Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **491,** 1980–1997 (2020)
Advance Access publication 2019 October 24

doi:10.1093/mnras/stz2988

Combining high-z galaxy luminosity functions with Bayesian evidence

Nicolas J. F. Gillet ,★ Andrei Mesinger and Jaehong Park

Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

Accepted 2019 October 16. Received 2019 October 14; in original form 2019 June 14

ABSTRACT

Galaxy formation during the first billion years of our Universe remains a challenging problem at the forefront of astrophysical cosmology. Although these $z \gtrsim 6$ galaxies are likely responsible for the last major phase change of our Universe, the epoch of reionization (EoR), detailed studies are possible only for relatively rare, bright objects. Characterizing the fainter galaxies that are more representative of the population as a whole is currently done mainly through their non-ionizing UV luminosity function (LF). Observing the faint end of the UV LFs is nevertheless challenging, and current estimates can differ by orders of magnitude. Here we propose a methodology to combine disparate high-z UV LF estimates in a Bayesian framework: Bayesian Data-analysis Averaging (BDA). Using a flexible, physically motivated galaxy model, we compute the relative evidence of various $z = 6$ UV LFs within the magnitude range $-20 \leq M_{\rm UV} \leq -15$ which is common to the data sets. Our model, based primarily on power-law scalings of the halo mass function, naturally penalizes systematically jagged points as well as misestimated errors. We then use the relative evidence to weigh the posteriors obtained from disparate LF data sets during the EoR, $6 \leq z \leq 10$. The resulting LF posteriors suggest that the star formation rate density (SFRD) integrated down to a UV magnitude of -17 represent $60.9^{+11.3}_{-9.6}$ per cent / $28.2^{+9.3}_{-10.1}$ per cent / $5.7^{+4.5}_{-4.7}$ per cent of the total SFRD at redshifts 6 / 10 / 15. The BDA framework we introduce enables galaxy models to leverage multiple, analogous LF estimates when constraining their free parameters.

Key words: galaxies: high-redshift – early Universe.

1 INTRODUCTION

The first billion years of the Universe remain a compelling cosmological mystery, mostly due to the fact that observations of this period remain challenging (e.g. Barkana & Loeb 2007; Loeb & Furlanetto 2013; Mesinger 2016; Dayal & Ferrara 2018). One of the simplest and most powerful observations are the non-ionizing (∼1500 Å rest frame) ultra-violet luminosity functions (UV LFs). These can be obtained with relatively straightforward broad-band photometric drop-out techniques (Steidel et al. 1999) and are thus useful in constraining the abundance of galaxies too faint to be studied with spectroscopy.

Nevertheless, pushing the UV LFs towards the fainter galaxies that are the dominant population during the first billion years is quite difficult. Lensing magnification has been shown to be a powerful tool for this purpose; however, the systematics quickly become significant going towards magnification factors of beyond $\mu \gtrsim 10$ (e.g. Bouwens et al. 2017; Atek et al. 2018). Various observational estimates of the faint end of the LF (where the bulk of the galaxies lie) can disagree by orders of magnitude.

How do we choose which LF estimates to use when constraining galaxy formation models? If every competing LF estimate is analysed independently, the corresponding parameter constraints would need to be eventually combined somehow. Alternately, one could first combine different LF data sets in some fashion and then fit galaxy parameters to the *combined* data. For example, Finkelstein (2016) used a Schechter function form (Schechter 1976) to fit LF estimates from various studies independently at each redshift ($z = 4$–10). These were then combined ‘agnostically’: with each competing estimate being given the same weight, regardless of how consistent it was with the assumed Schechter form.

In principle, one should be able to improve on this by applying some basic, prior knowledge of what the UV LFs *should* look like. For example, sharp discontinuities in the LF would be very difficult to explain physically and could be an indication of an unaccounted for systematic in the observations. The commonly used, empirically motivated Schechter function requires additional, ad hoc parameters to account for both redshift evolution and a faint-end turnover, in order to be consistent with observations and physically motivated galaxy formation models of ultra-faint dwarf galaxies (e.g. Behroozi, Wechsler & Conroy 2013; Dayal et al. 2014; O’Shea et al. 2015a; Liu et al. 2016; Finlator et al. 2017). Indeed, Yue et al. (2018) use a physical galaxy model to supplement a Schechter function, and derive constraints on the presence of a faint-end turnover in the LF, based on galaxy number counts at $z = 6$.

★ E-mail: nicolas.gillet@sns.it

Here we use a flexible galaxy model to combine disparate high-z LF estimates in a Bayesian evidence-based framework. The parametrization of this model should encapsulate the general, physical trends we expect from high-z LFs, while still being able to accommodate the unknown details of galaxy formation. We apply this Bayesian Data-analysis Averaging (BDA) framework to current LF data sets, resulting in combined LF constraints even at redshifts and magnitudes not probed by current observations.

We note that the differences between current LFs mostly arise from different analysis methods applied to the same *Hubble* fields. For example, different groups use different lensing models, completeness corrections, Eddington bias corrections, etc. Therefore, LFs are not raw data, but derived data products. We therefore adopt the nomenclature BDA.

This paper is organized as follows. In Section 2, we describe the BDA method, demonstrating its use on toy LFs. In Section 3, we introduce the LF data sets and the analytic model used to discriminate between data sets. In Section 4, we apply BDA on the $z = 6$ LFs, and we use the resulting weights to combine LF data across $z \sim 6$–10, presenting the resulting 'concordance' LFs. In Section 5, we state our conclusions. Unless stated otherwise, we use comoving units, and assume the following ΛCDM cosmological parameters ($\Omega_\mathrm{m} = 0.3175$, $\Omega_\Lambda = 0.6825$, $h = 0.6711$, $\Omega_\mathrm{b} = 0.049$, $n_\mathrm{s} = 0.9677$, and $\sigma_8 = 0.83$), consistent with the latest results from the *Planck* satellite (Planck Collaboration et al. 2018), and magnitudes are given in the AB system.

2 COMBINING DIFFERENT DATA SETS

Our methodology to combine the estimated LFs is inspired by Bayesian Model Averaging (BMA; e.g. Parkinson & Liddle 2013); however, we reverse 'model' and 'data'. Instead of comparing different models using a given observable, we compare different observables using a given model. This comparison is done with the Bayesian evidence, which allows us to weigh the relative posteriors from different data sets and combine them using this weight. We describe the procedure in detail below.

We note that alternative Bayesian methods have been proposed to combine data sets, taking advantage of Bayesian hierarchical modelling and/or hyper-parameters. A common approach is to add hyper-parameters to account for mis-estimated errors / systematics of each observation, which are then marginalized over to obtain the posterior of the desired quantities (e.g. Lahav et al. 2000; Hobson, Bridle & Lahav 2002; Ma & Berndsen 2014; Bernal & Peacock 2018). Such an approach relies on knowing how to parametrize these uncertainties and the additional parameters make the likelihood calculation more expensive. The procedure we propose avoids this but at the cost of relying on a parametrization of the 'truth'.

Below we briefly review BMA, before introducing our reversed application of it: BDA. We then demonstrate its use using toy models for LFs.

2.1 Bayes' equation and model averaging

Let $\mathcal{D}$ be a data catalogue composed of several data sets (raw or derived observations) and $\mathcal{M}$ an analytic model with parameters θ. Bayes' theorem permits us to compute the posterior: the probability distribution of the parameters θ given a specific data set $\mathcal{D}_i$:

$$\mathrm{P}(\theta|\mathcal{D}_\mathrm{i}, \mathcal{M}) = \frac{\mathrm{P}(\mathcal{D}_\mathrm{i}|\theta, \mathcal{M})\mathrm{P}(\theta|\mathcal{M})}{\mathrm{P}(\mathcal{D}_\mathrm{i}|\mathcal{M}) = \int_\theta \mathrm{P}(\mathcal{D}_\mathrm{i}|\theta, \mathcal{M})\mathrm{P}(\theta|\mathcal{M})\mathrm{d}\theta}, \quad (1)$$

where $\mathrm{P}(\theta|\mathcal{M})$ is the prior on the parameters, $\mathrm{P}(\mathcal{D}_\mathrm{i}|\theta, \mathcal{M})$ is the likelihood (commonly based on χ^2), and $\mathrm{P}(\mathcal{D}_\mathrm{i}|\mathcal{M})$ is probability of the data given the model (also known as the evidence).

In general, the posterior is just the normalized likelihood distribution, weighted by the priors. The evidence is commonly used only as a normalization factor because one is interested in the relative probabilities across the parameter space of θ. However, if one has various competing models, $\mathcal{M}_i$, then the relative evidence can be used to discriminate among them, answering the question: 'which model is preferred by the data?'. Additionally, the evidence can be used to average over parameters common to the various models. This is referred to as Bayesian model averaging (BMA).

2.2 Bayesian Data-analysis Averaging

In this work, we invert 'data' and 'model', asking the question: 'which data set is preferred by our model?' Given a model $\mathcal{M}$, we can compute the relative evidence of the data sets:

$$\mathrm{P}(\mathcal{D}_\mathrm{i}|\mathcal{D}, \mathcal{M}) = \frac{\mathrm{P}(\mathcal{D}_\mathrm{i}|\mathcal{M})}{\sum_\mathrm{j} \mathrm{P}(\mathcal{D}_\mathrm{j}|\mathcal{M})}. \quad (2)$$

The prior on the data set $\pi(\mathcal{D}_i)$ is taken to be uniform. This relative evidence can be used to compare the data sets between each other, given the model. We use the relative evidence from each data set as a weight of the resulting posterior for our model parameters:

$$\mathrm{P}(\theta|\mathcal{D}, \mathcal{M}) = \sum_\mathrm{i} \mathrm{P}(\theta|\mathcal{D}_\mathrm{i}, \mathcal{M}) \times \mathrm{P}(\mathcal{D}_\mathrm{i}|\mathcal{D}, \mathcal{M}), \quad (3)$$

where $\mathrm{P}(\theta|\mathcal{D}, \mathcal{M})$ is the final constrained posterior distribution. The corresponding 'concordance' LF is then obtained by sampling this combined posterior.

It is important to keep in mind that this procedure is model dependent. Ideally, one should choose a model with a parametrization capable of capturing the general trends we expect from the data, yet flexible enough to accommodate the large range of uncertainties. Conceptually, this is analogous to putting a (conservative) prior on what is expected from the observations. The model we use for this purpose is described in Section 3.2.

2.3 Demonstration on toy models

Here we illustrate the use of BDA, applied on toy LFs. Our mock LFs consist of nine points, generated by different methods of sampling a fiducial parameter combination (Section 3.2):

(i) Mock observation (A) was generated by sampling the expectation values from this model, assuming Gaussian errors with a standard deviation of 20 per cent, for each magnitude bin. The reported errors on these points also have a standard deviation of 20 per cent. Thus, the samples are consistent with the underlying model and the reported uncertainty corresponds to the true uncertainty. Hence, model (A) represents an accurate data set (c.f. top left panel in Fig. 1).

(i) Mock observation (B) was generated by sampling the same analytic model as (A), also taking Gaussian errors with a standard deviation of 20 per cent. However, here the reported errors are underestimated to be only 10 per cent (c.f. top right panel in Fig. 1).

