



MAGAZ3NE: Massive, Extremely Dusty Galaxies at $z \sim 2$ Lead to Photometric Overestimation of Number Densities of the Most Massive Galaxies at $3 < z < 4^*$

Ben Forrest¹ , M. C. Cooper² , Adam Muzzin³ , Gillian Wilson⁴ , Danilo Marchesini⁵ , Ian McConachie⁶ , Percy Gomez⁷ , Marianna Annunziatella⁸ , Z. Cemile Marsan³ , Joey Braspenning⁹, Wenjun Chang⁶ , Gabriella de Lucia^{10,11} , Fabio Fontanot^{10,11} , Michaela Hirschmann^{10,12} , Dylan Nelson¹³ , Annalisa Pillepich¹⁴ , Joop Schaye⁹ , Stephanie M. Urbano Stawinski² , Mauro Stefanon^{15,16} , and Lizhi Xie¹⁷

¹ Department of Physics and Astronomy, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA; bforrest@ucdavis.edu

² Department of Physics and Astronomy, University of California, Irvine, 4129 Frederick Reines Hall, Irvine, CA 92697, USA

³ Department of Physics and Astronomy, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

⁴ Department of Physics, University of California, Merced, 5200 North Lake Road, Merced, CA 95343, USA

⁵ Department of Physics and Astronomy, Tufts University, 574 Boston Avenue, Medford, MA 02155, USA

⁶ Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA

⁷ W.M. Keck Observatory, 65-1120 Mamalahoa Hwy., Kamuela, HI 96743, USA

⁸ Centro de Astrobiología (CSIC-INTA), Ctra de Torrejón a Ajalvir, km 4, E-28850 Torrejón de Ardoz, Madrid, Spain

⁹ Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

¹⁰ INAF—Astronomical Observatory of Trieste, Via G.B. Tiepolo 11, Trieste, Italy

¹¹ IFPU—Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Trieste, Italy

¹² Institute of Physics, Lab for Galaxy Evolution, EPFL, Observatoire de Sauverny, Chemin Pegasi 51, 1290, Switzerland

¹³ Universität Heidelberg, Institut für Theoretische Astrophysik, ZAH, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany

¹⁴ Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

¹⁵ Departament d'Astronomia i Astrofísica, Universitat de València, C. Dr. Moliner 50, E-46100 Burjassot, València, Spain

¹⁶ Unidad Asociada CSIC, “Grupo de Astrofísica Extragaláctica y Cosmología,” Instituto de Física de Cantabria—Universitat de València, Spain

¹⁷ Tianjin Normal University, Binshuixidao 393, Xiqing 300387, Tianjin, People's Republic of China

Received 2024 April 26; revised 2024 October 5; accepted 2024 October 22; published 2024 December 2

Abstract

We present rest-frame optical spectra from Keck/MOSFIRE and Keck/NIRES of 16 candidate ultramassive galaxies targeted as part of the Massive Ancient Galaxies at $z > 3$ Near-Infrared Survey (MAGAZ3NE). These candidates were selected to have photometric redshifts $3 \lesssim z_{\text{phot}} < 4$, photometric stellar masses $\log(M_*/M_\odot) > 11.7$, and well-sampled photometric spectral energy distributions (SEDs) from the UltraVISTA and VIDEO surveys. In contrast to previous spectroscopic observations of blue star-forming and poststarburst ultramassive galaxies, candidates in this sample have very red SEDs implying significant dust attenuation, old stellar ages, and/or active galactic nuclei (AGN). Of these galaxies, eight are revealed to be heavily dust-obscured $2.0 < z < 2.7$ galaxies with strong emission lines, some showing broad features indicative of AGN, three are Type I AGN hosts at $z > 3$, one is a $z \sim 1.2$ dusty galaxy, and four galaxies do not have a confirmed spectroscopic redshift. In fact, none of the sample has $|z_{\text{spec}} - z_{\text{phot}}| < 0.5$, suggesting difficulties for photometric redshift programs in fitting similarly red SEDs. The prevalence of these red interloper galaxies suggests that the number densities of high-mass galaxies are overestimated at $z \gtrsim 3$ in large photometric surveys, helping to resolve the “impossibly early galaxy problem” and leading to much better agreement with cosmological galaxy simulations. A more complete spectroscopic survey of ultramassive galaxies is required to pin down the uncertainties on their number densities in the early Universe.

Unified Astronomy Thesaurus concepts: Galaxy evolution (594); High-redshift galaxies (734)

1. Introduction

Measuring the change in the number density of galaxies versus stellar mass as a function of cosmic time provides insights into the buildup of stellar mass and, when considering star-forming and quiescent galaxies separately, the timescales over which galaxies quench their star formation. This galaxy stellar mass function (GSMF) has therefore become a standard tool for studying galaxy evolution.

The GSMF has been measured using optical surveys across a range of redshifts from the local Universe (e.g., S. Cole et al. 2001; E. F. Bell et al. 2003) out to $z \sim 1.5$ (e.g., A. Borch et al. 2006; K. Bundy et al. 2006). At these redshifts, the evolution of the GSMF appears to be small. The most massive galaxies are already in place and are typically only growing via minor mergers (i.e., slowly), while star formation rates (SFRs) in less massive galaxies are lower than at earlier epochs (e.g., L. Kawanwanichakij et al. 2020).

Wide and deep near-infrared imaging surveys have allowed for GSMF analyses out to considerably higher redshifts and have also probed down to lower stellar masses (e.g., D. Marchesini et al. 2009; O. Ilbert et al. 2013; A. Muzzin et al. 2013a; A. R. Tomczak et al. 2014; M. Stefanon et al. 2015; I. Davidzon et al. 2017; S. Sherman et al. 2020; D. J. McLeod et al. 2021; Z. C. Marsan et al. 2022; J. R. Weaver et al. 2023). These studies have shown that the GSMF evolves substantially in the first 4 Gyr as galaxies form new stars rapidly and the first galaxies begin quenching.

* The spectra presented herein were obtained at the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W. M. Keck Foundation.



Intriguingly, the number densities of the most massive galaxies ($\log(M_*/M_\odot) \gtrsim 11.5$) do not appear to change much, if at all, after this early epoch (D. Marchesini et al. 2010; L. Kawinwanichakij et al. 2020), implying that nearly all of these objects which exist today have built the majority of their stellar mass in the first ~ 3 Gyr of cosmic time and then ceased stellar mass growth almost completely. The number densities of these massive galaxies at high redshifts are significantly higher than those predicted by most simulations, and are close to, if not in excess of, the limits of hierarchical galaxy assembly and Lambda cold dark matter (Λ CDM; e.g., J. Antwi-Danso et al. 2023; K. Glazebrook et al. 2024). This has been termed the “impossibly early galaxy problem” (C. Steinhardt et al. 2016), though extremely efficient star formation in early, massive halos (e.g., M. Stefanon et al. 2021; J. R. Weaver et al. 2023) or a varying initial mass function (e.g., P. G. van Dokkum & C. Conroy 2012; J. T. Mendel et al. 2020) may help resolve the issue.

Spectroscopic confirmation of these most massive galaxies at early times with accurate measurements of their stellar mass is required to understand whether changes to the current paradigm of galaxy evolution are necessary. In the last decade, samples of ultramassive galaxies (UMGs; $\log(M_*/M_\odot) \gtrsim 11$) at $3 < z < 4$ have been confirmed with optical/near-infrared spectroscopy to the point where there can be no doubt that such galaxies exist in substantial numbers (e.g., Z. C. Marsan et al. 2015; K. Glazebrook et al. 2017; C. Schreiber et al. 2018a; M. Tanaka et al. 2019; F. Valentino et al. 2020; B. Forrest et al. 2020a, 2020b; P. Saracco et al. 2020; J. Antwi-Danso et al. 2023). Several have also been spectroscopically confirmed with longer wavelength observations (e.g., L. H. Jones et al. 2021). Importantly however, the difficulty of confirming these objects means that the brightest candidates are typically selected for spectroscopic follow-up. This biases the sample toward confirmation of blue star-forming and poststarburst galaxies at these epochs, which have rest-frame V-band magnitudes $\sim 10\times$ higher than older quiescent galaxies with similar stellar masses (e.g., G. Bruzual & S. Charlot 2003). Additionally, none of these UMGs confirmed with optical/near-infrared spectroscopy are highly dust-obscured galaxies (DOGs), which photometric surveys suggest contribute significantly to the population of high-mass galaxies at these redshifts (e.g., N. S. Martis et al. 2016; Z. C. Marsan et al. 2022; A. S. Long et al. 2023; though see also J. S. Dunlop et al. 2007). Such galaxies are extremely difficult to confirm spectroscopically in the optical/near-infrared even using the largest ground-based telescopes as dust obscuration reduces the strength of emission lines and absorption features in the rest-frame optical typically used for confirmation.

We have undertaken an ambitious campaign to spectroscopically confirm candidate UMGs with $\log(M_*/M_\odot) > 11.7$ and $z > 3$ (herein termed superultramassive galaxies (S-UMGs)) with Keck/MOSFIRE (I. S. McLean et al. 2010, 2012), many of which the photometry suggests may be dust obscured. In this work we describe the photometric sample, target selection, spectroscopic observations, and data reduction in Section 2, and analysis of the spectroscopy in Section 3. This is followed by a comparison of S-UMG candidate objects in the COSMOS field from different catalogs in Section 4 and discussion of their number densities in Section 5. Throughout this work we assume a Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$. We also adopt the AB magnitude system (J. B. Oke & J. E. Gunn 1983).

2. Data

2.1. Photometric Catalogs

This work builds on previous observations from the Massive Ancient Galaxies at $z > 3$ Near-Infrared (MAGAZ3NE) survey, which spectroscopically confirmed the largest sample to date of UMGs at $3 < z < 4$ using Keck/MOSFIRE (B. Forrest et al. 2020b). Previous targets were almost entirely either blue star-forming or poststarburst galaxies and showed excellent agreement between photometric and spectroscopic redshifts and stellar masses. This population also showed large velocity dispersions (B. Forrest et al. 2022) and the residence of at least some of the UMGs in confirmed overdense structures (L. Shen et al. 2021; I. McConachie et al. 2022; I. McConachie et al. 2024, submitted). In this work we target a sample which is considerably redder in color and contains the most massive candidate galaxies at $z > 3$.

As in previous MAGAZ3NE papers, we select targets for spectroscopic follow-up from wide and deep near-infrared selected photometric catalogs. Two catalogs in the COSMOS field (N. Scoville et al. 2007) based on deep near-infrared Y -, J -, H -, and K_s -band data from the UltraVISTA survey (H. J. McCracken et al. 2012) are considered. The first catalog (A. Muzzin et al. 2013b) uses the UltraVISTA Data Release 1 (DR1) over 1.62 deg^2 in addition to ground-based optical data (Subaru/SuprimeCam and CFHT/MegaCam; Y. Taniguchi et al. 2007; P. Capak et al. 2007) and space-based ultraviolet (Galaxy Evolution Explorer; D. C. Martin et al. 2005) and near-infrared (Spitzer/IRAC and Spitzer/MIPS; D. B. Sanders et al. 2007) data. Photometry is based on a K_s -band detection image with a 90% completeness of $K_s = 23.4 \text{ mag}$ and 30 photometric bandpasses. The second catalog (Z. C. Marsan et al. 2022; A. Muzzin 2024, private communication) is based on UltraVISTA Data Release 3 (DR3), which includes deeper observations ($K_s = 24.5 \text{ mag}$) in strips totaling 0.84 deg^2 . Additional observations from Spitzer/IRAC (V. Mehta et al. 2018; K. I. Caputi et al. 2017; M. L. N. Ashby et al. 2018) increased the depth in the 3.6 and $4.5 \mu\text{m}$ bands by $>1 \text{ mag}$, while additional data from Subaru/Hyper SuprimeCam (H. Aihara et al. 2019), CFHTLS-Deep, and the NEWFIRM Medium Band Survey (K. E. Whitaker et al. 2011) were also added for up to 49 photometric bandpasses per object.

Two catalogs based on Z -, Y -, J -, H -, and K_s -band imaging from the VISTA VIDEO survey (M. J. Jarvis et al. 2013) are also used (M. Annunziatella, private communication). These catalogs, one in the XMM-LSS (XMM) field and one in the Chandra Deep Field South (CDFS), were constructed following a similar process to the UltraVISTA catalogs. The VIDEO observations were combined with observations from the Dark Energy Survey (T. M. C. Abbott et al. 2018), CFHT-Deep/Wide Legacy Survey (S. D. Gwyn 2012), VOICE Survey (M. Vaccari et al. 2016), and Hyper-SuprimeCam (M. Hayashi et al. 2018). Deep IRAC observations from the SERVS (J.-C. Mauduit et al. 2012) and DeepDrill (M. Lacy et al. 2021) surveys were also included for a total of 22 photometric bandpasses. The XMM catalog covers 5.1 deg^2 while the CDFS catalog covers 3.3 deg^2 . Both have K_s -band detection images with 90% completeness limits of $K_s \sim 23.0 \text{ mag}$.

Galaxies in all catalogs have photometric redshift probability distributions from EAzy (G. B. Brammer et al. 2008) and stellar population properties derived using FAST (M. Kriek et al. 2009) assuming a G. Chabrier (2003) initial mass function.

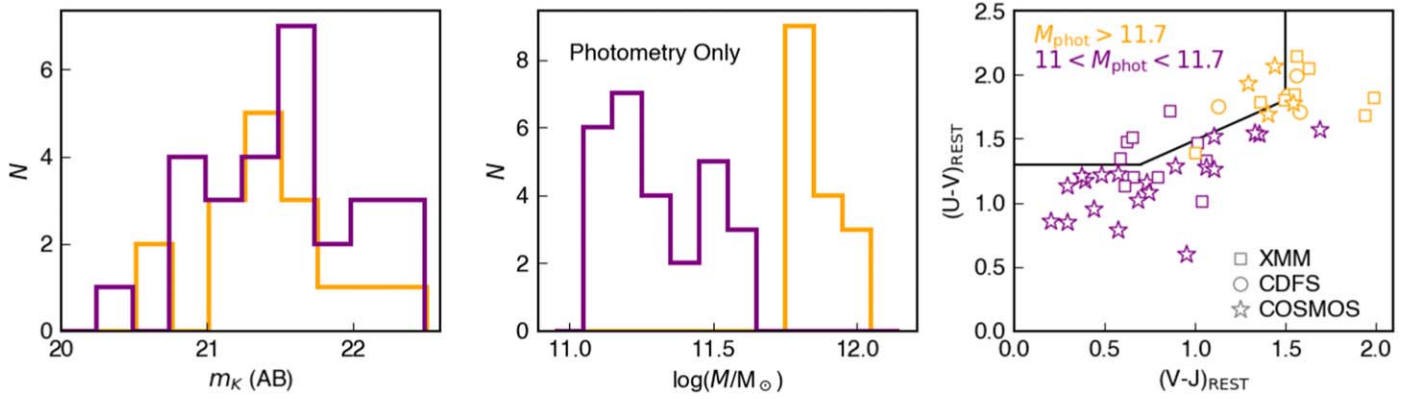


Figure 1. The K -band magnitudes (left), stellar masses derived from photometry (center), and UVJ rest-frame colors derived from photometry (right) of the MAGAZ3NE UMGs ($11.0 < \log(M_*/M_\odot) < 11.7$, purple) and S-UMGs ($\log(M_*/M_\odot) > 11.7$, orange) candidates. The shape of a symbol indicates the field, either VIDEO-XMM (square), VIDEO-CDFS (circle), or COSMOS-UltraVISTA (star).

2.2. Target Selection

We apply a photometric stellar mass cut of $\log(M_*/M_\odot) > 11.7$ and a photometric redshift cut of $3 < z_{\text{phot}} < 4$ to initially select galaxies in these catalogs. Each galaxy is visually inspected in an attempt to verify that it is a single object without contamination from a neighbor nearby in projection and that the best-fit spectral energy distribution (SED) model is a reasonable fit to the photometry. While a magnitude cut was not explicitly made for target selection, brighter targets generally had better fit SEDs and were more likely to be observed. Only two observed candidates, both in COSMOS, had K -band magnitudes fainter than the cut of $K_s < 21.7$ used in previous MAGAZ3NE works.

This results in a sample of 34 high-quality targets of which 16 were spectroscopically observed—8/16 in XMM, 4/8 in CDFS, and 4/10 in COSMOS. A comparison of the stellar masses and rest-frame colors from this sample and those targets from previous MAGAZ3NE work is shown in Figure 1.

