



# The Rest-frame Optical (900 nm) Galaxy Luminosity Function at $z \sim 4\text{--}7$ : Abundance Matching Points to Limited Evolution in the $M_{\text{STAR}}/M_{\text{HALO}}$ Ratio at $z \geq 4$

Mauro Stefanon<sup>1</sup>, Rychard J. Bouwens<sup>1</sup>, Ivo Labbé<sup>1</sup>, Adam Muzzin<sup>2</sup>, Danilo Marchesini<sup>3</sup>, Pascal Oesch<sup>4</sup>, and Valentino Gonzalez<sup>5</sup>

<sup>1</sup>Leiden Observatory, Leiden University, NL-2300 RA Leiden, Netherlands; [stefanon@strw.leidenuniv.nl](mailto:stefanon@strw.leidenuniv.nl)

<sup>2</sup>Department of Physics and Astronomy, York University, 4700 Keele St., Toronto, Ontario, M3J 1P3, Canada

<sup>3</sup>Physics and Astronomy Department, Tufts University, Robinson Hall, Room 257, Medford, MA 02155, USA

<sup>4</sup>Observatoire de Genève, 1290 Versoix, Switzerland

<sup>5</sup>Departamento de Astronomía, Universidad de Chile, Santiago, Chile

Received 2016 November 24; revised 2017 March 26; accepted 2017 May 10; published 2017 June 28

## Abstract

We present the first determination of the galaxy luminosity function (LF) at  $z \sim 4, 5, 6,$  and  $7$ , in the rest-frame optical at  $\lambda_{\text{rest}} \sim 900$  nm ( $z'$  band). The rest-frame optical light traces the content in low-mass evolved stars ( $\sim$ stellar mass— $M_*$ ), minimizing potential measurement biases for  $M_*$ . Moreover, it is less affected by nebular line emission contamination and dust attenuation, is independent of stellar population models, and can be probed up to  $z \sim 8$  through *Spitzer*/IRAC. Our analysis leverages the unique full-depth *Spitzer*/IRAC 3.6–8.0  $\mu\text{m}$  data over the CANDELS/GOODS-N, CANDELS/GOODS-S, and COSMOS/UltraVISTA fields. We find that, at absolute magnitudes where  $M_{z'}$  is fainter than  $\gtrsim -23$  mag,  $M_{z'}$  linearly correlates with  $M_{\text{UV},1600}$ . At brighter  $M_{z'}$ ,  $M_{\text{UV},1600}$  presents a turnover, suggesting that the stellar mass-to-light ratio  $M_*/L_{\text{UV},1600}$  could be characterized by a very broad range of values at high stellar masses. Median-stacking analyses recover an  $M_*/L_{z'}$  roughly independent on  $M_{z'}$  for  $M_{z'} \gtrsim -23$  mag, but exponentially increasing at brighter magnitudes. We find that the evolution of the LF marginally prefers a pure luminosity evolution over a pure density evolution, with the characteristic luminosity decreasing by a factor of  $\sim 5\times$  between  $z \sim 4$  and  $z \sim 7$ . Direct application of the recovered  $M_*/L_{z'}$  generates stellar mass functions consistent with average measurements from the literature. Measurements of the stellar-to-halo mass ratio at fixed cumulative number density show that it is roughly constant with redshift for  $M_h \gtrsim 10^{12} M_\odot$ . This is also supported by the fact that the evolution of the LF at  $4 \lesssim z \lesssim 7$  can be accounted for by a rigid displacement in luminosity, corresponding to the evolution of the halo mass from abundance matching.

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

## 1. Introduction

Stellar mass is one of the most fundamental parameters characterizing galaxies. This observable is driven by light emitted in the rest-frame optical/near-infrared (NIR) by lower-mass stars, and it correlates with the dynamical mass up to  $z \sim 2$  (e.g., Cappellari et al. 2009; Taylor et al. 2010; van de Sande et al. 2013; Barro et al. 2014), suggesting that it can be a robust estimate of the cumulative content of matter in galaxies. Stellar masses have been estimated for galaxies at redshifts as high as  $z \sim 7\text{--}8$  (e.g., Labbé et al. 2010a, 2013). Moreover, stellar mass estimates are readily available in the models of galaxy formation and evolution. For the above reasons, the stellar mass has been largely adopted in comparisons to the models.

The stellar mass function (SMF), i.e., the number density of galaxies per unit (log) stellar mass, provides a first census of a galaxy population, and it is therefore one of the most basic observables that need to be reproduced by any successful model of galaxy formation. Multi-wavelength photometric surveys like UltraVISTA (McCracken et al. 2012) and ZFOURGE (Straatman et al. 2016) have enabled SMF measurements up to  $z \sim 3\text{--}4$  (e.g., Ilbert et al. 2013; Muzzin et al. 2013b; Tomczak et al. 2014, but see also Stefanon et al. 2015; Caputi et al. 2015 for SMF of massive galaxies up to  $z \sim 7$ ), whereas *HST* surveys, like CANDELS (Grogin et al. 2011; Koekemoer et al. 2011), GOODS (Giavalisco et al. 2004), and HUDF (Beckwith et al. 2006; Bouwens

et al. 2011), complemented by *Spitzer*/IRAC observations (e.g., *Spitzer*/GOODS—PI Dickinson; Ashby et al. 2015), extended the study to the evolution of the SMF up to  $z \sim 7$  (e.g., Stark et al. 2009; González et al. 2011; Duncan et al. 2014; Grazian et al. 2015; Song et al. 2016).

Estimates of stellar mass, however, critically depend on quantities like the initial mass function (IMF), dust content, metallicity, and star-formation history (SFH) of each galaxy (see, e.g., Conroy & Wechsler 2009; Marchesini et al. 2009; Behroozi et al. 2010; Dunlop et al. 2012). Indeed, different sets of SED models characterized by, e.g., a different treatment of the TP-AGB phase, have been shown to potentially introduce systematics in stellar mass measurements as large as few decimal dex (Muzzin et al. 2009; Marchesini et al. 2010; Pforr et al. 2012; Mitchell et al. 2013; Grazian et al. 2015). Similarly, different dust laws (Cardelli et al. 1989; Calzetti et al. 2000; Charlot & Fall 2000) and SFHs can increase the systematics by up to  $\sim 0.2$  dex (Pérez-González et al. 2008). Furthermore, emission from nebular lines has recently been found to potentially bias stellar mass measurements (Schaerer & de Barros 2009; Stark et al. 2009). Stellar masses of high-redshift galaxies ( $z \gtrsim 4$ ) are particularly sensitive to contamination by nebular lines, because high- $z$  star-forming galaxies are likely to be characterized by emission lines with equivalent width in excess of  $\sim 1000$  Å (see e.g., Labbé et al. 2013; de Barros et al. 2014; Smit et al. 2014). Nonetheless, attempts to correct for this contamination can lead to results differing by factors of

a few (e.g., Stark et al. 2009; Labbé et al. 2010b; González et al. 2012; Stefanon et al. 2015).

Alternatively, one could directly study the rest-frame optical/NIR light emitted by the low-mass stars in galaxies, given its connection to stellar mass. Measurements of the optical/NIR luminosity have at least three major advantages over measurements of the stellar mass: (1) they are much less sensitive to assumptions about dust modeling; (2) estimates of luminosity are robust quantities, because luminosity can be recovered directly from the observed flux, with marginal-to-null dependence on the best-fitting SED templates, and hence on, e.g., SFH or the IMF; and (3) a careful choice of the rest-frame band reduces the contamination by nebular emission, minimizing the requirement of corrections to the fluxes (for instance, nebular emission could contribute up to 50% of the rest-frame  $R$ -band luminosity for galaxies at  $z \sim 8$ —Wilkins et al. 2017).

The wavelength range spanned by the rest-frame  $H$  and  $K_s$  bands is most sensitive to lower-mass stars. Light in bands redder than these is potentially contaminated by emission from the dust torus of AGNs, whereas light in bluer bands retains information about the recent SFH. The availability of data up to  $\lambda_{\text{obs}} \sim 8 \mu\text{m}$  from *Spitzer*/IRAC has enabled study of the evolution of the LF in rest-frame NIR bands ( $J$ ,  $H$ ,  $K_s$ ) up to  $z \sim 4$  (e.g., Cirasuolo et al. 2010; Stefanon & Marchesini 2013; Mortlock et al. 2017).

At higher redshifts, the choice of the rest-frame band that more closely correlates with the stellar mass must be a trade-off between including contamination from the recent star formation and performing the luminosity measurements on observational data, rather than relying on the extrapolation of SED templates.

In this context, the  $z'$  band ( $\lambda_{\text{eff}} \sim 0.9 \mu\text{m}$ ) emerges as a natural choice: it lays in the wavelength regime redder than the Balmer/4000 Å break, is free from contamination by strong nebular emission, and can be probed up to  $z \sim 8$  thanks to the *Spitzer*/IRAC 8.0  $\mu\text{m}$ -band data.

Recently, Labbé et al. (2015) assembled the first full-depth IRAC mosaics over the GOODS and UDF fields, combining IRAC observations from the IGOODS (PI: Oesch) and IRAC Ultra Deep Field (IUDF) (PI: Labbé) programs with all the available archival data over the two fields (GOODS, ERS, S-CANDELS, SEDS, and UDF2). Following the same procedure implemented by Labbé et al. (2015), full-depth IRAC mosaics have also now been generated for the GOODS-N and COSMOS/UltraVISTA fields. The GOODS-N mosaics double the area with the deepest IRAC imaging available, whereas the COSMOS/UltraVISTA data, shallower but covering a much larger field, are necessary to include the most massive galaxies at  $z \sim 4$ . The unique depth of IRAC 5.8 and 8.0  $\mu\text{m}$  mosaics in the GOODS fields (PID 194; PI Dickinson)—reaching  $\sim 24.5$  AB ( $5\sigma$ , 2''0 aperture diameter) allowed us to recover flux measurements with signal-to-noise ratio (S/N)  $\gtrsim 4$  in these two bands for galaxies at  $z \sim 7$ –8.

In this work, we leverage these characteristics to measure the evolution of the LF at  $4 \lesssim z \lesssim 7$  in the rest-frame  $z'$  band, providing a complementary approach to the determination of the evolution of the SMF at high redshift. Furthermore, we will show that the SMF can be recovered by applying a stellar mass-to-light ratio ( $M_*/L$ ) to the  $z'$ -band LF. Remarkably, a simple abundance matching reveals that the  $z'$ -band LF can also trace the halo mass function (HMF) and its evolution over  $4 \lesssim z \lesssim 7$ .

Our analysis is based on the photometric catalog of Bouwens et al. (2015), taken over the GOODS-N and GOODS-S fields. At  $z \sim 4$ , we complement our sample with a 37-bands 0.135–8.0  $\mu\text{m}$  photometric catalog, based on the second release (DR2) of the UltraVISTA Survey. The area covered by the DR2 data,  $\sim 0.75$  degrees<sup>2</sup>, is a factor of  $\sim 10\times$  larger than the cumulative area from the two GOODS fields ( $\sim 260$  arcmin<sup>2</sup>), enabling the recovery of the bright end of the  $z \sim 4$  LF with higher statistical significance.

This paper is organized as follows. In Section 2, we describe the adopted data sets and sample selection criteria. In Section 3, we present our results. More specifically, Section 3.2 presents the stellar mass-to-light ratios from stacking, and our LF and SMF measurements are presented in Sections 3.3 and 3.4, respectively. We discuss our LFs measurements with respect to the HMF in Section 4, and conclude in Section 5. Throughout this work, we have adopted a cosmological model with  $H_0 = 70 \text{ Km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ . All magnitudes refer to the AB system. We assumed a Chabrier (2003) IMF, unless otherwise noted.

## 2. The Sample of $4 < z < 7$ Galaxies

### 2.1. Data Sets

Our LF measurements are based on a composite sample of galaxies at  $4 \lesssim z \lesssim 7$ , selected in the rest-frame optical ( $z'$  band,  $\lambda_{\text{eff}} \sim 0.9 \mu\text{m}$ ; see Figure 1 for its transmission efficiency), with the bulk of our sample formed by Lyman break galaxies (LBGs) from the CANDELS/GOODS-N, CANDELS/GOODS-S—ERS fields. The  $z \sim 4$  bin is complemented by a sample of galaxies from a catalog based on the UltraVISTA DR2 mosaics.

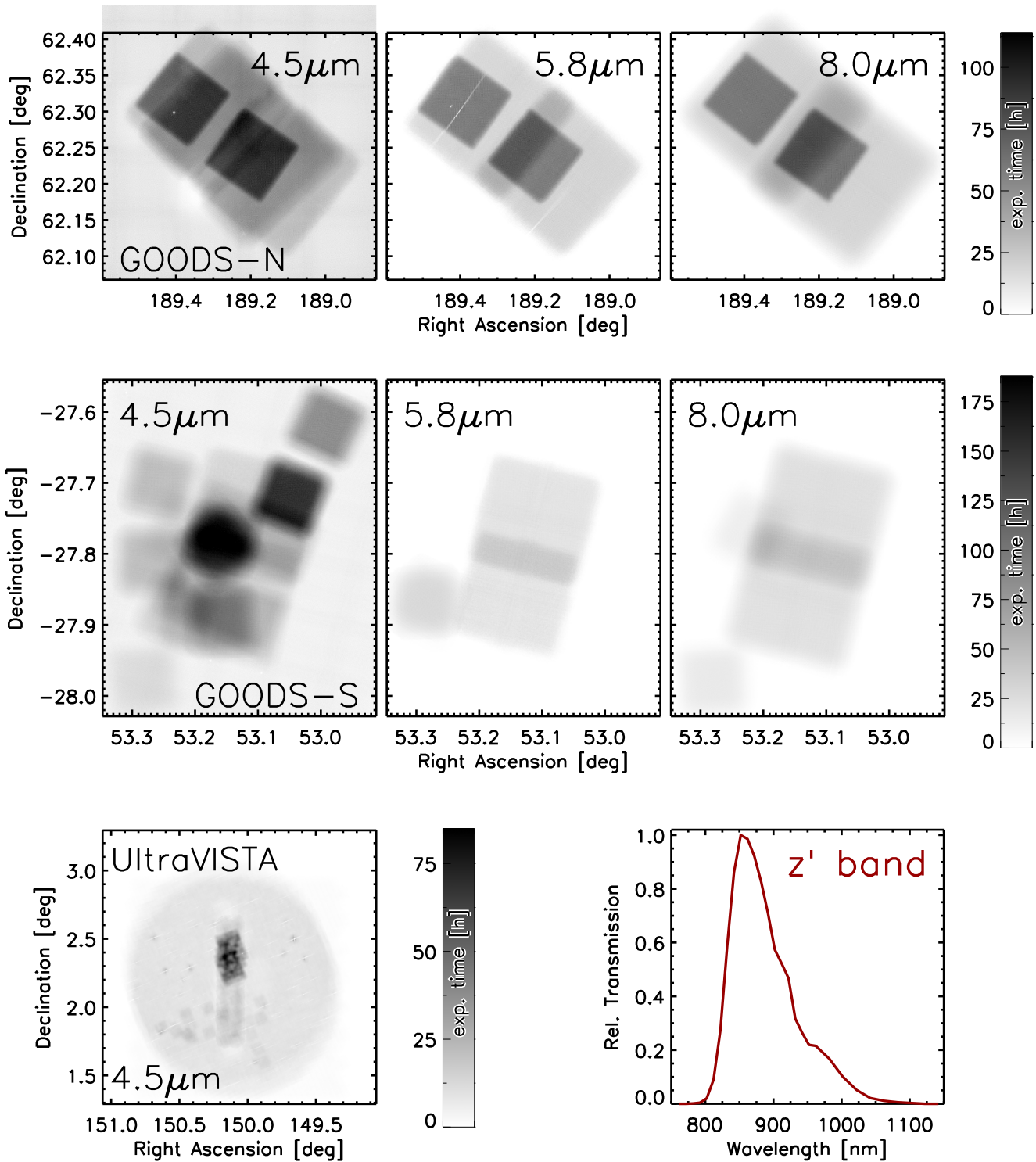
The LBG samples in the CANDELS GOODS-N/S and ERS fields rely on the multi-wavelength photometric catalogs of Bouwens et al. (2015). These are based on the re-reduction of public *HST* imaging, and are enhanced by proprietary full-depth *Spitzer*/IRAC mosaics. Specifically, they benefit from novel full depth IRAC 5.8  $\mu\text{m}$  and 8.0  $\mu\text{m}$  mosaics, not available in the original catalog of Bouwens et al. (2015). The UltraVISTA DR2 catalog is based on the most recent publicly available mosaics at UV-to-NIR wavelengths (including UltraVISTA DR2 data sets), and complemented by an internal release of full-depth *Spitzer*/IRAC mosaics.

In the following sections, we briefly describe these two parent catalogs and detail the criteria we adopted to assemble our final sample of galaxies.

#### 2.1.1. GOODS-N/S and ERS

For this work, we adopted the catalog assembled by Bouwens et al. (2015) over the CANDELS/GOODS-N, CANDELS/GOODS-S, and ERS fields. Here, we briefly summarize the main features; the reader should refer to Sections 2 and 3 of Bouwens et al. (2015) for full details.

The catalog contains the photometry in the *HST* ACS F435W, F606W, F775W, and F850LP bands (hereafter indicated by  $B_{435}$ ,  $V_{606}$ ,  $i_{775}$  and  $z_{850}$ ), together with *HST* WFC3 F105W, F125W, and F160W ( $Y_{105}$ ,  $J_{125}$ ,  $H_{160}$ ) data from CANDELS (Grogin et al. 2011), and WFC3 F140W band ( $JH_{140}$ ) from the 3D-*HST* (Brammer et al. 2012; Skelton et al. 2014) and AGHAST (Weiner & AGHAST Team 2014—<http://mingus.as.arizona.edu/~bjw/aghast/>).



**Figure 1.** Exposure maps of the full-depth IRAC mosaics used in this work for measurements of the  $z'$  band absolute magnitude. The maps are shown with inverted gray scale, and maintain the same scaling stretch across all panels in each row, to highlight the relative exposure times; the amount of exposure time is indicated by the vertical bar on the right. Each row refers to a different field: GOODS-N, GOODS-S/ERS, and UltraVISTA, top to bottom, respectively, labeled in the left-most panel. Left to right, the panels present the 4.5, 5.8, and 8.0  $\mu\text{m}$  mosaics. For the UltraVISTA field, we only show the 4.5  $\mu\text{m}$  mosaic, as we use this data set only at  $z \sim 4$ . The lower-right panel presents the filter transmission curve for the  $z'$  band adopted in this work.

The catalog takes also advantage of full-depth mosaics in the four *Spitzer* IRAC bands. The 3.6  $\mu\text{m}$  and 4.5  $\mu\text{m}$  mosaics were assembled by combining data from the IGOODS (PI: Oesch) and IUDF (PI: Labbé) programs with all the public archival data from either cryogenic or post-cryogenic programs over the GOODS-N and GOODS-S (GOODS, ERS, S-CANDELS,

SEDS, and UDF2). For the 5.8 and 8.0  $\mu\text{m}$  mosaics, however, only observations from the GOODS cryogenic program are available (PI: Dickinson, PID: 194). The mosaics were regenerated from the AORs, using the same procedure as Labbé et al. (2015). This procedure delivers the most accurate reconstruction of the point-spread function (PSF) at any



position across each mosaic, enabling a more accurate flux measurement in the IRAC bands (see below). The mosaics in the 4.5, 5.8, and 8.0  $\mu\text{m}$  bands are key to this work, as they probe the rest-frame  $z'$  band. Specifically, the 4.5  $\mu\text{m}$  band matches the rest-frame  $z'$  band at  $z \sim 4$ , whereas the 5.8 and 8.0  $\mu\text{m}$  bands are required for the rest-frame  $z'$  band at  $z \gtrsim 5$ .

Figure 1 presents the exposure time maps in the IRAC 4.5, 5.8, and 8.0  $\mu\text{m}$  for the GOODS-N and GOODS-S fields. As a result of the combination of data from different programs, the achieved depth across each field is highly inhomogeneous. This is particularly evident for the 4.5  $\mu\text{m}$  band, the depth of which ranges between 50 and 180 hr (corresponding to 25.1–25.8 AB, respectively, for  $5\sigma$ ,  $2''0$  aperture diameter). The GOODS-N field is characterized by the deepest 5.8 and 8.0  $\mu\text{m}$  data, reaching a depth of  $\sim 80$  hr ( $\sim 24.5$  AB,  $5\sigma$ ,  $2''0$  aperture).

The object detection was performed on the  $\chi^2$  image (Szalay et al. 1999) constructed from the  $Y_{105}$ ,  $J_{125}$ ,  $H_{160}$  band images. The detection mosaics have footprints of  $\sim 124$  and  $\sim 140$  arcmin<sup>2</sup>, respectively, for GOODS-N and GOODS-S, for a total of 264 arcmin<sup>2</sup>. Aperture photometry in the *HST* bands was performed in dual mode with SExtractor (Bertin & Arnouts 1996) on the mosaics matching the resolution of the  $H_{160}$  image. Fluxes were converted to total through the application of an aperture correction based on the Kron (1980) scalable apertures, and further corrected to take into account the flux losses of the scalable apertures compared to the PSF. Photometry of the IRAC mosaics was performed using a proprietary deblending code (Labbé et al. 2006, 2010a, 2010b, 2013). This code convolves the high-resolution *HST* mosaics with a kernel obtained from the highest S/N IRAC PSFs, to construct a model of the IRAC image. For each object,  $2''0$ -diameter aperture photometry is performed on the image, which was previously cleaned from neighbors by using the information from the model image. The aperture fluxes were then corrected to total by using the *HST* template specific of each source, convolved to match the *Spitzer* IRAC PSF.

Candidate LBGs at  $z \sim 4, 5, 6,$  and  $7$  were selected from among the  $B_{435}$ ,  $V_{606}$ ,  $i_{775}$  and  $z_{850}$  dropouts, respectively. For a complete list of criteria adopted to select each sample, see Table 2 of Bouwens et al. (2015). The sample included 8031 LBGs.

