Galaxies Going Bananas: Inferring the 3D Geometry of High-redshift Galaxies with JWST-CEERS

Viraj Pandya[1,2,52,54], Haowen Zhang (张昊文)[3], Marc Huertas-Company[2,4,5,6,7], Kartheik G. Iyer[1,2,52],
Elizabeth McGrath[8], Guillermo Barro[9], Steven L. Finkelstein[10], Martin Kümmel[11], William G. Hartley[12],
Henry C. Ferguson[13], Jeyhan S. Kartaltepe[14], Joel Primack[15], Avishai Dekel[16], Sandra M. Faber[17], David C. Koo[17],
Greg L. Bryan[1,2], Rachel S. Somerville[2], Ricardo O. Amorín[18,19], Pablo Arrabal Haro[20], Micaela B. Bagley[10],
Eric F. Bell[21], Emmanuel Bertin[22,23], Luca Costantin[24], Romeel Davé[25,26], Mark Dickinson[20], Robert Feldmann[27],
Adriano Fontana[28], Raphael Gavazzi[29,30], Mauro Giavalisco[31], Andrea Grazian[32], Norman A. Grogin[13],
Yuchen Guo[10], ChangHoon Hahn[33], Benne W. Holwerda[34], Lisa J. Kewley[35], Allison Kirkpatrick[36],
Dale D. Kocevski[8], Anton M. Koekemoer[13], Jennifer M. Lotz[37], Ray A. Lucas[13], Casey Papovich[38,39],
Laura Pentericci[28], Pablo G. Pérez-González[40], Nor Pirzkal[13], Swara Ravindranath[13], Caitlin Rose[14], Marc Schefer[41],
Raymond C. Simons[42], Amber N. Straughn[43], Sandro Tacchella[44,45], Jonathan R. Trump[42], Alexander de la Vega[46],
Stephen M. Wilkins[47,48], Stijn Wuyts[49], Guang Yang[50,51], and L. Y. Aaron Yung[43,53]
1 Columbia Astrophysics Laboratory, Columbia University, 550 West 120th Street, New York, NY 10027, USA; vgp2108@columbia.edu
2 Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA
3 Steward Observatory, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
4 Instituto de Astrofísica de Canarias (IAC), La Laguna, E-38205, Spain
5 Observatoire de Paris, LERMA, PSL University, 61 avenue de l'Observatoire, F-75014 Paris, France
6 Université Paris-Cité, 5 Rue Thomas Mann, F-75014 Paris, France
7 Universidad de La Laguna, Avda. Astrofísico Fco. Sanchez, La Laguna, Tenerife, Spain
8 Department of Physics and Astronomy, Colby College, Waterville, ME 04901, USA
9 University of the Pacific, Stockton, CA 90340, USA
10 Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA
11 LMU Faculty of Physics, Scheinerstr. 1, D-81679 München, Germany
12 Department of Astronomy, University of Geneva, ch. d'Ecogia 16, CH-1290 Versoix, Switzerland
13 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
14 Laboratory for Multiwavelength Astrophysics, School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA
15 Department of Physics, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
16 Center for Astrophysics and Planetary Science, Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel
17 UCO/Lick Observatory, Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
18 Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raul Bitrán 1305, La Serena 2204000, Chile
19 Departamento de Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena 1720236, Chile
20 NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Avenue, Tucson, AZ 85719, USA
21 Department of Astronomy, University of Michigan, 1085 S. University Avenue, Ann Arbor, MI 48109-1107, USA
22 Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, F-75014 Paris, France
23 CFHT, Kamuela, HI 96743, USA
24 Centro de Astrobiología (CAB), CSIC-INTA, Ctra de Ajalvir km 4, Torrejón de Ardoz, E-28850, Madrid, Spain
25 Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK
26 Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, Bellville, Cape Town 7535, South Africa
27 Institute for Computational Science, University of Zurich, Zurich, CH-8057, Switzerland
28 INAF—Osservatorio Astronomico di Roma, via Frascati 33, I-00078 Monte Porzio Catone, Italy
29 Laboratoire d'Astrophysique de Marseille, Aix-Marseille Université, CNRS, CNES, Marseille, France
30 Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, F-75014 Paris, France
31 University of Massachusetts Amherst, 710 North Pleasant Street, Amherst, MA 01003-9305, USA
32 INAF—Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy
33 Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
34 Department of Physics and Astronomy, University of Louisville, Louisville, KY 40292, USA
35 Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
36 Department of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA
37 Gemini Observatory/NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Avenue, Tucson, AZ 85719, USA
38 Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
39 George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, College Station, TX 77843-4242, USA
40 Centro de Astrobiología (CAB), CSIC-INTA, Ctra. de Ajalvir km 4, Torrejón de Ardoz, E-28850, Madrid, Spain
41 Department of Astronomy, University of Geneva, Chemin d'Ecogia 16, CH-1290, Versoix, Switzerland
42 Department of Physics, University of Connecticut, 196 Auditorium Road, Unit 3046, Storrs, CT 06269, USA
43 Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, USA
44 Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
45 Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge, CB3 0HE, UK
46 Department of Physics and Astronomy, University of California, 900 University Avenue, Riverside, CA 92521, USA
47 Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK
48 Institute of Space Sciences and Astronomy, University of Malta, Msida MSD 2080, Malta
49 Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK
50 Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

[51] SRON Netherlands Institute for Space Research, Postbus 800, 9700 AV Groningen, The Netherlands

Abstract

The 3D geometries of high-redshift galaxies remain poorly understood. We build a differentiable Bayesian model and use Hamiltonian Monte Carlo to efficiently and robustly infer the 3D shapes of star-forming galaxies in James Webb Space Telescope Cosmic Evolution Early Release Science observations with $\log M_*/M_\odot = 9.0$–10.5 at $z = 0.5$–8.0. We reproduce previous results from the Hubble Space Telescope Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey in a fraction of the computing time and constrain the mean ellipticity, triaxiality, size, and covariances with samples as small as ~50 galaxies. We find high 3D ellipticities for all mass–redshift bins, suggesting oblate (disky) or prolate (elongated) geometries. We break that degeneracy by constraining the mean triaxiality to be ~1 for $\log M_*/M_\odot = 9.0$–9.5 dwarfs at $z > 1$ (favoring the prolate scenario), with significantly lower triaxialities for higher masses and lower redshifts indicating the emergence of disks. The prolate population traces out a "banana" in the projected b/a–$\log a$ diagram with an excess of low-b/a, large-$\log a$ galaxies. The dwarf prolate fraction rises from ~25% at $z = 0.5$–1.0 to ~50%–80% at $z = 3$–8. Our results imply a second kind of disk settling from oval (triaxial) to more circular (axisymmetric) shapes with time. We simultaneously constrain the 3D size–mass relation and its dependence on 3D geometry. High-probability prolate and oblate candidates show remarkably similar Sérsic indices ($n \sim 1$), nonparametric morphological properties, and specific star formation rates. Both tend to be visually classified as disks or irregular, but edge-on oblate candidates show more dust attenuation. We discuss selection effects, follow-up prospects, and theoretical implications.

Unified Astronomy Thesaurus concepts: High-redshift galaxies (734); Galaxy classification systems (582); Dwarf galaxies (416); Galaxy structure (622); James Webb Space Telescope (2291); Galaxy disks (589); Galaxy spheroids (2032); Galaxy radii (617); Galaxy masses (607)

Supporting material: machine-readable table

1. Introduction

The 3D geometries of galaxies provide important clues about their formation history. There is a rich tradition of population studies tracing back almost a century that attempt to infer the 3D geometries of galaxies on the basis of their projected shapes. Hubble (1926) recognized that the distribution of projected ellipticities of local galaxies shows many more round objects than can be explained by randomly oriented disks alone. This served to justify, at least in part, a classification of nearby galaxies into a sequence of ellipticals and spirals. Subsequent interest in the degree of flattening in 3D of local galaxies contributed to the definitive establishment of an intermediate class of "lenticular" galaxies (de Vaucouleurs 1959; Sandage 1961). Sandage et al. (1970) later explored whether evolutionary connections can be made between these different subpopulations of galaxies using the distributions of their projected axis ratios. These and related studies of the time assumed that all galaxies in 3D can be thought of as a family of ellipsoids that can be classified into one of three extreme types: oblate (flattened in one direction like a disk), spheroidal (equally long in all three dimensions), or prolate (flattened in two directions and thus elongated in one direction). More general "triaxial" ellipsoid models were introduced soon after to explain the puzzling lack of rotation in local giant ellipticals and bulges (Contopoulos 1956; Stark 1977; Binney 1978).

It is now well accepted that nearby massive galaxies are a mixture of randomly oriented 3D oblate and spheroidal systems (Lambas et al. 1992; Alam & Ryden 2002; Ryden 2004; Vincent & Ryden 2005; Padilla & Strauss 2008; van der Wel et al. 2009, 2014a; Méndez-Abreu et al. 2010; Costantin et al. 2018). Stellar kinematics was crucial for establishing that, among the local giant elliptical population, it is only the the most massive ellipticals that are truly round or mildly triaxial whereas lower-mass ellipticals are more akin to lenticular galaxies (Binney 1985; Franx et al. 1991; Ryden 1992; Tremblay & Merritt 1995; Kormendy & Bender 1996; Emsellem et al. 2007, 2011; Cappellari 2016; Ene et al. 2018; Li et al. 2018). Nearby dwarfs also generally appear to be a combination of oblate and spheroidal 3D ellipsoids (Caldwell 1983; Ichikawa et al. 1986; Ferguson & Sandage 1989; Ichikawa 1989; Staveley-Smith et al. 1992; Ferguson & Binggeli 1994; Ryden & Terndrup 1994; Binggeli & Popescu 1995; Sung et al. 1998; Sánchez-Janssen et al. 2010; Roychowdhury et al. 2013; van der Wel et al. 2014a; Burkert 2017; Putko et al. 2019; Kado-Fong et al. 2020, 2021; Rong et al. 2020).

Connecting these constraints on the intrinsic shapes of nearby galaxies with those of their progenitors at high redshift became possible with the advent of the Hubble Space Telescope (HST). Numerous studies have established that the bright, massive population at high redshift already seems to have taken on oblate and spheroidal 3D shapes, albeit with smaller sizes and thicker minor-to-major-axis ratios (Reshetnikov et al. 2003; Elmegreen et al. 2004a, 2004b, 2005; Holden et al. 2012; Chang et al. 2013; van der Wel et al. 2014a; Satoh et al. 2019; Zhang et al. 2019, 2022; Hamilton-Campos et al. 2023). For these bright objects, constraints on gas kinematics through deep emission-line spectroscopy definitively showed the existence of large rotating disks with high random motions at early times (Förster Schreiber et al. 2006, 2009; Genzel et al. 2006; Law et al. 2009; Kassin et al. 2012; Glazebrook 2013; Wisnioski et al. 2015; Simons et al. 2016, 2017). There are also

indications of incredibly compact, massive, spheroidal "nuggets" that may be the precursors of present-day ellipticals (van Dokkum et al. 2008; Barro et al. 2013) or bulges in massive spirals (de la Rosa et al. 2016; Costantin et al. 2021, 2022).

The situation for dwarfs at high redshift is less clear.[55] It was, and still is, common practice to classify any distant, faint galaxies that are not obviously disks or ellipticals as "irregular" or "peculiar" (Driver et al. 1995; Glazebrook et al. 1995) and then to draw connections to mergers and interactions (e.g., Dressler et al. 1994). However, in one of the first deep fields with HST, Cowie et al. (1995) identified a new class of consistently elongated, linear, clumpy objects that they termed "chain" galaxies (see also Dickinson et al. 1995; van den Bergh et al. 1996). Dalcanton & Shectman (1996) argued that these chain galaxies are the edge-on projections of intrinsically oblate disk galaxies. Elmegreen et al. (2004a, 2004b, 2005) found more patterns among the peculiar/irregular population and grouped them into additional subclasses: chains, clump clusters, tadpoles, and double clumps (see also van den Bergh 2002; Conselice et al. 2004; Straughn et al. 2006). They, too, argued that chains were the edge-on versions of the rounder clump clusters, with the latter being harder to detect owing to surface brightness dimming. Hydrodynamical simulations of early clumpy star-forming galaxies by Immeli et al. (2004a, 2004b) and Bournaud et al. (2007) bolstered these claims.

But statistical analyses of projected axis ratios by Ferguson et al. (2004), Ravindranath et al. (2006), Yuma et al. (2011, 2012), and Law et al. (2012) also suggested another possibility: that high-redshift dwarfs may be intrinsically elongated (prolate) or triaxial rather than normal oblate disks. Soon after, the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) and 3D-HST Survey (Brammer et al. 2012; Skelton et al. 2014) provided much larger sample sizes, which allowed more robust 3D shape modeling. Van der Wel et al. (2014a) demonstrated that the asymmetric projected axis ratio distributions of high-redshift dwarfs peaking at small values can indeed be explained if they are a new class of preferentially elongated (prolate) systems. Zhang et al. (2019) additionally incorporated size information to further constrain 3D shapes and found that up to ∼70% of $\log M_*/M_\odot = $ 9.0–9.5 galaxies at $z = $ 1.5–2.0 and $z = $ 2.0–2.5 may be intrinsically prolate.

We do not consider the 3D nature of high-redshift dwarfs a resolved problem. With the ever-improving resolution of modern cosmological simulations, it has become possible to forward-model the 3D shapes of galaxies at different epochs. Ceverino et al. (2015) and Tomassetti et al. (2016) showed that, in their set of "zoom-in" simulations, low-mass galaxies at high redshift are indeed prolate and live in dark matter halos that are themselves prolate and aligned in the same direction as the stellar distribution. Pandya et al. (2019) used this to propose that intrinsic alignments of elongated high-redshift dwarfs may serve as tracers of cosmic web filaments, but they did not detect the expected signal, though they attributed this to severe spectroscopic incompleteness. On the other hand, Figures 8 and 9 of Pillepich et al. (2019) show far fewer galaxies with intrinsically prolate 3D stellar distributions in the TNG50

simulation compared to the observational constraints from van der Wel et al. (2014a) and Zhang et al. (2019). This potential discrepancy between different simulations with respect to each other and versus observations demands more detailed theoretical studies of 3D shapes. At the same time, the limited sensitivity, spatial resolution, and wavelength coverage of HST itself raise questions about the impact of completeness, blending, and light-weighting effects on the projected shape distributions of faint, distant galaxies. In particular, did HST miss round, face-on, oblate dwarfs with low surface brightness? Were groups of unresolved objects systematically blended together, leading to larger numbers of elongated sources? What do we infer about the 3D shapes of high-redshift galaxies by observing the bulk of their stellar population at even longer wavelengths?

The arrival of the James Webb Space Telescope (JWST) has the potential to transform our understanding of galaxy morphological evolution thanks to its high sensitivity and resolution at infrared wavelengths. Indeed, many exciting studies have already revealed the remarkable diversity of galaxy shapes at high redshift through a variety of methods. A combination of visual classifications and parametric and nonparametric morphological measurements suggested early on that the fraction of disk galaxies is higher in JWST imaging compared to HST (Ferreira et al. 2022; Kartaltepe et al. 2023). Robertson et al. (2023) used a deep-learning framework trained on previous HST imaging and CANDELS visual classifications to identify faint, distant disks in JWST imaging that were missed by HST (see also Huertas-Company et al. 2023; Tohill et al. 2023). But the visual classifications and metrics used in many of these studies so far do not necessarily distinguish between 3D shapes, namely prolate versus oblate geometries and the possibility of oval (triaxial) disks. Indeed, Vega-Ferrero et al. (2024) used a machine-learning method trained instead on the TNG50 simulations to classify galaxies observed with JWST and found that a substantial fraction of visually classified disks may instead be intrinsically elongated. Nelson et al. (2023) identify a sample of 12 massive, elongated, and surprisingly red galaxies at $z = $ 2–6, which they claim may be either oblate or prolate.

In this paper, we place new constraints on the 3D shapes of high-redshift galaxies using JWST observations from the Cosmic Evolution Early Release Science (CEERS) survey (Finkelstein et al. 2023). Our approach is distinct from and complementary to previous JWST studies of galaxy structure and morphology. We will use the distributions of projected axis ratios and sizes to fit a family of triaxial ellipsoid models to observed galaxies in different mass–redshift bins. We focus on star-forming galaxies since the fraction of quiescent galaxies drops dramatically toward high redshift (e.g., Pandya et al. 2017), which would limit our sample sizes for inference, and because quiescent galaxies likely occupy a different "mode" for 3D shapes (Chang et al. 2013). We will allow the data to speak for themselves using a Bayesian approach to constrain the relative fractions of oblate, spheroidal, and prolate galaxies, as well as triaxial systems more generally. In addition to constraining 3D shapes, we will also simultaneously derive the 3D size–mass relation and its redshift evolution, which is of great interest for constraining galaxy formation models. Although our sample sizes are small, we will show that the data have sufficient constraining power in many cases and that

---

[55] In this paper, when we say "high-redshift dwarfs," we mean the progenitors of present-day galaxies like our own Milky Way.

when this is not true, our posteriors reflect our uniform priors as they should.

This paper is organized as follows. In Section 2, we describe the new JWST data, as well as previous CANDELS observations for consistency checks. In Section 3, we detail our methods. We present our results on 3D shape evolution in Section 4 and compare some properties of high-probability prolate, oblate, and spheroidal candidates in Section 5. We discuss the implications of our findings in Section 6 and summarize our conclusions in Section 7. We assume a standard Planck Collaboration et al. (2016) cosmology with $h = 0.6774$, $\Omega_{m,0} = 0.3075$, $\Omega_{\Lambda,0} = 0.691$, and $\Omega_{b,0} = 0.0486$.

2. Data

2.1. CEERS JWST Observations

We use JWST observations from the CEERS survey (Program ID 1345; Finkelstein et al. 2023), which surveyed portions of the Extended Groth Strip (EGS; Davis et al. 2007) previously covered by CANDELS (Stefanon et al. 2017). Specifically, we focus on the 10 NIRCam pointings covering roughly 50% of the area of the CANDELS-EGS in six filters: F115W, F150W, F200W, F277W, F356W, and F444W. Four of the pointings were observed in 2022 June, while the other six were observed in 2022 December. Data reduction details are given in Bagley et al. (2023).[56]

We use two independent photometric and galaxy structural catalogs for the same CEERS imaging to ensure that our results are not driven by source detection and structural measurement methods. First, we use the internal CEERS team photometric catalog from Finkelstein et al. (2023), which uses the original Source Extractor code (Bertin & Arnouts 1996). Our stellar masses, star formation rates (SFRs), photometric redshifts, and dust attenuations are based on broadband spectral energy distribution (SED) fitting with EAZY (Brammer et al. 2008; more details specific to CEERS can be found in Barro et al. 2023).[57] We use galaxy structural measurements from GALFIT (Peng et al. 2002) based on single-component Sérsic (1963) model fits done independently in each filter (details will be given in E. McGrath et al. 2024, in preparation, but this closely follows van der Wel et al. 2012). Only sources with F356W < 28.5 AB mag were fit with EAZY and GALFIT, but our sample selection is generally brighter than ~27 AB mag (Appendix B). We only include sources with a GALFIT flag of 0, which indicates no problems during the fitting process.

As an alternative, we use the next-generation Source Extractor++ code (SE++; Bertin et al. 2020; Kümmel et al. 2022) to independently do source detection, characterization, and single-component Sérsic (1963) model fits jointly across all wavelengths, including previous HST imaging in ACS-F606W, ACS-F814W, WFC3-F125W, WFC3-F140W, and WFC3-F160W. SE++ starts with the same global background-subtracted image and fixes the local background around each source to zero just like Galfit, but it uses its own deblending, masking, and fitting algorithms. All Sérsic model parameters were allowed to vary across filters except the position angle. Simulated parameter recovery tests have shown that SE++

performs very well and, unlike GALFIT, provides meaningful uncertainties on Sérsic parameters (Euclid Collaboration et al. 2023). We only include sources whose sizes and axis ratios have fractional uncertainties <10%, which removes at most ~10% of the sample in each redshift bin. For every SE++ source, we cross-match to the nearest neighbor within $0\rlap{.}''25$ from the the CEERS team catalog described in the previous paragraph; this provides stellar mass, SFR, and redshift estimates from EAZY. We have a negligible number of SE++ sources without a cross-match, and we do not expect that our conclusions would have changed if we had done independent SED fits using the SE++ photometry itself. We visually inspected the Sérsic model fits and residuals for many of our objects and found that they generally look reasonable with very few catastrophic failures and with good agreement between SE++ and GALFIT. Appendix C shows cutouts of the data, Sérsic model, and residuals for some example galaxies fit by both SE++ and GALFIT.

Our definition of projected size is the Sérsic half-light radius, i.e., the semimajor-axis length of the ellipse that encloses half of the model light distribution. Importantly, all of our sizes and axis ratios from both GALFIT and SE++ are intrinsic, i.e., from the best-fitting Sérsic model before being convolved with the point-spread function (PSF). We use the same empirical PSFs described in Finkelstein et al. (2023) for both GALFIT and SE++. More goodness-of-fit details for GALFIT will be presented in E. McGrath et al. (2024, in preparation).

In this paper, we only consider star-forming galaxies with stellar masses in the range 10^9–$10^{10.5}M_\odot$ and with $z = 0.5$–8.0. These cuts reflect the completeness and sample sizes afforded by CEERS. The motivation for focusing on the star-forming population alone is threefold: (1) this is distinct from the quiescent population, which may have different 3D shapes; (2) it is easier to detect and hence be complete to star-forming galaxies in CEERS; and (3) the number of quenched galaxies drops rapidly at high redshift. We select star-forming galaxies following the procedure of Pandya et al. (2017): in each redshift bin, we find the median specific SFR (sSFR = SFR/M_*) of dwarfs with $M_* = 10^9$–$10^{9.5}M_\odot$, which should be overwhelmingly star-forming.[58] This gives us the median sSFR of galaxies on the star-forming main sequence (SFMS; for simplicity we assume zero slope for sSFR as a function of M_*, but this does not affect our conclusions). We only consider galaxies in each redshift bin whose sSFR is larger than the median main-sequence sSFR -0.45 dex. This cut crudely excludes any galaxies lying more than 1.5σ below the main-sequence ridge line, where $\sigma \sim 0.3$ dex is the typical main-sequence scatter.

Table 1 lists our mass–redshift bins and the number of galaxies with reliable structural measurements from both GALFIT and SE++. To ensure that we probe galaxy structure at roughly the same rest-frame optical wavelength (~4000–8000 Å) across redshift, we use a different NIRCam filter for each redshift interval as given in the table. We tried different filters and found that our conclusions are not sensitive to this. For both sets of catalogs, we only use galaxies whose semimajor-axis length from the Sérsic model is larger than the PSF FWHM in the filter corresponding to their redshift. In practice, this size cut mainly affects our $z = 3$–8 bin since

---

[57] Our broadband SED fits may not capture the impact of strong emission lines, which may bias our stellar masses toward larger values and lead to more uncertain photometric redshifts, SFRs, and dust attenuations. We do not expect this to affect our main conclusions.

[58] We will refer to high-redshift galaxies with $M_* = 10^9$–$10^{10}M_\odot$ as dwarfs generally. This includes Milky Way progenitors, which are expected to fall in this mass range at $z \sim 2$ (Papovich et al. 2015).

Table 1
Our CEERS Mass–Redshift Bins, Adopted NIRCam Filters that Probe Roughly the Same Rest-frame Optical Wavelength with Redshift, and the Number of CEERS Galaxies from GALFIT and SE++ that Satisfy Our Selection Criteria (Star-forming, Reliable Structural Measurements, Larger $\log a$ than PSF FWHM)

z	$\log M_*/M_\odot$	Filter	Galfit	SE++	CANDELS
0.5–1.0	9.0–9.5	F115W	199	242	2725
0.5–1.0	9.5–10.0	F115W	100	131	1464
0.5–1.0	10.0–10.5	F115W	48	57	600
1.0–1.5	9.0–9.5	F115W	313	388	3007
1.0–1.5	9.5–10.0	F115W	168	223	1594
1.0–1.5	10.0–10.5	F115W	90	111	741
1.5–2.0	9.0–9.5	F150W	379	461	3497
1.5–2.0	9.5–10.0	F150W	239	298	1741
1.5–2.0	10.0–10.5	F150W	75	90	780
2.0–2.5	9.0–9.5	F200W	417	530	2092
2.0–2.5	9.5–10.0	F200W	220	298	1159
2.0–2.5	10.0–10.5	F200W	72	90	550
2.5–3.0	9.0–9.5	F200W	196	250	⋯
2.5–3.0	9.5–10.0	F200W	104	138	⋯
2.5–3.0	10.0–10.5	F200W	28	41	⋯
3.0–8.0	9.0–9.5	F356W	453	493	⋯
3.0–8.0	9.5–10.0	F356W	257	312	⋯
3.0–8.0	10.0–10.5	F356W	93	106	⋯

Note. We also include the number of galaxies from all five CANDELS fields at $z < 2.5$ with reliable GALFIT measurements.

galaxies at lower redshifts are generally much larger than the PSF FWHM.

2.2. CANDELS HST Observations

Separately from CEERS, we also use the full five-field CANDELS data set to test our new 3D shape modeling code.[59] Since we will try to reproduce Zhang et al. (2019) as a consistency check, we apply the same cuts to the CANDELS data set as them, which include the standard F160W < 25.5 AB mag, PhotFlag = 0, CLASS_STAR < 0.8, and reliable GALFIT measurements (flag = 0 in the catalogs from van der Wel et al. 2012). The redshifts are a mixture of photometric, grism, and (when available) spectroscopic redshifts as described in Kodra et al. (2023). The stellar masses and SFRs are the medians from many different SED fitting codes (Mobasher et al. 2015; Santini et al. 2015). This is the same data set that was used for Pandya et al. (2019). To be consistent with Zhang et al. (2019), we select star-forming galaxies in each redshift bin using the sSFR–M_* relations provided in Fang et al. (2018). This is different from the Pandya et al. (2017) strategy we use for CEERS, but the conclusions are unaffected regardless of approach. We have verified through visual inspection that the resulting sample of star-forming galaxies roughly matches Zhang et al. (2019). Table 1 lists the number of CANDELS galaxies in each mass–redshift bin at $z < 2.5$. CANDELS was designed to be complete to $\log M^* \sim 9$ galaxies out to $z = 2.5$, so we do not attempt to extend beyond that for this part of the analysis.