2.3. Spectroscopy

2.3.1. Observations

Fifteen targets were followed up using the MOSFIRE instrument on the Keck I telescope. These were initially targeted in the K band, where the potentially strong [O III] $\lambda\lambda 4959, 5007$ and $H\beta$ emission lines lie across $3 < z < 3.8$. As with previous MAGAZ3NE follow-up, on-the-fly data reduction was performed to look for emission lines during the course of observations. If a redshift was considered confirmed, we moved on to the next target. As a result, the exposure times on each galaxy are nonuniform. When a target showed a single emission line in the K band whose nature could not reliably be determined, or when only continuum was detected, we followed up in H band.

We also targeted one S-UMG candidate (XMM-VID1-2433) with the NIRES instrument (J. C. Wilson et al. 2004) on Keck II. While NIRES is less sensitive than MOSFIRE in the K band and targets a single object at a time, it has the advantage of obtaining data in the Y, J, H , and K bandpasses simultaneously. The exposure times and conditions for S-UMG candidates are summarized in Table 1. A more complete table of MAGAZ3NE targets is provided in the Appendix.

2.3.2. Data Reduction

In contrast to previous MAGAZ3NE papers, which used the Keck-supported Data Reduction Pipeline (DRP) to reduce the

Table 1
Spectral Observations of Superluminous Galaxy Candidates

Galaxy	m_K	$t_{\text{exp},K} (m)$	$t_{\text{exp},H} (m)$	Avg. Seeing K''
COS-DR1-70455	21.85	312	...	0.58
COS-DR1-79837	21.10	120	216	0.74
COS-DR1-209435	22.03	252	...	0.68
COS-DR1-233254	21.10	69	...	0.77
CDFS-VID1-2420	22.39	132	...	0.74
CDFS-VID1-3091	20.59	114	40	0.60
CDFS-VID1-3536	21.44	204	...	0.67
CDFS-VID2-684	21.67	210	...	1.0
XMM-VID1-1415	21.34	66	...	0.70
XMM-VID1-2433*	20.66	260	=(260)	0.80
XMM-VID2-827	21.41	72	...	0.66
XMM-VID2-2379	21.71	96	...	0.64
XMM-VID3-657	21.66	360	64	0.69
XMM-VID3-1787	21.15	114	...	0.60
XMM-VID3-3517	21.42	296	...	0.85
XMM-VID3-3941	21.33	180	88	1.1

Note. All targets were observed with Keck/MOSFIRE with the exception of XMM-VID1-2433, which was observed with Keck/NIRES.

data, for this work we reduced the data using the MOSDEF 2D data reduction pipeline (M. Kriek et al. 2015, 2024). The main advantage of this reduction method over the DRP is the use of a bright star on a science slit to account for slight pointing offsets between exposures and weight individual frames by throughput and seeing, which results in a higher signal-to-noise ratio (SNR) 2D spectrum. We also find a slightly more accurate accounting of spectral errors. Previous MAGAZ3NE MOSFIRE data, which were initially reduced using the MOSFIRE DRP, have been rereduced using the MOSDEF pipeline as well. We note that in no case does this choice of reduction software significantly affect the redshift of an UMG in our sample.

Data from Keck/NIRES were reduced using v1.14 of Pyepit (J. Prochaska et al. 2020) and a telluric correction was calculated using Molecfit (W. Kausch et al. 2015; A. Smette et al. 2015).

A custom Python code was used to perform optimal extraction of the targets (K. Horne 1986). When one or more emission lines were present, the shape of the optimal aperture was calculated around the strongest emission line.

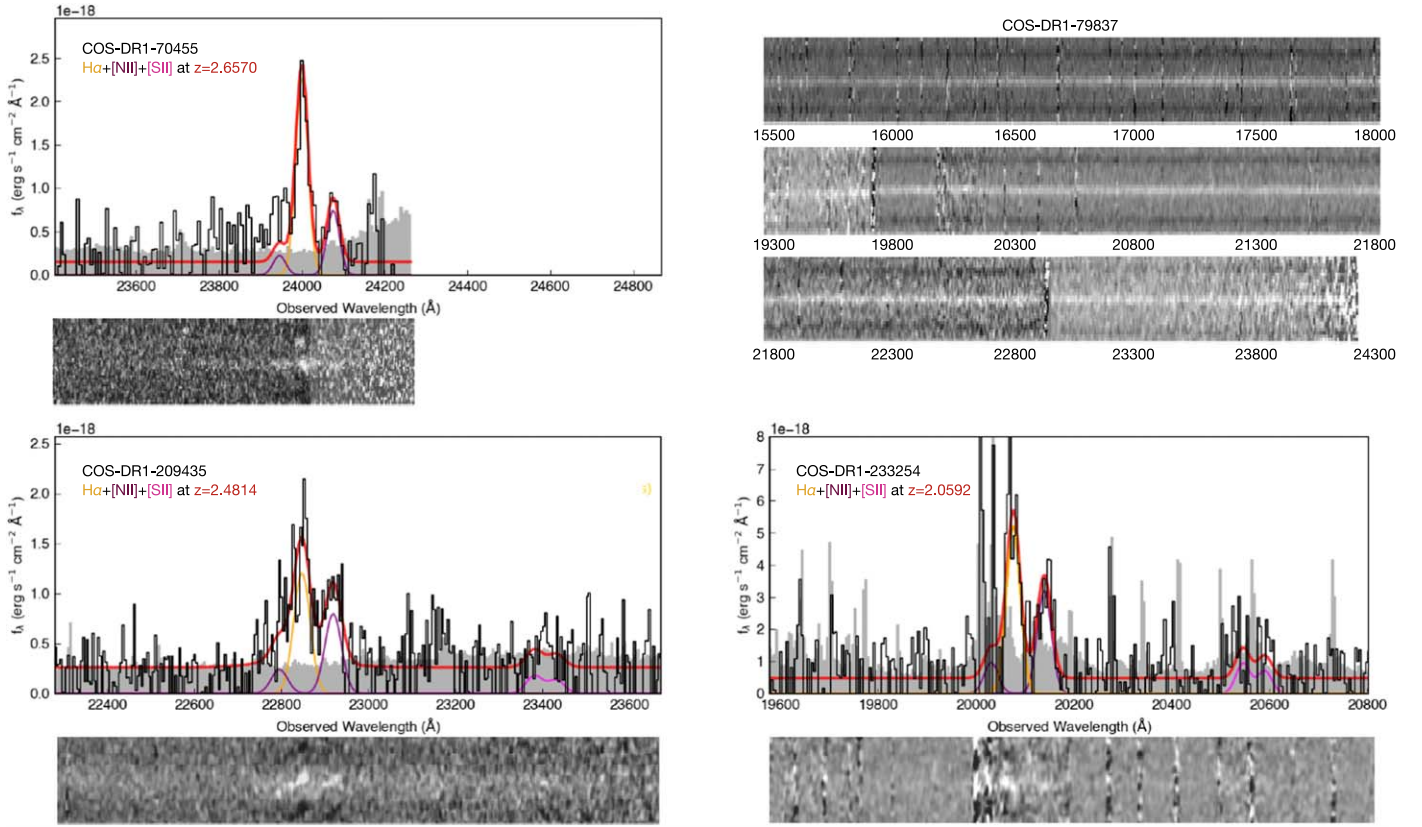


Figure 2. Line complex fitting for obtaining spectroscopic redshifts of candidate S-UMGs in the COSMOS field. For objects with confirmed redshifts, the regions of the 1D and 2D spectra with the confirming lines and fitted model are shown. The overall model is shown in red, while individual line components are shown in different colors ($H\alpha$ in orange, $[N II] \lambda\lambda 6548, 6584$ in purple, and $[S II] \lambda\lambda 6718, 6733$ in magenta). For objects without confirmed redshifts (e.g., COS-DR1-79837, top right), the entire wavelength range probed by H - and K -band spectroscopy is shown. Some targets were observed with multiple masks which were sensitive to different wavelengths; in such cases regions covered by fewer than the total number of masks are displayed as a lighter shade of gray. Similar plots for the targeted S-UMG candidates in the VIDEO fields are in the [Appendix](#).

3. Analysis of Spectra

3.1. Spectroscopic Redshifts

Of the 16 observed candidate S-UMGs, 12 show obvious emission lines; eight of these galaxies clearly display the $H\alpha + [N II] \lambda\lambda 6548, 6584$ lines (and therefore at $2.0 \lesssim z \lesssim 2.7$), while three show the $[O III] \lambda\lambda 4959, 5007 + H\beta$ lines at $z > 3$, and one target exhibits $[S III] \lambda 9531$ and $[C I] \lambda 9850$ at $z \sim 1.2$. These are fit with a multi-Gaussian model, with the $[N II] \lambda\lambda 6548, 6584$ doublet, $[S II] \lambda\lambda 6718, 6733$ doublet, and $[O III] \lambda\lambda 4959, 5007$ doublet line ratios fixed. The velocity width for the narrow component of all lines is fixed, but an additional broad-line component is also allowed. Figure 2 shows observed spectra and the model of the emission features in the COSMOS targets, which leave no uncertainty as to the galaxy redshift modulo small observational uncertainties (XMM and CDFS targets are similarly displayed in the [Appendix](#)).

The spectra of three of the remaining candidate S-UMGs (XMM-VID3-657, XMM-VID3-1787, and COS-DR1-79837) have clear continuum detections but no strong emission features or absorption features that could be used to determine a redshift, while the final candidate, XMM-VID1-1415, has only a very faint continuum detection. This implies that either the redshift of each galaxy is such that any strong emission lines fall out of the wavelength range to which the spectroscopic observations are sensitive, or that the galaxies do not have strong observable emission lines, either due to passive nature or high levels of dust obscuration. Based on the

photometric redshift probability distributions, the probability of these galaxies being located at a redshift where strong emission lines would not be observable for these targets is $\lesssim 10\%$. Fits to the combined photometry and spectroscopy suggest this may in fact be the case for both XMM-VID3-1787 and XMM-VID1-1415. XMM-VID3-657 has a low SNR detection at $\sim 2.36 \mu m$, which may correspond to $H\beta$ at $z = 3.85$, while a continuum break may be present in COS-DR1-79837. Deeper near-infrared spectroscopic observations from the ground are unlikely to yield additional information, though observations with JWST would likely result in confirmed redshifts.

3.2. Remodeling the Stellar Populations

The UltraVISTA and VIDEO catalogs use FAST to estimate stellar masses from photometry, which does not natively include emission lines in the models. While it has been known for some time that emission lines can contribute a significant fraction of flux in broadband measurements (e.g., M. Kriek et al. 2011; D. P. Stark et al. 2013; B. Salmon et al. 2015; R. Smit et al. 2015; B. Forrest et al. 2018), particularly for low-mass, blue, star-forming galaxies, recent results with JWST have shown that this is quite common at high redshifts (e.g., S. Withers et al. 2023; G. T. E. Sarrouh et al. 2024). Indeed, in this work we also see significant line-flux contribution in very red, massive galaxies.

Using the spectroscopic redshift, we remodel the stellar populations of the galaxy sample using FAST++ (C. Schreiber et al. 2018b) and employing the G. Bruzual & S. Charlot (2003)

stellar population synthesis models. FAST++ has the advantage of allowing for flexible star formation history parameterizations and including both photometric and spectroscopic data, though models still do not include emission lines. We therefore perform a correction of the photometry due to observed emission features as in B. Forrest et al. (2020b) with the line models described above and mask the emission line wavelengths in the spectra, but note that the possibility of other strong lines at spectrally unobserved wavelengths may additionally affect the modeling. The percentage contribution of flux from emission lines in the K band for the S-UMG candidate sample is $12.2 \pm 3.5\%$. The best-fit models overlaid on the observed photometry and spectra for objects with confirmed spectroscopic redshifts are shown in Figure 3, while those without confirmed spectroscopic redshifts are shown in Figure 4. While the reduced χ^2 value of the fit to the photometry alone is typically better than that of the fit to the combined photometry and spectroscopy, with $\Delta\chi^2 = 0.3$, this is partially due to the inclusion of the spectroscopic data points. The best-fit models to the photometry alone do not match the observed spectra.

As the z_{spec} values are more than 2σ discrepant from the z_{phot} predictions, other derived properties for these galaxies can also be significantly different from the photometric predictions. While the masses of all S-UMG candidates with spectroscopic redshifts decrease from the photometric estimate, the galaxies are certainly still very massive, with all but the $z \sim 1.2$ source having $\log(M_*/M_\odot) > 10.7$. Of note, these galaxies all have considerable amounts of dust, $A_V \sim 3$ mag. The derived properties from these fits are given in Table 2 and Figure 5.

Between their large amounts of dust and strong emission lines, the $z \sim 2$ objects are likely dusty star-forming galaxies and perhaps submillimeter galaxies, which photometric redshift codes can have difficulty accurately identifying (e.g., K. I. Caputi et al. 2012; S. Jin et al. 2024). Indeed one confirmed object, COS-DR1-209435, has a very strong $870 \mu\text{m}$ detection (J. M. Simpson et al. 2020). When far-infrared data are available, fitting with photometric redshift and SED modeling programs which include such information can lead to better constraints on galaxy properties (E. Da Cunha et al. 2015; U. Dudzeviciute et al. 2020; W. Chang et al. 2024, in preparation). Some of these objects may also be consistent with the hot DOG population (P. R. Eisenhardt et al. 2012; J. Wu et al. 2012; L. Finnerty et al. 2020). Objects in that sample are identified from Wide-field Infrared Survey Explorer colors, and often do not have extensive photometric observations similar to those in this work, which prevents a rigorous comparison of other parameters such as stellar mass or SFR.

The broad lines seen in five of the spectra, including all three confirmed at $z > 3$, indicate active galactic nucleus (AGN) activity. While the removal of this emission line flux from the photometry has reduced the stellar mass estimate, these galaxies all have stellar masses $\log(M_*/M_\odot) > 11.5$, with CDFS-VID1-2420 having an incredible stellar mass of $\log(M_*/M_\odot) \sim 11.9$. However, while the emission line flux has been removed, the potential exists for significant AGN continuum flux contribution to the photometry, which could artificially inflate the stellar mass (e.g., L. J. Furtak et al. 2024). The galaxies whose spectra show broad lines do appear compact in ground-based imaging, though several other objects with confirmed redshifts and no evidence for broad lines have similar morphologies, consistent with small sizes.

In an effort to account for this effect, we fit the combined photometric and spectroscopic data using BAGPIPES (A. C. Carnall et al. 2018, 2019), which includes functionality for fitting the AGN

continuum contribution. We also use a double power-law star formation history, a D. Calzetti et al. (2000) dust law, and nebular emission lines, and allow corrections for spectral calibration and noise underestimation issues. The fits from BAGPIPES are in some cases marginally better fits to the data compared to those from FAST++, but they retain the high stellar masses and dust attenuation levels, despite estimating the contribution of AGN continuum and emission line flux. As such we have confidence that these galaxies are indeed impressively massive, DOGs, though at lower redshifts than initially predicted from the photometry alone.

4. Comparison of Photometric Candidate Properties in COSMOS

Modeling of galaxy SEDs to obtain stellar masses requires both sufficient photometric sampling of the SED and flexible, accurate models to fit the observations. Extragalactic fields such as COSMOS, GOODS-S (CDFS), and UDS (XMM), among others, have sufficient multiwavelength observations from ultraviolet to near-infrared wavelengths to constrain redshifts and stellar masses well for most galaxies, though more bandpasses, and in particular medium and narrow bandpasses can improve the uncertainties on these measurements (e.g., K. E. Whitaker et al. 2011; D. Sobral et al. 2013; T. L. Suzuki et al. 2015; C. M. S. Straatman et al. 2016; G. T. E. Sarrouh et al. 2024). Also critical to accurate photometric redshifts and stellar masses is the modeling of strong emission lines from star formation and/or AGN (e.g., D. P. Stark et al. 2013; R. Smit et al. 2014; B. Salmon et al. 2015; B. Forrest et al. 2017; S. Withers et al. 2023).

Particularly at higher redshifts, near-infrared photometry is required to probe redward of the $D_n(4000)$ feature and constrain the mass. Deep IRAC data have historically been necessary to avoid the potential of a single line in the K band leading to overestimates in stellar mass. While the improved sensitivity and spatial resolution of JWST may eventually preclude the need for Spitzer data to constrain the near-infrared in some fields, JWST's smaller field of view means that deep IRAC data will still be a critical part of large area catalogs for the foreseeable future.

