

# Evolution of the galaxy stellar mass functions and UV luminosity functions at $z = 6\text{--}9$ in the Hubble Frontier Fields

Rachana Bhatawdekar<sup>1</sup>,<sup>1</sup> Christopher J. Conselice,<sup>1</sup> Berta Margalef-Bentabol<sup>2</sup> and Kenneth Duncan<sup>3</sup>

<sup>1</sup>School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK

<sup>2</sup>Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, F-75014 Paris, France

<sup>3</sup>Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands

Accepted 2019 March 20. Received 2019 March 20; in original form 2018 July 19

## ABSTRACT

We present new measurements of the evolution of the galaxy stellar mass functions (GSMFs) and ultraviolet luminosity functions (UV LFs) for galaxies from  $z = 6\text{--}9$  within the Frontier Field cluster MACSJ0416.1–2403 and its parallel field. To obtain these results, we derive the stellar masses of our sample by fitting synthetic stellar population models to their observed spectral energy distribution with the inclusion of nebular emission lines. This is the deepest and farthest in distance mass function measured to date and probes down to a level of  $M_* = 10^{6.8} M_\odot$ . The main result of this study is that the low-mass end of our GSMF to these limits and redshifts appears to become steeper from  $-1.98^{+0.07}_{-0.07}$  at  $z = 6$  to  $-2.38^{+0.72}_{-0.88}$  at  $z = 9$ , steeper than previously observed mass functions at slightly lower redshifts, and we find no evidence of turnover in the mass range probed. We furthermore demonstrate that the UV LF for these systems also appears to show a steepening at the highest redshifts, without any evidence of turnover in the luminosity range probed. Our  $M_{\text{UV}} - M_*$  relation exhibit shallower slopes than previously observed and are in accordance with a constant mass-to-light ratio. Integrating our GSMF, we find that the stellar mass density increases by a factor of  $\sim 15^{+21}_{-6}$  from  $z = 9$  to  $z = 6$ . We estimate the dust-corrected star formation rates (SFRs) to calculate the specific SFRs ( $\text{sSFR} = \text{SFR}/M_*$ ) of our sample, and find that for a fixed stellar mass of  $5 \times 10^9 M_\odot$ ,  $\text{sSFR} \propto (1+z)^{2.01 \pm 0.16}$ . Finally, from our new measurements, we estimate the UV luminosity density ( $\rho_{\text{UV}}$ ) and find that our results support a smooth decline of  $\rho_{\text{UV}}$  towards high redshifts.

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

## 1 INTRODUCTION

Exploring the very first galaxies is one of the major contemporary problems in astronomy. We do not know when the first galaxies formed, or how their formation occurred – two related but distinct questions. Until the *James Webb Space Telescope* (*JWST*) launches, these problems can be best addressed through deep imaging observations, particularly with the *Hubble Space Telescope* (*HST*). *Hubble* allows us to view the Universe back to within 500 million years of the big bang, and perhaps earlier. This allows us to address the question of how much of the current galaxy population was in place at these early times, and perhaps to also investigate the deeper question of how that formation occurred. The *HST* has spent several 1000s of orbits over two decades on the various *Hubble* (Ultra)

Deep Fields, to understand the formation and evolution of galaxies, which existed when the universe was 500 million years old until today. This has led to interesting results on early galaxy formation and cosmic dawn at redshifts  $z \sim 6\text{--}10$  (e.g. Bouwens et al. 2011, 2015; Finkelstein et al. 2012; McLure et al. 2013; Schenker et al. 2013; Duncan et al. 2014; Schmidt et al. 2014; Ishigaki et al. 2018)

To extend its reach even farther beyond its native technical capabilities, *HST* observed six massive clusters of galaxies as gravitational lenses as a part of Hubble Frontier Field (HFF) program (Lotz et al. 2017). Using Director’s Discretionary (DD) observing time, the HFF program (FF program 13495; P.I. Lotz, Co-PI: Mountain) imaged six massive clusters, Abell 2744 ( $z \sim 0.308$ ), MACSJ0416.1–2403 ( $z \sim 0.396$ ), MACSJ0717.5+3745 ( $z \sim 0.545$ ), MACSJ1149.5+2223 ( $z \sim 0.543$ ), Abell S1063 ( $z \sim 0.543$ ), Abell 370 ( $z \sim 0.543$ ), and their parallel fields, for 140 orbits each with the Advanced Camera for Surveys (ACS), and Wide Field

\* E-mail: [Rachana.Bhatawdekar@nottingham.ac.uk](mailto:Rachana.Bhatawdekar@nottingham.ac.uk)

Camera 3 (WFC3) onboard *HST* in three optical ( $F435W$ ,  $F606W$ , and  $F814W$ ) and four near-infrared ( $F105W$ ,  $F125W$ ,  $F140W$ , and  $F160W$ ) bands over two epochs. These clusters are being used as gravitational lenses to magnify faint distant galaxies, while the flanking fields are for deep observations of otherwise blank areas.

In fact, we now have a good understanding of the amount of light emitted by these early galaxies in the rest-frame UV. We are able to measure this for galaxies at a range of luminosities and masses back to this epoch of  $z \sim 7$  (e.g. Bouwens et al. 2007; Oesch et al. 2010; Grazian et al. 2011; McLure et al. 2013; Schenker et al. 2013). What this reveals is that there is a gradual steepening of the ultraviolet luminosity function (UV LF) as one goes back in time, such that there are more fainter galaxies per brighter galaxies as one probes from  $z \sim 3$  to  $z \sim 7$ . This is such that there are in fact more galaxies at these earlier times than there are today, as there are so many fainter and lower mass systems (e.g. Conselice et al. 2016). However, the issue with UV photometry and the UV LF is that these are a combination of a few physical processes. Namely, a higher star formation rate (SFR) will produce a higher UV luminosity, but this does not necessarily correlate with the underlying stellar mass of the galaxy. It is easy to imagine scenarios whereby the underlying mass is high, but because of dust or a low SFR, the UV luminosity is not as high. Therefore, it is essential to measure the stellar mass functions as well as LFs as they can be substantially different.

Whilst there is a significant amount of work done on the UV LF in these and other fields (e.g. Bouwens et al. 2015; Finkelstein et al. 2015; Laporte et al. 2016; McLeod, McLure & Dunlop 2016a; Livermore, Finkelstein & Lotz 2017; Atek et al. 2018), this reveals the ongoing star formation, while the stellar masses reveal the past formation, and is thus complementary. The stellar mass, unlike the UV luminosity, is a measure of the integrated formation and merging history of a galaxy. It includes all processes such as star formation and mergers, which contribute to building up the mass of a galaxy. In this sense, it is an excellent indicator for how galaxy formation has progressed over the epochs before it is measured, as it is the integral of all galaxy formation processes previous to the time in which we observe it. This is in contrast to the UV luminosity which is a good, albeit affected by dust, measure of the instantaneous SFR of a galaxy. This is in a sense the differential of the formation process, albeit with uncertainties due to dust. UV luminosities therefore as such do not tell us about the past history of a galaxy or how/when it formed.

There are many other reasons to search for these early galaxies and to study their stellar masses. The first is that by pushing back in time, we can find when the very first galaxies formed and therefore understand the physical causes of that early formation. Whilst this is still not possible due to technical limitations, we are able to see galaxies when they are forming at high SFRs back to this early epoch of around  $z \sim 7$ –8 and measure their stellar masses (e.g. Labb   et al. 2010; McLure et al. 2011; Duncan et al. 2014; Oesch et al. 2014; Laporte et al. 2016; Song et al. 2016).

Just as for the UV LF, the stellar mass function also evolves with time, and becomes steeper with higher redshift, such that there are more lower mass galaxies per massive/bright galaxy at higher redshift than at lower redshift (e.g. Gonz  lez et al. 2011; Duncan et al. 2014; Grazian et al. 2015; Song et al. 2016). The stellar masses of these galaxies allows us to probe the past star formation history (SFH) to reveal, and place constraints, on when the first epochs of star formation occurred and to probe whether galaxies could have caused reionization (e.g. Duncan & Conselice 2015).

As there are so many more faint and low-mass galaxies compare to massive ones, and the fact that the stellar mass function continues to be occupied by progressively lower mass galaxies, it is worth asking if, and how, this continues. At some point we may expect the mass function to turn over, such that at some mass limit we see a natural decline in the number of galaxies lower than some stellar mass limit. Stellar masses also provide information about the mass-to-light ratio of the lowest mass galaxies, which can only be probed through deep observations. Reaching these galaxies will be routine with the *JWST* (e.g. Gardner et al. 2006; Kalirai 2018), but until then our best chance to study these systems is through deep observations of lensing clusters (e.g. Bouwens et al. 2009; Kneib & Natarajan 2011; Zheng et al. 2012; Coe et al. 2013; Bradley et al. 2014; Zitrin et al. 2014). In this paper, we therefore study the MACSJ0416.1–2403 cluster and its parallel field, and combine these data to derive the first galaxy stellar mass function (GSMF) at  $z = 6$ –9 for the Frontier Fields program, by combining the *HST* imaging with *Spitzer* and ground-based VLT data.

Currently there are inconsistencies, and large uncertainties, in the best available measurements of galaxy number densities and LFs at  $z > 6$ . One of the most intriguing recent results from the deepest existing *HST* data is that there is an apparent decline of the number of galaxies at the highest redshifts,  $z \sim 6$ –11 (e.g. Bouwens et al. 2016). Based on extrapolating the UV LF at lower redshifts, we should have found more systems than the candidates discovered in the deepest *HST* data (e.g. Oesch et al. 2013). On the other hand, deep *HST* imaging of lensing clusters in the CLASH and HFF clusters have found a significant number of  $z > 6$  candidates (e.g. Atek et al. 2014; Livermore et al. 2017). This includes several lensed candidates at  $z \sim 9$  behind the HFFs (e.g. Zitrin et al. 2014). Another intriguing result from HFF studies is that the UV LF continues to remain unbroken to magnitudes as faint as  $M_{\text{UV}} = -12.5$  at  $z \sim 6$  (e.g. Livermore et al. 2017) and  $M_{\text{UV}} = -15$  at  $z \sim 9$  (e.g. Ishigaki et al. 2018), whereas more recently Bouwens et al. (2017b) and Atek et al. (2018) have reported a possible turnover in the faint-end of the UV LF beyond an absolute magnitude of  $M_{\text{UV}} = -15$  at  $z \sim 6$ . In this paper, we therefore revisit this issue by deriving the UV LF from  $z = 6$ –9 with our new measurements.

The structure of this paper is as follows: In Section 2, we describe the properties of the data used in this study. Section 3 describes in detail the method we developed to remove the massive cluster galaxies on the critical line of MACSJ0416.1–2403 cluster and the construction of multiwavelength catalogue, from 0.4 to 4.5  $\mu\text{m}$ , using *HST*, *Spitzer*, and ground-based VLT data. In this section, we also describe the photometric redshift determination and the selection criteria used to construct our sample of high-redshift galaxies at  $z = 6$ –9, along with the completeness simulations undertaken to take into account the effects of completeness and selection functions. We conclude this section describing the SED fitting method used to estimate the stellar masses for our sample. In Section 4, we present our results by including a derivation of the GSMF, UV LF, and an analysis of mass-to-light ratio of our sample of galaxies. We also present and discuss the estimated total stellar mass density (SMD), the specific SFR (sSFR) and the UV luminosity density in this section. Finally, we summarize our results and present the conclusions of this work in Section 5. Throughout this paper, we adopt a  $\Lambda$  $\text{CDM}$  cosmology with  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ ,  $\Omega_M = 0.3$ , and  $\Omega_\Lambda = 0.7$ . All magnitudes are quoted in the AB system (Oke & Gunn 1983), and a Chabrier (2003) stellar initial mass function (IMF) is used.

**Table 1.** Description of data set for the MACS0416 cluster and its parallel field. The  $5\sigma$  depths are calculated using 100s of 0.2 arcsec radius apertures in random positions for *HST* images, 0.4 arcsec radius apertures for HAWK-I image and 1.4 arcsec radius apertures for IRAC images.

Filter	MACS0416 cluster Depth ( $5\sigma$ )	MACS0416 parallel Depth ( $5\sigma$ )	Instrument
F435W	28.87	28.91	ACS
F606W	28.95	29.01	ACS
F814W	29.35	29.40	ACS
F105W	29.22	29.30	WFC3
F125W	28.95	28.02	WFC3
F140W	28.85	28.93	WFC3
F160W	28.65	28.75	WFC3
Hawk-I $K_s$	26.25	26.35	HAWK-I
IRAC 3.6	25.10	25.16	IRAC
IRAC 4.5	25.13	25.20	IRAC

## 2 THE DATA

### 2.1 *HST* data

Observations of MACSJ0416.1–2403, hereafter MACSJ0416 (RA: 04:16:08.9, Dec.: –24:04:28.7) and its parallel field (RA: 04:16:33.1, Dec.: –24:06:48.7) were carried out between 2014 January–2014 February (Epoch 1) and 2014 July–2014 September (Epoch 2) as a part of the HFF program. In this study, we use the final reduced and calibrated v1.0 mosaics and their associated weight and rms maps provided by the Space Telescope Science Institute (STScI) on the HFF website,<sup>1</sup> drizzled at 60 mas pixel scale. For a detailed description of data release, including the data reduction pipeline, exposure times, and calibration procedures, we refer the reader to the STScI data release documentation.<sup>2</sup>

We compute the depth of these *HST* images by placing 100s of 0.2 arcsec radius apertures in random positions on the images and measuring fluxes in them. Table 1 lists the resulting  $5\sigma$  limiting magnitudes for the seven bands using our detection methods. We find that the  $5\sigma$  limiting magnitudes of the cluster are lower than the parallel field (see Table 1). This is the result of the cluster field being dwarfed by the light from the massive foreground galaxies, making the effective raw depths shallower than the parallel field.

### 2.2 VLT data

HFF data by itself are inadequate for characterizing the galaxies we are interested in at  $z \gtrsim 6$ . Typically, the way to study the stellar masses and stellar populations of  $z \gtrsim 6$  galaxies is through the use of *Spitzer*/IRAC data at  $>3 \mu\text{m}$ . This is often included and used along with *HST* imaging by the use of the  $K_s$  band, the reddest ground-based filter we can obtain deep imaging data from. In fact, good-quality  $K_s$ -band data have been crucial to fully exploit *Spitzer*/IRAC data, which have poor resolution and are affected by blended sources. Finally,  $K_s$  band at  $2.2 \mu\text{m}$  also fills in the gap between the  $1.6 \mu\text{m}$  F160W band and the IRAC 3.6  $\mu\text{m}$  band.

Because of the aforementioned reasons, and in order to put better constraints on redshift estimates, we include longer wavelength  $K_s$ -band data in our analysis.  $K_s$ -band observations of MACSJ0416 fields were obtained between 2013 October and 2014 February with the High Acuity Wide-field K-band Imager (HAWK-I) on

the 8.2 m UT4 telescope at the ESO Very Large Telescope. We use the fully reduced  $K_s$ -band images that are made available to the public through the Phase 3 infrastructure of the ESO Science Archive Facility (ESO program 092.A-0472, P.I. Brammer). The full HAWK-I  $7.5 \text{ arcmin} \times 7.5 \text{ arcmin}$  field of view covers both the cluster and the parallel fields in a single pointing.

Similar to the *HST* bands, we calculate the depth of the image by placing 100s of 0.4 arcsec radius apertures in random positions on the image and measuring fluxes in them. Table 1 describes the  $5\sigma$  limiting magnitudes for the  $K_s$  band. We refer the reader to Brammer et al. (2016) for a detailed description of observations and data reduction process.

### 2.3 *Spitzer* data

The Balmer break/4000 Å break is crucial to estimate the age of stellar populations and subsequently in the measurement of galaxy stellar mass. This break is at observed wavelengths beyond  $2.4 \mu\text{m}$  at  $z > 5$ , and only the IRAC camera on board *Spitzer* can identify the break. IRAC data are also essential to put better constraints on redshift estimates in addition to obtaining robust stellar mass estimates. We therefore include *Spitzer* data in our analysis. In addition to *HST*, the *Spitzer Space Telescope* has dedicated  $\sim 1000 \text{ h}$  of DD time to observe the Frontier Fields at 3.6 and  $4.5 \mu\text{m}$  and has observed each field for  $\sim 50 \text{ h}$  in each channel. We use the final reduced mosaics made available to the public on IRSAs website<sup>3</sup> (Program ID 90258, P.I. T. Soifer). The depth of *Spitzer* data was again, similar to the *HST* and  $K_s$ -band data, computed by using a similar method of placing 100s of 1.4 arcsec apertures in random positions in the images. The  $5\sigma$  limiting magnitudes for both the channels are as listed in Table 1.

## 3 METHODS

### 3.1 Divide and Conquer

One aim of the HFF program is to find the faintest and the earliest galaxies in the Universe,  $\sim 10$ –100 times fainter than any previously studied, and examine their properties, using clusters as gravitational lenses. While the clusters provide a magnified boost to the light from background galaxies, providing us with a deeper view of the early Universe than we would obtain from just examining blank fields, the overwhelming luminosity of the brightest galaxies in the cluster impedes the detection of the faint galaxies. For this reason, it was imperative that we first model and subtract the foreground galaxies from the cluster. In this study, we choose to model and subtract galaxies on the critical line since the greatest magnification of distant objects occurs along the line of sight of the densest areas of the cluster.

Two other studies Livermore et al. (2017) and Merlin et al. (2016) have attempted to subtract the foreground light using wavelet decomposition method and GALFIT, respectively, delivering promising results. More recently, Shipley et al. (2018) have sought to model the light from the brightest cluster members that contribute significant light to the cluster using IRAF.

In this work, to subtract the massive foreground galaxies, we build upon the strategy used by Gu et al. (2013) to detect faint substructures in NGC 4889. As a first attempt, we try to model the brightest galaxies in the reddest band F160W, which is our

<sup>1</sup><http://www.stsci.edu/hst/campaigns/frontier-fields/FF-Data>

<sup>2</sup><https://archive.stsci.edu/pub/hlsp/frontier/>

<sup>3</sup><http://irsa.ipac.caltech.edu/data/SPITZER/Frontier/>

detection image, using GALAPAGOS (Barden et al. 2012) as a standalone application. GALAPAGOS is an IDL-based software that uses SExtractor (Bertin & Arnouts 1996) to detect the sources, and then fits a single Sérsic profile (Sérsic 1968) to the detected sources using GALFIT (Peng et al. 2002). Unfortunately, the resulting residuals from this procedure were unsatisfactory. This is primarily because GALAPAGOS uses only a single Sérsic component to model the galaxies. The galaxies in the MACS0416 cluster are very bright/massive and also lie in a crowded field. They therefore cannot be modelled with simply one or two Sérsic components. As the massive galaxies are embedded in a crowded field, we also need to fit and remove all the neighbouring galaxies at the same time.

To overcome these difficulties, we developed an iterative procedure, dubbed ‘Divide and Conquer’, in which we split the image into small regions, with the target bright galaxy in the centre, and make use of GALAPAGOS and GALFIT on those regions to model the small neighbouring galaxies first before attempting to model the big galaxies. This procedure is essentially a series of iterations between fitting away the small objects first in a small region, one by one, using one or more Sérsic models to simulate the galaxy of interest as well as the local sky background (whose parameters are thrown away afterwards as they are only needed to approximate the sky). We keep increasing the complexity until we get a reasonable residual (e.g. there are no oversubtracted regions in the residual). This process is repeated on all the small galaxies until we are left with the central galaxy, with all the neighbours subtracted away and the process is then repeated on the central galaxy. After we get a good fit, we do a massive simultaneous fit with all the neighbours together. We then create an image with these objects subtracted out from the original image and move on to fit the next bright galaxy.

The basic steps of our method are as follows:

(i) To subtract the bright galaxies, we select a rectangular patch covering the central bright galaxy, and its nearby region with smaller galaxies, and use GALAPAGOS to create postage stamps to acquire the first guess of the model parameters of each source in the stamp.

(ii) We then model the small neighbouring galaxies first using GALFIT, and once the small galaxies are fitted away cleanly, we model the central bright galaxy.

(iii) After all the galaxies in the rectangular patch are modelled accurately, we subtract the final model from the original image and move to the next rectangular patch and repeat the process in iteration.

In the following sections, we describe our procedure in detail.

## 3.2 Subtraction of brightest cluster galaxies

### 3.2.1 Creating postage stamp and obtaining initial guess of model parameters

Frontier Field images are sky subtracted and in units of counts s<sup>-1</sup>. The very first step that we perform, before we run GALAPAGOS, is to add sky background to the detection image (*F160W* band) and convert the image to counts. This step was necessary in order for GALFIT to produce a reliable sigma image. We then select a rectangular patch on the detection image that consists of the target central bright galaxy and the nearby region with smaller galaxies and run GALAPAGOS on it to get the postage stamp as well as the first guess of model parameters for all the sources in the stamp. For each object that was contributing light to the stamp, we obtain the initial guesses for all the model parameters such as the position within

the stamp, magnitude, effective radii, Sérsic index, axial ratio, and position angle required to produce a Sérsic model.

