

# The COSMOS2015 galaxy stellar mass function

## Thirteen billion years of stellar mass assembly in ten snapshots<sup>★</sup>

I. Davidzon<sup>1,2</sup>, O. Ilbert<sup>1</sup>, C. Laigle<sup>3</sup>, J. Coupon<sup>4</sup>, H. J. McCracken<sup>5</sup>, I. Delvecchio<sup>6</sup>, D. Masters<sup>7</sup>, P. Capak<sup>7</sup>, B. C. Hsieh<sup>8</sup>, O. Le Fèvre<sup>1</sup>, L. Tresse<sup>9</sup>, M. Bethermin<sup>1</sup>, Y.-Y. Chang<sup>8,10</sup>, A. L. Faisst<sup>7</sup>, E. Le Floc'h<sup>10</sup>, C. Steinhardt<sup>11</sup>, S. Toft<sup>11</sup>, H. Aussel<sup>10</sup>, C. Dubois<sup>1</sup>, G. Hasinger<sup>12</sup>, M. Salvato<sup>13</sup>, D. B. Sanders<sup>12</sup>, N. Scoville<sup>14</sup>, and J. D. Silverman<sup>15</sup>

(Affiliations can be found after the references)

Received 10 January 2017 / Accepted 17 May 2017

### ABSTRACT

We measure the stellar mass function (SMF) and stellar mass density of galaxies in the COSMOS field up to  $z \sim 6$ . We select them in the near-IR bands of the COSMOS2015 catalogue, which includes ultra-deep photometry from UltraVISTA-DR2, SPLASH, and Subaru/Hyper Suprime-Cam. At  $z > 2.5$  we use new precise photometric redshifts with error  $\sigma_z = 0.03(1+z)$  and an outlier fraction of 12%, estimated by means of the unique spectroscopic sample of COSMOS ( $\sim 100\,000$  spectroscopic measurements in total, more than one thousand having robust  $\zeta_{\text{spec}} > 2.5$ ). The increased exposure time in the DR2, along with our panchromatic detection strategy, allow us to improve the completeness at high  $z$  with respect to previous UltraVISTA catalogues (e.g. our sample is  $> 75\%$  complete at  $10^{10} M_\odot$  and  $z = 5$ ). We also identify passive galaxies through a robust colour–colour selection, extending their SMF estimate up to  $z = 4$ . Our work provides a comprehensive view of galaxy-stellar-mass assembly between  $z = 0.1$  and 6, for the first time using consistent estimates across the entire redshift range. We fit these measurements with a Schechter function, correcting for Eddington bias. We compare the SMF fit with the halo mass function predicted from  $\Lambda$ CDM simulations, finding that at  $z > 3$  both functions decline with a similar slope in the high-mass end. This feature could be explained assuming that mechanisms quenching star formation in massive haloes become less effective at high redshifts; however further work needs to be done to confirm this scenario. Concerning the SMF low-mass end, it shows a progressive steepening as it moves towards higher redshifts, with  $\alpha$  decreasing from  $-1.47^{+0.02}_{-0.02}$  at  $z \approx 0.1$  to  $-2.11^{+0.30}_{-0.13}$  at  $z \approx 5$ . This slope depends on the characterisation of the observational uncertainties, which is crucial to properly remove the Eddington bias. We show that there is currently no consensus on the method to quantify such errors: different error models result in different best-fit Schechter parameters.

**Key words.** galaxies: evolution – galaxies: luminosity function, mass function – galaxies: statistics – galaxies: high-redshift

### 1. Introduction

In recent years, improvements in observational techniques and new facilities have allowed us to capture images of the early universe when it was only a few billion years old. The *Hubble* Space Telescope (HST) has now provided samples of high- $z$  ( $z \gtrsim 3$ ) galaxies selected in stellar mass (Koekemoer et al. 2011; Grogin et al. 2011; Illingworth et al. 2013), which represent a breakthrough similar to the advent of spectroscopic surveys at  $z \sim 1$  more than a decade ago (Cimatti et al. 2002; Davis et al. 2003; Grazian et al. 2006). Indeed, they have a statistical power comparable to those pioneering studies, allowing for the same fundamental analyses such as the estimate of the observed galaxy stellar mass function (SMF). This statistical tool, providing a description of stellar mass assembly at a given epoch, plays a pivotal role in studying galaxy evolution.

We can distinguish between different modes of galaxy growth, for example by comparing the SMF of galaxies divided by morphological type or environment (e.g. Bolzonella et al. 2010; Vulcani et al. 2011; Mortlock et al. 2014; Moffett et al. 2016; Davidzon et al. 2016). Moreover, their rate of stellar mass

accretion changes as a function of  $z$ , being more vigorous at earlier epochs (e.g. Tasca et al. 2015; Faisst et al. 2016a). Thus, the SMF can give an overview of the whole galaxy population, at least down to the limit of stellar mass completeness, over cosmic time. Although more difficult to compute than the luminosity function (LF), the SMF is more closely related to the star formation history of the universe, with the integral of the latter being equal to the stellar mass density after accounting for mass loss (Arnouts et al. 2007; Wilkins et al. 2008; Ilbert et al. 2013; Madau & Dickinson 2014). Moreover, such a direct link to star formation rate (SFR) makes the observed SMF a basic comparison point for galaxy formation models. Both semi-analytical and hydrodynamical simulations are often (but not always, see e.g. Dubois et al. 2014) calibrated against the local SMF (e.g. Guo et al. 2011, 2013; Genel et al. 2014; Schaye et al. 2016); measurements at  $z > 0$  are then used to test theoretical predictions (Torrey et al. 2014; Furlong et al. 2015, and many others).

Deep HST surveys probe relatively small areas, resulting in sample variance significantly greater than ground-based observations conducted over larger fields (Trenti & Stiavelli 2008; Moster et al. 2011). Therefore it is difficult for them to make measurements at low-intermediate redshifts ( $z \lesssim 2$ ) where the corresponding volume is smaller. As a consequence, the literature lacks mass functions consistently measured from the local to the early universe. Such a coherent set of estimates would facilitate those studies probing a wide redshift range (e.g. Moster et al. 2013; Henriques et al. 2015; Volonteri et al. 2015), which at present have to combine miscellaneous datasets.

\* Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium (<http://ultravista.org/>). Based on data produced by the SPLASH team from observations made with the *Spitzer* Space Telescope (<http://splash.caltech.edu>).

To get a continuous view of galaxies' history, one has to combine low- $z$  estimates (e.g. Fontana et al. 2004, 2006; Pozzetti et al. 2010; Ilbert et al. 2010) with SMFs derived at  $z > 3$  (e.g. McLure et al. 2009; Caputi et al. 2011; Santini et al. 2012). Unfortunately, linking them is not an easy task. In particular, samples at different redshifts are built with heterogeneous photometry and selection effects. For instance, at high- $z$ , instead of using photometric redshifts, the widespread approach is based on the “drop-out” technique that selects Lyman-break galaxies (LBGs, Steidel et al. 1996). Even when photometric redshifts are used across the whole redshift range, differences, for example in the method to fit galaxies' spectral energy distribution (SED), may cause systematics in their redshift distribution, or in following steps of the analysis, such as the evaluation of stellar mass (for a comparison among various SED fitting code, we refer to Conroy 2013; Mitchell et al. 2013; Mobasher et al. 2015). Eventually, such inhomogeneity among the joined samples can produce spurious trends in the evolution of the SMF (Marchesini et al. 2009, for a critical assessment of SMF systematics).

Tackling these limitations is one of the main goals of the UltraVISTA survey (McCracken et al. 2012) and the *Spitzer* Large Area Survey with Hyper Suprime-Cam (SPLASH, Capak et al. 2012). These surveys cover the 2 square degrees of the COSMOS field (Scoville et al. 2007) in near and medium IR (NIR and MIR), respectively. With them, our collaboration built a catalogue of galaxies (COSMOS2015) from  $z = 0$  to 6. The COSMOS2015 catalogue has been presented in Laigle et al. (2016), where we showed the gain in terms of large-number statistics (due to the large volume probed) and improved depth (reaching  $K_s = 24.7$  and  $[3.6 \mu\text{m}] = 25.5$ , at  $3\sigma$  in 3'' diameter aperture). The deeper exposure translates to a higher completeness of the sample down to lower stellar masses with respect to previous versions of the catalogue.

In this paper we exploit the COSMOS2015 catalogue (together with exquisite ancillary data available in COSMOS) to derive a galaxy SMF up to  $z \sim 6$ , that is, when the universe was approximately 1 Gyr old. Following galaxy mass assembly across such a large time-span allows one to identify crucial stages of galaxies' lives, from the reionization era (see Robertson et al. 2015), through the “cosmic noon” at  $z \sim 2$  (Madau & Dickinson 2014), until more recent epochs when many galaxies have become red and dead (e.g. Faber et al. 2007). We aim at juxtaposing these key moments to get a global picture, also separating populations of active (i.e. star forming) and passive (quiescent) galaxies.

We organise our work as follows. First, we describe the COSMOS2015 catalogue and the other datasets that we use, with particular attention being paid to sample completeness (Sect. 2). At  $z \leq 2.5$  we rely on the original SED fitting estimates from Laigle et al. (2016), while at higher  $z$  we recompute photometric redshifts ( $z_{\text{phot}}$ , Sect. 3) and stellar masses ( $\mathcal{M}$ , Sect. 4) with an updated SED fitting setup optimised for the  $3 \leq z \leq 6$  range. Since the novelty of this work is the analysis between  $z = 2.5$  and 6 we present in Sect. 5 the SMFs at  $z > 2.5$ , while those at lower redshifts (directly derived from L16) can be found in the Appendix. The evolution of the SMF in the full redshift range, from  $z \sim 6$  down to 0.2, is then discussed in Sect. 6. Eventually, we summarise our work in Sect. 7.

Throughout this paper we assume a flat  $\Lambda\text{CDM}$  cosmology with  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ , and  $h_{70} \equiv H_0/(70 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 1$ . Galaxy stellar masses, when derived from SED fitting, scale as the square of the luminosity distance; therefore, there is a factor  $h_{70}^{-2}$  kept implicit throughout this paper (we refer to Croton 2013,

for an overview on cosmology conversions and their conventional notation). Magnitudes are in the AB system (Oke 1974).

## 2. Dataset

A description of our dataset is summarised in Sect. 2.1. Section 2.2 offers a comprehensive discussion about its completeness as a function of flux in IRAC channel 1 (i.e., at  $\sim 3.6 \mu\text{m}$ ). The core of our analysis is the COSMOS2015 catalogue, recently published in Laigle et al. (2016, L16 in the following); other COSMOS datasets provide additional information. A complete list of the surveys carried out by the collaboration can be found on the official COSMOS website<sup>1</sup>.

### 2.1. Photometry

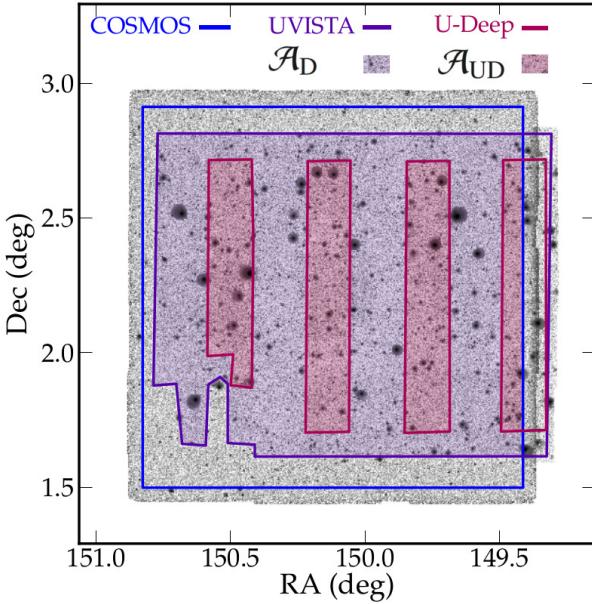
One cornerstone of the COSMOS2015 photometry comes from the new  $Y$ ,  $J$ ,  $H$ , and  $K_s$  images from the second data release (DR2) of the UltraVISTA survey (McCracken et al. 2012), along with the  $z^{++}$  band from Suprime-Cam at Subaru (Miyazaki et al. 2012). These images were added together to build a stacked detection image, as explained below.

The catalogue also includes the broadband optical filters  $u^*$ ,  $B$ ,  $V$ ,  $r$ ,  $i^+$ , and 14 intermediate and narrow bands. In NIR, UltraVISTA is complemented by the  $y$  band images from the Hyper Suprime-Cam (HSC), as well as  $H$  and  $K_s$  from WIRCam (at the Canada-France-Hawaii Telescope). The point-spread function (PSF) in each band from  $u^*$  to  $K_s$  is homogenised, so that the fraction of flux in a 3'' diameter aperture suffers from band-to-band seeing variations by less than 5% (see Fig. 4 in L16). Space-based facilities provided data in near-UV (from the GALEX satellite, Zamojski et al. 2007) and MIR (from IRAC on board the *Spitzer* Space Telescope), along with high-resolution optical images (ACS camera on board HST, see Sect. 3.2). Galaxies with an X-ray counterpart from XMM (Brusa et al. 2007) or *Chandra* (Marchesi et al. 2016) are excluded from the following analysis as their photometric redshifts, or the stellar mass estimates, would be likely corrupted by contamination of their active galactic nuclei (AGN). They represent less than 1% of the whole galaxy sample. The entire photometric baseline of COSMOS2015 is summarised in Table 1 of L16.

*Spitzer* data represent another pillar of this catalogue, probing the whole COSMOS area at  $3.6\text{--}8.0 \mu\text{m}$ , that is, the wavelength range where the redshifted optical spectrum of  $z \gtrsim 3$  galaxies is observed. Such a crucial piece of information mainly comes from SPLASH but other surveys are also included, in particular the *Spitzer*-Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (S-CANDELS, Ashby et al. 2015). Further details about how *Spitzer*/IRAC photometry was extracted and harmonised with the other datasets can be found in L16.

Compared with the previous version of the catalogue (Ilbert et al. 2013) the number of sources doubled because of the longer exposure time of the UltraVISTA DR2 in the so-called “Ultra-Deep” stripes (hereafter indicated with  $\mathcal{A}_{\text{UD}}$ , see Fig. 1). In that area of  $0.62 \text{ deg}^2$  we reach a  $3\sigma$  limiting magnitude (in a 3'' diameter aperture)  $K_{\text{lim,UD}} = 24.7$ , while in the remaining “Deep” area (dubbed  $\mathcal{A}_D = 1.08 \text{ deg}^2$ ) the limit is  $K_{\text{lim,D}} = 24.0$ . The larger number of detected sources in DR2 is also due to the new  $\chi^2$  stacked image produced in L16. Image stacking is a panchromatic approach for identification of galaxy sources, presented for the first time in Szalay et al. (1999). With respect to

<sup>1</sup> <http://cosmos.astro.caltech.edu>



**Fig. 1.** Layout of the COSMOS field. The image in the background is the  $\chi^2$ -stacked  $zYJHK_s$  image. The COSMOS 2 deg $^2$  field is enclosed by a blue line, while the UltraVISTA survey is inside the purple contour. UltraDeep stripes, where UltraVISTA exposure time is higher, are delimited by magenta lines. A purple (magenta) shaded area shows the Deep (UltraDeep) region used in this paper.

the previous UltraVISTA (DR1) stacking, in L16 we co-add not only NIR images but also the deeper  $z^{++}$  band, using the code SWarp (Bertin et al. 2002). Pixels in the resulting  $zYJHK$  image are the weighted mean of the flux in each stacked filter. As a result, the catalogue contains  $\sim 6 \times 10^5$  objects within 1.5 deg $^2$ , 190 650 of them in  $\mathcal{A}_{UD}$ . In L16 we also show the good agreement of colour distributions, number counts, and clustering with other state-of-the-art surveys.

For each entry of the COSMOS2015 catalogue we search for a counterpart in the four *Spitzer*-IRAC channels using the code IRACLEAN (Hsieh et al. 2012)<sup>2</sup>. The procedure is detailed in L16. In brief, positional and morphological information in the  $zYJHK_s$  detection image is used as a prior to identify IRAC sources and recover their total flux. In this latest version, IRACLEAN produces a weighing scheme from the surface brightness of the prior, to correctly deblend objects that are located at less than  $\sim 1$  FWHM of the IRAC PSF from one another. For each source, a flux error is estimated by means of the residual map, that is, the IRAC image obtained after subtracting the flux associated to detections.

## 2.2. Flux limits and sample completeness at high redshift

We aim to work with a flux-limited sample to restrict the analysis to a sample sufficiently complete with photometric errors sufficiently small. In L16 the completeness as a function of stellar mass has been derived in bins of  $K_s$  magnitudes, but this choice is not suitable for the present analysis, which extends to  $z \sim 6$ . Up to  $z \sim 4$ , a  $K_s$ -band selection is commonly used to derive a completeness limit in stellar mass (e.g. Ilbert et al. 2013; Muzzin et al. 2013a; Tomczak et al. 2014), but at higher redshifts this filter probes a rest-frame range of the galaxy spectrum

particularly sensitive to recent star formation. Indeed, the Balmer break moves to wavelengths larger than 2  $\mu\text{m}$  and most of the stellar light coming from K- and M-class stars is observed in the IRAC channels. This makes a [3.6] selection suitable at  $z > 4$ . In this paper we apply a selection in  $K_s$  or [3.6] depending on the redshift, always choosing the most direct link between stellar light and mass. Moreover, we show in Appendix B that even between  $z \sim 2$  and 4, where in principle both bands can be used, a cut in [3.6] is recommended. Thus, for our analysis at  $2.5 < z < 6$ , we extract from the parent catalogue a sample of galaxies with magnitude  $[3.6] < [3.6]_{lim}$ .

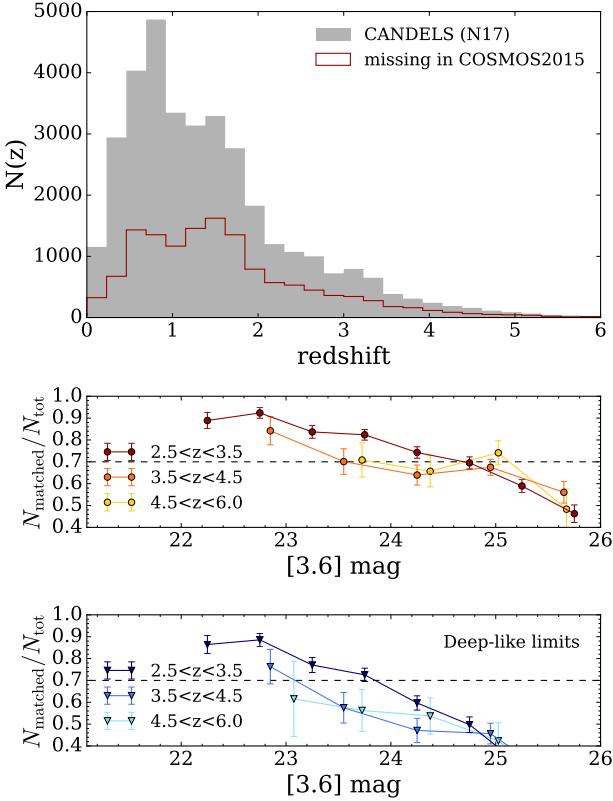
Determining  $[3.6]_{lim}$  is not as straightforward as for the  $K_s$  band. The nominal  $3\sigma$  depth (equal to 25.5 mag for a 3'' diameter aperture) has been calculated by means of the rms map of the [3.6] mosaic, after removing detected objects. However our sources were originally found in the co-added image, so the completeness of the final sample depends not only on possible issues in the IRAC photometric extraction (due e.g. to confusion noise) but also in  $zYJHK_s$ . For instance, we expect to miss red galaxies with  $[3.6] \ll 25.5$  but too faint to be detected in NIR. Their impact should not be underestimated, given the mounting evidence of strong dust extinction in high- $z$  galaxies (e.g. Casey et al. 2014a; Mancini et al. 2015). This is a limitation in any analysis that uses optical/NIR images as a prior to deblend IR sources (e.g. Ashby et al. 2013, 2015). Such an approach is somehow necessary, given the lower resolution of the IRAC camera, but exceptions do exist (e.g. Caputi et al. 2011, where IR photometry is extracted directly from 4.5  $\mu\text{m}$  images without any prior).

To estimate  $[3.6]_{lim}$  we make use of the catalogue built by Nayyeri et al. (2017, hereafter referred as N17) in the 216 arcmin $^2$  of the CANDELS-COSMOS field (Grogin et al. 2011; Koekemoer et al. 2011). Since CANDELS falls entirely in our Ultra-Deep area, N17 can be used to directly constrain  $[3.6]_{lim,UD}$ . The authors rely on  $F160W$  images ( $\sim 1.6 \mu\text{m}$ ) from the HST/WFC3 camera and extract IRAC sources using the software TFIT (Papovich et al. 2001; Laidler et al. 2007). Their approach is similar to Galametz et al. (2013), who derive the UV-to-IR photometry in another CANDELS field overlapping with the Ultra-Deep Survey (UDS) of UKIDSS. The  $5\sigma$  limiting magnitude in the  $F160W$  band is 26.5, while the data from *Spitzer* are the same as in L16. Then, we can test the effects of a different extraction algorithm and sensitivity depth of the prior.

First, we match galaxies from COSMOS2015 and N17 (within a searching radius of 0.8'') and compare their photometric redshift estimates and [3.6] magnitudes to check for possible bias. We confirm the absence of significant offsets ( $[3.6]_{L16} - [3.6]_{N17} < 0.03$  mag) as previously verified by Steinhardt et al. (2014). The [3.6] number counts of COSMOS2015 are in excellent agreement with CANDELS for magnitudes  $\lesssim 24.5$ ; after restricting the comparison inside the  $\mathcal{A}_{UD}$  region, counts agree with  $< 20\%$  difference until reaching  $[3.6] = 25$ , where the number of UltraDeep sources starts to decline compared to CANDELS. Despite such a small fraction of missing sources, our  $z \gtrsim 3$  statistical analysis would nonetheless suffer from severe incompleteness if most of them turned out to be at high redshift. For this reason, we inspect the  $z_{phot}$  distribution of 11 761 galaxies (out of 38 671) in N17 not matching any COSMOS2015 entry. They are extremely faint objects with  $F160W \gtrsim 26$ , most of them without a counterpart even in our IRAC residual maps (see below).

The CANDELS photometric redshifts ( $z_{phot,N17}$ ) have been computed independently by several authors, by means of different codes (see Dahlen et al. 2013). Here we use the median of

<sup>2</sup> The wavelength range of the four channels is centred respectively at 3.6, 4.5, 5.8, and 8.0  $\mu\text{m}$ ; in the following we refer to them as [3.6], [4.5], [5.8], and [8.0].



**Fig. 2.** *Upper panel:* redshift distribution of the whole CANDELS sample in the COSMOS field, taken from Nayyeri et al. (2017, N17, gray filled histogram). We also identify the N17 objects that do not have a counterpart (within 0.8") in the COSMOS2015 catalogue, showing their  $N(z)$  with a red histogram. *Middle panel:* ratio between the CANDELS objects with a match in COSMOS2015 ( $N_{\text{matched}}$ ) and all the CANDELS entries ( $N_{\text{tot}}$ ) in bins of [3.6] mag (filled circles). These estimates are divided into three redshift bins in the range  $2.5 < z_{\text{phot},\text{N17}} < 6$  (see colour-code inset); a dashed line marks the 70% completeness. *Lower panel:* similar to the middle panel, but the  $N_{\text{matched}}/N_{\text{tot}}$  ratio is estimated to reproduce the sensitivity depth of UltraVISTA-Deep.

those estimates, which is generally in good agreement with our  $z_{\text{phot}}$  for the objects in common. The upper panel of Fig. 2 shows that galaxies excluded from the CANDELS-COSMOS matching have a redshift distribution  $N(z)$  similar to the whole N17 sample. Restricting the analysis to  $2.5 < z_{\text{phot},\text{N17}} \leq 3.5$  galaxies, we clearly see that the fraction of N17 galaxies not detected in COSMOS2015 increases towards fainter magnitudes. The same trend, despite a larger shot noise due to the small-number statistics, is visible at higher redshifts. Taking CANDELS as a reference ‘parent sample’, the fraction of sources we recover is a proxy of the global completeness of COSMOS2015. As shown in Fig. 2 (middle panel) we can assume  $[3.6]_{\text{lim,UD}} = 25$  as a reliable  $>70\%$  completeness limit for our  $\mathcal{A}_{\text{UD}}$  sample up to  $z = 6$ .

We can also evaluate such a limit in  $\mathcal{A}_{\text{D}}$  ( $[3.6]_{\text{lim,D}}$ ) although that region does not overlap CANDELS. We repeat the procedure described above after applying a cut in  $z^+, Y, J, H$ , and  $K_s$  bands of N17, corresponding to the  $3\sigma$  limiting magnitudes in the  $\mathcal{A}_{\text{D}}$  area. The resulting threshold is almost 1 mag brighter than  $[3.6]_{\text{lim,UD}}$ , with a large scatter at  $z_{\text{phot},\text{N17}} > 3.5$  (Fig. 2, bottom panel). However, we warn that such a ‘mimicked’ selection is just an approximation, less efficient than the actual  $\mathcal{A}_{\text{D}}$  extraction. In fact, we restrict the N17 sample to be ‘Deep-like’ by simply considering as detected those objects whose flux is above the sensitivity limit in at least one of these five bands. Such an

approach differs from the actual  $\mathcal{A}_{\text{D}}$  also because doing this approximation we did not take into account the correction related to PSF homogenisation. With this caveat in mind, we suggest assuming  $[3.6]_{\text{lim,D}} = 24$  up to  $z \sim 4$ . However, the analysis in the present paper is restricted to the  $\mathcal{A}_{\text{UD}}$  sample and an accurate evaluation of  $[3.6]_{\text{lim,D}}$  is beyond its scope.

In addition, we run SExtractor (Bertin & Arnouts 1996) on the [3.6] and [4.5] residual maps to check whether the recovered sources coincide with those in CANDELS. Most of the latter ones are not found in the residual maps, because they are fainter than the SPLASH background noise ( $\gtrsim 25.5$  mag), with a low signal-to-noise ratio ( $S/N < 2$ ) that prevents us from effectively identifying them with SExtractor. For the 20% CANDELS unmatched objects that are brighter than  $[3.6]_{\text{lim,UD}}$ , their absence in the residual maps can be explained by blending effects; if a MIR source does not correspond to any COSMOS2015 detection, IRACLEAN may associate its flux to a nearby extended object. This highlights the capability of HST/WFC3 to correct for IRAC source confusion better than our ground-based images; although the deeper sensitivity we shall reach with the oncoming VISTA and HSC observations should dramatically reduce the gap. As a consequence of this blending issue, the IRAC flux of some of our bright galaxies (and stars) is expected to be overestimated, but less than 40% since the secondary blended source is generally  $>1$  mag fainter. Some of the CANDELS unmatched objects may also be corrupted detections, since for this test we did not apply any pre-selection using SExtractor quality flags. Eventually, we visually inspect 22 sources at  $24 < [3.6] < 25$  recovered from the IRAC residual map but not found in the N17 sample. These objects are not resolved in the  $F160W$  image, nor in UltraVISTA. They appear also in the IRAC [4.5] residual map suggesting that they should not be artefacts, but rather a peculiar type of  $3 < z < 5$  galaxy with a prominent  $D4000$  break (or less probably,  $z \sim 12$  galaxies) that we shall investigate in a future study.

