DRAFT VERSION JUNE 13, 2018
Typeset using LaTeX **twocolumn** style in AASTeX61

THE JWST EXTRAGALACTIC MOCK CATALOG: MODELING GALAXY POPULATIONS FROM THE UV THROUGH THE NEAR-IR OVER THIRTEEN BILLION YEARS OF COSMIC HISTORY

CHRISTINA C. WILLIAMS,[1,*] EMMA CURTIS-LAKE,[2] KEVIN N. HAINLINE,[1] JACOPO CHEVALLARD,[2] BRANT E. ROBERTSON,[3] STEPHANE CHARLOT,[2] RYAN ENDSLEY,[1] DANIEL P. STARK,[1] CHRISTOPHER N. A. WILLMER,[1] STACEY ALBERTS,[1] RICARDO AMORIN,[4,5] SANTIAGO ARRIBAS,[6] STEFI BAUM,[7] ANDREW BUNKER,[8] STEFANO CARNIANI,[4,5] SARA CRANDALL,[3] EIICHI EGAMI,[1] DANIEL J. EISENSTEIN,[9] PIERRE FERRUIT,[10] BERND HUSEMANN,[11] MICHAEL V. MASEDA,[12] ROBERTO MAIOLINO,[4,5] TIMOTHY D. RAWLE,[13] MARCIA RIEKE,[1] RENSKE SMIT,[4,5] SANDRO TACCHELLA,[9] AND CHRIS J. WILLOTT[14]

[1]*Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA*
[2]*Sorbonne Universités, UPMC-CNRS, UMR7095, Institut d'Astrophysique de Paris, F-75014, Paris, France*
[3]*Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA*
[4]*Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK*
[5]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[6]*Departamento de Astrofisica, Centro de Astrobiologia, CSIC-INTA, Cra. de Ajalvir, 28850-Madrid, Spain*
[7]*University of Manitoba, Dept. of Physics and Astronomy, Winnipeg, MB R3T 2N2, Canada*
[8]*Department of Physics, University of Oxford, Oxford, UK*
[9]*Harvard-Smithsonian Center for Astrophysics 60 Garden St., Cambridge, MA 02138*
[10]*Scientific Support Office, Directorate of Science, ESA/ESTEC, Keplerlaan 1, 2201AZ Noordwijk, The Netherlands*
[11]*Max Planck Institute for Astronomy, Konigstuhl 17, D-69117 Heidelberg, Germany*
[12]*Leiden Observatory, Leiden University, PO Box 9513, 2300 RA, Leiden, The Netherlands*
[13]*European Space Agency, c/o STScI, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[14]*NRC Herzberg, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada*

ABSTRACT

We present an original phenomenological model to describe the evolution of galaxy number counts, morphologies, and spectral energy distributions across a wide range of redshifts ($0.2 < z < 15$) and stellar masses [$\log(M/M_\odot) \geq 6$]. Our model follows observed mass and luminosity functions of both star-forming and quiescent galaxies, and reproduces the redshift evolution of colors, sizes, star-formation and chemical properties of the observed galaxy population. Unlike other existing approaches, our model includes a self-consistent treatment of stellar and photoionized gas emission and dust attenuation based on the BEAGLE tool. The mock galaxy catalogs generated with our new model can be used to simulate and optimize extragalactic surveys with future facilities such as the James Webb Space Telescope (*JWST*), and to enable critical assessments of analysis procedures, interpretation tools, and measurement systematics for both photometric and spectroscopic data. As a first application of this work, we make predictions for the upcoming JWST Advanced Deep Extragalactic Survey (JADES), a joint program of the *JWST*/NIRCam and NIRSpec Guaranteed Time Observations teams. We show that JADES will detect, with NIRCam imaging, thousands of galaxies at $z \gtrsim 6$, and tens at $z \gtrsim 10$ at $m_{AB} \lesssim 30$ (5σ) within the 236 arcmin^2 of the survey. The JADES data will enable accurate constraints on the evolution of the UV luminosity function at $z > 8$, and resolve the current debate about the rate of evolution of galaxies at $z \gtrsim 8$. Ready to use mock catalogs and software to generate new realizations are publicly available as the JAdes extraGalactic Ultradeep Artificial Realizations (JAGUAR) package.

Keywords: galaxies:evolution — galaxies:high-redshift — galaxies:photometry

ccwilliams@email.arizona.edu, curtis@iap.fr

* NSF Fellow

1. INTRODUCTION

Over the last two decades, deep extragalactic surveys with the *Hubble* (*HST*) and *Spitzer* Space Telescopes have revolutionized our understanding of galaxy evolution. These surveys measured the buildup of galaxy populations from the local Universe to the current redshift frontier at $z \sim 10$ (for a review, see, e.g. Stark 2016). Meanwhile, ground-based 8m- and 10m-class telescopes have characterized the physical conditions of galaxies even beyond $z \sim 2-3$, the peak in the cosmic star formation rate density (e.g. with Keck/MOSFIRE; Kriek et al. 2015, Steidel et al. 2014). Currently, further progress is hindered by the limited wavelength coverage of *HST*, relatively low sensitivity of *Spitzer*, and the atmospheric limitations that impede ground-based campaigns. However, the soon-to-launch *James Webb Space Telescope* (*JWST*; Gardner et al. 2006) will detect galaxies well beyond the current redshift frontier, below the magnitude and stellar mass limits currently achievable with existing facilities, while its high spatial resolution will image early galaxies in exquisite detail. Furthermore, the unprecedented spectroscopic capabilities of *JWST* will enable spectroscopic observations of even the faintest galaxies detected with *HST* to date (e.g. Chevallard et al. 2017)

This innovative telescope, hosting the largest mirror ever to fly in space and a suite of state-of-the-art near-infrared instruments, will provide unique data to answer key open questions about the formation and evolution of galaxies. Specifically, the wavelength coverage provides the opportunity, for the first time, to study the rest-frame optical properties of galaxies out to $z \sim 9$, and the rest-frame UV out to $z > 10$. Observations with *JWST* will enable precise constraints on the evolution of the stellar and chemical make up of galaxies, dust attenuation, and ionization sources across a broad range of redshift, stellar mass and luminosity (e.g. Mannucci et al. 2010; Reddy et al. 2015; Strom et al. 2017; Shapley et al. 2017). These data are fundamental for understanding the formation of the Hubble sequence, the emergence of quiescent galaxies, and the variety of observed scaling relations between galaxy properties (e.g. Faber & Jackson 1976; Tully & Fisher 1977; Kauffmann et al. 2003; Tremonti et al. 2004; Franx et al. 2008; Maiolino et al. 2008; Speagle et al. 2014; van der Wel et al. 2014; Glazebrook et al. 2017). In addition, *JWST* will be used to target the exact epoch and sources of cosmic reionization at high redshift (e.g. Bunker et al. 2004; Finkelstein et al. 2012a; Robertson et al. 2015; Stark 2016). Studies that address these topics will require large survey campaigns using multiple instruments on board *JWST* including, the Near Infrared Camera (NIRCam; Horner & Rieke 2004) and the Near Infrared Spectrograph, (NIRSpec; Bagnasco et al. 2007; Birkmann et al. 2016). These sensitive instruments will provide new space-based observation modes including parallel imaging and spectroscopic observations, simultaneous imaging enabled by the dichroic on NIRCam, as well as the choice of fixed slit, high-multiplex or integral field spectroscopy on NIRSpec.

Maximizing the scientific return of the innovative and complex instruments on board *JWST* will require the development of original analysis tools and space-based observing strategies. As an example, the advent of space-based multi-object spectroscopy (with the NIRSpec Micro-Shutter Array; MSA) initiates an era where spectroscopic follow up of *JWST*-selected targets will demand the rapid analysis of imaging data to create slit-mask designs. Meeting these future challenges requires physically-motivated simulations of *JWST* data that should ideally match existing observations, while also extending to the unprecedented depths and redshifts that will be attained by *JWST*. Such simulations enable critical tests of analysis procedures and processing tools, and aid the scientific interpretation by identifying potential observational biases on measured galaxy properties (e.g. galaxy sizes or UV continuum slope β; Dunlop et al. 2012; Finkelstein et al. 2012b; Rogers et al. 2013; Curtis-Lake et al. 2016; Bouwens et al. 2017a).

Physically-motivated *JWST* simulations will require mock galaxy catalogs, which can be built using semi-analytic galaxy formation models (e.g. Blaizot et al. 2005; Cai et al. 2009; Bernyk et al. 2016; Mirocha et al. 2017; Furlanetto et al. 2017) or hydrodynamical simulations (e.g. Torrey et al. 2015; McAlpine et al. 2016). However, such sophisticated approaches (e.g. Croton et al. 2006; Benson 2012; Vogelsberger et al. 2014; Schaye et al. 2015) are intrinsically model-dependent. As an example, semi-analytical models that match low-to-intermediate redshift stellar mass functions may provide widely different predictions for low-mass galaxies [$\log(M/M_\odot) \lesssim 8$] and at high redshifts ($z \gtrsim 4$, e.g. Lu et al. 2014), or underpredict the specific star-formation rates (sSFR) of sub-L^* galaxies (e.g. Somerville & Davé 2015; Fontanot et al. 2009; Weinmann et al. 2012; Somerville et al. 2015). In an effort to reduce the model-dependency of mock observation tools, empirically driven approaches have been developed based on observed galaxy distributions and relations among physical quantities that replicate deep extragalactic surveys as observed from current facilities (e.g. Schreiber et al. 2017).

As we look forward to future facilities that extend beyond current limitations, we must incorporate accurate descriptions of the spectral energy distributions (SEDs) of young, low-mass and high sSFR galaxies across cosmic time. These populations are of particular importance both as low-redshift interlopers, as well as the high-redshift galaxies which are the prime science targets for *JWST*, and are now known to produce strong nebular emission lines that can contribute significant excesses to broad-band photometric fluxes (Schaerer & de Barros 2009; Shim et al. 2011; Atek et al. 2011; Labbé

et al. 2013; Stark et al. 2013; Schenker et al. 2013; Smit et al. 2014, 2015; Roberts-Borsani et al. 2016; Rasappu et al. 2016). Thus the treatment of nebular emission in mock catalogs tailored to reproducing high-redshift galaxies is especially important. Currently, the treatment of nebular emission in mock catalogs based on galaxy formation models is often approximated in post-processing with subgrid prescriptions (e.g. Somerville & Davé 2015; Naab & Ostriker 2017), although more advanced ones have been recently proposed based on simplified prescriptions for the dependence of line emission on metallicity, ISM conditions or ionization parameter (e.g. Kewley et al. 2013; Orsi et al. 2014; Shimizu et al. 2016). A fully self-consistent treatment of stellar and nebular emission in hydrodynamical simulations is, however, still limited to small numbers of objects rather than full cosmological simulations (Hirschmann et al. 2017).

With this work, we present a new phenomenological model for the cosmic galaxy population designed to benefit future surveys with *JWST* and other forthcoming facilities targeting the UV to near-infrared emission of galaxies. Our model is designed to reproduce observations of galaxy properties from $0 < z < 10$, and enables extrapolations of galaxy distributions to $z \sim 15$, allowing for the generation of mock catalogs that include physically-motivated counts, luminosities, stellar masses, morphologies, photometry and spectroscopic properties down to arbitrarily low stellar mass. Importantly, we incorporate a self-consistent modeling of stellar and nebular emission using the models of Gutkin et al. (2016) teamed with the BEAGLE tool (Chevallard & Charlot 2016), which enables the inclusion of strong nebular emission lines and nebular continuum emission in mock galaxy spectra and photometric SED. These models cover the wide parameter space required to model the range of physical conditions expected in local and extremely high redshift galaxies ($z > 10$) without resorting to simple prescriptions of emission line ratios.

Simulations using our model have already proven invaluable to optimize the design of a large ($\sim$ 720 hours) observational program, the *JWST* Advanced Deep Extragalactic Survey (JADES), a joint program of the NIRCam and NIRSpec Guaranteed Time Observations (GTO) teams. In particular, mock catalogs produced using our model have been used to optimize the selection of photometric filters and spectral dispersers, the depth of the observations and area covered. This mock catalog tool, called JAdes extraGalactic Ultradeep Artificial Realizations (JAGUAR), and related *JWST* simulations will also provide a fundamental aid for the scientific interpretation of future *JWST* data, and has enabled us to make realistic science predictions for the future GTO survey.

The outline of this paper is as follows. In Section 2, we provide a conceptual overview of our procedure for producing mock galaxies and assigning their properties. In the subsequent sections, we describe the phenomenological model that underlies JAGUAR quantitatively. In Sections 3 and 4, we describe the procedure for producing star-forming and quiescent galaxies (respectively) across cosmic time, including their masses, redshifts, luminosities and SED properties. In Section 5 we describe the procedure for assigning morphological parameters to both star-forming and quiescent galaxies. In Section 6, we characterize a realization of our model (a JAGUAR mock catalog) by presenting comparisons to measurements made from current surveys between $0 < z < 10$. In Section 7, we present our predictions for the science results of JADES that are enabled by this tool. Finally, in Section 8 we summarize this work. We release ready-to-use realizations [1] as described below, as well as a PYTHON package for JAGUAR that can be used to generate catalogs to any area or depth. Throughout this work we assume a ΛCDM cosmology with H_0=70 km s^{-1} Mpc^{-1}, Ω_M = 0.3, Ω_Λ = 0.7. When necessary, we assume a Chabrier (2003) stellar initial mass function (IMF).

2. METHODS OVERVIEW

The foundation of our model consists of observed stellar mass and UV luminosity functions that have been measured from $0 < z < 10$. We use these observations to model the evolution of stellar mass functions for both star-forming and quiescent galaxies, which are then used to generate each mock population at all redshifts.[2] We assign integrated properties such as the UV absolute magnitude $M_{\rm UV}$ and UV continuum slope β (where $f_\lambda \propto \lambda^\beta$; for star-forming galaxies only), and structural properties based entirely on empirical relations or distributions. Finally, the model assigns spectra that are consistent with these integrated properties to each mock galaxy, which we use to produce the broadband photometry. Summaries of our overall procedure for star-forming galaxies are shown in Figures 1 and 2, which indicate the sections that describe the relevant quantitative procedures for assigning various properties to mock galaxies.

2.1. *Generating galaxy counts*

Here we describe the procedure we follow to generate galaxy number counts (i.e. the expected number of galaxies of a given mass, at fixed redshift and on-sky area). We first model the evolution of stellar mass functions across cosmic time using continuously-evolving Schechter functions for both star-forming and quiescent galaxies. We then generate the expected number of star-forming or quiescent galaxies for a given redshift bin over a given survey area by integrating their respective model mass function, multiplying by the co-moving volume and drawing from a Poisson distribution

[1] http://fenrir.as.arizona.edu/jaguar

[2] We note that we do not attempt to include galaxies composed of metal-free, 'PopIII' stars, since no empirical constraints exist on such objects.

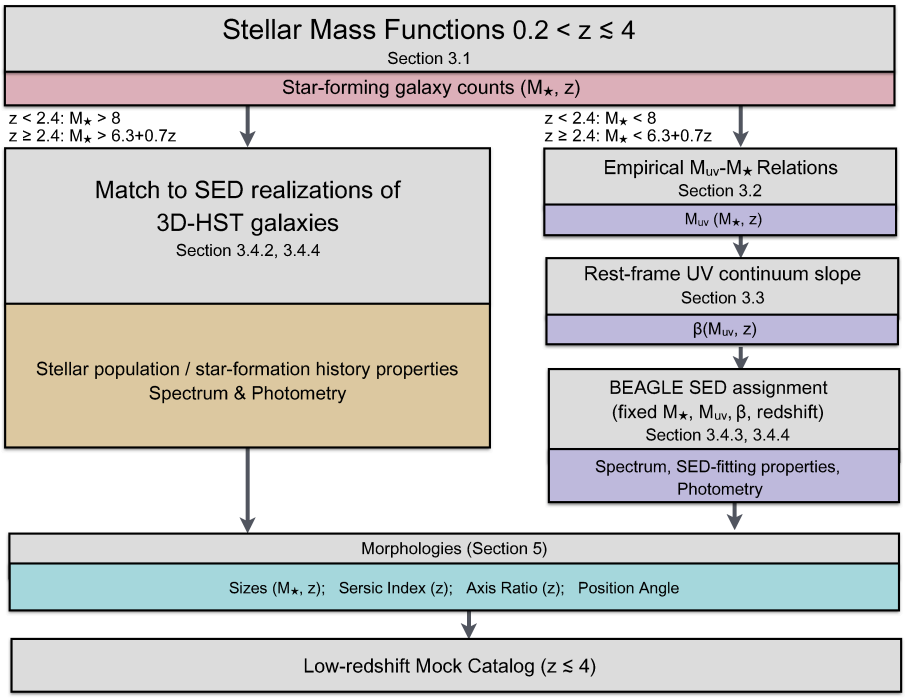

Figure 1. Diagram summarizing the procedures for generating star-forming galaxies at $z \lesssim 4$. $M_\star$ is defined as $\log(M/M_\odot)$. High-mass galaxies (defined as $M_\star > 8$ for $z < 2.4$ and $M_\star > 6.3+0.7z$ for $z \geq 2.4$; illustrated by the left pathway) and low-mass galaxies (right pathway) are generated differently as indicated. These criteria are defined in Section 3.4.4 as the approximate mass completeness limits in the 3D-HST catalog, which we use to assign real galaxy SEDs to high-mass mock galaxies. Gray boxes indicate the empirical relationships, distributions, or data on which mock galaxy properties are based, and colored boxes indicate the mock galaxy property which is generated in that step. Quiescent galaxies are generated at $z < 4$ following a different procedure which is described in Section 4 and illustrated in Figure 13.

with this mean. By computing the cumulative distribution function (CDF) of the stellar mass function we can then effectively draw this number of galaxies from the mass function using inverse transform sampling.

For star-forming galaxies at $z \gtrsim 4$, stellar mass function measurements become increasingly difficult and uncertain. With current facilities, this epoch represents a transition to rest-frame UV selections tracing young stars (with *HST*) instead of rest-frame optical selections that trace stellar mass (which would require *Spitzer*/IRAC whose sensitivity is lower). At $z \gtrsim 4$ the UV luminosity function becomes much more easily measurable than the stellar mass function with current facilities. Therefore, at redshifts $z \gtrsim 4$ we rely on observed UV luminosity functions to constrain the number densities of star forming galaxies. (Quiescent galaxies at $z > 4$ are instead based on an informed extrapolation, which is discussed in Section 4).

While generating galaxy counts from a stellar mass function is straightforward, using a UV luminosity function requires modeling a theoretical or an empirical relation linking a galaxy's UV luminosity, or $M_{\rm UV}$, to its stellar mass, hereafter $M_\star = \log(M/M_\odot)$. The observed connection between UV luminosity and stellar mass is not a simple monotonic relationship; galaxies exhibit a range of UV luminosities at fixed stellar mass that likely depends on other galaxy properties including stellar population age and metallicity, dust, and gas content. We will describe this distribution in terms of a Gaussian scatter about an average $M_{\rm UV}$–$M_\star$ relation, where the standard deviation of $M_{\rm UV}$ at fixed $M_\star$ is given by σ_{uv} (assumed independent of $M_\star$). We can then express the probability of a galaxy of stellar mass $M_\star$ to have a given $M_{\rm UV}$ as

$$\frac{dP}{dM_{\rm UV}}(M_\star, z) = \mathcal{N}[M_{\rm UV}, \overline{M}_{\rm UV}(M_\star, z), \sigma_{uv}] \quad (1)$$

where the mean relationship between UV luminosity at a given stellar mass and redshift is

$$\bar{M}_{UV}(M_\star, z) \equiv \int M_{\rm UV} \frac{dP}{dM_{\rm UV}}(M_\star, z)\, dM_{\rm UV}\,. \quad (2)$$

Once such a relation for $M_{\rm UV}$–$M_\star$ and its scatter has been adopted (see Section 3.2), the observed UV luminosity function $\Phi(M_{\rm UV}, z)$ can be modeled as the convolution of the stellar mass function $\Phi(M_\star, z)$ with the distribution of $M_{\rm UV}$

$$\Phi(M_{\rm UV}, z) = \int_0^\infty \Phi(M_\star, z)\, \mathcal{N}[M_{\rm UV}, \overline{M}_{\rm UV}(M_\star, z), \sigma_{uv}]\, dM_\star\,, \quad (3)$$

where $\Phi(M_{\rm UV}, z)$ represents the number of galaxies per comoving volume with UV absolute magnitude $M_{\rm UV}$ as a func-

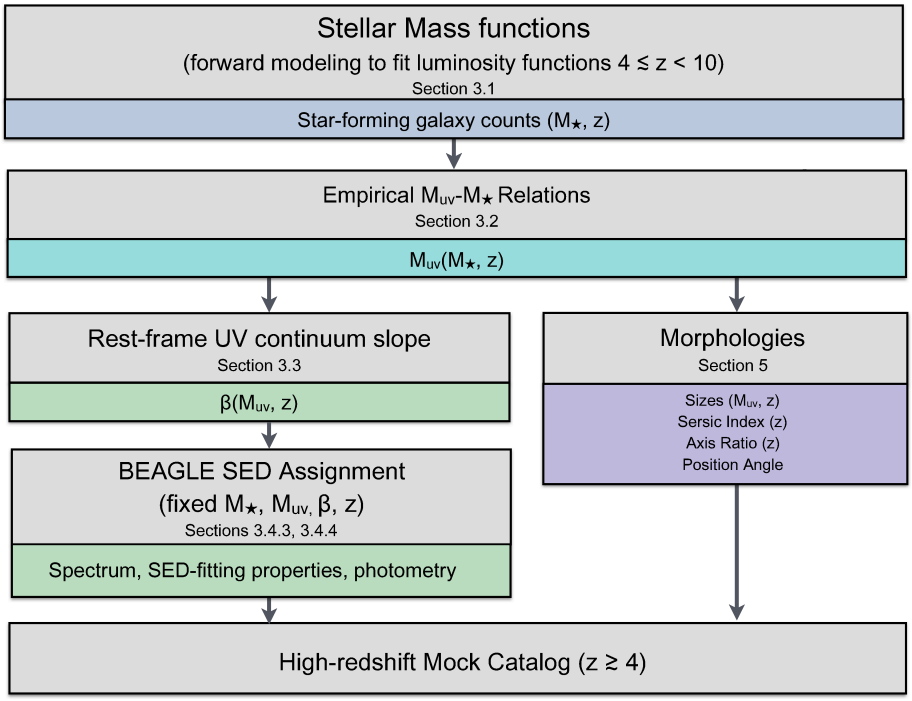

Figure 2. Diagram summarizing the procedures for generating the star-forming galaxies at $z \gtrsim 4$. $M_\star$ is defined as $\log(M/M_\odot)$. Gray boxes indicate the empirical relationships, distributions, or data on which mock galaxy properties are based, and colored boxes indicate the mock galaxy property which is generated in that step. All star-forming mock galaxies at $z > 4$ are generated following these procedures. Quiescent galaxies are generated at $z > 4$ following a different procedure which is described in Section 4 and illustrated in Figure 13.

tion of redshift. Therefore, to calculate star-forming galaxy counts at $z \gtrsim 4$ where we have the best constraints from the UV luminosity function, we forward model the continuously evolving stellar mass function, convolved with an empirical characterization of $\mathcal{N}[M_{\rm UV}, \overline{M}_{\rm UV}(M_\star, z), \sigma_{uv}]$ in order to fit with observed UV luminosity functions over the range $4 \lesssim z \lesssim 10$. [3]

This procedure enables us to produce one continuously evolving stellar mass function that, when sampled randomly as outlined above, produces star-forming galaxy counts that follow observed stellar mass functions at $z \lesssim 4$, the $M_{\rm UV}$–$M_\star$ distribution given by $\mathcal{N}[M_{\rm UV}, \overline{M}_{\rm UV}(M_\star, z), \sigma_{uv}]$, and the observed UV luminosity functions at $z \gtrsim 4$. We will describe the characterization of the empirical $M_{\rm UV}$–$M_\star$ distribution, $\mathcal{N}[M_{\rm UV}, \overline{M}_{\rm UV}(M_\star, z), \sigma_{uv}]$, that we use to forward model the stellar mass function in Section 3.2. In Section 3.1 we will describe our procedure to fit the observed evolving stellar mass function over $0.2 < z \lesssim 4$, and forward model by convolving the mass function with $\mathcal{N}[M_{\rm UV}, \overline{M}_{\rm UV}(M_\star, z), \sigma_{uv}]$ at $z \gtrsim 4$, to produce galaxy counts.

[3] Uncertainties in stellar masses complicate measurements of the stellar mass function, because the intrinsic stellar mass function must be convolved with the uncertainties in the stellar mass estimates. However, for this work we model the case where the intrinsic stellar masses are known perfectly.

2.2. *Generating integrated galaxy properties*

For each object generated in the mock galaxy population, we use redshift and stellar mass to assign other integrated galaxy properties including UV absolute magnitude $M_{\rm UV}$ and continuum slope β for star-forming galaxies, as well as type-dependent structural parameters. To assign the integrated properties we use empirical relations, plus appropriate scatter, to generate smoothly redshift-evolving distributions of $M_{\rm UV}$–$M_\star$ (Section 3.2), β–$M_{\rm UV}$ (Section 3.3), size-mass ($z < 4$) and size-UV luminosity (at $z > 4$; see Section 5). Figures 1 and 2 provide more details on this procedure. The integrated properties inform the assignment of a fully consistent SED to the mock galaxies, from which we derive *JWST*, *HST* and *Spitzer* filter photometry. These SEDs are created using BEAGLE and span a range of physical properties as described in the following section.

2.3. *Modeling galaxy SEDs with the* BEAGLE *tool*

BEAGLE (Chevallard & Charlot 2016; C16) is a new-generation tool for the modeling and interpretation of spectro-photometric galaxy SEDs based on a self-consistent approach to describe stellar emission and its transfer through the interstellar (ISM) and intergalactic (IGM) media. In this Section, we first describe the general characteristics of BEAGLE and the models integrated therein, and then summarize

our two methods for assigning SEDs to mock galaxies according to whether or not the realized properties overlap with those of observed galaxies from current surveys.

In BEAGLE, the emission from simple stellar populations of different ages, t' and metallicities, Z (the mass fraction of all elements heavier than Helium), is described by the latest version of the Bruzual & Charlot (2003) population synthesis code. Stellar emission is computed using the MILES stellar library (Sánchez-Blázquez et al. 2006) and includes new prescriptions for the evolution of massive stars (Bressan et al. 2012; Chen et al. 2015) and their spectra (Hamann & Gräfener 2004; Leitherer et al. 2010). We account for the (continuum+line) emission of gas photoionized by young stars by considering the large grid of photoionization models of Gutkin et al. (2016). These are based on the standard photoionization code CLOUDY (version 13.3; Ferland et al. 2013) and assume 'ionization bounded' nebulae, i.e. a zero escape fraction of H-ionizing photons. The models are described in terms of 'effective', i.e. galaxy-wide parameters following the prescription of Charlot & Longhetti (2001). Adjustable model parameters include the ionization parameter $\log U_S$, which sets the ratio of H-ionizing photons to H atoms at the edge of the Strömgren sphere, the interstellar metallicity Z_{ISM}, and the dust-to-metal (mass) ratio ξ_d, which traces metal depletion onto dust grains. Since the gas density n_H and depletion factor ξ_d do not significantly affect emission line ratios at sub-solar metallicities (see figure 3 and 5 of Gutkin et al. 2016), and most of our galaxies exhibit $\log(Z/Z_\odot) \lesssim -0.5$ (see Fig 12), we fix $n_H = 10^2\,\mathrm{cm}^{-3}$, the typical value measured in $z \sim 2-3$ galaxies (e.g. Sanders et al. 2016; Strom et al. 2017), and $\xi_d = 0.3$, a value similar to what measured in the Solar neighborhood (although, see Section 6.5). We account for attenuation by dust of the emission from stars and photoionized gas using the two-component model of Charlot & Fall (2000), parameterized in terms of the total attenuation optical depth $\hat{\tau}_V$, and the fraction of this arising in the diffuse ISM μ. The mean effects of intergalactic medium absorption are included following the model of Inoue et al. (2014).

For mock galaxies with properties that are observable using current facilities we use BEAGLE to generate a distribution of model SEDs consistent with the observations and assign these SEDs to the mock objects. To achieve this, we fit SED models from BEAGLE to the multi-band photometry of galaxies in two CANDELS fields using the 3D-HST catalog Skelton et al. (2014). When performing parameter estimation, BEAGLE employs the nested sampling algorithm (Skilling et al. 2006) as implemented in MULTINEST (Feroz et al. 2009). This procedure creates a range of statistically acceptable SED fits for each observed galaxy in a subset of the 3D-HST sources (see Sections 3.4.2 and 4.2, while for more detail of the BEAGLE output see C16, Section 3.3) which are then used to produce a parent catalog. This parent catalog is used to assign SEDs to mock objects with high stellar mass (i.e. those with mass above $\log(M/M_\odot) > 8$, or, above the mass-completeness of the 3D-HST catalog if larger in that redshift bin) and low redshift ($z < 4$), where the $\lambda \lesssim 4.5\mu$m photometry provides firm constraints on stellar mass. The SEDs are assigned by finding the closest match in stellar mass and redshift for each mock galaxy within the parent catalog, allowing us to encapsulate the observed diversity of galaxy SEDs at $z < 4$ with relatively few assumptions.