### 3.2.2 Modelling brightest cluster galaxies using GALFIT

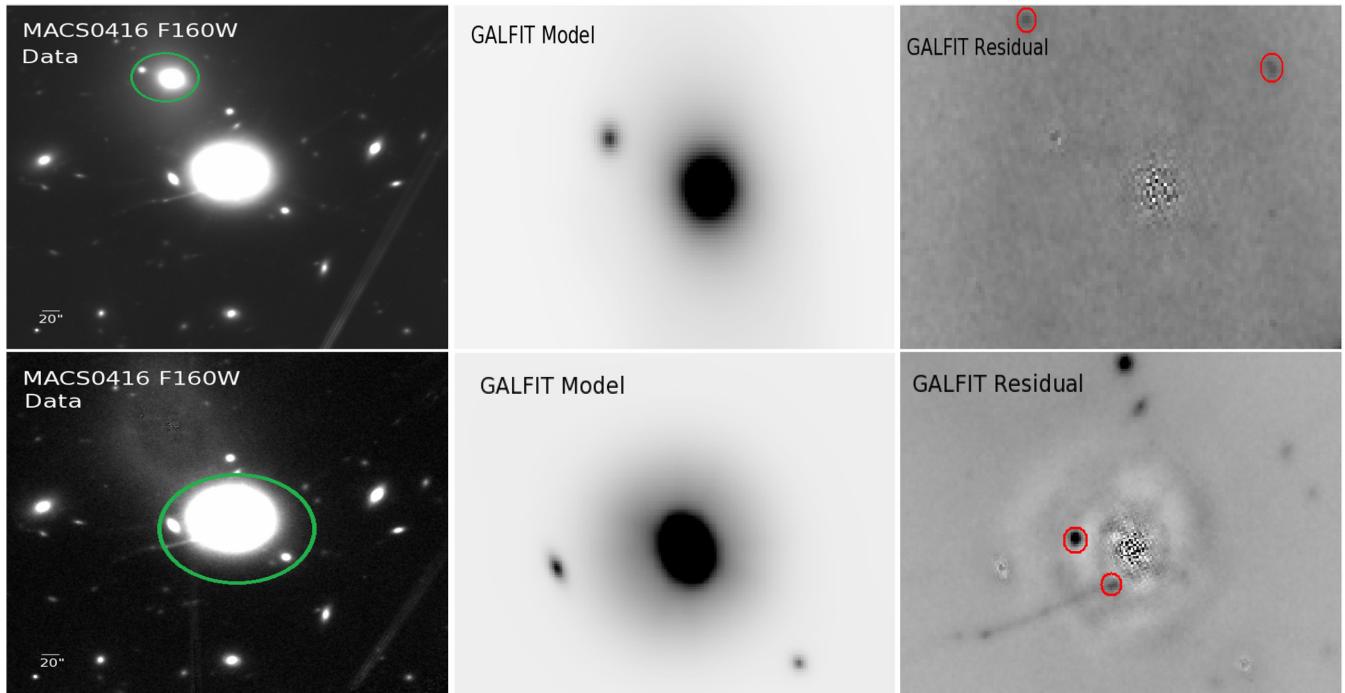
Once the postage stamp and the initial guess of model parameters are obtained, the next step is to model the galaxies. As described above, we first model the small neighbouring galaxies, along with the local background, as accurately as possible (and eventually all the bright galaxies on the critical line).

Every galaxy is modelled using a Sérsic profile described by the following expression for the surface brightness  $\Sigma(r)$  at radius  $r$ ,

$$\Sigma(r) = \sum_e \exp \left\{ -k \left[ \left( \frac{r}{r_e} \right)^{\frac{1}{n}} - 1 \right] \right\}, \quad (1)$$

where  $r_e$  is the half-light radius,  $\Sigma_e$  is the effective surface brightness,  $k$  is a normalization coefficient, and  $n$  is the Sérsic index. The de Vaucouleurs profile with index  $n = 4$  is a good fit for massive elliptical galaxies, and exponential profiles with  $n = 1$  to Gaussian  $n = 0.5$  tend to fit spiral galaxy discs and dwarf ellipticals. We start with the smallest/faintest galaxy that is contributing to the light in the stamp (with the neighbouring sources masked), with a single-component Sérsic model and the initial guesses obtained from GALAPAGOS. We run GALFIT on this system and visually inspect the residual. Although the reduced  $\chi^2$  value is the most reliable method to judge the goodness of fit, for our application, since we are not striving to obtain any physical information from the parameters, we judge the goodness of fit merely by looking at the quality of the residuals; the smoothness of the residual gives a good indication of how successful the model has been. To refine the fit, we slowly build the complexity by adding additional components and visually examining the fit and the model parameters after each trial. For example, in addition to the inadequate components required to model a galaxy, we also found that GALAPAGOS was unable to accurately estimate the radius for almost all the sources, defaulting to either a very small value or a very large value, resulting in poor residuals. As such, we found that GALFIT converged faster when we estimated the radius from the stamp image and changed this value manually. We note that this would be unsatisfactory were we interested in the properties of these galaxies. But since the aim is only to remove the foreground galaxies, this tuning by hand is not critical. This was similarly done for the other model parameters. We let all the fitting parameters free to vary but we fix them once we get a reasonable residual before moving on to the next galaxy.

This step is performed on all the neighbouring galaxies that contribute light to the stamp until we are left with the central target bright galaxy. We then repeat the process on the target bright galaxy once again until we get a reasonable residual. Finally, we do a massive simultaneous fit with all the neighbours together. Fig. 1 illustrates our method, showing the postage stamp for our target brightest cluster galaxy in the centre, surrounded by smaller neighbouring galaxies. For example, two components were required to model the smallest galaxy on the top and four components to model the bright galaxy below it (enclosed in green circle in the top left panel), revealing two faint sources (shown by red circles in the residual image). Similarly, for the massive galaxy in the middle (enclosed in green circle in the bottom left panel), six components were required, once again revealing two faint sources (shown by red circles in the residual image) that were obscured by the light of this massive galaxy. As shown in Fig. 1, our procedure clearly allows us to detect objects that are obscured by the bright galaxies once



**Figure 1.** Illustration of our multicomponent fitting procedure on the  $F160W$  image. Top row left: Postage stamp showing the target central bright galaxy to be subtracted along with neighbouring smaller galaxies. Middle: GALFIT models of the smaller systems on the top (shown by green circle in the top left panel) that we attempt to subtract first before subtracting the brightest galaxy in the centre. Two components were required to model the smallest galaxy at the top, and four components were required to model the brighter galaxy. Right: Residual after models subtraction, revealing two faint sources shown by red circles. Bottom row left: Postage stamp with the smaller systems on the top subtracted out. Middle: GALFIT model of the brightest central galaxy (shown by green circle in the bottom left panel) we want to subtract. Six components were required to model the brightest galaxy in the centre. Right: Residual image, clearly revealing the faint sources that were obscured behind it, shown by red circles.

the bright galaxies are subtracted from the cluster. We note that the number of Sérsic components that were required to model some of the galaxies are higher than that are studied usually. However, we attribute no physical significance to the total number of components; they were solely needed for the purpose of subtraction. In this work, we obtained GALFIT models in the range  $0.5 < n < 6.55$ , indicating that the brightest cluster galaxies in MACSJ0416 cluster vary greatly in profile, and hence structure and morphology.

### 3.2.3 Subtracting models from the original image

After obtaining reasonable models for all the secondary and primary sources in a stamp, we create a new object file such that the stamp pixel coordinates correspond to the pixel coordinates in the original image, along with all the other model parameters. GALFIT is then run on this new object file, resulting in an image with models only, which could be subtracted from the original image. This models-only image is then subtracted out from the original image before we move on to fit the next bright galaxy on the critical line. Steps 1 to 3 are then repeated in iteration until all the models for all the stamps are sequentially subtracted from the original image. Fig. 2 shows the final results of our procedure.

## 3.3 Multiwavelength photometry

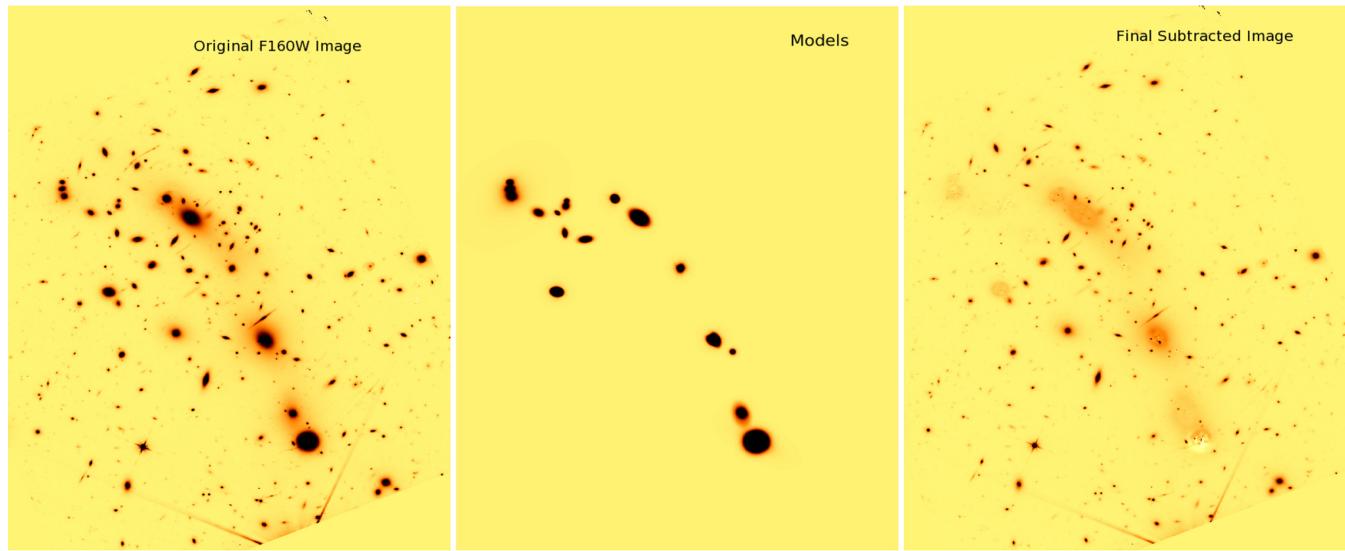
Having subtracted the massive galaxies on the critical line of the cluster in the  $H$  band, we proceed to construct a multiwavelength photometry catalogue from 0.4 to  $4.5 \mu\text{m}$ . In this section, we explain

how we obtained our photometric measurements for MACS0416 cluster and its parallel field starting from our subtracted  $H$  band.

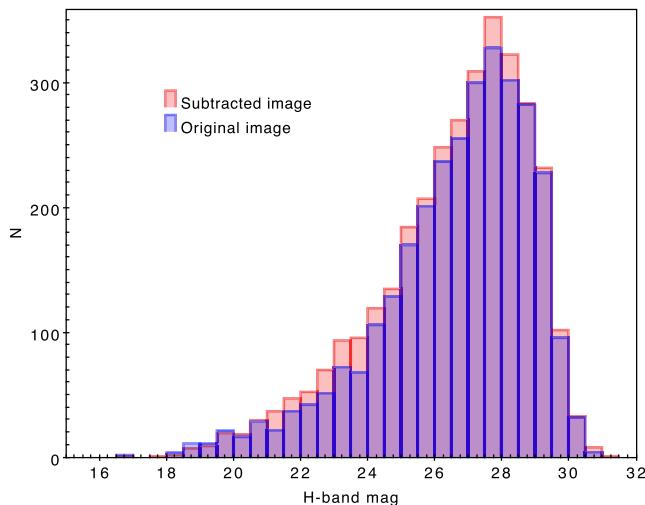
### 3.3.1 HST images

To measure accurate photometry, we first PSF-match all the other *HST* bands to the PSF of the lowest resolution  $H$  band ( $F160W$ ) using PSFMATCH task in IRAF. Empirical PSFs were generated by stacking the images of several isolated and unsaturated stars in the field. We then use SExtractor in dual image mode with our subtracted  $H$  band as detection image and use the same detections/apertures to perform photometry on the rest of the bands. SExtractor relies on the GAIN parameter to calculate accurate flux uncertainties. We therefore compute the effective gains for each band separately using  $\text{EFFECTIVE GAIN} = \text{INSTRUMENT GAIN} \times \text{EXPOSURE TIME}$  (the instrument gain is different for ACS and WFC3/IR; 2 for ACS and 2.5 for WFC3/IR) and use this value while performing photometry.

Although MAG\_AUTO (derived from flux in a flexible Kron-like elliptical aperture) generally gives the best estimate of the magnitude irrespective of the settings, it still underestimates the magnitudes for faint objects. On the other hand, MAG\_ISO computes the magnitude in an isophotal area roughly the same shape as the object, and when using separate detection and photometry images, gives you the most accurate colours (Bertin & Arnouts 1996). Following previous work (e.g. Galametz et al. 2013; Guo et al. 2013), we therefore adopt MAG\_ISO magnitudes in our analysis. Since our aim is to detect faint galaxies at  $z \gtrsim 6$ , we employ an aggressive detection strategy with following SExtractor



**Figure 2.** Results of our subtraction procedure on the MACSJ0416 cluster. Left to right: Original  $F160W$  detection image on which we applied our ‘Divide and Conquer’ procedure, models of galaxies, final subtracted image. Some of the faint sources that were obscured by the light of bright galaxies are clearly seen after the subtraction procedure.



**Figure 3.** Comparison of detections in the original  $H$ -band image and the subtracted  $H$ -band image for the MACSJ0416 cluster. The subtraction procedure has enhanced the detection of faint sources in the cluster field by  $\sim 8$  per cent.

parameters: *Detect\_minarea* 4 pixels, *Detect\_thresh* 0.7, *Deblend\_nthresh* 64 and *Deblend\_mincont* 0.0001. We employ the same method on the parallel field, with the exception of subtraction procedure, as there are no bright foreground galaxies in the parallel field. To analyse how much our subtraction procedure has enhanced the detection of faint sources in the cluster field, we run SExtractor with the same parameters on the original  $H$ -band image of MACSJ0416 cluster. We detect 3051 objects in the original  $H$  band and 3293 objects in the subtracted  $H$  band, showing that the subtraction procedure has enhanced the detection of faint sources in the cluster field by  $\sim 8$  per cent (see Fig. 3).

In order to compute stellar mass from spectral energy distribution (SED) fitting, we need to estimate the total flux in all the bands. For this, we use the same approach as Guo et al. (2013), and first

derive an aperture correction to total flux for each source in  $H$  band as  $\text{aprcor} = \text{FLUX\_AUTO}_F160W / \text{FLUX\_ISO}_F160W$ . We then convert the SExtractor isophotal fluxes and their uncertainties into  $\mu\text{Jy}$  first and then into total fluxes and uncertainties as

$$\text{FLUX\_TOTAL} = \text{aprcor} \times \text{FLUX\_ISO},$$

$$\text{FLUXERR\_TOTAL} = \text{aprcor} \times \text{FLUXERR\_ISO}.$$

We then build a catalogue of total fluxes and the associated uncertainties that combines the photometry of the MACSJ0416 cluster and the parallel field.

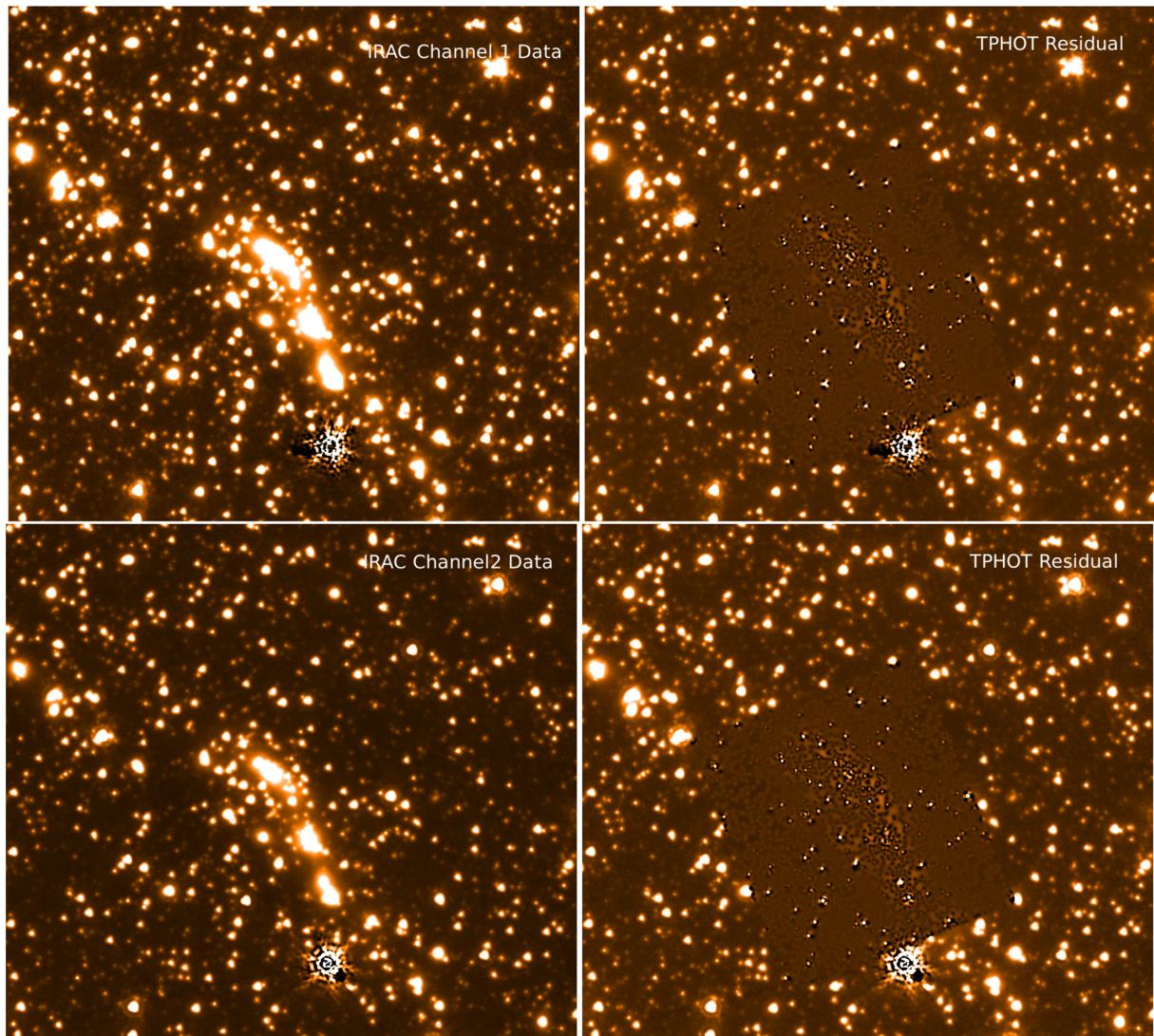
### 3.3.2 $K_s$ -band and Spitzer images

We also include longer wavelength  $K_s$ -band and *Spitzer* data in our multiwavelength photometry catalogue for reasons mentioned in Sections 2.2 and 2.3. However, extracting accurate photometry from low-resolution images ( $\sim 1.8$  arcsec FWHM PSF of *Spitzer* as opposed to  $\sim 0.18$  arcsec FWHM of WFC3) is challenging because the sources are severely blended (see Fig. 4), rendering aperture photometry and PSF fitting photometry unreliable. Also, PSF matching from *HST* to *Spitzer* data is not necessarily the best solution because we are then effectively throwing away all the useful spatial information from *HST* when we smooth it out.

To overcome these issues, we perform photometry on the  $K_s$ -band and *Spitzer* imaging using the T-PHOT code (Merlin et al. 2015), built on TFIT (Laidler et al. 2007) and CONVPHOT (De Santis et al. 2007). T-PHOT is a software designed to use information from the high-resolution image (such as source positions and morphologies), and use this information as priors to measure fluxes in the low-resolution image. T-PHOT uses a combination of input priors such as

- (i) a list of sources to obtain cut-outs from the high-resolution image,
- (ii) analytical models from codes such as GALFIT,
- (iii) unresolved, point-like sources.

In our study, we use the first method, which uses the true high-resolution priors. First, a list of source positions is obtained by SExtractor in order to obtain the high-resolution cut-outs of the



**Figure 4.** IRAC 3.4 and 4.5  $\mu\text{m}$  imaging of the MACS0416 cluster. Top left: Example of severely blended sources in IRAC channel 1 low-resolution image. Top right: Residual image from T-PHOT. Bottom left: Original IRAC channel 2 image. Bottom right: Residual image from T-PHOT.

sources as priors. The priors are then convolved with a suitable convolution kernel in order to degrade them to the resolution of the low-resolution image. These low-resolution (normalized) model templates are then placed at appropriate positions given by the high-resolution image source catalogue and scaled by a  $\chi^2$  minimization technique to give the measured fluxes of sources in the low-resolution image (see Merlin et al. 2015 for more details).