### 3. Photometric redshift and galaxy classification

We estimate the  $z_{\text{phot}}$  of COSMOS2015 sources by fitting synthetic SEDs to their multi-wavelength photometry. The COSMOS2015 catalogue already provides photometric redshifts and other physical quantities (e.g., galaxy stellar masses) derived through a SED fitting procedure detailed in L16. Here we follow the same approach, using the code LePhare (Arnouts et al. 2002; Ilbert et al. 2006) but with a configuration optimised for high- $z$  galaxies (Sect. 3.1).

The main reasons for a new SED fitting computation at  $z > 2.5$  are the following:

- L16 explored the parameter space between  $z = 0$  and 6, but to build an accurate  $\text{PDF}(z)$  for galaxies close to that upper limit one has to enlarge the redshift range. Therefore, we scan now a grid  $z = [0, 8]$  to select galaxies between  $z_{\text{phot}} = 2.5$  and 6.
- We have improved the method for removing stellar interlopers, which is now based on a combination of different star versus galaxy classifications, with particular attention to low-mass stars (see Sect. 3.2).
- With respect to L16, we included in the library additional high- $z$  templates, that is, SEDs of extremely active galaxies with rising star formation history (SFH) and highly attenuated galaxies with  $E(B - V) > 0.5$ .

Our results replace the original photometric redshifts of L16 only at  $z > 2.5$  (Sect. 3.3). Galaxies with a new  $z_{\text{phot}} < 2.5$  are not considered, so below  $z = 2.5$  the sample is the same as in L16. In

any case, the variation at low  $z$  is negligible, given the high percentage of galaxies that preserve their original redshift. A comparison between the original L16 SED fitting and our new results can be found in Appendix A.

### 3.1. Photometric redshift of $z > 2.5$ galaxies

We fit the multi-band photometry of the entire catalogue and then select galaxies with  $z_{\text{phot}} > 2.5$ . We apply zero-point offsets in all the bands as prescribed in L16. Also when  $S/N < 1$ , we consider the flux measured in that filter (and its uncertainty) without replacing it with an upper limit. This choice allows us to take into account non-detections without modifying the way in which the likelihood function is computed (whereas a different implementation is required to use upper limits, see Sawicki 2012).

Our SED fitting library includes early- and late-type galaxy templates from Polletta et al. (2007), together with 14 SEDs of star-forming galaxies from GALAXEV (Bruzual & Charlot 2003, see also Sect. 4.1). With this code we also produce templates of passive galaxies at 22 different ages (from 0.5 to 13 Gyr). These are the same templates used in L16. In addition, as mentioned above, we use two GALAXEV templates that represent starburst galaxies with an increasing SFH (Behroozi et al. 2013; da Cunha et al. 2015; Sparre et al. 2015). The age of both templates is 100 Myr. Instead of using an exponentially increasing SFH (Maraston et al. 2010) we opt for a multi-component parametrisation (Stark et al. 2014): a constant SFH is superimposed to a delayed  $\tau$  model with  $SFR \propto \tau^{-2} t e^{-t/\tau}$  (see Simha et al. 2014) where the  $e$ -folding time  $\tau$  is equal to 0.5 Gyr and  $t = 10\text{--}40$  Myr (Papovich et al. 2001, 2011; Smit et al. 2014).

We add to each synthetic SED the principal nebular emission lines: Ly $\alpha$   $\lambda 1216$ , [OII]  $\lambda 3727$ , H $\beta$   $\lambda 4861$ , [OIII] $\lambda\lambda 4959, 5007$ , H $\alpha$   $\lambda 6563$ . We calibrate the lines starting from the UV-[OII] relation of Kennicutt (1998), but we let the [OII] equivalent width vary by  $\pm 50\%$  with respect to what the equation prescribes. The approach is fully empirical, with line strength ratios based on Anders & Alvensleben (2003) and Moustakas et al. (2006). The addition of nebular emission lines has been discussed in several studies (see Sect. 5.2). In general, it is considered as an improvement; for example Ilbert et al. (2009), by including templates with emission lines, increase the  $z_{\text{phot}}$  accuracy by a factor  $\sim 2.5$ . Such a gain is due to the fact that strong optical lines (such as [OII] or H $\beta$ -[OIII]) can boost the measured flux and alter galaxy colours (e.g. Labb   et al. 2013).

We assume for nebular emission the same dust attenuation as for stars (Reddy et al. 2010; Kashino et al. 2013) although the issue is still debated (F  rster Schreiber et al. 2009; Wuyts et al. 2013). Moreover, we do not implement any specific prior to control the level of emission line fluxes, although recent studies indicate that their equivalent width (EW) and strength ratio evolve with redshift (e.g. Khostovan et al. 2016; Faisst et al. 2016a). Nevertheless, a stronger bias in the computation is produced by neglecting these lines, rather than roughly modelling them (Gonz  lez et al. 2011; Stark et al. 2013; Wilkins et al. 2013).

Attenuation by dust is implemented in the SED fitting choosing among the following extinction laws: Pr  v  ot et al. (1984, SMC-like), Calzetti et al. (2000), and two modified versions of Calzetti's law that include the characteristic absorbing feature at 2175 Å (the so-called "graphite bump", Fitzpatrick & Massa 1986) with different strength. The optical depth is free to vary from  $E(B - V) = 0$  to 0.8, to take into account massive and heavily obscured galaxies at  $z > 3$  (up to  $A_V \approx 3$ , e.g. Spitler et al. 2014).

Our initial sample at  $2.5 < z \leq 6$  includes 92 559 galaxies. The photometric redshift assigned to each of them is the median of the probability distribution function (PDF) obtained after scanning the whole template library. Hereafter, for sake of simplicity, for the reduced chi squared of a fit (often referred to as  $\chi^2_{\text{red}}$ ) we use the short notation  $\chi^2$ . The  $z_{\text{phot}}$  error ( $\sigma_z$ ) corresponds to the redshift interval around the median that delimits 68% of the integrated PDF area. The same definition of  $1\sigma$  error is adopted for stellar mass estimates as well as SFR, age, and rest-frame colours (see Sect. 4). As an exception, we prefer to use the best-fit redshift when the  $\text{PDF}(z)$  is excessively broad or spiky (i.e. there are a few peaks with similar likelihood) and the location of the median is thus highly uncertain; we identify 2442 galaxies in this peculiar condition, such that  $|z_{\text{median}} - z_{\text{best}}| > 0.3(1 + z_{\text{best}})$ .

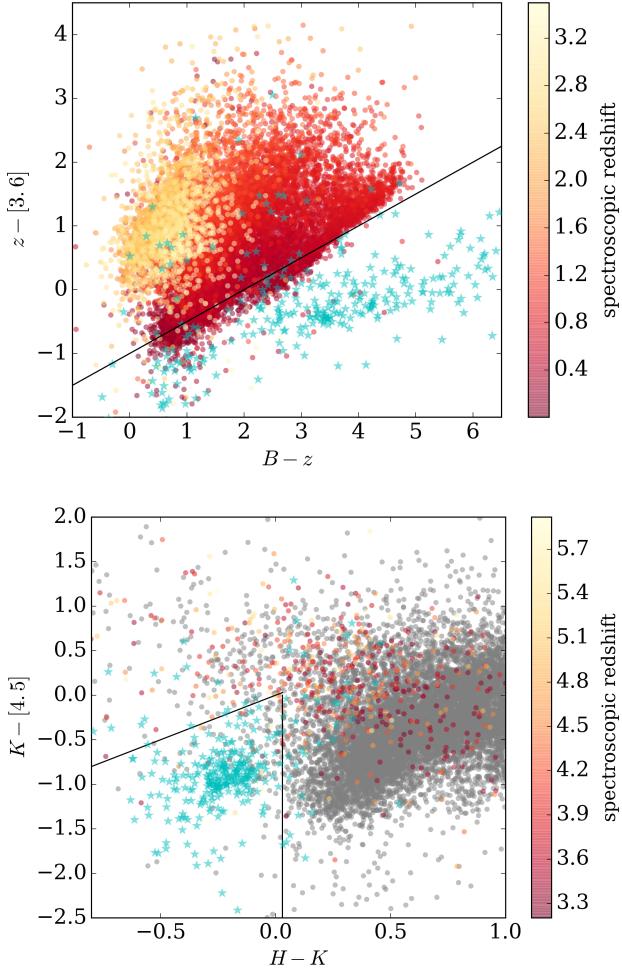
To secure our  $z_{\text{phot}} > 2.5$  sample, we apply additional selection criteria. In the redshift range of  $u$ -to- $V$  drop-outs ( $z_{\text{phot}} \gtrsim 3.2$ ) we require galaxies not to be detected in those optical bands centred at  $< 912(1 + z_{\text{phot}})$  Å. This condition is naturally satisfied by 97.3% of the sample. We also remove 249 sources with  $\chi^2 > 10$ . We prefer not to implement criteria based on visual inspection of the high- $z$  candidates to avoid subjective selections.

### 3.2. Stellar contamination

To remove stars from the  $z_{\text{phot}} > 2.5$  sample we adopt an approach similar to Moutard et al. (2016a), combining multiple selection criteria. First, we fit the multi-wavelength baseline with stellar spectra taken from different models and observations (Bixler et al. 1991; Pickles & J. 1998; Chabrier et al. 2000; Baraffe et al. 2015). We emphasise that the library contains a large number of low-mass stars of spectral classes from M to T, mainly from Baraffe et al. (2015). Unlike dwarf star spectra used in previous work (e.g. Ouchi et al. 2009; Bouwens et al. 2011; Bowler et al. 2014) those derived from Baraffe et al. (2015) extend to  $\lambda_{\text{r.f.}} > 2.5$  μm and therefore SPLASH photometry contributes to disentangling their degeneracy with distant galaxies (Wilkins et al. 2014).

We compare the  $\chi^2$  of stellar and galaxy fits, and flag an object as a star if  $\chi^2_{\text{gal}} - \chi^2_{\text{star}} > 1$ . When the  $\chi^2$  difference is smaller than this confidence threshold we use additional indicators, namely (i) the stellar locus in colour-colour diagrams and (ii) the maximum surface brightness ( $\mu_{\text{max}}$ ) above the local background level. For objects with  $0 < \chi^2_{\text{gal}} - \chi^2_{\text{star}} \leq 1$  we also set  $z_{\text{phot}} = 0$  when the criteria (i) or (ii) indicate that the source is a star.

The diagrams adopted for the diagnostic (i) are  $(z^{++} - [3.6])$  versus  $(B - z^{++})$  and  $(H - K_s)$  versus  $(K - [3.6])$ ; the former is analogous of the  $BzK$  by Daddi et al. (2004), the latter has been used, for example, in Caputi et al. (2015). The two diagrams are devised using the predicted colours of both stars and galaxy models, and tested by means of the  $z_{\text{spec}}$  sample (Fig. 3). The latter is used for galaxies not detected in the  $B$  band (mainly drop-outs at  $z \gtrsim 4$ ) for which the  $(z^{++} - [3.6])$  versus  $(B - z^{++})$  diagnostic breaks down (see L16, Fig. 15). In each colour-colour space, we trace a conservative boundary for the stellar locus, since photometric uncertainties increase the dispersion in the diagram and stars can be scattered out from the original sequence. Method (ii) is detailed in Leauthaud et al. (2007) and Moutard et al. (2016b). The surface brightness measurements in the wide  $I$ -filter of HST ( $F814W$ ) come from the Advanced Camera for Surveys (ACS) images analysed by Leauthaud et al. (2007). Stars are segregated



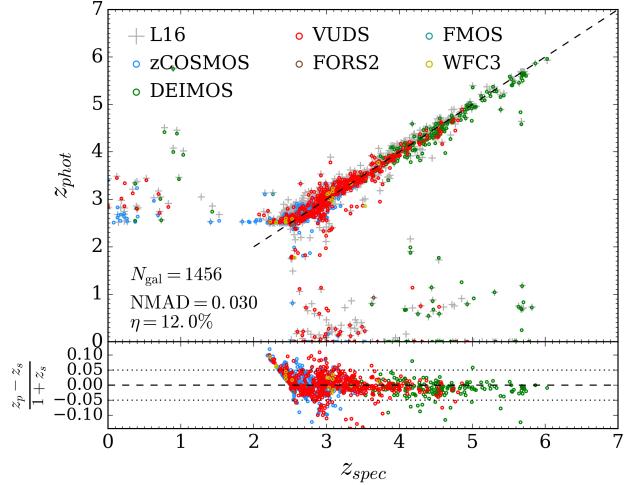
**Fig. 3.** Colour-colour diagrams for removing stellar contaminants. Only spectroscopic measurements with quality flags 3 or 4 ( $CL > 95\%$ ) are shown in the figure. *Upper panel:*  $(B - z^{++})$  versus  $(z^{++} - [3.6])$ . Galaxies detected in the  $B$  band are shown with filled circles coloured from red to yellow according to their  $z_{\text{spec}}$ . *Lower panel:*  $(H - K_s)$  versus  $(K_s - [4.5])$ . Circles with red-to-yellow colours are  $B$  drop-outs, while grey circles are the remaining spectroscopic galaxies at  $z \lesssim 3$ . In both panels, solid lines delimit the conservative boundaries we chose for the stellar locus. These are described by the following equations:  $(z^{++} - [3.6]) < 0.5(B - z^{++}) - 1$  in the upper panel, and  $(K_s - [3.6]) < (H - K_s) \wedge (H - K_s) < 0.03$  in the lower panel. Stars spectroscopically confirmed are plotted with cyan symbols. Typical photometric errors are  $\lesssim 0.05$  mag for object with  $[3.6] < 24$ , and increase up to  $0.08$ – $0.12$  mag for fainter ones.

in the  $\mu_{\max}$ - $I$  plane, which has been shown to be reliable at  $I \lesssim 25$ <sup>3</sup>.

### 3.3. Validation through spectroscopy and self-organising map

We use a catalogue of almost 100 000 spectroscopic redshifts to quantify the uncertainties of our  $z_{\text{phot}}$  estimates. These data were obtained during several campaigns, with different instruments and observing strategies (for a summary, we refer to Table 4 of L16). They have been collected and harmonised in a single

<sup>3</sup> Leauthaud et al. also discuss the limitations of the “stellarity index”, another commonly used classification provided by SExtractor. This indicator is less accurate than the one based on  $\mu_{\max}$ , especially for faint compact galaxies (see Leauthaud et al. 2007, Fig. 4). We then decided not to add stellarity indexes to our set of criteria.



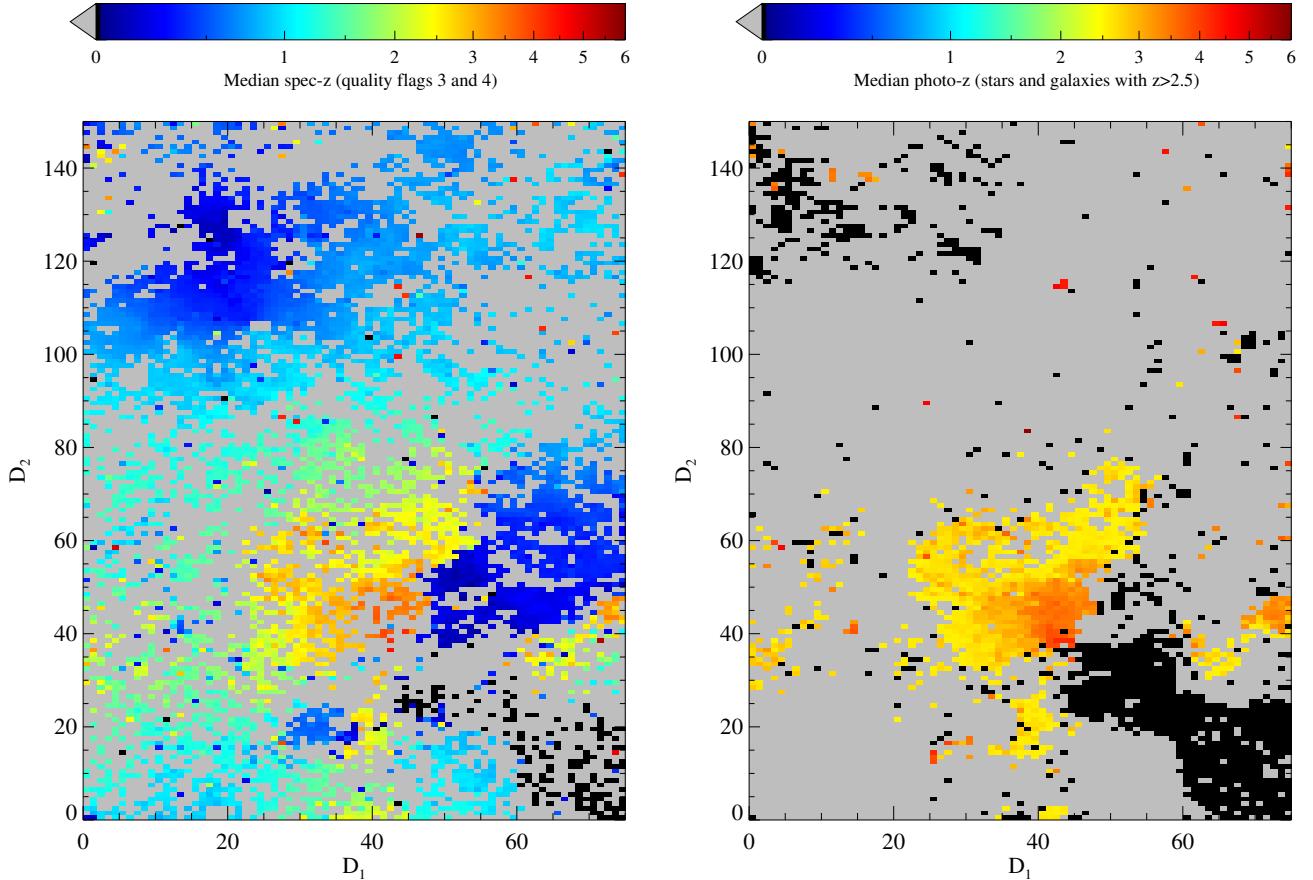
**Fig. 4.** Comparison between  $z_{\text{spec}}$  and  $z_{\text{phot}}$ , for spectroscopic galaxies (and stars) with  $[3.6] < 25$  (empty circles). Only robust spectroscopic measurements ( $CL > 95\%$ ) are plotted, and coloured according to their survey: zCOSMOS faint (Lilly et al., in prep.), VUDS (Le Fèvre et al. 2015), FMOS-COSMOS (Silverman et al. 2015), a survey with the FORS2 spectrograph at VLT (Comparat et al. 2015), a survey with DEIMOS at Keck II (Capak et al., in prep.), and grism spectroscopy from HST/WFC3 (Krogager et al. 2014). In the background we also show the comparison between  $z_{\text{spec}}$  and the original photometric redshifts of L16 (grey crosses). *Upper panel:* in addition to  $z_{\text{phot}}$  versus  $z_{\text{spec}}$ , in the bottom-left corner we report the number of objects considered in this test ( $N_{\text{gal}}$ ), the  $\sigma_z$  error defined as the NMAD, and the fraction of catastrophic outliers ( $\eta$ ). The dashed line is the  $z_{\text{phot}} = z_{\text{spec}}$  reference. *Lower panel:* scatter of the  $z_{\text{phot}} - z_{\text{spec}}$  values, with the same colour-code as in the upper panel. Horizontal lines mark differences (weighed by  $1 + z_{\text{spec}}$ ) equal to  $\pm 0.05$  (dotted lines) or null (dashed line).

catalogue by Salvato et al. (in prep.). Such a wealth of spectroscopic information represents an unequalled benefit of the COSMOS field.

The  $z_{\text{spec}}$  measurements used as a reference are those with the highest reliability, that is, a confidence level ( $CL > 95\%$ ) (equal to a selection of quality flags 3 and 4 according to the scheme introduced by Le Fèvre et al. 2005). We limit the comparison to sources brighter than  $[3.6] = 25$ , ignoring secure low- $z$  galaxies (those having both  $z_{\text{spec}}$  and  $z_{\text{phot}}$  below 2.5). Eventually, our test sample contains 1456 objects. The size of this sample is unique: with 350 galaxies at  $z_{\text{spec}} > 3.5$ , it is more than twice the number of robust spectroscopic redshifts available in CANDELS, GOODS-South and UDS, used in Grazian et al. (2015).

Among the 301 spectroscopic stars considered,  $> 90\%$  of them are correctly recovered by our method, with only three stellar interlopers with  $z_{\text{phot}} > 2.5$ . On the other hand, less than 1% of the spectroscopic galaxies are misclassified as stars. The catastrophic error rate is  $\eta = 12\%$ , considering any object with  $|z_{\text{phot}} - z_{\text{spec}}| > 0.15(1 + z_{\text{spec}})$  as an outlier. The precision of our photometric redshifts is  $\sigma_z = 0.03(1 + z)$ , according to the normalised median absolute deviation (NMAD, Hoaglin et al. 1983) defined as  $1.48 \times \text{median}\{|z_{\text{phot}} - z_{\text{spec}}|/(1 + z_{\text{spec}})\}$ . These results (Fig. 4) summarise the improvement with respect to the photometric redshifts of L16; we reduce the number of catastrophic errors by  $\sim 20\%$ , and we also observe a smaller bias at  $2.5 < z < 3.5$  (cf. Fig. 11 of L16).

The comparison between  $z_{\text{spec}}$  and  $z_{\text{phot}}$  is meaningful only if the spectroscopic sample is an unbiased representation of the “parent” photometric sample. Otherwise, we would test the reliability of a subcategory of galaxies only. We introduce a



**Fig. 5.** Bi-dimensional self-organising map of the COSMOS2015 catalogue (the two folded dimensions have generic labels  $D_1$  and  $D_2$ ). In the *left panel*, only robust spectroscopic objects ( $CL > 95\%$ ) are shown. In the *right panel*, the SOM is filled with photometric objects (stars and  $z_{\text{phot}} > 2.5$  galaxies). In both cases, each cell of the map is colour-coded according to the median redshift of the objects inside the cell (empty cells are grey, cells filled by stars are black).

self-organising map (SOM, Kohonen 1982) to show that our spectroscopic catalogue provides a representative sample of the underlying colour and redshift distribution. The algorithm version we use, specifically implemented for astronomical purposes, is the one devised in Masters et al. (2015). The SOM allows us to reduce a high-dimensional dataset in a bi-dimensional grid, without losing essential topological information. In our case, the starting manifold is the panchromatic space (fifteen colours) resulting from the COSMOS2015 broad bands: (NUV –  $u$ ), ( $u - B$ ), ..., ([4.5] – [5.8]), ([5.8] – [8.0]). As a side note, we highlight that the SOM dimensions do not necessarily have to be colours; in principle the parameter space can be enlarged by including other properties like galaxy size or morphological parameters. Each coloured cell in the map (Fig. 5) corresponds to a point in the 15-dimensional space that is non-negligibly occupied by galaxies or stars from the survey. Since the topology is preserved, objects with very similar SEDs – close to one another in the high-dimensional space – will be linked to the same cell (or to adjacent cells).

We inspect the distribution of spectroscopic objects in the SOM, using them as a training sample to identify the region of high- $z$  galaxies (Fig. 5, left panel). The COSMOS spectroscopic catalogue samples well the portion of parameter space we are interested in, except for the top-left corner of the map where we expect, according to models, the bulk of low-mass stars. The lack of spectroscopic measurements in that region may affect the precise evaluation of the  $z_{\text{phot}}$  contaminant fraction. Other cells that are weakly constrained correspond to the SED of star-forming

galaxies with  $i \gtrsim 23$  mag and  $z < 2$  (Masters et al. 2015), which however are not pivotal for testing our estimates.

After the spectroscopic calibration, we insert  $z_{\text{phot}} > 2.5$  galaxies in the SOM along with photometric stars (Fig. 5, right panel). Given the larger size of the photometric catalogue, cell occupation is more continuous and extended (see e.g. the stellar region in the bottom-right corner). By comparing the two panels of Fig. 5, one can see that the  $z_{\text{phot}} > 2.5$  galaxies are concentrated in the SOM region that has been identified as high- $z$  by the spectroscopic training sample. Although the latter is more sparse, about 60% of the area covered by  $z_{\text{phot}} > 2.5$  galaxies is also sampled by spectroscopic measurements, which are quantitatively in good agreement; in 82% of those cells that contain both photometric and spectroscopic redshifts, the median of the former is within  $1\sigma_z$  from the median of the  $z_{\text{spec}}$  objects laying in the same cell. Moreover, by plotting individual galaxies (not shown in the figure) one can verify that catastrophic  $z_{\text{phot}}$  errors are randomly spread across the SOM, not biasing any specific galaxy class.

#### 4. Stellar mass estimate and completeness

After building a  $z_{\text{phot}} > 2.5$  galaxy sample, we run LePhare to estimate their stellar mass, SFR, and other physical parameters such as rest frame colours. This is described in Sect. 4.1, while in Sect. 4.2 we compute stellar mass completeness limits and argue in favour of a [3.6] selection to work with a mass-complete sample up to  $z \sim 6$  (see also Appendix B). By means of their

rest-frame colours we then identify reliable quiescent galaxies (Sect. 4.3).

#### 4.1. Galaxy stellar mass

We estimate stellar mass and other physical properties of the COSMOS2015 galaxies always at fixed  $z \equiv z_{\text{phot}}$ . We fit their multi-wavelength photometry with a library of SEDs built using the stellar population synthesis model of Bruzual & Charlot (2003, hereafter BC03). The  $\mathcal{M}$  estimates are the median of the PDF marginalised over the other parameters. This kind of estimate is in good agreement with the stellar mass derived from the PDF peak (i.e. the best-fit template). The difference between median and best-fit values is on average 0.02 dex, with a rms of 0.11 dex.

The galaxy templates given in input to LePhare are constructed by combining BC03 simple stellar populations (SSPs) according to a given SFH. Each SSP has an initial mass function (IMF) that follows Chabrier (2003), while the stellar metallicity can be  $Z = 0.02, 0.008$ , or  $0.004^4$ . These stellar metallicities have been chosen to encompass the range observed up to  $z \sim 4\text{--}5$  (e.g. Maiolino et al. 2008; Sommariva et al. 2012); they are also in agreement with hydrodynamical simulations (Ma et al. 2016). For each template we combine SSPs with the same metallicity (i.e. there is no chemical enrichment in the galaxy model, nor interpolation between the three given  $Z$  values).