2.3. CEERS Size–Axis Ratio Histograms

Figure 1 shows the distributions of projected axis ratio for CEERS galaxies in each of our mass–redshift bins as measured

with GALFIT and SE++. The two sets of measurements are generally consistent in showing that low-mass galaxies preferentially have low b/a (appearing edge-on), with a deficit of high-b/a (round projected) dwarfs. Thus, with more sensitive, higher-resolution, redder wavelength imaging from JWST, we still see the same asymmetric b/a distributions skewed toward low values as were seen in HST-CANDELS by van der Wel et al. (2014a) and Zhang et al. (2019). The $\log M_*/M_\odot = 10$–10.5 bins generally have smaller sample sizes, but the distributions are becoming more uniform, which we expect for a mixture of disks and spheroids seen from random viewing angles.

Figure 2 similarly compares the distributions of the semimajor-axis lengths for CEERS galaxies in our different mass–redshift bins from GALFIT and SE++. Here again we see that the two sets of measurements are generally consistent. However, SE++ finds more large-size galaxies than GALFIT. In Appendix C, we argue that this does not affect our overall conclusions and is likely due to SE++ genuinely detecting additional large, bright galaxies rather than finding systematically larger sizes than GALFIT.

Figures 3 and 4 show the 2D histograms of projected b/a versus $\log a$ for CEERS based on GALFIT and SE++, respectively. This is the joint parameter space that we will use to constrain our 3D shape model. In the two lower-mass bins, we generally have ∼100–450 objects, which, as we will show later, is sufficient for constraining 3D shapes. However, our higher-mass bin tends to be very noisy, with <50 galaxies for some redshift intervals. In this case, our Bayesian 3D shape modeling code may be unconstrained by the data, so the posteriors may reflect our priors.

We remind the reader that our sizes and axis ratios are all intrinsic, i.e., from the best-fitting Sérsic model before being convolved with the PSF. While we can recover intrinsic sizes and axis ratios below the resolution limit, we do not want our results to be driven by completely unresolved galaxies, so we require the intrinsic semimajor-axis length $a > $ PSF FWHM but do not impose any such cut for the intrinsic semiminor-axis length b. In many of the lower-mass, higher-redshift bins, there is an excess of low-b/a galaxies and a deficit of round (high-b/a) sources rather than a uniform distribution in b/a for a given $\log a$, particularly for larger (well-resolved) $\log a$. In Section 3.1 below, we will illustrate how this curved "banana-like" joint distribution of b/a–$\log a$ arises from ellipsoids with intrinsically elongated 3D shapes.

In Appendix B, we describe completeness simulations to understand the reasonably faintest extended sources that we would be sensitive to with the CEERS imaging. We injected 10^4 fake Sérsic profiles into the CEERS imaging and processed those mock images using our entire pipeline. In short, we find that for extended sources spanning a reasonable range of sizes and axis ratios, we are complete to objects as faint as ∼26.5 AB mag (F277W). This is ∼2 mag deeper than HST-CANDELS, for which studies of galaxy morphology are typically restricted to sources brighter than 24.5 AB mag in the F160W filter. It is not clear that sources even fainter than 27 AB mag would still satisfy our $\log M_*/M_\odot > 9$ cut.

3. Methods

3.1. Banana Diagram Decomposition

Following van der Wel et al. (2014a) and Zhang et al. (2019), we start by assuming that all galaxies can be
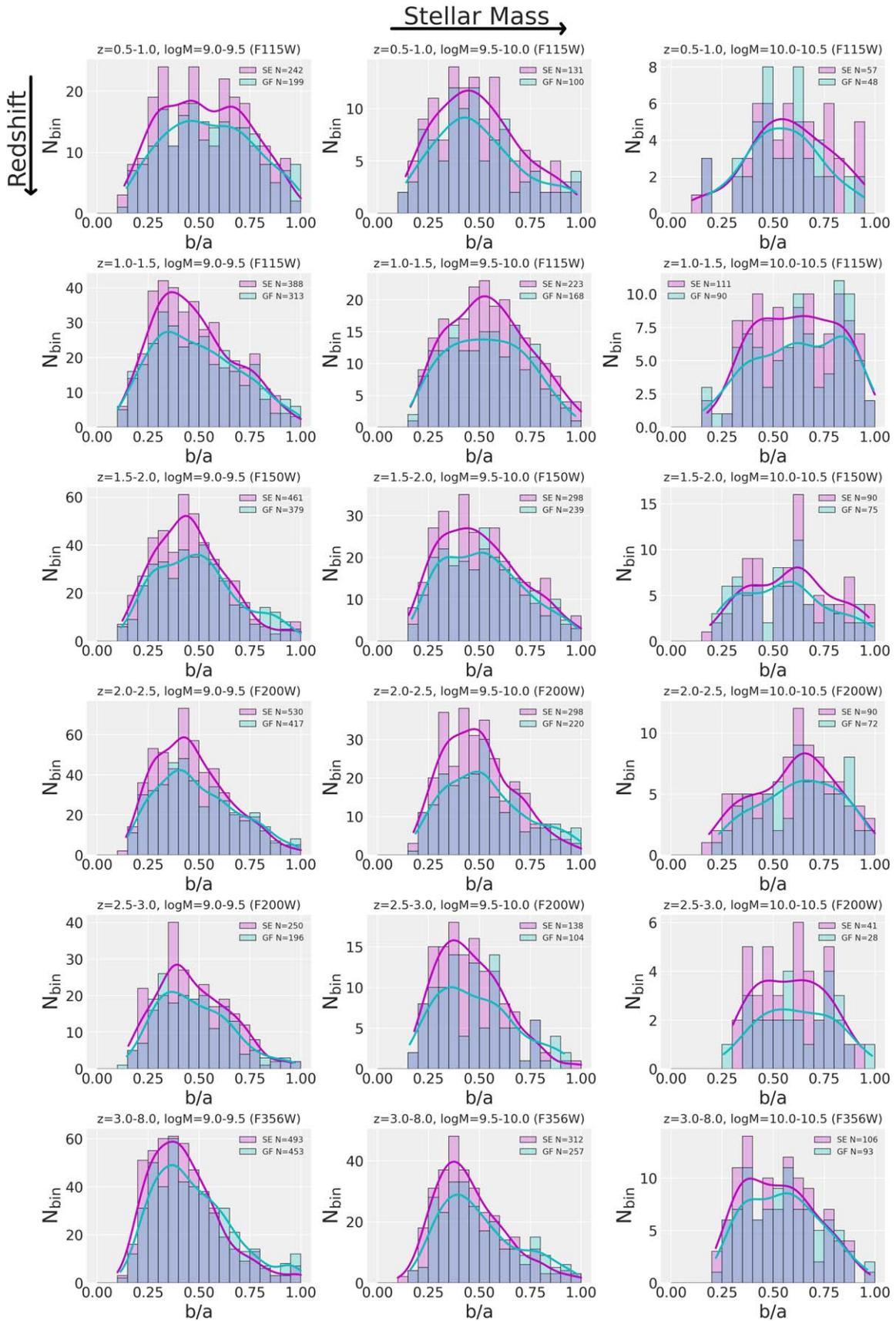
Figure 1. The distribution of projected axis ratios in our different mass–redshift bins as measured with GALFIT (cyan) and SE++ (magenta). The smooth curves are kernel density estimates. The two sets of measurements are generally consistent. We emphasize that the distributions for low-mass, high-redshift bins are asymmetric and skewed toward low b/a suggestive of prolate populations.
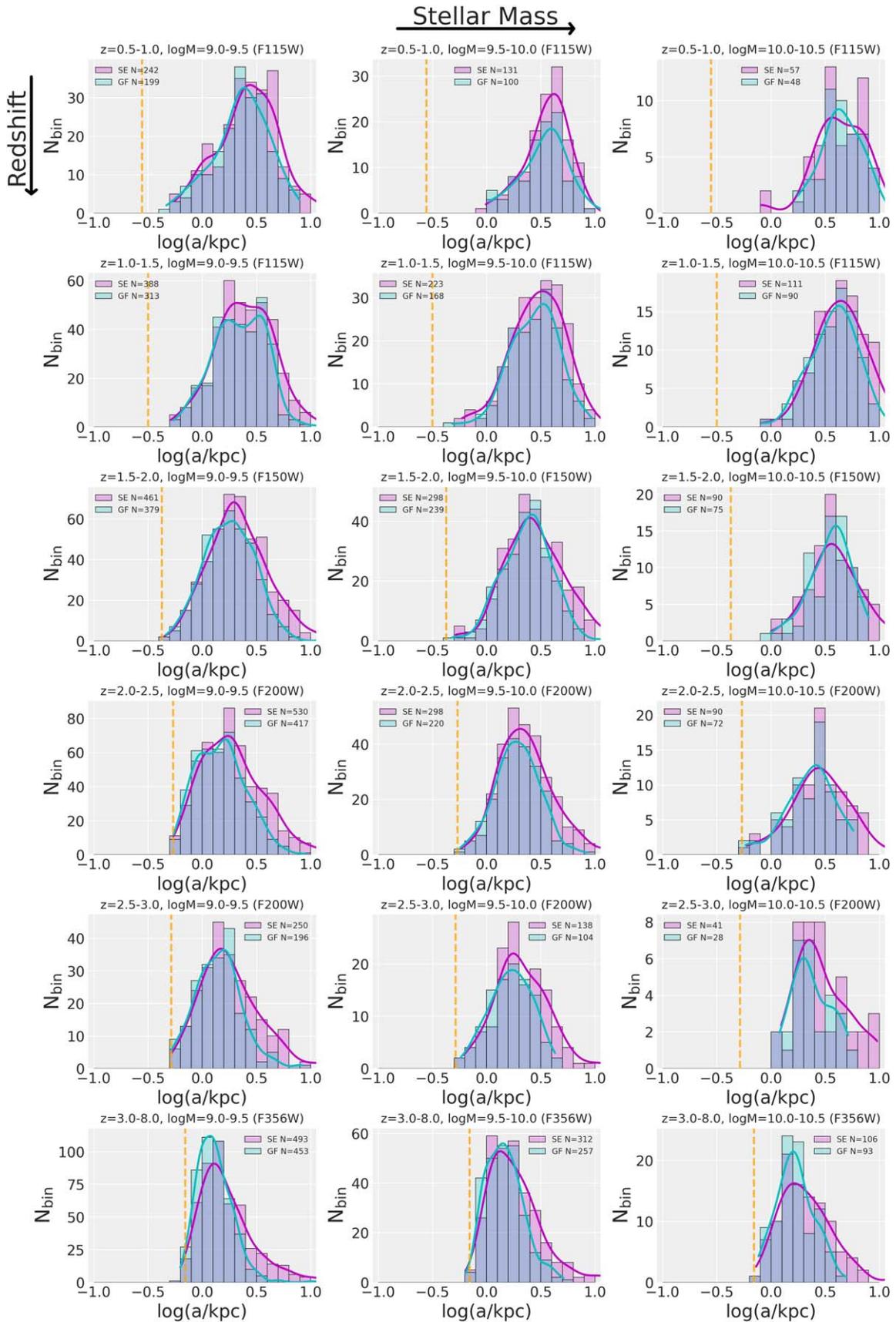
Figure 2. Similar to Figure 1, but now for the size distributions. The dashed vertical lines indicate the PSF FWHM converted to kpc at the midpoint of each redshift bin. As before, the two sets of measurements are similar, but SE++ finds more large-size objects than GALFIT (we discuss this more in Appendix C).
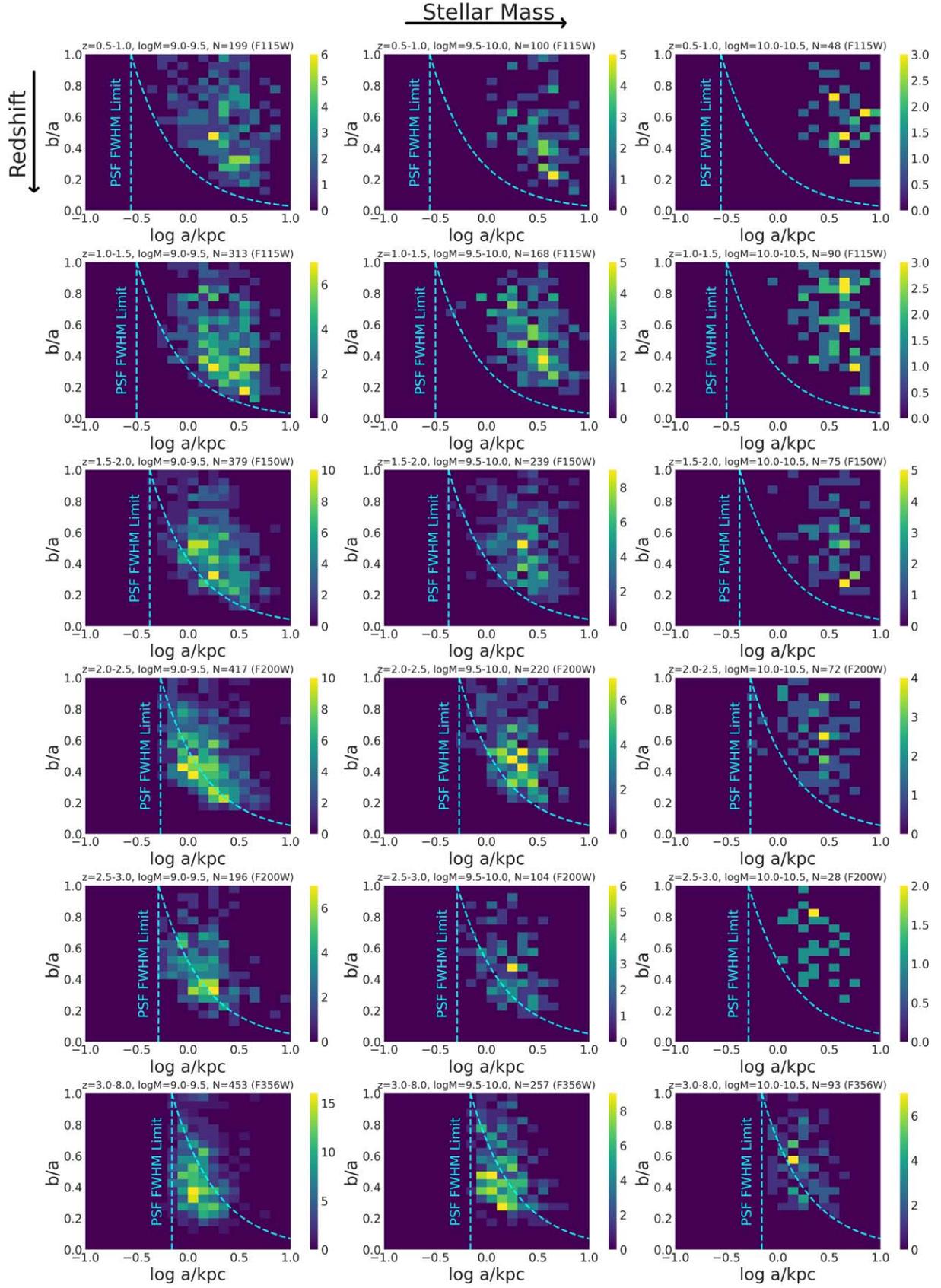
Figure 3. Two-dimensional histograms of projected b/a vs. $\log a$ from GALFIT for our different mass–redshift bins. The color bar denotes the number of galaxies in each histogram cell. The dashed cyan lines denote the PSF FWHM translated to proper kpc at the midpoint of each redshift bin, where the curved lines come from assuming $b =$ PSF FWHM. The sizes and axis ratios are intrinsic, i.e., from the best-fitting Sérsic model before being convolved with the PSF. The lower-mass bins reveal a striking trend in this diagram, namely an excess of low-b/a "edge-on" objects and a deficit of rounder objects. The histograms for some of the higher-mass bins are noisy owing to small sample sizes, so for these we expect our Bayesian model posteriors to reflect the priors.
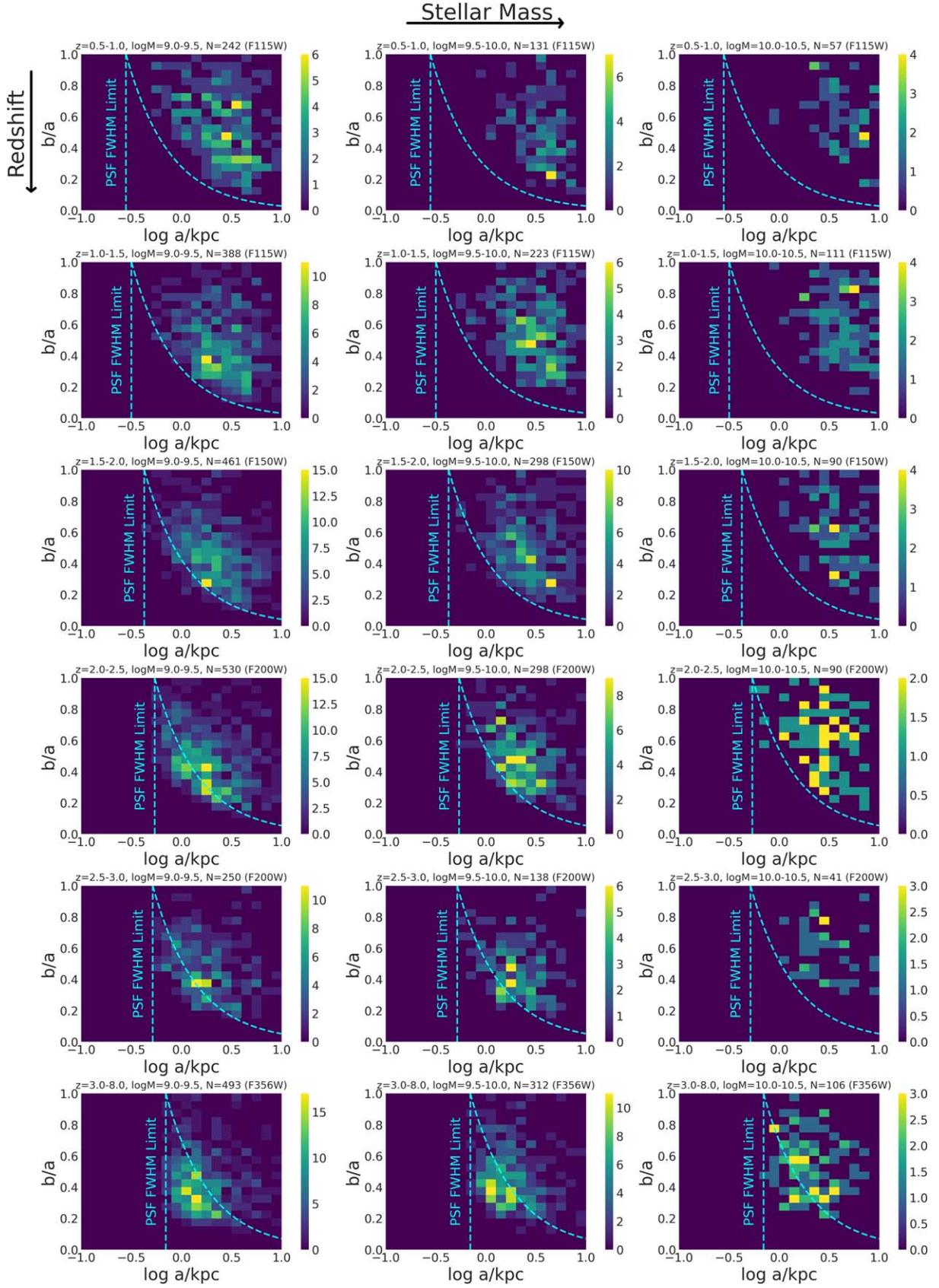
Figure 4. Similar to Figure 3, but now using the SE++ catalog. We see the same trends, namely that the lower-mass, higher-redshift bins have a deficit of rounder objects and an excess of low-b/a sources.

approximated as 3D ellipsoids.[60] All 3D ellipsoids are described by three numbers that relate their three axis lengths $A \geqslant B \geqslant C$: the ellipticity E, triaxiality T, and length of the largest axis in 3D, which we denote $\log A$. These are distinct from the projected quantities, which we denote using lowercase variables $\log a$ and b/a. The ellipticity and triaxiality are respectively defined as

$$E = 1 - C/A \tag{1}$$

and

$$T = \frac{A^2 - B^2}{A^2 - C^2}. \tag{2}$$

With these definitions, a spheroidal ellipsoid would have $E \sim 0$, whereas oblate and prolate ellipsoids would have $E \sim 1$. The triaxiality constrains the intermediate axis ratio B/A and in particular whether $B \sim A$ or $B \sim C$. For a nearly round spheroidal ellipsoid, the value of the triaxiality is not important. However, since oblate and prolate ellipsoids can have the same ellipticity, the triaxiality is needed to break the 3D shape degeneracy. Oblate ellipsoids have $B \sim A$, which means $T \to 0$. On the other hand, prolate ellipsoids have $B \sim C$, so $T \to 1$. Intuitively, oblate ellipsoids have one axis much shorter than the other two, which themselves are similar to each other ($B \sim A$), and so these can also be considered "disky" ellipsoids. In contrast, prolate ellipsoids have one axis much longer than the other two, which themselves are similar to each other ($B \sim C$). Note that if E is large and T is not at one of the extremes, then we have a triaxial ellipsoid where all three axis lengths are considerably different and the object looks like a flattened oval disk.[61]

In practice, the division between different kinds of ellipsoids is arbitrary except for the most extreme cases. We follow Figure 1 of Zhang et al. (2019; see also Figure 2 of van der Wel et al. 2014a) to classify 3D ellipsoids into spheroidal, oblate, or prolate based on where they land in the 3D C/A versus B/A diagram. We will show our exact boundaries later in Section 4.4.

For any 3D ellipsoid, it is straightforward to calculate its projected b/a and $\log a$ given a viewing angle, i.e., a combination of polar angle θ and azimuthal angle ϕ (Binney 1985; van de Ven & van der Wel 2021).[62] We first construct a library of 1 million 3D ellipsoids spanning a range of combinations of E, T, and $\log A$ on a uniform grid.[63] For

each of these 1 million 3D ellipsoids, we sample 100,000 random viewing angles, i.e., pairs of $(\cos \theta, \phi)$ drawn uniformly over the range $-1 < \cos \theta < 1$ and $0 < \phi < 2\pi$. By calculating the projected b/a and $\log a$ for each of these viewing angles, we can construct a 2D histogram of b/a versus $\log a$ for a single 3D ellipsoid, which can be thought of as a probability map for how that ellipsoid would appear in projection. The 2D histogram of projected b/a versus $\log a$ for a mixture of different kinds of 3D ellipsoids can be obtained by summing their individual corresponding 2D probability maps. We follow Section 5 of Chang et al. (2013) to incorporate typical observed uncertainties in b/a and $\log a$ when creating this library of 2D probability maps for all 1 million ellipsoids. Specifically, we assume a typical observational uncertainty on projected axis ratios of $\Delta(b/a) = 0.04$ and then use a Rice distribution (see Appendix C of Rix & Zaritsky 1995) to convert the true projected b/a for every ellipsoid seen from any viewing angle into a random measured b/a. We then use the ratio of the true and randomly drawn b/a to rescale the true projected $\log a$ into a randomly measured uncertain size. We further smear the predicted $\log a$ by a Gaussian of width 0.03 dex.

Figure 5 illustrates the differences between the 2D histograms of b/a versus $\log a$ for four different types of ellipsoids. Spheroidal ellipsoids would appear round from any viewing angle and thus lie at the upper part of this diagram. Oblate ellipsoids would appear round when viewed "face-on" and thin when viewed "edge-on," but they are equally likely to be observed from any viewing angle and thus show a uniform distribution in b/a. The finite thickness (intrinsic C/A) of oblate ellipsoids means that they will show an abrupt truncation at low projected b/a as seen in the figure.

Prolate ellipsoids are different. Since they have two short axes, they are more likely to be observed "edge-on," with the longest axis seen in projection. Thus, most of the projections of a prolate ellipsoid will have low b/a and large $\log a$. A similar trend is expected for flattened oval (triaxial) disks, which, even when seen face-on, would not appear circular and therefore also tend to have $b/a < 1$. The b/a–$\log a$ diagrams for both prolate ellipsoids and oval (triaxial) disks mimic the appearance of a banana, and so we refer to these as "banana diagrams" throughout this paper.

3.2. Multivariate Normal Model

In practice, a population of observed galaxies will be made up of a mixture of 3D ellipsoids with a variety of intrinsic shapes. Thus, we need to decompose the observed 2D histogram of b/a versus $\log a$ into a probability-weighted sum of the 2D projected histograms for all 1 million ellipsoids in our toy library. We use a Bayesian model to accomplish this.

Suppose we have N observed galaxies in a single mass–redshift bin. We say that these N observed galaxies are drawn from N 3D ellipsoids, each of which is characterized by its intrinsic shape vector $\boldsymbol{\theta} = (E, T, \log A)$. We further assume that this $\boldsymbol{\theta}$ vector is distributed as a 3D Gaussian with unknown mean vector

$$\boldsymbol{\mu} = (\mu_E, \mu_T, \mu_{\log A}) \tag{3}$$

---

[60] This standard assumption may not be a good one for high-redshift galaxies whose light profiles tend to be clumpy and asymmetric. However, even if such galaxies are not well described by 3D ellipsoids, the simplicity of the inferred parameters still makes them useful as well-defined summary statistics of the light distribution.

[61] We will use the term "oval disk" throughout since it is easier to visualize, but such an object can also be called a triaxial ellipsoid or nonaxisymmetric disk. In a circular disk, the stars are expected to be on circular orbits, but that cannot be the case in an oval disk, so the two are also kinematically different.