For mock galaxies with realized properties that extend beyond current measurements of real sources, we can leverage the capabilities of BEAGLE to produce theoretical SEDs and generate model spectra for the mock objects. In this second method, we generate a parent catalog built of theoretical SEDs covering a range of model parameters that can be matched to mock galaxy stellar mass, redshift, and, for star-forming galaxies, M_{UV}, and β (see Sections 3.4.3 and 4.2). We use this method at low stellar masses [$\log(M/M_\odot) < 8$] where current galaxy survey sampling of the population is less complete, and at $z \geq 4$ where SED coverage in the rest-frame optical is only available from imaging taken with IRAC, the $3.6-8\mu m$ camera on *Spitzer* (Fazio et al. 2004).

3. GENERATING STAR-FORMING GALAXIES ACROSS COSMIC TIME

Here we describe the phenomenological model and quantitative procedure for generating counts, redshifts, stellar masses, luminosities, and photometric and spectroscopic properties for mock star-forming galaxies. Galaxies are assigned masses and redshifts according to evolving stellar mass functions, as described in Section 3.1. In Sections 3.2 and 3.3 we describe the procedure for assigning integrated star-forming galaxy properties (M_{UV} and β) based on empirical distributions. Finally, in Section 3.4, we describe the procedure for assigning SEDs to star-forming galaxies.

3.1. *Generating star-forming galaxy counts*

In generating a mock galaxy catalog, we aim to reproduce measurements of the star-forming galaxy stellar mass functions at low redshift ($z \lesssim 4$) and the UV luminosity function at high redshift ($z \gtrsim 4$). Our primary mass function constraints come from Tomczak et al. (2014, hereafter T14), while our UV luminosity function constraints are adopted from Bouwens et al. (2015) at $4 \lesssim z \lesssim 8$ and the newest $z \sim 10$ estimate presented in Oesch et al. (2017).

T14 provide measurements of the stellar mass function of star-forming and quiescent galaxies in eight redshift bins in the range $0.2 < z < 3$. They employed imaging data from the FourStar galaxy evolution (ZFOURGE) survey (Straatman et al. 2016) covering the CDFS, COSMOS and UDS fields with 5 near-IR medium-bandwidth filters spanning the

J and H bands, as well as broad-band K_S imaging. Specifically they used the regions that also overlap with CANDELS J_{125} and H_{160} imaging (to ~ 26.5 depth to 5σ), covering a total area of $\sim$316 arcmin^2. Additionally, imaging from NEWFIRM Medium-band Survey (Whitaker et al. 2011) was used in the AEGIS and COSMOS fields, employing the same filter sets as the ZFOURGE survey to shallower depths but wider area to leverage better constraints of the high-mass end of the mass function. Each of the fields also benefit from further imaging that allows comprehensive sampling of galaxy SEDs over the wavelength range $0.3-8\mu$m, with the field-specific filter-sets and imaging programs summarized in Section 2.4 of Straatman et al. (2016).

T14 inferred photometric redshifts and rest-frame colors (used to separate galaxies into star-forming or quiescent based on the UVJ diagram of Whitaker et al. 2011) using the template-based EAZY code (Brammer et al. 2008), while stellar masses were estimated using FAST (Kriek et al. 2009). Within FAST, they used the original Bruzual & Charlot (2003) population synthesis code at fixed solar metallicity, employing a Chabrier (2003) IMF, and a declining exponential star-formation history. The 80% mass completeness limits of their sample increase from $\log(M/M_\odot) \sim 7.75$ at $z \sim 0.5$ to $\log(M/M_\odot) \sim 9.25$ at $z \sim 3$. T14 fit their resulting stellar mass functions with a sum of two Schechter (1976) functions:

$$\begin{aligned}\Phi(M_\star)dM_\star &= \Phi_1(M_\star)dM_\star + \Phi_2(M_\star)dM_\star \\ &= \ln 10\,\phi^*_{1,M}\,10^{(M_\star - M^*_{1,M})(1+\alpha_{1,M})}\exp(-10^{M_\star - M^*_{1,M}})dM_\star \\ &\quad + \ln 10\,\phi^*_{2,M}\,10^{(M_\star - M^*_{2,M})(1+\alpha_{2,M})}\exp(-10^{M_\star - M^*_{2,M}})dM_\star\,, \end{aligned} \quad (4)$$

where $M_\star = \log(M/M_\odot)$, as defined in Section 2.1, $\Phi(M_\star)$ indicates the number of galaxies per Mpc^3 with stellar masses between $M_\star$ and $M_\star + dM_\star$, and $M^*_{1,M}$, $M^*_{2,M}$, $\phi^*_{1,M}$, $\phi^*_{2,M}$, $\alpha_{1,M}$and $\alpha_{2,M}$ are the six free parameters of the function.[4] In a single Schechter function, M^*_M is the mass at the turnover, or "knee" of the mass function, ϕ^*_M is the characteristic number density of galaxies at the turnover, and α_M is the low-mass slope. In the double-Schechter function used in T14, they explicitly set $M^*_{1,M} = M^*_{2,M} = M^*_M$ meaning that they fit with a single "knee" but the different normalizations and faint-end slopes of each function enable them to fit the observed steepening of the mass function to low masses (see Figure 4).

At $z > 4$ stellar masses become progressively less well constrained from measurements, in part because the rest-frame optical SED (a key region containing the Balmer break at $\sim$ 3600 Å, and the 4000 Å break), shifts into the infrared where current facilities have low sensitivity. Additionally, high equivalent width (EW) emission lines can add to the flux in the reddest photometric bands, leading to an over-prediction of galaxy stellar masses (Schaerer & de Barros 2010, Stark et al. 2013, Curtis-Lake et al. 2013, de Barros et al. 2014). As a result, relative uncertainties on stellar mass measurements are high (e.g. 0.4 dex at $10^{10}M_\odot$ at $z = 4$, increasing with redshift and decreasing mass; Grazian et al. 2015, see also Mobasher et al. 2015) and may contribute to the large scatter of mass function measurements in the literature (nearly $\sim$1 dex in counts; see Figure 9 in Song et al. 2016, Figure 11 in Davidzon et al. 2017). Therefore, to generate galaxy counts at $z > 4$ we leverage the constraints provided by the observed UV luminosity function between $4 \lesssim z \lesssim 8$ from Bouwens et al. (2015) with luminosity function measurements with mean redshifts at $< z >= [3.8, 4.9, 5.9, 6.8, 7.9]$ using data from the HST Legacy Fields, as well as the $z \sim 10$ luminosity function of Oesch et al. (2017). The binned UV luminosity function measurements we use for this work are overall consistent with many other results in the literature at $M_{UV} < -17$ (e.g. McLure et al. 2013; Finkelstein et al. 2015; Atek et al. 2015; Laporte et al. 2015; Castellano et al. 2016; Yue et al. 2017; Livermore et al. 2017; Ono et al. 2017; Bouwens et al. 2017b).

We choose to model the redshift evolution of the six mass function parameters across the entire redshift range of the mock, i.e. $0.2 < z < 15$. This ensures a smooth evolution in number counts across the transition from mass to luminosity function-based constraints. At $z < 3.8$ (the mean redshift of the B-dropout sample used to produce the Bouwens et al. (2015) $z \sim 4$ luminosity function) we use the measured mass functions of T14 to directly constrain their redshift evolution, while at $z \geq 3.8$ we use our model of the redshift-evolving M_{UV}–$M_\star$ relation (see Section 3.2) to fit to the observed luminosity functions with mass function parameters. However, it is important to note that this is not a direct prediction of the shape or evolution of the $z \gtrsim 4$ mass functions that we expect to measure with *JWST*. Our $z \gtrsim 4$ mass functions are dependent on our model of the M_{UV}–$M_\star$ relation and additionally, we do not yet know how incomplete the current M_{UV}-selected samples at $z \gtrsim 4$ may be.

To determine a suitable form for the redshift-evolution of the Schechter function parameters, we first need to know what *mass* function parameters can reproduce the observed UV luminosity functions at $z \gtrsim 4$. The details of this fitting are given in Appendix A.1 and we plot the Schechter (mass) function parameters that best fit the $z \gtrsim 4$ luminosity function observations in Figure 3, as well as the individual maximum-likelihood estimates of T14 at $z < 4$. The estimates of M^*_M derived from the measured luminosity functions at $z \gtrsim 4$ are significantly lower than the T14 measurements. However, if we fit the evolution of a double Schechter function with different "knees", as in Equation 4, we can use $M^*_{1,M}$ to fit to the

[4] Schechter function parameters used to describe a mass function are suffixed by an 'M' to distinguish them from those used to describe a luminosity function.

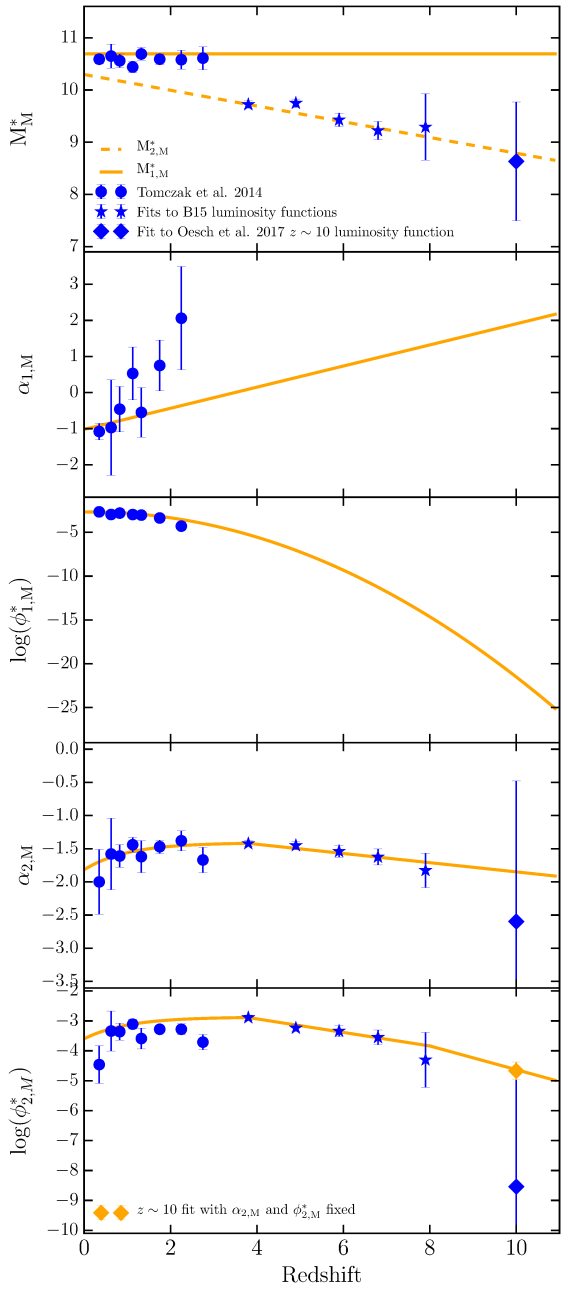

Figure 3. Redshift evolution of each parameter of the double Schechter function adopted in our model. The orange lines show the adopted evolution, while circles represent the original maximum likelihood estimates from T14, with $M^*_{1,M} = M^*_{2,M}$ set explicitly in the fitting. Stars (diamonds) show the median 68% confidence intervals for the parameter estimates from our MCMC fitting (described in Appendix A.1) to the Bouwens et al. (2015) luminosity functions at $z\sim4,5,6,7$ and 8 (Oesch et al. (2017) $z\sim10$, see text for details). The orange diamond in the lowest panel shows the $z\sim10$ luminosity function fit when fixing the values of $\alpha_{2,M}$ and M^*_M. The errors on the $z\sim10$ estimates are symmetric, so we choose to reduce the y-axis range of the panels displaying the redshift evolution of $\alpha_{2,M}$ and $\log(\phi^*_{2,M})$ for clarity.

high-mass end of the $z<3$ mass functions while $M^*_{2,M}$ (plus the fast evolution of $\phi^*_{1,M}$) can be used to account for the rapid evolution in the bright end required to fit to the $z\gtrsim4$ luminosity functions. We therefore choose to set the $z\geq3.8$ evolution of $M^*_{2,M}$, $\alpha_{2,M}$ and $\phi^*_{2,M}$ using a weighted least-squares linear regression to the luminosity function fits. We extrapolate the linear fit of $M^*_{2,M}$ to $z<3.8$ but re-fit the T14 measured mass functions allowing the other five Schechter function parameters to vary. In fact, this choice of $M^*_{2,M}$ evolution somewhat under-estimates the high-mass end of the $2<z<3$ mass functions (see Figure 4). It is entirely possible that the reason for the strong evolution in M^*_M seen between the $z<3.8$ and $z\geq3.8$ samples is due to the M_{UV}-selected samples missing a population of dusty, high mass star-forming galaxies. If they exist, these objects will be revealed by *JWST*, but currently we lack firm constraints on their number density evolution. We are basing this mock catalog on current observational constraints, and so choose to favor the fit to the $z\sim4$ luminosity function over the $2\lesssim z\lesssim3$ mass function at the high-mass end as it allows us to produce a model with number counts that vary relatively smoothly with redshift. As such, a caveat of our model is that we are not modeling the dusty star-forming galaxies currently missed in UV-selected samples, and mildly under-represent the high mass end of the $2\lesssim z\lesssim3$ mass functions. A model that simultaneously fits the $z\sim2.75$ T14 data and the $z\sim4$ Bouwens et al. (2015) luminosity function would require a strong gradient discontinuity in $M^*_{1,M}$ that would lead to a step discontinuity in the number counts of galaxies at high stellar masses.

When fixing the evolution of the Schechter function parameters at $z\gtrsim4$ we use a weighted least-squares linear regression excluding the point at $z\sim10$, which has noticeably lower number densities than can be accounted for by a simple linear relation in all three parameters. In fact, the exact form of the redshift evolution of the UV luminosity function, and associated cosmic star formation rate density (CSFRD) above $z\sim8$ has been an area of active debate in the literature, with e.g. McLeod et al. (2016) presenting measurements of the $z\sim9-10$ luminosity function that is consistent with a smooth decline in the CSFRD. For our fiducial mock catalog we choose to base the model on the Oesch et al. (2017) results in order to provide a conservative limit on $z\gtrsim8$ galaxy number counts likely to be detected with *JWST*. We defer further discussion of this issue to Section 7. The constraints at $z\sim10$ are not strong enough to constrain the likely evolution in $M^*_{2,M}$, $\phi^*_{2,M}$ *and* $\alpha_{2,M}$. We thus choose to re-fit the $z\sim10$ luminosity function with $\alpha_{2,M}$ and $M^*_{2,M}$ fixed to the values defined by the extrapolated linear fits at $z=10$, giving $\log(\phi^*_{2,M}/\mathrm{Mpc^3\,dex^{-1}}) = -4.67\pm0.3$. We then require the gradient of the evolution in $\log(\phi^*_{2,M})$ to decrease further at $z>8$ so that this value is reached by the relation at $z=10$.

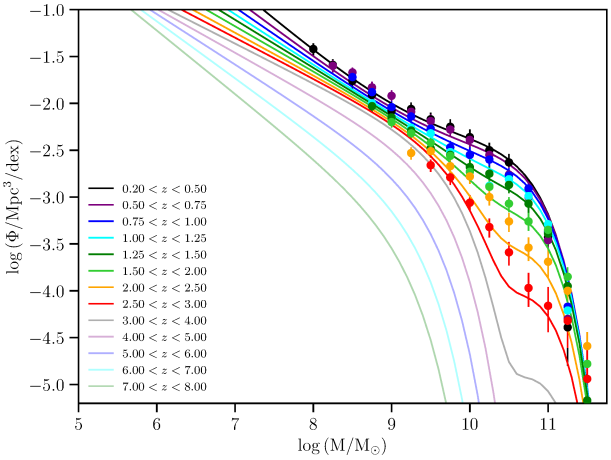

Figure 4. Evolution of the star-forming galaxy mass function from $0.2 < z < 8.0$ (lines), plotted with the observations from T14 (circles). The parameters of this fit to the MF evolution are given in Equations 5-10 and Table 4.

At $z < 3.8$ we re-fit to the T14 mass functions using a Bayesian multi-level modeling approach (see Appendix A.2), which allows us to derive the best-fit redshift evolution of the Schechter function parameters by fitting to the mass function measurements in each redshift bin simultaneously. This approach is more powerful than fitting a functional form to the published Schechter parameter estimates as it accounts for parameter covariance self-consistently. At $z < 3.8$ we choose a functional form for the redshift evolution for $\alpha_{2,\mathrm{M}}$ and $\phi^*_{2,\mathrm{M}}$ that asymptotically approaches the value of the best-fit linear relation at $z = 3.8$, but decreases rapidly at the lowest redshifts. Without this dip to low redshifts, the mass function is too shallow with too-high a normalization at low masses. We accept a mildly discontinuous evolution at $z = 3.8$ because allowing the functional form in either $\alpha_{2,\mathrm{M}}$ or $\phi^*_{2,\mathrm{M}}$ to increase and turn over by $z = 3.8$ (to give smooth evolution at $z = 3.8$) produces mass functions that cross over at low masses, a situation that we are trying to avoid by requiring that our model is monotonically increasing at given mass with decreasing redshift.

The redshift evolution of each Schechter function parameter is summarized below:

$$\mathrm{M}^*_{1,\mathrm{M}}(z) = a_1 \tag{5}$$

$$\log[\phi^*_{1,\mathrm{M}}(z)] = b_1 + b_2\,z + b_3\,z^2 \tag{6}$$

$$\alpha_{1,\mathrm{M}}(z) = c_1 + c_2\,z \tag{7}$$

$$\mathrm{M}^*_{2,\mathrm{M}}(z) = D_1 + D_2\,z \tag{8}$$

$$\begin{aligned} \log[\phi^*_{2,\mathrm{M}}(z)] &= e_1\,[1-\exp(-z)] + e_2 && z < 3.8 \\ &= E_1 + E_2\,z && 3.8 \le z < 8 \\ &= E'_1 + E'_2\,z && z \ge 8 \end{aligned} \tag{9}$$

$$\begin{aligned} \alpha_{2,\mathrm{M}}(z) &= f_1\,[1-\exp(-z)] + f_2 && z < 3.8 \\ &= F_1 + F_2\,z && z \ge 3.8 \end{aligned} \tag{10}$$

where the parameters D_1, D_2, E_1, E_2, F_1 and F_2 are all determined from the linear regression to the forward-modeled luminosity function fitting, and E'_1 and E'_2 are chosen to fit to the $z \sim 10$ luminosity function while maintaining continuous evolution in $\log[\phi^*_{2,\mathrm{M}}(z)]$ at $z = 8$. The parameters e_2 and f_2 are fixed to the values required to produce continuous evolution at $z = 3.8$ with $e_2 = E_1 + 3.8\,E_2 - e_1[1-\exp(-3.8)]$ and $f_2 = F_1 + 3.8\,F_2 - f_1[1-\exp(-3.8)]$. The remaining free parameters a_1, b_1, b_2, b_3, c_1, c_2, e_1 and f_1, are then constrained using the Multi-level modeling (see AppendixA.2) to the published T14 star-forming mass functions (their Table 1).

We report the median values and associated uncertainties along with the values for the model parameters defined by linear fits to the $z \gtrsim 4$ individual mass function estimates in Table 4. The chosen redshift evolution of each parameter is plotted as the orange lines in Figure 3. The resulting mass function comparisons to the T14 measurements at $z < 4$ are plotted in Figure 4, and the luminosity function comparisons are shown in Figure 5.

3.2. *The evolution of the M_{UV}–$\mathrm{M}_\star$ relation*

In this Section, we describe our method to characterize the relation (slope, intercept, scatter) between M_{UV} and $\mathrm{M}_\star$ of galaxies at redshifts $0.2 \le z \le 15$. Hereafter, we use the definition of M_{UV} adopted, e.g., in Robertson et al. (2013), as the average magnitude at rest-frame wavelength within a flat filter centered at 1500 Å and with a width of 100 Å, which is the definition adopted by BEAGLE (Chevallard & Charlot 2016). This definition of M_{UV} differs slightly from that used to measure the UV luminosity functions in Bouwens et al. (2015), which define M_{UV} to be at rest-frame wavelength of 1600Å. We calculate the typical color correction based on the mean β as a function of M_{UV} and redshift presented in Bouwens et al. (2014) and find that the typical difference in magnitudes between 1500 and 1600Å is negligible ($|\delta M_{\mathrm{UV}}| \lesssim 0.05$). This correction is significantly smaller than the k-correction applied to estimate M_{UV} at 1600Å from broad-band photometry

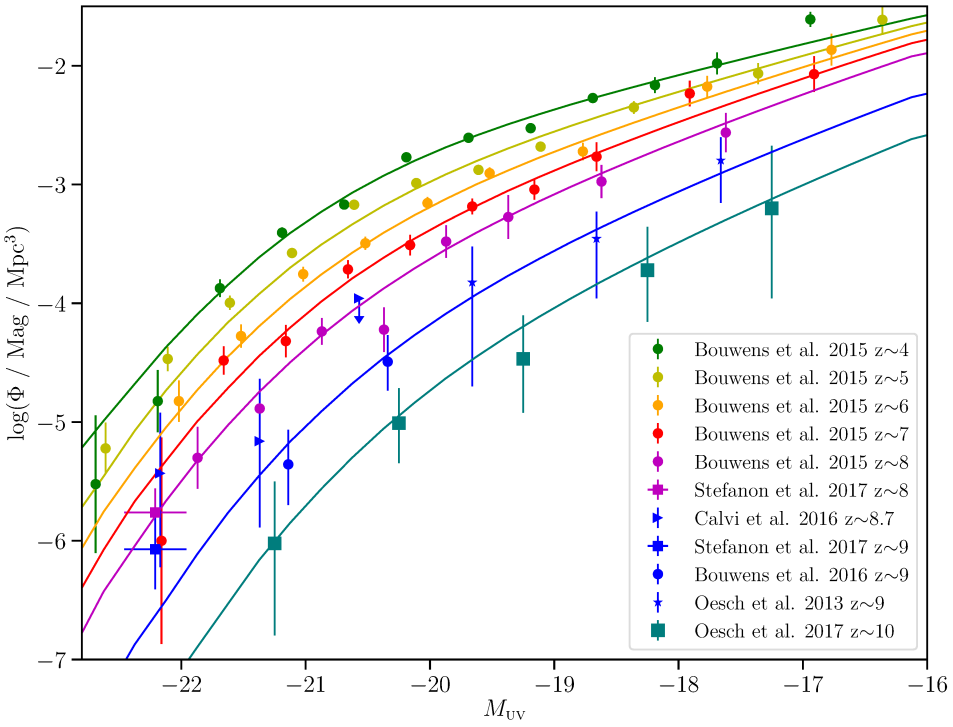

Figure 5. The UV luminosity function at $z \gtrsim 4$ of our continuously evolving phenomenological model (solid lines; described in Section 3), evaluated at the mean redshift of the dropout samples used in the fitting. Points are observations at the same mean redshifts as indicated by the colors (Bouwens et al. 2015; Stefanon et al. 2017b; Calvi et al. 2016; Bouwens et al. 2016b; Oesch et al. 2013, 2017). Our forward modeling approach explicitly fits to the binned UV luminosity functions of Bouwens et al. (2015); Oesch et al. (2017).

in the first place ($|\delta M_{\rm UV}| \lesssim 0.1$) and so we apply no conversion between rest-frame 1500Å and 1600Å $M_{\rm UV}$ values.

The $M_{\rm UV}$–M$_\star$ distribution and its evolution are critical components of our underlying phenomenological model, and are required to statistically assign UV luminosities to mock galaxies generated from our continuously evolving stellar mass function model. However, we note the following uncertainties to this procedure. At all redshifts, galaxies exhibit a diversity of mass-to-light ratios, which depend on the stellar population properties (age, metallicity), star-formation history, and dust content of a galaxy. As a result, the exact form of the relation between $M_{\rm UV}$ and M$_\star$ and its dependency on galaxy properties are largely unknown. In general, brighter galaxies at UV wavelengths correspond to more massive objects (e.g. Stark et al. 2009; Lee et al. 2011; González et al. 2011), and this holds out to $z \sim 7$ (Duncan et al. 2014; Salmon et al. 2015; Grazian et al. 2015; Song et al. 2016). Although the relation between $M_{\rm UV}$ and M$_\star$ follows a general trend of decreasing $M_{\rm UV}$ with M$_\star$ out to $\log(\mathrm{M}/\mathrm{M}_\odot) \sim 10$, at higher masses the average $M_{\rm UV}$ becomes fainter due to the appearance of a population of fainter objects. This trend could be attributable to several effects, such as increased dust content and older average stellar ages among massive galaxies (e.g. Spitler et al. 2014). Characterizing the relationship is further complicated by the difficulty of measuring stellar mass owing to emission line contamination at high-redshift (Labbé et al. 2013; Stark et al. 2013) and at low stellar masses (Whitaker et al. 2014), and the lack of a direct photometric probe of $M_{\rm UV}$ at intermediate redshifts ($0.6 < z < 1.5$). The procedure we outline here has a direct impact on the resulting UV luminosity functions (see Sections 2.1, 3.1 and 6.1). We have therefore developed a straightforward description of the $M_{\rm UV}$–M$_\star$ distribution and its evolution that is designed to encapsulate the diversity of real galaxies.

3.2.1. *Characterizing the evolution of $M_{\rm UV}$–M$_\star$ from observations*

We characterize the $M_{\rm UV}$–M$_\star$ relationship at $z \lesssim 4$ using measurements from the 3D-HST catalog (using SED-fitting with BEAGLE; see description in Section 3.4.1). As discussed extensively in Stefanon et al. (2017a), selection effects can heavily influence the observed shape of the $M_{\rm UV}$–M$_\star$ distribution. Therefore, we avoid including observed galaxies whose $M_{\rm UV}$ or M$_\star$ measurements are poorly constrained by the BEAGLE fits. Specifically we only use galaxies with

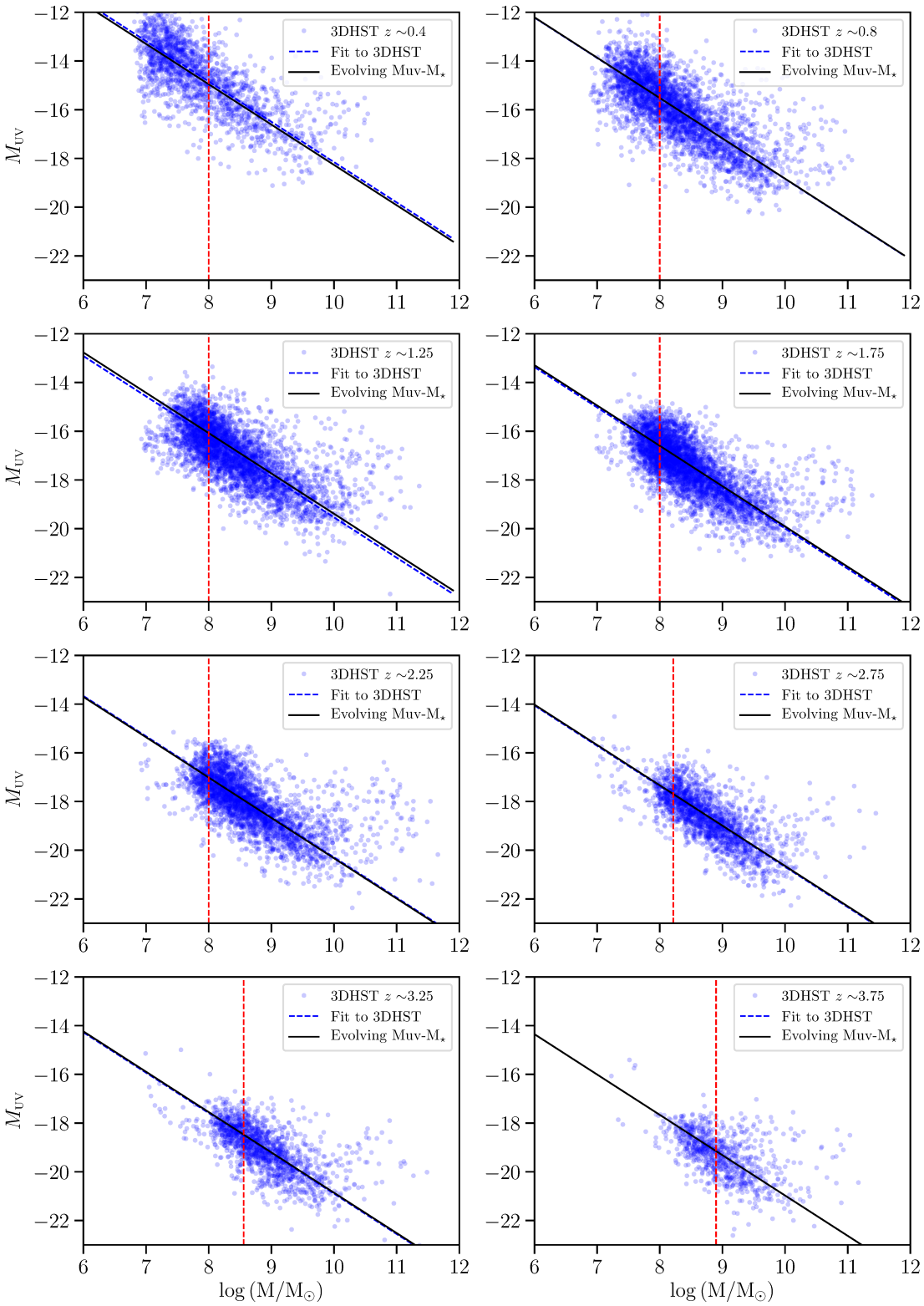

Figure 6. The $M_{\rm UV}$–M$_\star$ relation for realizations of the SED-fitting of observed star-forming galaxies ($\log(\mathrm{sSFR}) < -10$ Gyr^{-1}) in the 3D-HST survey (blue points) with high-confidence measurements of $M_{\rm UV}$, M, and redshift as described in Sections 3.2 and 3.4.1. Blue dashed lines indicate the best-fitting linear relationship under the assumption of a fixed slope as described in the text. Solid black lines indicate the smoothly evolving redshift-evolution of the $M_{\rm UV}$–M$_\star$ relation in our model as characterized in Figure 7 and Equation 11. The red dashed line indicates the stellar mass above which mock galaxies are matched to 3D-HST realizations, which is the larger between $\log(\mathrm{M}/\mathrm{M}_\odot) > 8$ and $\log(\mathrm{M}/\mathrm{M}_\odot) > 6.3+0.7z$ (the evolving mass limit exceeds $\log(\mathrm{M}/\mathrm{M}_\odot) > 8$ at $z > 2.4$).

