

Paper I: The Main Sequence of Simulated Galaxies

CHANGHOON HAHN,^{1,2,*} TJITSKE K. STARKENBURG,³ CLAIRE DICKEY,⁴ AND IQ COLLABORATORY (IN ALPHABETICAL ORDER)⁵

¹*Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley CA 94720, USA*

²*Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA*

³*Flatiron Institute, 162 Fifth Avenue, New York NY 10010, USA*

⁴*Yale University*

⁵*various*

ABSTRACT

Principled comparisons between the galaxy populations of simulations and observations play a pivotal role in validating our theories of galaxy formation and evolution. Key features in galaxy property-space like the star forming main sequence (SFMS) and a consistent way to identify them are critical for such comparisons. We present a data-driven approach to fitting the SFMS using Gaussian Mixture Models that can be flexibly applied to a wide range of star formation to stellar mass relations down to $M_* \sim 10^8 M_\odot$. Using this method, we identify the SFMS of central galaxies in the Illustris, EAGLE, and MUFASA hydrodynamic simulations, the Santa Cruz Semi-Analytic Model simulation, and observations from the Sloan Digital Sky Survey Data Release 7 and the NASA Sloan Atlas. Among our simulations, we find discrepancies on the order of a magnitude in the amplitudes of the SFMSs. In addition, our fitting method also identifies subpopulations that correspond to quenched, transitioning, and star-burst galaxies. We find overall consistent subpopulations among our hydrodynamic simulations but differences with the semi-analytic model. Using these subpopulations, we find that the quiescent fractions of the hydrodynamic simulations do not reproduce measurements from observations. Moreover, in *all* of the simulations we find a significant fraction of quenched galaxies at $M_* < 10^9 M_\odot$, in conflict with the literature. The SFMS fitting method we present provides a data-driven framework to consistently compare galaxy samples from both simulations and observations.

Keywords: cosmology: observations — galaxies: star formation — galaxies: statistics

1. INTRODUCTION

Large galaxy surveys of the past decade such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), have firmly established the major trends of galaxies in the local universe. Galaxies broadly fall into two populations: quiescent galaxies with little star formation that are red in color with elliptical

morphologies and star forming galaxies with significant star formation that are blue in color with disk-like morphologies (Kauffmann et al. 2003; Blanton et al. 2003; Baldry et al. 2006; Taylor et al. 2009; Moustakas et al. 2013; for a recent review see Blanton & Moustakas 2009). Star forming galaxies, furthermore, are found to have a tight relationship between their star formation rates (SFR) and stellar masses placing them on the so-called “star formation main sequence” (hereafter

* changhoonhahn@lbl.gov

SFMS) (*e.g.* Noeske et al. 2007; Daddi et al. 2007; Salim et al. 2007).

In fact, this sequence of star forming galaxies is found in observations well beyond the local universe out to $z > 2$ (Wang et al. 2013; Schreiber et al. 2015). But more than its persistence, the SFMS plays a crucial role in characterizing the evolving galaxy population (but see also Leja et al. 2015). The most dramatic transformations of galaxies over the past 10 Gyr can be described by the SFMS. For instance, the decline in the number density of massive star forming galaxies and the accompanying growth in number density of quiescent galaxies reflects the cessation of star formation in star forming galaxies migrating off of the SFMS (Blanton 2006; Borch et al. 2006; Bundy et al. 2006; Moustakas et al. 2013). Similarly, the cosmic decline in star formation (Hopkins & Beacom 2006; Behroozi et al. 2013a; Madau & Dickinson 2014) reflects the overall decline of star formation of the SFMS (Schreiber et al. 2015).

Numerical simulations today *qualitatively* reproduce the SFMS and similar global relations of galaxy properties and provide insights into the key physical processes governing those relations (*e.g.* Vogelsberger et al. 2014; Genel et al. 2014; Schaye et al. 2015; Davé et al. 2017a; for a recent review see Somerville & Davé 2015). These hydrodynamic and semi-analytic simulations each seek to capture the complex physics of gas heating and cooling, star formation, stellar feedback, chemical evolution, black hole formation and evolution, and AGN feedback using their distinct sub-grid model prescriptions. Both as an effort to shed light on the underlying physics and to validate their simulations and make predictions for new observations, many works have already compared simulations to observations of for example galaxy masses, colors, and star formation rates (*e.g.* Vogelsberger et al. 2014; Genel et al. 2014; Torrey et al. 2014; Sparre et al. 2015; Schaye et al. 2015; Bluck et al. 2016; Davé et al. 2017a). These works, however, primarily focus on comparing one spe-

cific simulated galaxy sample to one or a few observational datasets (which can differ significantly as well as demonstrated in Speagle et al. 2014). Extending such comparisons to include multiple simulations, observations, and a consistent framework for comparing the data-sets would allow us to make detailed comparison of the different sub-grid models and thereby provide key insights into the physics that govern galaxy formation and evolution.

The SFMS, given its prominence, naturally presents itself as a key feature in the data-space of galaxy properties to compare galaxy populations across both observations and simulations. Moreover, with the important role it plays for understanding galaxy evolution, the SFMS provides a way to understand the star forming and non-star forming galaxy populations and the processes that create them. Two main challenges lie in conducting a principled comparison of the SFMS in galaxy populations. First is the lack of a universal data-driven method for identifying the SFMS given a dataset of galaxy properties. The other is the difference in methodology for deriving galaxy properties (SFR, M_*) in different data-sets, which alone dramatically impacts the SFMS (*e.g.* Speagle et al. 2014). In this paper we address the first challenge by presenting a flexible, data-driven method for fitting the SFMS. Then we use this method to conduct a principled comparison between the central galaxy populations from the Illustris, EAGLE, and MUFASA hydrodynamic simulations, the Santa Cruz Semi-Analytic Model (SAM) simulation and observations from the SDSS and NASA-Sloan Atlas (hereafter NSA) catalogs.

In Section 2, we describe the simulated and observed galaxies of our dataset and how we specifically select our galaxy sample. Then in Section 4.1, we describe our data-driven SFMS fitting method, which makes use of Gaussian Mixture Modeling. We present the results from applying our SFMS fitting to our simulated and observed galaxies in Section 4 and compare the galaxy pop-

ulations of our simulations and observations. Finally, we conclude and summarize the results of our comparison in Section 5. This paper is the first in a series studying the star formation and quenching properties of galaxies. The series is initialized by the IQ (Isolated & Quenched) Collaboratory. In the subsequent paper we will address the discrepancies in measured galaxy properties by constructing mock observational spectra of simulated galaxies in our sample and measuring the properties of these galaxies in the same manner as observations (Starkenbug et al. in prep.).

2. OUR GALAXIES

In this work, we primarily focus on simulated galaxy samples from four large-scale cosmological simulations: three hydrodynamic (Illustris, EAGLE, and MUFASA) and one semi-analytic (Santa Cruz SAM). Although we mainly focus on the simulated galaxies, for reference, we also include two observed galaxy samples: a volume-limited sample from SDSS and a selected isolated dwarf galaxy sample from the NSA catalog. We briefly describe the four simulations in Section 2.1 and our observed galaxy sample in Section 2.2. Lastly in Section 2.3, we describe how we consistently identify the central galaxies in both our simulations and observations.

2.1. Simulated Galaxies

A consistent comparison of the different galaxy populations in our simulations requires consistently defined galaxy properties. Hence, for all of our simulated galaxies, we derive SFRs consistently on two timescales: instantaneous and averaged over 100 Myr. Our choice of these timescales is respectively motivated by the fact that observed SFR measurements based on $H\alpha$ correspond to formation of young stars with ages $\lesssim 10$ Myr and measurements based on UV brightness correspond to star formation in the last ~ 100 Myr (e.g. Kennicutt & Evans 2012).

More specifically, in our hydrodynamic simulations, we derive the 100 Myr averaged SFRs from

the ages, or formation times, of star particles in the simulated galaxies and instantaneous SFRs from the rate of star formation in the dense and cold gas. For the SAM, we derive the 100 Myr averaged SFRs from the total stellar mass formed in a galaxy, which is saved every 10 Myr, and the instantaneous SFR based on a Kennicutt-Schmidt relation for molecular hydrogen (Bigiel et al. 2008) and the derived H_2 surface density in radial bins. We note that the spatial, mass, and temporal resolution of our hydrodynamic simulation can significantly impact the averaged SFRs. In Appendix B, we discuss how we treat these resolution effects in our analysis. The total stellar mass of each (central) galaxy is defined as the total stellar mass within its halo, discounting stellar mass in any sub-halo of that halo. Even though halos are found and defined in different ways in the different simulations, as almost exclusively all stellar mass will be in the center of that halo the stellar mass definitions are consistent from simulation to simulation.

In the rest of this section we provide a brief description of the Illustris, EAGLE, MUFASA, and Santa-Cruz SAM simulations and their key sub-grid and feedback prescriptions.

2.1.1. The Illustris simulation

The Illustris simulation (Vogelsberger et al. 2014; Genel et al. 2014) evolves a cosmological volume of $(106 \text{ Mpc})^3$ with a uniform baryonic mass resolution of $1.6 \times 10^6 M_\odot$ using the Arepo moving-mesh code (Springel 2010). The full halo mass range of 2×10^8 to $3 \times 10^{14} M_\odot$ is set by resolution (32 particles) at the low-mass end and by volume at the high-mass end (e.g. including 10 halos with $M > 10^{14} M_\odot$). It employs sub-grid models for star-formation (Springel & Hernquist 2003), Bondi-like SMBH accretion, a phenomenological model for galactic winds, and two main modes for energy injection from SMBHs (see Vogelsberger et al. 2013). When accretion occurs at Eddington ratios > 0.05 , thermal energy is injected continuously in the local environment of the SMBH, while at lower accretion rates, the energy injection

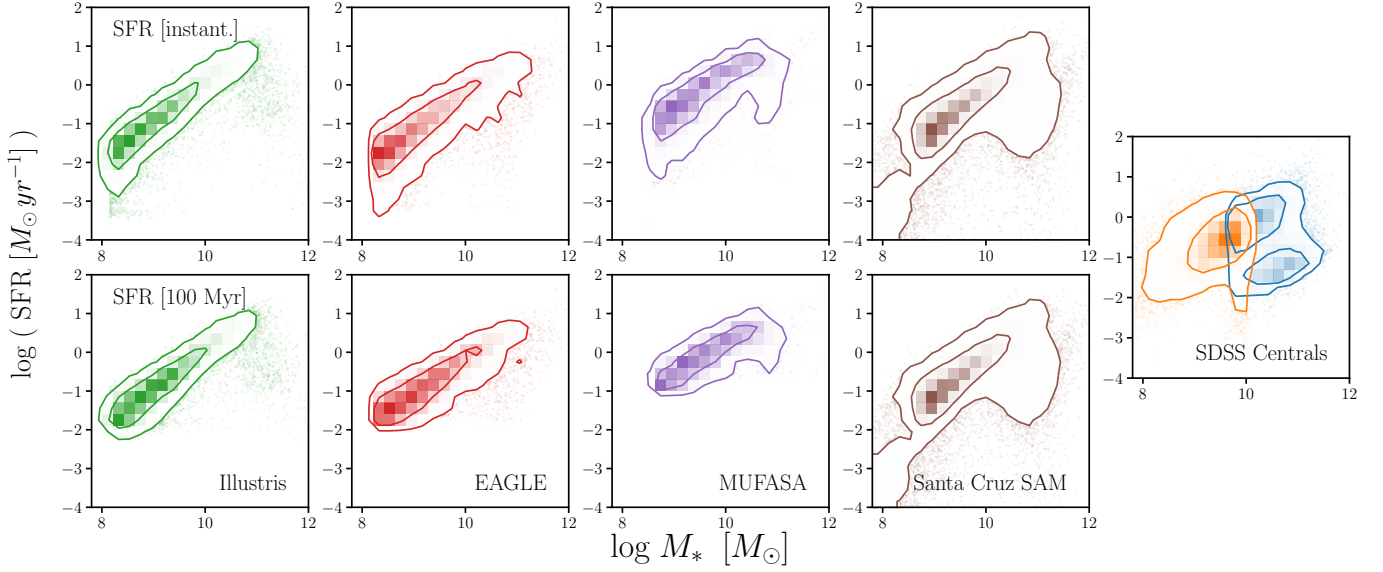


Figure 1. The SFR– M_* relations of central galaxies from the Illustris, EAGLE, MUFASA, and Santa Cruz SAM simulations (left to right). The top panels use instantaneous SFRs while the bottom panels use SFRs averaged over 100 Myr. The simulations and how they derive the SFRs are described in Section 2.1. Although a direct comparison to observations is tenuous due to the fact that the SFRs and M_* s of the observed SDSS galaxies are *not* derived consistently as simulations, we include, for reference, the observed SDSS galaxies (Section 2.2) on the right. *The SFR – M_* relations in every panel reveals a clear star forming main sequence.*

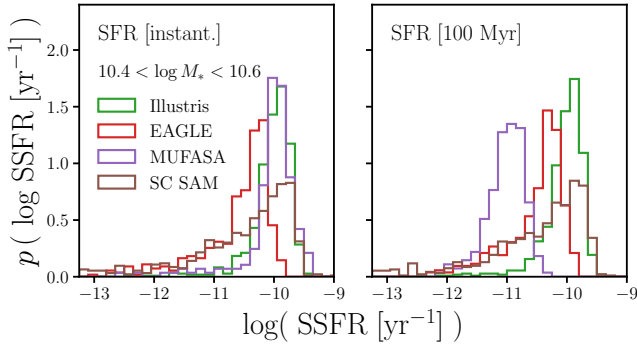


Figure 2. The SSFR distributions, $p(\log \text{SSFR})$, of the central galaxies in the Illustris (green), EAGLE (red), MUFASA (purple), and Santa Cruz SAM (brown) simulations with $10.4 < \log M_* < 10.6$. We use instantaneous SFRs on the left and SFRs averaged over 100 Myr on the right. Although the SFMS is universal (Figure 1), *the significant discrepancies among the $p(\log \text{SSFR})$ s, and thus SFR– M_* relations, make the SFMS difficult to consistently quantify.*

occurs in bursts at large distances from the SMBH, generating hot bubbles in the ICM (Sijacki et al.

