arcsec). Full details of the EGG simulations are given in Schreiber et al. (2016); we provide a brief description here. The mock catalogue is constructed from stellar mass functions based on data in the GOODS-S field (Schreiber et al. 2015; Grazian et al. 2015) for star-forming and quiescent galaxies separately, in redshift bins between $0.3 < z < 7.5$, with extrapolation up to $z = 11$, and down to $M_* = 10^8 M_\odot$. Stellar SEDs are assigned by measuring rest-frame colours of galaxies in CANDELS GOODS-S and defining mass-and-redshift dependent sequences for star-forming and quiescent galaxies in the $U - V$, $V - J$ (hereafter, UVJ) colour-colour diagram. Galaxies are modeled with disc and bulge components, which are assigned stellar SEDs according to their position in the rest-frame UVJ diagram, using templates from FAST with a Bruzual & Charlot (2003) stellar population. Star-forming galaxies are assigned a SFR based on the dual-mode model of Sargent et al. (2012); "main-sequence" SFRs are based on stellar mass and redshift using the fits from Schreiber et al. (2015), with 0.3 dex scatter, and a randomly-selected 3 per cent of galaxies are placed in the "starburst' mode by enhancing their SFR by a factor of 5.24. Quiescent galaxies are also assigned a SFR based on their stellar mass, in order to reproduce dust emission in these galaxies. Each galaxy's IR luminosity is based on the SFR by assuming an obscuration fraction depending on stellar mass (Schreiber et al. 2015), and the IR SED, which depends on redshift and mass, is based on the model described in Schreiber et al. (2016). Sky positions are assigned with a built-in angular correlation function with a power-law index of -1 and a normalisation which depends on redshift and stellar mass.

We selected a sample from the mock catalogue in an identical way to our true sample, fitted the simulated submm images with T-PHOT, and analysed results in an identical way. Comparing output versus input fluxes indicates very good agreement with no bias in stacked results; a small fraction of individual flux measurements could be boosted by blending (as might be expected due to incomplete priors at the lowest stellar masses), but this does not bias average results in stacks. Measuring the difference between average output and input SFRs in bins of $M_{\rm UV}$, M_* and redshift, we find that the distribution is consistent with the measurement errors: the distribution of $({\rm SFR_{out}} - {\rm SFR_{in}})/\Delta{\rm SFR_{out}}$ has a mean of 0.2 and a standard deviation of 0.7, and the scatter is uncorrelated with either M_*, $M_{\rm UV}$, or redshift. The input relationships between mass, $M_{\rm UV}$ and SFR are recovered with good fidelity, and the SFRD is also recovered without bias. We can therefore trust that the results described in Section 5 are valid and not subject to bias or systematic effects as a result of our method of measuring average SFR in bins.

4.5 Infrared spectral energy distributions

Having applied our technique to the 450-μm images, we completed an identical analysis on the images at 100, 160, 250 and 850μm, in order to measure stacked FIR SEDs. We used Gaussian PSFs with FWHM = 9, 11, 18, 14 arcsec in these four bands respectively. These images have greater levels of confusion than the 450-μm images, so the constraints at high redshifts are generally weaker (especially from *Herschel*), but at $z < 2$ the *Herschel* data provide valuable information at wavelengths close to the peak of the SED, which reduces the systematic uncertainty on the total IR luminosity. We have opted not to include the 350 and 500-μm SPIRE bands due to the greatly increased confusion at these wavelengths, which means that a much smaller number of priors can be reliably fitted, and for this particular study they do not offer significant additional information to constrain the SED shape.

Fig. 6 shows the stacked fluxes as a function of observed-frame wavelength in each of our bins. The uncertainties on these measurements include three components. Firstly, the error on the variance-weighted mean of the stack includes the T-PHOT measurement errors (from the diagonal of the covariance matrix, which accounts for covariances with other priors). To each of these, we add in quadrature the residual confusion noise due to sources not accounted for in the priors, which is estimated from the standard deviation of pixels in the T-PHOT residual map. Finally we add in quadrature the flux calibration uncertainty for each band. These are assumed to be 5.5 per cent at 100 and 160μm,[7] 5 per cent at 250μm,[8] 12 per cent at 450μm and 8 per cent at 850μm (Dempsey et al. 2013).

In each bin, we obtained least-squares fits to several SED templates that have been previously used to describe SEDs of high-redshift star-forming galaxies. We used the average submillimetre galaxy (SMG) template from Michałowski et al. (2010), the SMG template from Pope et al. (2008), and the Arp220 and M82 templates from Silva et al. (1998). The templates are redshifted into the observed frame at the mean redshift of each $z, M_*, M_{\rm UV}$ bin. We opted to exclude the PACS data from the fitting in bins at $z > 2.5$, considering that galaxies at these high redshifts are likely to be faint in these bands, and any detection is therefore more likely to be biased by contamination from any lower-redshift objects that may be missing from the priors (for example the 160μm detections in some bins at $2.5 < z < 4.0$ in Fig. 6). We found, however, that our results were not significantly changed if we included all data in the fitting at $z > 2.5$.

Results from the fitting indicate generally very consistent SED shapes in all bins, as far as can be determined from the stacked measurements. In each bin, we compared the χ^2 obtained by fits to the four models, and found that the lowest χ^2 was generally obtained with either the Michałowski et al. (2010) or the Pope et al. (2008) SMG templates. In cases where the Pope et al. template was the best fit, this was only a marginal improvement over the Michałowski et al. template. We therefore decided to use the Michałowski et al. in fitting SEDs in all bins. This template has been shown to provide a good description of the SEDs of SMGs over a wide range of stellar masses ($> 10^{10} M_\odot$; Dunlop et al. 2016). Furthermore, the effective cold-dust temperature of this SED (i.e. the form at $\gtrsim 100\mu$m) is close to a modified blackbody at 30 K with $\beta \approx 1.5$, which has been shown to be an appropriate description of the SEDs of K-selected and colour-selected samples at high redshifts (e.g. Hilton et al. 2012; Decarli et al. 2014). As described in the next section, we

[7] herschel.esac.esa.int/Docs/PACS/html/pacs_om.html
[8] herschel.esac.esa.int/Docs/SPIRE/html/spire_om.html

conservatively allow for the systematic uncertainty in the SED template in our obscured SFR calibration.

4.6 Calibrating obscured and unobscured SFR

We estimate the total IR luminosity ($L_{\rm IR}$) in each bin by integrating the best-fitting model over the rest-frame wavelength range 8–1000μm. We estimate the uncertainty on $L_{\rm IR}$ by combining two components in quadrature. The first of these is the formal fitting error from the least-squares fit, which accounts for uncertainties in the average flux densities measured at each wavelength (see Sections 4.1, 4.3). Secondly, we include the full range of luminosities given by the fits to the four SED templates described in Section 4.5, to conservatively allow for variation in the effective dust temperatures within our bins, between cold SMG-like SEDs (Michałowski et al. 2010; Pope et al. 2008), moderate Arp220-like SEDs, and hot M82-like SEDs (see for example Magnelli et al. 2013; Béthermin et al. 2015). We calibrate obscured SFRs by assuming the relationship

$$\mathrm{SFR_{IR}/M_{\odot}yr^{-1}} = 3.88\times10^{-44}\,L_{\rm IR}/\mathrm{erg\,s^{-1}} \quad (2)$$

(Murphy et al. 2011). We also estimate the unobscured SFR from the rest-frame far-UV luminosity at 1600Å (without extinction corrections), which is provided in the 3D-HST catalogue based on the EAZY templates fitted for the photometric redshifts. The rest-frame luminosity is converted to the unobscured SFR using the relationship

$$\mathrm{SFR_{UV}/M_{\odot}yr^{-1}} = 4.42\times10^{-44}\,\nu L_{\nu}(FUV)/\mathrm{erg\,s^{-1}} \quad (3)$$

(Hao et al. 2011; Murphy et al. 2011).[9] To estimate the *total* SFR, we simply take the sum of the obscured and unobscured SFRs within a given sample, which is an established method of accurately recovering total SFR (e.g. Bell 2003; Bell et al. 2005; Barro et al. 2011; Hao et al. 2011; Murphy et al. 2011; Davies et al. 2016). Uncertainties in the measurements of $L_{\rm IR}$ and $L_{\rm UV}$ are propagated through.

5 RESULTS AND DISCUSSION

5.1 The dependence of SFR on stellar mass and UV luminosity

In Fig. 7 we plot the total SFR measured from the stacked FIR+UV data as a function of stellar mass, divided into bins of $M_{\rm UV}$ and redshift, and including all objects irrespective of measured S/N (as described in Section 4.3). The numbers of objects in each bin are shown for reference as histograms in the upper panels. In the lower panels, large coloured symbols show the mass/$M_{\rm UV}$-binned stacks, while large black squares show the full mass-binned stacks with no $M_{\rm UV}$ binning (note that we plot these stacks down to $M_*=10^9\mathrm{M_{\odot}}$, although bins below $10^{10}\mathrm{M_{\odot}}$ are incomplete at $z>1.5$ as shown in Fig 4, hence average SFRs in those bins may be biased). Small coloured points indicate 450-μm detections, whose FIR SFR is estimated by scaling the Michałowski et al. (2010) SED template to the measured 450-μm flux and integrating the template in the range 8–1000μm (see Section 4.6). On this and subsequent scatter plots, the detected objects are coloured according to *maxCvRatio*, an output parameter from T-PHOT defined as the ratio of maximum covariance to variance on the flux measurement. This parametrizes the level of blending uncertainty on each object; individual detections with *maxCvRatio* $\gtrsim 1$ are heavily blended with another prior, although the resulting uncertainty in their flux is accounted for in their error estimate. It is clear that some of the high measured SFRs of detected objects could be attributed to high blending, especially for priors with stellar mass $<10^{10}\mathrm{M_{\odot}}$; overall, however, the T-PHOT measurements account for this blending and average (stacked) results are not affected.

Fig. 7 shows that the total stacked SFR is correlated with stellar mass, both in the full mass-limited samples and in narrow $M_{\rm UV}$ bins. The one exception to this relationship is that galaxies with $M_*>10^{11}\mathrm{M_{\odot}}$ and $M_{\rm UV}>-17.4$ at $z<1.5$ have lower SFRs than lower-mass galaxies at the same redshift, suggesting a significant fraction of passive/quenched galaxies in this bin (we will return to this issue in Section 5.2). In the lower panels of Fig 7, the obscuration of star formation is revealed by the fraction $\mathrm{SFR_{IR}/SFR_{UV}}$, which increases strongly with increasing stellar mass and with decreasing UV luminosity. These results indicate that UV luminosity (relative to $M^*_{\rm UV}$) is strongly anti-correlated with dust obscuration at all redshifts and stellar masses $>10^{10}\mathrm{M_{\odot}}$, and are consistent with previous studies covering similar redshift ranges (Heinis et al. 2014; Coppin et al. 2015). Fig. 8 shows the stacked SFRs plotted as a function of $M_{\rm UV}$. Total SFR is relatively weakly correlated with UV luminosity at $z<2.5$, rising by a factor of 2–3 over a range of 2.5 magnitudes (a factor of 10 in UV luminosity); in constrast, the SFR-mass relation is approximately linear in Fig. 7 (echoing the well-known "main sequence" of star formation; Noeske et al. 2007; Elbaz et al. 2011; see also Section 5.2). At $z>2.5$ in Fig. 8, there is no discernable dependence of SFR on $M_{\rm UV}$, and the SFR in the most UV-luminous bins is consistent with the average mass-limited SFR in Fig. 7.

The 450-μm-detected sample, shown by small coloured points in Figs. 7 and 8, samples a wide range of both stellar masses and $M_{\rm UV}$ at all redshifts $z<4$, with a limiting SFR $\sim 100\ \mathrm{M_{\odot}yr^{-1}}$ at $1.5<z<4$. At all redshifts, the most luminous FIR sources (represented by the 450-μm detections) sample the galaxies with the highest SFRs. These galaxies tend to be highly obscured, although they have a wide range of obscuration factors spanning roughly two orders of magnitude (lower panels of Fig. 7). In contrast, the most luminous UV sources (red symbols) tend to sample higher-than-average SFRs at $z<2.5$ (in comparison with the mass-selected bins; black squares), while at $z>2.5$ they appear to have similar SFRs, but are biased towards the least obscured systems at all redshifts.

5.2 The evolution of star-forming galaxies at high redshift

Specific SFR (SSFR = SFR/M_*) is commonly used as an indicator of the evolutionary state of individual galaxies and galaxy samples, since it is defined as the inverse of the time required to double the current stellar mass while sustaining the current SFR. Most star-forming galaxies form a tight

[9] Both $\mathrm{SFR_{IR}}$ and $\mathrm{SFR_{UV}}$ are calibrated to a Kroupa & Weidner (2003) IMF.

