The $z \sim 6$ Luminosity Function Fainter than $-15\,\mathrm{mag}$ from the Hubble Frontier Fields: The Impact of Magnification Uncertainties

R. J. Bouwens[1], P. A. Oesch[2,3], G. D. Illingworth[4], R. S. Ellis[5,6], and M. Stefanon[1]

[1] Leiden Observatory, Leiden University, NL-2300 RA Leiden, The Netherlands
[2] Department of Astronomy, Yale University, New Haven, CT 06520, USA
[3] Observatoire de Genève, 1290 Versoix, Switzerland
[4] UCO/Lick Observatory, University of California, Santa Cruz, CA 95064, USA
[5] European Southern Observatory (ESO), Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany
[6] Department of Physics and Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

Abstract

We use the largest sample of $z \sim 6$ galaxies to date from the first four Hubble Frontier Fields clusters to set constraints on the shape of the $z \sim 6$ luminosity functions (LFs) to fainter than $M_{UV,AB} = -14$ mag. We quantify, for the first time, the impact of magnification uncertainties on LF results and thus provide more realistic constraints than other recent work. Our simulations reveal that, for the highly magnified sources, the systematic uncertainties can become extremely large fainter than -14 mag, reaching several orders of magnitude at 95% confidence at approximately -12 mag. Our new forward-modeling formalism incorporates the impact of magnification uncertainties into the LF results by exploiting the availability of many independent magnification models for the same cluster. One public magnification model is used to construct a mock high-redshift galaxy sample that is then analyzed using the other magnification models to construct an LF. Large systematic errors occur at high magnifications ($\mu \gtrsim 30$) because of differences between the models. The volume densities we derive for faint ($\gtrsim -17$ mag) sources are ~ 3–$4\times$ lower than one recent report and give a faint-end slope $\alpha = -1.92 \pm 0.04$, which is 3.0–3.5σ shallower (including or not including the size uncertainties, respectively). We introduce a new curvature parameter δ to model the faint end of the LF and demonstrate that the observations permit (at 68% confidence) a turn-over at $z \sim 6$ in the range of -15.3 to -14.2 mag, depending on the assumed lensing model. The present consideration of magnification errors and new size determinations raise doubts about previous reports regarding the form of the LF at > -14 mag. We discuss the implications of our turn-over constraints in the context of recent theoretical predictions.

Key words: galaxies: evolution – galaxies: high-redshift

1. Introduction

One of the most important open questions in extragalactic studies regards cosmic reionization and clarifying which sources drive this important phase transition in the early universe. While much evidence suggests that the process might be driven by galaxies (e.g., Robertson et al. 2013, 2015; Bouwens et al. 2015b; Mitra et al. 2015), others have suggested that quasars could provide the dominant contribution (Giallongo et al. 2015; Madau & Haardt 2015; Mitra et al. 2016).

The important issues appear to be whether large numbers of faint quasars exist at high redshift (e.g., Willott et al. 2010; McGreer et al. 2013), whether faint galaxies show an appreciable (>5%) escape fraction (e.g., Siana et al. 2010, 2015; Vanzella et al. 2012, 2016; Nestor et al. 2013), and what the total emissivity is in the rest-frame UV in faint galaxies beyond the limits of current surveys in the Hubble Ultra Deep Field (HUDF: Beckwith et al. 2006; Ellis et al. 2013; Illingworth et al. 2013). Important issues for the latter question are the precise values of the faint-end slopes and the faint-end cut-off to the UV luminosity function (LF). Depending on the value of the faint-end slope and the luminosity where a cut-off in the LF occurs (Bouwens et al. 2012; Kuhlen & Faucher-Giguère 2012; Robertson et al. 2013; Bouwens 2016), the total emissivity from galaxies in the UV can vary by factors of ~ 2–10.

One potentially promising way to constrain the total luminosity density in the rest-frame UV is by taking advantage of the impact of gravitational lensing by galaxy clusters for magnifying individual sources. This can bring extremely faint galaxies into view such that they can be detected with current telescopes (e.g., Bradač et al. 2009; Maizy et al. 2010; Coe et al. 2015). There has been a significant investment in this approach by *HST* in the form of the Hubble Frontier Fields program (Coe et al. 2015; Lotz et al. 2017), which is investing 840 orbits into reaching ~ 29 mag in seven optical+near-IR bands, as well as two UVIS channels in a supporting effort (Siana 2013, 2015; Alavi et al. 2016).

Already, analyses of sources behind the HFF clusters have resulted in the identification of $z \sim 6$–8 sources first to -15 mag (Atek et al. 2014, 2015a, 2015b) and later to approximately -13 mag (Castellano et al. 2016a, 2016b; Kawamata et al. 2016; Livermore et al. 2017 (hereinafter, L17)). At $z \sim 2$–3, it has been similarly possible (Alavi et al. 2014, 2016) to probe to approximately -13 mag taking advantage of very deep WFC3/UVIS observations over Abell 1689 and various clusters in the HFF program. Based on these deep searches, the volume density of galaxies at > -16 mag have been estimated, with quoted faint-end slopes for $z \sim 2$–3 LFs that range from -1.6 to -1.9 (Alavi et al. 2014, 2016) and from -1.9 to -2.1 for $z \sim 6$–8 LFs (Atek et al. 2015a, 2015b; Ishigaki et al. 2015; Castellano et al. 2016b; Laporte et al. 2016; L17), respectively.

In spite of the great potential that lensing clusters have for probing the faint end of the UV LF, successfully making use of data over these clusters to perform this task in an accurate manner is not trivial. The entire enterprise is fraught with sources of systematic error. One of these sources of systematic error concerns the assumed size distribution of extremely faint galaxies (Grazian et al. 2011; Oesch et al. 2015; Bouwens et al. 2017), an issue that also impacts LF determinations from blank fields like the HUDF (but to a lesser degree since the faintest sources asymptote toward being entirely unresolved). Small differences in the assumed half-light radii have the potential to change the inferred faint-end slopes by large factors, i.e., $\Delta\alpha \gtrsim 0.3$ depending on whether the mean size of extremely faint galaxies is 120 mas, 30 mas, or 7.5 mas (e.g., see Figure 2 from Bouwens et al. 2017). Fortunately, we found that most of the extremely faint sources seem consistent with being almost unresolved, i.e., with intrinsic sizes of <10–30 mas (Bouwens et al. 2017; see also Kawamata et al. 2015; Laporte et al. 2016), making this issue much more manageable in terms of its impact; but it still remains an uncertainty. A second source of systematic error arises from errors in the magnification maps, since this can have a profound impact on the LFs derived. Finally, there are issues related to the subtraction of the foreground cluster light, contamination from individual sources in the clusters (e.g., globular clusters), and from other less important systematic effects that affect determinations of the volume densities in the field versus the cluster.[7]

Even without such considerations, it is easy to see that systematics could be a concern for LF studies from lensing clusters, simply by comparing several recent LF results from clusters with similar results based on deep field studies using the HUDF. To give one recent example, Alavi et al. (2016) reported a faint-end slope α of -1.94 ± 0.06 for the UV LF at $z \sim 3$ based on an analysis of sources behind three lensing clusters, while Parsa et al. (2016) reported a faint-end slope of -1.31 ± 0.04 based on a deep $z \sim 3$ search over the HUDF. These results differ at a significance level of $\sim 9\sigma$ taking at face value the quoted statistical errors. This is but one example of the large differences frequently present between LF results derived from deep field studies and those derived on the basis of lensing clusters (see Figure 1 for several other examples).[8]

In addition to the clear scientific importance of the faint-end slope α for computing the total ionizing emissivity from faint galaxies, the observations also allow us to test for a possible flattening or turn-over of the UV LF at low luminosities. Many cosmological hydrodynamic simulations of galaxy formation predict a flattening in the UV LF at approximately −13 or approximately −15 mag due to less efficient atomic and molecular hydrogen cooling, respectively (Muñoz & Loeb 2011; Krumholz & Dekel 2012; Jaacks et al. 2013; Kuhlen et al. 2013; Wise et al. 2014; Finlator et al. 2016; Gnedin 2016; Liu et al. 2016), while other simulations predict a flattening in the range approximately −16 to approximately −13 mag due
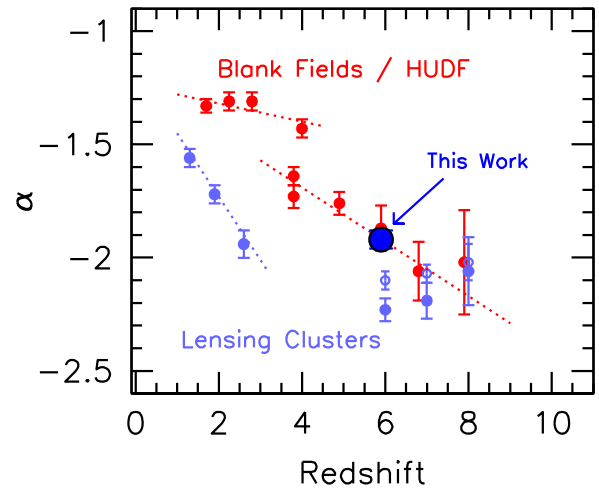
Figure 1. Some recent measurements of the faint-end slope α vs. redshift from the literature using deep fields (red solid circles) and using lensing clusters (light blue solid circles). The field LF results are from Parsa et al. (2016) at $z \leqslant 4$, Bouwens et al. (2007) at $z \sim 4$, and Bouwens et al. (2015b) at $z \geqslant 4$. The $z \leqslant 3$ cluster LF results are from Alavi et al. (2016). The dotted lines show the approximate trends in faint-end slope from each of these studies. The $z = 6$–8 cluster LF results shown are based on a fit to the L17 cluster stepwise LFs anchored to one point (−20 mag) at the bright end of the field LF (see Appendix E). This ensures that the presented faint-end slope α results from L17 are almost entirely independent of field constraints; the nominal faint-end slope results from L17 (including constraints from the field) are shown with the open circles. The large solid dark blue circle shows the faint-end slope α we estimate from our $z \sim 6$ HFF cluster sample in Section 4. As field and lensed LFs potentially probe different luminosity regimes in the UV LF (bright and fainter, respectively), it is possible there would be slight differences in the derived slopes; however, the differences run in the opposite direction normally predicted in simulations (e.g., see the right panel in Figure 1 from Gnedin 2016). Given that the differences between the derived α's are often much larger than the plotted statistical error bars, systematic errors must clearly contribute substantially to some of the determinations plotted here.

to the impact of radiative feedback (O'Shea et al. 2015; Ocvirk et al. 2016). Meanwhile, by combining abundance matching and detailed studies of the color–magnitude diagram of low-luminosity dwarfs in the local universe, evidence for a low-mass turn-over in the LF has been reported at −13 mag (Boylan-Kolchin et al. 2015 see also Boylan-Kolchin et al. 2014). Current observations likely provide us with some constraints in this regime. However, given the significant systematics that appear to be present in current determinations of the faint-end slope α from lensing clusters (versus field results), it is not at all clear that current constraints on the form of the UV LF at >−15 mag are reliable, particularly at $z > 4$.

In the present paper, we take the next step in our examination of the impact of systematic errors on derived LF results from lensing clusters, after our previous paper on this subject, i.e., Bouwens et al. (2017), where the emphasis was on the uncertain sizes of faint sources. Here the focus will be more on the uncertainties in LF results that arise from errors in the gravitational lensing models. As we will demonstrate explicitly, the recovered LF from a straightforward analysis tends to migrate toward a faint-end slope of approximately −2 (or slightly steeper), if uncertainties in the magnification factor are large. The impact of these uncertainties is to wash out features in the LF, particularly at low luminosities. Given that magnification factors μ necessarily become uncertain when these factors are high, i.e., $\mu > 10$ and especially $\mu > 50$, accurately constraining the shape of the LF at extremely low

---

[7] For example, the HFF program does not feature deep observations in the z_{850}-band, which is useful for discriminating between $z \sim 6$ and $z \sim 7$ galaxies, while the HUDF and CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) programs do feature deep integrations in this filter. The availability or not of deep observations in the z_{850} band could impact the $z \sim 6$ and $z \sim 7$ samples and LF results derived from these data sets in different ways.

[8] We plan to both investigate and try to resolve these large differences in a future work (R. J. Bouwens et al. 2017, in preparation).

Table 1
Magnification Models Used Here[a]

Model Name	Mass-Traces-Light	Dark-Matter	Code	Parametric[b]	Resolution ($''$)	References
			"Parametric" Models[b]			
GLAFIC	Y	Y	GLAFIC	Y	$0.''03$	Oguri (2010), Ishigaki et al. (2015), Kawamata et al. (2016)
CATS	Y	Y	LENSTOOL	Y	$0.''1$	Jullo & Kneib (2009), Richard et al. (2014), Jauzac et al. (2015a, 2015b)
Sharon/Johnson	Y	Y	LENSTOOL	Y	$0.''06$	Johnson et al. (2014)
Zitrin-NFW	Y	Y	Zitrin	Y	$0.''06$	Zitrin et al. (2013, 2015)
			"Non-Parametric" Models[b]			
GRALE	N	Y	GRALE	N	$0.''22$	Liesenborgs et al. (2006), Sebesta et al. (2016)
Bradac	N	Y	Bradac	N	$0.''2$	Bradač et al. (2009)
Zitrin-LTM	Y	N	Zitrin	N	$0.''06$	Zitrin et al. (2012, 2015).

Notes.
[a] This includes all publicly available lensing models that have high-resolution mass maps and are generally available for the first four HFF clusters. Our analyses therefore do not include the public HFF models of Diego et al. (2015) and Merten et al. (2015).
[b] Parametric models assume that mass in the cluster is in the form of one or more dark matter components with an ellipsoidal Navarro–Frenk–White (NFW: Navarro et al. 1997) form and to include a contribution from galaxies following specific mass-to-light scalings. Two well-known parametric modeling codes are LENSTOOL (Jullo & Kneib 2009) and GLAFIC (Oguri 2010). For the non-parametric models, both assumptions are typically relaxed, and the mass distributions considered in the models typically allow for much more flexibility than with the parametric models.

luminosities and also detecting a turn-over or flattening is very challenging.

The purpose of this paper is to look at the constraints we can set on the faint end of the $z \sim 6$ UV LF with a thorough assessment of the possible systematic errors. In doing so, we will look for possible evidence of a turn-over in the LF at very low luminosities and if not present, what constraints can be placed on the luminosity of a turn-over. Evidence for a turn-over will be evaluated through the introduction of a curvature parameter, which we constrain through extensive Markov-Chain Monte-Carlo (MCMC) trials. The confidence intervals we obtain on the shape of the UV LF at faint magnitudes will provide theorists with some important constraints for comparison with models and cosmological hydrodynamic simulations. Most importantly, these results provide balance to some discussion in the literature, where premature claims appear to have possibly been made regarding the LF's. To keep the focus of this paper on our new techniques, we restrict our analysis to just the $z \sim 6$ LF from the first four HFF clusters.

The plan for this paper is as follows. Section 2 summarizes the data sets we use to select our $z \sim 6$ sample and derive constraints on the $z \sim 6$ LF. Section 3 provides some useful context for the issue of errors in the magnification models and shows the general impact it would have on LF results. Sections 4 and 5 present new LF results at $z \sim 6$ using our new forward-modeling methodology. Section 6 compares our new results with previous reported LF results, as well as results from various theoretical models. Finally, in Section 7, we summarize and conclude. We refer to the *HST* F814W, F850LP, F105W, F125W, F140W, and F160W bands as I_{814}, z_{850}, Y_{105}, J_{125}, JH_{140}, and H_{160}, respectively, for simplicity. Estimates of the UV luminosities are made at ~ 1800 Å for the typical source in the sample. Through this paper, a standard "concordance" cosmology with $H_0 = 70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$ is assumed. This is in good agreement with recent cosmological constraints (Planck Collaboration et al. 2015). Magnitudes are in the AB system (Oke & Gunn 1983).

2. Data Sets and the $z \sim 6$ Sample

In our selection of $z \sim 6$ galaxies, we make use of the v1.0 reductions of the deep *HST* optical and near-IR *HST* observations available over the first four clusters in the HFF program: Abell 2744, MACS 0416, MACS 0717, and MACS 1149 (A. Koekemoer et al. 2016, in preparation; Lotz et al. 2017). The optical observations include ~ 18, ~ 10, and ~ 42 orbits of ACS observations in the F435W, F606W, and F814W bands from 0.4 to 0.9 μm. Near-IR observations over these fields total 34, 12, 10, and 24 orbits in the F105W, F125W, F140W, and F160W, reaching a roughly 5σ limiting magnitude of 28.8–29.0 mag.

Subtraction of foreground light from cluster galaxies and cluster galaxies was performed using galfit (Peng et al. 2002) and the median-filtering algorithm of SExtractor (Bertin & Arnouts 1996) run at two different grid scales. There are many similarities of our procedure to that from Merlin et al. (2016). The only areas clearly inaccessible to us in our searches for faint $z \sim 6$ galaxies occur directly under the cores of bright stars or galaxies in the cluster. Our procedure performs at least as well as any other procedure currently in use (Merlin et al. 2016; L17). Relative to the approaches of Merlin et al. (2016) or L17, our procedure appears to perform comparably well. One measure of this is the number of $z = 6$–8 galaxies we identify behind Abell 2744 and MACS 0416 (considered in both previous studies) for the current analysis. Our samples are $\gtrsim 10\%$ larger than that utilized in either previous study and could be enlarged further by 10%–20% by making use of different detection images (Appendix A).

A complete description of both our photometric procedure and selection criteria for identifying $z \sim 6$ galaxies is provided in R. J. Bouwens et al. (2017, in preparation). In most respects, our procedures are similar to that done in many of our previous papers (e.g., Bouwens et al. 2015), but we do note that we perform our photometric measurements after subtraction of the intracluster and bright elliptical galaxy light. While other procedures report sizeable differences between the total magnitude measurements on the original
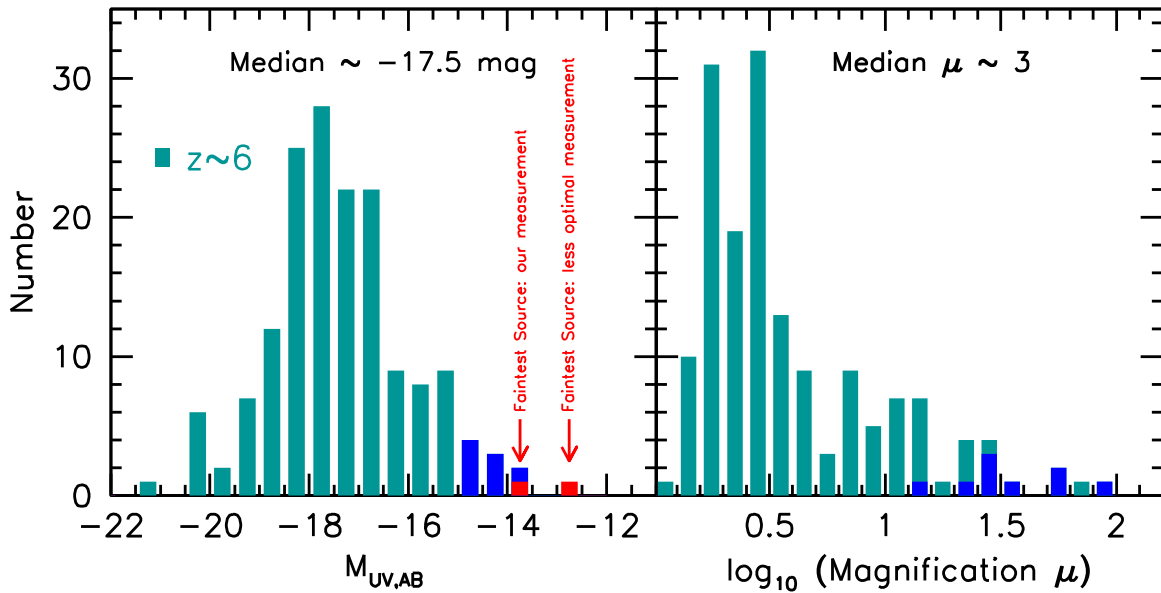
Figure 2. Number of galaxies found in our conservative selection of $z \sim 6$ galaxies behind the first four HFF clusters vs. their inferred M_{UV} luminosity (left panel) and magnification factor (right panel). We take the magnification factor to be the median of those derived from the four parametric models (GLAFIC, CATS, Sharon/Johnson, and Zitrin-NFW), enforcing a maximum value of 100 (due to the much weaker predictive power for the models at such high magnification factors: see Figure 3). The one source over our fields with a magnification factor in excess of 100 and is M0416I-6118103480 (04:16:11.81, −24:03:48.1) with a nominal magnification of 145 (nominally implying an absolute magnitude of −13.4 mag). The nine sources with the faintest intrinsic luminosities are shown in blue in each panel. The faintest source in our probe is sensitive to how total magnitude measurements are made and which magnification models are used. The two red squares show the luminosity of our faintest source, as we measure it with our total flux approach (left red square) and also (right red square) consistent with the way that L17 measure luminosities for many of the sources in their $z \sim 6$–8 samples (see Sections 2 and 6.1.2). The luminosity shifts $\sim$0.7 mag faintward for these sources in the latter approach.

and subtracted images (e.g., L17), with measurements on the original images giving brighter magnitudes, we only find a 0.03 ± 0.07 mag difference for the median source in these measurements. Further evidence for the fact that our procedures do not underestimate the total flux in sources can be seen by comparing our photometry with other groups (Sections 6.1.1 and 6.1.2). Our magnitude measurements are typically $\sim$0.1–0.3 mag brighter than other groups for the same sources.

We briefly summarize our criteria here for selecting a robust and large sample of $z \sim 6$ galaxies. We select all sources that satisfy the following I_{814}-dropout color criteria

$$(I_{814} - Y_{105} > 0.6) \wedge (Y_{105} - H_{160} < 0.45)$$
$$\wedge (I_{814} - Y_{105} > 0.6(Y_{105} - H_{160}))$$
$$\wedge (Y_{105} - H_{160} < 0.52 + 0.75(J_{125} - H_{160}))$$

and that are detected at $>6.5\sigma$, adding in quadrature the S/N of sources in the Y_{105}, J_{125}, JH_{140}, and H_{160} band images measured in a 0″35-diameter aperture. The above color selection criterion also explicitly excludes the inclusion of $z \sim 8$ Y_{105}-dropout galaxies. Because the above criteria identify sources at both $z \sim 6$ and $z \sim 7$, we compute the redshift likelihood function $P(z)$ for each source and only include those sources where the best-fit photometric redshift is less than 6.3. Sources are further required to have a cumulative probability of $<35\%$ at $z < 4$ to keep contamination to a minimum in our high-redshift samples.

Our sample of 160 $z \sim 6$ candidate galaxies is the largest compilation reported to date. Each of the HFF clusters we examine in this study have at least seven independent lensing models available, with both convergence κ and shear γ maps

(Table 1). We estimate the magnification of sources based on publicly available models by first multiplying the κ and γ maps of each cluster by D_{ls}/D_s and then computing the magnification μ as follows.

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

where D_{ls} and D_s represent the angular-diameter distances from the lensing cluster to the magnified galaxy and the angular-diameter distances to the source, respectively. For our magnification estimates for individual sources, we take the median of the model magnifications from the CATS (Jullo & Kneib 2009; Richard et al. 2014; Jauzac et al. 2015a, 2015b), GLAFIC (Oguri 2010; Ishigaki et al. 2015; Kawamata et al. 2016), Sharon/Johnson (Johnson et al. 2014), and Zitrin parametric NFW models (Zitrin-NFW: Zitrin et al. 2013, 2015). The parametric models generally provided the best estimates of the magnification for individual sources in the HFF comparison project (Meneghetti et al. 2016), but we emphasize that many non-parametric magnification models also performed very well.