### 2.1.2. UltraVISTA DR2

For the sample of  $z \sim 4$  galaxies, we also considered detections in the COSMOS/UltraVISTA field, the relatively larger field of which (compared to GOODS-N/S) allowed us to probe higher luminosities.

The UltraVISTA catalog used for this work is based on the *ultradeep* stripes of the second data release (DR2) of the UltraVISTA survey (McCracken et al. 2012). This release is characterized by  $5\sigma$  depth of  $\sim 25.6, 25.1, 24.8, 24.8$  AB ( $2''0$  aperture diameter) in  $Y, J, H,$  and  $K_s$ , respectively ( $\sim 0.8$ – $1.2$  mag deeper than DR1), and extends over an area of  $\sim 0.75$  square degrees in four stripes over the COSMOS field (Scoville et al. 2007). The 37-band catalog was constructed following the same procedure presented in Muzzin et al. (2013a) for the DR1. Briefly, the detection was performed in the  $K_s$  band; 33-band far-UV-to- $K_s$  aperture fluxes were measured with SExtractor (Bertin & Arnouts 1996) in dual mode, focused on the mosaics matching the PSF resolution of the  $H$ -band image. An aperture correction recovered from the Kron ellipsoid was applied on a per-object basis; total fluxes were finally computed by applying a further

aperture correction obtained from the PSF curve of growth. This new catalog also includes flux measurements in the Subaru narrow bands NB711, NB816, the UltraVISTA narrow band NB118, as well as the CFHTLS  $u^*, g', r', i',$  and  $z'$ , which are not available in the DR1 catalog of Muzzin et al. (2013a).

The COSMOS field benefits from several hundreds of hours of integration time with *Spitzer* IRAC. Similarly to what was done for the GOODS-N/S fields, full-depth mosaics were constructed following the procedure of Labbé et al. (2015). Specifically, full depth 3.6 and 4.5  $\mu\text{m}$  mosaics were reconstructed by combining data from the S-COSMOS (Sanders et al. 2007), S-CANDELS (Ashby et al. 2015) and SPLASH (PI: Capak, Steinhardt et al. 2014). The resulting coverage map for the 4.5  $\mu\text{m}$  band is shown in the lower panel of Figure 1. The depth ranges from  $\sim 4$  to  $\sim 90$  hr, which corresponds to  $\sim 23.8$ – $25.4$  AB ( $5\sigma$  in a  $2''0$  aperture).

Observations in the 5.8 and 8.0  $\mu\text{m}$  channels are only available from the S-COSMOS *Spitzer* cryogenic program. These data have a much shallower depth compared to the 3.6 and 4.5  $\mu\text{m}$  bands, with an average limit of  $\sim 22.2$  AB ( $5\sigma$ ,  $2''0$  aperture). For this reason, we only considered galaxies from the GOODS-N/S fields for the  $z \geq 5$  samples.

Fluxes in the four IRAC bands were measured using the template fitting procedure of Labbé et al. (2006, 2010a, 2010b, 2013), adopting the  $K_s$  band as the high-resolution template image to deblend the IRAC photometry.

Photometric redshifts were computed using EAZY (Brammer et al. 2008) on the 37 bands photometric catalog, complementing the standard EAZY template set with a maximally red template SED, i.e., an old (1.5 Gyr) and dusty ( $A_V = 2.5$  mag) SED template. We only considered objects whose fluxes were not contaminated by bright nearby stars, with extended morphology on the  $K_s$  image. Less than five bands were excluded, as the associated flux measurements were contaminated by NaN values. Galaxies in the  $z \sim 4$  redshift bin were selected among those with photometric redshift  $3.5 < z_{\text{phot}} < 4.5$ . The initial  $z \sim 4$  sample included 1208 objects.

The photometric redshift selection allowed us to consider objects that could have been missed by a pure LBG selection. The large area offered by the UltraVISTA DR2 footprint enabled the selection of bright/luminous sources whose surface density would be too low to be probed over the GOODS-N/S fields area. Such luminous systems could be intrinsically redder than normal LBGs, either because they are more (massive) evolved systems and/or they contain a higher fraction of dust. On the other side, the LBG selection at fainter luminosities from the GOODS-N/S samples is expected to suffer only limited selection bias against intrinsic red sources, as galaxies in this range of luminosities are mostly blue star-forming systems, with low dust content.

## 2.2. Sample Assembly

The first step consists of applying a cut in the  $K_s$  flux of the galaxies from the UltraVISTA sample, in order to control the detection completeness. The DR2 data is  $\sim 1$  mag deeper than DR1. Therefore, we set the threshold to  $K_s = 24.4$  mag, corresponding to the 90% completeness in detection of point-sources (Muzzin et al. 2013a). Instead, we did not apply any cut in the detection band of the GOODS-N/S sample, as the method adopted to estimate the co-moving volumes already takes into account the incompleteness from the detection stage.

Successively, we excluded from our sample those galaxies with poor flux measurements in those IRAC bands used to compute the rest-frame  $z'$  luminosity. The variation in depth across each IRAC mosaic prevented us from applying a single value of flux threshold at this stage. Instead, we applied a cut in S/N to the flux in the IRAC band closest to the rest-frame  $z'$  band (i.e., IRAC 4.5  $\mu\text{m}$  at  $z \sim 4$  and IRAC 5.8  $\mu\text{m}$  and 8.0  $\mu\text{m}$  at  $z \gtrsim 5$ ).

Considering the gap in the photometric depth probed by the 4.5  $\mu\text{m}$  mosaics compared to that reached by the 5.8–8.0  $\mu\text{m}$  data, we opted for applying a distinct S/N cut depending on the considered redshift bin. The sample at  $z \sim 4$  was selected by applying the cut of  $S/N > 5$  to the 4.5  $\mu\text{m}$  flux; the samples at  $z \sim 5, 6, 7$  were assembled by considering the cumulative flux in the 5.8 and 8.0  $\mu\text{m}$  bands as the inverse-variance weighted sum of the flux in these two bands. We then applied a cut to the corresponding S/N such that:

$$S/N_{5.8+8.0} \equiv \frac{S_{5.8}w_{5.8} + S_{8.0}w_{8.0}}{\sqrt{w_{5.8} + w_{8.0}}} > 4, \quad (1)$$

where  $S_\kappa$  is the flux measurement in band  $\kappa$  and  $w_\kappa$  is the weight defined as  $1/\sigma_\kappa^2$ , with  $\sigma_\kappa$  the corresponding flux uncertainty. The application of the S/N cut reduced the number of galaxies to 2644 (2040/604 for GOODS/UltraVISTA, respectively), 96, 17 and 4 at  $z \sim 4, z \sim 5, z \sim 6$ , and  $z \sim 7$ , respectively.

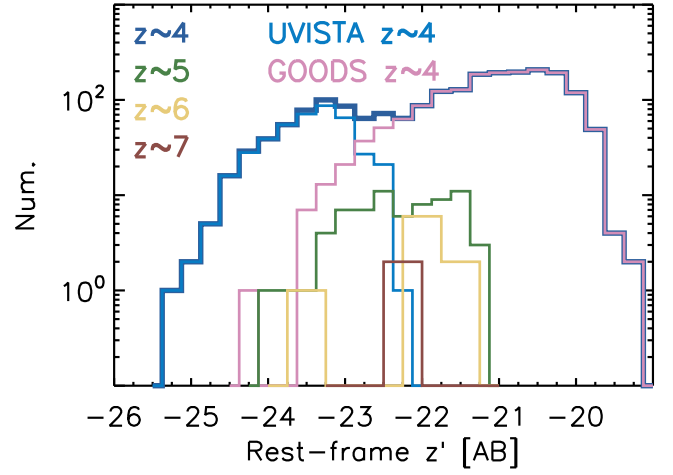
We further cleaned our sample, excluding those objects satisfying any of the following conditions: (1) the contribution to the 5.8 and 8.0  $\mu\text{m}$  flux from neighboring objects is excessively high; (2) the source morphology is very uncertain or confused making IRAC photometry undetermined; (3) the source is detected at X-rays wavelengths, suggesting it is a lower redshift AGN; (4) the source is at higher redshift, but its SED is dominated by AGN light; (5) LBGs with a likely  $z < 3.5$  solution from photometric redshift analysis. In Appendix A, we detail our application of these additional criteria in cleaning our sample.

Our final sample consists of 2098 galaxies at  $z \sim 4$  (1680 from the LBG sample and 418 from the UltraVISTA sample), 72 at  $z \sim 5$ , 10 at  $z \sim 6$ , and 2 objects at  $z \sim 7$ . The distribution of the absolute magnitudes in the  $z'$  band for the sample is presented in Figure 2, for the four different redshift bins. It is noteworthy how the GOODS-N/S and UltraVISTA samples complement each other at  $z \sim 4$ , allowing to fully exploit these data with little redundancy.

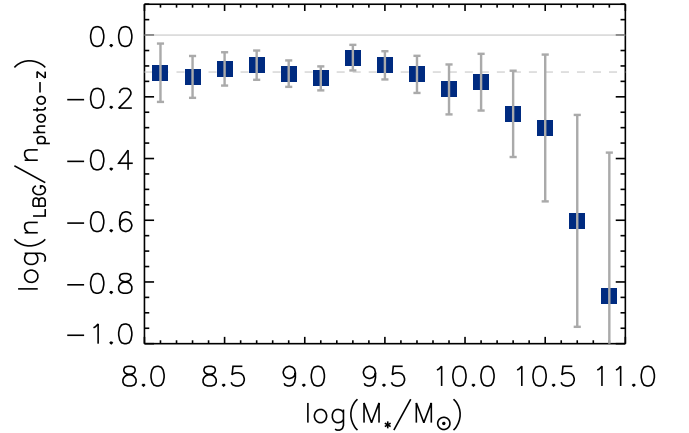
### 2.3. Selection Biases

The samples adopted in this work rely on LBG selection criteria, complemented at  $z \sim 4$  by a photometric redshift selected sample based on the UltraVISTA DR2 catalog.

The criteria adopted for the assembly of our samples introduce two potential biases to our estimates of LF,  $M_*/L$  ratio and SMF (Fontana et al. 2006; Grazian et al. 2015). The Lyman Break criteria select, by construction, blue star-forming galaxies, and may thus exclude a greater fraction of red objects compared to photometric-redshift selections. Furthermore, even samples based on photometric redshifts can suffer incompleteness from very red sources, too faint to appear in the detection bands (usually  $H$  or  $K_s$  or a combination of NIR bands), but that emerge at redder wavelengths (e.g., IRAC). In the following we attempt to evaluate the impact of these biases on our sample.



**Figure 2.** Distribution of the absolute magnitudes in the  $z'$  band of our composite sample of galaxies at  $z \sim 4, 5, 6$ , and  $7$ , as labeled in the figure. For the  $z \sim 4$  sample, the histograms for the GOODS-N/S and UltraVISTA samples are also presented separately, showing the complementarity in  $M_{z'}$  of the two data sets.



**Figure 3.** Fraction of  $z \sim 4$  galaxies recovered using LBG criteria relative to the underlying sample of galaxies selected to have photometric redshifts  $3.5 < z_{\text{phot}} < 4.5$ , shown as a function of stellar mass. The horizontal dashed line marks the mean of the recovered fraction for  $\log(M_*/M_\odot) \sim 10.2$ . The LBG criteria recover  $\gtrsim 75\%$  of all the sources up to  $\log(M_*/M_\odot) \leq 10.2$ .

From the stellar mass catalog of CANDELS/GOODS-S (Santini et al. 2015), we extracted those galaxies with photometric redshift  $3.5 < z_{\text{phot}} < 4.5$ . Successively, we applied the LBG criteria to their flux measurements from the multi-wavelength catalog of Guo et al. (2013). In Figure 3 we present, as a function of stellar mass, the ratio between the number of galaxies recovered through the LBG criteria and the number of galaxies in the photo- $z$  sample. The plot shows that the LBG selection is able to recover  $\gtrsim 75\%$  of galaxies with stellar mass  $\log(M_*/M_\odot) \lesssim 10$  (corresponding to  $\sim -0.12$  dex offset in number density measurements); at  $\log(M_*/M_\odot) \lesssim 10.5$  the galaxies missed by the LBG criteria amount to about 35% ( $\sim 0.2$  dex). At higher stellar masses, the fraction of galaxies not entering the Lyman Break selection increases to  $\gtrsim 60\%$ – $70\%$  ( $\sim 0.5$ – $0.6$  dex) consistent with Grazian et al. (2015). Duncan et al. (2014) showed that photometric uncertainties scatter a large fraction of the measurements outside the LBG selection box; specifically, the LBG criteria recover only  $\sim 1/4$  (equivalent to  $\sim -0.6$  dex offset) of the galaxies recovered through photo- $z$  (see also Dahlen et al. 2010). However, once selection criteria on

the redshift probability distribution are introduced, excluding from the sample poorly constrained photometric redshifts, the resulting photo- $z$  sample largely overlaps with the LBG one, as demonstrated by the fact that the resulting photometric redshift UV LFs agree well, usually within  $1 - \sigma$ , with the LBG UV LF (Duncan et al. 2014; Finkelstein et al. 2015a).

The photometric depth of the UltraVISTA DR2 catalog,  $K_s = 24.4$  mag, corresponds to a stellar mass completeness limit for a passively evolving simple stellar population of  $\log(M_*/M_\odot) \sim 10.6$  at  $z \sim 4$  and 11.2 at  $z \sim 5$ . The depth in the GOODS-Deep fields correspond to limits in stellar mass of  $\log(M_*/M_\odot) \sim 10.3$  at  $z \sim 4$  and 10.6 at  $z \sim 5$ , respectively. We would like to remark that our analysis for the  $z \sim 4$  sample at stellar masses  $\log(M_*/M_\odot) \gtrsim 10-10.5$  is dominated by the photometric redshift sample from UltraVISTA, covering a larger volume for bright sources than the GOODS fields. Therefore, our composite  $z \sim 4$  sample is only marginally affected by the LBG selection bias.

A number of works have studied the so called extremely red objects, characterized by very red ( $\gtrsim 2-3$  mag) rest-UV/optical colors, making them more elusive in high- $z$  samples (e.g., Yan et al. 2004; Huang et al. 2011; Caputi et al. 2012, 2015; Stefanon et al. 2015; Wang et al. 2016). Samples detected in IRAC bands suggest that many of these objects could be consistent with being  $z \gtrsim 3$  massive galaxies.

Recently, Wang et al. (2016) analyzed the properties of  $H - [4.5] > 2.25$  mag over the CANDELS/GOODS-N and GOODS-S fields. Interestingly they identified 18 sources not present in the  $H_{160}$ -band catalog, but included in the IRAC catalog of Ashby et al. (2013). Of these, 5 sources have an estimated photometric redshift  $3.5 < z_{\text{phot}} < 4.5$  and have a stellar mass  $10.5 \lesssim \log(M_*/M_\odot) \lesssim 11$ . Since their analysis refers to the same fields we consider in our work (although likely the configurations at the detection stage are different), we can use their result to obtain a rough estimate of the fraction of objects missed by our selection. Assuming a  $M_*/L_{z'} \sim 0.2M_\odot/L_\odot$ , quite typical for these masses and redshifts (as we show in Section 3.2), the stellar mass range of these galaxies would correspond to luminosities  $-24.7 \lesssim M_{z'} \lesssim -23.4$  AB. This sample would then constitute  $\sim 65\%$  of the objects in our  $z \sim 4$  LBG sample with similar luminosities. This fraction drops to  $\sim 8\%$  when comparing the 5 sources to the  $\sim 60$  galaxies with the same photometric redshift and stellar mass over the CANDELS/GOODS fields.

Caputi et al. (2015) presented SMF measurements at  $z \sim 3-5$  obtained complementing the SMF from a photometric-redshift,  $K_s$ -selected sample based on UltraVISTA DR2 data to SMF measurements from photometric redshift samples of  $K_s$ -dropouts detected in IRAC bands. The main result is that  $K_s$  dropouts can account for as high as  $\sim 0.5$  dex in number densities. However, Stefanon et al. (2015) showed that samples similar to those of Caputi et al. (2015) likely suffer from degeneracies in the measurement of photometric redshifts (and consequently stellar masses), and therefore the above estimate is still uncertain.

The depth of current IRAC data, however, is not sufficient to systematically inspect passively evolving stellar population with stellar masses below  $\sim 10^{10-10.5}M_\odot$  at  $z \gtrsim 4$ . We therefore caution the reader that any sample currently available dealing with stellar mass below the  $\sim 10^{10}M_\odot$  limit may still be biased against dusty and/or old galaxies. Forthcoming projects, like *Spitzer*/GREATS (I. Labbé et al. 2017, in preparation) and the

*JWST* will allow us to obtain a more complete picture by probing the lower mass regime.

#### 2.4. Selection Efficiency and Completeness

We implemented a Monte Carlo simulation based on real data to estimate the effects that our selection criteria in S/N and contamination polishing have on the sample of galaxies used in this work. For this simulation, we did not consider the effects of selection in the detection band, because the UltraVISTA sample is 90% or more complete in  $K_s$  by construction, whereas the effects of detection completeness in the GOODS-N/S sample have been taken into account when estimating the co-moving volumes adopted for the LF measurements.

We first defined a grid in apparent magnitude of width 0.20 mag. Given the small sizes of the galaxies compared to the IRAC PSF, for each magnitude value in the grid, we injected 100 point sources randomly distributed across a region of uniform depth in the 4.5, 5.8, and 8.0  $\mu\text{m}$  mosaics of the GOODS-N field. We chose the GOODS-N because this field is characterized by the deeper *Spitzer*/IRAC 4.5, 5.8, and 8.0  $\mu\text{m}$  band data among the fields considered for this work. We successively replicated the flux measurement, using the same procedure adopted for the actual photometry.

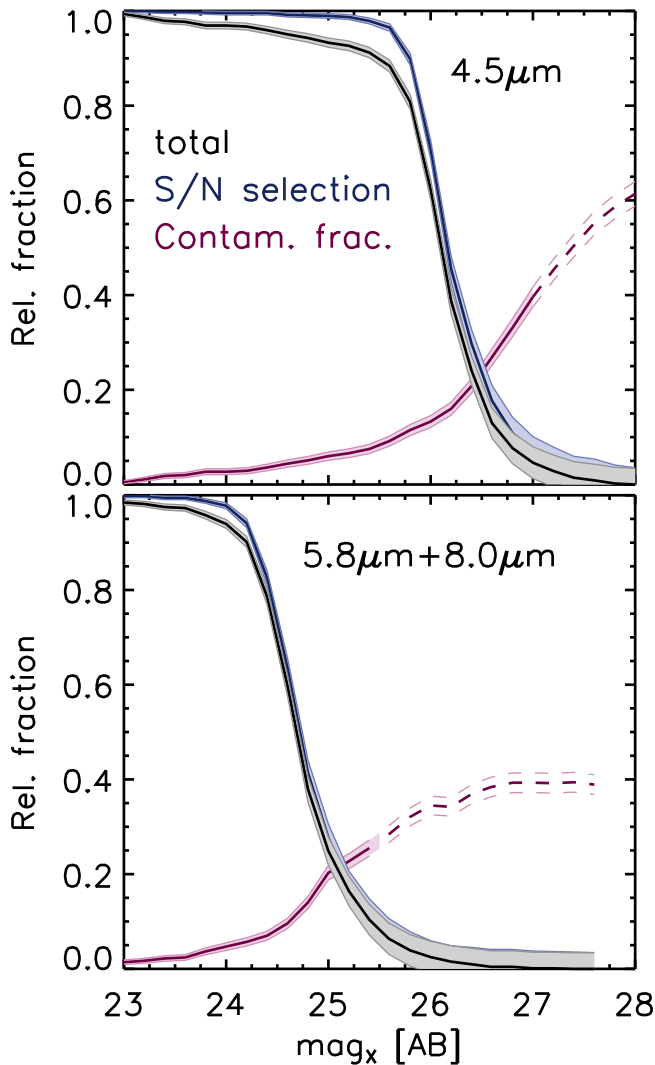
The completeness fraction in each magnitude bin was computed by comparing the number of objects satisfying our selection criteria (Section 2.2) to the number of objects initially injected into the simulation. For the completeness of the  $z \sim 4$  sample, the above process was applied to the 4.5  $\mu\text{m}$  mosaic only. For the completeness of the samples at  $z \geq 5$ , the point sources were added at matching positions in the 5.8 and 8.0  $\mu\text{m}$  mosaics. The selection of the S/N and contamination was then recovered by applying the corresponding criteria and assuming the SED to be flat in  $f_\nu$  in the observed 5–9  $\mu\text{m}$  region. This is a reasonable approximation because, as we show in Section 3.2, the median SEDs do not substantially deviate from a flat  $f_\nu$  SED in the wavelength range covered by IRAC observations. The whole process was repeated  $10\times$  in each band, to increase their statistical significance. The global completeness (i.e., the cumulative effects of S/N and contamination selection) at the different depths of the IRAC mosaics was obtained by rescaling the completeness in S/N selection to match the depth of the relevant region.

The results from our completeness simulation for the 4.5  $\mu\text{m}$  and the 5.8  $\mu\text{m} + 8.0 \mu\text{m}$  samples are presented in Figure 4. We first discuss the recovery of the contamination fraction; we then consider these results in the budget of our global completeness estimates.

The fraction of objects in the 4.5  $\mu\text{m}$  band contaminated by neighbors<sup>6</sup> is negligible for objects brighter than  $\sim 24$  AB, and increases exponentially up to  $\sim 27$  AB, where it starts to flatten out. A similar behavior is observed for the 5.8  $\mu\text{m} + 8.0 \mu\text{m}$  simulation, although shifted at brighter magnitudes due to the shallower depth of the 5.8 and 8.0  $\mu\text{m}$  compared to the 4.5  $\mu\text{m}$ . The flattening at the faint end is caused by a strong incompleteness in the data at such faint magnitudes, and likely does not reflect the true behavior. In what follows and our analysis, we do not consider the completeness for magnitudes fainter than those corresponding to the onset of the flattening,

<sup>6</sup> We consider a flux measurement to be contaminated when 65% or more of the flux at the position of the source comes from neighboring objects. See Appendix A for further details.