2007). Previous works discussing aspects of the star-formation main-sequence and/or quenching in the Illustris simulation include Vogelsberger et al. (2014); Sparre et al. (2015); Bluck et al. (2016); Terrazas et al. (2017).

2.1.2. EAGLE

The Virgo Consortium’s Evolution and Assembly of GaLaxies and their Environment (EAGLE) project (Schaye et al. 2015; Crain et al. 2015) exists as a suite of cosmological, hydrodynamic simulations of a standard Λ cold dark matter universe using ANARCHY (Dalla Vecchia et al. in prep.; see also Appendix A of Schaye et al. 2015 and Schaller et al. 2015), a modified version of the Gadget 3 N -body/SPH code (Springel 2005). Modifications include the SPH formulation, time stepping, and subgrid physics. The subgrid model for feedback from massive stars and AGN is based on thermal energy injection in the ISM without the need to turn-off cooling or hydrodynamic decoupling of winds (Dalla Vecchia & Schaye 2012). Similar to

semi-analytical models the subgrid parameters for stellar feedback and BH accretion are calibrated based on present-day galaxy stellar mass function while also requiring reasonable galaxy sizes. Additionally the the AGN feedback efficiency is constrained by the central black hole-galaxy mass relation. The simulations resolve galaxies over the stellar mass range $10^8 < M_*/M_\odot \lesssim 10^{11}$ and reproduce the galaxy stellar mass function to $\lesssim 0.2$ dex over this range. Of the EAGLE project simulations, we use L0100Ref, which has a volume of $(100 \text{ comoving Mpc})^3$ and baryonic mass resolution of $1.81 \times 10^6 M_\odot$. The SFR- M_* relation and quenched fractions in the EAGLE simulations have been previously discussed in [Furlong et al. \(2015\)](#); [Trayford et al. \(2015, 2017\)](#).

2.1.3. MUFASA

The MUFASA simulation uses GIZMO, a code built upon GADGET where we employ the Meshless Finite Mass (MFM) hydrodynamics method (?) rather than SPH. Additionally, it includes star formation via a Kennicutt-Schmidt law based on the molecular hydrogen density as computed using the subgrid recipe in (?); two-phase kinetic outflows with scalings as predicted in the Feedback in Realistic Environments (FIRE) simulations (?); and quenching in massive galaxies by keeping all non-self shielded gas within halos above a mass of $M_q > (1 + 0.48z)10^{12} M_\odot$ (?) near the halos' virial temperature (?). These prescriptions yield predictions that are in good agreement with a range of observations across cosmic time, including the galaxy stellar mass function evolution (?), the gas and metal content of galaxies ([Davé et al. 2017b](#)) and the colour-mass diagram (?). It has some difficulties in over-quenching satellite galaxies in massive halos ([Rafiee-Varhaei & Davé 2018](#)) and likely producing too high X-ray luminosities (Robson et al., in prep.). Modulo these caveats, MUFASA represents a state of the art model for studying galaxy assembly, particularly for central galaxies.

2.1.4. the Santa-Cruz Semi-Analytic Model [\[Rachel, Viraj\]](#)

2.2. Observed SDSS Galaxies

Our main focus in this work is to compare galaxies from our simulation. However, since the ultimate goal of the simulations are to reproduce observations, we include in our analysis galaxies from the SDSS volume-limited sample and the NSA low-luminosity galaxy sample. We provide a brief description of the two observed galaxy samples below.

For the SDSS sample, we follow the sample selection of [Tinker et al. \(2011\)](#). We construct the SDSS volume-limited galaxy sample with $M_r - 5 \log(h) < -18$ and complete in $M_* > 10^{9.7} M_\odot$ from the NYU Value-Added Galaxy Catalog (VAGC; [Blanton et al. 2005](#)) which corresponds to the SDSS Data Release 7 (DR7; [Abazajian et al. 2009](#)) at redshift $z \approx 0.04$. For further details on the sample, we refer readers to [Tinker et al. \(2011\)](#); [Wetzel et al. \(2013\)](#); [Hahn et al. \(2017\)](#).

At lower stellar masses, we use an isolated dwarf galaxy sample selected from the NSA catalog as described in [Geha et al. \(2012\)](#). Briefly, the NSA catalog is a reprocessing of SDSS DR8, optimized for low-luminosity objects. It relies on the improved background subtraction technique of [Blanton et al. \(2011\)](#). The catalog extends to $z \approx 0.055$ and includes re-calibrated spectroscopy ([Yan 2011](#); [Yan & Blanton 2012](#)) with much smaller errors. However, this recalibration is mostly relevant only at small equivalent width values and hence galaxies on the star formation main sequence will be largely unaffected by this. Dwarf galaxies are considered isolated, and selected, when the distance from a more massive host is $> 1.5 \text{ Mpc}$ ([Geha et al. 2012](#)).

For both galaxy samples, the stellar masses are estimated using the [Blanton & Roweis \(2007\)](#) kcorrect code, which assumes a [Chabrier \(2003\)](#) IMF. The SFRs are from the current release

of Brinchmann et al. (2004)¹, where they are derived using the Bruzual A. & Charlot (1993) model with the Charlot & Fall (2000) dust prescription and CLOUDY (version C90.04; Ferland 1996) emission line modeling. For galaxies classified as having an AGN or a composite spectrum, the SFR is measured from the D_n4000 index (Balogh et al. 1998). Additionally, for starforming galaxies that have low S/N spectra, the SFR is derived from the $H\alpha$ luminosity (Brinchmann et al. 2004). We emphasize that SSFRs $\lesssim 10^{-12}\text{yr}^{-1}$ should only be considered upper limits to the true value (Salim et al. 2007).

2.3. Identifying Isolated/Central Galaxies

Galaxies and their detailed properties carry the imprint of their environment (Hubble 1936; Oemler 1974; Dressler 1980; Guzzo et al. 1997, for a recent review see Blanton & Moustakas 2009). The environment dependence of, for instance, the quiescent fraction (*e.g.* Baldry et al. 2006; Peng et al. 2010; Hahn et al. 2015) and the timescale of star formation quenching (Wetzel et al. 2013; Hahn et al. 2017), suggest that different physical mechanisms act in different environments to impact galaxy star formation. Also, there may be significant discrepancies between the SFMSs of central versus satellite galaxies (Wang et al. 2018). Without delving further into the environment dependence, *we focus solely on the central galaxies, which constitute the majority of massive galaxies ($M_* > 10^{9.5}M_\odot$) at $z \sim 0$.*

Despite its long recognized importance, galaxy environment is often heterogeneously defined in the literature Muldrew et al. (2012). Even among our simulations, the classification of centrals depends on the definition of halo properties, and thus on the underlying halo finders. EAGLE and Illustris use SUBFIND (Springel et al. 2001), where halos are defined as locally overdense, gravitationally bound (sub)structures within a connected

region selected through a friend-of-friends (FOF; Davis et al. 1985) group finder. On the other hand MUFASA and the Santa Cruz SAM are based on halos found using ROCKSTAR (Behroozi et al. 2013b), which defines halos using a hierarchical phase-space based FOF technique and endeavors to maximize the consistency of the halo through time. Regardless of their differences, these central classifications use information of the underlying dark matter. Such classifications cannot be reproduced on observed galaxy samples and thus complicate comparison to observations. Therefore, we identify central galaxies in all of our simulations consistently using an extended version of the Tinker et al. (2011) group finder, designed to identify satellite/centrals in observations.

- **paragraph describing Jeremy's new group finder code**

Galaxies with $P_{\text{sat}} < 0.01$ are designated as centrals.

Overall, we find good agreement between the central classifications of the group finder and simulations. Using purity and completeness as defined in Eqs. 13 and 15 of Campbell et al. (2015), we find central galaxy classification purities of 99%, 93%, 84%, and 90% and completenesses of 86%, 89%, 91%, and 86% for the Illustris, EAGLE, MUFASA and Santa Cruz SAM simulations respectively. Differences in the purity and completeness for the simulations is likely due to the different halo finders used in the simulations as described above. For the hydrodynamic simulations, we find no significant stellar mass dependence in the purities. However, the Santa Cruz SAM has purity of 44% at M_* below $10^{8.5}M_\odot$ and 97% above. Consequently, we impose a stellar mass limit of $M_* > 10^{8.5}M_\odot$ for the Santa Cruz SAM. In addition to the overall agreement in central classifications, we confirm that the group finder classification does *not* impact the results of this paper. In the next section, we proceed to fitting the star formation main sequence of simulated galaxies identified as centrals.

¹ <http://www.mpa-garching.mpg.de/SDSS/DR7/>

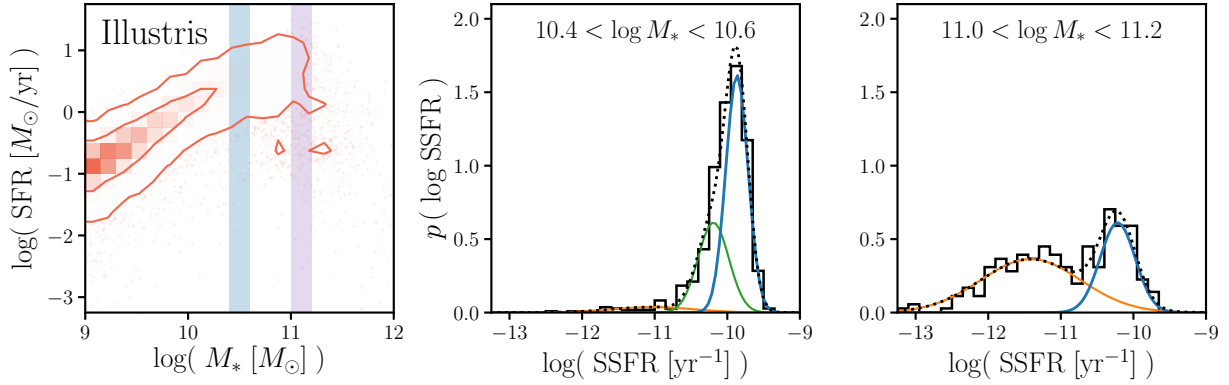


Figure 3. We illustrate our GMM SFMS fitting method for Illustris central galaxies in two stellar mass bins highlighted on the SFR– M_* relation of the left panel: $10.4 < \log M_* < 10.6$ and $11.0 < \log M_* < 11.2$. On the right, we compare the SSFR distributions, $p(\log \text{SSFR})$, in the two stellar mass bins to their best-fit GMMs. The $p(\log \text{SSFR})$ in the center panel is best described by a GMM with three components (orange, green, and blue) while the $p(\log \text{SSFR})$ in the right panel is best described by a GMM with two components (orange and blue). The SFMS components of the best-fit GMMs are plotted in blue. *Our GMM SFMS fitting provides a flexible and data-driven way to identify the SFMS for a wide variety of SSFR distributions without hard assumptions or cuts to the sample.*

3. FITTING THE STAR FORMATION MAIN SEQUENCE

We present the SFR– M_* relation of central galaxies identified from the simulations and observations of Section 2 in Figure 1. Regardless of SFR timescale (top/bottom), for both simulations and observations, and over four orders of magnitude in stellar mass, *the SFR and M_* of star-forming galaxies lie tightly correlated on the SFMS*. A well-defined SFMS can be found in each panel of Figure 1.