Table 1. The values of the parameters used in our model of the mass function evolution, as described in Equations 5-10. For those parameters determined using the multi-level model fitting to $z < 4$ mass functions, we report the median of the posterior distribution function, its 1σ confidence interval, as well as the prior used in the fitting.

parameter	median	1σ uncertainty	prior/source of fits
a_1	10.69	0.04	$\mathcal{N}(0,50)$
b_1	-2.68	0.16	$\mathcal{N}(0,50)$
b_2	0.06	0.24	$\mathcal{N}(0,50)$
b_3	-0.19	0.08	$\mathcal{N}(0,50), \in [-\infty,0]$
c_1	-1.02	0.16	$\mathcal{N}(0,50)$
c_2	0.29	0.13	$\mathcal{N}(0,50)$
D_1	10.30	0.10	Linear fitting $4 \lesssim z < 8$
D_2	-0.15	0.02	Linear fitting $4 \lesssim z < 8$
e_1	0.73	0.26	$\mathcal{N}(0,50), \in [0,\infty]$
e_2	-3.60	-	$= E_1 + 3.8E_2 - e_1[1 - \exp(-3.8)]$
E_1	-2.03	0.41	Linear fitting $4 \lesssim z < 8$
E_2	-0.23	0.09	Linear fitting $4 \lesssim z < 8$
E_1'	-0.67	-	$\phi^*_{2,\mathrm{M}}$ fit to $z \sim 10$ LF
E_2'	-0.40	-	$\phi^*_{2,\mathrm{M}}$ fit to $z \sim 10$ LF
f_1	0.41	0.17	$\mathcal{N}(0,50), \in [0,\infty]$
f_2	-1.82	-	$= F_1 + 3.8F_2 - f_1[1 - \exp(-3.8)]$
F_1	-1.16	0.10	Linear fitting $4 \lesssim z < 8$
F_2	-0.07	0.02	Linear fitting $4 \lesssim z < 8$

$\delta\log(\mathrm{M}/\mathrm{M}_\odot) < 1$, $\delta z < 1$ and $\delta M_{\mathrm{UV}} < 1$ (where e.g. δz is the 68% credibility interval on redshift.) The limits imposed were chosen to avoid biasing the characterization of the M_{UV}–$\mathrm{M}_\star$ distribution with overly strict M_{UV} or $\mathrm{M}_\star$ cuts, which we discuss further below.

In Figure 6, we plot the M_{UV}–$\mathrm{M}_\star$ distributions for the 3D-HST galaxies with well-constrained M_{UV}, $\mathrm{M}_\star$ and redshift measurements. As discussed above, these distributions show a trend of increasing stellar mass with decreasing M_{UV} at low stellar mass. At high stellar mass the M_{UV} values tend to be fainter than the linear relation, as observed in Spitler et al. (2014). Rather than attempting to fully model this mass-dependent behavior, especially given the Malmquist biases that begin to affect the higher-redshift bins, we adopt the following two-step procedure to describe the M_{UV}–$\mathrm{M}_\star$ distributions at $z \lesssim 4$. We fit the observed M_{UV}–$\mathrm{M}_\star$ distribution under the simplest assumption of a linear relationship to extrapolate to low masses, while at higher masses ($\log(\mathrm{M}/\mathrm{M}_\odot) \gtrsim$8-8.5, depending on the mass limit at a given redshift), we assign M_{UV} values by sampling from real galaxies of the same mass. The matching procedure allows us to maintain the observed flattening of the distribution at high masses, and is fully described in Section 3.4.2 below.

Figure 6 illustrates the substantial scatter in the observed M_{UV}–$\mathrm{M}_\star$ distributions at $z \lesssim 4$. Owing to the large scatter, the best fitting slope will depend strongly on the uncertainties on the data points, and the size of the uncertainties may depend on M_{UV}, $\mathrm{M}_\star$, and also plausibly on redshift. Indeed, we find that when fitting with both slope and normalization as free parameters neither parameter is well constrained, and the best fitting slope is highly variable between redshift bins. Therefore, we adopt a fixed slope for the M_{UV}–$\mathrm{M}_\star$ relation at all redshifts and fit only the intercept at each redshift. This procedure essentially fits the average redshift-dependent mass to light ratio, which has lower uncertainty and is less dependent on the error on individual galaxy measurements and stellar mass-dependent systematics. Several studies have reported the measurement of constant slope for UV-selected galaxies, with normalization evolving in redshift (Duncan et al. 2014; Salmon et al. 2015; Grazian et al. 2015; Song et al. 2016; Stefanon et al. 2017a), and find a reasonable description of the data. The blue dashed line in Figure 6 shows our best fit relation to the M_{UV}–$\mathrm{M}_\star$ distribution in each redshift bin, where the slope is fixed to a value of -1.66. We find excellent agreement with the observed distribution at all redshifts. For reference we also indicate the stellar mass limits in each redshift bin above which we assign M_{UV} values by sampling from real galaxies (red dashed lines). Fitting the M_{UV}–$\mathrm{M}_\star$ distribution only above these mass limits instead has a negligible effect on the result at $z < 3$. At $z \sim 3.75$ where there are fewer well-constrained measurements, fitting above this mass limit would increase the M_{UV}–$\mathrm{M}_\star$ intercept by ~ 0.1 mag, an indication that fitting only at the high-mass end biases the characterization of the M_{UV}–$\mathrm{M}_\star$ due to the high-mass end flattening. Therefore, we choose to proceed using all galaxies with well-characterized stellar mass, redshift and M_{UV}.

To set the full redshift evolution of the M_{UV}–$\mathrm{M}_\star$ relation, we combine the intercept values for the best fit relations with fixed slope at each redshift $z \lesssim 4$ with measurements of the M_{UV}–$\mathrm{M}_\star$ intercept at $z > 4$. We use the average observed value of stellar mass for bright (M_{UV}= -20) galaxies at $4 < z \leq 7$ to set the overall normalization in each $z > 4$ redshift bin, while assuming the same constant M_{UV}–$\mathrm{M}_\star$ slope. We utilize the high-redshift stellar mass measurements shown in figure 7 of Stark et al. (2013), where the measured stellar masses were fit while including the contribution to the SED from nebular emission lines. The normalization value at $M_{\mathrm{UV}} = -20$ shows an overall decline between

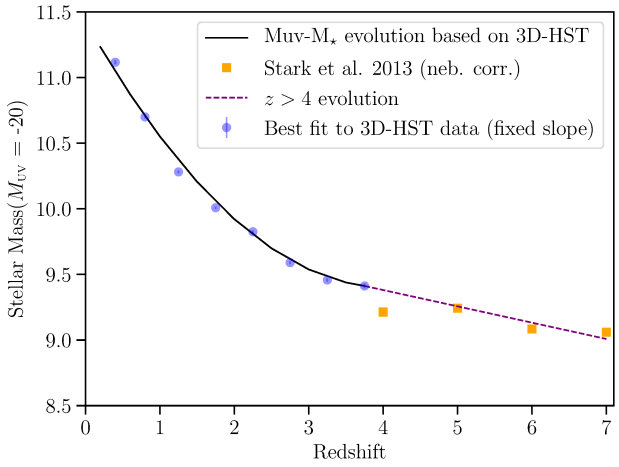

Figure 7. The redshift evolution of the intercept of the $M_{\rm UV}$–M$_\star$ distribution, $M^0_{M_{\rm UV}-{\rm M}_\star}(z)$, defined at $M_{\rm UV} = -20$. Blue points indicate the best fit intercept to the 3D-HST data presented in Figure 6 assuming a fixed slope. Error bars are smaller than the size of the symbol. Orange points are based on relation presented in Stark et al. (2013) that includes the correction for nebular emission lines.

$4 \lesssim z < 7$, indicating a decrease in the average mass to light ratios of galaxies with increasing redshift. The measured values for the $M_{\rm UV}$–M$_\star$ intercepts at all redshift bins, evaluated at $M_{\rm UV} = -20$, are shown as points in Figure 7.

3.2.2. *Continuous redshift evolution of $M_{\rm UV}$–M$_\star$*

Our model of the redshift evolution of the $M_{\rm UV}$–M$_\star$ relation is defined by fitting to the evolving intercept values shown in Figure 7. These intercepts show a rapid decline from $z \sim 0-3$, with a shallower decline at $z > 4$. We find that at $z \leq 4$ the intercept measurements are adequately described by a quadratic function, and a linear function at $z > 4$. To ensure these two functions remain continuous at $z \sim 4$, we include the boundary condition that the derivatives of the two functions are equal at the center of the highest redshift bin where we use the 3D-HST data ($z = 3.75$). This constraint results in the following function to describe the intercept of the $M_{\rm UV}$–M$_\star$ relation, $M^0_{M_{\rm UV}-{\rm M}_\star}(z)$, evaluated at $M_{\rm UV} = -20$ and $z \leq 3.75$:

$$M^0_{M_{\rm UV}-{\rm M}_\star}(z) = a(z-3.75)^2 + [b - 2a(z-3.75)](z-3.75) + c \quad (11)$$

where $a = 0.12$, $b = 0.08$, and $c = 9.41$. This evolution of the $M_{\rm UV}$–M$_\star$ intercept is shown as the black solid curve in Figure 7, along with the observed data (blue points). The resulting linear $M_{\rm UV}$–M$_\star$ relations in each redshift bin, according to the smoothly evolving intercept function defined in Equation 11, are also shown as black solid lines in Figure 6.

At $z > 3.75$, the $M_{\rm UV}$–M$_\star$ intercept evaluated at $M_{\rm UV} = -20$ evolves approximately linearly with the following form:

$$\begin{aligned} M^0_{M_{\rm UV}-{\rm M}_\star}(3.75 < z < 8) &= -0.12 \times z + 9.88 \\ M^0_{M_{\rm UV}-{\rm M}_\star}(z \geq 8) &= 8.92 \end{aligned} \quad (12)$$

We use Equation 11 to assign $M_{\rm UV}$ values to galaxies of a given stellar mass and redshift at $z \leq 3.75$, and Equation 12 to assign $M_{\rm UV}$ values at $z > 3.75$, assigned randomly within the scatter observed in 3D-HST data in Figure 6. We have characterized this scatter in both stellar mass, $M_{\rm UV}$, and redshift, and find that the scatter in $M_{\rm UV}$ is remarkably constant in both stellar mass and redshift, with an average value of $\sigma_{uv} \sim 0.7$ magnitudes. Therefore, at all redshifts, we assign $M_{\rm UV}$ values randomly according to this relation and that assumed Gaussian scatter σ_{uv}. This direct assignment of $M_{\rm UV}$ applies only to low-redshift low-mass star-forming mock galaxies ($z \leq 4$ and have $\log({\rm M}/{\rm M}_\odot) \leq 8$), or high-redshift star-forming galaxies at $z > 4$. Massive, low-redshift mock galaxies (with $\log({\rm M}/{\rm M}_\odot) > 8$ and $z \leq 4$) are assigned $M_{\rm UV}$ values according to their matched 3D-HST realization. The $M_{\rm UV}$–M$_\star$ relation and scatter as described here are used as priors when drawing realizations from fits to 3D-HST galaxies to ensure a smooth transition in the mock catalog $M_{\rm UV}$–M$_\star$ relation at $\log({\rm M}/{\rm M}_\odot) = 8$. This procedure is detailed in Section 3.4.2.

3.2.3. *Theoretical limits on the mass to light ratios of galaxies*

As mass to light ratios continue to decrease with increasing redshift and decreasing stellar mass according to our model, the mass to light ratios approach a theoretical limit of the stellar population models we generate with BEAGLE. This limit is set by the UV luminosities of individual massive stars, and represents the minimum mass to light ratio possible for an instantaneous burst of star-formation for any given IMF (with no dust attenuation, the lowest metallicity, and corresponding nebular continuum emission). Any IMF choice will result in such a limit in the possible $M_{\rm UV}$ given a stellar mass, with more top-heavy or bottom-light IMFs allowing for brighter limiting $M_{\rm UV}$ and bottom heavy IMFs producing fainter limiting $M_{\rm UV}$. The Chabrier IMF that we use in this work is relatively bottom-light and has a larger mass to light ratio parameter space than a more bottom heavy IMF (e.g. Kroupa 2001; Salpeter 1955). For a Chabrier IMF with our assumed high-mass cutoff of 100 M$_\odot$, we find that the stellar plus nebular continuum emission results in a theoretical mass to light ratio given by $M_{\rm UV} \approx -2.45 \log({\rm M}/{\rm M}_\odot) - 1.3$.

To accommodate this theoretical limit in the $M_{\rm UV}$–M$_\star$ evolution of our phenomenological model, we truncate the Gaussian distribution that we use to assign $M_{\rm UV}$ values. For galaxies at the detection limit of future blank surveys (e.g. apparent magnitude $m_{app} \sim 31$) this truncation has a negligible effect on the overall $M_{\rm UV}$–M$_\star$ distribution at $z < 4$. At $z > 4$, we find that scatter using the truncated Gaussian at

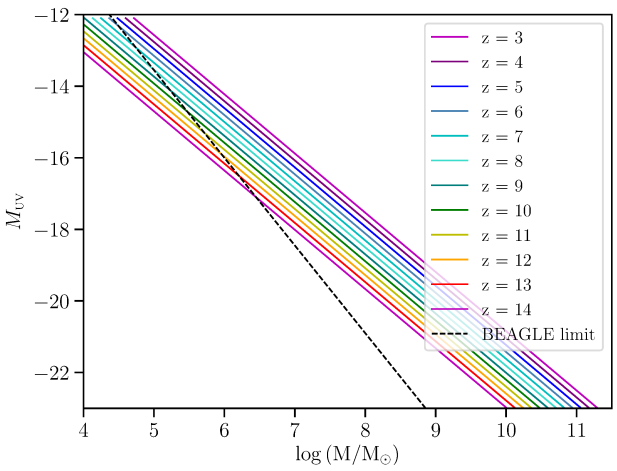

Figure 8. Demonstration of the evolution of the mean M_{UV}–M$_\star$ relation with redshift, (if mass to light ratios were allowed to continue decreasing above $z > 8$), in comparison to the mass-to-light ratio limit imposed by the stellar population modeling with BEAGLE (which implicitly assumes a Chabrier IMF with a high mass cut-off of 100 M$_\odot$.) In our model, all $z > 8$ galaxies follow the M_{UV}–M$_\star$ relation at $z = 8$ to avoid a scenario where significant numbers of galaxies exceed the theoretical mass-to-light ratio limit set by BEAGLE.

the limit changes the overall shape of the M_{UV}–M$_\star$ distribution by steepening the low-mass end. The effect becomes significant by $z \sim 8$. We therefore additionally halt the redshift evolution of the mean M_{UV}–M$_\star$ relation parametrized in Equation 12 at $z = 8$. A demonstration of the BEAGLE mass to light ratio limit is shown in Figure 8, compared with the projected evolution of the mean M_{UV}–M$_\star$ relation allowed to evolve past z∼8.

Although we make every effort to choose reasonable constraints where available, the overall shape of the M_{UV}–M$_\star$ distribution at $z > 4$ is still an extrapolation that impacts various evolutionary relations at high-redshift in the model, including the UV luminosity function and specific star-formation rate (sSFR). We will discuss these issues in depth in Section 6.

3.3. *UV continuum slope -M_{UV} relationship*

The spectral slope of the UV-continuum (β where $f_\lambda \propto \lambda^\beta$) of galaxies is sensitive to the properties of stellar populations (e.g. metallicity and age), star-formation history, and dust attenuation. Population studies of star-forming galaxies indicate well-characterized relationships between β and UV luminosity. Bright, massive galaxies tend to have red (i.e., shallower) UV continua, which likely owes to a combination of old stars, higher stellar metallicities, and a larger dust content. The bluer (i.e. steeper) UV continua of lower luminosity galaxies are often associated with younger, less metal-rich stellar populations, and less dust attenuation (e.g. Stanway et al. 2005; Labbé et al. 2007; Bouwens et al. 2009; Labbé et al. 2010; Bouwens et al. 2012b; Rogers et al. 2013, 2014). The detailed relations between β and M_{UV}, scatter, and evolution out to redshift $z \sim 8$ are still areas of active research, but many studies are consistent with a linear relationship between β and M_{UV} (e.g. Bouwens et al. 2012b; Kurczynski et al. 2014; Bouwens et al. 2014; Alavi et al. 2014; Rogers et al. 2014) with an average evolutionary trend towards bluer β with increasing redshift (e.g. Labbé et al. 2007; Bouwens et al. 2009; Finkelstein et al. 2012b; Castellano et al. 2012; Wilkins et al. 2011; Bouwens et al. 2014).

3.3.1. *Mean β–M_{UV} relation across cosmic time*

We use a compilation of measured β–M_{UV} relations and their scatter across redshifts to assign the rest-frame UV SEDs of mock galaxies. Following several studies (see previous Section), we model the average β–M_{UV} relation with a linear function, where the slope $d\beta(z)/dM_{UV}$ and intercept $\beta(M_{UV} = -19.5, z)$ of the function vary with redshift. We consider sets of β–M_{UV} relations at $1 \leq z \leq 8$ obtained from HST/ACS (Kurczynski et al. 2014; Alavi et al. 2014; Mehta et al. 2017) and HST/WFC3 imaging (Bouwens et al. 2009, 2014). The relationships describing β–M_{UV} at the high redshifts we model are broadly consistent with fits measured in other studies (e.g. Bouwens et al. 2012b; Finkelstein et al. 2015; Rogers et al. 2014).

We find that the slope of the β–M_{UV} relation shows little evolution at redshifts $1 \leq z \leq 8$, as already found in previous works (e.g. Bouwens et al. 2012b; Kurczynski et al. 2014; Bouwens et al. 2014). The intercept of the relation increases significantly from redshift $z \sim 1-8$, reflecting the evolutionary trend that galaxies have older ages and higher metallicities at later cosmic times (e.g. Labbé et al. 2007). We perform least-squares linear fits to measurements of both $d\beta(z)/dM_{UV}$ and $\beta(M_{UV} = -19.5, z)$ and their errors from the literature to produce a mean relation that smoothly evolves with redshift (see Figure 9), described by

$$\begin{aligned} d\beta(z)/dM_{UV} &= -0.007\,z - 0.09 \\ \beta(M_{UV} = -19.5, z) &= -0.09\,z - 1.49. \end{aligned} \tag{13}$$

For mock galaxies at $z \leq 1$, we extrapolate this relationship to lower redshifts to assign β values. At $z > 8$ we use the β–M_{UV} relationship at $z = 8$ to assign β to mock galaxies. There exist several motivations to curb the evolution of β–M_{UV}, with the foremost being the existence of a theoretical limit on the steepness of the UV spectrum emitted by non-Pop. III stars (see discussion in Section 3.3.2 below). Further, the data do not yet constrain evolutionary trends at the highest redshifts currently accessible ($z \sim 5-8$). While evolutionary trends with redshift are observed in most analyses (Finkelstein et al. 2012b; Bouwens et al. 2014), the combination of high-redshift color selections with flux boosting from noise

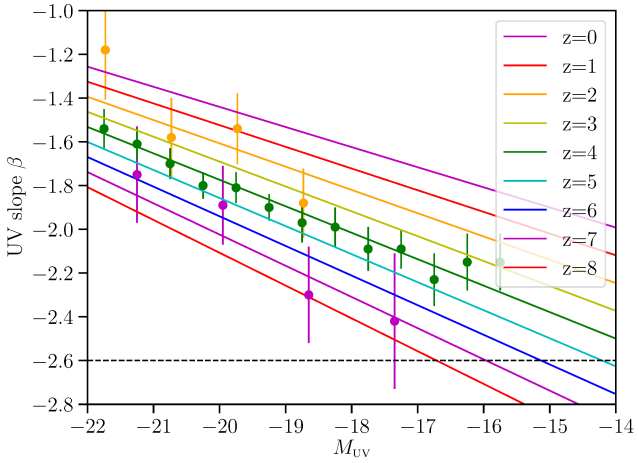

Figure 9. Redshift evolution of the mean β–$M_{\rm UV}$ relation. Intrinsic scatter is included as described in Section 3.3.3. Points indicate a selection of binned measurements from Bouwens et al. (2009, 2014, circles) at similar redshifts of the colored lines (orange at $z \sim 2.5$; green at $z \sim 4$; magenta at $z \sim 7$). The black dashed line indicates the theoretical limit in the BEAGLE models as discussed in Section 3.3.2.

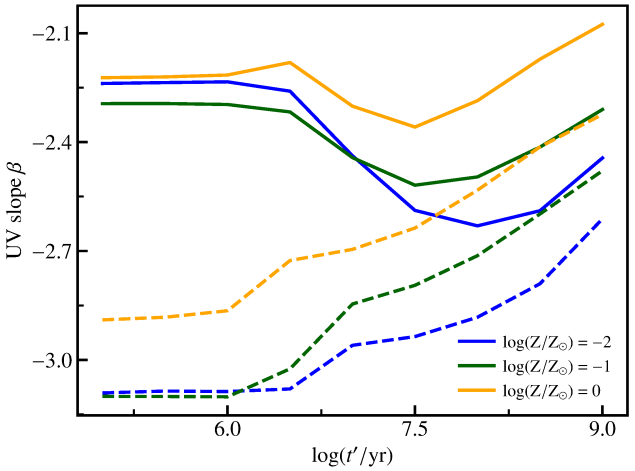

Figure 10. UV-continuum slopes predicted by the spectral evolution model adopted in this work. We show model predictions for a constant SFH of different ages and three metallicities, $\log Z/Z_\odot = -2$ (blue), -1 (green), and 0 (orange). Dashed lines indicate predictions for stellar emission only, while solid lines for stellar and nebular continuum emission. In this latter case, we consider a photoionization model with ionization parameter $\log U_{\rm S} = -2.5$ and depletion factor $\xi_{\rm d} = 0.3$.

in the filters used to measure β are still likely causing statistical studies to be biased against redder β measurements at $z > 5$ (Dunlop et al. 2012; Rogers et al. 2013). At the very highest redshifts currently accessible ($z \sim 7-10$) the data do not provide strong evidence for or against any evolutionary trend in β with redshift or $M_{\rm UV}$ (Dunlop et al. 2013; Wilkins et al. 2016), although the dynamic range in $M_{\rm UV}$ is relatively small at such early times. While evolution cannot be excluded by current data, deep imaging surveys with *JWST* will enable more robust characterization of the evolution beyond $z \sim 8$.

In our model, all mock galaxies are assigned a β according to the mean relationship and intrinsic, photometric bias-corrected scatter described in Section 3.3.3. We now detail the theoretical predictions for the UV continuum slope of our combined stellar population and photoionization model.

3.3.2. *Theoretical limits on UV-continuum slopes*

Model predictions on the shape of the UV-continuum emission of galaxies depend on the assumed properties of stellar populations and ISM (gas and dust). Young, massive stars show blue UV continua, which become bluer with decreasing metallicity. The effects of stellar age and metallicity on the UV-continuum emission over time in our spectral evolution model are illustrated in Fig. 10, which shows that the bluest UV spectra are obtained for very young ($\lesssim 1$ Myr) stellar populations with sub-solar metallicities (dashed lines). Dust reddens the emitted stellar spectrum, and leads to a relation between the attenuation suffered by a galaxy and its UV slope β (Meurer et al. 1999). Recombination-continuum from ionized hydrogen also reddens the UV continuum emission emerging from a galaxy. This effect is shown by the solid lines in Fig. 10, which illustrates how our combined stellar population + photoionization model predicts redder β slopes than a model accounting only for stellar emission.

In order to avoid unphysical values for the β slopes associated to our mock galaxies through the redshift-dependent β–$M_{\rm UV}$ relation presented in Section 3.3.1 above, we impose a limit $\beta_{\rm min} = -2.6$ for the bluest possible value. This limit corresponds, approximately, to the bluest β value obtained with BEAGLE for a model with constant SFH, sub-solar metallicity, low depletion factor ($\xi_{\rm d} = 0.1$), and low ionization parameter ($\log U_{\rm S} = -3.5$). We note that models with non-zero escape fraction of H-ionizing photons can reach bluer values.

3.3.3. *Scatter in the β–$M_{\rm UV}$ relation across cosmic time*

The scatter in the β–$M_{\rm UV}$ distribution discussed in Section 3.3.1 encodes the intrinsic diversity in age, metallicity, and dust attenuation of the galaxy population at fixed redshift and UV luminosity. We aim to assign β values to mock galaxies following the intrinsic scatter (i.e., corrected for photometric biases) of the β–$M_{\rm UV}$ distributions over cosmic time. Much effort has been put into characterizing the scatter, and how it might change with redshift, UV luminosity, or other galaxy properties (e.g. Bouwens et al. 2012b). The intrinsic scatter in β at fixed UV luminosity, σ_β, is surprisingly uniform across all redshifts from $z \sim 1-6$, and rel-

atively independent of UV luminosity with values between $\sigma_\beta \sim 0.3-0.4$ (Bouwens et al. 2012b, 2014; Kurczynski et al. 2014; Mehta et al. 2017). We note that there is some evidence for the intrinsic scatter of the distribution increasing with UV luminosity, such that populations of brighter galaxies will have larger intrinsic scatter in β (Rogers et al. 2014). This evidence comes from a careful analysis at $z \sim 5$ only, however, and such luminosity dependence in the σ_β is not characterized sufficiently across cosmic time to be incorporated in our model. We correspondingly adopt an intrinsic scatter of $\sigma_\beta \sim 0.35$ in our model uniformly across all redshifts and UV luminosities.

To include intrinsic scatter $\sigma_\beta \sim 0.35$ in our model, we assign β values to mock galaxies according to a Gaussian distribution with a mean defined at a given redshift and $M_{\rm UV}$ according to Equation 13 with $\sigma_\beta = 0.35$. To avoid values of β bluer than the theoretical limits described in the previous section (for which we would not be able to associate a BEAGLE spectrum), we truncate the distribution at $\beta = -2.6$. With this truncated Gaussian scatter, at the very highest redshifts and faintest $M_{\rm UV}$ the mean value of β reddens slightly so as to cause a mild flattening of the linear relations shown in Figure 9. However, we accept this feature as more favorable than artificially fixing to the bluest value of β and reducing the galaxy diversity in the mock. In addition, it mimics the behavior of β–$M_{\rm UV}$ seen in some studies that indicate an apparent flattening of the linear relation at faint luminosities (e.g. $M_{\rm UV} \gtrsim -19$; Bouwens et al. 2014). Future measurements from *JWST* imaging and spectroscopy will help to inform our stellar population synthesis models as we uncover the full range and distribution of UV continuum slopes in the early universe.

As described further in the following section, when possible we use 3D-HST galaxies to provide the constraints on the shape of the SED for mock galaxies. However, when this is not possible, we use the β slope that is assigned to each galaxy to match to a parent catalog of SEDs produced by BEAGLE (see Section 3.4.3). This ensures that our catalog will follow observed trends in β–$M_{\rm UV}$.

3.4. *Assigning Galaxy SEDs and spectroscopic properties*

We assign a set of spectral properties to each mock galaxy, allowing us to provide filter photometry as well as a full spectrum for each object. The general method is to produce a parent catalog of spectra that can be matched to galaxies in the mock. Where possible we produce this parent catalog from the results of SED fitting to galaxies in the 3D-HST catalog, allowing the observed photometry to provide the diversity of observed SEDs at given stellar mass and redshift (see Section 2.3). We limit the use of these empirical SEDs to $z \leq 4$ galaxies, as beyond this redshift the rest-frame optical is only sampled by the Spitzer IRAC bands where the poor resolution leads to a significant confusion of sources at faint magnitudes.

For galaxies at redshifts $z > 4$, or $z \leq 4$ but below the mass completeness limits of 3D-HST, we rely on extrapolations of observed relationships between $M_{\rm UV}$–$M_\star$ and β–$M_{\rm UV}$ to provide constraints on the galaxy SEDs. These constraints are used to match to a parent catalog, produced using BEAGLE in mock catalog mode. When generating this parent catalog we have the full parameter space of the stellar and nebular templates to choose from (described in Section 2.3) and so we use observed trends in galaxy physical parameters, albeit with large scatter, to restrict this parameter space. Specifically we use three observed relations: $M_\star - Z - \psi$, where ψ is the SFR of the object (the 'fundamental metallicity relation'); $\psi - Z - \hat{\tau}_{\rm V}$, to provide physically motivated constraints on dust attenuation (where $\hat{\tau}_{\rm V}$is the effective V-band optical depth); and $Z - \log U_{\rm S}$.