We present in Figure 2 the distribution of absolute magnitudes and magnification factors we estimate for sources in our $z \sim 6$ sample. Absolute magnitudes for our $z \sim 6$ sample are taken to equal the inverse-weighted mean of the fluxes measured in the F105W, F125W, F140W, and F160W bands (such that the rest-frame wavelength for our absolute magnitude measurements is $\sim$1800 Å). We set an arbitrary maximum magnification factor of 100, given the lack of

predictive power for magnification maps at such high values (see Section 3.1).[9] Our selection includes sources ranging from −22 to −13.5 mag and with magnification factors ranging from 1.2 to 145, with the bulk of the sources having absolute magnitudes of −18 and magnification factors of ∼2.

We should emphasize that the inferred luminosities and total magnitudes we report for sources are intended to provide a rather complete accounting for light in individual sources. They are based on scaled-aperture photometry following the Kron (1980) method, with a correction for flux on the wings of the PSF (e.g., see Bouwens et al. 2015a). However, in comparing our total magnitude measurements with the magnitude measurements from other groups (e.g., L17: see Section 6.1.2), we have found that some sources have been reported to have apparent magnitude measurements fainter by ∼0.3–0.5 mag than what we measure for the same sources. In addition, other teams occasionally report 1.3–1.8× higher values for the magnification factor of individual sources than we calculate based on the same models, e.g., the faintest source in L17 (Appendix F).

If we quote the luminosities of sources in our study using a similar procedure as to what L17 appear to use—where individual sources are often ∼0.4 mag fainter than we find—and adopt 1.3–1.8× higher magnification factors, our probe would extend to −12.6 mag (indicated in Figure 2 with the red bin), essentially identical to that claimed by L17 (see also Castellano et al. (2016a) and Kawamata et al. 2016). We emphasize, however, that the low luminosities claimed by measuring magnitudes in this way (and computing magnification factors in this way: see Appendix F) likely exaggerate how faint the HFF program probes. We discuss this further in Sections 6.1.2 and 6.2. We prefer our photometric scheme for accounting for the total light in faint sources.

3. Impact of Magnification Errors on the Derived LFs

An important aspect of the present efforts to provide constraints on the $z \sim 6$ LF will be our explicit efforts to include a full accounting of the uncertainties present in the magnification models we utilize. We begin by looking first at the general size of errors in the magnification models and second at how the errors would impact LFs derived from lensing clusters.

3.1. Predictive Value of the Public Magnification Models

In making use of various gravitational lensing models to derive constraints on the prevalence of extremely faint galaxies at high redshift, it is important to obtain an estimate of how predictive the lensing models are for the true magnification factors.

One way of addressing this issue is the fully end-to-end approach pursued by Meneghetti et al. (2016) and involves constructing highly realistic mock data sets, analyzing the mock data sets using exactly the same approach as are used on the real observations, and then quantifying the performance of the different methods by comparing with the actual magnification maps. While each of the methods did fairly well in

reproducing the magnification maps to magnification factors of ∼10, the best performing methods for reconstructing the magnification maps of clusters were the parametric models, with perhaps the best reconstructions achieved by the GLAFIC models, the Sharon/Johnson models, and the CATS models.

An alternate way of addressing this issue is by comparing the public lensing models against each other. Here we pursue such an approach. We treat one of the models as the truth and then to quantify the effectiveness of the other magnification models taken as a set for predicting that model's magnifiction map. We consider both cases in which the true mass profile of the HFF clusters is considered (1) to lie among the parametric class of models built on NFW-type mass profiles and (2) to lie among the non-parametric class of models, which allows for more freedom in the modeling process. We take the former models to include the GLAFIC, CATS, Sharon/Johnson, and Zitrin-NFW models, and the latter to include the Bradac (2009), GRALE (Liesenborgs et al. 2006; Sebesta et al. 2016), and Zitrin-LTM (Zitrin et al. 2012, 2015).[10] A brief description of the general properties of the public lensing models can be found in Table 1. In performing this test, we assume that the median of the magnification models provides our best means for predicting magnifications in the model we are treating as the truth. The truth model is always excluded when constructing the median magnification map for this test.

Alternatively treating each of the magnification models as the truth, we then quantify what the median magnification factor is in the truth model as a function of the median magnification factors from the other models. For perfectly predictive models, the magnification factors in the truth model would be precisely centered around the median magnification factors from the other models. In practice, this is not true, given the difficulty in predicting the precise locations of the rare regions around the cluster with the highest magnification factors. While one can control for these uncertainties through the use of quantities like the median, even the median will overpredict the true magnification, due to the impact of model "noise" on the medians and the possibility for chance overlap in the high-magnification regions across the models.

For the most general results, we take a geometric mean of the median magnification factors considering each model as the truth and then plot the results in the upper panel of Figure 3. Results on the predictive power of the parametric (GLAFIC, CATS, Sharon/Johnson, Zitrin-NFW) and non-parametric (GRALE, Bradac, Zitrin-LTM) models are presented separately with magenta and blue colored lines. The dashed and dotted lines give the "true" magnifications recovered versus median magnification factors for best and worst performing clusters. Meanwhile, the solid line between the dashed and dotted lines gives the geometric mean of the "true" magnifications across all four clusters considered here. The lower panel of Figure 3 shows the position-to-position scatter around the median magnification in the model treated as the truth. From this exercise, it is clear that the magnification maps have excellent predictive power to magnification factors of ∼10 in all cases and perhaps to even higher magnification factors assuming that the magnification profiles of HFF clusters are as well behaved as in the parametric models. The scatter, however, is already

---

[9] Our use of an upper limit on the magnification factors only affects one source, i.e., M0416I-6118103480 (04:16:11.81, −24:03:48.1) with a nominal magnification of 145 (nominally implying an absolute magnitude of −13.4 mag) and only has a minor impact on the parameters we derive for the $z \sim 6$ LF in Section 5 (changing α, δ, and ϕ^* by ⩽0.01, ⩽0.1, and less than 2%, respectively).

[10] Zitrin-LTM does not technically qualify as parametric or non-parametric, since the mass profile is governed by the distribution of light in a cluster. However, since the model shows a greater dispersion relative to the parametric models, we include it in the non-parametric group.
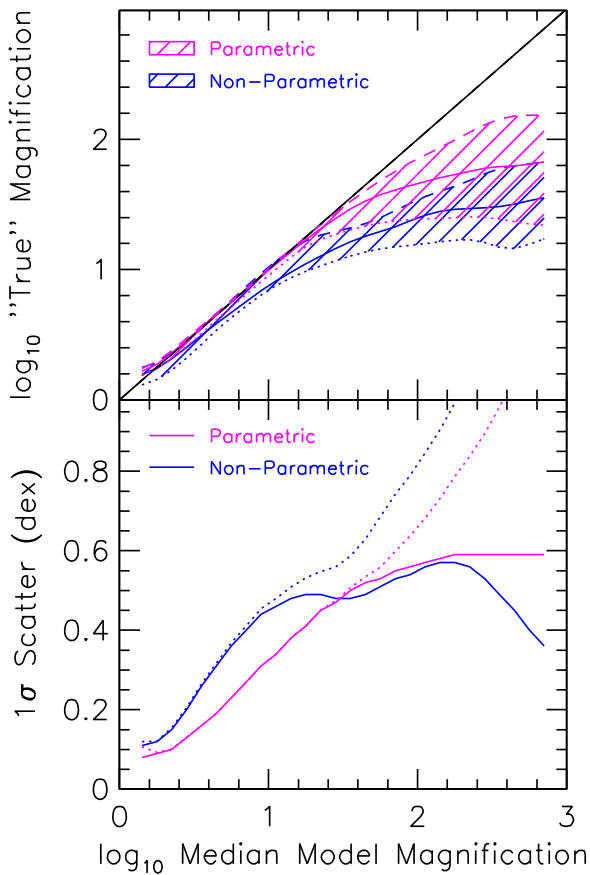
Figure 3. Evaluation of the predictive power of the lensing models and the median magnification maps. (Upper panel) Illustration of how well the median magnification factor from all the magnification models but one (variable on the x-axis) predicts the median magnification factor for the excluded magnification model, i.e., the "truth" model (variable on the y-axis). The plotted magnification plotted along the y-axis shows the geometric mean of the results, alternatively taking each model to be the truth. The dashed and dotted magenta lines show the recovered magnification factors for the parametric magnification models (i.e., GLAFIC, CATS, Sharon/Johnson, Zitrin-NFW) from the best and worst performing clusters as well. The solid magenta line shows the geometric mean of the recovered magnification factor across all clusters considered here. The blue dashed and dotted lines show the equivalent results excluding the non-parametric magnification models (GRALE, Bradac, and Zitrin-LTM) from the process. Again the solid blue shows the geometric mean of the recovered magnification factors for all clusters. For perfectly predictive magnification models, the plotted lines would follow the black diagonal line with a slope of 1. (Lower panel) Scatter in the magnification factors vs. median magnification factor for the parametric magnification models (magenta solid line). The blue solid line gives the results for the non-parametric models. The dotted lines are the same as the solid lines but also add in quadrature the logarithmic differences between the actual magnification factors in a model and that predicted from a median of the other models. From this figure, it is clear that the median magnification model has largely lost its predictive power by magnification factors of ∼10 and ∼30 assuming that the available non-parametric and parametric models, respectively, are representative of reality.

very large at magnification factors of 10. We will extend this exercise in a future work (R. J. Bouwens et al. 2017, in preparation).

The exercise we perform in this section shows similarities in philosophy to the analyses that Priewe et al. (2017) pursue, in comparing magnification models over the HFF clusters with each other to determine the probable errors in the individual magnification maps. One prominent conclusion from that study was that differences between the magnification maps were almost always larger than the estimated errors in the

magnification for a given map, pointing to large systematics in the construction of some subset of the individual maps. This provides some motivation for the tests we perform here and in future sections in this paper and confirmation of the importance of this study. Other powerful tests of the predictive power of the magnification maps, and the challenges, were provided by observations of SNe Ia (Rodney et al. 2015).

3.2. Impact of Magnification Errors on the Recovered LFs

The purpose of this subsection is to illustrate the impact of magnification errors on the derived LFs from the HFF clusters. Two different example LFs are considered for this exercise: (1) one with a faint-end slope of −2 and a turn-over at −15 and (2) another with a fixed faint-end slope α of −2 and no turn-over.

How well can we recover these LFs given uncertainties in the magnification maps? We can evaluate this by generating a mock catalog of sources for each of the first four clusters from the HFF program using one set of magnification models ("input" models) and then attempting to recover the LF using another set of magnification models ("recovery" models). These catalogs include positions and magnitudes for all the individual sources in each cluster. In computing the impact of lensing, the redshifts are fixed to $z = 6$ for all sources. The input magnification models are taken to be the GLAFIC models for this exercise. Following previous work (e.g., Ishigaki et al. 2015; Oesch et al. 2015), each galaxy in the image plane is treated as coming from an independent volume of the universe, allowing us to construct the input catalogs from the model magnification maps alone (and therefore not requiring use of the deflection maps). This choice does not bias the LF results in our analysis relative to analyses that account for multiple imaging of the same galaxies (from the source plane), since both the cosmological volume and total number of background galaxies is increased in proportion to the overcounting. The selection efficiencies of sources are accounted for when creating the mock catalogs, as estimated in Appendix B. In performing this exercise, we ignore errors in our estimates of the selection efficiencies and small number statistics at the faint end of the LF.

One can try to recover the input LFs from these mock catalogs, using various magnification models. Sources are binned according to luminosity using the "recovery" magnification model. The selection volumes available in each luminosity bin are also estimated as described in Appendix B using this "recovery" magnification model. To demonstrate the overall self-consistency in our approach, we show the recovered LFs using the same magnification model as we used to construct the input catalogs in the top two panels in Figure 4.

What is the impact if different lensing models are used to recover the LF than those used to construct the mock catalogs? The lowest two rows of panels in Figure 4 show the results using the latest magnification models by GRALE, CATS, Zitrin-LTM, and the median of the CATS, Sharon/Johnson, and Zitrin-NFW models where available.

These simulation results demonstrate that the recovery process appears to work very well for input LFs with faint-end slopes of −2 (left panels in Figure 4) independent of the magnification model, with all recovered LFs showing a form that is very similar to that of the input LFs.

Very different results are, however, obtained in our attempts to recover input LFs with a turn-over at −15 mag (right panels
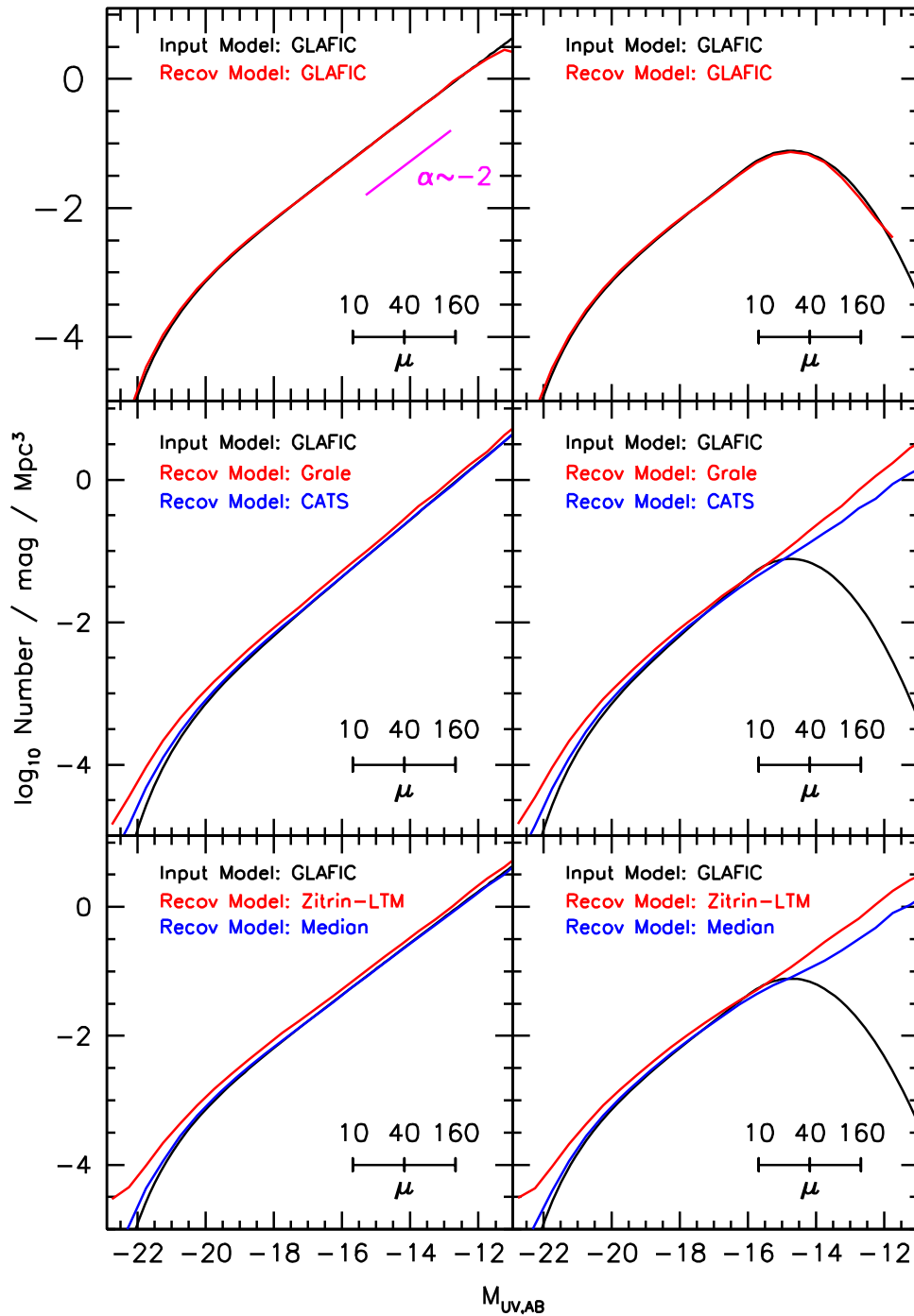
Figure 4. Comparison of the input LFs (black lines) into our forward-modeling simulations and the recovered LFs when using the same magnification models (top panels) and when using four different magnification models, including GRALE (red lines: middle panels), CATS (blue lines: middle panels), Zitrin-LTM (red lines: lower panels), and the median parametric model (blue lines: lower panels). A ticked horizontal bar is added to the panels to indicate the approximate luminosities probed by sources of a given magnification factor near the faint end of the HFF data set, i.e., 28.5 mag. Two input LFs are considered: one where the LF exhibits a faint-end slope of −2 with no turn-over at low luminosities (left panels) and a second also exibiting a faint-end slope of −2 but with a turn-over at −15 mag (right panels). In the first case, the recovered LFs show a faint-end slope α of −2 to very low luminosities, in agreement with the input LF. However, for the second case, the recovered LFs again show a faint-end slope α of −2 to very low luminosities, in significant contrast to the input LF. As a result, interpreting the LF results from lensing clusters can potentially be tricky, as the detection of a turn-over in the LF at > −15 mag is very challenging (see Section 3.2). This is due to the weaker predictive power of the magnification models at high magnification factors $\mu > 10$ and especially $\mu > 30$ (Figure 3). See also Figures 16 and 17 from Appendices B and C.

in Figure 4) using magnification models that are different from the input model. For all four magnification models we consider, the recovered LFs look very similar to the LF example we just considered. All recovered LFs show a steep faint-end slope to −11 mag. What is striking is that they do not reproduce the

turn-over present in the input model at −15 mag. There are some differences in the recovered LFs depending on how similar the input magnification model is to the recovery model, with effective faint-end slopes of −2, −1.8, and −1.7 achieved with the GRALE and Zitrin-LTM models, the CATS models,
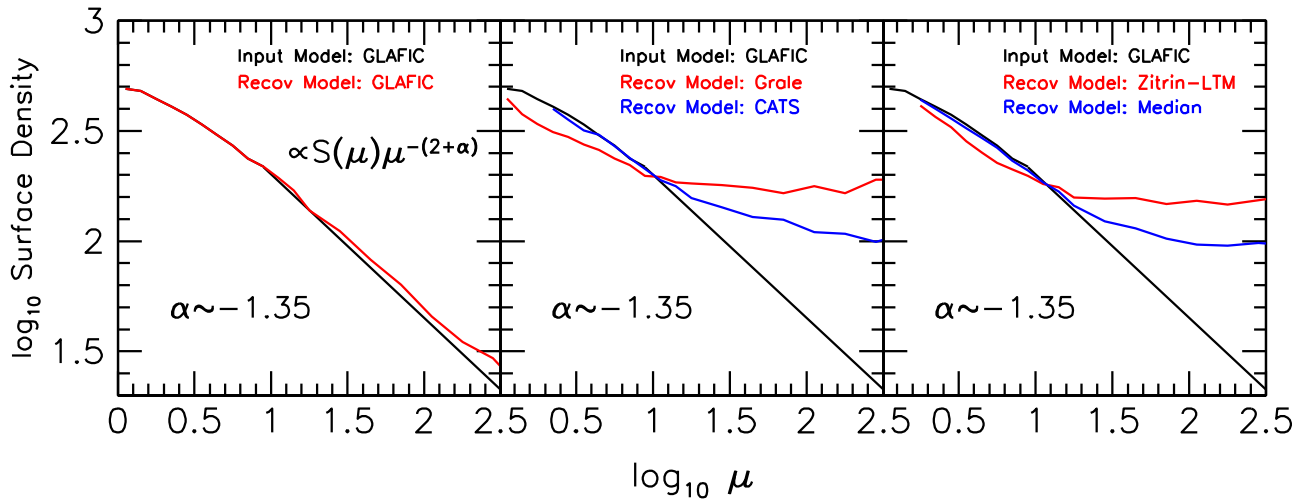
Figure 5. Use of forward modeling to demonstrate the expected dependence of the recovered surface densities of $z \sim 6$ sources on the model magnification factor μ. The input distribution of sources around the first four HFF clusters is generated using the GLAFIC model and a faint end slope of -1.35 and then recovered using the GLAFIC, GRALE, CATS, Zitrin-LTM, and median parametric model. In the case of perfect magnification maps, the surface density of sources is expected to depend on magnification μ as $S(\mu)\mu^{-(2+\alpha)}$, where $S(\mu)$ is the magnification-dependent selection efficiency. At sufficiently high magnifications, the predictive power of the lensing models breaks down and one would expect there to be no correlation between the surface density of galaxies and the model magnification factor, as the present forward model results illustrate in the center and right panels. In such a case, the recovered LF has a faint-end slope that asymptotes toward the value that implies a fixed surface density of sources above some magnification factor. In the case in which the selection efficiency does not depend on the magnification factor, this faint-end slope would be -2. However, in the more general case presented in Appendix C, the faint-end slope asymptotes to $-2 + d(\ln S(\mu))/d(\ln \mu)$.

and the median parametric models, respectively. The GLAFIC magnification model is not used when constructing the median parametric magnification model.

Both examples demonstrate that the faint-end slope for the recovered LFs tends to gravitate toward a value of -2. It is useful to provide a brief explanation as to why. For a power-law LF, i.e., L^{α}, and ignoring any dependence of the selection efficiency on magnification factor, one expects the surface density of sources on the sky to depend on magnification factor μ as $L^{\alpha}/\mu \, dL|_{L=L_{obs}/\mu} \propto \mu^{-\alpha-2} \propto \mu^{-(2+\alpha)}$, where L and L_{obs} represent the intrinsic and observed luminosities, respectively. For faint-end slopes shallower than -2, one therefore expects a systematic decrease in the surface density of sources on the sky as the magnification increases; for faint-end slopes of -2, one expects no dependence on source magnification; and for faint-end slopes steeper than -2, one expects a systematic increase in the surface density of sources as the magnification increases. All of the above statements are for the intrinsic surface densities. The observed surface densities will be impacted by the magnification-dependent selection efficiencies $S(\mu)$.

We illustrate this expected dependence on the magnification factor in Figure 5 for an LF with a faint-end slope of -1.35 by laying down sources behind the HFF clusters using the GLAFIC magnification model. The surface density of the sources versus magnification factor can then be recovered using a variety of other models. At sufficiently high magnification factors, the uncertainties in the magnification factors become large, washing out any dependence on the magnification factor. This results in a relatively constant surface density of sources and a faint-end slope of -2.

Two other examples of the impact of large magnification errors on LF results are presented in Figures 16 and 17 in Appendices C and D, utilizing an input LF with a faint-end slope of -1.3. For each of these examples, the recovered LF closely matches the input LF; dramatically, however, faintward of -15 mag (and even -16 mag for some models), the recovered LFs steepen and asymptote again toward a

faint-end slope of approximately -2 (or steeper if sources are resolved), even if the actual slope of the LF is much shallower (or the LF turns over!).

Each of these examples demonstrate that, regardless of the input LF, a faint-end slope α of approximately -2 will be recovered whenever the magnification uncertainties are large. One cannot, therefore, use the consistent recovery of a steep faint-end slope based on a large suite of lensing models to argue that the actual LF maintains a steep form to extremely low luminosities (as was done by L17 using their Figure 12). The presented examples show that this is not a valid argument.

How then can one interpret LF results from lensing clusters when a steep LF $\alpha \sim -2$ is found? As we have demonstrated, such a result could be indicative of the LFs truly being steep or simply an artifact of large magnification uncertainties. To determine which is the case, the safest course of action is to simulate all steps in the LF recovery process, to determine the impact of magnification uncertainties on the shape of the LF, and finally to interpret the recovered LFs from the observations. While we showed a few examples here, we formalize the process in the next section.

4. New Forward-modeling Methodology to Derive LF Results

The purpose of the present section is to describe a new methodology for quantifying the constraints on the UV LF to very low luminosities, given the uncertainties in the magnification maps. The development of such a procedure is useful given the challenges presented in the previous section. We will apply this formalism in Section 5.