Several steps had to be performed before running T-PHOT as explained below:

(i) The units of IRAC images are in  $\text{MJy sr}^{-1}$ , whereas the units of  $K_s$  band are in  $\text{counts s}^{-1}$ . We first convert all the low-resolution images to the same flux units of  $\mu\text{Jy}$  so that the fluxes from T-PHOT can be directly fed to the photometric redshift fitting code EAZY (Brammer, van Dokkum & Coppi 2008).

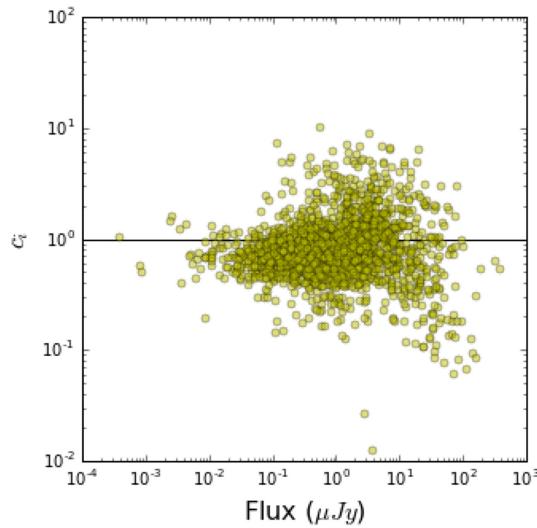
(ii) Since the resolution of IRAC is about 10 times lower than that of  $F160W$ , we use the IRAC PSF itself as the kernel to convolve the  $F160W$  images. We assume that the actual kernels would not be so different from the IRAC PSFs. For  $K_s$ -band images, however, we

run IRAF/PSFMATCH to compute the kernel. We derive empirical PSFs for both IRAC and  $K_s$  band via neighbour-masked median-scaled stacks of isolated stars.

(iii) T-PHOT requires that the lower resolution image (as well as the rms maps) be at the same orientation as the high-resolution image. We therefore resample all IRAC and  $K_s$ -band images and the rms maps to the  $F160W$  pixel scale of  $\sim 0.06$  arcsec and reproject to  $HST$  astrometry using SWARP (Bertin et al. 2002).

To correct the spatial distortion/misregistration and to obtain more astrometrically precise results, we run T-PHOT in two passes. The first pass cross-correlates the model image and the low-resolution image to compute the shifts and build a new set of kernels for regions under consideration. The second pass uses these shifted kernels to reduce the misalignment between the templates and low-resolution images to obtain more precise results (see Merlin et al. 2015).

The accuracy of this procedure with T-PHOT on IRAC images is illustrated in Fig. 4. A comparison between the original image and the residual image demonstrates that T-PHOT does a remarkably good



**Figure 5.** Covariance index  $c_i$  as a function of measured fluxes from T-PHOT for the MACS0416 cluster. 527 (16 per cent) sources from our catalogue lie above  $c_i > 1$  and are treated with caution.

job at fitting sources in the low-resolution image, with the exception of the very bright sources in the cluster. We notice that if the bright sources lie close to the faint sources, then their residuals affect the photometry of nearby faint sources significantly. For example, for such faint sources the S/N was significantly high (or negative) in the output catalogue generated by T-PHOT even if those faint sources were physically absent in the low-resolution IRAC image (e.g. Song et al. 2016 have noticed the same effect). We found that such contaminated sources caused some unfortunate high- $z$  solutions. T-PHOT helps to identify unreliable measurements for sources, especially contaminated from neighbours, with the help of a flag designated as ‘ $c_i$ ’, the covariance index (which is a ratio between a source’s maximum covariance term and its variance in the covariance matrix) flag and suggests users to treat sources with  $c_i > 1$  with caution (see Merlin et al. 2015). Fig. 5 shows the values of covariance index as a function of estimated fluxes in IRAC 3.4  $\mu m$  in the MACS0416 cluster. Out of 3293 sources in our catalogue, 527 (16 per cent) objects have  $c_i > 1$ . We therefore visually inspected each of these sources carefully to ensure that their photometric measurements are reliable.

### 3.4 Photometric redshifts and source selection

#### 3.4.1 Photometric redshifts

There are two widely used methods for computing photometric redshifts: (i) The Lyman Break (LB) (Steidel et al. 1996) also known as the ‘drop-out’ technique that relies on the large break in the continuum flux from an object in bands blueward of the Lyman break and two-colour selection in bands redward of the break and (ii) template fitting method in which photometric redshifts are derived by fitting synthetic template spectra to the observed photometry. There are several codes, such as HYPERZ (Bolzonella, Miralles & Pelló 2000), EAZY (Brammer et al. 2008), BPZ (Benítez 2000), LEPHARE<sup>4</sup> (Arnouts & Ilbert 2011), in place that use their own methods to calculate photometric redshifts. In our work,

we determine the photometric redshifts for our multiwavelength catalogue using EAZY (Brammer et al. 2008). We use the default reduced template set provided with EAZY, based on the PEGASE stellar population synthesis models of Fioc & Rocca-Volmerange (1997), which includes contribution from emission lines, and also an additional template based on the spectrum of Erb et al. (2010) that includes features such as strong optical emission lines and a high Ly $\alpha$  equivalent width; characteristics peculiar to young, unreddened, low-metallicity galaxies at high redshift, similar to Duncan et al. (2014).

#### 3.4.2 Sample selection criteria

After computing photometric redshifts, we construct a sample of galaxies in the redshift range  $5.5 \leq z \leq 9.5$ . To do this, instead of simply relying on the best-fitting redshift value, we take advantage of the full redshift probability distribution function (PDF) ( $P(z) \propto \exp(-\chi^2/2)$ , using the  $\chi^2$  distribution from EAZY, similar to previous work (e.g. Finkelstein et al. 2012, 2015; Duncan et al. 2014). We then form galaxy samples in four redshift bins centred at  $z \sim 6, 7, 8$ , and 9 with  $\Delta z = 1$ , for both the MACS0416 cluster and the parallel field, by applying a set of additional selection criteria following Duncan et al. (2014) as

$$\int_{z_s-0.5}^{z_s+0.5} p(z) dz > 0.4, \quad (2)$$

$$\int_{z_p-0.5}^{z_p+0.5} p(z) dz > 0.6, \quad (3)$$

$$(\chi^2_{\min}/N_{\text{filters}} - 1) < 3, \quad (4)$$

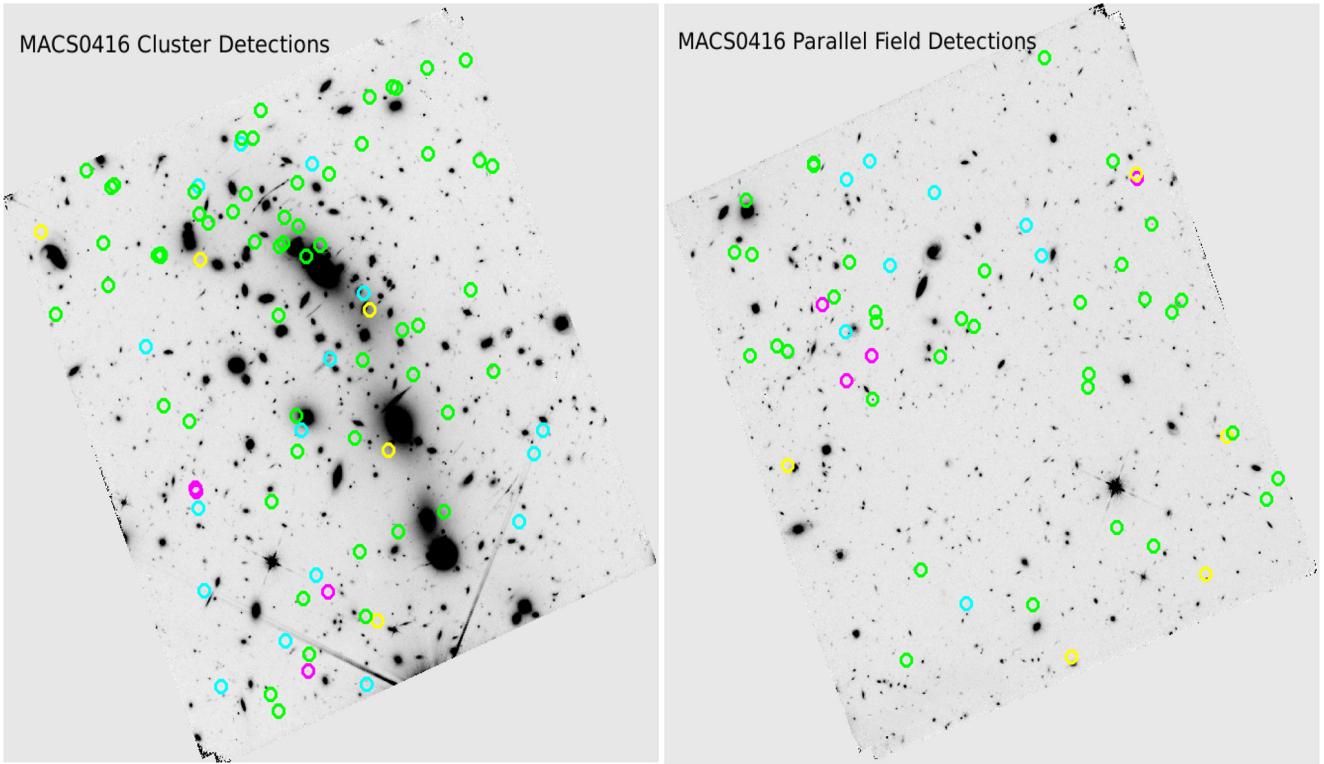
where  $z_s = 6, 7, 8$ , and 9 for the respective bins and  $z_p$  is the primary redshift peak.

With the first criterion we ensure that a significant area of the PDF occupies the redshift range of our interest. The second criterion ensures that at least 60 per cent of the PDF lies under the peak of the distribution, making sure if the high-redshift solution is picked, then it is the dominant one. With the third criterion we ensure that EAZY provides a reasonable fit. In addition to the above three criteria, we also place an S/N cut such that  $S/N(J_{125}) > 3.5$  and  $S/N(H_{160}) > 5$ . This ensures the secure detection of candidates in primary filters with very high significance, excluding spurious detections.

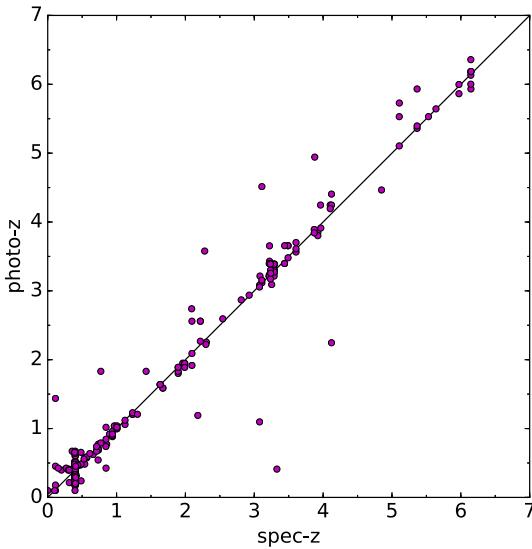
Once we have the sample of high-redshift candidates in each redshift bin for both the cluster and the parallel field, we inspect each object thoroughly to eliminate the potential contaminants such as stars, stellar diffraction spikes, sources at the edge of the images, sources with flagged photometry, etc. Our final samples then contain 134 galaxies: 82 in the MACS0416 cluster and 52 in the parallel field, out of which, 92 are at  $z \sim 6$ , 24 are at  $z \sim 7$ , 10 are at  $z \sim 8$ , and 8 are at  $z \sim 9$  (cluster and parallel field combined), as shown in Fig. 6.

In order to do comparisons between our estimated photometric redshifts and available spectroscopic redshifts, we use the published redshift catalogue of the MACS0416 cluster from the combination of the VIMOS CLASH-VLT campaign (Balestra et al. 2016) and the MUSE spectroscopic study presented in Caminha et al. (2017) [VLT programme IDs 186.A-0798, 094.A-0115(B), 094.A-0525(A)]. We match our photometric redshifts catalogue and the spectroscopic redshifts catalogue within 1 arcsec and include only those redshifts with QualityFlag = 3, 4, 5, and 9. Following Dahlen et al. (2013), we

<sup>4</sup><http://www.cfht.hawaii.edu/arnouts/LEPHARE/lephare.html>



**Figure 6.**  $H$ -band image of the MACS0416 cluster (left) and the parallel field (right). The positions of our high- $z$  galaxy sample at  $z = 6, 7, 8$ , and  $9$  are shown by green, cyan, yellow, and magenta circles, respectively.



**Figure 7.** Photometric redshift and spectroscopic redshift comparison for sources in our catalogue for the MACS0416 cluster, with spectroscopic redshifts from the published redshift catalogue of the MACS0416 cluster from the combination of the VIMOS CLASH-VLT campaign (Balestra et al. 2016) and the MUSE spectroscopic study presented in Caminha et al. (2017).

define outliers as  $|\Delta z/(1 + z_{\text{spec}})| \geq 0.15$ , where  $\Delta z = (z_{\text{spec}} - z_{\text{phot}})$ . After excluding the outliers, we compute  $\Delta z/(1 + z_{\text{spec}})$  and find  $\sigma_{\Delta z/(1+z_{\text{spec}})} = 0.041$ . The comparison between photometric and spectroscopic redshifts for the MACS0416 cluster are shown in Fig. 7.

### 3.4.3 Comparison of $8.5 \leq z \leq 9.5$ sample with previous work

Following the comparison of our photometric redshifts with available spectroscopic redshifts, we now compare our sample of high-redshift candidates with previous studies. While the number counts of our lower redshift sources are in general agreement with previous work such as Livermore et al. (2017), given the strong interest in the very high redshift candidates, in this section we only directly compare our highest redshift sample at  $8.5 \leq z \leq 9.5$  with previous works of Laporte et al. (2015), McLeod et al. (2016), and Castellano et al. (2016), who study MACS0416 and its parallel field using all the bands ( $HST$ , VLT, and IRAC) available for the HFF.

Laporte et al. (2015) report four candidates at  $z \sim 8$  behind the MACS0416 lensing cluster. Our results are in excellent agreement with theirs, in that we are able to recover all four of their sources in our work (our IDs 126, 1040, 1065, 393, see Tables A1 and A2), and the photometric redshift estimates agree within the error bars. McLeod et al. (2016) furthermore report two sources (their HFF2C-9-3 and HFF2C-9-5) in addition to the three sources (so five in total) found by Laporte et al. (2015) (their Y1, Y2, and Y4) at  $z \sim 8.4$ . However, these two sources are rejected in our defined selection criteria (see Section 3.4.2) and are therefore not included in our sample. Examining the parallel field, McLeod et al. (2016) report five candidates at  $z \sim 8.4$  and we recover three of these (ours IDs 1524, 1957, and 2660) with the correct photometric redshifts within the error bars. However, their HFF2P-9-2 and HFF2P-9-5 sources get rejected in our defined selection criteria.

Finally, we compare our detections with Castellano et al. (2016), who determine the photometric redshifts of their high- $z$  sample from the ASTRODEEP catalogue (Merlin et al. 2016) constructed by subtracting the massive galaxies in the cluster, similar to this

work. They report four sources at  $8.5 \leq z \leq 9.5$  in the MACS0416 cluster and we are able to recover two of them (their ID 99 and 743, our IDs 126 and 1065). Of the other two sources, one source (their ID 2385, which is their HFF2C-9-3 in McLeod et al. 2016) is rejected in our sample selection criteria and the other source (their ID 538 reported at  $z = 8.7^{+3.6}_{-3.6}$ ) appears to be a possible low- $z$  interloper at  $z = 0.27^{+1.581}_{-0.143}$  based on our analysis. In addition, Castellano et al. (2016) report eight candidates at  $z \geq 9.5$  in the MACS0416 cluster, out of which five appear to be possible noisy detections, one appears to be a potential low- $z$  interloper (at  $z = 0.14^{+0.001}_{-0.037}$  in our analysis), and the other two (their IDs 2362 and 2543) get rejected in our sample selection criteria, and therefore we are not able to recover any of their reported sources. In the parallel field, Castellano et al. (2016) find seven sources at  $8.5 \leq z \leq 9.5$ . We recover two of these (their ID 1272 and 1656, our IDs 1524 and 1957), while three of these sources get rejected by our sample selection criteria, and two appear to be potential low- $z$  interlopers at  $z = 1.51^{+0.943}_{-0.70}$  and  $z = 0.654^{+0.044}_{-0.135}$ , respectively, in our analysis. Lastly, they report two candidates at  $z \geq 9.5$ , one of which appears to be a possible noisy detection in our analysis, and the other one again seems to be a potential low- $z$  interloper at  $z = 2.22^{+7.437}_{-0.374}$ . However, because of the large uncertainty on the photometric redshift, it is difficult to ascertain the exact nature of this source.

### 3.5 Completeness simulations

Completeness correction helps us better constrain the low-mass end of the GSMF as well as the faint-end of UV LF. However, an overcorrection can lead to an artificial steepening of slope. We therefore perform an extensive set of simulations to estimate the completeness of our high-redshift sample by accounting for both the image incompleteness and the sample selection effects.

The gold standard way of doing this is by inserting 1000s of mock galaxies into the imaging data and perform the same analysis for source detection and recovery, photometric redshift estimation, and sample selection as for real data. This can be separated into two parts: (a) completeness from the method used to extract the objects – Here, we insert 1000s of fake galaxies to a blank image in the source place, assuming different intrinsic source sizes, positions, and magnitudes, and attempt to recover them with the same SExtractor parameters as for building the original catalogue and (b) selection efficiency of an object in a given sample of particular redshift because of the employed selection criteria – This involves calculating the photometric redshifts of a sample of galaxies and passing them through the same sample selection criteria as our real catalogue and estimating their recovery fraction, which is then folded into completeness from part a).

#### 3.5.1 Detection completeness

This is determined by inserting thousands of synthetic sources in the detection image ( $H$  band in our case) and attempting to recover them with the same SExtractor parameters as used for building the original catalogue.

For this, we first make an artificial list of galaxies using the IRAF/GALLIST function. GALLIST produces a list of  $x$  and  $y$  coordinates, magnitudes, morphological types, half-light radii, axial ratios, and position angles for a sample of galaxies based on user defined input. Various studies have looked into the evolution of sizes of galaxies at high redshift. For example, Grazian et al. (2012) and Ono et al. (2013) have found that faint galaxies at  $z \geq 7$

have extremely small half-light radii of 0.3–0.5 kpc. Kawamata et al. (2015) report the detection of a few lensed galaxies at  $z \sim 6$ –8 in Abell 2744 with sizes as small as 0.08 kpc. More recently, Bouwens et al. (2017a) have shown that the detection of highly magnified galaxies as a function of shear is highly dependent on galaxy sizes. With the help of simulations, they have demonstrated that only the most compact galaxies can be detected in high-shear regions, indicating that extremely faint  $z \sim 2$ –8 galaxies have near point-source profiles, resulting in smaller completeness corrections and hence shallower faint-end slopes than reported in recent HFF studies.

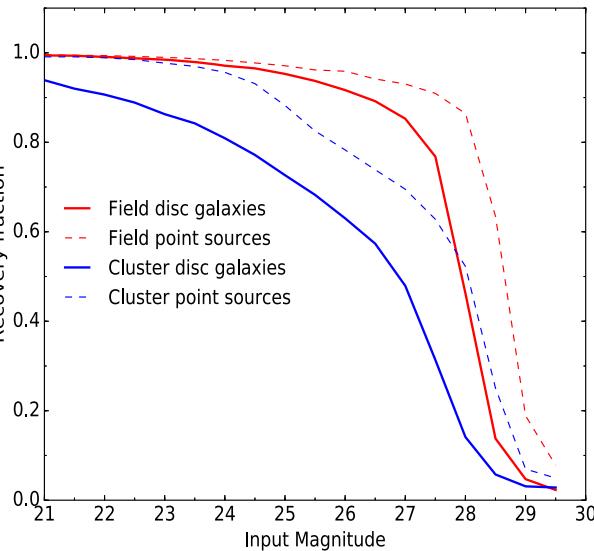
To quantify this, we perform our completeness simulations with two different methods: (i) Prompted by previous work such as Ferguson et al. (2004), Grazian et al. (2011), and Oesch et al. (2010), we draw galaxy sizes from a lognormal distribution of mean = 0.15 arcsec and  $\sigma = 0.075$  and (ii) with galaxies having near point-source profiles i.e. unresolved point sources. For both the methods, the Sérsic indices are chosen from a lognormal distribution in the range  $0.5 \leq n \leq 4.0$  with the majority of the galaxies having disc-like morphologies with  $n \leq 2$  (Ravindranath et al. 2006), whereas the position angle is selected from a uniform distribution between  $0^\circ$  and  $360^\circ$ . The axial ratio we use is also lognormal with a peak at 0.8 and the  $H$ -band magnitudes range from  $21 < H < 35$ . In each iteration of the simulation, the redshift is selected from a uniform distribution of  $z = 5$ –10. To avoid confusion due to excessive number of sources, and also due to blending with nearby sources in the field, we insert only 200 fake sources in each iteration.