[62] We are assuming that the ellipsoids are transparent, but this may not be the case for real galaxies since dust attenuation likely depends on galactocentric distance and viewing angle and because high-redshift systems typically do not have uniformly smooth light distributions.

[63] Uniformly sampling in the $(E, T, \log A)$ space does not translate to a uniform sample in $(C/A, B/A, \log A)$ space. The former leads to many more spheroids with high C/A and B/A. These two sampling strategies can be thought of as different priors for generating the library of toy ellipsoids and can influence posteriors in the limit of small sample sizes. This may impact our massive bins but not our key results regarding low-mass bins.
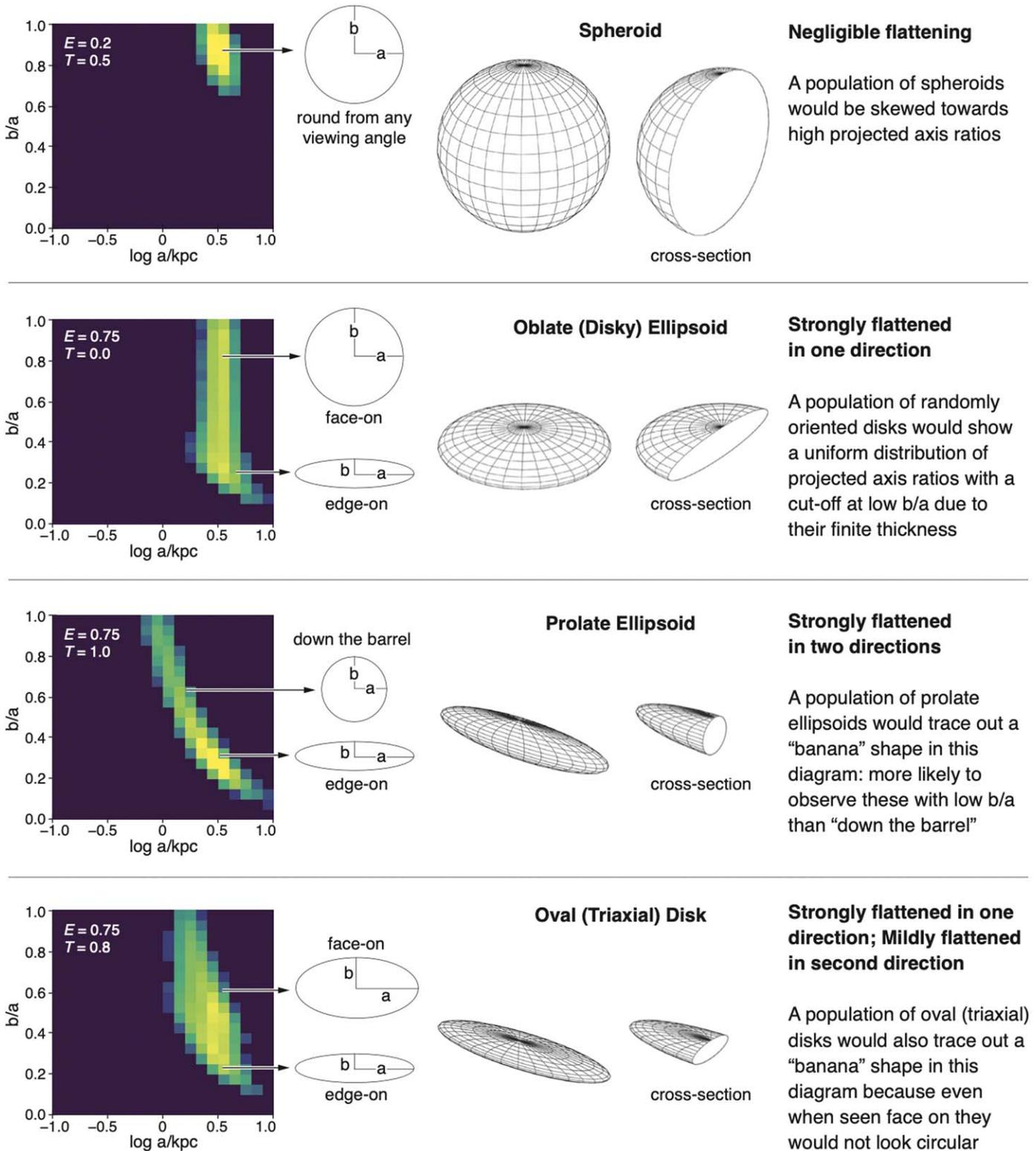
Figure 5. An illustration of how a population of extreme ellipsoids would appear in projection on the b/a–$\log a$ diagram. The histograms on the left are for a single ellipsoid seen from many different viewing angles accounting for measurement errors. All four ellipsoids have the same intrinsic $\log A/\mathrm{kpc} = 0.5$. The histogram color corresponds to the fraction of projections that end up in a given b/a–$\log a$ cell, with purple being very low and bright yellow being very high. For each type of ellipsoid, we show a 3D visualization along with a cross section, where the latter clearly differentiates between prolate systems and circular vs. oval disks. Face-on and edge-on projections are also shown corresponding to locations in the histograms. Top: spheroids would be concentrated at large projected b/a because they appear round from any viewing angle. Second from top: oblate/disky objects trace out a uniform distribution in b/a because they are equally likely to be seen from any viewing angle. They have a cutoff at low axis ratios due to their finite thickness. Third from top: prolate ellipsoids would trace out a banana in this diagram since they have two short axes, leading to most projections appearing "edge-on" with small b/a and large size. Bottom: oval (triaxial) disks would also preferentially have lower b/a in projection because, even when seen face-on, they would not be perfectly circular.
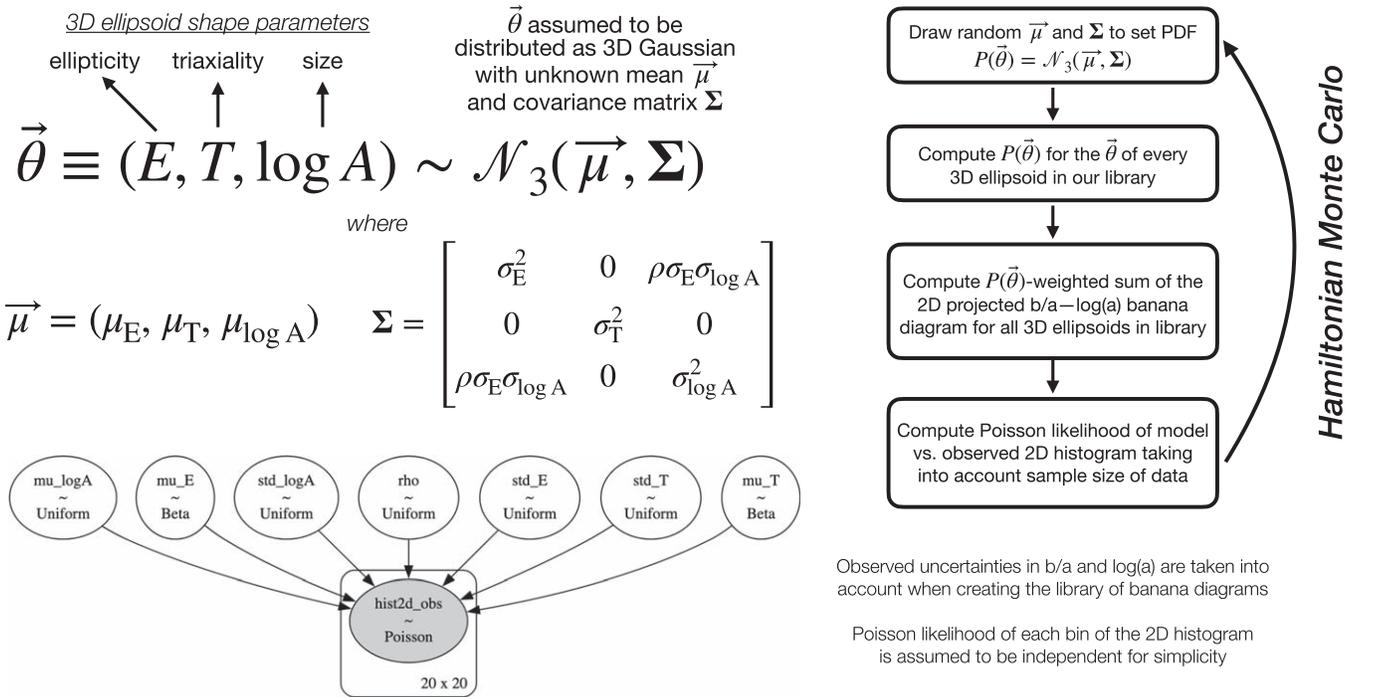
11

3D ellipsoid shape parameters

ellipticity   triaxiality   size

$$\vec{\theta} \equiv (E, T, \log A) \sim \mathcal{N}_3(\overrightarrow{\mu}, \mathbf{\Sigma})$$

$\vec{\theta}$ assumed to be distributed as 3D Gaussian with unknown mean $\overrightarrow{\mu}$ and covariance matrix $\mathbf{\Sigma}$

where

$$\overrightarrow{\mu} = (\mu_E, \mu_T, \mu_{\log A}) \qquad \mathbf{\Sigma} = \begin{bmatrix} \sigma_E^2 & 0 & \rho\sigma_E\sigma_{\log A} \\ 0 & \sigma_T^2 & 0 \\ \rho\sigma_E\sigma_{\log A} & 0 & \sigma_{\log A}^2 \end{bmatrix}$$

mu_logA ~ Uniform | mu_E ~ Beta | std_logA ~ Uniform | rho ~ Uniform | std_E ~ Uniform | std_T ~ Uniform | mu_T ~ Beta

hist2d_obs ~ Poisson   20 x 20

Hamiltonian Monte Carlo

Draw random $\overrightarrow{\mu}$ and $\mathbf{\Sigma}$ to set PDF $P(\vec{\theta}) = \mathcal{N}_3(\overrightarrow{\mu}, \mathbf{\Sigma})$

Compute $P(\vec{\theta})$ for the $\vec{\theta}$ of every 3D ellipsoid in our library

Compute $P(\vec{\theta})$-weighted sum of the 2D projected b/a—log(a) banana diagram for all 3D ellipsoids in library

Compute Poisson likelihood of model vs. observed 2D histogram taking into account sample size of data

Observed uncertainties in b/a and log(a) are taken into account when creating the library of banana diagrams

Poisson likelihood of each bin of the 2D histogram is assumed to be independent for simplicity

Figure 6. A visualization of our Bayesian model. For observed galaxies in any given mass–redshift bin, the corresponding distribution of 3D ellipsoid shape properties is assumed to follow a 3D multivariate normal with seven unknown parameters describing the mean ellipticity, triaxiality, and size, as well as the covariance matrix. The directed graph in the lower left corner shows our priors for these seven parameters and our assumption of a Poisson likelihood for fitting the model to the observed 2D histogram of b/a vs. $\log a$. The flowchart on the right illustrates our procedure for Hamiltonian Monte Carlo.

and unknown covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_E^2 & 0 & \rho\sigma_E\sigma_{\log A} \\ 0 & \sigma_T^2 & 0 \\ \rho\sigma_E\sigma_{\log A} & 0 & \sigma_{\log A}^2 \end{bmatrix} \tag{4}$$

with standard deviations σ_E, σ_T, and $\sigma_{\log A}$. Following previous work, we allow for a covariance ρ between the ellipticity and the size. In practice, we can also introduce covariances between the ellipticity and triaxiality and between the triaxiality and size. This does not appear to change our results, though it does require more sophisticated sampling methods and can slow down convergence as discussed in the next subsection. We believe that the current model is flexible enough to encompass a wide spread in 3D shapes despite treating all galaxies in a given mass–redshift bin as a single population, but we will discuss alternative approaches in Section 6.3.

Intuitively, this model says that our N observed galaxies are drawn randomly from a population of 3D ellipsoids that are characterized by their mean shape parameters along with intrinsic scatter around those means (as well as covariance between the mean ellipticity and mean size). As shown in the flowchart on the right side of Figure 6, for a given choice of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, we can assign a probability to each of the 1 million ellipsoids in our toy library. Then, we can do a probability-weighted sum of their corresponding 2D histograms of b/a versus $\log a$. If the data have sufficient constraining power (and if the model is flexible and robust enough), then we can use the observed 2D histogram of b/a versus $\log a$ to infer the relative contributions of 3D ellipsoids of different types to the observed population.

3.3. Hamiltonian Monte Carlo

We implement the above model in the probabilistic programming framework PyMC (Salvatier et al. 2015), which provides many different samplers for Bayesian parameter inference. In particular, we take advantage of its "automatic differentiation" capability to rapidly and exactly compute the gradient of our model likelihood with respect to all seven of its free parameters. Unlike symbolic differentiation or finite-difference methods, automatic differentiation involves translating code into a computational graph and keeping track of not only the results of mathematical operations but also their partial derivatives for use with the chain rule (see Baydin et al. 2018, for a recent review). This allows us to leverage a powerful sampling technique known as Hamiltonian Monte Carlo (HMC), which uses the gradient of the likelihood to more efficiently explore high-dimensional parameter spaces compared to traditional Markov Chain Monte Carlo methods. We use a specific implementation of HMC called the No U-Turn Sampler (NUTS; Hoffman & Gelman 2014).

We assume uniform priors for all seven free parameters. Specifically, μ_E and μ_T both have a uniform prior between 0 and 1 (for numerical robustness, we use a Beta distribution with $\alpha = \beta = 1$, which is just the uniform distribution). For $\mu_{\log A}$ we assume a uniform distribution between -1 and 1 dex. For the three standard deviations, we assume a uniform prior between 0 and 1. Finally, for the correlation coefficient $\rho(E, \log A)$, we assume a uniform prior between -1 and 1. As mentioned in the previous subsection, we experimented with fitting for the full covariance matrix directly using the sophisticated "Lewandowski–Kurowicka–Joe" (LKJ; Lewandowski et al. 2009) prior, which is optimized for Bayesian sampling methods. The LKJ prior samples the Cholesky decomposition (lower triangular matrix) of the covariance matrix and specifies how

much Σ deviates from the identity matrix. The advantage of the LKJ prior is that we can also fit for $\rho(E, T)$ and $\rho(T, \log A)$. In practice, for the few cases we tried, the LKJ prior gave similar results to our fiducial model but was slower to converge (especially with our relatively small sample sizes).

We assume a Poisson likelihood for comparing the observed and model 2D histograms of b/a versus $\log a$. For simplicity, we assume that all of the cells of the observed 2D histogram are independent, so that we can add the log-likelihood of all of them together to estimate the goodness of fit for any individual model realization. In principle, we should allow for the possibility that uncertainties in b/a and $\log a$ can mean that galaxies may contribute to other nearby cells of the 2D histogram compared to the one they are assigned to. However, we already accounted for the typical observed errors in b/a and $\log a$ when constructing our library of banana diagrams, so this potential smearing is already included in the model. In addition, the typical errors in b/a and $\log a$ should be smaller than our bin widths of $\Delta b/a = 0.05$ and $\Delta \log a = 0.1$ dex, especially for SE++, where we imposed a fractional uncertainty threshold of 10%.

We fit each mass–redshift bin independently. We use 5000 tuning (burn-in) draws and 2000 sampling draws with four chains in parallel. This is typically more than adequate for HMC/NUTS. We use the recommended `target_accept = 0.95` value for NUTS, which means smaller adaptive step sizes and makes it easier for the sampler to explore the potentially complicated posterior (particularly important in cases where we have only a few tens of objects). With these options, we do not get any catastrophic divergences during NUTS sampling and all four independent chains converge to the same posteriors, even for the mass–redshift bins with small sample sizes (though, of course, in these cases the posteriors for some parameters can be relatively unconstrained). In Appendix A, we show mock tests where NUTS succeeds in constraining all parameters for sample sizes >50. For smaller sample sizes, all parameters except μ_T and σ_T can still be constrained, with the latter showing very broad posteriors.

Figure 7 shows an example corner plot for the $z = 2.0$–2.5 and $\log M_*/M_\odot = 9.0$–9.5 population with constraints from the full five-field CANDELS data set and CEERS using either GALFIT or SE++. We see that all three models are very well constrained. For this mass–redshift bin, the mean ellipticity is constrained to be high, with $\mu_E \sim 0.75$, and there is a relatively small scatter, $\sigma_E \sim 0.1$, which means that the galaxies are consistent with either disks or prolate systems. The mean triaxiality is also constrained to be $\gg 0.75$ (with CANDELS hitting up against ~ 1), which strongly favors the prolate interpretation. The corresponding scatter is $\sigma_T \sim 0.25$–0.75, which means that there must also be some contribution from disks. Although the three models seem to show discrepancies with each other, we stress that these are relatively minor: the differences between their constrained posteriors are much smaller than the ranges of the uniform priors for all parameters. In other words, the different models reach similar regions of the enormous 7D parameter space despite discrepancies in the data, which suggests that our conclusions are robust. We also show that the observed histogram and mean model histogram agree well with each other and that the residual histograms are relatively featureless regardless of data set used. This also holds for our other mass–redshift bins, although some of the massive bins can show greater residuals owing to their small sample

sizes and perhaps the limited flexibility of the model (we discuss this more in Section 6.3).

3.4. Individual Ellipsoid Classification Probabilities

Figure 8 demonstrates that we can use our constrained model parameters to assign 3D ellipsoid classification probabilities for individual observed galaxies (see also Section 5.1 of Zhang et al. 2019). For illustrative purposes, we use the means of the posteriors of each parameter to reconstruct the mean model 2D histogram of b/a versus $\log a$. We can decompose the total mean model histogram into the relative contributions from prolate, oblate, and spheroidal ellipsoids using the boundaries on the C/A versus B/A diagram described in Section 4.4 (these are adapted from van der Wel et al. 2014a; Zhang et al. 2019).

For the particular low-mass, high-redshift bin shown in Figure 8 (see also analogous figures for the other mass–redshift bins in the corresponding figure set), prolate ellipsoids dominate at low b/a, oblate ellipsoids dominate for large b/a and large $\log a$, and spheroids are negligible. As a result, most of the individual observed galaxies with low b/a have >75% probability of being prolate in 3D. In contrast, observed galaxies in the upper right corner are given very high probability of being (face-on) disks. Of course, we cannot say for sure whether any individual galaxy is indeed prolate from this kind of statistical imaging-based analysis alone, but it is a first step toward more detailed comparative analyses and facilitating follow-up observational campaigns. Lastly, here we are using the means of the posteriors, but instead we could also do random draws of model parameter combinations from the posterior and then construct the mean model histogram that way (this would also provide a way to assign uncertainties on classifications).

Given the small sample sizes for many of our massive $\log M_*/M_\odot = 10$–10.5 bins, the posteriors for μ_T and σ_T tend not to be constrained. As a result, in these bins, the classification probabilities for most galaxies are not very high and may be split roughly equally between all three ellipsoid classes. In any comparative analyses in Section 5, we will only take galaxies with a >75% probability of being assigned to one of the three ellipsoid classes. This is, of course, arbitrary, but it allows us to focus on objects in mass–redshift bins with the greatest available constraints. After making this cut on our model based on SE++, we end up with 1806 prolate candidates, 220 oblate candidates, and 73 spheroidal candidates irrespective of mass and redshift. We remind the reader that we are only considering star-forming galaxies. As sample sizes increase, we can expect a larger fraction of observed galaxies to be assigned higher class probabilities with this technique.

3.5. Nonparametric Morphological Measurements

For individual observed galaxies with >75% probability of being assigned to one of the three ellipsoid classes, we will want to look for other signatures that may discriminate between prolate, oblate, and spheroidal objects. To this end, we use the publicly available `statmorph` Python package (Rodriguez-Gomez et al. 2019) to measure several nonparametric morphological properties. For each of our high-probability candidates, we create $3'' \times 3''$ cutouts of the science and error images in the relevant filter that probes rest-frame optical wavelengths (see Table 1). We also read in the empirical PSF created from stacking stars in the NIRCam fields for the relevant filter (Finkelstein et al. 2023). We create
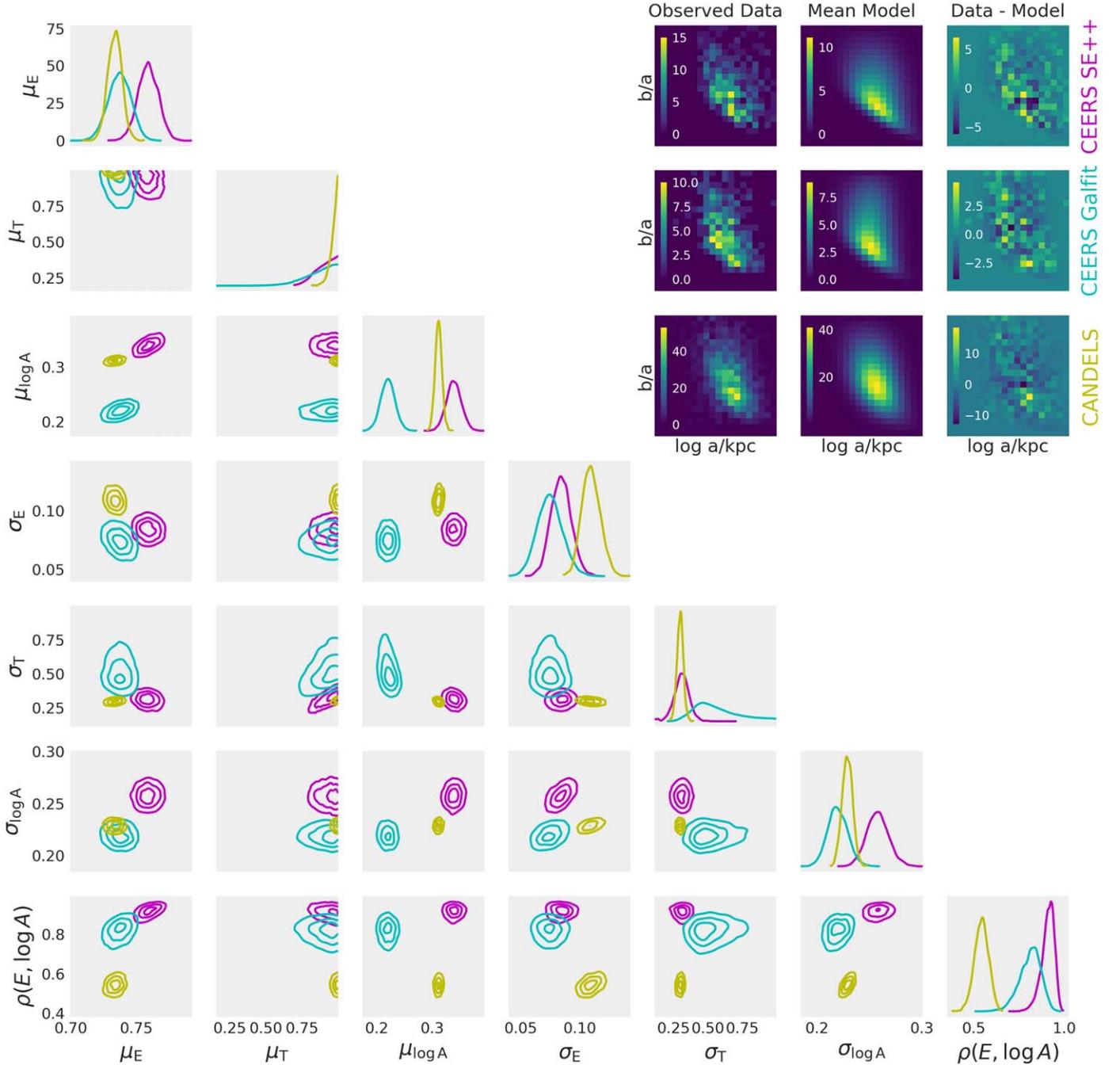
Figure 7. Example corner plot from our HMC for the $z = 2.0$–2.5 and $\log M_*/M_\odot = 9.0$–9.5 bin. The different colors correspond to the different data sets used for the fitting: SE++ (magenta), GALFIT (cyan), and CANDELS (yellow). Results from the four individual HMC chains for each run have been combined since the chains were all converged. The CEERS posteriors agree relatively well with each other and with CANDELS (i.e., the models are constrained to be in similar regions of the large 7D parameter space). The grid of histograms in the upper right corner shows that the mean model matches each observed data set well and that the residual map is relatively featureless (the inset color bars show the number of galaxies per histogram bin). CANDELS gives the tightest constraints because of the much larger sample size at $z < 2.5$. For this mass–redshift bin, both the mean ellipticity and mean triaxiality are high, which suggests a predominantly prolate population. Analogous figures for the other mass–redshift bins can be found at the Harvard Dataverse: doi:10.7910/DVN/SWTKVA .

our own regularized segmentation map with `photutils` (Bradley et al. 2022) in just the cutout region using a pixel detection threshold of 1.5σ above the error map while ensuring that only the main object of interest in the center of the cutout will be fit. Removing objects with statmorph flag > 1, we are left with 1766/1806 prolate candidates, 201/220 oblate candidates, and 73/73 spheroidal candidates using the SE++-based model. Visual inspection of

the failed fits reveals bright neighbors or artifacts, while the successful fits all look reasonable.