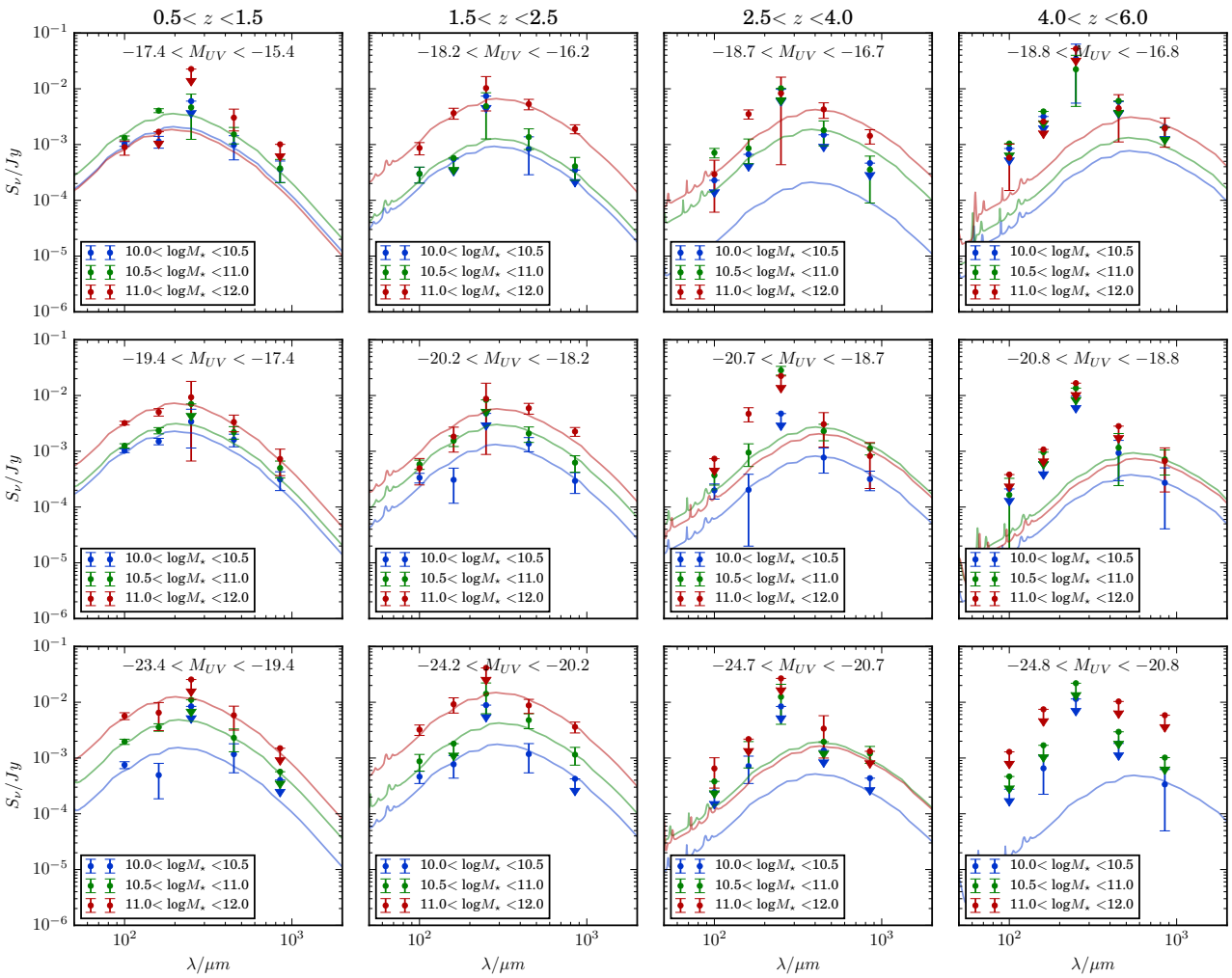

Figure 6. Stacked flux measurements at 100, 160, 250, 450, 850 μm of stellar-mass- and $M_{\rm UV}$-selected galaxies in four redshift bins (left to right) and three $M_{\rm UV}$ bins (top to bottom), relative to the redshift-dependent value of $M^*_{\rm UV}$. Error bars include the T-PHOT measurement uncertainty, the residual confusion noise and the flux calibration uncertainties as detailed in the text. Bins with stacked $S/N \leq 1$ are plotted as 2-σ upper limits. The best-fitting models using the Michałowski et al. (2010) SED templates are also shown.

"main sequence" in SSFR as a function of stellar mass and redshift, although a minority have much higher SSFRs (i.e. starbursts), while passive galaxies (which are common at high masses and low redshifts especially) fall well below the sequence (e.g. Noeske et al. 2007; Elbaz et al. 2011; Whitaker et al. 2012; Speagle et al. 2014; Schreiber et al. 2015). In order to define the sequence it is therefore particularly important to exclude potentially passive galaxies from the sample so that average SSFRs are not biased low.

In Fig. 9 we plot the average SSFR of star-forming galaxies in our sample (black squares), as a function of redshift, in the three bins of stellar mass that are complete. We have excluded passive galaxies from each bin using the *UVJ* colour criteria defined by Whitaker et al. (2011; see also Williams et al. 2009), taking the rest-frame *UVJ* magnitudes from the SED-fitting results in the 3D-HST data release (Skelton et al. 2014).[10] The criteria for passive galaxies are

$$
\begin{aligned}
&z < 0.5: \\
&\quad U-V > 0.69 + 0.88(V-J),\ U-V > 1.3,\ V-J < 1.6 \\
&0.5 < z < 1.5: \\
&\quad U-V > 0.59 + 0.88(V-J),\ U-V > 1.3,\ V-J < 1.6 \\
&1.5 < z < 2.0: \\
&\quad U-V > 0.59 + 0.88(V-J),\ U-V > 1.3,\ V-J < 1.5 \\
&z > 2.0: \\
&\quad U-V > 0.59 + 0.88(V-J),\ U-V > 1.2,\ V-J < 1.4
\end{aligned}
\qquad (4)
$$

Although the *UVJ* criteria from Whitaker et al. (2011) were

[10] The effects of these cuts on the SFR results in Figs. 7 and 8 are shown in Appendix A.

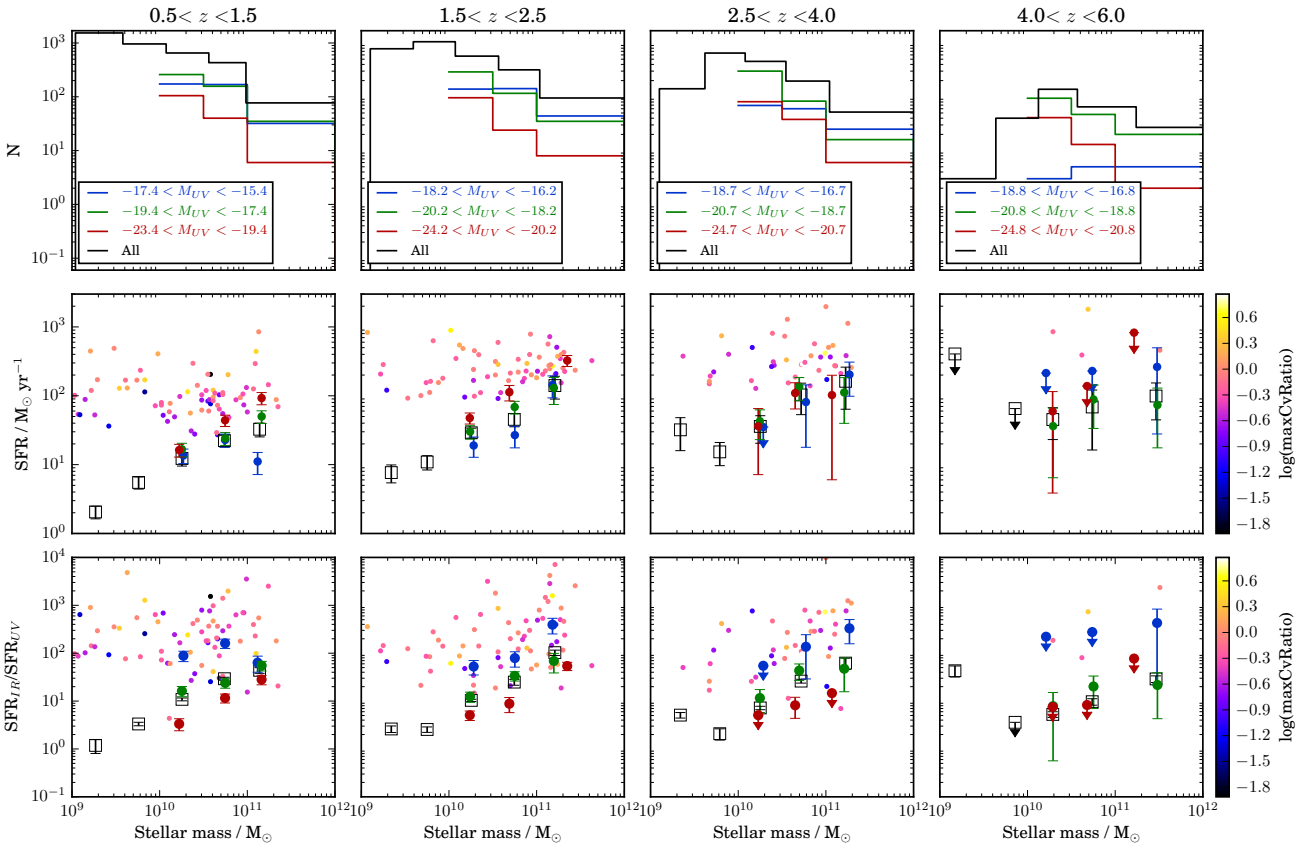

Figure 7. Stacked and detected samples as a function of stellar mass for each redshift bin. From top to bottom: number N per bin; total $\mathrm{SFR_{UV}} + \mathrm{SFR_{IR}}$; and obscuration ratio $\mathrm{SFR_{IR}}/\mathrm{SFR_{UV}}$. Large black squares show the full mass-binned stacks, while large filled symbols with error bars show the stacks divided into bins of M_{UV}, defined relative to M^*_{UV} at the appropriate redshift (Parsa et al. 2016). In bins where the stacked FIR emission is not detected with $S/N > 1$, the 2-σ upper limit is shown as a downward arrow. FIR detections ($S/N > 3$) are shown by small coloured points in which the colour coding indicates *maxCvRatio*, a T-PHOT output parameter defined as the ratio of maximum covariance to variance on the flux measurement. Individual detections with *maxCvRatio* $\gtrsim 1$ are heavily blended with another prior, which therefore dominates their uncertainty. Note that the high $\mathrm{SFR_{IR}}$ in bins with $M_* \sim 10^9 \mathrm{M}_\odot$ at $z > 1.5$ can be attributed to incompleteness and bias of the sample at low stellar masses.

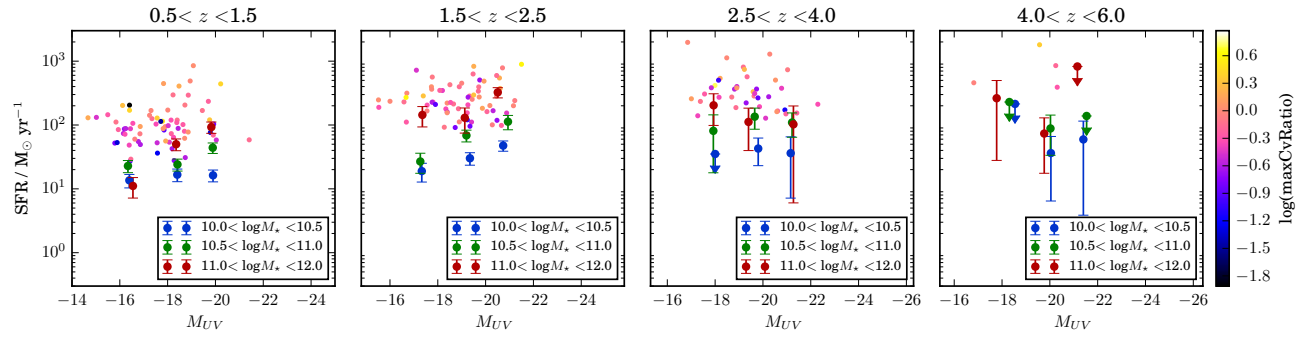

Figure 8. Total $\mathrm{SFR_{UV}} + \mathrm{SFR_{IR}}$ plotted as a function of M_{UV} for each redshift bin. The data shown are the same as in Fig 7, except that colours here indicate the stellar-mass bin.

only defined for $z < 3.5$, we extend their $2.0 < z < 3.5$ criteria to $z > 3.5$ on the basis that massive, passive galaxies may be present even at these early times (Marchesini et al. 2010; Ilbert et al. 2013; Muzzin et al. 2013; Nayyeri et al. 2014; Spitler et al. 2014; Straatman et al. 2014). The SSFR is calculated by taking the variance-weighted mean within each bin of all individual $(SFR_{UV}+SFR_{IR})/M_*$ measurements, where SFR_{IR} is measured by scaling the 450-μm flux measurement to the SMG SED template from Section 4.5 (since it is not possible to fit SEDs to individual flux measurements which are not detected). Note that this method is less prone to bias than taking the overall SFR_{IR} from the stacked SED fit and combining it with the mean SFR_{UV} and mean M_*. We have not binned by M_{UV}, since we have shown this is not a strong indicator of total SFR at a given stellar mass. We also show in Fig. 9 the functional fits to the main sequence of star-forming galaxies as a function of stellar mass and redshift from the compilation of heterogeneous results from the literature by Speagle et al. (2014), and also the fit to stacked *Herschel* data in GOODS, UDS and COSMOS by Schreiber et al. (2015). The two relations broadly agree despite differing methodologies. It is interesting to note that the 450-μm-detected sample (small points) have SSFRs much higher than the average mass-selected galaxies at low masses, while at high masses the detected sample reaches similar SSFRs to the stacks, and is fully consistent with the main-sequence fits in the literature. Similar conclusions have previously been noted in relation to SMGs selected at similar wavelengths, from these same data-sets and others (Michałowski et al. 2012; Roseboom et al. 2013; Koprowski et al. 2014, 2016).

