Extreme value statistics of the halo and stellar mass distributions at high redshift: are *JWST* results in tension with ΛCDM?

Christopher C. Lovell [1,2]★ Ian Harrison,[3] Yuichi Harikane,[4] Sandro Tacchella [5,6]
and Stephen M. Wilkins [7]

[1]*Institute of Cosmology and Gravitation, University of Portsmouth, Burnaby Road, Portsmouth PO1 3FX, UK*
[2]*Centre for Astrophysics Research, School of Physics, Engineering & Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK*
[3]*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK*
[4]*Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan*
[5]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[6]*Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[7]*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK*

ABSTRACT
The distribution of dark matter halo masses can be accurately predicted in the lambda cold dark matter (ΛCDM) cosmology. The presence of a single massive halo or galaxy at a particular redshift, assuming some baryon and stellar fraction for the latter, can therefore be used to test the underlying cosmological model. A number of recent measurements of very large galaxy stellar masses at high redshift ($z > 8$) motivate an investigation into whether any of these objects are in tension with ΛCDM. We use extreme value statistics to generate confidence regions in the mass–redshift plane for the most extreme mass haloes and galaxies. Tests against numerical models show no tension, neither in their dark matter halo masses nor their galaxy stellar masses. However, we find tentative $>3\sigma$ tension with recent observational determinations of galaxy masses at high redshift from both *Hubble Space Telescope* and *James Webb Space Telescope*, despite using conservative estimates for the stellar fraction ($f_\star \sim 1$). Either these galaxies are in tension with ΛCDM, or there are unaccounted for uncertainties in their stellar mass or redshift estimates.

Key words: galaxies: abundances – galaxies: haloes – galaxies: high-redshift.

1 INTRODUCTION

In the lambda cold dark matter (ΛCDM) paradigm, structure forms hierarchically in a bottom-up fashion, whereby density perturbations in the matter distribution at the time of inflation collapse first, then merge to form larger and larger structures. Within this framework, baryons fall in to virialized dark matter haloes and form galaxies (Somerville & Davé [2015]). At late times ($z < 2$), the largest overdensities collapse to form galaxy clusters, $>10^{14}$ $M_\odot$ haloes hosting a hot, X-ray emitting intracluster medium and hundreds, sometimes thousands of galaxies. At earlier times ($z > 2$) clusters have yet to form; galaxies and their host haloes are the largest virialized objects in the universe.

In this standard 'concordance' cosmology, the predicted halo mass distribution is trivial to calculate. It can then be used to constrain deviations from this concordance picture, for example the effect of non-Gaussian initial conditions (e.g. Matarrese, Verde & Jimenez [2000]; Jimenez & Verde [2009]). One approach exploits extreme value statistics (EVS; Gumbel [1958]; Katz & Nadarajah [2000]), which seeks to make predictions for the greatest (or least) valued random variable drawn from an underlying distribution. The power of EVS is that

it allows a test of the underlying cosmology from the observation of a *single* extreme object. It also provides both upper and lower limits on the mass of that object. Harrison & Coles ([2011]) applied EVS to the predicted halo mass function to generate the probability density function (PDF) of the most massive halo at a given redshift. They extended this in Harrison & Coles ([2012]) to survey volumes in order to assess whether any observed high redshift clusters ($1 < z < 2$) exceeded the maximum expected mass according to ΛCDM, finding no tension between observations and theory (see also Chongchitnan & Silk [2012]; Waizmann, Ettori & Moscardini [2012]). The approach has also been applied to the distribution of void sizes (Chongchitnan [2015]; Sahlén, Zubeldía & Silk [2016]).

In order to extend this approach to higher redshifts we require measurements of much lower halo masses than those hosting galaxy clusters. Unfortunately, such measurements are difficult, particularly at high redshift. Halo masses can be inferred from galaxy clustering, which has the benefit of not needing to assume the underlying baryonic physics, but cannot be used to measure the masses of individual objects. Abundance matching fixes the knee of the halo mass function to the knee of an observed luminosity function, but this explicitly uses features of the dark matter model to infer the halo masses. Another method is to measure the direct emission from the baryonic components of a galaxy, and assume some scaling with the total mass, or use spectral energy distribution (SED) modelling to

★ E-mail: christopher.lovell@port.ac.uk

estimate the baryonic masses. These masses can then be combined with the cosmological baryon fraction, f_b, and subsequent fractions of the relevant components, e.g. the stellar mass fraction $f_\star$, to derive the latent halo mass.

Steinhardt et al. (2016) first explored this approach, assuming a mass-to-light ratio measured at $z = 4$ and a fixed stellar–halo mass relation, finding some tension with observations at $4 < z < 8$. Behroozi & Silk (2018) also implement this latter method, using the cosmic baryon fraction as an absolute upper limit on the ratio of galaxy stellar mass to halo mass ($f_b \sim 0.16$; Planck Collaboration XIII 2016), whilst also allowing for redshift evolution in the stellar–halo mass relation.

This relationship can be inverted to place an upper limit on the halo mass for an observed stellar mass, and then compared with predicted halo mass functions. They found that, at the time, no observed galaxies exceeded these conservative upper limits. More recently, Boylan-Kolchin (2022) used a similar approach to Behroozi & Silk (2018) to test whether any recent high-mass, high-z candidates discovered in the first *James Webb Space Telescope* (*JWST*) data (Adams et al. 2022; Donnan et al. 2022; Finkelstein et al. 2022; Harikane et al. 2022a; Labbe et al. 2022; Naidu et al. 2022a; Rodighiero et al. 2022) exceed the limits set by ΛCDM. Uniquely, they test both the number density of galaxies above some stellar mass at early epochs, as well as limits placed on the stellar mass *density*. They find strong tension, particularly with the latter, for the candidates presented by Labbe et al. (2022) at $z \sim 10$, but less tension with other studies. Menci et al. (2022) have also used the abundance of high redshift *JWST* candidates to place constraints on dark energy models.

The EVS approach has a number of advantages over previous approaches. EVS mitigates the problem of the selection function in galaxy surveys; the most massive object acts as a lower limit on the most massive object one could have seen in a given survey. For estimates using full samples, uncertainty in the selection function can propagate into population measurements, e.g. the mean mass of your sample. As mentioned above, EVS also provides two-sided constraints (upper and lower limits) on the most massive object, and naturally considers the uncertainty in the mass of that object for a given survey volume/area. By combining with realistic functional forms for the stellar and baryon fraction, EVS naturally incorporates uncertainty in these parameters.

In this paper, we use EVS to calculate the full PDF of the mass of the most massive halo. We first compare to numerical simulations, computing the EVS PDF on fixed redshift hypersurfaces and compare to individual snapshots taken from these simulations (Section 2). We then proceed to calculate the statistics for observational survey volumes, and compare to recent observational measurements of galaxy masses, as well as make predictions for upcoming wide field surveys with *JWST*, Roman, and *Euclid* (Section 3). Observations of galaxy or halo masses significantly greater than the expected values for the *most massive* object would imply tension with ΛCDM.

Equally, by computing the full PDF with EVS, we can evaluate the *minimum* mass of the most massive halo or galaxy; if the largest observed object has a mass significantly lower than that predicted by EVS, this will place equally high tension on ΛCDM. We discuss our results and present our conclusions in Section 4. We also present a python package for computing confidence intervals for arbitrary survey areas (github.com/christopherlovell/evstats). We assume a flat ΛCDM cosmology with Planck Collaboration XIII (2016) parameters: $\Omega_M = 0.309$, $\Omega_b = 0.0486$, $\sigma_8 = 0.816$, and $h = 0.678$.