We focus on the concentration C, asymmetry A, clumpiness (smoothness S), Gini coefficient G, and second-order moment of the 20% brightest pixels M_{20}. These are defined in Lotz et al. (2004; see also Abraham et al. 2003; Conselice 2003), but we briefly summarize here. The concentration C reflects the ratio of the circular radii containing 80% and 20% of the light,
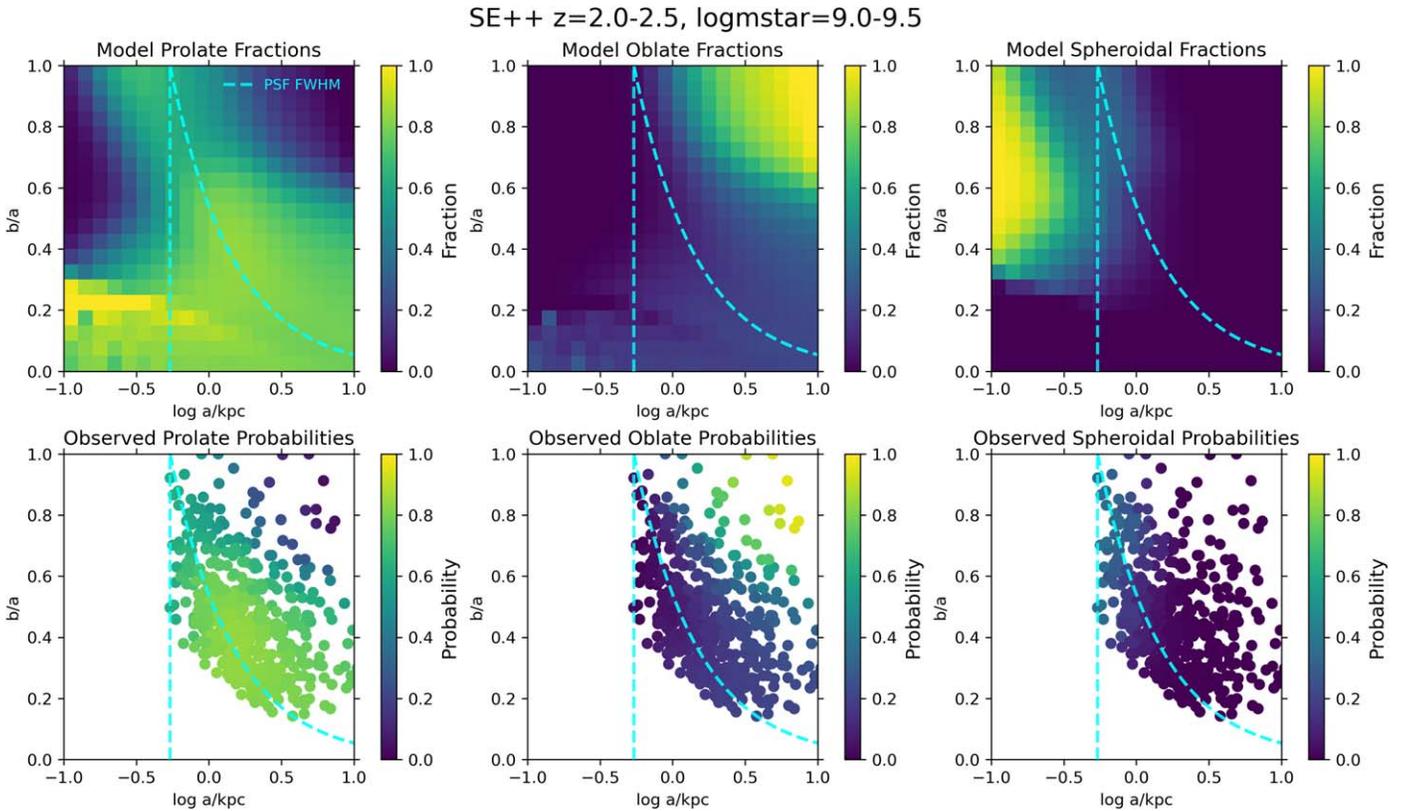
Figure 8. Illustration of how we assign 3D ellipsoid classification probabilities to individual observed galaxies using the $z = 2.0$–2.5, $\log M_*/M_\odot$ SE++ model as an example. Top row: the fractional contribution of each 3D ellipsoid across the b/a–$\log a$ diagram. Prolate ellipsoids dominate in the lower right and oblate ellipsoids in the upper right, whereas spheroids are negligible in this mass–redshift bin. Bottom row: 3D ellipsoid classification probabilities for individual observed galaxies depending on what region of the 2D histogram they fall in. Many galaxies with low b/a have >75% of being prolate, and there are also some high-probability disks in the upper right. Note that any features to the left of the PSF FWHM resolution limit (vertical cyan line) are numerical artifacts and we have no observed galaxies there. Analogous figures for the other mass–redshift bins based on SE++ and Galfit can be found at the Harvard Dataverse: doi:10.7910/DVN/SWTKVA .

respectively (it is another way to measure how concentrated the light profile is akin to the Sérsic index n). The asymmetry A is computed by summing over the residuals after subtracting a $180°$ rotated image from the original image. The clumpiness (smoothness S) similarly sums over the residuals after subtracting a boxcar-smoothed image from the original image with a smoothing scale of $0.25 r_P$ where r_P is the Petrosian radius estimated by statmorph. Thus, lower values of A and S correspond to more symmetric, smooth light distributions. The Gini coefficient G measures how different the distribution of pixel fluxes is from being uniform, with $G = 1$ corresponding to a single pixel containing all of the flux and $G = 0$ meaning that every pixel has the same flux. Finally, the second-order moment of the 20% brightest pixels M_{20} tracks the spatial distribution of the brightest regions relative to the total underlying flux. It is computed by multiplying the flux in the 20% brightest pixels by the square of their distances from the galaxy center and then dividing by the same calculation for all of the galaxy's pixels. The combination of G–M_{20} has been used to separate low-redshift galaxies into mergers, ellipticals, and bulge-dominated systems (Lotz et al. 2008) and has also recently been explored in the context of CEERS visual classification morphologies (Kartaltepe et al. 2023).

4. Results on 3D Shape Evolution

Here we present our constraints on the 3D shapes of JWST-CEERS galaxies as a function of stellar mass and redshift. Table 2 at the end of this section tabulates our results.

4.1. Ellipticity Evolution

In Figure 9, we start by showing the mass–redshift evolution of the mean and standard deviation model ellipticity parameters μ_E and σ_E. For all of our mass–redshift bins, the mean ellipticity is high, with $\mu_E \gtrsim 0.75$, and the scatter is generally small, with $\sigma_E \lesssim 0.1$. These high ellipticities translate to $C/A \sim 0.25$, which is thicker than local disks by a factor of ~ 2–3 (e.g., Elmegreen et al. 2005). We do not see evidence of strong evolution in C/A over the wide range of redshift and mass considered in Figure 9. These results indicate that the majority of the star-forming population that we see in CEERS may be either oblate or prolate in 3D since both configurations would have high ellipticity. The way to break this degeneracy is to constrain the triaxiality parameter, which we will explore below.

But first we show the correlation coefficient between the mean ellipticity and mean 3D size, $\rho(E, \log A)$, in Figure 10. The correlation coefficient is strongly positive in all mass–redshift bins, except perhaps the low-mass bin at $z = 3$–8 based on GALFIT. This means that larger galaxies tend to have higher ellipticity, i.e., larger galaxies have a greater likelihood of being either prolate or oblate. It makes sense that the largest galaxies we see would be disks or prolate since we are studying star-forming galaxies and are not going to such high masses that we would be in the quiescent elliptical regime. In addition, while we believe that this result is robust against PSF resolution effects, we will discuss this caveat in Section 6.3.

Table 2
Our 3D Shape Modeling Results Based on SE++ Structural Measurements

z	$\log M_*/M_\odot$	μ_E	μ_T	$\mu_{\log A}$	σ_E	σ_T	$\sigma_{\log A}$	$\rho(E, \log A)$	f_{prolate}	f_{oblate}	$f_{\mathrm{spheroidal}}$
0.5–1.0	9.0–9.5	0.795 ± 0.055	0.500 ± 0.258	0.542 ± 0.063	0.197 ± 0.038	0.720 ± 0.187	0.318 ± 0.036	0.886 ± 0.051	0.312 ± 0.055	0.447 ± 0.061	0.241 ± 0.037
0.5–1.0	9.5–10.0	0.789 ± 0.034	0.628 ± 0.245	0.626 ± 0.037	0.087 ± 0.037	0.700 ± 0.190	0.193 ± 0.028	0.826 ± 0.129	0.425 ± 0.072	0.519 ± 0.073	0.056 ± 0.043
0.5–1.0	10.0–10.5	0.742 ± 0.099	0.567 ± 0.280	0.836 ± 0.087	0.258 ± 0.118	0.520 ± 0.276	0.326 ± 0.061	0.521 ± 0.256	0.269 ± 0.119	0.280 ± 0.155	0.451 ± 0.103
1.0–1.5	9.0–9.5	0.782 ± 0.011	0.848 ± 0.109	0.458 ± 0.017	0.085 ± 0.012	0.523 ± 0.160	0.251 ± 0.014	0.857 ± 0.059	0.536 ± 0.065	0.412 ± 0.063	0.051 ± 0.020
1.0–1.5	9.5–10.0	0.707 ± 0.017	0.502 ± 0.253	0.537 ± 0.023	0.102 ± 0.023	0.717 ± 0.188	0.253 ± 0.019	0.645 ± 0.121	0.342 ± 0.060	0.430 ± 0.072	0.228 ± 0.062
1.0–1.5	10.0–10.5	0.686 ± 0.060	0.260 ± 0.221	0.741 ± 0.073	0.173 ± 0.080	0.390 ± 0.286	0.281 ± 0.044	0.592 ± 0.198	0.106 ± 0.103	0.467 ± 0.165	0.426 ± 0.101
1.5–2.0	9.0–9.5	0.772 ± 0.013	0.834 ± 0.079	0.410 ± 0.015	0.098 ± 0.013	0.199 ± 0.074	0.252 ± 0.012	0.794 ± 0.071	0.811 ± 0.072	0.121 ± 0.058	0.068 ± 0.027
1.5–2.0	9.5–10.0	0.751 ± 0.011	0.602 ± 0.243	0.488 ± 0.018	0.069 ± 0.016	0.768 ± 0.158	0.254 ± 0.015	0.788 ± 0.091	0.417 ± 0.051	0.516 ± 0.057	0.068 ± 0.038
1.5–2.0	10.0–10.5	0.725 ± 0.094	0.418 ± 0.293	0.669 ± 0.071	0.249 ± 0.124	0.576 ± 0.264	0.260 ± 0.048	0.693 ± 0.158	0.212 ± 0.106	0.367 ± 0.153	0.421 ± 0.093
2.0–2.5	9.0–9.5	0.759 ± 0.008	0.905 ± 0.066	0.339 ± 0.013	0.084 ± 0.009	0.309 ± 0.056	0.258 ± 0.011	0.908 ± 0.036	0.726 ± 0.051	0.204 ± 0.045	0.069 ± 0.018
2.0–2.5	9.5–10.0	0.721 ± 0.011	0.872 ± 0.083	0.427 ± 0.015	0.063 ± 0.015	0.314 ± 0.082	0.214 ± 0.012	0.767 ± 0.103	0.700 ± 0.078	0.200 ± 0.070	0.100 ± 0.040
2.0–2.5	10.0–10.5	0.722 ± 0.092	0.374 ± 0.253	0.538 ± 0.073	0.244 ± 0.102	0.482 ± 0.269	0.296 ± 0.051	0.750 ± 0.163	0.163 ± 0.102	0.411 ± 0.149	0.425 ± 0.092
2.5–3.0	9.0–9.5	0.756 ± 0.011	0.868 ± 0.084	0.313 ± 0.018	0.076 ± 0.014	0.305 ± 0.087	0.244 ± 0.014	0.852 ± 0.078	0.715 ± 0.077	0.221 ± 0.072	0.064 ± 0.028
2.5–3.0	9.5–10.0	0.753 ± 0.011	0.886 ± 0.083	0.380 ± 0.021	0.035 ± 0.015	0.320 ± 0.106	0.211 ± 0.017	0.680 ± 0.262	0.758 ± 0.101	0.229 ± 0.100	0.014 ± 0.018
2.5–3.0	10.0–10.5	0.644 ± 0.034	0.474 ± 0.259	0.533 ± 0.057	0.057 ± 0.048	0.555 ± 0.264	0.261 ± 0.048	0.451 ± 0.431	0.263 ± 0.120	0.323 ± 0.149	0.414 ± 0.145
3.0–8.0	9.0–9.5	0.792 ± 0.007	0.916 ± 0.060	0.286 ± 0.012	0.023 ± 0.009	0.281 ± 0.050	0.214 ± 0.008	0.576 ± 0.258	0.796 ± 0.051	0.203 ± 0.051	0.000 ± 0.001
3.0–8.0	9.5–10.0	0.755 ± 0.009	0.912 ± 0.065	0.311 ± 0.014	0.042 ± 0.012	0.317 ± 0.063	0.201 ± 0.010	0.565 ± 0.166	0.761 ± 0.066	0.220 ± 0.065	0.018 ± 0.015
3.0–8.0	10.0–10.5	0.716 ± 0.020	0.568 ± 0.238	0.355 ± 0.024	0.044 ± 0.026	0.559 ± 0.261	0.216 ± 0.018	–0.087 ± 0.461	0.438 ± 0.118	0.471 ± 0.134	0.091 ± 0.077

Note. For all quantities, we report the mean and standard deviation of random draws from the posterior. This table and an analogous one based on the GALFIT structural measurements can be downloaded at the Harvard Dataverse, doi:10.7910/DVN/SWTKVA, or in the online journal in the machine-readable format.

(This table is available in its entirety in machine-readable form.)
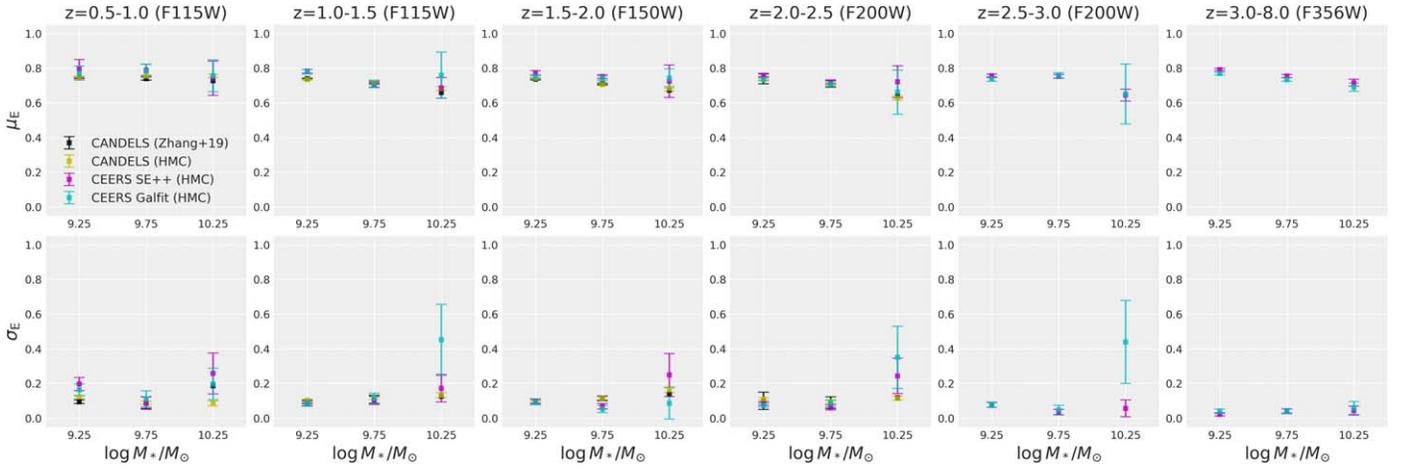
Figure 9. The evolution of the mean ellipticity (top row) and its standard deviation (bottom row). Each column corresponds to a different redshift bin increasing from left to right as indicated by the subplot titles. The black and yellow points show results at $z \leqslant 2.5$ from Zhang et al. (2019) and our new HMC code applied to all five CANDELS fields combined, respectively. The magenta and cyan points show results from our code applied to JWST-CEERS shape catalogs from SE++ and GALFIT, respectively. The error bars denote the 1σ width of the marginalized posterior from our code. Note how the mean ellipticity is well constrained to be high with $\sigma_E \lesssim 0.3$ for all mass–redshift bins that we consider, indicating that either disks or prolate galaxies dominate.
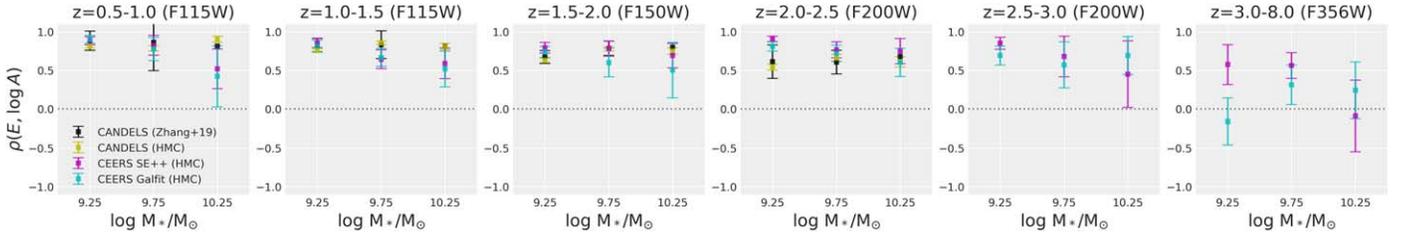
Figure 10. Similar to Figure 9, but now for the evolution of the correlation coefficient between ellipticity and 3D size. In general, the correlation coefficient is positive and consistent with $\gtrsim 0.5$, indicating that larger galaxies tend to have higher ellipticity, which in turn means that larger galaxies are either disky or prolate rather than spheroidal.
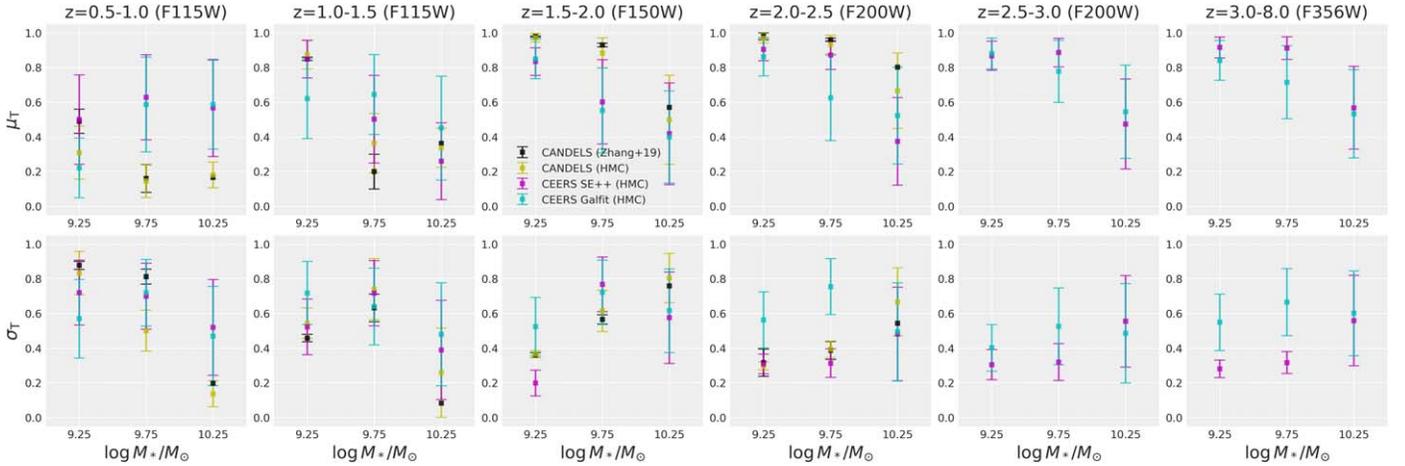
Figure 11. Similar to Figure 9, but now for the evolution of the mean (top) and standard deviation (bottom) of the triaxiality. The value of the triaxiality can break the degeneracy between interpreting high-ellipticity objects as either disky (low triaxiality) or prolate (high triaxiality). Our CEERS modeling extends the CANDELS trend of low-mass, high-redshift galaxies having high triaxialities and thus prolate shapes albeit with larger error bars. Likewise, higher-mass and/or lower-redshift galaxies are consistent with lower triaxialities and thus disky shapes. The standard deviations are generally consistent with $\sigma_T \sim 0.5$.

4.2. Triaxiality Evolution

Figure 11 answers the key question about the mass–redshift evolution of the mean triaxiality μ_T and its standard deviation σ_T. Recall from Equation (2) that disks that are nearly oblate/axisymmetric have low triaxiality ($T \approx 0$), whereas nearly prolate systems have high triaxiality ($T \approx 1$). All of the low-mass ($\log M_*/M_\odot = 9.0$–9.5) bins are consistent with $\mu_T \gg 0.8$

at $z > 1$, strongly favoring the prolate interpretation. Several of the intermediate-mass ($\log M_*/M_\odot = 9.5$–10.0) bins are also consistent with high triaxiality, and this becomes more pronounced at higher redshift ($z > 2$). The massive ($\log M_*/M_\odot = 10.0$–10.5) bins tend to have lower triaxiality and especially at lower redshifts are consistent with $\mu_T \lesssim 0.2$, indicative of oblate/disky 3D geometries. The scatter in the triaxiality is relatively high, with $\sigma_T \gtrsim 0.5$ in many mass–
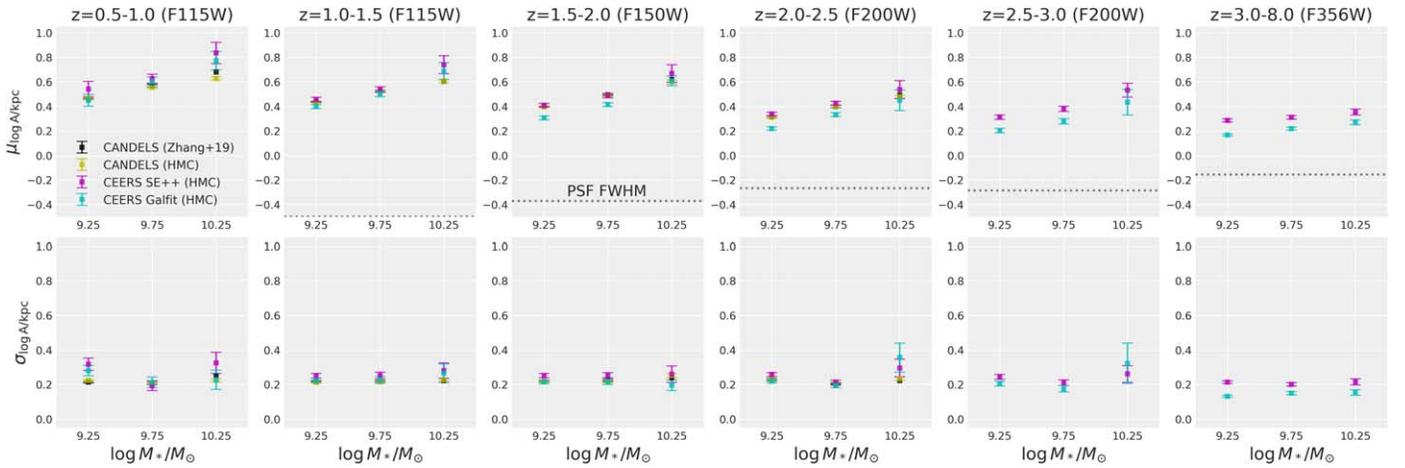
Figure 12. Similar to Figure 9, but now for the evolution of the mean (top) and standard deviation (bottom) of the 3D size–mass relation. The dotted horizontal black lines in the top panels show the PSF FWHM in the relevant filter at the midpoint of each redshift bin. There is a clear evolution in the 3D size–mass relation such that more massive galaxies in a given redshift bin are larger and the size–mass relation overall decreases toward high redshift. The scatter in the 3D size–mass relation is remarkably constant at $\sigma_{\log A} \sim 0.2$ with both mass and redshift.

redshift bins. This means that even in cases where μ_T is at one of the extremes (0 or 1), there can still be a substantial contribution from other types of ellipsoids. Thus, we need to combine the joint constraints on ellipticity and triaxiality to infer the relative fractions of ellipsoids of different types as we will show later. Nevertheless, the fact that in some bins we are seeing very high μ_T strongly suggests that there is a pattern in the data driving us toward high-redshift dwarfs being prolate in 3D.

4.3. 3D Size–Mass Relations

Figure 12 shows that our approach automatically also gives us the 3D size–mass relations, which are otherwise difficult to retrieve observationally. By 3D size, we mean the longest axis of the ellipsoid that corresponds to the deprojected 2D ellipse that encloses half of the Sérsic model light distribution. We see that the mean 3D size $\mu_{\log A}$ is larger for higher-mass galaxies at fixed redshift. We also see that as one goes to higher redshifts, the mean sizes decrease systematically: naturally, galaxies are getting smaller in size at fixed mass at earlier times. In other words, we are recovering the growth of galaxy size, but now with 3D size–mass relations based on JWST data. There is a hint that the 3D size–mass relation flattens out for $z = 3$–8. The scatter in the 3D size–mass relation is remarkably small and constant with mass and redshift at $\sigma_{\log A} \sim 0.2$ dex, which consistent with previous work (van der Wel et al. 2014b).

As has also been shown by Suess et al. (2022), our GALFIT-based model recovers the striking trend that galaxies on average appear smaller in JWST NIRCam imaging than they did in HST WFC3. The discrepancy becomes worse for lower-mass, higher-redshift bins, which would correspond to fainter galaxies for whom deblending and related issues would be preferentially important. We discuss this more in Appendix C. We stress that despite the ~0.1 dex offset in the 3D size–mass relation derived from SE++ and Galfit measurements, overall our Bayesian model is still reaching roughly similar regions of its enormous 7D parameter space, and so our conclusions about 3D shapes should be robust.

4.4. 3D Axis Ratios

Figure 13 shows the distribution of 3D axis ratios C/A versus B/A for model ellipsoids in each mass–redshift bin. These 2D histograms are the means of 500 random draws from the posterior for each mass–redshift bin. The yellow curves mark the (arbitrary) boundaries between prolate, oblate, and spheroidal systems following van der Wel et al. (2014a) and Zhang et al. (2019). It is immediately obvious that galaxies are clustered near the lower right oblate region at low redshift and that many of the massive bins are unconstrained owing to the small sample sizes in CEERS. However, the key trend is that galaxies are clearly in the lower left prolate region at low masses and high redshifts. This is another way to visualize that the model strongly prefers predominantly 3D prolate geometries for low-mass dwarfs at high redshift. We will discuss possible evolutionary connections across mass and redshift later.