3.4.1. *SED fitting to 3D-HST catalogs*

The photometric catalogs produced by the 3D-HST team (Skelton et al. 2014) are selected from the noise-equalized combination of HST/WFC3 J_{125}, JH_{140} and H_{160} images taken from an extensive set of publicly available imaging data over 5 fields (AEGIS, COSMOS, GOODS-North, GOODS-South, and the UKIDSS UDS) covering $\simeq 900$ arcmin^2. From these catalogs we use the data in the deeper regions of the CANDELS (Grogin 2011, Koekemoer et al. 2011) GOODS-South and GOODS-North fields to provide a parent catalog of redshift and mass-dependent SEDs that can be assigned to our mock catalog galaxies.

We use version 4.1 of the 3D-HST photometric catalogs (Momcheva et al. 2016). These catalogs include selection from mosaics that include HUDF-09 (11563; PI:Illingworth) and HUDF-12 (12498; PI: Ellis) WFC3 imaging in the HUDF and parallels (11563; PI: Illingworth) that was performed as part of release 3.0[5]. The catalogs do not include deeper HUDF ACS imaging of Beckwith et al. (2006) and so ACS photometry across the GOODS-South is at the depths of the original GOODS imaging (Giavalisco et al. 2004).

From the GOODS-South and GOODS-North 3D-HST catalogs we fit to the broad-band HST fluxes (B_{435}, V_{606}, i_{775}, z_{850}, J_{125}, JH_{140} and H_{160}), as well as the *Spitzer*/IRAC Channel 1 (3.6μm) and Channel 2 (4.5μm) imaging from SEDS (Ashby et al. 2013) to provide constraints in the rest-frame optical at high redshifts. We also use a subset of the ground-based filters that required small photometric zeropoint corrections in the SED fitting analysis of Skelton et al. (2014) compared to the H_{160} band (see their Table 11). In the

[5] as outlined in 3D-HST v3.0 release documentation http://monoceros.astro.yale.edu/RELEASE_V3.0/Photometry/3dhst_v3.0_readme.pdf

GOODS-South field we use photometry from VLT/ISAAC J, H, and K_s band imaging from the ESO/GOODS and FIRE-WORKS surveys (Wuyts et al. 2008; Retzlaff et al. 2010), and VLT/VIMOS U-band imaging from the ESO/GOODS survey (Nonino et al. 2009). In the GOODS-North field we use Subaru/MOIRCS imaging in J, H and Ks bands from the MODS survey (Kajisawa et al. 2011)[6]. We do not apply the zeropoint offsets reported in Skelton et al. (2014), Table 11 after verifying that applying these corrections does not improve the accuracy of photometric redshifts output by BEAGLE.

We fit the broad-band photometry using BEAGLE (see Section 2.3 for details on the model). We use a delayed star formation history $\psi(t) \propto t\exp(-t/\tau_{\rm SFR})$, where $\tau_{\rm SFR}$ is the star formation timescale and t the age of the galaxy, taken to lie between 10^7 yr and the maximum time allowed since the onset of star formation at the galaxy redshift. This parameterization gives a star formation history that rises at early times and declines exponentially at later times. This star formation history is shown to better reproduce the colors and mass-to-light ratios of galaxies in the smoothed particle hydrodynamics (SPH) simulations of Simha et al. (2014) than the widely-used exponentially decreasing star-formation histories. Additionally, simulations have been shown to predict that high-redshift galaxies have rising star-formation histories (Finlator et al. 2011), a scenario that is naturally achieved using this parameterization. To ensure that galaxies are not fitted with models that are older than the age of the Universe, we set an upper limit of $z_{\rm form}^{\rm max} = 15$ to the redshift of onset of star formation. We further employ a weakly informative Gaussian prior on $\log t$, with mean at $\log(t/{\rm yr}) = 9.3$ and $\sigma = 0.7$.[7] We approximate the distributions of stellar and interstellar metallicities in a galaxy with a single metallicity $Z_{\rm ISM} = Z$. We use an exponential prior for $\hat{\tau}_V$ and fix $\mu = 0.4$. The free parameters in the model fitting are summarized in Table 2, and see Section 2.3 for a general overview of the individual parameters.

Table 2. Parameters allowed to vary in the BEAGLE fitting to galaxies in the 3D-HST catalog with their priors.

Parameter	Prior	Description
z	Uniform $\in [0, 15]$	redshift
$\log(M_{tot}/M_\odot)^a$	Uniform $\in [7, 13]$	Integrated SFH
$\log(t/{\rm yr})$	Gaussian $\mathcal{N}(9.3; 0.7)$ truncated $\in [7, 10.15]$	Age of oldest stars in the galaxy
$\log(\tau_{\rm SFR}/{\rm yr})$	Uniform $\in [7, 12]$	Timescale of star formation
$\log(Z/Z_\odot)$	Uniform $\in [-2.2, 0.24]$	Stellar (and interstellar) metallicity ($Z=Z_{\rm ISM}$)
$\hat{\tau}_V$	Exponential $\exp(-\hat{\tau}_V)$ truncated $\in [0, 4]$	V-band attenuation optical depth
μ	Fixed 0.4	Fraction of attenuation arising in the diffuse ISM
$\log U_S$	Uniform $\in [-4, -1]$	Effective gas ionization parameter
ξ_d	Fixed 0.3	Dust-to-metal mass ratio

[a] BEAGLE samples over the integral of the past star formation history of the galaxy (M_{tot}). It returns the stellar mass ($M_\star$), which accounts for the mass returned by evolved stars to the ISM.

3.4.2. *Parent SED catalog of galaxies based on 3D-HST catalog for $z \leq 4$, $\log(M_\star/M_\odot) > 8$*

To generate our parent SED catalog for high mass galaxies at $z < 4$, we use fits to the broad-band photometry of star-forming galaxies in the 3D-HST catalog using BEAGLE. Since T14 use rest-frame $U-V$ vs $V-J$ colors to separate galaxies into star-forming and quiescent galaxies before measuring the type-dependent mass functions, we select star-forming galaxies from the 3D-HST catalog using a similar star-forming/quiescent (SF/Q) classification scheme. Specifically, we select objects in the star-forming region of $U-V$ vs $V-J$ color space as defined by Whitaker et al. (2011) (see their Figure 17 and Equations 14 and 25) using the reported U, V and J-band absolute rest-frame magnitudes supplied in the 3D-HST catalog.

As described in Section 2.3, BEAGLE uses MULTINEST (Feroz et al. 2009) to sample the parameter space, and records the associated SEDs and MULTINEST weights. This information can be used to produce samples drawn from the corresponding posterior probability distribution, and we use these samples to populate a parent catalog with a range of statistically acceptable SED fits (plus associated physical parameters) for each object.

At low masses, constraints on the rest-frame UV, metallicity and dust attenuation can suffer from poor photometric constraints. For each of these parameters we therefore impose conditional priors on $M_{\rm UV}$–$M_\star$, $M_\star$–Z–ψ and ψ–Z–$\hat{\tau}_V$. These are additional priors to those already set in the SED-fitting (listed in Table 2) that we must apply after the fact, as BEAGLE does not accept conditional priors. Luckily it is rela-

[6] We choose not to fit using the KPNO U-band data in GN as the imaging is significantly shallower than the VLT/VIMOS U-band imaging in GS, and a large zeropoint offset was measured in Skelton et al. (2014)

[7] Since the galaxy age is only weakly constrained by broad-band data alone, the resulting stellar masses are sensitive to the choice of the age prior (e.g. Pacifici et al. 2015). Empirically, we find that adopting a uniform prior on $\log t$ overweights young ages, therefore leading to underestimated stellar masses with respect to those derived by Pacifici et al. (2015).

tively simple to apply these priors as one effectively needs to adjust the MULTINEST weights derived from the initial BEAGLE fitting. This method was presented in Chevallard et al. (2017) (see their Section 2.3, where they describe drawing SEDs that match the observed main sequence of star-forming galaxies). Essentially we perform weighted draws from the MULTINEST output for each galaxy fitted to in the 3D-HST catalog such that the probability of a galaxy entering into the parent catalog follows:

$$\begin{aligned} &\mathrm{P}(M_{\mathrm{UV}}, \mathrm{Z}, \hat{\tau}_{\mathrm{V}} \mid \mathrm{M}_\star, \psi, z) \propto \\ &\mathrm{P}(M_{\mathrm{UV}} \mid \mathrm{M}_\star, z)\,\mathrm{P}(\mathrm{Z} \mid \mathrm{M}_\star, \psi)\,\mathrm{P}(\hat{\tau}_{\mathrm{V}} \mid \mathrm{Z}, \psi)\,\mathrm{P}(\boldsymbol{\Theta} \mid \mathbf{D}) \end{aligned} \quad (14)$$

where $\mathrm{P}(\boldsymbol{\Theta} \mid \mathbf{D})$ is the posterior probability of a set of parameters sampled over in the fitting ($\boldsymbol{\Theta} = [z, \mathrm{M}_{tot}, t, \tau_{\mathrm{SFR}}, \mathrm{Z}, \hat{\tau}_{\mathrm{V}}, \mu, \log U_{\mathrm{S}}, \xi_{\mathrm{d}}]$) given the data, $\mathbf{D}$. The shape of the prior in M_{UV}–$\mathrm{M}_\star$ is set by the observational constraints detailed in Section 3.2, and for $\mathrm{P}(\mathrm{Z} \mid \mathrm{M}_\star, \psi)$ and $\mathrm{P}(\hat{\tau}_{\mathrm{V}} \mid \mathrm{Z}, \psi)$ we use the same priors imposed when constructing the parent catalog not based on 3D-HST photometry (as described in the following sub-section, specifically equations 15 - 16 describe the prior in $\mathrm{P}(\mathrm{Z} \mid \mathrm{M}_\star, \psi)$ and equations 19 - 22 describe the prior in $\mathrm{P}(\hat{\tau}_{\mathrm{V}} \mid \mathrm{Z}, \psi)$) [8] The prior on M_{UV}–$\mathrm{M}_\star$ ensures a smooth transition in observables between galaxies assigned from the 3D-HST catalog constraints and those at lower masses or higher redshift that lie outside the parameter space covered by the observed galaxies. For objects in the 3D-HST catalog with firm M_{UV} constraints the prior changes the sampling very little, allowing the final mock M_{UV}–$\mathrm{M}_\star$ relation to display the wide diversity of SEDs seen in the observed population.

When drawing SEDs from the fits to 3D-HST galaxies, we account for the varying sensitivity limits of the adopted datasets. The HUDF field and parallels provide the faintest H_{160} objects in the catalogs, while the sensitivity in CANDELS varies between the Deep and Wide regions. To avoid favoring SEDs from the numerous H_{160}-bright objects found in shallow, wider survey areas, we limit the shallower area to that in the CANDELS deep regions within GOODS-N and GOODS-S. Figure 11 displays the residual difference in depths between the deeper HUDF and parallel fields and the CANDELS deep regions of GOODS-S and GOODS-N (plotted together). The number counts at given total H_{160} magnitude start to turn over at $H_{160} \sim 27$ for the GOODS-N+GOODS-S region, while the number counts for the HUDF and parallels remain flat to $H_{160} \sim 27.5$. We apply a conservative cut of $H_{160} = 26.8$ above which, only objects in the deeper HUDF and parallels are included in the parent catalog

[8] Please note that, although we impose a prior on the $\mathrm{Z} - \log U_{\mathrm{S}}$ relation in the parent catalog *not* produced using SED fits to galaxies in the 3D-HST catalog, we do not apply the same prior here.

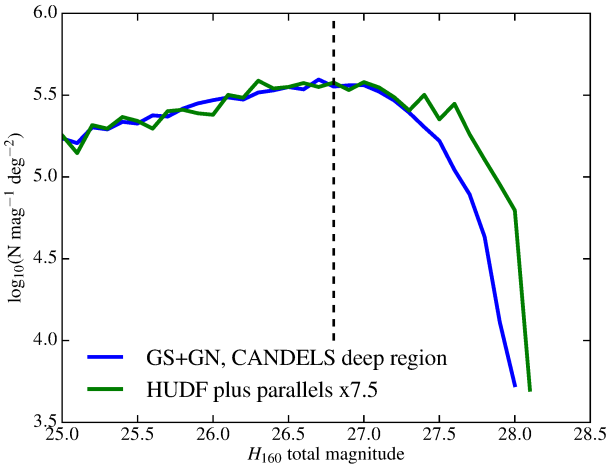

Figure 11. Number density of galaxies as a function of total H_{160} AB magnitude for the combined GOODS-S and GOODS-N CANDELS deep regions (blue line) and HUDF and parallel fields (green line). The number counts for the HUDF and parallel fields have been scaled by a factor of 7.5 to match the number counts of the GOODS-S+GOODS-N region (see text for details). The vertical dashed line shows the conservative limit we assign such that objects brighter than this limit are only included in the parent catalog if they are in the GOODS-S+GOODS-N region, and objects fainter than this limit are from the HUDF and parallel fields only.

and below which, only objects from GOODS-N+GOODS-S are included. We also require that the transition in number counts across this magnitude limit is smooth, ensuring that the distribution of H_{160} magnitudes at a given mass is not more heavily weighted by the more numerous objects in the shallower region. The ratio between cumulative number counts at $H_{160} < 26.7$ for the two regions is equal to 7.5, which essentially accounts for the difference in area between the two regions (this factor has already been applied to the number counts in the HUDF+parallels in the figure). We therefore draw 45 realizations per object from the HUDF+parallel regions, while only 5 random draws are included in the parent catalog for each object in the shallower region.

3.4.3. *Parent catalog for galaxies at $z \leq 4$ and with $\log(\mathrm{M}_\star/\mathrm{M}_\odot) < 8$, and for all $z > 4$ galaxies*

For mock galaxies outside of the stellar mass and redshift range of 3D-HST galaxies, we generate a parent catalog by using BEAGLE to produce SEDs covering a wide range of parameter values (e.g., z, $\mathrm{M}_\star$, τ, $\hat{\tau}_{\mathrm{V}}$, $\log U_{\mathrm{S}}$, Z, and t). We impose constraints to avoid populating the parent catalog with objects in unphysical regions of parameter space, such as galaxies with high metallicity and high $\log U_{\mathrm{S}}$ or high-mass galaxies with extremely small metallicities. To constrain

these parameters, we utilize the distributions of $M_\star - Z - \psi$ and $Z - \log U_S$ inferred from observations.

We use delayed SFHs to produce the SEDs, where the τ and t values are chosen to ensure that the galaxies would be classified as star-forming, with $\log(\psi_S/yr^{-1}) > -10$. Here we allow $\log(t/yr)$ to vary between 6 and the age of the Universe. The lower limit in age is lower than that introduced in the prior on $\log(t/yr)$ used in the fitting to objects in the 3D-HST catalog (see Table 2) as this parent catalog is going to be used to match to lower mass/higher redshift objects. BEAGLE provides the stellar masses accounting for mass returned to the ISM as stars evolve and die, as well as the current SFR which can be used to assign metallicity, ionisation parameter and V-band optical depth due to attenuation by dust.

To constrain the metallicities of galaxies, we use the fundamental metallicity relation between $M_\star - Z - \psi$ measured by Hunt et al. (2016) from a compilation of ~ 1000 galaxies covering a wide range in ψ and stellar mass, with oxygen abundance estimates derived from consistent calibrations, and, crucially, covering a wide range in redshifts up to $z \sim 3.7$. The redshift distribution of Hunt et al. (2016) can be viewed in their Figure 1, which can be compared with the fundamental metallicity relation measured in Mannucci et al. (2010) from galaxies in the Sloan Digital Sky Survey for objects with redshifts between 0.07 and 0.3. A fit to this fundamental metallicity relation is given by

$$12+\log(O/H) = -0.14\log(\psi/M_\odot\,yr^{-1}) + 0.37\log(M/M_\odot) + 4.82. \quad (15)$$

This relation is a fit to the gas-phase oxygen abundance, while we require a relation based on the nebular metallicity $Z_{ISM} = Z$ associated with our models. From the grid of $\xi_d = 0.3$ models used here, we infer the approximate relation $12+\log(O/H) \simeq \log(Z_{ISM}/Z_\odot)+8.7$. While this approximation is not suitable to determine accurately $12+\log(O/H)$ from the output metallicity for individual nebular models (see Gutkin et al. 2016, Table 2 for values of $12+\log(O/H)$ for different model metallicities), the errors it generates are much smaller than the scatter we introduce below in the fundamental metallicity relation.

It is important to highlight here that the Hunt et al. (2016) relation only models a linear dependence between oxygen abundance and stellar mass, whereas we know the mass-metallicity relation is *not* linear at high stellar masses (e.g. Tremonti et al. 2004). In fact the upper limit in metallicity of $Z/Z_\odot = 0.24$ (also the upper limit of the prior in metallicity employed when fitting to objects in the 3DHST catalogs) introduces a turnover in the catalog mass-metallicity relation at low redshifts and high masses (see Section 6.4).

We wish to instill a broad diversity in our parent catalog SEDs and avoid an over-representation of unphysical parameters in our resulting mock catalogs. We therefore apply a broad scatter to the fundamental metallicity relation, and do not attempt to predict the form of the $M_\star - Z$ relation or $M_\star - Z - \psi$ plane to high redshifts. The broad range of spectral parameters will enable investigations of selection effects in future observations, especially for redshift and magnitude regimes where current measurements cannot yet reach. We characterize the scatter with a Student's-t distribution:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (16)$$

where ν is the number of degrees of freedom and Γ is the gamma function and:

$$x = \frac{\log(Z_{ISM}/Z_\odot) - \log(\bar{Z}_{ISM}/Z_\odot)}{\sigma_x}. \quad (17)$$

We set $\sigma_x = 0.3$ and $\nu = 3$, where $\nu = 3$ has been chosen to provide a distribution with more weight in the tails compared to a Gaussian.

To constrain the ionization parameter of these galaxies we use a linear fit between metallicity and $\log U_S$ measurements at low redshift from Carton et al. (2017, see their figure 2):

$$\log U_S(Z) = -0.8\log(Z/Z_\odot) - 3.58. \quad (18)$$

We again use the Student's-t distribution with 3 degrees of freedom to introduce scatter in this relation.

We account for dust attenuation by using an approach commonly featured in semi-analytic models of galaxy formation (e.g. Guiderdoni & Rocca-Volmerange 1987; De Lucia & Blaizot 2007; Fontanot et al. 2009). Following Devriendt et al. (1999, their equation 6), we estimate the V-band, face-on attenuation optical depth $\hat{\tau}_{V\perp}$ using the relation

$$\hat{\tau}_{V\perp} = \left(\frac{Z_{ISM}}{Z_\odot}\right)^{1.6}\left(\frac{N_H}{2.1\times10^{21}\,cm^{-2}}\right) \quad (19)$$

where Z_{ISM} is the interstellar metallicity and N_H the mean hydrogen column density. As in Devriendt et al. (1999), we compute N_H from the (cold) gas fraction

$$N_H = 6.8\times10^{21}\frac{M_{gas}}{M_\star + M_{gas}}, \quad (20)$$

where the cold gas mass M_{gas} is computed by inverting the Schmidt-Kennicutt relation (Schmidt 1959; Kennicutt 1998). In practice, we consider the size and SFR of each galaxy to compute the SFR density $\Sigma_\psi = \psi/(\pi r^2)$, in units of $M_\odot\,yr^{-1}\,pc^{-2}$, where r is the galaxy effective radius. We can then compute the cold gas surface density Σ_{gas} from the Schmidt-Kennicutt relation, and estimate the cold gas mass as $M_{gas} = \Sigma_{gas}\,\pi\,r^2$.

Equation (19) provides us with the face-on attenuation optical depth, from which we can derive an angle-averaged attenuation optical depth by assuming a spatial distribution of

dust and stars. Following Devriendt et al. (1999), we approximate our galaxies as oblate ellipsoids where dust and stars are homogeneously mixed. The V-band attenuation optical depth averaged over all galaxy inclinations i can then be written as

$$\langle \hat{\tau}_{\rm V} \rangle_i = -2.5 \log \left(\frac{a_{\rm V}}{1-\omega_{\rm V}+\omega_{\rm V} a_{\rm V}} \right) \frac{1}{1.086}, \qquad (21)$$

where $\omega_{\rm V} = 0.87$ is the albedo at 5500 Å for dust grains with properties as in the Small Magellanic Cloud (Pei 1992), and $a_{\rm V}$ is computed as

$$a_{\rm V} = \frac{3}{4\,\hat{\tau}'_{{\rm V}\perp}} \left[1 - \frac{1}{2\,\hat{\tau}'^2_{{\rm V}\perp}} + \left(\frac{1}{\hat{\tau}'_{{\rm V}\perp}} + \frac{1}{2\,\hat{\tau}'^2_{{\rm V}\perp}} \right) \exp(-2\,\hat{\tau}'_{{\rm V}\perp}) \right],$$

where $\hat{\tau}'_{{\rm V}\perp} = 2.62\,\hat{\tau}_{{\rm V}\perp}$.

Equations (19)–(21) enable us to associate to each mock galaxy a physically-motivated value for the angle-averaged V-band attenuation optical depth, which depends on the galaxy SFR, size, and metallicity. We then account for the effect of galaxy inclination on the V-band attenuation optical depth $\hat{\tau}_{\rm V}$ by randomly drawing $\hat{\tau}_{\rm V}$ from a Gaussian distribution centered at $\langle \hat{\tau}_{\rm V} \rangle_i$ and truncated at $\hat{\tau}_{\rm V} = 0$. The width of the Gaussian used to draw the $\hat{\tau}_{\rm V}$ value is chosen to be dependent on $\langle \hat{\tau}_{\rm V} \rangle_i$ and Z, according to:

$$\sigma = 0.2 + {\rm Z}/{\rm Z}_\odot - \langle \hat{\tau}_{\rm V} \rangle_i \qquad (22)$$

This function ensures that at low metallicities, there is a smaller scatter in $\hat{\tau}_{\rm V}$, limiting range of dust attenuation in the regime where we expect low dust-to-gas ratios, while the minimum value of 0.2 prevents the values of $\hat{\tau}_{\rm V}$ being too constrained at the lowest metallicities. We note that non-negligible attenuation by dust even at very low metallicities is not unreasonable, as even if they have low gas-to-dust ratios they may be gas-rich, allowing for non-negligible dust-to-stellar mass ratios (e.g. da Cunha et al. 2010). Our choice of the negative dependence of σ on $\langle \hat{\tau}_{\rm V} \rangle_i$ mimics results obtained from radiative transfer calculations of dust attenuation in galaxies (e.g. Tuffs et al. 2004; Pierini et al. 2004), which show that galaxies with low angle-averaged attenuation optical depths $\langle \hat{\tau}_{\rm V} \rangle_i \lesssim 0.1$ exhibit a larger fractional range of inclination-dependent attenuations than galaxies with larger $\langle \hat{\tau}_{\rm V} \rangle_i$.

3.4.4. *Matching mock galaxies to the parent catalog*

To assign SEDs from the parent catalog to mock galaxies, we find the closest match between the mock galaxy parameters and the physical parameters of the parent catalog galaxies. Which properties are matched differ when assigning SEDs drawn from fits to 3D-HST galaxies or from the wide grid of galaxies produced by BEAGLE. For those objects with $z < 4$ and stellar mass higher than $\log({\rm M}/{\rm M}_\odot) = 8$ or the mass completeness limit of the 3D-HST catalog ($\log({\rm M}/{\rm M}_\odot) > 0.7z + 0.63$)[9], red vertical lines shown in figure 6), we find the closest match in $\rm M_\star$ and redshift, as we rely on observed broad-band photometry to constrain the expected SED shape of an object at a given stellar mass. For objects at $z > 4$ or with lower stellar masses at $z \leq 4$, the expected SED shape is based on extrapolations of the $M_{\rm UV}$–$\rm M_\star$ and β–$M_{\rm UV}$ relations, as described in Sections 3.2 and 3.3, respectively. These relations are used to assign $M_{\rm UV}$ and β values to each galaxy in the star-forming galaxy mock catalog. Each mock object is then assigned an SED based on the closest match in redshift, stellar mass, $M_{\rm UV}$ and UV-slope in the parent catalog.

For all matches between mock galaxies and parent catalog it is then possible to shift the redshift of the SED to the exact redshift of the mock galaxy. Figure 12 displays the distributions of physical parameters assigned to all star-forming galaxies in the mock catalog.

4. GENERATING QUIESCENT GALAXIES ACROSS COSMIC TIME

4.1. *Quiescent galaxy counts*

Our model for the redshift evolution of the stellar mass function of quiescent galaxies is based on observed stellar mass functions that have been measured in the redshift range $0.2 \leq z \leq 3.0$ by T14. These authors use the redshift-dependent UVJ color selection from Whitaker et al. (2011) to select quiescent galaxies from the ZFOURGE medium-band photometric survey (Straatman et al. 2016) and data from the CANDELS survey (see Section 3.1). T14 find that the quiescent galaxy stellar mass function is best fitted by a double-Schechter function (equation 4 above) in the redshift range $0.2 \leq z \leq 1.5$, and by a single-Schechter function at higher redshifts, $1.5 \leq z \leq 3.0$. The differing functional forms were chosen to match the observed upturn in the stellar mass function below $\log({\rm M}/{\rm M}_\odot) \leq 9.5$ at $z \leq 1$, a result in line with observations by Santini et al. (2012); Muzzin et al. (2013); Ilbert et al. (2013).

In this work, we use both the observed binned stellar mass functions and the fitted Schechter parameters in bins of redshift from T14 to construct a continuous model for the redshift evolution of the quiescent stellar mass function. To produce smooth evolution at all redshifts, we choose to adopt the double-Schechter function description for the mass function at all redshifts, even at $z > 1.5$ where observations find consistency with a single Schechter function. This double-

[9] The approximate mass completeness limits are estimated in bins of redshift and fit with a linear relation. Specifically, we randomly sample 100 SEDs from the posterior probability distribution of each galaxy in the HUDF portion of the 3D-HST catalog. These SEDs are binned in redshift and mass and the completeness limit calculated as the mass at which 95% of the SEDs are brighter than 27.6 in H_{160} (the magnitude at which the number counts in the UDF portion of the field start to turn over, see Figure 11).

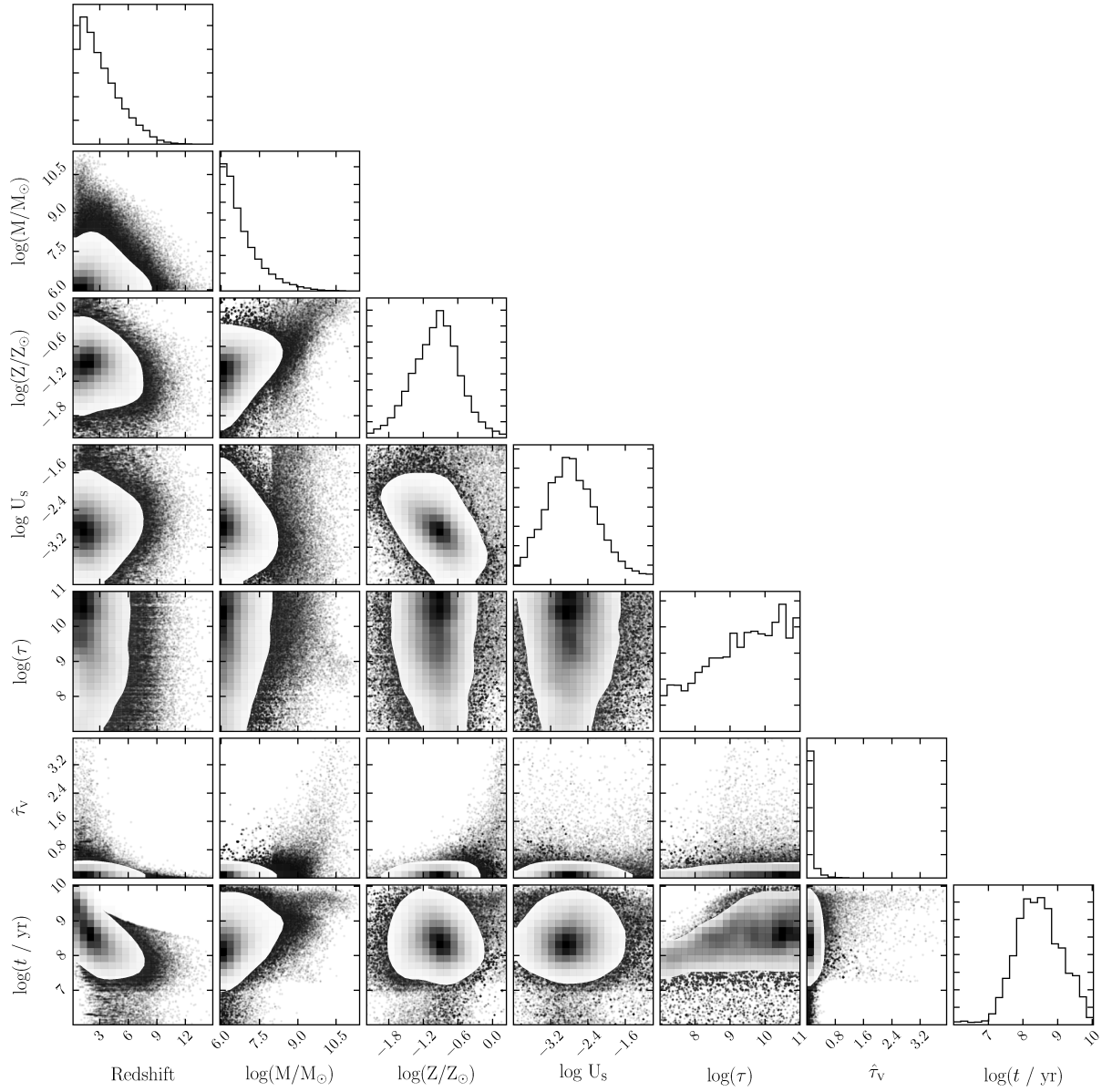

Figure 12. The parameters assigned to all star-forming galaxies in the mock catalog. Where the point density becomes high, a two-dimensional histogram indicates the density of points by the level of shading. The plot was made using the python package *corner* (Foreman-Mackey 2016)

.