**Figure 4.** Completeness fraction as a function of apparent magnitude, from our Monte Carlo simulation for selection of the  $4.5\ \mu\text{m}$  (top panel) and  $5.8\ \mu\text{m} + 8.0\ \mu\text{m}$  (bottom panel) bands. The pink curve presents the fraction of objects excluded because their flux measurements were highly contaminated by neighbors. The pink shaded area presents the associated  $1\sigma$  Poisson uncertainties. The blue curve and shaded area present the completeness fraction from the S/N cut in the flux of the corresponding band and associated Poisson uncertainty, respectively. The black curve and gray shaded area show the combined effect of S/N threshold and contamination cleaning, respectively.

i.e.,  $\sim 27$  AB and  $\sim 25$  AB for the  $4.5\ \mu\text{m}$  and  $5.8\ \mu\text{m} + 8.0\ \mu\text{m}$  data, respectively.

As it could intuitively be expected, the bright end of the global completeness curve is dominated by the (small) fraction of purged objects. This effect becomes less and less pronounced at fainter magnitudes, corresponding to lower S/N, where the effective selection is driven by the S/N itself.

### 2.5. Flux Boosting

The random noise from the background can positively combine at the location of a given source, introducing an increase in the measured flux (*flux boosting*, Eddington 1913). The amount of this boost is inversely correlated to the S/N of the source. The flux for sources with very high S/N will mostly be the result of the photons emitted by the source itself, with reduced contribution from the background; on the other hand, for sources with low S/N, the background level can be

typically just few factors smaller than the intrinsic signal from the source, making it sensitive to (positive) fluctuations of the background. Furthermore, sources do not uniformly distribute with flux, but rather follow an approximate power law, with fainter sources more numerous than brighter ones. Therefore, it is intrinsically more probable that fainter sources scatter to brighter fluxes than the reverse, giving origin to a net flux bias.

A second potential source of flux boosting comes from confusion noise: faint sources at apparent positions close to a brighter one are more likely to be blended into the brighter source, increasing the flux and decreasing the number of fainter objects. This effect is larger for flux measurement in those bands with wide PSF, like *Spitzer*/IRAC. However, in our case, the photometry in the IRAC bands was performed by adopting a higher-resolution morphological prior from *HST* mosaics (see Section 2.1). Furthermore, we applied a selection in flux contamination (see Section 2.2). Because these two factors drastically limit the potential contribution of confusion noise to the IRAC fluxes in our sample, we do not further consider its effects to the flux-boosting budget.

For each source, we estimated the flux bias as the ratio between the expected intrinsic flux  $f_i$  and the measured flux  $f_o$ . Because no direct measurement is possible, the intrinsic flux was recovered as the average flux obtained from an estimate of its probability distribution. This was constructed considering two distinct contributions: (1) the probability  $\mathcal{P}(f_i, f_o)$  that the observed flux  $f_o$  is drawn from the distribution of intrinsic flux  $f_i$ , given the noise  $\sigma_o$ , and (2) the frequency  $\mathcal{P}(f_i)$  of occurrence of the intrinsic flux  $f_i$ . Assuming each probability is normalized to 1, the final probability distribution would then be  $\mathcal{P}_{\text{tot}}(f_i) = \mathcal{P}(f_i, f_o)\mathcal{P}(f_i)$ .<sup>7</sup>

Assuming a Gaussian noise,  $\mathcal{P}(f_i, f_o)$  can be written as:

$$\mathcal{P}(f_i, f_o) = \frac{1}{\sqrt{2\pi\sigma_o^2}} \exp[-(f_i - f_o)/2\sigma_o^2], \quad (2)$$

normalized to a total probability of one.

The frequency associated with the intrinsic flux can be recovered from the (intrinsic) differential number count of sources,  $dN(f_i)/df_i$ . This can usually be described by a power-law form with negative index, which is thus divergent for  $f_i \rightarrow 0$ , preventing it from being normalized<sup>8</sup> (see, e.g., Hogg & Turner 1998, who also discuss possible reasons why the divergence at  $f_i \rightarrow 0$  is likely non-physical).

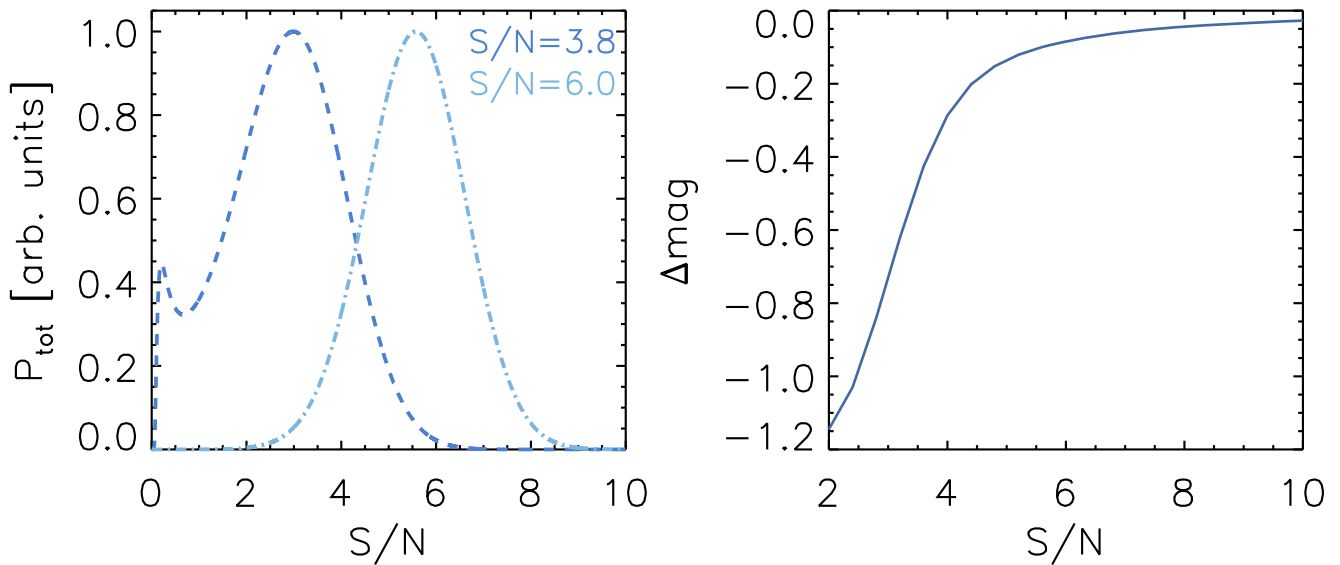
We therefore followed the formalism of Crawford et al. (2010), who introduced, as further constraint, the Poissonian probability that no sources brighter than  $f_i$  exist at the same location of the observed object. The expression for  $\mathcal{P}(f_i)$  then becomes:

$$\mathcal{P}(f_i) = \frac{dN(f_i)}{df_i} \times \exp\left(-\Delta\Omega_o \int_{f_i}^{+\infty} dN/df\right), \quad (3)$$

where  $\Delta\Omega_o$  corresponds to the area occupied by the source. In the left panel of Figure 5, we show examples of reconstructed  $(\mathcal{P})_{\text{tot}}$ , for the cases of S/N = 6 and S/N = 3.8, where it is evident that the contribution of the faint source population to

<sup>7</sup> A similar expression could also be recovered, modulo a normalization factor, by applying Bayes' theorem—see, e.g., Hogg & Turner (1998).

<sup>8</sup> For S/N  $\gtrsim 5$ , the product of  $\mathcal{P}(f_i, f_o)$  and  $dN/df$  does not diverge for  $f_i \rightarrow 0$ . However, this is no longer the case for lower S/N values, where the non-negligible probability of the low-flux tail from the Gaussian makes the divergent power-law dominate the Gaussian.



**Figure 5.** Left panel: examples of probability distribution of the intrinsic flux  $\mathcal{P}(f_i)$ , presented as a function of  $S/N$ , for two cases of observed  $S/N = 3.8$  and  $6.0$ , as labeled at the top right. The probabilities have been arbitrarily re-normalized to a maximum of one, for ease of readability. Right panel: expected flux boosting as a function of  $S/N$ , resulting from Equation (3). The flux boosting for  $S/N \lesssim 2.5$  is  $\gtrsim 1$  mag, suggesting that the recovery of the intrinsic flux for these cases can be very uncertain.

the expected intrinsic flux increases as the  $S/N$  decreases. The right panel of Figure 5 shows the expected flux boost as a function of  $S/N$ . For  $S/N \gtrsim 4.5$ , the flux boost is roughly the same amount as the flux uncertainty. However, for lower  $S/N$ , the estimated flux boost increases abruptly. For  $S/N \lesssim 2$ , the expected flux boost is  $\gtrsim 1.5$  mag, meaning that the recovery of the intrinsic flux for such low  $S/N$  data is highly uncertain.

The  $S/N$  in the *HST* bands for the galaxies in our sample is  $> 10$ . At  $z \sim 4$ , the  $S/N$  in the  $4.5 \mu\text{m}$  band, adopted for the selection of the  $z \sim 4$  sample, is  $\geq 5$  by construction; the  $S/N$  in the  $3.6 \mu\text{m}$  band is  $\geq 5$  as well, consistent with the nearly flat SEDs in that wavelength range. At  $z \geq 5$ , the selection was performed in a combination of  $5.8 \mu\text{m}$  and  $8.0 \mu\text{m}$  fluxes, adopting a  $S/N > 4$  threshold. Figure 5 shows that the expected flux boost for  $S/N > 5$  is  $\lesssim 0.1$  mag. However, for lower  $S/N$ , typical of the selection of samples at  $z \geq 5$ , the correction can be as high as  $\sim 0.8$ – $1.0$  mag.

We therefore applied the above correction to the fluxes in the  $5.8$  and  $8.0 \mu\text{m}$  of the  $z \geq 5$  samples. The average flux boost was  $\sim 0.19$  mag and  $\sim 0.25$  mag in the IRAC  $5.8 \mu\text{m}$  and  $8.0 \mu\text{m}$  bands, respectively.

In Appendix B, we present the SEDs of the 12 most luminous galaxies in the  $z \sim 5$  sample, along with the SEDs of the  $z \sim 6$  and  $z \sim 7$  samples, before and after applying the flux boost correction.

### 3. Results

#### 3.1. UV to Optical Luminosities

In the last few years, a number of works have studied the relation between the rest-frame UV luminosity and the stellar mass of high-redshift galaxies (e.g., Stark et al. 2009; González et al. 2011; Lee et al. 2011; McLure et al. 2011; Duncan et al. 2014; Spitler et al. 2014; Grazian et al. 2015, V. González et al. 2017, in preparation; Song et al. 2016). Indeed, a relation between the stellar mass and the UV luminosity is to be expected when considering a continuous star formation. Deviations from such a relation would then provide information on the

age and metallicity of the stellar population and dust content of the considered galaxies.

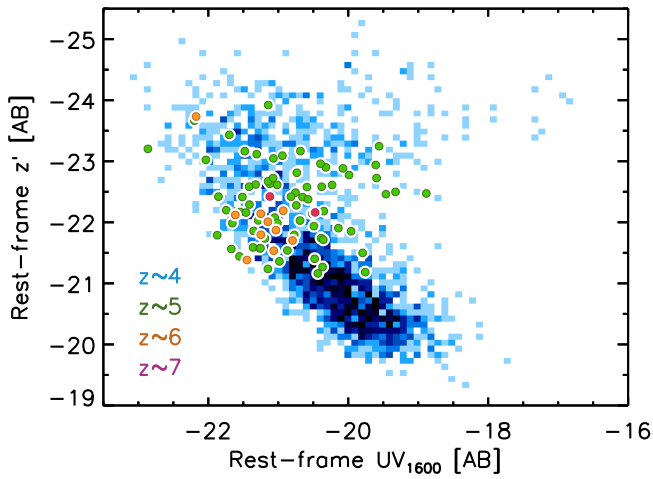
The emerging picture is that, at  $z \sim 4$  and for stellar masses  $\log(M_*/M_\odot) \lesssim 10$ , the stellar mass increases monotonically with increasing UV luminosity; however, at stellar masses higher than  $\log(M_*/M_\odot) \sim 10$ , the trend becomes more uncertain: Spitler et al. (2014), using a sample of  $K_s$ -based photometric redshift selected galaxies, found indication of a turnover of the UV luminosity, with the more massive galaxies ( $10.5 \lesssim \log(M_*/M_\odot) \lesssim 11$ ) spanning a wide range in UV luminosities (see also Oesch et al. 2013). Lee et al. (2011), on the other hand, using an LBG sample, found a linear relation between UV luminosities and stellar masses up to  $\log(M_*/M_\odot) \sim 11$ . Considering the different criteria adopted by the two teams for the assembly of their samples, selection effects might be the main reason for the observed tension.

This observed discrepancy could, however, just be the tip of the proverbial iceberg. Indeed, current high- $z$  surveys might still be missing lower-mass, intrinsically red galaxies (dusty and/or old), which could populate the  $M_*$ – $M_{\text{UV}}$  plane outside the main sequence (Grazian et al. 2015 and our discussion in Section 2.3). The depth of the current NIR surveys does not allow us to further inspect this, which will likely remain an open issue until *JWST*.

A monotonic relation between the UV luminosity and the stellar mass has also been found at  $z \sim 5$  and  $z \sim 6$  for  $\log(M_*/M_\odot) \lesssim 10$  (e.g., Stark et al. 2009; González et al. 2011; Duncan et al. 2014; Salmon et al. 2015; Song et al. 2016), with approximately the same slope and dispersion, but with an evolving normalization factor (but see, e.g., Figure 5 of Song et al. 2016 for further hints as to the existence of massive galaxies with faint UV luminosities).

Figure 6 presents the absolute magnitude in the rest-frame  $z'$  band ( $M_{z'}$ ) as a function of the absolute magnitude in the UV ( $M_{\text{UV}1600}$ ), for our sample in the four redshift bins ( $z \sim 4, 5, 6$  and  $7$ ). In Figure 7, we present the binned median in the  $M_{\text{UV}}$ – $M_{z'}$  plane for the  $z \sim 4$  and  $z \sim 5$  samples.





**Figure 6.** The  $z'$  absolute magnitudes vs. UV absolute magnitudes for our composite sample, color-coded according to the considered redshift bin. The  $z \sim 4$  data are presented as a density plot, with denser regions identified by a darker color, whereas the points for the  $z \geq 5$  samples are shown individually. The UV- $z'$  relation shows a turnover for  $M_{z'} \lesssim -22.5$ .

The  $z \sim 4$  sample shows a clear correlation between the luminosities in the rest-frame UV and  $z'$  bands for  $M_{z'} \gtrsim -22$  mag, which can be described by the following best-fitting linear relation:

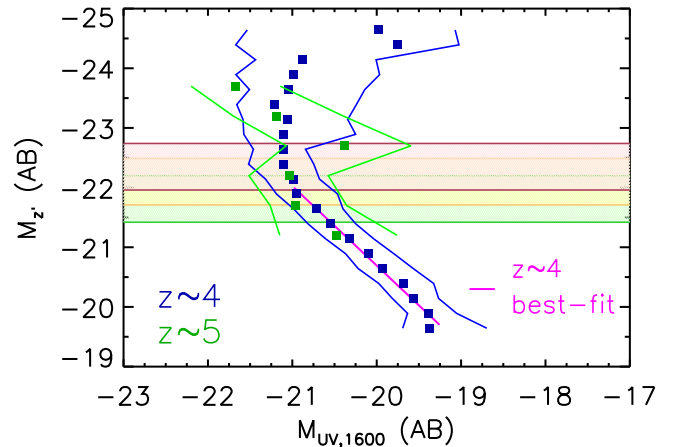
$$M_{UV} = (-3.58 \pm 1.49) + (0.79 \pm 0.07) \times M_{z'}. \quad (4)$$

The above best fit is marked by the magenta line in Figure 7, where we also indicate the  $3\sigma$  limits corresponding to our  $5.8 \mu\text{m} + 8.0 \mu\text{m}$  selection. At  $z \sim 4$  and  $z \sim 5$ , the depth of the IRAC data allows us to not only probe the bright end, where the relation between  $M_{UV}$  and  $M_{z'}$  breaks, but to also explore the regime of the linear correlation expressed by Equation (4). Slopes of 0.4–0.5 in the  $\log(M_*)-M_{UV}$  plane (with nominal  $1\sigma$  uncertainties of  $\sim 0.05-0.1$ ) have been reported by, e.g., Duncan et al. (2014) and Song et al. (2016). Assuming a constant  $M_*/L_{z'}$  ratio (see Section 3.2), our measurements correspond to a slope of 0.5 in the  $\log(M_*)-M_{UV}$  plane, consistent with previous measurements.

Assuming that the SFR mostly comes from the UV light, and that  $M_{z'}$  is a good proxy for stellar mass measurements, we can also compare the slope we derived for the  $M_{UV}-M_{z'}$  relation to that of the  $\log(\text{SFR})-\log(M_*)$  from the literature. Indeed, the observed UV slopes of  $M_{z'} \gtrsim -22$  galaxies in our sample are  $\beta \sim -2$  (see also Bouwens et al. 2010), consistent with star-forming galaxies and little-to-no dust extinction. Our measurement is perfectly consistent with the  $\log(\text{SFR})-\log(M_*)$  slope of  $\sim 0.8 \pm 0.1$  recently measured by, e.g., Whitaker et al. (2015) for  $z \lesssim 2.5$  star-forming galaxies with  $\log(M_*/M_\odot) \lesssim 10.5$ , and it has been shown to evolve little over the redshift range  $0.5 < z < 2.5$ .

For absolute magnitudes brighter than  $M_{z'} \sim -22$  mag, the linear relation expressed by Equation (4) breaks as we observe the beginning of a turnover in the absolute UV- $z'$  magnitude relation. Remarkably, this behavior is visible also for the  $z \sim 5$  sample, which is entirely based on LBG selection. This fact has important consequences for, e.g., SMF measurements; samples of galaxies selected at fixed rest-frame UV luminosity are potentially characterized by a wide range of stellar mass.

The absolute UV magnitude of galaxies with  $M_{z'} \sim -23.5 \pm 0.8$  mag spans the full range of values observed for



**Figure 7.** Median of the  $M_{UV}$  vs.  $M_{z'}$  relation in bins of  $M_{z'}$ . The blue points mark the median at  $z \sim 4$ , whereas the green points mark the median at  $z \sim 5$ . The left (blue) and right (green) curves represent the 25th and 75th percentiles, respectively, of the points at  $z \sim 4$  ( $z \sim 5$ ). The horizontal green, yellow, and red shaded regions indicate the limiting magnitude, corresponding to our S/N cuts at  $z \sim 5$ ,  $z \sim 6$ , and  $z \sim 7$ , respectively. The magenta line represents the best-fit relation for the  $z \sim 4$  sample in the range  $-22 < M_{z'} < -19.75$  mag.

$M_{z'} \lesssim -22.5$  mag. However, the bulk of values aggregates around the  $[M_{UV}, M_{z'}] \sim [-21.4, -23.5]$  mag region, and it is characterized by a large dispersion in  $M_{UV}$  ( $\gtrsim 3$  mag). This result is qualitatively consistent with the findings of Spitler et al. (2014), assuming a correlation between the absolute magnitude  $M_{z'}$  and the stellar mass. Most of the galaxies with  $M_{z'} \sim -23.5$  come from the photometric redshift sample selected from the UltraVISTA catalog. As we will present in Section 3.2, our measurements of the mass-to-light ratios from stacking analysis show that galaxies with  $M_{z'} \lesssim -23.5$  statistically have stellar masses  $\log(M_*/M_\odot) \gtrsim 10.6$ . The above result then underlines the bias that LBG selections may introduce against massive systems.

At  $z \sim 5$ , our data allow us to inspect the relation only for  $M_{z'}$  fainter than  $-23$  and  $M_{UV,1600}$  fainter than  $\sim -22$ . In this range of luminosities, our  $z \sim 5$  measurements are roughly consistent with the  $z \sim 4$  measurements in the same range of luminosities. The measurements for the  $z \sim 6$  and  $z \sim 7$  samples are still consistent with the trends observed at  $z \sim 4$ , although the low number of objects does not allow us to derive any statistically significant conclusion.

### 3.2. Stellar Mass-to-light Ratios from Stacking Analysis

Thus far, determinations of the stellar mass-to-light ratios ( $M_*/L$ ) for galaxies at  $z > 4$  have involved the  $M_*/L_{UV}$  ratio. This quantity is fundamental to our understanding of galaxy formation and evolution, as it combines information on the recent (through the UV luminosity) and integrated (through the stellar mass) SFH (e.g., Stark et al. 2009). Nonetheless, the  $M_*/L_{UV}$  has been used to recover the stellar mass and SMF of high redshift galaxies with alternating success (see, e.g., González et al. 2011; Song et al. 2016). In this section, instead, we explore for the first time the  $M_*/L_{z'}$  properties of galaxies at  $z \gtrsim 4$ . The rest-frame  $z'$  luminosity is more sensitive to the stellar mass, compared to the UV luminosity, for two reasons. Although the UV light is emitted by massive, short-lived stars, and thus traces the SFH in the past few hundred Myr, the luminosity in the rest-frame optical region mostly originates from lower-mass, longer-lived stars, which

constitute most of the stellar mass of galaxies. Furthermore, it is less sensitive to the dust extinction, and hence to the uncertainties in its determination, compared to the UV; for a Calzetti et al. (2000) extinction curve, an  $A_V = 1$  mag gives  $A_{\lambda 1600} \sim 2.5$  mag, compared to  $A_{\lambda 9000} \sim 0.5$  mag.

Because we are interested more on average trends in the  $M_*/L_{z'}$  ratios here, rather than studying it for specific galaxies, we performed our analysis using the median stacked SEDs constructed from our composite sample. Due to the different photometric bands in the catalogs, we performed the stacking of sources separately for sources in the GOODS-N/S and UltraVISTA samples.