We assumed various SFHs, namely “exponentially declining” and “delayed declining”. The former ones have  $SFR(t) \propto e^{-t/\tau}$ , while the shape of the latter is  $\propto t e^{-t/\tau}$ . For the exponentially declining profiles, the  $e$ -folding time ranges from  $\tau = 0.1$  to 30 Gyr, while for delayed SFHs the  $\tau$  parameter, which also marks the peak of SFR, is equal to 1 or 3 Gyr. We post-process the BC03 templates obtained in this way by adding nebular emission lines as described in Sect. 3.1. Dust extinction is implemented assuming  $0 \leq E(B-V) \leq 0.8$ . We allow for only one attenuation law, that is, Calzetti et al. (2000) with the addition of the 2175 Å feature (see Scoville et al. 2015).

We have tried a few alternate configurations to quantify the amplitude of possible systematics (see Sect. 5.2 for a detailed discussion). We added, for example, a second attenuation curve with slope proportional to  $\lambda^{-0.9}$  (Arnouts et al. 2013). Such a choice increases the number of degenerate best-fit solutions without introducing any significant bias; Calzetti’s law is still preferred (in terms of  $\chi^2$ ) by most of the objects at  $z \gtrsim 3$ . Other modifications, like the expansion of the metallicity grid, have a larger impact, as also found in other studies (e.g. Mitchell et al. 2013). Simplifying assumptions are somehow unavoidable in the SED fitting, not only for computational reasons but also because the available information (i.e. the multi-wavelength baseline) cannot constrain the parameter space beyond a certain number of degrees of freedom. This translates to systematic offsets when comparing different SED fitting recipes. The impact of these systematics on the SMF is clearly visible in Conselice et al. (2016, Fig. 1), where the authors overplot a wide collection of measurements from the literature: already at  $z < 1$ , where data are more precise, the various SMF estimates can differ even by a factor  $\sim 3$ . We emphasise that one advantage of our work, whose goal is to connect the SMFs at different epochs, is to be less affected by SED fitting uncertainties than analyses that combine measurements from different papers. In our case, SED fitting systematics

(unless they have a strong redshift or galaxy-type dependence) will cancel out in the differential quantities we want to derive.

As mentioned above, the 68% of the integrated  $\text{PDF}(\mathcal{M})$  area gives an error to each stellar mass estimate. However, the PDF is obtained from  $\chi^2$ -fit templates at fixed redshift. To compute stellar mass errors ( $\sigma_m$ ) including the additional uncertainty inherited from  $\sigma_z$ , we proceed in a way similar to Ilbert et al. (2013). They generate a mock galaxy catalogue by perturbing the original photometry and redshifts proportionally to their errors. After recomputing the stellar mass of each galaxy, the authors define its uncertainty as the difference between new  $\mathcal{M}$  and the original estimate, namely  $\Delta\mathcal{M}_i \equiv \log(\mathcal{M}_{\text{MC},i}/\mathcal{M}_i)$  (for the  $i$ th galaxy).

We implement a few modifications with respect to Ilbert et al. (2013). Instead of adding noise to photometry and  $z_{\text{phot}}$ , we exploit the PDFs. Moreover, we produce 100 mock catalogues, instead of a single one. We perform a Monte Carlo simulation, re-extracting the  $z_{\text{phot}}$  of each galaxy 100 times, according to its  $\text{PDF}(z)$ . Each time, we run again LePhare with the redshift fixed at the new value ( $z_{\text{MC}}$ ) to compute the galaxy stellar mass  $\mathcal{M}_{\text{MC}}$  and the offset from the original value. We group galaxies in bins of redshift and mass to obtain an estimate of  $\sigma_m$  from the distribution of their  $\Delta\mathcal{M}$ . As in Ilbert et al. (2013), this is well fit by a Gaussian multiplied by a Lorentzian distribution:

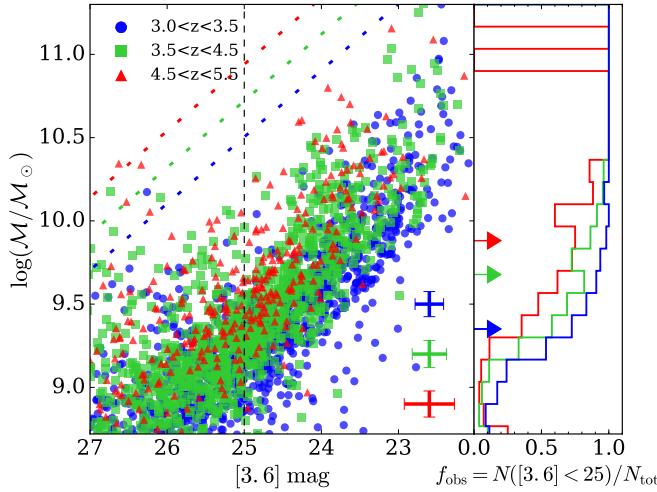
$$\mathcal{D}(\mathcal{M}_0, z) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-\mathcal{M}_0^2}{2\sigma^2}\right) \times \frac{\tau}{2\pi\left[\left(\frac{\tau}{2}\right)^2 + \mathcal{M}_0^2\right]}, \quad (1)$$

where  $\mathcal{M}_0$  is the centre of the considered stellar mass bin. The parameters  $\sigma$  and  $\tau$  are in principle functions of  $\mathcal{M}_0$  and  $z$ , left implicit in Eq. (1) for sake of clarity. We find  $\sigma \approx 0.35$  dex, without a strong dependence on  $\mathcal{M}$ , nor on  $z$ , at least for  $9.5 < \log(\mathcal{M}/\mathcal{M}_\odot) < 11.5$ . Also,  $\tau$  does not depend significantly on stellar mass but increases as a function of redshift. The relation assumed in Ilbert et al. (2013), namely  $\tau(z) = 0.04(1+z)$ , is still valid for our sample. We note that the value of  $\sigma$  is instead smaller (it was 0.5 dex in Ilbert et al. 2013), reflecting the increased quality of the new data. At face value, one can assume  $\sigma_m(\mathcal{M}) = 0.35$  dex (neglecting the Lorentzian “wings” of  $\mathcal{D}$ ); however a careful treatment of stellar mass uncertainties requires taking into account the whole Eq. (1). This computation gives us an idea of the impact of  $\sigma_z$  on the stellar mass estimate: The errors resulting from the  $\text{PDF}(\mathcal{M})$ , after fixing the redshift, are usually much smaller (e.g.  $< 0.3$  dex for 90% of the galaxies at  $3.5 < z \leq 4.5$ ). Further details about the impact of  $\sigma_m$  on the SMF are provided in Sect. 5.5.

#### 4.2. Stellar-mass-limited sample

To estimate the stellar-mass completeness ( $\mathcal{M}_{\text{lim}}$ ) as a function of redshift, we apply the technique introduced by Pozzetti et al. (2010). In each  $z$ -bin, a set of “boundary masses” ( $\mathcal{M}_{\text{resc}}$ ) is obtained by taking the most-used best-fit templates (those with the minimum  $\chi^2$  for 90% of the galaxies in the  $z$ -bin) and rescaling them to the magnitude limit:  $\log \mathcal{M}_{\text{resc}} = \log \mathcal{M} + 0.4([3.6] - [3.6]_{\text{lim,UD}})$ . Then,  $\mathcal{M}_{\text{lim}}$  is defined as the 90th percentile of the  $\mathcal{M}_{\text{resc}}$  distribution. Also taking into account the incompleteness due to the [3.6] detection strategy (see Fig. 2), we expect that in the lowest stellar mass bins of our SMF ( $\mathcal{M} \geq \mathcal{M}_{\text{lim}}$ ) there could be a sample incompleteness of 30% at most, caused by objects previously excluded from the IRAC-selection (i.e. with  $[3.6] > [3.6]_{\text{lim,UD}}$ ). Eventually we interpolate the  $\mathcal{M}_{\text{lim}}$  values found in different  $z$ -bins to describe the evolution of the limiting mass:  $\mathcal{M}_{\text{lim}}(z) = 6.3 \times 10^7 (1+z)^{2.7} \mathcal{M}_\odot$ .

<sup>4</sup> We avoid to use  $Z_\odot$  units since recent work suggests that solar metallicity is lower than the “canonical” value of 0.02 (e.g.,  $Z_\odot = 0.0134$  in Asplund et al. 2009).



**Fig. 6.** Stellar mass completeness as a function of redshift. Blue circles, green squares, and red triangles represent CANDELS galaxies at  $2.5 < z_{\text{phot},\text{N17}} \leq 3.5$ ,  $3.5 < z_{\text{phot},\text{N17}} \leq 4.5$ , and  $4.5 < z_{\text{phot},\text{N17}} \leq 6$ , respectively. The cut in apparent magnitude of our sample ( $[3.6]_{\text{lim}}$ ) is marked with a vertical dashed line. Slant dotted lines show a conservative estimate of the stellar mass limit, corresponding to the  $M/L$  ratio of an old SSP with  $A_V = 2$  mag. Three crosses in the bottom-right corner of the main panel show the average  $x$ - and  $y$ -axis uncertainties in the corresponding bin of redshift. The histograms on the right (same colours and  $z$ -bins of the scattered points) show the ratio of N17 galaxies with  $[3.6] < [3.6]_{\text{lim}}$  over the total N17 sample. This fraction is named  $f_{\text{obs}}$  since they are the objects that would be observed within the magnitude limit of COSMOS2015. Below each histogram, an arrow indicates the stellar mass threshold  $M_{\text{lim}}$  (see Sect. 4.2).

To verify our calculation, we consider the  $\mathcal{M}$  versus  $[3.6]$  distribution of CANDELS-COSMOS galaxies (N17, see Sect. 2.2) after recomputing their stellar mass with LePhare to be consistent with our catalogue. By using these deeper data, we verify that the  $M_{\text{lim}}$  values we found correspond to a completeness of 70–80% (Fig. 6). By means of the N17 sample, we can account for stellar mass incompleteness below  $M_{\text{lim}}$  (cf. Fontana et al. 2004). The factor we need for such a correction in the low-mass regime is  $f_{\text{obs}}(z, \mathcal{M})$ , that is, the fraction of objects at a given redshift and stellar mass that are brighter than  $[3.6]_{\text{lim,UD}}$ . This is obtained by fitting the histograms shown in Fig. 6 (right-hand panel). In doing so we assume that the CANDELS sub-field, which is  $\sim 10\%$  of  $\mathcal{A}_{\text{UD}}$ , is large enough to represent the parent COSMOS2015 volume and sufficiently deep enough to probe the full  $M/L$  range. The function  $f_{\text{obs}}(z, \mathcal{M})$  can be used to correct the SMF of COSMOS2015 at  $\mathcal{M} < M_{\text{lim}}$ , where the  $1/V_{\max}$  weights start to be biased (see Sect. 5).

We also include in Fig. 6 a comparison between our method and a more conservative one, based on the maximum  $M/L$  physically allowed at a given flux limit and cosmic time (see e.g. Pérez-González et al. 2008). The SED used for this purpose is the one of a galaxy that formed stars in a single initial burst at  $z = 20$ , and passively evolves until the desired redshift. Substantial extinction ( $A_V = 2$  mag) is added to further enlarge  $M/L$ . However, in our redshift range, the statistical relevance of such an extreme galaxy type is small, as one infers from the few CANDELS objects sparse in the upper-left corner of Fig. 6 (we refer also to the discussion in Marchesini et al. 2009, Appendix C). With the maximal  $M/L$  ratio we would overestimate the stellar mass completeness threshold by at least a factor 5.

At redshifts between 2.5 and 4, we could, in principle, evaluate  $M_{\text{lim}}$  also as a function of  $K_s$ , with the same empirical approach used above. The cut to be used in this case is  $K_s < 24.7$  (see Fig. 16 of L16). However, the  $M_{\text{lim}}$  resulting from the  $K_s$ -band selection is 0.2–0.4 dex higher (depending on the redshift) than the threshold derived using the  $[3.6]$  band. Such an offset reflects a real difference in the stellar mass distribution of the  $K_s$ -selected sample with respect to the  $[3.6]$ -selected one. The latter is anchored to a  $\chi^2$ -stacked image that includes bands deeper than  $K_s$ , so that the sample is unbiased down to lower masses (more details are provided in Appendix B).

#### 4.3. Quiescent galaxy classification

We also aim to derive the SMF of passive and active galaxies separately. As shown in the following, the quality of our dataset allows us to extend the classification up to  $z = 4$ . Galaxies that stopped their star formation occupy a specific region in the colour–colour diagrams ( $U - V$ ) versus ( $V - J$ ), ( $NUV - r$ ) versus ( $r - K$ ), or ( $NUV - r$ ) versus ( $r - J$ ), as shown in previous work (Williams et al. 2009; Arnouts et al. 2013; Ilbert et al. 2013, respectively). We adopted the last one, dubbed hereafter  $NUVrJ$ . Compared to ( $U - V$ ) versus ( $V - J$ ), which is often referred to as  $UVJ$ , the use of ( $NUV - r$ ) makes  $NUVrJ$  more sensitive to recent star formation (on  $10^6$ – $10^8$  yr scales Salim et al. 2005; Arnouts et al. 2007). This property results in a better distinction between fully quiescent galaxies ( $sSFR < 10^{-11}$  yr $^{-1}$ ) and those with residual star formation (typically with  $sSFR \simeq 10^{-10}$  yr $^{-1}$ ). With  $UVJ$ , these two kinds of galaxies occupy the same place in the diagram, as we verified using BC03 models. On the contrary, galaxies with negligible “frostings” of star formation are correctly classified as passive in the  $NUVrJ$ .

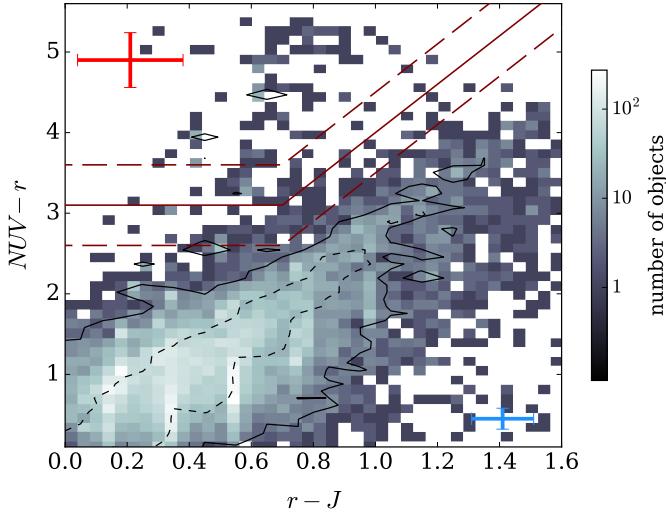
The  $NUVrJ$  indicator is similar to ( $NUV - r$ ) versus ( $r - K$ ), with the advantage that at high redshifts the absolute magnitude  $M_J$  is more robust than  $M_K$ , since for the latter the  $k$ -correction is generally more uncertain. LePhare calculates the absolute magnitude at a given wavelength ( $\lambda_{\text{r.f.}}$ ) by starting from the apparent magnitude in the band closest to  $\lambda_{\text{r.f.}}(1+z)$ . For example at  $z = 3$  the nearest observed filter to compute  $M_J$  is [4.5], whereas  $M_K$  falls beyond the MIR window of the four IRAC channels.

The  $NUVrJ$  diagram is shown in Fig. 7. The density map of our  $z_{\text{phot}} > 2.5$  sample highlights the “blue cloud” of star forming galaxies as well as an early “red sequence”. We also show in the figure how the  $NUVrJ$  distribution changes when colours are derived directly from the template SEDs, without using the nearest observed filter as a proxy of the absolute magnitude. This alternative method is commonly used in the literature, so it is worth showing the different galaxy classification that it yields. The distribution from pure template colours is much narrower (cf. solid and dashed lines in Fig. 7) but potentially biased: the SED library spans a limited range of slopes (colours), whereas the nearest filter method – by taking into account the observed flux – naturally includes a larger variety of SFHs.

The boundary of the passive locus is

$$(NUV - r) > 3(r - J) + 1 \quad \text{and} \quad (NUV - r) > 3.1, \quad (2)$$

defined empirically according to the bimodality of the two-dimensional galaxy distribution (Ilbert et al. 2013). This border is also physically justified; the slant line resulting from Eq. (2) runs perpendicular to the direction of increasing specific SFR ( $sSFR \equiv SFR/\mathcal{M}$ ). On the other hand, dust absorption moves galaxies parallel to the border, effectively breaking the degeneracy between genuine quiescent and dusty star forming galaxies.



**Fig. 7.**  $NUVrJ$  diagram of galaxies between  $z = 2.5$  and  $6$  (with their density distribution colour-coded in grey shades). The red solid line divides active and passive regions (see Eq. (2)), while red dashed lines are shifted by  $\pm 0.5$  mag from that border, to give a rough estimate of the width of the green valley, that is, the separation between active and passive clumps. The red (blue) cross in the top-left (bottom-right) corner shows the typical uncertainties  $\sigma_{NUV-r}$  and  $\sigma_{r-J}$  of passive (active) galaxies at  $3 < z \ll 4$ . We also compare our fiducial estimates (based on the nearest observed filter) with colours directly derived from the template SEDs: a solid contour encloses 90% of the former distribution, while a dashed line represents the 90% envelope for the latter.

The same properties are observed in  $(NUV - r)$  versus  $(r - K)$  and  $(U - V)$  versus  $(V - J)$ , as demonstrated in Arnouts et al. (2013) and Forrest et al. (2016).

Rest-frame colour selections have been successfully used up to  $z \sim 3$  (e.g. Ilbert et al. 2013; Moutard et al. 2016a; Ownsworth et al. 2016). The reliability of this technique at  $z > 3$  was recently called into question by Schreiber et al. (2017). They warn that the IRAC photometry in the COSMOS field could be too shallow to derive robust rest-frame optical colours for galaxies at those redshifts. In order to take into account such uncertainties, our algorithm rejects a filter if its error is larger than 0.3 mag. Most importantly, we evaluate  $(NUV - r)$  and  $(r - J)$  uncertainties, to quantify the accuracy of our  $NUVrJ$  selection. For each galaxy, a given rest-frame colour error ( $\sigma_{\text{colour}}$ ) is derived from the marginalised PDF, by considering 68% of the area around the median (similarly to  $\sigma_z$ ). In such a process, the main contributions to  $\sigma_{\text{color}}$  are the photometric uncertainties and the model  $k$ -correction. At  $3 < z \leq 4$ , the mean errors are  $\sigma_{NUV-r} = 0.13$  and  $\sigma_{r-J} = 0.10$  for the active sample, while  $\sigma_{NUV-r} = 0.34$  and  $\sigma_{r-J} = 0.17$  for objects in the passive locus (see Fig. 7). These values confirm that our classification is reliable up to  $z = 4$ ; even the relatively large  $(NUV - r)$  uncertainty does not affect it, given the scale of the  $y$  axis in Fig. 7. Despite the larger uncertainties, we also apply the  $NUVrJ$  classification at  $z > 4$ , identifying 13 potential passive galaxies (4 of them at  $z_{\text{phot}} > 4.5$ ). In that redshift range, a strong H $\alpha$  line can contaminate the [3.6] band (which is used to estimate  $M_r$ ). Comparing their observed ([3.6] – [4.5]) colour with BC03 models (as done in Faisst et al. 2016a) we find that six passive candidates at  $z > 4$  could have non-negligible H $\alpha$  emission ( $EW = 100$ – $500$  Å) while for the others their ([3.6] – [4.5]) is compatible with  $EW < 50$  Å. Nonetheless these  $z > 4$  galaxies need deeper data to be confirmed as passive.

Galaxies close to the border defined by Eq. (2) may be misclassified, but the fraction of objects in that intermediate corridor (encompassed by dashed lines in Fig. 7) is small and the bulk of the passive sample should not be significantly contaminated. We emphasise that by using  $(NUV - r)$  instead of  $(U - V)$  the larger dynamical scale on the  $y$  axis drastically reduces the contamination at the edge of the passive locus. Inside that intermediate region, galaxies are expected to be in transition from the blue cloud to the red sequence (Moutard et al. 2016a). Therefore, their classification within the active versus passive scheme is not straightforward even from a physical point of view. We come back to discussing green valley galaxies in Sect. 5.4.

## 5. Results

### 5.1. Galaxy stellar mass function at $z > 2.5$

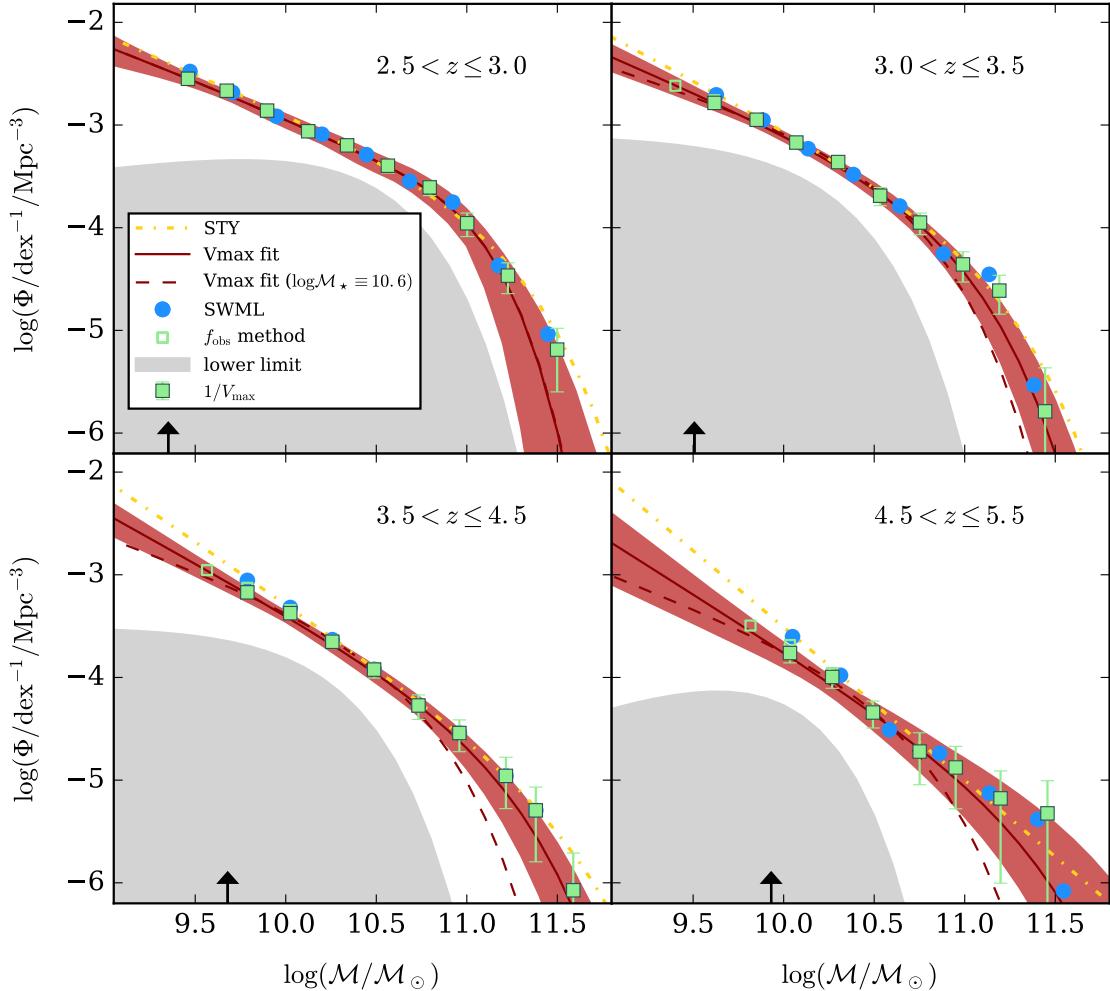
We estimate the galaxy SMF by means of three different methods (as implemented in the code ALF, Ilbert et al. 2005) and compare the results in Fig. 8. The  $1/V_{\max}$  (Schmidt 1968) and the stepwise maximum-likelihood (SWML, Efstathiou et al. 1988) are two techniques that do not impose any a priori model for the SMF, while the maximum likelihood method devised by Sandage et al. (1979, hereafter STY) is parametric and assumes that the SMF is described by a Schechter (1976) function:

$$\Phi(\mathcal{M})d\mathcal{M} = \Phi_* \left( \frac{\mathcal{M}}{\mathcal{M}_*} \right)^\alpha \exp \left( -\frac{\mathcal{M}}{\mathcal{M}_*} \right) \frac{d\mathcal{M}}{\mathcal{M}_*}. \quad (3)$$

A detailed description of the three methods, with their strengths and weaknesses, can be found, for example, in Takeuchi et al. (2000) and Weigel et al. (2016). Our principal estimator is the  $1/V_{\max}$ , widely used in the literature (see Sect. 5.3) due to its simplicity. In particular, the  $1/V_{\max}$  technique assumes uniform spatial distribution of the sources, which is expected to be more robust in our high-redshift bins.

In addition to these methods, at  $z > 3$  we experiment with an empirical approach that corrects for source incompleteness by means of the statistical weight  $f_{\text{obs}}$  (see Sect. 4.2). This weight plays the same role of the  $V_{\max}$  correction, as it accounts for the fraction of missing objects below the [3.6] detection limit. The difference is that  $f_{\text{obs}}$  is recovered empirically from a deeper parent sample (N17), instead of the accessible observable volume as in the case of  $V_{\max}$ . Obviously, the empirical method works under the hypothesis that the parent sample is complete. Not to rely on  $f_{\text{obs}}(z, \mathcal{M})$  in the range where it is too uncertain, we use it only until the correction exceeds a factor two (i.e. 0.2–0.3 dex below  $\mathcal{M}_{\text{lim}}$ ), whereas the previous methods stop at that threshold. For sake of clarity, in Fig. 8 we add error bars only to the  $1/V_{\max}$  estimates, noticing that they are of the same order of magnitude as the SWML uncertainties. A description of how these errors have been calculated is provided in Sect. 5.2. The three classical estimators coincide in the whole stellar mass range. Such an agreement validates the completeness limits we have chosen, because the estimators would diverge at  $\mathcal{M} > \mathcal{M}_{\text{lim}}$  if some galaxy population were missing (see Ilbert et al. 2004). The  $f_{\text{obs}}$  method is also consistent with the others, in the stellar mass range where they overlap, confirming the validity of our empirical approach.