Our stacked measurements of the average UV+IR SSFR in star-forming galaxies at $M_* > 10^{10}M_\odot$ are generally consistent with the plotted relations from Speagle et al. (2014) and Schreiber et al. (2015), as well as many other measurements, not shown, covering the full redshift range (e.g Whitaker et al. 2012; González et al. 2014; Koprowski et al. 2014, 2016; Steinhardt et al. 2014; Tasca et al. 2015). The average UV+IR SSFR is between $1-2$ Gyr^{-1} for stellar masses $> 10^{10}M_\odot$ at $z \sim 2-4$, and roughly $3-4$ Gyr^{-1} at $z \sim 5$, which is consistent with recent submm studies at these redshifts such as Koprowski et al. (2014, 2016) from SCUBA-2, and Dunlop et al. (2016) and Schreiber et al. (2016) from ALMA. The evolution with redshift appears stronger at higher stellar masses, so that at $z \sim 1$ there is a strong fall in SSFR with increasing mass. This is confirmed by modelling the redshift dependence as a power law $SSFR = a(1+z)^b$ in each mass bin, obtaining the best-fitting parameters listed in Table 1 (fits shown by the thin grey dashed lines in Fig. 9). The weaker evolution of lower-mass galaxies compared with higher-mass galaxies is consistent with the results of Speagle et al. (2014) and Schreiber et al. (2015).

To constrain these trends more precisely, we plot the average UV+IR SSFR as a function of redshift (left) and stellar mass (right) in Fig. 10. The results in all bins are shown, although some bins are known to be incomplete (see Fig. 4); these are indicated by small symbols linked by dotted lines, and these bins appear to be subject to strong bias towards high SSFRs. The data in the complete bins appear to show a smooth dependence of SSFR on mass and redshift, which we can model as a power-law in mass with an index and normalization which both evolve with redshift (e.g. Whitaker et al. 2012):

$$\log SSFR(M_*, z)/\mathrm{Gyr}^{-1} = a(z)[\log(M_*/M_\odot) - 10.5] + b(z)$$
$$a(z) = a_0 + a_1 \log(1+z)$$
$$b(z) = b_0 + b_1 \log(1+z) \quad (5)$$

Table 1. Best-fitting power-law parameters to the model $SSFR/\mathrm{Gyr}^{-1} = a\,(1+z)^b$ in stellar-mass bins as shown in Fig. 10.

IR+UV SSFR:

$\log M_*$	a	b
10.0 – 10.5	0.45 ± 0.19	1.3 ± 0.4
10.5 – 11.0	0.14 ± 0.05	1.9 ± 0.3
11.0 – 12.0	0.11 ± 0.04	1.9 ± 0.4

UV-corrected SSFR:

$\log M_*$	a	b
10.0 – 10.5	0.27 ± 0.03	1.3 ± 0.1
10.5 – 11.0	0.23 ± 0.04	0.9 ± 0.1
11.0 – 12.0	0.08 ± 0.04	1.5 ± 0.6

The dashed lines in Fig. 10 show the result of a least-squares fit to the data in the complete bins (large symbols, solid lines), with the parameters $a_0 = -0.64 \pm 0.19$, $a_1 = 0.76 \pm 0.45$, $b_0 = -9.57 \pm 0.11$, $b_1 = 1.59 \pm 0.24$. The reduced χ^2 of this fit is 0.8. This model describes a negative slope in log SSFR as a function of log M_* at low redshifts, which tends towards a flat relation as redshift increases towards $z \sim 5$ (in other words, a sub-linear slope in log SFR at low redshifts, which tends towards a linear relation at $z \sim 5$. Similar conclusions were drawn by Schreiber et al. (2015) and Tomczak et al. (2016), both at $0 < z < 4$, and Whitaker et al. (2014) at $0 < z < 2.5$. In contrast, Whitaker et al. (2012) found that the slope evolved in the opposite sense over $0 < z < 2.5$; however, this may have been influenced by a redshift-dependent mass completeness limit in their sample. The slope of the main sequence at high redshifts remains a matter of active debate and much disagreement in the literature; for recent overviews see Speagle et al. (2014) and Johnston et al. (2015).

The evolution of the normalization of the main sequence at $\log(M_*/M_\odot) = 10.5$ is described by $b(z)$, which indicates a relatively weak evolution as $(1+z)^{1.6}$, compared with some recent studies at high redshifts; e.g. $(1+z)^{2.9}$ $(0 < z < 2.5$; Whitaker et al. 2014), $(1+z)^{\sim 2.8}$ $(0 < z < 6$; Speagle et al. 2014), $(1+z)^{2.6}$ $(0 < z < 3$; Johnston et al. 2015), $(1+z)^{\sim 2.5}$ $(3.5 < z < 6.5$; Salmon et al. 2015b). However, Schreiber et al. (2015) find the best-fitting normalization evolves as $(1+z)^{1.5}$ over $0 < z < 4$, while Tasca et al. (2015) parametrize the evolution as a broken power law with $(1+z)^{2.8}$ at $z < 2.3$ and $(1+z)^{1.2}$ at $2.3 < z < 5.5$. Béthermin et al. (2015) similarly suggest a break in the trend, between $(1+z)^{2.8}$ at $z < 2$ and $(1+z)^{2.2}$ at $2 < z < 4$, while Duncan et al. (2014) fit a trend of $(1+z)^{2.1}$ at $4 < z < 7$. An even weaker evolution at high redshifts is supported by Marmol-Queralto et al. (2015), who report a trend of $(1+z)^{1.0}$ based on Hα equivalent widths at $1 < z < 5$. In comparison, several theoretical models predict an evolution that closely traces the gas or dark-matter accretion rate, i.e. $(1+z)^{2.25}$, which is intermediate between

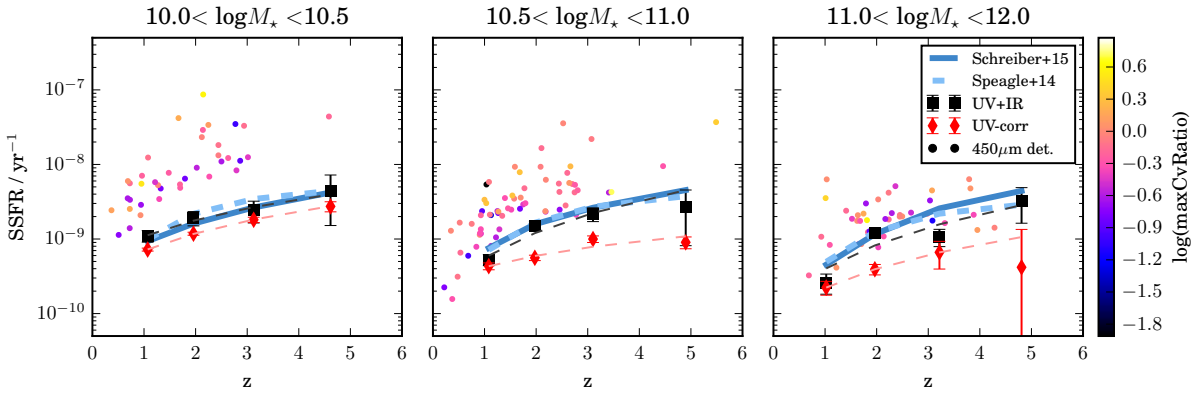

Figure 9. Specific SFR as a function of redshift in bins of stellar mass, for *UVJ*-selected star-forming galaxies. Black squares are the average of the mass-limited sample, ($M_* > 10^{10}$M$_\odot$) using the total IR+UV SFR; red diamonds are the dust-corrected SFR estimates from the UV SED fitting alone; small points show individual 450-μm detections. Thin dashed grey and red lines show the power-law fits to the binned UV+IR and dust-corrected UV data (respectively) as a function of $(1+z)$. Thick dashed and solid lines indicate fits to the main-sequence of star-forming galaxies from Speagle et al. (2014) and Schreiber et al. (2015).

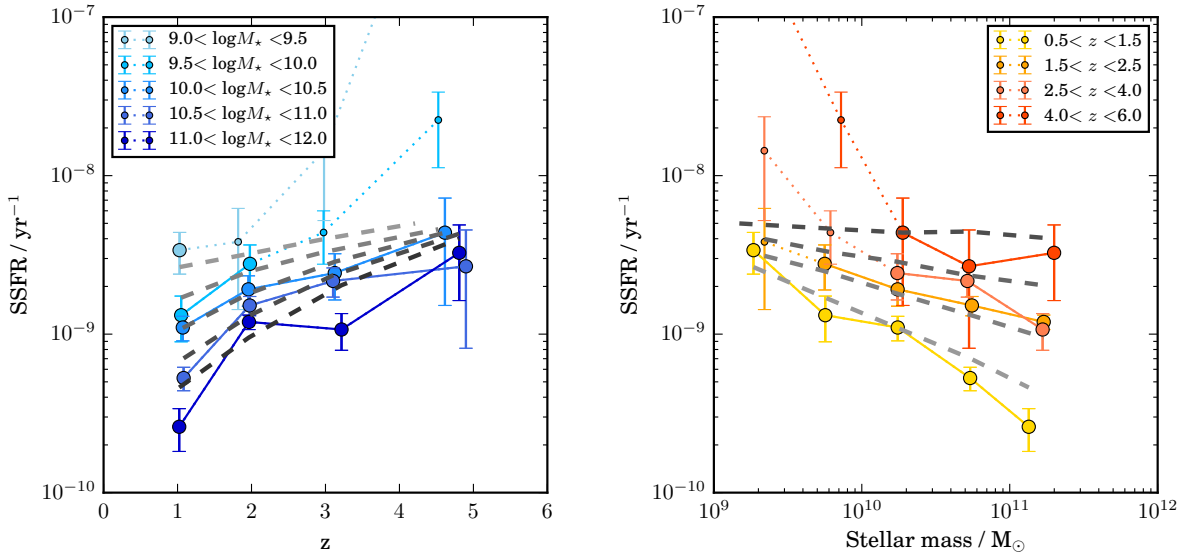

Figure 10. Stacked specific SFR in bins of redshift and stellar mass, plotted as a function of redshift (left) and stellar mass (right). Large symbols joined by solid lines indicate bins in which the input sample is complete; small symbols joined with dotted lines are incomplete bins in which the SSFR is biased towards high values. Dashed lines indicate the best-fitting power-law models to SSFR$(1+z)$ in each stellar mass bin (left) and SSFR(M_*) in each redshift bin (right), fitting only to complete bins. The parameters of these models are listed in Table 1.

the various observed trends (Dutton et al. 2010; Behroozi et al. 2013a; Guo et al. 2015). Lehnert et al. (2015) propose a model in which SSFR rises as $(1+z)^3$ at $0 < z < 2$, due to an increase with redshift of gas surface densities and gas accretion rates of galaxies, but at $z > 2$, self-regulating feedback maintains roughly constant (or very slowly rising) SSFRs with increasing redshift, which is consitent with some of the literature cited above. Our data suggest a moderate increase in SSFR(z) at $z > 2$, although the constraints on this slope from our data alone are relatively weak, given the large error bars and broad redshift bins.