However, looking more carefully at the low-mass, high-redshift, prolate-favoring bins, we see that the peaks of the distributions are not in the extreme lower left corner. The typical $C/A \sim 0.25$, but the B/A peaks at a ~2× larger value of ~0.5. We also see this in CANDELS (including in Figure 12 of Zhang et al. 2019). Thus, we cannot rule out another interpretation that is just as tantalizing: that we are indeed seeing flattened disks at these low masses and high redshifts, but they are unusually oval (triaxial, i.e., nonaxisymmetric) with 3D $B/A \sim 0.5$ instead of 3D $B/A \sim 1$ like the nearly round disks we see today. We will discuss this more later, along with the puzzle that some modern large-volume cosmological simulations with sufficient resolution do not reproduce these observational constraints on the 3D C/A–B/A diagram (e.g., Figure 8 of Pillepich et al. 2019).

4.5. Dependence of 3D Size on 3D Geometry

Figure 14 shows that the 3D sizes of galaxies depend on their 3D geometry on average. We use the dividing lines in the previous Figure 13 to separately compute the mean 3D size–mass relation for prolate, oblate, and spheroidal systems based on our SE++ model. Spheroids are systematically smaller than prolate and oblate ellipsoids. The latter are similar to each other, with oblate systems being slightly
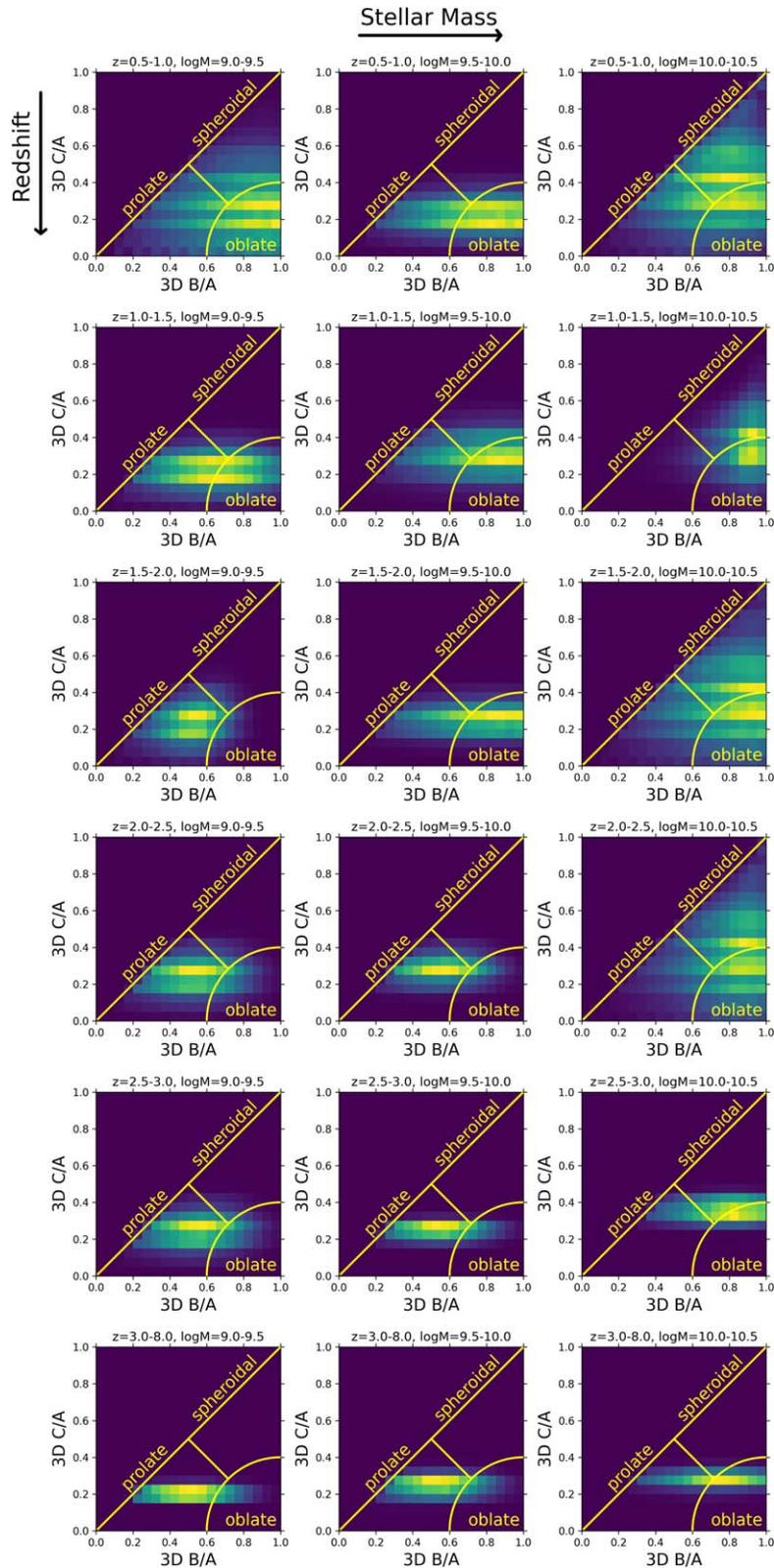
Figure 13. The distribution of 3D axis ratios in each mass–redshift bin from the model fit to the SE++ catalogs. Each panel shows the average of 500 histograms constructed randomly from the posterior for each mass–redshift bin. The yellow boundaries classify the 3D ellipsoids into the prolate, oblate, and spheroidal shapes following van der Wel et al. (2014a) and Zhang et al. (2019). Note how the distribution shifts toward the prolate bin at low masses and high redshifts but peaks at $B/A \sim 2 \times C/A$, which implies unusually oval (triaxial) disks. For the massive bins with small sample sizes, the distributions are less well constrained and thus broader. An analogous version of this figure based on Galfit can be found at the Harvard Dataverse: doi:10.7910/DVN/SWTKVA.

larger. This is consistent with our finding in Figure 10 that larger galaxies tend to have a higher ellipticity. Note that the small sizes of the spheroids are still in the well-resolved regime except for perhaps our highest-redshift, lowest-mass bin. We will discuss the possibility of unresolved small disks in Section 6.3.
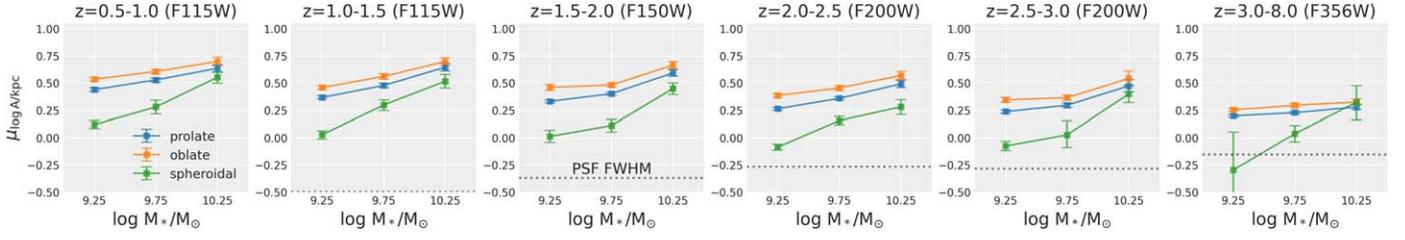
Figure 14. Mean 3D size–mass relations separately for prolate (blue), oblate (orange), and spheroidal (green) ellipsoids. The split by 3D geometry is based on the dividing lines in the previous Figure 13. Spheroids tend to have systematically smaller 3D sizes than prolate and oblate ellipsoids, which themselves are similar, with hints that oblate systems are slightly larger. These constraints are based on our SE++ model.
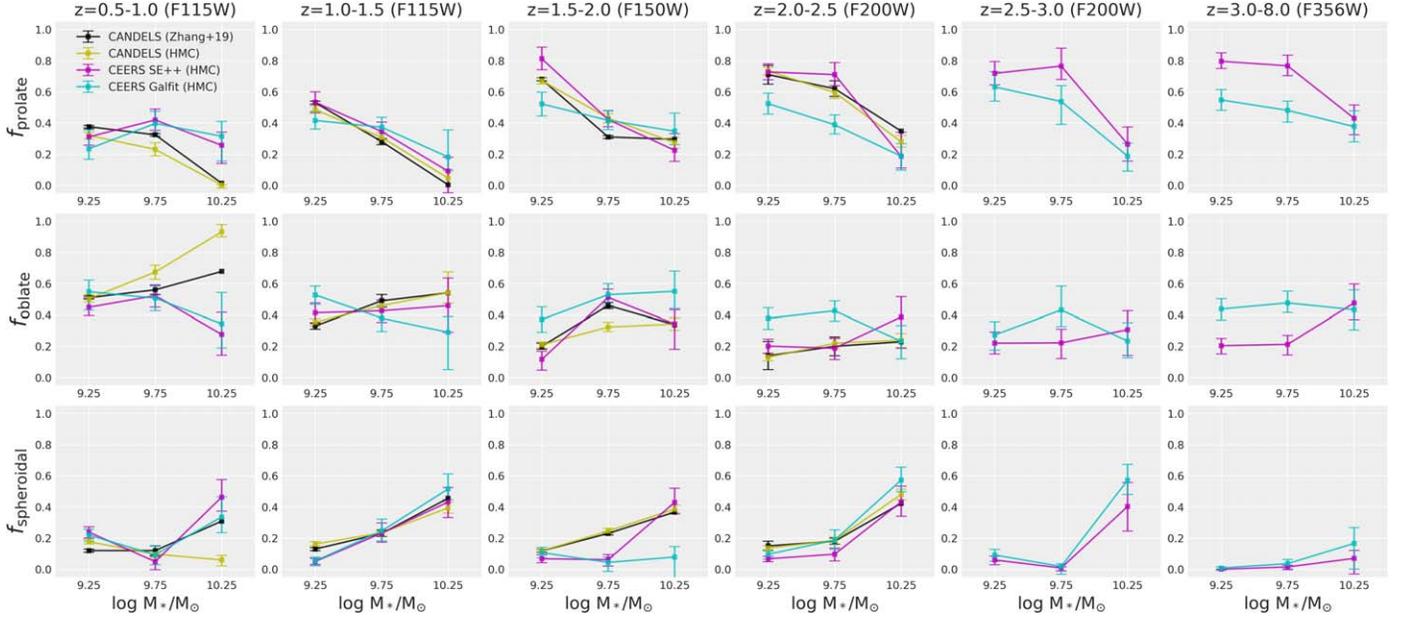
Figure 15. Similar to Figure 9, but now for the evolution of the prolate (top row), oblate (middle row), and spheroidal (bottom row) class fractions. Our new HMC modeling is generally consistent with the Zhang et al. (2019) fractions at $z < 2.5$ from all five CANDELS fields. With CEERS, we find that the prolate fractions of low-mass dwarfs continue to remain $\gtrsim 50\%$ out to $z = 8$.

4.6. Class Fraction Evolution

Figure 15 shows the mass–redshift evolution of the three ellipsoid class fractions. Given the joint constraints on ellipticity (always high) and triaxiality (high for low-mass high-redshift dwarfs, lower for more massive galaxies at later times), we can estimate the relative contributions of prolate, oblate, and spheroidal ellipsoids to the observed population in each mass–redshift bin. We see the striking trend that the prolate fraction goes from $\sim 25\%$ in our $z = 0.5$–1.0 bin up to 80% in the $z = 3$–8 bin. Thus, these galaxies are not an insignificant population, and the majority of low-mass, high-redshift dwarfs may start out as prolate.

The oblate (disky) fraction remains relatively constant at $\sim 20\%$–60% across cosmic time, with the lower fractions based on the SE++ model. There is a hint that dwarf disk fractions increase toward low redshift. The error bars from the CEERS modeling are larger in our massive bins owing to the small sample sizes, but the results are still generally consistent with CANDELS at $z < 2.5$. Since we are focusing only on star-forming galaxies, it is perhaps not surprising that we find very low 3D spheroidal fractions for dwarfs. However, there are puzzling exceptions for the high-redshift, high-mass population, where the 3D (star-forming) spheroidal fraction can rise to $\sim 50\%$. Given the small sample sizes of this high-mass bin at different redshifts, these high massive spheroid fractions may

be influenced by the implicit prior for how we generate our library of toy ellipsoids.

Inspired by Figure 4 of van der Wel et al. (2014a) and Figure 13 of Zhang et al. (2019), our Figure 16 shows a stacked bar chart with the mass–redshift dependence of our average class fractions using the SE++ model. This clearly illustrates the dominance of prolate ellipsoids at low mass and the emergence of disks at lower redshifts. Star-forming spheroids are negligible for these low-mass bins. The $\log M_*/M_\odot = 10$–10.5 class fractions may be influenced by the implicit prior for how we generate our library of toy ellipsoids owing to the small sample sizes.

5. Results from Comparative Analyses

In this section, we explore whether high-probability prolate, oblate, and spheroidal candidates in JWST-CEERS show differences in any other properties besides their 3D shape. These comparative analyses are meant to motivate future work with larger sample sizes and more detailed observational modeling.

5.1. Images of High-probability Candidates

We begin with Figure 17, which shows postage stamps of representative example galaxies with a high probability ($>75\%$) of being prolate, oblate, or spheroidal in 3D using
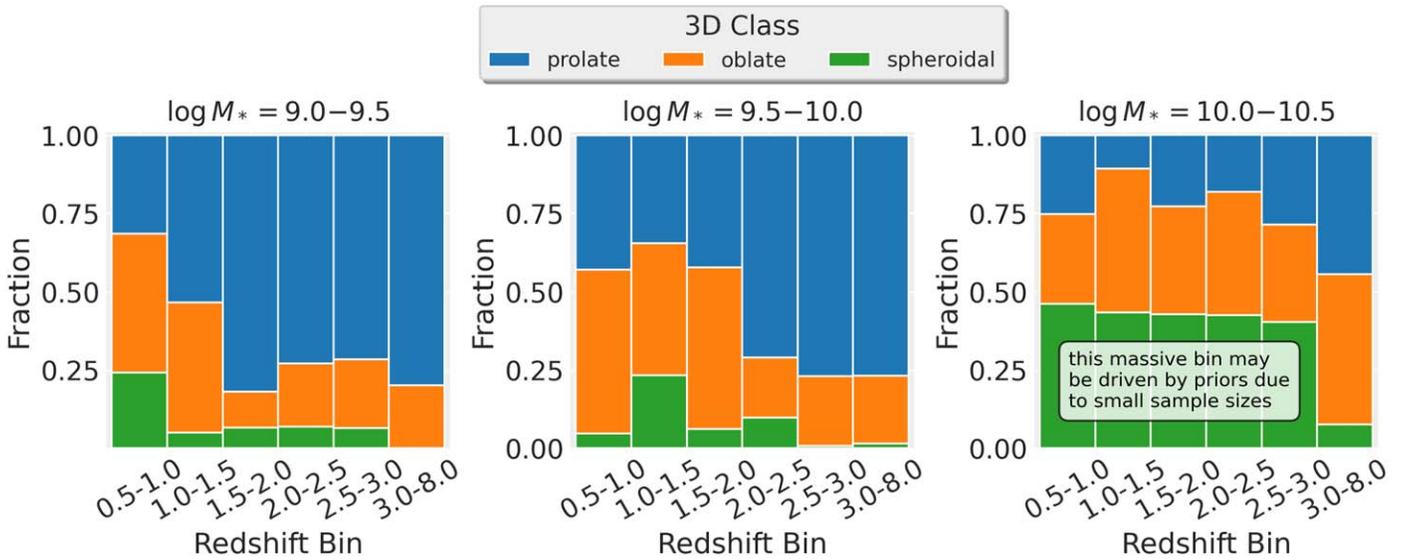
Figure 16. Alternative visualization of Figure 15 as a stacked bar chart using our average model class fractions with SE++. Prolate fractions are shown in blue, oblate in orange, and spheroidal in green. Prolate galaxies dominate at low masses at all $z > 1$. Oblate disks are found at the ~20%–50% level and increase toward low redshift. Spheroids are negligible at these masses. The $\log M_*/M_\odot = 10$–10.5 class fractions may be influenced by the implicit prior for how we generate our library of toy ellipsoids owing to the small sample sizes.

the model based on SE++. All postage stamps are false-color RGB (F356W+F200W+F115W) images and probe rest-frame optical wavelengths. These galaxies fall in mass–redshift bins where the model was able to strongly constrain how different kinds of 3D ellipsoids populate the observed projected b/a versus $\log a$ diagram. Note the striking diversity of the example galaxies both within and between class definitions in terms of their colors and substructures. Some of the example prolate candidates have multiple bright clumps, whereas others are smooth, and a few even show hints of warps or bends, all of which is reminiscent of previous works that subclassify chain galaxies into different morphological types (Elmegreen et al. 2005).

5.2. Sérsic Index and Nonparametric Measures

Motivated by our Bayesian classification scheme and by the diversity of galactic substructure seen in the previous subsection, here we present a statistical comparison of the Sérsic index and nonparametric morphological properties for objects with a high probability ($>75\%$) of falling in one of the three 3D classes of ellipsoids. Given our small sample size, we do not attempt to probe mass/redshift evolution in this paper and defer that to future work.

Figure 18 shows the distributions of the Sérsic index from statmorph for objects with $>75\%$ probability of being assigned to one of the 3D ellipsoid classes. The distributions overlap substantially, which means that the shape of the projected light profile cannot be used by itself to infer the 3D shape of individual galaxies. In particular, note that both prolate and oblate candidates have $n \sim 1$, which implies that not all high-redshift galaxies with exponential light profiles are automatically disks, but may instead be prolate or triaxial. It may seem surprising that what we call spheroidal galaxies have only modestly higher Sérsic indices of $n \sim 2$. This likely reflects our mass and SFR cuts to select only star-forming galaxies with $M_{star} < 10^{10.5} M_\odot$, thus preferentially removing quenched, massive ellipticals, which are expected to have de Vaucouleurs ($n \sim 4$) light profiles.

Figure 19 shows the concentration–asymmetry diagram, Gini–M_{20} diagram, and distributions of clumpiness (smoothness parameter S) as violins. Following Kartaltepe et al. (2023), we show the concentration–asymmetry divisors for nearby galaxies from Bershady et al. (2000) and Conselice (2003). Our high-probability prolate, oblate, and spheroidal candidates do not appear isolated in the concentration–asymmetry plane. Our spheroidal candidates do seem to be near the $z \sim 1$ elliptical region on the G–M_{20} diagram from Lotz et al. (2008), but the prolate and oblate candidates are scattered. Surprisingly, the three sets of candidates show overlapping distributions of clumpiness. This indicates that it may be difficult to distinguish 3D shapes using these traditional nonparametric features. Note that we have only included star-forming galaxies here, and only the highest-probability candidates were pulled from an already small sample size, so we cannot comment on mass and redshift dependence.

5.3. Visual Classifications

Figure 20 shows the fraction of our high-probability prolate, oblate, and spheroidal candidates that have visual classifications as spheroidal, bulge-dominated, disk, and irregular systems based on the deep-learning approach from Huertas-Company et al. (2023). Our 3D spheroid candidates also tend to be visually classified as spheroidal or bulge-dominated systems. Our 3D oblate candidates are predominantly visually classified as irregular, likely due to their clumpy structure. Finally, our 3D prolate candidates tend to be visually classified as irregular or disk objects. As before, this suggests that visual classifications alone may not be able to differentiate between 3D prolate and oblate classifications.

5.4. Specific Star Formation Rates

Figure 21 shows the deviation of the sSFR of high-probability prolate, oblate, and spheroidal candidates with respect to the SFMS in their respective redshift intervals. We do not see any trends such that different types of ellipsoids live
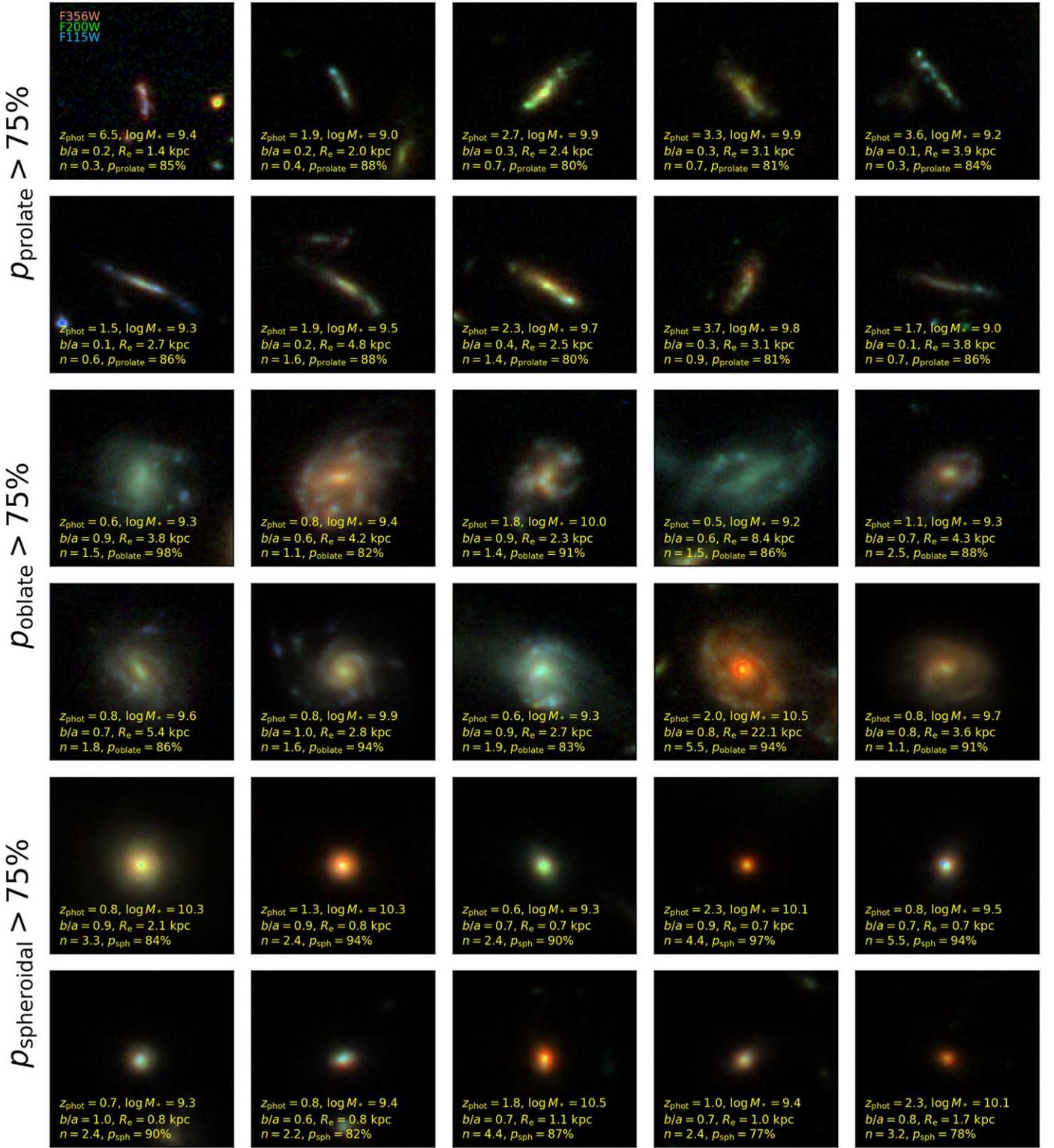
Figure 17. Example $3'' \times 3''$ false-color RGB (F356W+F200W+F115W) postage stamps of galaxies with a high ($>75\%$) probability of being prolate (top two rows), oblate (middle two rows), or spheroidal (bottom two rows). The inset text shows the photometric redshift, stellar mass, F200W b/a, R_e, and n, and the highest class probability. These galaxies fall in mass–redshift bins where our model was able to assign confident classification probabilities to different regions of the b/a vs. $\log a$ diagram.

on different parts of the SFMS. This may have been expected if, e.g., prolate candidates were preferentially undergoing gas-rich mergers that cause starbursts, hence leading to elevated sSFRs. Our sample sizes are currently too small to look for dependence on stellar mass and/or redshift.

5.5. Dust Attenuation

Figure 22 shows the dust attenuation A_V inferred from SED fitting for observed galaxies with $>50\%$ probability of being prolate, oblate, or spheroidal in 3D. We use a less stringent cut
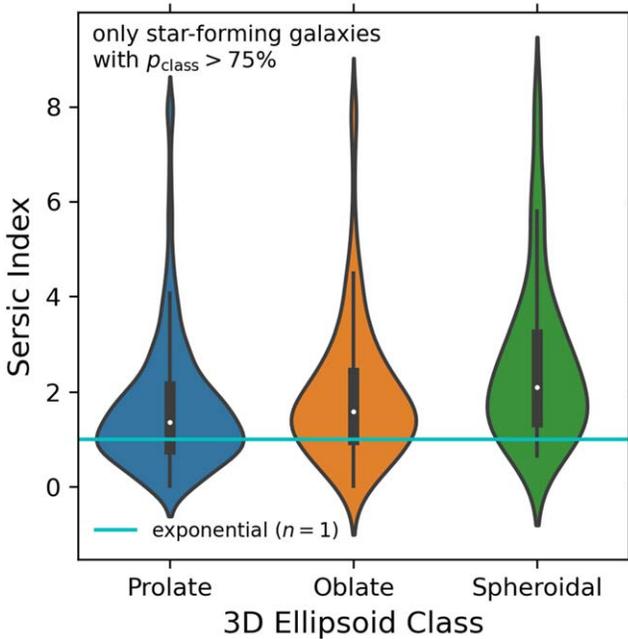
Figure 18. Violin plot showing the distribution of Sérsic index from statmorph for star-forming galaxies with >75% probability of being assigned to one of the 3D ellipsoid classes. The distributions overlap considerably, which means that the shape of the projected light profile alone cannot be used to assign galaxies to a given 3D ellipsoid class. Many oblate and prolate candidates both have exponential light profiles with $n \approx 1$ (cyan line). The Sérsic indices of spheroids are only marginally higher, but recall that our star-forming cut removes the typical quenched $n \sim 4$ objects.

of 50% instead of 75% since our sample sizes are too small to assign high probabilities to edge-on oblate systems, which we otherwise expect to dominate at low redshift and high mass based on CANDELS. We see that our prolate candidates tend to have very low dust attenuation even though they primarily have low projected b/a. On the other hand, we see a hint that edge-on oblate candidates with low b/a tend to have higher dust attenuation compared to more face-on oblate candidates.