Despite its universality among multiple observations and simulations, different datasets give rise to significantly SFR– M_* distributions. This makes the SFMS difficult to consistently quantify. So far in the literature, a wide variety of fitting methods have been applied to data. For instance, Bluck et al. (2016) fit the SFMS using median log SFRs of galaxies with $M_* < 10^{10} M_\odot$ and extrapolate to higher masses. This method, however, assumes that all $M_* < 10^{10} M_\odot$ galaxies lie on the SFMS and that there is no variation in the slope of the SFMS at higher stellar masses. Alternatively, Lee et al. (2015) fit the SFMS using median log SFRs of galaxies in the sample after some color-color

cut to identify SF galaxies. Other recent works in the literature have opted for more sophisticated methods such as fitting a three-component Gaussian (Bisigello et al. 2018) or a zero-inflated negative binomial distribution (Feldmann 2017).

All of these methods require arbitrary assumptions or hard cuts on the sample. Moreover, these methods struggle with fitting the SFMS over a wide SFR or M_* range and for the wide variety of SSFR distributions we see in simulations and observations. Even for fixed a stellar mass bin ($10.4 < \log M_* < 10.6$), Figure 2 reveals the significant discrepancies in the SSFR distributions of our four simulations. Therefore, in an effort to better fit a wide variety of star formation–stellar mass distributions and to relax the assumptions and cuts imposed on the data, *we present a flexible and data-driven method for fitting the SFMS that makes use of Gaussian Mixture Models* (Press et al. 1992; McLachlan & Peel 2000).

3.1. Using Gaussian Mixture Models

Gaussian mixture models (hereafter GMM), and mixture models in general, provide a probabilistic way of describing the distribution of a population by identifying subpopulations from the data.

Besides their extensive use in machine learning and statistics, GMMs has also been used in wide range of astronomical analyses (Bovy et al. 2011; Lee et al. 2012; Taylor et al. 2015). Since identifying the subpopulation of star forming galaxies from the overall galaxy population is equivalent to fitting the SFMS, GMMs provides a well-motivated, data-driven, and effective method to tackle the problem.

A GMM, more precisely, is a weighted sum of k Gaussian component densities

$$\hat{p}(x; \theta) = \sum_{i=1}^k \pi_i \mathcal{N}(x; \theta_i), \quad (1)$$

to estimate the density. The weights, π_i , mean, and variance $\theta_i = \{\mu_i, \sigma_i\}$ of the components are free parameters in the GMM. For a given data set $\{x_1, \dots, x_n\}$, these parameters are most commonly estimated through the expectation-maximization algorithm (Dempster et al. 1977; Neal & Hinton 1998, EM;).

Starting with randomly assigned θ_i^0 to the k GMM components, the EM algorithm iterates between two steps. First, for every data point, x_i , the algorithm computes for a probability of x_i being generated by each GMM component. These probabilities act as assignment weights to each of the components. Next, based on these weights, θ_i' of the components are updated to θ_i^{t+1} to maximize the likelihood of the assigned data. π_i are also updated by summing up the assignment weights and normalizing the sum by the total number of data points. These steps are repeated until convergence — *i.e.* when $p(\{x_1, \dots, x_n\}; \theta_t)$ converges. Instead of starting with randomly assigning θ_i^0 , we initiate our EM algorithm using a k -means clustering algorithm (Lloyd 1982). More specifically, for our Gaussian mixture density estimation we use the k -means++ algorithm of Arthur & Vassilvitskii (2007).

For our actual SFMS fitting method, we first divide the galaxy sample into stellar mass bins of width $\Delta \log M$. In this paper we use bins of

$\Delta \log M = 0.2$ dex; however, this choice does not significantly impact the final fit. For each stellar mass bin, if there are more than $N_{\text{thresh}} = 100$ galaxies in the bin, we fit the SSFR distribution using Gaussian mixture models (GMMs) with 1 to 3 components with parameters determined from the EM algorithm described above. Our restriction to models with a maximum of 3 components is motivated by the three main galaxy classifications: quiescent, star-forming, and transitioning populations. Furthermore, for our observed galaxy samples, even when we allow for more than 3 components, the best-fit GMMs have $k \leq 3$. We confirm that restricting ourselves to 3 components does not significantly impact the results of this work (Appendix A).

Out of the three ($k \leq 3$) GMMs, we select the “best-fit” model with the lowest Bayesian Information Criteria (BIC; Schwarz 1978). BIC is often used in conjunction with GMMs (*e.g.* Leroux 1992; Roeder & Wasserman 1997; Fraley & Raftery 1998; Steele & Raftery 2010) and also more generally for model selection in astronomy (*e.g.* Liddle 2007; Broderick et al. 2011; Vakili & Hahn 2016). In addition to the likelihood, BIC introduces a penalty term for the number of parameters in the model. This way, using BIC not only finds a good fit to the data, but it also addresses the concern of over-fitting.

In the best-fit GMM, we designate the Gaussian component of the best-fit GMM with mean $\log \text{SSFR} > 11$ as the SFMS component. In the case when more than one GMM component has mean $\log \text{SSFR} > 11$, the component with the larger weight (*i.e.* the mode) is designated as the SFMS component. Typically in such a situation, the SFRs of the SFMS is not well described by a log-normal distribution. In observations, we do not encounter this situation. In simulations, however, we encounter this situation in lower stellar mass bins ($\log M_* < 10^9 M_\odot$) of the Illustris and EAGLE simulations.

In Figure 3, we illustrate our GMM SFMS fitting for the central galaxies of the Illustris simulation in two stellar mass ranges highlighted in the left panel: $10.4 < \log M_* < 10.6$ (center) and $11.0 < \log M_* < 11.2$ (right). For the two stellar mass bins, we compare the SSFR distributions of the bins to the components of the best-fit GMMs derived from our SFMS fitting method. The SFMS component is plotted in blue. The SSFR distribution of the center panel is best described by a GMM with three components while the SSFR distribution in the right panel is best described by a GMM with only two components. These comparisons highlight the flexibility and effectiveness of our fitting method in identifying the SFMS for different SSFR distributions. All the code used for our SFMS fitting is publicly available at <https://github.com/changhoonhahn/LetsTalkAboutQuench>.

4. RESULTS

4.1. SFMS of simulated galaxies

Now with our method for fitting the SFMS, which can be flexibly applied a wide range of SSFR distributions, we identify the SFMSs of our simulated central galaxies from Section 2. For the instantaneous and 100 Myr SFR timescales, we present the best-fit SFMSs of our simulated galaxies from the Illustris, EAGLE, MUFASA, and Santa Cruz SAM simulations in Figures 4 and 5, respectively.

The best-fit SFMSs of the Illustris, EAGLE, and Santa Cruz SAM simulations exhibit a similar monotonic relation between SFR and M_* out to $M_* \gtrsim 10^{11} M_\odot$, for both SFR timescales. The best-fit SFMSs for MUFASA, however, turn over at $M_* \sim 10^{10.5} M_\odot$ and decreases in SFR at high stellar masses. We confirm that this turnover is *not* caused by misidentification of the SFMS or some systematic effect of the fitting by examining the best-fit GMM components and the $p(\log \text{SSFR})$ in these higher stellar mass bins. Between the two SFR timescales, we find little difference between the best-fit SFMSs for each simulation. However,

when we compare the best-fit SFMSs of the simulations altogether in more detail, we find *order of magnitude discrepancies in SFR among the best-fits throughout the stellar mass range of the simulations for both SFR timescales* (bottom right panels of Figures 4 and 5).

The uncertainties for the best-fit SFMSs in Figures 4 and 5 are derived from bootstrap resampling (Efron 1979) in each stellar mass bin of the fitting. These uncertainties are *underestimates* because they do not account for “cosmic variance” — *i.e.* we only have one finite volume realization of each simulation. Furthermore, uncertainty of the best-fit SFMS corresponds to the uncertainty of the means of the SFMS GMM component, which is only one of the parameters in the GMM. Uncertainties from bootstrap resampling, however, do *not* account for the correlations between the mean of the SFMS GMM and other parameters of the GMM (Eq. 1). A more robust estimate of the uncertainties would involve estimating the marginalized posterior distribution of the SFMS GMM component mean using a method like MCMC. Since this still does not account for cosmic variance, we use bootstrap uncertainties.

From the best-fit SFMSs, we can further parameterize the SFMS to some functional form as often done in the literature — *e.g.* power-law (Speagle et al. 2014) or broken power-law (Lee et al. 2015). With little evidence of a turnover in the SFMS for most of our galaxies, we fit a power-law of the form

$$\log \text{SFR}_{\text{MS}} = m (\log M_* - 10.5) + b \quad (2)$$

to the SFMSs in Figure 6. We list the best-fit (maximum likelihood) parameter values of Eq. 2 in Table 1. The power-law parameterizations accentuate the stellar mass dependence of the SFMSs and can reveal the mass dependence in the SFMS discrepancies. Putting aside MUFASA, whose power-law fit is impacted by the turnover at high stellar masses, we find significant discrepancies in the slopes of the SFMSs of the simulations with Santa

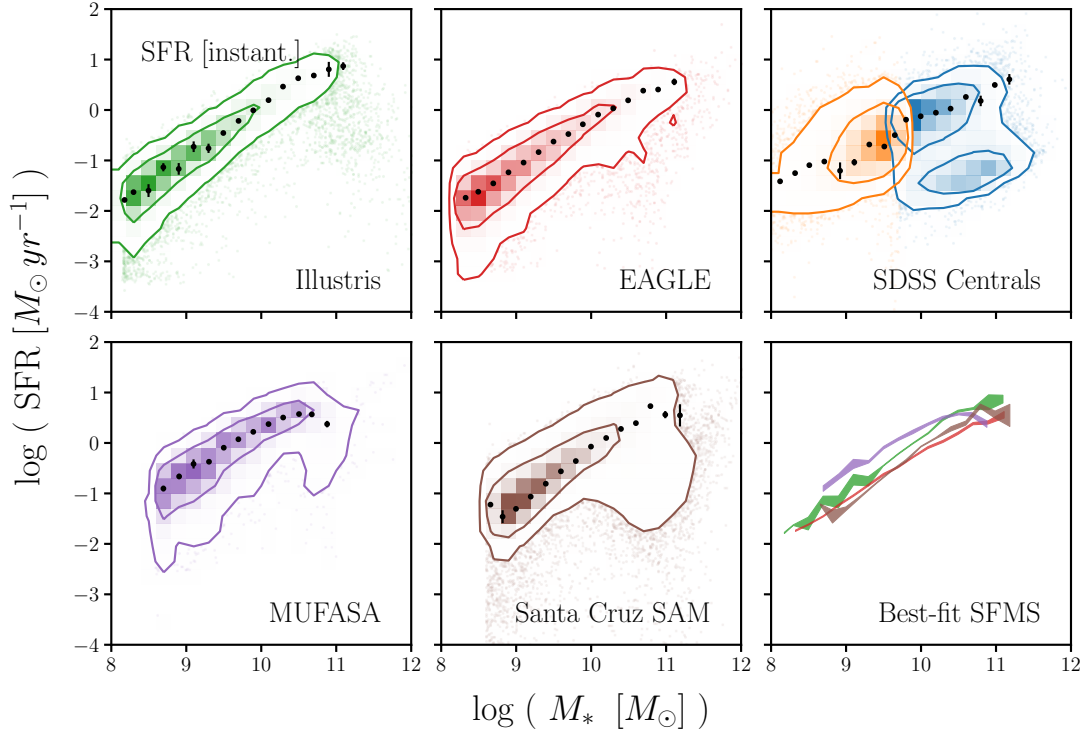


Figure 4. Best-fit SFMS of the central galaxies in the Illustris, EAGLE, MUFASA, and Santa Cruz SAM simulations as identified by our SFMS GMM fitting method (Section 4.1). The SFMSs above are fit from the instantaneous SFR– M_* relation. For reference, we include the best-fit SFMS of the SDSS sample in the top right panel. The uncertainties of the best-fit SFMS are derived using bootstrap resampling. When we compare the SFMS fits (bottom right panel), *the SFMSs of the simulations have similar overall stellar mass dependence, but their amplitude vary roughly by an order of magnitude.*

Cruz SAM having the steepest SFMS and EAGLE having the shallowest SFMS. The power-law SFMS fits, however, reveal little mass dependence in the discrepancies of the SFMSs. For instantaneous SFR, the discrepancies between the SFMSs of Illustris, EAGLE, and Santa Cruz SAM are ~ 0.5 dex throughout the stellar mass range. Meanwhile, for 100 Myr SFR, we find slightly greater discrepancies among Illustris, EAGLE, and Santa Cruz SAM throughout the stellar mass range.