The next step is to insert the mock sources in the source plane and incorporate the effect of lensing on them. Selecting the galaxy positions randomly in the source plane and then lensing them back to the image plane caused the sources to fall out of the edge of the image. We therefore choose the galaxy positions randomly in the image plane and then use IRAF/MKOBJECTS function to insert fake sources to a blank image at those positions in the source plane. This image is lensed back to the image plane using LENSTOOL (Kneib 1993; Jullo et al. 2007; Jullo & Kneib 2009) with the help of the latest lens models by the Sharon team (Johnson et al. 2014) made available to the Frontier Fields project.<sup>5</sup> The image is then convolved with the  $H$ -band PSF after its been transformed to the image plane; convolving it before the transformation made the galaxies look much bigger than they should. SExtractor is then run on this convolved image to recover the lensed galaxies and the parameters such as magnitudes, positions, half-light radii, etc. are stored in an input catalogue. These lensed, fake sources are then added to the subtracted  $H$ -band image and SExtractor (with the same parameters as used for building the original catalogue) is used to recover them and to construct a new catalogue. Finally, this catalogue is matched to the input catalogue with fake sources within a  $\sim 0.2$  arcsec matching radius.

We repeat the above process 2000 times and ultimately combine the results to calculate the recovery fraction as a function of position, input magnitude, and profile type. The same process was repeated for the parallel field, with the exception of the lensing equations.

Fig. 8 shows the recovery fraction as a function of input magnitudes for the MACS0416 cluster and the parallel field both with disc-like profile galaxies and unresolved point sources. As expected, we find that with point sources the completeness goes deeper. On the other hand, not all the sources are recovered even

<sup>5</sup><https://archive.stsci.edu/prepds/frontier/lensmodels/>



**Figure 8.** Recovery fraction as a function of  $H$ -band input magnitude for unresolved point-like (blue and red dashed lines for cluster and field, respectively) and disc-like sources (solid blue and solid red line for cluster and field, respectively) in the MACS0416 cluster and the parallel field.

at brighter magnitudes due to confusion and blending with nearby sources.

### 3.5.2 Selection function

To determine the selection function efficiency, we build a custom mock catalogue of high-redshift galaxies using the Theoretical Astrophysical Observatory (TAO; Bernyk et al. 2016). The TAO is an online cloud-based virtual laboratory that allows web-based access to mock observations of extragalactic survey data. To build our mock photometry catalogue, we use the existing CANDELS mock cone from redshift  $z = 0$  to  $z = 9$  on the TAO and generate SEDs from the single stellar populations of Bruzual & Charlot (2003), using the IMF of Chabrier (2003). Dust is then applied using the dust model of Calzetti et al. (2000) and the final catalogue is made with an  $H$ -band distribution of magnitudes in the range  $21 < H_{160} < 35$ , such that the magnitudes in the rest of the filters are determined based on the range of  $H$ -band magnitudes.

In order to generate errors for the mock photometry catalogue, we bin the fluxes of sources from our real catalogue and calculate the standard deviation and mean of the flux errors in those bins, giving us a Gaussian distribution of the errors. Photometric errors for each filter in the mock catalogue were then simulated by picking random errors each time from the Gaussian distribution in the corresponding bins. Furthermore, we ensure that the colours of the mock galaxies are a sufficiently close representation of the colours of real galaxies in our sample. For example, at  $z = 9$ , our catalogue had brighter galaxies as compared to the mock catalogue. Therefore, as there were not enough bright galaxies intrinsic to the catalogue, we rescaled the faint sources of the mock catalogue to the magnitudes we need so that the mock data samples our real data very well [see also Duncan et al. (2014), who use CANDELS mock cones similar to us and compare the colour distributions of the sources]. Following these steps, we run EAZY on this mock catalogue and compute the photometric redshifts.

Finally, to determine the selection efficiency for each of the redshift bins, we pass the high-redshift sources from the mock catalogue through the same sample selection criteria (see Section 3.4.2) as our real sample. From this, we measure the fraction of simulated galaxies that pass the selection criteria. This is then folded into the detection completeness computed in Section 3.5.1. Fig. 9 shows the selection efficiencies for the MACS0416 cluster.

## 3.6 Stellar masses, rest-frame magnitudes, and star formation rates

### 3.6.1 Stellar masses and rest-frame magnitudes

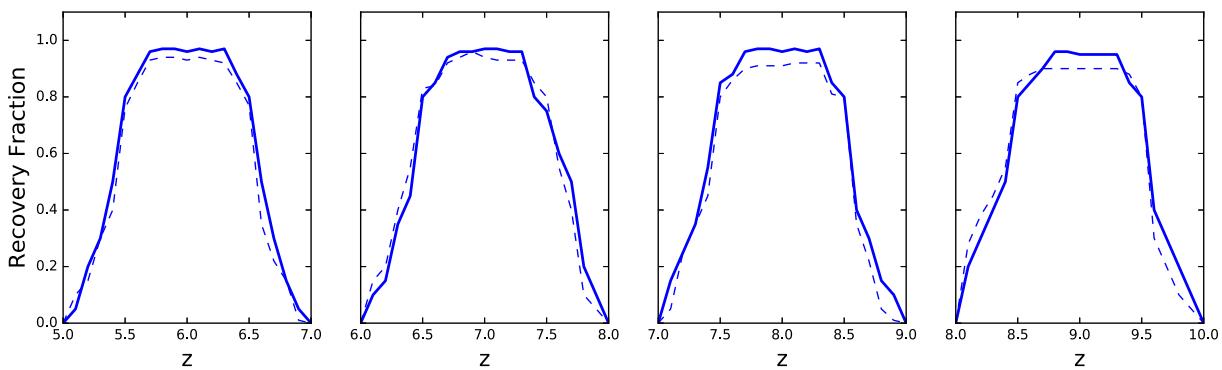
To examine the evolution of the GSF and UV LF from  $z = 6-9$ , we first estimate the stellar masses and rest-frame magnitudes for our sample in different redshift bins for both the cluster and the parallel field using a stellar population fitting technique described in detail in Duncan et al. (2014).

Briefly, we compute the stellar masses and rest-frame magnitudes using a custom template-fitting routine SMPY<sup>6</sup> (Duncan et al. 2014), which matches the observed SED of each source to one of the models it generates. First, single stellar population models of Bruzual & Charlot (2003) are used to generate synthetic SEDs for a user-defined combination of parameters such as metallicity, age, and SFH, assuming a Chabrier IMF (Chabrier 2003). The ages are allowed to vary from 5 Myr up to the age of the Universe at the redshift step being fit, dust attenuation is varied in the range  $0 \leq A_v \leq 2$ , and metallicities of 0.02, 0.2, and  $1 Z_\odot$  are used. The widely adopted parametrization of the SFH ( $SFR \propto e^{-t/\tau}$ ) is used with  $\tau = 0.05, 0.25, 0.5, 1, 2.5, 5, 10, -0.25, -0.5, -1, -2.5, -5, -10$ , and 1000 (constant SFR) Gyr, (similar to Duncan et al. 2014), where negative  $\tau$  values represent exponentially increasing histories.

Galaxy properties such as age, mass, and SFR deduced from SED fitting are affected to different degrees depending on the inclusion (or not) of nebular emission. Schaerer & de Barros (2012) discussed the importance of accounting for nebular emission in the SEDs of high-redshift galaxies while analysing a large sample of Lyman break galaxies from  $z \sim 3-6$  and found that the majority of objects were better fit with SEDs that include nebular emission, irrespective of SFH. Several other studies have also shown that significantly younger ages and lower masses could be obtained when nebular emission lines are included in SED fitting (e.g. Schaerer & de Barros 2009, 2010; Ono et al. 2010; McLure et al. 2011; Duncan et al. 2014). It has therefore become evident that nebular emission must be taken into account while understanding the photometric measurements of the SEDs of star-forming galaxies at high redshift.

Hence, we apply nebular emission lines on the model SEDs in this work, assuming an escape fraction of  $f_{\text{esc}} = 0$  (see Duncan et al. 2014 for a detailed description of the method for including nebular emission lines). Dust extinction is then applied using the law described by Calzetti et al. (2000). We then redshift each model SED in the range  $0 \leq z \leq 11$  in steps of  $\Delta z = 0.02$  and apply the attenuation by neutral hydrogen according to Madau (1995). Finally, each model spectrum is convolved through the photometric filters and the resulting SED grid is fit to the observed photometry. For each model, we compute the absolute magnitude at 1500 Å by fitting a 100 Å wide top-hat filter centred on 1500 Å. The model SEDs are fitted to the observed photometry using a Bayesian-like approach,

<sup>6</sup><https://github.com/dunkenj/smpy>



**Figure 9.** Selection efficiencies for the MACS0416 cluster, showing recovery fraction as a function of redshift and input magnitude. The solid line is for objects with  $H_{160} = 25$  and the dashed line is for  $H_{160} = 27$ , finding very similar results.

resulting in a likelihood distribution of stellar mass and rest-frame magnitudes. The stellar masses and rest-frame magnitudes thus are computed by summing the likelihoods.

The uncertainties of the stellar masses and rest-frame magnitudes are computed via a Monte Carlo analysis. For this, the observed flux of each source was perturbed by randomly choosing a point from a Gaussian distribution with standard deviation equal to the  $1\sigma$  uncertainty on the flux in a given filter. The stellar mass and rest-frame magnitude for each source is then estimated with the simulated photometry. This process is repeated 500 times and the final uncertainty is the standard deviation of the distribution of these 500 values. We employ the same method to account for errors on photometric redshifts on the estimation of stellar masses. The uncertainty due to photometric redshifts is then folded into the uncertainty from flux measurement by summing both errors in quadrature. This is then taken as the final uncertainty on stellar mass, which are found to be  $\sim 0.2$ – $0.3$  dex.

### 3.6.2 Star formation rates

To calculate the UV SFR, we use the rest-frame absolute magnitudes ( $M_{1500}$ ) obtained from the SED fitting in Section 3.6.1. We first convert the absolute magnitudes to UV luminosities, and then use the Kennicutt (1998) relation that converts the UV luminosities to SFRs assuming a Chabrier IMF:

$$\text{SFR}_{\text{UV}}(\text{M}_\odot \text{yr}^{-1}) = 1.4 \times 10^{-28} L_{\text{UV}}(\text{ergs}^{-1}\text{Hz}^{-1}). \quad (5)$$

This, however, does not take into account the dimming of light caused by dust, which can significantly affect the measurements of SFRs. In order to study the effects of dust on the UV continuum, Meurer, Heckman & Calzetti (1999) fitted the observed UV continuum of a sample of local starburst galaxies to a power law expressed by

$$f(\lambda) \propto \lambda^\beta, \quad (6)$$

where  $f(\lambda)$  is the observed flux density and  $\beta$  is the power-law index. Likewise, we measure the UV slope ( $\beta$ ) by fitting a power law to each model spectrum (see Section 3.6.1) using the 10 windows defined by Calzetti, Kinney & Storchi-Bergmann (1994) and calculate the dust extinction using the Meurer et al. (1999) relation as

$$A_{1600} = 4.43 + 1.99\beta, \quad (7)$$

which relates the UV slope  $\beta$  and the dust extinction at 1600 Å. We then use this to calculate the final dust-corrected SFRs.

### 3.7 Lensing magnification

Reliable lensing models are required in order to interpret many of the properties of background lensed galaxies accurately. For the HFF, the lens models were produced by seven independent teams for all six clusters using different methods, assumptions, and software. The primary difference between the lensing models is that some assume that the cluster mass substructure is traced by the luminous cluster galaxies, while others make no assumption about light tracing mass, and their models are instead solely constrained by lensing observables, and thus probe a broader range of possible mass distributions. We refer the reader to the MAST website<sup>7</sup> for more details of the lensing models. Five models that assume that light-traces-mass are CATS (Jauzac et al. 2014), Sharon (Johnson et al. 2014), GLAFIC (Oguri 2010; Ishigaki et al. 2015), Zitrin-NFW and Zitrin-LTM (Zitrin et al. 2013). The rest of the models that work without this assumption are provided by Bradac (Bradac et al. 2009), Williams (Grillo et al. 2015), and Merten (Merten et al. 2011).

In order to demagnify the masses and the rest-frame magnitudes, we first need to compute the per pixel magnification factor. Each lens modelling team has provided maps of mass surface density ( $\kappa$ ), and weak lensing shear ( $\gamma$ ) from which magnifications at any redshift can be derived. As a first step, we regrid the  $\gamma$  and  $\kappa$  maps to match the *HST* pixel scale and then use these maps to calculate the magnification values at each pixel using,

$$\mu = \frac{1}{(1 - \kappa)^2 - \gamma^2}, \quad (8)$$

where  $\kappa$  and  $\gamma$  both scale with the distance ratio  $D_{\text{LS}}/D_s$ ,  $D_{\text{LS}}$  being the angular diameter distance between the source and the cluster, and  $D_s$  the angular diameter distance between the observer and the source. Finally, we calculate the median magnification value,  $\mu_{\text{med}}$ , for all eight different lensing models in order to exclude possible outliers. This median magnification value was used to demagnify the masses and the rest-frame magnitudes for our sample in each redshift bin for the MACS0416 cluster (we do not demagnify the masses for parallel field).

We find that our subtraction procedure and the lensing effect allows us to probe stellar masses as low as  $10^{6.8} \text{ M}_\odot$ ; lower than previous studies (e.g. González et al. 2011; Duncan et al. 2014; Grazian et al. 2015; Song et al. 2016) at  $z \sim 6$  (see Fig. 13). Similarly, we are probing masses as low as  $10^{7.4}$ ,  $10^{7.8}$ , and  $10^{8.3} \text{ M}_\odot$  at  $z \sim 7$ ,

$z \sim 8$ , and  $z \sim 9$ , respectively. With these methods we also find that we are able to probe magnitudes as faint as  $M_{\text{UV}} = -13.5$ ,  $M_{\text{UV}} = -15.5$ ,  $M_{\text{UV}} = -15.6$ , and  $M_{\text{UV}} = -18.8$  at  $z \sim 6$ ,  $z \sim 7$ ,  $z \sim 8$ , and  $z \sim 9$ , respectively.

### 3.8 Number densities

#### 3.8.1 Volume estimation

To estimate the volume for our luminosity and mass functions we use an enhancement of the  $1/V_{\max}$  method (Schmidt 1968) introduced by Avni & Bahcall (1980) and treat our high-redshift galaxies as a ‘coherent’ sample (nomenclature of Avni & Bahcall 1980) including the cluster and the parallel field with their corresponding depths, following Avni & Bahcall (1980), Eales (1993), Ilbert et al. (2005), and Duncan et al. (2014).

Strong gravitational lensing with massive clusters enable the detection of faint galaxies by providing a magnified boost to the light from background sources. However, the disadvantage is that a high-magnification region will essentially reduce the survey volume. To account for this, we calculate the maximum observable comoving volume in which a galaxy can remain detectable as

$$V_{\text{obs},i} = \sum_k^{N_{\text{fields}}} \int_{z_{1,k}}^{z_{2,k}} \frac{dV}{dz} f(z, m, \mu) dA_k(\mu, z) dz, \quad (9)$$

where the summation,  $k$ , is over the cluster and the parallel field with their corresponding survey area,  $dA_k$ ,  $z_{1,k}$ ,  $z_{2,k}$  are the integration limits,  $f$  is the completeness function, and  $dV/dz$  is the comoving volume. For the cluster, the survey area  $dA$  is the delensed survey area in the source plane, which is a function of the magnification computed in Section 3.7 and redshift, whereas for the parallel field the survey area is just the survey area without the effects of lensing. Similarly, the completeness function  $f$ , which includes the corrections for incompleteness and the selection function calculated in Section 3.5, is a function of redshift, magnitude  $m$  and magnification factor  $\mu$  for the cluster, but does not incorporate the effects of lensing for the parallel field. The integration limits are given by  $z_{1,k} = z_{\min}$  and  $z_{2,k} = \min\{z_{\max}, z(z_j, m_j, m_{\max,k})\}$ , where  $z_{\min}$  and  $z_{\max}$  are the minimum and maximum redshift to which the source could be pushed and still be included in the sample (e.g. Avni & Bahcall 1980; Eales 1993; Ilbert et al. 2005; Duncan et al. 2014). The function,  $z(z_j, m_j, m_{\max,k})$ , gives the maximum redshift at which a source of apparent magnitude  $m_j$ , observed redshift  $z_j$ , could be observed and included in the sample given the depth  $m_{\max,k}$  of the field/cluster. So for e.g. for the  $z = 9$  sample ( $8.5 < z < 9.5$ ),  $z_{\min} = 8.5$ , and  $z_{\max} = 9.5$  in the case of the object being still sufficiently bright to be detected beyond this redshift.

Once we have the estimated volume, the number density in each magnitude (or mass) bin is calculated as

$$\phi(M) dM = \sum_i^{N_{\text{gal}}} \frac{N_i}{V_{\text{obs},i}}. \quad (10)$$

#### 3.8.2 Errors on number densities

To calculate the errors on number densities, we perform a Monte Carlo analysis in which the stellar mass (or rest-frame UV magnitude) and redshift of each galaxy is varied along a Gaussian distribution with standard deviation equal to the  $1\sigma$  uncertainty of the stellar masses (or rest-frame UV magnitudes) and redshifts. We then repeat the number density calculation explained in this

section on 500 simulated redshifts and stellar masses (or rest-frame UV magnitudes) to obtain simulated number densities. The standard deviation of these values then is the uncertainty due to measurement error. We then calculate the Poisson uncertainty for the number counts and add these in quadrature to the uncertainty from the simulations.

Furthermore, the choice of lensing maps impacts the magnification factor, the survey area and hence the effective volume. This can introduce large uncertainties, in particular at the very faint-end of the UV LF/GSMF where magnification factors exceed  $10\times$ . These uncertainties can reach a factor  $>2\times$  on the selection volume as shown in Atek et al. (2018) and therefore needs to be accounted for. For this, we calculate the effective volume and hence the number densities explained in this section using all eight lensing models used in this study. The standard deviation of these values is then the uncertainty due to lensing maps and is added in quadrature to the uncertainties from simulations and the Poisson uncertainty calculated above, which is then taken as the final uncertainty on number densities.

## 4 RESULTS

### 4.1 Ultraviolet luminosity function

One of the primary and popular methods for comparing and understanding galaxy evolution is to use the rest-frame UV LF and to measure its evolution (e.g. McLure et al. 2013; Bouwens et al. 2015; Finkelstein et al. 2015). This is largely because the UV LF is the most straightforward way of measuring a galaxy property and determining how it changes with time. The UV LF, which gives the number densities of galaxies at different UV luminosities, is the simplest measure of galaxies at these redshifts, which requires the fewest assumptions. Furthermore, the rest-frame UV LF (as well as the stellar mass function) allows us to answer fundamental questions about the way the first galaxies and stars formed. This is becoming even more true with the advent of specific predictions for the shape and normalization of the  $z > 8$  UV LF (and MF).

There are many different physical processes that are responsible for the formation of the first galaxies, and thus the creation of UV light, including the metallicity of the gas and stars, the dust content and form of extinction laws, various forms of feedback, the density of gas, as well as perhaps magnetic fields. The predictions for the formation of these first galaxies can vary significantly depending on the assumptions, and the LF is a powerful approach for understanding this issue, as models of its distribution are degenerate at lower redshifts, but differ significantly at higher redshifts.

Previous studies of LFs at a wide range of redshifts have shown that the number densities of galaxies as a function of luminosity can be characterized by a Schechter function (Schechter 1976) of the form:

$$\phi(L) = \Phi^* \left( \frac{L}{L^*} \right)^\alpha \exp \left( -\frac{L}{L^*} \right). \quad (11)$$

The Schechter function can also be expressed in log space as,

$$\phi(M) = 0.4 \ln(10) \phi^* 10^{-0.4(M-M^*)(\alpha+1)} e^{-10^{-0.4(M-M^*)}}, \quad (12)$$

where  $\phi^*$  provides the normalization,  $M^*$  (referred to as the knee of the Schechter function) corresponds to the characteristic magnitude at which the function turns over from a power law into an exponential form and  $\alpha$  is the faint-end slope.

#### 4.1.1 Best-fitting Schechter parameters and their uncertainties for UV LF

Once we have the number densities (see Section 3.8) in each luminosity bin, we proceed to determine the rest-frame UV LF at  $z = 6, 7, 8$ , and  $9$  by fitting a Schechter function as defined above to the number densities in each redshift bin.

To determine the best-fitting Schechter parameters and their uncertainties for UV LF, we use a pure-PYTHON implementation of Goodman and Weare's Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler (Goodman & Weare 2010; Foreman-Mackey et al. 2013) that examines the three-dimensional parameter space of Schechter parameters. For each redshift, we use  $10^2$  MCMC chains of  $10^4$  steps each to explore the full parameter space, building a distribution of  $M_{\text{UV}}^*$ ,  $\Phi^*$ , and  $\alpha$ . For the priors, we limit the parameter space to  $-24 < M_{\text{UV}}^* < -17$ ,  $\log(\phi^*/\text{Mpc}^{-3}) > -8$  and  $\alpha > -4$ . However at  $z = 9$ , where the sample size is small, we fix  $M_{\text{UV}}^*$  to the value estimated at  $z = 8$ .