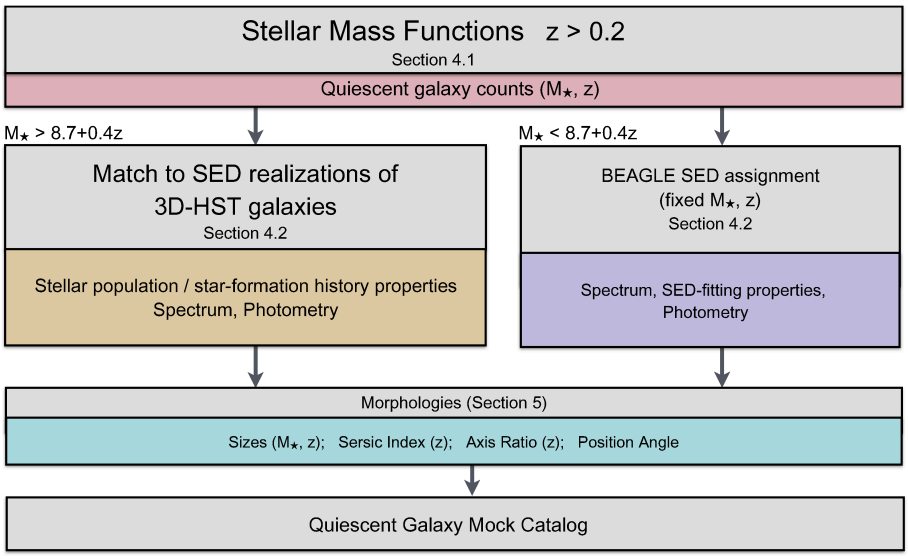

Figure 13. Diagram summarizing the procedures for generating the quiescent galaxies. $M_\star$ is defined as $\log(M/M_\odot)$. Gray boxes indicate the empirical relationships, distributions, or data on which mock galaxy properties are based, and colored boxes indicate the mock galaxy property which is generated in that step. All quiescent mock galaxies are generated following these procedures.

Schechter function has five parameters, M^*_M, $\alpha_{1,M}$, $\phi^*_{1,M}$, $\alpha_{2,M}$, and $\phi^*_{2,M}$. We fit the double-Schechter parameters from T14 at $z < 1.5$, but substituted the single-Schechter parameters (M^*, α, and ϕ^*) for the double-Schechter parameters (M^*_M, $\alpha_{1,M}$, $\phi^*_{1,M}$) at $z > 1.5$. In a double-Schechter function, $\alpha_{2,M}$ and $\phi^*_{2,M}$ control the slope and overall normalization at low-masses.

We choose to fix M^*_M to a value of $\log(M/M_\odot) \sim 10.6$ at all redshifts, owing to the lack of significant shift in the observed evolution in M^*_M with redshift from the T14 observations, and we fit the $\alpha_{1,M}$ and $\phi^*_{1,M}$ evolution with a cubic function with redshift. We fit a linear function to the $\alpha_{2,M}$ evolution and a quadratic function to the $\phi^*_{2,M}$ evolution.

Given the chosen forms for the evolution in the parameters, in order to prevent the cubic functions from diverging at low-redshift and high-redshift we stop the evolution for some of the parameters. For $\alpha_{1,M}$, $\phi^*_{1,M}$, $\alpha_{2,M}$, and $\phi^*_{2,M}$ we stop the redshift evolution at $z < 0.5$, $z < 0.75$, $z < 0.5$, and $z < 0.5$, respectively. We require $\phi^*_{1,M}$ to stay constant below $z < 0.75$ because the cubic fit to the T14 Schechter parameters would otherwise over-predict by an order of magnitude the number of objects as compared to observations at low redshift. We fix $\alpha_{1,M}$ at $z > 1.75$ to the value of the cubic fit at that redshift, and at $z > 1.88$ we impose a linear decline on the evolution of $\phi^*_{1,M}$ to extrapolate the quiescent galaxy counts to $z > 3$, beyond the range probed by the observations. This projected decline in $\phi^*_{1,M}$ predicts a number density of $2.16 \pm 0.8 \times 10^{-5}$ Mpc^{-3} quiescent galaxies at $3.4 \leq z \leq 4.2$, consistent with the number density of quiescent galaxies identified in Straatman et al. (2014, 2015, $1.8 \pm 0.8 \times 10^{-5}$ Mpc^{-3}) to their stellar mass limit of $\log(M/M_\odot) \sim 10.6$ and the ZFOURGE survey area of $\sim$363 arcmin^2. Thus our extrapolation is in broad agreement with the few constraints available on counts of quiescent galaxies at $z > 3.5$.

The resulting quiescent galaxy stellar mass function is compared with the T14 data in Figure 14, and we provide the detailed functional fit parameters in Table 3. Our model for the stellar mass function evolution broadly agrees within the uncertainties of the binned stellar mass function observations from T14. We note that the observations imply more low-mass [$\log(M/M_\odot) \leq 9.5$] quiescent galaxies at $0.5 \leq z \leq 0.75$ than below $z \leq 0.5$. The continuously varying stellar mass function evolution model we describe here monotonically increases in counts with decreasing redshift, and therefore it does not replicate this rise and fall of the low-mass end that is seen in the observations (Figure 14).

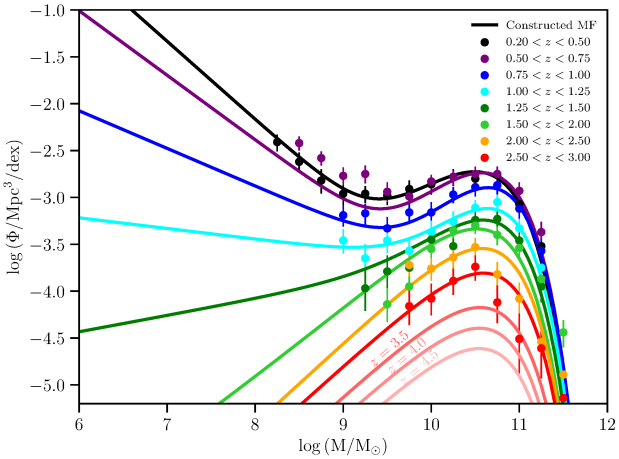

Figure 14. Evolution of the quiescent galaxy stellar mass function (dashed), plotted with the observations from T14 (circles). The parameters of this fit to the MF evolution are given in Table 3.

4.2. *Quiescent galaxy SEDs*

We apply the same SED assignment technique for quiescent galaxies as that for star-forming galaxies, described in Section 3.4, whereby we produce a large parent catalog of SEDs and associated physical parameters that can then be used to associate an SED with each mock galaxy. Where possible, we use observed galaxies to guide the allowed range of mock galaxy SEDs at a given stellar mass. Outside of the parameter space covered by 3D-HST objects we produce theoretical mock SEDs using BEAGLE.

To produce a parent catalog from observed objects, we use BEAGLE fits to 3D-HST galaxies classified as quiescent using the rest-frame U, V and J band absolute magnitudes and the Whitaker et al. (2011) $U-V$ vs. $V-J$ color criteria. Following the method described in Section 3.4.2, we draw SEDs (and associated physical parameters) from the BEAGLE fits to produce the parent catalog. Specifically, the stellar mass and associated 68% central credible interval are estimated for each galaxy, and the SEDs are randomly drawn from the BEAGLE output files if their stellar mass lies within this interval, without further weighting based on other physical parameters.

Uncertainties associated with the star-forming/quiescent separation have to be dealt with carefully when producing this parent catalog. More details are given in Appendix B, where we define a redshift-dependent H_{160} magnitude cut of $H_{160} < 24.5 + z$, brighter than which quiescent objects from the 3D-HST catalog can be used to provide SEDs for the parent catalog. This limit lies well above the sensitivity limit of the CANDELS deep region of the 3D-HST catalog (and so we do not encounter the situation we found for the star-forming galaxies where the faintest objects were only drawn from the smaller area HUDF portion of the catalog). We therefore skip the areal correction described in Section 3.4.2, and populate the 3D-HST parent catalog of quiescent galaxies with 5 random draws from the BEAGLE fit output for each quiescent galaxy in the 3D-HST catalog satisfying the

Table 3. Quiescent Mass Function double-Schechter parameters (Figure 14) and their evolution. The parameter is shown on the left column, the functional form of the parameter (with associated constants) on the right column, and the redshift-range of that functional form in the middle column.

Parameter	Redshift range	Functional Form
M^*_M	$z \geq 0.2$	$M^*_M = 10.617$
$\alpha_{1,M}$	$z < 0.5$	$\alpha_{1,M} = -0.225$
	$0.5 \leq z < 1.75$	$\alpha_{1,M} = b_0 \times z^3 + b_1 \times z^2 + b_2 \times z + b_3$ $b_0 = 0.43$, $b_1 = -2.33$, $b_2 = 3.49$, $b_3 = -1.44$
	$z \geq 1.75$	$\alpha_{1,M} = -0.150$
$\phi^*_{1,M}$	$z < 0.75$	$\log(\phi^*_{1,M}) = -2.67$
	$0.75 \leq z < 1.877$	$\log(\phi^*_{1,M}) = c_0 \times z^3 + c_1 \times z^2 + c_2 \times z + c_3$ $c_0 = -0.35$, $c_1 = 1.92$, $c_2 = -3.77$, $c_3 = -0.78$
	$z \geq 1.877$	$\log(\phi^*_{1,M}) = c_4 \times z + c_5$ $c_4 = - = 0.43$, $c_5 = -2.59$
$\alpha_{2,M}$	$z < 0.5$	$\alpha_{2,M} = -1.83$
	$z \geq 0.5$	$\alpha_{2,M} = d_0 \times z + d_1$ $d_0 = 1.15$, $d_1 = -2.41$
$\phi^*_{2,M}$	$z < 0.5$	$\log(\phi^*_{2,M}) = -4.71$
	$z \geq 0.5$	$\log(\phi^*_{2,M}) = e_0 \times z^2 + e_1 \times z + e_2$ $e_0 = -0.59$, $e_1 = 1.93$, $e_2 = -5.52$

$H_{160} < 24.5 + z$ criterion. The H_{160} limit translates to an approximate mass limit of $\log(M/M_\odot) > 8.7 + 0.4z$. Thus only objects in the mock catalog with $\log(M/M_\odot) > 8.7 + 0.4z$ are paired with the closest match in redshift and stellar mass among parent catalog galaxies.

For objects with $\log(M/M_\odot) < 8.7 + 0.4z$, we produce a parent catalog of SEDs and associated physical parameters using BEAGLE. For quiescent galaxies, we do not need to assign nebular H II-region parameters (e.g. ξ_d, $\log U_s$), and we neglect dust attenuation. We therefore vary only the galaxy age t, star formation timescale τ_{SFR}, and metallicity Z to generate SEDs for the parent galaxy catalog. Using a delayed SFH (Section 3.4.1) with $\log(\tau_{SFR}) < 1.11 \times \log(t) - 2.02$ ensures that the specific star formation rate of objects be less than $\log(\psi_s/yr^{-1}) = -10$. The parameters t and τ_{SFR} are assigned to each mock catalog galaxy from uniform distributions. The parameter t is allowed to vary between 30 Myr and the age of the Universe at the redshift of the object. We allow τ_{SFR} to vary between 10 Myr and the maximum value required to produce $\log(\psi_s/yr^{-1}) < -10$.

Measuring the stellar metallicities of quiescent galaxies and their evolution is technically challenging, requiring deep rest-frame optical spectra to measure stellar absorption-line indices. (*JWST* will provide new opportunities to probe stellar metallicities of quiescent galaxies at high-redshift). However, with existing data, most stellar metallicity measurements for quiescent galaxies exist only for high mass field galaxies ($\log(M/M_\odot) \gtrsim 9.5$) to moderate redshifts (e.g. Gallazzi et al. 2006, $z \sim 0.1$ and Gallazzi et al. 2014, $z \sim 0.7$) or cluster galaxies (e.g. Sánchez-Blázquez et al. 2009, Jørgensen & Chiboucas 2013). The highest mass galaxies in our mock have their physical properties assigned from SED fits to 3D-HST galaxies, and so, with the lack of current constraints on the stellar metallicities of low-mass and high-redshift quiescent galaxies, we assign metallicities with a uniform weight between the limits of our templates ($-2.2 < \log(Z/Z_\odot) < 0.24$).

After each galaxy in the mock catalog is assigned t, τ_{SFR}, and Z values, BEAGLE computes the fraction of mass $1 - M_\star/M_{tot}$ returned by evolved stars to the ISM, and hence the scaling required to generate an SED with the corresponding stellar mass assigned to the mock catalog object. BEAGLE then generates the SEDs in "mock mode" using the assigned parameters. As with the star-forming catalog, the SEDs assigned to each of the quiescent mock catalog galaxies are used to generate NIRCam filter fluxes.

5. GALAXY MORPHOLOGIES FROM $0.2 \leq Z \leq 15$

The evolution of galaxy morphologies with cosmic time represents one of many key insights into the galaxy formation process that *JWST* will provide. Perhaps more practically, galaxy shapes and light distributions affect detectability and measurability of other galaxy properties, and fully understanding these systematics in future *JWST* data will be

important for characterizing uncertainties. Anticipated applications of this mock catalog include *JWST* NIRCam image and NIRSpec spectroscopic simulations, as well as NIRSpec MSA slit assignment. Therefore, we assign simple morphologies to mock galaxies to enable these types of analyses. All mock galaxies are modeled as simple Sersic profiles (Sersic 1968), and follow the redshift evolution of the relevant morphological parameters that has been characterized using deep extragalactic surveys with *HST*. In the following sections, we describe our method for producing continuous evolutionary models and realistic distributions for galaxy sizes, shapes, and light profiles. Below, we describe the procedure for assigning half-light radii to both star-forming and quiescent galaxies at $z \leq 4$ where WFC3 has provided accurate rest-frame optical morphologies, and at $z > 4$ where HST has characterized rest-frame UV morphologies, and shapes, light profiles, and orientations for star-forming galaxies.

5.1. *Galaxy Sizes at $z < 4$*

We aim to generate a continuously evolving model in stellar mass, UV luminosity, and redshift, using the observed size-mass relationships that have been measured for galaxies at $0 \leq z \leq 4$. For this purpose, we use the relationships measured in van der Wel et al. (2014) for both star-forming and quiescent galaxies using CANDELS data, and extrapolate their behavior down to $\log(M/M_\odot) \sim 6$. These relationships have also been shown to agree with the measured size-mass relation of local galaxies in SDSS (Shen et al. 2003; Guo et al. 2009).

van der Wel et al. (2014) parametrize the redshift evolution of the half-light semi-major radius in kpc, $R_{eff,maj}$, as a power-law function of the Hubble parameter at a given redshift, H(z), in bins of stellar mass. The parametrization has the form

$$R_{\rm eff,maj}(z) = B_H \left(\frac{H(z)}{H_0}\right)^{\beta_H} \tag{23}$$

where both B_H and β_H (respectively, the coefficient and power-law slope of the redshift evolution) vary with stellar mass. To generate a smoothly evolving model, we generalize this evolution of $R_{eff,maj}(z)$ between stellar mass bins by fitting both B_H and β_H as functions of stellar mass, to produce one smooth function in both redshift and stellar mass.

The behavior of B_H and β_H with stellar mass differs between star-forming galaxies and quiescent, and therefore we model the stellar mass dependence differently between the two samples. For star-forming galaxies, both B_H and β_H appear linear with $M_\star = \log(M/M_\odot)$. The best-fitting linear relationships are

$$\begin{aligned} B_H(M_\star) &= 0.23M_\star - 1.61 \\ \beta_H(M_\star) &= -0.08M_\star + 0.25, \end{aligned} \tag{24}$$

and the shape of the resulting function $R_{eff,maj}(z, M_\star)$ for star-forming galaxies is shown in the left panel of Figure 15. We extrapolate the relationship out to $z \sim 4$, and assign sizes to all star-forming galaxies in this redshift range using this method.

For quiescent galaxies, the behavior of both B_H and β_H with stellar mass are not linear. We find that B_H ($M_\star$) is well described by either a quadratic or exponentially declining function of mass, however we choose to parametrize B_H using the exponential to avoid the undesirable quadratic feature that galaxy size increases unphysically to low masses. We find that β_H is well fit by an exponentially increasing function with decreasing stellar mass; however, increasingly large values of this power-law exponent at low masses produce unphysical size evolution at low mass. Therefore, we fix the value of β_H below $\log(M/M_\odot) < 9.75$. The relationships we use for quiescent galaxies are:

$$\begin{aligned} B_H(M_\star) &= 3.8\text{e}{-4}e^{0.71M_\star} - 0.11 && \text{(at all masses)} \\ \beta_H(M_\star) &= 1.38\text{e}12e^{-2.87M_\star} - 1.21 && (M_\star \geq 9.75) \\ \beta_H(M_\star) &= -0.19 && (M_\star < 9.75). \end{aligned} \tag{25}$$

The resulting function $R_{eff,maj}(z, M_\star)$ for quiescent galaxies is shown in Figure 15. The significant size evolution among massive quiescent galaxies ($\log(M/M_\odot) \gtrsim 10$) is in agreement with other studies (e.g. Cassata et al. 2013). The flattening of the quiescent galaxy size-mass relation evolution at lower stellar mass is consistent with the expectations of environmental effects due to satellite quenching (e.g. van der Wel et al. 2010; Kawinwanichakij et al. 2017) and the observations that quiescent galaxies in high-density regions typically have larger sizes (e.g. Cooper et al. 2012; Delaye et al. 2014).

As we have outlined in Section 4, our model extrapolates the evolution of the stellar mass function for quiescent galaxies at $z > 4$, and therefore mock catalogs will contain samples of quiescent galaxies at redshifts beyond where current surveys can identify them or measure their morphologies. Although the expected number of quiescent galaxies at $z > 4$ in mock surveys would be small, we estimate that mock catalogs of comparable area to one GOODS field ($\sim$150 arcmin^2) will contain quiescent galaxies out to $z \sim 6$. Therefore we note here that sizes for such objects come from an extrapolation of the relationship presented in Equation 25 which can be calculated for arbitrarily large redshift. We show the extrapolation out to $z \sim 6$ in the right panel of Figure 15.

The data presented in van der Wel et al. (2014) indicate that the scatter in galaxy sizes within quiescent and star-forming galaxy samples is approximately uniform in both redshift and stellar mass. The size distributions within redshift and mass bins for quiescent galaxies are log-normal with $\sigma = 0.16$ dex, whereas the sizes of star-forming galaxies follow a skewed distribution better fit by a Gauss-Hermite polynomial ex-

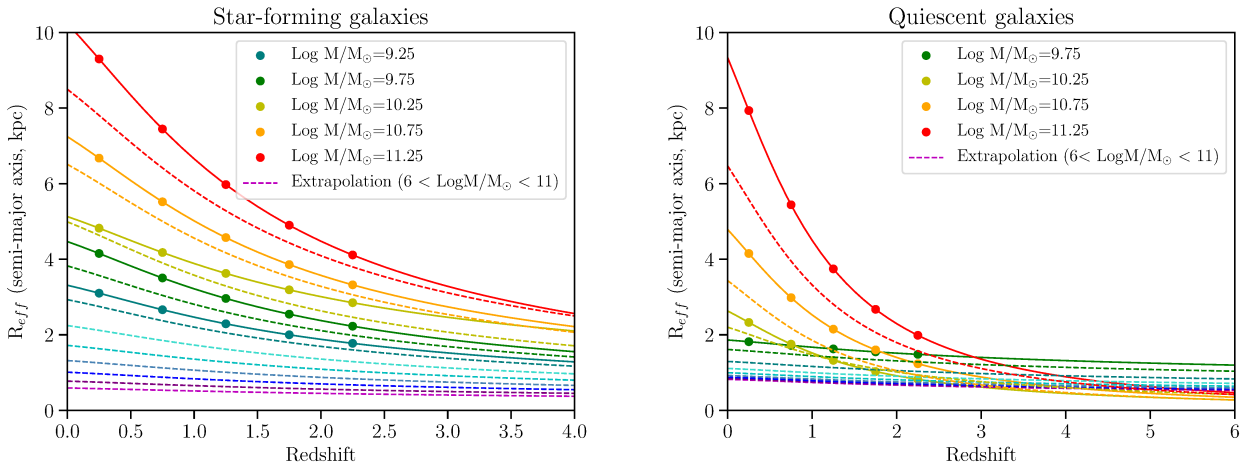

Figure 15. Model describing the variation of average $R_{cff,maj}$ for star-forming galaxies (left) and quiescent galaxies (right) as a function of both stellar mass and redshift. Points with solid lines indicate the observations in stellar mass bins and best-fit redshift evolution as published in van der Wel et al. (2014). Dashed lines illustrate the behavior of our continuously varying model with stellar mass, as defined in equations 24 and 25. Dashed lines go from $\log(M/M_\odot) = 6$ (magenta) to $\log(M/M_\odot) = 11$ (red) in increments of $\Delta\log(M/M_\odot) = 0.5$.

pansion in $\log_{10}R_{eff,maj}$ (van der Marel & Franx 1993) with fixed dispersion of 0.18 dex and skewness $h3 = -0.15$. The skewed distribution is motivated by the observation that the star-forming galaxy size distribution has a tail of small-sized galaxies (e.g. Williams et al. 2014, 2015; van Dokkum et al. 2015). The distribution has the form

$$P(R_{mock}|M*,z) \propto e^{-\frac{1}{2}y^2(1+h3(\frac{1}{\sqrt{6}}(2\sqrt{2}y^3-3\sqrt{(2)}y)} \qquad y = \frac{1}{\sigma}(\log_{10}R_{mock} - \log_{10}R_{eff,maj}). \tag{26}$$

To assign sizes to mock galaxies of each type, we draw a size R_{mock} from random distributions of these forms, where $R_{eff,maj}$ is the median size in kpc for each mock galaxy's redshift and stellar mass from equations 24 and 25 (Figure 15).

5.2. *Star-forming galaxy sizes at $z > 4$*

At $z > 3.5$ the WFC3 H-band moves out of the rest-frame optical. At higher redshifts, the current best imaging data set for morphological measures, the CANDELS H-band imaging, will then become a probe of the rest-frame UV morphology of high-redshift galaxies. For mock star-forming galaxies at $z > 4$ we therefore assign morphologies according to the redshift evolution of the observed M_{UV}-size relationships presented in (Shibuya et al. 2015). (As mentioned in the last section, quiescent galaxies at $z > 4$ follow the extrapolated relation presented in Equation 25). The size evolution in that work is parametrized by circularized half-light radius, defined as

$$R_{eff,circ} = R_{eff,maj}\sqrt{b/a} \tag{27}$$

where b/a is the axis ratio (ratio of the semi-minor to semi-major axis size). Shibuya et al. (2015) find that median $R_{eff,circ}$ of galaxies at $z > 4$ is correlated with UV luminosity, and at fixed UV luminosity, galaxy size significantly decreases with increasing redshift. Sizes at $z > 4$ in Shibuya et al. (2015) are measured from imaging at $1500 < \lambda_{rest} < 3000$Å. We generate sizes according to their size parameterizations with UV luminosity:

$$R_{eff,circ} = r_o\left(\frac{L_{UV}}{L_o}\right)^{0.27} \tag{28}$$

where $r_o = 6.9(1+z)^{-1.2}$ represents the effective radius in kpc at a characteristic UV luminosity L_o (corresponding to $M_{UV} = -21$). This relation is generally consistent with other measurements at $z \sim 6-7$ (Ono et al. 2013; Grazian et al. 2012; Oesch et al. 2010; Curtis-Lake et al. 2016; Bowler et al. 2017; Kawamata et al. 2015; Holwerda et al. 2015). The observed distribution of sizes is log-normal at all redshifts, with a mean size according to the above parametrization and a constant scatter with redshift and luminosity, $\sigma_{lnR_{eff,circ}} = 0.5$.

5.3. *Axis Ratios, Sersic Indices, and position angles across cosmic time*

In addition to sizes, galaxy shapes are characterized by the following additional properties: projected axis ratio (semi-minor half-light size / semi-major half-light size; b/a), Sersic index n (setting the overall concentration of the light profile), and position angle on the sky. The distribution of observed axis ratios and Sersic indices are known to correlate with star-formation (e.g. Franx et al. 2008; Bell et al. 2012; Mortlock et al. 2013). Star-forming galaxies often exhibit extended, disk-like light profiles characterized by $n < 2$ and lower values of b/a, whereas quiescent galaxies exhibit centrally concentrated light profiles with $n > 2$ and higher b/a. These properties are known to evolve within each star-formation

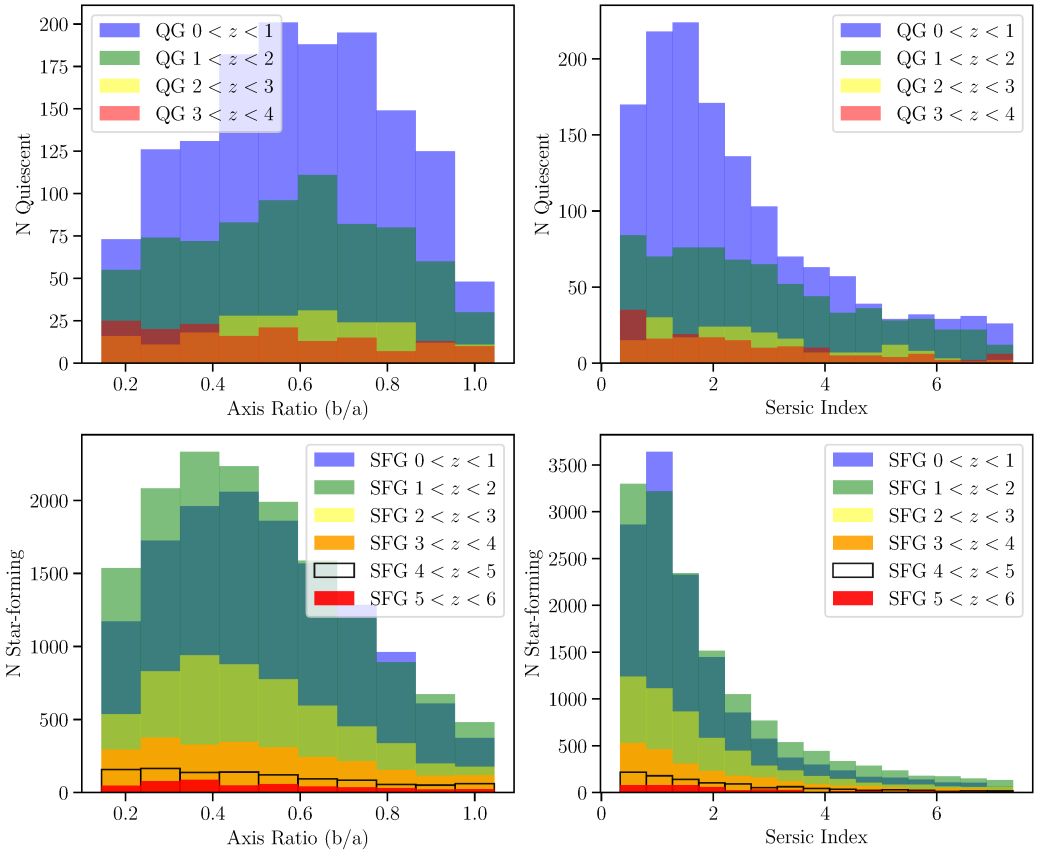

Figure 16. Observed distributions of axis ratios (left panels) and Sersic index (right panels) for observed 3D-HST galaxies as measured by van der Wel et al. (2012). Only galaxies with robust measurements ("flag = 0") are plotted. Top row shows the distributions for quiescent galaxies split by redshift, and bottom panels show distributions for star-forming galaxies split by redshift. The distributions are used to randomly assign mock galaxies axis ratio and sersic indices as a function of star-forming class and redshift.

sub-class, in that the n and b/a of quiescent galaxies tend to decrease with increasing redshift (e.g. van der Wel et al. 2011), while at high-redshift star-forming galaxies exhibit more clumpy concentrated morphologies in contrast to the extended disks seen at low-redshift (e.g. Elmegreen et al. 2009; Förster Schreiber et al. 2011; Lee et al. 2013; Conselice 2014; Guo et al. 2015). To capture the differing evolution within each class of galaxy, we use the observed axis ratios and Sersic indices that have been measured using GALFIT (Peng et al. 2002) from the deep HST/WFC3 F160W CANDELS imaging in both GOODS-North and GOODS-South (van der Wel et al. 2012). To produce the observed distribution of these parameters for each galaxy class and binned in redshift ($\Delta z = 1$), we match the real CANDELS galaxies in those morphology catalogs with robust morphology measurements ("flag = 0") to the photometric redshifts and rest-frame UVJ colors published by 3D-HST (Skelton et al. 2014). The resulting evolution with redshift of the distribution of parameters for star-forming and quiescent galaxies are shown in Figure 16.