2 EXTREME VALUE STATISTICS ON A FIXED REDSHIFT HYPERSURFACE

EVS (Gumbel 1958; Katz & Nadarajah 2000) is concerned with the most extreme deviations from the median of a probability distribution. Consider a sequence of N random variates $\{M_i\}$ drawn from a cumulative distribution function (CDF), $F(m)$. There will be a largest value of the sequence, $M_{max} \equiv \sup\{M_1...M_N\}$. Assuming all variables are mutually independent and identically distributed (IID), the probability all deviates are less than or equal to some value m is given by

$$\Phi(M_{max} \leqslant m; N) = F_1(M_1 \leqslant m) \ldots F_N(M_N \leqslant m) \qquad (1)$$

$$= F^N(M). \qquad (2)$$

By differentiating equation (2), we find the PDF of the distribution,

$$\Phi(M_{max} = m; N) = N F'(m)[F(m)]^{N-1} \qquad (3)$$

$$= N f(m)[F(m)]^{N-1}, \qquad (4)$$

where $f(m)$ is the PDF of the original distribution ($f(m) = dF(m)/dm$), and $\Phi(M_{max} = m; N)$ is the *exact* extreme value PDF for N observations drawn from the known distribution $F(m)$.[1] We apply this general result to the case of the halo mass function, where $n(M)$ is the number density of haloes of mass M, and derive $f(M)$ and $F(M)$,

$$f(m) = \frac{1}{n_{tot}} \frac{dn(m)}{dm}, \qquad (5)$$

$$F(m) = \frac{1}{n_{tot}} \int_{-\infty}^{m} dM \frac{dn(M)}{dM}, \qquad (6)$$

where n_{tot} is a normalization factor giving the total (co-moving) number density of haloes

$$n_{tot} = \int_{-\infty}^{\infty} dM \frac{dn(M)}{dM}. \qquad (7)$$

Together, these equations can be used to estimate the EVS PDF for a constant redshift co-moving volume V, where the total number of haloes $N = n_{tot} V$.[2]

The IID assumption will be broken where haloes are significantly clustred. Where the volume probed is sufficiently large the distribution is essentially homogeneous. A number of studies have shown that this volume limit is achieved above $\sim (100 \, \text{Mpc})^3$ (Gelb & Bertschinger 1994; Power & Knebe 2006; Reed et al. 2007; Davis et al. 2011). Another consideration when comparing to periodic simulations is the impact of finite-volume effects on the abundance of galaxies. In a given periodic volume there is a maximum fundamental mode that can be represented, and large-scale power on scales greater than the size of the simulation volume will not be captured. Additionally, only discrete modes can be represented in periodic volumes. These approximations can impact the halo mass function, particularly at the high-mass end where the effect of these large modes is more pronounced (Reed et al. 2007; DeRose et al. 2019). However, the volume at which these effects become pronounced has been shown to be $< (100 \, \text{Mpc})^3$; Lukić et al. (2007) show that the effect in boxes of this volume is <10 per cent on the normalization of the halo mass function (<0.05 dex), subdominant to statistical error.

---

[1]For more details on the advantages of using the exact EVS statistics rather than those employing asymptotic theory (see Harrison & Coles 2012).

[2]It is impractical and unnecessary to integrate between infinite endpoints. We use conservative finite limits of $10^6 \leq m \leq 10^{17}$ at all redshifts; the choice of these makes no difference to our results.
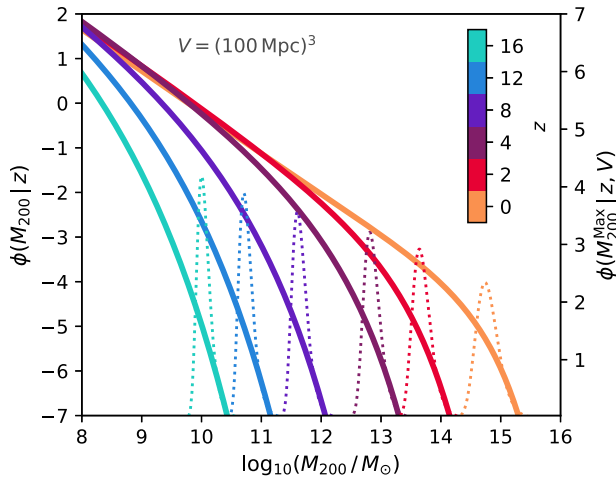
Figure 1. Halo mass function (solid lines; Behroozi, Wechsler & Conroy [2013]) in the range $0 \leq z \leq 20$, in units Mpc^{-3} dex^{-1}. The PDF, calculated using EVS, for the highest mass halo on a fixed hypersurface with volume $(100\,\text{Mpc})^3$ at each redshift is shown as a dotted line.

In this study, we only analyse simulated and observational volumes above this limit.

2.1 Halo masses

To calculate the EVS PDF, we must first assume a form for the halo mass function. We use the Behroozi et al. ([2013]) halo mass functions, which are calibrated using *N*-body simulations to the redshift range 2 $\leq z \leq 8$ based on the Tinker et al. ([2008]) mass functions (themselves derived from $0 < z \leq 2.5$ data). The redshift evolution of the halo mass function parameters from Behroozi et al. ([2013]) has smoothed off at $z > 8$, so we assume it is safe to extrapolate to these redshifts; we also note that we have tested using other forms for the mass function, and found little impact on our conclusions. Fig. [1] shows the halo mass function for a range of redshifts, along with the PDF for the highest mass halo on a constant redshift hypersurface, with volume $(100\,\text{Mpc})^3$, predicted by EVS. The peak of the PDF corresponds to the most probable mass of the most massive halo in the volume at that redshift.

A number of comparisons of the predictions of EVS with numerical simulations have been carried out in the past (Harrison & Coles [2012]; Watson et al. [2014]), all showing consistency. To test our results, we compare to predictions for the most massive halo from two hydrodynamic cosmological simulations. The fiducial EAGLE simulation (Crain et al. [2015]; Schaye et al. [2015]) is a $(100\,\text{Mpc})^3$ cosmological volume evolving both dark matter and baryons self-consistently. The FLARES simulations (Lovell et al. [2021]; Vijayan et al. [2021]) use the EAGLE physics model to resimulate zooms of a range of overdensities during the epoch of reionization, extending the dynamic range over periodic cosmological volumes. Since FLARES is not a continuous periodic box, one must calculate the 'effective volume' of the combined zoom regions, which is dependent on the mass/luminosity of the selected galaxies. We use a fixed effective volume of $(550\,\text{Mpc})^3$, which roughly corresponds to that for the most massive halo/galaxy at all redshifts. The underlying halo mass function in both of these simulations is not identical to that presented by Behroozi et al. ([2013]), but is in reasonably good agreement at the redshifts shown.

In this example, as well as in the sections below, we wish to show the EVS PDF for a range of redshifts simultaneously. To do this,
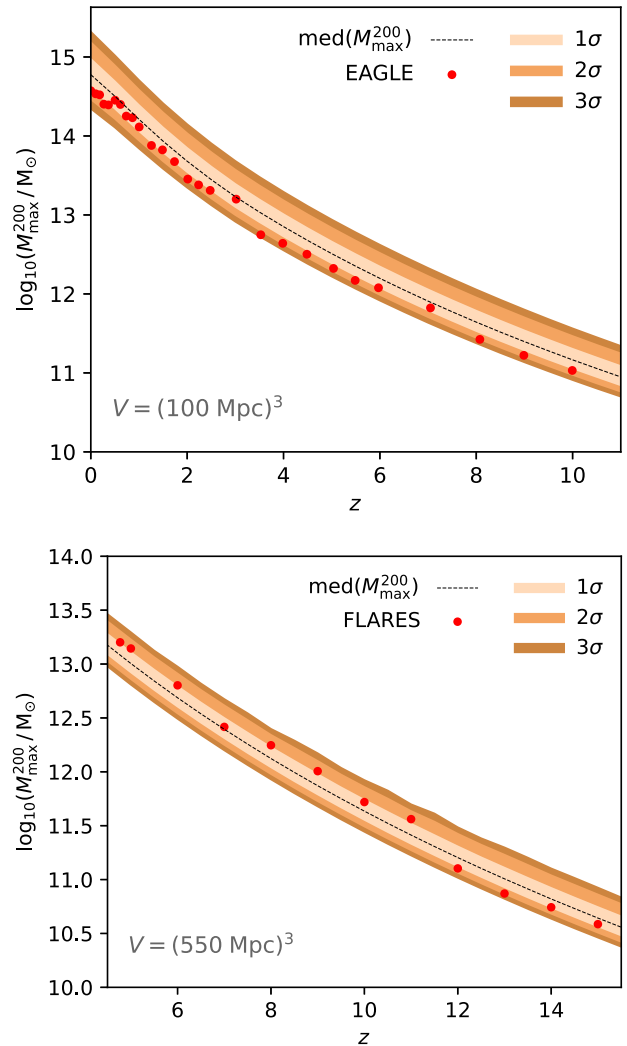
Figure 2. *Top*: confidence intervals for the EVS PDF on a fixed redshift hypersurface, evaluated at a range of redshifts, for haloes taken from a $(100\,\text{Mpc})^3$ volume. The most massive halo from each available snapshot in the fiducial EAGLE simulation (with identical volume) is shown in red. *Bottom*: as above, but for the $(550\,\text{Mpc})^3$ effective volume of FLARES, and showing the most massive halo at each available snapshot.