As a first step, we compute the likelihood function, which is simply a Gaussian, and numerically optimize it. The starting position of the chain(s) is then a small Gaussian ball around the maximum likelihood result. The chains start in small distributions around the maximum likelihood values and then they quickly diverge and start exploring the full posterior distribution. To minimize the dependence of the posterior distribution on the starting point, we discard the first 10 per cent of steps in the burn-in phase before running each chain. For our final result, we join the chains together giving a distribution of  $10^6$  values of Schechter function parameters at each redshift. The best-fitting values for each Schechter function parameter are the median of this distribution, with the uncertainties covering the central 68 per cent of the distribution.

In order to investigate the effect of galaxy sizes on completeness corrections and hence on the faint-end slopes, we derive the best-fitting Schechter parameters and their uncertainties for our UV LFs in this way using the completeness curves described in Section 3.5, with both disc-like galaxies and point sources. Our results, along with the values from the literature, are listed in Table 2 and the resulting UV LFs plotted along with previous work are shown Fig. 10. At  $z = 9$ , we also show the best-fitting Schechter function derived using a simple chi squared minimization technique using disc-like galaxies. The error bars on our data points in Fig. 10 take into account the errors on photometric redshifts, errors on magnitudes, Poisson errors and uncertainties due to lensing maps, but does not include the errors due to cosmic variance.

To estimate the fractional uncertainty in the number densities due to cosmic variance, we use the cosmic variance calculator tool by Trenti & Stiavelli (2008) using the estimated completeness and survey area as inputs. The average excess uncertainty in the moderately lensed area (e.g.  $\sim 1\text{--}10 \times$  magnification) is typically only 10–30 per cent higher than an equivalent blank field (Brant Robertson, private discussion). We therefore estimate the cosmic variance of our sample by computing the blank field cosmic variance and then scaling the rms cosmic variance by  $\sim 1.2$ . For the highest magnification objects, we compute the effective area and volume for their magnification and then estimate the cosmic variance as if the sources were in a blank field of that (much smaller) size. Finally, since we are observing two fields, we estimate the joint uncertainty on the combined sample by adding the uncertainties in quadrature. We find that the cosmic variance errors span from 30 to 90 per cent from the low luminosity/mass end to the high luminosity/mass end in all redshift bins. In Fig. 10, we show the measured error on number densities for galaxies of  $M_{\text{UV}} \approx -19$ .

We find through these Schechter fits that the faint-end slope  $\alpha$  of the UV LF becomes steeper at higher redshifts, such that the ratio of lower to higher luminosity galaxies increases back to where we can measure these systems. The faint-end slope  $\alpha$  changes from  $-2.03^{+0.12}_{-0.10}$  at  $z = 6$  to  $-2.20^{+0.51}_{-0.47}$  at  $z = 9$  using disc-like galaxies in our completeness simulations. These measurements are consistent with a continuation of the trends seen from lower redshift, however more data are needed to confirm the trend at these redshifts. Bouwens et al. (2017a) suggest shallower faint-end slopes than what has been derived in some recent studies (by  $\Delta\alpha \gtrsim 0.1\text{--}0.3$ ) when simulated galaxies for completeness have near point-source profiles. When point sources are considered in our analysis, we find that the faint-end slopes are indeed shallower ( $\Delta\alpha \sim 0.09\text{--}0.12$ , see Table 2) in the redshift range probed. We note that these systematic offsets in slope are less than those estimated by Bouwens et al. (2017a) but we can attribute this to the difference in assumed profiles for input galaxies with our broader range of assumed Sérsic indices resulting in a higher completeness for the same half-light radii (i.e.  $0.5 \leq n \leq 4.0$ , versus  $n = 1$  for Bouwens et al. 2017a). We also find a decrease in  $\Phi^*$  with increasing redshift at  $z = 6\text{--}9$ , and a slight evolution in  $M_{\text{UV}}^*$  is also observed, with it decreasing with redshift from  $z = 6\text{--}8$ .

#### 4.1.2 Comparison with previous work

As can be seen from Fig. 10, there are some differences and similarities between our work and previous work<sup>8</sup>. First, at  $z \sim 6$  we observe a disagreement in the binned LF values at various luminosities for the number densities across the entire luminosity range. We, for example, do not see the upturn in the LF near  $M_{\text{UV}} = -18$  as seen by Livermore et al. (2017). We, however, also do not find the possible downturn that Atek et al. (2018) and Bouwens et al. (2017b) find in the UV LF near  $M_{\text{UV}} = -15$ . Instead, we find that at  $z \sim 6$  the LF is well fit by a Schechter function, as well as a single power law, up to the faintest limits we go,  $M_{\text{UV}} = -13.5$ . Similarly, looking at the faint-end of the LF, we find a difference in the estimated error bars between our work and Atek et al. (2018) and Bouwens et al. (2017b). Our errors bars are likely underestimated due to the fact that we have only used the median magnification factors to demagnify the magnitudes for individual sources and also have not included the errors due to cosmic variance in our analysis.

At this redshift, we generally find that our binned LF values are lower than previous studies. This can be attributed to the significantly smaller volume that we are probing as compared to Finkelstein et al. (2015), Bouwens et al. (2015, 2017b), and Atek et al. (2018) [Bouwens et al. (2015) utilize data sets from CANDELS, HUDF09, HUDF12, ERS, and the BoRG/HIPPIES programs (Trenti et al. 2011), Bouwens et al. (2017b) combine the data from the first four HFF clusters; Abell 2744, MACS 0416, MACS 0717, and MACS 1149, Finkelstein et al. (2015) use data sets from CANDELS/GOODS, HUDF, the parallel fields near MACS0416 and Abell 2744 from the HFF program, whereas Atek et al. (2018) combine the data sets of all six clusters from the HFF program]. The difference in binned LF values with Livermore et al. (2017) is, however, unclear since they are probing a similar

<sup>8</sup>Note: We restrict the comparison of our results with previous work to the best-fitting Schechter parameters derived using disc-galaxies in our completeness simulations, and do not compare our results derived using point sources.

**Table 2.** Best-fitting Schechter function parameters and their uncertainties for our UV LFs. The quoted best-fitting values and  $1\sigma$  errors of the Schechter parameters constitute the median and the central 68 per cent of posterior distribution of each parameter from our MCMC analysis.

Redshift	$M_{\text{UV}}^*$	$\alpha$	$\phi*(10^{-4} \text{Mpc}^{-3})$
$z \sim 6$			
This work (disc galaxies)	$-20.94^{+0.31}_{-0.26}$	$-2.03^{+0.12}_{-0.10}$	$2.01^{+0.85}_{-0.60}$
This work (point sources)	$-21.00^{+0.36}_{-0.30}$	$-1.93^{+0.11}_{-0.11}$	$1.61^{+0.88}_{-0.62}$
Livermore et al. (2017)	$-20.83^{+0.05}_{-0.04}$	$-2.10^{+0.04}_{-0.03}$	$2.23^{+0.273}_{-0.100}$
Finkelstein et al. (2015)	$-21.13^{+0.25}_{-0.31}$	$-2.02^{+0.10}_{-0.10}$	$1.86^{+0.94}_{-0.80}$
Bouwens et al. (2015)	$-20.94^{+0.20}_{-0.20}$	$-1.87^{+0.10}_{-0.10}$	$5.00^{+2.2}_{-1.6}$
Bouwens et al. (2017a) <sup>a,b</sup>	$-20.94$ (fixed)	$-1.91^{+0.02}_{-0.02}$	$6.6^{+0.4}_{-0.4}$
Atek et al. (2018) <sup>a</sup>	$-20.74^{+0.21}_{-0.20}$	$-1.98^{+0.11}_{-0.09}$	$3.72^{+2.31}_{-1.43}$
$z \sim 7$			
This work (disc galaxies)	$-20.85^{+0.40}_{-0.38}$	$-2.06^{+0.17}_{-0.15}$	$1.70^{+1.20}_{-0.70}$
This work (point sources)	$-21.04^{+0.45}_{-0.40}$	$-1.95^{+0.20}_{-0.18}$	$1.20^{+1.00}_{-0.51}$
Livermore et al. (2017)	$-20.80^{+0.06}_{-0.05}$	$-2.06^{+0.05}_{-0.05}$	$2.13^{+0.260}_{-0.188}$
Finkelstein et al. (2015)	$-21.03^{+0.37}_{-0.50}$	$-2.03^{+0.21}_{-0.20}$	$1.57^{+1.49}_{-0.95}$
Bouwens et al. (2015)	$-20.87^{+0.26}_{-0.26}$	$-2.06^{+0.13}_{-0.13}$	$2.9^{+2.1}_{-1.2}$
Atek et al. (2015)	$-20.89^{+0.60}_{-0.72}$	$-2.04^{+0.17}_{-0.13}$	$2.88^{+5.82}_{-1.86}$
Laporte et al. (2016)	$-20.33^{+0.37}_{-0.47}$	$-1.91^{+0.26}_{-0.27}$	$3.7^{+1.2}_{-1.1}$
Ishigaki et al. (2018)	$-20.89^{+0.17}_{-0.13}$	$-2.15^{+0.08}_{-0.06}$	$1.65^{+0.684}_{-0.484}$
$z \sim 8$			
This work (disc galaxies)	$-20.40^{+0.52}_{-0.54}$	$-2.14^{+0.37}_{-0.32}$	$1.49^{+1.83}_{-0.83}$
This work (point sources)	$-20.58^{+0.57}_{-0.58}$	$-2.02^{+0.40}_{-0.36}$	$1.09^{+1.58}_{-0.62}$
Livermore et al. (2017)	$-20.72^{+0.18}_{-0.14}$	$-2.01^{+0.08}_{-0.08}$	$1.69^{+0.592}_{-0.439}$
Finkelstein et al. (2015)	$-20.89^{+0.74}_{-1.08}$	$-2.36^{+0.54}_{-0.40}$	$0.72^{+2.52}_{-0.65}$
Bouwens et al. (2015)	$-20.63^{+0.36}_{-0.36}$	$-2.02^{+0.23}_{-0.23}$	$2.08^{+2.3}_{-1.1}$
Laporte et al. (2016)	$-20.32^{+0.49}_{-0.26}$	$-1.95^{+0.43}_{-0.40}$	$3.01^{+8.5}_{-1.9}$
Ishigaki et al. (2018)	$-20.35^{+0.20}_{-0.30}$	$-1.96^{+0.18}_{-0.15}$	$2.5^{+1.03}_{-1.25}$
$z \sim 9$			
This work (disc galaxies)	$-20.40$ (fixed)	$-2.20^{+0.51}_{-0.47}$	$0.98^{+1.65}_{-0.60}$
This work (point sources)	$-20.58$ (fixed)	$-2.11^{+0.56}_{-0.47}$	$0.55^{+1.10}_{-0.35}$
Laporte et al. (2016)	$-20.45$ (fixed)	$-2.17^{+0.41}_{-0.43}$	$0.70^{+0.30}_{-0.30}$
Ishigaki et al. (2018)	$-20.35$ (fixed)	$-1.96$ (fixed)	$1.31^{+0.266}_{-0.318}$
Mcleod et al. (2016)	$-20.1$ (fixed)	$-2.02$ (fixed)	$2.51^{+1.46}_{-1.39}$

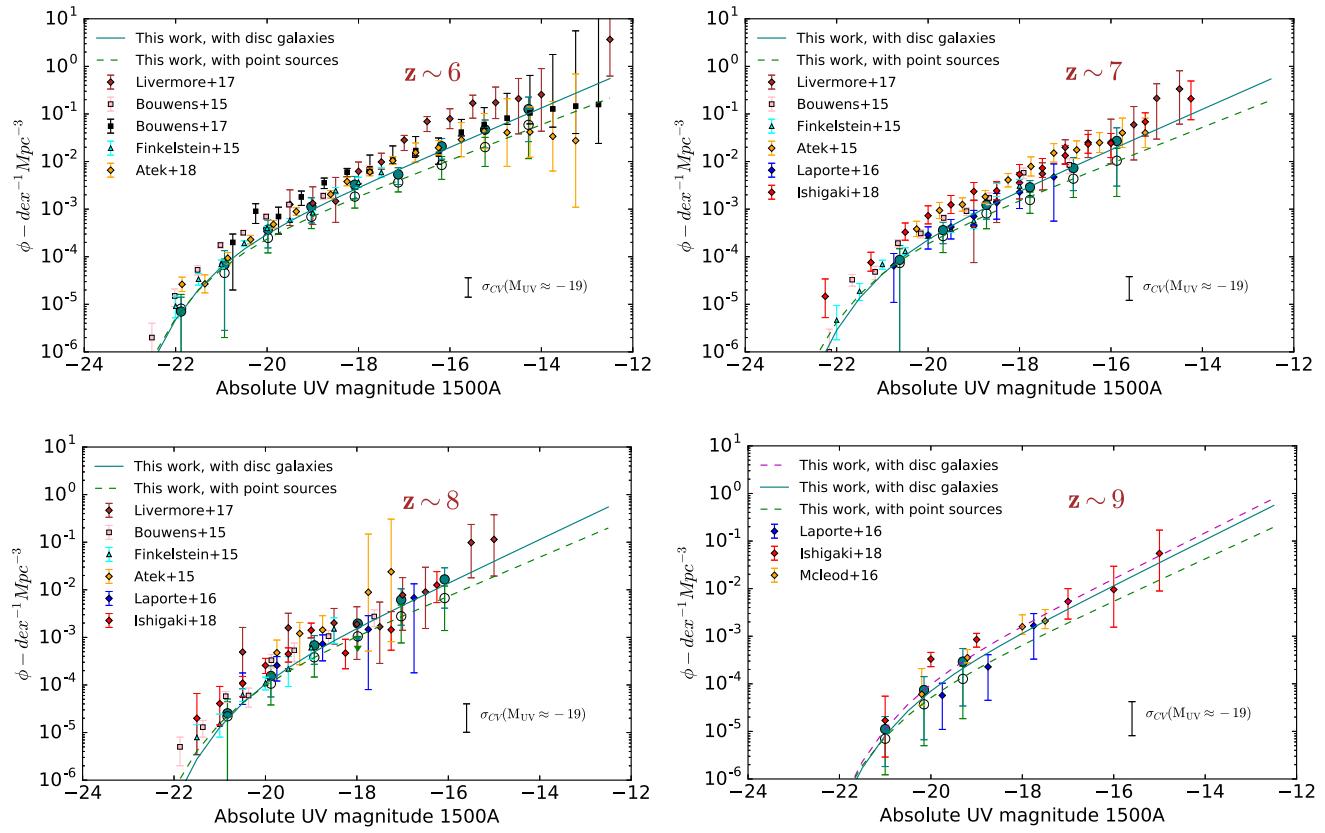
*Notes.* We only show the Schechter function parameters and their uncertainties derived using a classical Schechter function fit for Atek et al. (2018) and Bouwens et al. (2017b) and do not show the values they derive with a modified Schechter function that allows for a curvature at very faint magnitudes.

<sup>a</sup>Using classical Schechter function.

<sup>b</sup>Using the CATS model.

volume as our study. (They combine data from the Abell 2744 and the MACS0416 clusters.) Further discrepancy is observed in the Schechter function parameters. Livermore et al. (2017) are able to probe magnitudes as faint as  $M_{\text{UV}} = -12.5$ , with higher number densities resulting in a steeper faint-end slope than ours. Comparing our detections with Livermore et al. (2017), we find that their faintest source at  $M_{\text{UV}} = -12.5$  lie in the Abell 2744 cluster, which we have not included in our analysis in this study, and the next faintest source at  $M_{\text{UV}} = -14$  is in the MACS0416 cluster, but is not recovered in our analysis (it appears to be a possible noisy detection). In the case of  $M_{\text{UV}}^*$  values, we are unable to put robust constraints on  $M_{\text{UV}}^*$  as there are very few galaxies in the brightest bins ( $\sim 1-3$ ), given the smaller volume we are probing in this study.

At  $z \sim 7$ , our binned LF values are in excellent agreement with Laporte et al. (2016), who combine data from the HFF MACSJ0717 cluster and its parallel field. However, our binned LF values are lower as compared to other studies, once again possibly because of the smaller volume that we are probing. We also find that our faint-end slope is generally in agreement with previous studies, except Laporte et al. (2016) and Ishigaki et al. (2018), who are finding significantly shallower and steeper faint-end slopes, respectively. This is because we are going fainter than Laporte et al. (2016) at this redshift, whereas Ishigaki et al. (2018) are probing  $\sim 2$  magnitudes fainter than us and finding higher number densities with the complete HFF data, resulting in observed differences in the measured values of the faint-end slopes.



**Figure 10.** The UV LF at  $z = 6\text{--}9$  in the MACS0416 cluster and its parallel field. The solid green line is our best-fitting Schechter function derived from the MCMC analysis to the green circles, using disc-like galaxies in completeness simulations. The dashed green line is our best-fitting Schechter function from the MCMC analysis to the open green circles, using point sources. At  $z = 9$ , the dashed magenta line shows the best-fitting Schechter function derived using a simple chi squared minimization technique, using disc-like galaxies. The error bars on our data points take into account the errors on photometric redshifts, errors on magnitudes, Poisson errors, and uncertainties due to lensing maps, but does not include the errors due to cosmic variance. The representative error bar in the bottom right corner of each plot shows the measured error on number densities for galaxies of  $M_{\text{UV}} \approx -19$  due to cosmic variance.

Comparing to previous studies at  $z \sim 8$ , we find large discrepancies due to smaller number statistics in all the studies. For example, there are inconsistencies in the binned LF values at various luminosities. Our number densities are in agreement with Finkelstein et al. (2015) in brighter bins up to  $M_{\text{UV}} = -19$ , after which they find higher number densities in their faintest bin at  $M_{\text{UV}} = -18.5$ , possibly leading to a steeper faint-end slope than ours. Similarly, the number densities of Ishigaki et al. (2018) are higher than ours, except that they find a drastic drop in their number densities at  $M_{\text{UV}} = -18.25$ . Similarly, there is disagreement in the measured values of faint-end slopes at this redshift, in that our faint-end slope is steeper, except Finkelstein et al. (2015).

Finally, at  $z \sim 9$ , we find that our faint-end slope is in agreement with Laporte et al. (2016). McLeod et al. (2016) and Ishigaki et al. (2018) are finding shallower faint-end slopes than ours, however, we note that they held the value of faint-end slope  $\alpha$  to be fixed while fitting their data points.

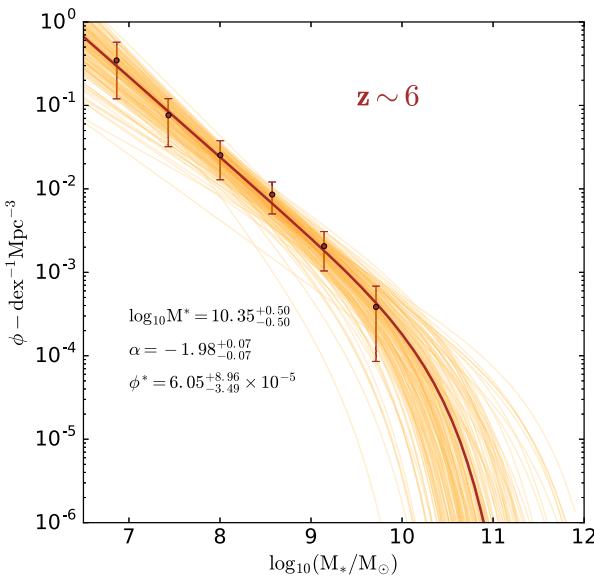
In general, at all redshifts, the difference in errors bars on the Schechter parameters between our work and previous work (e.g. Finkelstein et al. 2015; Bouwens et al. 2015; Atek et al. 2018) can be attributed to the fact that we do not include the errors due to cosmic variance on our number densities.

#### 4.2 Galaxy stellar mass function

The GMSF is a valuable mechanism to probe the growth of stellar mass, as it includes all processes such as star formation and mergers, which contribute to building up the mass of a galaxy. Cosmological simulations produce a distribution of dark matter haloes by hierarchical assembly following the initial tiny perturbations in the early young Universe. The mass distribution of these haloes follows the Schechter form and for this reason a Schechter function for the galaxy mass distribution is also expected. We therefore fit our number densities in each redshift bin derived in Section 3.8 with a Schechter function of the form:

$$\Phi(M) = \ln(10)\Phi^* 10^{(M-M^*)(\alpha+1)} e^{-10^{(M-M^*)}}, \quad (13)$$

where  $\Phi^*$  provides the normalization,  $M^*$  corresponds to the characteristic stellar mass at which the function turns over from a power law into an exponential form, and  $\alpha$  is the slope of the low-mass end.  $\alpha$  is usually negative, implying large numbers of galaxies with low masses.



**Figure 11.** Distribution of 100 random samples drawn from the chain at  $z \sim 6$  from our MCMC analysis and projected into the space of the observed data.