5.3 Obscuration and UV extinction corrections

In Fig. 9 (red diamonds) we compare our IR+UV SSFRs to those estimated from the UV continuum luminosity following the Meurer et al. (1999) prescription, using the UV slope (β) fitted to the rest-frame UV data by Skelton et al. (2014). As with the IR+UV measurements described in Section 5.2, we plot the mean of these "extinction-corrected" UV SSFRs for *UVJ*-selected star-forming galaxies in each bin of stellar mass and redshift. The results are generally lower than the "total" IR+UV SSFRs, and the discrepancy increases with both mass and redshift. The weaker evolution is indicated by the power-law fits to SSFR$(1+z)$ in each mass bin as

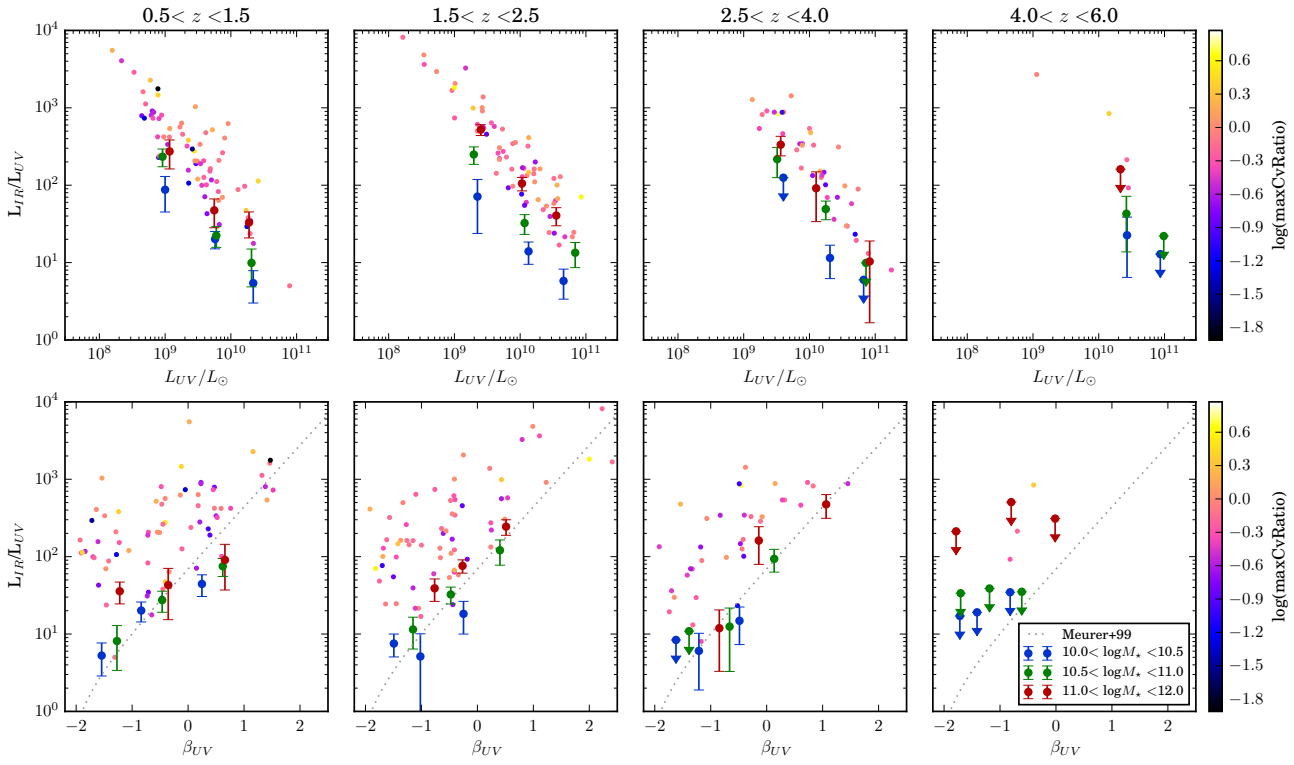

Figure 11. Average IRX= $L_{\rm IR}/L_{\rm UV}$ as a function of UV luminosity $L_{\rm UV} = \nu L_\nu(1600\text{Å})$ (upper panels) and UV slope β (lower panels) from binned data and individual detections. The dotted line is the relationship for local starbursts from Meurer et al. (1999).

shown in Fig. 9 and Table 1. Overall these results are in much worse agreement with the literature data plotted in in Fig. 9, indicating that these UV slope corrections may not be appropriate for stellar-mass-selected samples. We have already seen (Fig. 7; Section 5.1) that the ratio $\rm SFR_{\it IR}/SFR_{\it UV}$ is correlated with stellar mass and redshift, as well as being strongly anti-correlated with the UV luminosity. Since the Meurer et al. (1999) relation is calibrated for UV-bright starbursts, it is not surprising that, as our sample becomes increasingly dominated by obscured star formation at high masses and high redshifts, the UV dust corrections become less effective.

We can investigate this further by directly exploring the $L_{\rm IR}/L_{\rm UV}$ ratio (hereafter "IRX") as a function of UV luminosity [$L_{\rm UV} = \nu L_\nu(1600\text{Å})$] and β, as shown in Fig. 11. Here we have again excluded passive galaxies as defined by equation (4). In the upper panels of Fig. 11 we have binned the data by $M_{\rm UV}$ as in Fig. 7 and 8, but in the lower panels we have divided the sample equally into bins of β within each redshift and stellar-mass bin. Stacked measurements are the average within each bin of all IRX ratios, where the individual $L_{\rm IR}$ values are estimated by scaling the 450-μm flux to the same SED template used above (since we cannot fit SEDs to individual sources that are undetected). In the top panels, IRX is a strong function of UV luminosity in each stellar mass bin (see also Fig. 7). In the lower panels, the bins with higher β (which in general corresponds to lower $L_{\rm UV}$) generally have higher obscuration. Most of the bins are close to the Meurer et al. (1999) relation, but there is evidence for deviations in some regimes, which would account for the discrepancies between UV-corrected and IR+UV SSFRs in Fig. 9. At $z < 1.5$ and $M_* > 10^{11}\rm M_\odot$ the IRX–β relation appears flatter, possibly as a result of contamination of the sample from passive galaxies (in spite of our *UVJ* selection), although IRX is also below the Meurer et al. prediction for $\beta > 0$ at lower masses, which are more likely to be star-forming. At $z \sim 2$ there is evidence for a stellar-mass dependence in the relationship, with $M_* > 10^{11}\rm M_\odot$ galaxies falling above the Meurer et al. relation, and $M_* < 10^{10.5}\rm M_\odot$ galaxies falling below it. This trend may persist at $z > 2.5$, but the stacked detections are too weak in our *UVJ*-selected sample to be certain. The dependence of this relation on stellar mass has been previously noted in several studies, using stacking of high-redshift LBGs (Coppin et al. 2015; Álvarez-Márquez et al. 2016; Bouwens et al. 2016). Meanwhile, FIR-detected sources in Fig. 11 show a weak correlation between β and obscuration, although most appear to have much higher obscuration than predicted by the Meurer et al. relation, perhaps unsurprisingly given that the selection band is proportional to obscured SFR. The *maxCvRatio* values are mostly $\ll 1$, indicating that this cannot be explained by blending-induced errors in the IR luminosities.

The overall relationship between IRX and $L_{\rm UV}$ (in the upper panels of Fig. 11) appears to be independent of red-

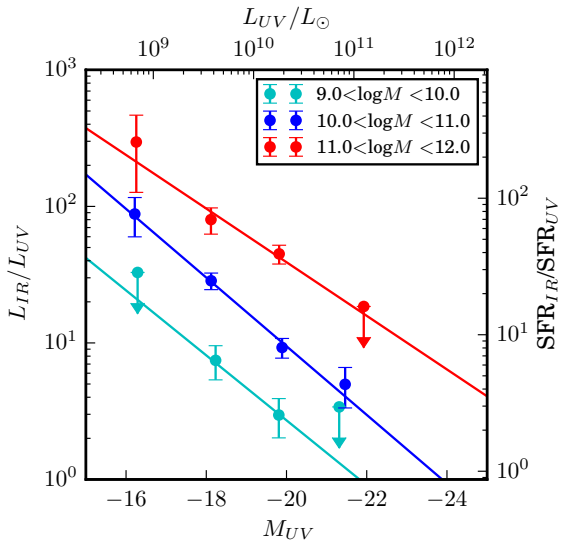

Figure 12. Average IRX as a function of UV absolute magnitude in stacked data across all redshifts. Lines show the best-fitting linear models described in Table 2.

Table 2. Best-fitting parameters of the linear model $\log(L_{IR}/L_{UV}) = p_0 + p_1(M_{\rm UV} + 18)$ in three stellar mass bins as shown in Fig. 12.

$\log M_*/\mathrm{M}_\odot$	p_0	p_1
9.0 – 10.0	0.91 ± 0.13	0.24 ± 0.10
10.0 – 11.0	1.48 ± 0.05	0.25 ± 0.03
11.0 – 12.0	1.98 ± 0.08	0.20 ± 0.05

shift, but strongly dependent on stellar mass. We therefore plot stacked IRX in bins of $M_{\rm UV}$ and stellar mass, but with no redshift binning, in Fig. 12. This shows a strong and smooth dependence of obscuration on UV luminosity in each stellar-mass bin. Linear regression fits for $\log(L_{\rm IR}/L_{\rm UV})$ as a function of $M_{\rm UV}$ in each mass bin are shown in Fig. 12 and Table 2. The slope of this relationship is consistent in all stellar mass bins, while the normalization increases roughly as $0.5 \log M_*$, indicating that IRX is proportional to $0.23 M_{\rm UV}$, or equivalently $(L_{\rm UV})^{-0.6}$; and ${M_*}^{0.5}$.

Alternatively, if we assume a smooth power-law dependence on both mass and luminosity, we can combine all binned data in a single model,

$$\log\left(\frac{L_{\rm IR}}{L_{\rm UV}}\right) = a + b\log\left(\frac{L_{\rm UV}}{10^9 \mathrm{L}_\odot}\right) + c\log\left(\frac{M_*}{10^9 \mathrm{M}_\odot}\right) \quad (6)$$

and find a least-squares fit with $a = 5.9 \pm 1.8$, $b = -0.56 \pm 0.07$, $c = 0.70 \pm 0.07$ (reduced $\chi^2 = 0.2$). These results are qualitatively consistent with previous *Herschel* studies at high redshift (e.g. Buat et al. 2012; Hilton et al. 2012; Heinis et al. 2014).

5.4 Dissecting the IRX-β relation

Fig. 13 shows the stacked IRX–β relation of *UVJ*-selected star-forming galaxies binned by stellar mass, but without redshift binning, assuming there is no evolution in the redshift-binned results shown in Fig 11. We now include all galaxies with stellar masses $> 10^9 \mathrm{M}_\odot$ (since assuming no redshift dependence, the results should be unaffected by the incompleteness at high redshifts), and can therefore better constrain the IRX-β relationship and its dependence on stellar mass. We find that lower-mass ($< 10^{11} \mathrm{M}_\odot$) star-forming galaxies are generally consistent with either the Meurer et al. (1999) or the Overzier et al. (2011) relation, except in the reddest bins ($\beta > 0$). Galaxies with $M_* > 10^{11} \mathrm{M}_\odot$ appear to have much higher obscuration than predicted by either of these relations, but are closer to the relation of Kong et al. (2004). Following these earlier works, we model the IRX–β relation in each of our stellar-mass bins as

$$\frac{L_{IR}}{L_{UV}} = \left(10^{0.4 A_{FUV}} - 1.0\right) \frac{BC_{FUV}}{BC_{FIR}} \quad (7)$$

where $A_{FUV} = p_0 + p_1\beta$ is the attenuation at 1600Å, and BC_{FUV} and BC_{FIR} are the bolometric corrections in the two wavebands, taken to be 1.68 and 1.0 respectively (Meurer et al. 1999; Overzier et al. 2011). Best-fitting parameters are shown in Table 3, and the fits are shown in Fig. 13. We note that the lowest-mass bin is entirely consistent with the Meurer et al. (1999) relation ($A_{FUV} = 4.43 + 1.99\beta$) and is essentially identical to the Overzier et al. (2011) relation ($A_{FUV} = 4.01 + 1.81\beta$). The fit to the intermediate-mass bin is not formally consistent with either of these, although it is apparent from Fig. 13 that it would be consistent if the single data point at $\beta > 1$ were excluded from the fit (since we noted earlier that the reddest bins at low redshift may be contaminated by passive galaxies despite the *UVJ* selection). The highest-mass bin ($M_* > 10^{11} \mathrm{M}_\odot$) is inconsistent with these relations, since it is both flatter and higher in normalization: high-mass galaxies have higher obscuration than would be predicted from their UV slope, and this is particularly true for bluer UV slopes. However, as with the intermediate-mass sample, if we excluded the reddest bin from this fit we would obtain a slope similar to that of the Meurer et al. relation, and the overall relation for this highest-mass bin would be similar to that of Kong et al. (2004). Deviations from the Meurer et al. relation can be explained by several possible variables, for example star-formation history (Kong et al. 2004; Casey et al. 2014), metallicity (Castellano et al. 2014), the bolometric correction (Calzetti et al. 2000) or the UV extinction curve (Overzier et al. 2011; Bouwens et al. 2012; Buat et al. 2012; Salmon et al. 2015a; Safarzadeh et al. 2016). The Meurer et al. and Kong et al. relations are calibrated on local starburst galaxies, but Kong et al. showed that the relation becomes less tight and has lower normalization with decreasing ratio of present/past-averaged SFR (i.e. more quiescent, less burst-dominated star-formation). This may indicate that the more massive star-forming galaxies in the sample are more likely to be starbursts. However, a lower value of BC_{FIR} in more massive galaxies could also explain the deviation; this would mean that a fraction of the energy emitted in the FIR is not associated with star formation, but may be heated by older stellar populations or even AGN.

Buat et al. (2012) studied the IRX–β relationship at high redshift using *Herschel* data, and also found that the relationship was broadly consistent with that from Overzier et al. (2011), although with broad scatter. They did not investigate whether the deviations correlated with mass,

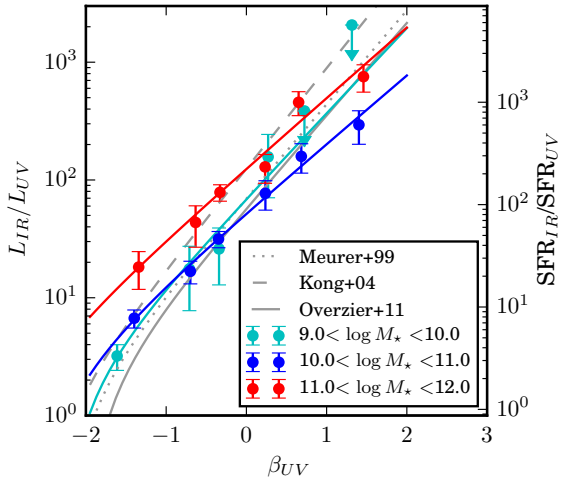

Figure 13. Average IRX as a function of UV slope β in stacked data across all redshifts, binned by stellar mass. Coloured lines show the best-fitting linear models described in Table 3, and grey dotted, dashed and dot-dashed lines show the relationships fitted by Meurer et al. (1999), Kong et al. (2004) and Overzier et al. (2011).