4.1. Basic Idea and Utility

The LF recovery results presented in Section 3 (Figure 4) illustrate the impact that errors in the magnification maps can have on the recovered LFs. Input LFs, of very different forms, can be driven toward a faint-end slope α of -2 at the faint end,
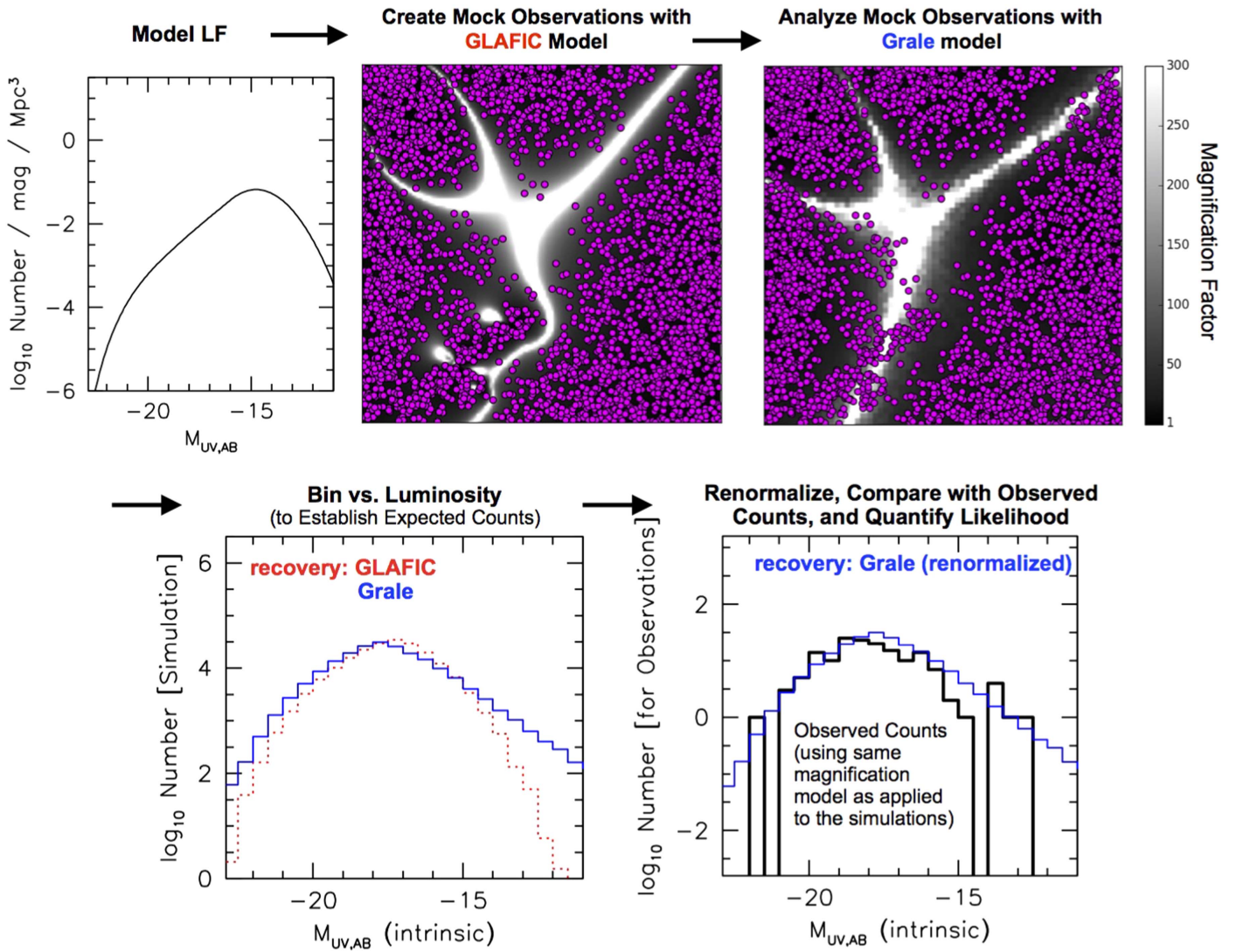
Figure 6. Illustration of the steps in our forward-modeling approach to determine the impact of errors in the lensing models on the derived LF results (Section 4.2: see also Sections 3.2 and 4.1). The upper middle panel shows the positions of the faint $H_{160,AB} > 28$ sources (violet circles) from a mock catalog created over a $14'' \times 14''$ region in the image plane near the center of Abell 2744 based on a model LF (shown in the upper left panel) and the GLAFIC lensing model, with the color bar at the top right providing the magnification scale for various shades of black (low) and white (high). (Note that sources are distributed uniformly over the source plane for the construction of the mock catalog.) The upper right panel shows where this same catalog of sources lies in the image plane relative to the critical lines in the GRALE lensing model over the same region in Abell 2744. The lower left panel shows histograms of the number of sources in our mock catalogs vs. luminosity, using both the original GLAFIC model used to construct the mock catalogs (dotted red histogram) and GRALE model used for recovery (blue histogram). We use these simulations to derive the expected number of galaxies per luminosity bin for a given LF and compare this with the observed numbers (where the intrinsic M_{UV} is calculated using the GRALE model) to estimate the likelihood of a given LF model (lower right panel). In the presented example, the turn-over in the LF at the faint end translates into a significant deficit of sources near the critical lines using the input magnification model. However, when interpreting this same catalog using a different lensing model, i.e., GRALE in this case, many sources nevertheless lie very close to the critical lines. As a result of the uncertain position of the critical curves, it can be challenging to detect a turn-over in the LF at > -15 mag.

after accounting for the impact of magnification errors. The results from Section 3.2 demonstrate the importance of forward modeling the entire LF recovery process to ensure that both the results and uncertainties are reliable.

We then utilize our forward-modeling approach to derive constraints on the $z \sim 6$ LF. The basic idea behind our approach follows closely from the simulations we ran in the previous section and is illustrated in Figure 6. We begin by treating one of the public magnification models as providing an exact representation of reality. In conjunction with an input LF, those models are used to create a mock data set for the four HFF clusters considered. The mock data set is then interpreted using other magnification models for the clusters to determine the distribution of sources versus UV

luminosity M_{UV} and also to recover the UV LF. As illustrated by the LF recovery experiments presented in Section 3.2, the recovery could be done with the models individually or by using some combination of models like the median.

There are many advantages to using the present procedure to derive accurate errors on the overall shape of the UV LF. Probably the most significant of these is inherent in the end-to-end nature of the present procedure and our relying significantly on forward modeling to arrive at accurate errors on the observational results. Through the construction of many mock data sets using plausible magnification models and recovery using other similarly plausible models, the proposed procedure allows us to determine the full range of allowed LFs.

In addition, the advocated procedure provides us with a natural means to account for "noise" in the lensing model magnification maps. The presence of lensing model "noise" is obvious looking at the range in magnification factors across the various models (e.g., compare the GLAFIC and GRALE critical lines in Figure 6 or see the lower panel in Figure 3). Such noise can even be present in a median magnification model created from the combination of many individual models because there will be regions where the high-magnification regions of the lensing maps simply overlap due to chance coincidence. Analogous to considerations of low-significance sources in imaging observations, the robustness of specific magnification factors in the median map can be assessed, by considering comparisons with independent determinations of the same map. Through the treatment of one of the public magnification models as the truth, the present forward-modeling approach effectively formalizes such a technique to determine the robustness of specific features in the magnification maps. The advantage of the current procedure is that this robustness can be determined using the full magnification maps available for each cluster (and not just at a limited number of positions where candidate high-magnification $\mu > 10$ sources happen to be found in the real observations), while also allowing us to derive more reliable likelihood distributions (with realistic errors).

Another advantage of our forward modeling procedure is that it explicitly incorporates source selection. This is important, since the selection efficiency S could depend on the magnification factor μ in the sense that the most magnified sources would also be the most incomplete, as is likely the case for the brightest and most extended objects, due to the impact of lensing shear on source detection (Oesch et al. 2015). If the same situation applied to the faintest sources in the HFFs (and one did not utilize a procedure that included forward modeling or an explicit correction), the recovered LFs would be biased. This is due to the fact that the actual surface density of the sources on the sky is proportional to $S(\mu_{\rm true})$, but it is assumed to be proportional to $S(\mu_{\rm model})$ and $\mu_{\rm true} < \mu_{\rm model}$ at high magnifications $\mu > 10$ (Figure 3). The recovered LF would therefore be higher by the factor $S(\mu_{\rm true})/S(\mu_{\rm model})$.

We would expect such an issue to affect recovered LFs, in all cases where sources have non-zero size (since the selection efficiency would then depend on the magnification factor). For example, if the completeness of sources shows an inverse correlation with the magnification factor as Oesch et al. (2015) find, i.e., $S(\mu) \propto \mu^{-0.3}$ (e.g., in their Figure 3), a direct approach would recover a faint-end slope that is $\Delta\alpha \sim 0.3$ steeper than in reality (Figure 16 from Appendix C), at very low luminosities, where $\mu > 10$ (where $\langle\mu_{\rm true}\rangle$ is typically less than $\langle\mu_{\rm model}\rangle$).

Remarkably, there is no evidence that this issue is even considered in many recently derived LFs, which is worrisome given the size assumptions that were made. This is most problematic for analyses pushing to very low luminosities, i.e., >-15 mag, while quoting tiny statistical uncertainties (e.g., L17 who quote statistical uncertainties on α of ± 0.04 versus this bias, which is $\sim 8\times$ larger).

4.2. Procedure

We perform our forward modeling simulations at the catalog level, to ensure that the time requirements on these simulations are manageable. This involves the construction of catalogs of sources with precise positions and apparent magnitudes. Both in the construction of the mock catalogs and in recovering the LF from these catalogs, the selection efficiency S must be accounted for, which is, in general, a function of the apparent magnitude m and magnification factor μ, i.e., $S(m, \mu)$. For the lowest luminosity $z \sim 5$–8 galaxies, there is little evidence to suggest that these sources show significant spatial extension (Bouwens et al. 2017), which implies that we can credibly treat their selection efficiencies as just a function of the apparent magnitudes, i.e., $S(m)$. We describe our procedure for estimating $S(m)$ in this case in Appendix B.[11]

In putting together the mock observed catalogs for each LF parameter set we are considering, i.e., ϕ^*, α, and a third parameter δ to be introduced in the next section, $\sim 2 \times 10^5$ sources are inserted at random positions (yet uniformly in the source plane) across the four HFF cluster fields we are considering (Figure 6). Sources are included in the catalogs in proportion to their estimated selection efficiencies $S(m, \mu)$ (Appendix B), their implied volume densities (according to the model LF), and cosmological volume element (proportional to the area divided by the magnification factor). It is the inclusion of sources in the catalogs in inverse proportion to the magnification factor that ensures that galaxies are distributed uniformly within the source plane (since our catalog construction process does not consider the deflection maps from the lensing models or multiple imaging).[12] All sources in the input catalogs are assumed to have the same input redshift $z = 6$.

During the recovery process, sources are placed into individual bins in UV luminosity using a "recovery" magnification model, which we take to be the median of the parametric magnification models.[13] During the recovery process, the redshift is taken to have the same mean value as assumed in constructing the mock catalogs, i.e., $z = 6.0$, but with a 1σ uncertainty of 0.3. This is to account for the impact of uncertainties in the estimated redshifts of individual sources on the recovered LFs.

We use the results of the simulations we run for each parameter set (each with $\sim 2 \times 10^5$ sources) to establish the expectation values for the number of sources per luminosity bin (in the same 0.5 mag intervals used in the previous section). We then compute the likelihood of a given parameterization of the LF by comparing the observed number of galaxies per bin in luminosity (considering all four clusters at the same time) with the expected numbers assuming a Poissonian distribution, as

$$\Pi_i e^{-N_{\rm exp,i}} \frac{(N_{\rm exp,i})^{N_{\rm obs,i}}}{(N_{\rm obs,i})!}$$

where $N_{\rm exp,i}$ is the expected number of sources in a given bin in intrinsic UV luminosity. The observed number of sources in a given bin in UV luminosity $N_{\rm obs,i}$ is computed from the median parametric magnification maps, as done in Section 2 (Figure 2). A flow chart showing one example of our forward-modeling approach is provided in Figure 6.

---

[11] Of course, we also recognize the value in understanding the impact on the results if the sizes of sources are larger, and this is discussed in Appendix C. The outcome is similar but leads to an even bigger disconnect from the real LF shape.

[12] We verified that sources in the mock catalogs produced by our procedure showed a uniform volume density of galaxies (to some limiting luminosities) in the source plane, independent of the magnification factor tying some region of the image plane to the source plane.

[13] The input magnification model is always excluded when constructing the median magnification map (used for recovery) to keep the process fair.
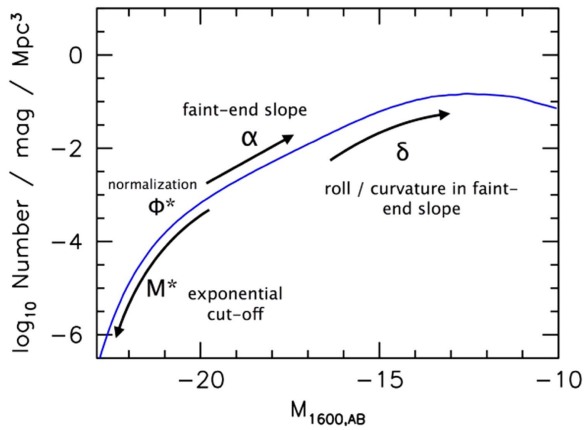
Figure 7. Illustration of the parameterization we utilize for the UV LF while assessing the possibility that it may turn over at low luminosities (Section 4.3). In addition to the standard Schechter parameters, M^*, ϕ^*, and α, we also allow for curvature in the effective slope of the LF using a fourth parameter δ—which can be used either to represent a roll-over or to indicate a possible steepening in the slope toward lower luminosities. We include such a curvature at >-16 mag, coincident with the luminosity range where magnification uncertainties become larger for individual sources. -16 mag also corresponds to that expected for a typically faint source ($\sim$28.5 mag) magnified by a factor of 10.

4.3. Parameterization of the LF Model

Use of a parametric form to the LF is particularly useful for examining the overall LF constraints on the shape of the UV LF in lensed fields, due to uncertainties that exist on the magnification factors, and hence luminosities, of individual sources used in the construction of the LF. This makes it difficult to place sources in specific bins of the UV LF (resulting in each bin showing a larger error).

For the parametric modeling we do of the LF, we start with a general Schechter form for the LF:

$$\phi^*(\ln(10)/2.5)\,10^{-0.4(M-M^*)(\alpha+1)}e^{-10^{-0.4(M-M^*)}}.$$

However, we modify the basic form of the Schechter function by multiplying the general Schechter form by the following expression faintward of -16 mag:

$$10^{-0.4\delta(M+16)^2}.$$

Positive values of δ result in the LF turning over faintward of -16 mag, while negative values of δ result in the LF becoming steeper faintward of -16 mag. An illustration of this parameterization is provided in Figure 7.

Our use of -16 mag allows us to test for possible curvature in the shape of the LF at >-16 mag, as predicted by some models (e.g., Jaacks et al. 2013; Kuhlen et al. 2013; Ocvirk et al. 2016). -16 mag is also just faintward of luminosities probed in field studies (i.e., -16.77 mag: Bouwens et al. 2015a). Finally, since >-16 mag corresponds to the luminosity of $\mu > 10$ faint sources in the HFFs, our fitting for a curvature parameter δ allows us to investigate how well the shape of the LF can be recovered in the regime where the magnification factors are large.

With this parameterization, the turn-over luminosity M_T (i.e., where $(d\phi/dM)_{M=M_T} = 0$) can be easily shown to be

$$M_T = -16 - \frac{\alpha+1}{2\delta} \qquad (2)$$

assuming that $\delta > 0$. For sufficiently small values for δ, i.e., $\delta \sim 0.05$, a turn-over in the LF would be so faint as to be impractical to confirm, and below what would be expected theoretically (Section 6.3). For typical models (Section 6.3), δ is expected to be $\delta \gtrsim 0.08$, resulting in turnover magnitudes <-10.

5. LF Results at $z \sim 6$

We now make use of the formalism we presented above and our selection of 160 $z \sim 6$ galaxies behind the first four HFF clusters by R. J. Bouwens et al. (2017, in preparation) to set constraints on the form of the UV LF at extremely faint magnitudes.

5.1. Using the HFF Observations Alone

We begin by looking at the constraints we can set on the shape of the $z \sim 6$ LF by restricting our analysis to $z \sim 6$ samples found behind the first four HFF clusters. Such an exercise is useful since it allows us to examine the LF constraints we obtain from our HFF search results, entirely independent of results in the field.

In deriving best-fit LF results, we use the following approach. We fix M^* to the same value Bouwens (2016) found at $z \sim 6$, i.e., -20.94 mag, set the curvature δ to be zero, and then fit for ϕ^* and α. We use a Markov Chain Monte Carlo (MCMC) algorithm where we start with the field LF results from Bouwens et al. (2015a), i.e., $\phi^* = 0.0005\,\mathrm{Mpc}^{-3}$ and $\alpha = -1.87$, to determine how the likelihood of various parameter combinations varies as a function of ϕ^* and α. For each set of parameters α and ϕ^*, we repeat the simulations described in Section 4.2 to calculate the likelihood of those parameters. For those simulation, we consistently use magnification maps from one team to create the mock observations and then recover the results using the median magnification maps formed from the parametric models. On the basis of the grid of likelihoods we derive, we determine the most likely values for ϕ^* and α, while also determining the covariance matrix that best fits the same likelihood grid. From the covariance matrix, we estimate errors on ϕ^* and α.

To determine the impact that errors in the magnification maps can have on the derived values for ϕ^* and α, we repeat the exercise from the above paragraph seven times. In each case, we treat the magnification maps from a different team as the truth and proceed to derive constraints on ϕ^* and α using the results from the first four HFF clusters. Because two of the HFF clusters we examine do not have Zitrin-NFW magnification maps available, i.e., MACS 0717 and MACS 1149, we make use of the Zitrin-LTM-Gauss models instead. The results are presented in Table 2.

The faint-end slope α results we obtain from the HFF clusters alone inhabit the range from -1.89 to -1.98 depending on which magnification model we treat as reality. The faint-end slope we estimate averaging the results from all of the models is -1.92 ± 0.04, while the faint-end slope α we find using the parametric models is -1.91 ± 0.03. The quoted uncertainty includes median statistical error added in quadrature with the standard deviation among the faint-end slope determinations for the different magnification models.

These results are interesting in that they are consistent with our own results over the field, i.e., Bouwens et al. (2015a), where $\alpha = -1.87 \pm 0.10$, as well as other estimates in the

Table 2
Best-fit Constraints on the $z \sim 6$ UV LF

Input Model	M_{UV}^{*} [a]	ϕ^{*} (10^{-3} Mpc^{-3})	α [b]	δ [c]	M_T [d]
HFF Observations Alone + M^* from field LF results (Section 5.1)					
GLAFIC	−20.94	0.69 ± 0.04	−1.90 ± 0.03	0.00	⋯
CATS	−20.94	0.66 ± 0.04	−1.91 ± 0.02	0.00	⋯
GRALE	−20.94	0.68 ± 0.06	−1.98 ± 0.03	0.00	⋯
Bradac	−20.94	0.70 ± 0.05	−1.89 ± 0.03	0.00	⋯
Sharon/Johnson	−20.94	0.68 ± 0.04	−1.91 ± 0.02	0.00	⋯
Zitrin-NFW	−20.94	0.58 ± 0.01	−1.92 ± 0.03	0.00	⋯
Zitrin-LTM	−20.94	0.67 ± 0.05	−1.95 ± 0.02	0.00	⋯
Mean	−20.94	0.66 ± 0.06	−1.92 ± 0.04	0.00	⋯
Mean Parametric	−20.94	0.65 ± 0.06	−1.91 ± 0.03	0.00	⋯
Fiducial (Section 5.2): HFF Observations + CANDELS/HUDF/HUDF-Parallel (Bouwens et al. 2015)					
GLAFIC	−20.94	0.57 ± 0.05	−1.92 ± 0.04	0.07 ± 0.16	>−14.2
CATS	−20.94	0.58 ± 0.05	−1.91 ± 0.04	0.17 ± 0.20	>−14.9
GRALE	−20.94	0.63 ± 0.07	−1.95 ± 0.03	0.16 ± 0.30	>−15.2
Bradac	−20.94	0.57 ± 0.05	−1.92 ± 0.04	0.21 ± 0.32	>−15.3
Sharon/Johnson	−20.94	0.57 ± 0.05	−1.92 ± 0.03	0.12 ± 0.21	>−14.9
Zitrin-NFW	−20.94	0.56 ± 0.06	−1.91 ± 0.03	0.07 ± 0.20	>−14.6
Zitrin-LTM	−20.94	0.58 ± 0.05	−1.93 ± 0.03	0.14 ± 0.25	>−15.1
Mean	−20.94	0.58 ± 0.06	−1.92 ± 0.04	0.14 ± 0.24	⋯
Mean Parametric	−20.94	0.57 ± 0.05	−1.92 ± 0.04	0.11 ± 0.20	⋯
Idem (but Estimating Completeness from Conventional Size–Luminosity Relations: Section 5.4)					
GLAFIC	−20.94	0.54 ± 0.06	−1.93 ± 0.04	−0.08 ± 0.18	⋯
CATS	−20.94	0.54 ± 0.06	−1.93 ± 0.04	−0.08 ± 0.22	⋯
GRALE	−20.94	0.58 ± 0.06	−1.94 ± 0.03	−0.27 ± 0.27	⋯
Bradac	−20.94	0.52 ± 0.06	−1.91 ± 0.04	−0.25 ± 0.28	⋯
Sharon/Johnson	−20.94	0.54 ± 0.06	−1.93 ± 0.04	−0.03 ± 0.22	⋯
Zitrin-NFW	−20.94	0.53 ± 0.06	−1.91 ± 0.04	−0.21 ± 0.23	⋯
Zitrin-LTM	−20.94	0.55 ± 0.05	−1.93 ± 0.04	−0.25 ± 0.27	⋯
Mean	−20.94	0.54 ± 0.06	−1.93 ± 0.04	−0.17 ± 0.26	⋯
Mean Parametric	−20.94	0.54 ± 0.06	−1.92 ± 0.04	−0.10 ± 0.22	⋯
Literature Including HFF Observations					
Atek et al. (2015b)	−20.9 ± 0.7	$0.28^{+0.59}_{-0.18}$	$-2.04^{+0.17}_{-0.13}$ [f]		
Livermore et al. (2017)	$-20.82^{+0.04}_{-0.05}$	$0.23^{+0.02}_{-0.02}$	$-2.10^{+0.08}_{-0.03}$	⋯	$>-11.1^{+0.4}_{-0.8}$ [e]
Literature Before HFF Observations					
Bouwens et al. (2015a)	−20.94 ± 0.20	$0.50^{+0.22}_{-0.16}$	−1.87 ± 0.10		
Finkelstein et al. (2015)	$-21.13^{+0.25}_{-0.31}$	$0.19^{+0.09}_{-0.08}$	−2.02 ± 0.10		

Notes.
[a] Fixed
[b] The faint-end slopes we derive are moderately dependent on the assumptions we make about the intrinsic size distribution of very low luminosity galaxies. Nevertheless, motivated by the results from a companion paper (Bouwens et al. 2017), where extremely faint $z \sim 5$–8 galaxies were found to have a size distribution consistent with point sources, we used this assumption in deriving results for the faint end of the UV LF. However, since obtaining direct constraints on the size distribution and hence completeness of extremely faint galaxies over the HFF clusters is difficult, we could underestimate the volume density of faint sources. This could result in faint end slopes that are steeper by $\Delta\alpha \sim 0.08$ (if we adopt 20% larger sizes than the Shibuya et al. (2015) size–luminosity relation instead of assuming galaxies to be point sources).
[c] Best-fit curvature in the shape of the UV LF faintward of −16 mag (see Figure 7). For HFF-only determinations, the curvature is fixed to 0 for simplicity.
[d] Brightest luminosity at which the current constraints from the HFF permit a turn-over in the $z \sim 6$ LF (within the 68% confidence intervals).
[e] This is the luminosity where according to Figure 12 of L17, L17 find $\Delta(\mathrm{BIC}) = 2$, where BIC denotes the Bayesian Information Criteria (analogous to $\Delta\chi^2$ for their usage). Strictly speaking, it is closer to an 84% confidence limit than a 68% confidence limit.
[f] LF constraints obtained for a combined $z \sim 6$–7 sample.

literature (Yan & Windhorst 2004; Bouwens et al. 2007; Calvi et al. 2013; Bowler et al. 2015; Finkelstein et al. 2015) which generally lie in the range from approximately −1.8 to approximately −2.0.