#### 4.2.1 Best-fitting Schechter parameters and their uncertainties for GSMF

To determine the best-fitting Schechter parameters and their uncertainties for GSMF, we perform an MCMC analysis described in 4.1.1 that examines the three-dimensional parameter space of Schechter parameters. Just as for the UV LF, for each redshift, we use  $10^2$  MCMC chains of  $10^4$  steps each to explore fully the parameter space, building a distribution of  $M^*$ ,  $\Phi^*$ , and  $\alpha$ . For the priors, we limit the parameter space to  $8 < \log(M^*/M_\odot) < 13$ ,  $\log(\phi^*/\text{Mpc}^{-3}) > -8$  and  $\alpha > -4$ . However at  $z = 9$ , where the sample size is small, we fix  $M^*$  to the value estimated at  $z = 8$ . Fig. 11 shows 100 random samples from the chain plotted on the top of our data points for our GSMF at  $z \sim 6$ .

For our final result, just as for the UV LF, we join the chains together giving a distribution of  $10^6$  values of Schechter function parameters at each redshift. The best-fitting values for each Schechter function parameter are the median of this distribution, with the uncertainties covering the central 68 per cent of the distribution. Fig. 12 shows the two-dimensional posterior probability distributions of characteristic stellar mass  $M^*$  and faint-end slope  $\alpha$  at  $z = 6-8$ .

Similar to the UV LF, we derive the best-fitting Schechter parameters and their uncertainties for our GSMF using the completeness curves described in 3.5, with both disc-like galaxies and point sources, in order to investigate the effect of galaxy sizes on completeness corrections, and hence on the faint-end slopes. Our results, along with the values from the literature, are listed in Table 3 and the resulting GSMFs plotted alongside previous work are shown in Fig. 13. Similar to UV LF, at  $z = 9$  we also show the best-fitting Schechter function derived using a simple chi squared minimization technique using disc-like galaxies. The error bars on our data points in Fig. 13 take into account the errors on photometric redshifts, errors on magnitudes, Poisson errors, and uncertainties due to lensing maps, but does not include the errors due to cosmic variance. We estimate the fractional uncertainty in number densities due to cosmic variance using the method described in Section 4.1.1

and in Fig. 13 we show the measured error on number densities for galaxies of  $10^{8.5} M_\odot$ .

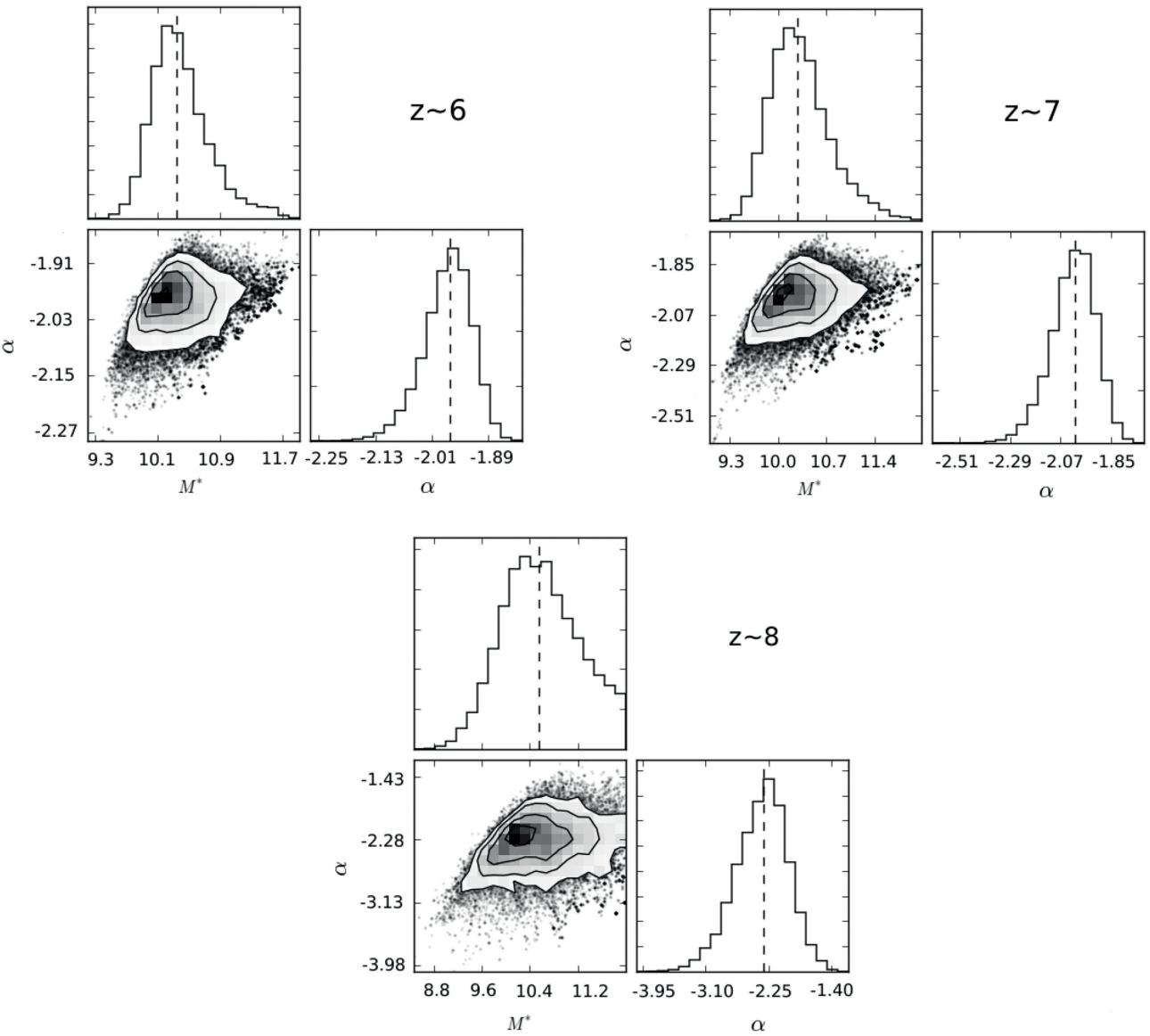
Examining our best-fitting Schechter parameters, our results show a steepening of the low-mass end slope  $\alpha$  with increasing redshift ( $-1.98_{-0.07}^{+0.07}$ ,  $-2.01_{-0.13}^{+0.17}$ ,  $-2.30_{-0.46}^{+0.51}$ , and  $-2.38_{-0.88}^{+0.72}$  at  $z = 6, 7, 8$ , and  $9$ , respectively). However, we cannot rule out a constant  $\alpha$  between  $z = 6-9$  and more data are needed to confirm the trend at these redshifts. There is also a decrease in  $\Phi^*$  with increasing redshift, but no evolution in  $M^*$  is observed. We notice a similar trend for the low-mass end slope  $\alpha$  (steepening with increasing redshift) when point sources are considered in our completeness simulations, albeit with a shallower slope as compared to disc galaxies ( $\Delta\alpha \sim 0.09-0.13$ ). A decrease in  $\Phi^*$  with increasing redshift is also observed in the case of point sources, but once again no evolution in  $M^*$  is observed. Fig. 14 shows the redshift evolution of the Schechter parameters for the stellar mass functions along with the values from the literature.

#### 4.2.2 Comparison with previous work

Fig. 13 shows the comparison of our GSMF results at  $z = 6-9$  with previous literature work (see footnote 8). The first thing that we notice here is the dissimilarity of normalization values between our work and the literature values at all redshifts. We are finding higher normalization values than González et al. (2011), Duncan et al. (2014), Grazian et al. (2015), and Song et al. (2016) at all redshifts, except at  $z \sim 6$  where the normalization values of Grazian et al. (2015) are higher than ours. The reason for this discrepancy is unclear, however, it could be attributed to the shallower slopes of our best fitted  $\log_10(M_*/M_\odot) - M_{\text{UV}}$  relation that we are finding (see Section 4.3) as compared to previous studies (see Table 4). A shallower slope of  $M_* - M_{\text{UV}}$  results into higher normalization values and steeper low-mass end slope (Song et al. 2016). Although, we do not use our  $M_* - M_{\text{UV}}$  relation to calculate our GSMF, we point out that our results of higher normalization values and steep low-mass end slope do strengthen the argument.

Inspecting the high-mass end of our GSMFs, we find a deficit of bright/massive galaxies in comparison to the literature, making it difficult for us to put any robust constraints on the high-mass end of the GSMF. The lack of brighter galaxies can be attributed to the smaller survey area that we are probing in our work. Interestingly, we notice that the massive galaxies that we find in our study all come from the parallel field alone (considering the relatively larger volume that it allows us to probe as compared to the cluster) i.e. the cluster area is devoid of massive galaxies. At  $z \sim 6$ , the number densities of our highest mass bin ( $\log(M_*/M_\odot) = 9.7-10.3$ ) are in agreement with Duncan et al. (2014) and Grazian et al. (2015), but higher than Song et al. (2016) and González et al. (2011). At  $z \sim 7$ , the number densities in our highest mass bin are in agreement with González et al. (2011), are lower than Duncan et al. (2014), but once again higher than Song et al. (2016).

At the low-mass end, even though our survey volume is significantly smaller than those of other studies, the strong gravitational lensing effect allows us to probe masses as low as  $10^{6.8} M_\odot$ , enabling us to put robust constraints on the low-mass end slope  $\alpha$ . Since the lowest mass bin of Grazian et al. (2015) is  $\log(M_*/M_\odot) \sim 9$ , we compare our low-mass end of the GSMF with González et al. (2011), Duncan et al. (2014), and Song et al. (2016), who are able to reach lower in masses. At  $z \sim 6$ , our number densities are higher than these studies. We also find a steep faint-end slope  $\alpha$  of  $-1.98_{-0.07}^{+0.07}$ , which is steeper than all the literature values except Duncan et al.



**Figure 12.** Corner plots showing the two-dimensional posterior probability distributions of characteristic stellar mass  $M^*$  and faint-end slope  $\alpha$  at  $z = 6, 7$ , and  $8$  with contours shown at  $0.5, 1, 1.5$ , and  $2\sigma$ . Top left: Posterior probability distribution at  $z \sim 6$ . Top right: Posterior probability distribution at  $z \sim 7$ . Bottom: Posterior probability distribution at  $z \sim 8$ . The marginalized distribution for each parameter is shown independently in the histograms, with the dashed line being the median value of the distribution.

(2014), who are finding a slope of  $-2.00^{+0.57}_{-0.40}$ , albeit with large error bars. At  $z \sim 7$ , we are again finding higher number densities and a steeper slope than previous studies. This could be attributed to the shallower slopes of our best fitted  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation that we are finding as mentioned earlier in this section. Finally, at  $z \sim 8$ , we are finding higher number densities and steeper slope than Song et al. (2016).

#### 4.3 $M_*$ – $M_{\text{UV}}$ relation

Scaling relations that describe the connection between the physical properties of galaxies such as metallicity, SFR, stellar mass, luminosity, etc. provide meaningful insights on galaxy formation and evolution. Using the stellar mass and the rest-frame absolute UV magnitudes estimated in Section 3.6.1, we now examine the mass-to-light ratios of our sample for each of our redshift bins to

determine the scaling relation between these properties and how it extends to the lowest masses.

For this, we derive the best-fitting  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation by fitting a linear function to the median masses of the sample in each rest-frame absolute UV magnitude bin of 0.5 mag. Fig. 15 shows the best-fitting relation and Table 4 lists the best-fitting values for our sample in each redshift bin. Whilst we notice a few high-mass galaxies with faint UV magnitudes, there is a clear positive linear trend of increasing stellar mass with increasing rest-frame absolute magnitude (with a large scatter in all the redshift bins).

At  $z \sim 6$ , our results are in agreement with Duncan et al. (2014) at the bright end ( $M_{\text{UV}} \leq -20$ ) when they include nebular emission lines while estimating their masses, but their stellar masses at the faint-end ( $M_{\text{UV}} \geq -19$ ) are lower than ours, resulting in a steeper slope. Comparing with Song et al. (2016), their estimated masses with the inclusion of nebular emission lines at the bright end ( $M_{\text{UV}} \leq$

**Table 3.** Best-fitting Schechter function parameters and their uncertainties for our GSMFs. The quoted best-fitting values and  $1\sigma$  errors of the Schechter parameters constitute the median and the central 68 per cent of posterior distribution of each parameter from our MCMC analysis.

Redshift	$\log_{10}M^*$	$\alpha$	$\phi*(10^{-5} \text{ Mpc}^{-3})$
$z \sim 6$			
This work (disc galaxies)	$10.35^{+0.50}_{-0.50}$	$-1.98^{+0.07}_{-0.07}$	$6.05^{+8.96}_{-3.49}$
This work (point sources)	$10.29^{+0.65}_{-0.67}$	$-1.89^{+0.09}_{-0.10}$	$5.43^{+8.16}_{-3.29}$
Song et al. (2016)	$10.72^{+0.29}_{-0.30}$	$-1.91^{+0.09}_{-0.09}$	$1.35^{+1.66}_{-0.75}$
Duncan et al. (2014)	$10.87^{+1.13}_{-0.54}$	$-2.00^{+0.57}_{-0.40}$	$1.40^{+41.1}_{-1.4}$
Grazian et al. (2015)	$10.49^{+0.32}_{-0.32}$	$-1.55^{+0.19}_{-0.19}$	$6.91^{+13.5}_{-4.57}$
$z \sim 7$			
This work (disc galaxies)	$10.27^{+0.60}_{-0.67}$	$-2.01^{+0.17}_{-0.13}$	$3.90^{+9.20}_{-2.85}$
This work (point sources)	$10.25^{+0.67}_{-0.71}$	$-1.91^{+0.18}_{-0.14}$	$2.93^{+6.29}_{-1.99}$
Song et al. (2016)	$10.78^{+0.29}_{-0.28}$	$-1.95^{+0.18}_{-0.18}$	$0.53^{+1.10}_{-0.38}$
Duncan et al. (2014)	$10.51^{+0.36}_{-0.32}$	$-1.89^{+1.39}_{-0.61}$	$3.6^{+3.01}_{-0.35}$
Grazian et al. (2015)	$10.69^{+1.58}_{-1.58}$	$-1.88^{+0.36}_{-0.36}$	$0.57^{+59.68}_{-0.56}$
$z \sim 8$			
This work (disc galaxies)	$10.54^{+1.00}_{-0.94}$	$-2.30^{+0.51}_{-0.46}$	$0.095^{+0.56}_{-0.080}$
This work (point sources)	$10.48^{+1.19}_{-0.92}$	$-2.17^{+0.55}_{-0.53}$	$0.090^{+0.51}_{-0.078}$
Song et al. (2016)	$10.72^{+0.29}_{-0.29}$	$-2.25^{+0.72}_{-0.35}$	$0.035^{+0.246}_{-0.030}$
$z \sim 9$			
This work (disc galaxies)	10.54 (fixed)	$-2.38^{+0.72}_{-0.88}$	$0.057^{+0.65}_{-0.050}$
This work (point sources)	10.48 (fixed)	$-2.28^{+0.83}_{-0.98}$	$0.045^{+0.63}_{-0.040}$

$-20$ ) are higher than ours, but they are finding lower measurements of stellar masses at the faint-end ( $M_{\text{UV}} \geq 19$ ), resulting in a steeper slope than ours. González et al. (2011) estimate their masses without the inclusion of nebular emission lines and find higher estimates of stellar masses at  $M_{\text{UV}} \leq -20$ , but obtain lower masses at  $M_{\text{UV}} \geq -19$ , once again resulting in a steeper slope than we find in this study. We discover a similar trend when comparing our results to previous studies at  $z \sim 7$  and  $z \sim 8$ , in that previous studies (e.g. Duncan et al. 2014; Song et al. 2016) are finding lower masses in faint UV bins at  $M_{\text{UV}} \geq -19$ , resulting in steeper slopes than ours.

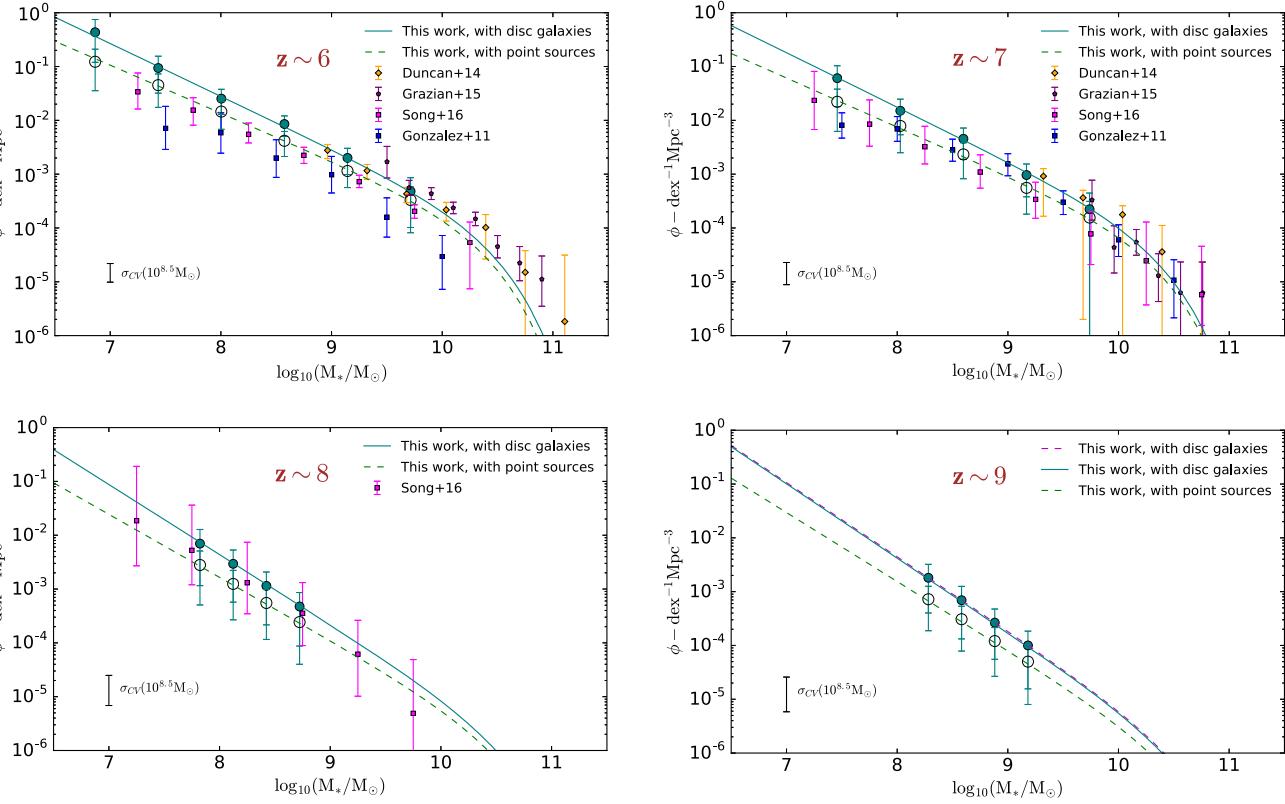
In general, we find that at fainter magnitudes, our estimated masses are  $\sim 0.2$  dex higher than previous studies. It is not clear why this is the case but this could be due to a couple of reasons: The discrepancy is likely a result of different fitting methods for stellar masses giving results which can differ by  $\sim 0.2$  dex between otherwise similar methods (e.g. Mobasher et al. 2015). The inconsistency could also be a result of IRAC deblending. Although we rule out the possibility of IRAC deblending by carefully inspecting our sample, we do notice that in spite of our deep IRAC data, the S/N of galaxies in faint UV bins is usually low ( $\lesssim 3\sigma$ ). A detailed stacking analysis by Song et al. (2016) reveals that at faint UV luminosities, the stacked points are usually lower than the median values, reflecting that the stellar masses of low S/N galaxies in the faint UV bins are on average biased towards higher masses. However, see also Behroozi et al. (2018) for a further discussion on systematic differences in the UV mass-to-light ratios, where they recalculate the median UV-stellar mass relations from the Song et al. (2016) SED stacks for  $z = 4-8$  galaxies, and argue that the masses found by them are too low, possibly a result of implausibly young ages being fit while estimating the stellar masses.