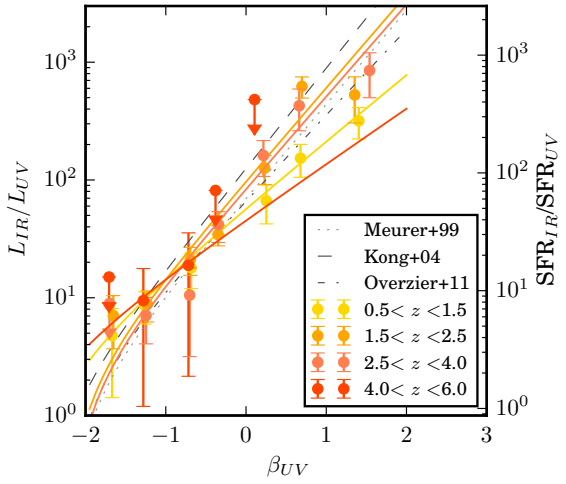

Figure 14. Average IRX as a function of UV slope β in stacked data across all stellar masses, binned by redshift. Coloured lines show the best-fitting linear models described in Table 3, and grey lines are as in Fig. 13.

although they did show that it correlated with L_{IR}, and depended on the best-fitting slope of the dust-attenuation law from their SED fitting. Talia et al. (2015) measured a lower and flatter relationship in a spectroscopic sample at $1 < z < 3$, but their sample showed a large dispersion and covered a narrow range of β, hence the uncertainties are large. Oteo (2014) used stacked *Herschel* data to investigate the IRX–β relationship in UV-selected galaxies at $0 < z < 3$, and found a strong evolution in the normalization with redshift, which is also supported by Pannella et al. (2015) in *Herschel* stacking of a mass-selected sample at $0.5 < z < 4$. Fig. 14 shows the IRX–β relations that we measure as a function of redshift for all $M_* > 10^{10}\mathrm{M}_\odot$ galaxies stacked in four redshift bins. Best-fitting parameters are listed in Table 3. There does appear to be a steepening of the slope between $z < 1.5$ and $z > 1.5$, but there is no evidence for any evolution at higher redshifts (although the constraint at $z > 4$ is very weak). We note that the flatter slope at $z < 1.5$ could also be a sign of contamination of this bin by passive galaxies.

Current evidence for evolution in the IRX–β relation appears contradictory. Our results appear to suggest weak, if any, evolution in a stellar-mass-selected sample, and are consistent with results from *Herschel* in the UV-selected samples of Buat et al. (2012) at $1 < z < 2$ and Reddy et al. (2012) at $1.5 < z < 2.6$ (which were found to be consistent with the low-redshift relations of Overzier et al. (2011) and Meurer et al. (1999) respectively). However, they do not support the strong evolution over $0 < z < 3$ in the UV-selected sample of Oteo (2014). A recent study of K-band-selected galaxies at $1 < z < 3$ by Forrest et al. (2016) indicated a relationship that is higher and steeper than both Meurer et al. and our own relations ($A_{FUV} = 5.05 + 2.39\beta$, c.f. Table 3), which may be attributed to differences in the analysis, in particular their use of composite SEDs fitted to individual *Spitzer* and *Herschel* FIR measurements, in contrast to our stacking technique. Conversely, Bouwens et al. (2016) used stacking in ALMA continuum imaging and found that LBGs with $\log M_*/\mathrm{M}_\odot > 9.75$ at $z = 2$–3 lie below the Meurer et al. relation, indicating an SMC-like dust extinction law, and that LBGs with lower masses lie even lower than this. Turning to high-redshift Lyman-break-selected samples, Coppin et al. (2015) stacked *Herschel* and SCUBA-2 data for $z \sim 3$ LBGs and found that their IRX–β relation was slightly above, but broadly in agreement with the Meurer et al. relation (within the statistical error bars). They also found that LBGs with higher stellar masses were offset further above the Meurer et al. relation; their high-mass bin ($M_* > 10^{10}\mathrm{M}_\odot$) lies systematically above our data in the same mass range in Fig. 13, but is broadly consistent given the error bars. Álvarez-Márquez et al. (2016) also used stacking of $z \sim 3$ LBGs to indicate that they lie close to the low-redshift relations. At higher redshifts, Capak et al. (2015) observed LBGs at $z = 5$–6 with ALMA, and found that they lie well below the local relations (in a sample of 9 LBGs, selected to probe

Table 3. Best-fitting parameters of the model for IRX(β) given by equation (7), with $A_{FUV} = p_0 + p_1\beta$, in the three stellar mass bins shown in Fig. 13, and the four redshift bins shown in Fig. 14.

$\log M_*/\mathrm{M}_\odot$	p_0	p_1
9.0 – 10.0	4.0 ± 0.4	1.8 ± 0.3
10.0 – 11.0	3.7 ± 0.1	1.5 ± 0.1
11.0 – 12.0	4.7 ± 0.1	1.5 ± 0.1
z	p_0	p_1
0.5 – 1.5	3.9 ± 0.1	1.4 ± 0.1
1.5 – 2.5	4.4 ± 0.1	2.0 ± 0.1
2.5 – 4.0	4.3 ± 0.2	2.0 ± 0.2
4.0 – 6.0	3.6 ± 1.2	1.2 ± 1.0

a range of UV luminosities above L^*), while Schaerer et al. (2015) found that $z \sim 7$ LBGs lie on or below the low-redshift relations (based on upper limits on IRX from ALMA and PdBI observations). In contrast, Smit et al. (2015) inferred an IRX–β relation that lies above the Meurer et al. relation, and also above our own ($A_{FUV} = 4.98 + 1.99\beta$, c.f. Table 3), by comparing Hα and UV data in a spectroscopically-selected sample at $3.8 < z < 5$. Their result is more in agreement with the results of Forrest et al. (2016).

Overall, the results indicate a general disparity between different samples, and there is certainly a large amount of scatter in the relation at all redshifts (e.g. Kong et al. 2004; Overzier et al. 2011; Casey et al. 2014), so that differences between reported correlations probably owe much to the effects of sample selection. For example, Casey et al. (2014) showed that *Herschel*-selected dusty star-forming galaxies at $0 < z < 3.5$ lie above the IRX–β relation of local star-forming galaxies, with a large amount of scatter (similar to our FIR-detected sample), and are therefore bluer than expected for their measured obscuration. Similar results have been discussed by Howell et al. (2010), and the evidence points towards a dependence on the ratio of present to past-averaged star-formation rates (Kong et al. 2004), and on the effects of geometry, which can lead to decoupling between the UV- and FIR-bright regions within galaxies (Howell et al. 2010; Casey et al. 2014). Samples selected by UV luminosity, stellar mass, or Lyman-break colours may therefore show variations with mass or redshift simply as a result of differing sample biases.

5.5 The total SFR density of massive galaxies

We estimate the total SFR density (SFRD) of galaxies in each of the mass/luminosity/redshift bins (described in Section 4.3) by taking the sum of SFRs of galaxies in the bin and dividing by the volume of the bin, given by the difference in comoving volumes at the upper and lower redshift boundary, scaled to the sampled sky area. We compute this only for galaxies with $M_* > 10^{10}\mathrm{M}_\odot$, since the sample is complete in that range (see Section 3.5), and we include all galaxies regardless of *UVJ* colours, in order to measure the full integrated SFRD. The results are shown in Fig. 15, and are divided into unobscured SFRD (from raw UV luminosity; blue squares) and obscured SFRD (from FIR luminosity; red circles), as well as the total (large black diamonds). The data points for the full mass-selected sample are linked by solid lines, but in the lower panel of Fig. 15 we also show the contribution to the SFRD from the UV-luminous subset ($M_{\mathrm{UV}} < M^*_{\mathrm{UV}}$; dashed lines), and from the FIR-luminous subset (450μm $S/N > 3$; dotted lines).

Among massive galaxies, the SFRD is dominated by obscured star formation at all redshifts by a factor of 10 or more (red versus blue solid lines). The FIR-luminous subset that is detected at 450μm contributes close to a third of this SFRD (black dotted line). This roughly corresponds to sources with $S_{450} > 3$ mJy and SFR $\gtrsim 100\mathrm{M}_\odot\mathrm{yr}^{-1}$, although the limiting SFR of this sample does evolve with redshift. The UV-luminous subset ($M_{\mathrm{UV}} < M^*_{\mathrm{UV}}$; black dashed line) contributes a slightly smaller fraction of around one fifth at all redshifts. This is partly because these objects are rare, but also because they do not generally have high SFRs compared with typical galaxies of a similar mass (see Fig. 7). In fact the FIR-bright galaxies (black dotted line) are much rarer, but contribute a significantly higher fraction of the total SFRD, at least at $z < 4$ (see also Le Floc'h et al. 2005; Murphy et al. 2011; Casey et al. 2012, 2013). Notably, even in the UV-luminous sample, the SFRD is dominated by the obscured portion (red versus blue dashed lines), although this becomes less dominant with increasing redshift. For the FIR-detected sample, we find that the SFRD is almost completely obscured; less than one per cent of their SFRD is detected in the UV (not shown in the figure).

In the upper panel of Fig. 15 we show SFRD measurements from the literature compiled by Madau & Dickinson (2014); these are indicated by the light blue and pink symbols (representing UV and FIR measurements respectively). These have been recalibrated from a Salpeter (1955) to a Kroupa & Weidner (2003) IMF for comparison with our results, by multiplying by a factor 0.61. Our data provide a new, direct measurement of the obscured SFRD of stellar-mass-selected galaxies at $z \sim 5$, highlighting the importance of our combination of high-resolution SCUBA-2 maps and high-quality multi-wavelength prior catalogues. Previous studies with *Herschel* have generally reached only up to $z \lesssim 4$ due to the much higher confusion limit, either using stacking (e.g. Viero et al. 2013; Schreiber et al. 2015) or by extrapolating and integrating measured luminosity functions of detected sources (e.g. Burgarella et al. 2013; Gruppioni et al. 2013). The selection and identification of bright, red submm sources can provide lower limits on the obscured SFRD at $z > 4$ (Dowell et al. 2014; Shu et al. 2016), but it is impossible to reliably estimate the total SFRD from these results which only probe the very bright end of the IR LF. However, it has already been shown that SCUBA-2 can offer a view of the $z > 4$ obscured SFRD in SMGs (Casey et al. 2013)and stacked LBGs (Coppin et al. 2015), and we have extended this view to include a wider sample of stellar-mass-selected galaxies. We will directly compare our results to the literature in the next section.

5.6 Recovering the full SFR density

In Fig. 15 we plotted our direct measurements of the SFRD using a complete sample of massive galaxies with $M_* > 10^{10}\mathrm{M}_\odot$. We now attempt to recover the full SFRD by estimating the additional contribution from galaxies with lower masses. The full unobscured SFRD up to high redshifts is well-characterised in the literature without explicit mass limits, by integration of the UV luminosity function (LF; Bouwens et al. 2007, 2012, 2015; Cucciati et al. 2012; Smit et al. 2012; McLure et al. 2013; Duncan et al. 2014; Finkelstein et al. 2015; Mashian et al. 2015; Parsa et al. 2016; McLeod et al. 2016). We use the UV LFs from Parsa et al. (2016) in CANDELS GOODS-S and the Hubble Ultra-Deep Field (HUDF) to correct for the fraction of the unobscured SFRD that is missing from our direct measurements in Fig. 15. To estimate this, we match the catalogue of UV absolute magnitudes from Parsa et al. (2016) with the catalogue of stellar masses in the same field from Mortlock et al. (2016, in preparation). The catalogue from Parsa et al. contains photometric redshifts in the range $0 < z < 6$ (although only $1.5 < z < 4.5$ are included in their analysis). In each of our redshift bins, we then measure (i) the integrated luminosity density of galaxies with $M_* > 10^{10}\mathrm{M}_\odot$, down to