The stacked SEDs were constructed as follows (see also González et al. 2011). At each redshift interval, we divided the galaxies into sub-samples according to their  $M_{z'}$ . The different depths reached by the  $4.5 \mu\text{m}$  and  $5.8 \mu\text{m} + 8.0 \mu\text{m}$  samples resulted in different numbers of subsamples across the redshift bins. Under the working assumption of limited variation in both redshift and SED shape in each bin of  $M_{z'}$ , as well as for each *HST* band, we took the median of the individual flux measurements. Our assumption is also supported by the fact that the SEDs from stacking are generally characterized by a flat  $f_\nu$  continuum at both rest-frame UV and optical regimes. Uncertainties on the median were computed from bootstrap techniques, drawing with replacement the same number of flux measurements as the number of galaxies in the considered absolute magnitude bin. Before median-combining, the fluxes were perturbed according to their associated uncertainty. The process was repeated 1000 times, and the standard deviation of the median values was taken as the final uncertainty. For the IRAC bands, median stacking was performed on the mosaic cutouts centered at the position of each source, previously cleaned from neighbors. Photometry was performed on the median of the images in apertures of  $2''.5$  diameter. Total fluxes were then recovered through the PSF growth curve. Uncertainties were computed by applying to the image cutouts the same bootstrap technique adopted for the median stacking of the fluxes, as described above. In randomly drawing the image cutouts, we preserved the total exposure time.

Photometric redshifts and  $z'$ -band luminosities were obtained from EAzy (Brammer et al. 2008) on the stacked SEDs. Briefly, EAzy initially selects the two SED templates that provide the closest match to the observed color in the two filters bracketing the rest-frame band of interest. The luminosity is then computed from the interpolation of the two colors, relative to the rest frame band of interest, obtained from the two selected SEDs (for full details, see Appendix C of Rudnick et al. 2003). Stellar masses were computed by running FAST (Kriek et al. 2009), adopting the Bruzual & Charlot (2003) template SEDs, a Chabrier (2003) IMF, solar metallicity, and a delayed-exponential SFH. The bands potentially contaminated by nebular emission were excluded from the fit. Because we performed the stacking in each band individually, assuming the same redshift for all sources, the flux in those bands close to the Lyman and the Balmer breaks potentially suffers from high scatter introduced by the range of redshifts of the galaxies in each sub-sample, depending on whether the break enters the band. Fluxes in these bands were therefore excluded from the fit with FAST. Specifically, for the  $z \sim 4$  LBG stacks, we excluded the  $B_{435}$ ,  $V_{606}$  and  $H_{160}$  bands; for the  $z \sim 4$  UltraVISTA stacks, we excluded the  $B$ , IA427, IA464, IA484, IA505, IA527, IA574, IA624, IA679,  $g'$ ,  $g^+$ ,  $V$ ,  $H$

bands; for the  $z \sim 5$  stacks, we excluded the  $V_{606}$  and the  $i_{775}$  bands; for the  $z \sim 6$  stack, we excluded the  $i_{775}$  band; for the  $z \sim 7$  stack, we excluded the  $I_{814}$  band. The stacked SEDs, along with the best-fit templates from FAST, are presented in Figure 8.

The photometric redshifts measured from the stacked SED are all consistent with the values of the corresponding redshift bin; the difference of the photometric redshifts of the median stacked SEDs and the median of the photometric redshifts of the individual sources in each subsample is  $\Delta z/(1+z) \lesssim 0.05$ , i.e., within the uncertainties expected for photometric redshifts.

The inset in the left panel of Figure 8 presents the slope of the UV continuum ( $\beta$ ), measured on the stacked photometry, as a function of absolute magnitude  $M_{z'}$ . The stacked SEDs at  $z \sim 4$  and  $z \sim 5$  are characterized by a trend in the UV slope, with bluer slopes for low-luminosity galaxies, particularly evident for the  $z \sim 4$  stacks, and qualitatively consistent with the results of, e.g., González et al. (2011) and Oesch et al. (2013). Our measurements do not present evidence for evolution of  $\beta$  with redshift at fixed luminosity ( $\approx M_*$ ), although the large uncertainties in  $\beta$  (especially for the higher redshift bins) may be blurring trends. Furthermore, the SEDs in the rest-frame UV wavelength region of the four or five brightest  $z \sim 4$  stacks do not differ too much one from each other, whereas they differ substantially at wavelengths redder than the Balmer/4000 Å break.

Recently, Oesch et al. (2013) presented stacked SEDs at  $z \sim 4$  in bins of  $z$ -band absolute magnitudes from a sample of galaxies based on CANDELS/GOODS-S, HUDF, and HUDF09-2. The sample benefits from deep IRAC  $3.6 \mu\text{m}$  and  $4.5 \mu\text{m}$  imaging from the IUDF program. The stacked SEDs (see, e.g., their Figure 2) show a clear trend of redder colors with increasing rest-frame  $z$ -band luminosity, in particular for  $M_z < -21.5$ . Our stacked SEDs are in qualitative agreement with those of Oesch et al. (2013), confirming the observed trend with luminosity. Furthermore, thanks to the wide area offered by UltraVISTA, which provides coverage for even brighter sources, we are able to extend the trend to even more luminous galaxies.

In the top panel of Figure 9, we present the  $M_*/L_{z'}$  measured<sup>9</sup> from the stacked SEDs as a function of  $z'$  absolute magnitude. Total uncertainties were obtained by propagating the 68% confidence intervals in stellar mass generated by FAST and the uncertainties in luminosity, taken as the flux uncertainties with stacking. At  $z \sim 4$ , the  $M_*/L_{z'}$  ratio is consistent with being constant for  $M_{z'}$  fainter than  $\sim -22.5$  mag. We find:

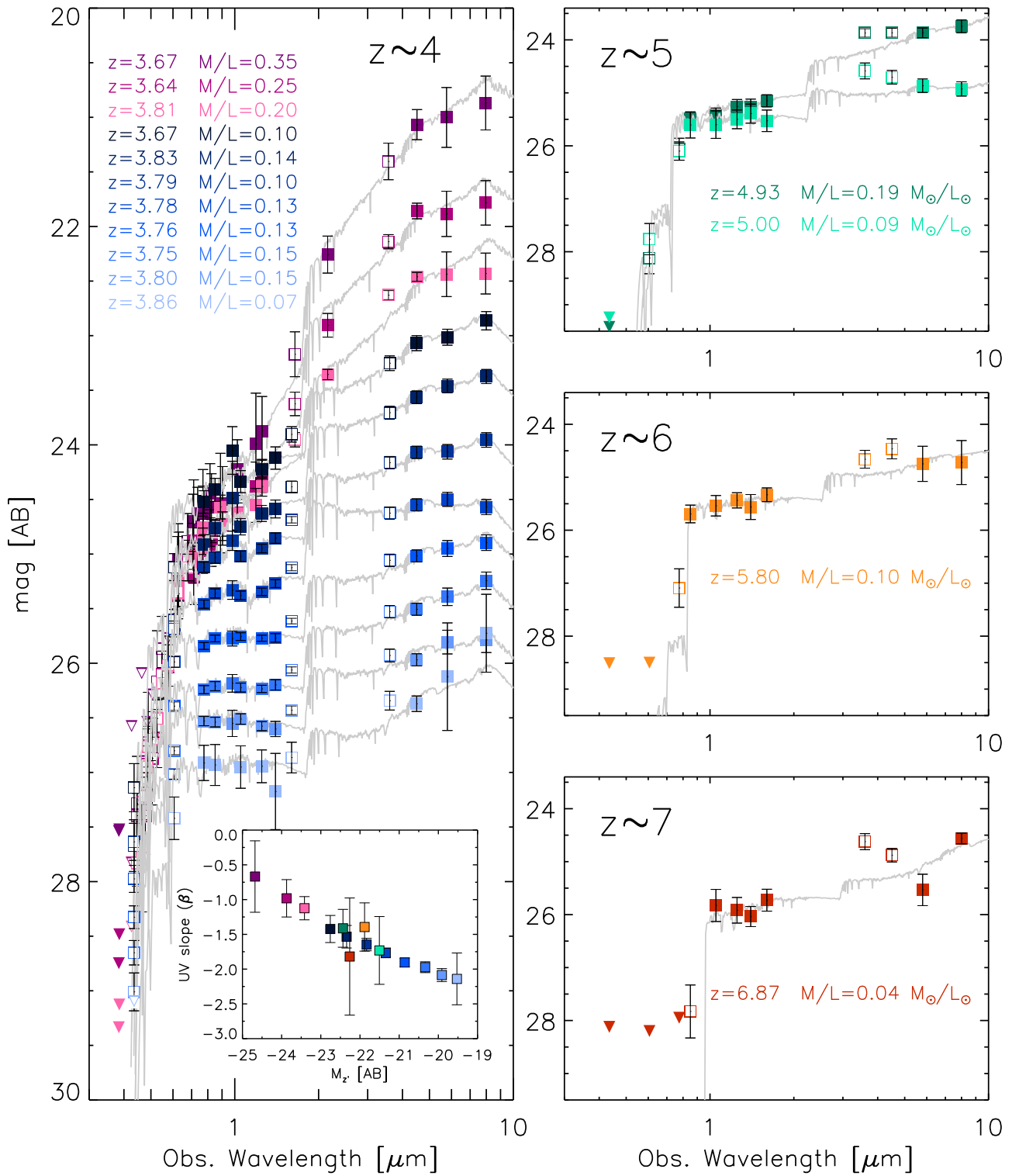
$$\log(M_*/L_{z'}) = -0.87 \pm 0.09, \text{ for } M_{z'} \gtrsim -22.5. \quad (5)$$

For  $M_{z'} \lesssim -22.5$  mag, there is indication of  $M_*/L_{z'}$  increasing with the luminosity, although the error bars are large. The best-fit SEDs of the most luminous stacks have a nearly constant age ( $\sim 10^{8.8}$  yr), and show an  $A_V$  slightly increasing with stellar mass (from 1.0 to 1.2 mag). A linear fit of the  $\log(M_*/L_{z'})$  values for  $M_{z'} \lesssim -22.5$  mag resulted in the following relation:

$$\log(M_*/L_{z'}) = (-5.1 \pm 4.7) - (0.19 \pm 0.18) \times M_{z'}, \quad (6)$$

for  $M_{z'} \lesssim -22.5$ .

<sup>9</sup> We adopted  $M_{z',\odot} = 4.52$  AB.

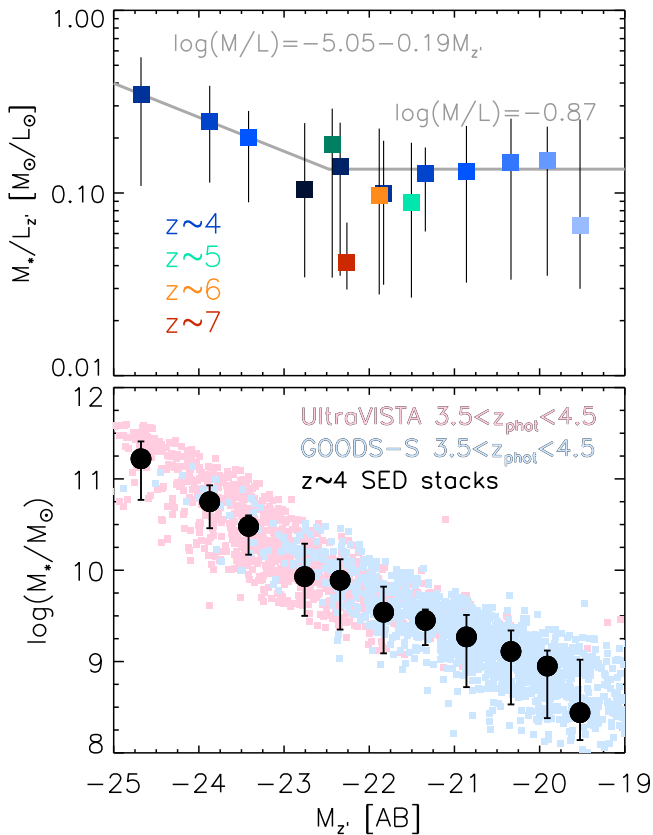


**Figure 8.** Stacked SEDs. Each panel refers to a redshift bin, as indicated by the labels. In each panel, the filled colored squares with error bars represent the SED from the stacking analysis. The gray curve marks the best-fitting FAST template. The open symbols mark those bands excluded from the fit due to potential contamination by nebular emission, or whose stacked measurement was considered unreliable due to the presence of either the Lyman or Balmer break (see the main text for details). For the  $z \sim 4$  sample, the stacked SEDs from the UltraVISTA catalog are plotted with shades of magenta. The stacks from the GOODS-N/S sample are plotted with shades of blue. In each panel, the photometric redshift from EAZY and the mass-to-light ratio ( $M/L$ ), in units of  $M_{\odot}/L_{\odot}$ , for each stacked SED, are also reported and share the color of the corresponding SED. The inset in the  $z \sim 4$  panel presents the relation between the UV slope ( $\beta$ ) and the absolute magnitude  $M_z$  for the stacked SEDs in the four redshift bins. Colors match the redshift and luminosity bin.

Our linear relation recovers a constant value of  $\log(M_*/L_{z'}) = -0.87$  at  $M_{z'} \sim -22.5$ . However, the uncertainties on the fit parameters make the above relation also consistent with a constant value.

The  $M_*/L_{z'}$  ratios for the  $z \sim 5$  and  $z \sim 6$  are consistent with  $M_*/L_{z'}$  measurements of the  $z \sim 4$ ,  $M_z > -22.5$  AB stacks. The  $M_*/L_{z'}$  for the  $z \sim 7$  bin is consistent with the average  $M_*/L_{z'}$  only at the  $\sim 3\sigma$  level. We note, however, that





**Figure 9.** Top panel: the color-filled squares with error bars present the mass-to-light ratio ( $M_*/L_{z'}$ ) from our stacking analysis, as a function of absolute magnitude in the  $z'$  band ( $M_{z'}$ ). The color codes in the legend identify the four redshift bins considered in this work. The assumed constant and best fit log-linear relations to the  $z \sim 4$  points are displayed by the gray lines, and expressed by the gray labels. Bottom panel: relation between the stellar mass and luminosity in the rest-frame  $z'$  band, expressed in terms of absolute magnitude, obtained from samples selected through photometric redshifts (i.e., no LBG selection) over the COSMOS/UltraVISTA and CANDELS/GOODS-S fields (pink and light blue points, respectively). The black solid circles with error bars mark our  $z \sim 4$  estimates of stellar mass and luminosity from the stacking analysis. Unlike the  $M_{UV}-M_{z'}$  plane (Figure 6), the  $M_*-M_{z'}$  follows a monotonic relation, with virtually no outliers over a wide range in luminosity and stellar mass. No significant offset is observed between our stacking measurements and those from the photo- $z$  selected sample.

the  $z \sim 7$  sample only includes two sources, thus reducing the statistical significance of the observed disagreement.

The above results are consistent with what is observed in Figure 6. In Section 3.1, we showed that galaxies more luminous than  $M_{z'} \sim -22.5$  mag form a cloud in the rest-frame UV- $z'$  plane around  $M_{UV} \sim -21.4$  mag. From our stacking analysis, the average apparent magnitude at  $\lambda_{\text{obs}} \sim 8000 \text{ \AA}$  (i.e., the rest-frame UV<sub>1600</sub>) for the stacked SEDs with  $[4.5 \mu\text{m}] < 23.5$  AB is  $\sim 24.7$  AB, which corresponds at  $z \sim 4$  to an absolute magnitude  $M_{UV,1600} \sim -21.4$ . According to the above relation, the stellar mass corresponding to  $M_{z'} \sim -23$  mag is  $\log(M_*/M_\odot) \sim 10.3$ . This behavior raises concern about potential biases that can occur when adopting the UV luminosity and  $M_*/L_{UV}$  in the measurement of stellar masses, particularly for massive galaxies.

A constant  $M_*/L$  is equivalent to a slope of  $-0.4$  in the  $\log$  (stellar mass)—absolute magnitude plane. Our result at  $z \sim 4$ , obtained for galaxies with  $M_{z'} < -23$  mag, is consistent with the  $\sim -0.4$  slopes found in the stellar mass— $M_{UV}$  plane (see, e.g., Duncan et al. 2014; Grazian et al. 2015). Steeper slopes,

such as those found by Stark et al. (2009), González et al. (2011), Lee et al. (2011), McLure et al. (2011), and Song et al. (2016), require the  $\log(M_*/L)$  to decrease for fainter galaxies, increase for brighter galaxies, or a combination of both effects. The origin for this is still unclear, as it could be a mix between selection effects (see, e.g., Grazian et al. 2015) and nebular emission contamination, which could boost the stellar masses of the more luminous galaxies (e.g., Song et al. 2016).

In the bottom panel of Figure 9, we compare the measurements of stellar mass and luminosity recovered from our stacking analysis to a *control* sample. This sample is composed of individual measurements selected from the COSMOS/UltraVISTA and CANDELS/GOODS-S catalogs to have photometric redshifts  $3.5 < z_{\text{phot}} < 4.5$ .

The individual measurements of the control sample distribute according to a monotonic relation defining a *main sequence*, with a scatter of about  $\sim 0.7$  dex. This correlation holds over a wide range of values, both in stellar mass and luminosity. In particular, we notice the absence of any turnover, as instead is observed when considering  $M_{UV}-M_*$ , e.g., Spitzer et al. (2014) or equivalently, our Figure 6. Furthermore, there are virtually no measurements outside the main sequence.

The relation between the stellar mass and luminosity recovered from the stacking analysis is in excellent agreement with the values of the control sample. We remark here that the control sample was selected based exclusively on photometric redshifts criteria. The agreement between our stacking results in the GOODS field, then, indicates that the stacking does not suffer any major bias from the LBG selection. This result is not unexpected, however, given the low fraction of objects missed by the LBG selection for stellar masses  $\log(M_*/M_\odot) \lesssim 10.2$ , as we showed in Section 2.3.

Together, these two results increase our confidence in the reliability of the  $z'$  band as a proxy for the stellar mass, as well as the robustness of our stacking analysis.

### 3.3. Evolution of the $z'$ -band Luminosity Function

The LFs were measured adopting the  $1/V_{\text{max}}$  estimator (Schmidt 1968). Although this method is intrinsically sensitive to local overdensities of galaxies, the clustering is expected to be negligible at  $z > 4$ . On the other hand, the  $1/V_{\text{max}}$  method directly provides the normalization of the LF. Furthermore, and most importantly, the *coherent analysis* extension developed by Avni & Bahcall (1980) is key to this work.

As we showed in Section 2.1, our composite sample is based on a dual-band flux selection, corresponding to a double flux threshold. The detection process introduces the first flux cut in the corresponding band ( $K_s$  or  $\chi^2$  image built from the *HST* NIR bands, for the UltraVISTA and GOODS-N/S sample, respectively). The S/N cut on the flux in the IRAC band closest to the rest-frame  $z'$  is responsible for the second flux threshold in the relevant IRAC band.

For each galaxy in the sample, each flux threshold generates an upper limit to the redshift the specific galaxy can have and still be included in the sample. These different upper limits in redshift correspond to different comoving volumes for each object that could potentially enter the  $V_{\text{max}}$  computation. The coherent approach allowed us to take this double selection into consideration in a consistent way: the upper limit in redshift, used to compute the comoving volume, was taken to be the smaller of the two redshift upper limits, computed based on the threshold in the corresponding selection band. Furthermore, as

we showed in Figure 1, the depth of the IRAC mosaics is highly inhomogeneous. Therefore, for the computation of the comoving volumes in each field, we divided the IRAC footprint into a number of sub-fields, such that each sub-field was characterized by nearly homogeneous depths in both the detection<sup>10</sup> and relevant IRAC band. Again, the Avni & Bahcall (1980) prescription allowed us to coherently analyze the different sub-samples.

Comoving volumes were computed differently, depending on the field and on the band driving the selection. For the galaxies in the GOODS-N/S fields, we used the comoving volumes computed by Bouwens et al. (2015). These volumes were estimated using an extensive Monte Carlo simulation based on real data. Sources were added to the different mosaics and recovered following the same procedure applied for the assembly of the LBG sample. Such volume estimates natively take into account the selection effects at the detection stage, and correct for flux-boosting effects and contamination by lower redshift interlopers and brown dwarfs. The volumes  $V_i$  for those objects  $i$  in the GOODS-N/S fields whose redshift upper limit  $z_{\text{up}}$  was driven by the IRAC S/N threshold ( $z_{\text{up}} \equiv z_{\text{up,IRAC}}$ ) were rescaled by the ratio between the volume associated with the redshift upper limit of the IRAC band  $V_i(z_{\text{up,IRAC}})$  and that of the  $\chi^2$  image  $V_i(z_{\text{up},\chi^2})$ :

$$V_{i,\text{IRAC}} = V_{i,\text{GOODS}} \times \frac{V_i(z_{\text{up,IRAC}})}{V_i(z_{\text{up},\chi^2})}. \quad (7)$$

For the UltraVISTA sample, the volumes were computed directly from the limits in redshift that correspond to the flux limits in the  $K_s$  and  $4.5 \mu\text{m}$  bands.

We computed the LF in four redshift bins centered at  $z \sim 4$ ,  $z \sim 5$ ,  $z \sim 6$ , and  $z \sim 7$ . Although the IRAC data potentially allowed us to consider galaxies at  $z \sim 8$ , we did not find any candidate with reliable flux measurement in the IRAC 5.8 and  $8.0 \mu\text{m}$ . Uncertainties on the LF measurements were derived by combining in quadrature the Poisson noise in the approximation of Gehrels (1986) to an estimate of cosmic variance from the recipe of Moster et al. (2011). The average cosmic variance value obtained for the  $z \sim 4$  UltraVISTA sample was  $\sim 0.43$ ; the average cosmic variance estimates for the GOODS-N/S sample were  $\sim 0.27$ ,  $\sim 0.41$ ,  $\sim 0.58$ , and  $\sim 0.80$ , respectively, for the  $z \sim 4$ , 5, 6, and  $z \sim 7$  redshift bins. The high values of the cosmic variance registered for all redshifts and luminosities are the dominant source of stochastic uncertainties in our LF measurements.