It is worthwhile to emphasize the value of the test we perform in the previous paragraph. Because we consider the use of searches behind lensing clusters for constraining the faint end of the UV LFs, it is essential that we derive constraints from the lensing clusters in isolation of those obtained from field searches to verify that no major systematics appear to be present in the LF results from the lensing clusters. This is relevant, since recent determinations of the faint-end slope α to the LFs from field and cluster search results seem to show a substantial discrepancy (Figure 1).

5.2. Using Results from the HFF Clusters and the Field

We now proceed to derive constraints on the overall form of the UV LF at $z \sim 6$ combining constraints from the field with those available from the HFF clusters. For simplicity, we keep the characteristic luminosity M^* fixed to the value -20.94 mag that we found in our earlier wide-area field study (Bouwens et al. 2015a), given the lack of substantial information in the HFF cluster program for constraining this parameter due to the small volume probed.

In combining constraints from the field and from the HFF clusters, we need to allow for some error in the normalization of both the field and HFF cluster results, as a result of large-scale structure variations ("cosmic variance": Robertson et al. 2014; see also Somerville et al. 2004 and Trenti & Stiavelli 2008) and also small systematic errors in the estimates of the volume densities of galaxies in each of our probes. We assume an uncertainty of $\sim 20\%$ in the volume density in both the field and HFF LF results.

The 20% uncertainty we assume to be present in the normalization of the LF results at both the bright and faint ends includes an $\sim 10\%$ uncertainty in our estimates of the selection volume and $\sim 10\%$ systematic error due to uncertainties in the total magnitude measurements (reflecting a ~ 0.1 mag systematic error in the magnitude measurements). The inclusion of such an error is relevant given the existence of real errors in the estimated volume densities of galaxies using photometric criteria. Small differences appear to be guaranteed, given that the filters available for the selection of galaxies from the field are different (in particular, including a deep "z"-band filter) from those available over the HFF clusters (which do not include a deep "z" band filter). Important factors contributing to these uncertainties are (1) likely differences between the assumed sizes and SEDs of galaxies versus redshift in the observations versus those in the simulations and (2) uncertainties in the contamination rate of observed samples.[14]

Similar to our LF results using the HFF observations alone, we derive confidence regions on the $z \sim 6$ LF results using an MCMC-type procedure where we explore a limited region in the ϕ^*-α-δ parameter space and calculate the likelihood of each point in parameter space using the forward modeling simulations we describe in Section 4.1. Our calculated likelihoods explicitly include a marginalization across the 20% volume density uncertainties we assume. These likelihoods are then multiplied by the likelihoods on the same Schechter parameters derived by Bouwens et al. (2015a) at $z \sim 6$ using results from the full CANDELS program, the HUDF, and the HUDF parallels (again after marginalizing the Bouwens et al. (2015a) over ϕ^* to account for a 20% uncertainty in the volume density of sources). Finally, this 3D likelihood grid is fit to derive the most likely values for ϕ^*, α, and δ and also the covariance matrix.

As in our determinations of the LF parameters from the HFF programs alone, we repeat this exercise seven different times, treating each of the magnification models

from different teams as reality and recovering the LF results using the median magnification map. We present our constraints on each of the LF parameters in Table 2. The faint-end slope α we estimate averaging all of our models and just the parametric models are -1.92 ± 0.04 and -1.92 ± 0.04, respectively.

As in our determinations using only the HFF observations themselves, the faint-end slopes α we derive at $z \sim 6$ are fairly consistent with LF results in the field. Our obtaining consistent results for all seven of the magnification model families we consider is not especially surprising, given the fact that individual sources would be expected to scatter in almost the same direction as the dominant slope of the LF, i.e., approximately -2 (the expected slope of the LF from scatter) versus ~ -1.9 (the actual slope of the LF).

Despite general agreement on the most likely value for α at $z \sim 6$, we find a broad range of values for the curvature parameter δ, from 0.07 to 0.21, with 1σ uncertainties ranging from 0.16 to 0.32. None of the magnification models we considered point toward our having even modest evidence, i.e., $\delta > 0$ at $>2\sigma$, for the $z \sim 6$ LF showing a turn-over at the faint end.

We find the smallest uncertainties on δ assuming that the GLAFIC magnification models represent reality. Slightly larger uncertainties on δ are found assuming that the CATS, Sharon/Johnson, and Zitrin-NFW models represent reality, while the largest uncertainties on δ are found, assuming that the Zitrin-LTM, Bradac, and GRALE models represent reality. The size of the uncertainty on δ is a function of how similar the various magnification models are to the median magnification model formed from the parametric models. Given the similarity of assumptions utilized in the GLAFIC, CATS, Sharon/Johnson, and Zitrin-NFW models, it is not surprising that their magnification maps agree best with median magnification maps constructed using similar assumptions.

The results in Table 2 for the different magnification models indicate the general range of constraints one could obtain on the form of the $z \sim 6$ LF: the results for the non-parametric models indicate the errorbars on the LFs if the mass distribution in clusters do not strictly follow the assumptions made in the parametric models, while results for the parametric models indicate the likely error bars, if the mass profiles in the HFF clusters generally do adhere to those assumptions.

Uncertainties in the redshifts of the lensed $z \sim 6$ galaxies also contribute to the error in δ. To estimate the impact, we kept the redshifts of lensed background souces fixed while rerunning the forward-modeling simulations from our MCMC chain. Comparing the uncertainties we derive in δ to the uncertainties we derive including errors in the redshift, we find a typical increase of 0.01 in the uncertainty on δ, i.e., from 0.15 to 0.16 in the case of the GLAFIC simulations. If we assume that uncertainties in the deflection maps and redshifts both add in quadrature, this suggests that errors in the photometric redshift errors contribute $\sim 12\%$ of the fractional error in δ, i.e., $(0.16^2 - 0.15^2)/0.16^2 \sim 0.12$.

To help visualize what our present LF results mean, we have made use of our parametrized constraints to derive 68% and 95% confidence intervals on the volume density of galaxies as a function of the UV luminosity M_{UV}. These results are presented both in Figure 8 and also in Table 3.

---

[14] Measurements of the total magnitudes typically differ at the ~ 0.1 mag level, as is evident looking at the broad range of magnitude measurements in Skelton et al. (2014). See, e.g., their Figures 35–36. The situation is likely even more challenging for galaxies behind lensing clusters due to the substantial foreground light from the clusters themselves, and in fact the total magnitudes measured with different procedures and by different groups are found to differ at the ~ 0.2–0.25 mag level. See Section 6.1.2. This translates into normalization differences of $\sim 20\%$ to $\sim 25\%$ assuming a faint-end slope of approximately -2 for LF results.
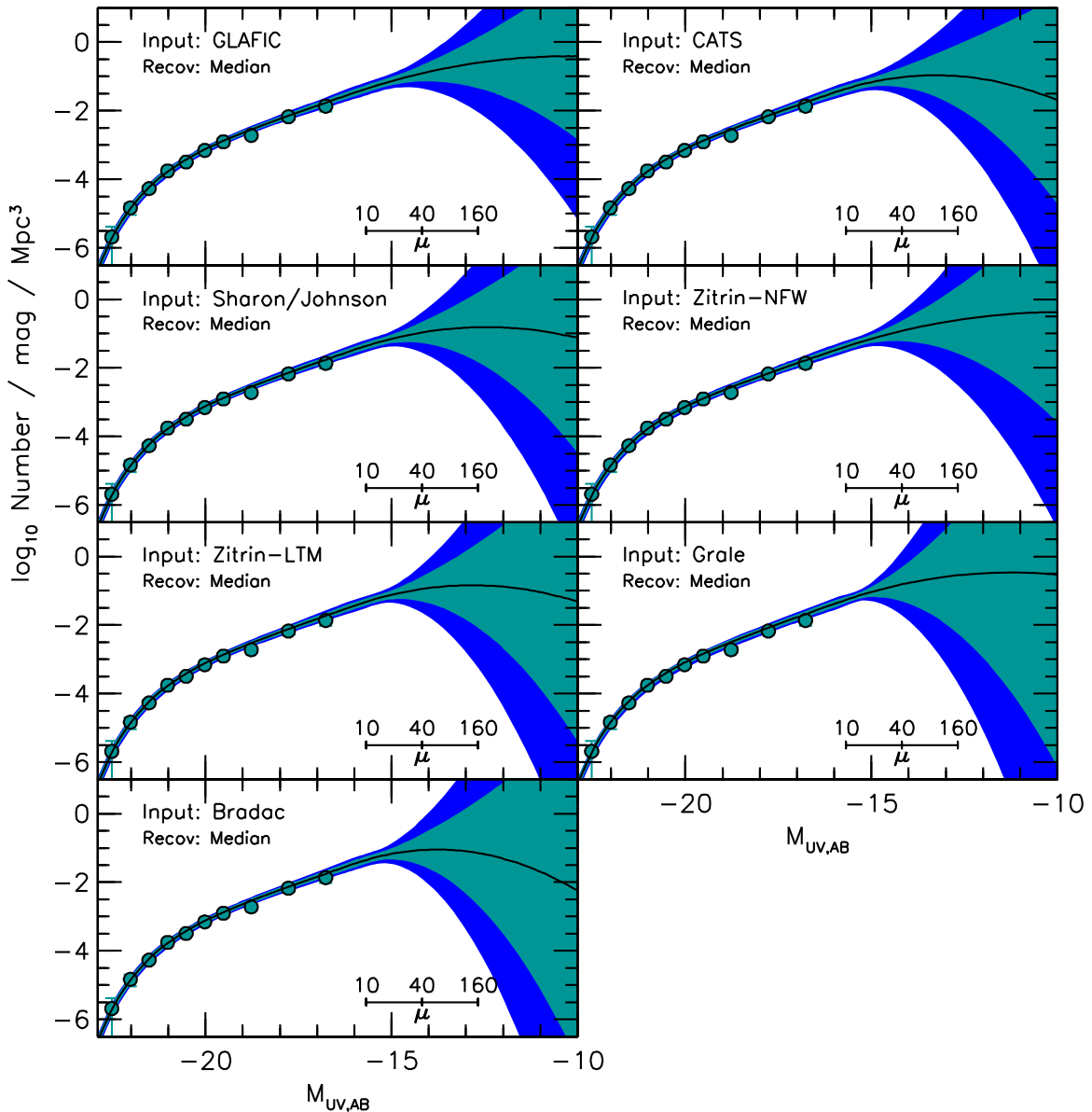
Figure 8. Determination of the 68% and 95% confidence intervals (shaded in cyan and blue, respectively, see Section 5.2) on the overall shape of the $z \sim 6$ LF. The LF here combines constraints from the Bouwens et al. (2015a) $z \sim 6$ study with the HFF observations. We then alternatively assume that the GLAFIC, CATS, Sharon/Johnson, Zitrin-NFW, Zitrin-LTM, Bradac, and GRALE lensing models represent reality and recover the LF using the median of the GLAFIC, CATS, Sharon/Johnson, and Zitrin-NFW models (when available and excluding models from the median when treated as reality to make the assessment fair). The cyan solid circles show the binned $z \sim 6$ LF from the HUDF, HUDF-parallel fields, and CANDELS (Bouwens et al. 2015a). The black line indicates the nominal best-fit LF. The ticked line showing the magnification factors is the same as in Figure 4. The large range of allowed LFs (shaded regions) is a direct result of the impact of magnification uncertainties as illustrated in Figure 4. The plotted confidence intervals are tabulated assuming the GLAFIC and Bradac model as inputs in Table 3. If differences between this median magnification map and the non-parametric magnification maps are representative of the actual uncertainties, the present results suggest that we cannot rule out a turn-over in the LF at ~ -15 mag. Even taking as the alternate case the assumption that the GLAFIC magnification models represent reality, the present results suggest that a turn-over in the LF is permitted at ~ -14.2 mag within the 68% confidence intervals.

5.3. Constraints on a Possible Turn-over in the $z \sim 6$ LF

One question that has recently been of significant interest in the literature regards whether there is a flattening or turn-over in the UV LF at the faint end. This question is important because the answer could indicate to us whether there is a sufficient number of extremely faint galaxies to produce the photons necessary to drive the reionization of the universe.

Fortunately, using the likelihood contours for δ-α-ϕ^* and Equation (2), we can directly determine the brightest point in the LF where a turn-over is permitted (with the 68% confidence intervals). The results do depend somewhat on which magnification model we assume to be representative of reality

(and therefore which of the panels we consider from Figure 8). Nevertheless, we find that the HFF observations allow for a turn-over in the LF as bright as -14.2 mag to -15.3 mag (within the 68% confidence intervals). The allowed turn-over luminosities we estimate assuming different magnification models are presented in Table 2 and also in Figure 9.

5.4. LF Results: The Impact of Galaxy Size

In a companion paper (Bouwens et al. 2017), we showed that the slope of the LF at low luminosities is strongly dependent on the size of very faint galaxies (see Figure 2 from that paper). We constrained the sizes of faint galaxies (> -16 mag) by

Table 3
68% and 95% Confidence Intervals on the UV LF at $z \sim 6$ (Section 5.2)
Adopting the Functional Form in Section 4.3

| | $\phi(M)$ [$\log_{10}$(#/Mpc^3/mag)] | | | |
| | Lower Bound | | Upper Bound | |
$M_{UV,AB}$	95%	68%	68%[c]	95%[c]
	Case 1 (GLAFIC)[a]			
−16.75	−1.88	−1.82	−1.69	−1.63
−16.25	−1.71	−1.64	−1.50	−1.42
−15.75	−1.54	−1.47	−1.30	−1.23
−15.25	−1.38	−1.30	−1.13	−1.05
−14.75	−1.29	−1.18	−0.95	−0.84
−14.25	−1.31	−1.12	−0.73	−0.54
−13.75	−1.43	−1.13	−0.49	−0.18
−13.25	−1.65	−1.18	−0.22	0.25
−12.75	−1.95	−1.29	0.07	0.73
−12.25	−2.34	−1.45	0.37	1.26
−11.75	−2.82	−1.66	0.70	1.84
−11.25	−3.37	−1.93	1.04	2.48
−10.75	−4.01	−2.24	1.40	3.17
−10.25	−4.74	−2.60	1.79	3.91
	Case 2 (Bradac)[b]			
−16.75	−1.85	−1.79	−1.66	−1.60
−16.25	−1.67	−1.60	−1.46	−1.39
−15.75	−1.50	−1.42	−1.27	−1.20
−15.25	−1.34	−1.27	−1.11	−1.04
−14.75	−1.36	−1.21	−0.90	−0.75
−14.25	−1.57	−1.27	−0.65	−0.35
−13.75	−1.93	−1.42	−0.36	0.16
−13.25	−2.43	−1.66	−0.04	0.75
−12.75	−3.08	−1.99	0.31	1.42
−12.25	−3.88	−2.40	0.68	2.18
−11.75	−4.81	−2.90	1.09	3.02
−11.25	−5.88	−3.48	1.52	3.95
−10.75	−7.10	−4.17	1.98	4.95
−10.25	−8.46	−4.93	2.47	6.04

Notes.
[a] For case 1, we assume that the differences between the magnifications in the median parametric model and the GLAFIC model are a good representation of the typical errors in the magnification models we utilize.
[b] For case 2, we assume that differences between the magnifications in the median parametric model and the Bradac model are a good representation of the typical errors in the magnification models we utilize. Similar confidence regions are obtained if one uses the GRALE model instead of the Bradac model.
[c] If 50% of faint sources at $z \sim 6$–8 are significantly spatially extended (intrinsic half-light radii >30 mas), the 68%-likelihood upper bounds on the implied LF constraints could increase by $\sim$0.3 dex (Bouwens 2017). The actual upper bound on the volume density could be much higher if the completeness is substantially less than 50%.

taking advantage of the large samples of such galaxies available behind the first four HFF clusters at $z \sim 2$–3 and $z \sim 5$–8. We found no evidence to indicate that these galaxies were significantly resolved, looking at (1) the prevalence of high magnification sources as a function of the predicted shear and (2) their stacked profile along the expected shear axis. The slope varied dramatically from $\alpha \sim -1.9$ for the small sizes we found (taking 7.5 mas to be representative of the half-light radius) to $\alpha \sim -2.65$ (taking 120 mas as representative). The resulting luminosity density from integrating the LF to -14 changed by a factor of 40. Clearly the size assumed for faint galaxies is a key parameter that is central to a reliable determination of the LF.
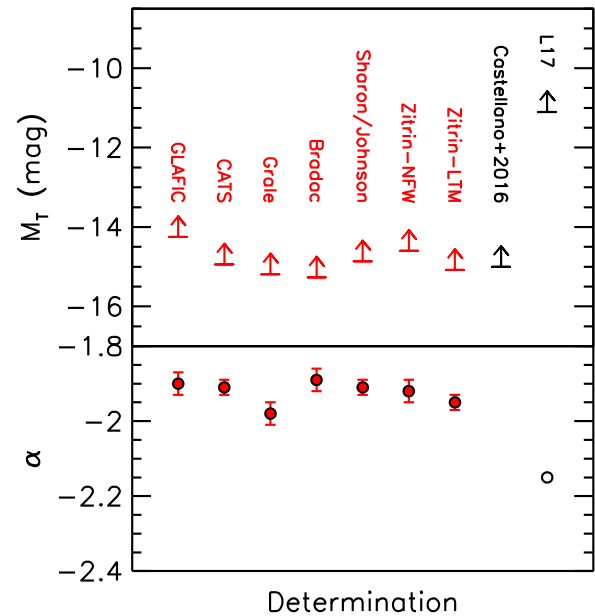
Figure 9. (Upper panel) Brightest luminosity allowed for a potential turn-over in the $z \sim 6$ UV LF (within our 68% confidence intervals), using faint $z \sim 6$ galaxies identified behind the first four HFF clusters and assuming different magnification models represent reality. The reported constraints from Castellano et al. (2016b) and L17 are also presented, with the L17 constraints plotted at a Δ(BIC) value of two. The substantially fainter allowed turn-over luminosity reported by L17 (significantly discrepant with the other estimates) is likely an artifact of the very large sizes that L17 assume (see Figure 11, Sections 6.1.2, 6.2, and also Bouwens et al. 2017) and a large number of sources close to the detection limit of their selection (Figure 12). (Lower panel) Faint-end slope α determinations using only the HFF cluster search results. See Table 2 for a tabulation of these results.

On the basis of our findings in a companion study, can we assume that >-16 mag galaxies are all entirely unresolved? Unfortunately, we cannot due to the impact of surface brightness selection effects on the recovered samples. The impact is sufficient that we cannot rule out (87% confidence) sources having half-light radii of 30 mas, which is approximately the size we would predict for faint sources extrapolating conventional size–luminosity relations.

Given this fact, it is therefore quite logical to repeat the present determination of the LF, but this time assuming a conventional size–luminosity relation. We will then treat the results as providing an upper bound on the $z \sim 6$ LF results, given uncertainties in the size distribution. To this end, we suppose that the median half-light radii of galaxies follow the following correlation with luminosity $r_{hl} = (0''.14)(L/L^{*}_{z=3})^{0.27}$, which is in good agreement with the size–luminosity relation found by Shibuya et al. (2015). In addition, we adopt the sizes of galaxies that exhibit a log-normal distribution with 0.3 dex 1σ scatter. We assume galaxies to have a Sérsic profile and for the Sérsic indices to have a log-normal distribution with a median of 1.5 and scatter of 0.3 dex. The axial ratio is also assumed to be log-normal with a median value of 1.8 and a scatter of 0.3 dex. A random position angle is adopted for sources. Finally, the pixel-by-pixel profiles for all sources in the Monte-Carlo catalogs are computed. The impact of gravitational lensing is included using the latest deflection maps from the CATS team.

The simulated galaxies are then added to the real observations, and we utilize the same procedure for catalog creation and

source selection as we use on the real data. We then rederive the selection volumes in the same way as before (i.e., see Appendix B) and repeat the determination of the LF results using the outlined forward-modeling procedure. Compared to the situation where point-source sizes are assumed, the selection volumes we derive are lower, increasing the overall volume density of sources inferred at lower luminosities >-18 mag.

The approximate impact of this use of larger sizes for faint sources is to increase the volume density of sources at approximately -17, -15, and -14 by factors of ~ 1.6, ~ 2, and ~ 3.3, respectively. The amplitude of the correction increases toward lower luminosities due to the correlation between surface brightness and luminosity implied by conventional size–luminosity relations (where $R \propto L^{0.27}$: e.g., Shibuya et al. 2015), i.e., $L/R^2 \propto L/(L^{0.27})^2 \propto L^{0.46}$.[15]

Based on this scaling, one would expect $0.01\,L^*$ and $0.001\,L^*$ galaxies to have $\sim 8\times$ and $\sim 24\times$ lower surface brightnesses, respectively, than more luminous L^* galaxies. Since gravitational lensing preserves surface brightness, it should not be easy to select extremely low luminosity galaxies behind lensing clusters, if conventional relations held. We should emphasize, however, that it is not clear, however, that conventional relations hold down to such low luminosities, i.e., $M_{\rm UV} > -16$. The light profile in galaxies may be dominated by just a single super star cluster or two in this regime, as suggested by our results in Bouwens et al. (2017).

The derived LF results we derive for the stated size assumptions are presented in Table 2. The results are similar to our fiducial results. However, they do nevertheless give faint-end slopes α that are $\Delta\alpha \sim 0.01$ steeper and curvature parameters δ that are approximately 0.2 lower. With the present size assumptions, δ inferred for the $z \sim 6$ LF is formally negative for all of the magnification models we consider. As in Section 5.3, we emphasize that a possible upturn in the LF (i.e., $\delta < 0$) is not a statistically robust result. If we force δ to be 0, the faint-end slope we derive is $\Delta\alpha \sim 0.03$ steeper for the typical lensing model. If we allow for such a change in α, the present tension in the faint-end slope α versus L17 would decrease to 3σ.

The exercise in this section demonstrates the sensitivity of the curvature parameter in the LF δ—and in fact the whole question as to where or if the UV LF turns over—to the form of the size–luminosity relation. We caution that the conclusions here are based on an extrapolation of sizes seen at significantly higher luminosities and that the indications from our recent work on sizes in the HFF clusters (Bouwens et al. 2017) suggest that galaxies at luminosities >-16 may be very small. In such a case, the completeness corrections will be much smaller.

5.5. Implied Constraints on the $z \sim 6$ UV Luminosity Density

To determine if galaxies produce enough ionizing photons to drive the reionization of the universe, we require constraints on the luminosity density in the rest-frame UV that include the contribution from all galaxies.

We can compute confidence intervals on the luminosity densities in the rest-frame UV by using the constraints from the analyses

performed in the previous section and then marginalizing over ϕ^*, α, and δ. We compute the results to a number of different limiting luminosities $M_{\rm UV}$, i.e., -17, -15, -13, -10, and -3 mag,[16] the first four of which commonly appear in the literature, in considering whether galaxies might drive cosmic reionization, particularly including the contribution at very low luminosities. We have presented the results we obtain in Figure 10 and also in Table 4 to these faint-end limits. Our results imply a luminosity density of $10^{26.38\pm0.05}\,{\rm erg\,s^{-1}\,Hz^{-1}\,Mpc^{-3}}$ to -15 mag.

Not surprisingly, our LF results allow for essentially an arbitrarily high contribution from very faint galaxies to the UV luminosity density, particularly including a contribution from galaxies to -3 mag. These results also provide fairly firm 1σ and 2σ lower bounds on the luminosity density. We find 1σ and 2σ lower limits on luminosity density in the rest-frame UV of $\sim 10^{26.40}\,{\rm erg\,s^{-1}\,Hz^{-1}\,Mpc^{-3}}$ and $\sim 10^{26.35}\,{\rm erg\,s^{-1}\,Hz^{-1}\,Mpc^{-3}}$, respectively.