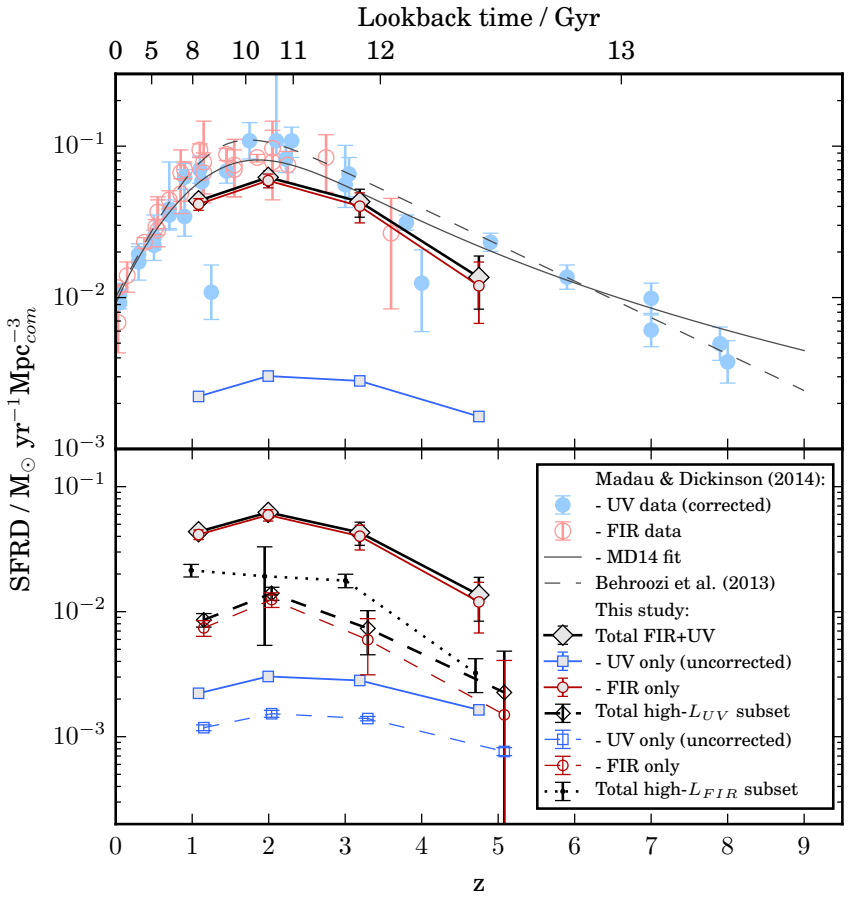

Figure 15. Evolution with redshift of the comoving cosmic SFR density (SFRD) of galaxies with $M_* > 10^{10}$M$_\odot$, as estimated from the luminosities in the rest-frame UV (blue), FIR (red) and the total (black). The SFRD is estimated from the total SFR of all objects in a given bin of each sample. In the lower panel, these are broken down as follows: solid lines are the mass-limited sample ($M_* > 10^{10}$M$_\odot$); dashed are UV-luminosity-limited ($M_{\rm UV} < M^*$); and dotted are detected at 450μm (approximately FIR-luminosity-limited at SFR$\gtrsim 100$ M$_\odot$yr^{-1}). In the upper panel, pastel coloured symbols represent results from the literature compilation by Madau & Dickinson (2014) from rest-frame UV data (blue solid symbols) and FIR data (pink open symbols). These have been converted from a Salpeter IMF to a Kroupa IMF by multiplying by 0.61. Thin solid and dashed lines represent the best-fitting models to the data compiled by Madau & Dickinson (2014) and by Behroozi et al. (2013b) respectively (also scaled to a Kroupa IMF).

the same $M_{\rm UV}$ limit as our bins (see Section 4.3); and compare this with (ii) the integrated luminosity density in the full UV LF from Parsa et al. (2016), down to a constant limit of $M_{\rm UV} = -15$ at all redshifts. The ratio (ii)/(i) gives the correction factor to our direct measurement of the unobscured SFRD in our data set. Assuming a constant limiting $M_{\rm UV} = -15$ at all redshifts facilitates comparisons across all redshifts. This limit was chosen because Parsa et al. (2016) showed that it effectively captures the full luminosity density. The derived correction factors are 5.6, 8.0, 10.8 and 3.2 in the four redshift bins ($0.5 < z < 1.5$, $1.5 < z < 2.5$, $2.5 < z < 4$ and $4 < z < 6$) respectively. We apply an analogous technique to correct our measurements of the unobscured SFRD in the UV-luminous ($M_{\rm UV} < M^*_{\rm UV}$) bins. The ratio of the total UV density at $M_{\rm UV} < M^*_{\rm UV}$ to that in massive galaxies ($M_* > 10^{10}$M$_\odot$) provides the correction factors of 3.3, 5.2, 7.5, 5.3 in the four redshift bins respectively.

The results, in Fig. 16, show that accounting for galaxies with $M_* < 10^{10}$M$_\odot$ brings the unobscured SFRD in line with the obscured SFRD at $z > 3$. This is because most of the unobscured star formation in the Universe occurs within low-mass galaxies, while the star formation in more massive galaxies is mostly obscured. Now we see that while the unobscured SFRD peaks at $z \sim 3$, the obscured (and therefore the total) SFRD peaks later, at $z \sim 2$. While the obscured SFRD dominates at $z < 3$, the obscured fraction appears to be close to 50 per cent at higher redshifts, and in UV-luminous galaxies the majority of the star formation is unobscured at $z > 3$.

However, we have not yet accounted for the obscured SFRD at lower masses. The monotonic trends in Figs. 7 and 12 (notwithstanding incompleteness-related bias at low masses) indicate that the average obscuration fraction (SFR$_{\rm IR}$/SFR$_{\rm UV}$) at $M_* < 10^{10}$M$_\odot$ is $\lesssim 1$ at all redshifts, and therefore that the SFRD at low masses is predominantly unobscured. While our sample is incomplete at low masses and high redshifts, we can nevertheless test this assertion by including all galaxies down to 10^9M$_\odot$ in our SFRD measurements, and calculating obscured SFRD from the sum of

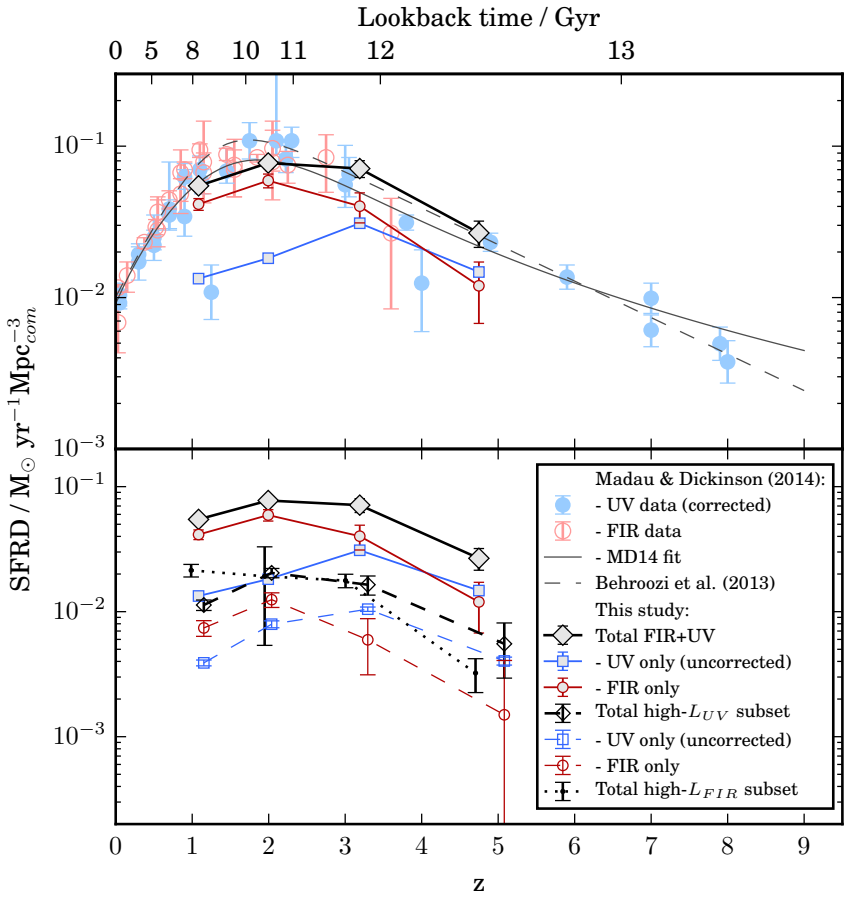

Figure 16. Estimated total SFRD as a function of redshift, as in Fig. 15, with corrections to account for the contribution from galaxies beneath the stellar-mass cut. Both panels show the total SFRD (black solid lines), the unobscured portion from the rest-frame UV data (blue), and the obscured portion from the FIR (red). The lower panel also shows the contribution from the UV-luminosity-limited sample ($M_{\rm UV} < M^*$; black dashed lines), broken down into unobscured (blue dashed) and obscured (red dashed); and the contribution from 450μm-detected sources (black dotted line), which are almost completely obscured.

$\rm SFR_{IR}/V_{max}$ (where $V_{\rm max}$ is defined as the maximum volume within which that galaxy could reside and have been selected in our sample). Doing so, our obscured SFR values increase by a factor of roughly 1.25 in each redshift bin, a very small change to the data shown in Fig. 16 (note the logarithmic scale). By extension, the contribution to obscured SFRD from galaxies with $M_* < 10^9 \rm M_\odot$ can be assumed to be negligible.

Fig. 16 reveals a transition from an early Universe dominated by unobscured star formation to a Universe dominated by obscured star formation at "cosmic noon" (the peak of SFRD at $z \sim 2.5$); yet Fig. 7 indicates that obscuration at a given stellar mass is not redshift-dependent. The reason for this is the stellar-mass dependence of obscuration. At $z > 3$ a greater fraction of the SFRD is contributed by lower-mass galaxies, which are less obscured, but by $z \sim 2$ the high-mass end of the stellar mass function has built up and the SFRD is dominated by galaxies with $M_* > 10^{10} \rm M_\odot$, which have heavily obscured star formation.

Several interesting comparisons can be made between our results in Fig. 16 and data from the literature. At $z < 3$, our measurement of the total SFRD is very close to the fit to the literature compilation by Madau & Dickinson (2014), although at $z > 3$ we measure slightly higher values that are closer to the functional form of Behroozi et al. (2013b), and significantly lower than the earlier compilation of (Hopkins & Beacom 2006; not shown here).[11] A great deal of scatter is seen between different studies around the peak of the cosmic star formation history (for a review, see Behroozi et al. 2013b; Madau & Dickinson 2014), as a result of the broad range of observational techniques that have been employed (IR+UV; dust-corrected UV; 1.4GHz; Hα; etc). In contrast, the chief tracer that has been available at $z > 4$ is the UV LF, which relies on an assumed universal IRX–β relationship. In Section 5.4 we showed that this relationship may be mass-dependent and is not necessarily always consistent with the Meurer et al. (1999) relation that is commonly assumed (an issue raised by many other high-redshift studies,

[11] Note that the offset between the Behroozi et al. (2013b) curve and the Madau & Dickinson (2014) curve may be explained by the fact that the former is estimated from the integral over all halo masses in their model, while the latter is estimated by integrating the IR and UV LFs down to $0.03\,L^*$.

e.g. Buat et al. 2012; Casey et al. 2014; Oteo 2014; Pannella et al. 2015; Talia et al. 2015). In spite of this, and in spite of the significant deviations between the dust-corrected UV SSFRs and the UV+IR SSFRs in Fig. 9, we find that the total UV+IR SFRD in Fig. 16 is surprisingly close to the UV-based determinations at $z > 4$ from studies such as Bouwens et al. (2012), which rely on an assumed dust correction based on Meurer et al. (1999). In fact, the increasing dominance of the unobscured SFRD between $z \sim 3$ and $z \sim 5$ can only strengthen our confidence in determining the continuing evolution of the SFRD towards $z \sim 10$.

On the other hand, accounting for obscuration is key to accurately measuring the SFRD at $1 < z < 4$. In comparison to our results, Burgarella et al. (2013) measured a higher obscured SFRD (and a higher obscured/unobscured ratio) at $2 < z < 4$, by integrating FIR LFs from *Herschel* (Gruppioni et al. 2013) and UV LFs from Cucciati et al. (2012). Their results are particularly subject to uncertainties in the faint end of both LFs at $z > 2$, and are consistent with our results within these uncertainties. Viero et al. (2013) and Schreiber et al. (2015) both measured the obscured SFRD in stacked *Herschel* data at $0 < z < 4$. Viero et al. found that it closely followed the Behroozi et al. (2013b) model, which is slightly above our estimates for the obscured SFRD (but is consistent with our estimate of the *total* SFRD). This slight discrepancy may result from systematic uncertainties in the FIR SEDs, since Viero et al. fitted their data with SEDs that had temperatures increasing with redshift, while we assumed a non-evolving SED based on our best-fitting templates (our error bars account for this uncertainty). Schreiber et al. (2015) similarly measured the total SFRD by stacking *Herschel* data at $0 < z < 4$, and adding in the unobscured UV contribution. Their results closely followed the fits from Madau & Dickinson (2014), and are consistent with our own. Our results also agree with the measurements of the obscured+unobscured SFRD at $1 < z < 5$ from stacking in the ALMA map of the HUDF (Dunlop et al. 2016). Furthermore, independent measurements of the total SFRD at $2 < z < 5$ from OII emission-line surveys (Khostovan et al. 2015) are consistent with our determinations (once accounting for the different IMF assumed in that study).

6 CONCLUSIONS

In this paper we have demonstrated how statistical information about faint, high-redshift source populations can be extracted from confused, single-dish, submm surveys (from S2CLS) with a combination of deep, value-added positional prior catalogues (from CANDELS/3D-HST) and the computational deconfusion technique offered by T-PHOT (see Sections 3–4). Applying these techniques to 230 arcmin^2 of the deepest 450-μm imaging over the AEGIS, COSMOS and UDS CANDELS fields, supported by additional data at 100–850μm, we obtained the following main results:

(i) We detect 165 galaxies at 450-μm with $S/N > 3$ at $S_{450} \gtrsim 3$ mJy, similar to published 450-μm samples from SCUBA-2. The detected sources have a broad redshift distribution at $0 < z < 4$ (median $z = 1.68$), although we also detect four sources at $4 < z < 6$. They span a wide range of stellar masses (typically $9.5 < \log(M_*/\mathrm{M}_\odot) < 11.5$) and UV luminosities (typically $-21 < M_{\rm UV} < -16$). This sample generally traces the highest SFRs at $z < 4$, but exhibit a wide range of obscuration fractions, with $1 < \log(L_{\rm IR}/L_{\rm UV}) < 4$ (see Section 5.1).