We also fit a Schechter function to the  $1/V_{\max}$  points, accounting for the so-called Eddington bias (Eddington 1913). This systematic bias is caused by stellar mass uncertainties, which make galaxies scatter from one bin to another in the observed SMF. We remove the Eddington bias as done in Ilbert et al. (2013), by convolving Eq. (3) with a description of



**Fig. 8.** SMF of COSMOS2015 galaxies, in four redshift bins between  $z = 2.5$  and  $5.5$ . In each panel, the  $1/V_{\max}$  determination is shown by green squares, while blue circles represent the SWML. Error bars of  $1/V_{\max}$  points include Poisson noise, sample variance, and the scatter due to SED fitting uncertainties (see definition of  $\sigma_\Phi$  in the text). The yellow dot-dashed line represents the STY fitting function, which is a Schechter function at  $z < 4.5$  and a power law in the bottom right panel. Empty squares are obtained from an empirical method where, instead of the  $V_{\max}$  correction, we apply to each galaxy the statistical weight  $f_{\text{obs}}(z, M)$  obtained from a deeper reference sample (see Sect. 4.2). At  $M > M_{\text{lim}}$  the empty squares are not visible since the  $f_{\text{obs}}$  method coincides with  $1/V_{\max}$ . The  $1/V_{\max}$  determinations are fit by a double Schechter function (Eq. (4)) at  $2.5 < z \leq 3$ , and a single Schechter (Eq. (3)) in the other bins (in all the cases, the fit is shown by a red solid line, while the red shaded area is its  $1\sigma$  uncertainty). Another Schechter fit (red dashed line) to the  $1/V_{\max}$  points is made by assuming that the parameter  $\log(M_\star/M_\odot)$  is equal to 10.6. By considering only the most secure  $z_{\text{phot}}$ , we compute a lower limit for the SMF, below which we colour the plot area in grey. An arrow in the bottom part of each panel marks the observational limit in stellar mass (see Sect. 4.2).

the observational uncertainties (Eq. (1)) and using the resulting function to fit data points. At  $2.5 < z \leq 3.0$ , we find that a double Schechter fit, that is,

$$\Phi(M)dM = \left[ \Phi_1^* \left( \frac{M}{M_\star} \right)^{\alpha_1} + \Phi_2^* \left( \frac{M}{M_\star} \right)^{\alpha_2} \right] \exp \left( -\frac{M}{M_\star} \right) \frac{dM}{M_\star}, \quad (4)$$

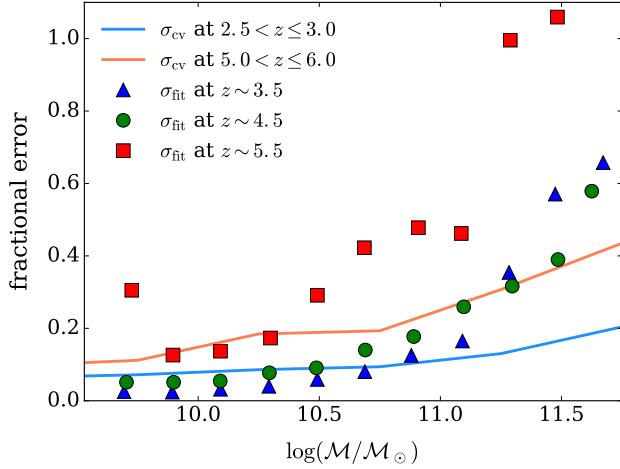
is preferred in the  $\chi^2$  fitting. At  $4.5 < z \leq 5.5$  the STY algorithm does not converge unless assuming unreasonable values for the turnover mass ( $M_\star \gg 10^{11} M_\odot$ ). A simple power law fits well through the points, so for the STY calculation in this  $z$ -bin we replace the Schechter function with

$$\Phi(M) = A \left( \frac{M}{10^{10} M_\odot} \right)^B, \quad (5)$$

where  $A = -3.42 \pm 0.06$  and  $B = -1.57 \pm 0.13$  (given the logarithmic scale, the latter coefficient corresponds to  $\alpha = -2.57$

for a Schechter function). This should be considered as an upper limit of the  $z \simeq 5$  SMF. On the other hand, the fit to the  $1/V_{\max}$  points, taking into account stellar mass errors, recovers a Schechter profile.

We report in Table 1 the Schechter parameters fitting the  $1/V_{\max}$  points at various redshifts. Those are the fiducial values obtained without imposing any constraint (i.e. the fit assumes a flat prior on  $\alpha$ ,  $M_\star$ , and  $\Phi_\star$ ). To deal with the SMF uncertainties, Song et al. (2016) impose in their fitting algorithm a prior for each Schechter parameter, in particular a lognormal PDF( $\log M_\star$ ) centred at 10.75 with  $\sigma = 0.3$  dex. Duncan et al. (2014) perform a fit at  $z \sim 5-7$  with  $M_\star$  fixed to the value they find at  $z \simeq 4$ . In the same vein, we fit again Eq. (3) to the  $1/V_{\max}$  points, but this time with  $\log(M_\star/M_\odot) = 10.6$ , in accordance with the Schechter parameter we find at  $z < 3$  (see Table 2). We adopt this solution consistent with phenomenological models claiming that  $M_\star$  is a redshift independent parameter (e.g. Peng et al. 2010). Observations at  $z < 2$  have confirmed



**Fig. 9.** SMF uncertainties due to cosmic variance (expressed as a fractional error:  $\sigma_{\text{cv}}/\Phi$ ) and SED fitting ( $\sigma_{\text{fit}}/\Phi$ ), as a function of redshift and stellar mass. We show the impact of cosmic variance on the two extreme bins  $2.5 < z \leq 3$  and  $5 < z \leq 6$  (blue and red line, respectively). For  $\sigma_{\text{fit}}$  we take, as an example, three redshift bins:  $2.5 < z \leq 3.5$  (blue triangles),  $3.5 < z \leq 4.5$  (green circles), and  $4.5 < z \leq 5.5$  (red squares).

this turnover mass to be between  $2 \times 10^{10}$  and  $10^{11} M_\odot$  (e.g. Kauffmann et al. 2003; Bundy et al. 2006; Haines et al. 2017). This second fitting function shows how  $\alpha$  and  $M_\star$  are coupled: once the “knee” of the SMF is fixed, the low-mass slope is forced to be shallower ( $\alpha$  increases by  $\sim 0.15$ ). We discuss in detail a few caveats related to the fitting procedure in Sect. 5.5.

In addition, we compute a lower limit for the SMF by considering only the most robust  $z_{\text{phot}}$  (Fig. 8). The selection of such a “pure sample” is done by removing (i) objects with bimodal  $\text{PDF}(z)$ , having a secondary (often low- $z$ ) solution with a non-negligible Bayesian probability; (ii) objects for which the stellar fit, albeit worse than the galaxy best fit, is still a reasonable interpolation of their photometry ( $0 < \chi^2_{\text{star}} - \chi^2_{\text{gal}} < 1$ ).

## 5.2. Sources of uncertainty

In the statistical error budget of the COSMOS2015 mass function, we take into account Poisson noise ( $\sigma_{\text{Poi}}$ ), cosmic variance ( $\sigma_{\text{cv}}$ ), and the scatter due to SED fitting uncertainties ( $\sigma_{\text{fit}}$ , not to be confused with the SED fitting systematics discussed below). The total statistical error is  $\sigma_\Phi = (\sigma_{\text{Poi}}^2 + \sigma_{\text{cv}}^2 + \sigma_{\text{fit}}^2)^{1/2}$ .

Cosmic variance is estimated by means of a modified version of the “cosmic variance calculator” by Moster et al. (2011), for the geometry of our survey and the cosmology we assumed. The main contribution to  $\sigma_{\text{fit}}$  comes from photometric errors and degeneracies between different SEDs. Starting from the Monte Carlo mock samples described in Sect. 4.1, we obtain 100 realisations of our SMF, whose dispersion in a given stellar mass bin is taken as the  $\sigma_{\text{fit}}$  at that mass. A summary of these sources of uncertainty at  $z \geq 3$  is shown in Fig. 9.

In addition to random errors, the SED fitting may introduce systematic offsets in the measured SMF, depending on the adopted recipe. An example is the IMF: assuming Salpeter (1955) the logarithmic stellar masses will be on average 0.24 dex larger than the ones obtained with an IMF as in Chabrier (2003). At  $z > 3$ , the largest biases are generally expected from the  $z_{\text{phot}}$  estimates, since an entire galaxy class may be systematically put in a different  $z$ -bin if the code misinterprets that SED shape

(e.g. because of the degeneracy between Lyman and Balmer breaks).

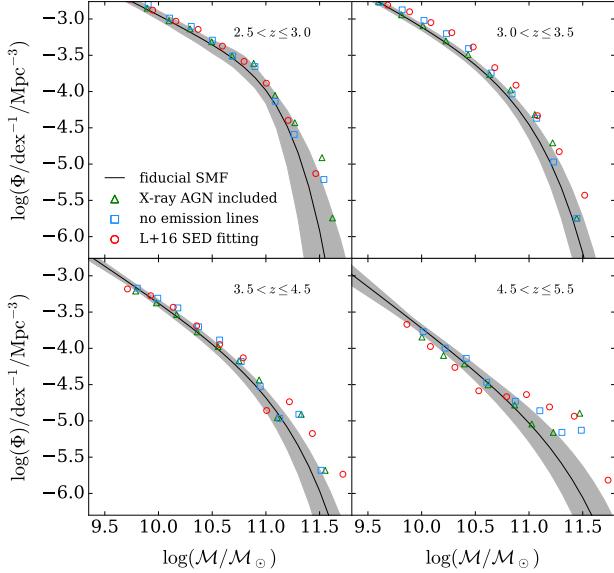
Figure 10 contains three flavours of the COSMOS2015 SMF obtained by modifying the SED fitting recipe. One of these alternate estimates is based on the photometric redshifts and masses provided by L16 ( $z_{\text{phot,L16}}$  and  $M_{\text{L16}}$ ). For the stellar mass computation, the set-up to build BC03 templates includes not only Calzetti et al. (2000, this time without the graphite bump) but also the extinction law  $\propto \lambda^{-0.9}$  described in Arnouts et al. (2013). The metallicity grid of L16 templates is narrower than in the present work, not including  $Z = 0.004$ . The recipe to add nebular emission lines to the synthetic SEDs, although conceptually identical to the one described in Sect. 3.1, assumed slightly different values for the line strength ratios. The most evident feature in this version, contrasting it with our fiducial SMF, is the excess of massive galaxies at  $z > 4$  due to higher stellar contamination (we remove more interlopers with the additional criteria described in Sect. 3.2). We also observe an enhanced number density (but less than a factor of two) at  $3 < z \leq 3.5$ ; the origin of this offset likely resides in the different  $z_{\text{phot}}$  estimates, as discussed in Appendix A.

Another source of systematic error is the addition of nebular emission lines to the BC03 templates. This issue is debated in the literature, with various authors finding from marginal to substantial variations, depending on the code and galaxy sample used. For instance Stark et al. (2013) find that with the addition of emission lines  $M$  decreases by a factor from 1.2 to 2, from  $z \approx 4$  to 6. The offset found by de Barros et al. (2014) in a similar redshift range is on average 0.4 dex, up to 0.9 for the stronger LBG emitters<sup>5</sup>. On the other hand, a recent study on  $H\alpha$  emission in galaxies at  $3 < z_{\text{spec}} < 6$  suggests that previous SED-dependent analyses may overestimate  $H\alpha$  equivalent width (Faisst et al. 2016a, their Fig. 5). Stefanon et al. (2015) compared three different SED fitting methods: using their standard  $z < 3$  calibration, they find no tension with the  $M$  estimates obtained without emission lines. On the contrary, assuming EWs evolving with redshift (Smit et al. 2014), their stellar masses decrease on average by 0.2 dex; although for  $M \lesssim 10^{10} M_\odot$  galaxies at  $4 < z < 5$  they find an opposite trend. In general, SED fitting stellar masses are less sensitive to this bias when a large number of bands are used (Whitaker et al. 2014), especially when the estimates are derived through a Bayesian approach rather than best-fit templates (Salmon et al. 2015). Neglecting lines in our SED fitting procedure,  $M$  estimates increase on average by  $\lesssim 0.05$  dex or less, with noticeable exceptions for some individual galaxies.

We also investigate how the SMF changes if we re-introduce the X-ray sources that have been excluded from the sample in Sect. 2.1. First we verify that their  $z_{\text{phot}}$  are in sufficiently good agreement with estimates derived from a more accurate fitting done with AGN templates (Salvato et al., in prep.). Then, we recompute the stellar mass of each X-ray emitter by means of a three-component SED fitting code (sed3fit, Berta et al. 2013)<sup>6</sup>. This tool relies on the energy balance between dust-absorbed UV stellar continuum and the reprocessed emission in the IR (like MAGPHYS, da Cunha et al. 2008), and also accounts for an additional AGN component from the torus library of Feltre et al. (2012; see also Fritz et al. 2006). X-ray

<sup>5</sup> However de Barros et al. find such a large difference not only by introducing nebular emission lines but also changing other SED fitting parameters like the SFH. When considering only the impact of emission lines, the stellar mass offset is much smaller, especially at  $\log(M/M_\odot) > 10$  (de Barros et al. 2014, top-right panel of Fig. 13).

<sup>6</sup> <http://cosmos.astro.caltech.edu/page/other-tools>



**Fig. 10.** Alternate versions of the COSMOS2015 galaxy stellar mass function: the SMF after the reintroduction of X-ray sources (green triangles), a modified version without emission lines in the synthetic galaxy templates (blue squares), and the SMF based on the original SED fitting of L16 (red circles). The fiducial estimates, already shown in Fig. 8, are reproduced here with solid lines and shaded areas.

luminous ( $L_x > 10^{44} \text{ erg s}^{-1}$ ) AGN are usually hosted in massive ( $\mathcal{M} \gtrsim 10^{11} \mathcal{M}_\odot$ ) galaxies (e.g. Bundy et al. 2008; Brusa et al. 2009; Hickox et al. 2009). They increase the exponential tail of our SMF at least at  $z \leq 3$ , while at higher redshift the number of sources detected by *Chandra* or XMM is too small (Fig. 10). At  $z > 3$ , massive galaxies can also host an active black hole and disregarding its contribution in the SED fitting may cause a stellar mass overestimate of 0.1–0.3 dex (Hainline et al. 2012; Marsan et al. 2017). The impact of different AGN populations on the SMF shall be discussed in a future publication (Delvecchio et al., in prep.). We do not add the various systematic errors together, as done for the statistical ones, because their combined effect is different from the sum of the offsets measured by changing one parameter at a time.

### 5.3. Comparison with previous work

We compare the SMF of COSMOS2015 galaxies with the literature (Fig. 11) recomputing it in the same redshift bins used by other authors (masses are rescaled to Chabrier’s IMF when required). We plot our  $1/V_{\max}$  estimates only, as this is the same estimator used by most of the other authors. Our error bars include  $\sigma_\Phi$  errors defined in Sect. 5.2.

From the literature, we select papers published in the last five years. Some of them, like the present work, collect a galaxy sample where photometric redshifts and stellar mass are derived via SED fitting (“ $\mathcal{M}$ -selected” SMFs: Santini et al. 2012; Ilbert et al. 2013; Muzzin et al. 2013a; Duncan et al. 2014; Caputi et al. 2015; Stefanon et al. 2015; Grazian et al. 2015). Others use rest-frame optical colours to select LBGs; after determining their  $\mathcal{M}$  versus  $L_{\text{UV}}$  distribution, they convolve this with a LF estimate in the UV to derive the SMF (“ $L_{\text{UV}}$ -selected” SMFs: González et al. 2011; Lee et al. 2012; Song et al. 2016; Stefanon et al. 2017). Below we show that our estimates are in excellent agreement with  $\mathcal{M}$ -selected SMFs derived from deep space-based surveys. On the other hand there are some

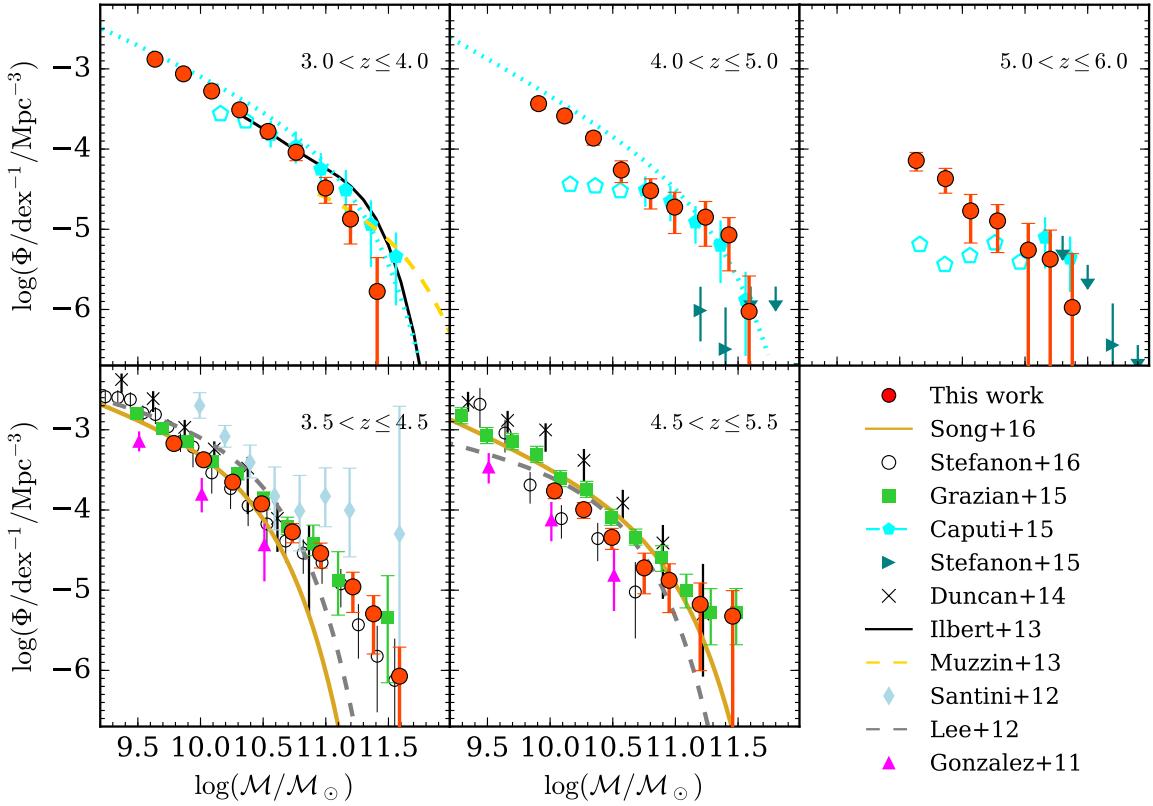
differences with the  $L_{\text{UV}}$ -selected SMFs, as LBGs samples are expected to miss quiescent and dust-obscured galaxies.

Among the  $\mathcal{M}$ -selected SMFs, those with NIR data from ground-based facilities (VISTA and UKIRT) are shown in the top panels of Fig. 11. Our results are in overall agreement with Caputi et al. (2015), which is an updated version of  $K$ -selected galaxy SMFs in UDS and COSMOS (Caputi et al. 2011; Ilbert et al. 2013; Muzzin et al. 2013a). These estimates are limited to  $K_s < 24$ , while Caputi et al. (2015) account for fainter NIR sources ( $24 < K_s \leq 24.7$ ) bright enough in MIR to result in  $\mathcal{M} \gtrsim 10^{11} \mathcal{M}_\odot$ . Including them in their SMF, Caputi et al. (2015) find a number density of massive galaxies comparable to ours, except at  $5 < z \leq 6$  where they have more galaxies at  $11 < \log(\mathcal{M}/\mathcal{M}_\odot) < 11.5$ . This discrepancy could be due to cosmic variance (at those redshifts, Caputi et al. rely mostly on UDS data) but we cannot rule out other explanations; for example, a difference in the  $z_{\text{phot}}$  calculation (more uncertain at such high redshifts, see their Fig. 10). We also note that the SMF of Caputi et al. (2011), one of the original results revised in Caputi et al. (2015), was computed with the 2007 model of Charlot & Bruzual and converted to BC03 by applying a constant 1.24 factor. Moreover, Caputi et al. (2015) do not use SPLASH, but shallower IRAC data (Sanders et al. 2007). Figure 11 also shows the STY fit of Caputi et al. (2015) to demonstrate that the extrapolation of the fit below their stellar mass limit is consistent with our data points.

At  $3 < z \leq 4$ , we also compare with Ilbert et al. (2013) and Muzzin et al. (2013a), two independent estimates both based on UltraVISTA DR1 images. Surprisingly, the results found by the two collaborations, despite the good agreement at lower redshifts, diverge in this bin; massive galaxies ( $> 10^{11} \mathcal{M}_\odot$ ) are relatively abundant in Muzzin et al., whereas the exponential tail in Ilbert et al. decreases more steeply. On the other hand, the latter found more low-mass galaxies and their SMF is defined down to a stellar mass limit that is 0.6 dex smaller than the limit of Muzzin et al. (2013a). In the latter the magnitude cut is  $K_s < 23.4$ , whereas in Ilbert et al. (2013) it is 0.6 mag brighter, but this should account for a 0.24 dex difference only. In fact, the two papers use a different definition of  $\mathcal{M}_{\text{lim}}$  (the one adopted by Ilbert et al. 2013, is the same as in our analysis). With respect to both papers, we reach smaller masses thanks to the deeper  $\mathcal{A}_{\text{UD}}$  images and the panchromatic selection discussed in Appendix B. The differences between Ilbert et al. (2013) and Muzzin et al. (2013a) cannot be explained by cosmic variance, since both SMFs are derived in the same field (using the same raw data). On the contrary, we find that the main reason for the discrepancy between Ilbert et al. (2013) and our SMF at  $\mathcal{M} > 10^{11} \mathcal{M}_\odot$  is our smaller (Ultra-Deep) volume. If we derive the SMF over the whole COSMOS area ( $\mathcal{A}_{\text{D}} + \mathcal{A}_{\text{UD}}$ , see Fig. 1) we find a better agreement. Also the new MIR coverage from SPLASH plays a role (see Appendix B).

To get an insight on the SED fitting uncertainties at  $z > 3$ , we inspect Ilbert et al. (2013) and Muzzin et al. (2013b) catalogues; both publicly available. In the latter sample, among 165 galaxies with  $3 < z_{\text{phot},\text{M13}} < 4$  and  $\log(\mathcal{M}/\mathcal{M}_\odot) > 10.94$ , only 25% remain in the same  $z$ -bin if one replaces Muzzin et al. photometric redshifts with Ilbert et al. estimates<sup>7</sup>. We compare with the COSMOS spectroscopic redshifts introduced in Sect. 3.3, which were not available when both catalogues were built. Among the galaxies that are at  $3 < z_{\text{phot},\text{M13}} \leq 4$  according to Muzzin et al. (2013b), 94 have a match in our COSMOS

<sup>7</sup> Here we consider only objects cross-matched between the two catalogues, with a 0.6'' searching radius.



**Fig. 11.** Comparison to galaxy SMFs from the literature. Our  $1/V_{\max}$  measurements are shown as red circles. If needed, we converted estimates from the literature to Chabrier (2003) IMF. *Upper panels:* we plot the SMF estimates from Caputi et al. (2015), with filled (empty) blue pentagons above (below) their mass completeness limit (in addition, their STY fit is shown by a blue dotted line). For sake of clarity, we plot only the Schechter functions (not the  $1/V_{\max}$  points that have been fit) from Ilbert et al. (2013) and Muzzin et al. (2013a); in both cases the Schechter function (black solid and yellow dashed line, respectively) is truncated at their stellar mass completeness limit. Rightward triangles and downward arrows show  $1/V_{\max}$  and upper limits from a conservative estimate by Stefanon et al. (2015). *Lower panels:*  $M$ -selected SMFs published in Santini et al. (2012, light blue diamonds), Duncan et al. (2014, black crosses), and Grazian et al. (2015, green squares). Other SMFs of LBG samples are taken from González et al. (2011, upward triangles), Lee et al. (2012, grey dashed lines), Song et al. (2016, orange solid line), Stefanon et al. (2017, empty circles).

spectroscopic catalogue. We find that 69 of them are catastrophic errors, defined as  $|z_{\text{phot},\text{M13}} - z_{\text{spec}}| > 0.15(1+z_{\text{spec}})$ . The number of robust estimates in the same  $z$ -bin rises to 87 (out of 107 galaxies spectroscopically observed) when repeating this test using Ilbert et al. (2013) photometric redshift.

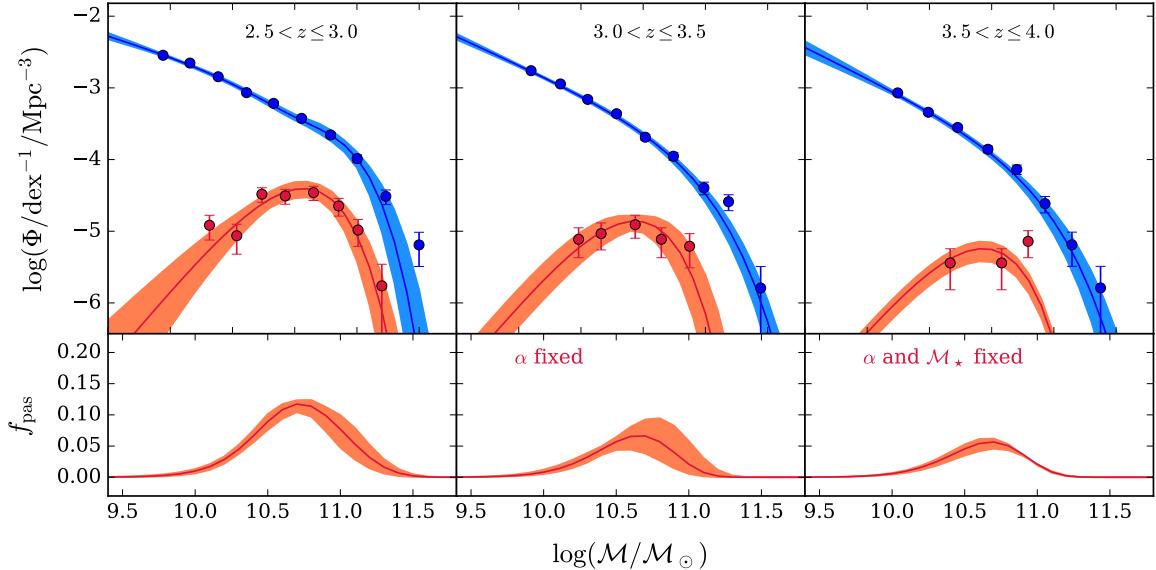
To show the impact of this kind of uncertainty, in Fig. 11 we plot also the  $1/V_{\max}$  estimates (and upper limits) derived by Stefanon et al. (2015) from the subsample of their most robust galaxies, that is, those objects (detected in UltraVISTA DR1) for which the  $z_{\text{phot}}$  satisfies a series of strict reliability criteria. The drop in number density with respect to the other measurements gives an idea of the uncertainties that still affect the high- $z$  SMF.