Our LF measurements are presented in Figure 10 and Table 1. The LF at  $z \sim 7$  consists of a single-bin measurement, and is characterized by large uncertainties that do not allow us to properly constrain its shape. The absolute magnitude range of the  $z \sim 4$  LF spans  $\sim 5$  magnitudes,  $\gtrsim 3\times$  more than the magnitude range of the  $z \geq 5$  LFs. The larger absolute magnitude range available at  $z \sim 4$  is the result of a number of distinct factors. First, the increased depth in the  $4.5 \mu\text{m}$  band from the combination of *Spitzer*/IRAC cryogenic and post-cryogenic epochs enables us to reach fainter absolute magnitudes than the cryogenic  $5.8 + 8.0 \mu\text{m}$  data alone at  $z \geq 5$ . Second, the smaller PSF size of the  $4.5 \mu\text{m}$  data, compared to the  $5.8 \mu\text{m}$  and  $8.0 \mu\text{m}$  bands, allows to reach

fainter fluxes for the same exposure time and detector efficiency. Third, the availability of COSMOS/UltraVISTA data over an area  $\sim 4\times$  larger than the GOODS-N/S footprint allowed us to recover the exponential decline of the bright end of the  $z \sim 4$  LF, which is otherwise inaccessible by the small footprint of the GOODS-N/S mosaics.

In order to verify the consistency of the  $z \sim 4$  LF with respect to the GOODS-N/S and UltraVISTA data, we also computed the LF separately on each one of these two data sets. The resulting LFs are marked in Figure 10 by the open squares, and show a good agreement with the LF from the composite sample.

The large uncertainties associated to the number density measurements at  $z \sim 5$ , 6, and 7 do not allow us to disentangle whether the evolution is in luminosity, number density, or both. In Section 3.5, we attempt to analyze this in a more quantitative way.

The dashed gray curve in Figure 10 marks the  $z \sim 0$  LF of Kelvin et al. (2014), measured with data from the Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2009). Compared to our lowest redshift LF measurement, the  $z \sim 0$  LF is characterized by a steeper decay for  $M_{z'} \lesssim -24.5$ . Fully understanding the evolution of the LF from  $z \sim 4$  to  $z \sim 0$  goes beyond the scope of the present work. However, qualitatively, a decrease in luminosity at  $z \sim 0$  compared to  $z \sim 4$  is expected, considering the lower values of the star formation rate density at  $z \sim 0$  than at  $z \sim 4$  (e.g., Madau & Dickinson 2014), and that the  $z'$  band may retain the effects of the recent SFH. This is particularly true for the bright-end of the LF: indeed, the brightest (i.e., most massive) galaxies that are still forming stars at  $z \sim 4$  are likely to become quenched by  $z \sim 0$  (e.g., Muzzin et al. 2013b).

### 3.4. Evolution of the SMF

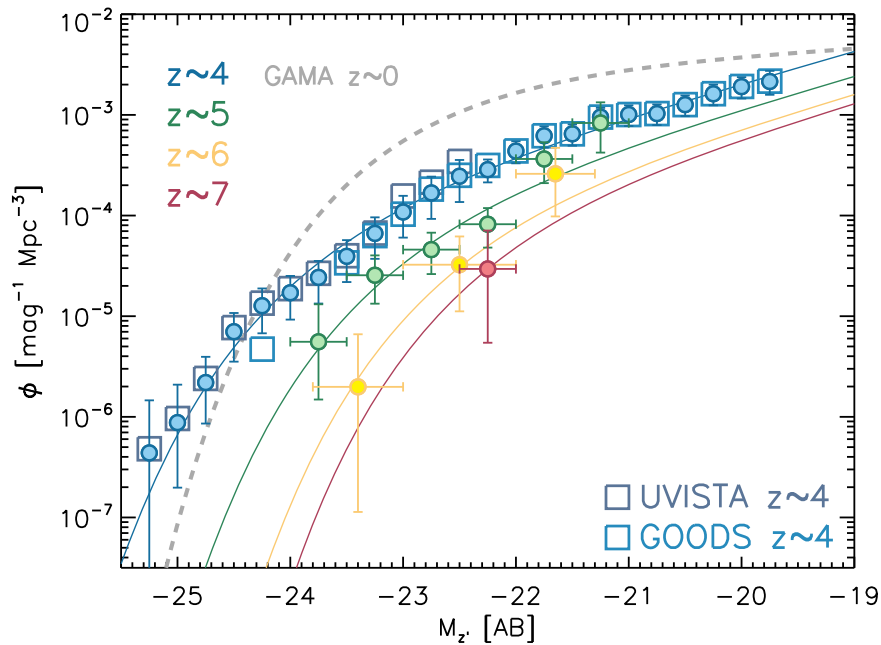
We generated SMF measurements by taking advantage of the mass-to-light ratios we measured in Section 3.2. The fact that  $M_*/L_{z'}$  does not decrease with luminosity makes the shape of the LF resemble that of the SMF, allowing us to attempt a simple and straightforward conversion of the LF into the SMF. Other  $M/L$  relations allow for the recovery of the SMF from the LF, although in a less straightforward way. We further discuss this in Section 4.1.

We adopt the following very simple procedure. We assume that the constant  $\log(M_*/L_{z'})$  and the linear relation observed at  $z \sim 4$  (Equations (5) and (6)) are valid at all redshifts. The absolute magnitudes corresponding to the bin centers of the LFs are converted into stellar mass by applying the relevant  $\log(M_*/L_{z'})$  relation, depending on the  $M_{z'}$  value (see Equations (5) and (6)). We then differentiate the two relations, solving for  $dM_*$ . The obtained values, specific for each  $M_*$  bin are used to rescale the LF normalization, to take into account the change in units from  $\text{mag}^{-1}$  to  $\text{dex}^{-1}$ .

Our SMF measurements are presented in Figure 11 and Table 2. Unsurprisingly, the  $z \sim 4$  SMF covers a range in stellar mass wider than the  $z \sim 5$ , 6, and 7 SMFs, for the same reasons we described for the LF.

In Figure 11, we also plot a compilation of SMF measurements from the literature (Pérez-González et al. 2008; Marchesini et al. 2009, 2010; Stark et al. 2009; González et al. 2011; Lee et al. 2012; Santini et al. 2012; Ilbert et al. 2013; Muzzin et al. 2013b; Duncan et al. 2014; Stefanon et al. 2015; Caputi et al. 2015; Grazian et al. 2015; Song et al. 2016; Davidzon et al. 2017). At  $z \sim 4$ , starting from the low-mass

<sup>10</sup> Because it is not straightforward to associate a limiting magnitude with a  $\chi^2$  image from the combination of different filters, we considered WFC3/ $H_{160}$  to be the relevant band for the depth of the detection in the GOODS-N/S fields.



**Figure 10.** Color-filled circles mark our measurements of the  $1/V_{\max}$  LF in the four redshift bins, as detailed by the legend in the top-left corner. Error bars include the contribution from Poisson noise and cosmic variance. For the  $z \sim 4$  bin, we also present the individual LF from UltraVISTA (darker blue open squares) and GOODS-N/S (lighter blue open squares). These two latter measurements are consistent with each other where they overlap, and with the LF from the composite  $z \sim 4$  sample. For ease of representation, we omit the uncertainties of the individual UltraVISTA and GOODS-N/S LFs. The colored solid curves mark the best-fit Schechter functions at the corresponding redshift. The gray dashed line represents the  $z \sim 0$  LF from GAMA (Kelvin et al. 2014).

end where the measurements are generally quite consistent with each other, the discrepancies increase with increasing stellar mass. One possible reason for the increased dispersion at higher masses is that galaxies constituting the low-mass end are mostly star-forming. Their redshift can then be assessed through the location of the observed Lyman break (either from dropouts or photometric redshift selections). The massive end, instead, possibly also includes more evolved and/or dusty systems, and it is therefore more sensitive to the degeneracy in identifying the observed break as either the Balmer/4000 Å break or the Lyman break.

Our  $z \sim 4$  SMF determination is in good agreement with the SMFs of Stark et al. (2009), Lee et al. (2011), Stefanon et al. (2015), Caputi et al. (2015), Grazian et al. (2015), Song et al. (2016), and Davidzon et al. (2017). This is quite remarkable, because these SMFs have been recovered from different selection techniques. Specifically, Stark et al. (2009) and Lee et al. (2011) have measurements based on dropouts samples from *HST*/WFC3 data; the SMF of Grazian et al. (2015) was built from a  $H_{160}$ -detected photometric redshift sample over the CANDELS/GOODS-S and CANDELS/UDS fields; Stefanon et al. (2015) assembled a composite sample that complements a  $K_s$ -detected catalog from UltraVISTA data with detections in IRAC 3.6 and 4.5  $\mu\text{m}$  bands; Caputi et al. (2015) use measurements based on a  $K_s$ -detected SMF complemented by SMF measurements from detections in IRAC 4.5  $\mu\text{m}$ . The sample selection of both Stefanon et al. (2015) and Caputi et al. (2015) relies on photometric-redshift measurements. Song et al. (2016) recovered the SMF measurements by converting the UV LF into SMF through a linear  $M_*-M_{\text{UV}}$  relation for  $M_* < 10^{10} M_{\odot}$ , complemented by bootstrapped estimates at  $M_* > 10^{10} M_{\odot}$ . Finally, the SMF of Davidzon et al. (2017) was based on a photometric-redshift sample from  $K_s$ -detection in COSMOS/UltraVISTA DR2 mosaics. On the other side, the

normalization of our SMF is higher than González et al. (2011) SMFs; these measurements were obtained by converting the observed UV LF into SMF through  $M_*/L_{\text{UV}}$  measurements. The discrepancy between our SMF (and the bulk of the other SMF determinations) could be due to a steeper  $M_*/L_{\text{UV}}$  relation found by González et al. (2011), and consequent lower normalization term.

At the massive end ( $\log(M_*/M_{\odot}) \gtrsim 11.2$ ), we observe a discrepancy between our  $z \sim 4$  SMF and some of the corresponding measurements from the literature (e.g., Ilbert et al. 2013; Muzzin et al. 2013b). This discrepancy could, at least in part, be explained by our SMF lacking any scatter in  $M_*/L_{z'}$  for a given  $L_{z'}$ . Our stacking analysis, by construction, recovers median  $M_*/L_{z'}$  ratios, potentially excluding such extreme cases as very dusty/old systems with very high stellar masses. However, the bottom panel of Figure 6, and our discussion in Section 2.3, show that our sample is not strongly biased against this class of objects in the limits of current data. Nonetheless, they do not allow us to properly ascertain their existence at lower stellar masses. Aside from that, because of the Eddington bias, a distribution in the observed  $M_*/L_{z'}$  values for a specific luminosity would introduce a higher fraction of lower stellar mass objects that are scattered to higher stellar masses than the opposite, increasing the number density of the massive objects. One additional potential reason for this discrepancy is that the current and previous estimates of photometric redshifts do not agree. Compared to Muzzin et al. (2013b) or Ilbert et al. (2013), the DR2 version of the UltraVISTA catalog benefits from deeper NIR and IRAC data, providing improved photometric redshift constraints. The agreement between our estimates and those of Grazian et al. (2015) and Davidzon et al. (2017), both obtained from photometric redshift samples, is noteworthy—suggesting that



**Table 1**  
 $V_{\max}$  Measurements of the Luminosity Functions

$z$ bin	$M_z'$ (mag)	$\Delta M_z'$ (mag)	$\Phi$ ( $10^{-5} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )	# gal.
$z \sim 4$	-25.25	0.25	$0.044^{+0.102}_{-0.041}$	1
	-25.00	0.25	$0.088^{+0.121}_{-0.068}$	2
	-24.75	0.25	$0.22^{+0.18}_{-0.13}$	5
	-24.50	0.25	$0.70^{+0.37}_{-0.35}$	16
	-24.25	0.25	$1.27^{+0.62}_{-0.59}$	29
	-24.00	0.25	$1.71^{+0.80}_{-0.78}$	39
	-23.75	0.25	$2.4^{+1.1}_{-1.1}$	55
	-23.50	0.25	$3.9^{+1.8}_{-1.7}$	78
	-23.25	0.25	$6.6^{+3.0}_{-2.9}$	100
	-23.00	0.25	$10.8^{+4.8}_{-4.8}$	86
	-22.75	0.25	$16.8^{+7.6}_{-7.3}$	64
	-22.50	0.25	$25^{+11}_{-11}$	72
	-22.25	0.25	$28.6^{+7.5}_{-7.2}$	64
	-22.00	0.25	$44^{+11}_{-11}$	87
	-21.75	0.25	$62^{+15}_{-15}$	123
	-21.50	0.25	$65^{+16}_{-15}$	128
	-21.25	0.25	$95^{+22}_{-22}$	186
-21.00	0.25	$101^{+24}_{-23}$	194	
-20.75	0.25	$103^{+24}_{-24}$	196	
-20.50	0.25	$126^{+29}_{-29}$	207	
-20.25	0.25	$162^{+38}_{-37}$	193	
-20.00	0.25	$191^{+46}_{-45}$	119	
-19.75	0.25	$215^{+59}_{-56}$	49	
$z \sim 5$	-23.75	0.50	$0.56^{+0.76}_{-0.41}$	2
	-23.25	0.50	$2.6^{+1.5}_{-1.2}$	9
	-22.75	0.50	$4.6^{+2.2}_{-2.0}$	16
	-22.25	0.50	$8.2^{+3.7}_{-3.4}$	20
	-21.75	0.50	$36^{+17}_{-15}$	17
	-21.25	0.50	$83^{+50}_{-41}$	8
$z \sim 6$	-23.40	0.80	$0.20^{+0.46}_{-0.19}$	1
	-22.50	1.00	$3.2^{+2.9}_{-2.1}$	4
	-21.65	0.70	$26^{+21}_{-16}$	5
$z \sim 7$	-22.25	0.50	$2.9^{+4.1}_{-2.4}$	2

our composite sample suffers from small selection bias, even at higher masses.

At  $z \sim 5$ , the measurements of the low-mass end ( $\log(M_*/M_\odot) \sim 9\text{--}9.5$ ) of the SMF are characterized by a larger scatter than for  $\log(M_*/M_\odot) \sim 10\text{--}10.5$ . At higher stellar masses, measurements are broadly consistent with each other—mainly because of the large uncertainties on the number densities. Our SMF determination overlaps with the measurements of Stark et al. (2009), González et al. (2011), Lee et al. (2011), Grazian et al. (2015), Song et al. (2016), and Davidzon et al. (2017). However, it lies below the SMF measurements of Duncan et al. (2014).

At even higher redshift, the Poisson uncertainties on the number densities start to be of the same order as those from cosmic variance. The small number of galaxies at  $z \sim 6$  and  $z \sim 7$  (10 and 2 galaxies, respectively) generates large Poisson uncertainties, resulting in a broad agreement among the different SMF determinations. Our SMF overlaps with the measurements of Stark et al. (2009), González et al. (2011), Grazian et al. (2015), and Song et al. (2016). The large uncertainties also make our SMF measurements roughly consistent with those of Duncan et al. (2014).

### 3.5. Schechter Best-fit to the LF and SMF

The shape of the LF and SMF of galaxies at high redshift is usually well-described by a Schechter (1976) function over a wide range of redshifts ( $2 \lesssim z \lesssim 10$ , but see also, e.g., Bowler et al. 2014). We therefore fitted our LF measurements with a Schechter function:

$$\Phi(M) = \Phi^* \frac{\ln(10)}{2.5} 10^{-0.4(M-\mathcal{M}^*)(\alpha+1)} \exp(-10^{-0.4(M-\mathcal{M}^*)}), \quad (8)$$

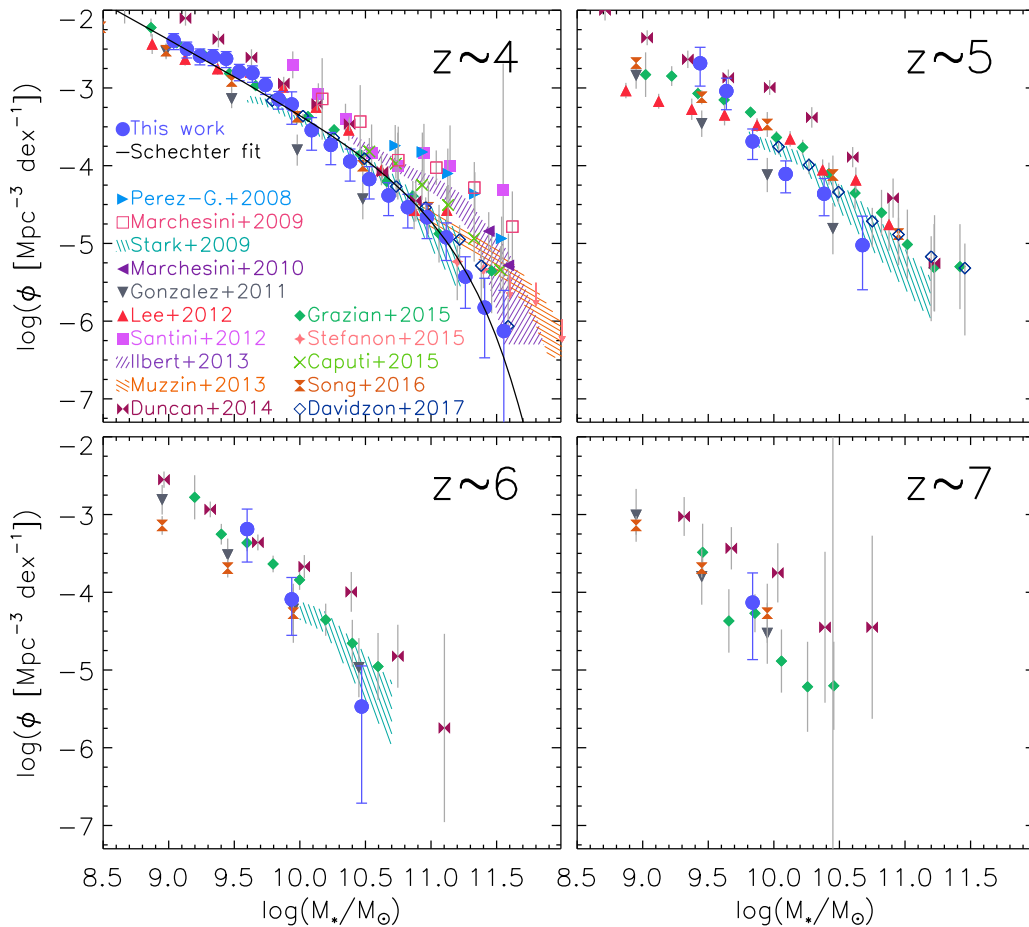
where  $\mathcal{M}^*$  is the characteristic magnitude, corresponding to the knee of the density distribution. Here,  $\alpha$  is the faint end slope and  $\Phi^*$  is a global normalization factor. The single-bin measurement at  $z \sim 7$  does not allow us to properly inspect the shape of the LF. However, LF estimates at similar redshift, although in different bands, show that it is reasonably well-determined and consistent with a single Schechter form (Bouwens et al. 2015; Finkelstein et al. 2015a), or a double power law (Bowler et al. 2014, 2017).

We performed a best fit to our LFs' measurements, using the Levenberg-Marquardt method. For the  $z \sim 4$  LF, we left the three Schechter parameters free to vary. Because the faint-end slope of the  $z \geq 5$  LFs is poorly sampled, the best fits to the  $z \geq 5$  LFs were done by fixing  $\alpha$  to the value obtained for the  $z \sim 4$  LF. The fit to the  $z \sim 5$  and  $z \sim 6$  LFs were done in three different configurations: both  $\mathcal{M}^*$  or  $\Phi^*$  as free parameters, with  $\mathcal{M}^*$  as unique free parameter, or with  $\Phi^*$  as the only free parameter. We did not fit the  $z \sim 7$  LF, leaving both  $\mathcal{M}^*$  or  $\Phi^*$  as free parameters. The resulting best-fit Schechter functions for the case of pure luminosity evolution are marked in Figure 10 with solid curves. Table 3 lists the recovered parameters and their uncertainties.

Visual inspection of the best-fitting Schechter function shows that, overall, there is a preference for a pure luminosity evolution against a pure density evolution. Moreover, when both  $\mathcal{M}^*$  and  $\Phi^*$  were left free to vary, the values of  $\Phi^*$  were characterized by large uncertainties, making them consistent with the value of  $\Phi^*$  at  $z \sim 4$ , i.e., corresponding to no evolution with redshift. On the contrary, the values of  $\mathcal{M}^*$  showed a more clear trend with redshift, increasing our confidence in the luminosity evolution. We note, however, that these results potentially suffer from the limited coverage of  $\mathcal{M}^*$ .

The luminosity evolution registered here is in contrast with the most recent measurements of the evolution of the UV LF at  $z > 4$ , where indication is found that the characteristic magnitude evolves very little with redshift (likely constrained by the impact of dust extinction at high masses—Bouwens et al. 2009; Reddy et al. 2010), as most of the evolution seems to be driven by variation in the overall density (see, e.g., van der Burg et al. 2010; Bouwens et al. 2015; Finkelstein et al. 2015a).

The LFs presented in our work are the first determination of the rest-frame optical LF at  $z \gtrsim 4$ , and therefore direct comparisons to previous estimates are not possible. However, recent works have recovered the measurement of the characteristic magnitude in the rest-frame  $z'$ -band at  $z \sim 4$  (Oesch et al. 2013), or have studied the evolution of the LF up to  $z \lesssim 4$  in rest-frame optical bands close to  $z'$  (Marchesini et al. 2012; Stefanon & Marchesini 2013). In the following paragraphs, we will compare our determination of the  $z \sim 4$  characteristic magnitude to the estimates from the above three works.



**Figure 11.** Our estimates for the SMF are marked by the filled blue circle with error bars. Top to bottom, left to right, the panels present the SMF at  $z \sim 4, 5, 6,$  and  $7,$  respectively. Measurements of the SMF from the literature are also shown, with symbols following the legend in the top-left panel. The same plotting conventions are applied to all panels. Our SMFs are in good agreement with previous determinations.

Oesch et al. (2013) estimated the characteristic magnitude of the  $z \sim 4$  LF by applying a correction based on the average  $i_{775} - [4.5]$  color to the characteristic magnitude of the  $z \sim 4$  UV LF. The obtained value,  $\mathcal{M}_z^* = -21.7$  AB, is  $\sim 1.7$  mag fainter than what we found in this work. A possible reason for this large discrepancy is the lack of galaxies brighter than  $M_z < -23$  from the sample of Oesch et al. (2013), which instead have become accessible through the deep and wide area UltraVISTA data.