(ii) In the stacked results, total SFR (from IR+UV) is strongly correlated with stellar mass at all redshifts, while the raw UV luminosity is a relatively weak indicator of total SFR, especially at $z \gtrsim 2$ (Section 5.1).

(iii) Instead, UV luminosity primarily indicates the level of SFR obscuration, which is also strongly correlated with stellar mass, but does not evolve significantly with redshift at a fixed stellar mass (Sections 5.1, 5.3). We find that the obscuration can be determined from the stellar mass and UV luminosity as $\log(L_{\rm IR}/L_{\rm UV}) = a + b(L_{\rm UV}/10^9\mathrm{L}_\odot) + c(M_*/10^9\mathrm{M}_\odot)$, where $a = 5.9 \pm 1.8$, $b = -0.56 \pm 0.07$, $c = 0.70 \pm 0.07$.

(iv) The average UV+IR SSFRs of *UVJ*-selected star-forming galaxies rise with redshift, and the evolution is steeper for more massive galaxies, indicating that, on average, they stop forming stars earlier. Massive galaxies ($M_* > 10^{10}\mathrm{M}_\odot$) have average SSFRs $\sim 1-2$ Gyr^{-1} at $z \sim 2-4$, and $\sim 3-4$ Gyr^{-1} at $z \sim 5$, in agreement with the most recent studies from both SCUBA-2 and ALMA. We fit the binned data with a bivariate model: $\log SSFR/\mathrm{Gyr}^{-1} = a(z)[\log(M_*/\mathrm{M}_\odot) - 10.5] + b(z)$. The evolution of the slope of the SSFR(M_*) main sequence is given by $a(z) = (-0.64 \pm 0.19) + (0.76 \pm 0.45)\log(1+z)$, while the evolution of the normalization (at $\log(M_*/\mathrm{M}_\odot) = 10.5$) is given by $b(z) = (-9.57 \pm 0.11) + (1.59 \pm 0.24)\log(1+z)$ (Section 5.2).

(v) Dust-corrected SFRs from the UV luminosity and spectral slope (β) can under-estimate the total SFR and overall predict a weaker evolution in the average SSFR as a function of redshift (Section 5.3).

(vi) By stacking *UVJ*-selected star-forming galaxies, we find that massive galaxies ($M_* > 10^{11}\mathrm{M}_\odot$) tend to have higher obscuration for a given UV slope β in comparison with lower-mass galaxies. It is also possible that the IRX–β relation evolves with redshift, although this result could be influenced by contamination of our $z < 1.5$ bins with passive galaxies (Section 5.4).

(vii) Our results provide homogeneous measurements of the obscured SFR density (SFRD) in a highly-complete sample of massive galaxies ($M_* > 10^{10}\mathrm{M}_\odot$) over the redshift range $0.5 < z < 6$, extending beyond what has been possible in previous studies using *Herschel* data. We show that obscured star formation dominates the total SFRD in massive galaxies at all redshifts, and exceeds unobscured star formation by a factor of > 10 (Section 5.5).

(viii) The FIR-detected sample, which is effectively flux-limited at $S_{450} > 3$ mJy and samples the highest star-formation rates at all redshifts (SFR $\gtrsim 100\mathrm{M}_\odot\mathrm{yr}^{-1}$), accounts for approximately one third of the total SFRD over the redshift range $0.5 < z < 6$.

(ix) The most UV-luminous massive galaxies, defined as those with $M_{\rm UV} < M^*_{\rm UV}$ and $M_* > 10^{10}\mathrm{M}_\odot$, account for around one fifth of the total SFRD over the same redshift range, but even in these the majority of the SFRD is obscured at $z \lesssim 3$.

(x) After correcting for the contributions from lower-mass galaxies, the full SFRD from UV+IR data is in good agreement with previous literature estimates both from UV+IR at $z \lesssim 3$ and from UV-only data at $z \sim 5$. This indicates

that UV-selected samples with β dust corrections are successful for calibrating total SFRD at the highest redshifts ($z > 3$), in spite of variations in the IRX–β relation and the lower SSFRs estimated from the dust-corrected UV alone. This is due to the increasing dominance of unobscured star formation at $z \gtrsim 3$ (Section 5.6).

(xi) When accounting for all stellar masses, the SFRD at $z \lesssim 3$ remains dominated by obscured star formation, but at higher redshifts the total obscured and unobscured SFRD are equal, while the most UV-luminous galaxies are predominantly unobscured at $z \gtrsim 3$. The SFRD contribution from the most UV-luminous and the FIR-detected galaxies are approximately equal at $z \sim 2$–3 when including all stellar masses.

(xii) We conclude that the reason for this transition from an early Universe ($z > 3$) dominated by unobscured star formation to a Universe dominated by obscured star formation at cosmic noon ($z \approx 2$) is explained by the increasing contribution of massive galaxies as the high-mass end of the stellar mass function is built up at around $z \sim 2$–3. This is consistent with our observation that the obscuration of star formation at a fixed stellar mass is independent of redshift.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 312725, and also from the European Research Council via the award of an Advanced Grant (JSD). KC acknowledges support from STFC (grant number ST/M001008/1). KK acknowledges support from the Swedish Research Council. The submillimetre data used in this paper were obtained from the JCMT SCUBA-2 Cosmology Legacy Survey. The James Clerk Maxwell Telescope has historically been operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the United Kingdom, the National Research Council of Canada and the Netherlands Organisation for Scientific Research. Additional funds for the construction of SCUBA-2 were provided by the Canada Foundation for Innovation. This research has also made use of data from HerMES project (http://hermes.sussex.ac.uk/). HerMES is a Herschel Key Programme utilising Guaranteed Time from the SPIRE instrument team, ESAC scientists and a mission scientist. The HerMES data was accessed through the Herschel Database in Marseille (HeDaM - http://hedam.lam.fr) operated by CeSAM and hosted by the Laboratoire d'Astrophysique de Marseille. This work uses observations taken by the 3D-HST Treasury Program (GO 12177 and 12328) with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. The analysis in this paper has made use of the open-source Python packages MATPLOTLIB (Hunter 2007) and ASTROPY, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013).

REFERENCES

Álvarez-Márquez J., et al., 2016, A&A, 587, A122
Ashby M. L. N., et al., 2015, ApJS, 218, 33
Astropy Collaboration et al., 2013, A&A, 558, A33
Barro G., et al., 2011, ApJS, 193, 30
Behroozi P. S., Wechsler R. H., Conroy C., 2013a, ApJ, 762, L31
Behroozi P. S., Wechsler R. H., Conroy C., 2013b, ApJ, 770, 57
Bell E. F., 2003, ApJ, 586, 794
Bell E. F., et al., 2005, ApJ, 625, 23
Béthermin M., Dole H., Cousin M., Bavouzet N., 2010, A&A, 516, A43
Béthermin M., et al., 2015, A&A, 573, A113
Blain A. W., Smail I., Ivison R. J., Kneib J.-P., Frayer D. T., 2002, Physics Reports, 369, 111
Boissier S., et al., 2007, ApJS, 173, 524
Bourne N., et al., 2012, MNRAS, 421, 3027
Bouwens R. J., et al., 2012, ApJ, 754, 83
Bouwens R. J., et al., 2015, ApJ, 803, 34
Bouwens R. J., et al., 2016, ArXiv:1606.05280
Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, ApJ, 670, 928
Brammer G. B., et al., 2012, ApJS, 200, 13
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buat V., et al., 2012, A&A, 545, A141
Burgarella D., et al., 2013, A&A, 554, A70
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Capak P. L., et al., 2015, Nature, 522, 455
Casey C. M., et al., 2012, ApJ, 761, 140
Casey C. M., et al., 2013, MNRAS, 436, 1919
Casey C. M., et al., 2014, ApJ, 796, 95
Castellano M., et al., 2014, A&A, 566, A19
Chabrier G., 2003, PASP, 115, 763
Chapin E. L., Berry D. S., Gibb A. G., Jenness T., Scott D., Tilanus R. P. J., Economou F., Holland W. S., 2013, MNRAS, 430, 2545
Chapman S. C., et al., 2000, MNRAS, 319, 318
Chen C.-C., et al., 2015, ApJ, 799, 194
Chomiuk L., Povich M. S., 2011, AJ, 142, 197
Coppin K. E. K., et al., 2015, MNRAS, 446, 1293
Cucciati O., et al., 2012, A&A, 539, A31
Davies L. J. M., Bremer M. N., Stanway E. R., Lehnert M. D., 2013, MNRAS, 433, 2588
Davies L. J. M., Bremer M. N., Stanway E. R., Mannering E., Lehnert M. D., Omont A., 2012, MNRAS, 425, 153
Davies L. J. M., et al., 2016, MNRAS
Decarli R., et al., 2014, ApJ, 780, 115
Dempsey J. T., et al., 2013, MNRAS, 430, 2534
Desert F.-X., Boulanger F., Puget J. L., 1990, A&A, 237, 215
Dobos L., et al., 2012, MNRAS, 420, 1217
Dowell C. D., et al., 2014, ApJ, 780, 75
Duncan K., et al., 2014, MNRAS, 444, 2960
Dunlop J. S., et al., 2016, ArXiv:1606.00227
Dutton A. A., van den Bosch F. C., Dekel A., 2010, MNRAS, 405, 1690
Elbaz D., et al., 2011, A&A, 533, A119
Finkelstein S. L., et al., 2015, ApJ, 810, 71
Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950
Forrest B., et al., 2016, ApJ, 818, L26
Furlong M., et al., 2015, MNRAS, 450, 4486
Geach J. E., et al., 2013, MNRAS, 432, 53
Geach J. E., et al., 2016, submitted
Genel S., et al., 2014, MNRAS, 445, 175
González V., Bouwens R., Illingworth G., Labbé I., Oesch P., Franx M., Magee D., 2014, ApJ, 781, 34
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Frenk C. S., Wilkins S. M., 2013, MNRAS, 429, 1609
Grazian A., et al., 2015, A&A, 575, A96