(i) Mock observation (C) is statistically the same as (A) for the brightest six points; however, the faintest three data points are systematically offset from the underlying analytic model, showing an upturn for $M_\mathrm{UV} > -16$ of 15 per cent, 30 per cent,

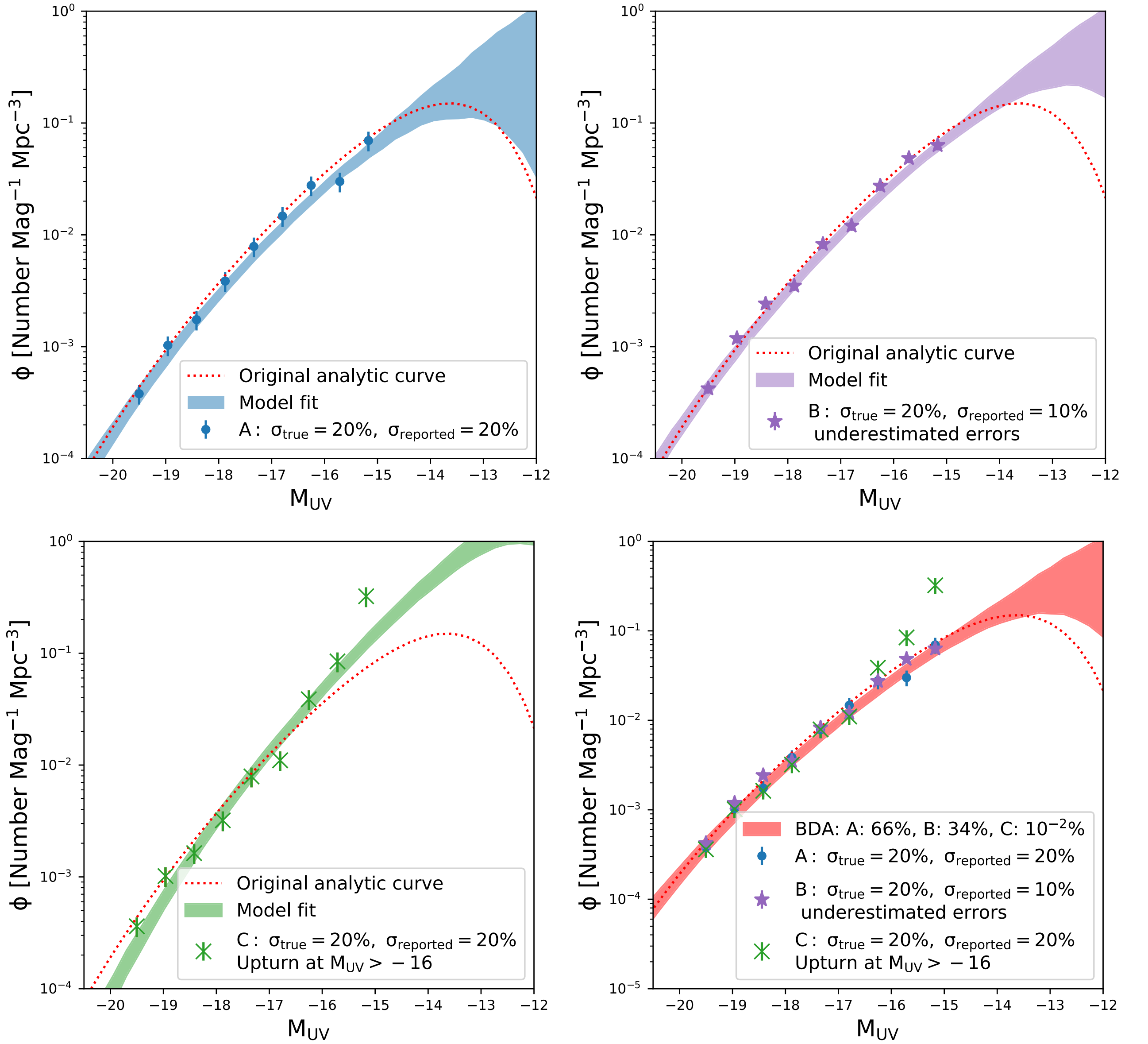

Figure 1. Mock luminosity functions to demonstrate the application of BDA. The three mock data sets correspond to the dots (stars and X) with the reported errors. The red dashed line illustrates the true underlying LF used to create the mocks. The shaded areas represent the 68 per cent confidence interval from the posterior. On the bottom right, the same three mocks are shown with coloured dots (stars and X) and the shaded area corresponds to the BDA combined posterior, with the relative weights of the three data sets shown in the legend.

and 50 per cent, respectively. This observation is illustrative of an unknown systematic in the data, which cannot be captured by our model (c.f. bottom left panel in Fig. 1)

We show the three mock data sets and 68 per cent confidence interval (C.I.) on the posteriors in the first three panels of Fig. 1. As expected, the posteriors of data sets A and B are comparable, given that they only differ in the error estimates. However model C prefers a much steeper LF posterior. This is because our galaxy model (described in detail in Section 3.2) does not allow for upturns,[1] and so the last three points steepen the LF posterior, despite the fact that the first six points are statistically the same as for model A.

[1]Mirocha & Furlanetto (2019) use an empirical model to show that *if* the recent EDGES claimed detection (Bowman et al. 2018) was cosmological, then one would need either an upturn in the LFs or a new population of faint, transient galaxies at $z \sim 20$. Allowing for an upturn in LFs could alter some conclusions here, and we defer it to future work. We point out however, that an upturn at $z \sim 6$ is very contrived. The most plausible physical motivation of such an upturn could be ultra-faint dwarfs, which get their gas through molecular hydrogen cooling (so-called minihaloes). These form from pristine gas and could have different properties from the observed galaxies (e.g. Tumlinson & Shull 2000; Schaerer 2002; Yoshida, Omukai & Hernquist 2008; Xu et al. 2016b; Koh & Wise 2018), which could result in an upturn. However, these minihaloes likely lie at fainter magnitudes than those accessible with current observations (e.g. O'Shea et al. (2015b) and Qin et al. in prep). Moreover, minihaloes are expected to disappear well before z 6, since they get sterilized quickly by a Lyman Werner background (Holzbauer & Furlanetto 2012; Fialkov et al. 2013; Mebane et al., in preparation).

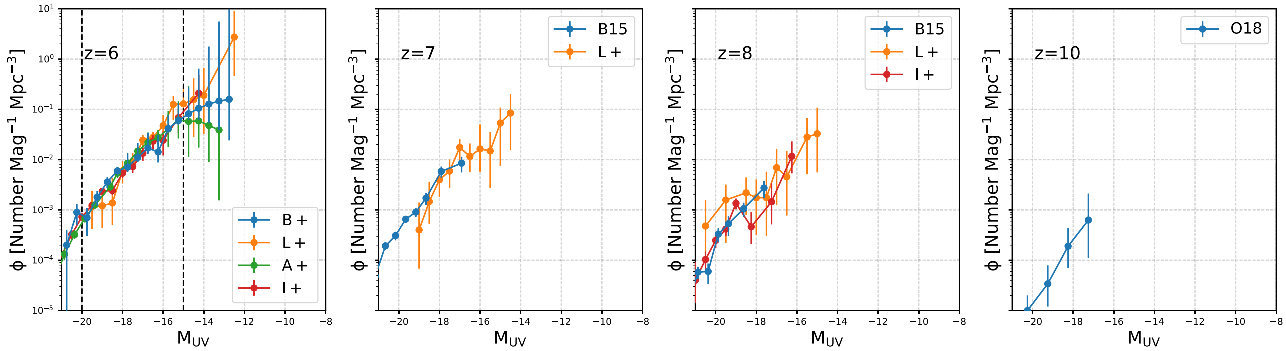

Figure 2. LF determinations at redshifts 6, 7, 8, and 10 from left to right (see text for details and references). The vertical black dash lines delimit the 'faint' and 'ultra-faint' end.

In the final panel, we show the combined LF posteriors, obtained after using the relative evidence to weight the posteriors of A, B, and C (equation 3). The relative evidence from BDA is shown in the legend: 66 per cent, 34 per cent, 10^{-2} per cent, for data sets A, B, and C, respectively. BDA down-weighs the posterior of data set C quite strongly, and so it does not really contribute to the combined posterior. This 'penalty' is due to our *belief* (qualified in terms of our analytic model; see Section 3.2), that upturns in LFs are non-physical.

Data set A provides the most constraining power, as the error bars of the data are estimated properly. BDA prefers A over B by a factor of 2, even though the only difference between the two data sets is that the later data set underestimated the errors of its data points.[2]

3 THE NON-IONIZING LUMINOSITY FUNCTION AT HIGH REDSHIFT

We now wish to apply BDA on actual LF estimates. We first discuss the data sets we use, then our analytic model which is used to weigh them, before specifying how we compute the evidence.

3.1 Luminosity function estimation

In this study, four sets of high-z LFs are used, from redshift 6 and above when available. In the rest of the paper, we define 'faint end' to be magnitudes fainter than −20 (the dominant population we are interested in characterizing) and 'ultra-faint end' to be magnitude fainter than −15 for which lensing uncertainties increase dramatically (c.f. Finkelstein 2016; Bouwens et al. 2017; Atek et al. 2018). These four data sets are as follows:

(i) The 'Bouwens et al. data set' (B+): consisting of the $z = 6$ LF from Bouwens et al. (2017), the $z = 7$ and 8 LFs from Bouwens et al. (2015), and the $z = 10$ LF from Oesch et al. (2018). The estimates at $z = 6$ are based on the four first clusters of the Hubble Frontier Field programme (HFF): Abell 2744, MACS0416, MACS0717, and MACS1149.

(ii) The 'Atek et al. data set' (A+): we take the reported LF from Atek et al. (2018), adjusted according to their prescription to correspond to $z = 6$. This data set used the six clusters of HFF: Abell 2744, MACS0416, MACS0717, MACS1149, AS1063, and A370 and in addition they use the bright part of the LF from Bouwens et al. (2015).

(iii) The 'Ishigaki et al. data set' (I+): consisting of the $z = 6$ and 8 LFs from Ishigaki et al. (2018). They use the four first HFF clusters, as well as the LF, extracted from blank fields from Bouwens et al. (2015).

(iv) The 'Livermore et al. data set' (L+): consisting of the $z =$ 6, 7, 8 LFs from Livermore (private communication; Finkelstein, in preparation). The LFs correspond to those in Livermore, Finkelstein & Lotz (2017), but corrected for Eddington bias, which reduces the implied number densities, most notably at the faint end by up to a factor of ∼2 [though we note that the amplitude of Eddington bias corrections is debatable, with Bouwens et al. (2017) arguing they should be no larger than 10 per cent]. These Eddington-bias adjusted LFs have also been used in Yung et al. (2019). The two first HFF clusters are used to derive the faint-end LF: Abell 2744 and MACS0416.

We assume a minimum fractional uncertainty of 20 per cent (in linear scale), as suggested in Bouwens et al. (2017), increasing the error of all the data points if the reported error is smaller. Fig. 2 presents these four data sets, at redshift 6, 7, 8, and 10 from left to right. As seen in the panels, the implied galaxy density can vary by orders of magnitude, especially in the ultra-faint end when lensing uncertainties such as completeness corrections dominate the systematics.

To compute the relative evidence as described above, we need data at the same magnitude and redshift bins. For this purpose, we use the 10 points in the magnitude range $-20 \leq M_{\rm UV} \leq -15$ at $z = 6$ (c.f. Fig. 3). The bright limit of this range is still faint enough to be relatively free from dust and active galactic nucleus (AGN) feedback, which are not accounted for in our model. Indeed the slope of the UV continuum β seems to change around this value above redshift 6 (e.g. Finkelstein et al. 2012; Bouwens et al. 2014), roughly consistent with simulation results that suggest that at fainter magnitudes the impact of dust starts becoming negligible (e.g. Wilkins et al. 2016, 2017; Cullen et al. 2017; Ma et al. 2019), and AGN feedback can be neglected (e.g. Wilkins et al. 2017; Yung et al. 2019).[3] The faint limit, although in the lensing regime

[2]We repeat this experiment with 1000 different realizations, finding that data set A consistently contributes the most to the combined posterior, at the level of 70 per cent on average.

[3]We test the impact of the bright end limit on our results by removing the brightest two magnitude bins and re-computing the posteriors. The resulting posteriors are consistent with our fiducial ones, with a somewhat broader PDF for the slope parameter, α_*, due to the removal of points with comparably small errors. Thus we do not find evidence that the bright end limit changes the implied slope of the stellar mass to halo mass relation, and

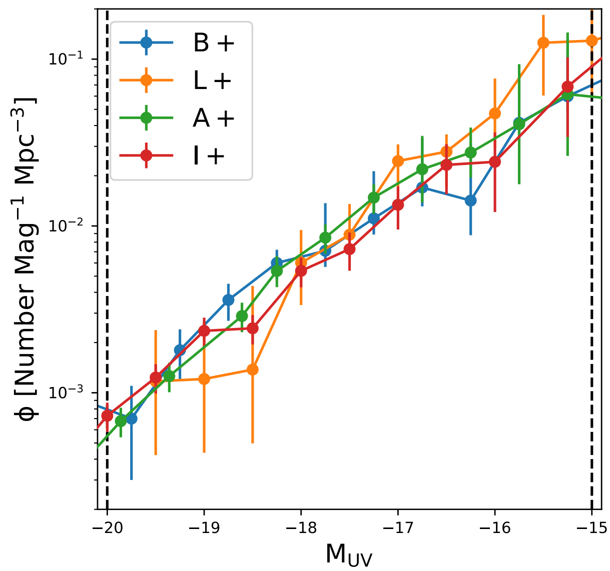

Figure 3. Zoom-in on the 10 $z = 6$ LFs points that are common to all data sets, and which we use when computing the BDA evidence.

($M_{\rm UV} \gtrsim -17$) is sourced by relatively modest magnification factors, with correspondingly well-behaved uncertainties (Finkelstein 2016; Bouwens et al. 2017; Atek et al. 2018). Most importantly, this range is common to all four data sets, which is necessary in order to compare their corresponding Bayesian evidence.