The catalogs used for target selection in this work satisfy these requirements and show excellent agreement between photometric and spectroscopic redshifts for galaxy populations as a whole. Additionally, they have reported completeness limits ($\geq 80\%$) down to at least 1 mag fainter in K_s than the faintest target observed in this program, yet a significant fraction of the S-UMG candidates do not have redshifts and masses consistent with the photometric estimates. In what follows we attempt to uncover the root of this issue by comparing objects in the COSMOS field, for which numerous deep photometric catalogs have been constructed.

4.1. Catalogs Considered

The stellar masses of the galaxies in question determined from modeling the photometry alone are exceptionally high. In order to confirm that this is not a result of the particular methods used in the generation of the UltraVISTA catalogs, we compare results between the UltraVISTA DR1 (A. Muzzin et al. 2013b), UltraVISTA DR3 (A. Muzzin 2024, private communication), COSMOS2015 (C. Laigle et al. 2016), and both sets of COSMOS2020 (CLASSIC and FARMER; J. R. Weaver et al. 2022) photometric catalogs (five in total). This allows for analysis

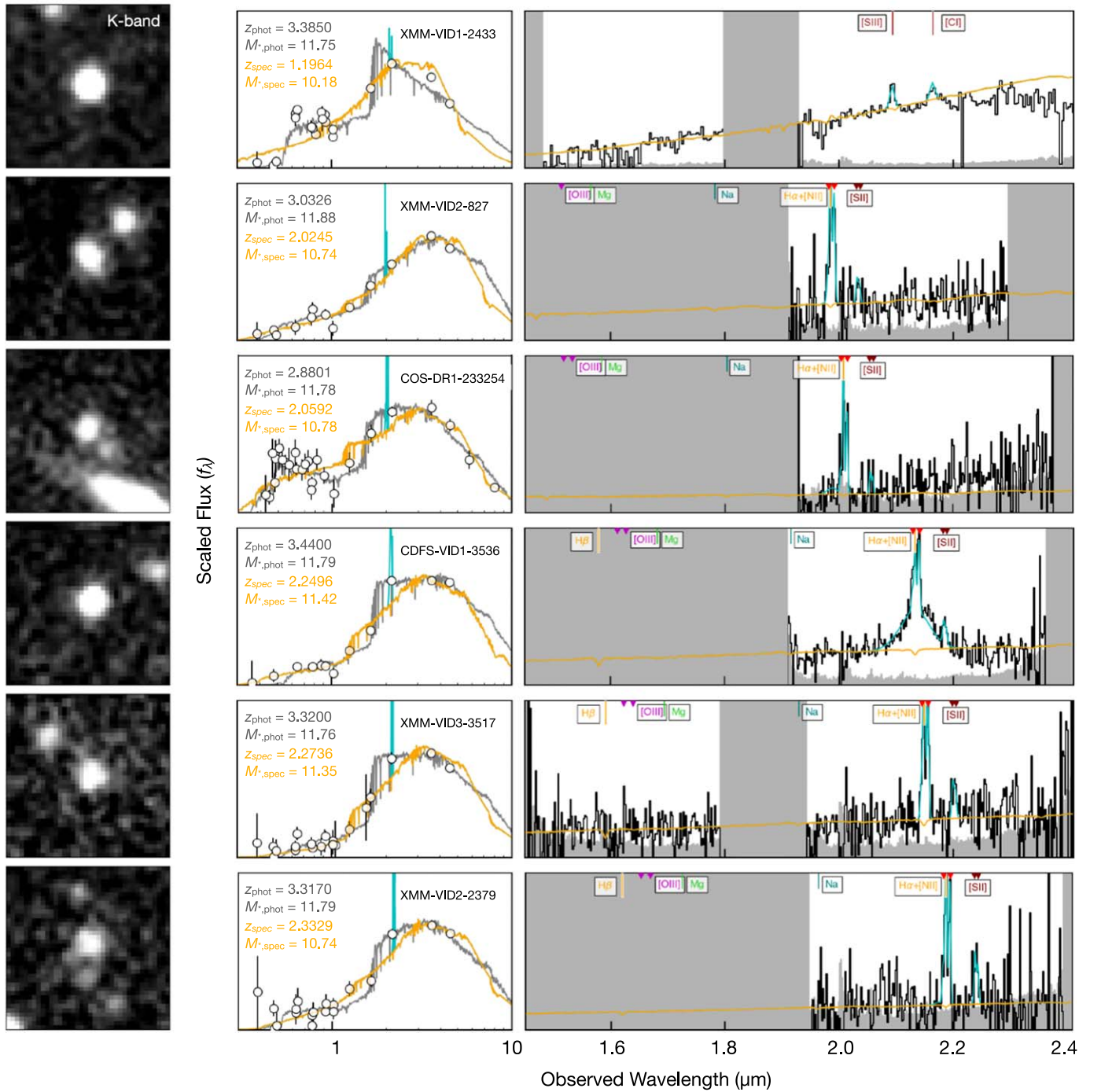


Figure 3. S-UMG candidates with confirmed spectroscopic redshifts, in order of increasing z_{spec} . Left: K-band imaging cutouts of the target, $7''$ on a side. Center: the observed photometry of the galaxy (open circles) with best-fit models to the photometry alone (gray line) and photometric and spectroscopic data (gold line with cyan emission lines). Right: the observed spectrum binned in ~ 19.5 Å bins (12 pixels in H band, 9 pixels in K band) is shown in black, with uncertainties in gray. Completely gray regions were not observed. The emission line model is again cyan added to the best-fit continuum model in orange. Note that the vertical ranges in the center and right panels are not the same.

of differences in galaxy characterization due to different imaging bandpasses and depths (different publications), different methods of photometric extraction using the same images (COSMOS2020 CLASSIC versus FARMER), and different SED modeling codes on the same photometry (COSMOS2020 EAZY-py and LePhare characterizations).

An in depth discussion of the differences of the methodologies involved in the construction of these catalogs is beyond the scope of this work, and we refer the interested reader to the

relevant publications for more details. Briefly however, the UltraVISTA DR1 and UltraVISTA DR3 catalogs use SExtractor (E. Bertin & S. Arnouts 1996) on deep K -band images to identify sources and measure photometry and MOPHONGO (I. Labbe et al. 2006, 2010) to deblend IRAC photometry based on the K -band image, assuming no color gradients from the K band to the Spitzer bandpasses. The COSMOS2015 and COSMOS2020 CLASSIC catalogs use SExtractor on a combined $zYJHKs$ and $izYJHKs$ detection

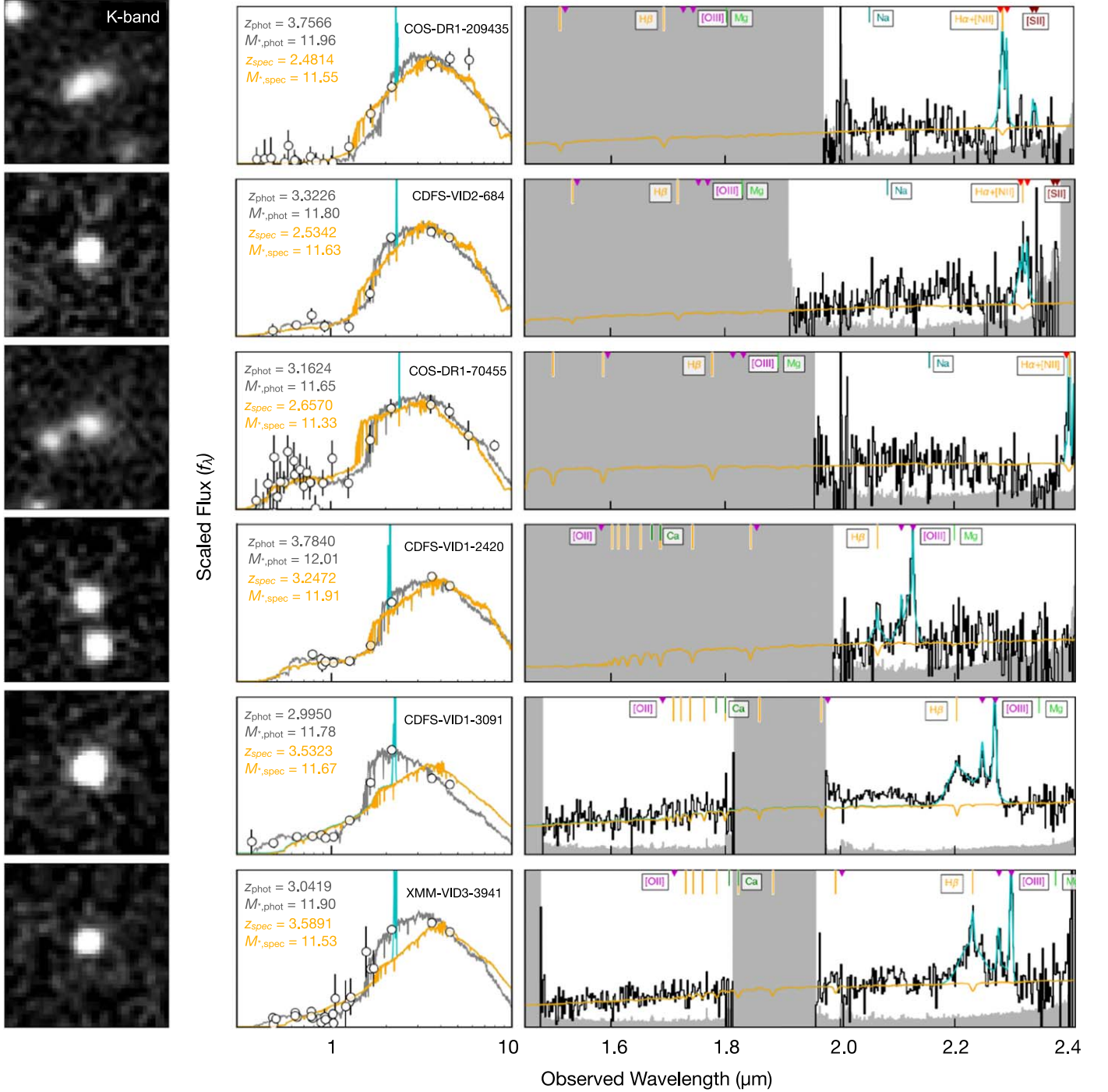


Figure 3. (Continued.)

image, respectively, for the source detection and IRACLEAN (B.-C. Hsieh et al. 2012) to deblend IRAC photometry. The COSMOS2020 FARMER catalog uses The Tractor (D. Lang et al. 2016) to extract photometry from images in different bandpasses simultaneously. Differences in detection image bandpass and depth, as well as the specifics of source detection inputs, can influence not only whether a source is detected or not, but also the total integrated magnitude of the source in each bandpass.

Photometric redshifts for the UltraVISTA DR1 and UltraVISTA DR3 catalogs are derived using EAzy (G. B. Brammer et al. 2008)

and stellar population parameters are derived using FAST (M. Kriek et al. 2009). One set of photometric redshifts in both the COSMOS2020 CLASSIC and FARMER catalogs are derived using the similar EAzy-py code (G. Brammer 2021), which also models stellar population parameters. COSMOS2015 and a second set of photometric redshifts and stellar population parameters in both COSMOS2020 catalogs are derived using LePhare (S. Arnouts et al. 1999; O. Ilbert et al. 2006), incorporating the BC03 models (G. Bruzual & S. Charlot 2003). As a result, for an object detected in all five catalogs, there are seven characterizations of its redshift and stellar populations.

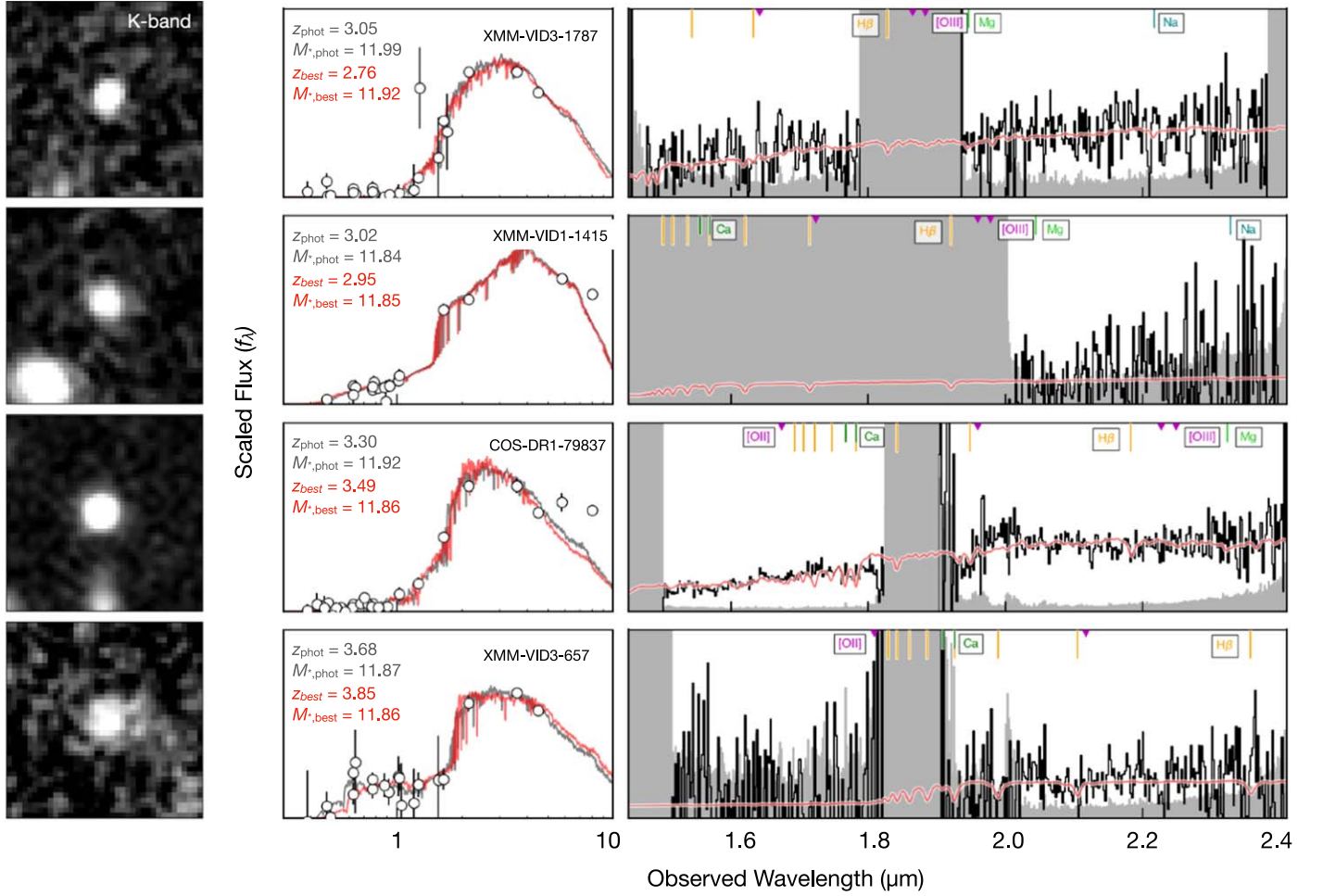


Figure 4. Same as Figure 4 but for objects without confirmed spectroscopic redshifts. Best-fit models to the combined photometry and spectroscopy are shown in red.

4.2. Observed S-UMG Candidates

In addition to many of the UMGs confirmed in B. Forrest et al. (2020b), four of the S-UMG candidates which were targeted spectroscopically are in the COSMOS field, and all were selected from the UltraVISTA DR1 catalog (they are covered by the ultra-deep stripes in the UltraVISTA DR3 catalog). We search for the counterparts of these objects in the COSMOS2015 and COSMOS2020 catalogs using a projected radius of $0''.5$ and, in the case of multiple such matches, select the neighbor within that radius with the most similar K -band magnitude. The four S-UMG candidates are detected in the COSMOS2015 and COSMOS2020 CLASSIC catalogs, while only two of them (COS-DR1-79837 and COS-DR1-209435) are detected in the COSMOS2020 FARMER catalog (COS-DR1-70455 and COS-DR1-233254 are not). The associated derived properties for these objects in the various catalogs are given in Table 3.