These findings are consistent with Figure 2 of Zhang et al. (2019), although we cannot yet explore mass and redshift dependence owing to our small sample sizes (but this is motivation for future work). In the presence of volume-filling diffuse dust, the higher A_V we observe for edge-on oblate candidates is expected since edge-on disks would have a larger path length for dust attenuation. In contrast, prolate systems with small 3D $B/A = C/A$ would have small path lengths for dust attenuation even when seen with small projected b/a. Our spheroidal candidates are also surprisingly dusty, which makes us wonder whether some of them are prolate objects seen down the barrel (thus appearing round) with a large path length for dust attenuation. In detail, these arguments depend on the clumpiness of the dust as shown by Zhang et al. (2023), as well as the degeneracies between dust attenuation, stellar population age and metallicity, and photometric redshift.

6. Discussion

6.1. Connection to Previous Work

Prior to JWST, there was already evidence from HST that massive, high-redshift galaxies were largely consistent with the intrinsically oblate and spheroidal 3D shapes of massive systems we see today (Holden et al. 2012; Chang et al. 2013;

van der Wel et al. 2014a; Zhang et al. 2019, 2022). There was also evidence that fainter galaxies, despite their overwhelmingly peculiar/irregular appearance, may still be sorted into distinct and comprehensible classes such as chains, clump clusters, and tadpoles (Cowie et al. 1995; van den Bergh et al. 1996; Elmegreen et al. 2005). However, the interpretation of these faint objects has been puzzling. One side of the argument is that these apparently exotic faint star-forming galaxies are consistent with underlying oblate geometries viewed edge-on and that we do not observe as many round, face-on objects because of surface brightness detection biases (Dalcanton & Shectman 1996). On the other hand, statistical 3D shape modeling of the deepest, largest surveys from HST, such as CANDELS, suggests a real paucity of round dwarfs at high redshift and proposes predominantly prolate 3D shapes as the solution (van der Wel et al. 2014a; Zhang et al. 2019). Our own completeness simulations in Appendix B demonstrate that we are complete to large face-on disks down to ~26.5 AB mag (F277W), which is ~2 mag deeper than HST-CANDELS, for which studies of galaxy morphology are typically restricted to <24.5 AB mag (F160W; van der Wel et al. 2012).

Many other studies have also used JWST to analyze the evolution of galaxy structure and morphology. As emphasized by Huertas-Company et al. (2023), JWST-CEERS goes significantly deeper than HST-CANDELS, but only a relatively small fraction of objects that were classified as irregular in HST-CANDELS now have diffuse, extended emission from faint disks newly detected by JWST. Ferreira et al. (2022) and Kartaltepe et al. (2023) both used visual classifications along with parametric and nonparametric modeling to show that there are more disks in JWST imaging than seen by HST. Robertson et al. (2023) also showed that deep-learning methods trained on HST-CANDELS visual classifications recover fainter disks in new JWST imaging. They also used the distribution of projected axis ratios to place an upper limit of 57% on the pure disk fraction, though they did not comment on mass and redshift dependence. It is important to realize that our results are not necessarily inconsistent with these previous studies since visual classifications and exponential light profiles (Sérsic index $n \sim 1$) alone cannot distinguish between prolate and oblate 3D geometries. Indeed, Vega-Ferrero et al. (2024) find that many visually classified disks in JWST seem to have peculiar features, and more detailed follow-up is required. The machine-learning study by Tohill et al. (2023) also identified several distinct morphological classes for high-redshift galaxies observed with JWST, one of which is a set of consistently elongated systems. While our own analysis does identify disks at the ~20%–50% level, in the dwarf regime we find that prolate or triaxial ellipsoids outnumber oblate disks by a factor of several.

The 3D shapes of galaxies are intimately related to their sizes, and the latter are of great interest for galaxy formation theory (Ferguson et al. 2004; Somerville et al. 2018). Studies with JWST are now beginning to measure projected size–mass or size–luminosity relations at high redshift (Yang et al. 2022; Ito et al. 2023; Ward et al. 2023). Our modeling approach simultaneously constrains the 3D size–mass relation and its evolution with redshift to $z = 8$. This appears broadly consistent with previous work with HST, but it will be interesting to compare to future compilations. We find that the scatter in this relation is remarkably constant and small with $\sigma_{\log A} \sim 0.2$ dex, which is in agreement with van der Wel et al. (2014b). We also
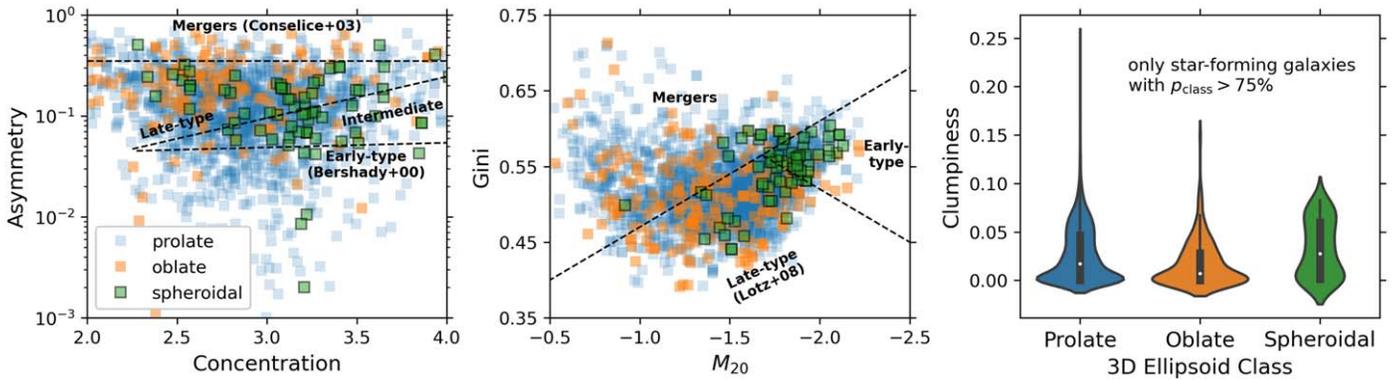
Figure 19. Nonparametric morphological properties for high-probability prolate (blue), oblate (orange), and spheroidal (green) candidates. Left: concentration–asymmetry diagram with the $A = 0.35$ divisor from Conselice (2003), above which nearby galaxies are mergers, and with the two sets of divisors from Bershady et al. (2000) used to separate nearby early- and late-type galaxies. The prolate candidates do not occupy a special place in this diagram. Middle: Gini–M_{20} diagram with divisors from Lotz et al. (2008) for $z \sim 1$ galaxies. While our spheroidal candidates are near the early-type region, the oblate and prolate candidates are not isolated. Right: violin plot comparing the distribution of clumpiness (smoothness parameter S). Surprisingly, all three sets of candidates have similar and overlapping distributions. All of these panels indicate that prolate galaxies seem to be difficult to characterize using these traditional nonparametric features.
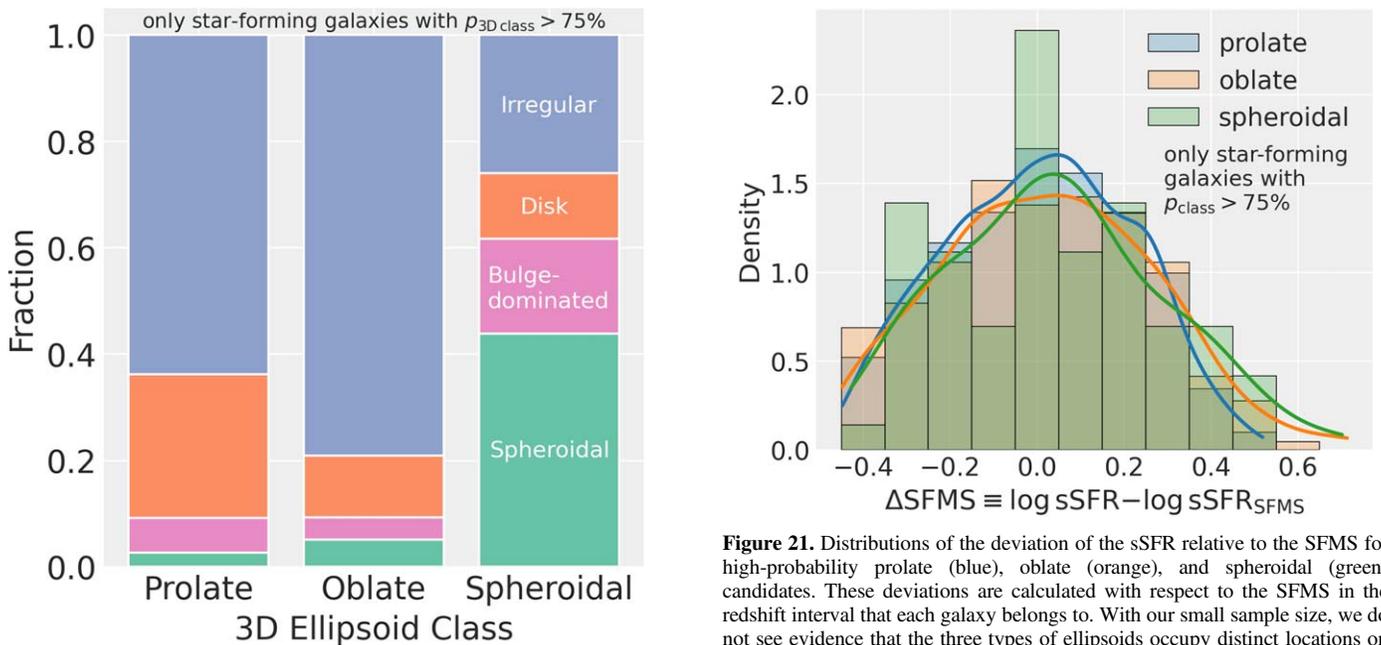
Figure 20. Visual classifications from the deep-learning approach of Huertas-Company et al. (2023) for our high-probability prolate, oblate, and spheroidal star-forming galaxies. The colors indicate different visual classifications: spheroidal (green), bulge-dominated (pink), disk (orange), and irregular (purple). Our 3D spheroid candidates also tend to be visually classified as spheroids and bulge-dominated systems. Many of our 3D oblate candidates are classified as irregular, probably due to their clumpy structure. Our 3D prolate candidates tend also to be visually classified as irregular or disk systems.

Figure 21. Distributions of the deviation of the sSFR relative to the SFMS for high-probability prolate (blue), oblate (orange), and spheroidal (green) candidates. These deviations are calculated with respect to the SFMS in the redshift interval that each galaxy belongs to. With our small sample size, we do not see evidence that the three types of ellipsoids occupy distinct locations on the SFMS.

saw in Figure 14 the dependence of 3D size on 3D geometry, namely that high-redshift star-forming spheroids tend to be much smaller than the oblate and prolate populations, particularly at dwarf scales. Combined with the class fraction evolution in Figure 15, this invokes a basic picture of the high-redshift star-forming dwarf population as comprising relatively rare, small-size spheroids and mostly large prolate systems, while oblate geometries emerge later. Finally, we see systematically smaller sizes in JWST-CEERS relative to HST-CANDELS when using GALFIT as was also shown by Suess et al. (2022), but this discrepancy goes away when we use SE++. We discuss this more in Appendix C.

Finally, our analysis directly builds on the HST-CANDELS work by Zhang et al. (2019), which itself generalized the study by van der Wel et al. (2014a). We showed that our new code reproduces the results of Zhang et al. (2019) using all five CANDELS fields in a fraction of the computing time. Our JWST results are also roughly consistent at $z < 2.5$, where CEERS and CANDELS overlap. One subtlety that is worth commenting on is that both van der Wel et al. (2014a) and Zhang et al. (2019), and by extension we, assume rather arbitrary boundaries for dividing 3D ellipsoids into one of the three extreme classes (oblate, spheroidal, or prolate). In reality, our models are based on general triaxial ellipsoids such that if $E \gg 0$ and $T \neq 0$ or $T \neq 1$, then the three axes have different lengths, so we cannot say that we are clearly in one of the three extreme shape scenarios. Our Figure 13 and Figure 12 of Zhang et al. (2019) show that the 3D C/A–B/A model distributions in what we call the prolate-dominated mass–redshift bins seem to be in the more ambiguous triaxial
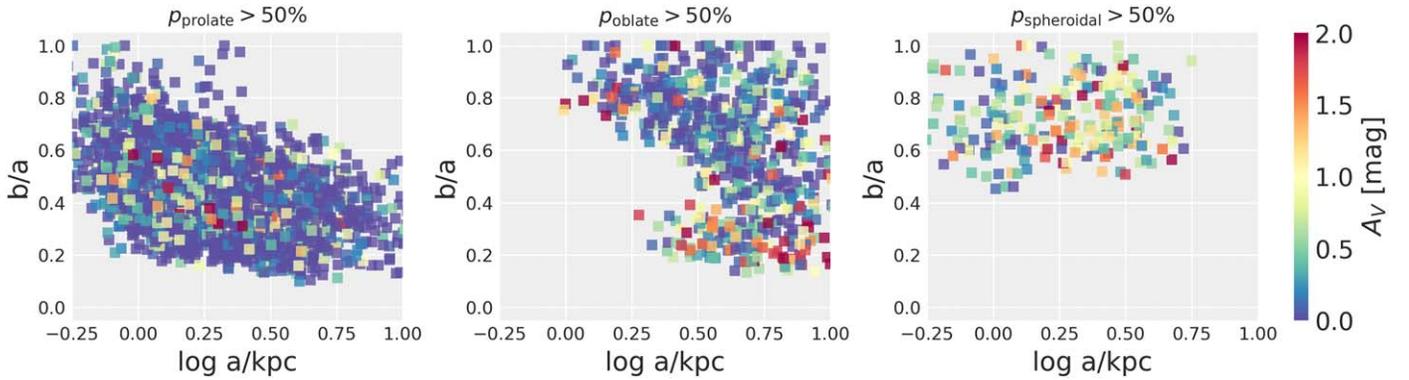
Figure 22. Dust attenuation A_V across the $b/a - \log a$ diagram for observed galaxies with >50% of being prolate (left), oblate (middle), or spheroidal (right) according to our 3D shape modeling. The prolate candidates are generally dust-free. In contrast, oblate candidates with lower b/a tend to have higher dust attenuation, which is expected for edge-on disks and consistent with Figure 2 of Zhang et al. (2019). Our 3D spheroidal star-forming candidates are also surprisingly dusty.

category. This is because $C/A \sim 0.25$ and $B/A \sim 2 \times C/A \approx 0.5$, which evokes an unusually oval, flattened disk compared to the rounder disks seen in the local Universe. This triaxial rather than extreme prolate conclusion was also considered by Elmegreen et al. (2005), Ravindranath et al. (2006), Yuma et al. (2011, 2012), and Law et al. (2012). It calls to mind the notion of disk settling in terms of thickness (Kassin et al. 2012), but here we argue that there may also be another kind of disk settling from oval to circular shapes.

6.2. Astrophysical and Cosmological Implications

We will now discuss the implications of our results for each class of 3D ellipsoids, starting with star-forming spheroids. Previous work with HST has shown that star-forming spheroids are rare (Brennan et al. 2015), and indeed our Figure 14 shows their fractions at $\lesssim 20\%$ for $\log M_*/M_\odot < 10$. However, they rise to the $\sim 40\%$ level for $\log M_*/M_\odot = 10.0$–10.5 at $z = 0.5$–3. Given the small sample sizes of this massive bin at different redshifts, these high massive spheroid fractions may be influenced by the implicit prior in how we generate our library of toy ellipsoids. If real, this abrupt rise in the spheroidal fraction at these masses and redshifts may be related to the formation of compact star-forming galaxies at high redshift and their eventual transformation into compact quiescent galaxies, which are thought to be the progenitors of local giant ellipticals (van Dokkum et al. 2008; Barro et al. 2013) or bulges in massive spirals (de la Rosa et al. 2016; Costantin et al. 2021, 2022). These compact star-forming spheroids may arise as a result of what is called a "compaction" event, where gas-rich mergers, disk instabilities, or cold gas inflows cause a galaxy to rapidly increase in mass surface density (Dekel & Burkert 2014; Zolotov et al. 2015; Tacchella et al. 2016; Lapiner et al. 2023).

The origin of galactic disks remains a complicated problem, and it is therefore important to identify high-redshift disks and understand their properties in the context of local spirals (see the review by van der Kruit & Freeman 2011). Many authors have shown that selecting objects on the basis of exponential light profiles alone may not be sufficient for identifying genuine rotating disks at high redshift (Law et al. 2007, 2012; van der Wel et al. 2014a), and our own Figure 16 reveals that both oblate and prolate candidates can have remarkably similar Sérsic indices of $n \sim 1$. Thus, 3D shape modeling like ours is crucial to identify oblate disk candidates in addition to visual classifications and parametric and nonparametric

morphological measurements. We find oblate fractions of $\sim 20\%$–60% over our full mass and redshift ranges of $\log M_*/M_\odot = 9.0$–10.5 and $z = 0.5$–8.0 from JWST-CEERS, suggesting that disks were indeed already in place at very early times. Our Figure 13 shows that, on average, for mass–redshift bins where the population is predominantly in the lower right oblate corner, we find a typical intrinsic $C/A \sim 0.2$–0.3, which, according to Elmegreen et al. (2005), is ~ 2–$3\times$ thicker than local spirals. The mean 3D sizes and remarkably small ~ 0.2 dex scatter of our highest-redshift disk candidates may provide new constraints for modeling the relative roles of angular momentum conservation, gas accretion and outflows, mergers, and halo properties like concentration, virial radius, and assembly history in governing disk formation and evolution.

Perhaps the most startling aspect of our results is the high prolate dwarf fractions, rising to $\sim 50\%$–80% at $z = 3$–8 for $\log M_*/M_\odot = 9.0$–9.5 galaxies. Ceverino et al. (2015) and Tomassetti et al. (2016) have argued, based on cosmological zoom-in simulations, that prolate dwarfs at high redshift can be explained if they form within host halos that are themselves elongated along their host dark matter filament. These objects would continually undergo mergers along the direction of the filament, hence causing their elongation and possibly intrinsic alignments on large scales (as originally proposed by Pandya et al. 2019). At lower redshifts, as filaments become diffuse and the continuous merger process slows down, the observed prolate fraction may decrease in accordance with our Figure 14. In other words, our high prolate fractions at high redshift may be telling us something about the hierarchical merger-driven process of galaxy formation. Now, in this case, we may expect merger-driven starbursts, but Figure 21 shows that high-probability prolate candidates do not occupy a special place on the SFMS, at least with our small sample size. Ceverino et al. (2015) and Tomassetti et al. (2016) have also suggested that the centers of halos hosting prolate galaxies are dark matter dominated, and as they undergo accretion and "compaction," they become baryon dominated and capable of supporting disks (see also Lapiner et al. 2023). This prolate-to-oblate transition is thought to involve the formation of compact "blue nuggets" in a characteristic mass range of $\log M_*/M_\odot \sim 9.2$–10.3 as observed in HST-CANDELS (Huertas-Company et al. 2018). We do not yet have large enough sample sizes with JWST to explore such an evolutionary connection across mass and redshift.

In the prolate phase, the stellar motions should be velocity dispersion supported and rotating gaseous disks may not be a stable configuration. Existing spectroscopic constraints do already show a decline in the fraction of rotationally supported galaxies with increasing redshift and decreasing mass (e.g., Figure 4 of Simons et al. 2017), but prohibitively deep stellar spectroscopy will be needed to definitively test this picture. We have also discussed the possibility that what we are calling prolate dwarfs are actually unusually oval (triaxial) ellipsoids. If true, then just as galaxy disks "settle" from thick to thin over cosmic time (Kassin et al. 2012), they must also settle from oval to circular shapes toward low redshift. This oval-to-circular transition may be driven by a changing mode of gas accretion or mergers wherein at earlier times it is more clumpy, leading to episodic star formation, and at later times it is smoother. Alternatively, we may interpret these as stellar bars (Gullberg et al. 2019), though they may form differently from normal bars since there is no obvious, bright, extended stellar disk and since the dark matter likely dominates over the self-gravity of the stars in these dwarfs. Nevertheless, we point out that progenitors of Milky Way–mass galaxies at $z \gtrsim 2$ are thought to have $\log M_*/M_\odot \sim 9$–10 (e.g., Papovich et al. 2015), which is where we find consistently high fractions of prolate and/or triaxial ellipsoids. This implies that our own Galaxy may have gone through a prolate or triaxial morphological phase in its past.

Lastly, it is worth commenting on hydrodynamical simulations. Early on during the debates about the nature of chain galaxies, Immeli et al. (2004a, 2004b) and Bournaud et al. (2007) used simulations to argue that chain galaxies are edge-on manifestations of intrinsically oblate clumpy star-forming galaxies. As far as we are aware, only Ceverino et al. (2015), Tomassetti et al. (2016), and Lapiner et al. (2023) claim to have found unambiguously prolate or triaxial high-redshift dwarfs in their zoom-in simulations. Pillepich et al. (2019) show in their Figures 8 and 9 that there are far fewer prolate galaxies in the TNG50 simulations compared to the observational estimates from van der Wel et al. (2014a) and Zhang et al. (2019), and now our paper as well. Their high-redshift dwarfs seem to be predominantly spheroidal or oblate with relatively high C/A and/or B/A compared to our Figure 13 (see also Zhang et al. 2022). This discrepancy between at least two sets of simulations with respect to each other and versus observational constraints demands that 3D shapes be analyzed in more detail in modern simulations. We note in passing that the uncertain nature of dark matter has motivated simulations of alternatives to cold dark matter, such as fuzzy dark matter, in which filaments and galaxies may naturally be more elongated (Mocz et al. 2020; Dome et al. 2023).

6.3. Limitations of Our Study

Our analysis, like any, is subject to uncertainties. First and foremost, our sample sizes from JWST-CEERS are rather small. We argue that our key results regarding low-mass galaxies being predominantly prolate or triaxial are robust given the sufficiently large sample sizes from CEERS alone. However, our massive $\log M_*/M_\odot = 10$–10.5 bins have small sample sizes, and so their posteriors on 3D shape parameters may be influenced by our priors (in particular, our choice to uniformly sample toy ellipsoids in the $E-T-\log A$ space rather than $C/A-B/A-\log A$ space). We also see that these massive bins tend to have greater residuals between the observed and mean model histograms, which may arise from both the small sample sizes and perhaps the limited flexibility of the model. This can be remedied in the future by combining data sets from different JWST surveys and trying more sophisticated models. We plan to pursue this in the future since our differentiable Bayesian approach is uniquely fast and robust. By combining data sets from different areas of the sky, we can also address the issue of cosmic variance. Relatedly, our analysis may suffer from detection or measurement biases, though in Appendix B we showed that CEERS is complete to disks with a range of sizes and axis ratios as faint as $\sim$26.5 AB mag (F277W), which is $\sim$2 mag deeper than HST-CANDELS (F160W). We cannot rule out fainter disks, but it is unclear whether they would satisfy our $\log M_*/M_\odot > 9$ sample selection limit. Future parameter recovery tests for mock Sérsic profiles and fake 3D ellipsoids inserted into the real imaging will help address these questions (see also E. McGrath et al. 2024, in preparation).

We found systematic size differences between SE++ and GALFIT, with the latter producing systematically smaller size measurements for the same galaxies seen in HST-CANDELS (see also Suess et al. 2022). Figure 12 shows that this discrepancy is worse for lower-mass, higher-redshift objects, which would also be the faintest and thus preferentially susceptible to issues with deblending and masking. Both codes use the same global background-subtracted images, fix the local background to zero, and have the same empirical PSFs, so these cannot be the causes of the size discrepancies. We discuss this in more detail in Appendix C. However, the key point is that, regardless of the discrepancies between the SE++ and Galfit measurements, our two sets of 3D shape modeling results using the different data sets still lead to similar conclusions. In other words, despite any differences between the two codes, our model still converges to roughly the same region of its enormous 7D parameter space when using either data set.

We showed in Figures 3 and 4 that, for many of the low-mass high-redshift bins, there is a deficit of high-b/a objects and an excess of low-b/a objects, especially at large $\log a$. This is the crux of our argument in favor of high-redshift dwarfs being prolate or triaxial. However, the impact of the PSF means that we cannot rule out the possible existence of a population of small-size disks with radii close to the PSF FWHM limit since it would be difficult to resolve their low or even intermediate b/a projections. This could be another explanation for why smaller galaxies tend to have higher b/a: they may be intrinsically spheroidal as Zhang et al. (2019) and we suggest, or we may only be observing the face-on projections of small disks. Now, it is possible to recover the projected shapes of small galaxies if the PSF is well understood, and indeed many of our $z = 3$–8 low-mass, small-size galaxies are constrained to have b/a well below the PSF limit. This caveat also does not explain away our main finding that larger (well-resolved) dwarfs preferentially show up with small b/a, indicating prolate or triaxial geometries. Ultimately, we argue that telescopes with even better resolution than JWST will be needed to definitively constrain the existence of small-size disks. In the meantime, larger data sets will allow more sophisticated modeling approaches, such as fitting for multiple populations in a single mass–redshift bin (e.g., using Gaussian mixture models), which may also give us a better handle on PSF-related limitations. Somewhat related is that our analysis assumes that SE++ and GALFIT are measuring the true (intrinsic) projected axis ratios of galaxies after accounting for

instrumental effects, but we have not considered the impact of weak lensing on distorting the shapes of distant dwarfs, namely making them appear more systematically elongated.