Marchesini et al. (2012) presented estimates of the rest-frame V-band LF at  $z \lesssim 4$ , obtained from a composite sample that includes wide-area data from the NMBS, FIRES, FIREWORKS, HDFN, HUDF, and GOODS-S programs. The characteristic magnitude recovered from a maximum likelihood analysis is  $\mathcal{M}_V^* = -22.76^{+0.40}_{-0.63}$ . This value is brighter than that found by Oesch et al. (2013), although it is still  $\sim 0.6$  mag fainter than the estimate in our work. Nonetheless, it is consistent at  $1\sigma$  with our estimate, considering the associated uncertainties. Our stacking analysis (see Figure 8) shows that the brighter galaxies have redder ( $V - z'$ ) colors, reaching  $(V - z') \sim 1$  mag for the brightest stacks. Using our stacked SED at  $M_{z'} \sim -23.5$ , close to the value of the characteristic magnitude of our  $z \sim 4$  LF, we find a rest-frame color  $(V - z') = +0.53$  mag; applied to the  $\mathcal{M}_V^*$ , it gives a  $\mathcal{M}_{z'}^* = -23.29$  mag, very close to our best-fit  $\mathcal{M}_{z'}^* = -23.38$  mag.

Stefanon & Marchesini (2013) measured the evolution of the rest-frame  $J$ - and  $H$ -band LF up to  $z \sim 3.5$  by using a

composite sample of galaxies from the MUSYC, FIRES, and GOODS-CDFS programs. From a maximum likelihood analysis, the characteristic magnitude of the  $z \sim 3.25$  rest-frame  $J$ -band LF was estimated to be  $\mathcal{M}_J^* = -23.28^{+0.33}_{-0.29}$  mag. Applying the same analysis adopted for the comparison to Marchesini et al. (2012), we find a rest-frame color  $(z' - J) = +0.11$  mag, and a corresponding  $\mathcal{M}_{z'}^* = -23.17$  mag, still consistent with our estimate at  $1\sigma$  level.

We also performed a Schechter fit to the SMFs. However, the differential  $M_*/L_{z'}$  we measured at  $z \sim 4$  and applied to the  $z \geq 5$  LFs has the effect of *stretching* the original LFs. The consequence of this stretching, along with the limited number of measurements available at each redshift, is that the  $z \geq 5$  SMFs do not show any robust evidence for the exponential cut at the massive end, preventing any reliable estimate of the characteristic stellar mass. A similar result was found by Grazian et al. (2015), even though the issue is not yet settled (see, e.g., Caputi et al. 2015). Therefore, we performed the Schechter fit only on the  $z \sim 4$  SMF, and obtained the following results:  $\alpha = -1.93 \pm 0.24$ ,  $\Phi_*^* = (2.72 \pm 1.81) \times 10^{-5} \text{ Mpc}^{-3} \text{ dex}^{-1}$  and  $\log(M_*/M_\odot) = 10.96 \pm 0.33$ . The corresponding best-fit Schechter function is presented in Figure 11.

The large uncertainties associated with the Schechter parameters make them consistent with most of the measurements from the literature. The low number of massive galaxies from the exponential part of the Schechter function, at the massive end, suffer from high uncertainties from both (relative)

**Table 2**  
Stellar Mass Function Measurements

$z$ bin	Stellar mass $\log(M_*/M_\odot)$	$\Phi$ ( $10^{-5} \text{ Mpc}^{-3} \text{ dex}^{-1}$ )
$z \sim 4$	11.55	$0.075^{+0.174}_{-0.070}$
	11.41	$0.15^{+0.21}_{-0.12}$
	11.26	$0.37^{+0.30}_{-0.23}$
	11.12	$1.20^{+0.64}_{-0.59}$
	10.97	$2.2^{+1.0}_{-1.0}$
	10.82	$2.9^{+1.4}_{-1.3}$
	10.68	$4.1^{+1.9}_{-1.9}$
	10.53	$6.7^{+3.0}_{-3.0}$
	10.38	$11.3^{+5.0}_{-5.0}$
	10.24	$18.5^{+8.3}_{-8.2}$
	10.09	$29^{+13}_{-13}$
	9.94	$61^{+28}_{-27}$
	9.84	$71^{+19}_{-18}$
	9.74	$109^{+27}_{-27}$
	9.64	$155^{+37}_{-37}$
	9.54	$162^{+39}_{-38}$
	9.44	$238^{+56}_{-55}$
9.34	$252^{+59}_{-58}$	
9.24	$258^{+60}_{-60}$	
9.14	$316^{+73}_{-73}$	
9.04	$405^{+94}_{-94}$	
8.94	$477^{+115}_{-114}$	
8.84	$537^{+147}_{-141}$	
$z \sim 5$	10.68	$0.95^{+1.29}_{-0.70}$
	10.38	$4.4^{+2.5}_{-2.1}$
	10.09	$7.8^{+3.7}_{-3.4}$
	9.84	$20.5^{+9.2}_{-8.5}$
	9.64	$91^{+42}_{-39}$
9.44	$208^{+125}_{-102}$	
$z \sim 6$	10.47	$0.34^{+0.79}_{-0.32}$
	9.94	$8.1^{+7.4}_{-5.3}$
	9.60	$65^{+53}_{-40}$
$z \sim 7$	9.84	$7.4^{+10.4}_{-6.0}$

**Table 3**  
Luminosity Function Best-fit Schechter Parameters

Redshift bin	$\Phi^*$ ( $10^{-5} \text{ Mpc}^{-3} \text{ mag}^{-1}$ )	$\mathcal{M}^*$ (mag)	$\alpha$
4	$19.4 \pm 7.1$	$-23.38 \pm 0.19$	$-1.79 \pm 0.09$
5	$9.1 \pm 7.7$	$-22.99 \pm 0.51$	$-1.79$
	19.4	$-22.62 \pm 0.12$	$-1.79$
	$6.85 \pm 1.36$	$-23.38$	$-1.79$
6	$51.1 \pm 36.0$	$-21.77 \pm 0.42$	$-1.79$
	19.4	$-22.09 \pm 0.17$	$-1.79$
	$8.59 \pm 5.40$	$-23.38$	$-1.79$
7	19.4	$-21.82 \pm 0.39$	$-1.79$
	$1.99 \pm 1.62$	$-23.38$	$-1.79$

Poisson noise and cosmic variance. The massive end of our SMF is based on the UltraVISTA sample, suggesting a more adequate coverage of massive galaxies.

## 4. Discussion

### 4.1. Rest-frame Optical LF versus SMF Measurements

One of the main aims of this paper is to test the LF in rest-frame optical bands, as a proxy for SMF measurements at high redshifts ( $z \gtrsim 4$ ). The compilation of SMF measurements presented in Figure 11 shows that, for  $z \geq 4$ , they are characterized by a scatter that can be as large as  $\gtrsim 1$  dex. This has immediate consequences for our understanding of more fundamental and global properties, like the evolution with cosmic time of the stellar mass density or the stellar-to-halo mass relation. The likely main reason for this large scatter is sample selection. In addition to this, further systematics may arise from our limited knowledge of some of the specific aspects characterizing the stellar population of each galaxy, including contamination from emission lines (see, e.g., Conroy & Wechsler 2009; Stark et al. 2009; Behroozi et al. 2010).

Reliable stellar mass measurements require coverage of rest-frame optical/NIR wavelengths, probed by *Spitzer*/IRAC at  $z \gtrsim 4$ . In particular, current measurements of the low-mass end of the  $z > 4$  SMFs rely on IRAC data over the GOODS-N/S fields, as these are the only fields with sufficient depth that provide photometric coverage over a sufficient area. The 3.6 and 4.5  $\mu\text{m}$  bands, although characterized by deeper data than the 5.8 and 8.0  $\mu\text{m}$  bands at  $z \gtrsim 4$ , are potentially contaminated by nebular lines. Previous work has attempted to estimate the impact that nebular emission can have on the recovery of stellar mass, yet no concordance has been found so far. Specifically, estimates based on relations between the EW of H $\beta$  and the ionizing properties of the best-fit SEDs generally predict systematics  $\lesssim 0.2$  dex, and are nearly independent of redshift (e.g., Duncan et al. 2014; Grazian et al. 2015; Salmon et al. 2015). On the other side, estimates based on the measurements of the evolution of the H $\alpha$  EW with redshift indicate an increasing contribution of nebular lines to the stellar mass measurements, ranging from  $\sim 0$  dex at  $z \sim 4$ , to  $z \sim 0.6$  dex at  $z \sim 7$ , as seen in Stark et al. 2013. Those authors presented, among other estimates, an extrapolation to  $z \sim 7$ , later supported by other observations (e.g., Smit et al. 2014; Faisst et al. 2016; Rasappu et al. 2016). Such systematics in the measurement of stellar mass introduce up to  $\sim 0.5$  dex offset in the number densities for the higher redshift bins. To further complicate the picture, recent works have shown that, when the correction for nebular line contamination is applied on a statistical basis, irrespective of the specific SEDs, it can even *boost* the stellar mass measurements by up to  $\sim 0.2$ – $0.3$  dex (Stefanon et al. 2015; Nayyeri et al. 2017; Stefanon et al. 2017).

One possible way to circumvent this problem is estimating the stellar mass exclusively from the 5.8 and 8.0  $\mu\text{m}$  bands. These bands cover a region of the SEDs of  $z \gtrsim 4$  galaxies free from contamination by strong nebular lines. However, the larger PSF FWHM of *Spitzer*/IRAC 5.8 and 8.0  $\mu\text{m}$  bands, compared to the 3.6 and 4.5  $\mu\text{m}$  bands, may introduce blending effects in the flux measurements. Labbé et al. (2015) have shown that *mophongo*, the software we adopted to measure the fluxes in the IRAC bands, does not introduce any substantial systematics in the flux measurement, even in very crowded regions. Nonetheless, given the larger FWHM and the lower S/N characterizing the 5.8  $\mu\text{m}$  and 8.0  $\mu\text{m}$  data, one could expect an increased scatter in the flux measurements, compared to the bluer IRAC bands.



Through the  $M_*/L_{z'}$  we derived in the present work, the S/N cuts we applied to the flux in the IRAC 5.8 and 8.0  $\mu\text{m}$  bands identify a range in stellar mass where their measurements can be considered reliable. The LF and SMF from this work can thus be regarded as a indicative of most of the current SMF measurements at  $z \geq 4$ , as they are based on the subsample of objects with the highest S/N measurements in those bands more sensitive to the stellar mass and with reduced contamination from nebular emission. This is visible in Figure 11; at  $z \sim 5$  and above the lowest stellar mass over which our SMFs are defined, we see  $S/N \approx 5\text{--}10\times$  higher than most current SMF determinations. Specifically, this also means that stellar masses below our low-mass limits are necessarily based on very low S/N measurements in the IRAC 5.8 and 8.0  $\mu\text{m}$  bands, (still uncertain) correction for nebular emission contamination, or a combination of the two.

Our analysis showed that SMFs consistent with the average determinations from the literature could be recovered by applying a simple  $M_*/L_{z'}$  relation to the observed  $z'$ -band LF, with stellar masses measured from common stellar population parameters (e.g., delayed exponential SFH, solar metallicity, Chabrier IMF). Our simple transformation of the LF into SMF was supported by the non-decreasing  $M_*/L_{z'}$  ratio for increasing luminosities observed at  $z \sim 4$ . Relations between the  $M_*/L$  and  $L$  other than that (e.g., if the  $M_*/L$  presented a minimum for some value of  $L$ ) would still allow the conversion, but would require us to consider, for specific bins of  $M_*$ , the contributions to the density originating from different bins of  $L$ . A non-decreasing  $M_*/L_{z'}$ , instead, constitutes an injective mapping between luminosity and stellar mass; galaxies with higher luminosity will always have higher stellar mass, and galaxies with lower luminosity will always have lower stellar mass.

A different scenario arises when multiple  $M_*/L$  exist in correspondence with a single value of  $L$ , as can be the case, for instance, for UV luminosity versus stellar mass. Figure 6 (along with the bottom panel of Figure 9) shows that, for  $M_{\text{UV}} \gtrsim -20$  mag, there are broadly two very different  $M_*/L_{\text{UV}}$  values. More specifically, this means that, if galaxies with a higher  $M_*/L_{\text{UV}}$  ratio are included in the sample, those bins of luminosity will include (and mix) the contribution from both high- and low-mass galaxies. If this effect is not properly taken into account, it introduces an overestimate of the low-mass-end slope and an underestimate of the massive end of the SMF. The only way to deal with this problem is to directly count the number of galaxies in each of the two  $M_*/L$  bins.

Finally, the ideal rest-frame band for this kind of study is probably one for which the  $M_*/L$  ratio would not depend on the luminosity, as any luminosity dependence could potentially hide effects from, e.g., SFH. Our measurements of the  $M_*/L_{z'}$  relation suggest that, for a wide range in luminosity, they are consistent with a constant value. A log-linear relation arises at the bright (massive) end of the LF (SMF), suggesting that the  $z'$ -band at these luminosities holds the signature of the stellar population age and/or the dust content, at some level.

#### 4.2. Tracking the Assembly of DM Halos through the Evolution of the Rest-frame Optical LF

Given the potential systematics on the SMF measurements discussed above, LF estimates can provide a valid alternative for recovering the halo masses ( $M_h$ ) for high-redshift galaxies through abundance matching techniques (e.g., Behroozi et al. 2013b;

Finkelstein et al. 2015b; Steinhardt et al. 2016). To date, measurements of  $z \geq 4$  LFs are mostly available in the rest-frame UV. The adoption of UV LFs in the  $M_h/L$  estimates provides information on the relative importance of star formation processes (e.g., gas cooling, stellar ejections, SFR timescales) versus the hierarchical growth of the dark matter halos (Bouwens et al. 2015). However, stellar masses are likely to be more strongly correlated with the halo masses than UV luminosities.

The Schechter fits performed in Section 3.5 suggest that the evolution of the  $z'$ -band LFs can be accounted for by an increase of luminosity with cosmic time uniformly across luminosities at a given redshift. Because we have shown that the  $z'$ -band LF is a reasonable proxy for the SMF, it is tempting to analyze the evolution of the rest-frame optical LF obtained in the present work in terms of the evolution of the dark-matter HMF.

We perform a first analysis as follows. We apply a simple abundance-matching technique (Vale & Ostriker 2004) consisting of matching the cumulative number density of the LF to that of Behroozi et al. (2013b) HMF obtained from `HMFcalc`<sup>11</sup> (Murray et al. 2013) and recover the evolution in mass of the HMF at  $z \geq 5$  relative to  $z \sim 4$ . Given the rapid evolution of the HMF in this range of redshift, we adopt the HMFs at  $z = 3.78, 4.95, 5.76, \text{ and } 6.87$ , corresponding to the (median) photometric redshifts of the stacked SEDs. The HMFs assume  $\sigma_8 = 0.81$ . We find a relative displacement in halo mass of 0.76, 0.52, and 0.35 dex, corresponding to  $\sim 1.9, \sim 1.3, \text{ and } \sim 0.9$  mag from  $z \sim 7, 6 \text{ and } 5$  to  $z \sim 4$ , respectively. Figure 12 shows the result of applying the above offsets to the  $\mathcal{M}_{z'}^*$  of the  $z \sim 4$  Schechter parameterization. The solid blue curve marks the best-fit Schechter function at  $z \sim 4$ , whereas the dashed curves represent the  $z \sim 4$  LF rigidly shifted by the corresponding amount at  $z \geq 5$ . The agreement between the predicted and observed LF is very good at all redshifts, suggesting that the  $z'$ -band LF could trace the evolution of the HMF.

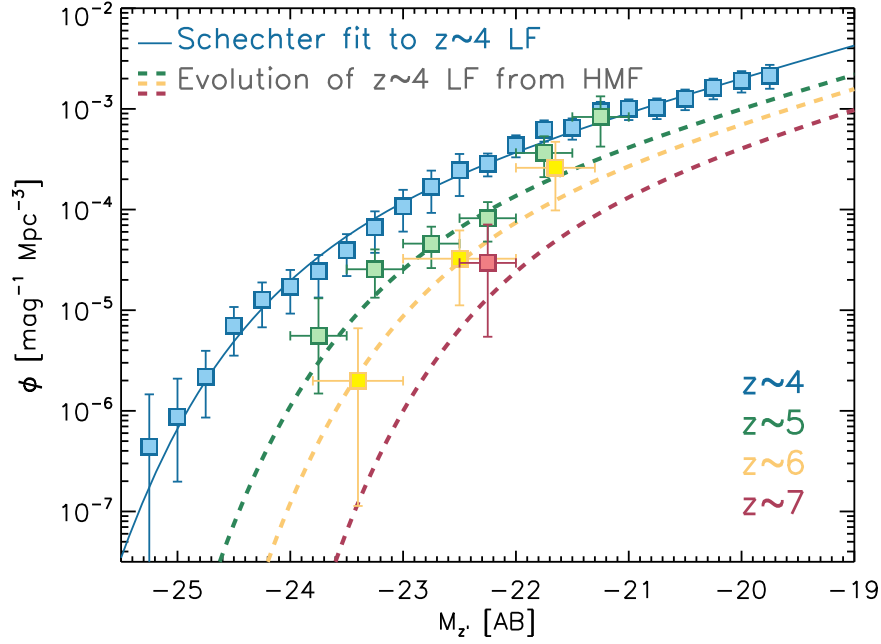
#### 4.3. Evolution of $M_h^*$

In this section, we discuss the evolution of those halos associated with a constant cumulative number density of  $3.1 \times 10^{-5} \text{ Mpc}^{-3}$  over  $4 \lesssim z \lesssim 7$ . This value corresponds to the cumulative number density of  $\mathcal{M}_{z'}^*$  galaxies at  $z \sim 4$  ( $M_{z'} = -23.38$  mag), and allows us to recover the corresponding absolute magnitude up to  $z \sim 7$  with strongly reduced dependence (i.e.,  $\lesssim 0.5$  mag) on the extrapolation of the LFs to magnitudes fainter than actually observed. Table 4 lists a compilation of the values of the main parameters recovered with our analysis.

The abundance matching performed through cumulative number density implicitly assumes that each halo contains one and only one galaxy, and that halos of the same mass contain galaxies of the same  $z'$  luminosity ( $\sim$ stellar mass). Indeed, recent measurements have shown that the scatter between halo mass-stellar mass relation is quite small,  $\sim 0.15\text{--}0.20$  dex (e.g., Conroy & Wechsler 2009; Moster et al. 2010; Tinker et al. 2017; but see also Gu et al. 2016, who found scatter of up to 0.32 dex).

The cumulative number densities of the LFs were computed adopting the Schechter parameterization presented in Section 3.5. For the LFs at  $z \geq 5$ , we adopted the best-fit Schechter functions obtained when the characteristic magnitude was assumed to be

<sup>11</sup> <http://hmf.icrar.org/>—we used the python implementation from <https://github.com/steven-murray/hmf>.



**Figure 12.** Filled colored squares mark our measurements of the  $1/V_{\max}$  LF in the four redshift bins, as detailed by the legend in the bottom-right corner. Error bars include the contribution from Poisson noise and cosmic variance. The solid blue curve marks the best-fit Schechter function at  $z \sim 4$ , and the dashed curves present the  $z \sim 4$  Schechter function evolved in luminosity following the evolution in mass of the halo mass function relative to the  $z \sim 4$  HMF.

**Table 4**  
Values of the Main Observables from our Abundance Matching Analysis

	$z$	Cum. Den. <sup>a</sup> [log(Mpc <sup>-3</sup> )]	$M_{z'}$ [mag]	$M_h$ [log( $M_h/M_\odot$ )]	$L_{z'}/M_h$ [ $L_\odot/M_\odot$ ]	$M_*$ [log( $M_*/M_\odot$ )]	$M_*/M_h$	ISFE <sup>b</sup>
Fixed den.	3.78	-4.51	$-23.39 \pm 0.27$	12.32	$0.069^{+0.019}_{-0.015}$	$10.46 \pm 0.16$	$0.014^{+0.006}_{-0.004}$	$0.087^{+0.038}_{-0.027}$
	4.95	-4.51	$-22.63 \pm 0.23$	11.91	$0.089^{+0.021}_{-0.017}$	$10.02 \pm 0.13$	$0.013^{+0.005}_{-0.003}$	$0.081^{+0.029}_{-0.021}$
	5.76	-4.51	$-22.09 \pm 0.26$	11.67	$0.095^{+0.025}_{-0.020}$	$9.77 \pm 0.10$	$0.013^{+0.003}_{-0.003}$	$0.082^{+0.022}_{-0.017}$
	6.87	-4.51	$-21.82 \pm 0.43$	11.40	$0.138^{+0.067}_{-0.045}$	$9.67 \pm 0.17$	$0.019^{+0.009}_{-0.006}$	$0.118^{+0.058}_{-0.039}$
Evol. den.	3.78	-4.51	$-23.39 \pm 0.27$	12.32	$0.069^{+0.019}_{-0.015}$	$10.46 \pm 0.16$	$0.014^{+0.006}_{-0.004}$	$0.087^{+0.038}_{-0.027}$
	4.95	-4.32	$-22.41 \pm 0.23$	11.83	$0.088^{+0.020}_{-0.017}$	$9.90 \pm 0.12$	$0.012^{+0.004}_{-0.002}$	$0.075^{+0.024}_{-0.015}$
	5.76	-4.14	$-21.65 \pm 0.26$	11.51	$0.090^{+0.024}_{-0.019}$	$9.60 \pm 0.10$	$0.012^{+0.003}_{-0.003}$	$0.078^{+0.021}_{-0.016}$
	6.87	-3.88	$-21.02 \pm 0.43$	11.14	$0.119^{+0.058}_{-0.039}$	$9.34 \pm 0.17$	$0.016^{+0.008}_{-0.005}$	$0.102^{+0.050}_{-0.034}$

**Notes.**

<sup>a</sup> Cumulative number density adopted for the abundance matching.

<sup>b</sup> Integrated Star Formation Efficiency  $\equiv (M_*/M_h)/(\Omega_b/\Omega_m)$ ,  $\Omega_b/\Omega_m \equiv 0.157$ .

the only free parameter of the fit, coinciding with the case of pure luminosity evolution.