The FARMER catalog does not detect COS-DR1-70455 and COS-DR1-233254 due to their proximity to bright stars, the areas around which are masked more conservatively than in other catalogs. It is notable that these two objects show considerably more variance in their reported K -band magnitudes between catalogs. While COS-DR1-79837 and COS-DR1-209435 have a range of $\Delta m_K \lesssim 0.1$ mag between catalogs, COS-DR1-70455 and COS-DR1-233254 have ranges of $\Delta m_K = 0.3$ and 1.6 mag, respectively. This suggests that

differences in accounting for flux from nearby stars is contributing to variance in the extracted magnitudes, though in the case of the COS-DR1-233254 several somewhat close neighbors are visible in imaging, which may also have been subsumed into the flux totals for the counterpart of this object in some catalogs—reproducing and analyzing this aspect of source detection is beyond the scope of this work. Both characterizations of COS-DR1-79837 and COS-DR1-209435 in the FARMER catalog yield significantly lower redshifts (and therefore lower stellar mass estimates) than the other parameterizations, and in the case of COS-DR1-209435, considerably lower than the z_{spec} .

While some of these objects have redshifts and stellar population parameters derived from the photometry which are similar to those derived once the redshift is fixed to the spectroscopic redshift, these values are more often discrepant, and no catalog produces consistently better agreement than the others. Given that in general the photometry for these objects is roughly consistent between catalogs, this suggests that the main difficulty with these objects is obtaining an accurate photometric redshift, which then propagates into determinations of stellar mass, etc. Of course, a small sample size is considered here, and photometric redshift codes may not be optimized for describing rare objects accurately when generating large catalogs. Also, we note that these large discrepancies were not seen for the lower-mass (and generally bluer) UMGs described in B. Forrest et al. (2020b).

Table 2
Properties of the 16 Spectroscopically Targeted Candidate Superultramassive Galaxies

Galaxy	z_{phot} z_{spec}	$M_{*,\text{phot}}$ $M_{*,\text{spec}}$ $\log(M_{\odot})$	SFR_{phot} SFR_{spec} $\log(M_{\odot} \text{ yr}^{-1})$	$A_{V,\text{phot}}$ $A_{V,\text{spec}}$ mag	$t_{50,\text{phot}}$ $t_{50,\text{spec}}$ $\log(\text{yr})$	$t_{q,\text{phot}}$ $t_{q,\text{spec}}$ $\log(\text{yr})$
XMM-VID1-1415	$3.02^{+0.09}_{-0.10}$...	$11.84^{+0.11}_{-0.08}$...	$2.56^{+0.29}_{-1.34}$...	$2.8^{+0.2}_{-0.1}$...	$8.13^{+0.33}_{-0.17}$...	$7.69^{+0.20}_{-7.69}$...
XMM-VID1-2433	$3.38^{+0.04}_{-0.04}$ 1.1964	$11.75^{+0.03}_{-0.05}$ $10.18^{+0.05}_{-0.02}$	$1.52^{+0.28}_{-0.25}$ $2.46^{+0.04}_{-0.08}$	$1.3^{+0.2}_{-0.2}$ $3.4^{+0.1}_{-0.0}$	$8.36^{+0.04}_{-0.14}$ $7.24^{+0.08}_{-0.02}$	$8.08^{+0.01}_{-0.09}$ SF
XMM-VID2-0827	$3.03^{+0.22}_{-0.21}$ 2.0245	$11.88^{+0.10}_{-0.58}$ $10.74^{+0.13}_{-0.06}$	$2.98^{+0.18}_{-0.52}$ $2.89^{+0.51}_{-0.24}$	$2.5^{+0.7}_{-0.2}$ $3.4^{+0.3}_{-0.1}$	$8.81^{+0.08}_{-1.11}$ $7.32^{+0.20}_{-0.29}$	SF SF
XMM-VID2-2379	$3.32^{+0.25}_{-0.25}$ 2.3329	$11.79^{+0.01}_{-0.39}$ $10.74^{+0.68}_{-0.04}$	$2.01^{+2.57}_{-0.35}$ $3.74^{+0.13}_{-1.61}$	$1.7^{+0.7}_{-0.2}$ $4.0^{+0.1}_{-1.2}$	$8.90^{+0.16}_{-2.31}$ $6.73^{+2.19}_{-0.05}$	$8.01^{+0.36}_{-8.01}$ SF
XMM-VID3-0657	$3.68^{+0.44}_{-0.32}$...	$11.87^{+0.04}_{-0.10}$...	$1.76^{+0.08}_{-0.14}$...	$1.1^{+0.6}_{-0.3}$...	$9.07^{+0.03}_{-0.11}$...	$8.47^{+0.12}_{-0.36}$...
XMM-VID3-1787	$3.05^{+0.07}_{-0.39}$...	$11.99^{+0.00}_{-0.11}$...	<0 ...	$1.0^{+0.4}_{-0.1}$...	$9.25^{+0.01}_{-0.06}$...	$9.18^{+0.06}_{-0.02}$...
XMM-VID3-3517	$3.27^{+0.18}_{-0.19}$ 2.2736	$11.76^{+0.05}_{-0.10}$ $11.35^{+0.02}_{-0.04}$	<0 $0.70^{+0.78}_{-0.07}$	$2.0^{+0.5}_{-0.2}$ $3.3^{+0.1}_{-0.0}$	$8.39^{+0.07}_{-0.24}$ $7.98^{+0.04}_{-0.06}$	$8.25^{+0.05}_{-0.27}$ $7.78^{+0.01}_{-0.10}$
XMM-VID3-3941*	$3.04^{+0.12}_{-0.11}$ 3.5891	$11.90^{+0.01}_{-0.10}$ $11.53^{+0.02}_{-0.01}$	$1.50^{+0.04}_{-0.31}$ $4.75^{+0.08}_{-0.03}$	$1.4^{+0.0}_{-0.3}$ $4.0^{+0.0}_{-0.0}$	$9.13^{+0.00}_{-0.10}$ $6.53^{+0.02}_{-0.04}$	$8.69^{+0.06}_{-0.10}$ SF
CDFS-VID1-2420*	$3.78^{+0.21}_{-0.27}$ 3.2472	$12.01^{+0.15}_{-0.05}$ $11.91^{+0.00}_{-0.05}$	$2.14^{+0.70}_{-0.24}$ $2.66^{+0.22}_{-0.18}$	$1.5^{+0.8}_{-0.2}$ $2.3^{+0.2}_{-0.1}$	$9.00^{+0.07}_{-0.13}$ $8.97^{+0.05}_{-0.11}$	$8.11^{+0.36}_{-8.11}$ SF
CDFS-VID1-3091*	$3.00^{+0.17}_{-0.21}$ 3.5323	$11.78^{+0.06}_{-0.04}$ $11.67^{+0.00}_{-0.00}$	$1.26^{+0.19}_{-0.37}$ $4.99^{+0.00}_{-0.02}$	$0.8^{+0.2}_{-0.4}$ $3.8^{+0.0}_{-0.0}$	$9.00^{+0.13}_{-0.03}$ $6.47^{+0.01}_{-0.00}$	$8.65^{+0.13}_{-0.06}$ SF
CDFS-VID1-3536*	$3.44^{+0.19}_{-0.40}$ 2.2496	$11.79^{+0.08}_{-0.24}$ $11.42^{+0.04}_{-0.08}$	$0.43^{+2.05}_{-1.31}$ $2.53^{+0.18}_{-0.20}$	$2.2^{+1.7}_{-0.2}$ $2.9^{+0.2}_{-0.2}$	$8.24^{+0.68}_{-1.76}$ $8.81^{+0.12}_{-0.26}$	$8.07^{+0.62}_{-8.07}$ SF
CDFS-VID2-0684*	$3.32^{+0.28}_{-0.28}$ 2.5342	$11.80^{+0.11}_{-0.08}$ $11.63^{+0.03}_{-0.03}$	$1.51^{+0.17}_{-0.10}$ $1.89^{+0.26}_{-0.21}$	$1.4^{+0.6}_{-0.4}$ $2.4^{+0.2}_{-0.2}$	$9.03^{+0.10}_{-0.50}$ $9.03^{+0.07}_{-0.10}$	$8.59^{+0.16}_{-0.52}$ SF
COS-DR1-70455	$3.15^{+0.26}_{-0.23}$ 2.6570	$11.75^{+0.03}_{-0.16}$ $11.33^{+0.16}_{-0.16}$	$1.91^{+0.26}_{-0.15}$ $2.28^{+0.30}_{-0.33}$	$1.5^{+0.5}_{-0.3}$ $2.2^{+0.3}_{-0.4}$	$9.03^{+0.07}_{-0.23}$ $8.62^{+0.50}_{-0.49}$	SF SF
COS-DR1-79837	$3.32^{+0.17}_{-0.22}$...	$11.92^{+0.06}_{-0.02}$...	$0.63^{+0.65}_{-0.03}$...	$0.8^{+0.7}_{-0.2}$...	$9.12^{+0.00}_{-0.30}$...	$8.89^{+0.00}_{-0.21}$...
COS-DR1-209435	$3.77^{+0.33}_{-0.37}$ 2.4814	$11.98^{+0.02}_{-0.20}$ $11.55^{+0.04}_{-0.09}$	<0 <1.1	$0.9^{+0.4}_{-0.2}$ $2.9^{+0.1}_{-0.3}$	$9.20^{+0.05}_{-0.21}$ $8.63^{+0.18}_{-0.17}$	$9.19^{+0.05}_{-0.20}$ $8.58^{+0.18}_{-0.30}$
COS-DR1-233254	$2.88^{+0.17}_{-0.10}$ 2.0592	$11.81^{+0.03}_{-0.12}$ $10.78^{+0.08}_{-0.03}$	$2.08^{+0.28}_{-0.13}$ $2.42^{+0.55}_{-0.19}$	$1.3^{+0.5}_{-0.3}$ $2.6^{+0.2}_{-0.1}$	$9.13^{+0.00}_{-0.13}$ $7.53^{+0.14}_{-0.07}$	SF SF

Note. The first row for each object reports values derived from photometry alone (redshift, stellar mass, SFR, dust attenuation, lookback time to when 50% of the stellar mass had formed, and lookback time to when the galaxy had quenched)—“SF” indicates not yet quenched, while the second row shows the same variables recalculated using the spectroscopic redshift and line-flux-corrected photometry. Galaxies marked with an asterisk are those showing broad lines from AGN activity.

4.3. Broader Comparison of (Super)Ultramassive Galaxy Candidates between Catalogs

This raises the question of not only how reliably, but also how consistently, high-mass galaxy candidates at these redshifts can be identified photometrically. To explore this, in each characterization we find all UMG ($\log(M_{*}/M_{\odot}) > 11.0$, $3 < z_{\text{phot}} < 4$) and S-UMG candidates ($\log(M_{*}/M_{\odot}) > 11.7$, $3 < z_{\text{phot}} < 4$) with $m_K < 23.4$, the brightest reported completeness limit of the catalogs (80%). Note that in these cases no further cuts are made for model–SED agreement, neighbor contamination, etc.

In Tables 4 and 5 we show the number of identified UMG and S-UMG candidates in each catalog. We then match these objects to their sources in the other catalogs as above, finding neighbors with coordinates within $0''.5$ of those in the initial catalog and choosing the one which is most similar in K -band magnitude when there are multiple matches. We report the number of these candidates which have a counterpart in at least one other catalog,

and the number which have a counterpart in at least three other catalogs (out of the five considered). This removes the requirement of a target lying in the UltraVISTA DR3 footprint, which covers roughly half the area of the other catalogs. Similarly, we report the number of those counterparts which also have photometric redshifts and stellar masses satisfying either UMG or S-UMG cuts in at least one other stellar population characterization, and the number of counterparts satisfying these cuts in at least five other characterizations (out of the seven considered).

Of all the UMG candidates, 96.9% have counterparts in at least one other catalog and 53.0% are characterized as UMGs in at least one other characterization. For the more restrictive comparison, 84.5% have counterparts in at least three other catalogs, but only 5.0% are characterized as UMGs in at least five other characterizations. Of all the S-UMG candidates, 87.9% have counterparts in at least one other catalog and 24.8% are characterized as UMGs in at least one other characterization. Similarly, 73.0% have counterparts in at least three other catalogs, but none are characterized as UMGs in at

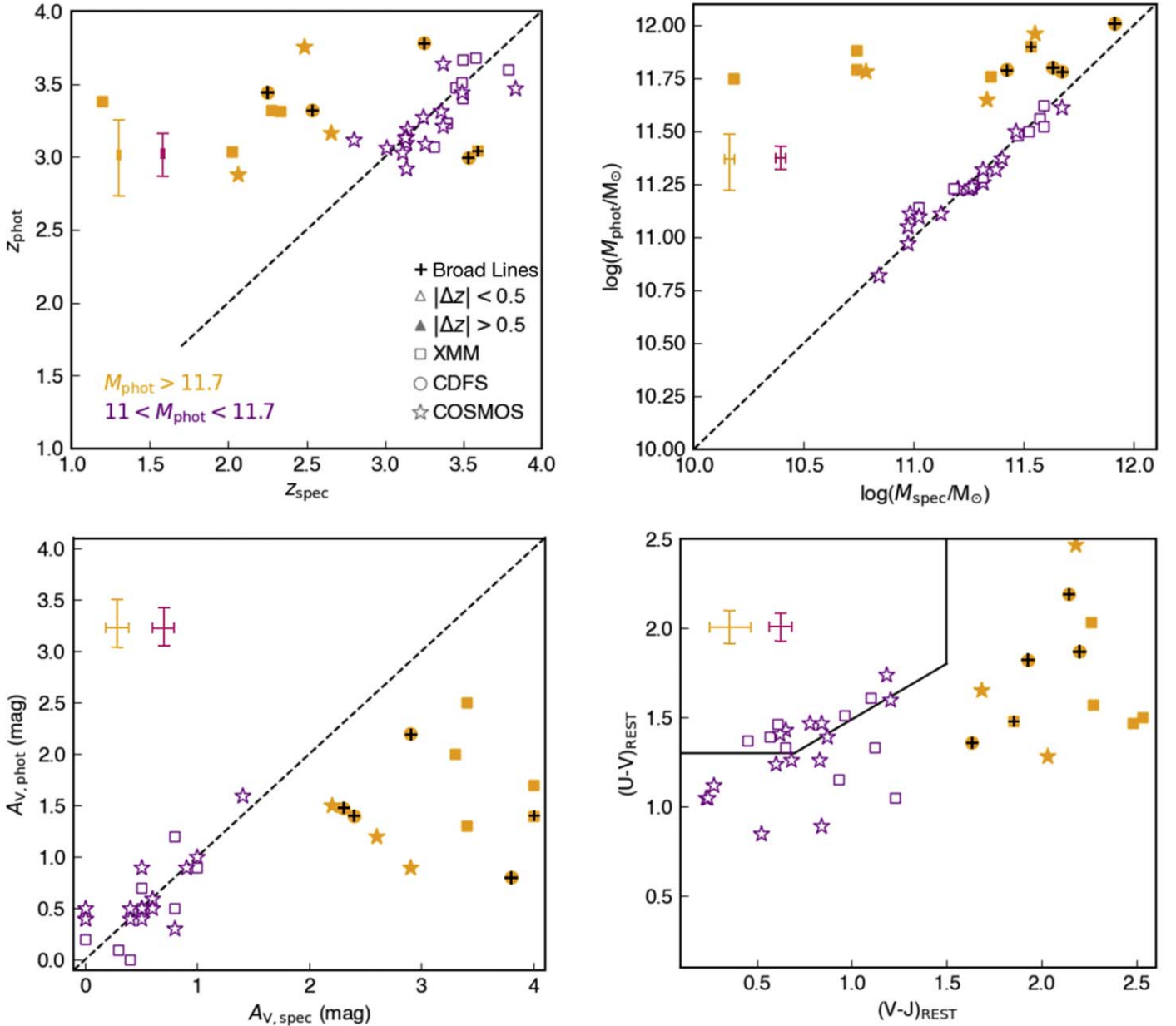


Figure 5. Comparison of galaxy properties with and without spectroscopic information—redshifts (top left), stellar masses (top right), and dust attenuation values (bottom left) are shown. Marker shape indicates the field, color indicates the stellar mass estimated from photometry alone, and a filled symbol represents objects with $|z_{\text{phot}} - z_{\text{spec}}| > 0.5$. A clear discrepancy is seen between the UMG candidates (studied in previous works; purple points), which have generally good agreement, and the S-UMG candidates (the focus of this paper; yellow markers), which typically have different redshifts, lower stellar masses, and greater dust attenuation than predicted from photometry. Spectra exhibiting broad lines from AGN activity are marked with a cross. The bottom right panel shows the spectroscopically derived rest-frame colors for comparison with Figure 1. The redshift interlopers have extremely red rest-frame colors.

least five other characterizations. COSMOS2020 tends to predict more high-mass galaxies than other catalogs, but the COSMOS2020 CLASSIC LePhare characterization yields lower numbers than UltraVISTA DR1 for instance. The fraction of candidates in other catalogs is also higher for COSMOS2020 since both the CLASSIC and FARMER catalogs used the same detection image.