A comparison with  $M$ -selected SMFs derived from HST/WFC3 detections is shown in the bottom panels of Fig. 11. An estimate using the Early Release Science in the GOODS-South field (complementing WFC3 data with a deep Hawk-I survey) has been provided by Santini et al. (2012). Looking to the error bars of their  $1/V_{\max}$  points, one can appreciate the progress made by the latest studies in terms of statistics. HST data available to Santini et al. (2012) covered about  $33 \text{ arcmin}^2$ , with less than 50 galaxies located between  $z = 3.5$  and 4.5. Duncan et al. (2014) provide another SMF estimate in the GOODS-South field, but using more recent data from CANDELS. Their  $1/V_{\max}$  points are consistent with ours at  $z \approx 4$ , while at  $z \approx 5$  their results are systematically higher by 0.4–0.5 dex (although the

discrepancy is smaller than  $2\sigma$  because of their large sample variance).

We find a remarkably good agreement with the SMF measured by Grazian et al. (2015) over three CANDELS fields (GOODS-South, UDS, HUDF). We underline that the authors have not used the CANDELS-COSMOS field, so their estimate is completely independent from ours. At very high masses ( $> 5 \times 10^{11} M_\odot$ ) the  $1/V_{\max}$  points of Grazian et al. seem to be affected by a similar level of uncertainty of our SMF, confirming the excellent quality of our ground-based data. Given the good agreement between CANDELS and COSMOS2015  $1/V_{\max}$  mass functions, one would expect that the Schechter functions derived from the data points are also similar. However, the two fits differ from each other, as we discuss in Sect. 5.5.

Figure 11 also shows the comparison with the  $L_{\text{UV}}$ -selected SMFs (González et al. 2011; Lee et al. 2012; Song et al. 2016; Stefanon et al. 2017). At  $3.5 < z < 4.5$  the LBG sample of Stefanon et al. (2017) is complemented with UltraVISTA DR2 galaxies that they select according to their  $z_{\text{phot}}$  estimates. Apart from that distinction, these studies should be considered as an estimate of the abundance of UV-bright active galaxies, rather than a census of the entire high- $z$  galaxy population (as clearly stated e.g. in Lee et al. 2012). Moreover, LBGs are usually selected by means of colour criteria that photometric errors can impair more heavily than  $z_{\text{phot}}$  estimates (see Duncan et al. 2014). Le Fèvre et al. (2015) and Thomas et al. (2017) also show that



**Fig. 12.** Active and passive SMFs (blue and red symbols, respectively) in the same redshift bins of Fig. 8. Filled circles are the  $1/V_{\max}$  points, while solid lines represent the Schechter functions fitting to them (shaded areas being the  $1\sigma$  uncertainty of the fit). In the bottom panels we show the fraction of passive galaxies ( $f_{\text{pas}}$ ) as a function of stellar mass in the same  $z$ -bins.

dust and inter-galactic medium extinction move a significant fraction of galaxies outside the standard LBG regions in colour-colour diagrams. Discrepancies with  $M$ -selected SMFs are also due to the different method used to recover stellar masses; galaxies that are outliers in the  $M$  versus  $L_{\text{UV}}$  distribution can be biased when an average  $M(L_{\text{UV}})$  relation is assumed.

Nonetheless, the agreement we find at  $4.5 < z < 5.5$  with Lee et al. (2012) and Song et al. (2016) may suggest that most of the galaxies at those redshifts are going through a star-forming phase. At  $3.5 < z < 4.5$ , their high-mass end is much lower than ours, an indication that the bulk of dusty massive galaxies starts to form already at  $z \sim 4$ . In addition, we include also the SMF of González et al. (2011), which has a normalisation lower by a factor  $\sim 2$ , to show that these analyses also have to deal with severe uncertainties. In general, the SMF of LBGs is derived from their UV luminosity function by assuming a  $L_{\text{UV}}-M$  relation, but such a conversion can hide a number of systematic effects (Song et al. 2016; Harikane et al. 2016).

#### 5.4. Build-up of the quiescent SMF at high redshift

After dividing our sample into active and passive galaxies through the  $NUVrJ$  criterion (Sect. 4.3), we derive the SMF of each galaxy type up to  $z = 4$ . The method of Pozzetti et al. (2010) is applied to each sub-sample in order to compute the corresponding limiting mass  $M_{\text{lim}}$ , which differs for active and passive galaxies as the range of  $M/L$  ratios is not the same (e.g. Davidzon et al. 2013). The SMFs are shown in Fig. 12 (top panels) and their Schechter parameters are reported in Table 1. For the passive mass functions at  $z > 3$  we kept  $\alpha$  fixed to the value found at  $2.5 < z \leq 3$ , and at  $3.5 < z < 4$  we also fix  $M_*$  because otherwise the small number of  $1/V_{\max}$  points cannot constrain the fit effectively. At  $z > 4$  the uncertainties are still too large to perform a robust  $NUVrJ$  classification. However, the small number of passive candidates we found (which could potentially include several interlopers) indicates that the SMF at  $z > 4$  is essentially composed of active galaxies only.

At  $2.5 < z \leq 3$ , similarly to the total sample, the active SMF shows a mild double Schechter profile (Eq. (4)), with the

characteristic bump between  $10^{10}$  and  $10^{11} M_\odot$ . Most of the  $NUVrJ$ -passive galaxies have a stellar mass within that range, while in the massive end of the SMF red galaxies are exclusively dusty objects belonging to the star-forming population (confirming the trend that Martis et al. 2016 find at  $z < 3$ ). Thus, our  $NUVrJ$  technique does not misclassify galaxies reddened by dust as part of the passive sample. Using *Spitzer*/MIPS and *Herschel* data to identify this kind of contaminant, we find only five  $NUVrJ$ -passive objects with far-IR emission. For instance at  $2.5 < z \leq 3$ ,  $\sim 70\%$  of the active galaxies have  $E(B-V) > 0.25$  ( $A_V \gtrsim 1$ ) in the bin  $10.5 < \log(M/M_\odot) \leq 11$ , and almost the entire active sample is heavily attenuated at  $\log(M/M_\odot) > 11$ .

To compare with Spitler et al. (2014, ZFOURGE survey), we can merge the bins  $3 < z \leq 3.5$  and  $3.5 < z \leq 4$  to get the galaxy number density in the same redshift range ( $3 < z \leq 4$ ). By integrating the SMF at  $\log(M/M_\odot) > 10$  we find  $\rho_N = 6.6_{-0.4}^{+1.6} \times 10^{-6} \text{ Mpc}^{-3}$ , a factor 3–4 lower than the quiescent galaxies of ZFOURGE, for which  $\rho_N = (2.3 \pm 0.6) \times 10^{-5} \text{ Mpc}^{-3}$ . Also at  $2.5 < z < 3$  ZFOURGE quiescent galaxies are more numerous than ours, with their SMF being higher, especially in the low-mass regime (cf. Tomczak et al. 2014). The difference is mainly due to the classification method. Using  $UVJ$ , the ZFOURGE passive sample includes galaxies with  $sSFR \simeq 10^{-10} \text{ yr}^{-1}$  (especially at  $M < M_*$ ) that are excluded in our selection. Another reason for such a discrepancy may be the AGN contamination. For example, in Spitler et al. (2014), 9 out of their 26 quiescent galaxies at  $3 < z < 4$  are potential AGN.

We also compute the passive galaxy fraction ( $f_{\text{pas}}$ ) defined as the ratio between passive and total SMFs (Fig. 12, bottom panels). The peak of  $f_{\text{pas}}$  is always located at  $10.5 < \log(M/M_\odot) < 10.9$ , with values decreasing from 12 to 5%. The evaluation at  $3.5 < z \leq 4$  is more uncertain if we let Schechter parameters free in the fit, nonetheless  $f_{\text{pas}}$  remains below 10%.

#### 5.5. Impact of the Eddington bias

Describing the low-mass end of the SMF (or better, its slope) by means of the Schechter parameter  $\alpha$  is crucial to obtaining a comprehensive view of stellar mass assembly. The probe

**Table 1.** Schechter parameters of the COSMOS2015 galaxy SMF (also dividing the sample into active and passive galaxies).

Redshift	$\log M_\star$ [ $h_{70}^{-2} \mathcal{M}_\odot$ ]	$\alpha_1$	$\Phi_{\star 1}$ [ $10^{-3} h_{70}^2 \text{Mpc}^{-3}$ ]	$\alpha_2$	$\Phi_{\star 2}$ [ $10^{-3} h_{70}^2 \text{Mpc}^{-3}$ ]
Total sample					
$0.2 < z \leq 0.5$	$10.78^{+0.13}_{-0.14}$	$-1.38^{+0.08}_{-0.25}$	$1.187^{+0.633}_{-0.969}$	$-0.43^{+0.62}_{-0.60}$	$1.92^{+0.73}_{-0.78}$
$0.5 < z \leq 0.8$	$10.77^{+0.09}_{-0.08}$	$-1.36^{+0.05}_{-0.06}$	$1.070^{+0.287}_{-0.315}$	$0.03^{+0.43}_{-0.43}$	$1.68^{+0.33}_{-0.33}$
$0.8 < z \leq 1.1$	$10.56^{+0.05}_{-0.05}$	$-1.31^{+0.05}_{-0.06}$	$1.428^{+0.306}_{-0.308}$	$0.51^{+0.35}_{-0.34}$	$2.19^{+0.40}_{-0.41}$
$1.1 < z \leq 1.5$	$10.62^{+0.08}_{-0.07}$	$-1.28^{+0.05}_{-0.05}$	$1.069^{+0.222}_{-0.240}$	$0.29^{+0.40}_{-0.42}$	$1.21^{+0.23}_{-0.22}$
$1.5 < z \leq 2.0$	$10.51^{+0.08}_{-0.07}$	$-1.28^{+0.06}_{-0.06}$	$0.969^{+0.202}_{-0.208}$	$0.82^{+0.48}_{-0.52}$	$0.64^{+0.18}_{-0.17}$
$2.0 < z \leq 2.5$	$10.60^{+0.15}_{-0.12}$	$-1.57^{+0.12}_{-0.21}$	$0.295^{+0.173}_{-0.177}$	$0.07^{+0.70}_{-0.74}$	$0.45^{+0.12}_{-0.12}$
$2.5 < z \leq 3.0$	$10.59^{+0.36}_{-0.36}$	$-1.67^{+0.26}_{-0.26}$	$0.228^{+0.300}_{-0.300}$	$-0.08^{+1.73}_{-1.73}$	$0.21^{+0.14}_{-0.38}$
$3.0 < z \leq 3.5$	$10.83^{+0.15}_{-0.15}$	$-1.76^{+0.13}_{-0.11}$	$0.090^{+0.064}_{-0.039}$		
$3.5 < z \leq 4.5$	$11.10^{+0.21}_{-0.21}$	$-1.98^{+0.14}_{-0.13}$	$0.016^{+0.020}_{-0.009}$		
$4.5 < z \leq 5.5$	$11.30^{+1.22}_{-1.22}$	$-2.11^{+0.34}_{-0.22}$	$0.003^{+0.002}_{-0.002}$		
Active sample					
$0.2 < z \leq 0.5$	$10.26^{+0.07}_{-0.06}$	$-1.29^{+0.03}_{-0.03}$	$2.410^{+0.341}_{-0.337}$	$1.01^{+0.34}_{-0.36}$	$1.30^{+0.25}_{-0.25}$
$0.5 < z \leq 0.8$	$10.40^{+0.06}_{-0.06}$	$-1.32^{+0.02}_{-0.02}$	$1.661^{+0.188}_{-0.176}$	$0.84^{+0.25}_{-0.26}$	$0.86^{+0.13}_{-0.12}$
$0.8 < z \leq 1.1$	$10.35^{+0.05}_{-0.05}$	$-1.29^{+0.02}_{-0.02}$	$1.739^{+0.166}_{-0.164}$	$0.81^{+0.22}_{-0.24}$	$0.95^{+0.12}_{-0.11}$
$1.1 < z \leq 1.5$	$10.42^{+0.05}_{-0.05}$	$-1.21^{+0.02}_{-0.02}$	$1.542^{+0.127}_{-0.122}$	$1.11^{+0.20}_{-0.21}$	$0.49^{+0.07}_{-0.07}$
$1.5 < z \leq 2.0$	$10.40^{+0.05}_{-0.05}$	$-1.24^{+0.02}_{-0.02}$	$1.156^{+0.107}_{-0.105}$	$0.90^{+0.23}_{-0.24}$	$0.46^{+0.07}_{-0.06}$
$2.0 < z \leq 2.5$	$10.45^{+0.07}_{-0.07}$	$-1.50^{+0.05}_{-0.05}$	$0.441^{+0.093}_{-0.088}$	$0.59^{+0.34}_{-0.35}$	$0.38^{+0.06}_{-0.06}$
$2.5 < z \leq 3.0$	$10.39^{+0.14}_{-0.10}$	$-1.52^{+0.07}_{-0.08}$	$0.441^{+0.149}_{-0.137}$	$1.05^{+0.54}_{-0.61}$	$0.13^{+0.04}_{-0.04}$
$3.0 < z \leq 3.5$	$10.83^{+0.08}_{-0.08}$	$-1.78^{+0.05}_{-0.05}$	$0.086^{+0.027}_{-0.021}$		
$3.5 < z \leq 4.0$	$10.77^{+0.12}_{-0.11}$	$-1.84^{+0.10}_{-0.09}$	$0.052^{+0.028}_{-0.020}$		
$4.0 < z \leq 6.0$	$11.30^{+0.15}_{-0.15}$	$-2.12^{+0.05}_{-0.05}$	$0.003^{+0.001}_{-0.000}$		
Passive sample					
$0.2 < z \leq 0.5$	$10.83^{+0.07}_{-0.08}$	$-1.30^{+0.26}_{-0.43}$	$0.098^{+0.177}_{-0.177}$	$-0.39^{+0.34}_{-0.20}$	$1.58^{+0.16}_{-0.21}$
$0.5 < z \leq 0.8$	$10.83^{+0.04}_{-0.04}$	$-1.46^{+0.36}_{-0.48}$	$0.012^{+0.024}_{-0.024}$	$-0.21^{+0.14}_{-0.09}$	$1.44^{+0.08}_{-0.08}$
$0.8 < z \leq 1.1$	$10.75^{+0.02}_{-0.02}$	$-0.07^{+0.04}_{-0.04}$	$1.724^{+0.059}_{-0.060}$		
$1.1 < z \leq 1.5$	$10.56^{+0.03}_{-0.03}$	$0.53^{+0.08}_{-0.08}$	$0.757^{+0.023}_{-0.024}$		
$1.5 < z \leq 2.0$	$10.54^{+0.03}_{-0.03}$	$0.93^{+0.13}_{-0.12}$	$0.251^{+0.015}_{-0.017}$		
$2.0 < z \leq 2.5$	$10.69^{+0.07}_{-0.07}$	$0.17^{+0.19}_{-0.16}$	$0.068^{+0.006}_{-0.007}$		
$2.5 < z \leq 3.0$	$10.24^{+0.11}_{-0.11}$	$1.15^{+0.63}_{-0.50}$	$0.028^{+0.007}_{-0.010}$		
$3.0 < z \leq 3.5$	$10.10^{+0.10}_{-0.09}$	$1.15$	$0.010^{+0.002}_{-0.002}$		
$3.5 < z \leq 4.0$	$10.10$	$1.15$	$0.004^{+0.001}_{-0.001}$		

**Notes.** A double Schechter function (Eq. (4)) is used up to  $2.5 < z \leq 3$ ; a single Schechter function (Eq. (3)) beyond that bin.

of low-mass galaxies can be impeded by the survey depth, as it can result in an overly high stellar mass limit (we refer to Weigel et al. 2016; and Parsa et al. 2016, for an analogue discussion on the LF). In our case, the extrapolation of the low-mass end seems reliable, as the estimates below the limiting mass (derived through the  $f_{\text{obs}}$  weights) are in agreement with the fit to  $1/V_{\text{max}}$  points, which stops at  $M_{\text{lim}}$ .

Besides this caveat, another fundamental problem is related to observational errors that alter the galaxy distribution (i.e. the Eddington bias). It has already been shown that the Eddington bias significantly modifies the SMF shape (Ilbert et al. 2013; Caputi et al. 2015; Grazian et al. 2015). Observational (photometric) errors affect the SMF calculation in two ways. First, they introduce an error in  $z_{\text{phot}}$  that spreads out galaxies from the

**Table 2.** Schechter parameters of the COSMOS2015 galaxy SMF at  $z > 3$ , resulting from a fit in which  $\mathcal{M}_\star$  was fixed to  $10^{10.6} \mathcal{M}_\odot$ .

Redshift	$\log \mathcal{M}_\star$ [ $h_{70}^{-2} \mathcal{M}_\odot$ ]	$\alpha_1$	$\Phi_{\star 1}$ [ $10^{-3} h_{70}^3 \text{Mpc}^{-3}$ ]
$3.0 < z \leq 3.5$	10.60	$-1.58^{+0.07}_{-0.07}$	$0.204^{+0.024}_{-0.023}$
$3.5 < z \leq 4.5$	10.60	$-1.65^{+0.10}_{-0.11}$	$0.094^{+0.013}_{-0.013}$
$4.5 < z \leq 5.5$	10.60	$-1.69^{+0.26}_{-0.28}$	$0.038^{+0.011}_{-0.010}$

intrinsic  $N(z)$  galaxy distribution. Moreover, even if the redshift of an object is known precisely, photometric errors can still affect the  $\mathcal{M}$  estimates as they allow a certain number of SEDs (with different  $\mathcal{M}/L$  normalisation) to fit the data reasonably well.

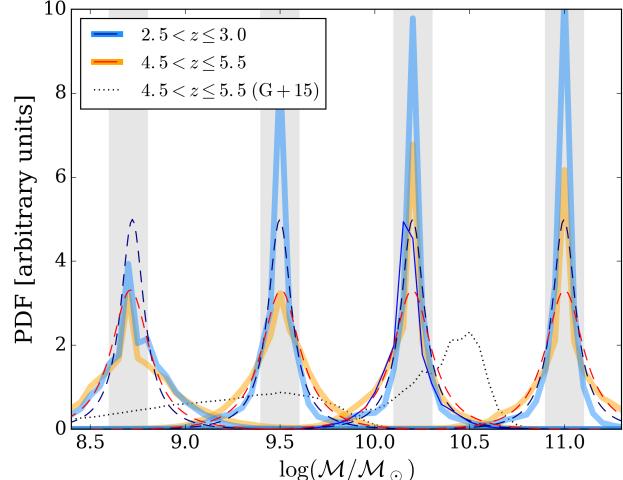
Here, we stress that there is no unique approach to account for such an effect, and the shape of the final bias-free SMF strongly depends on the capability of accurately describing the observational uncertainties ( $\sigma_m$ ). Despite the different implementations, the basic concept is to convolve a pure Schechter function (see Eq. (3)) with  $\sigma_m$ . In this way we construct a function

$$\Phi_{\text{obs}}(z, \mathcal{M}) = \int \Phi(m) \sigma_m(z, m) dm, \quad (6)$$

that can be used in the STY method, or to fit non-parametric estimates like the  $1/V_{\text{max}}$  points. As can be noticed, the main distinction among different correction techniques is how  $\sigma_m$  is defined. This difference can be summarised by comparing Grazian et al. (2015, hereafter G15) and the present analysis. We show below that the two studies, starting from data points that are in excellent agreement, lead to discrepant Schechter functions after the Eddington bias correction, because of different  $\sigma_m$  estimates.

The method we used to estimate the stellar mass uncertainty is fully described in Sect. 4.1. The function  $\sigma_m$  required in Eq. (6) is the one of Eq. (1), with a Lorentzian component that increases with redshift, while the standard deviation of the Gaussian distribution stays constant. Both components are independent of the logarithmic stellar mass, therefore  $\sigma_m$  is symmetric, affecting the SMF mainly in the high-mass end because of the exponential decline of number counts.

On the other hand, the correction implemented by G15 in CANDELS is derived in the following way. For the photometry of each galaxy in a given  $z$  bin (say,  $z_1 < z < z_2$ ) they scan the whole BC03 library, after fixing the redshift, to obtain the PDF( $\mathcal{M}, z = z_i$ ). The fixed redshift  $z_i$  is taken from an equally spaced grid that covers the full range between  $z_1$  and  $z_2$ . The procedure is repeated for each step of the grid. In other words the conditional PDFs are initially computed assuming a flat prior on  $z$ . To include the uncertainty on photometric redshift and obtain PDF( $\mathcal{M}|z$ ), G15 multiply each PDF( $\mathcal{M}, z = z_i$ ) by PDF( $z_i$ ), where the latter is the redshift probability coming from the PDF( $z$ ) that Dahlen et al. (2013) provide for each CANDELS galaxy. The global uncertainty  $\sigma_m(z_1 < z < z_2, \mathcal{M}_j)$  is then obtained by adding the PDFs of all the objects in the given bin of redshift and stellar mass ( $\mathcal{M}_j$ ). Before adding them, each PDF in the stellar mass bin is re-aligned in the centre of the bin. Note that this approach implies a discretised version of Eq. (6):  $\Phi_{\text{obs}}(\mathcal{M}) = \sum_j \Phi(\mathcal{M}_j) \sigma_m(\mathcal{M}_j) \Delta \mathcal{M}_j$ , that is inaccurate if the mass binning is too coarse, much larger than the average PDF (this is not the case of G15, who assume  $\Delta \mathcal{M}_j = 0.2$  dex). Such an estimate of  $\sigma_m$ , unlike ours, depends also on the stellar mass: the



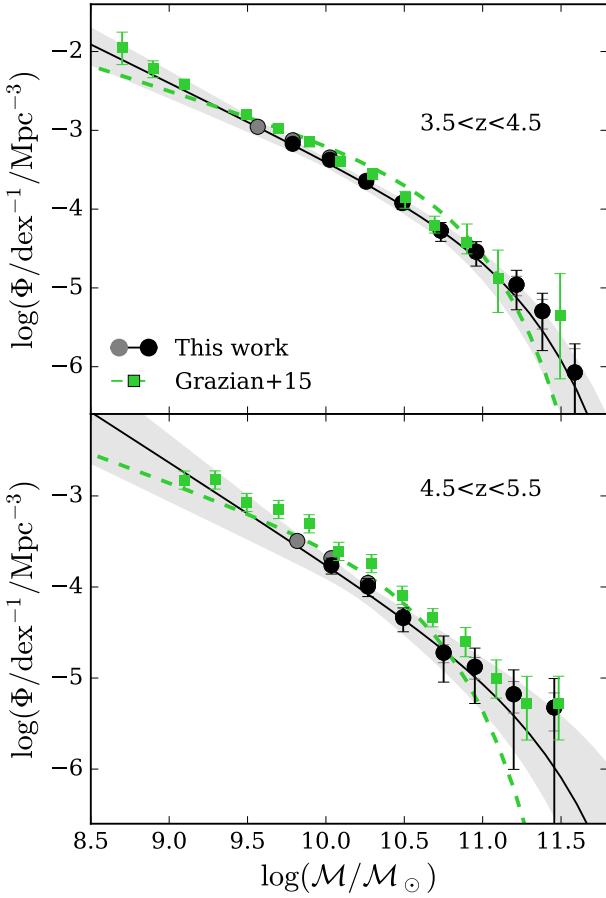
**Fig. 13.** Average PDF( $\mathcal{M}|z$ ) for galaxies in four bins of stellar mass (highlighted with grey vertical bands), at redshift  $2.5 < z \leq 3$  (blue solid lines) and  $4.5 < z \leq 5.5$  (orange). The plot also shows the  $\sigma_m(z)$  uncertainty (see Eq. (1), with dark blue and red dashed lines, in the low and high  $z$ -bin respectively. Two examples of  $z \approx 5$  PDFs from Grazian et al. (2015) are shown with dotted lines; they are obtained by stacking the PDF( $\mathcal{M}|z$ ) of CANDELS galaxies at  $9.4 < \log(\mathcal{M}/\mathcal{M}_\odot) < 9.6$  and  $10.4 < \log(\mathcal{M}/\mathcal{M}_\odot) < 10.6$  respectively (Chabrier's IMF). All the PDFs in the figure have been normalised to unity.

larger the galaxy mass is, the narrower the error (because of the higher  $S/N$  in the photometry). Moreover, the typical PDF( $\mathcal{M}|z$ ), especially below  $10^{10} \mathcal{M}_\odot$ , is very skewed, with a prominent tail towards lower masses (see Fig. 13, and Fig. B.1 of G15).

G15 clearly illustrate the difference between their treatment and Ilbert et al. (2013, whose approach is similar to ours): Starting from an intrinsic (i.e. unbiased) SMF and convolving it by the  $\sigma_m(z)$  used by Ilbert et al., the observed SMF will be a Schechter function with an enhanced number density of massive galaxies, but substantially unchanged at lower masses. Conversely, by applying the  $\sigma_m(z, \mathcal{M})$  computed in G15 to the same intrinsic mass function,  $\Phi_{\text{obs}}$  will have a steeper low-mass end. This latter method should not significantly modify the high-mass end, where the PDF( $\mathcal{M}|z$ ) is expected to be narrower (G15).

Thus,  $\sigma_m$  estimates are different in CANDELS and COSMOS. To find the reason for the discrepancy, we recompute  $\sigma_m$  with a procedure similar to G15, deriving the galaxy PDF( $\mathcal{M}|z$ ) in bins of stellar mass (Fig. 13). Interestingly, our PDF( $\mathcal{M}|z$ ) is well described by the same distribution (Eq. (1)) we found with the method of Ilbert et al. (2013). There is a trend with stellar mass like in G15, but so mild that for galaxies with  $\mathcal{M} \gtrsim \mathcal{M}_{\text{lim}}$  the PDF can be considered constant (at a given redshift, see Fig. 13). So, even with this alternate computation, the stellar mass uncertainties of CANDELS and COSMOS2015 remain at variance. The fact that at a given redshift our  $\sigma_m$  (at least above  $10^9 \mathcal{M}_\odot$ ) does not depend on mass can appear counterintuitive. The SED fitting should be better constrained for massive galaxies since they generally have higher  $S/N$  photometry. This is true for the  $\mathcal{M}$  estimates, when the SED fitting is performed after fixing the redshift. However the  $z_{\text{phot}}$  uncertainty, which is the dominant component in  $\sigma_m$ , is nearly constant in this stellar mass range (see the discussion in Caputi et al. 2015).