The 1σ and 2σ lower bounds we find on the luminosity density integrating to arbitrarily faint luminosities is not especially higher than what we find integrating to -15 mag. These results indicate that it is not yet possible to argue for the discovery of a significant additional reservoir of photons from galaxies faintward of -15 mag (Atek et al. 2015a, 2015b), as has been the claim in one recent study (L17).

5.6. Binned Determinations of $z \sim 6$ LF Using Direct Method

Finally, to conclude this section, we derive a binned representation of the results from the first four HFF clusters. Binned representations of the LF results from the HFFs should be very reliable at high luminosities, where errors in the magnification maps are smaller. Binned representations retain the advantage that they are much more model independent probes of the LF shape as a function of luminosity.

In our binned representation of the LF, we adopt bins of width 0.5 mag and determine the binned LF results ϕ_m to be as follows.

$$\phi_m = \frac{N_m}{V_m}, \qquad (3)$$

where N_m is the number of sources in absolute magnitude bin m after correcting for the estimated magnification.

We derive the selection volumes V_m in a given magnitude bin using the following equation.

$$V_m = \int_A \int_{dz} C(z, m, \mu) \frac{1}{\mu(A)} \frac{dV(z)}{dA} dz\, dA, \qquad (4)$$

where A denotes the area, V denotes the volume, C denotes the estimated completeness, and μ denotes the magnification factor. Our estimates of the completeness are provided in Appendix B; the completeness C does not appear to be a strong function of the magnification factor μ in data sets as deep as the HFFs, if we take the results of Bouwens et al. (2017) to be indicative.

We derived simple stepwise constraints on the UV LF brightward of -16 mag, by dividing the number of sources in each absolute magnitude bin by the computed selection volume. The results are presented in Table 5 and Figure 11.

---

[15] In Bouwens et al. (2017), we presented evidence that the scaling may be steeper than this, i.e., $R \propto L^{0.5\pm0.07}$, at lower luminosities based on the sizes and luminosities of $z \sim 6$ sources in the HFFs. However, we caution that correcting for completeness successfully in the HFF data is sufficiently challenging that the true relation could be shallower than what we found.

[16] The faintest limit here, -3 mag, is only included for illustrative value and takes as its inspiration results from O'Shea et al. (2015) and Ocvirk et al. (2016), which predict sources to such faint magnitudes.
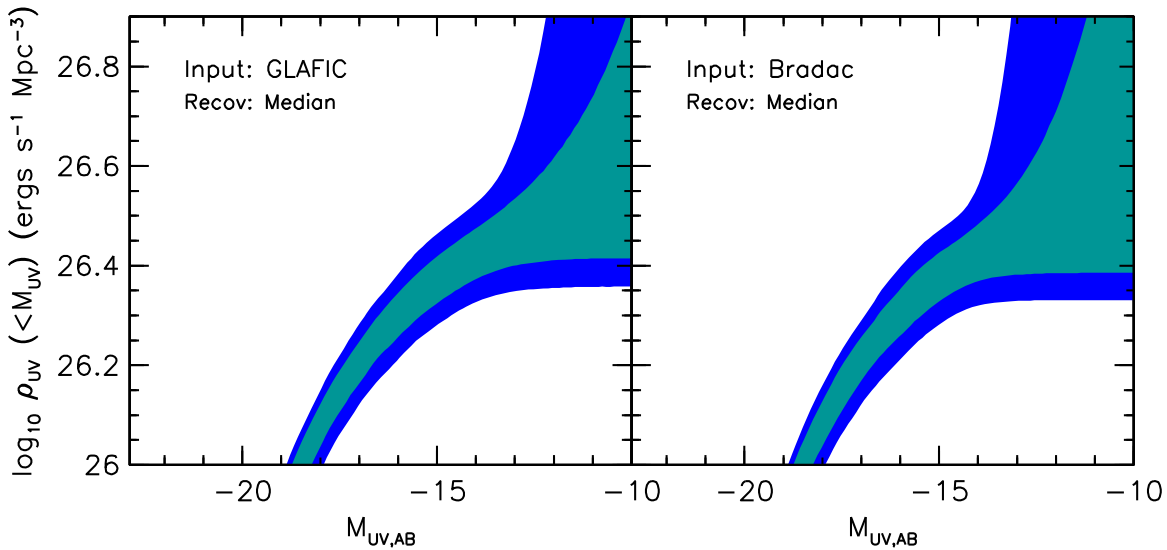
Figure 10. 68% and 95% confidence intervals (shaded in cyan and blue, respectively) on the estimated UV luminosity density for all sources brighter than some M_{UV} (Section 5.5). Shown are the results for two different assumptions about the size of the magnification errors in the lensing models. As we integrate further down the UV LF to derive the luminosity densities, the lower boundaries allowed by our analysis show a clear monotonic increase. However, faintward of approximately -15 mag, the lower boundaries cease to show an increase that is highly significant (i.e., $>1.5\sigma$). This demonstrates that it is not yet possible to make strong claims that >-15 mag galaxies provide an additional reservoir of photons to reionize the universe.

Table 4
68% and 95% Confidence Intervals on the UV Luminosity Density at $z \sim 6$ to Various Limiting Luminosities (Section 5.5)

		$\log_{10}\rho_{UV}$ (UV Luminosity Density)		
		(erg s^{-1} Hz^{-1} Mpc^{-3})		
	Lower Bound		Upper Bound	
Faint-end Limit	95%	68%	68%	95%
	Case 1 (GLAFIC)[a]			
$M_{UV} < -17$	26.13	26.17	26.25	26.28
$M_{UV} < -15$	26.28	26.33	26.42	26.47
$M_{UV} < -13$	26.35	26.40	26.54	26.65
$M_{UV} < -10$	26.36	26.42	26.93	28.65
$M_{UV} < -3$[b]	26.36	26.42	31.10	41.23
	Case 2 (Bradac)[a]			
$M_{UV} < -17$	26.13	26.18	26.26	26.29
$M_{UV} < -15$	26.29	26.33	26.43	26.47
$M_{UV} < -13$	26.33	26.39	26.57	27.00
$M_{UV} < -10$	26.33	26.39	27.39	31.62
$M_{UV} < -3$[b]	26.33	26.39	34.35	55.95

Notes.
[a] Same assumptions as in Table 3.
[b] The -3 mag limit is included here for illustrative value and takes as its inspiration results from O'Shea et al. (2015) and Ocvirk et al. (2016), which predict sources to such faint magnitudes.

The upper limits we quote on the volume densities of sources include a possible $1.6\times$ underestimate, if the sizes of sources follow the size–luminosity relation presented in Section 5.4.

Faintward of -16 mag, we include results in Table 5 and Figure 11 by taking the geometric mean of the best-fit LF results using the GLAFIC, CATS, and Sharon/Johnson models. We present those results to luminosities plausibly probed by the present study. 1σ errors on the results at the faint end of the $z \sim 6$ LF are similarly taken to be the geometric mean of the confidence intervals on the LF fits for the same parametric magnification models. The upper limits we quote on

Table 5
Binned Determination of the Rest-frame UV LF at $z \sim 6$ (Section 5.6)[a]

| $M_{UV,AB}$[b]	ϕ_k (10^{-3} Mpc^{-3} mag^{-1})
-20.75	$0.0002^{+0.0002}_{-0.0002}$
-20.25	$0.0009^{+0.0004}_{-0.0004}$
-19.75	$0.0007^{+0.0004}_{-0.0004}$
-19.25	$0.0018^{+0.0006}_{-0.0006}$
-18.75	$0.0036^{+0.0009}_{-0.0009}$
-18.25	$0.0060^{+0.0012}_{-0.0012}$
-17.75	$0.0071^{+0.0066c}_{-0.0014}$
-17.25	$0.0111^{+0.0102c}_{-0.0022}$
-16.75	$0.0170^{+0.0165c}_{-0.0039}$
-16.25	$0.0142^{+0.0171c}_{-0.0054}$
-15.75	$0.0415^{+0.0354c}_{-0.0069}$
-15.25	$0.0599^{+0.0757d}_{-0.0106}$
-14.75	$0.0817^{+0.1902d}_{-0.0210}$
-14.25	$0.1052^{+0.5414d}_{-0.0434}$
-13.75	$0.1275^{+1.6479d}_{-0.0747}$
-13.25	$0.1464^{+5.4369d}_{-0.1077}$
-12.75	$0.1584^{+19.8047d}_{-0.1343}$

Notes.
[a] These LF results are simple estimates, representing the number of sources at a given luminosity (using the median magnification maps) after division by the selection volumes. No account is made for scatter resulting from errors in the magnification maps. Errors in the magnification maps can be best handled using the forward modeling simulations and methodology we consider in Section 4.2, leading to the results presented in Table 3.
[b] Upper limits are 1σ.
[c] The 1σ upper limits indicate the upper limits if one adopts the larger size for sources assumed in Section 5.4.
[d] 68% confidence intervals on the $z \sim 6$ UV LF at >-16 mag we achieve using forward modeling and observations of the first four HFF clusters in Section 4.3. The quoted constraints give the geometric mean of our results using the GLAFIC, CATS, and Sharon/Johnson parametric models as inputs. The 1σ upper limits indicate the upper limits if one adopts the larger size for sources assumed in Section 5.4 (resulting in ~ 0.01 and ~ 0.2–0.3 more negative values for α and δ). The error bars are not independent.
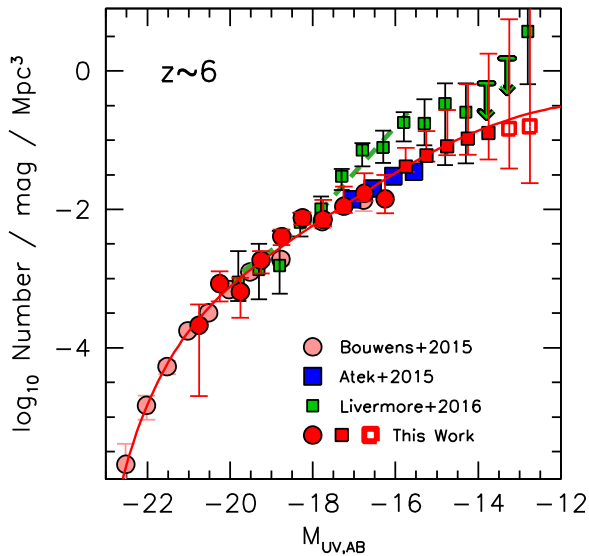
Figure 11. Comparison of the present stepwise UV LF from the HFF program at $z \sim 6$ (dark red circles: see Section 5.6) with previous determinations by Atek et al. 2015b: blue squares), L17 (green squares), and Bouwens et al. (2015a), which just come from using the HUDF, HUDF-parallel, and CANDELS fields (light red circles). All error bars and upper limits are 1σ. The dark red squares and open red squares give the results from our full forward-modeling procedure, as given in Section 4 (but where the error bars are not independent: see Section 5.6). We use open red squares at > -13.5 mag where there are no $z \sim 6$ sources in our selection to indicate a greater uncertainty. The upper 1σ error bars include uncertainties in the size distribution (Section 5.4). See Table 5 for a tabulation of the present constraints shown here. The red line shows our best-fit LF that we derive by doing a forward-modeling analysis using the GLAFIC magnification models as inputs. The luminosities of the individual points in the L17 and Atek et al. (2015) LFs have been corrected brightward by ~ 0.3 to ensure better consistency with the luminosities (and total magnitudes) measured in our own study (see Section 6.1.2). The dashed green line schematically indicates the upward break in the $z \sim 6$ LF results of L17 (i.e., an apparent steepening) that likely impacts their interpretation of the shape of the LF at lower luminosities, i.e., their claim that there is no turn-over in the $z \sim 6$ LF until ~ -12 mag (see Section 6.2).

the LF results include the impact of the larger sizes quoted in Section 5.4, which we find to result in a 0.01–0.03 steeper value of the faint-end slope α and to lower the inferred δ by ~ 0.2–0.3.

6. Discussion

The purpose of this paper is to present new constraints on the form of the $z \sim 6$ LF to low luminosities utilizing new constraints from the first four clusters available from the HFF program.

6.1. Comparison with Previous Observational Constraints

Before looking into comparisons of our new observational constraints with theory, it is important first to compare the present results with previous observational results where available to evaluate and understand any differences. Doing so helps clarify the gains that are made with our newer analysis and enables others to also understand what new factors led to the changes from prior work.

6.1.1. Atek et al. (2015b)

We first consider a comparison with the most recent results of Atek et al. (2015b) who make use of observations available

over the first three HFF clusters Abell 2744, MACS 0416, and MACS 0717 and selected galaxies using an I_{814}-dropout selection criteria, which would identify galaxies from $z \sim 6$ to $z \sim 7$.

A comparison with the most recent determination of the $z \sim 6$–7 LF from Atek et al. (2015b) is provided in Figure 11. In comparing against the Atek et al. (2015b) LF determinations, we incorporate a ~ 0.3 mag brightward shift of the Atek et al. (2015b) LF to correct for differences in our apparent magnitude measurements for individual sources. As already noted in one of the companion papers to this study (Bouwens et al. 2017), overall the agreement appears to be quite good, at least insofar as the stepwise points are concerned.

The best-fit ϕ^* and luminosity density that Atek et al. (2015b) estimate to -15 mag, i.e., $\sim 10^{26.20 \pm 0.13}$ erg s^{-1} Hz^{-1} Mpc^{-3}, is ~ 0.18 dex lower than what we find. This is a small but readily understandable difference that arises because Atek et al. (2015a) provide a constraint on the LF at a higher mean redshift than we do, i.e., $z \sim 6.5$ versus $z \sim 6$, and also include in their determinations results from field surveys, i.e., CANDELS or the HUDF, which probe $z \sim 7$ versus our $z \sim 6$ probe. Given that the integrated luminosity density to a limit of -17 mag changes by ~ 0.2 dex per unit redshift, our larger luminosity density estimate is entirely expected.

6.1.2. L17

As can be seen from Figure 11, the differences between our results and those from L17 are more substantial than those with Atek et al. (2015b) discussed above. These differences with L17 are particularly large for the LFs at the fainter magnitudes > -17, where much of the current interest lies since this is a region that is uniquely accessible using the HFF clusters. As a result, the discussion of the reasons for these differences is necessarily much more extensive than that for Atek et al. (2015b).

To ensure that comparisons with the LF results from L17 were made using a consistent luminosity scheme, we carefully cross-matched sources from our catalogs with those from L17. We also computed apparent magnitudes for individual sources using the tabulated absolute magnitudes, redshifts, and magnification factors that L17 provide in their Tables A1–A3. It is these derived apparent magnitudes that we compare to our own photometry and that of other groups.

Comparing the total magnitudes, we derive for sources using our scaled aperture scheme to the L17 apparent magnitudes, we find a 0.43 mag median difference, with the L17 apparent magnitudes being fainter than ours, both for relatively bright $H_{160,AB} < 28$ sources and also for the fainter $H_{160,AB} > 28$ sources. If we instead estimate total magnitudes for sources by taking the flux in fixed apertures that would enclose 70% of the flux for point sources, as performed by the HUDF12 team (McLure et al. 2013; Schenker et al. 2013) and derive an inverse variance-weighted total magnitude from the Y_{105}, J_{125}, JH_{140}, and H_{160} bands, we find differences of 0.2 mag in the median, with the L17-inferred magnitudes being fainter, comparing magnitudes for the faintest sources (i.e., > 28 mag). The L17 magnitudes show a similar offset relative to the published photometry of Atek et al. (2015b).

Given that the HUDF12 apparent magnitude measurement scheme should give a fairly conservative lower limit on the total fluxes for individual candidates, these comparisons suggest that L17 systematically underestimate the luminosity
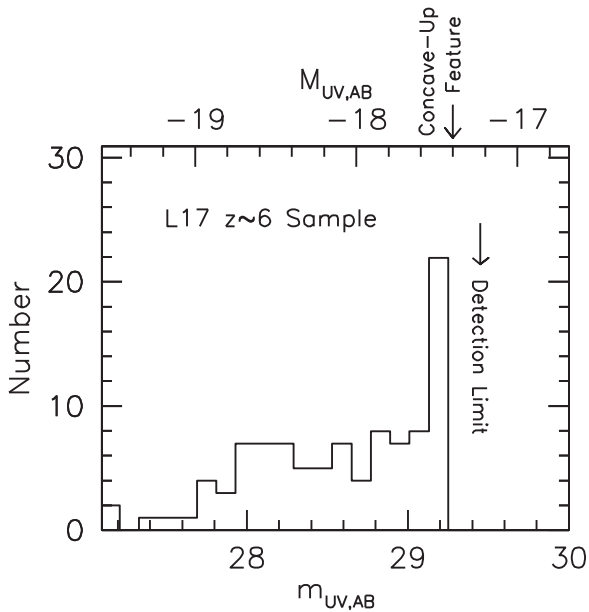
Figure 12. Number of sources per apparent magnitude bin ($\Delta m = 0.12$ mag) in the L17 $z \sim 6$ sample shown with respect to the approximate detection limit in the HFF data. The upper horizontal axis shows the equivalent absolute magnitude for a $z \sim 6$ source assuming a magnification of 1. Apparent magnitudes are derived from the absolute magnitudes, redshifts, and magnification factors given in Table A1 of L17. We note 22 sources in 29.25 mag bin just brightward of the detection limit vs. ~ 7 sources in the typical ~ 0.1 mag bin. This large pile-up of sources at the $z \sim 6$ magnitude limit is not apparent in Figure 9 of L17, since L17 set the upper limit on the vertical axis to 30—even though there were actually 45 sources in their faintest bin. Given that is also where their recovery fraction (and sample completeness) is approaching zero (Figure 4 of L17), one would expect their LF to show a substantial increase in the volume density of sources over the entire absolute magnitude range where sources in their ~ 29.25 mag bin impacts the LF, i.e., from -17.5 mag (where the magnification factor is ~ 1) to -12.5 mag (where the magnification factor is 100). This apparent upturn in the L17 LF is illustrated with the dashed green line in Figure 11. The impact of the 19 sources on the $z \sim 6$ LF would be exacerbated by L17's assuming larger sizes than are found in most observational studies (Kawamata et al. 2015; Laporte et al. 2016; Bouwens et al. 2017: see Section 6.1.2).

of individual sources in their catalog by at least ~ 0.2 mag, if not more (taking our scaled-aperture magnitudes as the baseline).

Given this range in values, we adopt a shift of the binned $z \sim 6$ LF of L17 brightward by 0.3 mag to compare volume density measurements at luminosities closer to what we measure. After doing so, we find that the L17 stepwise results appear to be a factor of ~ 3–$4\times$ higher than our own results in the luminosity range -17 to -14.5 mag (see Figure 11). After considering various explanations for these differences, it seems likely that the sizeable discrepancies arise due to the excess of sources L17 at the completeness limit, i.e., there are some ~ 22 sources in the 29.25 mag bin ($\Delta m = 0.12$ mag) versus ~ 7 sources per bin brightward of 29.2 mag (this can be clearly seen in Figure 12, which is adapted from Figure 13 of Bouwens et al. 2017). When one combines such an excess with the large intrinsic half-light radii that L17 assume (median of $0.\!''09$) results, it seems clear that L17 may significantly overestimate the volume density implied by galaxies at the faint end of their probe. In Bouwens et al. (2017), we demonstrated through extensive simulations that the assumptions of such large sizes for faint galaxies would result in much higher inferred volume densities for sources (by factors of >5) than if smaller sizes

(i.e., <10 mas) were used. The effect of using large sizes for faint galaxies can be seen in Figure 2 from Bouwens et al. (2017).

Based on recent results from a number of papers (Bouwens et al. 2017; see also Kawamata al. 2015; Laporte et al. 2016), it is now clear that the use of such large sizes for very faint galaxies is not realistic. Several lines of evidence indicate that the intrinsic half-light radii of very low luminosity sources (i.e., >-16 mag) are very small, i.e., $\lesssim 0.\!''03$, with many sources appearing to have sizes $\lesssim 0.\!''01$ (Bouwens et al. 2017). Small sizes imply small completeness corrections and therefore much lower volume densities for faint galaxies. It is important to verify that the sizes assumed in selection volume simulations are realistic, given how sensitive the selection volume estimates (and hence LFs) for faint galaxies are to the sizes assumed for faint galaxies (see Figure 13 of Bouwens et al. 2017).

The very high volume densities L17 find for >-17.5 mag galaxies in their $z \sim 6$ LF seem likely to have impacted the analyses they performed regarding a possible turn-over at the faint end, as we explain in Section 6.2. For similar reasons, the present constraints on the faint-end slope α that we determine, i.e., $\alpha = -1.92 \pm 0.04$ are shallower than those obtained by L17. Given that the constraints can also depend on the gravitational lensing model assumed, it is best to compare faint-end slopes assuming the same gravitational lensing model. If we take their formal constraints as measured assuming the GLAFIC gravitational lensing model, our estimated faint-end slope α is -1.92 ± 0.04 versus the L17 faint-end slope at $z \sim 6$ of -2.10 ± 0.03. The difference in the derived slope is 3.5σ, combining the errors from both measurements of the slope.

The actual differences between our faint-end slope α estimates and L17's estimates are likely even larger than 3.5σ, if we compare like measurements with like. L17 combine their HFF measurements with the Finkelstein et al. (2015) field constraints, which prefer -2.02 ± 0.10. Re-estimating the faint-end slope α from the L17 HFF results alone, we derive their faint-end slopes to be -2.15 to -2.3 (see Appendix E here or Section 6.2 of Bouwens et al. 2017). This is significantly ($\geqslant 4.5\sigma$) steeper than our mean estimate using the HFF data alone of -1.92 ± 0.04 (Section 5.2; Table 2).

6.2. Observational Constraints on a Possible Turn-over at Very Low Luminosities

One issue we examined in this study concerned the existence of a possible turn-over in the $z \sim 6$ UV LF at the faint end. This question is of great interest for the theoretical models, as we shall see in the next section, and so any observational claims regarding where a turnover occurs—or does not occur—require a very high degree of careful analysis and credibility if they are to be of real value to the theoretical modelers.

We noted in Section 5.3 that we were not able to find clear and compelling evidence for a turn-over at the faint end of the LF. Using our likelihood contours for δ-α-ϕ^*, we concluded that a turn-over in the LF is allowed faintward of -14.2 mag or -15.3 mag (at 68% confidence), depending on whether we assumed smaller or larger uncertainties in the magnification maps.

Since we could not find compelling evidence for a turn-over at the faint end of the LF (Section 5.3), we proceeded to examine what constraints we can place on the presence of a turn-over as well as the luminosity at which a turn-over could occur. Using our search results, we concluded that any possible

turn-over in the LF would need to occur at $\gtrsim -14.2$ mag or $\gtrsim -15.3$ mag (within the 68% confidence intervals), depending on the assumptions we made about errors in the magnification maps.

The present conclusions parallel to those of Castellano et al. (2016b), who also concluded that the UV LF at $z \sim 6$ appeared unlikely to show a turn-over at < -15 mag (68% confidence), after factoring in the uncertainties in the magnification maps. We now proceed to examine in more detail what constraints we can place on the presence of a turn-over as well as the luminosity at which a turn-over could occur. We also look in more detail at the L17 result since they claim strong evidence against a turn-over to very low luminosities.

L17 concluded based on their analysis of $z \sim 6$ galaxies behind the first two HFF clusters that they have "positive" and "strong" evidence that any turn-over in the LF must be fainter than -11.1 mag and -12 mag using the criteria $\Delta(\mathrm{BIC}) = 2$ and $\Delta(\mathrm{BIC}) = 6$, respectively (where we draw these numbers from their Figure 12). BIC denotes the Bayesian Information Criterion (Schwarz 1978). In the way L17 apply BIC, BIC is effectively just equal to $\Delta\chi^2$, and the above limits on the turn-over luminosity translate to nominal confidence levels of 84% and 98.5%, respectively. On the basis of the L17 fit results assuming different magnification models, L17 reported an uncertainty in their allowed turn-over luminosity of $^{+0.4}_{-0.8}$ mag and $^{+0.3}_{-0.6}$ mag, respectively. While our conclusions regarding the lack of a bright turnover (at ~ -15 or ~ -14 mag) are consistent with the much lower limit claimed by L17, there are a number of reasons for being concerned about the validity of their constraints on the luminosity of a possible turn-over and their strong statements that it cannot occur at brighter magnitudes.