The photometric redshifts, stellar masses, and rest-frame colors of UMG and S-UMG candidates selected from UltraVISTA DR1 and the COSMOS2020 EAZY-py characterizations from the CLASSIC and FARMER catalogs are shown in Figures 6 and 7, along with those of their counterparts in the other characterizations. When UMG and S-UMG candidates from one characterization are not classified as such in another characterization, they generally

have slightly lower photometric redshifts ($2 < z < 3$) and stellar masses ($\log(M_*/M_\odot) > 10.5$), though the rest-frame UVJ colors are often quite red. Notably, the UltraVISTA DR1 EAZY template set did not contain templates as red as the COSMOS2020 EAZY-py template set, limiting derived rest-frame colors to $(V-J) < 2$. We note that for all of the S-UMG candidates spectroscopically observed, even in the most extreme cases of effects that can confound photometric redshift codes such as high dust attenuation and strong emission line contamination, all objects but one with spectroscopic redshifts still have $\log(M_*/M_\odot) > 10.7$ and $z_{\text{spec}} > 2$. There are a handful of $z > 3$ photometric candidates which other characterizations place at $z < 1$ as well, potentially due to the choice of interpretation of a photometric break as either a $z > 3$ Lyman break or a $z < 1$ Balmer break.

Table 3
Comparison of Photometric Properties for Four Targeted Superultramassive Galaxies in the COSMOS Field

Property	Source				
ID	UVDR1	70455	79837	209435	233254
	C15	440292	530898	798016	985682
	C20C	567661	718745	1152024	1450899
	C20F	...	768111	670662	...
$\Delta_{\text{proj,UVDR1}}$ (")	C15	0.10	0.02	0.08	0.31
	C20C	0.18	0.14	0.12	0.32
	C20F	...	0.13	0.05	...
m_K	UVDR1	21.85	21.10	22.03	21.10
	C15	22.13	21.05	21.99	21.76
	C20C	21.94	21.10	21.97	20.14
	C20F	...	21.03	21.92	...
z	UVDR1 EAzY	$3.149^{+0.259}_{-0.230}$	$3.318^{+0.168}_{-0.214}$	$3.769^{+0.332}_{-0.365}$	$2.870^{+0.177}_{-0.152}$
	C15 LePhare	$2.924^{+0.379}_{-0.240}$	$2.783^{+0.118}_{-0.114}$	$3.259^{+0.314}_{-0.139}$	$2.773^{+0.295}_{-0.642}$
	C20C EAzY-py	$3.401^{+0.064}_{-0.070}$	$2.452^{+0.387}_{-0.066}$	$2.439^{+0.207}_{-0.240}$	$11.83^{+0.076}_{-0.116}$
	C20C LePhare	$2.639^{+0.171}_{-0.178}$	$2.883^{+0.123}_{-0.138}$	$3.460^{+0.232}_{-0.298}$	$2.783^{+0.092}_{-0.302}$
	C20F EAzY-py	...	$1.238^{+0.014}_{-0.029}$	$0.050^{+0.084}_{-0.060}$...
	C20F LePhare	...	$1.201^{+0.106}_{-0.136}$	$0.722^{+0.018}_{-0.040}$...
	Spec FAST++	2.6570	...	2.4814	2.0592
	Spec FAST++
$\log(M/M_\odot)$	UVDR1 FAST	11.75	11.99	11.98	11.81
	C15 LePhare	11.32	12.32	11.92	11.60
	C20C EAzY-py	11.41	11.57	11.75	NaN
	C20C LePhare	11.26	12.02	11.95	11.47
	C20F EAzY-py	...	10.09	5.94	...
	C20F LePhare	...	9.75	7.99	...
	Spec FAST++	11.33	...	11.55	10.78
$\log(\text{SFR}/(M_\odot \text{ yr}^{-1}))$	UVDR1 FAST	1.82	1.36	<0	2.08
	C15 LePhare	2.30	2.83	2.45	2.13
	C20C EAzY-py	3.04	3.21	2.94	NaN
	C20C LePhare	1.96	2.43	2.55	<0
	C20F EAzY-py	...	0.12	<0	...
	C20F LePhare	...	0.72	0.18	...
	Spec FAST++	2.28	...	<1.1	2.42
A_V	UVDR1 FAST	1.4	1.5	0.9	1.3
	C15 LePhare
	C20C EAzY-py	2.4	3.6	3.8	NaN
	C20C LePhare
	C20F EAzY-py	...	0.6	1.5	...
	C20F LePhare
	Spec FAST++	2.2	...	2.9	2.6

Note. Catalogs considered are the UltraVISTA DR1 (UVDR1; A. Muzzin et al. 2013a), COSMOS2015 (C15; C. Laigle et al. 2016), COSMOS2020 CLASSIC (C20C), and COSMOS2020 FARMER (C20F; J. R. Weaver et al. 2022). None of these objects are in the ultra-deep stripes of the proprietary UltraVISTA DR3 catalog.

From this analysis there does not appear to be one catalog and SED modeling method which is definitively better than the others in terms of UMG candidate selection. This suggests that while different methods of source detection, masking of regions around stars, etc., do play a role in the differences between catalogs, different photometric redshift modeling codes and their associated fitting templates are the most substantial source of discrepancy in massive galaxy candidate identification.

5. Implications for the Number Densities of Massive Galaxies

5.1. Number Density Corrections

No S-UMG candidates have $|z_{\text{spec}} - z_{\text{phot}}| < 0.5$ in our primary selection catalogs, which appears to be due to difficulty in

accurately modeling galaxies with very red SEDs. Such SEDs have degeneracies between redshift, dust attenuation, AGN activity, and emission line strength, which are not consistently and reliably fit by the different modeling codes tested herein. This 0/16 success rate illustrates the extreme difficulty in accurately identifying such massive galaxies at these early epochs from ultraviolet, optical, and near-infrared photometry alone. To test the significance of this result, we assume that the 34 total high-quality S-UMG candidates from the parent catalogs consist of a range of “true” S-UMGs (i.e., $3 < z < 4$ and $\log(M_*/M_\odot) > 11.7$) from $N_{\text{real}} = 0$ to $N_{\text{real}} = 34$. For each value of N_{real} we then randomly select 16 objects 1000 times to assess how often zero are real. We confirm $N_{\text{real}} < 2$ at 1σ significance and $N_{\text{real}} < 8$ at 3σ significance. This implies that photometric catalogs overestimate the number of S-UMGs by a factor > 4 (99% confidence) and

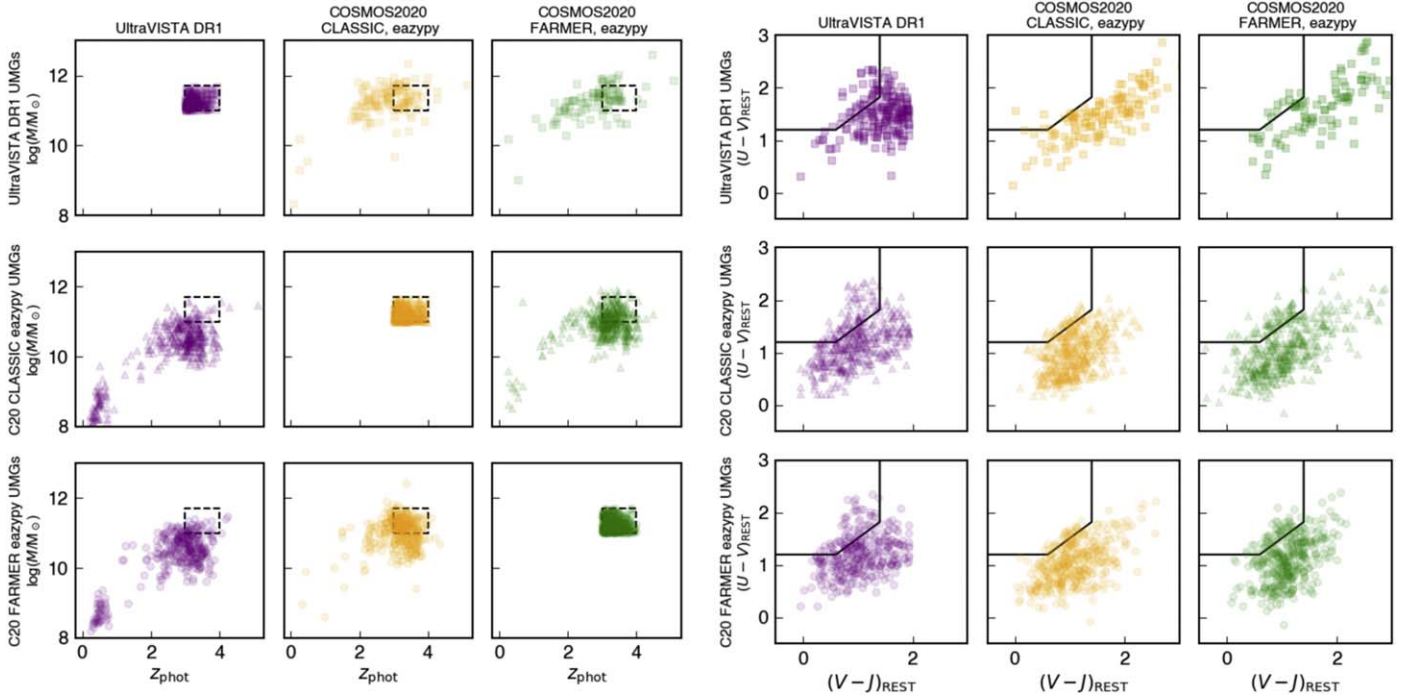


Figure 6. Comparison of UMGS selected from different photometric catalogs. Left side: the photometric redshifts and stellar masses of UMGS. The selection is represented by a black dashed box for comparison. In each row, either the UltraVISTA DR1 (top), COSMOS2020 CLASSIC EAZY-py (middle), or COSMOS2020 FARMER EAZY-py (bottom) catalog is used to select UMGS. The properties of the best matches to these objects in the other catalogs are shown in the other panels in that row. Objects with $z_{\text{phot}} < 1$ are ostensibly due to photometric confusion between the Lyman- and Balmer-break features. Note that not all sources have a match in all catalogs. Right side: the location of UMGS on the rest-frame UVJ color-color diagram following the same scheme as the left side.

Table 4
Comparison of Candidate Ultramassive Galaxies in the COSMOS Field between Different Catalogs

Catalog	Characterization	Total UMG Candidates	Counterparts in 1+ Other Catalogs	Counterparts in 3+ Other Catalogs	Counterparts which are UMGS in 1+ Other Characterizations	Counterparts which are UMGS in 5+ Other Characterizations
DR1	EaZY+FAST	201	155	108	85	14
DR3	EaZY+FAST	97	91	76	70	9
C15	LePhare	188	183	150	114	16
C20C	EaZY-py	476	476	422	220	12
C20C	LePhare	148	148	133	116	15
C20F	EaZY-py	375	375	338	202	11
C20F	LePhare	369	369	339	176	15

Note. Here we consider objects with $3 < z_{\text{phot}} < 4$, $11.0 < \log(M_*/M_\odot) < 11.7$, and $m_K < 23.4$. There are five catalogs considered for counterpart detection (UltraVISTA DR1 and DR3, COSMOS2015, and COSMOS2020 CLASSIC and FARMER) and seven object characterizations (the EaZY-py and LePhare characterizations for both COSMOS2020 catalogs).

Table 5
Comparison of Massive Galaxies in the COSMOS Field between Different Catalogs

Catalog	Characterization	Total S-UMG Candidates	Counterparts in 1+ Other Catalogs	Counterparts in 3+ Other Catalogs	Counterparts which are S-UMGs in 1+ Other Characterizations	Counterparts which are S-UMGs in 5+ Other Characterizations
DR1	EaZY+FAST	23	13	8	3	0
DR3	EaZY+FAST	6	4	2	0	0
C15	LePhare	4	2	2	1	0
C20C	EaZY-py	45	45	39	14	0
C20C	LePhare	4	4	3	3	0
C20F	EaZY-py	49	46	41	12	0
C20F	LePhare	10	10	8	2	0

Note. Here we consider objects with $3 < z_{\text{phot}} < 4$ and $m_K < 23.4$, with UMGS having $\log(M_*/M_\odot) \geq 11.0$ and S-UMGs having $\log(M_*/M_\odot) \geq 11.7$. There are five catalogs considered for counterpart detection (UltraVISTA DR1 and DR3, COSMOS2015, and COSMOS2020 CLASSIC and FARMER) and seven object characterizations (the EaZY-py and LePhare characterizations for both COSMOS2020 catalogs).

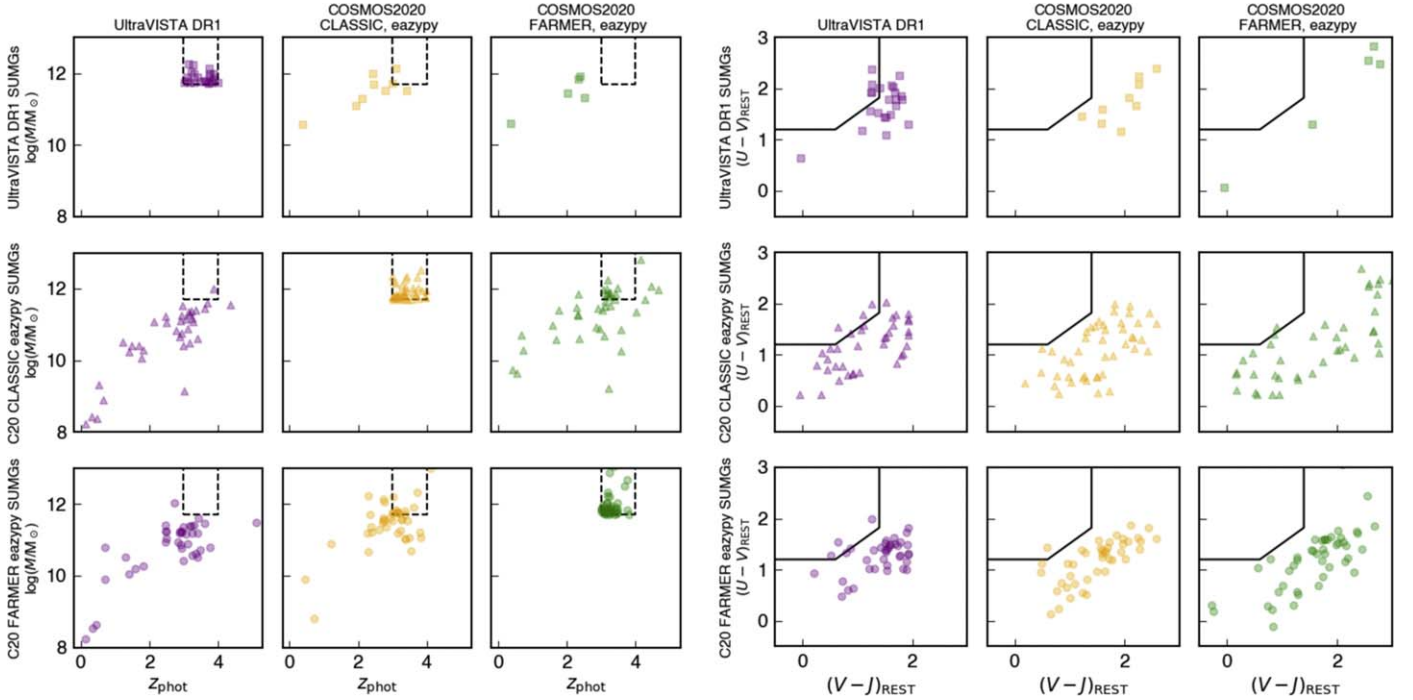


Figure 7. The same as Figure 6 but for the S-UMG sample selection ($\log(M_*/M_\odot) > 11.7$).

perhaps by a factor > 17 (68% confidence). A similar analysis of spectroscopically targeted MAGAZ3NE UMGs shows a much smaller correction of around $\lesssim 10\%$, suggesting that such a correction is only significant for the most massive and/or the reddest candidates.