we calculate the PDF for narrow redshift intervals ($\Delta z = 0.2$, where the number of bins is chosen so that $N_{\text{bins}} >> N_{\text{galaxies}}$).[3] Each bin can be thought of as a Bernoulli trial, therefore there is a non-zero probability of exceeding a given contour threshold; we have tested and found that, for our chosen binning, this probability is negligible. Further discussion on this effect is provided in Appendix [A]. We integrate over these PDFs to find the $[1,2,3]\sigma$ confidence intervals, and plot these along with the median of each distribution. Fig. [2] shows the PDF of M_{200} evaluated at a range of redshifts, along with the value of M_{200} from the most massive halo selected from each available simulation snapshot in EAGLE and FLARES. All of the simulated haloes lie within the reasonably tight 3σ confidence intervals. The level of agreement is very good, and gives us confidence that our

---

[3]The choice of Δz has negligible impact on the results. For a single bin, the maximum mass in this bin is the same as that measured over multiple bins; due to hierarchical and positive structure formation, this tends to be biased towards those objects at lower redshifts.

EVS scheme is correctly able to produce realistic contours in the halo mass–redshift plane, despite assuming a slightly different halo mass function to that produced in the simulations. We now introduce astrophysical effects to predict the stellar mass distribution.

2.2 Stellar masses

To convert our PDF for the halo distribution to a PDF for the galaxy stellar mass, we must account for both the baryon fraction, f_b, and the stellar fraction, $f_\star$. The baryon fraction is set by our assumed cosmology ($f_b = 0.16$; Planck Collaboration XIII 2016). We assume a fixed value, though we note that this value can vary in different cosmic environments (e.g. lower than the universal value in local galaxy clusters (Gonzalez et al. 2013). This is an effect that could be accounted for by using a functional distribution for f_b, though we note that at high redshifts deviations from the universal value are not expected to be as large, due to the shorter time for feedback effects to have imprinted on baryon distributions.

The stellar fraction is dependent on the astrophysics that converts cold gas into stars. A conservative upper limit is to assume *all* baryons are converted into stars, $f_\star = 1$, and simply multiply the halo PDF by the product of the baryon and stellar fractions

$$\Phi(M_*) = \Phi(M_{DM}) f_b f_\star . \tag{8}$$

The bottom panel of Fig. 3 shows an example of the halo EVS PDF, as well as the stellar mass PDF obtained using a fixed stellar fraction of unity. In reality, measurements of the stellar fraction suggest much lower values, particularly in the most massive haloes (e.g. Giodini et al. 2009). To account for this, we assume a (truncated; $0 \leq f_\star \leq 1$) lognormal distribution of $f_\star$,

$$f_\star = \ln N(\mu, \sigma^2), \tag{9}$$

where $\mu = e^{-2}$ and $\sigma = 1$. This simple model ignores the dependence of the stellar fraction on redshift and halo mass, but incorporates the range of values inferred from simple halo models (Tacchella, Trenti & Carollo 2013; Tacchella et al. 2018) and observations (e.g. Harikane et al. 2016, 2018; Stefanon et al. 2021). However, it is worth noting that in the pre-reionization epoch ($z > 10$) high-star formation efficiencies, close to the cosmic baryon fraction, have been predicted from theoretical models (Susa & Umemura 2004). We then calculate the product of this PDF with the halo PDF, whilst assuming the same fixed baryon fraction. Fig. 3 shows the lognormal form of $f_\star$ as well as other parametric forms, and an example of the stellar mass PDF obtained using these different distributions. We also present upper 3σ limits based on assuming $f_s = 1$ and applying directly to the halo EVS PDF, which can be interpreted as conservative upper limits.

Fig. 4 shows the stellar mass PDF for a fixed redshift hypersurface, with volume $(100\,\text{Mpc})^3$ and $(550\,\text{Mpc})^3$, assuming this lognormal distribution of $f_\star$. The uncertainties are larger than for the halo PDF as expected. Results from the EAGLE and FLARES simulation are shown, and all lie within the uncertainties. There is a noticeable plateau in the maximum stellar mass in EAGLE as we go to lower redshifts, which demonstrates the redshift and halo-mass dependent evolution of $f_\star$, particularly at low-z. For now we ignore this redshift and halo-mass dependence, and assume a fixed distribution, however one could incorporate these effects.

3 EXTREME VALUE STATISTICS FOR AN OBSERVATIONAL SURVEY VOLUME

In a galaxy survey, we do not observe galaxies at a fixed redshift, and must therefore take account of the change in volume with redshift in
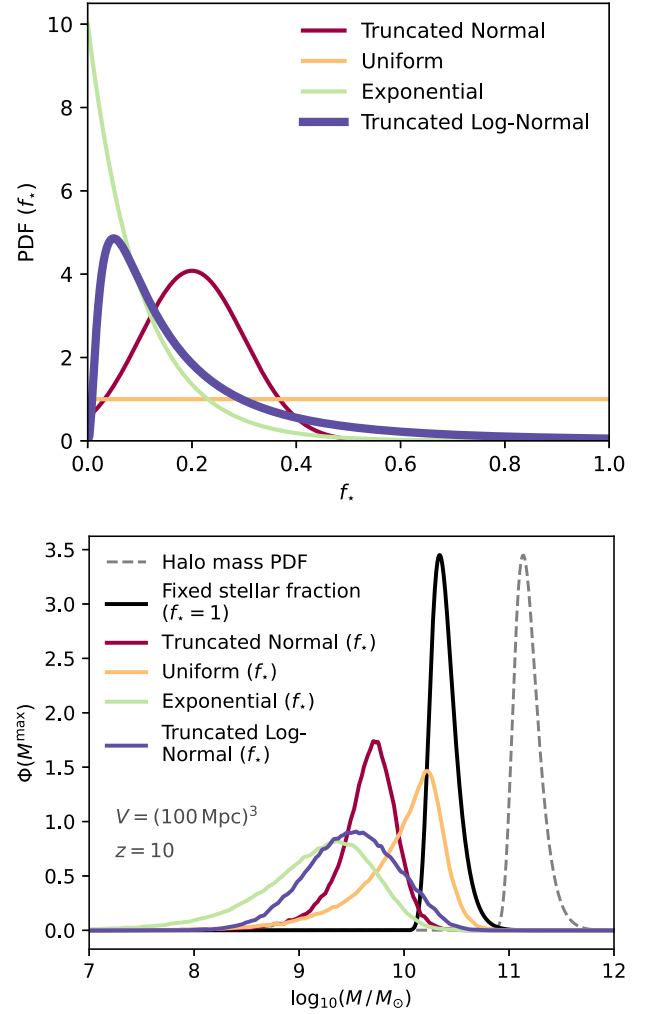
Figure 3. *Top*: Parametric forms for the distribution of $f_\star$. *Bottom*: stellar mass EVS PDF as a function of mass for a fixed redshift hypersurface at $z = 10$, assuming different parametric forms for $f_\star$. The halo mass PDF is shown by the grey dashed line. The black solid line shows the stellar mass PDF assuming $f_\star = 1$.