The redshifts and stellar masses we use to group galaxies into different bins and the SFRs we use to select only star-forming galaxies are all based on SED fitting. It is well known that these can all be uncertain depending on the assumptions and methods used for SED fitting (Pacifici et al. 2023). Along these lines, we found hints of a trend in Figure 22 such that high-probability oblate candidates seen edge-on have a higher dust attenuation A_V, whereas prolate candidates appear to be relatively dust-free. This requires follow-up since the detailed geometry of dust (i.e., whether it is clumpy or diffuse) will also affect attenuation in addition to the viewing angle dependence (e.g., as was recently shown by the joint analysis of 3D shapes and dust attenuation for HST galaxies by Zhang et al. 2023; see also Padilla & Strauss 2008; Zhang et al. 2019).

Finally, we have not explored the possibility that the changing mass-to-light ratios (and radial gradients thereof) of disks can masquerade as geometric evolution. In other words, without any 3D shape modeling, can the asymmetric distributions of projected b/a and the banana-shaped b/a–$\log a$ 2D histograms be reproduced via disk color-related selection effects? Are we simply seeing different parts of faint underlying disks at high redshift such as bars and star-forming knots? In a related sense, could strong emission lines from the gas in these early systems affect their appearance even in broadband imaging (e.g., Amorín et al. 2015)? It is beyond the scope of this paper to address these questions, but they would be fruitful avenues for future work.

7. Conclusions

We have developed a differentiable Bayesian model and used HMC to constrain the 3D shapes of high-redshift star-forming galaxies from JWST-CEERS observations. To ensure that our results are not driven by source detection and shape characterization methods, we have used two different catalogs: internal CEERS team catalogs based on GALFIT (E. McGrath et al. 2024, in preparation), and independent catalogs from the next-generation SE++ (Bertin et al. 2020; Kümmel et al. 2022). We run our efficient and robust model on CANDELS data to reproduce previous results from Zhang et al. (2019) in a fraction of the computing time, and we also use mock tests to show that the model and data have constraining power for sample sizes as small as ∼50 galaxies. For the new JWST-CEERS data, our model is able to constrain the mean ellipticity, mean triaxiality, mean size, and covariances of these quantities as a function of stellar mass and redshift over the ranges $\log M_*/M_\odot = 9.0$–10.5 and $z = 0.5$–8.0.

Our findings can be summarized as follows:

1. With the better spatial resolution and sensitivity of JWST NIRCam imaging, we still find peculiarly asymmetric distributions of projected axis ratios peaking at low values of $b/a \sim 0.3$–0.4 for galaxies with $\log M_*/M_\odot = 9.0$–10.0 at $z > 1$. This confirms previous findings from HST but now alleviates concerns about blending, leading to an overabundance of elongated objects, as well as incompleteness to faint face-on disks (Figures 1–4).

2. We assume that galaxies can be described as 3D ellipsoids (of which the extreme types are oblate, prolate,

and spheroidal) and demonstrate how random projections of these types of systems would trace out different paths on the projected b/a–$\log a$ plane. Spheroids would preferentially show up with large b/a, axisymmetric disks would show a uniform vertical stripe, and prolate and oval (triaxial, i.e., nonaxisymmetric) disks would both trace out a curved "banana" trajectory on this diagram (Figure 5).

3. Using 2D histograms of projected b/a–$\log a$ as observational constraints, our Bayesian model combined with HMC (Figure 6) finds high mean ellipticities with small scatter for all mass–redshift bins we consider. This means that the galaxies we observe are either disks or prolate. In many mass–redshift bins, our model is able to break that degeneracy by constraining the triaxiality. For $\log M_*/M_\odot = 9.0$–9.5 at $z > 1$, the mean triaxiality tends to be very high, strongly favoring the prolate interpretation. The mean triaxiality tends to be lower for higher-mass galaxies at lower redshifts, suggesting the emergence of disks (Figures 9, 11).

4. Our model also automatically constrains the 3D size–mass relation and its mass–redshift evolution. At a fixed redshift, higher-mass galaxies have larger sizes. As one goes to higher redshift, the size–mass relation drops in normalization: more distant galaxies are naturally smaller in 3D at fixed mass. The scatter in the size–mass relation is remarkably small and constant with the mass–redshift relation at $\sigma_{\log A} \sim 0.2$ dex. The 3D size–mass relation depends on 3D geometry in the sense that high-redshift star-forming dwarfs tend to be systematically smaller than prolate and oblate ellipsoids (Figures 12, 14).

5. The fraction of prolate galaxies rises from ∼25% at $z = 0.5$–1.0 up to ∼50%–80% at $z = 3$–8 for $\log M_*/M_\odot = 9.0$–9.5 dwarfs. The prolate fraction decreases toward higher masses at all redshifts. The dwarf disk fraction tends to rise from ∼20%–40% to ∼40%–60% toward low redshift. We find surprisingly high (∼40%) spheroid fractions for massive galaxies with $\log M_*/M_\odot = 10$–10.5, but this may be influenced by small sample sizes and the implicit prior for generating our library of toy model ellipsoids (Figures 15, 16).

6. If low-mass, high-redshift dwarfs are indeed disks, they cannot be axisymmetric but instead must be unusually oval (triaxial) compared to local circular disks, implying that their stars cannot move on circular orbits. This is supported by our model, which suggests that they have 3D axis ratios of $C/A \sim 0.25$ but $B/A \sim 2 \times C/A \approx 0.5$ as opposed to $B/A \sim 1$ observed for the nearly round disks that we see today. This interpretation suggests that disks may "settle" not only from thick to thin but also from oval to circular shapes with cosmic time (Figure 13).

7. We can assign high-probability (>75%) classifications of 3D ellipsoid class to ∼2000 galaxies irrespective of mass–redshift. For these, color postage stamps reveal remarkably linear, large, and thin systems classified as prolate, as well as obvious cases of face-on disks and compact spheroids. Some of the prolate candidates are reminiscent of the long-forgotten class of "chain galaxies" identified in deep Hubble images (Cowie et al. 1995; van den Bergh et al. 1996; Elmegreen et al. 2005; Figure 17).

8. The high-probability prolate and oblate candidates have similar Sérsic indices of $n \sim 1$, meaning that this alone cannot be used to infer their 3D geometry. Both tend to be visually classified as disks or irregular. Surprisingly, the three classes of high-probability candidates do not reveal significant differences in their nonparametric morphological properties like concentration, asymmetry, clumpiness (smoothness), Gini coefficient, and the second-order moment of the 20% brightest regions. The three classes also do not differ in their distribution of deviations from the SFMS. However, we find hints that edge-on oblate candidates have higher dust attenuation A_V, whereas prolate candidates are generally blue and dust-free (Figures 18–22).

Appendix A
Mock HMC Parameter Recovery Tests

In order to understand the robustness of our model and the HMC sampler, we perform mock parameter recovery tests. In general, we start by picking values of our seven model parameters, generating the corresponding true probability map of the projected b/a versus $\log a$ 2D histogram, and then random Poisson sampling N "observed" objects from that 2D probability map. Since we are mainly interested in our ability to distinguish prolate-, oblate-, and spheroid-dominated populations, we focus on varying the mean ellipticity and mean triaxiality parameters while fixing

$\mu_{\log A} = 0.3$, $\sigma_E = 0.1$, $\sigma_T = 0.3$, $\rho(E, \log A) = 0.8$ motivated by the observational constraints.

Figure 23 shows an example corner plot from one of many mock HMC parameter recovery tests that we did. In this case, we chose parameters reflective of a prolate-dominated mock population: $\mu_E = 0.9$ and $\mu_T = 0.9$, with the other parameters given above. We tried four different sample sizes similar to our CEERS observations: $N = 500, 100, 50, 25$. We find that for all of these sample sizes we are always able to constrain μ_E, $\mu_{\log A}$, σ_E, and $\sigma_{\log A}$ very well. The correlation coefficient $\rho(E, \log A)$ also tends to be recovered even for $N = 25$ for this and other mock parameter combinations that we tried. However, the key parameters that distinguish oblate from prolate populations, μ_T and σ_T, generally require sample sizes of $N > 50$. For smaller sample sizes such as $N = 25$, the posteriors are unconstrained and reflect the broad uniform prior.

The same conclusions for our prolate-dominated mock also apply to oblate-dominated ($\mu_E = 0.9$, $\mu_T = 0.1$) and spheroid-dominated ($\mu_E = 0.1$, $\mu_T = 0.1$) mocks as can be seen in the downloadable figure set corresponding to Figure 23. However, in the latter case, the triaxiality of nearly round spheroids is meaningless since it does not matter whether $b \sim a$ or $b \sim c$ since $c \sim a$. Thus, the triaxiality posteriors for spheroid-dominated mocks are always broad, even for $N = 500$ or even larger sample sizes that we tried, but this is to be expected.

Perhaps the most interesting case is a combination of intermediate μ_E and μ_T that is not dominated by any one of the three ellipsoid classes. Thus, it is a more ambiguous ellipsoid mixture population and a stronger mock test for our algorithm. Here again we find that all parameters except μ_T and σ_T can be recovered for any sample size down to $N = 25$. The correlation coefficient $\rho(E, \log A)$ in the $N = 50$ mock has a broader posterior but still peaks at the true value. As for μ_T and σ_T, it is perhaps not surprising that larger sample sizes are needed to constrain this since intermediate triaxiality values combined with intermediate ellipticity values lead to subtler variations in the 2D projected b/a versus $\log a$ diagram. However, our key point is that we are apparently not in this ambiguous population regime for many of our observed CEERS mass–redshift bins with $N \sim 500$, so this is not a concern.

Appendix B
Completeness of CEERS to Faint Face-on Disks

Figure 24 shows the distributions of apparent magnitude in the relevant filters for our sample across the mass–redshift grid. Recall that our sample is restricted to $\log M_*/M_\odot > 9$, and so our objects are generally brighter than ~ 27 AB mag even for the highest-redshift, lowest-mass bin. We verified that if we plot the apparent magnitude as a function of stellar mass for all galaxies in the CEERS source catalog, there are very few sources fainter than ~ 27 AB mag that also have $\log M_*/M_\odot > 9$.

In order to make sure that our results are not driven by incompleteness to faint, face-on disks (i.e., that the deficit of galaxies in the upper right corner of the b/a–$\log a$ banana diagram is real), we run mock completeness simulations. Following Section 7.1 of Finkelstein et al. (2023), we inject 10^4 fake Sérsic profiles into the CEERS imaging. We assume $n = 1$ for all sources since we are mainly interested in completeness to disks and since our actual sample clusters around $n = 1$ as
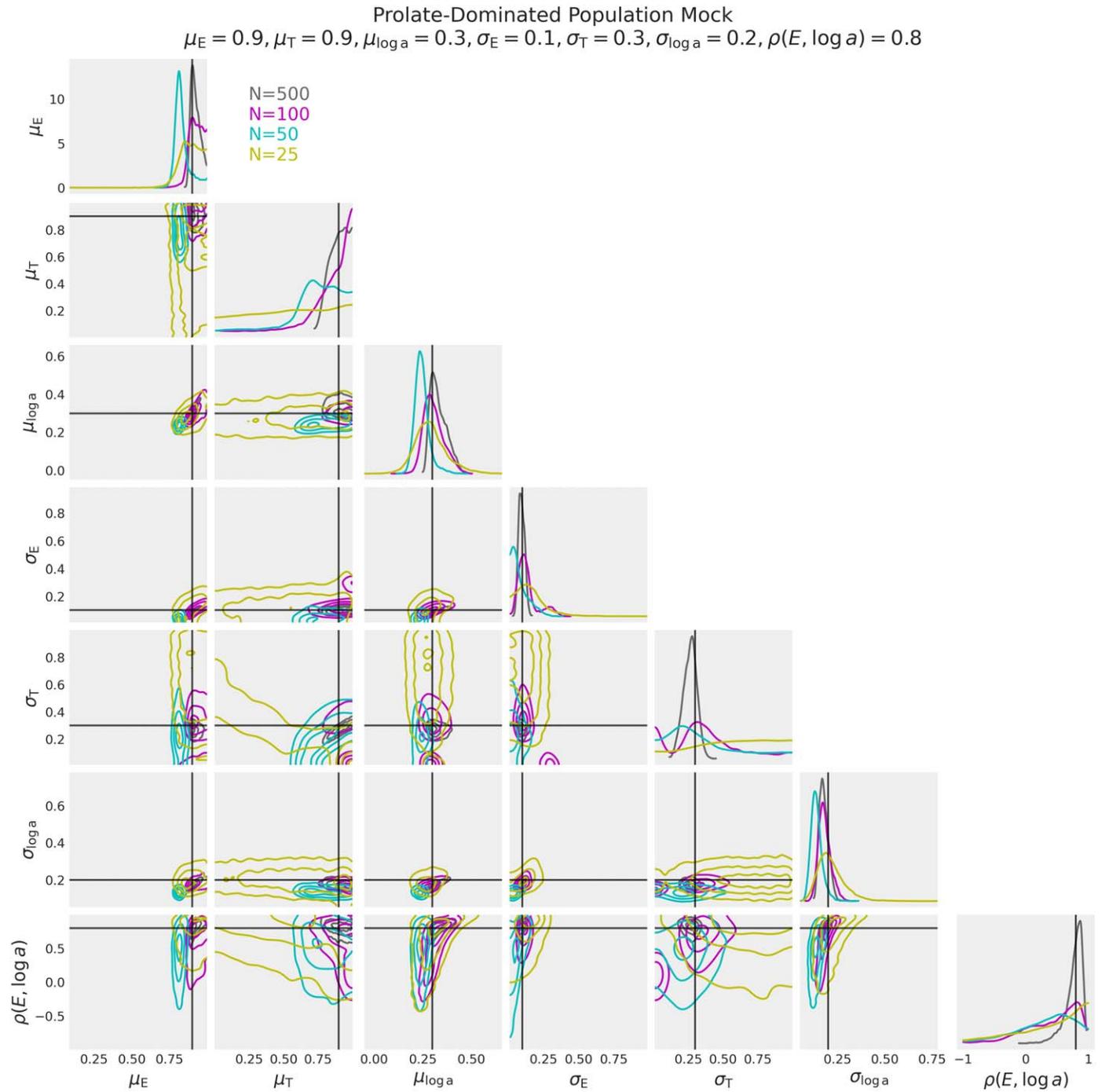
Figure 23. Corner plot showing parameter recovery tests for a prolate-dominated mock ellipsoid population. The different colors show results from our HMC for mocks with different sample sizes based on random Poisson sampling of the underlying true 2D projected b/a vs. $\log a$ histogram: $N = 500$ (gray), 100 (magenta), 50 (cyan), and 25 (yellow). The black vertical and horizontal lines mark the true parameter values. The HMC is successful at recovering the true parameter values for $N = 500$ and $N = 100$, and even for $N = 50$, albeit with broader posteriors for μ_T and σ_T. However, for $N = 25$ the posteriors for the triaxiality parameters are broad and the model cannot distinguish between prolate and oblate populations, though it does get the ellipticity, size, and correlation coefficient parameters right. This is generally true for other combinations of input mock parameters that we tried, for which analogous figures can be found at the Harvard Dataverse: doi:10.7910/DVN/SWTKVA.

well. To each mock source, we randomly assign a uniformly drawn F277W magnitude between 22 and 28.5 AB mag (going no fainter since that is already the completeness limit for compact point sources; Finkelstein et al. 2023), uniformly drawn axis ratio between $b/a = 0.2$ and 1.0, and uniformly drawn half-light radius between $0\!''\!05$ and $1\!''\!0$. These combinations of integrated Sérsic profile magnitude, half-light

radius and axis ratio allow us to explore completeness to exponential disks over a reasonably wide range of surface brightnesses and orientations. We also assign each source a random redshift uniformly drawn between $z = 0.5$ and 8.0 that sets its redshifted SED (though this does not have a big effect since JWST-NIRCAM probes redward of the Lyman-break for star-forming galaxies at these redshifts). The fake Sérsic
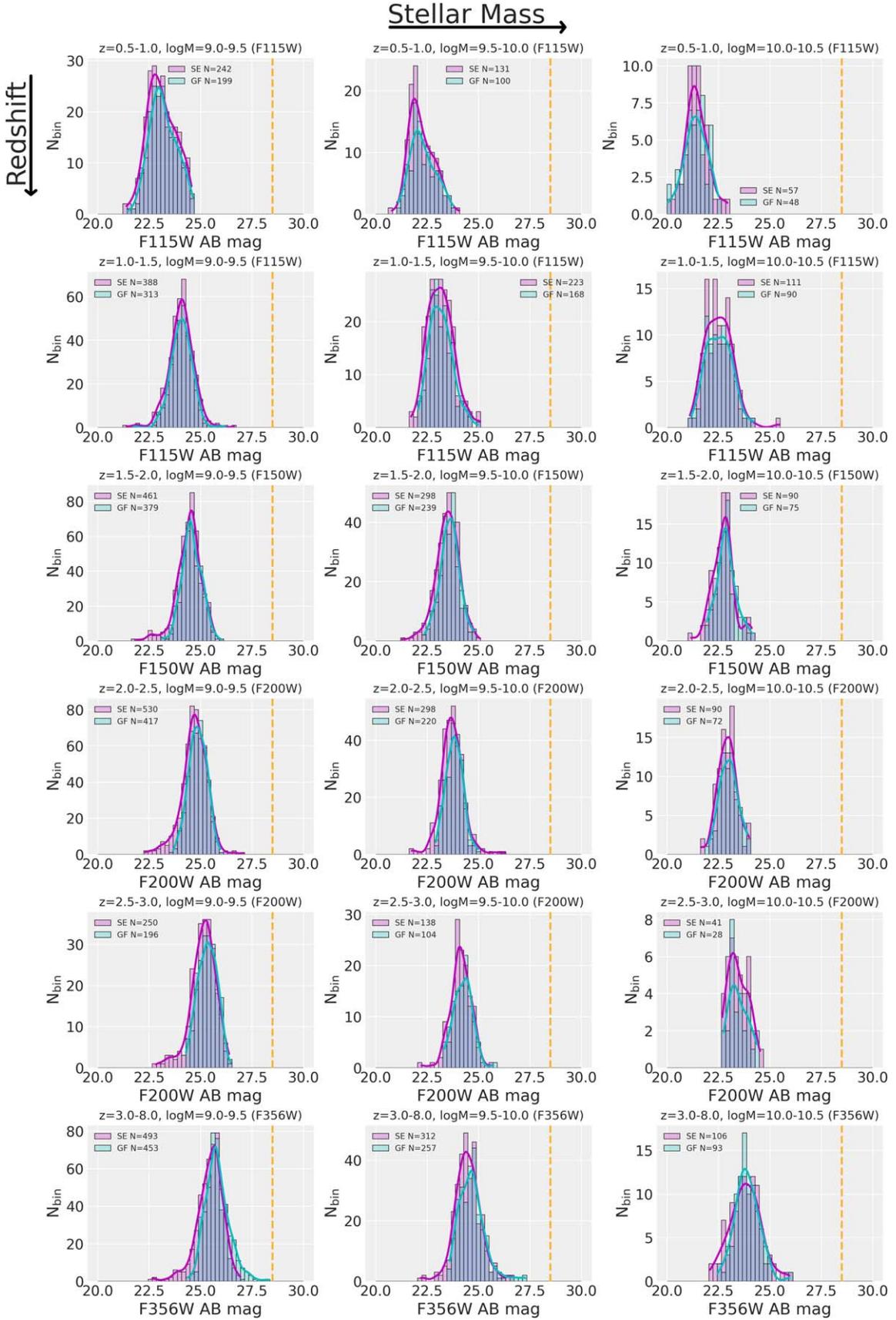
Figure 24. Distributions of apparent magnitude in the relevant filter for all sources included in our analysis across the mass–redshift grid. This shows that our sample, which is restricted to $\log M_*/M_\odot > 9$, is generally brighter than ~27 AB mag. The vertical orange line denotes the CEERS completeness limit for compact sources (28.5 AB mag; Finkelstein et al. 2023).
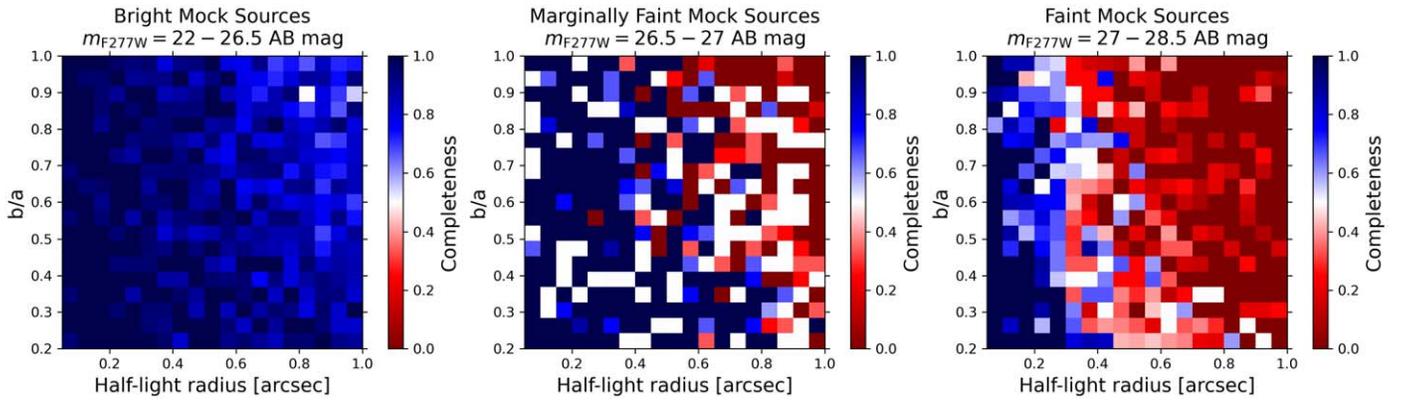
Figure 25. Completeness across the b/a–size diagram based on mock simulations where we injected 10^4 fake Sérsic profiles into the CEERS imaging with uniformly randomly assigned F277W magnitude, b/a, and half-light radius (fixing $n = 1$). Left: for galaxies brighter than 26.5 AB mag (corresponding to most of our selected sample), we are nearly complete even to large, face-on disks. Middle: for marginally faint sources (26.5–27 AB mag), we start to see hints of incompleteness to large face-on disks, but these would be at the extremely faint end of our sample. Right: for galaxies fainter than 27 AB mag, we are severely incomplete, but it is not clear whether these galaxies would satisfy our $\log M_*/M_\odot > 9$ cut. These simulations demonstrate that, for extended sources, JWST-CEERS goes ~ 2 mag deeper than HST-CANDELS, for which a cut of 24.5 AB mag is usually made in the F160W filter when studying galaxy morphology.

profiles are generated with GALFIT and added to the CEERS images, and those images are then run through the entire analysis pipeline including the same source detection setup with Source Extractor. Of the 10^4 mock input sources, 7648 were recovered by Source Extractor, but what we are interested in is the recovery fraction across the b/a–size diagram.

Figure 25 shows the completeness (i.e., the detection fraction of mock input sources) across the b/a–size plane. We consider the completeness as a function of three brightness bins: galaxies brighter than 26.5 AB mag in F277W (corresponding to most of our selected sample), marginally faint galaxies with 26.5–27 AB mag in F277W (at the extremely faint end of our sample), and truly faint galaxies with 27–28.5 AB mag in F277W. For the bright sample, we are nearly fully complete even to large face-on disks, so if these existed, CEERS should have detected them. For marginally faint galaxies, we start to see hints of incompleteness for large galaxies, particularly face-on ones, but again these are at the extremes of our sample selection. Finally, we are severely incomplete for the faint sample, but it is not clear whether such faint galaxies would satisfy our rather conservative mass cut of $\log M_*/M_\odot > 9$. Regardless, these simulations demonstrate that, for extended sources spanning a reasonable range of sizes and axis ratios, the CEERS survey is complete to ~ 26.5 AB mag, which is ~ 2 mag deeper than HST-CANDELS, for which studies of galaxy morphology are typically restricted to sources brighter than 24.5 in the F160W filter.

While the above is encouraging, we argue that more work is needed to fully appreciate the impact of completeness on our results. There are at least two other ways to address completeness that are beyond the scope of this paper. First, as shown in Appendix A of Zhang et al. (2019), it is possible to deproject and then reproject toy 2D Sérsic models to assess how much fainter the face-on version of an observed edge-on disk might be (see also van de Ven & van der Wel 2021). Zhang et al. (2019) use this approach to show that at most $\sim 20\%$ of face-on disks could have been missed by HST-CANDELS, which is not adequate to explain the $\sim 70\%$ prolate fractions found for high-redshift, low-mass galaxies by those authors. Second, one can take disk galaxies in hydrodynamical simulations and insert them into empty areas of the imaging

with different viewing angles and progressively larger distances until they become lost in the noise due to surface brightness dimming as $(1 + z)^4$. We suggest that creating mock images from hydrodynamical simulations will be a fruitful avenue for future completeness-related tests in the context of our work. Finally, we stress that these are simply photometric completeness simulations, not Sérsic parameter recovery tests, which are equally important and will be presented in E. McGrath et al. (2024, in preparation).

Appendix C
Sérsic Model Residuals

Figure 26 shows postage stamps of imaging data, GALFIT Sérsic model, GALFIT residuals, SE++ Sérsic model, and SE ++ residuals for some example galaxies fit by both SE++ and GALFIT. The fits are generally sensible for these and most of the other objects that we visually inspected. There were very few catastrophic failures. More goodness-of-fit details for GALFIT will be presented by E. McGrath et al. (2024, in preparation).