We assign axis ratios and Sersic indices to mock star-forming and quiescent galaxies at a given redshift by generating random variates from those redshift-dependent distributions. We choose to draw directly from the binned distribution, rather than fit with an assumed functional shape for the evolution of the distributions. We first bin axis ratio measurements in bins of $\Delta(b/a) = 0.1$ and Sersic index in bins of $\Delta n = 0.5$, and treat these distributions as probability distribution functions. Then we draw random variates from these distributions to assign to mock galaxies, assuming uniform probability within each bin. Quiescent galaxies at $z > 4$ are assigned morphologies according to the $3 \leq z \leq 4$ distribu-

tions, and star-forming galaxies at $z > 6$ follow the $5 \leq z \leq 6$ distributions.

The resulting axis ratio for each mock galaxy is used to convert the semi-major axis at $z \leq 4$ into both semi-minor axis and circularized half-light radius, and to convert circularized half-light radius into semi-major and minor axes at $z > 4$. Position angles for all galaxies are assigned randomly from a uniform distribution.

6. MOCK GALAXY PROPERTIES

To assess the performance and possible limitations of our phenomenological model, we compare mock galaxy properties, distributions, and relations to observations that were not used to inform our methodology. For this purpose we use a single realization (i.e. a JAGUAR mock catalog) on an area of 11×11 square arcminutes containing both star-forming and quiescent galaxies with $\log(M/M_{\odot}) = 6-12$ and at $z = 0.2-15$.[10] In the following sections, we compare the mock galaxies from this realization to the redshift evolution of observed quantities including galaxy UV luminosity functions, star-formation rate densities and average specific star-formation rates, the mass–metallicity relation, emission line diagnostic diagrams, and observed infrared galaxy colors.

6.1. *UV Luminosity Function Evolution*

We compare the UV luminosity function at $z \leq 4$ computed from the mock catalog with measurements from the literature, as this enables us to test the adopted evolutionary model of both the star-forming galaxy stellar mass function (Section 3.1) and the $M_{\rm UV}$–$M_{\star}$ relation (Section 3.2).

To compare with observations of the UV luminosity function, we use the compilation of literature measurements analyzed in Parsa et al. (2016) from $0.4 < z < 4$. Parsa et al. (2016) provide a compilation of Schechter function parameters with errorbars but do not quantify the covariance(s) between these parameters, which are known to be strong. Because of the degeneracies among the parameters, we avoid comparing directly to individually measured Schechter functions, and rather convert these into step-wise binned luminosity functions with equal magnitudes and bin widths. We then average the binned luminosity functions to produce a mean step-wise luminosity function at each redshift. We quantify the scatter in the literature as the standard deviation of the galaxy counts in each luminosity bin, divided by the square root of the number of measured luminosity functions contributing to the average. The binned averages and scatter from the literature are presented in Table 4.

Figure 17 shows the comparison of the $0.4 < z < 4$ mock galaxy UV luminosity functions (black circles) with the average literature measurements (blue squares).[11] At $z > 1.5$, the mock catalog exhibits excellent agreement with the observations, while at $z < 1.5$ the agreement is less robust. We note, however, that the mock catalog never overpredicts the measured number density of galaxies by more than ~ 0.36 dex (at $M_{\rm UV} \sim -19$ at $0.6 < z < 1$ and $1 < z < 1.5$). This overprediction is likely caused by the poor observational constraints on the rest-frame UV emission of galaxies at $z < 1.5$, where photometry in the 3D-HST catalog no-longer provides coverage at 1500Å (rest wavelength). This affects both the characterization of the $M_{\rm UV}$–$M_{\star}$ relation, which we use to directly assign $M_{\rm UV}$ values to low-mass mock galaxies at these redshifts, as well as the $M_{\rm UV}$ values of high-mass galaxies that are derived directly from fits to the 3D-HST photometry. The same uncertainties affect the observed UV luminosity functions compiled from the literature between $z \sim 0$, where GALEX data are available, and $z \sim 1$, where *HST* near-UV bands probe the rest-frame 1500 Å flux (e.g. Oesch et al. 2010; Windhorst et al. 2011; Teplitz et al. 2013; Rafelski et al. 2015).

6.2. *Star-formation rate density evolution*

In Figure 18, we compare the cosmic star-formation rate density (CSFRD) of mock galaxies with the average CSFRD evolution presented in Madau & Dickinson (2014, based on a uniform analysis of luminosity function measurements in the literature), converted to a Chabrier IMF. To approximate the same UV luminosity limits imposed by Madau & Dickinson (2014), we estimate the limiting $M_{\rm UV}$ corresponding to $0.03\,L^*$ in each redshift bin. These values are taken directly from the Schechter UV luminosity function parameter M^*_{UV} which is used to estimate the equivalent limit in $M_{\rm UV}$ from $0 < z < 10$. The equivalent of $0.03\,L^*$, as published in the following studies, corresponds to $M_{\rm UV} \sim -14.5$ at $1 < z$ (Cucciati et al. 2012); $M_{\rm UV} \sim -15.5$ at $1 < z < 2$ (Cucciati et al. 2012); $M_{\rm UV} \sim -16.89$ at $1.9 < z < 2.7$ (Reddy & Steidel 2009); and $M_{\rm UV} \sim -17$ at $z > 2.7$ (Bouwens et al. 2015; Finkelstein et al. 2015; Bouwens et al. 2016b; Oesch et al. 2017).

To calculate the CSFRD of mock galaxies in each redshift bin, we identify all mock galaxies above the limiting $M_{\rm UV}$ corresponding to $0.03\,L^*$ and take the sum of their star-formation rates (averaged over the past 100 Myr) per comoving volume. We find that the evolution of the CSFRD estimated from the mock catalog qualitatively reproduces the overall shape of the observed CSFRD; however, at $4 < z < 6$ mock galaxies show a slight excess with respect to the results of Madau & Dickinson (2014), who derive their $z > 4$ points

[10] Available at http://fenrir.as.arizona.edu/jwstmock/

[11] We exclude mock galaxies at $z < 0.4$, as the volume probed in the realization is small, and the empirical constraints on $M_{\rm UV}$–$M_{\star}$ and observed UV luminosity functions are less robust.

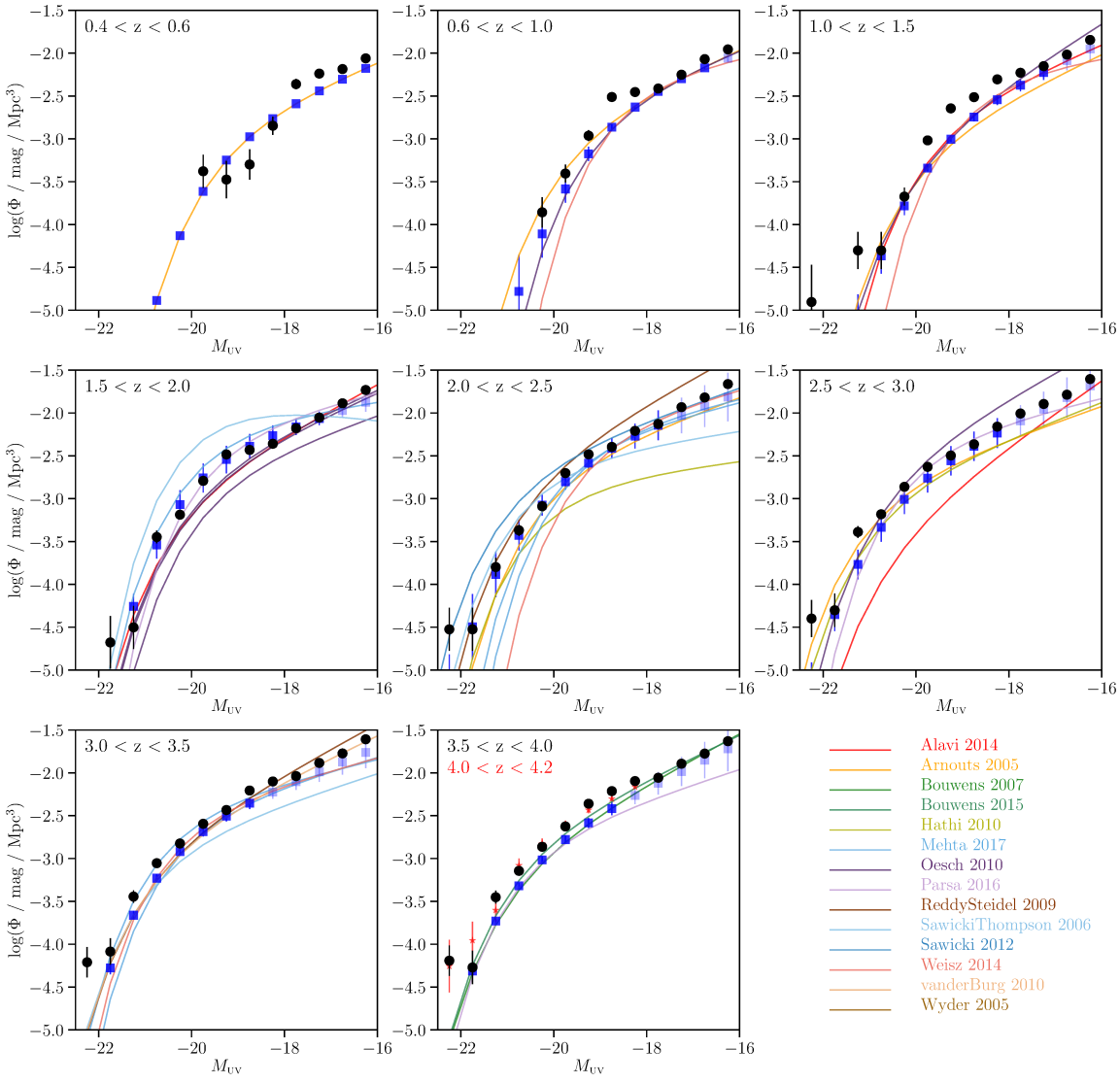

Figure 17. Comparison of the $z < 4$ UV luminosity functions of mock and observed galaxies. Black points indicate mock galaxies, where the error bars represent Poisson errors. Colored lines indicate UV luminosity functions from the literature (Wyder et al. 2005; Arnouts et al. 2005; Hathi et al. 2010; Bouwens et al. 2007, 2015; Alavi et al. 2014; Mehta et al. 2017; Oesch et al. 2010; Reddy & Steidel 2009; Sawicki & Thompson 2006; Sawicki 2012; Weisz et al. 2014; van der Burg et al. 2010), while the average (step-wise) luminosity functions from the literature is shown by the blue squares (blue error-bars indicate the scatter of the observations as described in the text). Observed points below the incompleteness limit are indicated by light-blue points. Red stars in the highest redshift bin indicate the luminosity function measured from mock galaxies at $4 < z < 4.2$.

Table 4. Stepwise average binned luminosity functions from the literature in the redshift bins presented in Figure 17

$M_{\rm UV}$	z$\sim$ 0.5	z$\sim$ 0.8	z$\sim$ 1.25	z$\sim$ 1.75	z$\sim$ 2.25	z$\sim$ 2.75	z$\sim$ 3.75
	LogΦ[a]	LogΦ	LogΦ	LogΦ	LogΦ	LogΦ	LogΦ
-22.75	-14.59	-12.02 $\pm$ 1.34	-10.38 $\pm$ 0.88	-9.06 $\pm$ 0.21	-6.55 $\pm$ 0.70	-6.27 $\pm$ 0.30	-6.44 $\pm$ 0.17
-22.25	-10.42	-8.96 $\pm$ 1.10	-7.91 $\pm$ 0.69	-6.92 $\pm$ 0.17	-5.34 $\pm$ 0.53	-5.15 $\pm$ 0.24	-5.16 $\pm$ 0.11
-21.75	-7.74	-6.98 $\pm$ 0.85	-6.27 $\pm$ 0.50	-5.35 $\pm$ 0.15	-4.49 $\pm$ 0.38	-4.35 $\pm$ 0.19	-4.28 $\pm$ 0.07
-21.25	-6.02	-5.68 $\pm$ 0.62	-5.15 $\pm$ 0.34	-4.26 $\pm$ 0.15	-3.88 $\pm$ 0.27	-3.77 $\pm$ 0.17	-3.66 $\pm$ 0.05
-20.75	-4.88	-4.78 $\pm$ 0.43	-4.37 $\pm$ 0.21	-3.54 $\pm$ 0.16	-3.43 $\pm$ 0.18	-3.33 $\pm$ 0.17	-3.23 $\pm$ 0.05
-20.25	-4.13	-4.11 $\pm$ 0.27	-3.78 $\pm$ 0.11	-3.07 $\pm$ 0.17	-3.08 $\pm$ 0.12	-3.01 $\pm$ 0.17	-2.92 $\pm$ 0.05
-19.75	-3.61	-3.58 $\pm$ 0.16	-3.34 $\pm$ 0.04	-2.76 $\pm$ 0.17	-2.8 $\pm$ 0.09	-2.76 $\pm$ 0.17	-2.69 $\pm$ 0.05
-19.25	-3.25	-3.17 $\pm$ 0.08	-3.0 $\pm$ 0.03	-2.54 $\pm$ 0.16	-2.59 $\pm$ 0.10	-2.56 $\pm$ 0.17	-2.51 $\pm$ 0.06
-18.75	-2.98	-2.86 $\pm$ 0.03	-2.75 $\pm$ 0.05	-2.39 $\pm$ 0.14	-2.41 $\pm$ 0.12	-2.39 $\pm$ 0.17	-2.36 $\pm$ 0.07
-18.25	-2.76	-2.63 $\pm$ 0.01	-2.54 $\pm$ 0.06	-2.27 $\pm$ 0.12	-2.27 $\pm$ 0.15	-2.23 $\pm$ 0.17	-2.22 $\pm$ 0.08
-17.75	-2.59	-2.45 $\pm$ 0.01	-2.37 $\pm$ 0.07	-2.16 $\pm$ 0.10	-2.14 $\pm$ 0.17	-2.09 $\pm$ 0.18	-2.10 $\pm$ 0.10
-17.25	-2.44	-2.3 $\pm$ 0.00	-2.23 $\pm$ 0.09	-2.06 $\pm$ 0.08	-2.03 $\pm$ 0.21	-1.95 $\pm$ 0.20	-1.99 $\pm$ 0.12

[a] The $z \sim 0.5$ binned luminosity function only includes one measurement from the literature, therefore the observed scatter in that bin is zero.

exclusively from the UV luminosity functions of Bouwens et al. (2012b,a). Bouwens et al. (2012a) provided measurements of the $z \sim 6$ luminosity function, along with a compilation of earlier measurements from at $z \sim 4-5$ and $z \sim 7-8$ (Bouwens et al. 2007, 2011). However, our model is based on the UV luminosity function of Bouwens et al. (2015), who find higher number counts than in previous works, in particular at the bright end at $M_{\rm UV} < -18$ and $z \sim 4-5$, and at $M_{\rm UV} < -20$ and $z > 6$, which likely explains our excess relative to Madau & Dickinson (2014). At $1 < z < 3$, the CSFRD measured from mock galaxies is slightly lower than the one of Madau & Dickinson (2014). A reason for this may be that our phenomenological model does not explicitly include extremely dust-obscured galaxies which may be missing from the *HST*-selected samples that we use to characterize the stellar mass function evolution, whereas Madau & Dickinson (2014) do incorporate far-infrared measurements of dust-obscured objects. To validate this explanation, we include individual measurements of the CSFRD from UV-selected samples at $z < 3$ and their uncertainties (Schiminovich et al. 2005; Cucciati et al. 2012; Reddy & Steidel 2009). These indicate a better agreement with our mock galaxies, within the scatter of binned observations, than implied by the Madau & Dickinson (2014) curve.

We additionally plot measurements of the CSFRD at $z > 4$ from Bouwens et al. (2015); Finkelstein et al. (2015); Bouwens et al. (2016a); McLeod et al. (2016); Oesch et al. (2017, converted to Chabrier IMF when necessary). The CSFRD evolution of mock galaxies at $z > 4$ is in better agreement with these later measurements. The $M_{\rm UV}$ integration limits in these studies match those used to estimate the CSFRD from the mock catalog with the exception of McLeod et al. (2016), whose integration limit is $M_{\rm UV} \sim -17.7$. The CSFRD of McLeod et al. (2016) would be higher if measured to the same limiting $M_{\rm UV}$ used for the mock catalog. The higher UV counts (and thus CSFRD) measured in McLeod et al. (2016) have been interpreted as evidence for a slower rate of decline in galaxy number counts with redshift, in contrast to the "accelerated" evolution that is seen in, e.g., Bouwens et al. (2015); Oesch et al. (2017). Since our model follows the more rapidly evolving luminosity function of Oesch et al. (2017), the CSFRD in the mock follows more closely the CSFRD measured in that work. We find that our mock realization generally reproduces the $z > 4$ evolution of the CSFRD based on the latest results from current extragalactic surveys, and in particular those we have used to develop our model for the evolution of the stellar mass function.

6.3. *Evolution of the average specific star-formation rate*

We compare the redshift evolution of the specific star-formation rate (sSFR) from the mock catalog with literature measurements at $0 < z < 7$, the highest redshift at which observational constraints are available. We consider the median sSFR of galaxies with masses $8.8 < \log(M/M_\odot) < 10$ as computed by Noeske et al. (2007); Damen et al. (2009); Daddi et al. (2007); Reddy & Steidel (2009); Stark et al. (2013); González et al. (2014); Tasca et al. (2015); Salmon et al. (2015). To calculate the median sSFR of mock galaxies, we use the BEAGLE-assigned SFR averaged over the last 100 Myr divided by stellar mass of mock galaxies in the same stellar mass range as the observations. We plot this comparison in Figure 19, and show that the sSFR values for our mock are in excellent agreement with the observations within their uncertainties.

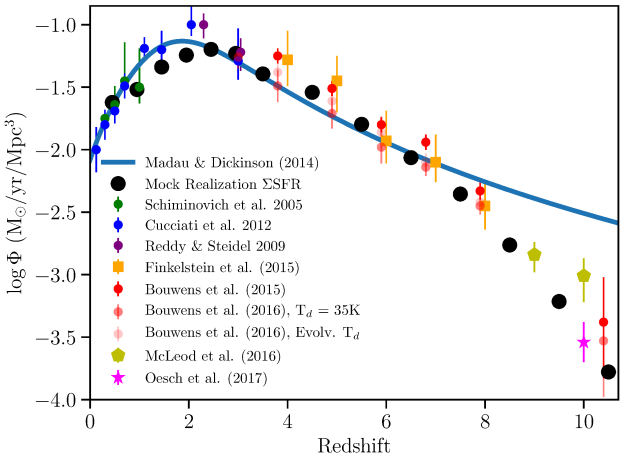

Figure 18. Evolution of the cosmic star formation rate density of the Universe as compiled by Madau & Dickinson (2014) (blue curve) and as computed from our mock catalogue (black points). Colored points indicate individual measurements from the literature.

As discussed in Section 3.2, the sSFR of galaxies at $z \gtrsim 4$ and the redshift dependence of the average sSFR are directly related to the extrapolation we adopted for the redshift evolution and normalization of the $M_{\rm UV}$–$M_\star$ relation. This is in contrast to mock galaxy sSFRs at $z < 4$, where the $M_{\rm UV}$–$M_\star$ relation is based directly on more reliable stellar mass measurements across a range of UV luminosities. Therefore, the agreement between the mock galaxy sSFRs with measurements in the literature at $z > 4$ shown in Figure 19 is a validation of the adopted extrapolations of the $M_{\rm UV}$–$M_\star$ relation at high redshift. We note, however, that the evolution of the sSFR of galaxies at $z > 4$ is still a matter of active research; most studies support an increasing average sSFR rather than a plateau at $z > 4$ (see discussion in Stark 2016). An accurate characterization of the sSFR evolution at $z > 4$ will await improved measurements that will be made possible by *JWST*.

6.4. *Evolution of the Mass-Metallicity Relation*

The mass-metallicity relation of galaxies is known to evolve, such that galaxies at higher redshifts are more metal-

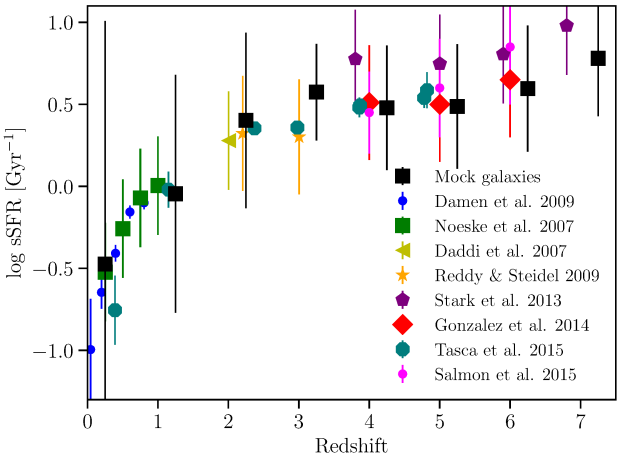

Figure 19. Redshift evolution of the (median) specific star formation rate of mock galaxies with $8.8 < (\log M_\star/M_\odot) < 10$ (black squares) compared to measurements from the literature (colored points).

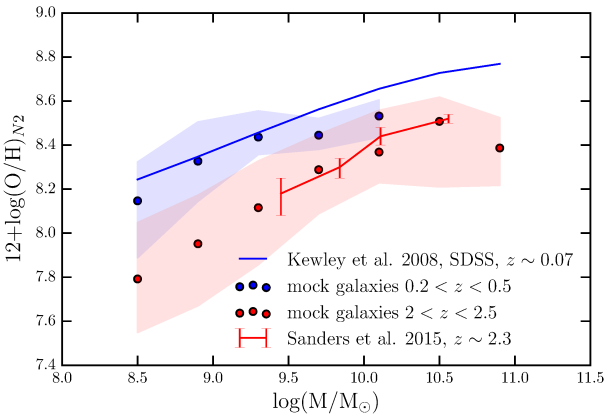

Figure 20. Mass-metallicity relation of mock galaxies compared to observations in two bins of redshift, $0.2 < z < 0.5$ (blue) and $2 < z < 2.5$ (red). The points show the median values of metallicity for mock galaxies in those redshift ranges, estimated using the NII calibration (see text for details), while the shaded regions encompass the 25% and 75% percentiles for mock galaxies. Only bins containing at least 5 galaxies are plotted. The blue solid line indicates the observed relation of Kewley & Ellison (2008) ($z \sim 0.07$), while the red line and error bars show the observations of Sanders et al. (2015) ($z \sim 2.3$).

poor at a given stellar mass (e.g. Maiolino et al. 2008). When assigning SEDs (and associated physical properties) to each mock galaxy, we use a parent catalog that has a built-in redshift-independent prior linking $M_\star - Z - \psi$. based on the fundamental metallicity relation as measured by Hunt et al. (2016). The redshift-evolving M_{UV}–$M_\star$ relation, however, produces a redshift evolution of the $M_\star - \psi$ relation which, in turn, should generate an evolving mass-metallicity relation.

Figure 20 shows the mass-metallicity relation in two redshift bins, $0.2 \leq z \leq 0.5$, and $2 \leq z \leq 2.5$. Given that different metallicity indicators provide systematically different metallicity estimates (Kewley & Ellison 2008), we choose to estimate the mock galaxy metallicities using a metallicity calibration often used in the literature, based on the ratio between the lines [NII]λ6584 and Hα (the Pettini & Pagel (2004) N2 metallicity calibration):

$$12+\log(O/H)_{N2} = 8.90 + 0.57 \times \log([NII]\lambda6584/H\alpha) \quad (29)$$

This allows us to compare directly to observationally-derived mass-metallicity relations.

Figure 20 indicates that the mass-metallicity relation derived from the mock catalog does show a turn-over at high stellar masses, as seen in the observations, despite the Hunt et al. (2016) fundamental metallicity relation only being linearly dependent on $\log(Z/Z_\odot)$. The turnover is caused by the presence of a maximum metallicity $\log(Z/Z_\odot) = 0.24$ that can be assigned to the mock galaxies (see Sections 3.4.2 and 3.4.3). Figure 20 demonstrates a good agreement among our mock galaxies and the mass-metallicity relation of Kewley & Ellison (2008) (at $z \sim 0.07$) and of Sanders et al. (2015) (at $z \sim 2.3$). As discussed above, the evolution of the mass-metallicity relation in our mock catalog is a result of the priors we impose on both $M_\star - Z - \psi$ and M_{UV}–$M_\star$.

6.5. *Emission line diagnostic diagrams*

Emission line diagnostic diagrams are commonly used in the literature to identify the ionization sources in galaxies and characterize their physical properties. Baldwin et al. (1981) (BPT hereafter) pioneered the use of the ratios of [OIII]λ5007/Hβ and [NII]λ6584/Hα to separate star-forming galaxies and AGN-dominated galaxies into two distinct regions of the plot, with composite galaxies (those with both significant star-formation and AGN contribution) spanning the region in between. Recent observations of rest-frame optical spectra at redshifts out to $z \sim 2.5$ have revealed that the locus of star-forming galaxies is slightly shifted to higher [OIII]λ5007/Hβ ratios, at fixed [NII]λ6584/Hα, than in the local Universe (e.g. Shapley et al. 2005; Brinchmann et al. 2008; Hainline et al. 2009; Kewley et al. 2013; Steidel et al. 2014; Kashino et al. 2017). This shift can been explained if the physical conditions of the ionized gas are evolving with redshift. For example, evolution in the ionization parameter (e.g. Kashino et al. 2017), the incident ionizing spectrum of stars (e.g. Steidel et al. 2014, 2016), gas-phase nitrogen abundance (e.g. Masters et al. 2016), and hydrogen density (e.g. Sanders et al. 2016) can all introduce a shift in observed [OIII]λ5007/Hβ ratio. However, it is also possible that selection effects play a role (e.g. Juneau et al. 2014). More likely, the observed evolution of the location

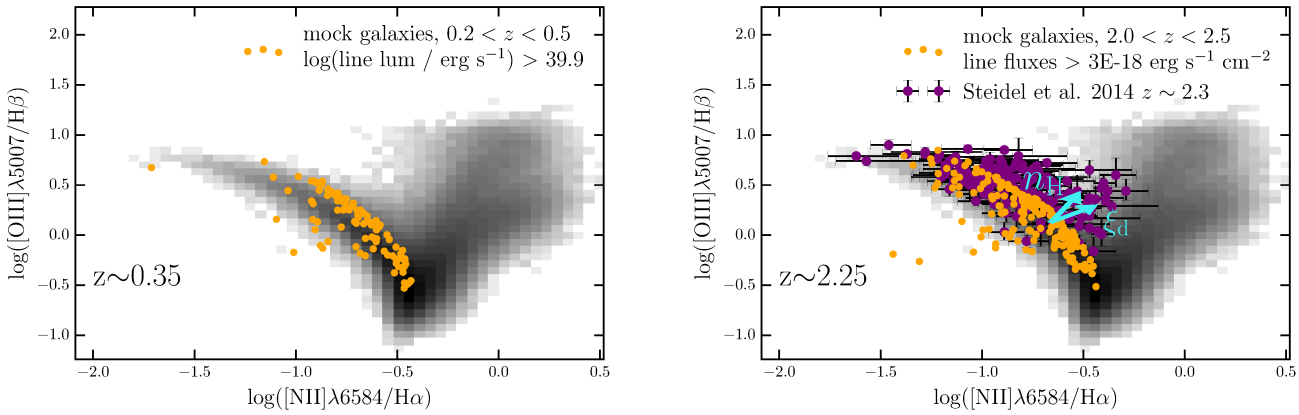

Figure 21. The BPT diagram of mock galaxies compared to observations in two different redshift bins, $0.2 < z < 0.5$ in the right panel and $2 < z < 2.5$ in the left panel. The mock galaxies are plotted as orange points in each panel, and we display the observations from SDSS as the grey 2D histogram. Purple points and error-bars in the right panel are the observations of Steidel et al. (2014), excluding those objects classified as AGN. In the left panel, only mock galaxies with line luminosities above log(line lum / erg s^{-1}) > 39.9 are plotted, to approximate the limits in the SDSS sample, while in the right panel a flux limit of $> 3 \times 10^{-18}$ erg s $^{-1}$ was applied to mock galaxies plotted at $2 < z < 2.5$, chosen to mimic the flux limit of the Steidel et al. (2014) sample. The two blue vectors plotted on the right panel display the direction that objects move in the diagram with changes to $n_{\rm H}$ (changing $n_{\rm H}$ from 100 to 1000 cm^{-3}) or $\xi_{\rm d}$ (changing $\xi_{\rm d}$ from 0.3 to 0.5).

of star-forming galaxies in the BPT diagram is caused by a combination of the effects above, but current data do not allow us to quantify their respective roles.