Unfortunately, by calculating the evidence over a limited range, we end up implicitly assuming that this range characterizes also the systematic biases in the other magnitude bins used when combining the data sets. This is unlikely to be true; however our fiducial choice of $-20 \le M_{\rm UV} \le -15$ at $z = 6$ corresponds to the maximum range all data sets have in common. In the absence of a compelling a priori reason to perform further data cuts, it is reasonable to use the largest range possible (see however the test in footnote 3). Moreover, we note that these data are the most constraining, as the fainter magnitudes and higher redshift data points contribute only modestly to the posterior. As discussed in detail in Appendix F and shown in Fig. F1, the addition of fainter magnitudes mainly results in a slight peak in the posterior for the turnover scale and as a small bi-modality in the star formation rate to halo mass scaling, both driven by the A+ data set.

3.2 Analytic model

The analytic model, $\mathcal{M}$, used in this study is the same as in Park et al. (2018). This model characterizes UV LFs using four, fairly empirical parameters. It is physically motivated in the sense that it scales the LF from the halo mass function (HMF), assuming power-law scalings. Specifically, the typical stellar mass, M_*, of galaxies residing in haloes of total mass, $M_{\rm h}$, is assumed to (on average) follow a power law with arbitrary amplitude and power-law index (see Behroozi et al. 2013; Behroozi & Silk 2015):

$$M_*(M_{\rm h}) = f_{*,10} \left(\frac{M_{\rm h}}{10^{10}{\rm M}_\odot}\right)^{\alpha_*} \left(\frac{\Omega_{\rm b}}{\Omega_{\rm m}}\right) M_{\rm h}. \quad (4)$$

The typical star formation rate (SFR) in a given halo mass bin is taken to be the total stellar mass divided by some fraction of the Hubble time:

$$\dot{M}_*(M_{\rm h}, z) = \frac{M_*}{t_* H^{-1}(z)}. \quad (5)$$

The SFR is then converted to a UV luminosity assuming a simple conversion factor:

$$\dot{M}_* = \kappa_{\rm UV} \times L_{\rm UV}, \quad (6)$$

where $\kappa_{\rm UV} = 1.15 \times 10^{-28}\,{\rm M}_\odot\,{\rm yr}^{-1}/{\rm erg}^{-1}\,{\rm Hz}^{-1}$ (Sun & Furlanetto 2016; see also Kennicutt 1998; Bouwens et al. 2012; Madau & Dickinson 2014) is determined by the IMF (and is degenerate with our SFR parameters) and the UV magnitude is computed from the UV luminosity:

$$\log_{10}(L_{\rm UV}) = 0.4 \times (51.63 - M_{\rm UV}). \quad (7)$$

Star formation in low-mass haloes is suppressed via a 'duty cycle', motivated by inefficient gas accretion and/or strong feedback (e.g. Okamoto, Gao & Theuns 2008; Sobacchi & Mesinger 2013, 2014; Dayal et al. 2014; O'Shea et al. 2015a; Ocvirk et al. 2016, 2018; Yue, Ferrara & Xu 2016). Specifically, we assume that only a fraction $f_{\rm duty}$ of haloes of mass $M_{\rm h}$ can host star-forming galaxies, with

$$f_{\rm duty}(M_{\rm h}) = \exp\left(-\frac{M_{\rm t}}{M_{\rm h}}\right). \quad (8)$$

Here, $M_{\rm t}$ is the characteristic halo mass scale below which star formation is inefficient. Our results are not very sensitive to the exact functional form of this duty cycle, since most of the observations probe galaxies inside more massive haloes, as we shall see below.

Finally, the LF is computed from the HMF and the relation between the halo mass and the UV magnitude:

$$\phi(M_{\rm UV}) = \left(f_{\rm duty}\frac{{\rm d}n}{{\rm d}M_{\rm h}}\right)\frac{{\rm d}M_{\rm h}}{{\rm d}M_{\rm UV}}. \quad (9)$$

The model has four free parameters: $f_{*,10}, \alpha_*, t_*$, and $M_{\rm t}$. Hereafter we call a parameter point: $\theta = (f_{*,10}, \alpha_*, t_*, M_{\rm t})$. Fig. 4 presents a schematic view of the influence of each parameter on the LF.[4] The turnover mass $M_{\rm t}$ shifts the peak of the LF towards fainter and brighter magnitudes. The parameter α_* controls the slope, rotating the LFs around the normalization value of $M_{\rm UV}(M_{\rm h} = 10^{10}M_\odot)$, which depends on $f_{*,10}$ and t_*. It also changes the location of the turnover, because it affects the conversion of halo mass to magnitude (note that the model parameters are all defined in terms of halo mass, not directly a UV magnitude). Finally, the parameters $f_{*,10}$ and t_* translate the LF curves horizontally. Note that these two parameters are completely degenerate, and the LF is only sensitive to the ratio $t_*/f_{*,10}$, though ancillary observations such at 21-cm can mitigate this degeneracy. We refer the reader to Park et al. (2018) for a detailed analysis of the influence of each parameter on the luminosity function.

as a consequence that we would need additional parameters characterizing dust or AGN feedback. The relative evidence does change somewhat for this reduced data set, with 19 per cent / 37 per cent / 41.5 per cent / 2.5 per cent attributed to B+ / I+ / A+ / L+ (c.f. values in Table 1). This reflects the fact that the I+ data set has very small errors for those two bins, and the implies counts are consistent with our parametrization. Thus their removal shifts some of the corresponding relative evidence to B +. Selecting sub-samples of the data is, in any case, ad hoc, so we use the largest range which is common to the data sets and over which our galaxy parametrization is reasonable.

[4] An animation showing how the parameters influence the LFs and the 21-cm signal is available at http://homepage.sns.it/mesinger/Videos/parameter_variation.mp4.

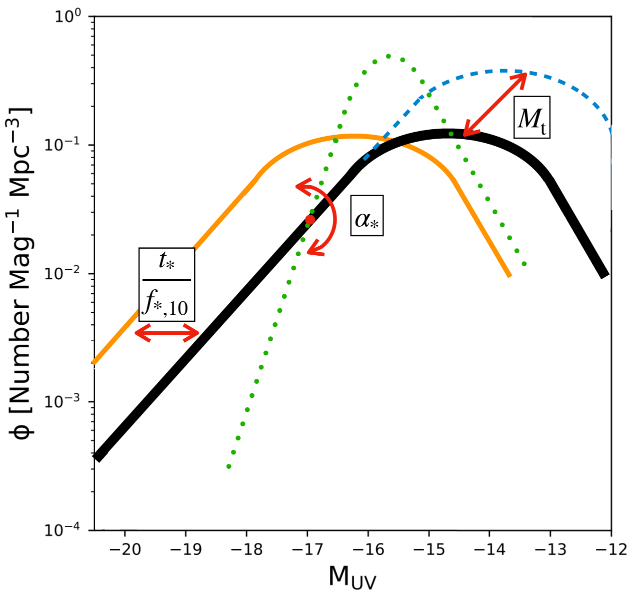

Figure 4. Schematic view of the influence of the model's parameters on the LF. See the discussion in the text and Park et al. (2018) for more details.

The important point for this study is that (i) this model is physically motivated: the galaxy density is directly linked to the dark matter halo density allowing us to penalize extreme LF shapes that are difficult to obtain from HMFs; and (ii) the model is flexible enough to fit reasonably well a large variety of estimated luminosity functions as well as those from hydrodynamic cosmological simulations (see appendix 1 in Park et al. 2018) and SAMs (Greig et al., in preparation).

3.3 Computing the likelihood and the evidence

Computing the evidence can be computationally challenging in high-dimensional parameter space (e.g. Trotta 2008), since the likelihood has to be integrated over the whole space (see the denominator of equation 1).

Our likelihood at each parameter sample θ is computed by comparing the corresponding (logarithmic) LF to the data at every magnitude and redshift bin considered. We use a split-norm probability density function around each data point to take into account of the asymmetrical errors in most of the data sets (c.f. Appendix A). Explicitly, we have

$$P(\mathcal{D}_i|\theta) = \prod_{M_{\rm UV}\ bins} \mathcal{S}(x(\theta), \mu, \sigma_1, \sigma_2), \qquad (10)$$

where $\mathcal{S}$ is the split-norm likelihood (see equation A1), $x(\theta)$ is the modelled galaxy density in a given bin and (μ, σ_1, σ_2) the data galaxy density and its positive and negative errors in the same bin.

We have to compute the likelihood over the whole 4D parameter space $\theta = [f_{*,10}, \alpha_*, t_*, M_{\rm t}]$. Running an MCMC like in Park et al. (2018) for a data set takes 2 weeks. In this study we want to compute the likelihood distribution for four datasets, and explore different configurations. To aid in this computation, we pre-compute the z = 6, 7, 8, and 10 LFs corresponding to a grid of 4×10^5 parameter space samples. This grid is the concatenation of eight Latin Hyperbolic Samples (LHS) of 50 000 points each. The construction of this pre-computed grid of LFs takes around 2 weeks, but the subsequent likelihood calculation is very fast: it just takes a dozen minutes to obtain the likelihood distribution over 400 000 points for each data set on a single core, while an MCMC with the same chain length could take days on several cores. We check that this approximation of the likelihood distribution is converged by comparing with the MCMC results from Park et al. (2018) (see Appendix D). We also check that the posterior is unchanged when computed using only half of the grid samples (i.e. 200 000 points). The discreteness of the sampling results in some noticeable noisiness in the *marginalized* posteriors; however, the parameter estimation and the evidence is converged (see Appendix D).

Finally, once the likelihood samples are computed, the pieces of evidence can be estimated. We interpolate the 400 000 likelihood samples to an evenly spaced grid, consisting of 50 bins per axis, and compute the evidence by summing over this grid. We consider uniform priors for all the parameters over the ranges ($\log_{10}(f_{*,10}) \in [-2.5, -1]$, $\alpha_* \in [-0.5, 1]$, $t_* \in [0, 1]$, $\log_{10}(M_{\rm t}) \in [8, 10]$).

Table 1. Relative evidence of the data sets in the magnitude range [−20, −15] at redshift 6, given the analytic model (in per cent).

(B+)	(I+)	(A+)	(L+)
3.5	52.9	43.4	0.2

4 RESULTS

We apply BDA on the four data sets in order to compare them and create a combined LF. As explained above, the relative evidence is computed from the 10 data points in the magnitude range [−20, −15] at redshift 6 for each data set. These data are illustrated in Fig. 3. Note that the ultra-faint end, where the difference between observational teams is maximal, is not used for the relative evidence. The restriction to this smaller range of magnitudes and redshifts does slightly impact the final posterior distribution, as discussed in Section 3.1, but it is necessary in order to apply the BDA method.

Table 1 gives the resulting relative evidence of the data sets. The I+ and A+ data sets are preferred by our model compared to the two others. This preference is mostly due to the combination of (i) smoothness of the points over the reference range and (ii) small error[5] bars that are still consistent with our parametric model. The L+ data set is disfavoured because it has a plateau at $M_{\rm UV} = -19.5$ to −18.5 and a steepening at the faint end; these features are difficult to fit with our model which relies on smooth functions on top of the HMFs. B+ also has small relative evidence, mainly because of the non-monotonic feature at $M_{\rm UV} = -16$, and the comparably large error bars at the bright end of the range.

We can now combine the posteriors of each individual data sets, weighted by this relative evidence (equation 3). We note that, although the relative weights are computed using only the 10 LF points at z = 6 common to every data set, *each individual posterior is then re-computed using all the data available in the*

[5]As demonstrated in Section 2.3, errors which are too small are naturally penalized by BDA. We can however explicitly check if the data sets have underestimated errors by computing their χ^2 to the corresponding ML model. The resulting χ^2 are 2.5 / 3.7 / 0.9 / 1.7 for B+ / I+ / A+ and L+. Although A+ has the smallest chi-squared, it is consistent with a χ^2 distribution with three effective degrees of freedom (like our model). I+ has the largest chi-squared (within 71 per cent C.L. of the chi-squared distribution), which is even higher if one uses the quoted errors instead of the 20 per cent minimum errors that we applied (χ^2 of 5.9 at 88 per cent C.L.). This is weakly suggestive that the errors in the I+ data set could be underestimated.