Grogin N. A., et al., 2011, ApJS, 197, 35
Gruppioni C., et al., 2013, MNRAS, 432, 23
Guo Q., et al., 2015, ArXiv:1512.00015
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Hatsukade B., et al., 2016, PASJ, 68, 36
Heinis S., et al., 2014, MNRAS, 437, 1268
Hilton M., et al., 2012, MNRAS, 425, 540
Holland W. S., et al., 2013, MNRAS, 430, 2513
Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142
Howell J. H., et al., 2010, ApJ, 715, 572
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Hurley P. D., et al., 2016, ArXiv:1606.05770
Ilbert O., et al., 2013, A&A, 556, A55
Johnston R., Vaccari M., Jarvis M., Smith M., Giovannoli E., Häußler B., Prescott M., 2015, MNRAS, 453, 2540
Kajisawa M., et al., 2009, ApJ, 702, 1393
Kauffmann G., et al., 2003, MNRAS, 341, 33
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Khostovan A. A., Sobral D., Mobasher B., Best P. N., Smail I., Stott J. P., Hemmati S., Nayyeri H., 2015, MNRAS, 452, 3948
Kobayashi M. A. R., Inoue Y., Inoue A. K., 2013, ApJ, 763, 3
Koekemoer A. M., et al., 2011, ApJS, 197, 36
Kohno K., et al., 2016, ArXiv:1601.00195
Kong X., Charlot S., Brinchmann J., Fall S. M., 2004, MNRAS, 349, 769
Koprowski M. P., Dunlop J. S., Michalowski M. J., Cirasuolo M., Bowler R. A. A., 2014, MNRAS, 444, 117
Koprowski M. P., et al., 2016, MNRAS, 458, 4321
Kriek M., et al., 2009, ApJ, 700, 221
Kroupa P., Weidner C., 2003, ApJ, 598, 1076
Kurczynski P., Gawiser E., 2010, AJ, 139, 1592
Le Floc'h E., et al., 2005, ApJ, 632, 169
Lehnert M. D., van Driel W., Le Tiran L., Di Matteo P., Haywood M., 2015, A&A, 577, A112
Lutz D., et al., 2011, A&A, 532, A90
MacKenzie T., Scott D., Swinbank M., 2015, ArXiv:1512.04615
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Magnelli B., et al., 2013, A&A, 553, A132
Marchesini D., et al., 2010, ApJ, 725, 1277
Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765
Marmol-Queralto E., McLure R. J., Cullen F., Dunlop J. S., Fontana A., McLeod D. J., 2015, ArXiv:1511.01911
Mashian N., Oesch P., Loeb A., 2015, ArXiv:1507.00999
McLeod D. J., McLure R. J., Dunlop J. S., 2016, ArXiv e-prints
McLeod D. J., McLure R. J., Dunlop J. S., Robertson B. E., Ellis R. S., Targett T. A., 2015, MNRAS, 450, 3032
McLure R. J., et al., 2013, MNRAS, 432, 2696
Merlin E., et al., 2015, A&A, 582, A15
Merlin E., et al., 2016, A&Asubmitted
Meurer G. R., Heckman T. M., Calzetti D., 1999, ApJ, 521, 64
Michałowski M., Hjorth J., Watson D., 2010, A&A, 514, A67
Michałowski M. J., Dunlop J. S., Cirasuolo M., Hjorth J., Hayward C. C., Watson D., 2012, A&A, 541, A85
Momcheva I. G., et al., 2015, ArXiv:1510.02106
Mortlock A., Conselice C. J., Bluck A. F. L., Bauer A. E., Grützbauch R., Buitrago F., Ownsworth J., 2011, MNRAS, 413, 2845
Muñoz-Mateos J. C., et al., 2009, ApJ, 701, 1965
Murphy E. J., Chary R.-R., Dickinson M., Pope A., Frayer D. T., Lin L., 2011, ApJ, 732, 126
Murphy E. J., et al., 2011, ApJ, 737, 67
Muzzin A., et al., 2013, ApJ, 777, 18
Nayyeri H., et al., 2014, ApJ, 794, 68
Noeske K. G., et al., 2007, ApJ, 660, L43
Oke J. B., 1974, ApJS, 27, 21
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Oliver S. J., et al., 2012, MNRAS, 424, 1614
Oteo I., 2014, A&A, 572, L4
Overzier R. A., et al., 2011, ApJL, 726, L7
Pannella M., et al., 2015, ApJ, 807, 141
Parsa S., Dunlop J. S., McLure R. J., Mortlock A., 2016, MNRAS, 456, 3194
Peacock J. A., et al., 2000, MNRAS, 318, 535
Pope A., et al., 2008, ApJ, 675, 1171
Price S. H., et al., 2014, ApJ, 788, 86
Reddy N., et al., 2012, ApJ, 744, 154
Reddy N. A., Steidel C. C., Pettini M., Adelberger K. L., Shapley A. E., Erb D. K., Dickinson M., 2008, ApJS, 175, 48
Roseboom I. G., et al., 2010, MNRAS, 409, 48
Roseboom I. G., et al., 2013, MNRAS, 436, 430
Safarzadeh M., Ferguson H. C., Lu Y., Inami H., Somerville R. S., 2015, ApJ, 798, 91
Safarzadeh M., Hayward C. C., Ferguson H. C., 2016, ArXiv:1604.07402
Salmon B., et al., 2015a, ArXiv:1512.05396
Salmon B., et al., 2015b, ApJ, 799, 183
Salpeter E. E., 1955, ApJ, 121, 161
Sargent M. T., Béthermin M., Daddi E., Elbaz D., 2012, ApJ, 747, L31
Schaerer D., Boone F., Zamojski M., Staguhn J., Dessauges-Zavadsky M., Finkelstein S., Combes F., 2015, A&A, 574, A19
Schreiber C., et al., 2015, A&A, 575, A74
Schreiber C., et al., 2016, ArXiv:1606.05354
Schreiber C., Pannella M., Leiton R., Elbaz D., Wang T., Okumura K., Labbé I., 2016, ArXiv:1606.06252
Shu X. W., et al., 2016, ApJS, 222, 4
Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103
Skelton R. E., et al., 2014, ApJS, 214, 24
Smit R., Bouwens R. J., Franx M., Illingworth G. D., Labbé I., Oesch P. A., van Dokkum P. G., 2012, ApJ, 756, 14
Smit R., Bouwens R. J., Labbé I., Franx M., Wilkins S. M., Oesch P. A., 2015, ArXiv:1511.08808
Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, MNRAS, 423, 1992
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Spitler L. R., et al., 2014, ApJ, 787, L36
Stanway E. R., Bremer M. N., Davies L. J. M., Lehnert M. D., 2010, MNRAS, 407, L94
Steinhardt C. L., et al., 2014, ApJ, 791, L25
Straatman C. M. S., et al., 2014, ApJ, 783, L14
Takeuchi T. T., Yuan F.-T., Ikeyama A., Murata K. L., Inoue A. K., 2012, ApJ, 755, 144
Talia M., et al., 2015, A&A, 582, A80
Tasca L. A. M., et al., 2015, A&A, 581, A54
Tomczak A. R., et al., 2016, ApJ, 817, 118
Viero M. P., et al., 2013, ApJ, 779, 32
Wang L., et al., 2014, MNRAS, 444, 2870
Watson D., Christensen L., Knudsen K. K., Richard J., Gallazzi A., Michałowski M. J., 2015, Nature, 519, 327
Whitaker K. E., et al., 2011, ApJ, 735, 86
Whitaker K. E., et al., 2014, ApJ, 795, 104
Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, ApJ, 754, L29
Wilkins S. M., Trentham N., Hopkins A. M., 2008, MNRAS, 385, 687
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879
Wright A. H., et al., 2016, MNRAS, 460, 765
Yajima H., Nagamine K., Thompson R., Choi J.-H., 2014, MNRAS, 439, 3073
Zeimann G. R., et al., 2015, ApJ, 814, 162

APPENDIX A: STACKED SFRS OF STAR-FORMING GALAXIES

In Section 5.1 we showed the results of stacked SFRs from IR+UV data encompassing all galaxies in the sample, divided into bins of redshift, stellar mass and UV absolute magnitude. However, in Section 5.2 we used *UVJ* colour criteria to remove passive galaxies from the sample in order to investigate the SSFR and IR/UV luminosity ratio of star-forming galaxies as a function of stellar mass and redshift. To show the effects of removing passive galaxies on the average SFR as a function of redshift, stellar mass and $M_{\rm UV}$, we reproduce Figs. 7 and 8 after applying the *UVJ* cuts in equation (4) to remove passive galaxies. The results are shown as a function of stellar mass in Fig. A1, and as a function of $M_{\rm UV}$ in Fig. A2.

APPENDIX B: EXAMPLE T-PHOT PARAMETER FILE

```
# TPHOT PARAMETER FILE

#------------------- PIPELINE --------------------#
# Select the desired one or create another:
# 1st PASS:
order           positions, fit, diags, archive

#----------------- CUTOUT STAGE ------------------#
poscat          3DHST_xy_positions.dat
culling         False
relscale        1
cutoutdir       cutouts
cutoutcat       cutouts/_cutouts.cat
normalize       true

#---------------- CONVOLUTION STAGE -------------#

loresfile       AEGIS-CANDELS-4_flux.fits
loreserr        AEGIS-CANDELS-4_rms.fits
errtype         rms
rmsconstant     1
bgconstant      0
maxflag         64
FFTconv true
multikernels false
kernelfile      psf-4.fits
psffile       psf-4.fits
kernellookup    ch1_dancecard.txt
posframe        hires
templatedir     templates
templatecat templates/_templates.cat

#---------------- FITTING STAGE ------------------#

# Filenames:
fitpars          tpipe_tphot.param
tphotcat         lores_tphot.cat_pass1
tphotcell        lores_tphot.cell_pass1
tphotcovar       lores_tphot.covar_pass1
# Control parameters:
fitting  single
dithercell      true
cellmask        true
maskfloor       1e-9
fitbackground   true
writecovar      true
threshold 0.0
linsyssolver lu # [options: ibg, cholesky, lu]
clip false

#----------------- DIAGNOSTICS STAGES ------------#

modelfile       lores_collage_pass1.fits
# Dance:
dzonesize       100
maxshift        1.0
ddiagfile       ddiags.txt
dlogfile        dlog.txt
dancefft true
```

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

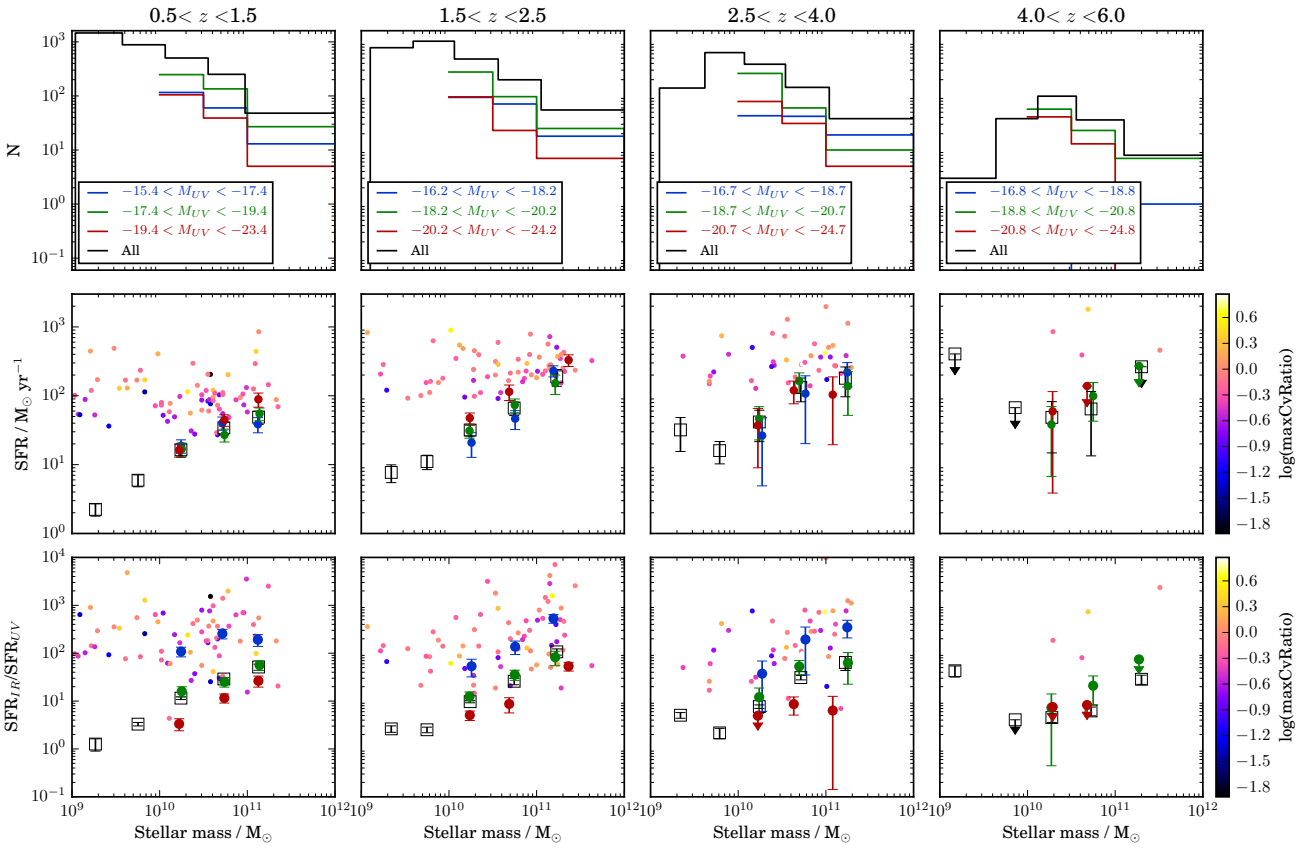

Figure A1. Same as Fig. 7, but excluding passive galaxies as defined by equation (4). From top to bottom: Number N per bin; total $\mathrm{SFR_{UV}} + \mathrm{SFR_{IR}}$; and obscuration ratio $\mathrm{SFR_{IR}/SFR_{UV}}$. Large black squares show the full mass-binned stacks, while large filled symbols with error bars show the stacks divided into bins of M_{UV}, defined relative to M^*_{UV} at the appropriate redshift (Parsa et al. 2016). FIR detections ($S/N > 3$) are shown by small coloured points in which the colour coding indicates *maxCvRatio*.

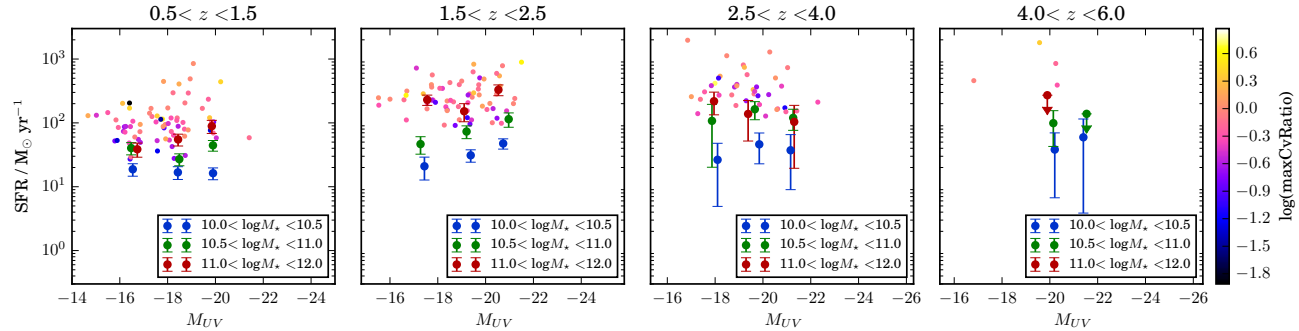

Figure A2. Same as Fig. 8, but excluding passive galaxies as defined by equation (4). The data plotted are the same as in Fig A1, except that colours here indicate the stellar-mass bin.