an expanding universe, as well as the change in the number density of haloes with redshift due to the growth of structure. The PDF and CDF for haloes in a fixed fraction of the sky, f_{sky}, between redshifts z_{min} and z_{max} is then given by

$$f(m) = \frac{f_{sky}}{n_{tot}} \left[\int_{z_{min}}^{z_{max}} dz \frac{dV}{dz} \frac{dn(m, z)}{dm} \right] \tag{10}$$

$$F(m) = \frac{f_{sky}}{n_{tot}} \left[\int_{z_{min}}^{z_{max}} \int_{-\infty}^{m} dz\, dM \frac{dV}{dz} \frac{dn(M, z)}{dM} \right], \tag{11}$$

where

$$n_{tot} = f_{sky} \left[\int_{z_{min}}^{z_{max}} \int_{-\infty}^{\infty} dz\, dM \frac{dV}{dz} \frac{dn(M, z)}{dM} \right]. \tag{12}$$

We can then use these with equation (4) to find the halo EVS for a given survey. As in Section 2, we assume the Behroozi et al. (2013) halo mass functions, a fixed baryon fraction, $f_b = 0.16$, and a truncated lognormal distribution for the stellar fraction.
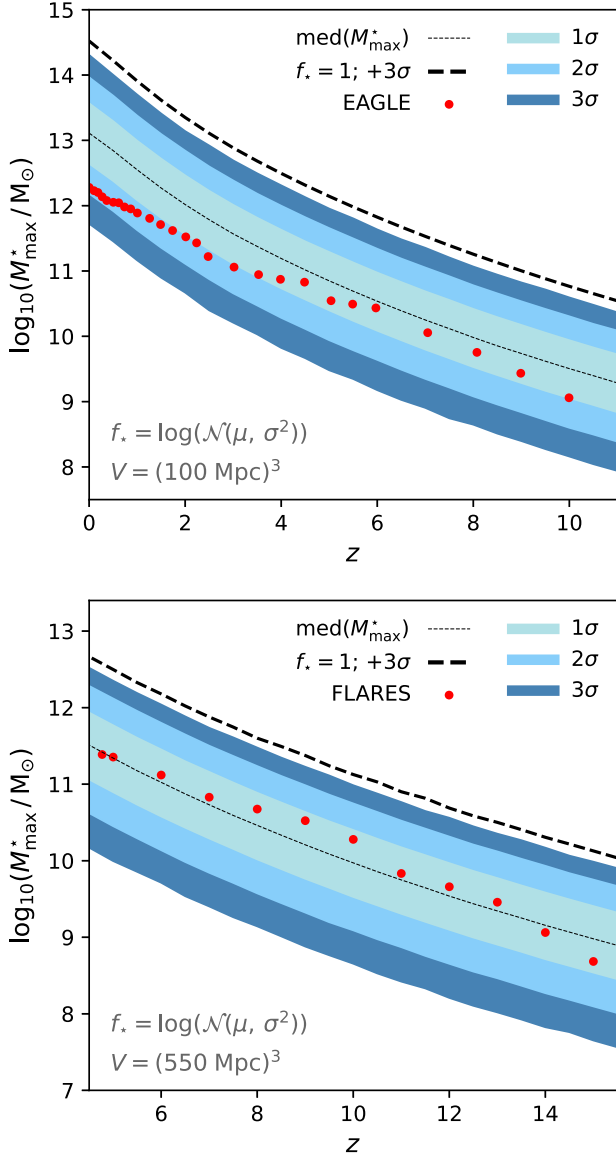
Figure 4. *Top*: the stellar mass EVS confidence intervals on a fixed redshift hypersurface, evaluated at a range of redshifts, for galaxies taken from a $(100\,\mathrm{Mpc})^3$ volume. The dashed line shows the 3σ upper limit assuming a stellar fraction of unity. The most massive galaxy from each available snapshot in the fiducial EAGLE simulation (with identical volume) is shown in red. *Bottom*: as above, but for the $(550\,\mathrm{Mpc})^3$ effective volume of FLARES, showing the most massive galaxy at each available snapshot from all resimulations.

3.1 Eddington bias

To compare our theoretical mass functions with observations, we need to correct for Eddington Bias (Eddington 1913). For haloes, the steepness of the mass function means there are significantly more low-mass haloes than high mass, so there is greater upscatter of low-mass halo measurements than down scatter of higher mass halo measurements, boosting the apparent number of higher mass haloes. The same effect applies to galaxy stellar masses. We correct using the following:

$$\ln M_{\mathrm{edd}} = \ln M_{\mathrm{obs}} + \frac{1}{2}\,\epsilon\,\sigma_{\ln M}^2 \,, \tag{13}$$

where ϵ is the local slope of the underlying halo mass function, and $\sigma_{\ln M}$ is the uncertainty in the halo/stellar mass estimate. We choose

to correct the observations, using quoted uncertainties on the stellar mass. For stellar masses, we derive ϵ from the halo mass function, but use the halo mass given by the observed stellar mass multiplied by the inverse baryon fraction. The true steepness of the galaxy stellar mass function is known to be steeper at low-z due to AGN feedback, but this effect is expected to be less extreme at the high redshifts considered here.

3.2 Observational comparison (pre-*JWST*)

A number of recent studies have presented estimates of galaxy stellar masses at high redshift ($z > 8$). Tacchella et al. (2022b) present an analysis of a number of bright galaxies selected from *HST* CANDELS fields, with associated *Spitzer*/IRAC fluxes (Finkelstein et al. 2021). They use the PROSPECTOR (Leja et al. 2017; Johnson et al. 2021) Bayesian SED fitting code to obtain stellar mass estimates from the photometric data, then present an analysis of the likelihood of the stellar mass estimates in ΛCDM using the methodology of Behroozi & Silk (2018). They probe down to some limiting number density $\Phi > 10^{-6}\,\mathrm{Mpc}^{-3}$, approximately that expected for a similar survey area, and assume a baryon fraction $f_{\mathrm{b}} = 0.16$ and a conservative stellar fraction of unity, $f_\star = 1$. Two galaxies in their sample are in tension with these constraints, COSMOS-20646 and UDS-18697, at 3σ and 4.6σ, respectively. However, they argue that cosmic variance (Trenti & Stiavelli 2008), observational uncertainties (particularly the contribution of near neighbours contaminating the IRAC photometry), and measurement uncertainty (related to SED modelling assumptions, such as the choice of prior on the star formation history) reduce this tension significantly.

Here, we repeat their analysis using the EVS framework. The top panel of Fig. 5 shows the galaxy stellar mass EVS PDF for the combined survey area of Finkelstein et al. (2015), with the stellar mass estimates for the selected galaxies from Tacchella et al. (2022b) after correcting for Eddington bias. The majority of the galaxies are within the 3σ uncertainties, however the same two galaxies identified in Tacchella et al. (2022b) lie outside these bounds, even after correcting for Eddington bias. They lie on the 3σ upper limit assuming a stellar fraction of unity.