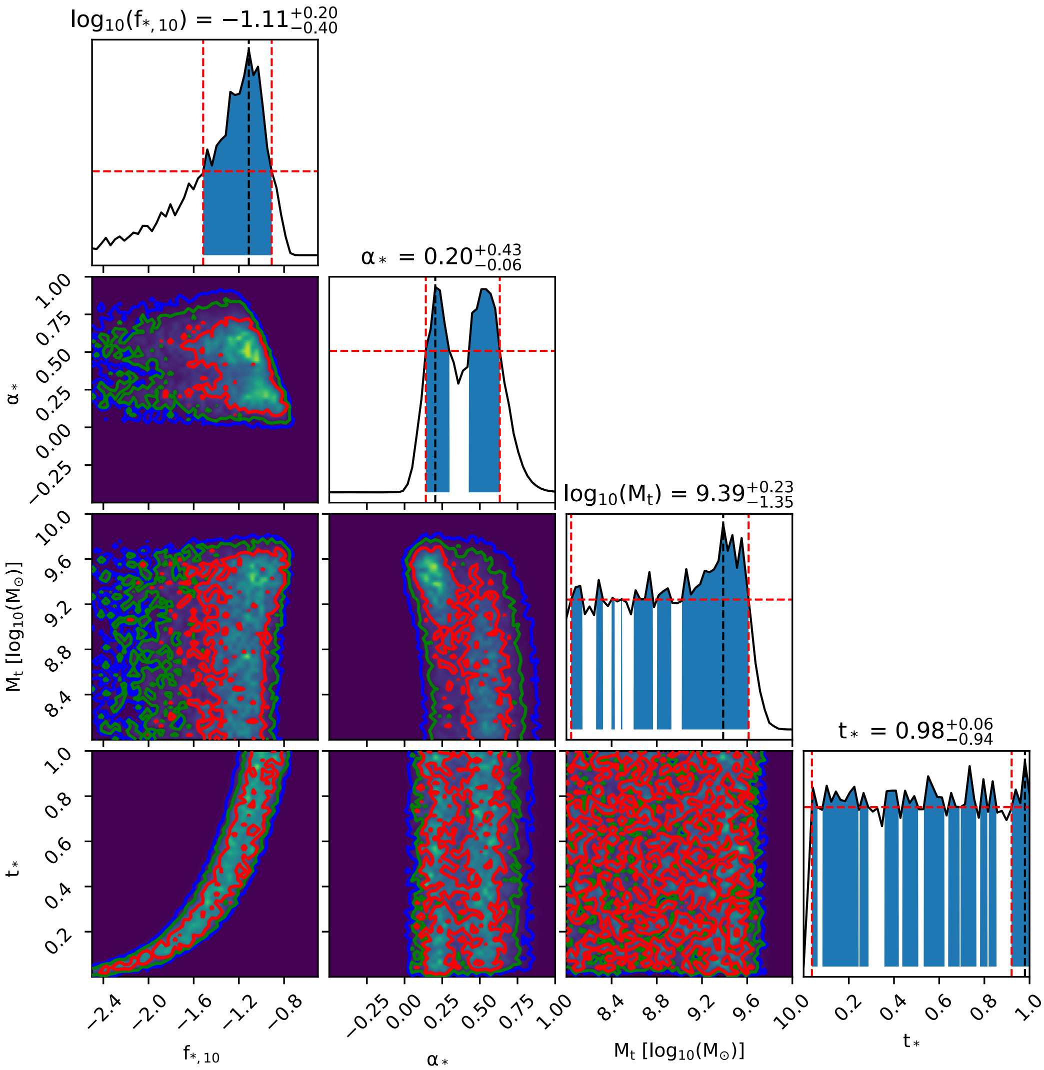

Figure 5. 1D and 2D marginalized posterior distributions of galaxy parameters resulting from the BDA weighing of the posteriors from each data set. The relative weights are listed in Table 1. Although the relative weights are computed using only the 10 LF points at $z = 6$ common to every data set (see Fig. 3), the final posteriors that are averaged are then re-computed using all data points (see text for details). The resulting distribution is mostly the average of the posteriors of A+ and I+. In the 1D marginalized figures, the reported values are the maximum and 68 per cent of highest posterior density (HPD) (illustrated by the blue shaded area).

data set, i.e. including the ultra-faint end and all redshifts (see Fig. E1). It is these posteriors resulting from all data points which are averaged using the relative weights in Table 1, resulting in the combined posteriors shown in Fig. 5. To summarize, the weights are computed on comparable data, at redshift 6 in the magnitude range [−20, −15] and are applied on the posterior computed using all the data available.

There are several trends evident in Fig. 5. First, we note the degeneracy between $f_{*,10}$ and t_*, as the ratio of the two ($r_* = t_*/f_{*,10}$) is relevant for the LFs (see Appendix C). Following Park et al. (2018), we use a linear prior over $\log(f_{*,10})$ and t_*; as a result, the latter is not constrained, showing a flat distribution over the full range.

The double peak in the 1D marginalized posterior of α_* comes from the fact that the two data sets driving the combined posteriors (A+ and I+) favour two different values for this slope of the M_*–M_h relation (see Appendix E). A+ in particular favours a steeper LF (smaller α_*), resulting in a marginalized one sigma constraint of $\alpha_* = 0.2^{+0.09}_{-0.07}$. This can be understood since the data points that are most constraining are those with the smallest errors. For A+ as for I+ , the error is minimum at the bright end of the range we use (see Fig. 3), and for A+ these points have a steeper slope.

The combined marginalized posterior also shows some constraints on M_t, which peaks at $9.39^{+0.23}_{-1.35}$ [$\log_{10}(M_\odot)$]. This peak is entirely driven by A+ ($9.55^{+0.13}_{-0.55}$ [$\log_{10}(M_\odot)$]), with all of the other data sets only providing an upper limit (see Fig. D1). However, the statistical significance of this peak is down-weighted by the BDA combined posteriors, resulting in only an upper limit on the turnover scale (see also Yue et al. 2016, where they look for evidence of a feedback-induced turnover in the LF).

4.1 The combined luminosity functions

The posterior over the parameter space is sampled to obtain the corresponding constraints on the LFs. In Fig. 6 we present the LF constraints corresponding to the 68 per cent C.L. range of the BDA

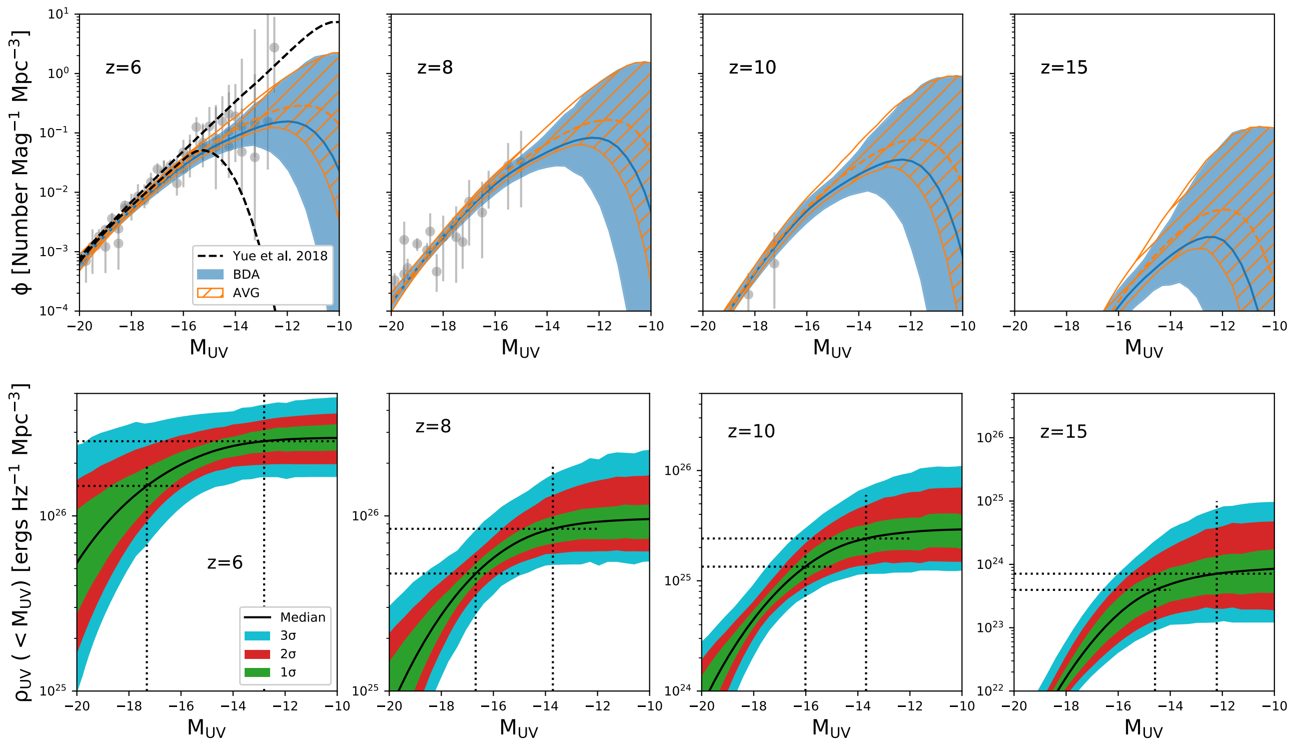

Figure 6. *First row*: the 68 per cent confidence interval of the combined LFs corresponding to the BDA posteriors from Fig. 5 (*blue shaded regions*). For comparison, the orange shaded regions show LF constraints if instead of BDA weights (see Table 1), the posterior of each data set was given an equal weight (i.e. an average of the posteriors) (orange hatched area). In this later case, the relative down-weighting of A+ evidenced by the turnover scale shifting towards fainter magnitudes. All data points used in this work are shown as the grey dots. In the first panel, the 68 per cent C.L. from the data-constrained model of Yue et al. (2018) are shown with the dashed black lines.
Second row: the 68 per cent, 95 per cent, and 99 per cent confidence limits of the cumulative UV luminosity density corresponding to the BDA LFs. The dashed lines correspond to the magnitude limit below which brighter galaxies contribute 50 per cent and 90 per cent of the total UV luminosity density.

posteriors *blue shaded areas*. One nice result from this procedure is the forecast of LFs at even higher redshifts at which we currently have no data (c.f. $z = 15$ LFs in the rightmost panel); although we caution that as our model is mostly constrained by the $z = 6$ points, these extensions to higher redshifts are even more model-dependent. We provide the numerical values for these LF constraints in Tables G1, G2, and G3.

In this figure, we also compare the BDA LFs with those resulting from a uniform weighing of the estimated data sets, i.e. a simple average of each individual posterior, giving a relative weight of 25 per cent to all data sets. The 68 per cent C.L. of the LFs obtained through this simple averaging are shown with the orange shaded regions in Fig. 6. Comparing the orange and the blue shaded regions, we see that the posteriors obtained with BDA are broader, allowing for a turnover at brighter magnitudes. This is driven by the fact that the A+ data set, which is the only one showing evidence of a turnover, has a larger relative contribution in the BDA posterior (43 per cent compared to 25 per cent). Specifically, we note that BDA LFs do not start to flatten or turn over until *at least* $M_{\rm UV} \gtrsim -14$ (1σ). The corresponding scale is shifted fainter by 1 dex for the uniform weighted LFs, to $M_{\rm UV} \gtrsim -13$.

We can also compare our BDA combined LFs to those presented in Yue et al. (2018), who use redshift 6 blank field data from Bouwens et al. (2015), complemented with their own lensed galaxy estimates obtained by taking a mean probability of the number of galaxies per bin implied by different lensing models. The resulting LFs are presented in terms of confidence limits, obtained by sampling a Schecter function modified to allow for a turnover, and shown in the first panel of Fig. 6. Their LFs at the bright end of the range are in agreement with our BDA combined LFs; however, their 68 per cent contours for magnitudes fainter than $M_{\rm UV} \gtrsim -15$ are broader than the ones resulting from BDA.

The corresponding *cumulative* UV luminosity densities for the BDA LFs are shown in the bottom row of Fig. 6, with the dotted lines denoting 50 per cent and 90 per cent of the total UV luminosity density (see also the bottom two rows in Table 2). At redshift 6, galaxies brighter than −17.3 (Understanding the Epoch of Cosmic Reionization 12.8) contribute to 50 per cent (90 per cent) of the total UV luminosity. The 50 per cent limit magnitude increases with redshift, increasing the contribution of fainter galaxies in the total UV budget. But at the same time, the 90 per cent limit magnitude does not significantly evolve with redshift.

It is important to note that the distribution of the *ionizing* photon number density (relevant for reionization) is likely shifted even further towards fainter galaxies than the non-ionizing UV luminosity density. This is because the ionizing escape fraction is expected to increase towards smaller, fainter galaxies, in which it is easier for feedback to evacuate low column density channels facilitating the escape of ionizing photons (e.g. Razoumov & Sommer-Larsen 2010; Yajima, Choi & Nagamine 2011; Ferrara & Loeb 2013; Paardekooper, Khochfar & Dalla Vecchia 2015; Xu et al. 2016a; Kimm et al. 2017). Therefore, when it comes to the total ionizing photon budget, faint galaxies are likely even more important than implied by the 1500 Å CDFs shown in the bottom row of Fig. 6.