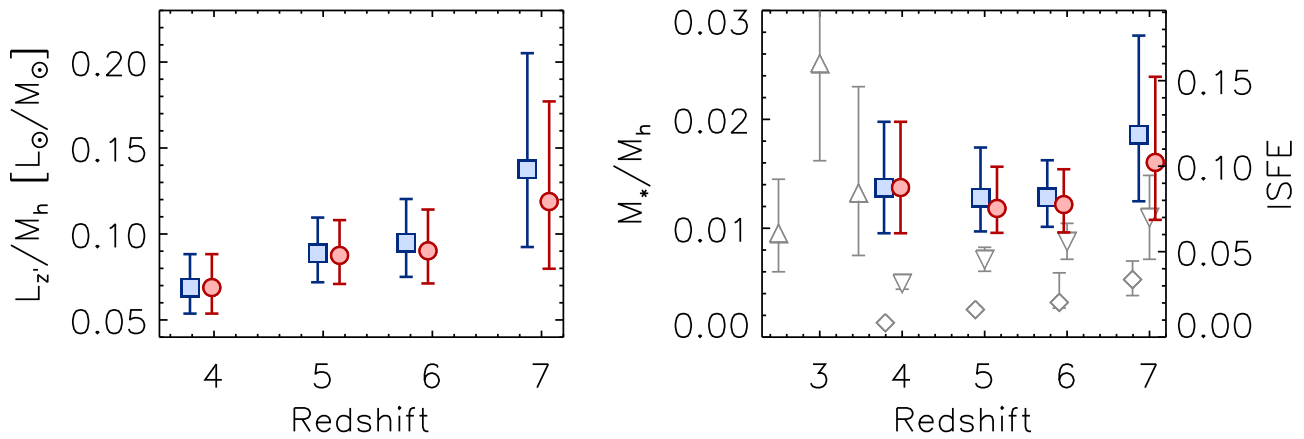
The values of the absolute magnitudes we obtain from our procedure are  $M_{z'} \sim -23.39, -22.63, -22.09, -21.82$  mag for the  $z \sim 4, z \sim 5, z \sim 6,$  and  $z \sim 7$  cases, respectively. These values are consistent within  $1\sigma$ , with the characteristic magnitudes of our Schechter fit. This is not surprising, considering the pure luminosity evolution of the Schechter fits themselves.

The matches to the cumulative number density performed on the HMFs resulted in halo masses  $\log(M_h/M_\odot) = 12.32, 11.91, 11.67, 11.40$  for the  $z \sim 4, z \sim 5, z \sim 6,$  and  $z \sim 7$  cases, respectively. These displacements correspond to an evolution in the halo mass of  $\mathcal{M}_{z'}$  galaxies of  $\sim 0.9, \sim 0.6,$  and  $\sim 0.3$  dex from  $z \sim 7, 6, 5$  to  $z \sim 4$ , respectively. We note, however, that a rigid displacement in mass of the HMF is sufficient to reproduce the HMF evolution at these redshifts only for halo masses  $\log(M_h/M_\odot) \gtrsim 12$ ; at lower halo masses, the

displacement in mass must be coupled to a steepening with redshift of the low-mass end slope.

We can now use the above results on the evolution of the luminosity and of the halo mass to recover the evolution with redshift of the light-to-halo mass for galaxies at fixed cumulative number density. Combining the two, we obtain  $L_{z'}/M_h \sim 0.069, 0.089, 0.095, 0.138$  in units of  $L_\odot/M_\odot$ . These values are also presented in the left panel of Figure 13 and suggest a mild increase with redshift (a factor  $\lesssim 2\times$ ), although the large uncertainties make them consistent with a constant value across the 800 Myr of cosmic time, from  $z \sim 4$  to  $z \sim 7$ .

Using the results on the  $M_*/L_{z'}$  from Section 3.2, we can convert the  $L_{z'}/M_h$  into  $M_*/M_h$ . The result of this is shown in the right panel of Figure 13. The corresponding values are listed in Table 4. The stellar-to-halo mass ratio does not present any significant evolution with redshift. In the same panel, we convert the  $M_*/M_h$  into the integrated star formation efficiency (ISFE),



**Figure 13.** Left panel: ratio between the luminosity and the halo mass from abundance matching at (1) constant cumulative number density (blue squares with error bars) and (2) cumulative number density evolving following Behroozi et al. (2013a, red circles with error bars). The red points have been arbitrarily shifted by  $\delta z = 0.2$  to improve readability. The points show a mild indication of increase with redshift of the  $L_z/M_h$  ratio, although the large uncertainties make it also consistent with no evolution. Right panel: stellar-to-halo mass ratios (Chabrier 2003 IMF), recovered from the values presented in the left panel, by applying the  $M_*/L_z$  relation in Section 3.2 (same plotting conventions of the left panel). The gray points present stellar-to-halo mass ratios from the literature: Durkalec et al. (2015, open upward triangles), Finkelstein et al. (2015a, open downward triangles, converted to Chabrier 2003 IMF), and Harikane et al. (2016, open diamonds). We warn the reader that all these estimates refer to different  $M_h$ , making straightforward comparisons difficult to interpret (see text for more details). The y-axis on the right presents the integrated star formation efficiency (ISFE—also called the stellar baryon fraction), i.e.,  $M_*/M_h$  in units of  $\Omega_b/\Omega_m$ . No clear evidence is found for an evolution of the ISFE with redshift.

i.e.,  $M_*/M_h$  in units of  $\Omega_b/\Omega_m$  (Conroy & Wechsler 2009), which is equivalent to the stellar baryon fraction (Finkelstein et al. 2015b), using  $\Omega_b/\Omega_m = 0.157$  (Planck Collaboration et al. 2016). Our measurements are consistent with the ISFE being constant with redshift. We stress here that this result refers to  $M_h \gtrsim 10^{12} M_\odot$  and does not exclude the existence of evolution with redshift at lower halo masses. We defer a more complete analysis of the dependence of the  $M_*/M_h$  ratio on halo mass to a future work. Furthermore, our samples at  $z \sim 5-7$  are entirely based on LBG selection. If non-negligible numbers of redder (dustier/more evolved) galaxies exist at these epochs, they would affect SMF (and likely its massive end, e.g., Caputi et al. 2015; Stefanon et al. 2015) and, consequently, the recovered  $M_h$ .

Matching galaxies at a constant cumulative number density, however, does not consider the effect of major mergers in the galaxy ranking. We therefore repeated the same analysis using a cumulative number density that evolves with redshift, following the recipe of Behroozi et al. (2013a). The results are listed in Table 4, and plotted as red circles in Figure 13. No significant difference with the constant cumulative number density match is observed. We also note that the values for the  $z \sim 7$  bin rely on the extrapolation of the LF to luminosities below those currently probed by our sample. Those measurements should then be treated with caution.

In the right panel of Figure 13, we also plot recent estimates of the  $M_*/M_h$  from the literature: Durkalec et al. (2015), Finkelstein et al. (2015a, converted to a Chabrier 2003 IMF by applying a factor 0.55) and Harikane et al. (2016). Durkalec et al. (2015) applied the measurements of the two-point correlation function to a halo occupation model to recover the halo mass of samples of galaxies at  $z \sim 2-5$  with spectroscopic redshift from the VIMOS Ultra Deep Survey (Le Fèvre et al. 2015). Finkelstein et al. (2015b) measured the evolution of the  $M_*/M_h$  by abundance matching the  $z \sim 4-7$  UV LF. Harikane et al. (2016) recovered  $M_*/M_h$  from the clustering of LBGs selected at  $z \sim 4-7$  from a variety of programs, including CANDELS, the *Hubble* Frontier Fields (PI: J. Lotz), and Subaru Hyper-Suprime Cam Subaru Strategic Program (PI: S. Miyazaki).

At face value, our measurements are consistent with those of Durkalec et al. (2015) at  $z \sim 4$ , and with those of Finkelstein et al. (2015b) at  $z \sim 6-7$ ; however, they are inconsistent with those of Harikane et al. (2016) over the full range of redshift, and with those of Finkelstein et al. (2015b) at  $z \sim 4-5$ . Recently, Mancuso et al. (2016), applying abundance matching to the evolution of the SFR function recovered from UV+far-IR data, found indication for a non-evolving  $M_*/M_h$  ratio at  $z \geq 4$ .

The  $M_*/M_h$  measurements presented in the right panel of Figure 13 were obtained from a variety of methods, and ultimately refer to different  $M_h$  estimates, making a straightforward comparison difficult to interpret. Specifically, our measurements of the halo mass are based on a constant cumulative number-density match with  $\log(M_h/M_\odot) \sim 12.3, 11.9, 11.7, 11.4$  at  $z \sim 4, 5, 6, 7$ , respectively. The increase with redshift of the  $M_*/M_h$  of Harikane et al. (2016) refers to a fixed  $\log(M_h/M_\odot) = 11$  across  $z \sim 4-7$ . Finkelstein et al. (2015b) report an increase with redshift of the ISFE  $\propto (0.024 \pm 0.07) \times z$ . However, this trend is most likely driven by the point at  $z \sim 4$ , and originates from limited evolution of the UV LF between  $z \sim 4$  and  $z \sim 5$  ( $\Delta \mathcal{M}_{UV}^* \sim 0.08 \pm 0.16$  mag) observed by Finkelstein et al. (2015a), in contrast to the large-luminosity evolution ( $\Delta \mathcal{M}_{UV}^* \sim 0.76 \pm 0.22$  mag) observed in our work over the same redshift interval. This, along with the constant characteristic magnitude of the UV LF (i.e.,  $\approx$ constant SFR) assumed as a criterion for the abundance matching, generates a reference cumulative number density that decreases with redshift, and halo masses  $\log(M_h/M_\odot) \sim 11.9, 11.7, 11.6, 11.3$ . We note here that our measurements are consistent with those of Finkelstein et al. (2015b) at  $z \sim 6-7$ , i.e., where the  $M_h$  recovered by the two teams are more similar, whereas they are inconsistent at  $z \sim 4-5$ , where the  $M_h$  differ. Finally Durkalec et al. (2015) estimates refer to a diversity of halo masses and redshift ranges:  $\log(M_h/M_\odot) \sim 11.1, 11.5, 11.2$  for  $z_{\text{mean}} \sim 2.5, 3.0, 3.5$ , respectively.

Our finding of  $M_*/M_h$  independent of redshift is also qualitatively in agreement with recent estimates of the galaxy bias, observed to be nearly constant over  $z \sim 4-6$



(Barone-Nugent et al. 2014), although a measurement at  $z \sim 7$  from the same work seems to suggest a potential change of the star formation efficiency at earlier epochs.

The picture is not settled, even from a theoretical perspective. Indeed, some of the models predict an *increase* with cosmic time of the  $M_*/M_h$  ratio for a fixed  $M_h$  (e.g., Somerville et al. 2015). Other models find that  $M_*/M_h$  decreases with cosmic time at fixed halo mass (Moster et al. 2013; Behroozi & Silk 2015). However, when considering the evolution of the same population of galaxies (through an evolving number density), Behroozi & Silk (2015) find that  $M_*/M_h$  is nearly independent on redshift. Finally, other models, instead, find  $M_*/M_h$  to be insensitive to redshift (e.g., O’Shea et al. 2015; Mutch et al. 2016). Some semi-empirical models have been able to reproduce the evolution of the UV LF from  $z \sim 2$  to  $z \sim 10$  under the assumption that, for star-forming galaxies, the  $M_*/M_h$  depended on  $M_h$  but not on redshift (Trenti et al. 2010; Tacchella et al. 2013; Mason et al. 2015).

## 5. Summary and Conclusions

The main aim of this work was to measure the rest-frame  $z'$ -band luminosity function (LF) of field galaxies, and study its evolution at  $z \geq 4$ . The rest-frame  $z'$  band was selected for three reasons: 1) it is not contaminated by strong emission from nebular lines; 2) light in this wavelength range is dominated by lower-mass, long-living stars; and 3) it can be probed up to  $z \sim 8$  using the current *Spitzer*/IRAC data. These characteristics suggest that it can provide a complementary basis for dealing with stellar mass measurements at high redshift, minimizing the potential systematic effects that can affect stellar mass measurements.

We therefore assembled samples of LBGs at  $z \sim 4$ ,  $z \sim 5$ ,  $z \sim 6$ , and  $z \sim 7$ , selected over the GOODS-N and GOODS-S fields. The  $z \sim 4$  sample was complemented by galaxies with photometric redshifts  $3.5 < z_{\text{phot}} < 4.5$ , extracted from a 37-band far-UV-to- $8.0 \mu\text{m}$   $K_s$ -detected photometric catalog based on UltraVISTA DR2. The larger  $z \sim 4$  co-moving volume provided by the UltraVISTA data allowed us to gain statistics on the rarer, more luminous, and/or redder galaxies.

The GOODS-N/S sample takes advantage of the recently released, full-depth IRAC maps (Labbé et al. 2015), obtained from the combination of all the IRAC programs carried out so far over these fields; namely, IGOODS, IUDF, GOODS, ERS, S-CANDELS, SEDS, and UFD2. These maps reach a depth of  $\sim 25.8$  and  $\sim 24.5$  mag in the  $4.5 \mu\text{m}$  and  $5.8 \mu\text{m}$  bands, respectively ( $2''0$  diameter aperture,  $5\sigma$ ), although the coverage is highly inhomogeneous. Similarly, the UltraVISTA catalog benefits from IRAC 3.6 and  $4.5 \mu\text{m}$  mosaics that combine the S-COSMOS, S-CANDELS, and SPLASH programs, and reach a depth of  $\sim 22.5$  mag ( $2''0$  diameter aperture,  $5\sigma$ ).

We further selected our sample based on the S/N in the IRAC band, or bands closer to the rest-frame  $z'$  band. Specifically, the final  $z \sim 4$  sample was selected to have  $S/N > 5$  in the  $4.5 \mu\text{m}$ , whereas the  $5 < z < 7$  samples were selected to have  $S/N > 4$  in the inverse-variance weighted combination of S/N in the  $5.8$  and  $8.0 \mu\text{m}$  bands. Our final composite sample included 2098, 72, 10, and 2 objects, for the  $z \sim 4, 5, 6,$  and  $7$  redshift bins, respectively. Although the  $z'$  band is covered by the  $5.8$  and  $8.0 \mu\text{m}$  IRAC data up to  $z \sim 8$ , we do not register any LBG galaxy at  $z \sim 8$ , which also satisfies our selection criteria on the S/N of IRAC fluxes.

Our main results are as follows.

1. At  $z \sim 4$ , and for absolute magnitudes  $M_{z'}$  fainter than  $\sim -23$  AB, galaxies follow a linear relation on the  $M_{\text{UV}}-M_{z'}$  plane, with slope  $\sim 0.8$ . This correlation breaks at  $M_{z'} \lesssim -23$  AB: the  $M_{\text{UV}}$  of these galaxies covers the full range of values observed for galaxies with fainter  $M_{z'}$  (Figure 6).
2. We performed stacking analysis and measured the  $M_*/L_{z'}$  of galaxies segregated according to their redshift and absolute magnitude  $M_{z'}$ . The  $M_*/L_{z'}$  at  $z \sim 4$  is independent of  $M_{z'}$  for  $M_{z'} \gtrsim -22.5$ ; at brighter  $M_{z'}$ , the  $M_*/L_{z'}$  increases with luminosity following a power law. The  $M_*/L_{z'}$  at  $z \geq 5$  are consistent with those observed at  $z \sim 4$ , although the associated large uncertainties may hide a different behavior (Figures 8 and 9).
3. We computed the LF in the rest-frame  $z'$  band, using the  $V_{\text{max}}$  estimator, in four different redshift bins:  $z \sim 4$ ,  $z \sim 5$ ,  $z \sim 6$ , and  $z \sim 7$ . We admit freely that our single bin measurement at  $z \sim 7$  does not allow us to set stringent limits on the shape of the  $z \sim 7$  LF. The LF shows evolution from  $z \sim 7$  to  $z \sim 4$ . Schechter fits to the  $V_{\text{max}}$  LF marginally prefer pure evolution in luminosity over one in density (Figure 10).
4. The non-decreasing  $M_*/L_{z'}$  with luminosity (corresponding to an injective mapping) allowed us to apply a simple conversion from luminosity to stellar mass. We therefore converted our LF measurements into SMF using the  $M_*/L_{z'}$  recovered from the stacking analysis at  $z \sim 4$ . The obtained SMFs are consistent with the average SMF determination from the literature. Despite the relaxed S/N cuts in IRAC flux applied to our samples, the lower stellar mass over which we recover our SMFs is  $\sim 5-10\times$  larger than typical lower limits from the literature (Figure 11).
5. Evolution in the halo mass relative to  $z \sim 4$ , recovered from abundance matching the HMFs, reproduces the luminosity evolution of the LF at  $z \gtrsim 4$  (Figure 12). The stellar-to-halo mass ratio at fixed cumulative number density shows no strong evidence for evolution with redshift over  $4 < z < 7$  (Figure 13).

The above results allow us to draw the following conclusions.

1. The current depth of *Spitzer*/IRAC data are sufficient to probe the regimes in rest-frame UV luminosities, both where the rest-UV luminosities are correlated with stellar mass and where they are not.
2. The existence, at  $z \sim 4$  and  $z \sim 5$  of LBGs luminous in the rest-frame  $z'$  band and spanning a broad range of UV luminosities, suggests that the adoption of the UV LF for SMF estimates may be affected by systematics. Specifically, samples of galaxies selected to have a narrow range in UV luminosities potentially include a combination of high- and low-mass objects, and ultimately can introduce an overestimate of the low-mass end slope, and an underestimate of the densities at the high-mass end.
3. The higher values of the lower stellar mass bin in our SMFs, compared to recent determination from the literature, arising from the S/N cuts applied to the IRAC fluxes, suggests that the current low-mass end of the SMFs at  $z \gtrsim 4$  might be based on low S/N flux measurements ( $\sim 1-2\sigma$  upper limits) in the observed

IRAC bands that are most sensitive to the stellar mass (i.e., 5.8 and 8.0  $\mu\text{m}$ ). Higher S/N measurements are available from 3.6 to 4.5  $\mu\text{m}$  data. However, these bands at  $z \gtrsim 5$  are contaminated by nebular emission, which can potentially bias the stellar mass estimates, given our still-limited knowledge of the emission line intensities of high-redshift star-forming galaxies.

4. The rest-frame  $z'$  band LF can be a valid proxy for SMFs and HMF measurements at  $z \gtrsim 4$ , and complementary to SMF estimates based on individual stellar mass measurements. The nearly flat dependence of the  $M_*/L_{z'}$  on  $M_{z'}$  increases this confidence.

This work is largely based on data from the cryogenic programs of *Spitzer*/IRAC. Although the depth of the 3.6  $\mu\text{m}$ - and 4.5  $\mu\text{m}$ -band data can still be improved through non-cryogenic programs, the sensitivity of *JWST*/MIRI provides the only opportunity for increasing the depth at wavelengths  $\lambda > 5 \mu\text{m}$ , necessary for improving current estimates of stellar masses at  $z \gtrsim 5$ .

We are appreciative to Adriano Fontana for a very helpful referee report that greatly improved this paper. M.S. would like to thank Adriano Fontana for constructive discussions. M.S. and R.B. are grateful to Karina Caputi, Yuichi Harikane, Michele Trenti, and Stuart Wyithe for helpful feedback on an advanced draft of this manuscript. This work is 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. This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. This work is based on observations taken by the 3D-*HST* Treasury Program (GO 12177 and 12328) with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

## Appendix A Sample Selection Criteria

After applying the S/N cuts described in Section 2.2, we further cleaned our sample, excluding those objects satisfying any of the following conditions: (1) the contribution to the 5.8 and 8.0  $\mu\text{m}$  flux from neighboring objects is excessively high; (2) the source morphology is very uncertain or confused, making IRAC photometry undetermined; (3) the source is detected at X-ray wavelengths, suggesting that it is a lower-redshift AGN; (4) the source is at higher redshift, but its SED is dominated by AGN light; (5) LBGs with a likely  $z < 3.5$  solution from photometric redshift analysis. In the following paragraphs, we will describe in more detail the above criteria and their effects on the sample size.

The broad IRAC PSF, along with the unprecedented photometric depth of the IRAC mosaics in the GOODS-N, GOODS-S, and UltraVISTA fields, can result in flux measurements potentially affected by contamination from brighter nearby objects. The procedure we adopted for the flux measurements already deals with this problem by cleaning each source from its neighbors before performing the

photometry. However, in some cases, the flux at the position of the object of interest mostly comes from the bright neighbors, resulting in potentially very uncertain flux measurements. We therefore opted to further clean our sample by applying a cut to the maximum fraction of flux from neighbors contributing to the flux of each object, before the neighbor-cleaning process. Specifically, we excluded from our sample those sources whose neighbors were contributing more than 65% to the total flux at the position of each source in our sample. We also visually inspected the cutouts from the IRAC photometry, and further excluded those sources showing a residual contamination from bright nearby sources. In this step, we removed 280 galaxies (211/69, for GOODS-N/S and UltraVISTA, respectively; of the 211 GOODS galaxies, three were at  $z \sim 6$  and one was at  $z \sim 7$ ). In Section 2.4, we describe the Monte Carlo simulation that we implemented to statistically evaluate the selection effects introduced by the above selection criteria.

We visually inspected the cutouts of the GOODS-N/S sample in the WFC3/ $H_{160}$  band and of the UltraVISTA sample in the ACS/F814W, and excluded those objects with doubtful morphology, e.g., if one was the result of two or more distinct objects, or the deblending from SExtractor was deemed inconsistent. Furthermore, the visual inspection also allowed us to identify and exclude objects with point-source morphology as either potentially AGN dominated or brown dwarf contaminants. This was particularly important for the UltraVISTA sample, because the  $K_s$ -band detection image is characterized by a PSF FWHM  $\sim 0''.8$ , much broader than that of ACS or WFC3 (FWHM  $\sim 0''.12$  and  $\sim 0''.2$ , respectively). This class of objects is subject to very inaccurate photometry, redshift classification, and/or luminosity measurement. Through the above criteria, we excluded 53 objects from the GOODS-N/S sample (one at  $z \sim 7$ ) and 63 objects from the UltraVISTA sample.