Things to discuss at the workshop:

- **What’s causing the simulations to have SFMSs that are almost an order of magnitude different?**

Despite the SFMSs of our simulations vary by an order of magnitude we find relatively consistent cosmic star formation densities among the simula-

tions. For Illustris, EAGLE, MUFASA, and Santa Cruz SAM, respectively, we estimate cosmic star formation densities of -1.66 , -2.22 , -1.87 , and $-1.94 M_\odot \text{yr}^{-1} \text{Mpc}^{-3}$ using instantaneous SFRs and similarly -1.68 , -2.20 , -1.91 , and $-1.94 M_\odot \text{yr}^{-1} \text{Mpc}^{-3}$ using 100 Myr SFRs.

- **Why are the cosmic star formation density estimates consistent despite the discrepancy in the SFMS of the simulations.**
- **Why does MUFASA have a turnover in the SFMS? – Due to MUFASA’s quenching method being based on halo mass? YES, I suspect this causes a somewhat sharper cutoff than in the other more self-consistent AGN feedback models. – Romeel**

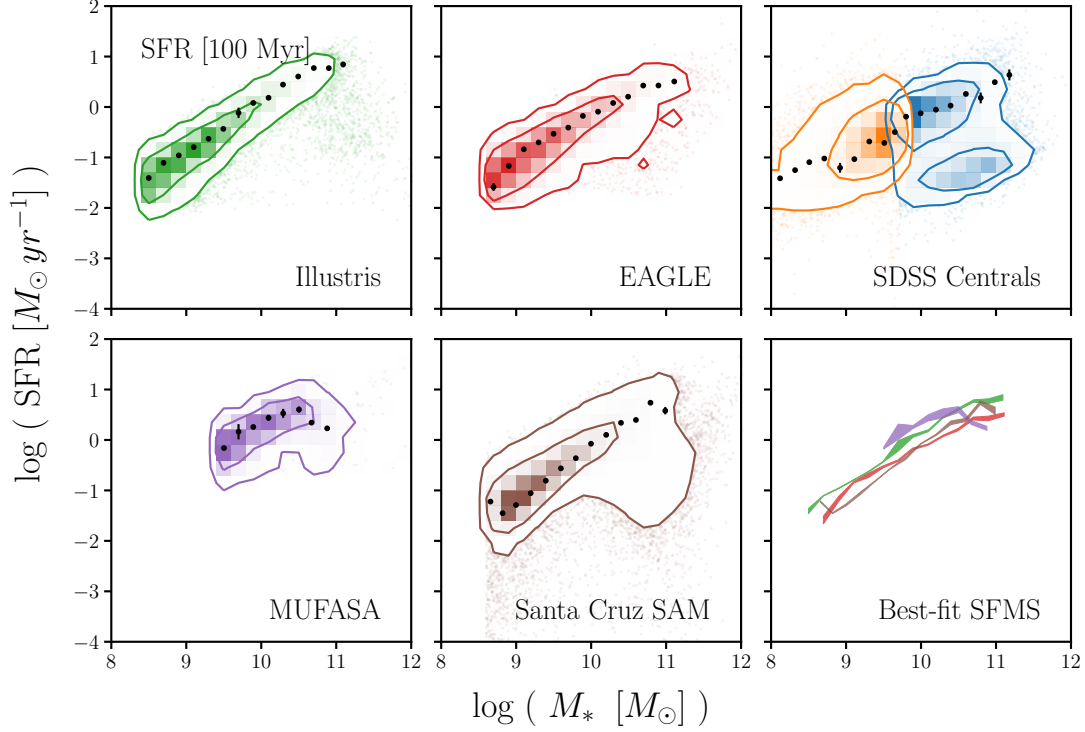


Figure 5. Same as Figure 4 but for the 100 Myr SFR– M_* relation. As in Figure 4, the SFMSs of the simulations have similar stellar mass dependence but vary roughly by an order of magnitude in amplitude.

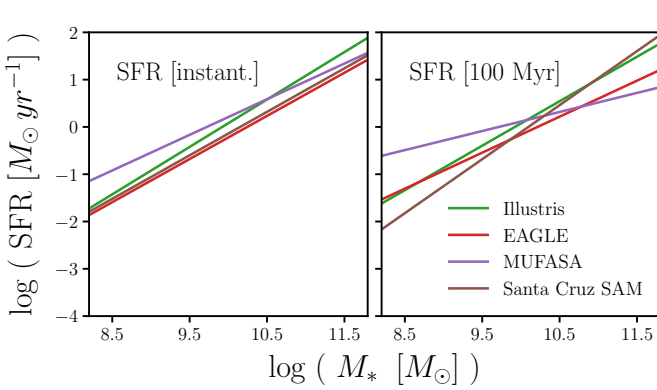


Figure 6. Power-law fits to the best-fit SFMS for the Illustris (green), EAGLE (red), MUFASA (purple), and Santa Cruz SAM (brown) simulations. The left and right panels use instantaneous SFR and 100 Myr SFR, respectively. We list the power-law fit parameters in Table 1. The power-law fit to the MUFASA SFMS is impacted by the high stellar mass turnover. The power-law fits for the Illustris, EAGLE, and Santa Cruz SAM SFMS differ by ~ 0.5 dex with no significant mass dependence.

mean
In addition to its position, μ_{SFMS} , the SFMS GMM component is also described by σ_{SFMS} — the width of the SFMS. Using σ_{SFMS} derived from the SFMS GMM fitting, we can compare the width of the SFMS among the simulations (Figure 7). The uncertainties for the widths are calculated through bootstrap resampling in the same way as the uncertainties for the SFMS fits. Overall, we find little stellar mass dependence in σ_{SFMS} for the simulations. For Illustris, EAGLE, MUFASA, and Santa Cruz SAM we respectively find $\sigma_{\text{SFMS}} \sim 0.16, 0.20, 0.21$, and 0.20 dex for instantaneous SFR and $\sigma_{\text{SFMS}} \sim 0.17, 0.24, 0.23$, and 0.24 dex for 100 Myr SFR. These σ_{SFMS} are narrower than the ~ 0.3 dex width measured in observations (*e.g.* Daddi et al. 2007; Noeske et al. 2007; Magdis et al. 2012; Whitaker et al. 2012). This difference, however, does not account for the uncertainties in the observed SFR measurements, which would reduce the discrepancy. We therefore conclude that *the width of the SFMS from our sim-*

Table 1. Power-law fit to the SFMS of our simulated central galaxies from the Illustris, EAGLE, MUFASA, and Santa Cruz SAM simulations.

Star Forming Main Sequence fit		
$\log \text{SFR}_{\text{MS}} = m (\log M_* - 10.5) + b$		
Simulation	m	b
Illustris [inst. SFR]	1.00	0.58
Illustris [100 Myr SFR]	0.94	0.54
EAGLE [inst. SFR]	0.91	0.23
EAGLE [100 Myr SFR]	0.76	0.22
MUFASA [inst. SFR]	0.75	0.59
MUFASA [100 Myr SFR]	0.40	0.32
Santa Cruz SAM [inst. SFR]	0.92	0.32
Santa Cruz SAM [100 Myr SFR]	1.14	0.46

ulations are in agreement with the observed SFMS width.

One factor that impacts our SFMS fits is the strict lower limit of the log SFRs caused by the resolution effects in the simulations. This is particularly evident in the 100 Myr SFR– M_* relations of the hydrodynamic simulations of Figure 1—especially MUFASA. As we describe in Section 2.1, the 100 Myr SFRs are calculated using the ages of all star particles in a galaxy. For a galaxy to have star formation (*i.e.* SFR > 0), it must *at least* form one star particle over the last 100 Myr. A single star particle forming over 100 Myr amounts to a SFR of $\sim 0.02 M_\odot \text{ yr}^{-1}$ for Illustris and EAGLE and $\sim 0.2 M_\odot \text{ yr}^{-1}$ for MUFASA. This resolution limit, ultimately impacts the SFMS fits at stellar masses below $10^{8.4}$, $10^{8.6}$, and $10^{9.4} M_\odot$ for Illustris, EAGLE, and MUFASA respectively. These stellar mass limits have accordingly been imposed on the SFMS fits and power-law SFMS fits in Figures 5 and 6. In Appendix B, we describe how we derive these stellar mass limits.

By using the SFMS fitting method we present in this paper, we’re able to conduct a principled data-driven comparison of the SFMSs of central galaxies from the Illustris, EAGLE, MUFASA, and

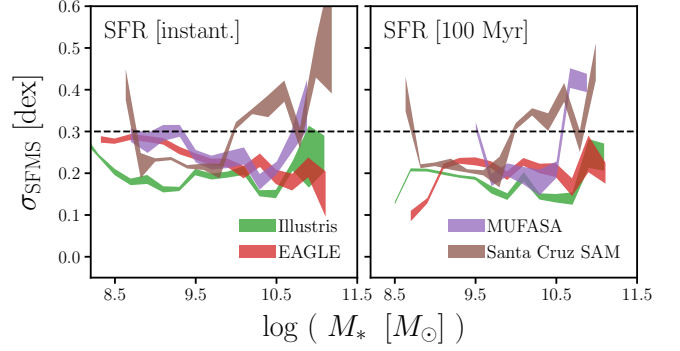


Figure 7. The width of the SFMS, σ_{SFMS} , for our simulated central galaxies from Illustris, EAGLE, MUFASA, and Santa Cruz SAM (green, red, purple, and brown respectively). The widths are derived from the GMM SFMS fitting and its uncertainties are estimated using bootstrap resampling in the same way as the SFMS fit uncertainties. σ_{SFMS} s in our simulations have little stellar mass dependence and, naively accounting for measurement errors in observed SFR, they are consistent with $\sigma_{\text{SFMS}} \sim 0.3$ dex from observations (black dashed).

Santa Cruz SAM simulations. From this comparison, we find that the amplitudes of the SFMSs differ from one another by an order of magnitude, with no significant stellar mass dependence. Furthermore, despite the discrepancies in their amplitude, the SFMSs of our simulations have similar widths, consistent with observations.

4.2. Beyond the SFMS of simulated galaxies

So far, we have focused solely on the SFMSs of our simulated galaxies—*i.e.* $\theta_{\text{SFMS}} = \{\mu_{\text{SFMS}} \text{ and } \sigma_{\text{SFMS}}\}$. Our GMM fitting method, however, also determines θ_i for components other than the SFMS. These GMM components provide extra features to compare our simulated galaxy samples and also offer interesting insights into the different populations in our simulated galaxy samples. When we examine the mean and variance θ_i of all the components from our fitting for our simulated galaxies we find correspondence between the components and the quiescent, transitioning, and star-burst galaxy populations (Figure 8). Sim-

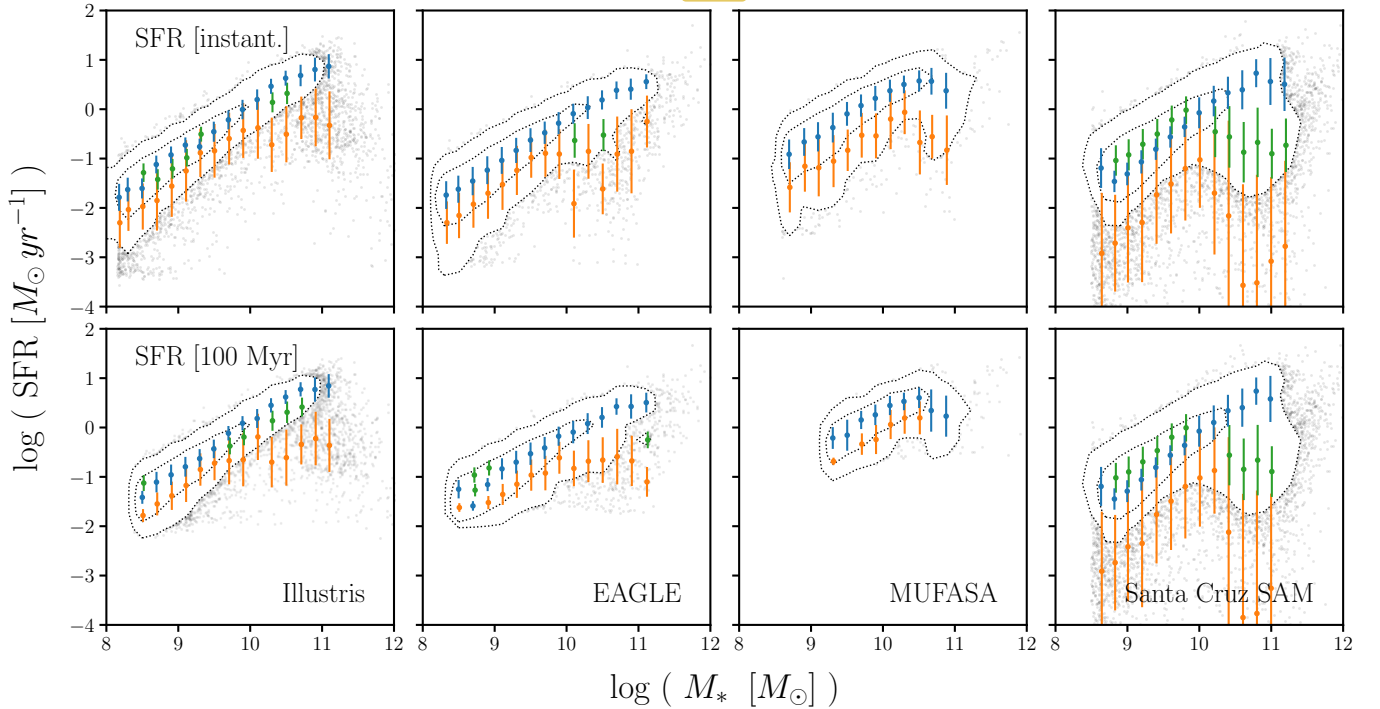


Figure 8. Components of the best-fit GMM from our SFMS fitting method for the SFR- M_* relations of central galaxies in the Illustris, EAGLE, MUFASA, and Santa Cruz SAM simulations (left to right). The top panels use instantaneous SFRs while the bottom panels use SFRs averaged over 100 Myr. We mark the SFMS components in blue, the components with lowest SFR in orange, and the other components in green. These components roughly correspond to the star-forming, quiescent, and transitioning or star-burst populations. Among the hydrodynamical simulations, despite the SFMS discrepancies, (Figures 4 and 5), the GMM components reveal overall similarities in the galaxy populations. Meanwhile for the Santa Cruz SAM, the GMM components reveal its broad quenched component that extends beyond $\text{SFR} < 10^{-4} M_\odot \text{yr}^{-1}$ and prominent transitioning components at $M_* > 10^{10} M_\odot$, different from the hydrodynamic simulations.