Recently, Harikane et al. (2022b) presented two bright galaxy candidates, HD1 and HD2, from Hyper Suprime-Cam, VISTA, and *Spitzer* observations of the COSMOS and SXDS fields. Their photometry suggests redshifts of $z = 15.2^{+1.2}_{-2.1}$ and $12.3^{+0.4}_{-0.3}$, and one of the sources (HD1) additionally has a tentative detection of [O III] 88 μm, giving a spectroscopic redshift of $z = 13.27$. Estimates of physical properties for these sources are not well constrained; however, they quote stellar masses in the range 10^9–10^{11} and $10^{9.8}$–10^{11} M$_\odot$ for HD1 and HD2, respectively. These ranges bound the stellar masses obtained using the $M_\star$–M_{UV} relation at $z = 8$ from Song et al. (2016).

The bottom panel of Fig. 5 shows the galaxy stellar mass EVS PDF for the combined survey area (2.3 deg^2), as well as the stellar mass estimates for HD1 and HD2 after correcting for Eddington bias. We show HD1 using both the photometric and spectroscopic redshift estimates. The stellar mass correction due to Eddington bias is quite large due to the significant uncertainties in the stellar masses, which brings the estimates within the 3σ contours for both objects.

3.3 Observational comparison (*JWST* first results)

In the short time since the first data from *JWST* was released there have been a number of studies estimating the redshifts and stellar masses of high-redshift galaxies (e.g. Adams et al. 2022; Donnan
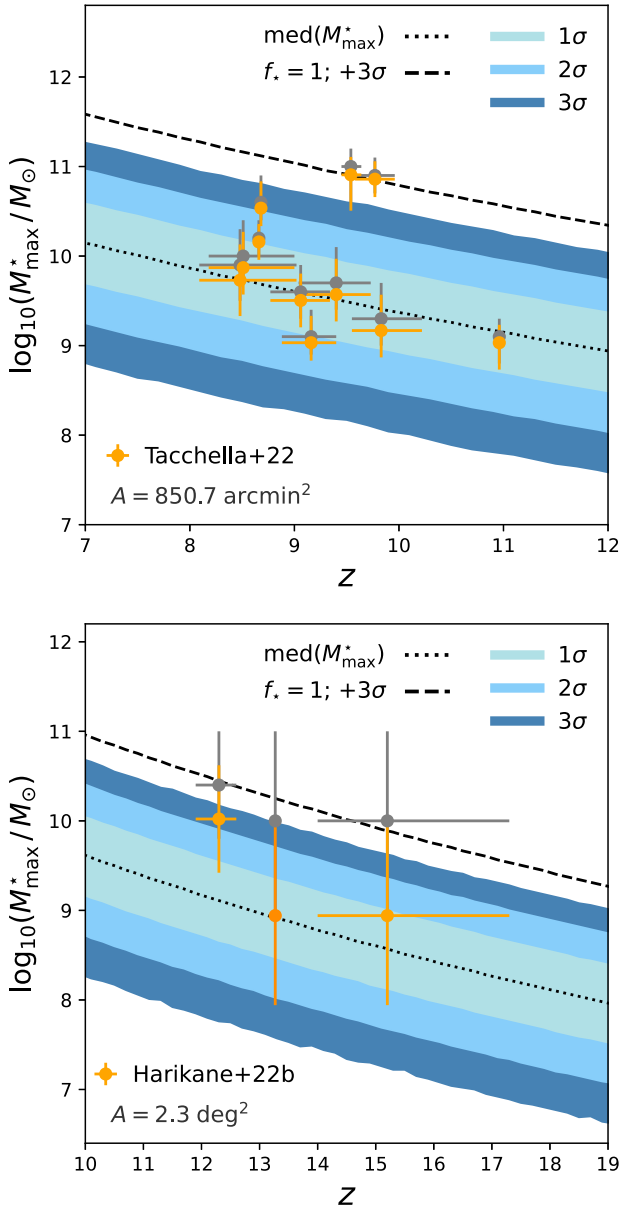
Figure 5. *Top*: The stellar mass EVS confidence intervals for an observational survey volume with area $850.7\,\mathrm{arcmin}^2$, evaluated at a range of redshifts. Stellar mass estimates from Tacchella et al. (2022b) for galaxies selected from the *HST*/CANDELS fields (Finkelstein et al. 2021), are shown in yellow, after correcting for Eddington bias. Grey points show the uncorrected stellar mass estimates. The dashed line shows the 3σ upper limit assuming a stellar fraction of unity. *Bottom*: as above, but showing an observational survey volume with area $2.3\,\mathrm{deg}^2$, and observational results from Harikane et al. (2022b) converted to a stellar mass estimate using a linear scaling relation.

et al. 2022; Finkelstein et al. 2022; Harikane et al. 2022a; Labbe et al. 2022; Naidu et al. 2022a; Rodighiero et al. 2022). Many of these have proposed candidates that lie at the extremes of the redshift–stellar mass plane. Here, we test using the EVS framework whether any of these candidates are in tension with ΛCDM.

Labbe et al. (2022) presented seven $> 10^{10}\,\mathrm{M}_\odot$ candidates at $7 < z < 11$ in a 40 arcmin^2 area, taken from the CEERS program, using EAZY and PROSPECTOR for the photometric redshift and stellar mass estimates, respectively. The top left panel of Fig. 6 shows these

candidates on the stellar mass–redshift plane, with the EVS PDF for an identical survey area. 5 out of the 7 candidates lie above the 3σ upper limits assuming a stellar fraction of unity. Since they do not provide error estimates on the stellar masses we cannot evaluate the effect of Eddington bias, but if we assume 0.3 dex errors this relieves the tension with 3 of these candidates, leaving only 2 outside of the 3σ contours assuming a lognormal stellar fraction.

A particularly exciting discovery in the early data is a potential $z \sim 17$ candidate, also identified in the CEERS data (40 arcmin^2), first presented by Donnan et al. (2022). There is some debate as to the accuracy of this photometric redshift estimate (e.g. Zavala et al. 2022), with a $z \sim 5$ solution potentially also capable of explaining the observed photometry. Naidu et al. (2022b) provide stellar mass and photometric redshift estimates for two potential $z \sim 5$ solutions, as well as the $z \sim 17$ solution; we present the higher redshift solution in the top left panel of Fig. 6. This solution is in significant tension with the EVS PDF, even after accounting for stellar mass errors and the resulting Eddington bias. We have checked the lower redshift solutions, and found that these are not in tension.

Harikane et al. (2022a) also identify galaxies out to $z \sim 17$ in the ERO and ERS programs, covering a total area of 90 arcmin^2. They use PROSPECTOR for photometric redshift and stellar mass estimates, and find good agreement in the stellar mass estimates for most of the objects from other studies that identified the same objects (Donnan et al. 2022; Finkelstein et al. 2022; Naidu et al. 2022a). The main exception being the $z \sim 17$ source mentioned above, for which Harikane et al. (2022a) and Donnan et al. (2022) predict lower stellar masses (by ~ -0.7 dex) than those obtained by Naidu et al. (2022b). The top right panel of Fig. 6 shows these candidates compared to our predicted EVS PDF; all objects lie within the contours, even at the most extreme redshifts.

Finally, in the bottom panel of Fig. 6 we show a selection of high-mass candidates from Rodighiero et al. (2022) over the 10 arcmin^2 area covering the SMACS0723 cluster. They particularly target those objects that are dark in UV–optical rest-frame wavelengths, which they cite as evidence for high levels of dust obscuration. We plot a number of their candidates, the majority of which are consistent with our EVS predictions. However, one candidate, at $z \sim 10$, is in significant tension. We additionally plot the 3σ upper limits, for a stellar fraction of unity, assuming a whole sky survey, and show that this candidate is even in tension with this highly conservative limit.

3.4 Predictions for future surveys

As well as comparing to results from existing surveys, we can also make predictions for a number of planned upcoming surveys.

There are a number of relatively wide area surveys planned with *JWST* Fig. 7 shows predictions for the full COSMOS Web survey area (0.6 deg^2; Kartaltepe et al. 2021), as well as the combined medium and deep survey areas in the JADES survey (136 arcmin^2; Rieke et al. 2019). *JWST*'s limited survey area is not expected to identify the rarest objects at lower redshifts, but its exceptional depth may be capable of discovering extreme objects in the epoch of very first star and galaxy formation ($z \gtrsim 15$), as shown in Section 3.3.