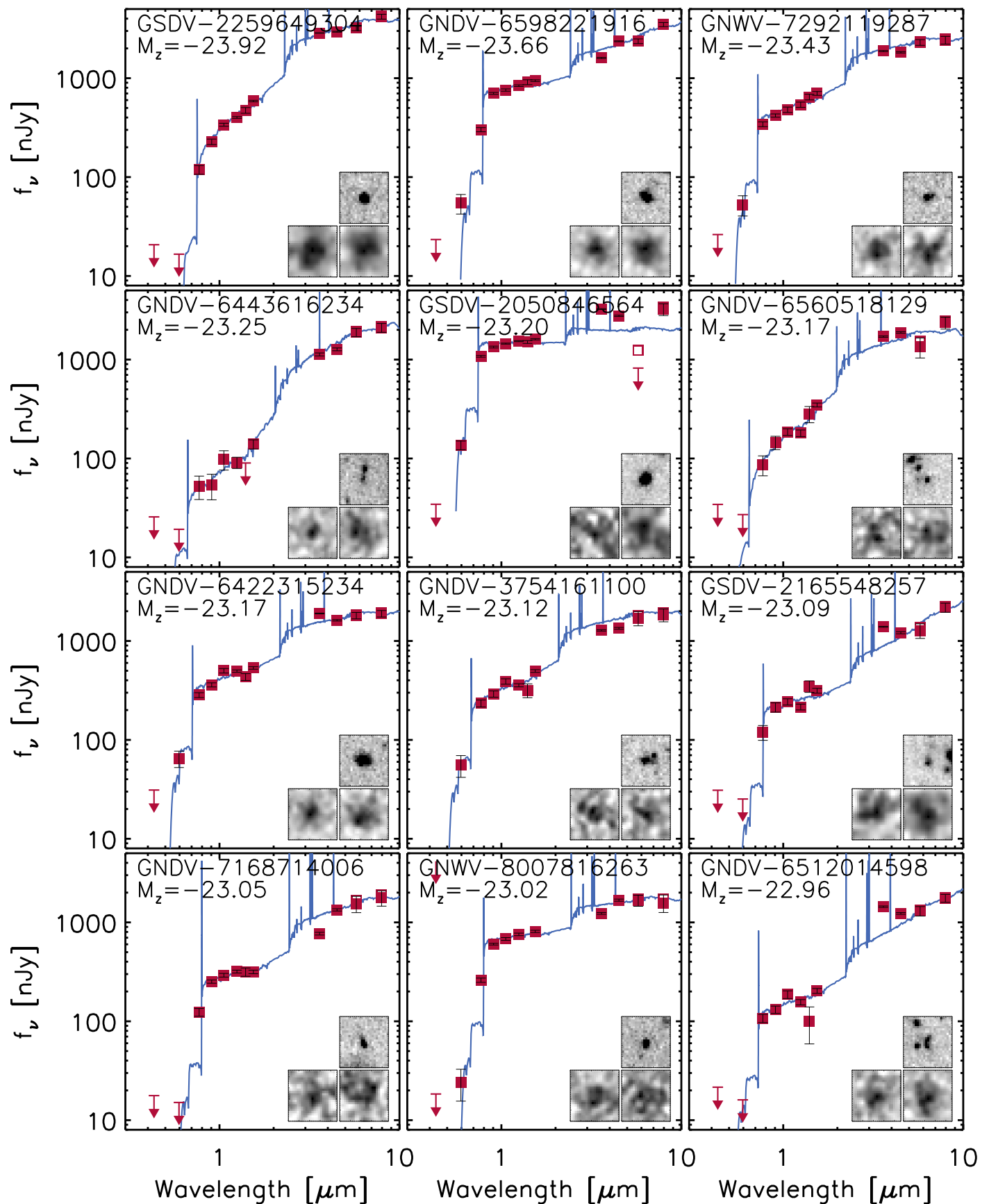
Successively, we cross-matched our sample to catalogs of X-ray sources in the GOODS-N/S (Alexander et al. 2003; Xue et al. 2011) and COSMOS fields (Cappelluti et al. 2009; Elvis et al. 2009; Pâris et al. 2012), and excluded all the matching sources, as these are potential lower-redshift AGN contaminants. We identified 34 sources with an X-ray counterpart matching our initial sample, most of which were at  $z \sim 4$  (29), and four at  $z \sim 5$ . Furthermore, we visually inspected all the observed SEDs, to exclude objects with very red, power-law-like rest-frame optical/NIR slopes that could be signature of Type-1 AGN. In this step, we flagged and removed from the sample a total of 74 sources (20/54).

Finally, we ran EAzy (Brammer et al. 2008) on the sample of LBGs, and excluded those galaxies with  $z_{\text{peak}} < 3.5$ , sources with prominent secondary lower- $z$  solution, and those with inconsistent SED. Through this step, we excluded 109 sources (4 of which at  $z \sim 6$ ).

The final sample consists of 2098 galaxies at  $z \sim 4$  (1680 from the LBG sample and 418 from the UltraVISTA sample), 72 at  $z \sim 5$ , 10 at  $z \sim 6$ , and 2 objects at  $z \sim 7$ .

## Appendix B SEDs of the $z \sim 5, 6$ and 7 Samples

In Figure 14, we present the SEDs of the 12 most luminous galaxies included in the  $z \sim 5$  sample, whereas Figures 15 and 16 show the full sample of galaxies at  $z \sim 6$  and  $z \sim 7$ , respectively.



**Figure 14.** SEDs of the 12 most luminous galaxies in the  $z \sim 5$  sample. Photometric measurements are marked by the solid boxes with error bars, with arrows for  $2\sigma$  upper limits. The measurements for the 5.8 and 8.0  $\mu\text{m}$  bands include the correction for the flux boosting. The original measurements for these two bands are shown as open boxes. The blue curve identifies the best-fit SED template from EAZY. In each panel, the three insets present the cutouts ( $\sim 3''$  side) in the WFC3  $H_{160}$  band (upper box) and neighbor-subtracted 5.8 and 8.0  $\mu\text{m}$  bands (lower boxes, left and right, respectively). Labeled in the top left corner is the absolute magnitude  $M_z$ .



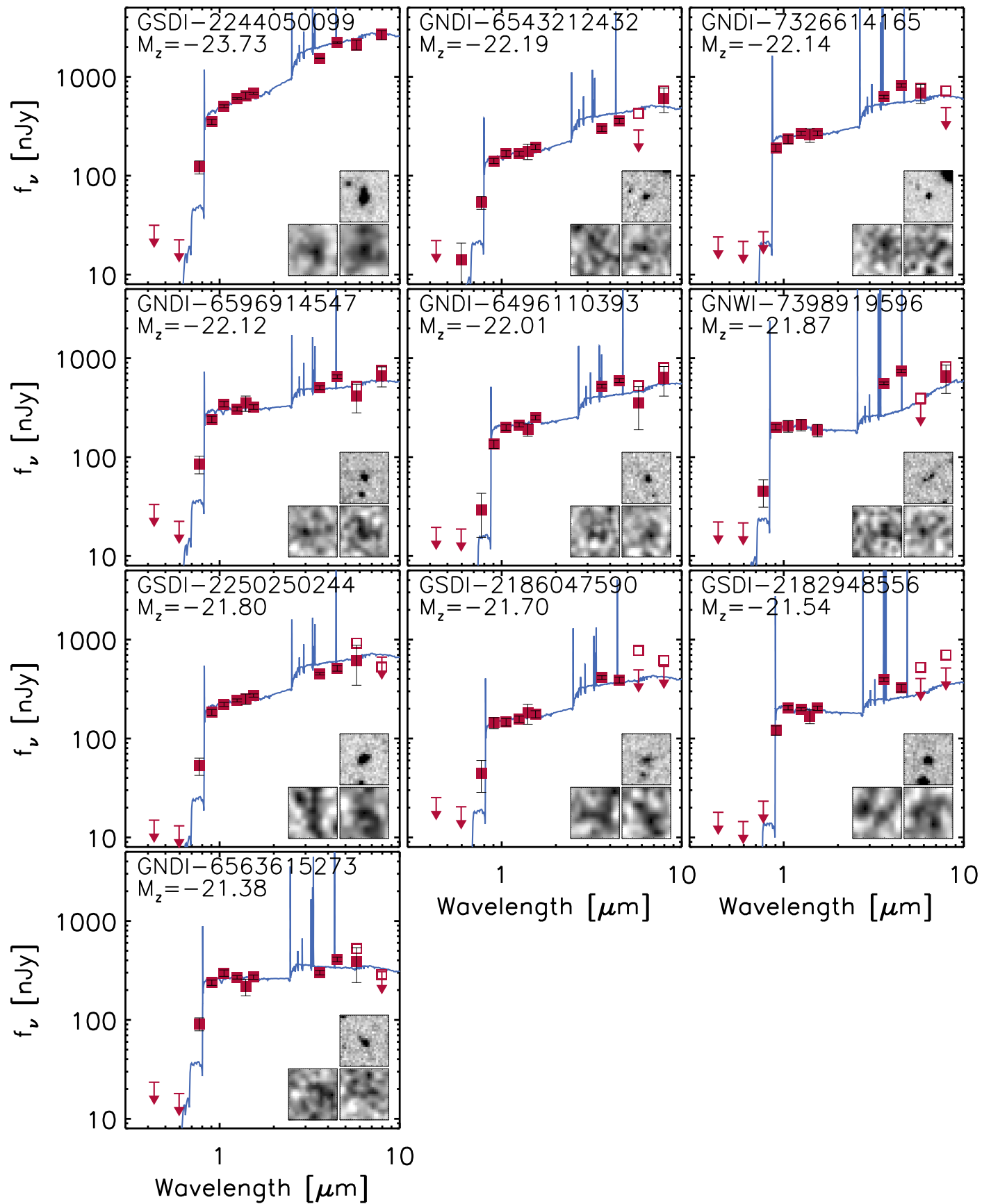


Figure 15. SEDs of the  $z \sim 6$  sample. Other plotting conventions as for Figure 14.

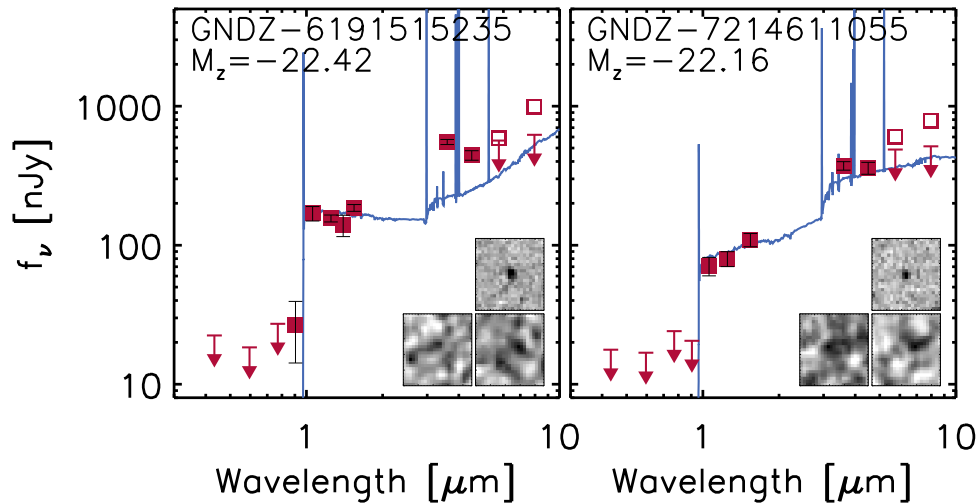


Figure 16. SEDs of the galaxies at  $z \sim 7$ . Other plotting conventions as for Figure 14.

Most of the galaxies are characterized by small, compact sizes in the  $H_{160}$ . Noticeably, GSDI-2244050099, at  $z \sim 6$ , is among the most luminous galaxies of the full sample, including  $z \sim 4$ . Its apparent size is larger than the average size of the galaxies in the  $z \sim 5, 6$ , and 7 samples. Careful inspection of the WFC3/ $H_{160}$  cutout does not show any indication of clumpiness. However, in ACS/F184W, we observe two possible components, separated by  $\sim 0''.45$  ( $\sim 2.6$  kpc at  $z = 6$ ). We opted to include it in our sample, given the small separation between the two components that is visible only in the ACS data, the consistency of the SED, and the fact that the IRAC flux appears to be centered at the position of the brighter component. This object constitutes a unique element of the highest luminosity bin of the  $z \sim 6$  LF. It is noteworthy that, even assuming a twofold overestimate of the IRAC flux, the resulting absolute magnitude would still be consistent with the highest-luminosity bin.

## References

- Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, *AJ*, 126, 539
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, *ApJ*, 769, 80
- Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, *ApJS*, 218, 33
- Avni, Y., & Bahcall, J. N. 1980, *ApJ*, 235, 694
- Barone-Nugent, R. L., Trenti, M., Wyithe, J. S. B., et al. 2014, *ApJ*, 793, 17
- Barro, G., Trump, J. R., Koo, D. C., et al. 2014, *ApJ*, 795, 145
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, 132, 1729
- Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, *ApJ*, 717, 379
- Behroozi, P. S., Marchesini, D., Wechsler, R. H., et al. 2013a, *ApJL*, 777, L10
- Behroozi, P. S., & Silk, J. 2015, *ApJ*, 799, 32
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013b, *ApJ*, 770, 57
- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, *ApJ*, 705, 936
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010, *ApJL*, 708, L69
- 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., & McLeod, D. J. 2017, *MNRAS*, 466, 3612
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
- Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, *ApJS*, 200, 13
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Cappellari, M., di Serego Alighieri, S., Cimatti, A., et al. 2009, *ApJL*, 704, L34
- Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, *A&A*, 497, 635
- Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, *ApJL*, 750, L20
- Caputi, K. I., Ilbert, O., Laigle, C., et al. 2015, *ApJ*, 810, 73
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
- Chabrier, G. 2003, *PASP*, 115, 763
- Charlot, S., & Fall, S. M. 2000, *ApJ*, 539, 718
- Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2010, *MNRAS*, 401, 1166
- Conroy, C., & Wechsler, R. H. 2009, *ApJ*, 696, 620
- Crawford, T. M., Switzer, E. R., Holzappel, W. L., et al. 2010, *ApJ*, 718, 513
- Dahlen, T., Mobasher, B., Dickinson, M., et al. 2010, *ApJ*, 724, 425
- Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, arXiv:1701.02734
- de Barros, S., Schaerer, D., & Stark, D. P. 2014, *A&A*, 563, A81
- Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *A&G*, 50, 5.12
- Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, *MNRAS*, 444, 2960
- Dunlop, J. S., McLure, R. J., Robertson, B. E., et al. 2012, *MNRAS*, 420, 901
- Durkalec, A., Le Fèvre, O., de la Torre, S., et al. 2015, *A&A*, 576, L7
- Eddington, A. S. 1913, *MNRAS*, 73, 359
- Elvis, M., Civano, F., Vignali, C., et al. 2009, *ApJS*, 184, 158
- Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016, *ApJ*, 821, 122
- Finkelstein, S. L., Ryan, R. E., Jr., Papovich, C., et al. 2015a, *ApJ*, 810, 71
- Finkelstein, S. L., Song, M., Behroozi, P., et al. 2015b, *ApJ*, 814, 95
- Fontana, A., Salimbeni, S., Grazian, A., et al. 2006, *A&A*, 459, 745
- Gehrels, N. 1986, *ApJ*, 303, 336
- Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJL*, 600, L93
- González, V., Bouwens, R. J., Labbé, I., et al. 2012, *ApJ*, 755, 148
- González, V., Labbé, I., Bouwens, R. J., et al. 2011, *ApJL*, 735, L34
- Grazian, A., Fontana, A., Santini, P., et al. 2015, *A&A*, 575, A96
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- Gu, M., Conroy, C., & Behroozi, P. 2016, *ApJ*, 833, 2
- Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24
- Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, *ApJ*, 821, 123
- Hogg, D. W., & Turner, E. L. 1998, *PASP*, 110, 727
- Huang, J.-S., Zheng, X. Z., Rigopoulou, D., et al. 2011, *ApJL*, 742, L13
- Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, *A&A*, 556, A55
- Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2014, *MNRAS*, 439, 1245
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36
- Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, *ApJ*, 700, 221
- Kron, R. G. 1980, *ApJS*, 43, 305
- Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, *ApJL*, 649, L67
- Labbé, I., González, V., Bouwens, R. J., et al. 2010a, *ApJL*, 716, L103
- Labbé, I., González, V., Bouwens, R. J., et al. 2010b, *ApJL*, 708, L26
- Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, *ApJL*, 777, L19
- Labbé, I., Oesch, P. A., Illingworth, G. D., et al. 2015, *ApJS*, 221, 23
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, 576, A79
- Lee, K.-S., Dey, A., Reddy, N., et al. 2011, *ApJ*, 733, 99
- Lee, K.-S., Ferguson, H. C., Wiklind, T., et al. 2012, *ApJ*, 752, 66
- Madau, P., & Dickinson, M. 2014, *ARA&A*, 52, 415
- Mancuso, C., Lapi, A., Shi, J., et al. 2016, *ApJ*, 823, 128

- Marchesini, D., Stefanon, M., Brammer, G. B., & Whitaker, K. E. 2012, *ApJ*, **748**, 126
- Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2009, *ApJ*, **701**, 1765
- Marchesini, D., Whitaker, K. E., Brammer, G., et al. 2010, *ApJ*, **725**, 1277
- Mason, C. A., Trenti, M., & Treu, T. 2015, *ApJ*, **813**, 21
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
- McLure, R. J., Dunlop, J. S., de Ravel, L., et al. 2011, *MNRAS*, **418**, 2074
- Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, *MNRAS*, **435**, 87
- Mortlock, A., McLure, R. J., Bowler, R. A. A., et al. 2017, *MNRAS*, **465**, 672
- Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, **428**, 3121
- Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, *ApJ*, **710**, 903
- Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, *ApJ*, **731**, 113
- Murray, S. G., Power, C., & Robotham, A. S. G. 2013, *A&C*, **3**, 23
- Mutch, S. J., Geil, P. M., Poole, G. B., et al. 2016, *MNRAS*, **462**, 250
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013a, *ApJS*, **206**, 8
- Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013b, *ApJ*, **777**, 18
- Muzzin, A., Marchesini, D., van Dokkum, P. G., et al. 2009, *ApJ*, **701**, 1839
- Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, *ApJS*, **228**, 7
- Oesch, P. A., Labbé, I., Bouwens, R. J., et al. 2013, *ApJ*, **772**, 136
- O'Shea, B. W., Wise, J. H., Xu, H., & Norman, M. L. 2015, *ApJL*, **807**, L12
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2012, *A&A*, **548**, A66
- Pérez-González, P. G., Rieke, G. H., Villar, V., et al. 2008, *ApJ*, **675**, 234
- Pforr, J., Maraston, C., & Tonini, C. 2012, *MNRAS*, **422**, 3285
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Rasappu, N., Smit, R., Labbé, I., et al. 2016, *MNRAS*, **461**, 3886
- Reddy, N. A., Erb, D. K., Pettini, M., Steidel, C. C., & Shapley, A. E. 2010, *ApJ*, **712**, 1070
- Rudnick, G., Rix, H.-W., Franx, M., et al. 2003, *ApJ*, **599**, 847
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, **799**, 183
- Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, *ApJS*, **172**, 86
- Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, *ApJ*, **801**, 97
- Santini, P., Fontana, A., Grazian, A., et al. 2012, *A&A*, **538**, A33
- Schaerer, D., & de Barros, S. 2009, *A&A*, **502**, 423
- Schechter, P. 1976, *ApJ*, **203**, 297
- Schmidt, M. 1968, *ApJ*, **151**, 393
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, **214**, 24
- Smit, R., Bouwens, R. J., Labbé, I., et al. 2014, *ApJ*, **784**, 58
- Somerville, R. S., Popping, G., & Trager, S. C. 2015, *MNRAS*, **453**, 4337
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, *ApJ*, **825**, 5
- Spitler, L. R., Straatman, C. M. S., Labbé, I., et al. 2014, *ApJL*, **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
- Stefanon, M., & Marchesini, D. 2013, *MNRAS*, **429**, 881
- Stefanon, M., Marchesini, D., Muzzin, A., et al. 2015, *ApJ*, **803**, 11
- Stefanon, M., Yan, H., Mobasher, B., et al. 2017, *ApJS*, **229**, 32
- Steinhardt, C. L., Capak, P., Masters, D., & Speagle, J. S. 2016, *ApJ*, **824**, 21
- Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, *ApJL*, **791**, L25
- Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, *ApJ*, **830**, 51
- Szalay, A. S., Connolly, A. J., & Szokoly, G. P. 1999, *AJ*, **117**, 68
- Tacchella, S., Trenti, M., & Carollo, C. M. 2013, *ApJL*, **768**, L37
- Taylor, E. N., Franx, M., Brinchmann, J., van der Wel, A., & van Dokkum, P. G. 2010, *ApJ*, **722**, 1
- Tinker, J. L., Brownstein, J. R., Guo, H., et al. 2017, *ApJ*, **839**, 121
- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, *ApJ*, **783**, 85
- Trenti, M., Stiavelli, M., Bouwens, R. J., et al. 2010, *ApJL*, **714**, L202
- Vale, A., & Ostriker, J. P. 2004, *MNRAS*, **353**, 189
- van de Sande, J., Kriek, M., Franx, M., et al. 2013, *ApJ*, **771**, 85
- van der Burg, R. F. J., Hildebrandt, H., & Erben, T. 2010, *A&A*, **523**, A74
- Wang, T., Elbaz, D., Schreiber, C., et al. 2016, *ApJ*, **816**, 84
- Weiner, B. J. & AGHAST Team 2014, American Astronomical Society Meeting, **223**, 227.07
- Whitaker, K. E., Franx, M., Bezanson, R., et al. 2015, *ApJL*, **811**, L12
- Wilkins, S. M., Feng, Y., Di-Matteo, T., et al. 2016, *MNRAS*, **460**, 3170
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, **195**, 10
- Yan, H., Dickinson, M., Eisenhardt, P. R. M., et al. 2004, *ApJ*, **616**, 63