ilar to how we designated the SFMS component of the GMM, we designate the component with the lowest SFR as the quiescent component, the component with SFR in between the SFMS and the quiescent components as transitioning, and the component with SFR higher than the SFMS component as the star-burst component. In Figure 8, we mark the SFMS components in blue, quiescent components in orange, and the other components in green.

All of the simulations have significant populations below the SFMS at all masses. The hydrodynamical simulations have relatively tight quiescent components ~ 1 dex below the SFMS. The quiescent components of the Santa Cruz SAM,

meanwhile, are much broader and extend below $\log \text{SFR} = -4$. Regardless of their position and width, at low stellar masses, the significant quiescent populations are in disagreement with observations (Geha et al. 2012), which find no isolated/central quenched galaxies below $M_* \sim 10^9 M_\odot$.

- **What is causing central galaxies to have low SFRs at $10^9 M_\odot$?**

This discrepancy between simulations and observations, however, must be taken with a grain of salt. Both the environment and quenched classifications in Geha et al. (2012) are defined differently than in the simulations. Geha et al. (2012) classifies as a galaxy as quenched if it does not have an $H\alpha$ emission and based on a D_n4000 criteria.

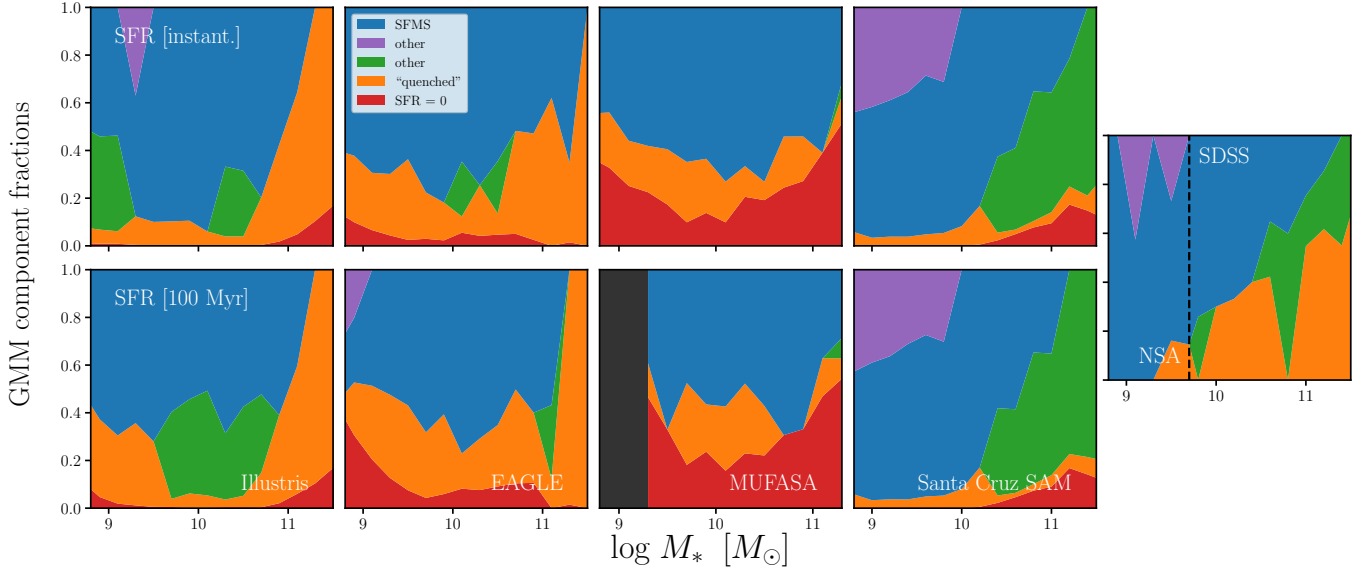


Figure 9. Fractional contributions, π_i , of the best-fit GMM components from our SFMS fitting of the central galaxies in the Illustris, EAGLE, MUFASA, and Santa Cruz SAM simulations (left to right). We highlight the SFMS component in blue, the quenched component in orange, galaxies with $\text{SFR} = 0$ in red, the transitioning components in green, and the star-burst components in purple. For reference, we include π_i of the observed centrals from SDSS and NSA in the rightmost panel. In all the simulations, a significant fraction of central galaxies lie below the SFMS at $M_* \lesssim 10^9 M_\odot$ in disagreement with observations. For the hydrodynamic simulations, the SFMS fraction has little stellar mass dependence at $M_* < 10^{11} M_\odot$ unlike observations. Meanwhile, a number of similar components are found between Santa Cruz SAM and observations: star-burst components at low M_* and transitioning components at high M_* .

Furthermore, the isolation criteria of [Geha et al. \(2012\)](#) is more stringent than the group finder at these mass scales.

In Illustris and the Santa Cruz SAM, we find a number of components between the SFMS and quiescent components —*i.e.* transitioning galaxies. For Illustris, transitioning components are identified at $10^9 M_\odot < M_* < 10^{11} M_\odot$. For the Santa Cruz SAM, transitioning components are identified at $M_* > 10^{10} M_\odot$. We also identify components with SFRs above the SFMS. These *star-burst* components are particularly evident in the Santa Cruz SAM, where a significant star-burst population is observed at $10^9 M_\odot < M_* < 10^{10} M_\odot$. Besides the Santa Cruz SAM, we also find star-burst components at the lowest stellar masses ($M_* < 10^9 M_\odot$) of the Illustris and EAGLE simulations for 100 Myr SFRs.

The different components we identify using the GMM fitting allows us to make a number of interesting comparisons of the simulated galaxies. Overall the galaxies from hydrodynamical simulations have similar components and features in the $\text{SFR}-M_*$ data space. Besides the difference in the position and width of the SFMS, the only significant discrepancy is the transitioning components found in Illustris. Between the hydrodynamical simulations and the Santa Cruz SAM, however, more significant differences are revealed by the non-SFMS components.

- **Why is the SAM so different overall?**
- **Why does Santa Cruz SAM have a broader range of SFRs?**
- **Why does Santa Cruz SAM have star-bursts in "intermediate" stellar masses**

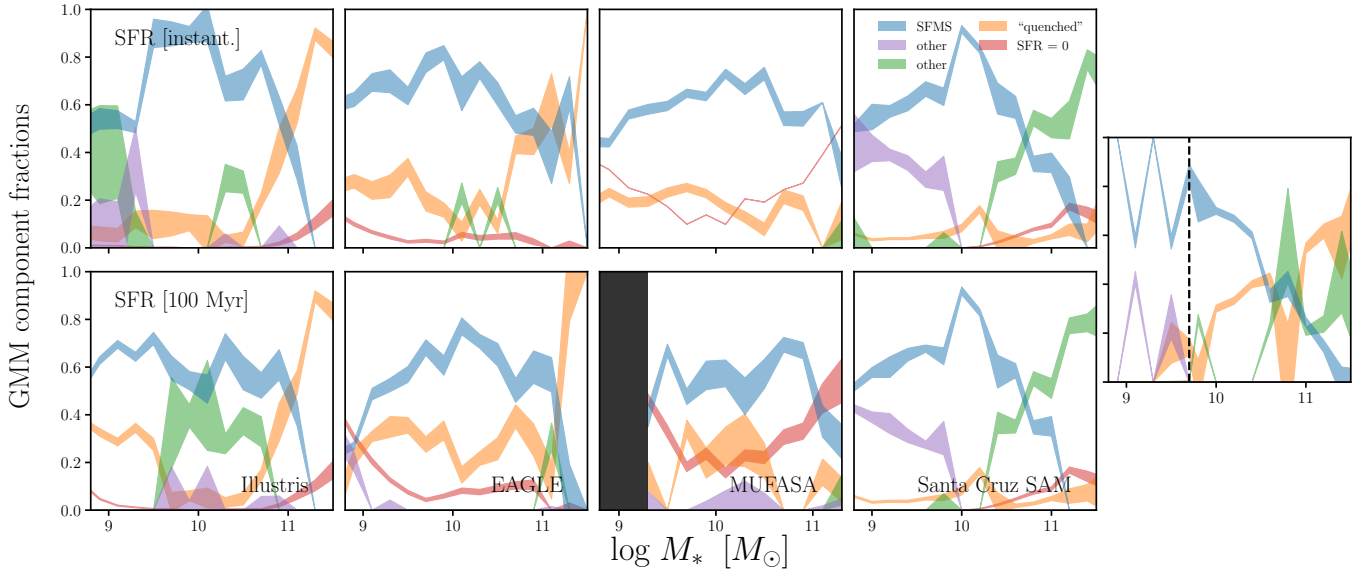


Figure 10. Uncertainties in the fractional contributions, π_i , of the best-fit GMM components from our SFMS fitting from Figure 9. The uncertainties are estimated through bootstrap resampling. We plot the SFMS component in blue, the quenched component in orange, galaxies with SFR= 0 in red, the transitioning components in green, and the star-burst components in purple. For reference, we include π_i of the observed centrals from SDSS and NSA in the rightmost panel.

Another set of parameters we infer from our GMM fitting is the weight of the GMM components— π_i in Eq. 1. Since we designated each component to a corresponding galaxy population, these weights correspond to the fractional contribution of the different populations. For example, the weight of the quenched component is an estimate of the quiescent fraction (*e.g.* Blanton & Moustakas 2009; Geha et al. 2012; Hahn et al. 2015). In Figure 9 we present the fractional contribution, as a function of stellar mass, for all the components from our GMM fitting : SFMS (blue), quenched (orange), SFR = 0 galaxies (red), transitioning (green), and star-burst (purple). We present the uncertainties of π_i , estimated through bootstrap resampling, in Figure 10. For every simulation, a significant fraction of galaxies have SFR= 0. In hydrodynamical simulations, a galaxy with SFR=0 can have an SFR below the resolution limit, or have a *true* SFR=0 on the measured timescales (Appendix B). In the Santa Cruz SAM, all the SFR=0 galaxies have truly no star formation on the timescales that we present here. Therefore, in both hydrodynamical

and SAM simulations, the SFR= 0 galaxies can be considered quiescent. Moreover, we note that SFR resolution does not significantly impact the fraction contributions of Figures 9 and 10 (see Appendix B and Figure 15).

The fractional contributions of the GMM components in Figures 9 and 10 reveal disagreements between our simulated galaxies and trends established from observations—especially the hydrodynamical simulations. As we mentioned earlier, observations have established a stellar mass lower bound for isolated/central quenched galaxies (Geha et al. 2012). From Figure 9, we find significant quiescent populations at $M_* < 10^9 M_\odot$, confirming Figure 8. Treating SFR= 0 galaxies as quiescent, the quiescent fraction for the EAGLE simulation is roughly 0.4 at $10^9 M_\odot$. Illustris with 100 Myr SFRs and MUFASA with instantaneous SFRs have similarly high quiescent fractions. Even in the Santa Cruz SAM, we find a non-negligible ($\sim 10\%$) quiescent fraction at $10^9 M_\odot$.