Wider field surveys are more likely to discover extreme objects that may challenge our cosmological and galaxy evolution models. The Roman High Latitude Wide Area Survey (HLWAS), consisting of spectroscopic and imaging components (Wang et al. 2022), will cover an area of 1700 deg^2; stellar mass EVS PDF predictions for such a survey are shown in Fig. 7. *Euclid* will also carry out a wide spectroscopic survey with the aim of constraining models of dark energy (1500 deg^2; Eulcid Collaboration XVII 2022), as well as two
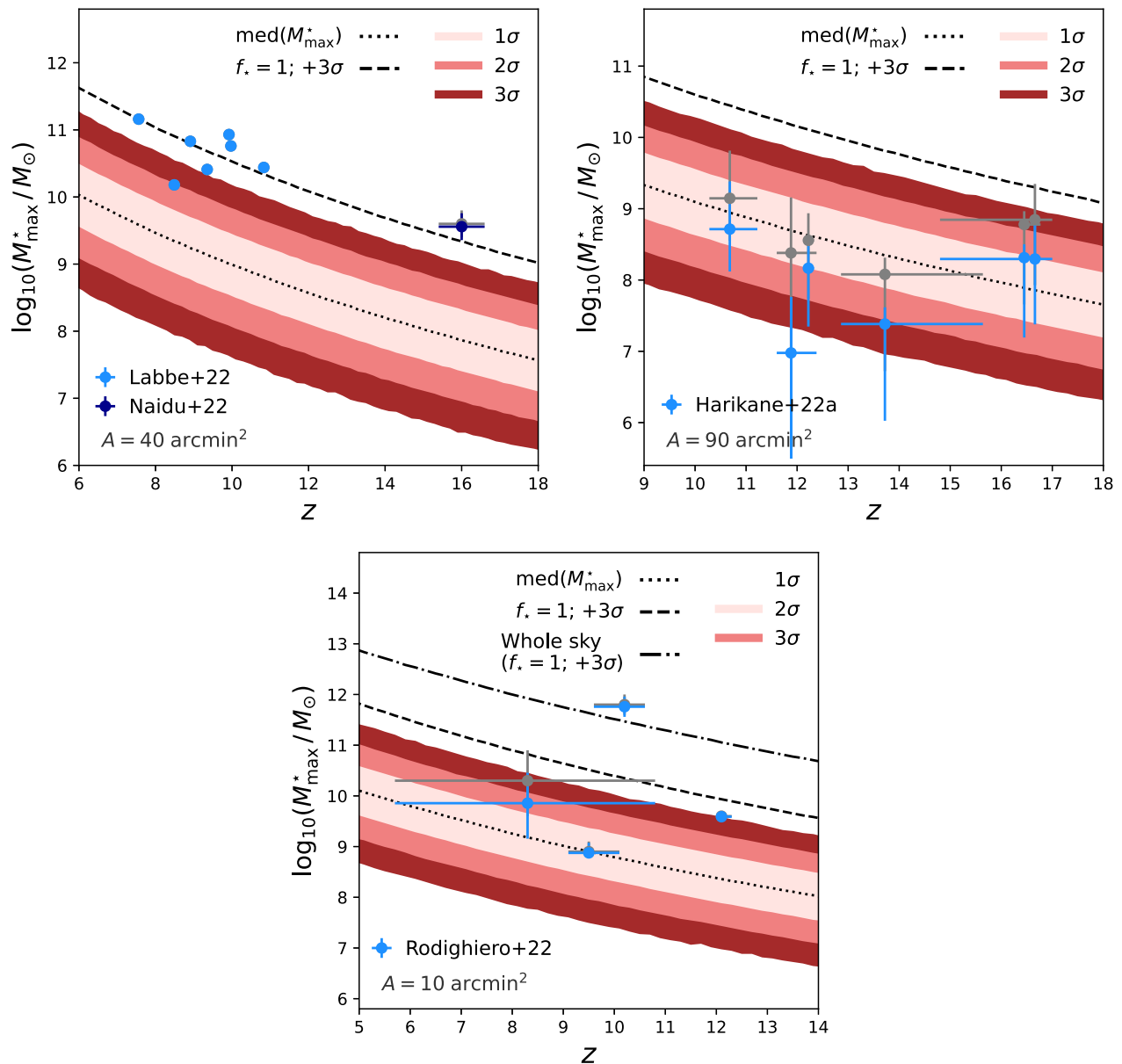
Figure 6. As for Fig. 5, but showing the latest high-z candidates from *JWST*. *Top left*: an observational survey volume with area 40 arcmin^2. Observational results from Labbe et al. (2022) at $z \sim 10$ are shown, as well as the Donnan et al. (2022) candidate, with stellar mass and redshift solutions at $z \sim 17$ derived by Naidu et al. (2022b) shown. *Top right*: an observational survey volume with area 90 arcmin^2, with stellar mass estimates from Harikane et al. (2022a). *Bottom*: stellar mass estimates from Rodighiero et al. (2022), assuming an observational survey with area 10 arcmin^2. The dashed–dotted black line shows the 3σ upper limit, assuming a stellar fraction of unity, for a *whole sky* survey. One of the Rodighiero et al. (2022) candidates exceeds even this most conservative upper limit.

deep fields; we show the stellar mass EVS PDF for the planned Wide and Deep Field North (20 deg^2) surveys.

Fig. 7 also shows the galaxy stellar mass EVS for a whole sky survey ($f_{sky} = 1$). Whole sky surveys at these high redshifts are inconceivable with current observational capabilities, however such a comparison avoids any possible *a posteriori* effects of region selection.

To enable the EVS approach to be applied to arbitrary surveys, we have made a python package, EVSTATS, available at github.com/christopherlovell/evstats, where interested users can find a simple to use Jupiter notebook detailing how to create your own confidence intervals in the stellar mass–redshift plane for a given survey area. We also provide output files in ECSV format for the future surveys presented in Fig. 7 at github.com/christopherlovell/evstats/tree/main/example/data.

4 DISCUSSION AND CONCLUSIONS

We have used EVS to predict the stellar mass of the most massive galaxy in a flat ΛCDM universe at high redshift ($z > 5$). Our results are as follows:

(i) Assuming some form for the halo mass function, we calculate the EVS PDF for the most massive halo on a fixed-redshift hypersurface. The most massive halo in the EAGLE and FLARES simulations is within the predicted 3σ confidence intervals.