To visualize these offsets, we calculate number densities based on the average from each of the seven COSMOS catalogs and SED modeling variations considered in Section 4. A comparison of the number density of UMG ($\log(M_*/M_\odot) > 11.0$) and S-UMG ($\log(M_*/M_\odot) > 11.7$) candidates from the nominal cuts to each catalog ($3 < z_{\text{phot}} < 4$, $m_K < 23.4$) show variations of up to ~ 0.5 dex from the values reported in the literature (A. Muzzin et al. 2013a; Z. C. Marsan et al. 2022). This is understandable, as some targets were removed from these studies due to, e.g., photometric contamination from neighbors, proximity to stars, other bad regions, etc. As such this can be thought of as an optimistic upper limit on the number densities of such targets. We take the average of these values as number density measurements and the asymmetric standard deviations as the uncertainties.

These values are shown as filled red stars in Figure 8 and agree well with other literature values. The effects of the correction factors derived above are shown as unfilled red symbols—a star for the UMGs and arrows (upper limits) for the S-UMGs.

5.2. Comparison with Simulations

The high-mass end of the GSMF has long shown discrepancies between observations and simulations at high redshifts, implying either inaccuracies or systematics when inferring the physical properties of such galaxies in observations or difficulties in building sufficient stellar mass at early times in simulations (or both; e.g., C. Steinhardt et al. 2016; S. Sherman et al. 2020).

Our results suggest that observational-related uncertainties play a significant role. It is also worth noting that such massive galaxies are very rare at early times, and thus require large volumes in both observations and simulations to include these

objects in statistically meaningful numbers. To this end, we compare the number densities of UMGs from photometric catalogs to those in the newest generation of models with large volumes, including: (1) the TNG300 hydrodynamical simulation of the IllustrisTNG project (D. Nelson et al. 2018, 2019; A. Pillepich et al. 2018; V. Springel et al. 2018), with a volume of $(205 \text{ Mpc } h^{-1})^3$ where $h = 0.6774$, (2) the FLAMINGO suite of simulations (J. Schaye et al. 2023; R. Kugel et al. 2023), considering the largest volume box, L2p8_m9, $(2.8 \text{ Gpc})^3$ in volume with $h = 0.681$, and (3) the GAEA semianalytic model (G. De Lucia et al. 2024), which builds upon the $(500 \text{ Mpc } h^{-1})^3$ box from the Millennium dark matter only simulation ($h = 0.73$; V. Springel et al. 2005).

We consider predictions over the redshift range $0.5 < z < 6$, and convolve the intrinsic stellar mass of each galaxy with a Gaussian of center unity and standard deviation 0.25 dex to simulate the systematic uncertainties from different stellar population modeling codes (e.g., B. Mobasher et al. 2015; B. Wang et al. 2024). We consider stellar mass inside a projected aperture of 30 kpc for FLAMINGO and within twice the half-stellar-mass radius for TNG300, while stellar masses from GAEA are “total” stellar masses. Comparison of these choices and other factors contributing to differences between simulations and observations are presented in Appendix B. We note however that there will still be discrepancies between these approximations and a full photometric extraction and catalog construction of the simulation data emulating the exact methods used for observations. Such a detailed comparison is beyond the scope of this work.

The number density comparison is shown in Figure 8. Broadly speaking, the agreement in number densities of UMGs between observations and simulations is good. Number densities from different observational works are consistent with each other out to $z \lesssim 4$, and diverge when moving to higher redshifts. TNG300 replicates the observational number densities over most of the $0.5 < z < 6$ redshift range considered, with a slight underprediction

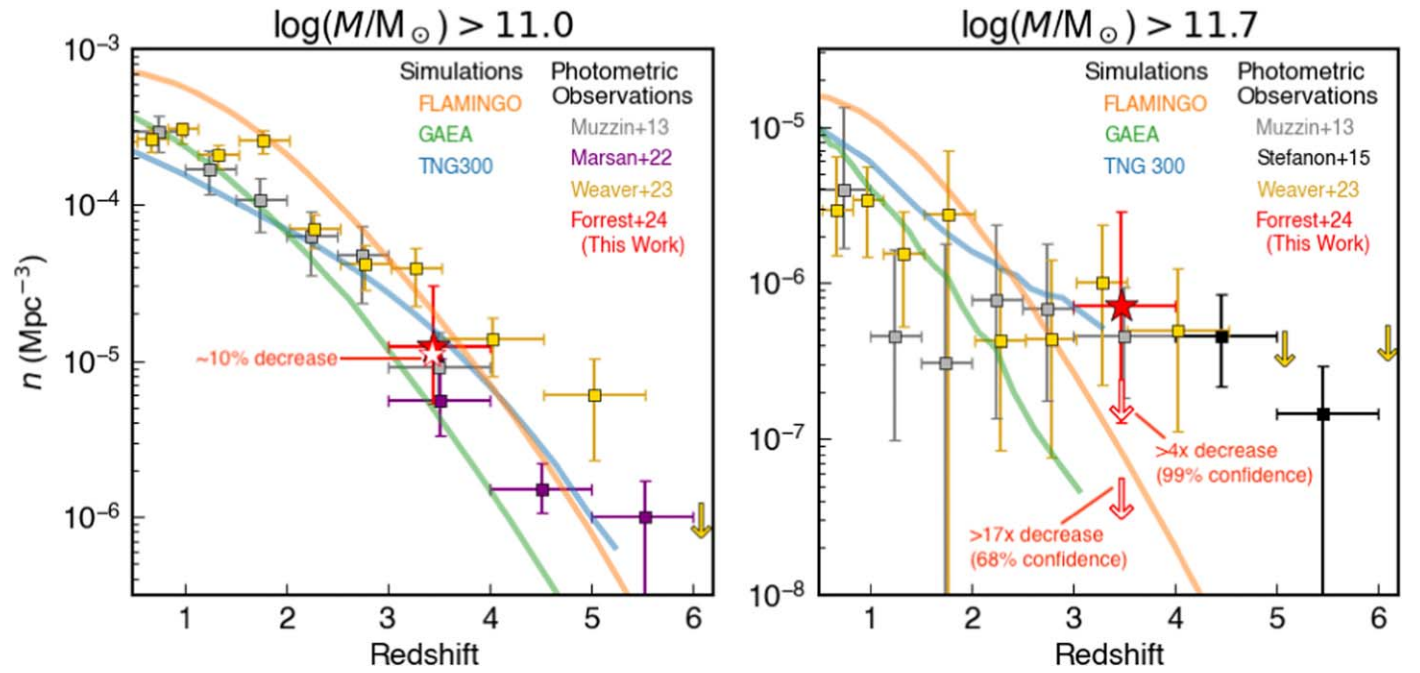


Figure 8. The number density of massive galaxy candidates in the COSMOS field. Left: literature values for galaxies with $\log(M_*/M_\odot) > 11$ from A. Muzzin et al. (2013a), Z. C. Marsan et al. (2022), and J. R. Weaver et al. (2022) are shown as gray, purple, and gold squares, respectively. Values from the FLAMINGO, GAEA, and TNG300 simulations convolved with a mass uncertainty of 0.25 dex are shown as orange, green, and blue lines, respectively, which truncate when the number of galaxies above the mass threshold drops below 10. The filled red star is the average number density from the seven COSMOS catalogs and SED modeling characterizations considered in Section 4, while the unfilled red star on the left panel shows the correction to this number density for galaxies with $\log(M_*/M_\odot) > 11.0$ based on the MAGAZ3NE spectroscopic success rate, a correction of $\sim 10\%$. Right: literature values for galaxies with $\log(M_*/M_\odot) > 11.7$ from A. Muzzin et al. (2013a), M. Stefanon et al. (2015), and J. R. Weaver et al. (2022) are shown as gray, black, and gold squares, respectively. The unfilled red arrows show the spectroscopic correction for galaxies with $\log(M_*/M_\odot) > 11.7$, which suggest a decrease of at least $4\times$ (99% confidence) and possibly more than $17\times$ (68% confidence).

at low redshifts. The growth of number densities over time from FLAMINGO and GAEA are similar, but there is a normalization offset of a few times in number density over all redshifts.

There is more uncertainty in the number densities of S-UMGs, much of which is driven by observations. After decreasing out to $z \sim 2$, observational number densities appear to flatten, which may be consistent with measurements being dominated by Eddington bias and Poissonian statistics. The lack of spectroscopically confirmed S-UMGs in this work suggests that the number densities of such objects does in fact continue to decrease to higher redshifts. While simulations do show a continued dropoff in S-UMG number density at $z > 2$, even the large volumes of the GAEA and TNG300 simulations do not predict statistically meaningful numbers of S-UMGs at high redshifts. The volume of the largest FLAMINGO simulation allows a prediction of S-UMG number densities below the observed values, but still within $\lesssim 2\sigma$ due to the large observational uncertainties, and also in good agreement with the observational correction derived from this work. Spectroscopic follow-up of larger and more complete samples of photometric S-UMG candidates is necessary to better constrain whether a discrepancy between simulations and observations is significant. With a larger study it could be quite possible that this tension disappears altogether.

5.3. Application to Massive Photometric High-redshift Candidates

We have attempted to use a combination of spectroscopy and photometry to generate SED models for use with photometric redshift fitting codes for similar galaxies. These efforts have shown only limited success, implying that the problem may not be simply

a lack of sufficiently similar models, but also the application of priors which do not accurately reflect the existence of this population of galaxies. Additionally, a stack of Keck/NIRES spectra of lower-redshift galaxies with extreme dust obscuration and strong rest-frame optical emission lines (hot DOGs; L. Finnerty et al. 2020) shows that such galaxies can have degenerate JWST NIRCcam colors with red, high-redshift galaxies (J. McKinney et al. 2023). As a result, it is possible that such misidentifications from photometric data occur over a range of redshifts.

Indeed, large numbers of candidate massive galaxies at surprisingly early times have been identified in JWST data, (e.g., R. P. Naidu et al. 2022a, 2022b; I. Labbe et al. 2023; G. Barro et al. 2024; D. D. Kocevski et al. 2024). A handful of these have been spectroscopically confirmed as early, massive galaxies (e.g., P. Arrabal Haro et al. 2023a; A. C. Carnall et al. 2023; S. Fujimoto et al. 2023). However, populations of red galaxies with significant AGN flux contributions have also been confirmed with JWST and longer wavelength data, some of which have significantly higher photometric redshifts than later confirmed (e.g., P. Arrabal Haro et al. 2023b; S. Fujimoto et al. 2023; R. A. Meyer et al. 2024). Further JWST/NIRSpec follow-up of these candidates will shed light on the hypothesis that a significant fraction of red, massive high-redshift candidates are in fact lower-redshift galaxies and/or less massive due to inaccurate photometric accounting of dust and AGN activity.

6. Conclusions

Deep, multipassband photometric catalogs point to the existence of a population of galaxies with redder colors and greater stellar masses than the most massive galaxies which

have previously been spectroscopically confirmed at $3 < z < 4$. In this work, the MAGAZ3NE survey spectroscopically targeted 16 such galaxy candidates with $z_{\text{phot}} \gtrsim 3$ and $\log(M_*/M_\odot) > 11.7$ (which we term S-UMGs) across three fields with some of the most extensive photometric data and high-quality photometric catalogs (UltraVISTA-COSMOS, VIDEO-CDFS, and VIDEO-XMM).

In sharp contrast to slightly lower-mass galaxies from these same catalogs, which show excellent agreement between spectroscopic and photometric redshifts, none of the S-UMG candidates has a redshift $|z_{\text{spec}} - z_{\text{phot}}| < 0.5$. Three galaxies show broad emission lines at $z > 3$ indicating Type I AGN contamination of the photometry biasing the photometric estimates of redshift and stellar mass. Half of the sample (9/16) are confirmed to be lower-redshift interlopers at $2.0 < z < 2.7$ (eight objects) and $z \sim 1.2$ (one object) but are still very massive. The remaining four targets do not have confirmed spectroscopic redshifts.

A comparison of catalogs in the COSMOS field suggests that none of the various construction methods or SED fitting methodologies used in these catalogs characterize the redshifts and stellar masses of these galaxies accurately. The underlying issue appears to be degeneracies among redshift, AGN contribution, and dust attenuation for objects with such red observed colors, which SED modeling templates have trouble disentangling.

We use the results of our spectroscopic observations to revise the number density of S-UMGs seen in photometric surveys. This correction is a reduction by a factor of >4 (99% confidence) and likely >17 (68% confidence). A comparison to some of the newest and largest volume models shows that this reduction reduces the severity of the discrepancy between simulations and observations at $3 < z < 4$. These misidentifications thus contribute substantially, and possibly entirely, to the discrepancy between observations and simulations in the number densities of the most massive galaxies at $3 < z < 4$, reducing the observed apparent flattening in number densities at high redshifts for these very massive objects.

The apparent failure of photometric redshift fitting codes to accurately identify these objects from fields with some of the most extensive and deepest photometry available suggests that improved optical and near-infrared photometry alone is unlikely to solve this problem over wide fields. While concentrated surveys with medium- and narrowband photometry may reduce the severity of this issue, as could further imaging from facilities such as the Atacama Large Millimeter/submillimeter Array and JWST, spectroscopic campaigns targeting such objects are the only way to truly quantify the scale of this effect.

Acknowledgments

The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Maunakea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

This research was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project “Understanding the evolution and transitioning of distant proto-clusters into clusters.” We are grateful for the support of ISSI and the use of their facilities. We also gratefully acknowledge the Lorentz Center in Leiden (NL) for facilitating discussions on this project.

The FLAMINGO simulation used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the Science and Technology Facilities Council (STFC) Distributed Research Utilizing Advanced Computing (DiRAC) High Performance Computing Facility (<http://www.dirac.ac.uk>). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operations grant ST/R000832/1.

G.W. gratefully acknowledges support from the National Science Foundation through grant AST-2205189 and from HST program number GO-16300. D.N. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) through an Emmy Noether Research Group (grant No. NE 2441/1-1).

B.F. thanks Brian C. Lemaux and Ian Smail, as well as the anonymous referee, for helpful comments on the manuscript.

This work has relied heavily upon code developed by other people, to whom we are quite thankful.

Software: Astropy (T. P. Robitaille et al. 2013; A. M. Price-Whelan et al. 2018; Astropy Collaboration et al. 2022), Bagpipes (A. C. Carnall et al. 2018), EAzy (G. B. Brammer et al. 2008), FAST (M. Kriek et al. 2009), FAST++ (C. Schreiber et al. 2018b), IPython (F. Perez & B. E. Granger 2007), LePhare (S. Arnouts et al. 1999; O. Ilbert et al. 2006), Matplotlib (J. D. Hunter 2007), and NumPy (T. E. Oliphant 2007).

Appendix A Additional Superultramassive Galaxy Candidate Information

Similar to Figure 2, the 2D and 1D spectra with emission line modeling for S-UMG candidates in the CDFS and XMM fields are shown in Figures A1 and A2, respectively. Table 6 reports the coordinates, observing conditions, and properties of all targeted MAGAZ3NE UMG and S-UMG candidates.