Table 2. The cosmic SFR density obtained by integrating our BDA LFs down to the commonly chosen limit of $M_{\rm UV} < -17$ (equivalent to SFR $\gtrsim 0.32\,{\rm M}_\odot\,{\rm yr}^{-1}$) (*top row*), compared with the total (cumulative) SFRD (*second row*). The third row shows the resulting completeness. Errors correspond to 68 per cent C.L. The bottom two rows denote the UV magnitude limit corresponding to 50 per cent and 90 per cent of the cumulative UV luminosity density (illustrated in Fig. 6 by the dashed black lines).

Redshifts	6	7	8	9	10	12	15
SFRD at $M_{\rm UV} < -17$	$-1.72^{+0.12}_{-0.10}$	$-2.01^{+0.07}_{-0.08}$	$-2.32^{+0.06}_{-0.06}$	$-2.66^{+0.06}_{-0.06}$	$-3.02^{+0.08}_{-0.07}$	$-3.82^{+0.14}_{-0.11}$	$-5.20^{+0.23}_{-0.21}$
SFRD total	$-1.49^{+0.07}_{-0.08}$	$-1.71^{+0.08}_{-0.07}$	$-1.96^{+0.11}_{-0.08}$	$-2.21^{+0.15}_{-0.10}$	$-2.47^{+0.19}_{-0.13}$	$-3.04^{+0.24}_{-0.20}$	$-4.00^{+0.38}_{-0.31}$
SFRD completeness at $M_{\rm UV} < -17$ (in per cent)	$60.9^{+11.3}_{-9.6}$	$52.6^{+11.0}_{-9.3}$	$44.1^{+10.6}_{-9.2}$	$36.0^{+10.3}_{-9.4}$	$28.2^{+9.3}_{-10.1}$	$16.0^{+7.4}_{-9.3}$	$5.7^{+4.5}_{-4.7}$
50% $\rho_{\rm UV}$ ($M_{\rm UV}$)	−17.3	−17.0	−16.7	−16.3	−16.0	−15.5	−14.6
90% $\rho_{\rm UV}$ ($M_{\rm UV}$)	−12.8	−13.5	−13.7	−13.8	−13.7	−13.2	−12.2

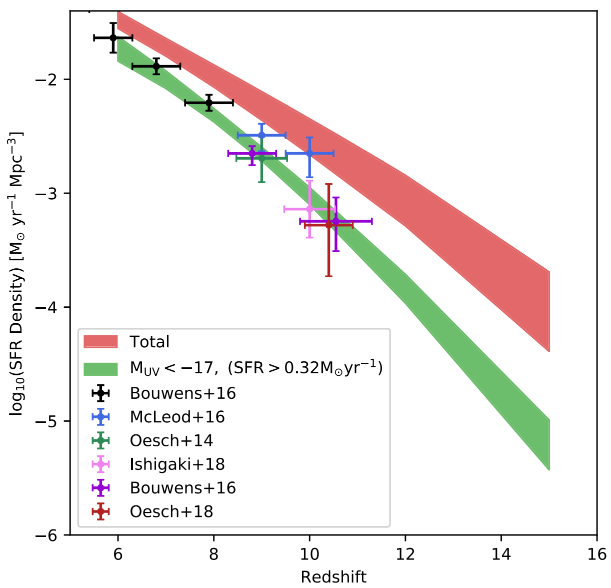

Figure 7. 68 per cent C.L. on the cosmic SFR density implied by our BDA LFs, integrated down to $M_{\rm UV} < -17$ corresponding to SFR $\gtrsim 0.32\,{\rm M}_\odot\,{\rm yr}^{-1}$ (*green shaded area*), as well as the total SFRD (*red shaded area*). The estimated data sets have been homogenized by considering the same SFR–luminosity relation (equation 6) and the same integration limit of −17 [data provided by Oesch (private communication) and published in Oesch et al. (2018)]. The derived SFRD for the two thresholds are given in the Table 2, as well as the completeness. The original data points are from Bouwens et al. (2014, 2016), McLeod, McLure & Dunlop (2016), Oesch et al. (2013, 2014), and Ishigaki et al. (2018).

4.2 Star formation rate density

Finally, in Fig. 7 we show the cosmic star formation rate density (SFRD) from the BDA LFs presented in the previous figure. The SFRD is shown for two integration limits, up to the magnitude of −17 (with 68 per cent C.L. in blue) and integrating over the whole population (68 per cent C.L. in orange). We see that the SFRD up to a magnitude limit of −17 is consistent with observational estimates over the corresponding range (homogenized to correspond to the same limit according to Oesch et al. 2018). However, accounting for star formation in fainter galaxies implies a less rapid decrease going towards higher redshifts. For example, the SFRD down to −17 drops by 3.5 dex going from redshifts 6–15, while the total SFRD only decreases by 2.5 dex over the same redshift interval.

We also quote the median and 68 per cent C.L. in Table 2 for these two integration limits (two first rows) as well as the completeness expressed in per cent of the total SFRD. At redshift 6, galaxies brighter than magnitude −17 account for 60 per cent of the total SFRD. However, this completeness drops rapidly as we go deeper into the EoR and cosmic dawn, becoming only 6 per cent at $z = 15$.[6]

5 CONCLUSIONS

High-redshift LFs provide an important constraint on galaxy formation in the first billion years of the Universe. However, the observations are very challenging, with some estimates disagreeing significantly.

Here we present a simple framework, BDA, to combine different high-z LF estimations. The approach relies on a simple analytic model to encapsulate what we expect from LFs (i.e. smoothness and dependence on HMFs) while allowing flexibility to account for the unknown physics behind them.

In principle, there are two uses of BDA: (i) to infer ‘true’ LFs from disparate data sets; and (ii) to allow disparate data sets to constrain a galaxy model. For the former, the inferred ‘true’ LFs depend on the parametrization of the galaxy model, which should hopefully be reasonable and flexible. For the later, the galaxy model is already assumed: if one is not doing model selection, the parametrization is implicitly the ‘truth’ and one is only interested in constraining the parameters with observations. In this case, BDA provides a way of combining disparate LF estimates self-consistently using the same galaxy parametrization that one wishes to constrain.

We apply BDA on four data sets of high-z ($z \geq 6$), faint-end $M_{\rm UV} > -20$ LFs. The resulting posteriors are mostly driven by two of the four data sets, showing a corresponding bimodality in the implied M_*–$M_{\rm halo}$ relation. The combined posterior also shows very weak evidence of a turnover at faint magnitudes, driven entirely by one data set.

We provide the BDA LFs corresponding to our combined posteriors, which could be used to constrain similar galaxy formation models. These LFs extend to high redshifts and faint objects, for which we currently have no data. However, those extrapolations are model dependent, implicitly relying on our galaxy parameters being able to characterize the true LFs. The approach we present can be applied to future data sets, such as those expected from *JWST*, as well as providing a framework for galaxy models to be informed by disparate data sets.

[6]The completeness is even lower at higher redshifts if there is a separate, transient population of molecularly cooled galaxies. We expect these molecularly cooled galaxies to have different properties compared with the galaxies we observe at $z \lesssim 10$ (e.g. So et al. 2014; Wise et al. 2014), and the framework we use here does not allow for disparate galaxy populations. We will return to this in future work, focused on the ultra-high redshifts in which such galaxies are expected to live.

ACKNOWLEDGEMENTS

We thank R. Bouwens, S. Finkelstein, and H. Atek for valuable comments on a draft version of this work. We also thank R. Livermore for providing us with the unpublished, Eddington bias-corrected data points and P. Oesch for sharing with us the homogenized star formation rate density data points. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 638809 – AIDA – PI: Mesinger). The results presented here reflect the authors' views; the ERC is not responsible for their use. We thank contributors to SCIPY,[7] MATPLOTLIB,[8] PYDOE,[9] and the PYTHON programming language.[10]

REFERENCES

Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, MNRAS, 479, 5184
Barkana R., Loeb A., 2007, Rep. Prog. Phys., 70, 627
Behroozi P. S., Silk J., 2015, ApJ, 799, 32
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bernal J. L., Peacock J. A., 2018, J. Cosmol. Astropart. Phys., 7, 002
Bouwens R. J. et al., 2012, ApJ, 754, 83
Bouwens R. J. et al., 2014, ApJ, 793, 115
Bouwens R. J. et al., 2015, ApJ, 803, 34
Bouwens R. J. et al., 2016, ApJ, 833, 72
Bouwens R. J., Oesch P. A., Illingworth G. D., Ellis R. S., Stefanon M., 2017, ApJ, 843, 129
Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67
Cullen F., McLure R. J., Khochfar S., Dunlop J. S., Dalla Vecchia C., 2017, MNRAS, 470, 3006
Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1
Dayal P., Ferrara A., Dunlop J. S., Pacucci F., 2014, MNRAS, 445, 2545
Ferrara A., Loeb A., 2013, MNRAS, 431, 2826
Fialkov A., Barkana R., Visbal E., Tseliakhovich D., Hirata C. M., 2013, MNRAS, 432, 2909
Finkelstein S. L., 2016, Publ. Astron. Soc. Aust., 33, e037
Finkelstein S. L. et al., 2012, ApJ, 758, 93
Finlator K. et al., 2017, MNRAS, 464, 1633
Hobson M. P., Bridle S. L., Lahav O., 2002, MNRAS, 335, 377
Holzbauer L. N., Furlanetto S. R., 2012, MNRAS, 419, 718
Ishigaki M., Kawamata R., Ouchi M., Oguri M.,Shimasaku K., Ono Y. , 2018, ApJ, 854, 73
Kennicutt Robert C. J., 1998, ARA&A, 36, 189
Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826

[7]http://www.scipy.org
[8]http://www.matplotlib.sourceforge.net
[9]https://pythonhosted.org/pyDOE/
[10]http://www.python.org

Koh D., Wise J. H., 2018, MNRAS, 474, 3817
Lahav O., Bridle S. L., Hobson M. P., Lasenby A. N., Sodré L., 2000, MNRAS, 315, L45
Liu C., Mutch S. J., Angel P. W., Duffy A. R., Geil P. M., Poole G. B., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 235
Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, ApJ, 835, 113
Loeb A., Furlanetto S. R., 2013, The First Galaxies in the Universe
Ma Y.-Z., Berndsen A., 2014, Astron. Comput., 5, 45
Ma X. et al., 2019, MNRAS, 487, 1844
Madau P., Dickinson M., 2014, ARA&A, 52, 415
McLeod D. J., McLure R. J., Dunlop J. S., 2016, MNRAS, 459, 3812
Mesinger A., ed., 2016, Astrophysics and Space Science Library, Vol. 423, Understanding the Epoch of Cosmic Reionization
Mirocha J., Furlanetto S. R., 2019, MNRAS, 483, 1980
O'Shea B. W., Wise J. H., Xu H., Norman M. L., 2015a, ApJ, 807, L12
O'Shea B. W., Wise J. H., Xu H., Norman M. L., 2015b, ApJ, 807, L12
Ocvirk P. et al., 2016, MNRAS, 463, 1462
Ocvirk P. et al., 2018, preprint (arXiv:1811.11192)
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
Okamoto T., Gao L., Theuns T., 2008, MNRAS, 390, 920
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Park J., Mesinger A., Greig B., Gillet N., MNRAS, 2018, 484, 933
Parkinson D., Liddle A. R., 2013, Stat. Anal. Data Min., 9, 3
Planck Collaboration et al., 2018, preprint (arXiv:1807.06209)
Razoumov A. O., Sommer-Larsen J., 2010, ApJ, 710, 1239
Schaerer D., 2002, A&A, 382, 28
Schechter P., 1976, ApJ, 203, 297
Sobacchi E., Mesinger A., 2013, MNRAS, 432, 3340
So G. C., Norman M. L., Reynolds D. R., Wise J. H., 2014, ApJ, 789, 149
Sobacchi E., Mesinger A., 2014, Mem. Soc. Astron. Ital., 85, 509
Steidel C. C., Adelberger K. L., Giavalisco M., Dickinson M., Pettini M., 1999, ApJ, 519, 1
Sun G., Furlanetto S. R., 2016, MNRAS, 460, 417
Trotta R., 2008, Contemp. Phys., 49, 71
Tumlinson J., Shull J. M., 2000, ApJ, 528, L65
Wallis K. F., 2014, Stat. Sci., 29, 106
Wilkins S. M., Bouwens R. J., Oesch P. A., Labbé I., Sargent M., Caruana J., Wardlow J., Clay S., 2016, MNRAS, 455, 659
Wilkins S. M., Feng Y., Di Matteo T., Croft R., Lovell C. C., Waters D., 2017, MNRAS, 469, 2517
Wise J. H., Demchenko V. G., Halicek M. T., Norman M. L., Turk M. J., Abel T., Smith B. D., 2014, MNRAS, 442, 2560
Xu H., Norman M. L., O'Shea B. W., Wise J. H., 2016a, ApJ, 823, 140
Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016b, ApJ, 833, 84
Yajima H., Choi J.-H., Nagamine K., 2011, MNRAS, 412, 411
Yoshida N., Omukai K., Hernquist L., 2008, Science, 321, 669
Yue B., Ferrara A., Xu Y., 2016, MNRAS, 463, 1968
Yue B. et al., 2018, ApJ, 868, 115
Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019, MNRAS, 483, 2983

APPENDIX A: SPLIT NORM

To take into account the asymmetric errors provided in the observations we used the split norm distribution (Wallis 2014). It is just the concatenation of two half-normal distributions, re-normalized to ensure continuity at the origin:

$$\mathcal{S}(\mathrm{x}) = \begin{cases} A\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_1^2}\right), x \leq \mu, \\ A\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma_2^2}\right), x \geq \mu, \\ A = (\sqrt{2\pi}\left(\frac{\sigma_1+\sigma_2}{2}\right)^2)^{-1}. \end{cases} \quad \text{(A1)}$$

For illustrative purposes, Fig. A1 presents two half normal distributions in blue and orange with two different standard deviations (respectively 0.30 and 0.10). The corresponding split-norm distribution is shown in red. For comparison, we also show in green the normal distribution obtained using the average of the variance of the two half normal distributions (i.e. a standard deviation of ∼0.224).