(ii) We model the stellar fraction as a log-normal distribution, and combine with a fixed baryon fraction to translate our halo mass EVS PDF into one for stellar mass. The most massive galaxy in the EAGLE and FLARES simulations is within the predicted 2σ confidence intervals.
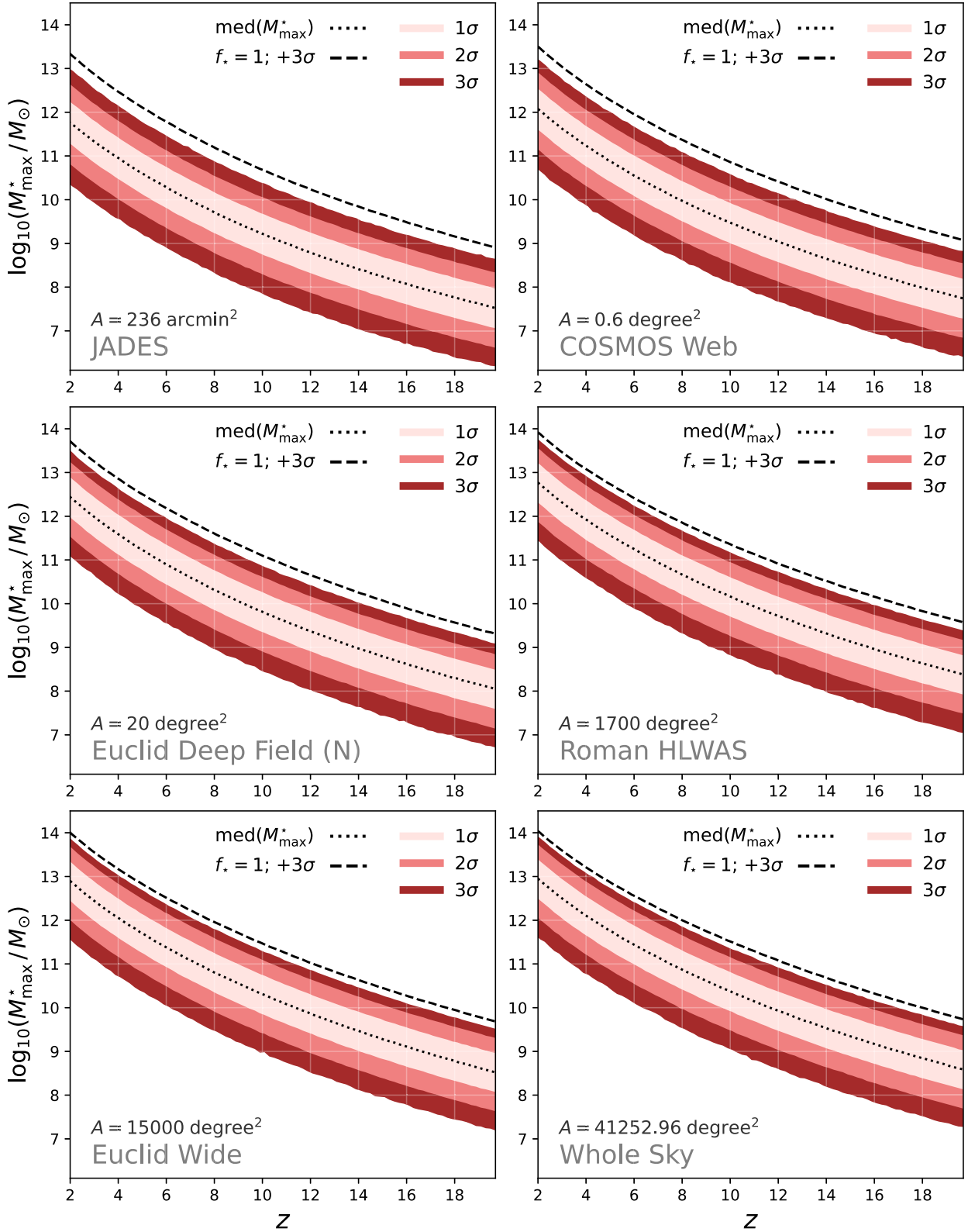
Figure 7. Predicted EVS PDF of the galaxy stellar mass distribution for a number of upcoming surveys. Clockwise from top left: the combined JADES medium and deep surveys (*JWST*), the COSMOS Web survey (*JWST*), Roman HLWAS, the theoretical prediction for a whole sky survey, Euclid Wide Survey, Euclid Deep Field North. The dashed line shows the 3σ upper limit assuming a stellar fraction of unity.

(iii) We calculate the stellar mass EVS PDF for an observational survey volume, and compare to recent pre-*JWST* observations of galaxies at $z > 8$. We find tension with predicted stellar masses for two objects from Tacchella et al.

(2022b), though no tension with results from Harikane et al. (2022b), mostly due to the significant uncertainties in the stellar mass estimates, which translate into a large Eddington bias correction.

(iv) We also compare to recent high redshift candidates from the first *JWST* data, and find significant tension with certain stellar mass estimates of a $z \sim 17$ candidate from Donnan et al. (2022), Harikane et al. (2022a), and Naidu et al. (2022b), as well as $z \sim 10$ candidates presented in Labbe et al. (2022) and Rodighiero et al. (2022).

(v) Finally, we present the stellar mass EVS PDF for a number of upcoming surveys from *JWST*, *Euclid*, and Roman between $2 < z < 20$.

The use of extreme value statistics is a powerful means of understanding the likelihood of the most massive objects in the Universe, complementing existing approaches (Steinhardt et al. 2016; Behroozi & Silk 2018; Boylan-Kolchin 2022). Already a number of objects detected in recent years (pre-*JWST*) are in tension with the predicted distributions (Tacchella et al. 2022b), assuming even conservative limits on the conversion of baryons into stars, and a number of the first candidates from early *JWST* data are also in significant tension (Donnan et al. 2022; Labbe et al. 2022; Naidu et al. 2022b; Rodighiero et al. 2022). We stress, however, that it is entirely plausible that these objects are not in tension with ΛCDM, and that instead there are unaccounted for uncertainties in their redshift or stellar mass estimates.

Redshift estimates of high-redshift sources are often multimodal, leaving the possibility that many high redshift candidates are, in fact, low-redshift interlopers (see Zavala et al. 2022) and the understanding of the necessary *JWST* instrument calibration is evolving (Adams et al. 2022). With regards to stellar mass, estimates are sensitive to a number of modelling assumptions during the SED fitting process, such as the assumed initial mass function (IMF) and stellar population synthesis (SPS) model. A clear example of this is the $z \sim 17$ object analysed by Donnan et al. (2022), Harikane et al. (2022a), and Naidu et al. (2022b); the estimates from these different studies, using a variety of different modelling assumptions, cover almost 1 dex in stellar mass.

At the highest redshifts, it is possible that Population III star formation may contribute up to 3–4 times the number of UV photons (Harikane et al. 2022a), boosting nebular emission in the rest-frame optical, which can bias stellar mass and SFR estimates. These first stars are also expected to have a significantly top heavy IMF, which can complicate their interpretation using standard SPS models. The assumed prior on the star formation history can also have a large effect on derived stellar mass estimates (Whitler et al. 2022; Tacchella et al. 2022a; Tacchella et al. 2022b). Models suggest high redshift galaxies have rising star formation histories (Finlator, Oppenheimer & Davé 2011; Wilkins et al. 2022); using incompatible functional forms can lead to significant biases (Carnall et al. 2019). Steinhardt et al. (2022) recently highlighted the impact of using templates calibrated or derived from lower redshift conditions, leading to offsets in stellar mass estimates of high-redshift sources of up to 1–1.6 dex. Mason, Trenti & Treu (2022) estimate the maximal UV luminosity assuming all gas in a halo is converted into stars over a time-scale that maximizes the UV emission (~ 10 Myr), and found that the upper limit derived is higher than that measured in recent HST & JWST results (Bouwens et al. 2021; Donnan et al. 2022). AGN contamination can also bias both stellar mass and redshift estimates (see Inayoshi et al. 2022), and may be a particularly pertinent contaminant in the analysis presented here; it is in the most massive haloes that the most massive central black holes are expected to reside. However, the EVS formalism presented here does allow us to place wide priors on these processes, producing self-consistent PDFs that take into account many of these uncertainties. As our understanding of the physics of galaxy formation at high redshift

improves, these priors can be narrowed, allowing for more precise limits on the maximum halo and stellar mass at a given redshift to be made.

We do not take account of the effect on our predictions of surveys taken from multiple areas of the sky. Behroozi & Silk (2018) argue that such a survey approach increases the chance that a single survey will contain an outlier. However, we note that the effect of observational errors leading to Eddington and Malmquist bias has a much larger effect on the predicted probabilities. Including this cosmic variance effect within the EVS framework is left for future work.

With upcoming wide field surveys a number of galaxies will be detected that may potentially be in tension with predictions from ΛCDM, or require extreme conversion rates of baryons into stars. We hope the predictions presented here, and the publicly accessible code (github.com/christopherlovell/evstats), will present a means of producing confidence intervals for any given survey, and allow observers to quickly evaluate the probability that a given source is in tension with a given cosmology.