Therefore, despite the agreement between the  $1/V_{\text{max}}$  estimates of G15 and ours, the correspondent Schechter functions diverge below  $10^9 \mathcal{M}_\odot$ , after accounting for the Eddington bias (Fig. 14). The flatter slope recovered by G15 is due to the

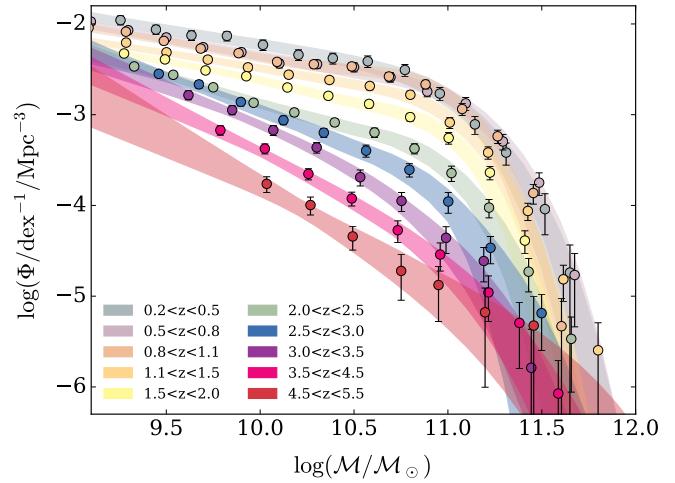


**Fig. 14.** Comparison between the SMF of COSMOS2015 galaxies and Grazian et al. (2015), in two redshift bins between  $z = 3.5$  and  $5.5$  (as indicated in the panels). Our  $1/V_{\max}$  estimates and the fit to those points are shown with black filled circles and solid lines (green squares and dashed lines show the same quantities for Grazian et al. 2015). Grey filled circles are the SMF estimates we obtain using the  $f_{\text{obs}}$  weights (see Sect. 5). In both  $z$ -bins, the  $1\sigma$  uncertainty of the COSMOS2015 Schechter function is encompassed in a shaded area.

mass-dependent  $\sigma_m$  that becomes much wider at lower  $M$ . On the other hand, their fit should be anchored to the more precise stellar mass estimates at  $M > M_*$ , but in practice their resulting Schechter exponential tail rests below the data points because of the correlation between  $\alpha$  and  $M_*$ . Vice versa, with our  $\sigma_m$ , the error deconvolution does not significantly modify the low-mass end while it decreases the number density in the high-mass end<sup>8</sup>. The bias-free Schechter function of G15 turns out to have a low-mass end nearly constant from  $z \sim 4$  to  $6$ , comparable to what is found at  $z \sim 2$ – $3$  (i.e.  $\alpha \simeq -1.6$ ). In particular they find  $\alpha = -1.63 \pm 0.05$  at  $z = 4$  and  $\alpha = -1.63 \pm 0.09$  at  $z = 5$ . Without removing the bias, the G15 Schechter function has  $\alpha = -1.77 \pm 0.05$  and  $-1.90 \pm 0.20$ , at  $z = 4$  and  $5$ , respectively. In the same redshift bins we find  $\alpha = -1.97^{+0.10}_{-0.09}$  and  $-2.11^{+0.30}_{-0.13}$ .

These tests indicate that the different characterisation of  $\sigma_m$  is the major aspect responsible for the contrasting Schechter fits in the two analyses. The  $z_{\text{phot}}$  uncertainties are the most plausible cause of discrepancy but at present we cannot establish whether

<sup>8</sup> Song et al. (2016) state that, after accounting for Eddington bias, the low-mass end of their SMF does not change significantly, as in our case. However, details about how the bias correction has been implemented are not provided in their paper.



**Fig. 15.** Evolution of the SMF between  $z = 0.2$  and  $5.5$ , for the COSMOS2015 galaxy sample. Filled circles show the  $1/V_{\max}$  estimates, and shaded areas show the  $1\sigma$  uncertainty of the best Schechter function fitting to them (as in Figs. 8 and C.2). Colours indicating the redshift bins are summarised in the bottom-left corner of the plot.

the PDFs( $z$ ) of COSMOS2015 are underestimated, or those in CANDELS overestimated. We aim at investigating this issue in future work.

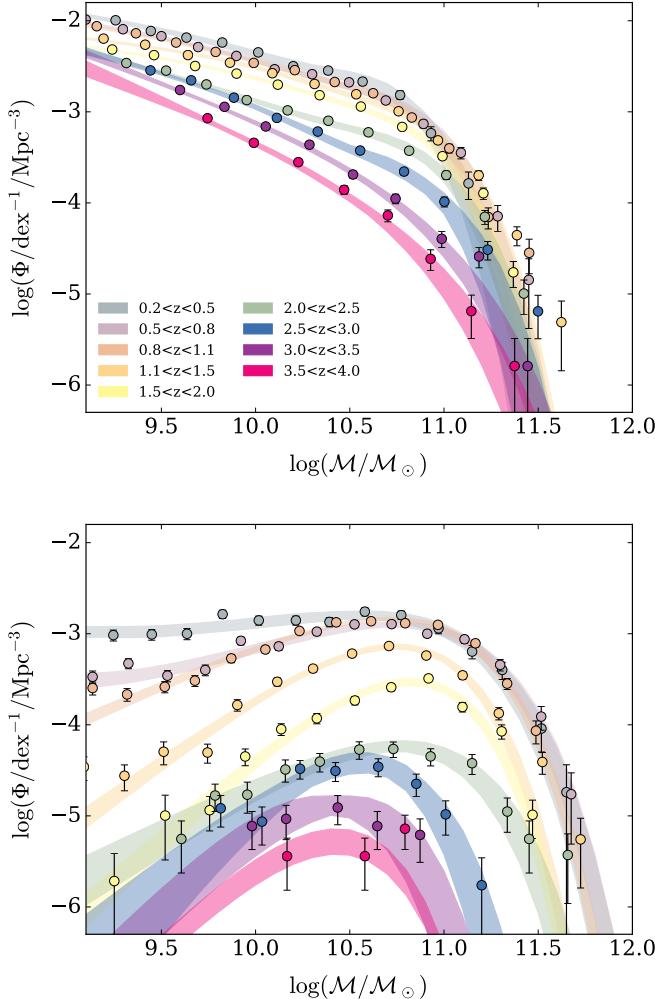
## 6. Discussion

### 6.1. Stellar mass assembly between $z \sim 0$ and $6$

To study the SMF evolution over a larger time interval, we combine our results at  $z > 2.5$  (Sect. 5) with the SMFs at lower redshifts. The latter ones are obtained using the SED fitting estimates of L16, whose main features have already been described (see Sects. 3.1 and 5.2). To have a comprehensive view, in Fig. 15 we overplot the estimates from  $z = 0.2$  to  $5.5$ . In Fig. 16 we also show the SMFs of star-forming and quiescent objects, up to  $z = 4$ . Additional plots of the  $0 < z \leq 4$  SMFs are shown in Appendix C.

The resulting picture shows a progressive build-up of galaxies at  $10.0 < \log(M/M_\odot) < 11.5$ , sharpening the knee of the SMF as time goes by. This feature becomes stable at  $z \lesssim 2$ , since the SMF grows in normalisation but the shape of the exponential tail remains nearly the same. In comparison, there is little increase in the number density of galaxies with  $\log(M/M_\odot) \leq 10.0$  across the whole redshift range. In the bin centred at  $\log(M/M_\odot) = 9.5$ , where our data provide a direct constraint in all the  $z$ -bins, the increase in number density is 6–7 times smaller than at  $10.5$ . The combined effect of such a differential evolution is a flattening of the low-mass end as one moves to lower redshifts. In the early universe (<2 Gyr old) the best fit is a single Schechter function, with the addition of a secondary component only at  $z < 3$ . We observe that this evolutionary trend does not imply necessarily a change of  $M_*$ . If galaxies below the turnover mass outnumber the most massive ones, the “dip” of the Schechter function is smoothed out even if  $M_*$  remains constant (see Tomczak et al. 2014; Song et al. 2016).

These observations are consistent with models in which the suppression of star formation (“quenching”) is particularly efficient when galaxies reach a given stellar mass threshold ( $\sim M_*$ ). In this way, star forming galaxies that accreted up to such a mass cannot easily grow further, so they accumulate at  $M_*$ . This



**Fig. 16.** Evolution of the SMF between  $z = 0.2$  and  $4$ , for active (upper panel) and passive (lower panel) galaxies. Same symbols as in Fig. 15.

effect is even more evident in the active SMF (Fig. 16, upper panel), which does not extend beyond  $\log(M/M_\odot) = 11.5$ . This kind of mass-dependent quenching could be caused by internal processes, for example, AGN feedback or heating via stable virial shocks (Gabor et al. 2010, and references therein). Without any assumption regarding the underlying physics, the empirical model of Peng et al. (2010) shows how a galaxy SMF that is a power-law function at  $z = 10$  will assume a Schechter profile at lower  $z$ , mainly because of the action of mass quenching. Potentially confirming this picture, the COSMOS2015 SMF, moving towards higher  $z$ , starts to resemble a power law (Eq. (5)). We caution that this clue may actually be the effect of galaxy interlopers on the high-mass end, however it is not implausible that the SMF departs from a Schechter function at  $z \gtrsim 6$ , to reproduce more closely the shape of the underlying dark matter (DM) distribution (see below). Similarly, Bowler et al. (2015, 2017) find evidence that the UV LF of  $z \approx 7$  galaxies is better fit with a double power law.

The SMF of  $NUVrJ$ -passive galaxies (Fig. 16, lower panel) agrees with this scenario, with a distinct  $\log(M_\star/M_\odot) = 10.5\text{--}10.8$  peak even at  $z > 3$ . The most significant growth of the passive sample, in terms of number density, happens from  $z = 2.5$  to  $1$ . A substantial increase (by a factor  $\times 4$ ) is observed in particular from  $2 < z < 2.5$  to  $1.5 < z < 2$  (i.e. in less than 1 Gyr). This is consistent with previous studies that indicate that

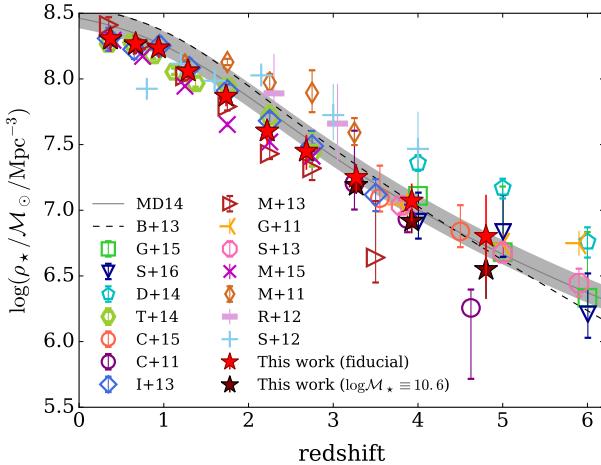
local early-type galaxies with  $11 < \log(M/M_\odot) < 12$  entered into their quiescent phase between  $z \approx 0.8$  and  $2.5$  (Thomas et al. 2010).

The build-up of passive galaxies corresponds to a transition of the total SMF from a single to a double Schechter function. This is only an approximate scheme, because the emerging secondary component cannot be fully ascribed to quenching. Also, the active SMF is better fit by a double Schechter function at least at  $z < 2.5$  (see also Ilbert et al. 2013; Tomczak et al. 2014). When the active sample is divided into two or more classes – for example, distinguishing between intermediate and high sSFR or different morphologies, as in (Ilbert et al. 2010) – each SMF is well described by a single Schechter function. From a morphological analysis of  $z < 0.06$  galaxies, Moffett et al. (2016, GAMA survey) find that the double Schechter profile of the active SMF is the sum of the SMF of Sd and irregular galaxies (dominant at the low-mass end) and the one of Sa to Sc types (which creates the dip at intermediate masses). With irregular galaxies being more common in earlier epochs, the result should be a single Schechter at high  $z$ , as observed. Moreover, Moffett et al. (but also Kelvin et al. 2014) find a precise decomposition of their local SMF in two Schechter functions by simply dividing disc- and bulge-dominated galaxies. Without speculating further, we simply remark that a similar morphological transformation, characterised by an “inside-out” quenching and bulge growth, is expected to begin at  $z \approx 2.5$  (according to recent simulations as Tacchella et al. 2016), that is, the epoch when we observe a secondary low- $z$  component emerge in the SMF.

We also determine the stellar mass density ( $\rho_*$ ) as a function of  $z$ . This is usually done by integrating the Schechter function between  $10^8$  and  $10^{13} M_\odot$ . Since our  $M_{\text{lim}}$  is larger than  $10^9 M_\odot$  at  $z > 2$ , the computation at high redshift is extremely sensitive to the extrapolation of the low-mass end below our data point (see Sect. 5.5). We show in Fig. 17 several  $\rho_*$  estimates from COSMOS2015 and other surveys, compared to the stellar mass density derived via integration of the SFR density (SFRD) function (as given in Behroozi et al. 2013; Madau & Dickinson 2014). The difference between the two methods is smaller than in the analogous plot shown in Madau & Dickinson (2014), where estimates derived from SED fitting are  $\sim 0.2$  dex lower than  $\rho_*$  from SFRD (their Fig. 11). As explained in that paper, the level of consistency also depends on the assumed IMF. Time integration of the SFRD takes into account the gas recycling fraction ( $f_{\text{return}}$ ), which is 0.41 for Chabrier’s and 0.27 for Salpeter’s IMF. Since we use the former, the resulting stellar mass density is  $\sim 0.1$  dex smaller than the one obtained by Madau & Dickinson (2014) starting from the same SFRD function.

In Fig. 17 we also see that our fiducial SMF at  $z \geq 4$  originates higher  $\rho_*$  values than the fit with fixed  $M_\star$ , whose main difference is indeed the flatter low-mass end. Both estimates are nonetheless consistent, at within  $1\sigma$  from each other, and in fairly good agreement with  $\rho_*$  from Behroozi et al. (2013) and Madau & Dickinson (2014). The tension with the SFRD predictions starts to be evident when considering, for example, Santini et al. (2012) or Duncan et al. (2014), whose SMFs are even steeper.

A precise determination of  $\alpha$  is also pivotal in the essential formalism of those empirical models (e.g. Peng et al. 2010; Boissier et al. 2010) that try to connect the SMF evolution to the main sequence (MS) of star forming galaxies (Noeske et al. 2007; Daddi et al. 2007; Elbaz et al. 2007). Reconciling the galaxy growth predicted by the MS with the redshift evolution of  $\alpha$  is an effective way of constraining stellar mass assembly and



**Fig. 17.** Redshift evolution of  $\rho_*$ , as measured in different papers by integration of the SMF: Caputi et al. (2011, C+11 in the legend), Caputi et al. (2015, C+15), Duncan et al. (2014, D+14), González et al. (2011, G+11), Grazian et al. (2015, G15), Ilbert et al. (2013, I+13), Mortlock et al. (2011, M+11), Mortlock et al. (2014, M+15), Muzzin et al. (2013a, M+13), Reddy et al. (2012, R+12), Santini et al. (2012, S+12), Song et al. (2016, S+16), and Tomczak et al. (2014, T+14). If  $\rho_*$  uncertainties are not quoted in the paper, we plot approximate error bars by considering the  $1\sigma$  error of the  $\alpha$  parameter. Red stars are the stellar mass density from our fiducial Schechter, brown stars are from the fit with fixed  $M_*$ . By integrating their SFRD functions, we can plot  $\rho_*(z)$  from Behroozi et al. (2013, black dashed line) and Madau & Dickinson (2014, grey solid line). In both integrations we assume  $f_{\text{return}} = 41\%$  (coherently with Chabrier's IMF). For Madau & Dickinson (2014) we also show with a shaded area the  $\rho_*$  range enclosed by  $f_{\text{return}} = 50\%$  and  $25\%$  (the latter value is similar to the one prescribed by Salpeter's IMF).

quenching mechanisms (e.g. Leja et al. 2014; Steinhardt et al. 2017).

## 6.2. Dark matter connection

To better understand the evolution of the SMF we investigate the relation between galaxy stellar mass and DM halo mass assembly. As pointed out by Lilly et al. (2013), galaxy sSFR and the specific mass increase rate of DM haloes (sMIR  $\equiv \dot{M}_h^{-1} dM_h/dt$ , see e.g. Neistein & Dekel 2008) evolve in a similar way, as expected if star formation is regulated by the amount of cold gas in the galaxy reservoir, which in turn depends on the inflow of DM into the halo (Lilly et al. 2013; Saintonge et al. 2013).

We compare the SMF of COSMOS2015 galaxies to the halo mass function (HMF) provided in Tinker et al. (2008)<sup>9</sup>. Recently, a discrepancy between these two quantities has been highlighted by Steinhardt et al. (2016): the most massive galaxies observed at  $z > 4$  seem to be too numerous compared to the haloes that should host them. Such an excess, if confirmed, would call into question either theoretical aspects of the  $\Lambda$ CDM model or some fundamental principle of galaxy formation (we refer to the discussion about these “impossibly early galaxies” in Steinhardt et al. 2016).

<sup>9</sup> The HMF has been computed in our  $z$ -bins and cosmological framework ( $\sigma_8 = 0.82$ ) by means of the code HMFCalc (Murray et al. 2013). The code allows us to choose among alternate models (e.g. Sheth et al. 2001; Tinker et al. 2010; Angulo et al. 2012) without any significant impact on our conclusions.

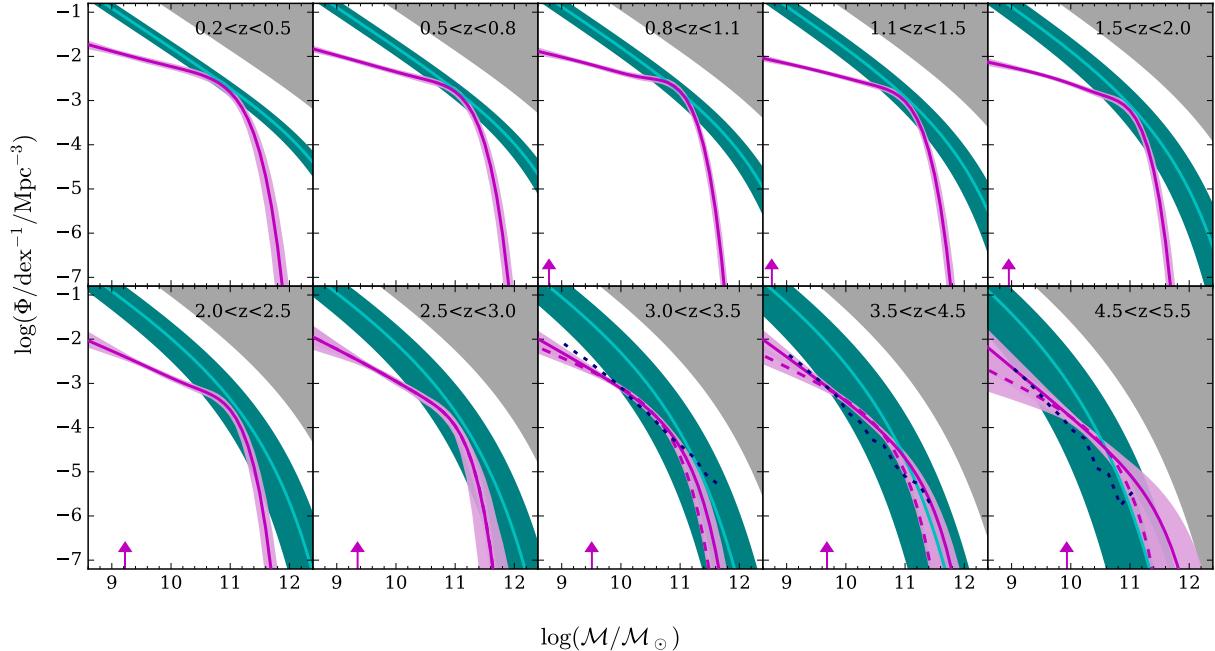
The co-evolution of SMF and HMF between  $z = 0.2$  and 5.5 is shown in Fig. 18. For a synoptic view, we superimpose the HMF on the SMF, rescaling  $M_h$  by a constant factor equal to 0.018. This scaling factor is the stellar-to-halo mass ratio (SMHR  $\equiv M/\bar{M}_h$ ) provided in Behroozi et al. (2013, see their Eq. (3)) for a typical  $M_h^*$  halo at  $z = 0$ <sup>10</sup>. We emphasise that the same rigid translation is applied in each  $z$ -bin, simply to ease the comparison between the HMF and the SMF shapes. A thorough link between haloes and galaxies, for example via abundance matching, is deferred to future work.

At  $z < 2$ , Fig. 18 (upper row of panels) illustrates a well-known result. The shape of stellar and halo mass functions do not coincide, at neither  $M < M_*$  nor  $M > M_*$ . Reconciling the observed SMF with the DM distribution has required the introduction of quenching mechanisms in galaxy formation models (Baugh 2006, for a review). Star formation of low-mass galaxies is assumed to be halted via stellar feedback, for example, stellar winds or supernova explosions that heat/eject gas (Larson 1974; Dekel & Silk 1986; Leitherer et al. 1999). In galaxies at  $M > M_*$ , hot halo gas is removed or prevented from cooling (e.g. by AGN outflows, Fabian 2012) or virial shock heating (Dekel & Birnboim 2006).

As for the transition from single to double Schechter function, the epoch of a key change is  $z = 2$ –3. In fact, the tension between SMF and HMF lessens at  $z > 2$  (Fig. 18, lower panels). In the high-mass regime, the SMF exponential tail moves closer to the rescaled HMF, until they overlap at  $z > 3$ . Considering the crude rescaling (i.e. the 0.018 factor) and the SMF uncertainties at high  $z$ , the match between massive galaxies and massive DM haloes is excellent. At  $4.5 < z \leq 5.5$ , the massive end of our fiducial SMF is slightly higher than the HMF, but still compatible within the errors. Such an excess of observed galaxies does not challenge the theoretical framework, since small modifications, for example, to the HMF scaling factor (which has been fixed to the SHMR at  $z = 0$ ) would be enough to reconcile the two functions.

To show that there is no substantial inconsistency between the two functions, we derive from the HMF an upper limit for the SMF. Starting from the present baryon density  $\Omega_{b,0} = 0.0486$  (Planck Collaboration XIII 2016), we assume  $\Omega_b/\Omega_m$  as a SHMR with a baryon-to-stellar mass conversion of 100%. Rescaling the HMF accordingly, we obtain the maximal SMF physically allowed (grey shaded area in Fig. 18). The observed SMF is always below this upper limit. In other words, we do not find any impossibly early galaxy, at least at  $z < 6$ . Steinhardt et al. (2016) discuss this critical issue relying on UV LFs up to  $z \sim 10$  (Bouwens et al. 2015; Bouwens 2016). In this respect, Mancuso et al. (2016) claim that the tension between the observed number density of massive galaxies and the predicted abundance of their host haloes is mainly due to the dust corrections applied to UV data. When including far-IR data to determine the SFR function, they find that the formation of stars in  $z \geq 4$  massive haloes is not required to start as early as argued in Steinhardt et al. (2016). This kind of bias does not affect our comparison, which however probes  $z < 6$ . Conclusive

<sup>10</sup>  $M_h^*$  is the characteristic halo mass that marks the peak at which the integrated star formation is most efficient. At  $z = 0$  it is about  $10^{12} M_\odot$  (Behroozi et al. 2013). Roughly speaking,  $M_h^*$  separates the SHMR behaviours at low and high halo masses. In Behroozi et al. (2013) the  $M$ - $M_h$  relation is calibrated against  $0 \leq z \leq 8$  data (several SMF, specific SFR, and cosmic SFR estimates) through a Markov chain Monte Carlo. We verified that adopting the SHMR of Moster et al. (2013), constrained via galaxy-halo abundance matching, differences are within the  $1\sigma$  error bars.



**Fig. 18.** Evolution of HMF (cyan line) vs. SMF (magenta) from  $z = 0.2$  to  $5.5$ . In each panel halo mass is multiplied by a factor 0.018, that is, the SHMR at  $z = 0$  and  $\mathcal{M}_h = \mathcal{M}_h^*$  (according to Behroozi et al. 2013). Shaded cyan regions show the uncertainties in the HMF shift by taking the  $1\sigma$  error of the SHMR parametrisation. At  $z > 2.5$ , the solid magenta line is our fiducial fit of the SMF, while the dashed magenta line is the Schechter function with  $\log(\mathcal{M}_*/\mathcal{M}_\odot)$  fixed to 10.6 (see Fig. 8). The shaded magenta regions combine the  $1\sigma$  CL of the two fits, to give a conservative estimate of the uncertainties. A grey shaded area highlights the “forbidden region” where, according to the  $\Omega_b/\Omega_m$  ratio, no galaxies are expected. Black dotted lines at  $z > 3$  are the SMFs predicted from the semi-analytical model of Garel et al. (2016), converted to the IMF of Chabrier (2003).

evidence on this issue will come from next-generation surveys, which shall provide direct measurements of the SMF at  $z > 7$  (i.e. the epoch when these discrepancies should be the largest Steinhardt et al. 2016).

The high-mass end of our SMF declines with a slope similar to the HMF. Such an agreement suggests that massive galaxies at  $z > 3$ , which resides in  $\gtrsim \mathcal{M}_h^*$  haloes, all have similar SHMR ( $\sim 2\%$ ). However, one should discriminate between the contribution of central and satellite galaxies before making conclusions about the star-formation efficiency in distinct haloes (see Coupon 2015). A nearly constant SHMR for  $\mathcal{M}_h > \mathcal{M}_h^*$  is at odds with Behroozi et al. (2013) and Moster et al. (2013) results, but a few caveats concerning those two studies should be noted. Moster et al. (2013) relation has been calibrated by means of an abundance matching technique up to  $z = 4$ . The SMFs they used (Santini et al. 2012; Pérez-González et al. 2008) are more plagued by sample variance, and at the massive end their technique is sensitive to the assumptions made for correcting observational errors. Behroozi et al. used a larger set of SMFs, but at  $z > 4$  these are all derived from LBG samples (Stark et al. 2009; Lee et al. 2012; Bouwens et al. 2011; Bradley et al. 2012), possibly biased at high masses as we discussed in Sect. 5.3. Interestingly, if the  $\mathcal{M}-\mathcal{M}_h$  relation of Behroozi et al. (2013) is recomputed without  $z > 4$  observational priors, the SHMR at  $z \approx 4$  and 5 is flatter (and only marginally consistent within the  $1\sigma$  errors of the original fit, see Behroozi & Silk 2015). Thus, the COSMOS2015 catalogue represents a novel opportunity to investigate the connection between DM and stellar content in a redshift and mass regime where previous SHMR estimates were lacking robust observational constraints (Coupon et al., in prep.).