In order to see the effect of  $\lesssim 3\sigma$  sources in faint UV bins on the best-fitting  $\log_{10}(M_*/\text{M}_\odot) - M_{\text{UV}}$  relation, we implement two

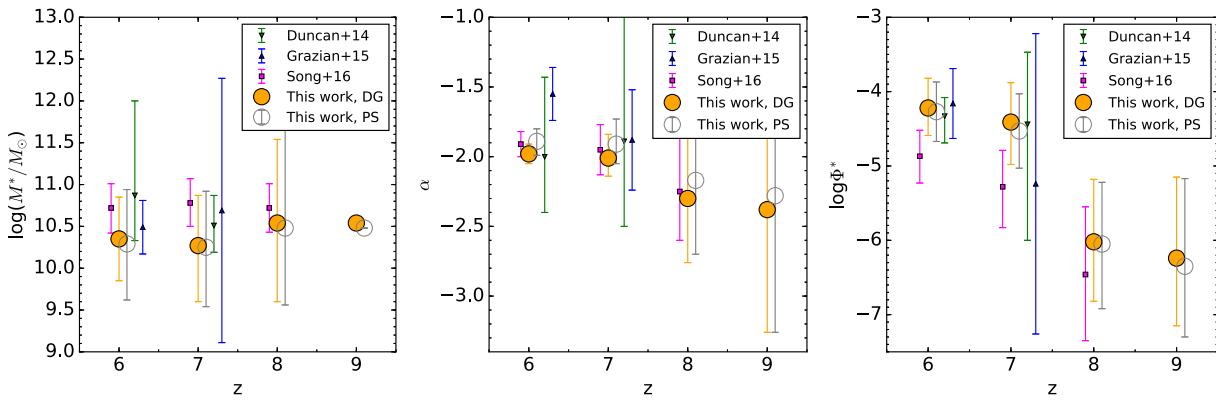
fits to our data. By performing a linear fit to the full luminosity range probed in our study, we find that we are observing shallower slopes in all the redshift bins than previous studies of González et al. (2011), Duncan et al. (2014), and Song et al. (2016). This could be the reason for the higher observed normalization values in our derived GSMFs at all redshifts as mentioned previously in Section 4.2.2. However, by restricting our analysis to brighter magnitudes ( $M_{\text{UV}} \leq -16$ ) at  $z \sim 6$ ,  $z \sim 7$ , and  $z \sim 8$  and excluding the  $\lesssim 3\sigma$  points from our sample, we observe a steepening of slopes in comparison (see Table 4), but these are still shallower than previous studies. This furthermore highlights the importance of deeper imaging with JWST to obtain accurate photometry of galaxies at  $> 2 \mu\text{m}$ . Nevertheless, in both the cases, we find that the slopes of our best fitted  $\log_{10}(M_*/\text{M}_\odot) - M_{\text{UV}}$  relation are close to a constant mass-to-light ratio of  $-0.40$ , suggesting no strong evolution of mass-to-light ratio with luminosity. We notice that normalization, on the other hand, evolves very weakly from  $z \sim 6$  to  $z \sim 9$ , with a decrease in normalization with increasing redshift.

#### 4.4 Stellar mass density

To estimate the total stellar mass density (SMD) at  $z = 6-9$ , we integrate the best-fitting Schechter function in each redshift bin from  $M_* = 10^8$  to  $10^{13} \text{ M}_\odot$ . These limits were chosen so as to allow us to compare with the SMD values in the literature. We estimate the  $1\sigma$  uncertainties as the minimum and maximum range of stellar mass densities within the  $1\sigma$  contours of Schechter parameters obtained from our MCMC analysis in Section 4.2.1. Table 5 lists our estimates of SMD along with their  $1\sigma$  uncertainties and Fig. 16 shows the evolution of the SMD. Our results are shown as solid orange points when disc-like galaxies are considered in our completeness simulations, and as open grey circles considering point sources. In



**Figure 13.** The GSMF at  $z = 6\text{--}9$  in the MACS0416 cluster and its parallel field. The solid green line is our best-fitting Schechter function from our MCMC analysis to the green circles, considering disc-like galaxies in completeness simulations. The dashed green line is our best-fitting Schechter function from MCMC analysis to the open green circles, using point sources. At  $z = 9$ , the dashed magenta line shows the best-fitting Schechter function derived using a simple chi squared minimization technique, using disc-like galaxies. The error bars on our data points take into account errors on photometric redshifts, errors on stellar mass, Poisson errors, and uncertainties due to lensing maps, but does not include errors due to cosmic variance. The representative error bar in the bottom left corner of each plot shows the measured error on number densities for galaxies of  $10^{8.5} M_\odot$  due to cosmic variance.



**Figure 14.** Redshift evolution of best-fitting Schechter function parameters for our stellar mass functions. The solid orange circles represent the best-fitting Schechter parameters considering disc-like galaxies in our completeness simulations, whereas the open circles are our results considering point sources. The low-mass end slope  $\alpha$  steepens with increasing redshift, number density  $\Phi^*$  decreases with increasing redshift, whereas no evolution in the characteristic mass  $M^*$  is observed. Note. The error bars on characteristic mass  $M^*$  at  $z \sim 9$  are not shown since we keep the value of  $M^*$  fix to the value estimated at  $z \sim 8$ .

Fig. 16, we also show the results from the literature for comparison, converted to Chabrier IMF where required.

We find that at  $z \sim 6$  and  $z \sim 7$ , our results are in agreement with González et al. (2011) and Duncan et al. (2014), but higher than

Stark et al. (2013), Grazian et al. (2015), and Song et al. (2016). At  $z \sim 6$ , our SMD estimates are higher by  $\sim 0.5$ ,  $\sim 0.6$ , and  $\sim 0.2$  dex compared to Grazian et al. (2015), Song et al. (2016), and Stark et al. (2013), respectively. Comparing at  $z \sim 7$ , the estimates are

**Table 4.** Best-fitting  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation. The values in parentheses are the best-fitting values when we restrict our analysis to brighter magnitudes ( $M_{\text{UV}} \leq -16$ ). The quoted errors represent the  $1\sigma$  uncertainties.

$z$	$\log M_{*(M_{\text{UV}}=-19.5)}$	Slope
6	$8.66 \pm 0.05$	$-0.38 \pm 0.07$
	$(8.66 \pm 0.07)$	$(-0.41 \pm 0.06)$
7	$8.56 \pm 0.08$	$-0.37 \pm 0.09$
	$(8.58 \pm 0.08)$	$(-0.40 \pm 0.05)$
8	$8.52 \pm 0.18$	$-0.38 \pm 0.14$
	$(8.50 \pm 0.16)$	$(-0.40 \pm 0.15)$
9	$8.49 \pm 0.28$	$-0.42 \pm 0.21$

higher by  $\sim 0.7$ ,  $\sim 0.6$ , and  $\sim 0.5$  dex by Grazian et al. (2015), Song et al. (2016), and Stark et al. (2013), respectively. Finally, at  $z \sim 8$ , the only previous SMD estimates are from Song et al. (2016), which are  $\sim 0.4$  dex lower than ours.

Our measurements reveal that the integrated stellar mass density has decreased by a factor of  $\sim 15^{+21}_{-6}$  from  $z \sim 6$  to  $z \sim 9$ , confirming that the process of ongoing star formation and merging increases the total stellar mass in the Universe with time. We also find that there is a surprisingly high stellar mass density for galaxies in the early universe up to  $z \sim 9$ . This is an indication that galaxies of masses around  $10^8 M_\odot$  have already formed a significant density by this time. This further indicates that the star formation and assembly history for galaxies is significant in the epochs  $z > 9$ , which we cannot probe in detail until the launch of the JWST.

#### 4.5 Specific star formation rates

Previous studies have shown that the inclusion of emission lines in the estimation of stellar masses results in higher values of sSFR at high redshift (e.g. Schaefer & de Barros 2009, 2010; Stark et al. 2013; Duncan et al. 2014; González et al. 2014). This is in contrast with the results from other studies such as Stark et al. (2009), González et al. (2010) and Bouwens et al. (2012), who show that the sSFR remains constant at  $\sim 2 \text{ Gyr}^{-1}$  with increasing redshift, and also with theoretical expectations of sSFR evolution (e.g. Davé 2008; Weinmann, Neistein & Dekel 2011). We therefore use our stellar mass and SFR estimates and investigate this further by calculating the sSFR (sSFR = SFR/ $M_*$ ) of our sample in different redshift bins.

In order to compare with previous studies and to avoid incompleteness, we calculate our sSFR in a fixed stellar mass bin of  $\log_{10}(M/M_\odot) = 9.7 \pm 0.3$ . Fig. 17 shows our results along with the values of sSFR estimated in the same mass bin from the literature and show a clear trend of increasing sSFR with redshift. We find that our sSFRs are in good agreement with Duncan et al. (2014) and González et al. (2014) at  $z \sim 6$  and  $z \sim 7$ , but are lower than Stark et al. (2013). However, it turns out that the scatter in the  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation plays a crucial role in the estimated values of sSFR, in that the sSFR will be overestimated if only the  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation is used without considering the numerous population of lower SFR sources with large mass-to-light ratios. Stark et al. (2013) do not take this into account and report their results considering zero scatter. Nevertheless, they discuss this and report that if the scatter of 0.5 dex described by González et al. (2012) is included, then this would lower the average sSFR by  $2.8 \times$  at  $z \sim 4$ .

To understand the evolution of sSFR with redshift, we fit a power law to the observed values of sSFR, giving us  $\text{sSFR} \propto (1 + z)^{2.01 \pm 0.16}$ . This is still a weaker trend than the theoretical expectations of Dekel et al. (2013), in which the specific accretion rate follows the form  $\text{sSFR} \propto (1 + z)^{2.5}$ , but is closer to that observed by Duncan et al. (2014) ( $\text{sSFR} \propto (1 + z)^{2.06 \pm 0.25}$ ).

In general, we find an increase in the sSFR when we probe galaxies at higher redshifts. This implies that there is a decline in the relative formation rate for galaxies at  $z < 9$ , and that the formation rate, relative to mass, increases at higher look back times.

#### 4.6 UV luminosity density

To calculate the UV luminosity density ( $\rho_{\text{UV}}$ ) in each redshift bin, we integrate the best-fitting Schechter parameters for our UV LFs down to a faint-end magnitude limit of  $M_{\text{UV}} = -13.5$ , which is the magnitude of the faintest galaxy in our sample. Since dust extinction decreases with both redshift and UV luminosity (Bouwens et al. 2014; Finkelstein et al. 2015), and hence has a negligible effect on estimated  $\rho_{\text{UV}}$  values, we do not dust correct the data while calculating the UV luminosity density. We then use the Kennicutt (1998) relation to convert the UV luminosity density values to cosmic SFR density (SFRD), assuming a Salpeter IMF and constant SFH over  $\geq 100$  Myr.

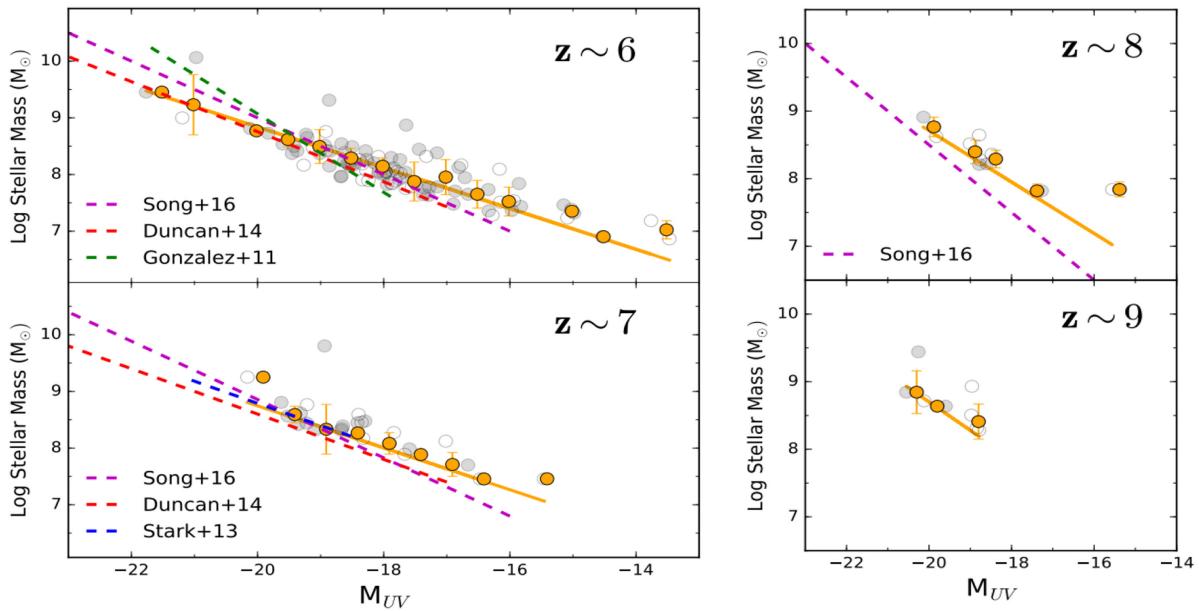
In order to facilitate comparison with other studies such as Bouwens et al. (2015), Finkelstein et al. (2015), McLeod et al. (2016), Ishigaki et al. (2018), and Oesch et al. (2018), we also estimate  $\rho_{\text{UV}}$  down to  $M_{\text{UV}} = -17$ . Fig. 18 shows the evolution of UV luminosity density as well as SFR density from  $z = 6-9$  and Table 5 shows the derived values of  $\rho_{\text{UV}}$  and  $\rho_{\text{SFR}}$  in each redshift bin. In both the cases ( $M_{\text{UV}} = -13.5$  and  $M_{\text{UV}} = -17$ ) our results support a smooth decline of  $\rho_{\text{UV}}$  towards high redshifts when disc-like galaxies are considered in our completeness simulations. However, there appears to be a slight accelerated decline of  $\rho_{\text{UV}}$  and  $\rho_{\text{SFR}}$  when using completeness results with point sources (see Fig. 18). The values estimated with point sources are shown in parentheses in Table 5. The quoted  $1\sigma$  uncertainties represent the minimum and maximum range of  $\rho_{\text{UV}}$  and  $\rho_{\text{SFR}}$  within the  $1\sigma$  contours of Schechter parameters obtained from our MCMC analysis in Section 4.1.1.

Performing a simple power-law fit to the data points estimated with disc-like galaxies at  $z \sim 6-9$ , we find that  $\rho_{\text{UV}} \propto (1 + z)^{-2.63}$  (shown by dotted yellow line in Fig. 18), whereas fitting a power law at  $z \geq 8$  values results in  $\rho_{\text{UV}} \propto (1 + z)^{-4.61}$ . This is even shallower than McLeod et al. (2016), who find that  $\rho_{\text{UV}} \propto (1 + z)^{-5.8}$  beyond  $z \geq 8$ , supporting a smooth decline in  $\rho_{\text{UV}}$ . This is in contrast with Oesch et al. (2014), Bouwens et al. (2015), Ishigaki et al. (2018), and Oesch et al. (2018), who find an accelerated decline in  $\rho_{\text{UV}}$  at  $z \geq 8$  when they integrate it down to  $M_{\text{UV}} = -17$ .

## 5 SUMMARY

In this paper, we have exploited the power of gravitational lensing of massive clusters and combined the *HST*, VLT, and *Spitzer* imaging of MACS0416 cluster and its parallel field to probe galaxy evolution with GSMFs and UV LFs out to  $z \sim 9$ . We have developed a novel method to subtract the massive foreground galaxies that lie close to the critical line from the MACS0416 cluster, allowing for a deeper and cleaner detection of the faintest systems at  $z \geq 6$ .

We have constructed a multiwavelength catalogue (from 0.4 to 4.5  $\mu\text{m}$ ) using all the bands (*HST*, *K*, and IRAC) available for the HFFs, allowing us to put better constraints on redshift estimates as well as obtain robust stellar mass estimates. From this, we have estimated the stellar masses of our high- $z$  sample through SED fitting with the



**Figure 15.** Stellar mass as a function of rest-frame absolute UV magnitude at  $1500 \text{ \AA}$  at  $z = 6\text{--}9$ . Grey filled circles represent sources with an S/N ratio  $>3\sigma$  at  $3.6 \mu\text{m}$  in IRAC, whereas grey open circles are those with an S/N ratio  $\lesssim 3\sigma$ . Filled yellow circles represent the median stellar masses in each rest-frame absolute UV magnitude bin of 0.5 mag, and the yellow error bars represent the standard deviation in stellar mass in each UV magnitude bin. The best-fitting  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation is shown by solid yellow line in each redshift bin. The best-fitting  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  lines of González et al. (2011), Stark et al. (2013), Duncan et al. (2014), and Song et al. (2016) are also shown by green, blue, red, and magenta dashed lines, respectively, for comparison.

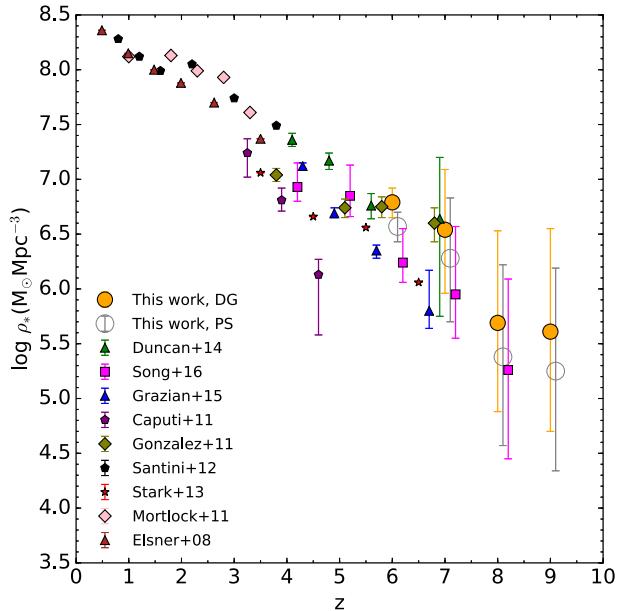
**Table 5.** Total stellar mass density estimates by integrating the best-fitting Schechter function from  $M_* = 10^8$  to  $10^{13} \text{ M}_\odot$  for our GSMFs. The quoted  $1\sigma$  error bars represent the minimum and maximum range of stellar mass densities within the  $1\sigma$  contours of Schechter parameters obtained from our MCMC analysis. Also shown in the table are values of  $\rho_{\text{UV}}$  and  $\rho_{\text{SFR}}$ . The  $\rho_{\text{UV}}$  values are integrated down to  $M_{\text{UV}} = -13.5$ , the magnitude of the faintest galaxy in our sample. The SFR densities are calculated using the (Kennicutt 1998) relation, assuming a Salpeter IMF and a constant SFH over  $\geq 100$  Myr. The values in the parentheses are computed using point sources in our completeness simulations.

$z$	$\log \rho_*$ ( $\text{M}_\odot \text{ Mpc}^{-3}$ )	$\log \rho_{\text{UV}}$ ( $\text{erg s}^{-1} \text{ Hz}^{-1} \text{ Mpc}^{-3}$ )	$\log \text{SFR density}$ ( $\text{M}_\odot \text{ yr}^{-1} \text{ Mpc}^{-3}$ )
6	$6.79^{+0.13}_{-0.12}$	$26.16^{+0.08}_{-0.07}$	$-1.69^{+0.08}_{-0.07}$
	$(6.57^{+0.13}_{-0.14})$	$(25.94^{+0.08}_{-0.07})$	$(-1.91^{+0.08}_{-0.07})$
7	$6.54^{+0.52}_{-0.55}$	$26.09^{+0.08}_{-0.08}$	$-1.76^{+0.08}_{-0.08}$
	$(6.28^{+0.55}_{-0.58})$	$(25.86^{+0.09}_{-0.09})$	$(-1.99^{+0.09}_{-0.09})$
8	$5.69^{+0.83}_{-0.81}$	$25.95^{+0.10}_{-0.09}$	$-1.90^{+0.10}_{-0.09}$
	$(5.38^{+0.84}_{-0.81})$	$(25.71^{+0.12}_{-0.10})$	$(-2.14^{+0.12}_{-0.10})$
9	$5.61^{+0.92}_{-0.90}$	$25.88^{+0.14}_{-0.13}$	$-1.97^{+0.14}_{-0.13}$
	$(5.25^{+0.94}_{-0.91})$	$(25.55^{+0.15}_{-0.14})$	$(-2.30^{+0.15}_{-0.14})$

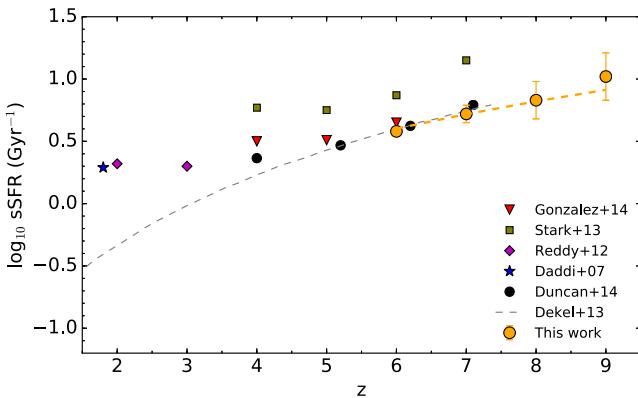
inclusion of nebular emission lines and have derived, for the first time, the GSMF at  $z = 6\text{--}9$  for the HFF program.