In Figure 21 we plot the BPT diagram of mock galaxies at low ($0.2 < z < 0.5$) and high ($2 < z < 2.5$) redshift. We compare the mock galaxies with SDSS galaxies ($z \sim 0.07$) and with a sample of star-forming galaxies from the KBSS survey (Steidel et al. 2014) ($z \sim 2.5$). For mock galaxies at $0.2 < z < 0.5$, we adopt the lowest luminosity threshold of Juneau et al. (2014), and only plot objects with log luminosities > 39.9 erg s^{-1} in the lines [OIII]λ5007, Hβ, Hα, [NII]λ6584. As Figure 21 shows, the mock galaxies sit close to the main star-forming galaxy locus, despite being at higher redshift than the SDSS sample. At $2 < z < 2.5$, we instead apply a flux limit close to that of Steidel et al. (2014), i.e. line flux $> 3 \times 10^{-18}$ erg s^{-1} cm^{-2}. At these higher redshifts, the mock galaxies cover similar regions to the Steidel et al. (2014) points at low [NII]λ6584/Hα, but not at high values of [NII]λ6584/Hα. This difference can be understood by appealing to the recent study of Hirschmann et al. (2017), where they self-consistently couple the nebular emission models of Gutkin et al. (2016) for star forming galaxies and narrow-line AGN-driven models of Feltre et al. (2016) to cosmological zoom-in hydrodynamic simulations. In their work, they consistently tie log$U_{\rm S}$ to the simulated galaxy properties, and fix $\xi_{\rm d}$ and $n_{\rm H}$ to the same values as fixed in this mock catalog. They reproduce the observed evolution of the line ratios, i.e. higher [OIII]λ5007/Hβ on average at higher redshifts, but their figure 12 indicates that covering the elevated [OIII]λ5007/Hβ values at high [NII]λ6584/Hα requires either a higher value of $\xi_{\rm d}$ and/or of $n_{\rm H}$, or some level of AGN activity. We add to the the right panel of Figure 21 two vectors indicating how variations of these two parameters modify the expected line ratios. Given that the possible reasons for the elevated [OIII]λ5007/Hβ ratio is far from determined, we choose to supply line ratios for mock galaxies with different values of $\xi_{\rm d}$ and $n_{\rm H}$. Since an elevated $\xi_{\rm d}$ ratio acts to increase the relative abundance of gas-phase nitrogen to oxygen (since oxygen is a refractory element, unlike nitrogen, and thus heavily depleted onto dust grains), we additionally provide line fluxes and EWs for a single mock realization with $\xi_{\rm d} = 0.5$ (where the fiducial mock has $\xi_{\rm d}$=0.3) which will be consistent with elevated nitrogen abundance (whether physically we expect this to be due to dust depletion or the relative importance of primary vs. secondary nitrogen production in stars at different epochs, the effect on these two line ratios will be similar). We also provide line fluxes and equivalent for galaxies in that realization with a higher hydrogen density. The Gutkin et al. (2016) models are evaluated in unit steps in log($n_{\rm H}$), so we are able to supply this line information for $n_{\rm H}$ = 1000 cm^{-3} (where the fiducial mock has $n_{\rm H}$ = 100 cm^{-3}).

6.6. *Rest-frame optical colors at $z \sim 4-7$*

Strong nebular emission lines are now known to contaminate the broad-band photometry of galaxies. In particular, colors from *Spitzer*/IRAC have been extensively used to infer the EW of strong optical emission lines (Shim et al. 2011; Smit et al. 2014, 2015, 2016; Mármol-Queraltó et al. 2016; Rasappu et al. 2016; Faisst et al. 2016; Castellano et al. 2017). In Figure 22, we compare the IRAC 3.6 - 4.5 μm color of mock galaxies with measurements at $z > 4$ from Smit et al. (2014, 2015, 2016) and Rasappu et al. (2016).The

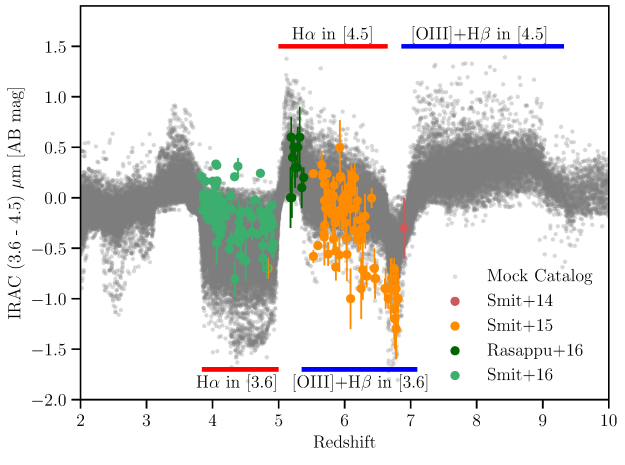

Figure 22. *Spitzer*/IRAC 3.6 and 4.5μm band colors for mock galaxies with log(M/M$_\odot$)>6 (grey points) compared to observations (colored points). Redshift ranges where Hα, [OIII] and Hβ enter the 3.6 (4.5)μm bands are shown as the horizontal bars on the bottom (top) of the figure. Our model produces mock galaxies whose rest-frame optical emission line properties span the observed color excesses from current surveys.

redshift evolution of the 3.6 - 4.5 μm color predicted by the mock catalog, caused by the emission lines Hα, [OIII], and Hβ entering and leaving the IRAC bands, is consistent with the observations. This suggests that the emission line strengths in the mock galaxies are compatible with the values inferred from IRAC observations. In Section 7.2 below, we will use the mock catalog to show how the sensitivity and wavelength coverage of *JWST*/NIRCam will constrain emission line EWs with greater accuracy than existing *Spitzer* data.

7. PREDICTIONS FOR THE FIRST NIRCAM/*JWST* IMAGING SURVEYS

We demonstrate the predictive power of JAGUAR using the realization of the phenomenological model analyzed in Section 6 in order to make predictions for the initial deep extragalactic surveys that will be produced with *JWST*. The NIRCam and NIRSpec Science Teams have proposed a joint GTO program, the JADES survey, which will encompass both imaging and spectroscopy over 236 arcmin^2 in two well-studied fields, GOODS-S and GOODS-N. The NIRCam imaging aspect of the survey is a two-tiered strategy composed of a 46 arcmin^2 deep sub-survey, and a 190 arcmin^2 medium-depth sub-survey, with ten photometric bands between 0.7 and 5 μm. The average point source detection limits for each subsurvey are summarized in Table 5. For illustration we plot in Figure 23 two example SEDs of mock high-redshift galaxies that would be detectable in the JADES survey.

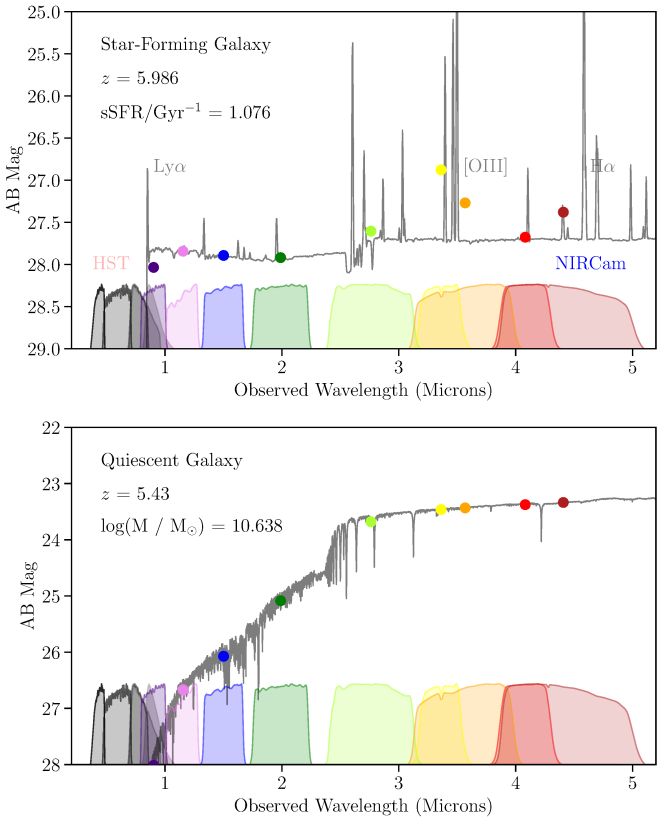

Figure 23. Example SEDs and photometry for two objects in the mock catalog, with HST and NIRCam photometric filter transmission curves plotted below. The bands used are, from left to right HST ACS F435W, F606W, F750W, F814W, and F850LP (shades of grey), and NIRCam F070W, F090W, F115W, F150W, F200W, F277W, F335M, F356W, F410M, and F444W (rainbow). (Top) A star-forming galaxy at $z = 5.986$, with strong emission lines labeled. (Bottom) A quiescent galaxy at $z = 5.43$. The full SED of each galaxy, as well as line fluxes and rest-frame EWs, are provided in the mock catalog, which contains over 300,000 model SEDs across 11×11 arcmin^2.

A visualization of the mock catalog can be seen in Figure 24, which shows a simulated NIRCam mosaic of the JADES deep subsurvey (see Table 5). The images in the mosaic were generated using *Guitarra* (C. Willmer, in preparation), a ray-tracing image simulator specifically designed to create mock *JWST*/NIRCam scenes. Photons in the simulated image that are associated with an object are added to the detector pixels after being convolved with the mock galaxy Sersic model, the point spread function and the intra-pixel capacitance (Rauscher et al. 2007). The simulated image also includes the contribution due to the zodiacal and telescope background light, cosmic rays, read noise and the detector signatures as measured from ground-based data. The scenes are made using the same read-out patterns that will be used in flight, and are reduced using the pipeline that will be ap-

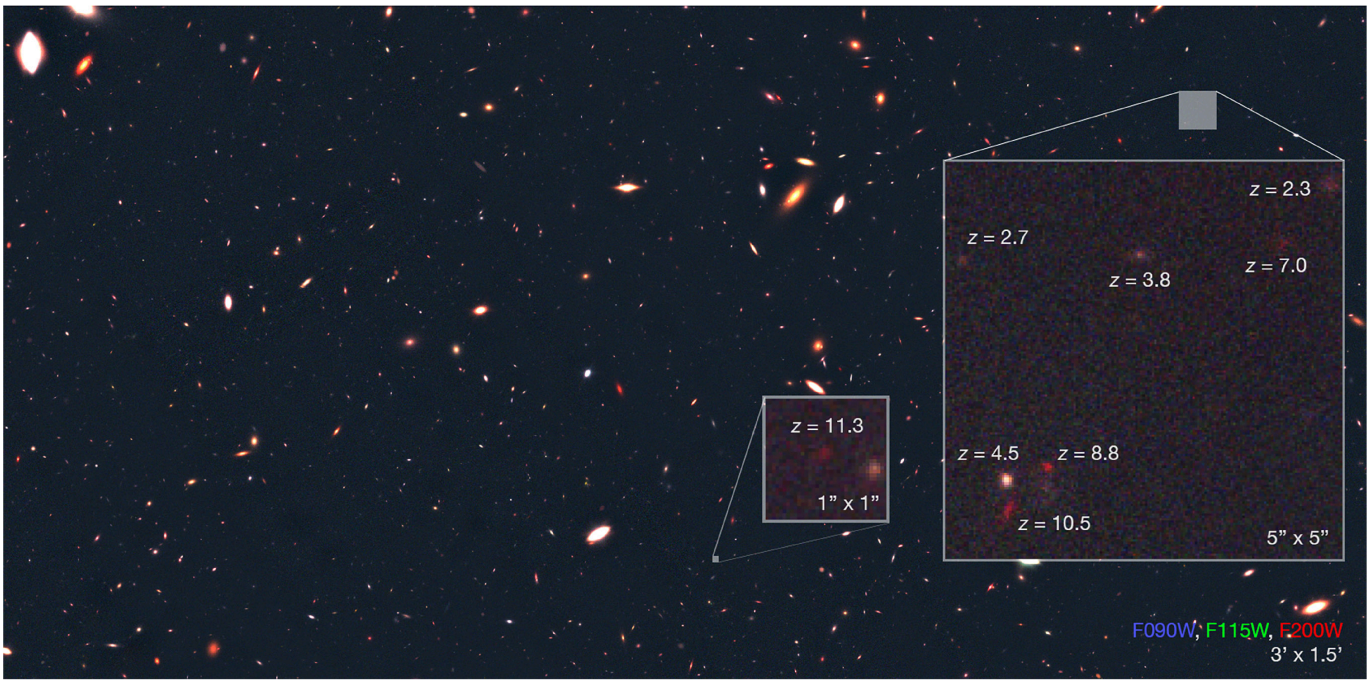

Figure 24. Simulated *JWST*/NIRCam mosaic generated using JAGUAR and the NIRCam image simulator *Guitarra* (C. Willmer, in preparation), at the depth of the JADES Deep program. This image is focused on a region of 3' by 1.5', and is a composite of the F090W (blue), F115W (green), and F200W (red) filters. The insets show a 5" by 5' region with multiple high-redshift galaxies, and a 1" by 1" region focused on a galaxy at $z = 11.3$.

plied to the JWST data once these become available. The image shown in Figure 24 is a composite of a total of 648 images in F090W, F115W and F356W, combined using *swarp* (Bertin et al. 2002), using the dither positions calculated by the JWST Astronomer's Proposal Tool (APT) for GTO proposal 1180.

7.1. *High-redshift galaxy counts*

In Figures 25 and 26, we show predictions for the high-redshift detections with NIRCam imaging in the JADES GTO survey. In these figures, we plot only those objects that are detected with at least 5σ (assuming point source detection limits) in two rest-frame UV photometric bands: that closest to 1500 Å, and the nearest band at a longer wavelength. The detection bands correspond to F115W and F150W at $6 < z < 7$; F150W and F200W at $7 < z < 9.6$; and F200W and F277W at $9.6 < z < 13$.

Figure 25 shows the expected number of objects detected as a function of $M_{\rm UV}$ and redshift in the JADES survey based on our phenomenological model. We find that JADES should detect several thousands of galaxies at $z > 6$, and several tens out to $z > 10$. These predictions are based on the galaxy number counts evolution of our phenomenological model, which follows observations supporting a more rapid ("accelerated") evolution of the UV luminosity function at $z \gtrsim 8$ (e.g. Bouwens et al. 2016b; Oesch et al. 2017), implying a factor of ~ 10 decrease of galaxy counts between $z \sim 8$ and $z \sim 10$. This is substantially faster than the factor of ~ 2 de-

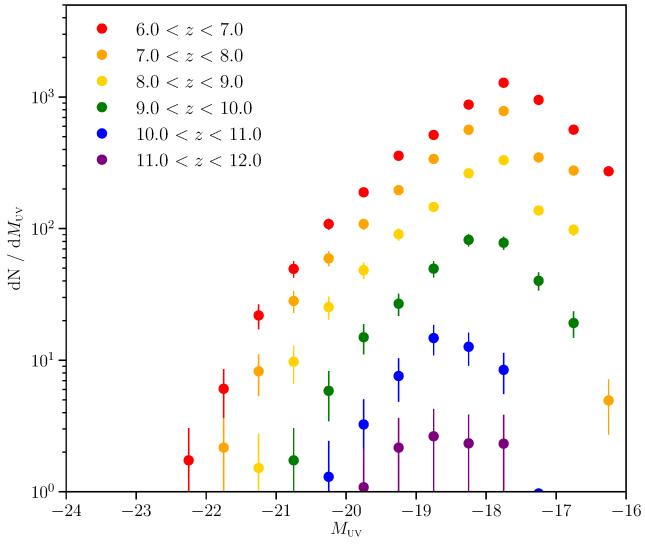

Figure 25. Predicted total number of objects detected as a function of $M_{\rm UV}$ in bins of redshift for JADES. Objects are selected as "detected" if they are brighter than the 5σ limits in two photometric bands corresponding to the rest-frame UV, similar to common LBG selection techniques.

Subsurvey	Area [arcmin^2]	5σ Point Source Magnitude (AB)									
		F070W	F090W	F115W	F150W	F200W	F277W	F335M	F356W	F410M	F444W
Deep	46	—	30.3	30.6	30.7	30.7	30.3	29.6	30.2	29.8	29.9
Medium	190	28.8[a]	29.4	29.6	29.7	29.8	29.4	28.8[a]	29.4	28.9	29.1

Table 5. Summary of NIRCam Imaging as part of JADES. The 5-σ depth for a point source corresponding to the average exposure times. We also include the area covered in square arcminutes. [a]The F070W and F335M areas of the Medium survey are only 93 square arcminutes.

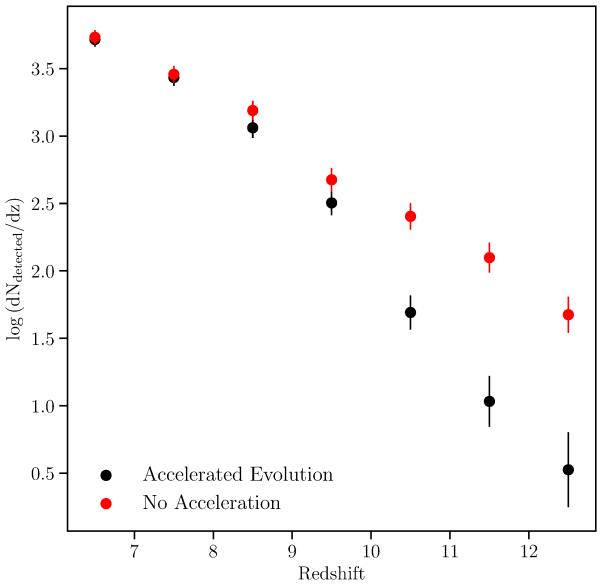

Figure 26. Predicted total number of galaxies as a function of redshift in the full JADES area detected at $\gtrsim 5\sigma$. We plot results from the mock with the accelerated evolution above $z = 8$ with black points against those with the evolution predicted by Bouwens et al. (2015) with red points. The uncertainties include both Poisson error, and cosmic variance estimated from the Trenti & Stiavelli (2008) "cosmic variance calculator". Data from the JWST GTO program will help to discern between the two evolutionary scenarios at high redshift.

crease of galaxy counts per unit redshift seen at $3 < z < 8$ (e.g. Bouwens et al. 2015; Finkelstein 2016) which are similar to the decrease measured by McLeod et al. (2015, 2016) at $z > 8$. However, due to large uncertainties from small sample sizes and survey volumes at $z \gtrsim 8$ in current observations, there remain discrepancies in the literature of how fast the luminosity function evolves (Oesch et al. 2012, 2014; Zheng et al. 2012; Coe et al. 2013; Ellis et al. 2013; McLeod et al. 2015, 2016; Bouwens et al. 2015, 2016b; Calvi et al. 2016; Stefanon et al. 2017b; Oesch et al. 2017; Ishigaki et al. 2017). Accurately measuring the evolution at $z > 8$ is an important goal for *JWST* because it will directly constrain the relationship between star-formation efficiency and the evolution of the halo mass function at early times (Trenti et al. 2010; Tacchella et al. 2013; Mason et al. 2015; Liu et al. 2016; Mashian et al. 2016; Sun & Furlanetto 2016; Ceverino et al. 2017; Cowley et al. 2018), and is critical for our understanding of reionization.

Therefore, we have used our mock catalog tool to predict if JADES will distinguish between the constant vs. "accelerated" evolutionary model at $z \gtrsim 8$. To do this, we have used the tool to produce a comparison mock realization based on the constant rate of evolution that was parametrized in Bouwens et al. (2015) at $4 < z < 8$, which we extrapolate to $z \gtrsim 8$. This parametrization is in good agreement with the observations at $z \sim 9-10$ by McLeod et al. (2015, 2016). For this secondary set of realizations, we follow the same procedures for assigning physical properties and SEDs to mock galaxies as outlined previously in this work; the only difference is the surface density of counts of a given $M_{\rm UV}$ predicted by the UV luminosity functions. In Figure 26, we additionally plot the total predicted number of galaxies that will be detected per redshift bin, expected for the constant rate of evolution from Bouwens et al. (2015) from the JADES survey (red points) in addition to the accelerated model (black points). The error bars in this figure represent both Poisson uncertainties as well as estimates for the uncertainty due to cosmic variance (Trenti & Stiavelli 2008), assuming the JADES survey area and a Press-Schechter formalism. From this figure it is clear that the future JADES survey will discriminate between the two evolutionary scenarios.

7.2. *Emission line predictions for NIRCam*

The inclusion of nebular emission lines in the SEDs of mock galaxies enables analyses of the photometric contamination that is likely to be observed in future NIRCam imaging. Identifying the exact level of contamination in current surveys remains difficult, because there are only two deep *Spitzer* bands that provide rest-frame optical coverage of SEDs above $z \sim 4.5$. These bands are broad and there are only limited, narrow redshift ranges where one or other band is free from emission line contamination, and therefore able to provide constraints on the stellar continuum. However, the extensive filter set of NIRCam, with three broad-band filters red-wards of current ground-based K-band imaging as well as a large array of medium- and narrow-band filters will offer the opportunity to better characterize both the typical and

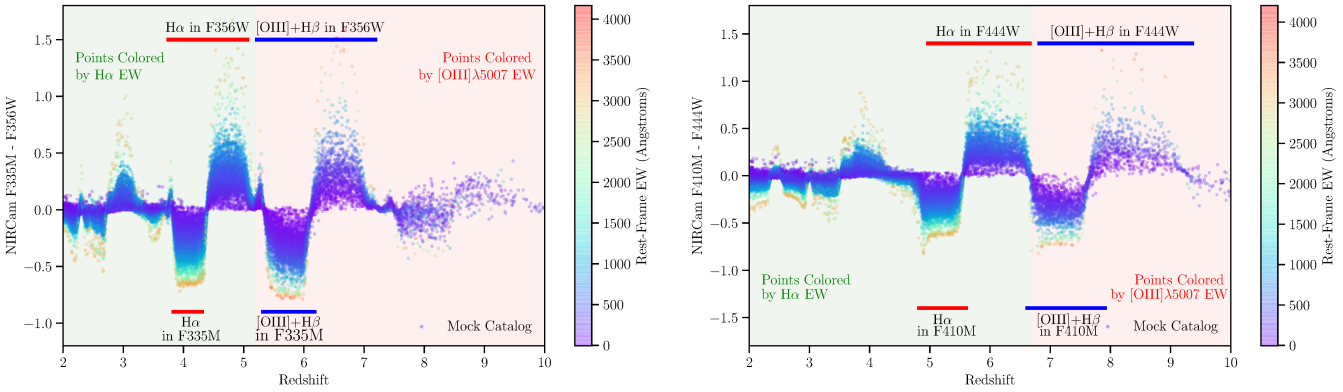

Figure 27. Predicted *JWST*/NIRCam colors for mock galaxies in the realization of our model as a function of redshift. In both panels, we only plot those objects with fluxes above the 5σ detection limit for that waveband in the Deep region of JADES as described in the text and Table 5. Left panel: F335M-F356W colors. Right panel: F410M-F444W colors. The points are colored by the rest-frame EW of the Hα emission line (light green region), and [OIII]λ5007 emission line (light red region). The redshift ranges where Hα, [OIII] and Hβ enter the medium (wide) band filters are shown as the horizontal bars on the bottom (top) of the figure. In both panels, the mock predicts that we will be able to identify objects with powerful emission lines based on simple NIRCam color cuts. In the right panel, the features at $z = 2-3.5$ result from higher-order Paschen lines in the near-IR moving into the F410M and F444W filters.

extreme emission line EWs, which in turn will provide much better constraints on stellar masses at $z > 4$. As summarized in Table 5, the JADES survey will provide imaging in two medium bands at $\lambda > 3\,\mu$m (F335M and F410M). These filters provide continuum anchor points when the emission line is outside the medium-band filter, or direct emission line contamination estimates of the broad-band fluxes when the line sits within the filter.

In Figure 27, we plot the predicted redshift evolution of the NIRCam colors F335M-F356W and F410M-F444W of mock galaxies that would be detected at $> 5\sigma$ in the deep region of the JADES survey. The galaxies are color-coded by the rest-frame EW of the emission lines that are contributing to the observed photometric flux at that redshift: Hα at $z < 5.2$ and [OIII]λ5007 at $z > 5.2$ ([OIII]λ5007 dominates the EW of the [OIII]$\lambda\lambda$4959,5007 doublet). This figure demonstrates how strong emission lines can impact the observed infrared colors of galaxies in specific redshift ranges. The multiple overlapping NIRCam bands means that not all bands are contaminated at a given time, and stellar mass estimates are likely to be better constrained than currently possible at $z \sim 6$, where both IRAC bands are affected by emission line contamination. Additionally, Figure 27 demonstrates how NIRCam colors can be used to select high EW emission line objects at specific redshifts. Of those galaxies in the mock catalog with 5σ Deep JADES detections, a color cut of F335M - F356W > 0.8 and < -0.5 will be successful at selecting a sample that is 99% comprised of objects with rest-frame EW$_{\rm H\alpha} \gtrsim 1000$Å at $z < 5.2$ (the number density of this sample is 3.85 objects per square arcminute). Similarly, a color cut of F335M - F356W > 0.8 and < -0.6 will be successful at selecting a sample that is 98% comprised of objects with rest-frame EW$_{\rm [OIII]\lambda5007} \gtrsim 1000$Å at $z > 5.2$ (the number density of this sample is 1.62 objects per square arcminute). Objects selected with such extreme equivalent width emission lines would be obvious targets for follow-up spectroscopy with NIRSpec, and provide an independent probe of the Hα-derived star-formation rates free from the effects of slit losses produced by the MSA on NIRSpec. They would additionally provide a consistent probe of the evolution of the number densities of extreme emission line galaxies from $z \sim 4.8-8.3$.

8. SUMMARY

We have developed a novel phenomenological model for the evolution of galaxies and their properties, based on empirical constraints from current surveys between $0.2 < z < 10$. Our model follows observed stellar mass functions, UV luminosity functions, integrated distributions including $M_{\rm UV}$–M$_\star$, β–$M_{\rm UV}$, and size-mass and size-$M_{\rm UV}$ distributions. Importantly, mock realizations of our model include galaxy SEDs that include strong nebular emission lines thanks to our self-consistent modeling of the stellar and nebular emission with BEAGLE. These allow us to realistically model emission-line contamination to broad and medium-band filters but also provide emission line properties and low-resolution spectra for each galaxy in our mock catalog. We have demonstrated that our phenomenological model is successful at matching the CSFRD, observations of the evolution of both the UV luminosity function and sSFR of galaxies over cosmic time as well as the evolution of the mass-metallicity relation.

We have created a mock catalog (our fiducial mock) using our phenomenological model and used this to make predictions for deep extragalactic surveys with NIRCam, in particular the joint NIRCam/NIRSpec GTO survey, JADES. We

find that this survey will detect 1000s of galaxies at $z > 6$ and 10s of galaxies at $z > 10$, and will put firm constraints on the evolution of galaxy counts at $z > 8$, resolving uncertainties on the rate of evolution that is currently debated in the literature. Additionally, we demonstrate how NIRCam colors can be used to select for high-equivalent width line emitters at high-redshift using the emission line information that is provided for the mock galaxies. We make JAGUAR available for use, including both ready to use mock catalogs, and software to produce additional mock catalogs[12].

Acknowledgments. We gratefully thank Darren Croton for thoughtful and constructive comments. Authors acknowledge helpful discussions with George Rieke, Michaela Hirschmann, and Jacob Magnusson, and gratefully thank Karl Misselt for computing and website assistance, and Pascal Oesch for providing his $z \sim 10$ luminosity function measurements. CCW acknowledges enlightening conversations with Ivo Labbé. ECL, JC and SCh acknowledge support from the European Research Council (ERC) via an Advanced Grant under grant agreement no. 321323-NEOGAL. CCW acknowledges support from the National Science Foundation Astronomy and Astrophysics Fellowship grant AST-1701546. SCh acknowledges financial support from the Science and Technology Facilities Council (STFC). All members of NIRCam (CCW, KNH, BER, RE, DPS, CNAW, SAl, SB, SCr, EE, DJE, MR) acknowledge funding from JWST/NIRCam contract to the University of Arizona, NAS5-02015. BER acknowledges partial support through NASA contract NNG16PJ25C, grants 17-ATP17-0034 and HST-GO-14747. SAr is funded by MINECO under grant ESP2015-68964-P. RM and RA acknowledge ERC Advanced Grant 695671 "QUENCH" and support by the Science and Technology Facilities Council (STFC). RS acknowledges a NWO Rubicon grant, project number 680-50-1518. This work is based on observations taken by the CANDELS Multi-cycle Treasury Program with the NASA/ESA HST. This research made use of Astropy, a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013).

APPENDIX

A. RE-FITTING OBSERVED UV-LUMINOSITY AND MASS FUNCTIONS

A.1. *Fitting mass function parameters to $z \gtrsim 4$ observed luminosity functions*

As described in Section 3.1, we re-fit the individual Bouwens et al. (2015) UV-luminosity functions with *mass* function Schechter parameters. To do this we convolve a given mass function with our model of the evolving $M_{\rm UV}$–$M_\star$ relation (described in Section 3.2) to produce the corresponding UV luminosity function (general procedure described in Section 2.1). We then use Markov-Chain Monte-Carlo (MCMC) sampling, employing the adaptive metropolis algorithm of Haario et al. (2001), to sample from the posterior probability distributions of the Schechter function parameters. At each iteration, Schechter function parameters are proposed and corresponding CDF function generated. The CDF function is used to randomly assign masses to a large population of objects, which can then be used to assign each object an $M_{\rm UV}$ value following our adopted redshift-evolving $M_{\rm UV}$–$M_\star$ model. The resulting UV luminosity function is measured within the $M_{\rm UV}$ bins of the published measurements. The likelihood of those given Schechter function parameters is evaluated using the published $\Phi(M_{\rm UV}, z)$ values and associated errors given in Bouwens et al. (2015), Table 5, and the modeled luminosity function values evaluated in the same $M_{\rm UV}$ bins.

The $z \gtrsim 4$ luminosity functions are well-described by a single Schechter function; thus fitting these data with a double Schechter function renders the parameters of $\Phi_1(M_\star)$ unconstrained at $z \gtrsim 4$. We therefore constrain the values of each $\Phi_1(M_\star)$ parameter while performing the luminosity function fits. The values of $\alpha_{1,M}$ and $\phi^*_{1,M}$ are fixed in each redshift bin using the extrapolated best-fit linear and quadratic relations to the published T14 maximum likelihood measurements, and we also tie the value of $M^*_{1,M}$ to that of $M^*_{2,M}$with $M^*_{1,M} = M^*_{2,M} = M^*_M$. The exact form of the redshift evolution of the parameters of $\Phi_1(M_\star)$ do not affect the results of the luminosity function fits as long as the number density has fallen significantly by $z \sim 4$.