Furthermore, at $M_* > 10^9 M_\odot$, we find little stellar mass dependence in the fraction of high SFR

components (SFMS and star-burst; blue and purple in Figure 9) in our hydrodynamical simulations at $M_* < 10^{11} M_\odot$. If we take the quiescent fraction to be $f_Q = 1 - \sum_{\text{high SFR}} \pi_i$, then the quiescent fraction of our central galaxies from hydrodynamical simulations have little stellar mass dependence over $10^9 < M_* < 10^{11} M_\odot$ — unlike f_Q measurements from Baldry et al. (2006); Peng et al. (2010); Hahn et al. (2015), which all find significant stellar mass dependence in the quiescent/red fraction of isolated galaxies in SDSS. In this regards, the quiescent fraction of the Santa Cruz SAM is in good agreement with the isolated galaxy quiescent fraction from SDSS (Baldry et al. 2006; Peng et al. 2010; Hahn et al. 2015).

As we have discussed in this section, our SFMS fitting provides additional features, more than just the SFMSs, to compare different galaxy samples. Moreover, the components identified by our fitting method offers insight into the distinct galaxy populations of our simulations. Based on the non-SFMS components/populations, we find consistency among the hydrodynamical simulations in SFR– M_* space, but not with the Santa Cruz SAM. When we compare to trends in the literature, we find that all of the simulations have a significant fraction of low SFR (quenched) central galaxies at $M_* \lesssim 10^9 M_\odot$ —in disagreement with observations. In particular, the hydrodynamical simulations, at intermediate stellar masses $M_* \lesssim 10^{11} M_\odot$, do not reproduce the quiescent fractions from observations.

4.3. Comparing to Observations

Our data-driven GMM SFMS fitting provides a principled way to compare galaxy populations through the major features in the data-space. These comparisons have so far revealed a number of agreements and discrepancies among the central galaxies in our simulations. The ultimate goal of these simulations, however, is to reproduce the observations, which means we ultimately want to compare them to observations. In this section, we extend our GMM SFMS fitting to the observed

galaxy sample from SDSS and NSA and compare our simulated galaxies to observation.

We first focus on the best-fit SFMS of our SDSS and NSA central galaxies in the top right panels of Figure 4 and 5. Similar to the Illustris, EAGLE, and Santa Cruz SAM simulations, the SFR of the SFMS monotonically increases with M_* . The SFMS has no turnover like MUFASA’s. Overall, the SDSS and NSA SFMS is shallower and has a lower amplitude than the SFMSs of our simulations.

- **Why are there significant discrepancies in the SFMS slopes of the simulation versus observations? What are the main sub-grid processes that dictate the slope of the SFMS?**

Looking beyond the SFMS, the other components of the SDSS and NSA central galaxy population further highlight the disagreements between the simulations and observation from the previous section (Figure 9). At $M_* \lesssim 10^9 M_\odot$, we find no quenched population, in agreement with Geha et al. (2012) and in disagreement with the simulations. Furthermore, in agreement with measurements in the literature, we find significant M_* dependence in the quiescent fraction, which the hydrodynamic simulations do not find. Aside from the low stellar mass quenched population, the components of the Santa Cruz SAM simulation are in good agreement with the SDSS and NSA components. They both identify star-burst populations at low masses and significant transitioning populations from $M_* > 10^{10} M_\odot$.

The comparison we make between our simulations and observations so far in this section only scratches the surface. However, we deliberately refrain from a more detailed comparison due to the limitations of such a comparison. The main bottleneck stems from the fact that the SFRs and M_* , the main galaxy properties considered in this paper, are defined and measured differently in simulations versus observations. SFRs from the SDSS and NSA catalogs are measured using a combina-

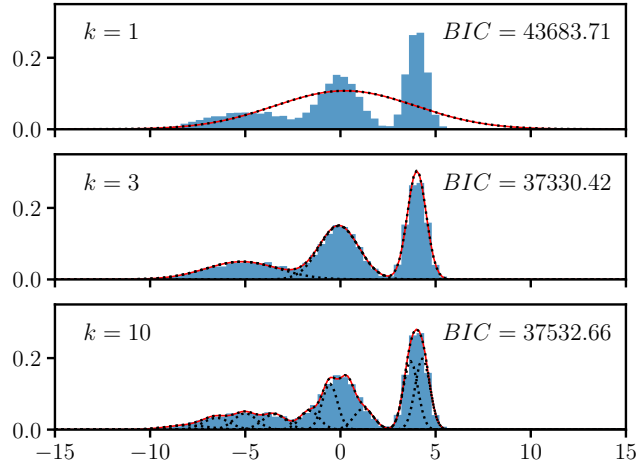


Figure 11. A pedagogical illustration of Gaussian mixture density estimation. We use GMMs with $k = 1$ (top), 3 (middle), 10 (bottom) components to estimate the distribution of data (blue) drawn from three Gaussian distributions. The GMM outputs of the EM algorithm are plotted in red with dotted lines representing each of their components. We also include the BIC of the GMMs, which we use to select the number of components k . Of the three panels, $k = 3$ has the lowest BIC and therefore best represents the data according to our selection scheme.

tion of $H\alpha$ and D_n4000 . Although we choose the SFR timescales to best reflect this observed SFR, as [Speagle et al. \(2014\)](#) find even for the same SDSS galaxies, different SFR indicators can produce large discrepancies in the slope and amplitude of the SFMS. Therefore, for further scrutiny we require a more apples to apples comparison. In the next paper we will bring this comparison on a more equal footing by deriving star formation rates from mock observations of the simulated galaxies (Starkenburg et al. in prep.).

5. SUMMARY AND CONCLUSIONS

conclusion!

ACKNOWLEDGEMENTS

It's a pleasure to thank Shirley Ho, Emanuel Schaan, **insert others** for valuable discussions. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under contract No. DE-AC02-05CH11231. The Isolated and Quenched galaxies (IQ) collaboratory thanks the Flatiron Institute for hosting the collaboratory and its meetings. The Flatiron Institute is supported by the Simons Foundation.

APPENDIX

A. GAUSSIAN MIXTURE MODELS: 1D AND BEYOND

So far we have assumed that one, two, or three populations would describe galaxies in the $\text{SFR}-M_*$ plane well. We have checked whether the possibility of adding more than three Gaussians in the GMM shows that more populations are present. This is the case only for the Santa Cruz semi-analytic model, where in a number of bins the presence of 4, 5, or (in one case) 6 Gaussians is preferred. In almost all cases the added Gaussians have low weight and form effectively additional intermediate populations. For both Illustris and EAGLE there is only one mass bin where a fourth population would be preferred. For all other mass bins in all the hydrodynamical simulations, and all mass bins in the SDSS 3 or less Gaussians are preferred.

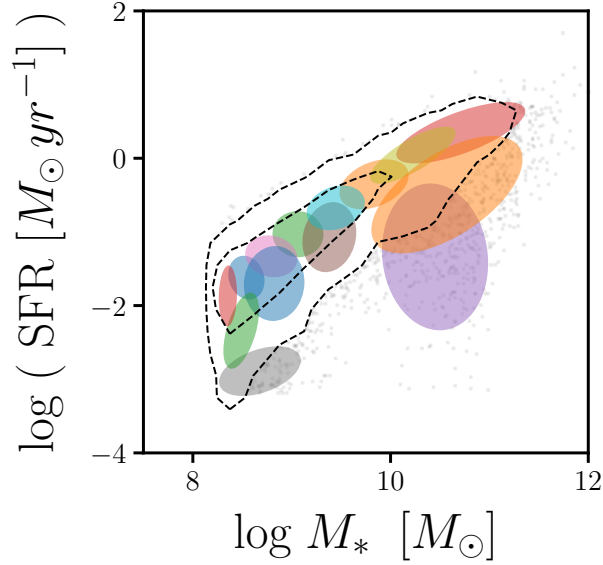


Figure 12. Two-dimensional GMM fit to the SFR- M_* relation of central galaxies of the EAGLE simulation. The two-dimensional GMM fitting is an extension of the SFMS fitting method we describe in Section 4.1. The colorful shaded ellipses over-plotted on the SFR- M_* relation (black) are the Gaussian components of the best-fit GMM. Although, identifying the SFMS from these Gaussian components is difficult, the 2D GMM is effective in capturing the features of the SFR- M_* relation and provides a good way comparing SFR- M_* relations from different data.

In addition to identifying the SFMS, the GMM fitting method described above can be extended to describe the entire SFR- M_* relation using a two-dimensional GMM. In Figure 12, we compare the instantaneous SFR to M_* relation of central galaxies in the EAGLE simulation with the best-fit two-dimensional GMM. The over-plotted shaded ellipses represent the two-dimensional Gaussian components of the best-fit GMM. Overall the best-fit 2D GMM captures the features in the EAGLE SFR- M_* relation. It also provides a straightforward way of comparing different SFR- M_* relations. However, as Figure 12 illustrates, specifically identifying the SFMS using the two-dimensional model is more challenging. Therefore in this paper, we do not discuss the 2D GMM further.

B. SFR RESOLUTION EFFECTS IN HYDRODYNAMICAL SIMULATIONS

In our analysis, we consistently derive SFRs for all of our simulated galaxies on two timescales: instantaneous and averaged over 100 Myr (Section 2.1). For our hydrodynamic simulations, SFR averaged over 100 Myr is derived using the formation times of the star particles in the simulation. This means that both the mass and temporal resolutions of the simulations cause resolution effects in the 100 Myr SFR. More specifically, in Illustris, EAGLE, and MUFASA 100 Myr SFR has a resolution of $\Delta_{\text{SFR}} = 0.016, 0.018,$ and $0.182 M_*/\text{yr}$, corresponding to baryon particle or cell masses of $1.6 \times 10^6 M_\odot$, $1.8 \times 10^6 M_\odot$, and $1.82 \times 10^7 M_\odot$, respectively.

For galaxies with high 100 Myr SFR, the resolution Δ_{SFR} is relatively small compared to their SFRs and therefore does not have a significant impact. However for low SFR galaxies, the resolution effect is more significant. For example at the lowest SFR end, galaxies, which would have SFR ranging $0 < \text{SFR} < \Delta_{\text{SFR}}$, can only have SFR= 0. These galaxies, consequently, are not included in the $\log \text{SFR}$ – $\log M_*$ plane or the SFMS fitting. In Figure 13, we present the impact of this resolution effect on the $P(\log \text{SSFR})$ distributions

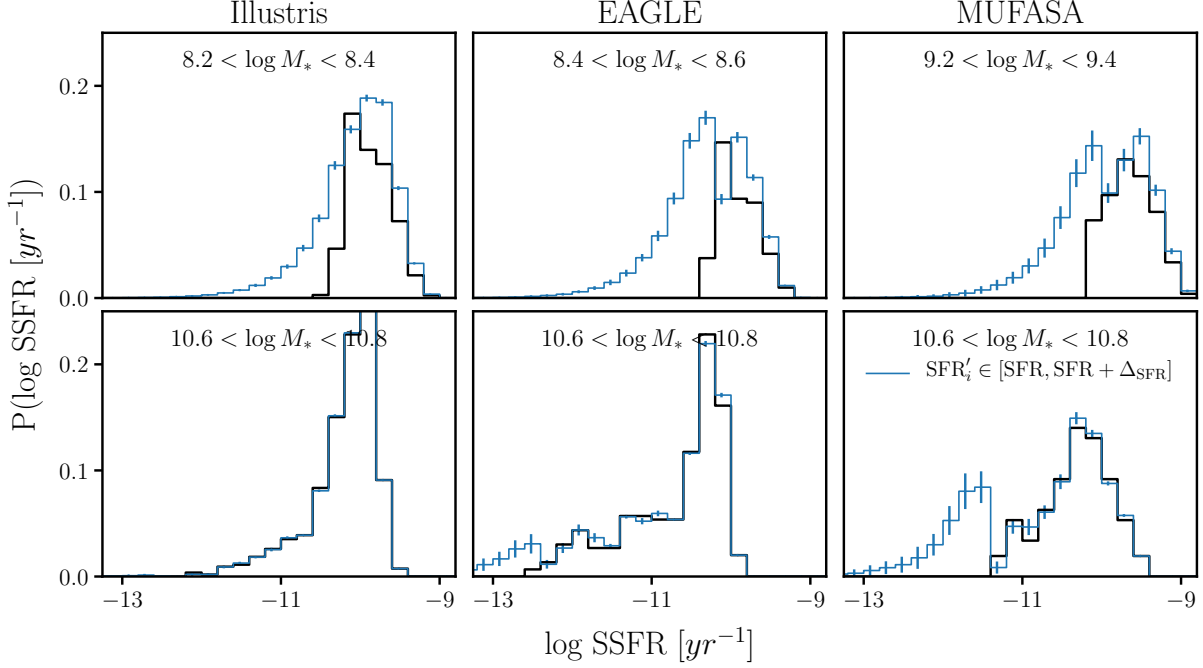


Figure 13. The impact of SFR resolution on the log SSFR distribution, $P(\text{SSFR})$, in two stellar mass bins of our hydrodynamical simulations: Illustris (left), EAGLE (center), and MUFASA (right). We plot the $P(\text{SSFR})$ distributions using the 100 Myr SFRs *with* resolution effects in black. In blue, we plot the $P(\log \text{SSFR})$ distributions of the same galaxies, but with each of their SFRs sampled uniformly within the SFR range spanned by the resolution. The uncertainties for the blue $P(\text{SSFR})$ s are estimated from re-sampling the SFR of each galaxy based on the SFR resolution. At low stellar masses (top) the SFR resolution significantly impacts the star-forming end of $P(\text{SSFR})$ s. At higher stellar masses, although the SFR resolution impacts the $P(\text{SSFR})$ s, the effect is limited to $\log \text{SSFR} < -11$.