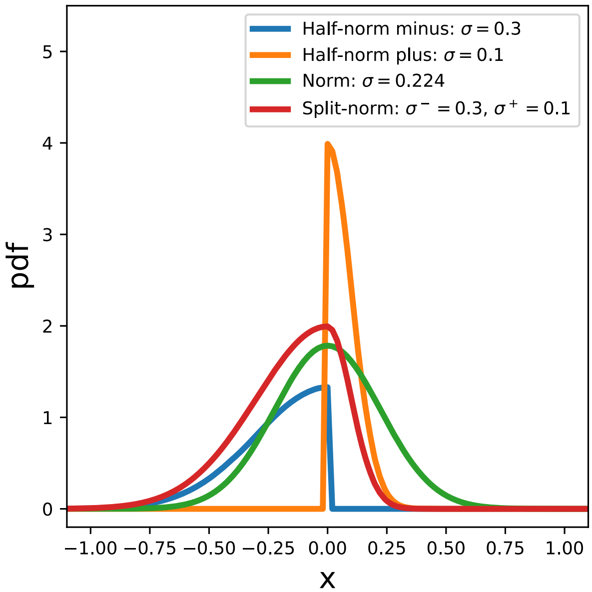

Figure A1. Example of application of the split normal distribution. Two half normal distributions are shown in blue and orange with two different standard deviations (respectively 0.30 and 0.10). The corresponding split-norm distribution is shown in red. For comparison, we also show in green the normal distribution obtained using the average of the variance of the two half normal distributions (i.e. a standard deviation of ∼0.224).

APPENDIX B: CONVERSION OF LOGARITHMIC TO LINEAR SCALE FOR ERRORS

Some studies give the estimated data points and errors in logarithmic base 10 while others do so in linear scale. In this study, we chose to work in logarithmic base 10. The transformation from linear to logarithmic scale for the errors are made as follows:

$$\begin{cases} \phi_{\mathrm{log}} = \log_{10}(\phi_{\mathrm{lin}}), \\ \sigma^+_{\mathrm{log}} = \log_{10}(\phi_{\mathrm{lin}} + \sigma^+_{\mathrm{lin}}) - \log_{10}(\phi_{\mathrm{lin}}), \\ \sigma^-_{\mathrm{log}} = \log_{10}(\phi_{\mathrm{lin}}) - \log_{10}(\phi_{\mathrm{lin}} - \sigma^-_{\mathrm{lin}}), \end{cases} \quad \text{(B1)}$$

$$\begin{cases} \phi_{\mathrm{lin}} = 10^{\phi_{\mathrm{log}}}, \\ \sigma^+_{\mathrm{lin}} = 10^{\phi_{\mathrm{log}}+\sigma^+_{\mathrm{log}}} - 10^{\phi_{\mathrm{log}}}, \\ \sigma^-_{\mathrm{lin}} = 10^{\phi_{\mathrm{log}}} - 10^{\phi_{\mathrm{log}}-\sigma^-_{\mathrm{log}}}. \end{cases} \quad \text{(B2)}$$

Note that symmetric errors in one scale become asymmetric in the other.

APPENDIX C: THE RATIO $t_*/f_{*,10}$

The model used in this study contains two parameters that are completely degenerate in predicting the LF. Although only the ratio $r_* = t_*/f_{*,10}$ is relevant for the LF, we explore the more general formulation by default in this work since EoR observations (or other data sets) can break this degeneracy (see Park et al. 2018).

In Fig. C1 we replace $f_{*,10}$ and t_* by r_* in the traditional corner plot of the posterior. It is the same posterior as presented in Fig. 5, i.e. it is derived from the BDA combination of the LF estimations. This ratio is strongly constrained by the LFs estimation, $\log_{10}(r_*) = 1.01^{+0.06}_{-0.15}$. It is degenerate with α_* and also slightly with M_t at large values of the latter. It is also noticeable that the sampling noise is reduced, due to the reduction of the parameter space dimensionality.

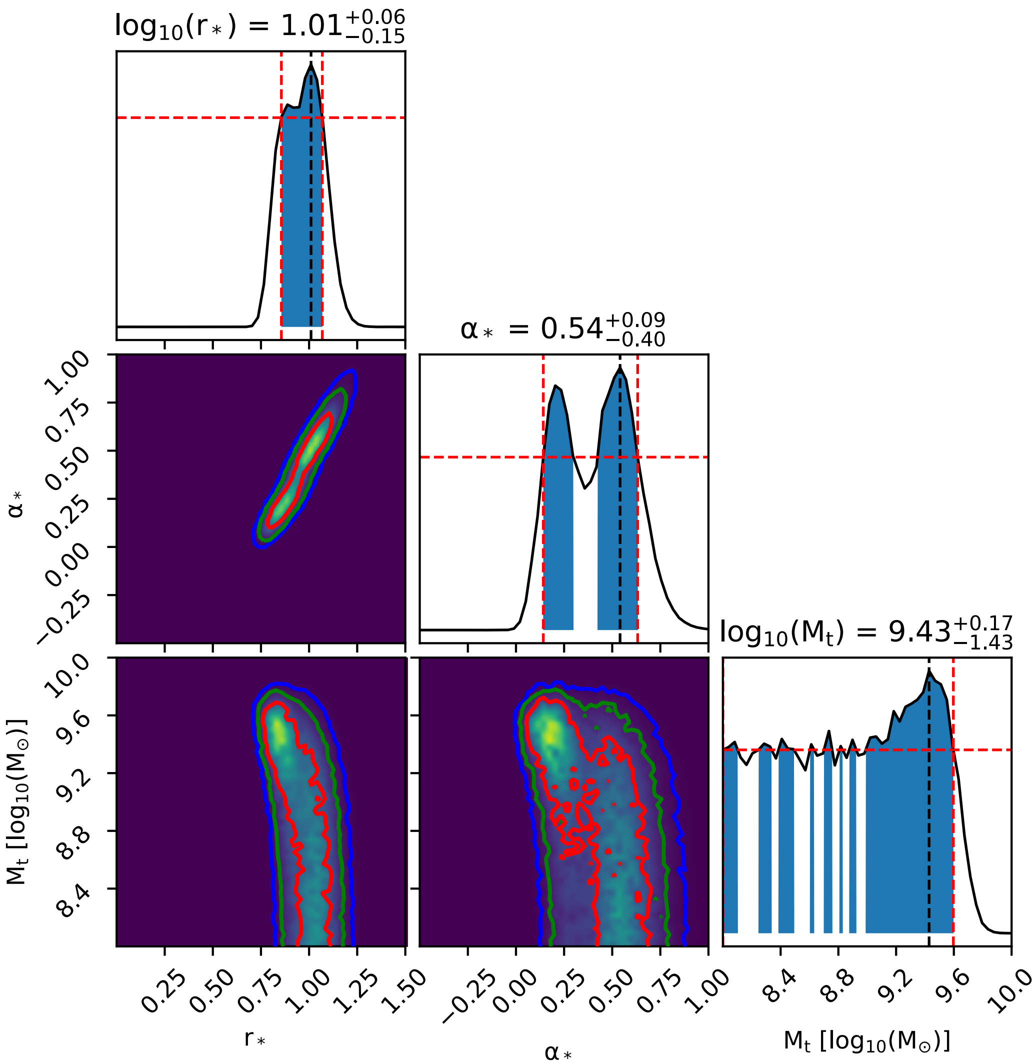

Figure C1. 1D and 2D marginalized posterior distributions of galaxy parameters resulting from the BDA weighing of the posteriors from each data set. The relative weights are listed in Table 1. This figure is the same as Fig. 5, but here the degenerate parameters $f_{*,10}$ and t_* are replaced by their ratio $r_* = t_*/f_{*,10}$ (in log scale). Note that the range of the ratio is zoomed, the original one derive from $f_{*,10}$ and t_* should be [−4, 2.5].

APPENDIX D: CONVERGENCE TEST

In this study, the likelihood is estimated on a grid of points sampled by LHS (200 000 points). To test the convergence of our estimation of the posterior distribution, we compare it with the posterior distribution generated with on-the-fly MCMC sampling. Note that the MCMC chain also contains 200 000 points and has converged. Fig. D1 presents the comparison of the marginalized posterior distributions obtained with the grid (red) and with MCMC (green). Both posteriors are generated using the B+ data set. The 2D contour is the marginalized 1σ. As expected, the marginalized distributions obtained using the grid sampling are noisier, but the final constraints are comparable. We note a slight shift on the estimation of the parameter α_*, due mostly to the difference in the treatment of the error: for computational simplicity, the MCMC code used (Park et al. 2018) treats LF error bars as symmetric, while here we allow for asymmetry (see A1).

We perform a second convergence test on the estimation of the pieces of evidence, by comparing the results obtain with a sample 200 000 points and one of 400 000 points. The results are identical in both cases proving the convergence of the estimation of the evidence.