Using the same sample, we have also derived the UV LF at  $z = 6\text{--}9$ . For our high- $z$  sample, we have estimated the dust-corrected SFRs from UV luminosities and UV continuum slopes. From this, we have calculated the sSFR and UV luminosity density in the redshift range probed. Our key conclusions are as follows:

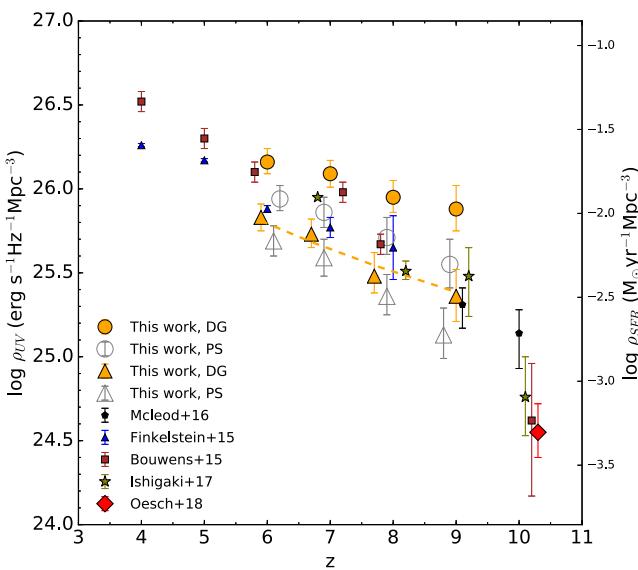
- (i) Our new measurements of the GSMF show an apparent steepening of the low-mass end slope  $\alpha$  with increasing redshift ( $-1.98^{+0.07}_{-0.07}$ ,  $-2.01^{+0.17}_{-0.13}$ ,  $-2.30^{+0.51}_{-0.46}$ , and  $-2.38^{+0.72}_{-0.88}$  at  $z = 6, 7, 8,$



**Figure 16.** Evolution of total SMD as a function of redshift. We calculate the total SMD by integrating the best-fitting Schechter function from  $M_* = 10^8$  to  $10^{13} \text{ M}_\odot$  for our GSMFs. The quoted  $1\sigma$  error bars represent the minimum and maximum range of stellar mass densities within the  $1\sigma$  contours of Schechter parameters obtained from our MCMC analysis. Our results are shown as solid orange points when disc-like galaxies are considered in our completeness simulations, and as open grey circles considering point sources. Shown also are the values from the literature converted to a Chabrier IMF, where necessary.



**Figure 17.** Biweight mean sSFR and error on the mean for galaxies in a fixed stellar mass bin of  $\log_{10}(M/M_\odot) = 9.7 \pm 0.3$  as a function of redshift. The dashed yellow line is the best-fitting power law to the observed values of sSFR from  $z \sim 6-9$ . The theoretical model of Dekel et al. (2013) in which the specific accretion rate follows the form  $s\text{SFR} \propto (1+z)^{2.5}$  is shown by dashed grey line.



**Figure 18.** Evolution of UV luminosity density as a function of redshift.  $\rho_{\text{UV}}$  values calculated with an  $M_{\text{UV}} = -13.5$  limit are shown in orange filled circles and  $\rho_{\text{UV}}$  calculated using an  $M_{\text{UV}} = -17$  limit are shown in orange filled triangles. Also shown are the  $\rho_{\text{UV}}$  estimates considering point sources in our completeness simulations with open grey circles and open grey triangles for  $M_{\text{UV}} = -13.5$  and  $M_{\text{UV}} = -17$ , respectively. The literature points shown are estimated with an  $M_{\text{UV}} = -17$  limit. The dashed yellow line is the power-law fit to the data points at  $z \sim 6-9$ , such that  $\rho_{\text{UV}} \propto (1+z)^{-2.63}$ . Fitting a power law at  $z \geq 8$  values results in  $\rho_{\text{UV}} \propto (1+z)^{-4.61}$  (not shown).

and 9, respectively) within the error bars, statistically steeper than previously observed mass functions at slightly lower redshifts, and we find no evidence of a turnover in the mass range probed. We also find a decrease in normalization  $\Phi^*$  with increasing redshift, but no evolution in characteristic mass  $M^*$  is observed.

(ii) The faint-end slope of the UV LF also exhibit an apparent steepening with increasing redshift (from  $-2.03^{+0.12}_{-0.10}$  at  $z = 6$  to  $-2.20^{+0.51}_{-0.47}$  at  $z = 9$ ), without any evidence of a turnover. These measurements are consistent with a continuation of the trends seen from lower redshift, however more data are needed to confirm

the trend at these redshifts. The normalization  $\Phi^*$  of UV LF decreases with increasing redshift and a weak evolution in  $M_{\text{UV}}^*$  is also observed, with it decreasing with redshift, implying that the evolution of the UV LF with redshift appears to be more consistent with an evolution of density.

(iii) The slopes of our best fitted  $\log_{10}(M_*/M_\odot) - M_{\text{UV}}$  relation are close to a constant mass-to-light ratio of  $-0.40$ , suggesting no strong evolution of mass-to-light ratio with luminosity. We notice that normalization, on the other hand, evolves very weakly from  $z \sim 6$  to  $z \sim 9$ , with a decrease in normalization with increasing redshift.

(iv) From our new measurements of the GSMF, we estimate the stellar mass density and find that the SMD increases by a factor of  $\sim 15^{+21}_{-6}$ , from  $\log_{10}\rho_* = 5.61^{+0.92}_{-0.90}$  at  $z = 9$  to  $\log_{10}\rho_* = 6.79^{+0.13}_{-0.12}$  at  $z = 6$ .

(v) We estimate the sSFR ( $\text{sSFR} = \text{SFR}/M_*$ ) of our sample, and find that for a fixed stellar mass of  $5 \times 10^9 M_\odot$ ,  $\text{sSFR} \propto (1+z)^{2.01 \pm 0.16}$ .

(vi) From our new measurements, we estimate the UV luminosity density ( $\rho_{\text{UV}}$ ) and the cosmic SFRD ( $\rho_{\text{SFR}}$ ), and find that our results support a smooth decline of  $\rho_{\text{SFR}}$  towards high redshifts.

Our study exhibits the power of gravitational lensing to probe the faintest and earliest galaxies in the early Universe. While the results from this study are very intriguing, the uncertainties are still large due to a small sample size. The analysis of the complete HFF data set comprising six clusters and six associated parallel fields will increase the signal to noise of these results, providing robust constraints in the future. This will, however, require a significant amount of effort as subtracting the foreground galaxies takes up the bulk of the effort when studying nearby massive lensing clusters. On the other hand, to probe even higher redshifts requires the advantage of *JWST*. The advent of *JWST* will allow us to extend this study all the way back to redshift of  $z = 12$ , if not even earlier. Ultimately, it will also allow us to probe far deeper down the luminosity and mass functions to determine if and where there is a turnover in the number counts as a function of mass and luminosity. Until then, any future studies of lensing clusters must take into account the careful subtraction of the foreground cluster galaxies such as we have done here, otherwise the results will be biased and galaxies at the faintest limits will be missed.

## ACKNOWLEDGEMENTS

The authors would like to thank the referee for their comments, which have greatly improved this paper. We would also like to thank Rachael Livermore, Rychard Bouwens, and Derek Mcleod for their suggestions and acknowledge the discussions that led to improvements in this paper. In addition, we would like to thank Emiliano Merlin for his support and guidance with the T-PHOT software. This work is based on the observations made with the NASA/ESA *Hubble Space Telescope*, obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (STScI), which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. This work is also based on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory (JPL), California Institute of Technology under a contract with NASA and uses the data taken with the Hawk-I instrument on the European Southern Observatory (ESO) Very Large telescope (VLT) from ESO programme 092.A-0472.

This work utilizes gravitational lensing models produced by PIs Bradac, Ebeling, Merten & Zitrin, Sharon, and Williams funded as part of the *HST* Frontier Fields program conducted by STScI. The lens models were obtained from the MAST. RB acknowledges funding from the Science and Technology Facilities Council (STFC). KD acknowledges support from the European Research Council (ERC) Advanced Investigator programme NewClusters 321271.

## REFERENCES

- Arnouts S., Ilbert O., 2011, Astrophysics Source Code Library, record ascl:1108.009
- Atek H. et al., 2014, *ApJ*, 786, 60
- Atek H. et al., 2015, *ApJ*, 800, 18A
- Atek H., Richard J., Kneib J.-P., Schaefer D., 2018, *MNRAS*, 479, 5184
- Avni Y., Bahcall J. N., 1980, *ApJ*, 235, 694
- Balestra I. et al., 2016, *ApJS*, 224, 33
- Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, *MNRAS*, 422, 449
- Behroozi P., Wechsler R., Hearin A., Conroy C., 2018, preprint ([arXiv: 1806.07893](https://arxiv.org/abs/1806.07893))
- Benitez N., 2000, *ApJ*, 536, 571
- Bernyk M. et al., 2016, *ApJS*, 223, 9
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Bertin E., Mellier Y., Radovich M., Missonnier G., Didelon P., Morin B., 2002, in Bohlander D. A., Durand D., Handley T. H., eds, *ASP Conf. Ser. Vol. 281, Astronomical Data Analysis Software and Systems XI*. Astron. Soc. Pac., San Francisco, p. 228
- Bolzonella M., Miralles J.-M., Pelló R., 2000, *A&A*, 363, 476
- Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, *ApJ*, 670, 928
- Bouwens R. J. et al., 2009, *ApJ*, 690, 1764
- Bouwens R. J. et al., 2011, *Nature*, 469, 504
- Bouwens R. J. et al., 2012, *ApJ*, 754, 83
- Bouwens R. J. et al., 2014, *ApJ*, 795, 126
- Bouwens R. J. et al., 2015, *ApJ*, 803, 34
- Bouwens R. J. et al., 2016, *ApJ*, 830, 67
- Bouwens R. J., Illingworth G. D., Oesch P. A., Atek H., Lam D., Stefanon M., 2017a, *ApJ*, 843, 41
- Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017b, *ApJ*, 843, 129
- Bradač M. et al., 2009, *ApJ*, 706, 1201
- Bradley L. D. et al., 2014, *ApJ*, 792, 76
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503
- Brammer G. B. et al., 2016, *ApJS*, 226, 6
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, *ApJ*, 429, 582
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Caminha G. B. et al., 2017, *A&A*, 600, A90
- Castellano M. et al., 2016, *A&A*, 590, A31
- Chabrier G., 2003, *PASP*, 115, 763
- Coe D. et al., 2013, *ApJ*, 762, 32
- Conselice C. J., Wilkinson A., Duncan K., Mortlock A., 2016, *ApJ*, 830, 83
- Dahlen T. et al., 2013, *ApJ*, 775, 93
- Davé R., 2008, *MNRAS*, 385, 147
- De Santis C., Grazian A., Fontana A., Santini P., 2007, *New Astron.*, 12, 271
- Dekel A., Zolotov A., Tweed D., Cacciato M., Cevarino D., Primack J. R., 2013, *MNRAS*, 435, 999
- Duncan K. et al., 2014, *MNRAS*, 444, 2960
- Duncan K., Conselice C. J., 2015, *MNRAS*, 451, 2030
- Eales S., 1993, *ApJ*, 404, 51
- Erb D. K., Pettini M., Shapley A. E., Steidel C. C., Law D. R., Reddy N. A., 2010, *ApJ*, 719, 1168
- Ferguson H. C. et al., 2004, *ApJ*, 600, L107
- Finkelstein S. L. et al., 2012, *ApJ*, 756, 164
- Finkelstein S. L. et al., 2015, *ApJ*, 810, 71
- Fioc M., Rocca-Volmerange B., 1997, *A&A*, 326, 950
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Galametz A. et al., 2013, *ApJS*, 206, 10
- Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
- González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., Brammer G. B., 2010, *ApJ*, 713, 115
- González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, *ApJ*, 735, L34
- González V., Bouwens R. J., Labbé I., Illingworth G., Oesch P., Franx M., Magee D., 2012, *ApJ*, 755, 148
- González V., Bouwens R., Illingworth G., Labbé I., Oesch P., Franx M., Magee D., 2014, *ApJ*, 781, 34
- Goodman J., Weare J., 2010, *Commun. Appl. Math. Comput. Sci.*, 5, 65
- Grazian A. et al., 2011, *A&A*, 532, A33
- Grazian A. et al., 2012, *A&A*, 547, A51
- Grazian A. et al., 2015, *A&A*, 575, A96
- Grillo C. et al., 2015, *ApJ*, 800, 38
- Gu M., Ho L. C., Peng C. Y., Huang S., 2013, *ApJ*, 773, 34
- Guo Y. et al., 2013, *ApJS*, 207, 24
- Ilbert O. et al., 2005, *A&A*, 439, 863
- Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2015, *ApJ*, 799, 12
- Ishigaki M., Kawamata R., Ouchi M., Oguri M., Shimasaku K., Ono Y., 2018, *ApJ*, 854, 73
- Jauzac M. et al., 2014, *MNRAS*, 443, 1549
- Johnson T. L., Sharon K., Bayliss M. B., Gladders M. D., Coe D., Ebeling H., 2014, *ApJ*, 797, 48
- Jullo E., Kneib J.-P., 2009, *MNRAS*, 395, 1319
- Jullo E., Kneib J.-P., Limousin M., Elfásdóttir Á., Marshall P. J., Verdugo T., 2007, *New J. Phys.*, 9, 447
- Kalirai J., 2018, *Contemp. Phys.*, 59, 251
- Kawamata R., Ishigaki M., Shimasaku K., Oguri M., Ouchi M., 2015, *ApJ*, 804, 103
- Kennicutt R. C., Jr, 1998, *ARA&A*, 36, 189
- Kneib J.-P., 1993, PhD thesis, Université Paul Sabatier, Toulouse
- Kneib J.-P., Natarajan P., 2011, *A&AR*, 19, 47
- Labbé I. et al., 2010, *ApJ*, 716, L103
- Laidler V. G. et al., 2007, *PASP*, 119, 1325
- Laporte N. et al., 2015, *A&A*, 575, A92
- Laporte N. et al., 2016, *ApJ*, 820, 98
- Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, *ApJ*, 835, 113
- Lotz J. M. et al., 2017, *ApJ*, 837, 97
- Madau P., 1995, *ApJ*, 441, 18
- McLeod D. J., McLure R. J., Dunlop J. S., 2016, *MNRAS*, 459, 3812
- McLure R. J. et al., 2011, *MNRAS*, 418, 2074
- McLure R. J. et al., 2013, *MNRAS*, 432, 2696
- Merlin E. et al., 2015, *A&A*, 582, A15
- Merlin E. et al., 2016, *A&A*, 590, A30
- Merten J. et al., 2011, *MNRAS*, 417, 333
- Meurer G. R., Heckman T. M., Calzetti D., 1999, *ApJ*, 521, 64
- Mobasher B. et al., 2015, *ApJ*, 808, 101
- Oesch P. A. et al., 2010, *ApJ*, 709, L21
- Oesch P. A. et al., 2013, *ApJ*, 773, 75
- Oesch P. A. et al., 2014, *ApJ*, 786, 108
- Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, *ApJ*, 855, 105
- Oguri M., 2010, *PASJ*, 62, 1017
- Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713
- Ono Y., Ouchi M., Shimasaku K., Dunlop J., Farrah D., McLure R., Okamura S., 2010, *ApJ*, 724, 1524
- Ono Y. et al., 2013, *ApJ*, 777, 155
- Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, *AJ*, 124, 266
- Ravindranath S. et al., 2006, *ApJ*, 652, 963
- Schaerer D., de Barros S., 2009, *A&A*, 502, 423
- Schaerer D., de Barros S., 2010, *A&A*, 515, A73

- Schaerer D., de Barros S., 2012, in Tuffs R. J., Popescu C. C., eds, IAU Symp. Vol. 284, The Spectral Energy Distribution of Galaxies – SED 2011, Proceedings of the International Astronomical Union, Observatoire de Geneve, p. 20
- Schechter P., 1976, *ApJ*, 203, 297
- Schenker M. A. et al., 2013, *ApJ*, 768, 196
- Schmidt M., 1968, *ApJ*, 151, 393
- Schmidt K. B. et al., 2014, *ApJ*, 786, 57
- Sersic J. L., 1968, Atlas de Galaxias Australes. Observatorio Astronomico, Cordoba, Argentina
- Shipley H. V. et al., 2018, *ApJS*, 235, 14
- Song M. et al., 2016, *ApJ*, 825, 5
- Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, *ApJ*, 697, 1493
- Stark D. P., Schenker M. A., Ellis R., Robertson B., McLure R., Dunlop J., 2013, *ApJ*, 763, 129
- Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, *ApJ*, 462, L17
- Trenti M., Stiavelli M., 2008, *ApJ*, 676, 767
- Trenti M. et al., 2011, *ApJ*, 727, L39
- Weinmann S. M., Neistein E., Dekel A., 2011, *MNRAS*, 417, 2737
- Zheng W. et al., 2012, *Nature*, 489, 406
- Zitrin A. et al., 2013, *ApJ*, 762, L30
- Zitrin A. et al., 2014, *ApJ*, 793, L12

## APPENDIX A: CATALOGUES

**Table A1.** Catalogue of the  $z \sim 8$  sample. Column (1) lists the source IDs, columns (2) and (3) list their coordinates, column (4) gives the photometric redshift, column (5) lists the median magnification, column (6) is the absolute magnitude at 1500 Å, and column (7) is the logarithm of the stellar mass. Both absolute magnitude and stellar mass values are corrected for magnification.

ID	RA	Dec.	$z_{\text{phot}}$	Magnification	$M_{1500}$	$\log M_*$ ( $M_\odot$ )
MACS0416 cluster						
2095	64.0342	-24.0699	$8.43^{+0.27}_{-0.66}$	3.12	$-18.78^{+0.45}_{-0.32}$	$8.21^{+0.34}_{-0.24}$
257	64.0336	-24.0900	$7.71^{+0.16}_{-0.71}$	2.21	$-18.60^{+0.28}_{-0.36}$	$8.29^{+0.45}_{-0.29}$
2365	64.0476	-24.0667	$8.06^{+0.18}_{-1.05}$	11.92	$-15.55^{+0.61}_{-0.86}$	$7.84^{+0.79}_{-1.7}$
2613	64.0603	-24.0649	$8.06^{+0.22}_{-0.36}$	1.61	$-18.47^{+0.42}_{-0.31}$	$8.36^{+0.44}_{-0.35}$
1288	64.0327	-24.0790	$8.15^{+0.49}_{-0.36}$	9.67	$-17.26^{+0.29}_{-0.26}$	$7.82^{+0.28}_{-0.25}$
Parallel field						
154	64.1319	-24.1311	$7.71^{+0.10}_{-1.09}$	-	$-18.76^{+0.22}_{-0.56}$	$8.27^{+0.32}_{-0.23}$
2674	64.1268	-24.1000	$8.25^{+0.12}_{-0.62}$	-	$-19.00^{+0.27}_{-0.35}$	$8.51^{+0.49}_{-0.31}$
495	64.1213	-24.1258	$7.88^{+0.10}_{-0.59}$	-	$-19.82^{+0.28}_{-0.32}$	$8.62^{+0.27}_{-0.21}$
1237	64.1196	-24.1169	$7.71^{+0.27}_{-0.66}$	-	$-18.78^{+0.27}_{-0.66}$	$8.64^{+0.27}_{-0.66}$
1084	64.1545	-24.1188	$8.25^{+0.27}_{-0.66}$	-	$-20.13^{+0.27}_{-0.66}$	$8.91^{+0.27}_{-0.66}$

**Table A2.** Catalogue of the  $z \sim 9$  sample. Column (1) lists the source IDs, columns (2) and (3) list their coordinates, column (4) gives the photometric redshift, column (5) lists the median magnification, column (6) is the absolute magnitude at 1500 Å, and column (7) is the logarithm of the stellar mass. Both absolute magnitude and stellar mass values are corrected for magnification.

ID	RA	Dec.	$z_{\text{phot}}$	Magnification	$M_{1500}$	$\log M_*$ ( $M_\odot$ )
MACS0416 cluster						
126	64.0391	-24.0931	$8.62^{+0.22}_{-0.18}$	1.72	$-20.55^{+0.47}_{-0.36}$	$8.84^{+0.20}_{-0.18}$
1040	64.0479	-24.0816	$8.72^{+0.12}_{-0.23}$	1.61	$-20.12^{+0.27}_{-0.16}$	$8.71^{+0.28}_{-0.21}$
1065	64.0480	-24.0814	$8.72^{+0.52}_{-0.22}$	1.56	$-20.25^{+0.18}_{-0.22}$	$9.44^{+0.42}_{-0.37}$
393	64.0375	-24.0881	$8.62^{+0.14}_{-0.55}$	1.58	$-18.78^{+0.27}_{-0.26}$	$8.28^{+0.42}_{-0.29}$
Parallel field						
1524	64.1498	-24.1133	$9.42^{+0.23}_{-0.76}$	-	$-19.59^{+0.23}_{-0.21}$	$8.63^{+0.51}_{-0.39}$
1957	64.1517	-24.1084	$9.01^{+0.52}_{-0.35}$	-	$-18.96^{+0.20}_{-0.18}$	$8.93^{+0.60}_{-0.48}$
2660	64.1267	-24.1003	$8.53^{+0.17}_{-0.62}$	-	$-18.97^{+0.27}_{-0.32}$	$8.50^{+0.46}_{-0.31}$
1646	64.1479	-24.1117	$8.72^{+0.09}_{-0.94}$	-	$-18.84^{+0.22}_{-0.18}$	$8.31^{+0.51}_{-0.43}$

This paper has been typeset from a TeX/LaTeX file prepared by the author.