of our hydrodynamical simulations in two stellar mass bins. In black, we plot the $P(\log \text{SSFR})$ distributions using the 100 Myr SFRs *with* resolution effects — galaxies with $\text{SFR} = 0$ are not included in this distribution. In blue, we plot the $P(\log \text{SSFR})$ distributions of the same galaxies, but with SFR'_i of each galaxy sampled uniformly within the SFR resolution range: $[\text{SFR}_i, \text{SFR}_i + \Delta_{\text{SFR}}]$. Uncertainties for the blue $P(\log \text{SSFR})$ s are derived from repeating this SFR sampling 100 times. For the low M_* bins (top), the SFR resolution affects the $P(\log \text{SSFR})$ s well above $\log \text{SSFR} = -11$, on the star-forming end of the distribution. Meanwhile, the impact at higher M_* (bottom), is limited to the low SSFR end.

In order to more explicitly quantify the impact of the SFR resolution effect on our SFMS fitting, in Figure 14 we compare the SFMS fits using 100 Myr SFRs *with* resolution effects (black) to the SFMS fits using 100 Myr SFRs sampled uniformly within the SFR range spanned by the resolution (orange; $\text{SFR}'_i \in [\text{SFR}_i, \text{SFR}_i + \Delta_{\text{SFR}}]$). The uncertainties of our SFMS fits in black are calculated using bootstrap resampling (Section 4.1). In agreement with Figure 13, we find that the SFR resolution significantly impacts SFMS fitting at low M_* . Moreover, using the comparison of Figure 14, we determine the stellar mass limit above which the SFR resolution does *not* significantly impact the SFMS fitting — *i.e.* the shift in best-fit SFMS is below 0.2 dex. For Illustris, EAGLE, and MUFASA we determine $\log M_{\text{lim}} = 8.4, 8.6$, and 9.4 , respectively.

In addition to its effect on the SFMS fits, we also examine the impact of SFR resolution on the non-SFMS components of our GMM fitting method. In Figure 15 we present the fraction contributions, π_i , of the best-fit components for the Illustris, EAGLE, and MUFASA simulations, where instead of the 100 Myr

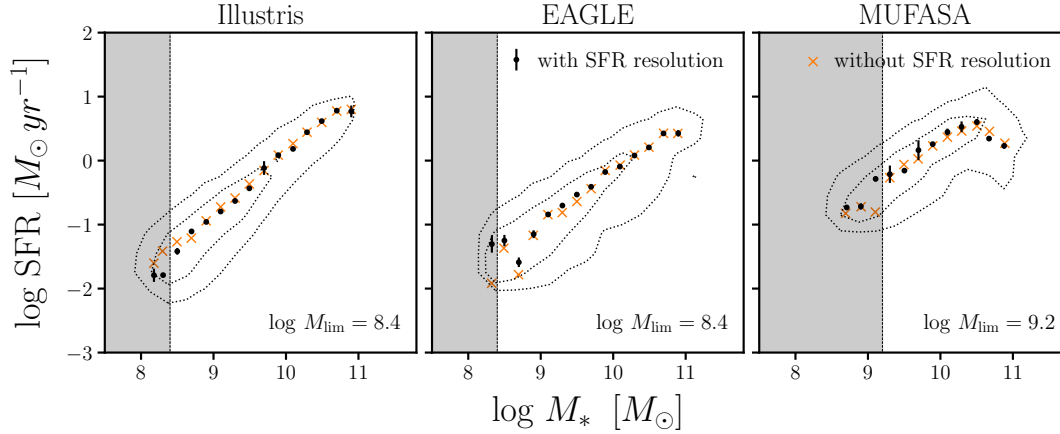


Figure 14. The resolution effect of 100 Myr SFRs in the hydrodynamic simulations (Illustris, EAGLE, and MUFASA) impact SFMS fitting at low stellar mass. In black we plot the best-fit SFMS with the resolution effects. In orange we plot the best-fit SFMS without resolution effects, where the SFR for each galaxy is sampled uniformly within the resolution range: $\text{SFR}'_i \in [\text{SFR}_i, \text{SFR}_i + \Delta_{\text{SFR}}]$. SFR resolution impacts SFMS fits at low stellar mass. Furthermore, based on the discrepancy between the SFMS fits, we determine stellar mass limits above which the SFR resolution does *not* significantly impact (< 0.2 dex) the SFMS fit. For Illustris, EAGLE, and MUFASA $\log M_{\text{lim}} = 8.4, 8.6$, and 9.4 , respectively.

SFR with resolution effects (Figures 9 and 10) we uniformly sample the SFRs of all the galaxies within the resolution range (same as Figure 13 and 14). Aside from the fact that we remove the $\text{SFR} = 0$ component with the SFR sampling, we find no significant change from π_i s of Figures 9 and 10) and, thus, the results of Section 4.2.

REFERENCES

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *The Astrophysical Journal Supplement Series*, 182, 543, doi: [10.1088/0067-0049/182/2/543](https://doi.org/10.1088/0067-0049/182/2/543)
- Arthur, D., & Vassilvitskii, S. 2007, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07* (Philadelphia, PA, USA: Society for Industrial and Applied Mathematics), 1027–1035
- Baldry, I. K., Balogh, M. L., Bower, R. G., et al. 2006, *Monthly Notices of the Royal Astronomical Society*, 373, 469, doi: [10.1111/j.1365-2966.2006.11081.x](https://doi.org/10.1111/j.1365-2966.2006.11081.x)
- Balogh, M. L., Schade, D., Morris, S. L., et al. 1998, *The Astrophysical Journal Letters*, 504, L75, doi: [10.1086/311576](https://doi.org/10.1086/311576)
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013a, *The Astrophysical Journal*, 770, 57, doi: [10.1088/0004-637X/770/1/57](https://doi.org/10.1088/0004-637X/770/1/57)
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013b, *The Astrophysical Journal*, 762, 109, doi: [10.1088/0004-637X/762/2/109](https://doi.org/10.1088/0004-637X/762/2/109)
- Bigiel, F., Leroy, A., Walter, F., et al. 2008, *The Astronomical Journal*, 136, 2846, doi: [10.1088/0004-6256/136/6/2846](https://doi.org/10.1088/0004-6256/136/6/2846)
- Bisigello, L., Caputi, K. I., Grogin, N., & Koekemoer, A. 2018, *Astronomy and Astrophysics*, 609, A82, doi: [10.1051/0004-6361/201731399](https://doi.org/10.1051/0004-6361/201731399)
- Blanton, M. R. 2006, *The Astrophysical Journal*, 648, 268, doi: [10.1086/505628](https://doi.org/10.1086/505628)
- Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, *The Astronomical Journal*, 142, 31, doi: [10.1088/0004-6256/142/1/31](https://doi.org/10.1088/0004-6256/142/1/31)
- Blanton, M. R., & Moustakas, J. 2009, *Annual Review of Astronomy and Astrophysics*, 47, 159, doi: [10.1146/annurev-astro-082708-101734](https://doi.org/10.1146/annurev-astro-082708-101734)
- Blanton, M. R., & Roweis, S. 2007, *The Astronomical Journal*, 133, 734, doi: [10.1086/510127](https://doi.org/10.1086/510127)

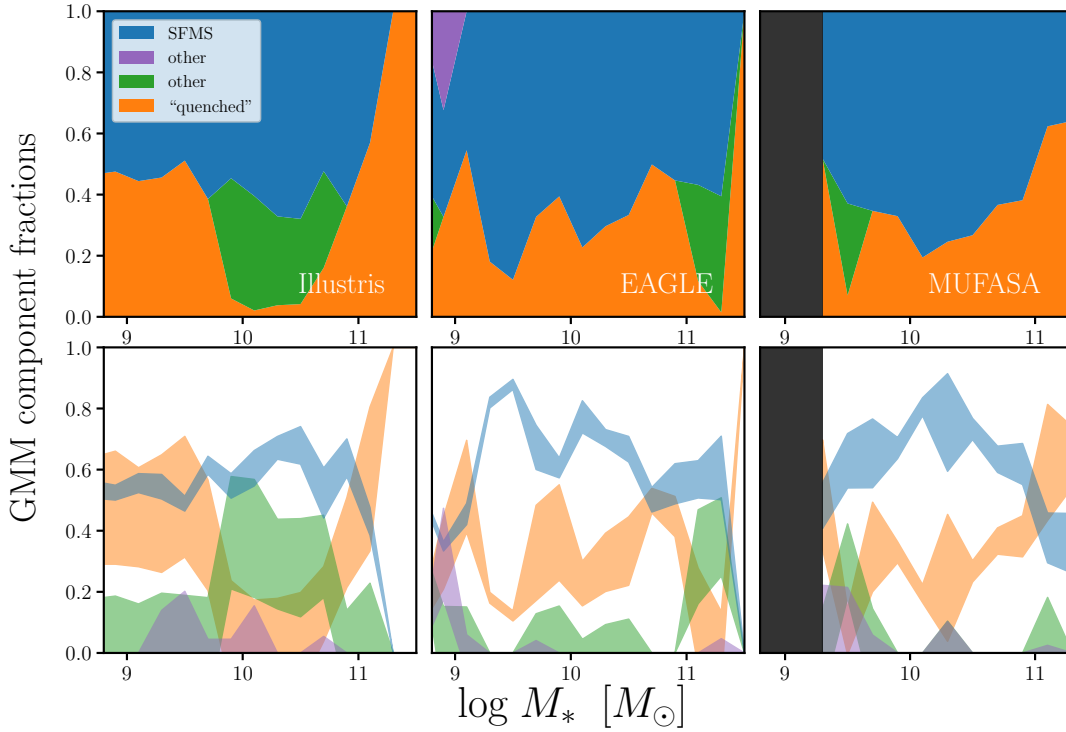


Figure 15. Fractional contributions, π_i , of the best-fit GMM components of our SFMS fitting for our hydrodynamic simulations (Illustris, EAGLE, and MUFASA), where the 100 Myr SFRs are sampled from the resolution range — $\text{SFR}'_i \in [\text{SFR}_i, \text{SFR}_i + \Delta_{\text{SFR}}]$. Comparing to Figures 9 and 10, we find SFR resolution has no significant impact on the qualitative results from Section 4.2.