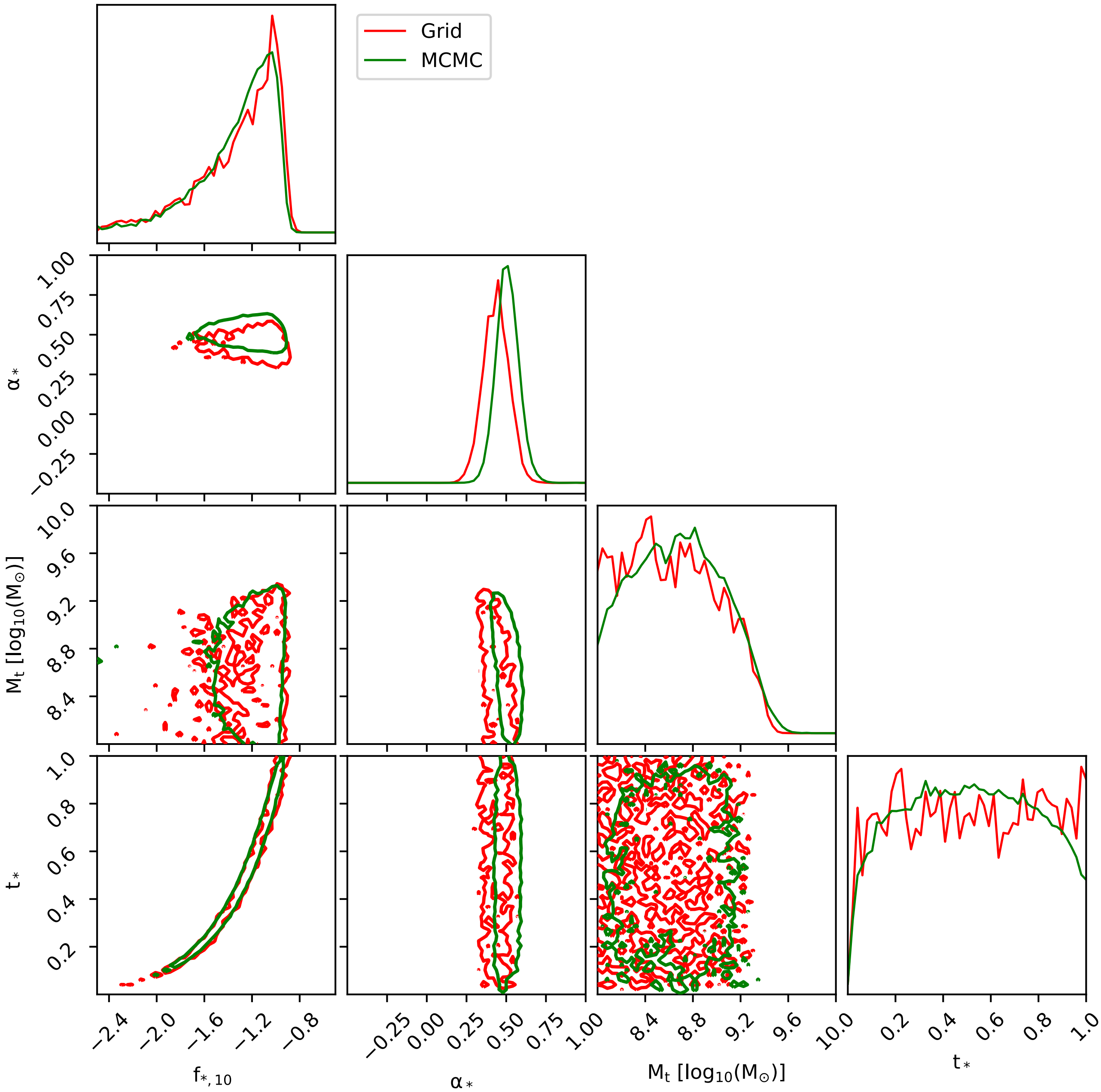

Figure D1. Comparison of the posterior distribution obtained with an on-the-fly MCMC and the pre-computed grid sampling of this study. Both the MCMC and the grid contain the same number of samples: 200 000 points. Those posteriors are generated using the B + data set. The slight shift in the marginalized 1D constraints on α_* is due to the different treatment of errors, as discussed in the text.

APPENDIX E: COMPARISON OF ALL POSTERIOR DISTRIBUTIONS

We compare the posterior distribution obtained with the four data sets in Fig. E1. These are generated using all data points for data sets. While figure E2 presents the same comparison but only using the 10 data points at redshift 6 with $-20 \leq M_{\rm UV} \leq -15$.

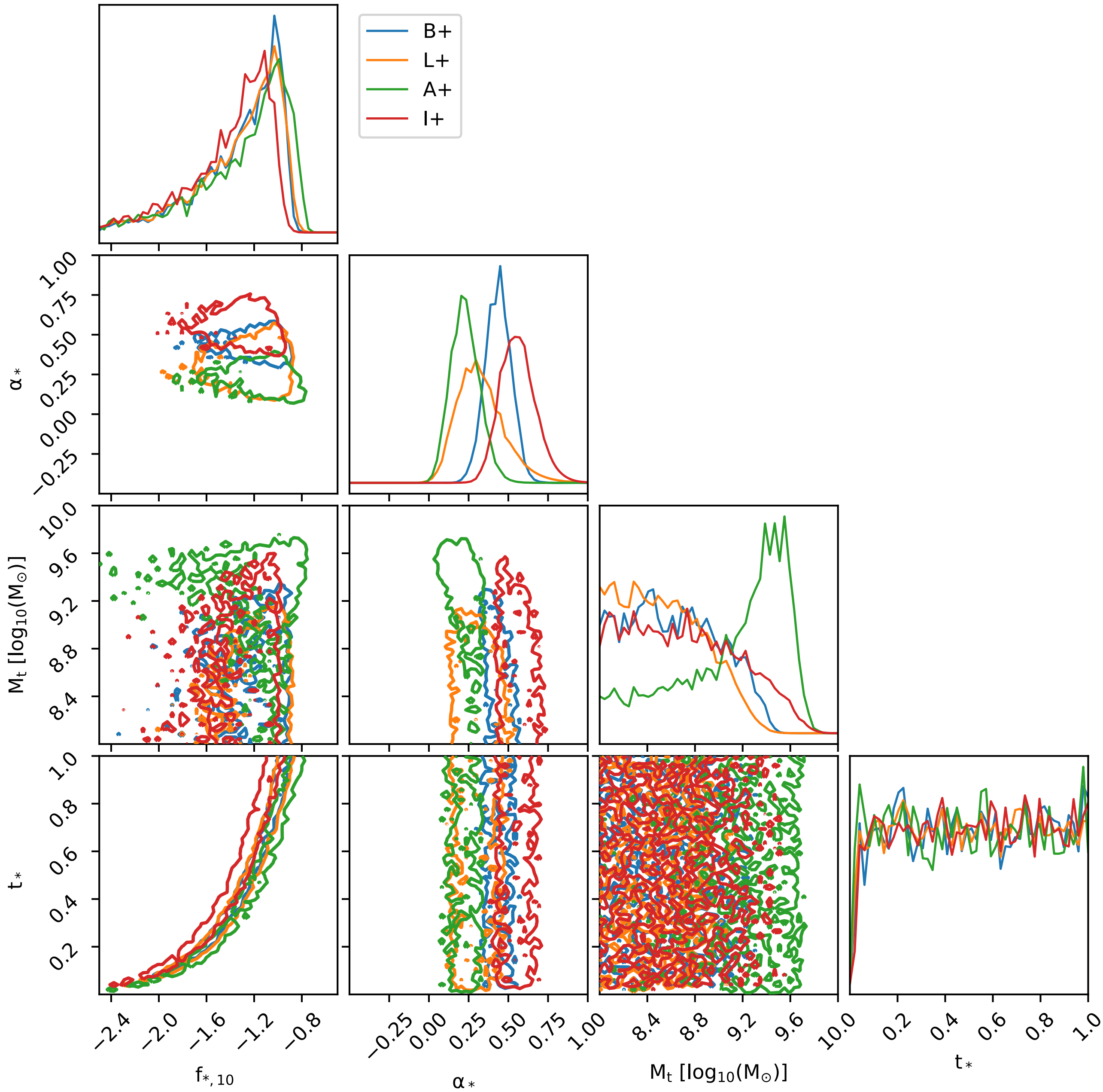

Figure E1. Comparison of the posterior distribution obtained with the four estimated data sets, using all the data available at every redshift.

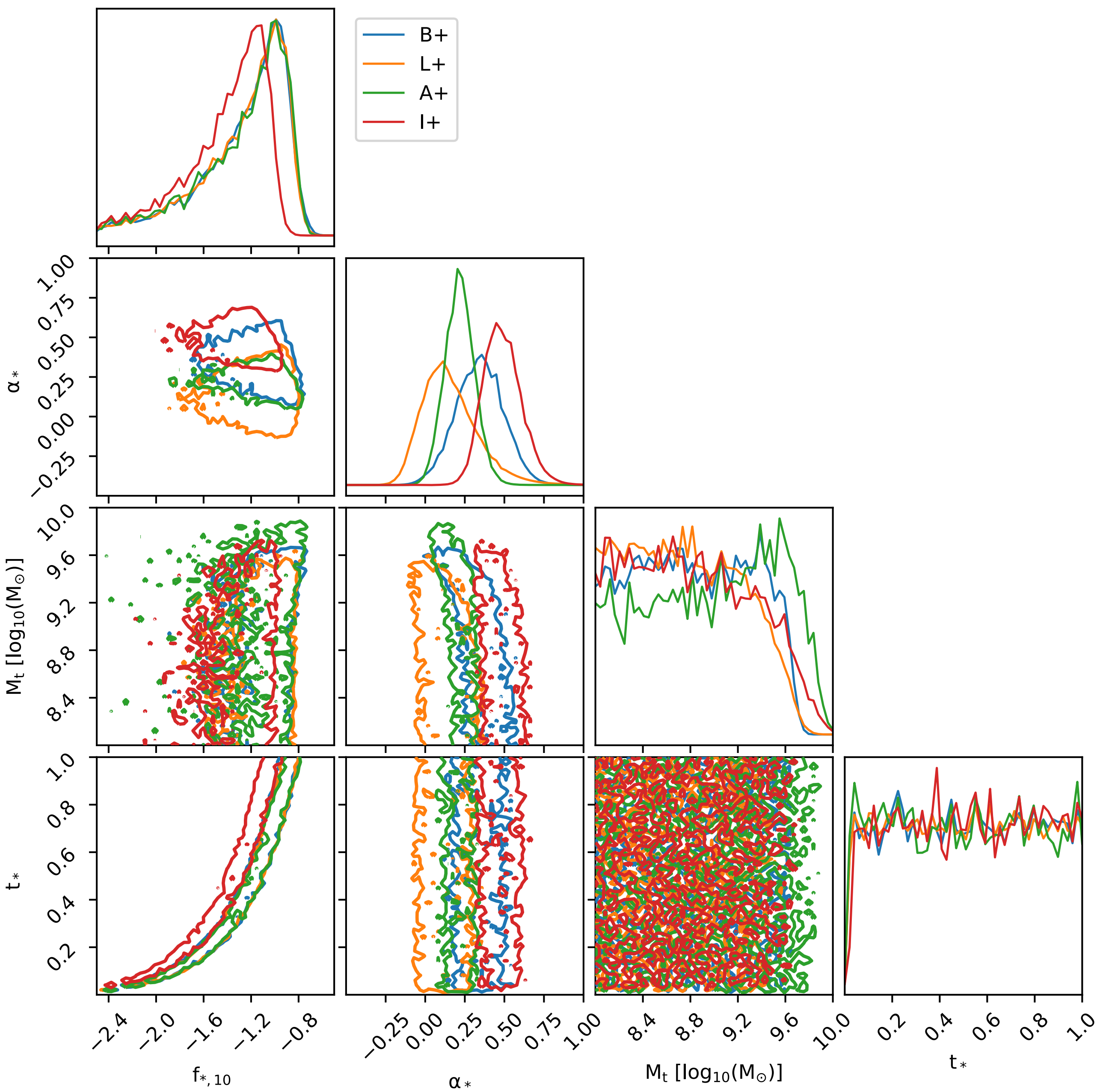

Figure E2. Comparison of the posterior distribution obtained with the four estimated data sets, using only the 10 data points at redshift 6 with $-20 \leq M_{\rm UV} \leq -15$.

APPENDIX F: COMPARISON BDA, AVERAGE, AND REDUCE DATA SETS

We compare the posterior distribution obtained with the BDA method with a simple average of all individual posterior. Fig. F1 present in blue the BDA posterior and in orange the average posterior. There are two noticeable differences. The first is on the parameter α_*, in the average case, the distribution has a more Gaussian shape. But this difference has no noticeable effect once projected on the LF space (see Fig. 6). The second difference is the parameter $M_{\rm t}$: in the case of the average, it is just a lower limit. This effect is visible in the LF space (c.f. Fig. 2 and associated discussion).

The posterior distribution using only the redshift 6 data in the magnitude range [−20, −15] is presented by the red curve. The bimodality of α_* is smoothed, because the constraints of this parameters are wider with the reduce data and the estimation of I + is slightly smaller. The second difference is for $M_{\rm t}$, where the peak at 9.4 is reduce, because it is driven by the ultra-faint end.