### 6.3. A reduced impact of star formation feedback

At  $z > 3$ , therefore, there is no need for additional quenching to reconcile the abundance of  $\mathcal{M} > \mathcal{M}_*$  galaxies with

that of  $\mathcal{M}_h > \mathcal{M}_h^*$  haloes. This finding supports results from simulations, in which massive systems at high redshifts are weakly affected by AGN activity. An example is the cosmological hydrodynamical simulation Horizon-AGN (Dubois et al. 2014): at  $z > 3$ , the SMF of the Horizon-AGN simulated galaxies does not change significantly if AGN feedback is switched off (Kaviraj et al. 2017). In addition, Horizon-AGN allows one to follow in detail the evolution of central black holes. By tracing their mass assembly as a function of time, Volonteri et al. (2016) find that black holes with Eddington ratio  $< 0.01$  (those responsible for radio-mode feedback in the simulation) are the dominant population at  $z \lesssim 2$ , while at  $z > 3$  most of the black holes are fast accretors (Eddington ratio  $> 0.1$ ), luminous enough to trigger a radiative feedback (the so-called “quasar mode”). Smaller occurrence of radio-mode feedback at high redshifts is also suggested by observational studies on radio-loud AGN, whose volume density decreases as a function of  $z$  (Padovani et al. 2015).

Radiative AGN feedback is also expected to be inefficient: In hydrodynamical simulations of  $z \sim 6$  galaxies, outflows generated by bright quasars tend to escape from the direction of least resistance, without interacting with the dense filamentary structure around the galaxy (Costa et al. 2014). Interferometric observations of high- $z$  targets are in line with these theoretical results. Their CO and [CII] mapping indicates that, although AGN are able to remove a large amount of molecular gas (Maiolino et al. 2012; Cicone et al. 2014), they do not prevent extended cold clouds from fueling star formation (Cicone et al. 2015).

Nevertheless,  $z \sim 5-6$  radiative outflows, despite their relatively weak impact on interstellar medium, can perturb the cold filamentary accretion over larger time scales. For instance, quasar energy injection can cause a sort of starvation in the later stages of a galaxy’s life (Dubois et al. 2013; Curtis & Sijacki 2016). Our results constrain the timescale of quasar-mode feedback. Even if central black holes are at work in

the early universe, their effect (i.e. a deviation of the SMF from the HMF high-mass end) is observed only at  $z \leq 3$ . This means that such a quenching mechanism is effective on timescales larger than 2 Gyr, likely after multiple outflow episodes.

We can also compare stellar mass and halo mass functions at  $\mathcal{M} < \mathcal{M}_\star$  (or equivalently  $\mathcal{M}_h < \mathcal{M}_h^\star$ ). A similar exercise has been made in [Song et al. \(2016\)](#) up to  $z \approx 7$ , finding that the low-mass end of their SMF has a slope similar to the HMF at  $z \gtrsim 7$ . In our SMF, the low-mass end becomes steeper, and slightly closer to the HMF, already at  $z > 3.5$ . In terms of Schechter parametrisation, we find that  $\alpha$  ranges between  $-1.4$  and  $-1.2$  at  $z \leq 2$ , becomes  $\approx -1.7$  at  $2 < z \leq 3.5$  and eventually is  $\lesssim -2$  at  $z \approx 4-5$  (while [Song et al. 2016](#) find  $\alpha < -1.9$  only beyond this redshift range). Although the difference in  $\alpha$ , the low-mass end of both SMFs diverges from the HMF, as expected (e.g. from the simulations in [Costa et al. 2014](#)) if stellar feedback remains efficient at least up to  $z \approx 6$ .

In addition, we compare to the semi-analytical model of [Garel et al. \(2015, 2016\)](#), specifically designed to study Ly $\alpha$  emitters and LBGs. At face value, the slope of the SMF predicted by [Garel et al. \(2016\)](#) is similar to ours (Fig. 18), suggesting that their model may be an effective description of stellar feedback in the early universe. A more detailed comparison with simulations is deferred to another paper of this series (Laigle et al., in prep.).

## 7. Summary and conclusions

Relying on the latest photometric catalogue in the COSMOS field (COSMOS2015, [Laigle et al. 2016](#)), we have estimated the SMF of galaxies between  $z = 0$  and  $6$ . A deep NIR coverage from UltraVISTA, and associated SPLASH images in MIR, allowed us to probe the high-mass end of SMF as well as  $\mathcal{M} \lesssim \mathcal{M}_\star$  across the whole redshift range. In particular our SMF reaches  $5-10 \times 10^9 \mathcal{M}_\odot$  at  $z \approx 5$ , an unprecedented mass regime at that redshift for a ground-based survey covering such a large area. One of the reasons for this achievement is the panchromatic detection technique we applied, based on a  $\chi^2$  stacking of images from  $z^{++}$  to  $K_s$ . Our stellar mass completeness limit  $\mathcal{M}_{lim}(z)$  is almost 0.5 dex smaller than what would result from a single (e.g.  $K$ ) band selection, as done in previous analyses. Deeper HST data available in the overlap with the CANDELS field ([Nayyeri et al. 2017](#)) have been used to confirm the absence of significant biases in our final sample, cut at  $[3.6] < 25$ .

A comparison with the literature showed the improvements in terms of statistics with respect to HST surveys at high  $z$  (e.g. [Santini et al. 2012](#); [Duncan et al. 2014](#)). The large volume of COSMOS has allowed us to collect rare massive galaxies, although most of those at  $z > 2$  are severely reddened by dust, increasing the uncertainties in their  $z_{phot}$  and  $\mathcal{M}$ . Comparing to the SMF of LBGs, where such massive and dusty galaxies may be totally missing, we stressed the importance of high-quality MIR data to build mass-selected galaxy samples at  $z > 4$ . Now in its final phase, eight years after the end of the cryogenic mission, the *Spitzer* program will soon be superseded by the *James Webb* Space Telescope (JWST), to shed light on this peculiar galaxy population.

Besides the emphasis on the early universe, we remark that our SMF covers in a coherent way a time interval larger than previous work, providing an overview of the last  $\sim 13$  Gyr of galaxy evolution. In addition we robustly selected (via the  $NUVrJ$  diagram) active and passive galaxies, deriving their SMF up to  $z = 4$ . At higher redshifts the contribution of the passive sample to the SMF becomes negligible, and the few passive galaxies we found need follow-up observations in order to be confirmed.

Concerning the whole evolutionary path from  $z \sim 6$  to  $\sim 0$ , our results are summarised in the following.

1. Considering the growth of the SMF with cosmic time, we marked  $z \approx 3$  as a key moment of galaxy evolution. At lower redshifts, the best fit to the SMF is a double Schechter function. At  $z > 3$  the SMF shows a smoother profile (especially at the knee of the function) and at  $z > 5$  it can be fit also by a power-law function with a cut-off at  $\sim 3 \times 10^{11} \mathcal{M}_\odot$ . To a first approximation the emergence of an additional component in the low- $z$  SMF is related to the assembly of the passive galaxy sample. However also the active SMF is fit by a double Schechter (a feature often ignored e.g. in some phenomenological models). The fact that star formation starts fading in the core of galaxies already at  $z \approx 2-2.5$  ([Taccella et al. 2015, 2016](#)) may be a hint that the change of SMF shape at  $z \approx 3$  is related to such inside-out quenching. Further evidence is needed to verify this hypothesis.
2. At  $z \gtrsim 3$  we also found a change in the relation between stellar mass and halo mass functions. While at  $z \lesssim 2$  the SMF shape diverges from the HMF in both mass regimes (above and below  $\mathcal{M}_\star$ ), at higher redshifts the massive end of the SMF has the same slope of the HMF. This implies that the  $\mathcal{M}/\mathcal{M}_h$  ratio is roughly constant at  $\mathcal{M} \gtrsim \mathcal{M}_\star$ . We interpret this trend as evidence of a reduced quenching for massive galaxies at  $z \gtrsim 3$ , such that the star formation process becomes dominated by simple baryonic cooling. Thus, according to our observations, AGN do not trigger significant feedback (either in radio or quasar mode) during the first  $\sim 2$  Gyr after the Big Bang. This is consistent to hydrodynamical simulations in which AGN ejecta at high- $z$  can hardly stop (or prevent) star formation.
3. There is a progressive flattening of the SMF low-mass end since  $z \sim 6$ . For massive galaxies ( $\gtrsim 5 \times 10^{10} \mathcal{M}_\odot$ ), number density increases about one order of magnitude more than for  $\sim 10^9 \mathcal{M}_\odot$  galaxies. Fitting our data points with a Schechter function, we found  $\alpha$  ranging from  $-2.11^{+0.30}_{-0.13}$  to  $-1.47^{+0.02}_{-0.02}$ , from  $z \approx 5$  to  $0.1$ . A similar slope at  $z > 4$  has been found by other authors (e.g. [Song et al. 2016](#)) and in hydrodynamical simulations ([Garel et al. 2016](#)). Other SMFs are shallower, with  $\alpha \approx -1.6$  up to  $z \sim 6-7$  (see e.g. [Grazian et al. 2015](#)). Such a disagreement is mainly due to the Eddington bias correction: Depending on the characterisation of stellar mass errors made by the authors, the effect of the bias correction on the Schechter function can vary significantly.

Our work shall gain additional momentum from the next wave of UltraVISTA (as the survey is still ongoing) and *Spitzer* images, and also from the development of new tools for data analysis (e.g. [Masters et al. 2015](#); [Speagle et al. 2016](#); [Morrison et al. 2017](#)). Spectro-photometric data from JWST will certainly be beneficial to answering some of the questions raised here, but to probe the massive end of high- $z$  SMFs, the role of large-area surveys such as COSMOS2015 will remain fundamental.

*Acknowledgements.* The authors warmly thank the anonymous referee for her/his constructive comments. The authors thank Shoubaneh Hemmati and Hooshang Nayyeri for providing us with the CANDELS Multiwavelength Catalog in the COSMOS field, and Andrea Grazian and Thibault Garel for sending their results in a convenient digitalised format. I.D. thanks Marta Volonteri, Jeremy Blaizot, Yohan Dubois, Andrea Grazian, Roberto Maiolino for very useful discussions. I.D. and O.I. acknowledge funding of the French Agence Nationale de la Recherche for the SAGACE project. C.L. acknowledges support from a Beecroft fellowship. I.D. acknowledges the European Union's Seventh Framework programme under grant agreement 337595 (ERC Starting Grant, "CoSMass"). A.F. acknowledges support from the Swiss National Science

Foundation. The COSMOS team in France acknowledges support from the Centre National d'Études Spatiales.

## References

- Anders, P., & Alvensleben, U. F. 2003, *A&A*, **401**, 1063
- Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, *MNRAS*, **329**, 355
- Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, *MNRAS*, **426**, 2046
- Arnouts, S., Walcher, C., Le Fèvre, O., et al. 2007, *A&A*, **476**, 137
- Arnouts, S., Le Floc'h, E., Chevallard, J., et al. 2013, *A&A*, **558**, A67
- Ashby, M., Stanford, S., Brodin, M., et al. 2013, *ApJS*, **209**, 22
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, *ApJS*, **218**, 22
- Asplund, M., Grevesse, N., Sauval, A., & Scott, P. 2009, *ARA&A*, **47**, 481
- Baldry, I., Driver, S., Loveday, J., et al. 2012, *MNRAS*, **421**, 621
- Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, *A&A*, **577**, A42
- Baugh, C. 2006, *Rep. Prog. Phys.*, **69**, 3101
- Behroozi, P. S., & Silk, J. 2015, *ApJ*, **799**, 32
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, **770**, 57
- Bell, E. F., & de Jong, R. S. 2001, *ApJ*, **550**, 212
- Berta, S., Lutz, D., Santini, P., et al. 2013, *A&A*, **551**, A100
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in *Astronomical Data Analysis Software and Systems XI*, eds. D. Bohlender, D. Durand, & T. Handley, *ASP Conf. Ser.*, **281**, 228
- Bixler, J. V., Bowyer, S., & Laget, M. 1991, *A&A*, **250**, 370
- Boissier, S., Buat, V., & Ilbert, O. 2010, *A&A*, **522**, A18
- Bolzonella, M., Kovač, K., Pozzetti, L., et al. 2010, *A&A*, **524**, A76
- Bouwens, R. J. 2016, in *Understanding the Epoch of Cosmic Reionization: Challenges and Progress*, ed. A. Mesinger (Springer International Publishing), 111
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, *ApJ*, **737**, 90
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, **803**, 34
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, *MNRAS*, **440**, 2810
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, *MNRAS*, **452**, 1817
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2017, *MNRAS*, **466**, 3612
- Bradley, L. D., Trenti, M., Oesch, P. A., et al. 2012, *ApJ*, **760**, 108
- Brusa, M., Zamorani, G., Comastri, A., et al. 2007, *ApJS*, **172**, 353
- Brusa, M., Fiore, F., Santini, P., et al. 2009, *A&A*, **507**, 1277
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Bundy, K., Ellis, R., Conselice, C., et al. 2006, *ApJ*, **651**, 120
- Bundy, K., Georgakakis, A., Nandra, K., et al. 2008, *ApJ*, **681**, 931
- Calzetti, D., Armus, L., Bohlin, R., et al. 2000, *ApJ*, **533**, 682
- Capak, P., Aussel, H., Bundy, K., et al. 2012, SPLASH: Spitzer Large Area Survey with Hyper-Suprime-Cam, Spitzer Proposal 90042
- Caputi, K. I., Cirasuolo, M., Dunlop, J. S., et al. 2011, *MNRAS*, **413**, 162
- Caputi, K. I., Ilbert, O., Laigle, C., et al. 2015, *ApJ*, **810**, 73
- Casey, C. M., Narayanan, D., & Cooray, A. 2014a, *Phys. Rep.*, **541**, 45
- Casey, C. M., Scoville, N. Z., Sanders, D. B., et al. 2014b, *ApJ*, **796**, 95
- Chabrier, G. 2003, *PASP*, **115**, 763
- Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, *ApJ*, **542**, 464
- Cicone, C., Maiolino, R., Gallerani, S., et al. 2014, *A&A*, **562**, A21
- Cicone, C., Maiolino, R., Gallerani, S., et al. 2015, *A&A*, **574**, A14
- Cimatti, A., Daddi, E., Mignoli, M., et al. 2002, *A&A*, **381**, L68
- Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, *A&A*, **575**, A40
- Conroy, C. 2013, *ARA&A*, **51**, 393
- Conselice, C., Wilkinson, A., Duncan, K., & Mortlock, A. 2016, *ApJ*, **830**, 83
- Costa, T., Sijacki, D., Trenti, M., & Haehnelt, M. 2014, *MNRAS*, **439**, 2146
- Coupon, J. 2015, *MNRAS*, **449**, 1352
- Croton, D. J. 2013, *PASA*, **30**, e052
- Curtis, M., & Sijacki, D. 2016, *MNRAS*, **457**, L34
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, **388**, 1595
- da Cunha, E., Walter, F., Smail, I., et al. 2015, *ApJ*, **806**, 110
- Daddi, E., Cimatti, A., Renzini, A., et al. 2004, *ApJ*, **617**, 746
- Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, **670**, 156
- Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, *ApJ*, **775**, 93
- Davidzon, I., Bolzonella, M., Coupon, J., et al. 2013, *A&A*, **558**, A23
- Davidzon, I., Cucciati, O., Bolzonella, M., et al. 2016, *A&A*, **586**, A23
- Davis, M., Faber, S., Newman, J., et al. 2003, in *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conf.*, ed. P. Guhathakurta, 4834, 161
- de Barros, S., Schaefer, D., & Stark, D. 2014, *A&A*, **563**, A81
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, **368**, 2
- Dekel, A., & Silk, J. 1986, *ApJ*, **303**, 39
- Dubois, Y., Pichon, C., Devriendt, J., et al. 2013, *MNRAS*, **428**, 2885
- Dubois, Y., Pichon, C., Welker, C., et al. 2014, *MNRAS*, **444**, 1453
- Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, *MNRAS*, **444**, 2960
- Eddington, A. 1913, *MNRAS*, **73**, 359
- Efstathiou, G., Ellis, R. S., & Peterson, B. A. 1988, *MNRAS*, **232**, 431
- Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, *A&A*, **468**, 33
- Faber, S. M., Willmer, C. N. A., Wolf, C., et al. 2007, *ApJ*, **665**, 265
- Fabian, A. C. 2012, *ARA&A*, **50**, 455
- Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016a, *ApJ*, **821**, 122
- Faisst, A. L., Capak, P. L., Davidzon, I., et al. 2016b, *ApJ*, **822**, 29
- Feltre, A., Hatziminaoglou, E., Fritz, J., & Franceschini, A. 2012, *MNRAS*, **426**, 120
- Fitzpatrick, E., & Massa, D. 1986, *ApJ*, **307**, 286
- Fontana, A., Pozzetti, L., Donnarumma, I., et al. 2004, *A&A*, **424**, 23
- Fontana, A., Salimbeni, S., Grazian, A., et al. 2006, *A&A*, **459**, 745
- Forrest, B., Tran, K.-V. H., Tomczak, A. R., et al. 2016, *ApJ*, **818**, L26
- Förster Schreiber, N., Genzel, R., Bouché, N., et al. 2009, *ApJ*, **706**, 1364
- Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, *MNRAS*, **366**, 767
- Furlong, M., Bower, R. G., Theuns, T., et al. 2015, *MNRAS*, **450**, 4486
- Gabor, J., Davé, R., Finlator, K., & Oppenheimer, B. 2010, *MNRAS*, **407**, 749
- Galametz, A., Grazian, A., Fontana, A., et al. 2013, *ApJS*, **206**, 10
- Garel, T., Blaizot, J., Guiderdoni, B., et al. 2015, *MNRAS*, **450**, 1279
- Garel, T., Guiderdoni, B., & Blaizot, J. 2016, *MNRAS*, **455**, 3436
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, **445**, 175
- González, V., Labbé, I., Bouwens, R. J., et al. 2011, *ApJ*, **735**, L34
- Grazian, A., Fontana, A., de Santis, C., et al. 2006, *A&A*, **449**, 951
- Grazian, A., Fontana, A., Santini, P., et al. 2015, *A&A*, **575**, A96
- Grogan, N., Kocevski, D., Faber, S., et al. 2011, *ApJS*, **197**, 35
- Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, *MNRAS*, **413**, 101
- Guo, Q., White, S., Angulo, R., et al. 2013, *MNRAS*, **428**, 1351
- Haines, C. P., Iovino, A., Krywult, J., et al. 2017, *A&A*, **605**, A4
- Hainline, K. N., Sharpley, A. E., Greene, J. E., et al. 2012, *ApJ*, **760**, 74
- Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, *ApJ*, **821**, 123
- Henriques, B., White, S., Thomas, P., et al. 2015, *MNRAS*, **451**, 2663
- Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, *ApJ*, **696**, 891
- Hoaglin, D., Mosteller, F., & Tukey, J. 1983, *Understanding robust and exploratory data analysis* (New York: Wiley)
- Hsieh, B.-C., Wang, W.-H., Hsieh, C.-C., et al. 2012, *ApJS*, **203**, 23
- Ilbert, O., Tresse, L., Arnouts, S., et al. 2004, *MNRAS*, **351**, 541
- Ilbert, O., Tresse, L., Zucca, E., et al. 2005, *A&A*, **439**, 863
- Ilbert, O., Arnouts, S., McCracken, H., et al. 2006, *A&A*, **457**, 841
- Ilbert, O., Capak, P., Salvato, M., et al. 2009, *ApJ*, **690**, 1236
- Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, *ApJ*, **709**, 644
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, **556**, A55
- Illingworth, G., Magee, D., Oesch, P., et al. 2013, *ApJS*, **209**, 6
- Kashino, D., Silverman, J. D., Rodighiero, G., et al. 2013, *ApJ*, **777**, L8
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, **341**, 54
- Kaviraj, S., Laigle, C., Kimm, T., et al. 2017, *MNRAS*, **467**, 4739
- Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2014, *MNRAS*, **444**, 1647
- Kennicutt, R. C. 1998, *ARA&A*, **36**, 189
- Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2016, *MNRAS*, **463**, 2363
- Koekemoer, A., Faber, S., Ferguson, H., et al. 2011, *ApJS*, **197**, 36
- Kohonen, T. 1982, *Biological Cybernetics*, **43**, 59
- Krogager, J.-K., Zirm, A. W., Toft, S., Man, A., & Brammer, G. 2014, *ApJ*, **797**, 17
- Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, *ApJ*, **777**, L19
- Laidler, V. G., Papovich, C., Grogan, N. A., et al. 2007, *PASP*, **119**, 1325
- Laigle, C., McCracken, H., Ilbert, O., et al. 2016, *ApJ*, **224**, 24
- Larson, R. B. 1974, *MNRAS*, **169**, 229
- Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, *A&A*, **439**, 845
- Le Fèvre, O., Tasca, L., Cassata, P., et al. 2015, *A&A*, **576**, A79
- Leauthaud, A., Massey, R., Kneib, J., et al. 2007, *ApJS*, **172**, 219
- Lee, K.-S., Ferguson, H. C., Wiklind, T., et al. 2012, *ApJ*, **752**, 66
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, **123**, 3
- Leja, J., van Dokkum, P., Franx, M., & Whitaker, K. E. 2014, *ApJ*, **798**, 115
- Li, C., & White, S. 2009, *MNRAS*, **398**, 2177
- Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, *ApJ*, **772**, 119
- Ma, X., Hopkins, P. F., Faucher-Giguere, C.-A., et al. 2016, *MNRAS*, **456**, 2140
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, **488**, 463
- Maiolino, R., Gallerani, S., Neri, R., et al. 2012, *MNRAS*, **425**, L66
- Mancini, M., Schneider, R., Graziani, L., et al. 2015, *MNRAS*, **451**, L70
- Mancuso, C., Lapi, A., Shi, J., et al. 2016, *ApJ*, **823**, 128
- Maraston, C., Pforr, J., Renzini, A., et al. 2010, *MNRAS*, **407**, 830
- Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, **817**, 34
- Marchesini, D., van Dokkum, P., Förster Schreiber, N., et al. 2009, *ApJ*, **701**, 1765
- Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2015, *ApJ*, **801**, 133
- Marsan, Z. C., Marchesini, D., Bedregal, A. G., et al. 2017, *ApJ*, **842**, 21

- Martis, N. S., Marchesini, D., Brammer, G. B., et al. 2016, [ApJ](#), **827**, L25
- Masters, D., Capak, P., Stern, D., et al. 2015, [ApJ](#), **813**, 53
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, [A&A](#), **544**, A156
- McLure, R. J., Cirasuolo, M., Dunlop, J. S., Foucaud, S., & Almaini, O. 2009, [MNRAS](#), **395**, 2196
- Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, [MNRAS](#), **435**, 87
- Miyazaki, S., Komiyama, Y., Nakaya, H., et al. 2012, in Ground-based and Airborne Instrumentation for Astronomy IV, [Proc. SPIE](#), **8446**, 84460Z
- Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, [ApJ](#), **808**, 101
- Moffett, A. J., Ingarfield, S. A., Driver, S. P., et al. 2016, [MNRAS](#), **457**, 1308
- Morrison, C. B., Hildebrandt, H., Schmidt, S. J., et al. 2017, [MNRAS](#), **467**, 3576
- Mortlock, A., Conselice, C., Bluck, A., et al. 2011, [MNRAS](#), **413**, 2845
- Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2014, [MNRAS](#), **447**, 2
- Moster, B., Somerville, R., Newman, J., & Rix, H.-W. 2011, [ApJ](#), **731**, 113
- Moster, B., Naab, T., & White, S. 2013, [MNRAS](#), **428**, 3121
- Moustakas, J., Kennicutt, Jr., R. C., & Tremonti, C. A. 2006, [ApJ](#), **642**, 775
- Moustakas, J., Coil, A., Aird, J., et al. 2013, [ApJ](#), **767**, 50
- Moutard, T., Arnouts, S., Ilbert, O., et al. 2016a, [A&A](#), **590**, A103
- Moutard, T., Arnouts, S., Ilbert, O., et al. 2016b, [A&A](#), **590**, A102
- Murray, S., Power, C., & Robotham, A. 2013, [Astronomy and Computing](#), **3**, 23
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013a, [ApJ](#), **777**, 18
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013b, [ApJS](#), **206**, 8
- Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, [ApJS](#), **228**, 7
- Neistein, E., & Dekel, A. 2008, [MNRAS](#), **383**, 615
- Noeske, K., Weiner, B., Faber, S., et al. 2007, [ApJ](#), **660**, L43
- Oke, J. 1974, [ApJS](#), **27**, 21
- Ouchi, M., Mobasher, B., Shimasaku, K., et al. 2009, [ApJ](#), **706**, 1136
- Ownsworth, J. R., Conselice, C. J., Mundy, C. J., et al. 2016, [MNRAS](#), **461**, 1112
- Padovani, P., Bonzini, M., Kellermann, K. I., et al. 2015, [MNRAS](#), **452**, 1263
- Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, [ApJ](#), **559**, 620
- Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, [MNRAS](#), **412**, 1123
- Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, [MNRAS](#), **456**, 3194
- Peng, Y.-j., Lilly, S., Kovač, K., et al. 2010, [ApJ](#), **721**, 193
- Pérez-González, P., Rieke, G., Villar, V., et al. 2008, [ApJ](#), **675**, 234
- Pickles, A., & J., A. 1998, [PASP](#), **110**, 863
- Planck Collaboration XIII. 2016, [A&A](#), **594**, A13
- Polletta, M., Tajer, M., Maraschi, L., et al. 2007, [ApJ](#), **663**, 81
- Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, [A&A](#), **523**, A13
- Prévot, M., Lequeux, J., Prevot, L., Maurice, E., & Rocca-Volmerange, B. 1984, [A&A](#), **132**, 389
- Reddy, N., Erb, D., Pettini, M., Steidel, C., & Shapley, A. 2010, [ApJ](#), **712**, 1070
- Reddy, N., Dickinson, M., Elbaz, D., et al. 2012, [ApJ](#), **744**, 154
- Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, [ApJ](#), **802**, L19
- Saintonge, A., Lutz, D., Genzel, R., et al. 2013, [ApJ](#), **778**, 2
- Salim, S., Charlot, S., Rich, R. M., et al. 2005, [ApJ](#), **619**, L39
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, [ApJ](#), **799**, 183
- Salpeter, E. 1955, [ApJ](#), **121**, 161
- Sandage, A., Tammann, G. A., & Yahil, A. 1979, [ApJ](#), **232**, 352
- Sanders, D., Salvato, M., Aussel, H., et al. 2007, [ApJS](#), **172**, 86
- Santini, P., Fontana, A., Grazian, A., et al. 2012, [A&A](#), **538**, A33
- Sawicki, M. 2012, [PASP](#), **124**, 1208
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2016, [MNRAS](#), **446**, 521
- Schechter, P. 1976, [ApJ](#), **203**, 297
- Schmidt, M. 1968, [ApJ](#), **151**, 393
- Schreiber, C., Pannella, M., Leiton, R., et al. 2017, [A&A](#), **599**, A134
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, [ApJS](#), **172**, 1
- Scoville, N., Faisst, A., Capak, P., et al. 2015, [ApJ](#), **800**, 108
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, [MNRAS](#), **323**, 1
- Silverman, J. J., Kashino, D., Sanders, D., et al. 2015, [ApJS](#), **220**, 12
- Simha, V., Weinberg, D. H., Conroy, C., et al. 2014, eprint arXiv [[arXiv:1404.0402](#)]
- Smit, R., Bouwens, R. J., Labbe, I., et al. 2014, [ApJ](#), **784**, 58
- Sommariva, V., Mannucci, F., Cresci, G., et al. 2012, [A&A](#), **539**, A136
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, [ApJ](#), **825**, 5
- Sparre, M., Hayward, C., Springel, V., et al. 2015, [MNRAS](#), **447**, 3548
- Speagle, J., Capak, P., Eisenstein, D., Masters, D., & Steinhardt, C. 2016, [MNRAS](#), **461**, 3432
- Spitler, L. R., Straatman, C. M. S., Labbe, I., et al. 2014, [ApJ](#), **787**, L36
- Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, [ApJ](#), **697**, 1493
- Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, [ApJ](#), **763**, 129
- Stark, D., Richard, J., Siana, B., et al. 2014, [MNRAS](#), **445**, 3200
- Stefanon, M., Marchesini, D., Muzzin, A., et al. 2015, [ApJ](#), **803**, 23
- Stefanon, M., Bouwens, R. J., Labb  , I., et al. 2017, [ApJ](#), **843**, 36
- Steidel, C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. 1996, [ApJ](#), **462**, L17
- Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, [ApJ](#), **791**, L25
- Steinhardt, C. L., Capak, P., Masters, D., & Speagle, J. S. 2016, [ApJ](#), **824**, 21
- Steinhardt, C. L., Yurk, D., & Capak, P. 2017, [MNRAS](#), **468**, 849
- Szalay, A. S., Connolly, A. J., & Szokoly, G. P. 1999, [AJ](#), **117**, 68
- Tacchella, S., Carollo, C. M., Renzini, A., et al. 2015, [Science](#), **348**, 314
- Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, [MNRAS](#), **458**, 242
- Takeuchi, T., Yoshikawa, K., & Ishii, T. 2000, [ApJS](#), **129**, 1
- Tasca, L. A. M., Le F  vre, O., Hathi, N. P., et al. 2015, [A&A](#), **581**, A54
- Thomas, D., Maraston, C., Schawinski, K., Sarzi, M., & Silk, J. 2010, [MNRAS](#), **404**, 1775
- Thomas, R., Le F  vre, O., Le Brun, V., et al. 2017, [A&A](#), **597**, A88
- Tinker, J. L., Kravtsov, A. V., Klypin, A., et al. 2008, [ApJ](#), **688**, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, [ApJ](#), **724**, 878
- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, [ApJ](#), **783**, 85
- Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, [MNRAS](#), **438**, 1985
- Trenti, M., & Stiavelli, M. 2008, [ApJ](#), **676**, 767
- Volonteri, M., Habouzit, M., Pacucci, F., & Tremmel, M. 2015, [MNRAS](#), **449**, 1470
- Volonteri, M., Dubois, Y., Pichon, C., & Devriendt, J. 2016, [MNRAS](#), **460**, 2979
- Vulcani, B., Poggianti, B. M., Dressler, A., et al. 2011, [MNRAS](#), **413**, 921
- Wang, T., Elbaz, D., Schreiber, C., et al. 2016, [ApJ](#), **816**, 84
- Weigel, A. K., Schawinski, K., & Bruderer, C. 2016, [MNRAS](#), **459**, 2150
- Whitaker, K. E., Franx, M., Leja, J., et al. 2014, [ApJ](#), **795**, 104
- Wilkins, S. M., Trentham, N., & Hopkins, A. M. 2008, [MNRAS](#), **385**, 687
- Wilkins, S. M., Coulton, W., Caruana, J., et al. 2013, [MNRAS](#), **435**, 2885
- Wilkins, S. M., Stanway, E. R., & Bremer, M. N. 2014, [MNRAS](#), **439**, 1038
- Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labb  , I. 2009, [ApJ](#), **691**, 1879
- Wuyts, S., F  rster Schreiber, N., Nelson, E., et al. 2013, [ApJ](#), **779**, 135
- Zamojski, M., Schiminovich, D., Rich, R., et al. 2007, [ApJS](#), **172**, 468