- Blanton, M. R., Hogg, D. W., Bahcall, N. A., et al. 2003, *The Astrophysical Journal*, 594, 186, doi: [10.1086/375528](https://doi.org/10.1086/375528)
- Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, *The Astronomical Journal*, 129, 2562, doi: [10.1086/429803](https://doi.org/10.1086/429803)
- Bluck, A. F. L., Mendel, J. T., Ellison, S. L., et al. 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 2559, doi: [10.1093/mnras/stw1665](https://doi.org/10.1093/mnras/stw1665)
- Borch, A., Meisenheimer, K., Bell, E. F., et al. 2006, *Astronomy and Astrophysics*, 453, 869, doi: [10.1051/0004-6361:20054376](https://doi.org/10.1051/0004-6361:20054376)
- Bovy, J., Hogg, D. W., & Roweis, S. T. 2011, *The Annals of Applied Statistics*, 5, 1657, doi: [10.1214/10-AOAS439](https://doi.org/10.1214/10-AOAS439)
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *Monthly Notices of the Royal Astronomical Society*, 351, 1151, doi: [10.1111/j.1365-2966.2004.07881.x](https://doi.org/10.1111/j.1365-2966.2004.07881.x)
- Broderick, A. E., Fish, V. L., Doeleman, S. S., & Loeb, A. 2011, *The Astrophysical Journal*, 735, 110, doi: [10.1088/0004-637X/735/2/110](https://doi.org/10.1088/0004-637X/735/2/110)
- Bruzual A., G., & Charlot, S. 1993, *The Astrophysical Journal*, 405, 538, doi: [10.1086/172385](https://doi.org/10.1086/172385)
- Bundy, K., Ellis, R. S., Conselice, C. J., et al. 2006, *The Astrophysical Journal*, 651, 120, doi: [10.1086/507456](https://doi.org/10.1086/507456)
- Campbell, D., van den Bosch, F. C., Hearin, A., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 444, doi: [10.1093/mnras/stv1091](https://doi.org/10.1093/mnras/stv1091)
- Chabrier, G. 2003, *Publications of the Astronomical Society of the Pacific*, 115, 763, doi: [10.1086/376392](https://doi.org/10.1086/376392)
- Charlot, S., & Fall, S. M. 2000, *The Astrophysical Journal*, 539, 718, doi: [10.1086/309250](https://doi.org/10.1086/309250)
- Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1937, doi: [10.1093/mnras/stv725](https://doi.org/10.1093/mnras/stv725)
- Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *The Astrophysical Journal*, 670, 156, doi: [10.1086/521818](https://doi.org/10.1086/521818)
- Dalla Vecchia, C., & Schaye, J. 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 140, doi: [10.1111/j.1365-2966.2012.21704.x](https://doi.org/10.1111/j.1365-2966.2012.21704.x)

- Davé, R., Rafieferantsoa, M. H., & Thompson, R. J. 2017a, arXiv:1704.01135 [astro-ph].
<https://arxiv.org/abs/1704.01135>
- . 2017b, *Monthly Notices of the Royal Astronomical Society*, 471, 1671, doi: [10.1093/mnras/stx1693](https://doi.org/10.1093/mnras/stx1693)
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *The Astrophysical Journal*, 292, 371, doi: [10.1086/163168](https://doi.org/10.1086/163168)
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1
- Dressler, A. 1980, *The Astrophysical Journal*, 236, 351, doi: [10.1086/157753](https://doi.org/10.1086/157753)
- Efron, B. 1979, *The Annals of Statistics*, 7, 1, doi: [10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552)
- Feldmann, R. 2017, *Monthly Notices of the Royal Astronomical Society*, 470, L59, doi: [10.1093/mnras/slx073](https://doi.org/10.1093/mnras/slx073)
- Ferland, G. J. 1996, *Hazy, A Brief Introduction to Cloudy* 90
- Fraley, C., & Raftery, A. E. 1998, *The Computer Journal*, 41, 578, doi: [10.1093/comjnl/41.8.578](https://doi.org/10.1093/comjnl/41.8.578)
- Furlong, M., Bower, R. G., Theuns, T., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 4486, doi: [10.1093/mnras/stv852](https://doi.org/10.1093/mnras/stv852)
- Geha, M., Blanton, M. R., Yan, R., & Tinker, J. L. 2012, *The Astrophysical Journal*, 757, 85, doi: [10.1088/0004-637X/757/1/85](https://doi.org/10.1088/0004-637X/757/1/85)
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 175, doi: [10.1093/mnras/stu1654](https://doi.org/10.1093/mnras/stu1654)
- Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, *The Astrophysical Journal*, 489, 37, doi: [10.1086/304788](https://doi.org/10.1086/304788)
- Hahn, C., Tinker, J. L., & Wetzel, A. R. 2017, *The Astrophysical Journal*, 841, 6, doi: [10.3847/1538-4357/aa6d6b](https://doi.org/10.3847/1538-4357/aa6d6b)
- Hahn, C., Blanton, M. R., Moustakas, J., et al. 2015, *The Astrophysical Journal*, 806, 162, doi: [10.1088/0004-637X/806/2/162](https://doi.org/10.1088/0004-637X/806/2/162)
- Hopkins, A. M., & Beacom, J. F. 2006, *The Astrophysical Journal*, 651, 142, doi: [10.1086/506610](https://doi.org/10.1086/506610)
- Hubble, E. P. 1936, *Realm of the Nebulae*
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *Monthly Notices of the Royal Astronomical Society*, 341, 54, doi: [10.1046/j.1365-8711.2003.06292.x](https://doi.org/10.1046/j.1365-8711.2003.06292.x)
- Kennicutt, R. C., & Evans, N. J. 2012, *Annual Review of Astronomy and Astrophysics*, 50, 531, doi: [10.1146/annurev-astro-081811-125610](https://doi.org/10.1146/annurev-astro-081811-125610)
- Lee, K. J., Guillemot, L., Yue, Y. L., Kramer, M., & Champion, D. J. 2012, *Monthly Notices of the Royal Astronomical Society*, 424, 2832, doi: [10.1111/j.1365-2966.2012.21413.x](https://doi.org/10.1111/j.1365-2966.2012.21413.x)
- Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, *The Astrophysical Journal*, 801, 80, doi: [10.1088/0004-637X/801/2/80](https://doi.org/10.1088/0004-637X/801/2/80)
- Leja, J., van Dokkum, P. G., Franx, M., & Whitaker, K. E. 2015, 798, 115, doi: [10.1088/0004-637X/798/2/115](https://doi.org/10.1088/0004-637X/798/2/115)
- Leroux, B. G. 1992, *The Annals of Statistics*, 20, 1350, doi: [10.1214/aos/1176348772](https://doi.org/10.1214/aos/1176348772)
- Liddle, A. R. 2007, *Monthly Notices of the Royal Astronomical Society*, 377, L74, doi: [10.1111/j.1745-3933.2007.00306.x](https://doi.org/10.1111/j.1745-3933.2007.00306.x)
- Lloyd, S. 1982, *IEEE Transactions on Information Theory*, 28, 129, doi: [10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489)
- Madau, P., & Dickinson, M. 2014, *Annual Review of Astronomy and Astrophysics*, 52, 415, doi: [10.1146/annurev-astro-081811-125615](https://doi.org/10.1146/annurev-astro-081811-125615)
- Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, *The Astrophysical Journal*, 760, 6, doi: [10.1088/0004-637X/760/1/6](https://doi.org/10.1088/0004-637X/760/1/6)
- McLachlan, G., & Peel, D. 2000, *Finite Mixture Models* (Wiley-Interscience)
- Moustakas, J., Coil, A. L., Aird, J., et al. 2013, *The Astrophysical Journal*, 767, 50, doi: [10.1088/0004-637X/767/1/50](https://doi.org/10.1088/0004-637X/767/1/50)
- Muldrew, S. I., Croton, D. J., Skibba, R. A., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 419, 2670, doi: [10.1111/j.1365-2966.2011.19922.x](https://doi.org/10.1111/j.1365-2966.2011.19922.x)
- Neal, R. M., & Hinton, G. E. 1998, in *Learning in Graphical Models*, NATO ASI Series (Springer, Dordrecht), 355–368
- Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, *The Astrophysical Journal Letters*, 660, L43, doi: [10.1086/517926](https://doi.org/10.1086/517926)
- Oemler, Jr., A. 1974, *The Astrophysical Journal*, 194, 1, doi: [10.1086/153216](https://doi.org/10.1086/153216)
- Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, *The Astrophysical Journal*, 721, 193, doi: [10.1088/0004-637X/721/1/193](https://doi.org/10.1088/0004-637X/721/1/193)
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing* (New York, NY, USA: Cambridge University Press)

- Rafieerantsoa, M., & Davé, R. 2018, *Monthly Notices of the Royal Astronomical Society*, 475, 955, doi: [10.1093/mnras/stx3293](https://doi.org/10.1093/mnras/stx3293)
- Roeder, K., & Wasserman, L. 1997, *Journal of the American Statistical Association*, 92, 894, doi: [10.1080/01621459.1997.10474044](https://doi.org/10.1080/01621459.1997.10474044)
- Salim, S., Rich, R. M., Charlot, S., et al. 2007, *The Astrophysical Journal Supplement Series*, 173, 267, doi: [10.1086/519218](https://doi.org/10.1086/519218)
- Schaller, M., Dalla Vecchia, C., Schaye, J., et al. 2015, 454, 2277, doi: [10.1093/mnras/stv2169](https://doi.org/10.1093/mnras/stv2169)
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 521, doi: [10.1093/mnras/stu2058](https://doi.org/10.1093/mnras/stu2058)
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *Astronomy and Astrophysics*, 575, A74, doi: [10.1051/0004-6361/201425017](https://doi.org/10.1051/0004-6361/201425017)
- Schwarz, G. 1978, *The Annals of Statistics*, 6, 461, doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
- Sijacki, D., Springel, V., Di Matteo, T., & Hernquist, L. 2007, *Monthly Notices of the Royal Astronomical Society*, 380, 877, doi: [10.1111/j.1365-2966.2007.12153.x](https://doi.org/10.1111/j.1365-2966.2007.12153.x)
- Somerville, R. S., & Davé, R. 2015, *Annual Review of Astronomy and Astrophysics*, 53, 51, doi: [10.1146/annurev-astro-082812-140951](https://doi.org/10.1146/annurev-astro-082812-140951)
- Sparre, M., Hayward, C. C., Springel, V., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 447, 3548, doi: [10.1093/mnras/stu2713](https://doi.org/10.1093/mnras/stu2713)
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *The Astrophysical Journal Supplement Series*, 214, 15, doi: [10.1088/0067-0049/214/2/15](https://doi.org/10.1088/0067-0049/214/2/15)
- Springel, V. 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 1105, doi: [10.1111/j.1365-2966.2005.09655.x](https://doi.org/10.1111/j.1365-2966.2005.09655.x)
- . 2010, 401, 791, doi: [10.1111/j.1365-2966.2009.15715.x](https://doi.org/10.1111/j.1365-2966.2009.15715.x)
- Springel, V., & Hernquist, L. 2003, *Monthly Notices of the Royal Astronomical Society*, 339, 289, doi: [10.1046/j.1365-8711.2003.06206.x](https://doi.org/10.1046/j.1365-8711.2003.06206.x)
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 726, doi: [10.1046/j.1365-8711.2001.04912.x](https://doi.org/10.1046/j.1365-8711.2001.04912.x)
- Steele, R. J., & Raftery, A. E. 2010
- Taylor, E. N., Franx, M., van Dokkum, P. G., et al. 2009, *The Astrophysical Journal*, 694, 1171, doi: [10.1088/0004-637X/694/2/1171](https://doi.org/10.1088/0004-637X/694/2/1171)
- Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 2144, doi: [10.1093/mnras/stu1900](https://doi.org/10.1093/mnras/stu1900)
- Terrazas, B. A., Bell, E. F., Woo, J., & Henriques, B. M. B. 2017, *The Astrophysical Journal*, 844, 170, doi: [10.3847/1538-4357/aa7d07](https://doi.org/10.3847/1538-4357/aa7d07)
- Tinker, J., Wetzel, A., & Conroy, C. 2011, *ArXiv e-prints*, 1107, arXiv:1107.5046
- Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 438, 1985, doi: [10.1093/mnras/stt2295](https://doi.org/10.1093/mnras/stt2295)
- Trayford, J. W., Theuns, T., Bower, R. G., et al. 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 2879, doi: [10.1093/mnras/stv1461](https://doi.org/10.1093/mnras/stv1461)
- Trayford, J. W., Camps, P., Theuns, T., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 771, doi: [10.1093/mnras/stx1051](https://doi.org/10.1093/mnras/stx1051)
- Vakili, M., & Hahn, C. H. 2016, arXiv:1610.01991 [astro-ph]. <https://arxiv.org/abs/1610.01991>
- Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 3031, doi: [10.1093/mnras/stt1789](https://doi.org/10.1093/mnras/stt1789)
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 1518, doi: [10.1093/mnras/stu1536](https://doi.org/10.1093/mnras/stu1536)
- Wang, L., Farrah, D., Oliver, S. J., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 431, 648, doi: [10.1093/mnras/stt190](https://doi.org/10.1093/mnras/stt190)
- Wang, L., Norberg, P., Brough, S., et al. 2018, arXiv:1802.08456 [astro-ph]. <https://arxiv.org/abs/1802.08456>
- Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, *Monthly Notices of the Royal Astronomical Society*, 432, 336, doi: [10.1093/mnras/stt469](https://doi.org/10.1093/mnras/stt469)
- Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, *The Astrophysical Journal Letters*, 754, L29, doi: [10.1088/2041-8205/754/2/L29](https://doi.org/10.1088/2041-8205/754/2/L29)
- Yan, R. 2011, *The Astronomical Journal*, 142, 153, doi: [10.1088/0004-6256/142/5/153](https://doi.org/10.1088/0004-6256/142/5/153)
- Yan, R., & Blanton, M. R. 2012, *The Astrophysical Journal*, 747, 61, doi: [10.1088/0004-637X/747/1/61](https://doi.org/10.1088/0004-637X/747/1/61)
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *The Astronomical Journal*, 120, 1579, doi: [10.1086/301513](https://doi.org/10.1086/301513)