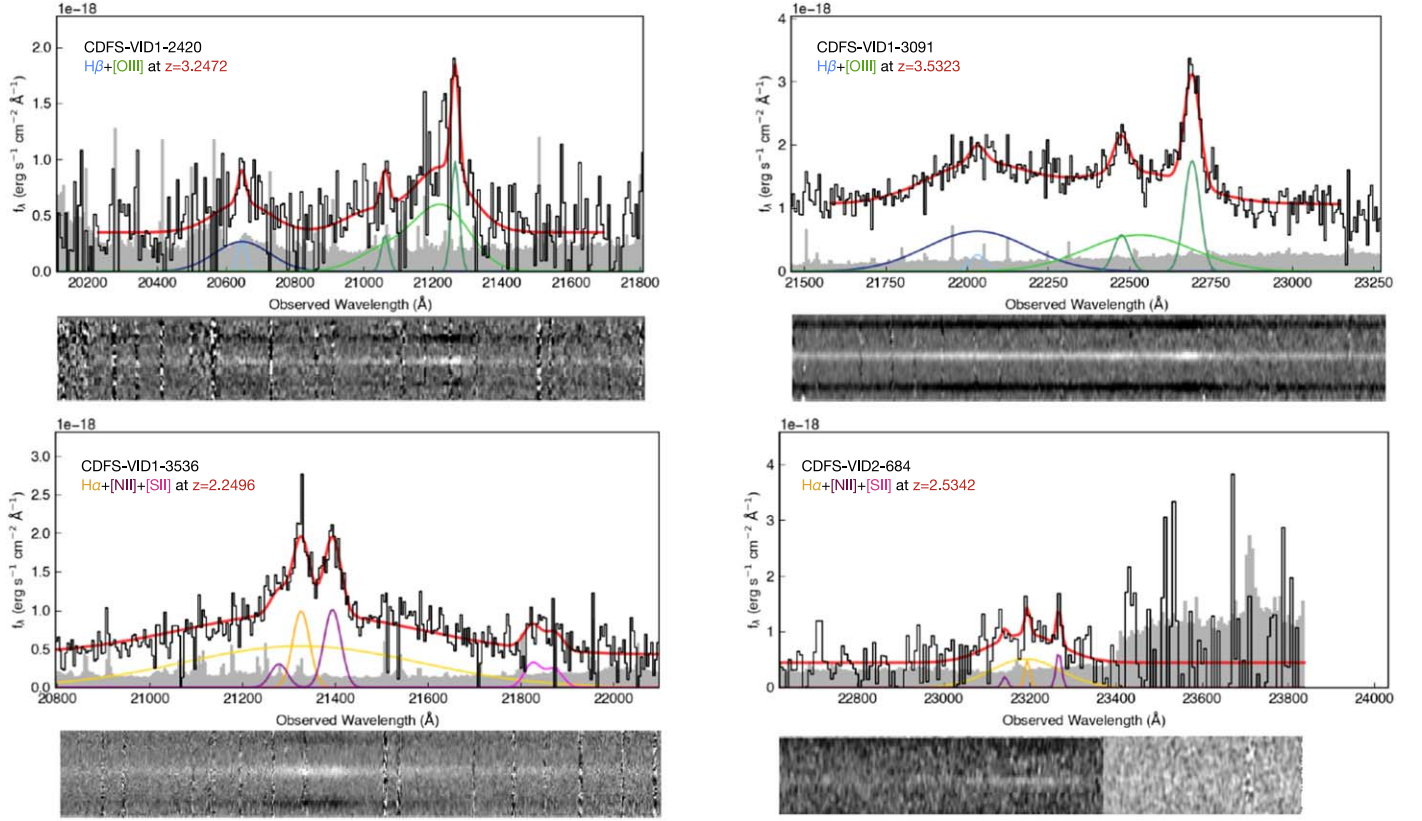


Figure A1. Line complex fitting for obtaining spectroscopic redshifts of candidate S-UMGs in the VIDEO-CDFS field with the same layout as Figure 2. For objects with confirmed redshifts, the region of the 1D and 2D spectrum with the confirming lines and fitted model are shown. The overall model is shown in red, while individual line components are shown in different colors ($H\beta$ in blue, $[O\text{ III}] \lambda\lambda 4959, 5007$ in green, $H\alpha$ in orange, $[N\text{ II}] \lambda\lambda 6548, 6584$ in purple, and $[S\text{ II}] \lambda\lambda 6718, 6733$ in magenta). For objects without confirmed redshifts, the entire wavelength range probed by H - and K -band spectroscopy is shown. Some targets were observed with multiple masks which were sensitive to different wavelengths; in such cases regions covered by fewer than the total number of masks are displayed as a lighter shade of gray.

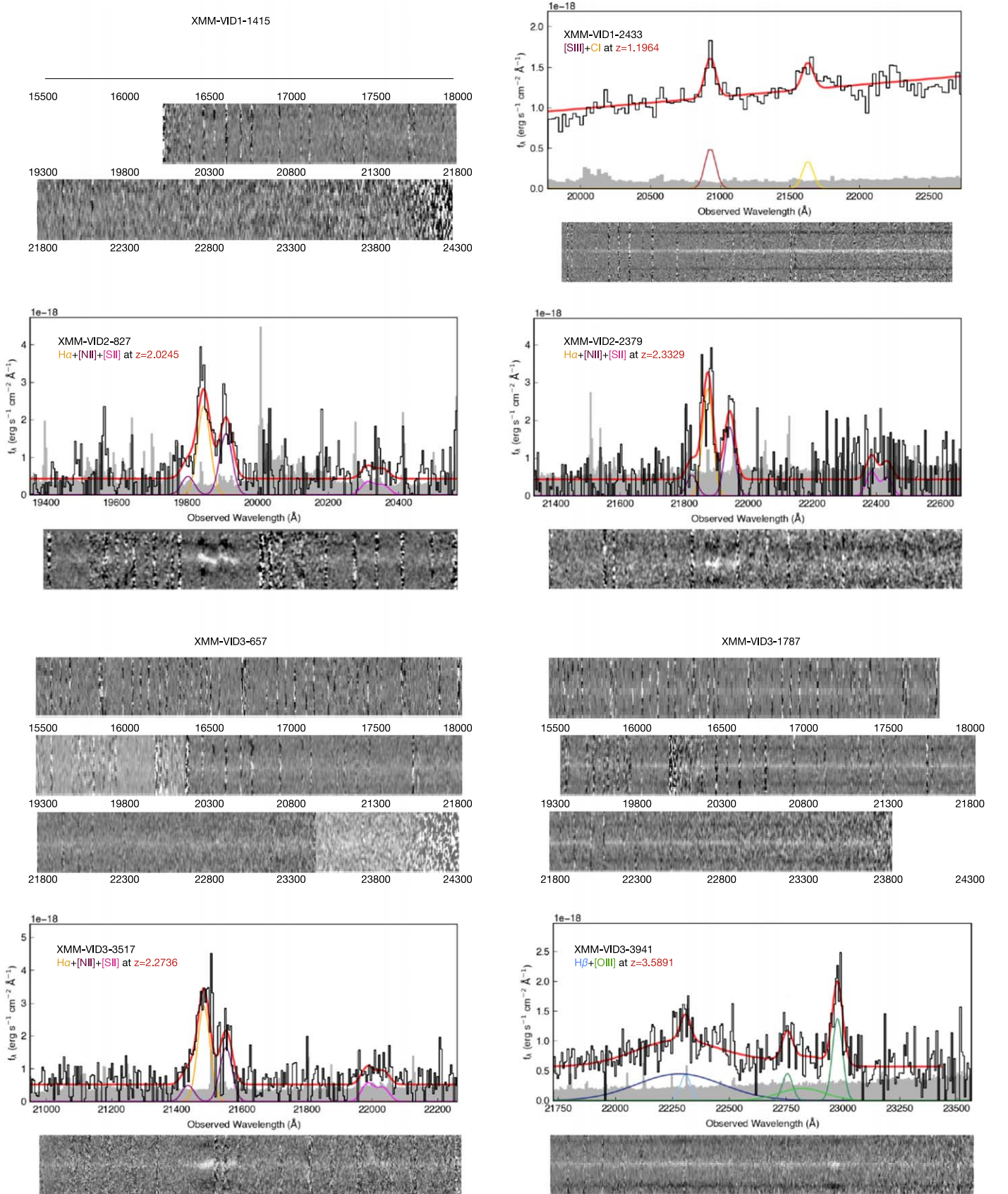


Figure A2. Line complex fitting for obtaining spectroscopic redshifts of candidate S-UMGs in the VIDEO-XMM field. Same layout as Figure 2.

Table 6

The List of Photometric Ultramassive Galaxy and Superultramassive Galaxy Candidates Targeted by the MAGAZ3NE Survey with Keck/MOSFIRE and Keck/NIRES

Galaxy	R.A. (hh mm ss)	Decl. (dd mm ss)	$t_{\text{exp},K}$ (H) (min)	Seeing $_K$ (H) ($''$)	m_K	z_{phot}	z_{spec}	M_{phot}^* $\log(M_{\odot})$	M_{spec}^* $\log(M_{\odot})$	References
XMM-VID1-896	02 15 36.408	−04 52 19.33	120	0.65	22.82	4.133	...	11.05	...	(k)
XMM-VID1-1126	02 16 46.821	−04 13 46.30	72	0.75	21.65	3.740	3.495	11.64	11.59	(d)
XMM-VID1-1415	02 19 01.864	−05 11 14.53	66	0.70	21.34	3.016	...	11.84	...	(d)
XMM-VID1-2075	02 17 32.863	−05 28 57.39	84	0.70	20.80	3.474	3.452	11.50	11.52	(b, c)
XMM-VID1-2399	02 18 01.830	−05 05 11.83	312	0.85	22.05	3.680	3.580	11.14	11.02	(b)
XMM-VID1-2433	02 17 05.599	−04 20 32.30	260 (=)	0.80	20.66	3.385	1.196	11.77	10.18	(d)
XMM-VID1-2761	02 19 21.082	−04 30 10.74	312	0.84	21.75	3.596	...	11.43	...	(d)
XMM-VID2-827	02 23 43.360	−04 50 33.85	72	0.66	21.41	3.033	2.024	11.88	10.74	(d)
XMM-VID2-2379	02 20 44.230	−05 06 27.35	96	0.64	21.71	3.317	2.333	11.77	10.75	(d)
XMM-VID3-657	02 25 40.887	−04 16 08.95	354 (96)	0.69 (0.75)	21.66	3.683	...	11.88	...	(d)
XMM-VID3-1120	02 27 10.098	−04 34 44.99	180 (376)	0.60 (1.2)	21.17	3.445	3.492	11.48	11.47	(a, b, c)
XMM-VID3-1638	02 24 50.993	−05 10 31.74	60	0.55	22.42	4.059	3.824	11.23	11.18	(k)
XMM-VID3-1787	02 26 22.333	−04 42 56.63	114 (80)	0.60 (0.85)	21.15	3.051	...	11.99	...	(d)
XMM-VID3-2293	02 26 34.076	−04 22 19.37	48	0.85	21.35	3.075	3.313	11.56	11.57	(b, l)
XMM-VID3-2457	02 26 56.043	−04 32 11.64	180	0.60	21.52	3.513	3.489	11.23	11.26	(b, c)
XMM-VID3-3517	02 24 12.920	−04 58 53.58	296 (128)	0.85 (0.85)	21.42	3.267	2.274	11.78	11.35	(d)
XMM-VID3-3941	02 24 13.691	−04 40 10.94	180	1.1	21.33	3.042	3.590	11.79	11.53	(b, d)
CDFS-VID1-2420	03 31 31.881	−28 04 10.32	132	0.74	21.69	3.784	3.245	12.07	11.91	(d)
CDFS-VID1-3091	03 28 00.346	−27 31 50.77	114 (40)	0.60 (0.70)	20.59	2.995	3.532	11.85	11.67	(d)
CDFS-VID1-3536	03 27 39.060	−28 02 09.25	204	0.67	21.44	3.274	2.25	11.87	11.42	(d)
CDFS-VID2-684	03 30 27.064	−28 21 13.22	210	1.0	21.67	3.323	2.533	11.80	11.63	(d)
COS-DR1-70455	09 58 02.923	01 57 55.62	312	0.58	21.85	3.149	2.655	11.75	11.33	(d)
COS-DR1-79837	10 02 32.244	02 06 32.18	120 (216)	0.74 (0.85)	21.10	3.318	...	11.99	...	(d)
COS-DR1-99209	10 02 56.366	02 24 06.00	54 (72)	0.90 (1.2)	20.90	3.071	2.983	11.16	11.22	(c)
COS-DR1-113684	09 59 43.891	02 07 24.28	204	0.75	21.39	3.473	3.831	11.23	11.20	(b, c)
COS-DR1-130749	10 01 22.382	02 20 02.73	144	0.85	21.54	2.984	...	11.29	...	(d)
COS-DR1-187247	10 02 49.205	02 32 55.76	162	0.65	22.04	3.314	...	11.46	...	(d)
COS-DR1-209435	10 00 04.793	02 30 45.28	252	0.68	22.03	3.769	2.481	11.98	11.55	(d, j)
COS-DR1-233254	10 01 03.578	02 48 10.22	69	0.77	21.10	2.880	2.059	11.78	10.78	(d)
COS-DR1-258857	09 58 18.120	02 45 36.20	120	0.70	21.66	3.269	...	11.39	...	(d)
COS-DR3-9705	10 00 13.481	01 37 04.76	84	0.65	21.92	3.108	...	11.47	...	(d)
COS-DR3-61518	10 02 05.695	02 13 50.12	180	0.75	22.26	3.233	3.367	11.33	11.31	(d)
COS-DR3-84674	10 00 43.759	02 10 28.68	0 (280)	−(0.9)	21.28	3.067	3.009	11.23	11.25	(b, c)
COS-DR3-111740	09 58 53.746	02 10 38.63	156	1.3	21.10	3.149	2.799	11.11	10.98	(b, c)
COS-DR3-131150	10 01 46.001	02 29 49.13	372	0.75	22.53	3.031	3.112	10.82	10.84	(i)
COS-DR3-131925	10 01 42.662	02 30 20.87	372	0.75	20.96	3.191	3.139	11.11	10.27	(b, i)
COS-DR3-133914	10 01 58.922	02 32 27.98	372	0.75	23.24	3.129	3.127	10.52	10.52	(i)
COS-DR3-134032	10 01 59.155	02 32 29.84	372	0.75	22.44	2.918	3.137	10.97	10.97	(i)
COS-DR3-160748	10 00 27.811	02 33 49.23	60	0.61	20.26	3.318	3.352	11.50	11.46	(b, e, f, g, h, i)
COS-DR3-179370	09 59 24.393	02 25 36.51	288	0.78	22.14	3.629	3.367	11.32	11.37	(b, e, f, h, i)
COS-DR3-186449	09 58 56.899	02 32 58.01	96	1.1	21.54	3.675	...	11.10	...	(d)
COS-DR3-195616	09 59 24.590	02 43 10.13	132	0.85	21.64	3.067	3.255	11.26	11.31	(b, j)
COS-DR3-201999	09 57 48.578	01 39 57.69	120 (352)	0.75 (1.0)	21.00	3.108	3.131	11.37	11.40	(b, c)
COS-DR3-202019	09 57 48.516	01 40 05.06	120 (352)	0.75 (1.0)	20.79	3.108	3.133	11.61	11.67	(b, c)
COS-DR3-208070	09 57 53.815	01 51 57.52	132	0.75	21.70	3.448	3.491	11.24	11.26	(b)
COS-DR3-226441	09 57 36.156	02 28 10.63	288	0.75	21.65	3.275	3.245	11.10	11.02	(b)

Publications listed are (a) B. Forrest et al. (2020a), (b) B. Forrest et al. (2020b), (c) B. Forrest et al. (2022), (d) this work, (e) Z. C. Marsan et al. (2015), (f) Z. C. Marsan et al. (2017), (g) P. Saracco et al. (2020), (h) I. McConachie et al. (2022), (i) I. McConachie et al. 2024, submitted, (j) W. Chang et al. 2024, in preparation, (k) S. M. Urbano Stawinski et al. 2024, in preparation, and (l) L. Shen et al. (2021).

Appendix B Simulation Considerations

Number densities from the FLAMINGO, GAEA, and TNG300 simulations can be derived in various ways.

Additionally, simulation data require application of observational effects in an attempt to perform fair comparisons. In Figure B1 we show comparisons of these choices and sources of uncertainty.