To be more specific, L17 use their LF results to claim evidence against a turn-over in the $z \sim 6$ LF 3.1 mag fainter than what we do, despite their examination of only half the HFF data set and with a resulting smaller $z \sim 6$ sample (we use four clusters versus their two). How could they claim stronger constraints? Part of this could be because we have recognized, and applied, the very large uncertainties that are inherent in using very highly magnified sources, as indicated in Figure 8, whereas their analysis does not include this very large source of systematic error. Remarkably, however, it appears that the primary reason for the difference lies elsewhere, and not with the faintest sources they report. It appears to arise from the very high volume densities they estimate for the UV LF in the range -17 to -15 (which lie in significant excess of our own determinations and those of Atek et al. 2015b, by factors of ~ 3–4: see Figure 11).

When compared to brighter points at < -18 in the LF, the volume density of sources L17 report over the luminosity range from -17.5 to -15 is sufficiently high (with small error bars) as to suggest an LF form, which steepens further at lower luminosities, i.e., $\delta < 0$, rather than one that retains a fixed faint-end slope and then flattens toward lower luminosities, i.e., $\delta > 0$ (using our formalism). Such a shape, with a "concave up" feature around approximately -17.5 mag is quite unusual (this region is indicated in Figure 11 with the dashed green line). Given this, we suspect that it would be very difficult indeed for L17 to find evidence for a turn-over at intermediate luminosities (since their LF is suggesting the opposite curvature). Given the statistical weight of this upward change of slope at approximately -17.5 to -15 mag, L17 would find

that they needed to probe very faint indeed to find a luminosity where a turn-over was allowed.

A probable explanation for this derives from what we found in the L17 apparent magnitude distribution in Figure 12. The faintest sources are subject to very large completeness corrections. Normally, such sources would only contribute to the faintest bin in the LF, but by virtue of a diversity of magnification factors relevant to sources in this bin, they impact all the fainter bins in the LF. The large number of sources in the 29.13–29.25 mag bin exacerbates this effect, and results in large contributions to the LF over a broad range. We will discuss this further in a future paper (R. J. Bouwens et al. 2017, in preparation), but the current result suggests the need for particular conservatism in selections near the detection limit, when taking advantage of gravitational lensing.

There is a second piece of evidence L17 present that could argue against a possible turn-over in the LF at ~ -15 to ~ -14 mag. This involves their $z \sim 6$ LF at approximately -12.4 mag (A2744_z6_2830). Based on this candidate, L17 estimate a volume density of ~ 6 galaxies per Mpc^3 at ~ -12.5 mag, which would disfavor a turn-over in the LF at any luminosity down to this limit. This candidate galaxy thus assumes a critically important role in their conclusions, and it is therefore very important both to consider and examine, as to its robustness. Interestingly, this $z \sim 6$ candidate from L17 is also detected in our catalogs, though we do not include it in our $z \sim 6$ sample, since the integrated likelihood of the candidate lying at $z < 4$ is $\sim 50\%$ and hence it does not meet our selection criteria. Nonetheless, given its importance, we need to give consideration to this object.

On the basis of our own photometry, we estimate the source to have a total apparent magnitude of 28.9 mag. Using the median magnification factor 61^{+44}_{-15} we derive for the source based on the latest publicly available magnification models (weighting each type equally) and a redshift of $z = 6.11^{+1.05}_{-1.22}$ (L17's estimate), we calculate an absolute magnitude of -13.4 mag for the source. This is ~ 1 mag brighter than what L17 estimate for the same source. L17's reported luminosity is based on a magnification factor of $110.0^{+129.0}_{-22.2}$. This high magnification factor appears to be at the high end of the publicly available models, as we discuss in Appendix F, even though L17's estimate is purportedly a median of those same model results. By contrast, our magnification estimates agree very well ($< 2\%$ difference) with L17's initial estimate of $60.4^{+129.0}_{-22.2}$. It is unclear why L17 changed their median magnification estimate from 60.4 (their version 1 value) to 110 (their published value) since the median magnification is clearly much closer to 60 than it is to 110 (as is obvious from Figure 19 and the entire discussion in Appendix F). We have verified that this is the case both from our own calculations and using the public magnification calculator.[17] Despite this described tension and this faint candidate not being in our $z \sim 6$ selection, we can examine the implications this source would have for our UV LF results if it is indeed at $z \sim 6$.

Taking this single source and dividing by the selection volume in the range $-13.5 < M_{\mathrm{UV},AB} < -13$, we estimate a volume density of $0.28^{+0.64}_{-0.22}$ Mpc^{-3} mag^{-1}. Interestingly, this is completely consistent, as illustrated in Figure 13 (left panel), with the LF constraints we find at ~ -13.4 mag treating the GLAFIC magnification model as reality and recovering the LF

---

[17] https://archive.stsci.edu/prepds/frontier/lensmodels/#magcalc
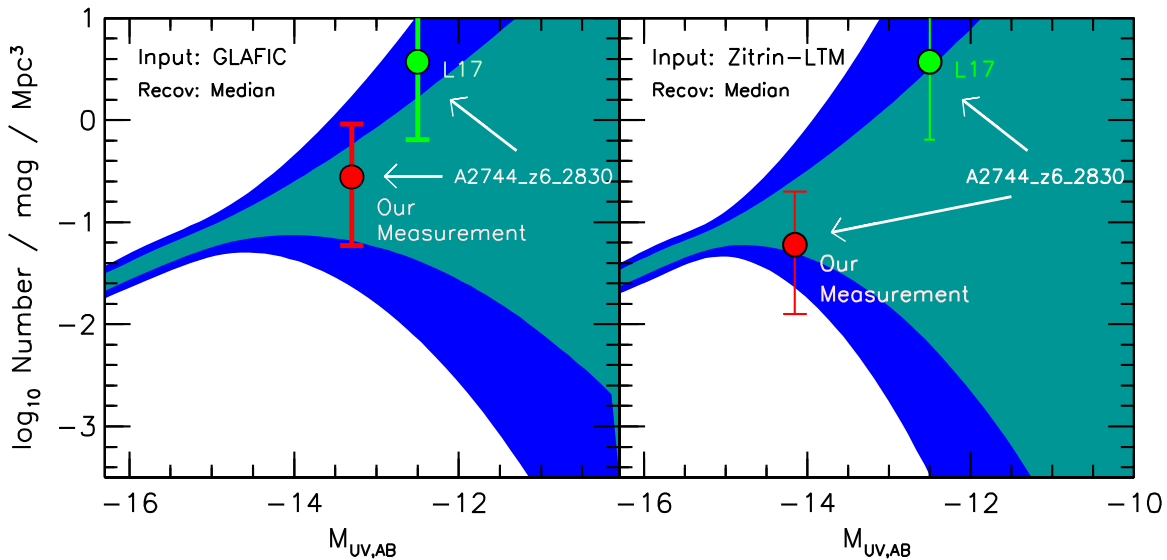
Figure 13. Comparison of the faintest point in the $z \sim 6$ LF from L17 (large green point with 1σ error bars) with the 68% and 95% likelihood intervals implied by our $z \sim 6$ LF results (shaded in cyan and blue, respectively) assuming that the GLAFIC and Zitrin-LTM magnification models represent reality (left and right panels, respectively). This point originates from just a single $z \sim 6$ candidate in the L17 catalog, but is important because it provides significant leverage in their discussion regarding a turn-over. While this source does not satisfy the criteria for our own $z \sim 6$ selection due to its having an estimated probability of $\sim$50% of lying at $z < 4$, we can nevertheless determine the LF constraint we would obtain if we had included it in our $z \sim 6$ sample. This is shown with the large red point in each panel. The absolute magnitude we estimate for this source is 1.0 mag brighter than what L17 estimate. The Zitrin-LTM magnification model implies that this source is another factor of three brighter than in the GLAFIC models. For either lensing model, we find no significant tension between our 68% and 95% likelihood contours and the volume densities we estimate for this candidate using our own photometry and selection volume constraints. While this point plays a significant role in L17's discussion regarding a turn-over, a reassessment of its luminosity and volume density indicates that it is consistent with other forms for the $z \sim 6$ LF, including one with a turn-over at approximately -15 mag.

results using the median magnification model. It is also consistent (right panel to Figure 13) with the LF results we derive using the other magnification models, e.g., Zitrin-LTM, which suggest a magnification factor of 19, with the associated volume density of $\sim 0.06\,\mathrm{Mpc}^{-3}\,\mathrm{mag}^{-1}$, instead of the magnification factor of $\sim$110 adopted by L17. It is not surprising that we recover lower volume densities than those recovered by L17, given our use of smaller sizes for the faint > -16 mag population as now appears to be appropriate (Bouwens et al. 2017). Smaller sizes translate into a higher completeness and selection volume.

Since this source is not included in our LF results, we also checked the impact it would have on our results if it had been included. We found that it only had a moderate impact on the allowed luminosity of a turn-over. We find our constraints on a possible turn-over in the UV LF change by 0.8 mag (becoming fainter) assuming that GLAFIC-versus-median model are typical of the true magnification errors and by $\sim$0.4 mag (becoming fainter) assuming that the Bradac-versus-median or GRALE-versus-median models are more typical of the errors. If we ask which luminosities we can exclude for a turn-over at 95% confidence, the excluded luminosities are < -14.7 mag for the GLAFIC-versus-median case (2.7 mag brighter than what L17 report for this limit).

The considerations discussed in this section suggest that it is not possible using current observational data to definitively rule out the presence of a turn-over in the LF as bright as -15 mag and especially at -14 mag. L17 (despite a smaller $z \sim 6$ sample) had previously claimed strong evidence against a turn-over brightward of ~ -12 mag. From the present discussion, it appears their conclusions were impacted by the large numbers of faint sources incorporated into their LF results near the completion limit (Figure 12) and their size assumptions. This

resulted in an apparent upturn in their $z \sim 6$ LF at $\gtrsim -17.5$ mag, strongly disfavoring a turn-over in their LF results until very low luminosities. A reassessment of L17's faintest source using the public magnification models shows a wide range of estimated magnifications, mostly lower than their value and many consistent with a turn-over at higher luminosities.

6.3. Comparison with Theoretical Expectations

The observational constraints we have obtained here are obviously of great value for comparing against the predictions for the form of the LF at the faint end, as provided by many different teams using simulations, theoretical models, and on the basis of observations of the nearby universe.

We compare our LF constraints with the following cosmological simulation or theoretical model results.

Renaissance (O'Shea et al. 2015): O'Shea et al. (2015) report some of the first results from the "Renaissance" simulations. The "Renaissance" simulations are zoom-in simulations of a $(28.4\,\mathrm{Mpc/h})^3$ volume of the universe, powered by the Enzo code (Bryan et al. 2014), and self-consistently following the evolution of gas and dark matter, including H_2 formation and destruction from photodissociation. Star-formation and supernovae physics are included and ionizing and UV radiation are produced as predicted by Starburst99 (Leitherer et al. 1999). Individual dark-matter particles in the simulations have masses of $2.9 \times 10^4\,M_\odot$, meaning that the smallest halos that are resolved in the simulation are $2 \times 10^6\,M_\odot$ ($\sim$70 particles/halo). Many details of the physical implementation of the implementation of the physics and also sub-grid recipes are provided in Xu et al. (2013, 2014) and Chen et al. (2014). In the "Renaissance" simulations, flattening in the UV LF is a direct result of the

decreasing fraction of baryons converted to stars in the lowest mass halos, due to the impact of radiative feedback and less efficient cooling processes. While it is not yet possible to follow the results of these simulations to $z \sim 6$, results are available at $z \sim 12$ and this is the redshift we use for comparisons.

CoDa (Ocvirk et al. 2016): the Cosmic Dawn (CoDa) simulations are full gravity + hydrodynamic simulations of a large $\sim$(100 Mpc)^3 volume of the universe using the RAMSES code (Teyssier 2002). The simulations include standard prescriptions for star formation and supernovae explosions following standard recipes (Ocvirk et al. 2008; Governato et al. 2009, 2010). One new feature of the CoDa simulations is the inclusion of radiative transfer into the simulations, in the sense that hydrodynamics and radiative transfer are now fully coupled. As a result, the effects of photoionization heating on low-mass galaxies are fully included in the CoDa simulations. Ocvirk et al. (2016) report that radiative feedback plays a big role in suppressing star formation in low-mass galaxies and modulating the faint end of the LF.

Cosmic Reionization On Computers (CROC; Gnedin 2014, 2016): the LF results for the CROC are based on gravity + hydrodynamical simulations using the Adaptive Refinement Treement code (Kravtsov 1999; Kravtsov et al. 2002; Rudd et al. 2008). The CROC simulations include a wide variety of physical processes, including gas cooling and heating processes, molecular hydrogen chemistry, star formation, stellar feedback, radiative transfer of ionizing and UV light from stars. These simulations are conducted in $20\,h^{-1}$ Mpc boxes at a variety of resolutions. The effective slope of CROC LFs continue to flatten toward fainter magnitudes and reach a peak at approximately -12 mag. However, the peak at approximately -12 mag is reported not to be a robust prediction of the simulation and to depend on the minimum particle size in the simulations. The impact of radiative feedback is less important in the CROC simulations than in CoDa.

Finlator et al. (2015, 2016, 2017): the Finlator et al. (2015, 2016, 2017) LF results are based on a cosmological simulation of galaxy formation in a $(7.5\,h^{-1})^3$ Mpc^3 volume of the universe including both gravity and hydrodynamics as implemented in the GADGET-3 code (Springel 2005). To this code, gas cooling is added through collisional excitation of hydrogen and helium as in Katz et al. (1996), and metal line cooling is implemented using the collisional ionization equilibrium tables of Sutherland & Dopita (1993). Star formation is added using the Kennicutt–Schmidt law, with supernovae feedback included following the "ezw" prescription from Davé et al. (2013) and metal enrichment from supernovae as implemented as in Oppenheimer & Davé (2008). Flattening in the Finlator et al. (2015, 2016, 2017) LFs occurs mostly due to less efficient gas cooling at lower halo masses.

DRAGONS (Liu et al. 2016): the LF results from Liu et al. (2016) are based on the Dark-ages Reionization And Galaxy-formation Observables from Numerical Simulations (DRA-GONS)[18] project, which builds semi-numerical models of galaxy formation on top of halo trees derived from *N*-body simulations done over different box sizes to probe a large dynamical range. The semi-numerical models include gas cooling physics, star-formation prescriptions, feedback, and

merging prescriptions, among other components of the model. The turn-over in the LF results of Liu et al. (2016) at approximately -12 mag correspond to the approximate halo masses $\sim 10^8\,M_\odot$ where the gas temperature is 10^4 K. Above this temperature, atomic cooling processes become efficient. In earlier data sets, Muñoz & Loeb (2011) had looked at what constraints could be placed on this mass using earlier LFs of Bouwens et al. (2007).

Jaacks et al. (2013): the simulation results in Jaacks et al. (2013) are powered by the GADGET-3 (Springel 2005) gravity +hydrodynamics code run in three box sizes (10, 33.75, and 100 h^{-1} Mpc) and three different particle sizes (9×10^5, 2×10^7, and $3 \times 10^8\,h^{-1}\,M_\odot$). Radiation cooling is included in the simulations by H, He, and metals (Choi & Nagamine 2009), a UV background self-shielding effect, and heating by a uniform UV background (Faucher-Giguère et al. 2009). Supernovae feedback is impleneted by a momentum-driven wind model (Choi & Nagamine 2011). Star formation in the simulations is governed by the SFR-versus-H$_2$ model of Krumholz et al. (2009) rather than in terms of the total density in cold gas. The implementation of this in the GADGET-3 code is as described in Thompson et al. (2014) and is similar to the implementation of the same recipe by Kuhlen et al. (2013) in the Enzo code. As a result of the lower gas density in molecular hydrogen H$_2$ in fainter, lower-mass galaxies, the LFs predicted by Jaacks et al. (2013) show a turn-over at approximately -15.4 ± 0.6 mag.

Dayal et al. (2014): the LF results from Dayal et al. (2014) are based on a semi-analytic model, which follows the evolution of galaxies in the merger tree constructed from the extended Press–Schechter theory (Lacey & Cole 1993). The star-formation rate in individual galaxies proceeds at such a rate as to balance the impact of supernovae feedback in expelling all the gas from a galaxy. Flattening in the UV LF is partially the result of a similar flattening in the halo mass function, as well as lower efficiency for star formation in the lower-mass halos that contribute to the low luminosity end of the LF.

Yue et al. (2016): Yue et al. (2016) derive their LF results assuming a non-evolving stellar mass–halo mass relation. Yue et al. (2016) adopt a very similar approach to what Mason et al. (2015) employ in predicting the galaxy LF (see also Trenti et al. 2010 and Tacchella et al. 2013). Yue et al. (2016) start with the halo mass function, break up the star-formation history of each halo into segments according to which the halo grows in mass by a factor of two, and then assume that the SFR must be such to maintain a constant stellar mass-halo mass relation which they calibrate to the $z \sim 5$ LF of Bouwens (2015a). Yue et al. (2016) then look into the impact that radiative feedback could have during the epoch of reionization. Yue et al. (2016) derive their LF results assuming that halos below some fixed circular velocity would have their star formation quenched.

Finally, we also include a comparison with the empirical results on the faint end of the LF at high redshift:

Boylan-Kolchin et al. (2015): Boylan-Kolchin et al. (2015) arrive at constraints on the faint-end of the LF at $z \sim 7$ by leveraging deep probes of the color–magnitude relationship of nearby dwarf galaxies, which allow one to estimate the luminosity of these sources at $z \sim 7$. By comparing the distribution of inferred luminosities of these dwarfs with the expected numbers extrapolating $z \sim 7$ LFs to -10 mag, Boylan-Kolchin et al. (2015)

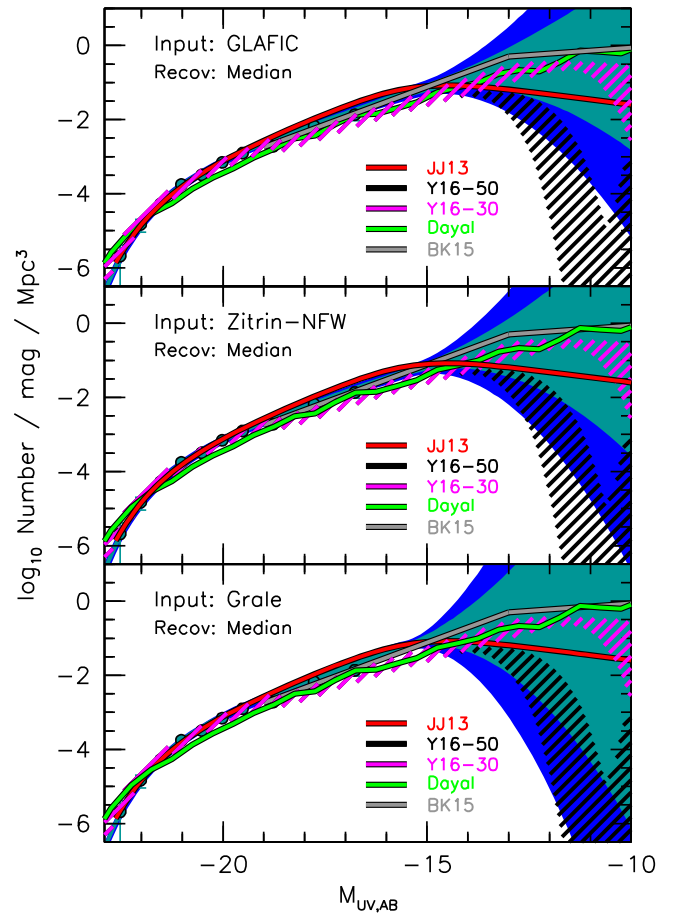---

[18] http://dragons.ph.unimelb.edu.au
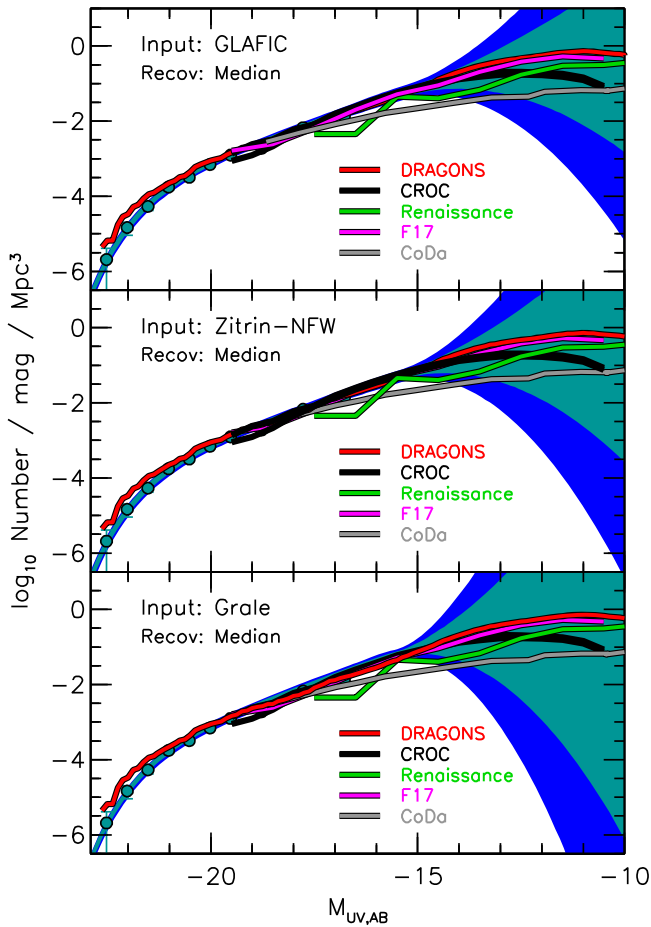
Figure 14. Comparison of the 68% and 95% confidence intervals we have derived on the shape of the $z \sim 6$ UV LF with the predictions for this LF. Confidence intervals are shown making different assumptions about the typical size of errors in the lensing models, assuming these errors to typically be as large as the differences between the median parametric model and the GLAFIC model, Zitrin-NFW models, and GRALE models. The plotted theoretical models include DRAGONS (red lines: Liu et al. 2016), CROC (black lines: Gnedin 2016), ENZO (green lines: O'Shea et al. 2015), CoDa (gray lines: Ocvirk et al. 2016), and Finlator et al. (2015, 2016, 2017) [F17]: purple lines). The LF results from O'Shea et al. 2015, rely on their $z \sim 12$ LF, since those simulations have not yet run down to $z \sim 6$. Overall, we find good agreement between the predicted LF results and the present observational constraints.

Figure 15. Identical to Figure 14 but showing the results for the Jaacks et al. (2013: JJ13) model, two different models from Yue et al. (2016), where radiative feedback becomes important at circular velocities of 30 and 50 km s^{-1}, and the Dayal et al. (2014) model. The dip at -11 mag in the 50 km s^{-1} Yue et al. (2016) model is due to the quenching of star formation in low-mass halos from radiative feedback. Also included among the presented results are the LF constraints implied from the abundance matching analysis Boylan-Kolchin et al. (2015: BK15) perform using dwarfs in the nearby universe.

infer a break in the LF at ~ -13 mag from a faint-end slope of approximately -2 to approximately -1.2.

We present comparisons of the predicted LFs from both sophisticated hydrodynamical simulations, various semi-analytic models, and the empirical results of Boylan-Kolchin et al. (2015) in Figures 14 and 15.

Overall, we find reasonable agreement between our observational results and the predicted LFs from both hydrodynamic simulations and various semi-empirical theoretical models. The predicted LFs from the CoDa simulation (Ocvirk et al. 2016) and from two semi-analytical models Yue et al. (2016) and Dayal et al. (2014), against which we compare, fall slightly below our observational constraints at -15.5 mag by $\sim$0.2–0.3 dex, but otherwise are in reasonable agreement with our results.