The measurements are displayed in Figure 3.

A.2. *Multi-level modeling to fit to Tomczak et al. (2014) star-forming galaxy stellar mass functions*

To constrain the parameters a_1, b_1, b_2, b_3, c_1, c_2, e_1 and f_1 of our model of the redshift evolution of the mass function (equations 5 - 10), we fit to the T14 star-forming mass functions using a Bayesian multi-level modeling approach. Whereas the Bayesian fitting to each individual luminosity function entails sampling from the posterior distribution of each Schechter function parameter, the multi-level modeling involves sampling over the posterior distribution of the hyper-parameters (as we shall now refer to the parameters a_1, b_1, b_2, b_3, c_1, c_2, e_1 and f_1) *and* the conditional probability distributions for each Schechter function

[12] available for download at http://fenrir.as.arizona.edu/jaguar/

parameter in each redshift bin. The posterior distribution of the model free parameters can be expressed as:

$$\mathrm{P}(\mathbf{A},\Phi \mid \mathbf{X}) \propto \prod_{i=1}^{8} \mathrm{P}(A_i) \prod_{z=1}^{8} \mathrm{P}(\mathbf{x}_z \mid \phi_z) \prod_{z=1}^{8} \mathrm{P}(\phi_z \mid \mathbf{A}), \tag{A1}$$

where $\mathbf{A} = [a_1, b_1, b_2, b_3, c_1, c_2, e_1, f_1]$ represents the free parameters describing the redshift evolution of the individual Schechter function parameters, $\mathrm{P}(A_i)$ is the prior on the ith free parameter, $\Phi = [\phi_\mathbf{1}, \phi_\mathbf{2}, ..., \phi_\mathbf{n}]$ the set of n mass function parameters, $\mathbf{X} = [\mathbf{x_0}, \mathbf{x_1}, ..., \mathbf{x_8}]$ the measurements of the mass function in each redshift bin, and $\phi_z = [\mathrm{M}^*_{1,\mathrm{M},z}, \phi^*_{1,\mathrm{M},z}, \alpha_{1,\mathrm{M},z}, \mathrm{M}^*_{2,\mathrm{M},z}, \phi^*_{2,\mathrm{M},z}, \alpha_{2,\mathrm{M},z}]$ the Schechter function parameters in each redshift bin. We fit with weakly informative priors (broad Gaussian distributions) on each hyper-parameter, with the added constraint that b_3 be negative to ensure that the normalization of $\Phi_1(\mathrm{M})$ decreases with redshift, and e_1 and f_1 be positive. All priors are reported in Table 4. We assume that the measurements represent true measurements of the mass function at the mid-point of each redshift bin ($\mathbf{z} = [0.35, 0.625, 0.875, 1.125, 1.375, 1.75, 2.25, 2.75]$), and so do not include any modeling of the redshift distribution of the underlying sources that make up the mass bin volume density measurements.

The mass function estimates reported in T14 are supplied as $\log[\Phi(\mathrm{M}_\star)]$ (see their Table 1), with no information of the number of galaxies entering each bin. In fact, the associated errors adopted from T14 account for Poisson noise, cosmic variance, the uncertainties arising from classifying galaxies as star-forming or quiescent and from the determination of stellar masses, and so are not simply Poissonian errors. Thus, in the absence of the information required to construct the correct function for our count distribution, we resort to a Gaussian assumption, giving:

$$\mathrm{P}(\mathbf{x}_z \mid \phi_z) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\Big(-\frac{(x_i - \hat{x_i}(\phi_z))^2}{2\sigma_i^2}\Big) \tag{A2}$$

where x_i and σ_i are the estimate and error of the mass function in bin i respectively and $\hat{x_i}(\phi_z)$ is the model prediction.

We sample the posterior distribution of model parameters with a Metropolis-within-Gibbs sampler (see, e.g., Sharma 2017). Gibbs sampling involves drawing directly from the conditional distribution for each parameter in turn. However, when the conditional probability is tricky to derive, it is possible to use the Metropolis update step, where the next step in the chain is sampled from a proposal distribution, often a Gaussian, and accepted or rejected based on comparison of the posterior probability between the current and last steps of the chain. The advantage of using a Gibbs sampler for this problem is that the conditional distribution of the model parameters in each redshift bin are independent of each other, meaning that $\mathrm{P}(\phi_i \mid ...\phi_{i-1}, \phi_{i+1}, ..., \mathbf{A}, \mathbf{X}) = \mathrm{P}(\mathbf{x}_i \mid \phi_i)\mathrm{P}(\phi_i \mid \mathbf{A})$ and so the mass function parameters for each redshift bin can be sampled in turn, before updating the values of the hyper-parameters.

During each iteration of the chain our multi-level modeling algorithm performs the following steps:

1. Values for the double Schechter function parameters are proposed for each individual redshift bin in turn, sampling from $\mathrm{P}(\phi_i \mid ...\phi_{i-1}, \phi_{i+1}, ..., \mathbf{A}, \mathbf{X})$, using a Metropolis update step.

2. Once the mass function parameters of each redshift bin has been updated, new values for the hyper parameters are proposed. They are updated simultaneously, sampling from $\mathrm{P}(\mathbf{A} \mid \Phi, \mathbf{X})$, again using a Metropolis update step.

After an initial burn-in stage of 10000 iterations, the width of the Gaussian proposal distribution is set using the covariance matrix of the past history of the chains, following the adaptive Metropolis algorithm of Haario et al. (2001). We run two different chains with independent starting points for 40000 iterations each, after which we test for convergence using the scale reduction factor, $\hat{R}$, which compares the within-chain and between-chain variance. We require $\hat{R} < 1.05$ in each free parameter.

After rejecting the initial 10000 iterations, we estimate each parameter value as the median of the values in the chain and uncertainties as their standard deviation.

B. UNCERTAINTIES IN STAR-FORMING/QUIESCENT SELECTION CRITERIA AND IMPACT ON CREATING PARENT CATALOG FOR HIGH-MASS QUIESCENT GALAXIES FROM 3D-HST DATA

When drawing multiple realizations from SED fits to individual quiescent galaxies we often find that the uncertainties on the U, V and J-band absolute magnitudes allow for solutions that would place the object in the star-forming region of the UVJ color space. This is because galaxies are selected from a combined J_{125}, J_{140} and H_{160} band image in the 3D-HST catalog (Section 3.4.1). Since quiescent galaxies have red SEDs, their fainter fluxes at shorter wavelengths can therefore suffer from large observational uncertainties. Although the uncertainties on the star-forming and quiescent galaxy classifications are included in

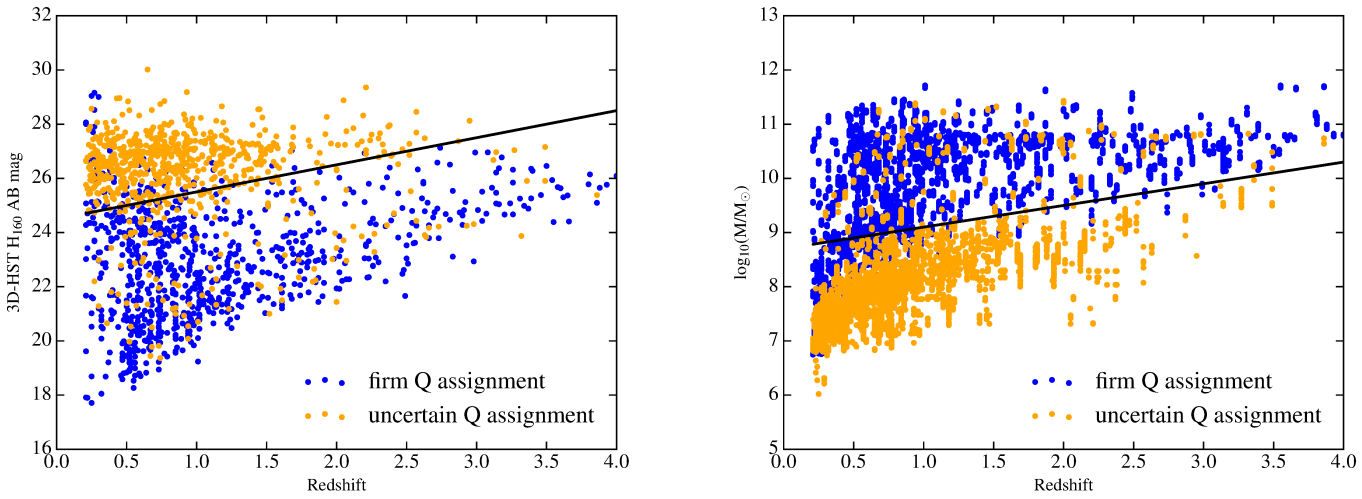

Figure 28. Left panel: 3D-HST H_{160} magnitude plotted against redshift for a subset of the galaxies classified as quiescent (Q) from their rest-frame $U-V$ and $V-J$ colors. The blue points are objects that have at least 45 realizations that are assigned as quiescent from a random sampling of 50 realizations within the 68% confidence contour on stellar mass. The orange points have between 20-30 realizations assigned as quiescent. The black line is $H_{160} = 24.5+z$ shows the approximate division between the two categories and is used to determine objects which are robustly assigned as quiescent. Right panel: model $M_\star$ vs redshift for each

the T14 stellar mass function estimates, we must ensure that galaxies designated as quiescent in our mock catalog are assigned quiescent SEDs. In the case of assigning SEDs to our star-forming galaxy mock, the prior on $M_{\rm UV}$–$M_\star$ used to weight the draws from the BEAGLE fitting prevents the assignment of quiescent SEDs. However, for the quiescent parent catalog we only require that the realizations are drawn from within the 68% contour on stellar mass and this requirement does not explicitly require the SED to also be quiescent.

To characterize the redshift-dependent limiting magnitude at which the quiescent assignment based on UVJ colors is robust, we randomly sample 50 solutions from within the 68% uncertainty in mass for each quiescent object in the 3D-HST catalog and re-measure the UVJ colors, providing a fraction of draws per object that reside in the quiescent region of UVJ color space. We compare in Figure 28 the distribution of H_{160} magnitudes of objects for which at least 45 of these 50 realizations fall in the quiescent region of the UVJ color space (in blue) to that of objects for which between 20 and 30 realizations fall outside of the quiescent region (in orange). Based on these results, we limit the objects that enter the 3D-HST-derived parent catalog based on their UVJ colors to $H_{160} < 24.5+z$ (black line in Figure 28). Given that the H_{160} band samples the SED red-wards of the 4000Å break for $z \lesssim 2.5$, we expect this limit in H_{160} to approximately correspond to a limit in stellar mass. We therefore plot the stellar mass as a function of redshift for those objects with firm and uncertain quiescent assignments in Figure 28, right panel. Using this figure we set a mass limit of $\log(M/M_\odot) > 8.7+0.4z$ for firm quiescent galaxy assignment. We note that this limit is likely to be conservative at $z > 2.5$, where the H_{160}-band magnitude no longer correlates strongly with stellar mass. In practice, only objects from our mock catalog with masses above this limit will be matched to the observationally-derived parent galaxy catalog. Objects below this mass will be matched to the parent galaxy catalog produced by BEAGLE.

REFERENCES

Alavi, A., Siana, B., Richard, J., et al. 2014, ApJ, 780, 143

Arnouts, S., Schiminovich, D., Ilbert, O., et al. 2005, ApJL, 619, L43

Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Atek, H., Siana, B., Scarlata, C., et al. 2011, ApJ, 743, 121

Atek, H., Richard, J., Kneib, J.-P., et al. 2015, ApJ, 800, 18

Bagnasco, G., Kolm, M., Ferruit, P., et al. 2007, in Proc. SPIE, Vol. 6692, Cryogenic Optical Systems and Instruments XII, 66920M

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729

Bell, E. F., van der Wel, A., Papovich, C., et al. 2012, ApJ, 753, 167

Benson, A. J. 2012, NewA, 17, 175

Bernyk, M., Croton, D. J., Tonini, C., et al. 2016, ApJS, 223, 9
Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228
Birkmann, S. M., Ferruit, P., Rawle, T., et al. 2016, in Proc. SPIE, Vol. 9904, Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave, 99040B
Blaizot, J., Wadadekar, Y., Guiderdoni, B., et al. 2005, MNRAS, 360, 159
Bouwens, R. J., Illingworth, G. D., Franx, M., & Ford, H. 2007, ApJ, 670, 928
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2017a, ApJ, 843, 41
Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017b, ApJ, 843, 129
Bouwens, R. J., Illingworth, G. D., Franx, M., et al. 2009, ApJ, 705, 936
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, ApJ, 737, 90
—. 2012a, ApJL, 752, L5
—. 2012b, ApJ, 754, 83
—. 2014, ApJ, 793, 115
—. 2015, ApJ, 803, 34
Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016a, ApJ, 833, 72
Bouwens, R. J., Oesch, P. A., Labbé, I., et al. 2016b, ApJ, 830, 67
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2017, MNRAS, 466, 3612
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Brinchmann, J., Pettini, M., & Charlot, S. 2008, MNRAS, 385, 769
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bunker, A. J., Stanway, E. R., Ellis, R. S., & McMahon, R. G. 2004, MNRAS, 355, 374
Cai, Y.-C., Angulo, R. E., Baugh, C. M., et al. 2009, MNRAS, 395, 1185
Calvi, V., Trenti, M., Stiavelli, M., et al. 2016, ApJ, 817, 120
Carton, D., Brinchmann, J., Shirazi, M., et al. 2017, MNRAS, 468, 2140
Cassata, P., Giavalisco, M., Williams, C. C., et al. 2013, ApJ, 775, 106
Castellano, M., Fontana, A., Grazian, A., et al. 2012, A&A, 540, A39
Castellano, M., Yue, B., Ferrara, A., et al. 2016, ApJL, 823, L40
Castellano, M., Pentericci, L., Fontana, A., et al. 2017, ApJ, 839, 73
Ceverino, D., Glover, S. C. O., & Klessen, R. S. 2017, MNRAS, 470, 2791
Chabrier, G. 2003, PASP, 115, 763
Charlot, S., & Fall, M. 2000, ApJ, 539, 718
Charlot, S., & Longhetti, M. 2001, MNRAS, 323, 887
Chen, Y., Bressan, A., Girardi, L., et al. 2015, MNRAS, 452, 1068
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
Chevallard, J., Curtis-Lake, E., Charlot, S., et al. 2017, ArXiv e-prints, arXiv:1711.07481
Coe, D., Zitrin, A., Carrasco, M., et al. 2013, ApJ, 762, 32
Conselice, C. J. 2014, ARA&A, 52, 291
Cooper, M. C., Griffith, R. L., Newman, J. A., et al. 2012, MNRAS, 419, 3018
Cowley, W. I., Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 2018, MNRAS, 474, 2352
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Cucciati, O., Tresse, L., Ilbert, O., et al. 2012, A&A, 539, A31
Curtis-Lake, E., McLure, R. J., Dunlop, J. S., et al. 2013, MNRAS, 429, 302
—. 2016, MNRAS, 457, 440
da Cunha, E., Charmandaris, V., Díaz-Santos, T., et al. 2010, A&A, 523, A78
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Damen, M., Labbé, I., Franx, M., et al. 2009, ApJ, 690, 937
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
de Barros, S., Schaerer, D., & Stark, D. P. 2014, A&A, 563, A81
De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2
Delaye, L., Huertas-Company, M., Mei, S., et al. 2014, MNRAS, 441, 203
Devriendt, J. E. G., Guiderdoni, B., & Sadat, R. 1999, A&A, 350, 381
Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960
Dunlop, J. S., McLure, R. J., Robertson, B. E., et al. 2012, MNRAS, 420, 901
Dunlop, J. S., Rogers, A. B., McLure, R. J., et al. 2013, MNRAS, 432, 3520
Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJL, 763, L7
Elmegreen, B. G., Elmegreen, D. M., Fernandez, M. X., & Lemonias, J. J. 2009, ApJ, 692, 12
Faber, S. M., & Jackson, R. E. 1976, ApJ, 204, 668
Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016, ApJ, 821, 122
Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, ApJS, 154, 10
Feltre, A., Charlot, S., & Gutkin, J. 2016, MNRAS, 456, 3354
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Finkelstein, S. L. 2016, PASA, 33, e037
Finkelstein, S. L., Papovich, C., Ryan, R. E., et al. 2012a, ApJ, 758, 93
Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012b, ApJ, 756, 164

Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., et al. 2015, ApJ, 810, 71
Finlator, K., Oppenheimer, B. D., & Davé, R. 2011, MNRAS, 410, 1703
Fontanot, F., Somerville, R. S., Silva, L., Monaco, P., & Skibba, R. 2009, MNRAS, 392, 553
Foreman-Mackey, D. 2016, The Journal of Open Source Software, 24
Förster Schreiber, N. M., Shapley, A. E., Erb, D. K., et al. 2011, ApJ, 731, 65
Franx, M., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2008, ApJ, 688, 770
Furlanetto, S. R., Mirocha, J., Mebane, R. H., & Sun, G. 2017, MNRAS, 472, 1576
Gallazzi, A., Bell, E. F., Zibetti, S., Brinchmann, J., & Kelson, D. D. 2014, ApJ, 788, 72
Gallazzi, A., Charlot, S., Brinchmann, J., & White, S. D. M. 2006, MNRAS, 370, 1106
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, SSRv, 123, 485
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJL, 600, L93
Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, Nature, 544, 71
González, V., Bouwens, R., Illingworth, G., et al. 2014, ApJ, 781, 34
González, V., Labbé, I., Bouwens, R. J., et al. 2011, ApJL, 735, L34
Grazian, A., Castellano, M., Fontana, A., et al. 2012, A&A, 547, A51
Grazian, A., Fontana, A., Santini, P., et al. 2015, A&A, 575, A96
Grogin, N. A. e. a. 2011, ApJS, 197, 35
Guiderdoni, B., & Rocca-Volmerange, B. 1987, A&A, 186, 1
Guo, Y., McIntosh, D. H., Mo, H. J., et al. 2009, MNRAS, 398, 1129
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
Gutkin, J., Charlot, S., & Bruzual, G. 2016, MNRAS, 462, 1757
Haario, H., Saksman, E., & Tamminen, J. 2001, Bernoulli, 7, 223
Hainline, K. N., Shapley, A. E., Kornei, K. A., et al. 2009, ApJ, 701, 52
Hamann, W.-R., & Gräfener, G. 2004, A&A, 427, 697
Hathi, N. P., Ryan, Jr., R. E., Cohen, S. H., et al. 2010, ApJ, 720, 1708
Hirschmann, M., Charlot, S., Feltre, A., et al. 2017, MNRAS, 472, 2468
Holwerda, B. W., Bouwens, R., Oesch, P., et al. 2015, ApJ, 808, 6
Horner, S. D., & Rieke, M. J. 2004, in Proc. SPIE, Vol. 5487, Optical, Infrared, and Millimeter Space Telescopes, ed. J. C. Mather, 628
Hunt, L., Dayal, P., Magrini, L., & Ferrara, A. 2016, MNRAS, 463, 2002
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, MNRAS, 442, 1805
Ishigaki, M., Kawamata, R., Ouchi, M., Oguri, M., & Shimasaku, K. 2017, ArXiv e-prints, arXiv:1702.04867
Jørgensen, I., & Chiboucas, K. 2013, AJ, 145, 77
Juneau, S., Bournaud, F., Charlot, S., et al. 2014, ApJ, 788, 88
Kajisawa, M., Ichikawa, T., Tanaka, I., et al. 2011, PASJ, 63, 379
Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, ApJ, 835, 88
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33
Kawamata, R., Ishigaki, M., Shimasaku, K., Oguri, M., & Ouchi, M. 2015, ApJ, 804, 103
Kawinwanichakij, L., Papovich, C., Quadri, R. F., et al. 2017, ApJ, 847, 134
Kennicutt, Jr., R. C. 1998, ApJ, 498, 541
Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013, ApJ, 774, 100
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221
Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, ApJS, 218, 15
Kroupa, P. 2001, MNRAS, 322, 231
Kurczynski, P., Gawiser, E., Rafelski, M., et al. 2014, ApJL, 793, L5
Labbé, I., Franx, M., Rudnick, G., et al. 2007, ApJ, 665, 944
Labbé, I., González, V., Bouwens, R. J., et al. 2010, ApJL, 716, L103
Labbé, I., Oesch, P. A., Bouwens, R. J., et al. 2013, ApJL, 777, L19
Laporte, N., Streblyanska, A., Kim, S., et al. 2015, A&A, 575, A92
Lee, B., Giavalisco, M., Williams, C. C., et al. 2013, ApJ, 774, 47
Lee, K.-S., Dey, A., Reddy, N., et al. 2011, ApJ, 733, 99
Leitherer, C., Ortiz Otálvaro, P. A., Bresolin, F., et al. 2010, ApJS, 189, 309
Liu, C., Mutch, S. J., Angel, P. W., et al. 2016, MNRAS, 462, 235
Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113
Lu, Y., Wechsler, R. H., Somerville, R. S., et al. 2014, ApJ, 795, 123
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Mármol-Queraltó, E., McLure, R. J., Cullen, F., et al. 2016, MNRAS, 460, 3587

Mashian, N., Oesch, P. A., & Loeb, A. 2016, MNRAS, 455, 2101
Mason, C. A., Trenti, M., & Treu, T. 2015, ApJ, 813, 21
Masters, D., Faisst, A., & Capak, P. 2016, ApJ, 828, 18
McAlpine, S., Helly, J. C., Schaller, M., et al. 2016, Astronomy and Computing, 15, 72
McLeod, D. J., McLure, R. J., & Dunlop, J. S. 2016, MNRAS, 459, 3812
McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2015, MNRAS, 450, 3032
McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, MNRAS, 432, 2696
Mehta, V., Scarlata, C., Rafelski, M., et al. 2017, ApJ, 838, 29
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Mirocha, J., Furlanetto, S. R., & Sun, G. 2017, MNRAS, 464, 1365
Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, ApJ, 808, 101
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Mortlock, A., Conselice, C. J., Hartley, W. G., et al. 2013, MNRAS, 433, 1185
Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18
Naab, T., & Ostriker, J. P. 2017, ARA&A, 55, 59
Noeske, K. G., Faber, S. M., Weiner, B. J., et al. 2007, ApJL, 660, L47
Nonino, M., Dickinson, M., Rosati, P., et al. 2009, ApJS, 183, 244
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbe, I., & Stefanon, M. 2017, ArXiv e-prints, arXiv:1710.11131
Oesch, P. A., Bouwens, R. J., Carollo, C. M., et al. 2010, ApJL, 725, L150
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2012, ApJ, 745, 110
—. 2013, ApJ, 773, 75
—. 2014, ApJ, 786, 108
Ono, Y., Ouchi, M., Curtis-Lake, E., et al. 2013, ApJ, 777, 155
Ono, Y., Ouchi, M., Harikane, Y., et al. 2017, PASJ, arXiv:1704.06004
Orsi, Á., Padilla, N., Groves, B., et al. 2014, MNRAS, 443, 799
Pacifici, C., da Cunha, E., Charlot, S., et al. 2015, MNRAS, 447, 786
Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, MNRAS, 456, 3194
Pei, Y. C. 1992, ApJ, 395, 130
Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pierini, D., Gordon, K. D., Witt, A. N., & Madsen, G. J. 2004, ApJ, 617, 1022
Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, AJ, 150, 31
Rasappu, N., Smit, R., Labbé, I., et al. 2016, MNRAS, 461, 3886
Rauscher, B. J., Fox, O., Ferruit, P., et al. 2007, PASP, 119, 768
Reddy, N. A., & Steidel, C. C. 2009, ApJ, 692, 778
Reddy, N. A., Kriek, M., Shapley, A. E., et al. 2015, ApJ, 806, 259
Retzlaff, J., Rosati, P., Dickinson, M., et al. 2010, A&A, 511, A50
Roberts-Borsani, G. W., Bouwens, R. J., Oesch, P. A., et al. 2016, ApJ, 823, 143
Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19
Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71
Rogers, A. B., McLure, R. J., & Dunlop, J. S. 2013, MNRAS, 429, 2456
Rogers, A. B., McLure, R. J., Dunlop, J. S., et al. 2014, MNRAS, 440, 3714
Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183
Salpeter, E. E. 1955, ApJ, 121, 161
Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703
Sánchez-Blázquez, P., Jablonka, P., Noll, S., et al. 2009, A&A, 499, 47
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, ApJ, 799, 138
—. 2016, ApJ, 816, 23
Santini, P., Fontana, A., Grazian, A., et al. 2012, A&A, 538, A33
Sawicki, M. 2012, MNRAS, 421, 2187
Sawicki, M., & Thompson, D. 2006, ApJ, 642, 653
Schaerer, D., & de Barros, S. 2009, A&A, 502, 423
—. 2010, A&A, 515, A73
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Schechter, P. 1976, ApJ, 203, 297
Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2013, ApJ, 777, 67
Schiminovich, D., Ilbert, O., Arnouts, S., et al. 2005, ApJL, 619, L47
Schmidt, M. 1959, ApJ, 129, 243
Schreiber, C., Elbaz, D., Pannella, M., et al. 2017, A&A, 602, A96
Sersic, J. L. 1968, Atlas de Galaxias Australes
Shapley, A. E., Coil, A. L., Ma, C.-P., & Bundy, K. 2005, ApJ, 635, 1006
Shapley, A. E., Sanders, R. L., Reddy, N. A., et al. 2017, ArXiv e-prints, arXiv:1707.05331
Sharma, S. 2017, ARA&A, 55, 213
Shen, S., Mo, H. J., White, S. D. M., et al. 2003, MNRAS, 343, 978
Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15
Shim, H., Chary, R.-R., Dickinson, M., et al. 2011, ApJ, 738, 69
Shimizu, I., Inoue, A. K., Okamoto, T., & Yoshida, N. 2016, MNRAS, 461, 3563
Simha, V., Weinberg, D. H., Conroy, C., et al. 2014, ArXiv e-prints, arXiv:1404.0402

Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24
Skilling, J., et al. 2006, Bayesian Analysis, 1, 833
Smit, R., Bouwens, R. J., Labbé, I., et al. 2016, ApJ, 833, 254
—. 2014, ApJ, 784, 58
Smit, R., Bouwens, R. J., Franx, M., et al. 2015, ApJ, 801, 122
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
Somerville, R. S., Popping, G., & Trager, S. C. 2015, MNRAS, 453, 4337
Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, ApJ, 825, 5
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Spitler, L. R., Straatman, C. M. S., Labbé, I., et al. 2014, ApJL, 787, L36
Stanway, E. R., McMahon, R. G., & Bunker, A. J. 2005, MNRAS, 359, 1184
Stark, D. P. 2016, ARA&A, 54, 761
Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, ApJ, 697, 1493
Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2017a, ApJ, 843, 36
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2017b, ArXiv e-prints, arXiv:1706.04613
Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJL, 783, L14
—. 2015, ApJL, 808, L29
Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, ApJ, 830, 51
Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164
Sun, G., & Furlanetto, S. R. 2016, MNRAS, 460, 417
Tacchella, S., Trenti, M., & Carollo, C. M. 2013, ApJL, 768, L37
Tasca, L. A. M., Le Fèvre, O., Hathi, N. P., et al. 2015, A&A, 581, A54
Teplitz, H. I., Rafelski, M., Kurczynski, P., et al. 2013, AJ, 146, 159
Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2014, ApJ, 783, 85
Torrey, P., Snyder, G. F., Vogelsberger, M., et al. 2015, MNRAS, 447, 2753
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Trenti, M., & Stiavelli, M. 2008, ApJ, 676, 767
Trenti, M., Stiavelli, M., Bouwens, R. J., et al. 2010, ApJL, 714, L202
Tuffs, R. J., Popescu, C. C., Völk, H. J., Kylafis, N. D., & Dopita, M. A. 2004, A&A, 419, 821
Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661
van der Burg, R. F. J., Hildebrandt, H., & Erben, T. 2010, A&A, 523, A74
van der Marel, R. P., & Franx, M. 1993, ApJ, 407, 525
van der Wel, A., Bell, E. F., Holden, B. P., Skibba, R. A., & Rix, H.-W. 2010, ApJ, 714, 1779
van der Wel, A., Rix, H.-W., Wuyts, S., et al. 2011, ApJ, 730, 38
van der Wel, A., Bell, E. F., Häussler, B., et al. 2012, ApJS, 203, 24
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van Dokkum, P. G., Nelson, E. J., Franx, M., et al. 2015, ApJ, 813, 23
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518
Weinmann, S. M., Pasquali, A., Oppenheimer, B. D., et al. 2012, MNRAS, 426, 2797
Weisz, D. R., Johnson, B. D., & Conroy, C. 2014, ApJL, 794, L3
Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, ApJ, 735, 86
Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104
Wilkins, S. M., Bouwens, R. J., Oesch, P. A., et al. 2016, MNRAS, 455, 659
Wilkins, S. M., Bunker, A. J., Stanway, E., Lorenzoni, S., & Caruana, J. 2011, MNRAS, 417, 717
Williams, C. C., Giavalisco, M., Cassata, P., et al. 2014, ApJ, 780, 1
Williams, C. C., Giavalisco, M., Lee, B., et al. 2015, ApJ, 800, 21
Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
Wuyts, S., Labbé, I., Förster Schreiber, N. M., et al. 2008, ApJ, 682, 985
Wyder, T. K., Treyer, M. A., Milliard, B., et al. 2005, ApJL, 619, L15
Yue, B., Castellano, M., Ferrara, A., et al. 2017, ArXiv e-prints, arXiv:1711.05130
Zheng, W., Postman, M., Zitrin, A., et al. 2012, Nature, 489, 406