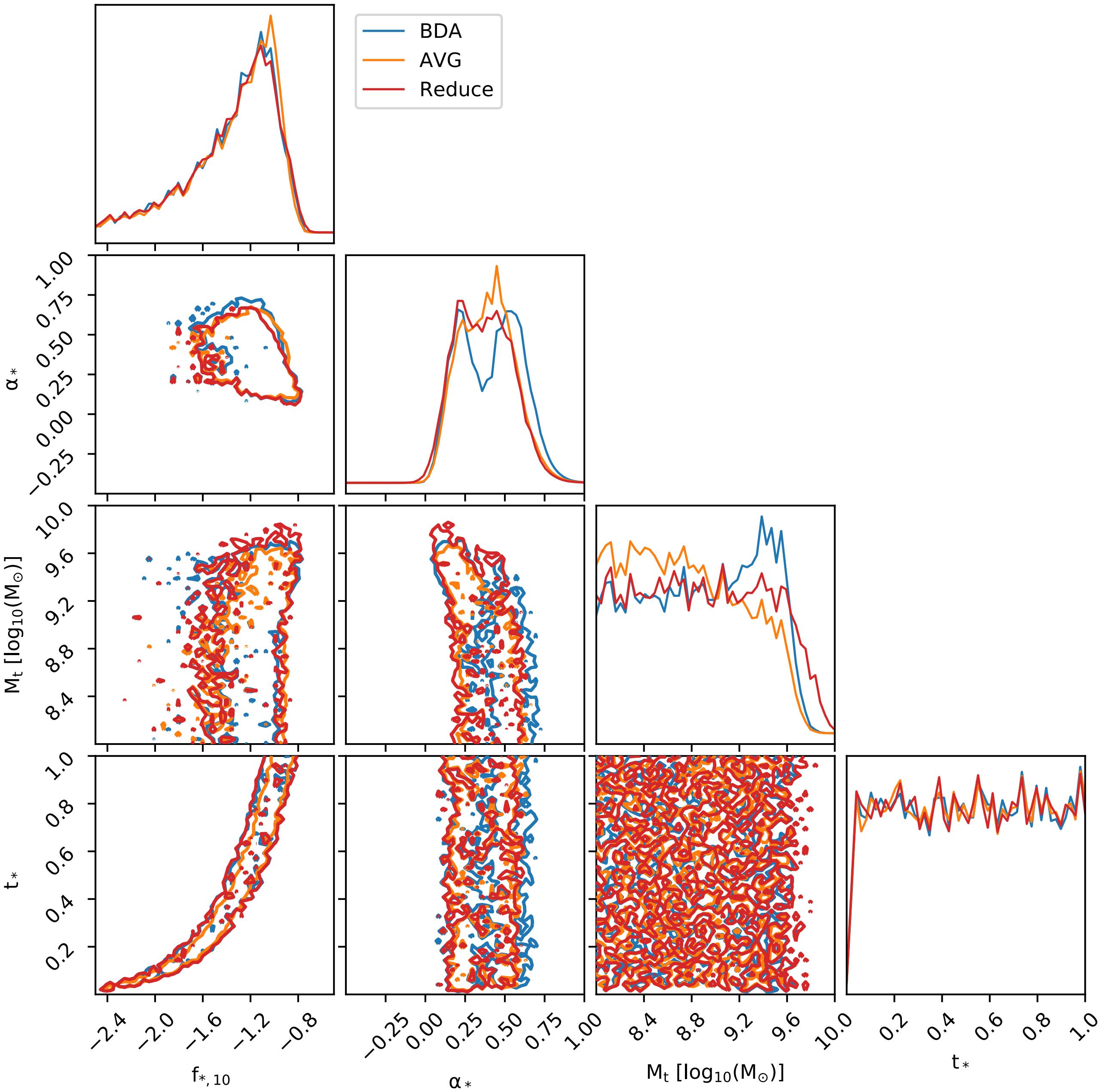

Figure F1. Comparison of the combined posterior distribution obtained with our fiducial BDA procedure (*blue*), a simple average of the individual posteriors (*orange*), and the BDA posterior obtained using only the data range used for computing the evidence (*red*; i.e. using only the redshift six points in the magnitude range [−20, −15]. The reduced range and the simple average both decrease the evidence for a turnover and the related bimodality in α_*, as these trends are driven by the ultra-faint, $M_{\rm uv} > -15$, points from the A + data set.

APPENDIX G: TABLES OF LUMINOSITY FUNCTIONS

Table G1. BDA determination of the UV LF at redshifts 6, 7, and 8. The values of LF are given in logarithmic scale: ϕ [$\log_{10}(M_\odot\ mag^{-1}\ Mpc^{-3})$], σ_{sup}, and σ_{inf} the superior and inferior 68 per cent C.I.

		$z = 6$			$z = 7$			$z = 8$	
M_{UV}	ϕ	σ_{sup}	σ_{inf}	ϕ	σ_{sup}	σ_{inf}	ϕ	σ_{sup}	σ_{inf}
−20.11	−3.26	0.15	0.12	−3.57	0.15	0.13	−3.95	0.17	0.14
−19.77	−3.09	0.12	0.10	−3.38	0.12	0.10	−3.73	0.13	0.11
−19.43	−2.92	0.09	0.08	−3.20	0.10	0.08	−3.53	0.10	0.09
−19.09	−2.77	0.08	0.07	−3.02	0.08	0.07	−3.33	0.08	0.08
−18.75	−2.61	0.07	0.06	−2.86	0.07	0.07	−3.14	0.08	0.08
−18.41	−2.47	0.06	0.06	−2.69	0.07	0.07	−2.96	0.07	0.09
−18.07	−2.32	0.07	0.07	−2.53	0.07	0.09	−2.79	0.09	0.09
−17.73	−2.19	0.07	0.08	−2.38	0.08	0.10	−2.62	0.09	0.11
−17.39	−2.05	0.09	0.09	−2.23	0.10	0.11	−2.45	0.11	0.12
−17.05	−1.92	0.10	0.11	−2.09	0.11	0.12	−2.30	0.13	0.14
−16.71	−1.79	0.11	0.12	−1.95	0.12	0.13	−2.15	0.14	0.14
−16.37	−1.67	0.12	0.12	−1.82	0.13	0.14	−2.01	0.15	0.15
−16.03	−1.56	0.12	0.13	−1.70	0.14	0.14	−1.88	0.15	0.16
−15.69	−1.45	0.13	0.14	−1.59	0.14	0.15	−1.76	0.16	0.16
−15.35	−1.36	0.14	0.14	−1.49	0.14	0.16	−1.66	0.16	0.17
−15.01	−1.27	0.14	0.15	−1.40	0.15	0.17	−1.56	0.17	0.19
−14.67	−1.19	0.15	0.17	−1.32	0.17	0.18	−1.47	0.20	0.19
−14.33	−1.12	0.19	0.18	−1.25	0.22	0.19	−1.40	0.26	0.21
−13.99	−1.06	0.24	0.20	−1.17	0.28	0.22	−1.32	0.32	0.25
−13.65	−0.99	0.31	0.23	−1.11	0.35	0.28	−1.25	0.41	0.30
−13.31	−0.94	0.40	0.27	−1.05	0.45	0.32	−1.19	0.50	0.37
−12.97	−0.89	0.45	0.37	−1.00	0.53	0.42	−1.14	0.56	0.52
−12.63	−0.84	0.55	0.48	−0.96	0.61	0.57	−1.10	0.71	0.64
−12.29	−0.82	0.65	0.63	−0.94	0.74	0.73	−1.08	0.81	0.87
−11.95	−0.81	0.76	0.83	−0.93	0.88	0.96	−1.09	0.95	1.14
−11.61	−0.83	0.89	1.10	−0.97	0.98	1.32	−1.14	1.10	1.51
−11.27	−0.88	1.07	1.41	−1.05	1.17	1.71	−1.24	1.30	1.97
−10.93	−0.99	1.27	1.84	−1.19	1.41	2.18	−1.41	1.54	2.56
−10.59	−1.16	1.47	2.42	−1.40	1.64	2.87	−1.65	1.81	3.32
−10.25	−1.40	1.72	3.18	−1.68	1.93	3.75	−1.99	2.19	4.30
−9.91	−1.72	2.08	4.10	−2.07	2.35	4.84	−2.44	2.61	5.59

Table G2. BDA determination of the UV LF at redshifts 9, 10, and 12. The values of LF are given in logarithmic scale: ϕ [$\log_{10}(M_\odot\ mag^{-1}\ Mpc^{-3})$], σ_{sup}, and σ_{inf} the superior and inferior 68 per cent C.I.

		$z = 9$			$z = 10$			$z = 12$	
M_{UV}	ϕ	σ_{sup}	σ_{inf}	ϕ	σ_{sup}	σ_{inf}	ϕ	σ_{sup}	σ_{inf}
−20.11	−4.37	0.18	0.14	−4.84	0.19	0.16	−5.90	0.21	0.17
−19.77	−4.13	0.14	0.12	−4.57	0.15	0.12	−5.58	0.16	0.15
−19.43	−3.90	0.11	0.09	−4.31	0.12	0.10	−5.27	0.13	0.13
−19.09	−3.68	0.09	0.09	−4.07	0.11	0.09	−4.97	0.12	0.12
−18.75	−3.47	0.09	0.09	−3.84	0.10	0.10	−4.69	0.12	0.14
−18.41	−3.27	0.09	0.10	−3.62	0.10	0.11	−4.42	0.14	0.14
−18.07	−3.08	0.10	0.11	−3.40	0.11	0.13	−4.17	0.14	0.17
−17.73	−2.89	0.12	0.12	−3.20	0.13	0.14	−3.92	0.17	0.19
−17.39	−2.71	0.13	0.13	−3.01	0.16	0.15	−3.69	0.20	0.19
−17.05	−2.54	0.14	0.16	−2.82	0.17	0.17	−3.47	0.21	0.21
−16.71	−2.38	0.16	0.16	−2.64	0.18	0.18	−3.26	0.21	0.22
−16.37	−2.23	0.17	0.17	−2.48	0.18	0.19	−3.07	0.22	0.23
−16.03	−2.09	0.16	0.18	−2.33	0.18	0.20	−2.89	0.23	0.23
−15.69	−1.96	0.17	0.18	−2.19	0.18	0.20	−2.74	0.22	0.24
−15.35	−1.85	0.18	0.18	−2.08	0.19	0.20	−2.61	0.22	0.25
−15.01	−1.75	0.19	0.20	−1.97	0.21	0.22	−2.49	0.24	0.27
−14.67	−1.66	0.23	0.22	−1.87	0.26	0.23	−2.38	0.33	0.27
−14.33	−1.58	0.29	0.24	−1.78	0.32	0.26	−2.27	0.39	0.31

Table G2 – *continued*

	$z = 9$			$z = 10$			$z = 12$		
M_{UV}	ϕ	σ_{sup}	σ_{inf}	ϕ	σ_{sup}	σ_{inf}	ϕ	σ_{sup}	σ_{inf}
−13.99	−1.50	0.36	0.28	−1.70	0.40	0.31	−2.17	0.48	0.39
−13.65	−1.42	0.46	0.33	−1.62	0.53	0.35	−2.09	0.63	0.43
−13.31	−1.36	0.55	0.43	−1.55	0.61	0.47	−2.01	0.71	0.59
−12.97	−1.30	0.64	0.57	−1.50	0.71	0.64	−1.95	0.84	0.79
−12.63	−1.27	0.75	0.76	−1.46	0.83	0.85	−1.92	1.01	1.02
−12.29	−1.25	0.90	0.99	−1.45	1.00	1.09	−1.91	1.17	1.35
−11.95	−1.27	1.03	1.32	−1.48	1.12	1.49	−1.95	1.33	1.81
−11.61	−1.34	1.23	1.71	−1.56	1.35	1.92	−2.06	1.56	2.37
−11.27	−1.46	1.47	2.21	−1.71	1.58	2.49	−2.26	1.84	3.08
−10.93	−1.66	1.70	2.90	−1.93	1.86	3.25	−2.53	2.16	4.01
−10.59	−1.94	1.99	3.80	−2.25	2.18	4.27	−2.92	2.60	5.18
−10.25	−2.32	2.42	4.89	−2.68	2.64	5.49	−3.43	3.07	6.68
−9.91	−2.82	2.88	6.31	−3.21	3.15	7.00	−4.00	3.63	8.20

Table G3. BDA determination of the UV LF at redshift 15. The values of LF are given in logarithmic scale: ϕ [$\log_{10}(\mathrm{M_\odot}\ \mathrm{mag}^{-1}\ \mathrm{Mpc}^{-3})$], σ_{sup}, and σ_{inf} the superior and inferior 68 per cent C.I.

M_{UV}	ϕ	σ_{sup}	σ_{inf}
−20.11	−7.82	0.25	0.20
−19.77	−7.39	0.21	0.17
−19.43	−6.98	0.17	0.17
−19.09	−6.60	0.16	0.17
−18.75	−6.23	0.16	0.19
−18.41	−5.88	0.18	0.22
−18.07	−5.55	0.21	0.24
−17.73	−5.24	0.25	0.25
−17.39	−4.94	0.26	0.28
−17.05	−4.66	0.27	0.30
−16.71	−4.40	0.30	0.29
−16.37	−4.16	0.29	0.31
−16.03	−3.94	0.29	0.31
−15.69	−3.76	0.28	0.31
−15.35	−3.60	0.28	0.32
−15.01	−3.45	0.32	0.35
−14.67	−3.32	0.43	0.34
−14.33	−3.18	0.52	0.39
−13.99	−3.06	0.62	0.49
−13.65	−2.96	0.79	0.56
−13.31	−2.86	0.91	0.75
−12.97	−2.79	1.08	0.98
−12.63	−2.75	1.25	1.32
−12.29	−2.75	1.42	1.76
−11.95	−2.81	1.65	2.33
−11.61	−2.97	1.97	3.00
−11.27	−3.21	2.26	3.95
−10.93	−3.58	2.63	5.18
−10.59	−4.06	3.17	6.61
−10.25	−4.60	3.69	8.16
−9.91	−5.10	4.17	9.38

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