In particular, we find that our observational constraints allow for the existence of a flattening or turn-over in the $z \sim 6$ LF at > -15 mag as predicted in the theoretical models, due to a variety of physical processes, including a greater role for

radiative feedback (O'Shea et al. 2015; Ocvirk et al. 2016; Yue et al. 2016) and less efficient cooling in lower mass halos where atomic cooling processes would be less important (Wise et al. 2014; Gnedin 2016; Liu et al. 2016). The present results suggest that these physical processes can impact the shape of the LF at > -15 mag, as is predicted, and there is no fundamental disagreement with observational results to > -14 mag (contrary to reports from L17).

Our observational results are also fully consistent with the abundance matching constraints obtained by Boylan-Kolchin et al. (2015), which suggest a break in the faint end slope of the LF at -13 mag. The toy LF from Boylan-Kolchin et al. (2015) is almost entirely contained with our 68% confidence intervals, suggesting that the flattening they infer from analyses of nearby dwarf galaxies is fully consistent with the HFF observations of faint $z \sim 6$ galaxies.

7. Summary

We have combined a large sample of 160 lensed $z \sim 6$ galaxies from the first four HFF clusters with a first-ever determination of the systematic uncertainties at high magnification in the massive lensing clusters. In so doing, we provide

the most realistic determination yet of the shape of the $z \sim 6$ LF to very low luminositis. This sample of lensed $z \sim 6$ galaxies represents the most comprehensive sample to date. The sample reaches to low luminosities comparable to others when different magnitude and magnification measurement approaches are considered (see Sections 2 and 6.1–6.2). Our analysis provides a much more realistic assessment of the impact of the large magnification uncertainties inherent at high magnifications. This allows us to set improved constraints on the faint-end slope α and also to investigate whether the UV LF shows a turn-over at very low luminosities.

One particular emphasis of this analysis was to include a full account of systematic errors in deriving accurate constraints on the shape of the $z \sim 6$ UV LF. We looked especially at the impact of errors in the magnification maps, but we also considered the impact of uncertainties in the estimated completeness based on the size distribution building upon results in a companion paper (Bouwens et al. 2017).

To explore the impact of errors in the magnification map on LF results, we have developed a new forward modeling approach that involves using one set of magnification models and a candidate LF to create mock catalogs over each of the HFF clusters. These catalogs are then analyzed using the same type of magnification models as are used to interpret the real observations (Figure 6). The likelihood of a given LF can then be assessed by comparing the observed counts with the expected counts derived from the simulations. Our quantification of the general form of the $z \sim 6$ LF includes not only the normal Schechter parameters, but also a curvature parameter δ, which we apply faintward of -16 mag to characterize the form of the LF to very low luminosities.

Our new simulation results using forward modeling demonstrate the substantial impact of magnification errors on the LF results. We show that scatter due to magnification errors results in the LF asymptoting to a faint-end slope of approximately -2 or steeper in the very low luminosity regime when the magnification factors are high $\mu > 10$–30. This occurs regardless of the true slope. This effect is so pervasive that it can eliminate any indication of a turn-over (even if present in reality) at the faint end of the LF (see Figure 4). At lower magnification factors, i.e., $\mu < 10$, where the predictive power of the magnification models is best (e.g., see Figure 3 and also Meneghetti et al. 2016), the impact appears to be most manageable in terms of the overall impact on the LF results.

For higher magnification factors, i.e., at $\mu > 10$ and especially at $\mu > 30$, the predictive power of the magnification models is much poorer, resulting in large uncertainties in the magnification factors. As a result, it can be difficult to determine whether the LF shows a turn-over at approximately -15 mag, whether it steepens further at approximately -15 mag, or whether it continues with a fixed faint-end slope to -12 mag (see Figure 4).

Taking advantage of our forward modeling procedure, we derive new constraints on the faint-end slope of the LF and arrive at a value of -1.92 ± 0.04 using the HFF observations alone (Table 2) and, rather coincidentally, also -1.92 ± 0.04 combining our constraints with the field results of Bouwens et al. (2015a). Both constraints are consistent with our previous determination $\alpha = -1.87 \pm 0.10$ (Bouwens et al. 2015a) using the HUDF+HUDF-parallel+CANDELS data alone. We nevertheless caution that the faint-end slope could be

steeper (by $\Delta \alpha \sim 0.01$–0.03) with a potentially more negative δ (less consistent with a possible turn-over at the faint end of the LF) if the sizes of faint sources are not so small as to be essentially point sources (Bouwens et al. 2017) and closer to conventional size–luminosity relations (e.g., Shibuya et al. 2015). In this case, when the galaxy sizes are substantially larger, the completeness of galaxies faintward of -14.5 mag could become quite large.

We use our new constraints to derive 68% and 95% confidence regions on the faint-end form of the $z \sim 6$ LF, presenting our results in Figure 8. We find no evidence for a turn-over in the LF at the faint end. Nevertheless, we can place constraints on how faint it must be, though the result does depend on the assumed size of the errors in the magnification models. If the true errors in the models are similar to the differences between the GLAFIC model and the median parametric model, our results strongly indicate that a turn-over cannot occur brightward of -14.2 mag (68% confidence). However, if differences between the non-parametric models and the median parametric models are typical, then a turn-over cannot occur brightward of -15.3 mag (68% confidence). Our results are fully consistent with recent observational results from Boylan-Kolchin et al. (2015) and theoretical models (O'Shea et al. 2015; Gnedin 2016; Liu et al. 2016; Ocvirk et al. 2016) predicting some flattening in the UV LFs at > -15 mag.

The faint-end slope α we derive at $z \sim 6$ is -1.92 ± 0.04 and 3.5σ shallower than the Livermore et al. (2017) faint-end slope $\alpha = -2.10 \pm 0.03^{+0.02}_{-0.01}$. The tension with the faint-end slope result of L17 decreases to 3σ, if we allow for larger source sizes (Section 5.4) and hence a steeper α by 0.03. Meanwhile, our constraints on the turn-over are consistent with the findings by Atek et al. (2015b) and Castellano et al. (2016b), but occur at much higher luminosities than what L17 report. Despite having larger samples than L17 and considering twice as many HFF clusters (while probing to comparably low luminosities), we only find evidence against a turn-over brightward of approximately -15.3 and approximately -14.2 mag at 68% confidence, versus the approximately -11.1 mag reported by L17 at nominally slightly higher confidence. We speculate that L17's stronger claims against a turn-over (and steeper faint-end slope results) arose as a result of artifacts in their determinations of the LFs resulting from likely inaccurate size assumptions (see Sections 6.1–6.2 and Bouwens et al. 2017) and a large number of sources near their completeness limit (Section 6.1.2, Figure 12, and Figure 13 of Bouwens et al. 2017). We show that these limitations likely led L17 to set constraints that we cannot reproduce through the analysis of current data sets.

The new formalism we have developed to derive LF results in the presence of errors in the magnification map has significant utility and can be applied to other *HST* observations that have been obtained with the HFF program. In the immediate future, we plan to make use of our new forward-modeling methodology to derive LF results at $z \sim 6$, $z \sim 7$, $z \sim 8$, and $z \sim 9$ from the full HFF program. These results would perhaps provide us with our most complete set of information on the faint-end form of the LFs before *JWST* and also provide us with clues as to how the overall ionizing emissivity evolves with cosmic time.

Appendix A
Performance of Our Intra-cluster Light Subtraction Technique

In this section, we briefly quantify the performance of our procedure for subtracting the intra-cluster light in galaxy clusters relative to that obtained by other groups.

One measure of performance regards the total number of $z = 6$–8 candidate galaxies that it is possible to recover from the observations, after subtraction of the intra-cluster light. We begin by comparison of the number of $z = 6$–7 galaxies in our own samples. When extracting these samples without our foreground cluster subtraction procedure, we find 61 $z = 6$–7 galaxies over Abell 2744, but 71 $z = 6$–7 galaxies when making use of images where the foreground light from the cluster has already been removed. This illustrates the basic increase in numbers one can achieve from a subtraction of the foreground light.

The published results of Merlin et al. (2016) and L17 provide us with a separate benchmark. Merlin et al. (2016) report 138 $z \sim 7$–8 galaxy candidates over the two clusters, while L17 report 161 $z \sim 6$, 7, and 8 candidates. With our procedure, we recover 176 candidates over the two clusters, which is slightly larger than what L17 obtain.

However, the numbers of $z = 6$–8 candidates L17 report are likely boosted by their considering selections generated from coadditions of the Y_{105}, J_{125}, JH_{140}, and H_{160} data in 14 different combinations (e.g., Y_{105}, J_{125}, $Y_{105} + J_{125}$, $Y_{105} + J_{125} + JH_{140}$). By considering selections from many different combinations of such images, it is possible to increase the completeness of one's selections. This occurs since SExtractor often defines the apertures of specific sources in ways that are not entirely ideal and by selecting candidates off many different detection images, one can improve the overall completeness of a selection.

An alternative way to achieve similar gains in sample size is by perturbing the detection image multiple times, rerunning the selection, and adding in to the main sample any new sources that are found. For this test, we perturb the detection image by adding to it a smoothed noise image the same rms as the data itself. The smoothing is a Gaussian kernel with FWHM of $0\farcs28$. Repeating the selection process four additional times and removing redundant sources over MACS 0416, we find that we can recover a 30% higher surface density of $z = 6$–8 sources than running the selection on just a single detection image. Assuming similar gains in numbers over Abell 2744, we estimate a total sample size of 228 $z \sim 6$–8 galaxies behind Abell 2744 and MACS 0416, 65% and 40% larger than claimed by Merlin et al. (2016) and L17, respectively.

Despite the demonstrated potential to make use of a larger sample of $z = 6$–8 sources, we only make use of 87 $z \sim 6$ galaxies (176 $z = 6$–8 galaxies) behind Abell 2744 and MACS 0416 in the LF analyses we conduct in this paper.

Appendix B
Selection Volume Estimates

Here we describe our procedure for estimating the effective selection volumes for faint galaxies behind the HFF clusters we examine.

Our baseline treatment is to model faint galaxies as point sources in estimating their selection volumes. This choice is motivated by our finding in Bouwens et al. (2017) that the faintest $z \sim 2$–8 galaxies behind the HFF clusters had properties consistent with point-source spatial profiles, with no discernible extension along the expected shear axes. Also, no discernible dependence was found for the measured surface densities as a function of the predicted shear in the high-magnification regions. Nonetheless, because such small sizes for galaxies are unexpected (e.g., Liu et al. 2017; Kravtsov 2013), we also consider the impact of larger sizes (and a larger incompleteness) on the LF results throughout the paper.

As point sources, the only quantity of importance from the lensing model is the magnification factor; the form of the deflection (or shear) map has no impact on the results. This simplifies the selection volume simulations enormously, since it means we can estimate the selection volumes for extremely faint sources in the presence of lensing in exactly the same way we would estimate the selection volumes in the absence of lensing. The only quantity of importance in estimating the selection volumes is the apparent magnitudes of the sources. Bouwens et al. (2017) discuss this in their Section 6.3.

We adopt a median value of -2.2 for the UV-continuum slope for galaxies in our simulations to match that found in the observations (e.g., Bouwens et al. 2014: see also Wilkins et al. 2011; Bouwens et al. 2012; Finkelstein et al. 2012; Dunlop et al. 2013; Rogers et al. 2014; Duncan & Conselice 2015).

Adopting these assumptions for the color of the sources and a point-source assumption for the size, we create artificial images for each source over the full suite of passbands and insert these images into the real observations. We then do object detection and photometry using the same procedure as we use in constructing our catalogs (Section 3) and then apply our selection criteria. In this way, we derive the completeness for sources in different regimes.

Selection volumes are computed by multiplying the cosmological volume element by the estimated completeness and integrating over redshift. Following previous work (e.g., Ishigaki et al. 2015; Oesch et al. 2015), we treat different multiple images of the same source as entirely independent for the purposes of our analysis.

Appendix C
Source Size Modulates the Impact Magnification Errors Have on the Inferred Shape of the LF

The impact of magnification errors on the derived LF can also depend on source size, as discussed in Section 4.1. This can occur as a result of the fact that higher magnification sources are generally more difficult to detect, if they are spatially extended, than if their magnification is lower. In other words, the selection efficiency S is a function of the magnification factor μ.
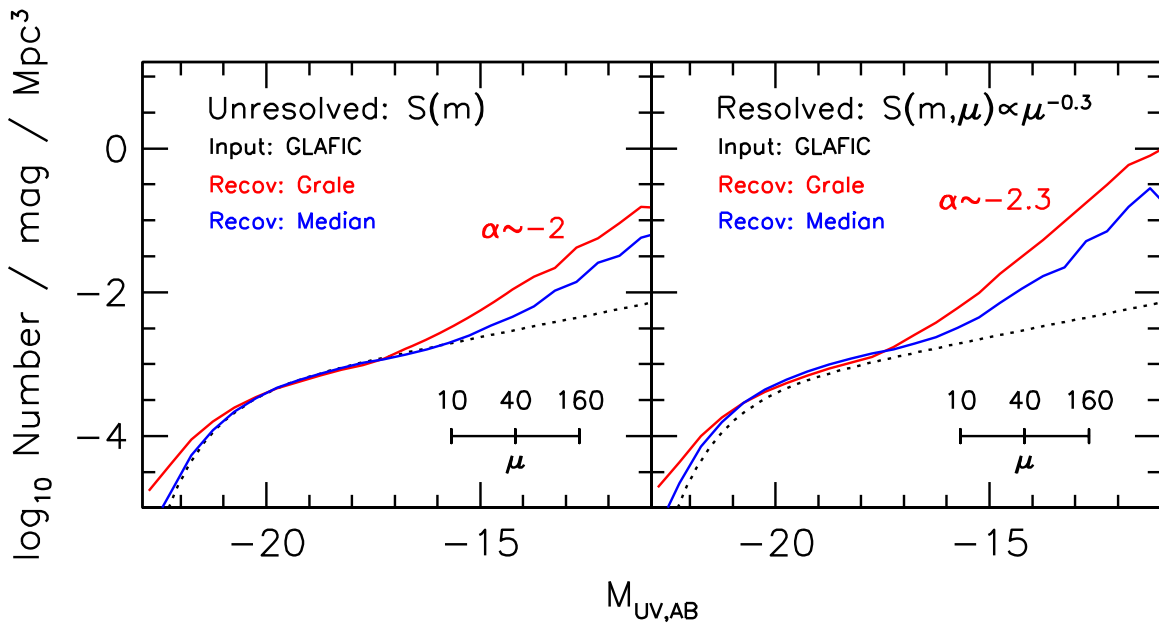
Figure 16. Illustration of how the impact of magnification errors on the recovered LFs depends on the sizes of sources (Appendix C). The magnification scale in the corner is as in Figure 4. An input LF with a faint-end slope of −1.3 and no turn-over at the faint end is assumed. In our forward modeling procedure, the GLAFIC magnification model is used to create the mock catalogs, while the GRALE and median magnification maps are used for recovering the LFs (red and blue lines, respectively). The left panel shows the recovered LFs assuming that point-source spatial profiles for all galaxies, while the right panel shows the recovered LFs assuming more extended sources and where the selection efficiency $S(\mu)$ decreases toward large magnification factors as $\mu^{-0.3}$ (as in Figure 3 from Oesch et al. 2015). The recovered LFs show a departure from the input LF at ∼−15.5 mag, asymptoting toward a slope of −2 for unresolved sources (left panel) and −2.3 for resolved sources (right panel). Previous studies appear to have completely ignored the issue raised in this figure.

The issue is that while the actual surface density of sources in our catalogs is proportional to $S(\mu_{\rm true})$, the selection volumes we estimate for these sources is $S(\mu_{\rm model})$. This results in the recovered volume density for these sources being higher than the true surface density by the factor $\langle S(\mu_{\rm true})\rangle/\langle S(\mu_{\rm model})\rangle$. When $\langle\mu_{\rm true}\rangle \sim \langle\mu_{\rm model}\rangle$, no bias is present in the recovery volume density of sources. However, when $\langle\mu_{\rm true}\rangle$ is less than $\langle\mu_{\rm model}\rangle$, as is often the case at higher magnifications $\mu > 10$ (Figure 3), the recovered surface density of sources can be significantly higher than reality. For example, assuming that $S(\mu) \propto \mu^{-0.3}$ as in Figure 3 of Oesch et al. (2015), the LF asymptotes toward a faint-end slope of −2.3.

To illustrate the impact of source size and different assumptions about $S(\mu)$, we consider two different cases: the first involving point sources where S is independent of μ and the second involving extended sources where $S(\mu)$ is proportional to $\mu^{-0.3}$. Similar to the simulations run in Section 3.2, we use the GLAFIC model to construct mock catalogs and then recover the LF using either the GRALE model or the median parametric magnification model. In the two cases, we incorporate the different dependencies of S on the magnification factor μ for both the catalog construction and recovery of the LF. The input LF for the simulations has a faint-end slope of −1.3, with no turn-over at the faint end.

The results are presented in Figure 16 with the red and blue lines indicating recovery by the Grale and median parametric magnification models, respectively. Differences between the two size cases are immediately obvious. In the point-source case (where $S(\mu)$ is independent of μ), the faint-end slope asymptotes to −2. However, in the case of extended sources (where $S(\mu)$ scales as, e.g., $\mu^{-0.3}$), the faint-end slope instead asymptotes to −2.3. In general, one expects the slope to equal $-2 + d(\ln S(\mu))/d(\ln\mu)$, as Figure 5 illustrates in Section 3.2. In either case, the LFs asymptotically approach these slopes

faintward of −16 mag assuming the GRALE model and faintward of −15 assuming the median parametric magnification model.

This example should reinforce how difficult it is to obtain accurate constraints on the shape of the LF at >−15 mag and thus to detect the existence of a flattening or turn-over in the LF. Not only do the results depend on the magnification level to which magnification maps retain their predictive power (e.g., see Figure 3), but the results also depend significantly (i.e., $\Delta\alpha \sim 0.3$) on the size distribution for faint sources.

We emphasize that the impact this has on the LF shape is distinct from the effect already discussed in the companion study to the present one (Bouwens et al. 2017), where the faint-end slope α of the LF could be biased if the sizes and hence selection volumes were improperly estimated. This bias explicitly arises because of errors in the magnification map and due to mismatches between $\langle S(\mu_{\rm true})\rangle$ and $\langle S(\mu_{\rm model})\rangle$.

If faint sources are slightly resolved (after magnification)—as assumed in many recent studies of faint galaxies—this bias has the potential to be quite significant at absolute magnitudes $M_{\rm UV}$ of >−15 mag, where $\mu > 20$. Amazingly, however, previous work appears to have neither recognized the importance of such a bias, nor made use of procedures that would allow for its correction, even though given the size assumptions in, e.g., L17, this bias would constitute an important effect.

Appendix D
Recovery of LFs with Shallower Faint-end Slopes

Errors in the magnification maps can have a substantial impact on the shape of the $z \sim 6$ LF faintward of −15 mag. We already illustrated this in Section 3.2 of this paper using an LF that turned over at −15 mag (Figure 4).
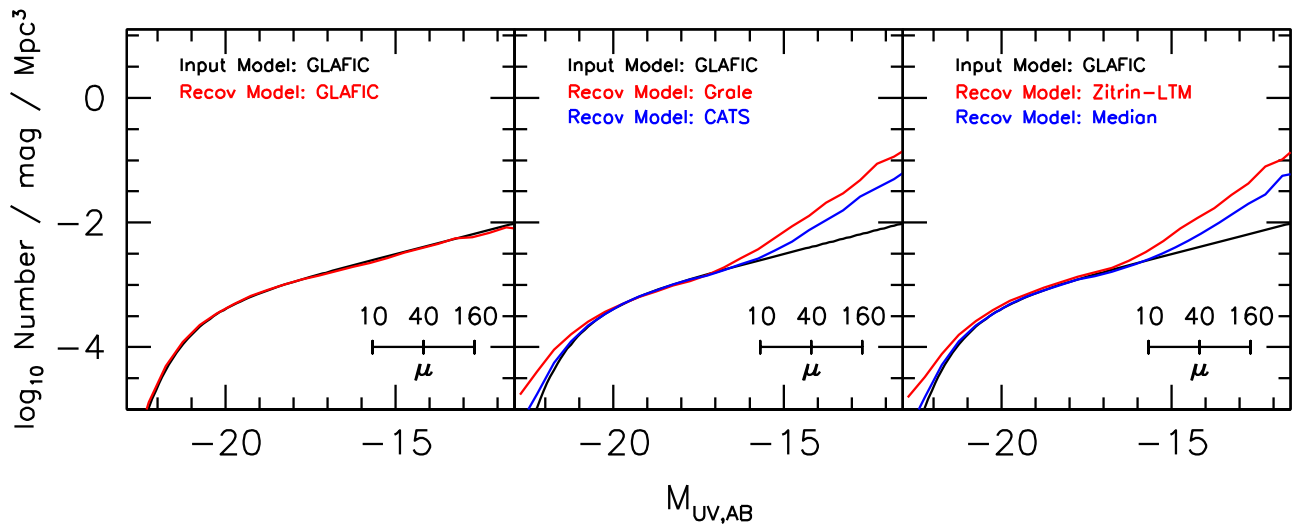
Figure 17. Comparison of an input LF with a shallower faint-end slope of −1.3 with the recovered LFs using a forward-modeling procedure where we create mock catalogs using the GLAFIC magnification maps and recover the LF results using the CATS, GRALE, Zitrin-LTM, and median parametric magnification maps (see Appendix D). The typical magnification levels of sources probing a given luminosity range are indicated by magnification scale in the corner. The recovered LFs show excellent agreement with the input LFs to −16.5 mag, but show a departure at −15 mag and rapidly asymptote toward a faint-end slope of −2. Interestingly, the absolute magnitudes M_{UV} where this departure occurs correspond to magnification factors where the models lose their predictive power (Figure 3).
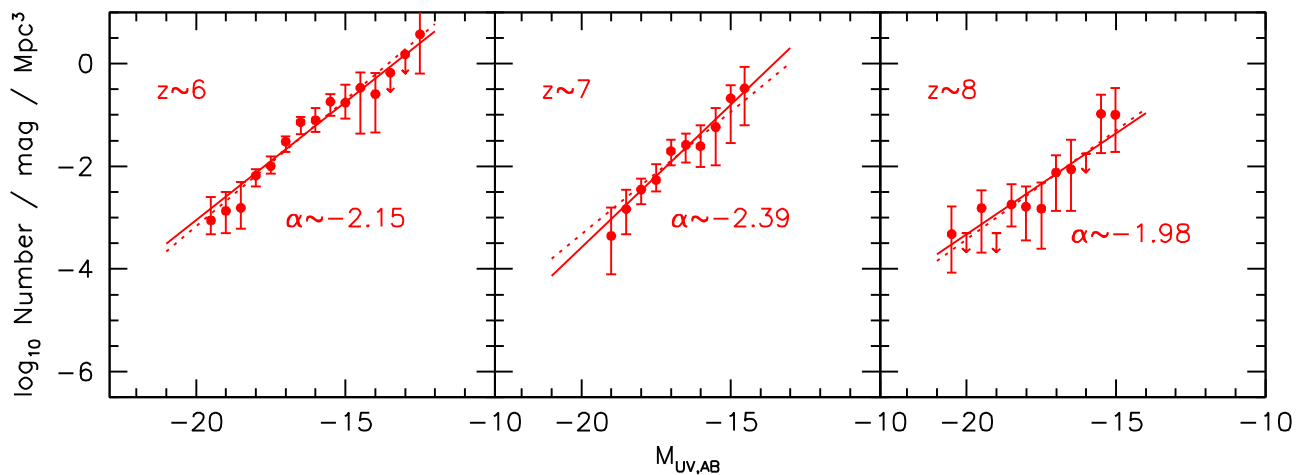
Figure 18. Power-law fits (solid lines) to the binned $z \sim 6$, 7, and 8 LF results from L17 in an effort to estimate the faint-end slope estimates from the HFF results alone. Also presented (dotted lines) are fits done to the L17 HFF LF results anchored to the field LF results from the same group (Finkelstein et al. 2015) at $M_{UV,AB} = -20$. The motivation for deriving the faint-end slopes α from the HFF results alone is to keep the derived results independent of those derived for the field. This makes it possible to compare the lensed LF results and field LF results in a fair way, as we do in Figure 1.