<sup>1</sup> Aix-Marseille Univ., CNRS, LAM, Laboratoire d’Astrophysique de Marseille, Marseille, France  
e-mail: iary.davidzon@lam.fr

<sup>2</sup> INAF–Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy

<sup>3</sup> Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

<sup>4</sup> Department of Astronomy, University of Geneva, Ch. d’Ecogia 16, 1290 Versoix, Switzerland

<sup>5</sup> Sorbonne Universit  s, UPMC University Paris 6 and CNRS, UMR 7095, Institut d’Astrophysique de Paris, 98bis bd Arago, 75014 Paris, France

<sup>6</sup> Department of Physics, University of Zagreb, Bijeni  ka cesta 32, 10002 Zagreb, Croatia

<sup>7</sup> Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA

<sup>8</sup> Academia Sinica Institute of Astronomy and Astrophysics, PO Box 23-141, 10617 Taipei, Taiwan, PR China

<sup>9</sup> Centre de Recherche Astrophysique de Lyon, Universit   de Lyon, Universit   Lyon 1, CNRS, Observatoire de Lyon, 9 avenue Charles Andr  , 69561 Saint-Genis Laval Cedex, France

<sup>10</sup> Laboratoire AIM Paris-Saclay, UMR 7158, CEA, CNRS, Universit   Paris VII, CE-SACLAY, B  t. 709, 91191 Gif-sur-Yvette, France

<sup>11</sup> Dark Cosmology Centre, Niels Bohr Institute, Copenhagen University, Juliane Maries Vej 30, 2100 Copenhagen O, Denmark

<sup>12</sup> Institute for Astronomy, 2680 Woodlawn Drive Honolulu, HI 96822-1839, USA

<sup>13</sup> Max-Planck-Institut f  r extraterrestrische Physik, Giessenbachstrasse, 85748 Garching, Germany

<sup>14</sup> Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA

<sup>15</sup> Kavli Institute for the Physics and Mathematics of the Universe (WPI), Todai Institutes for Advanced Study, The University of Tokyo, Kashiwa, 277-8583 Chiba, Japan

## Appendix A: Photometric redshifts in COSMOS2015

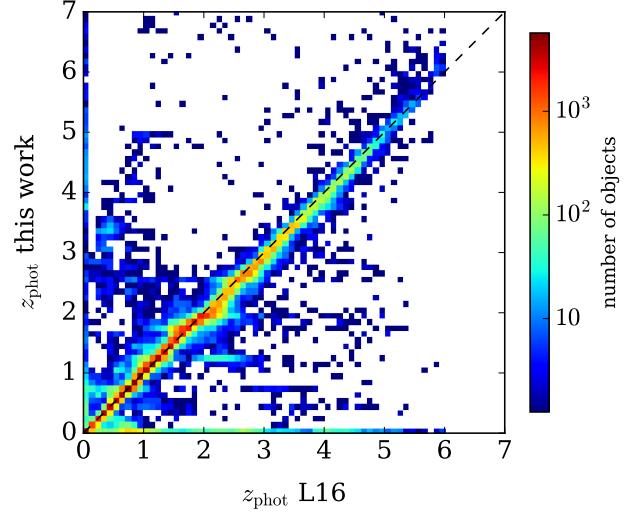
Despite the  $z_{\text{phot}}$  accuracy we reached in L16, in the present work we refined our SED fitting procedure with specific improvement for the high- $z$  analysis (see Sect. 3). It is worth noting that the photometry used here is exactly the same as that published in L16.

Figure A.1 shows the comparison between photometric redshifts computed in L16 ( $z_{\text{phot,L16}}$ ) and the new ones (which we name in the following simply  $z_{\text{phot}}$ ). The comparison includes 125 578 objects brighter than 25 in the IRAC [3.6] channel, from the UltraDeep region of the COSMOS2015 catalogue. Excluding stars, we find that for 68% of them the difference  $\Delta z \equiv |z_{\text{phot,L16}} - z_{\text{phot}}|$  is smaller than 0.05 (and  $\Delta z < 1$  for 99% of the galaxies). In addition to 10 013 photometric objects already classified as stars in L16, we identified a further 11 231 stars in the new SED fitting run (there are also 2116 galaxies that were previously labelled as stars).

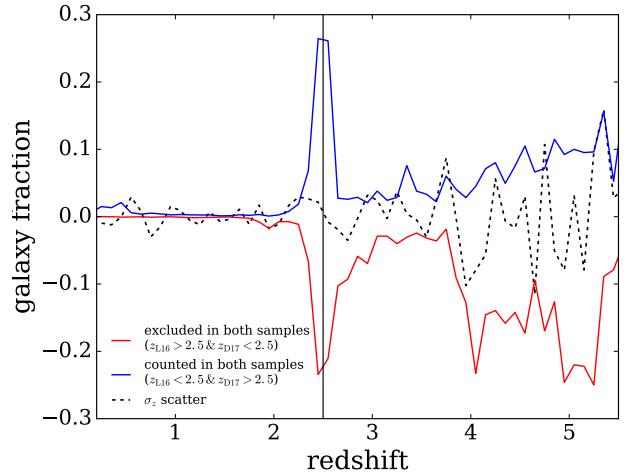
Despite the overall agreement, one can note in Fig. A.1 a subsample of objects that moved from  $z_{\text{phot,L16}} < 1$  to  $z_{\text{phot}} \simeq 3$ . These sources, with the larger  $E(B-V)$  range and the new SFHs we adopted, are now classified as dusty galaxies at high redshift. The  $B$  drop-out in their photometry can be interpreted either as a Balmer break (according to L16) or a strongly attenuated UV slope (in the new computation). Other groups of galaxies that significantly changed their redshift (e.g. from  $z \sim 4$  to  $\sim 1$ ) have no statistical impact on our analysis.

The most interesting galaxies, which also represent a challenge for SED fitting techniques, are those at  $z_{\text{phot}} > 2.5$ , more massive than  $\log(M/M_{\odot}) > 10.5$ , with a strong FIR re-emission ( $> 90 \mu\text{Jy}$  in MIPS  $24 \mu\text{m}$ ,  $S/N > 3$ ). They belong to the subsample of dusty galaxies; more than 80% of them have  $E(B-V) > 0.3$ , and an uncertain photometric redshift as mentioned above (see Spitler et al. 2014; Casey et al. 2014a; Martis et al. 2016, for further insight into massive passive galaxies at high  $z$ ). Although there are only 124 of them in the UltraDeep region, lying in the exponential tail of the SMF they can produce a non-negligible offset if their  $z_{\text{phot}}$  is wrong. We emphasise that their  $z_{\text{phot}}$  distribution agrees with that of the  $24 \mu\text{m}$  emitters studied by Wang et al. (2016) in CANDELS (i.e. in both cases there is a sharp drop of detections at  $z \sim 3$ ). For twelve of these “MIPS-bright” sources, for which we have also a spectroscopic measurement,  $z_{\text{phot}}$  is within  $\pm 0.1(1+z_{\text{spec}})$ . Most of the MIPS-bright galaxies are fit by templates with  $SFR > 100 M_{\odot} \text{ yr}^{-1}$ , although in reality their IR emission may be caused not only by heated dust, but also (at least partly) by an active galactic nucleus (e.g. Casey et al. 2014b; Marsan et al. 2015, 2017).

Figure A.2 summarises the changes described above, showing as a function of  $z$  the fraction of galaxies either excluded from the SMF computation (i.e.  $z_{\text{phot,L16}} > 2.5$  and  $z_{\text{phot}} < 2.5$ ) or counted twice ( $z_{\text{phot,L16}} < 2.5$  and  $z_{\text{phot}} > 2.5$ ). These galaxies have a negligible impact at  $z < 2$ . At  $z \simeq 2.5$ , where we join the two samples, we observe the scatter due to random errors (i.e.  $z_{\text{phot,L16}}$  and  $z_{\text{phot}}$  are compatible within  $1\sigma_z$ ); the number of galaxies totally neglected and the number of “duplicated” galaxies balance out in this  $z$ -bin. At  $z \gtrsim 3$  the fraction of low- $z$  galaxies (according to L16) that have been relocated at  $z > 2.5$  in our analysis is 5–10%. There is a larger number of objects with  $z_{\text{phot,L16}} > 2.5$  that we ruled out as interlopers from our high- $z$  sample, as discussed above. In Fig. A.2 we also plot an estimate of the scatter due to  $\sigma_z$ , obtained recomputing  $N(z)$  from our Monte Carlo simulation (Sect. 4.1). This comparison shows that changes related to the replacement of  $z_{\text{phot,L16}}$  with the new



**Fig. A.1.** Official photometric redshifts ( $z_{\text{phot,L16}}$ , Laigle et al. 2016) of COSMOS2015 compared to the new estimates from the present work. The sources have  $[3.6] < 25$  and are selected in the UltraDeep area. Red-orange pixels include 90% of the objects (whose total number is 125 578). A dashed line is a reference for the 1:1 correspondence.



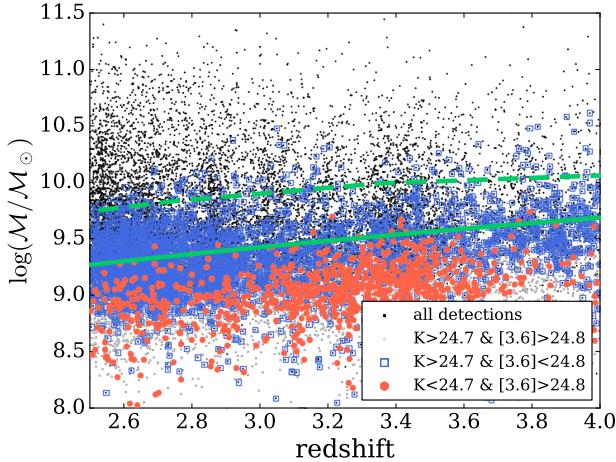
**Fig. A.2.** Fraction of galaxies neglected in the SMF computation (red line, negative values) and those included both in the L16 sample at low  $z$  and in the revised computation at  $z_{\text{phot}} > 2.5$  (blue line, positive values). A vertical solid line separates the two samples. Variations due to  $\sigma_z$  are shown for comparison (dotted line).

estimates are of the same order of magnitude of typical fluctuations due to  $z_{\text{phot}}$  statistical errors.

## Appendix B: Implications of a multi-band detection technique

To compute the SMF, first we have to assess the minimum stellar mass ( $M_{\text{lim}}$ ) below which the sample incompleteness can impair our measurement. We mentioned in Sect. 4.2 two possibilities that are feasible in COSMOS2015: a mass complete sample derived from a  $K_s$ -band selection, or one derived after a cut in [3.6]. A selection in IRAC, rather than in a single NIR band, is motivated by the fact that our detections are made by combining images from multiple bands, some of them deeper than  $K_s$ .

We can directly compare to Stefanon et al. (2015), who work with the UltraVISTA dataset of Muzzin et al. (2013b), with the same limiting magnitude of COSMOS2015, but in the shallower

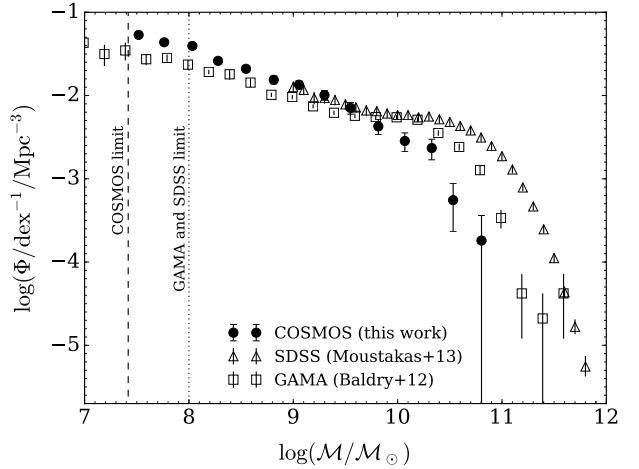


**Fig. B.1.** Galaxy stellar mass as a function of redshift, in a range where a cut in either  $K_s$  or [3.6] is applicable in order to build a flux limited sample. In the plot, small circles are all the COSMOS2015 galaxies in  $\mathcal{A}_{UD}$ , those detected in both  $K_s$  and [3.6] (black dots) and the ones that are not (grey dots). Red filled circle (blue empty squares) show galaxies fainter than the [3.6] ( $K_s$ ) magnitude limit, but detected in  $K_s$  ([3.6]). Green lines represent the stellar mass completeness limits  $M_{lim}(z)$  (see Sect. 4.2) resulting from either the [3.6]-selected sample (solid lines) or the  $K_s$ -selected (dashed line).

$\mathcal{A}_D$  region. Their source extraction is similar to ours although they used  $K_s$  only as a prior image. By the inspection of the IRAC residual maps, they find that their  $z_{phot} > 4$  sample increases by up to 38% (depending on the method used to estimate photometric redshifts). Among the 408 sources they recover from the residual maps (in the whole UltraVISTA area) 48% are naturally included in COSMOS2015 (M. Stefanon, priv. comm.). Caputi et al. (2015) make another test by comparing the IRAC sources detected in UltraVISTA DR1 to those from DR2 in the UltraDeep stripes, whose depth increased by 0.7 mag. Their work shows the difference between  $\mathcal{A}_D$  and  $\mathcal{A}_{UD}$  regions. Caputi et al. find 574 IR-bright ( $[4.5] < 23$ ) sources detected in DR2 but not in DR1 (i.e. with  $K_s > 24$ ). We note that about 75% of them are detected in  $z^{++}$ , confirming the convenience of adding this band in the  $\chi^2$ -stacked image not to suffer from such an incompleteness (see L16).

As already emphasised, our flux-limited sample with a cut at  $[3.6] < 25$  results in a lower  $M_{lim}(z)$  with respect to  $K_s < 24.7$ . In the  $zYJHK_s$  image we detect 17 319 galaxies between  $z = 2.5$  and 4 within  $\mathcal{A}_{UD}$ . Among them, 954 (9%) have  $K_s < 24.7$  but are not in the IRAC selected sample, being fainter than  $[3.6]_{lim,UD}$ . On the contrary, more than double (2182 objects) would be excluded by the cut in  $K_s$ , despite their  $[3.6] < [3.6]_{lim,UD}$ . Besides the percentage of missing objects, we stress that galaxies faint in IRAC are on average less massive than  $10^{9.4} M_\odot$ , while a selection in  $K_s$  would remove more massive objects (see Fig. B.1).

Why are the two selections so different in terms of stellar mass completeness? First, the  $(K_s - [3.6])$  colour of low- and intermediate-mass galaxies at  $2.5 < z < 4$  is on average red (median  $K_s - [3.6] = 0.2$ ) and tends to be redder moving to higher masses. For several galaxies the difference is sufficient to include them among the [3.6] detections but not in  $K_s$ . Another reason is that in such a faint regime the average  $S/N$  of  $K_s$  is lower than [3.6], also because flux extraction in the latter is improved by the  $\chi^2$ -stacking strategy. Therefore the  $K_s$  measurement for low-mass galaxies is more scattered, with a higher chance of being far below the threshold we imposed.



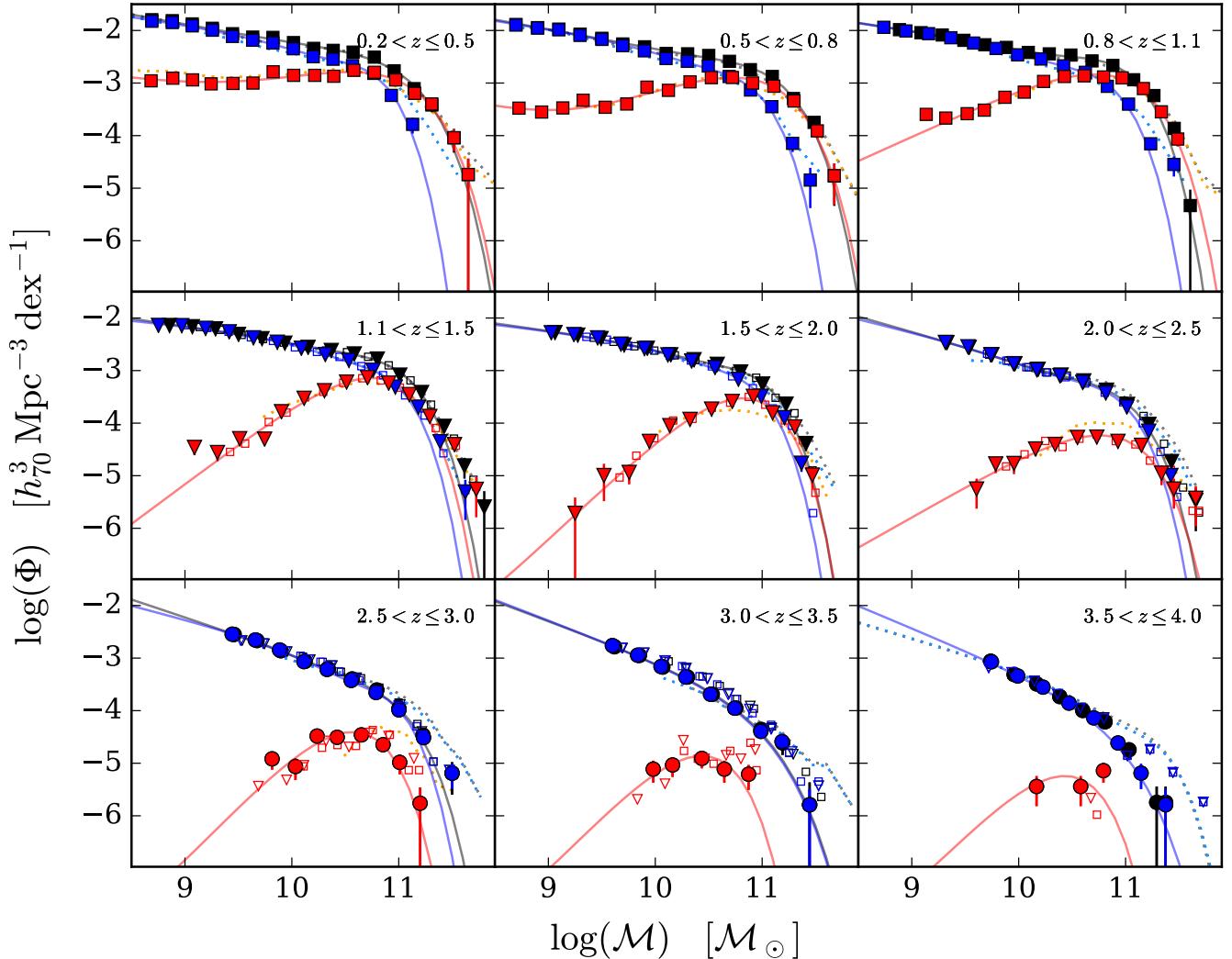
**Fig. C.1.** The galaxy SMF of COSMOS2015 (filled circles), SDSS (Moustakas et al. 2013,  $0 < z < 0.2$ , empty triangles), GAMA (Baldry et al. 2012,  $0 < z < 0.06$ , empty squares). Vertical lines show the stellar mass limit of each survey. A  $10^8 M_\odot$  limit is shown also for the SDSS survey, considering the lowest value (Li & White 2009) among the ones used in various studies. Other authors (like Moustakas et al.) adopt a more conservative threshold for their estimates ( $\sim 10^9 M_\odot$ ).

## Appendix C: COSMOS2015 stellar mass function at $0 < z < 4$

We show in this Appendix the SMF of galaxies at  $z < 3$  derived from the L16 SED fitting estimates, along to the new results at higher redshifts (see Sect. 5). The evolution of these SMFs has already been shown in Figs. 15 and 16. Here we plot the SMF in each  $z$ -bin separately.

The SMF at  $0 < z < 0.15$  (median redshift  $\langle z \rangle = 0.12$ ) is presented in Fig. C.1, together with the SMF of the Sloan Digital Sky Survey (SDSS, Moustakas et al. 2013,  $\langle z \rangle = 0.1$ ) and the Galaxy And Mass Assembly survey (GAMA, Baldry et al. 2012,  $z < 0.06$ ). Our survey probes a small volume in the local universe, so the high-mass end is highly incomplete (also considering the bias due to saturated sources). On the other hand the deeper exposure of COSMOS allows us to probe the SMF at lower masses than SDSS and GAMA: While the sample of Baldry et al. (2012) is complete above  $10^8 M_\odot$  (the limit is the same for SDSS, see Li & White 2009) our SMF extends down to  $\sim 2.5 \times 10^7 M_\odot$ . Fitting a Schechter function to the SMF we find that the low-mass end has a slope  $\alpha = -1.47 \pm 0.02$  (we note that Baldry et al. 2012, find  $-1.47 \pm 0.05$ ).

In Fig. C.2 we show the SMF of both  $K_s$ - and [3.6]-selected galaxies from  $z = 0.2$  to 4. The various estimates are overall in good agreement. In particular at  $2.5 < z \leq 3$  the L16 estimate joins well with the SMF derived from the new SED fitting run. Moving to higher redshifts there are small discrepancies (always  $\leq 0.2$  dex) in the comoving number density. The overestimate (observed mainly in the high-mass end) of the L16 galaxy SMF is likely due to the systematics (interlopers, different set of template) discussed in Appendix B. However, it is difficult to pinpoint a single cause for such a difference as several effects may act in combination. For example age-metallicity degeneracy, after removing subsolar templates, forces one to choose younger galaxy ages, underestimating the  $M/L$  ratio (see Bell & de Jong 2001); however the offset goes in the opposite direction when dust attenuation adds further degeneracy (see Davidzon et al. 2013).



**Fig. C.2.** Three different  $1/V_{\max}$  estimates of the COSMOS2015 galaxy SMF, from  $z = 0.2$  to 4. The estimates are derived from: A  $K_s$ -selected sample using the official SED fitting from L16 (squares); a [3.6]-selected sample that also relies on  $z_{\text{phot,L16}}$  and  $M$  of L16 (triangles); the high- $z$  [3.6]-selected sample presented in this paper (circles). Active and passive galaxies, classified in Sect. 4.3 by means of  $(NUV - r)$  and  $(r - J)$  colours, are shown with blue and red symbols, respectively. The SMFs used in Sect. 6 to discuss  $\sim 10$  billion years of galaxy evolution are shown with filled symbols, while smaller empty symbols are used in the  $z$ -bins where the reference SMF changed. In each bin, Schechter functions (solid lines, same colours for total, active, and passive galaxies) fit the data points of the reference sample (i.e. the filled symbols). We also plot the Schechter functions fitting the SMF of UltraVISTA DR1 galaxies (Ilbert et al. 2013) with dotted lines. All the fits are corrected for the Eddington bias.

In Fig. C.2 we also show the SMF of the UltraVISTA DR1 galaxies (Ilbert et al. 2013). Their estimates are in good agreement with COSMOS2015 up to  $z = 2$ . At higher redshifts we observe the same systematic effect pointed out in Faisst et al. (2016b): without the SPLASH coverage, many high- $z$  objects in

UltraVISTA DR1 were not provided with an accurate MIR photometry (if not at all), and their mass was underestimated (see Fig. 4 of Faisst et al. 2016b).