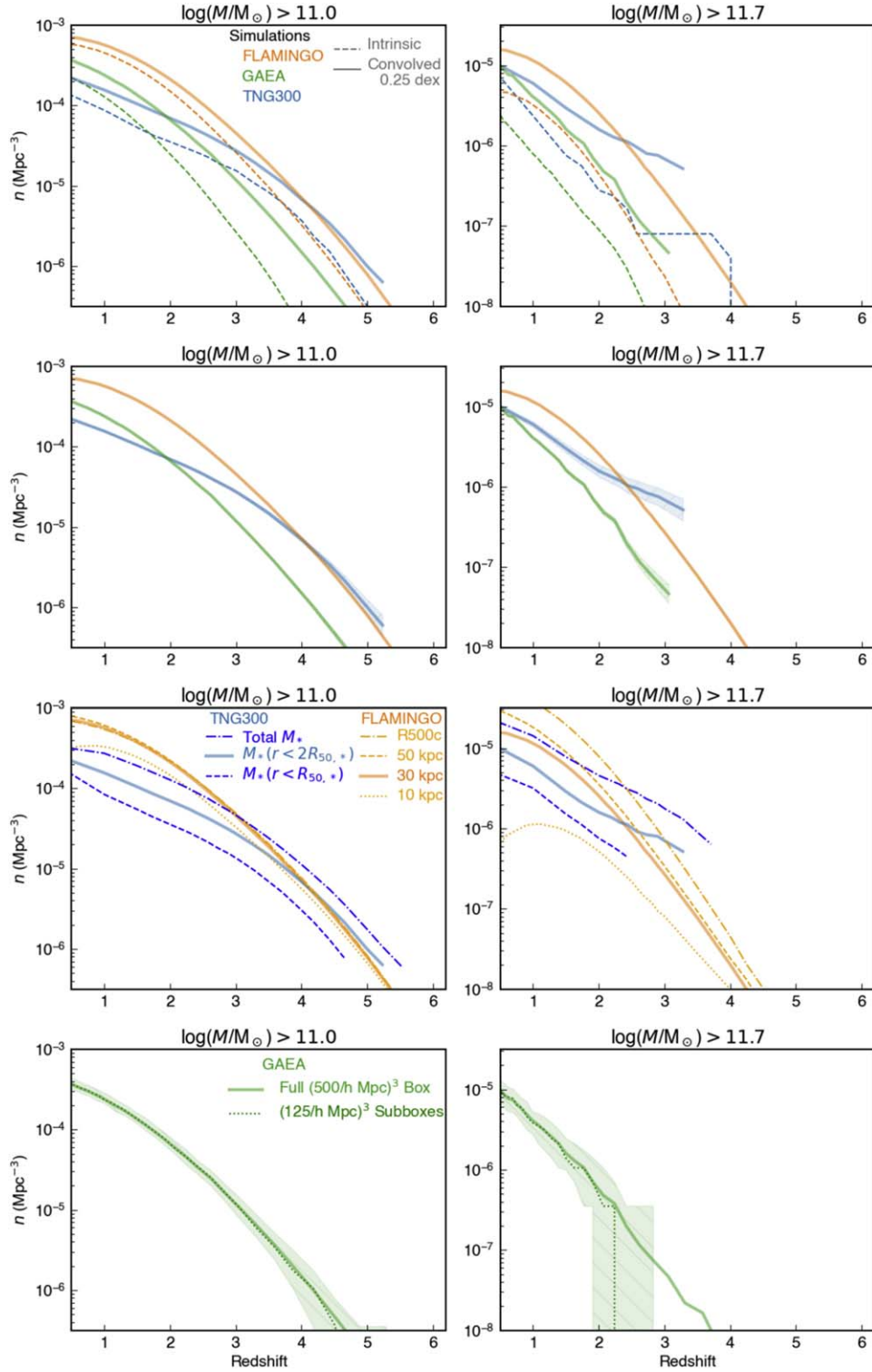


Figure B1. The effects of different simulation comparison uncertainties. Left panels show UMGs and right panels show S-UMGs. FLAMINGO, GAEA, and TNG300 are shown in orange, green, and blue, respectively. Top row: the intrinsic number densities from simulations (dashed lines) compared to the values after resampling each stellar mass using a Gaussian of width 0.25 dex (solid lines). Second row: the scale of Poissonian uncertainties in each simulation (shaded areas). Values are truncated when the number of objects in the simulation passing the mass cut drops below 10. Third row: number densities resulting from different aperture selection. The FLAMINGO comparison includes lines for three projected aperture radii in physical units, as well as a spherical aperture with radius R_{500c} , where the density is 500 times the critical density of the Universe. The TNG300 comparison includes lines considering the total stellar mass in a subhalo, all stellar mass within the stellar half-mass radius, and all stellar mass within twice the stellar half-mass radius. Bottom row: number densities in the entire GAEA ($500 h^{-1} \text{ Mpc}^3$) box (solid line), and those from $(125 h^{-1} \text{ Mpc})^3$ subboxes. The median is shown as a dotted line, with shading between the 16th and 84th percentile values. Values are shown even when the number of galaxies in a bin are fewer than 10.

ORCID iDs

Ben Forrest  <https://orcid.org/0000-0001-6003-0541>
 M. C. Cooper  <https://orcid.org/0000-0003-1371-6019>
 Adam Muzzin  <https://orcid.org/0000-0002-9330-9108>
 Gillian Wilson  <https://orcid.org/0000-0002-6572-7089>
 Danilo Marchesini  <https://orcid.org/0000-0001-9002-3502>
 Ian McConachie  <https://orcid.org/0000-0002-2446-8770>
 Percy Gomez  <https://orcid.org/0000-0003-0408-9850>
 Marianna Annunziatella  <https://orcid.org/0000-0002-8053-8040>
 Z. Cemile Marsan  <https://orcid.org/0000-0002-7248-1566>
 Wenjun Chang  <https://orcid.org/0000-0003-2144-2943>
 Gabriella de Lucia  <https://orcid.org/0000-0002-6220-9104>
 Fabio Fontanot  <https://orcid.org/0000-0003-4744-0188>
 Michaela Hirschmann  <https://orcid.org/0000-0002-3301-3321>
 Dylan Nelson  <https://orcid.org/0000-0001-8421-5890>
 Annalisa Pillepich  <https://orcid.org/0000-0003-1065-9274>
 Joop Schaye  <https://orcid.org/0000-0002-0668-5560>
 Stephanie M. Urbano Stawinski  <https://orcid.org/0000-0001-8169-7249>
 Mauro Stefanon  <https://orcid.org/0000-0001-7768-5309>
 Lizhi Xie  <https://orcid.org/0000-0003-3864-068X>

References

- Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, *ApJS*, **239**, 18
 Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, *PASJ*, **71**, 1
 Antwi-Danso, J., Papovich, C., Esdaile, J., et al. 2023, arXiv:2307.09590
 Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, **310**, 540
 Arrabal Haro, P., Dickinson, M., Finkelstein, S. L., et al. 2023a, *ApJL*, **951**, L22
 Arrabal Haro, P., Dickinson, M., Finkelstein, S. L., et al. 2023b, *Natur*, **622**, 707
 Ashby, M. L. N., Caputi, K. I., Cowley, W., et al. 2018, *ApJS*, **237**, 39
 Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, **935**, 167
 Barro, G., Pérez-González, P. G., Kocevski, D. D., et al. 2024, *ApJ*, **963**, 128
 Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, *ApJS*, **149**, 289
 Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
 Borch, A., Meisenheimer, K., Bell, E. F., et al. 2006, *A&A*, **453**, 869
 Brammer, G. 2021, eazy-py v0.5.2, Zenodo, doi:10.5281/zenodo.5012705
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
 Bundy, K., Ellis, R. S., Conselice, C. J., et al. 2006, *ApJ*, **651**, 120
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
 Capak, P., Aussel, H., Ajiki, M., et al. 2007, *ApJS*, **172**, 99
 Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, *ApJL*, **750**, 1
 Caputi, K. I., Deshmukh, S., Ashby, M. L. N., et al. 2017, *ApJ*, **849**, 45
 Carnall, A. C., McLure, R. J., Dunlop, J. S., & Dave, R. 2018, *MNRAS*, **480**, 4379
 Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2019, *MNRAS*, **490**, 417
 Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2023, *Natur*, **619**, 716
 Chabrier, G. 2003, *PASP*, **115**, 763
 Cole, S., Norberg, P., Baugh, C. M., et al. 2001, *MNRAS*, **326**, 255
 Da Cunha, E., Walter, F., Smail, I. R., et al. 2015, *ApJ*, **806**, 110
 Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, *A&A*, **605**, A70
 De Lucia, G., Fontanot, F., Xie, L., & Hirschmann, M. 2024, *A&A*, **687**, A68
 Dudzeviciute, U., Smail, I., Swinbank, A. M., et al. 2020, *MNRAS*, **494**, 3828
 Dunlop, J. S., Cirasuolo, M., & McLure, R. J. 2007, *MNRAS*, **376**, 1054
 Eisenhardt, P. R., Wu, J., Tsai, C. W., et al. 2012, *ApJ*, **755**, 173
 Finnerty, L., Larson, K., Soifer, B. T., et al. 2020, *ApJ*, **905**, 16
 Forrest, B., Tran, K.-V. H., Broussard, A., et al. 2017, *ApJ*, **838**, L12
 Forrest, B., Tran, K.-V. H., Broussard, A., et al. 2018, *ApJ*, **863**, 131
 Forrest, B., Annunziatella, M., Wilson, G., et al. 2020a, *ApJ*, **890**, L1
 Forrest, B., Marsan, Z. C., Annunziatella, M., et al. 2020b, *ApJ*, **903**, 47
 Forrest, B., Wilson, G., Muzzin, A., et al. 2022, *ApJ*, **938**, 109
 Fujimoto, S., Finkelstein, S. L., Burgarella, D., et al. 2023, *ApJ*, **955**, 130
 Furtak, L. J., Labbe, I., Zitirri, A., et al. 2024, *Natur*, **628**, 57
 Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, *Natur*, **544**, 71
 Glazebrook, K., Nanayakkara, T., Schreiber, C., et al. 2024, *Natur*, **628**, 277
 Gwyn, S. D. 2012, *AJ*, **143**, 38
 Hayashi, M., Tanaka, M., Shimakawa, R., et al. 2018, *PASJ*, **70**, 1
 Home, K. 1986, *PASP*, **98**, 609
 Hsieh, B.-C., Wang, W.-H., Hsieh, C.-C., et al. 2012, *ApJS*, **203**, 23
 Hunter, J. D. 2007, *CSE*, **9**, 90
 Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, **457**, 841
 Ilbert, O., McCracken, H. J., Le Fevre, O., et al. 2013, *A&A*, **556**, A55
 Jarvis, M. J., Bonfield, D. G., Bruce, V. A., et al. 2013, *MNRAS*, **428**, 1281
 Jin, S., Sillassen, N. B., Hodge, J., et al. 2024, *A&A*, **690**, L16
 Jones, L. H., Rosenthal, M. J., Barger, A. J., & Cowie, L. L. 2021, *ApJ*, **916**, 46
 Kausch, W., Noll, S., Smette, A., et al. 2015, *A&A*, **576**, A78
 Kawinwanichakij, L., Papovich, C., Ciardullo, R., et al. 2020, *ApJ*, **892**, 7
 Kocevski, D. D., Finkelstein, S. L., Barro, G., et al. 2024, arXiv:2404.03576
 Kriek, M., van Dokkum, P. G., Labbe, I., et al. 2009, *ApJ*, **700**, 221
 Kriek, M., van Dokkum, P. G., Whitaker, K. E., et al. 2011, *ApJ*, **743**, 168
 Kriek, M., Beverage, A. G., Price, S. H., et al. 2024, *ApJ*, **966**, 36
 Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, **218**, 15
 Kugel, R., Schaye, J., Schaller, M., et al. 2023, *MNRAS*, **526**, 6103
 Labbe, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, *ApJ*, **649**, L67
 Labbe, I., Gonzalez, V., Bouwens, R. J., et al. 2010, *ApJL*, **708**, 26
 Labbe, I., van Dokkum, P., Nelson, E., et al. 2023, *Natur*, **616**, 266
 Lacy, M., Surace, J. A., Farrah, D., et al. 2021, *MNRAS*, **501**, 892
 Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24
 Lang, D., Hogg, D. W., & Mykytyn, D., 2016 The Tractor: Probabilistic
 Astronomical Source Detection and Measurement, Astrophysics Source
 Code Library, ascl:1604.008
 Long, A. S., Casey, C. M., del P. Lagos, C., et al. 2023, *ApJ*, **953**, 11
 Marchesini, D., Van Dokkum, P. G., Förster Schreiber, N. M., et al. 2009, *ApJ*, **701**, 1765
 Marchesini, D., Whitaker, K. E., Brammer, G., et al. 2010, *ApJ*, **725**, 1277
 Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2017, *ApJ*, **842**, 21
 Marsan, Z. C., Marchesini, D., Brammer, G. B., et al. 2015, *ApJ*, **801**, 133
 Marsan, Z. C., Muzzin, A., Marchesini, D., et al. 2022, *ApJ*, **924**, 25
 Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, *ApJ*, **619**, L1
 Martis, N. S., Marchesini, D., Brammer, G. B., et al. 2016, *ApJ*, **827**, L25
 Mauduit, J.-C., Lacy, M., Farrah, D., et al. 2012, *PASP*, **124**, 1135
 McConachie, I., Wilson, G., Forrest, B., et al. 2022, *ApJ*, **926**, 37
 McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
 McKinney, J., Finnerty, L., Casey, C. M., et al. 2023, *ApJL*, **946**, L39
 McLean, I. S., Steidel, C. C., Epps, H., et al. 2010, *Proc. SPIE*, **7735**, 77351E
 McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, *Proc. SPIE*, **8446**, 84460J
 McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2021, *MNRAS*, **503**, 4413
 Mehta, V., Scarlata, C., Capak, P., et al. 2018, *ApJS*, **235**, 36
 Mendel, J. T., Beifiori, A., Saglia, R. P., et al. 2020, *ApJ*, **899**, 87
 Meyer, R. A., Barrufet, L., Boogaard, L. A., et al. 2024, *A&A*, **681**, L3
 Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, *ApJ*, **808**, 101
 Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013a, *ApJ*, **777**, 18
 Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013b, *ApJS*, **206**, 8
 Naidu, R. P., Oesch, P. A., Setton, D. J., et al. 2022a, arXiv:2208.02794
 Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022b, *ApJL*, **940**, L14
 Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, **475**, 624
 Nelson, D., Springel, V., Pillepich, A., et al. 2019, *ComAC*, **6**, 1
 Oke, J. B., & Gunn, J. E. 1983, *ApJ*, **266**, 713
 Oliphant, T. E. 2007, *CSE*, **9**, 10
 Perez, F., & Granger, B. E. 2007, *CSE*, **9**, 21
 Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, **475**, 648
 Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, **156**, 123
 Prochaska, J., Hennawi, J., Westfall, K., et al. 2020, *JOSS*, **5**, 2308
 Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, *A&A*, **558**, A33
 Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, **799**, 183
 Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, *ApJS*, **172**, 86
 Saracco, P., Marchesini, D., Barbera, F. L., et al. 2020, *ApJ*, **905**, 40
 Sarrouh, G. T. E., Muzzin, A., Iyer, K. G., et al. 2024, *ApJL*, **967**, L17
 Schaye, J., Kugel, R., Schaller, M., et al. 2023, *MNRAS*, **526**, 4978
 Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018a, *A&A*, **618**, A85
 Schreiber, C., Labbé, I., Glazebrook, K., et al. 2018b, *A&A*, **611**, A22
 Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
 Shen, L., Lemaux, B. C., Lubin, L. M., et al. 2021, *ApJ*, **912**, 60
 Sherman, S., Jogee, S., Florez, J., et al. 2020, *MNRAS*, **491**, 3318
 Simpson, J. M., Smail, I., Dudzeviciute, U., et al. 2020, *MNRAS*, **495**, 3409
 Smette, A., Sana, H., Noll, S., et al. 2015, *A&A*, **576**, A77
 Smit, R., Bouwens, R. J., Labbé, I., et al. 2015, *ApJ*, **833**, 254
 Smit, R., Bouwens, R. J., Labbe, I., et al. 2014, *ApJ*, **784**, 58

- Sobral, D., Smail, I., Best, P. N., et al. 2013, [MNRAS](#), **428**, 1128
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, [Natur](#), **435**, 629
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, [MNRAS](#), **475**, 676
- Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, [ApJ](#), **763**, 129
- Stefanon, M., Bouwens, R. J., Labbe, I., et al. 2021, [ApJ](#), **922**, 29
- Stefanon, M., Marchesini, D., Muzzin, A., et al. 2015, [ApJ](#), **803**, 11
- Steinhardt, C., Capak, P., Masters, D., & Speagle, J. S. 2016, [ApJ](#), **824**, 21
- Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, [ApJ](#), **830**, 51
- Suzuki, T. L., Kodama, T., Tadaki, K.-I., et al. 2015, [ApJ](#), **806**, 208
- Tanaka, M., Valentino, F., Toft, S., et al. 2019, [ApJ](#), **885**, L34
- Taniguchi, Y., Scoville, N., Murayama, T., et al. 2007, [ApJS](#), **172**, 9
- Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, [ApJ](#), **783**, 85
- Vaccari, M., Covone, G., Radovich, M., et al. 2017, in Proc. of Science 275, 4th Annual Conf. on High Energy Astrophysics in Southern Africa (HEASA 2016) (Trieste: SISSA), 26
- Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, [ApJ](#), **889**, 93
- van Dokkum, P. G., & Conroy, C. 2012, [ApJ](#), **760**, 70
- Wang, B., Leja, J., Atek, H., et al. 2024, [ApJ](#), **963**, 74
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, [ApJS](#), **258**, 11
- Weaver, J. R., Davidzon, I., Toft, S., et al. 2023, [A&A](#), **677**, A184
- Whitaker, K. E., Labbe, I., van Dokkum, P. G., et al. 2011, [ApJ](#), **735**, 86
- Wilson, J. C., Henderson, C. P., Herter, T. L., et al. 2004, [Proc. SPIE](#), **5492**, 1295
- Withers, S., Muzzin, A., Ravindranath, S., et al. 2023, [ApJL](#), **958**, L14
- Wu, J., Tsai, C. W., Sayers, J., et al. 2012, [ApJ](#), **756**, 96