DATA AVAILABILITY STATEMENT

All of the code and data used in the analysis is available at github.com/christopherlovell/evstats. Details on where to obtain the stellar and halo mass values from the EAGLE and FLARES simulations are provided in McAlpine et al. (2016) and Lovell et al. (2021), respectively.

REFERENCES

Adams N. J. et al., 2022, preprint (arXiv:2207.11217)
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Behroozi P., Silk J., 2018, MNRAS, 477, 5382
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bouwens R. J. et al., 2021, AJ, 162, 47
Boylan-Kolchin M., 2022, preprint (arXiv:2208.01611)

[4]https://credit.niso.org/

Carnall A. C., Leja J., Johnson B. D., McLure R. J., Dunlop J. S., Conroy C., 2019, ApJ, 873, 44

Chongchitnan S., 2015, J. Cosmol. Astropart. Phys., 2015, 062

Chongchitnan S., Silk J., 2012, Phys. Rev. D, 85, 063508

Crain R. A. et al., 2015, MNRAS, 450, 1937

Davis O., Devriendt J., Colombi S., Silk J., Pichon C., 2011, MNRAS, 413, 2087

DeRose J. et al., 2019, ApJ, 875, 69

Donnan C. T. et al., 2022, preprint (arXiv:2207.12356)

Eddington A. S., 1913, MNRAS, 73, 359

Euclid Collaboration XVII, 2022, A&A, 658, A126

Finkelstein S. L. et al., 2021, preprint (arXiv:2106.13813)

Finkelstein S. L. et al., 2022, preprint (arXiv:2207.12474)

Finkelstein S. L., Dunlop J., Fevre O. L., Wilkins S., 2015, preprint (arXiv:1512.04530)

Finlator K., Oppenheimer B. D., Davé R., 2011, MNRAS, 410, 1703

Gelb J. M., Bertschinger E., 1994, ApJ, 436, 491

Giodini S. et al., 2009, ApJ, 703, 982

Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, ApJ, 778, 14

Gumbel E. J., 1958, Statistics of Extremes. Columbia University Press, New York, US

Harikane Y. et al., 2016, ApJ, 821, 123

Harikane Y. et al., 2018, PASJ, 70, S11

Harikane Y. et al., 2022a, preprint (arXiv:2208.01612)

Harikane Y. et al., 2022b, ApJ, 929, 1

Harrison I., Coles P., 2011, MNRAS, 418, L20

Harrison I., Coles P., 2012, MNRAS, 421, L19

Harris C. R. et al., 2020, Nature, 585, 357

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

Inayoshi K., Harikane Y., Inoue A. K., Li W., Ho L. C., 2022, preprint (arXiv:2208.06872)

Jimenez R., Verde L., 2009, Phys. Rev. D, 80, 127302

Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021, ApJS, 254, 22

Kartaltepe J. et al., 2021, JWST Proposal. Cycle 1. p. 1727

Katz S., Nadarajah S., 2000, Extreme Value Distributions: Theory and Applications. Imperial College Press, London, UK

Labbe I. et al., 2022, preprint (arXiv:2207.12446)

Leja J., Johnson B. D., Conroy C., van Dokkum P. G., Byler N., 2017, ApJ, 837, 170

Lovell C. C., Vijayan A. P., Thomas P. A., Wilkins S. M., Barnes D. J., Irodotou D., Roper W., 2021, MNRAS, 500, 2127

Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160

Mason C. A., Trenti M., Treu T., 2022, Technical Report, The Brightest Galaxies at Cosmic Dawn, https://ui.adsabs.harvard.edu/abs/2022arXiv2 20714808M

Matarrese S., Verde L., Jimenez R., 2000, ApJ, 541, 10

McAlpine S. et al., 2016, A&C, 15, 72

Menci N., Castellano M., Santini P., Merlin E., Fontana A., Shankar F., 2022, Technical Report, High-Redshift Galaxies from Early JWST Observations: Constraints on Dark Energy Models, https://ui.adsabs.harv ard.edu/abs/2022arXiv220811471M

Naidu R. P. et al., 2022a, preprint (arXiv:2207.09434)

Naidu R. P. et al., 2022b, preprint (arXiv:2208.02794)

Planck Collaboration XIII, 2016, A&A, 594, A13

Power C., Knebe A., 2006, MNRAS, 370, 691

Reed D., Bower R., Frenk C., Jenkins A., Theuns T., 2007, MNRAS, 374, 2

Rieke M. et al., 2019, Bull. Am. Astron. Soc., 51, 45

Rodighiero G., Bisigello L., Iani E., Marasco A., Grazian A., Sinigaglia F., Cassata P., Gruppioni C., 2022, Technical Report, JWST Unveils Heavily Obscured (Active and Passive) Sources up to z ∼ 13, https://ui.adsabs.ha rvard.edu/abs/2022arXiv220802825R

Sahlén M., Zubeldía Í., Silk J., 2016, ApJ, 820, L7

Schaye J. et al., 2015, MNRAS, 446, 521

Somerville R. S., Davé R., 2015, ARAA, 53, 51

Song M. et al., 2016, ApJ, 825, 5

Stefanon M., Bouwens R. J., Labbé I., Illingworth G. D., Gonzalez V., Oesch P. A., 2021, ApJ, 922, 1

Steinhardt C. L., Capak P., Masters D., Speagle J. S., 2016, ApJ, 824, 21

Steinhardt C. L., Kokorev V., Rusakov V., Garcia E., Sneppen A., 2022, preprint (arXiv:2208.07879)

Susa H., Umemura M., 2004, ApJ, 610, L5

Tacchella S. et al., 2022a, preprint (arXiv:2208.03281)

Tacchella S. et al., 2022b, ApJ, 927, 170

Tacchella S., Trenti M., Carollo C. M., 2013, ApJ, 768, L37

Tacchella S., Bose S., Conroy C., Eisenstein D. J., Johnson B. D., 2018, ApJ, 868, 92

Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M. S., Yepes G., Gottlober S., Holz D. E., 2008, ApJ, 688, 709

Trenti M., Stiavelli M., 2008, ApJ, 676, 767

Vijayan A. P., Lovell C. C., Wilkins S. M., Thomas P. A., Barnes D. J., Irodotou D., Kuusisto J., Roper W. J., 2021, MNRAS, 501, 3289

Virtanen P. et al., 2020, Nat. Methods, 17, 261

Waizmann J.-C., Ettori S., Moscardini L., 2012, MNRAS, 420, 1754

Wang Y. et al., 2022, ApJ, 928, 1

Watson W. A., Iliev I. T., Diego J. M., Gottlöber S., Knebe A., Martínez-González E., Yepes G., 2014, MNRAS, 437, 3776

Whitler L., Stark D. P., Endsley R., Leja J., Charlot S., Chevallard J., 2022, preprint (arXiv:2206.05315)

Wilkins S. M. et al., 2022, preprint (arXiv:2208.00976)

Zavala J. A. et al., 2022, preprint (arXiv:2208.01816)

APPENDIX A: REDSHIFT BINNING

When drawing the confidence intervals plotted in *e.g.* Fig. 2, we sample multiple PDFs for each redshift slice. This is equivalent to carrying out N Bernoulli trials. There is therefore a non-zero probability of x events being above some threshold, which we can work out by calculating the binomial probability,

$$P_x = \binom{N}{x} p^x q^{N-x}, \tag{A1}$$

where p is the chosen probability threshold. We have tested this for the two objects from the Tacchella et al. (2022b) sample shown in the top panel of Fig. 5, and find that, for our fiducial bin spacing, the probability of exceeding the 3σ contour threshold is $P_x < 3\sigma$. These objects are therefore statistically unlikely to exceed this limit due to the number of trials, and are therefore still significant outliers.

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