We showed in Figures 2 and 12 that SE++ tends to find larger sizes than Galfit, particularly for lower-mass, higher-redshift objects, which would be faint and more susceptible to noise and fitting issues. Both codes use the same PSFs and the same global background-subtracted images, and both also fix the local background to zero, so these cannot be the causes of the discrepancies. One contributor is the different source detection strategies, where Finkelstein et al. (2023) use only the so-called "hot mode" configuration for the original SExtractor (Bertin & Arnouts 1996), which is optimized for detecting small, faint sources. As described by van der Wel et al. (2012), the additional "cold mode" is designed to pick up large sources without artificially fragmenting them into individual objects. SE++ does not have such a distinction. The fact that SE++ picks up more galaxies overall (Table 1), preferentially more larger galaxies (Figure 2) and also more galaxies at the bright end (Figure 24), strongly suggests that the CEERS catalog we are using is genuinely missing some large-size galaxies that may otherwise be detected in the "cold mode."
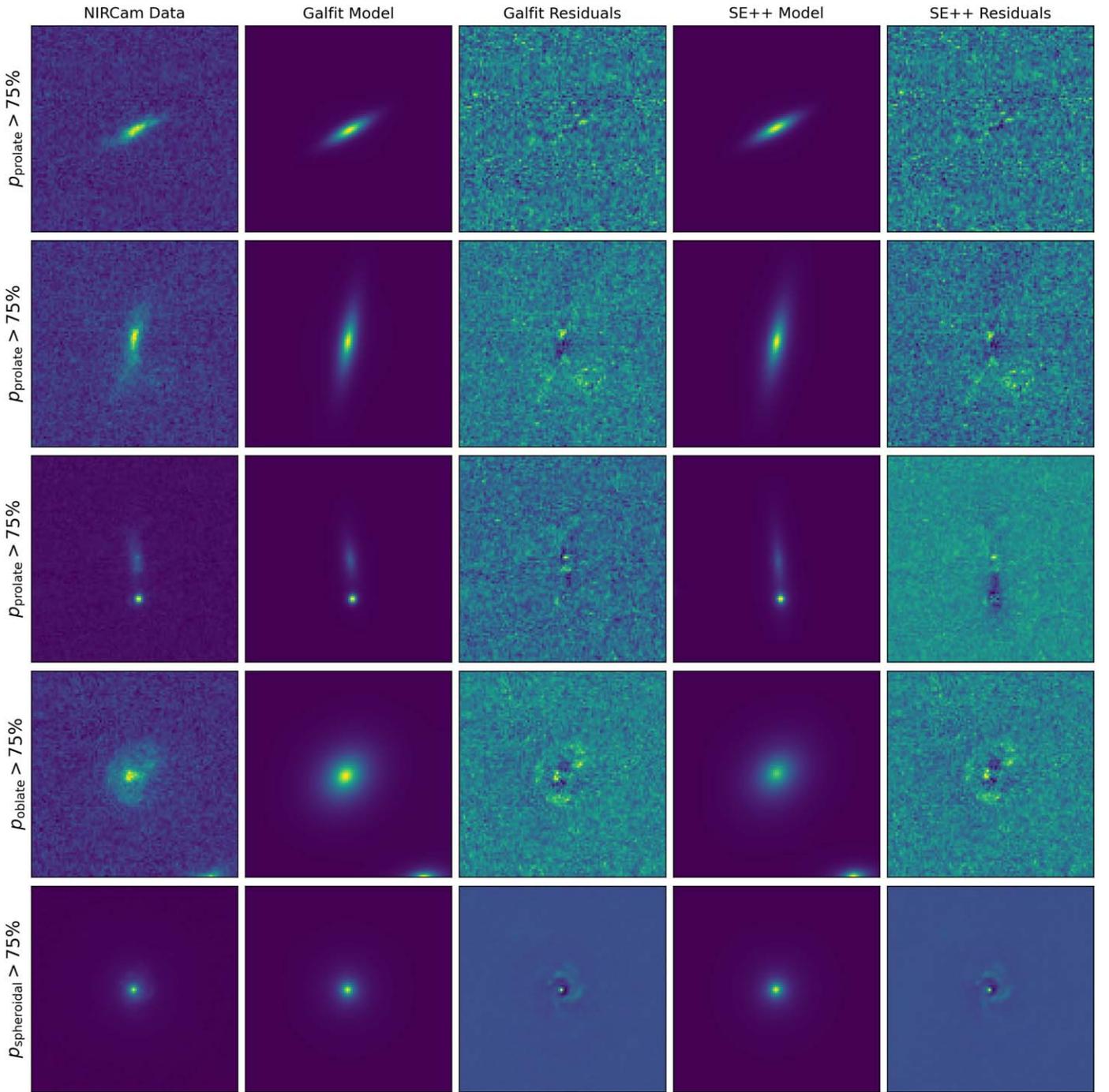
Figure 26. Sérsic model fits and residuals for some example objects with high probability of being prolate (top three rows), oblate (second from bottom row), and spheroidal (bottom row). These are all $3'' \times 3''$ cutouts in either F115W or F200W, which corresponds to rest-frame optical wavelengths given the redshifts of these galaxies. From left to right: cutout of NIRCam image, GALFIT model, GALFIT residuals, SE++ model, and SE++ residuals. These fits and residuals all look sensible, and this is also representative of many other galaxies that we visually inspected (including higher-redshift galaxies, for which we use redder filters).

However, that is not the whole story. The different deblending and masking algorithms of the two codes may also lead to different segmentation maps and hence structural parameters. In addition, the two codes may genuinely have ended up in different parts of the Sérsic model parameter space allowed for faint galaxies. Figure 27 shows that when SE++ finds a larger size than Galfit, it also tends to have a larger Sérsic index n than Galfit, and that is more pronounced for lower-mass (fainter) galaxies. When SE++ finds both a larger size and lower n than Galfit, it also tends to find a lower axis ratio. This may be due to the different fitting algorithms and/or the explicit or implicit priors used by the two codes, which would be important for low signal-to-noise ratio objects. Future mock Sérsic parameter recovery simulations for faint galaxies will help address these questions. For the purposes of our paper, the 3D shape model ends up in roughly the same region of its enormous 7D parameter space regardless of whether we use SE++ or Galfit measurements, which suggests that our overall conclusions are robust.
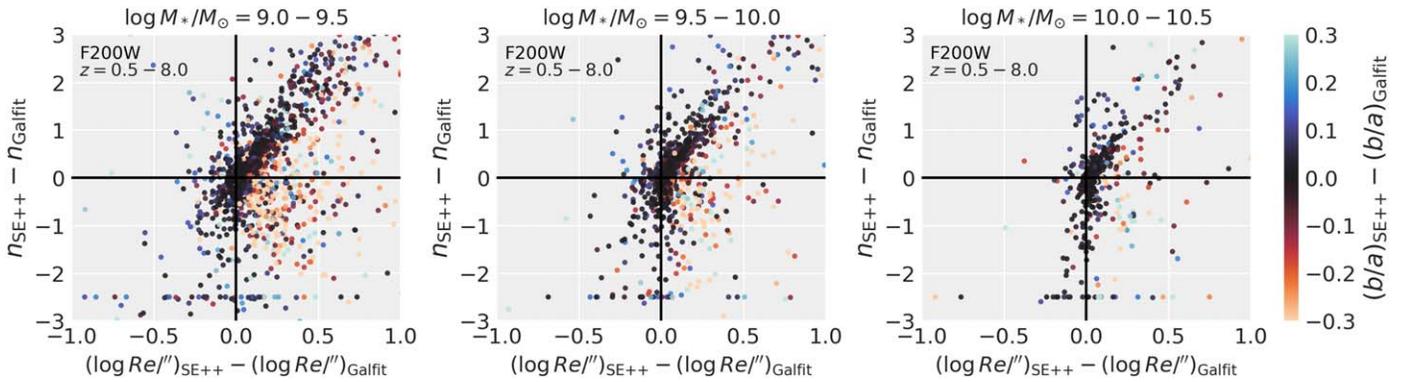
Figure 27. Illustration showing that SE++ and Galfit end up in different parts of the degenerate Sérsic n–R_e–b/a parameter space, particularly for low-mass galaxies. This is for all sources in CEERS that satisfy our selection cuts, namely $\log M_*/M_\odot = 9.0$–10.5 and $z = 0.5$–8.0, split into our mass bins increasing from left to right. When SE++ finds a larger size than Galfit, it also tends to find a higher Sérsic index than Galfit as evidenced by the large number of points in the upper right quadrant, especially for lower-mass (fainter) galaxies. The color bar shows how these n–R_e residuals correlate with b/a residuals, namely that when SE++ finds both a larger size and n than Galfit, it also tends to find a lower b/a. This is for F200W, but we see similar trends for the other filters.

ORCID iDs

Viraj Pandya ● https://orcid.org/0000-0002-2499-9205
Haowen Zhang (张昊文) ● https://orcid.org/0000-0002-4321-3538
Marc Huertas-Company ● https://orcid.org/0000-0002-1416-8483
Kartheik G. Iyer ● https://orcid.org/0000-0001-9298-3523
Elizabeth McGrath ● https://orcid.org/0000-0001-8688-2443
Guillermo Barro ● https://orcid.org/0000-0001-6813-875X
Steven L. Finkelstein ● https://orcid.org/0000-0001-8519-1130
Martin Kümmel ● https://orcid.org/0000-0003-2791-2117
Henry C. Ferguson ● https://orcid.org/0000-0001-7113-2738
Jeyhan S. Kartaltepe ● https://orcid.org/0000-0001-9187-3605
Joel Primack ● https://orcid.org/0000-0001-5091-5098
Avishai Dekel ● https://orcid.org/0000-0003-4174-0374
David C. Koo ● https://orcid.org/0000-0003-3385-6799
Greg L. Bryan ● https://orcid.org/0000-0003-2630-9228
Rachel S. Somerville ● https://orcid.org/0000-0002-6748-6821
Ricardo O. Amorín ● https://orcid.org/0000-0001-5758-1000
Pablo Arrabal Haro ● https://orcid.org/0000-0002-7959-8783
Micaela B. Bagley ● https://orcid.org/0000-0002-9921-9218
Eric F. Bell ● https://orcid.org/0000-0002-5564-9873
Emmanuel Bertin ● https://orcid.org/0000-0002-3602-3664
Luca Costantin ● https://orcid.org/0000-0001-6820-0015
Romeel Davé ● https://orcid.org/0000-0003-2842-9434
Mark Dickinson ● https://orcid.org/0000-0001-5414-5131
Robert Feldmann ● https://orcid.org/0000-0002-1109-1919
Adriano Fontana ● https://orcid.org/0000-0003-3820-2823
Raphael Gavazzi ● https://orcid.org/0000-0002-5540-6935
Mauro Giavalisco ● https://orcid.org/0000-0002-7831-8751
Andrea Grazian ● https://orcid.org/0000-0002-5688-0663
Norman A. Grogin ● https://orcid.org/0000-0001-9440-8872
Yuchen Guo ● https://orcid.org/0000-0002-4162-6523
ChangHoon Hahn ● https://orcid.org/0000-0003-1197-0902
Benne W. Holwerda ● https://orcid.org/0000-0002-4884-6756
Lisa J. Kewley ● https://orcid.org/0000-0001-8152-3943
Allison Kirkpatrick ● https://orcid.org/0000-0002-1306-1545
Dale D. Kocevski ● https://orcid.org/0000-0002-8360-3880
Anton M. Koekemoer ● https://orcid.org/0000-0002-6610-2048
Jennifer M. Lotz ● https://orcid.org/0000-0003-3130-5643
Ray A. Lucas ● https://orcid.org/0000-0003-1581-7825
Casey Papovich ● https://orcid.org/0000-0001-7503-8482
Laura Pentericci ● https://orcid.org/0000-0001-8940-6768

Pablo G. Pérez-González ● https://orcid.org/0000-0003-4528-5639
Nor Pirzkal ● https://orcid.org/0000-0003-3382-5941
Swara Ravindranath ● https://orcid.org/0000-0002-5269-6527
Caitlin Rose ● https://orcid.org/0000-0002-8018-3219
Raymond C. Simons ● https://orcid.org/0000-0002-6386-7299
Amber N. Straughn ● https://orcid.org/0000-0002-4772-7878
Sandro Tacchella ● https://orcid.org/0000-0002-8224-4505
Jonathan R. Trump ● https://orcid.org/0000-0002-1410-0470
Alexander de la Vega ● https://orcid.org/0000-0002-6219-5558
Stephen M. Wilkins ● https://orcid.org/0000-0003-3903-6935
Stijn Wuyts ● https://orcid.org/0000-0003-3735-1931
Guang Yang ● https://orcid.org/0000-0001-8835-7722
L. Y. Aaron Yung ● https://orcid.org/0000-0003-3466-035X

References

Abraham, R. G., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218
Alam, S. M. K., & Ryden, B. S. 2002, ApJ, 570, 610
Amorín, R., Pérez-Montero, E., Contini, T., et al. 2015, A&A, 578, A105
Bagley, M. B., Finkelstein, S. L., Koekemoer, A. M., et al. 2023, ApJL, 946, L12
Barro, G., Faber, S. M., Pérez-González, P. G., et al. 2013, ApJ, 765, 104
Barro, G., Perez-Gonzalez, P. G., Kocevski, D. D., et al. 2023, arXiv:2305.14418
Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. 2018, J. Mach. Learn. Res., 18, 1
Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bertin, E., Schefer, M., Apostolakos, N., et al. 2020, in ASP Conf. Ser. 527, Astronomical Data Analysis Software and Systems XXIX, ed. R. Pizzo et al. (San Francisco, CA: ASP), 461
Binggeli, B., & Popescu, C. C. 1995, A&A, 298, 63
Binney, J. 1978, MNRAS, 183, 501
Binney, J. 1985, MNRAS, 212, 767
Bournaud, F., Elmegreen, B. G., & Elmegreen, D. M. 2007, ApJ, 670, 237
Bradley, L., Sipőcz, B., Robitaille, T., et al. 2022, astropy/photutils: 1.5.0, v1.5.0, Zenodo, doi:10.5281/zenodo.6825092
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, ApJS, 200, 13
Brennan, R., Pandya, V., Somerville, R. S., et al. 2015, MNRAS, 451, 2933
Burkert, A. 2017, ApJ, 838, 93
Caldwell, N. 1983, AJ, 88, 804
Cappellari, M. 2016, ARA&A, 54, 597
Ceverino, D., Primack, J., & Dekel, A. 2015, MNRAS, 453, 408
Chang, Y.-Y., van der Wel, A., Rix, H.-W., et al. 2013, ApJ, 773, 149
Conselice, C. J. 2003, ApJS, 147, 1
Conselice, C. J., Grogin, N. A., Jogee, S., et al. 2004, ApJL, 600, L139
Contopoulos, G. 1956, ZA, 39, 126
Costantin, L., Méndez-Abreu, J., Corsini, E. M., et al. 2018, A&A, 609, A132

Costantin, L., Pérez-González, P. G., Méndez-Abreu, J., et al. 2021, ApJ, 913, 125

Costantin, L., Pérez-González, P. G., Méndez-Abreu, J., et al. 2022, ApJ, 929, 121

Cowie, L. L., Hu, E. M., & Songaila, A. 1995, AJ, 110, 1576

Dalcanton, J. J., & Shectman, S. A. 1996, ApJL, 465, L9

Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJL, 660, L1

Dekel, A., & Burkert, A. 2014, MNRAS, 438, 1870

de la Rosa, I. G., La Barbera, F., Ferreras, I., et al. 2016, MNRAS, 457, 1916

de Vaucouleurs, G. 1959, HDP, 53, 311

Dickinson, M., Dey, A., & Spinrad, H. 1995, in Galaxies in the Young Universe, ed. H. Hippelein, K. Meisenheimer, & H.-J. Röser (Berlin: Springer), 164

Dome, T., Fialkov, A., Mocz, P., et al. 2023, MNRAS, 519, 4183

Dressler, A., Oemler, A., Jr., Butcher, H. R., & Gunn, J. E. 1994, ApJ, 430, 107

Driver, S. P., Windhorst, R. A., & Griffiths, R. E. 1995, ApJ, 453, 48

Elmegreen, D. M., Elmegreen, B. G., & Hirst, A. C. 2004a, ApJL, 604, L21

Elmegreen, D. M., Elmegreen, B. G., Rubin, D. S., & Schaffer, M. A. 2005, ApJ, 631, 85

Elmegreen, D. M., Elmegreen, B. G., & Sheets, C. M. 2004b, ApJ, 603, 74

Emsellem, E., Cappellari, M., Krajnović, D., et al. 2007, MNRAS, 379, 401

Emsellem, E., Cappellari, M., Krajnović, D., et al. 2011, MNRAS, 414, 888

Ene, I., Ma, C.-P., Veale, M., et al. 2018, MNRAS, 479, 2810

Euclid Collaboration, Merlin, E., Castellano, M., et al. 2023, A&A, 671, A101

Fang, J. J., Faber, S. M., Koo, D. C., et al. 2018, ApJ, 858, 100

Ferguson, H. C., & Binggeli, B. 1994, A&ARv, 6, 67

Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, ApJL, 600, L107

Ferguson, H. C., & Sandage, A. 1989, ApJL, 346, L53

Ferreira, L., Adams, N., Conselice, C. J., et al. 2022, ApJL, 938, L2

Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, ApJL, 946, L13

Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, ApJ, 706, 1364

Förster Schreiber, N. M., Genzel, R., Lehnert, M. D., et al. 2006, ApJ, 645, 1062

Franx, M., Illingworth, G., & de Zeeuw, T. 1991, ApJ, 383, 112

Genzel, R., Tacconi, L. J., Eisenhauer, F., et al. 2006, Natur, 442, 786

Glazebrook, K. 2013, PASA, 30, e056

Glazebrook, K., Ellis, R., Santiago, B., & Griffiths, R. 1995, MNRAS, 275, L19

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Gullberg, B., Smail, I., Swinbank, A. M., et al. 2019, MNRAS, 490, 4956

Hamilton-Campos, K. A., Simons, R. C., Peeples, M. S., Snyder, G. F., & Heckman, T. M. 2023, ApJ, 956, 147

Hoffman, M. D., & Gelman, A. 2014, J. Mach. Learn. Res., 15, 1593

Holden, B. P., van der Wel, A., Rix, H.-W., & Franx, M. 2012, ApJ, 749, 96

Hubble, E. P. 1926, ApJ, 64, 321

Huertas-Company, M., Iyer, K. G., Angeloudi, E., et al. 2023, arXiv:2305.02478

Huertas-Company, M., Primack, J. R., Dekel, A., et al. 2018, ApJ, 858, 114

Ichikawa, S.-I. 1989, AJ, 97, 1600

Ichikawa, S. I., Wakamatsu, K. I., & Okamura, S. 1986, ApJS, 60, 475

Immeli, A., Samland, M., Gerhard, O., & Westera, P. 2004a, A&A, 413, 547

Immeli, A., Samland, M., Westera, P., & Gerhard, O. 2004b, ApJ, 611, 20

Ito, K., Valentino, F., Brammer, G., et al. 2023, arXiv:2307.06994

Kado-Fong, E., Greene, J. E., Huang, S., et al. 2020, ApJ, 900, 163

Kado-Fong, E., Petrescu, M., Mohammad, M., et al. 2021, ApJ, 920, 72

Kartaltepe, J. S., Rose, C., Vanderhoof, B. N., et al. 2023, ApJL, 946, L15

Kassin, S. A., Weiner, B. J., Faber, S. M., et al. 2012, ApJ, 758, 106

Kodra, D., Andrews, B. H., Newman, J. A., et al. 2023, ApJ, 942, 36

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Kormendy, J., & Bender, R. 1996, ApJL, 464, L119

Kümmel, M., Álvarez-Ayllón, A., Bertin, E., et al. 2022, arXiv:2212.02428

Lambas, D. G., Maddox, S. J., & Loveday, J. 1992, MNRAS, 258, 404

Lapiner, S., Dekel, A., Freundlich, J., et al. 2023, MNRAS, 522, 4515

Law, D. R., Steidel, C. C., Erb, D. K., et al. 2007, ApJ, 669, 929

Law, D. R., Steidel, C. C., Erb, D. K., et al. 2009, ApJ, 697, 2057

Law, D. R., Steidel, C. C., Shapley, A. E., et al. 2012, ApJ, 745, 85

Lewandowski, D., Kurowicka, D., & Joe, H. 2009, J. Multivar. Anal., 100, 1989

Li, H., Mao, S., Cappellari, M., et al. 2018, ApJL, 863, L19

Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177

Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163

Méndez-Abreu, J., Simonneau, E., Aguerri, J. A. L., & Corsini, E. M. 2010, A&A, 521, A71

Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, ApJ, 808, 101

Mocz, P., Fialkov, A., Vogelsberger, M., et al. 2020, MNRAS, 494, 2027

Nelson, E. J., Suess, K. A., Bezanson, R., et al. 2023, ApJL, 948, L18

Pacifici, C., Iyer, K. G., Mobasher, B., et al. 2023, ApJ, 944, 141

Padilla, N. D., & Strauss, M. A. 2008, MNRAS, 388, 1321

Pandya, V., Brennan, R., Somerville, R. S., et al. 2017, MNRAS, 472, 2054

Pandya, V., Primack, J., Behroozi, P., et al. 2019, MNRAS, 488, 5580

Papovich, C., Labbé, I., Quadri, R., et al. 2015, ApJ, 803, 26

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Pillepich, A., Nelson, D., Springel, V., et al. 2019, MNRAS, 490, 3196

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Putko, J., Sánchez Almeida, J., Muñoz-Tuñón, C., et al. 2019, ApJ, 883, 10

Ravindranath, S., Giavalisco, M., Ferguson, H. C., et al. 2006, ApJ, 652, 963

Reshetnikov, V. P., Dettmar, R. J., & Combes, F. 2003, A&A, 399, 879

Rix, H.-W., & Zaritsky, D. 1995, ApJ, 447, 82

Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, ApJL, 942, L42

Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, MNRAS, 483, 4140

Rong, Y., Dong, X.-Y., Puzia, T. H., et al. 2020, ApJ, 899, 78

Roychowdhury, S., Chengalur, J. N., Karachentsev, I. D., & Kaisina, E. I. 2013, MNRAS, 436, L104

Ryden, B. 1992, ApJ, 396, 445

Ryden, B. S. 2004, ApJ, 601, 214

Ryden, B. S., & Terndrup, D. M. 1994, ApJ, 425, 43

Salvatier, J., Wiecki, T., & Fonnesbeck, C. 2015, arXiv:1507.08050

Sánchez-Janssen, R., Méndez-Abreu, J., & Aguerri, J. A. L. 2010, MNRAS, 406, L65

Sandage, A. 1961, The Hubble Atlas of Galaxies (Washington, DC: Carnegie Institution)

Sandage, A., Freeman, K. C., & Stokes, N. R. 1970, ApJ, 160, 831

Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, ApJ, 801, 97

Satoh, Y. K., Kajisawa, M., & Himoto, K. G. 2019, ApJ, 885, 81

Sérsic, J. L. 1963, BAAA, 6, 41

Simons, R. C., Kassin, S. A., Trump, J. R., et al. 2016, ApJ, 830, 14

Simons, R. C., Kassin, S. A., Weiner, B. J., et al. 2017, ApJ, 843, 46

Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24

Somerville, R. S., Behroozi, P., Pandya, V., et al. 2018, MNRAS, 473, 2714

Stark, A. A. 1977, ApJ, 213, 368

Staveley-Smith, L., Davies, R. D., & Kinman, T. D. 1992, MNRAS, 258, 334

Stefanon, M., Yan, H., Mobasher, B., et al. 2017, ApJS, 229, 32

Straughn, A. N., Cohen, S. H., Ryan, R. E., et al. 2006, ApJ, 639, 724

Suess, K. A., Bezanson, R., Nelson, E. J., et al. 2022, ApJL, 937, L33

Sung, E.-C., Han, C., Ryden, B. S., et al. 1998, ApJ, 505, 199

Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, MNRAS, 458, 242

Tohill, C.-B., Bamford, S., Conselice, C., et al. 2023, arXiv:2306.17225

Tomassetti, M., Dekel, A., Mandelker, N., et al. 2016, MNRAS, 458, 4477

Tremblay, B., & Merritt, D. 1995, AJ, 110, 1039

van den Bergh, S. 2002, PASP, 114, 797

van den Bergh, S., Abraham, R. G., Ellis, R. S., et al. 1996, AJ, 112, 359

van der Kruit, P. C., & Freeman, K. C. 2011, ARA&A, 49, 301

van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24

van der Wel, A., Chang, Y.-Y., Bell, E. F., et al. 2014a, ApJL, 792, L6

van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014b, ApJ, 788, 28

van der Wel, A., Rix, H.-W., Holden, B. P., Bell, E. F., & Robaina, A. R. 2009, ApJL, 706, L120

van de Ven, G., & van der Wel, A. 2021, ApJ, 914, 45

van Dokkum, P. G., Franx, M., Kriek, M., et al. 2008, ApJL, 677, L5

Vega-Ferrero, J., Huertas-Company, M., Costantin, L., et al. 2024, ApJ, 961, 51

Vincent, R. A., & Ryden, B. S. 2005, ApJ, 623, 137

Ward, E. M., de la Vega, A., Mobasher, B., et al. 2023, arXiv:2311.02162

Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, ApJ, 799, 209

Yang, L., Morishita, T., Leethochawalit, N., et al. 2022, ApJL, 938, L17

Yuma, S., Ohta, K., & Yabe, K. 2012, ApJ, 761, 19

Yuma, S., Ohta, K., Yabe, K., Kajisawa, M., & Ichikawa, T. 2011, ApJ, 736, 92

Zhang, H., Primack, J. R., Faber, S. M., et al. 2019, MNRAS, 484, 5170

Zhang, J., Wuyts, S., Cutler, S. E., et al. 2023, MNRAS, 524, 4128

Zhang, J., Wuyts, S., Witten, C., et al. 2022, MNRAS, 513, 4814

Zolotov, A., Dekel, A., Mandelker, N., et al. 2015, MNRAS, 450, 2327