Here we show the impact of these magnification errors using a $z \sim 6$ LF with a faint-end slope −1.3, which we intentionally take to be substantially shallower than −2 (the direction in which magnification errors drive the apparent faint-end slope). Again, we use a forward modeling procedure where we create the mock catalogs using the GLAFIC magnification maps and then alternatively recover the LFs with the CATS, GRALE, Zitrin-LTM, and median parametric magnification maps.

The result is shown in Figure 17, and excellent recovery of the LF is observed brightward of −16.5 mag for all magnification maps. The best performance is achieved using the median magnification map; however, we note that faintward of −15 mag, the recovered LF still diverges from the input LF, rapidly transitioning faintward of −15 to a faint-end slope of −2.

Appendix E
Estimates of the Faint-end Slopes in Previous Work Using Only the HFF Data

In utilizing the data from the HFF clusters to map out the faint-end of the UV LF and derive faint-end slope results, it is valuable to perform this exercise using only the HFF samples to preserve the independence of the faint-end slope determinations from those derived from the field (i.e., the HUDF). By doing so, one can conduct fair comparisons between faint-end slope results α derived using lensing clusters and from the field to test for consistency.

Toward this end, we have taken the binned LF results from L17 on the HFF clusters and fit the results to a power law to estimate a faint-end slope α. The results are presented in Figure 18 as the solid lines and give slopes of -2.15 ± 0.09,
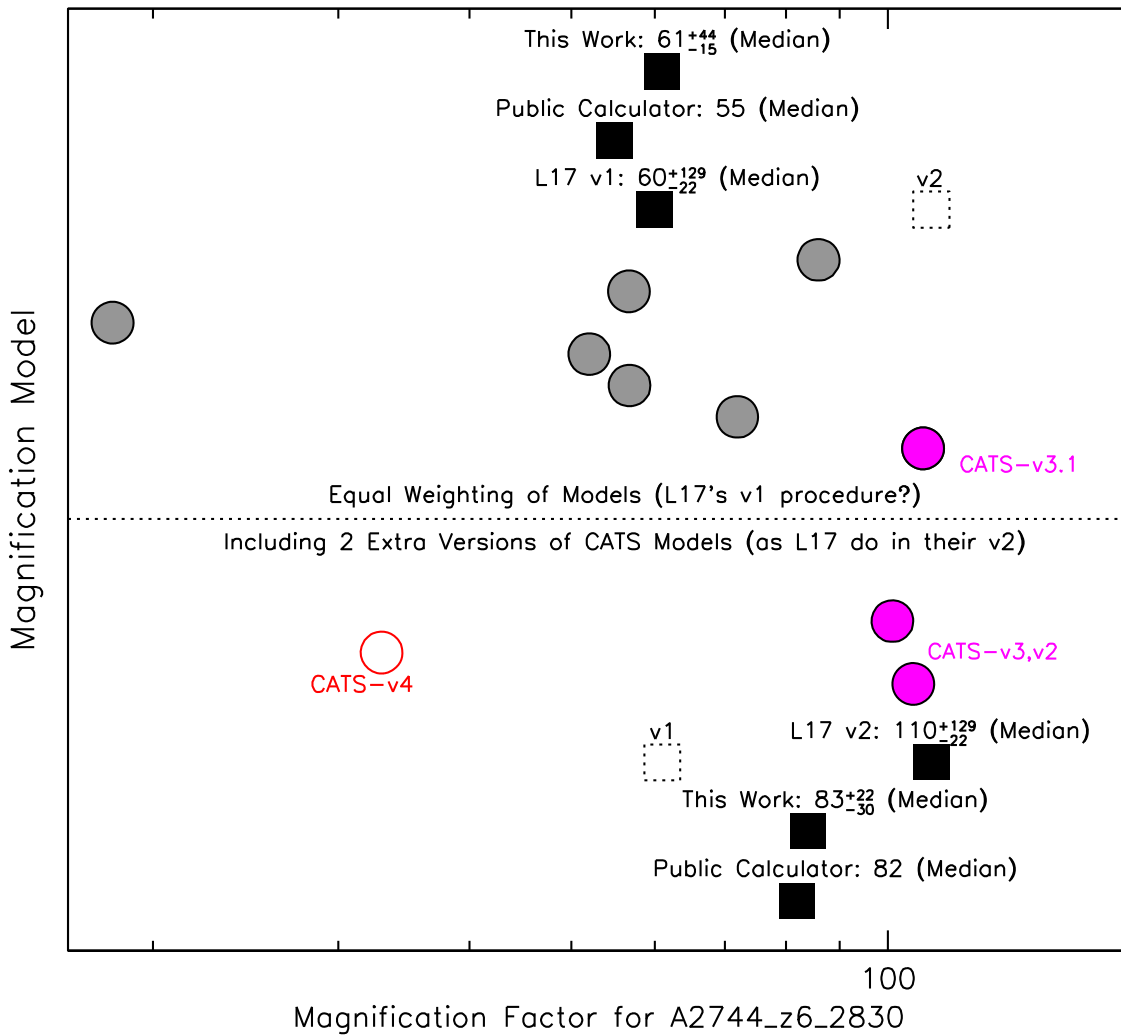
Figure 19. Magnification of the faintest source A2744_z6_2830 in L17 as computed using the latest version of each of the post-HFF models (filled circles above the dotted line) and as computed from just version 2.1 and 3.0 of the CATS models (filled circles below the dotted line). The gray circles give the median magnification estimates based on all magnification models of a given type, i.e., CATS, GLAFIC, SHARON/JOHNSON, ZITRIN-NFW, ZITRIN-LTM, BRADAC, and GRALE. The three CATS models given preferential weight in L17 are shown in magenta. The magnification estimate for A2744_z6_2830 from version 4 of the CATS team (Mahler et al. 2017) is shown with the red open circle. Interestingly, the latest CATS version 4 gives magnifications for A2744_z6_2830 that are $\sim$3.0–3.4$\times$ lower than the earlier version 2 and 3 CATS models. This indicates that the preferential weight given by L17 to the results from the V3 and V2 CATS models was not well-founded. The median magnification factors we compute, giving equal weight to the latest model of a given type (61^{+44}_{-15}), and including two additional versions of the CATS models (83^{+20}_{-32}), are shown with the solid black square above and below the dotted line, respectively. The median magnification factors are similarly estimated using the public calculator and are in excellent agreement with both of our own determinations. The median magnification factor L17 quote (110^{+129}_{-22}) in the final version is significantly higher than what both we and the public calculator compute based on the public models (83 and 82, respectively). The high magnification reported by L17 appears to lack a clear justification. In contrast, the reported magnification factor 60^{+129}_{-22} in version 1 of their paper (plotted in the upper panel as a black square) agrees very well with the median magnification factor from the public calculator and from our calculations. In both cases the magnifications were computed by weighting each model equally, suggesting this was how L17 had originally computed their magnification factor for the source. The magnification factor L17 quote in their version 1 was an excellent representation of the median magnification across the models. It is unclear why L17 subsequently gives preferential weight to the CATS models in their version 2 instead of using the more appropriate values nearer the median of all models, and why they adopted a value that was nearly double their original magnification factor, and even larger than any model estimate. All the indications are that a value around 60 for the magnification of A2744_z6_2830 is well-justified by the models.

-2.39 ± 0.24, and -1.98 ± 0.27 at $z \sim 6$, $z \sim 7$, and $z \sim 8$, respectively. Interestingly, these results are mostly steeper than the faint-end slopes inferred from field searches at the relevant redshifts. Part of this difference could be due to the modest bias toward steeper slopes as a result of the large sizes L17 use in estimating the selection volumes (Bouwens et al. 2017: see their Figure 2).

We can obtain even stronger constraints on the faint-end slope α results from the HFF clusters by having at least one point on the bright end of the LF from field searches to use as an anchor. There is not much search volume available behind clusters to constrain this part of the LF, and so including this information is useful. We therefore refit the $z \sim 6$, $z \sim 7$, and $z \sim 8$ LF results on the HFF clusters from L17, anchoring the fit results to the field LF results from the same group, i.e., Finkelstein et al. (2015), at -20 mag using their best-fit Schechter function. We chose -20 mag somewhat arbitrarily to be close enough to the knee of the LF, i.e., ~ -21 mag, while not being so bright as to be affected by uncertainties in the assumed characteristic luminosity or form of the bright end of the LF (Schechter versus power law, i.e., Bowler et al. 2015).

The fits are presented in Figure 18 as the dotted lines. The faint-end slopes in this case are -2.23 ± 0.05, -2.19 ± 0.08, and -2.06 ± 0.15 for the $z \sim 6$, $z \sim 7$, and $z \sim 8$ LFs,

respectively. These values are consistent with the faint-end slope α results we derive from the HFF results from L17 without a bright anchor point.

Appendix F
Magnification Factor for the Faintest Source in L17

The faintest $z \sim 6$ candidate in L17 plays a key role in anchoring their LF at the faint end, as we noted in Section 6.2. The magnification factor they used to establish the absolute magnitude was very large ($110.0^{+129.0}_{-22.2}$). Given the importance of this source in their analysis we wanted to check the high magnification quoted. We performed this test using the same set of magnification models that L17 listed as being used, i.e., GLAFIC, Sharon/Johnson, Zitrin-NFW, Bradac, Grale, Zitrin-LTM, and three versions of the CATS models, including both the best model and the other models in the MCMC chains. The CATS model versions were the version 3.0, 3.1, and the 2.1 models (the version 2.1 gives a higher magnification estimate than the version 2.2 model for the source).

For each model, a magnification factor can be computed using Equation (1) using an interpolation of the public γ and κ maps and using an assumed redshift to compute the D_{LS}/D_S factor, which we draw from the median redshift and uncertainties quoted for the faintest source by L17, i.e., $z_s = 6.11^{+1.05}_{-1.22}$. We derive the magnification factor from each magnification map separately in the same way. The results are shown in Figure 19. Both the estimated magnification factor using all models (black solid square) are plotted, as well as that estimated from each model individually (gray circles). Each of our magnification estimates matches that computed by the public calculator to typically within 4%, as one would expect given that the process of computing magnifications from the public models is well defined.

The spread in the magnification factor is very large, from <20 to just over 100, with the largest being the CATS version 2 and 3 models with magnifications around 100. The median magnification estimate we derive based on the 9 presented models is 83^{+22}_{-30}. Using the public calculator to calculate median magnification for the source based on the full range of MCMC models available and taking the median of the seven models, we find 82 (shown as a black solid square in Figure 19), almost identical to our own estimate as we would expect given that the procedure for the calculating magnification factor based on the public models is very well defined. In contrast, L17 estimate a magnification of 110^{+129}_{-22} for the same source by purportedly taking a median of the same models. It is unclear why the L17's magnification factor is higher than what we estimate based on the same public models given that the calculation is very well-defined. In addition, it is unclear how their estimate could be a median if it lies higher than for all but one model or why their determination for the lower interquartile range on the magnification factor, i.e., 88 (=110−22), lies higher than for six of the nine individual models.

One possible explanation would be if they treated the source as extended. While such a treatment is not particularly justified given the essentially point-like spatial profile of the source, it turns out that this does not matter for the final result. Because the average magnification factor would correspond to $(\int_A f(A) dA)/(\int_A f(A)/\mu(A) dA)$, where $f(A)$ is the flux profile of a source and A is the area of integration, particularly high magnification areas are de-weighted in computing the net magnification factor (biasing the net magnification lower not

higher). If we re-estimate the magnification factor assuming a source size of $0''.4$, we derive a median magnification factor of 83 (including three of the four versions of the CATS post-HFF models in the weighting as L17 report to do). This is essentially identical to the magnification factor we estimate if we assume that A2744_z6_2830 were a point source.

In providing context for their claimed median magnification factor of 110, we note also that L17 report a large magnification factor of 150 for A2744_z6_2830 from the v3 Bradac and v2 Sharon models. These estimates, however, are significantly higher than what we find and what is derived from the HFF public calculator. For the Bradac v3 models, we use our own procedures and the public calculator and find magnification factors of 52 and 56, respectively. For the Sharon v2 models, we compute a magnification factor of 109 and 107, respectively. Note that these v2 models predate the HFF and so it is unclear why L17 bring them into the discussion since they are not incorporated into the computed median magnification factor. L17's claimed magnification factors of 150 from both models are ~3x and $\sim1.4\times$ larger than both our own calculations and those from the public calculator. Thus, L17 unfortunately appear to have likely erred in computing the magnification factor for A2744_z6_2830 based on the public models.

Probably the most robust estimate for the magnification of A2744_z6_2680 can be obtained by weighting each of the modeling efforts equally (not weighting one modeling effort in excess of the others as L17 do) and repeating the above process. Following this procedure, we compute 61^{+44}_{-15}, which is lower than our earlier estimate including three versions of the CATS models in the median (instead of one). Making use of the public calculator, we compute a median magnification of 55 (shown as a black solid square). Interestingly enough, both of these estimates agree very well with the estimate L17 provide in their version 1 for this source ($60.4^{+129.0}_{-22.2}$), suggesting that this is how L17 had originally calculated the magnification factor for their faintest source. It is unclear why L17's quoted estimate approximately doubled from version 1 to 2, since the available models appear to best support a magnification factor of 60 for the source (Figure 19).

References

Alavi, A., Siana, B., Richard, J., et al. 2014, ApJ, 780, 143
Alavi, A., Siana, B., Richard, J., et al. 2016, ApJ, 832, 56
Atek, H., Richard, J., Jauzac, M., et al. 2015a, ApJ, 814, 69
Atek, H., Richard, J., Kneib, J.-P., et al. 2014, ApJ, 786, 60
Atek, H., Richard, J., Kneib, J.-P., et al. 2015b, ApJ, 800, 18
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 39
Bouwens, R. 2016, in Understanding the Epoch of Cosmic Reionization: Challenges and Progress, Vol. 423 (Basel: Springer), 111
Bouwens, R., Bradley, L., Zitrin, A., et al. 2014, ApJ, 795, 126
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012, ApJ, 754, 83
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015a, ApJ, 803, 34
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015b, ApJ, 811, 140
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2017, ApJ, 843, 41
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, MNRAS, 452, 1817
Boylan-Kolchin, M., Bullock, J. S., & Garrison-Kimmel, S. 2014, MNRAS, 443, L44
Boylan-Kolchin, M., Weisz, D. R., Johnson, B. D., et al. 2015, MNRAS, 453, 1503
Bradač, M., Treu, T., Applegate, D., et al. 2009, ApJ, 706, 1201
Bryan, G. L., Norman, M. L., O'Shea, B. W., et al. 2014, ApJS, 211, 19
Calvi, V., Pizzella, A., Stiavelli, M., et al. 2013, MNRAS, 432, 3474
Castellano, M., Amorín, R., Merlin, E., et al. 2016a, A&A, 590, A31

Castellano, M., Yue, B., Ferrara, A., et al. 2016b, ApJL, 823, L40
Chen, P., Wise, J. H., Norman, M. L., Xu, H., & O'Shea, B. W. 2014, ApJ, 795, 144
Choi, J.-H., & Nagamine, K. 2009, MNRAS, 393, 1595
Choi, J.-H., & Nagamine, K. 2011, MNRAS, 410, 2579
Coe, D., Bradley, L., & Zitrin, A. 2015, ApJ, 800, 84
Davé, R., Katz, N., Oppenheimer, B. D., Kollmeier, J. A., & Weinberg, D. H. 2013, MNRAS, 434, 2645
Dayal, P., Ferrara, A., Dunlop, J. S., & Pacucci, F. 2014, MNRAS, 445, 2545
Diego, J. M., Broadhurst, T., Molnar, S. M., Lam, D., & Lim, J. 2015, MNRAS, 447, 3130
Duncan, K., & Conselice, C. J. 2015, MNRAS, 451, 2030
Dunlop, J. S., Rogers, A. B., McLure, R. J., et al. 2013, MNRAS, 432, 3520
Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJL, 763, L7
Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, ApJ, 703, 1416
Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012, ApJ, 756, 164
Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., et al. 2015, ApJ, 810, 71
Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., Hernquist, L., et al. 2015, ApJ, 810, 71
Finlator, K., Oppenheimer, B. D., Davé, R., et al. 2016, MNRAS, 459, 2299
Finlator, K., Prescott, M. K. M., Oppenheimer, B. D., et al. 2017, MNRAS, 464, 1633
Finlator, K., Thompson, R., Huang, S., et al. 2015, MNRAS, 447, 2526
Giallongo, E., Grazian, A., Fiore, F., et al. 2015, A&A, 578, A83
Gnedin, N. Y. 2014, ApJ, 793, 29
Gnedin, N. Y. 2016, ApJL, 825, L17
Governato, F., Brook, C. B., Brooks, A. M., et al. 2009, MNRAS, 398, 312
Governato, F., Brook, C., Mayer, L., et al. 2010, Natur, 463, 203
Grazian, A., Castellano, M., Koekemoer, A. M., et al. 2011, A&A, 532, A33
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6
Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2015, ApJ, 799, 12
Jaacks, J., Thompson, R., & Nagamine, K. 2013, ApJ, 766, 94
Jauzac, M., Jullo, E., Eckert, D., et al. 2015a, MNRAS, 446, 4132
Jauzac, M., Richard, J., Jullo, E., et al. 2015b, MNRAS, 452, 1437
Johnson, T. L., Sharon, K., Bayliss, M. B., et al. 2014, ApJ, 797, 48
Jullo, E., & Kneib, J.-P. 2009, MNRAS, 395, 1319
Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19
Kawamata, R., Ishigaki, M., Shimasaku, K., Oguri, M., & Ouchi, M. 2015, ApJ, 804, 103
Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016, ApJ, 819, 114
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kravtsov, A. V. 1999, PhD thesis, New Mexico State Univ.
Kravtsov, A. V. 2013, ApJL, 764, L31
Kravtsov, A. V., Klypin, A., & Hoffman, Y. 2002, ApJ, 571, 563
Kron, R. G. 1980, ApJS, 43, 305
Krumholz, M. R., & Dekel, A. 2012, ApJ, 753, 16
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 699, 850
Kuhlen, M., & Faucher-Giguère, C.-A. 2012, MNRAS, 423, 862
Kuhlen, M., Madau, P., & Krumholz, M. R. 2013, ApJ, 776, 34
Lacey, C., & Cole, S. 1993, MNRAS, 262, 627
Laporte, N., Infante, L., Troncoso Iribarren, P., et al. 2016, ApJ, 820, 98
Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3
Liesenborgs, J., De Rijcke, S., & Dejonghe, H. 2006, MNRAS, 367, 1209
Liu, C., Mutch, S. J., Angel, P. W., et al. 2016, MNRAS, 462, 235
Liu, C., Mutch, S. J., Poole, G. B., et al. 2017, MNRAS, 465, 3134
Livermore, R., Finkelstein, S., & Lotz, J. 2017, ApJ, 835, 113 [L17]
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
Madau, P., & Haardt, F. 2015, ApJL, 813, L8
Mahler, G., Richard, J., Clément, B., et al. 2017, MNRAS, arXiv:1702.06962
Maizy, A., Richard, J., de Leo, M. A., Pelló, R., & Kneib, J. P. 2010, A&A, 509, A105
Mason, C. A., Trenti, M., & Treu, T. 2015, ApJ, 813, 21

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, ApJ, 768, 105
McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, MNRAS, 432, 2696
Meneghetti, M., Natarajan, P., Coe, D., et al. 2016, arXiv:1606.04548
Merlin, E., Amorín, R., Castellano, M., et al. 2016, A&A, 590, A30
Merten, J., Meneghetti, M., Postman, M., et al. 2015, ApJ, 806, 4
Mitra, S., Choudhury, T. R., & Ferrara, A. 2015, MNRAS, 454, L76
Mitra, S., Choudhury, T. R., & Ferrara, A. 2016, arXiv:1606.02719
Muñoz, J. A., & Loeb, A. 2011, ApJ, 729, 99
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Nestor, D. B., Shapley, A. E., Kornei, K. A., Steidel, C. C., & Siana, B. 2013, ApJ, 765, 47
Ocvirk, P., Gillet, N., Shapiro, P. R., et al. 2016, MNRAS, 463, 1462
Ocvirk, P., Pichon, C., & Teyssier, R. 2008, MNRAS, 390, 1326
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2015, ApJ, 808, 104
Oguri, M. 2010, PASJ, 62, 1017
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Oppenheimer, B. D., & Davé, R. 2008, MNRAS, 387, 577
O'Shea, B. W., Wise, J. H., Xu, H., & Norman, M. L. 2015, ApJL, 807, L12
Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, MNRAS, 456, 3194
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2015, arXiv:1502.01589 [PC15]
Priewe, J., Williams, L. L. R., Liesenborgs, J., Coe, D., & Rodney, S. A. 2017, MNRAS, 465, 1030
Richard, J., Jauzac, M., Limousin, M., et al. 2014, MNRAS, 444, 268
Robertson, B. E., Ellis, R. S., Dunlop, J. S., et al. 2014, ApJL, 796, L27
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19
Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71
Rodney, S. A., Patel, B., Scolnic, D., et al. 2015, ApJ, 811, 70
Rogers, A. B., McLure, R. J., Dunlop, J. S., et al. 2014, MNRAS, 440, 3714
Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, ApJ, 672, 19
Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, ApJ, 768, 196
Schwarz, G. 1978, AnSta, 6, 461
Sebesta, K., Williams, L. L. R., Mohammed, I., Saha, P., & Liesenborgs, J. 2016, MNRAS, 461, 2126
Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15
Siana, B. 2013, HST Proposal, 13389
Siana, B. 2015, HST Proposal, 14209
Siana, B., Shapley, A. E., Kulas, K. R., et al. 2015, ApJ, 804, 17
Siana, B., Teplitz, H. I., Ferguson, H. C., et al. 2010, ApJ, 723, 241
Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Somerville, R. S., Lee, K., Ferguson, H. C., et al. 2004, ApJL, 600, L171
Springel, V. 2005, MNRAS, 364, 1105
Sutherland, R. S., & Dopita, M. A. 1993, ApJS, 88, 253
Tacchella, S., Trenti, M., & Carollo, C. M. 2013, ApJL, 768, L37
Teyssier, R. 2002, A&A, 385, 337
Thompson, R., Nagamine, K., Jaacks, J., & Choi, J.-H. 2014, ApJ, 780, 145
Trenti, M., & Stiavelli, M. 2008, ApJ, 676, 767
Trenti, M., Stiavelli, M., Bouwens, R. J., et al. 2010, ApJL, 714, L202
Vanzella, E., de Barros, S., Vasei, K., et al. 2016, ApJ, 825, 41
Vanzella, E., Guo, Y., Giavalisco, M., et al. 2012, ApJ, 751, 70
Wilkins, S. M., Bunker, A. J., Stanway, E., Lorenzoni, S., & Caruana, J. 2011, MNRAS, 417, 717
Willott, C. J., Delorme, P., Reyle, C., et al. 2010, AJ, 139, 906
Wise, J. H., Demchenko, V. G., Halicek, M. T., et al. 2014, MNRAS, 442, 2560
Xu, H., Ahn, K., Wise, J. H., Norman, M. L., & O'Shea, B. W. 2014, ApJ, 791, 110
Xu, H., Wise, J. H., & Norman, M. L. 2013, ApJ, 773, 83
Yan, H., & Windhorst, R. A. 2004, ApJL, 612, L93
Yue, B., Ferrara, A., & Xu, Y. 2016, MNRAS, 463, 1968
Zitrin, A., Broadhurst, T., Bartelmann, M., et al. 2012, MNRAS, 423, 2308
Zitrin, A., Fabris, A., Merten, J., et al. 2015, ApJ, 801, 44
Zitrin, A., Meneghetti, M., Umetsu, K., et al. 2013, ApJL, 762, L30