A new census of the 0.2 < z < 3.0 universe. II. The star-forming sequence

Leja, J.; Speagle, J.S.; Ting, Y.-S.; Johnson, B.D.; Conroy, C.; Whitaker, K.E.; ... ; Franx, M.

Version:	Publisher's Version
License:	
Downloaded from:	

Note: To cite this publication please use the final published version (if applicable).

A New Census of the $0.2 < z < 3.0$ Universe. II. The Star-forming Sequence

Joel Leja[1,2,3] ⃝, Joshua S. Speagle (沈佳士)[4,5,6,18] ⃝, Yuan-Sen Ting (丁源森)[7,8,9,10,11,19] ⃝, Benjamin D. Johnson[12] ⃝,
Charlie Conroy[12] ⃝, Katherine E. Whitaker[13,14] ⃝, Erica J. Nelson[15] ⃝, Pieter van Dokkum[16] ⃝, and Marijn Franx[17] ⃝

[1] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA; joel.leja@psu.edu
[2] Institute for Computational & Data Sciences, The Pennsylvania State University, University Park, PA, USA
[3] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[4] Department of Statistical Sciences, University of Toronto, Toronto, ON M5S 3G3, Canada
[5] David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada
[6] Dunlap Institute for Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada
[7] Research School of Astronomy & Astrophysics, Australian National University, Cotter Rd., Weston, ACT 2611, Australia
[8] Institute for Advanced Study, Princeton, NJ 08540, USA
[9] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08540, USA
[10] Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA
[11] Research School of Computer Science, Australian National University, Acton ACT 2601, Australia
[12] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street Cambridge, MA 02138, USA
[13] Department of Astronomy, University of Massachusetts Amherst, 710 N Pleasant Street, Amherst, MA 01003, USA
[14] Cosmic Dawn Center (DAWN), Copenhagen, Denmark
[15] Astrophysical & Planetary Sciences, 391 UCB, 2000 Colorado Avenue, Duane Physics Building, Rm. E226, Boulder, CO 80309, USA
[16] Department of Astronomy, Yale University, New Haven, CT 06511, USA
[17] Leiden Observatory, Leiden University, NL-2300 RA Leiden, The Netherlands

Abstract

We use the panchromatic spectral energy distribution (SED)-fitting code Prospector to measure the galaxy $\log M^*$–logSFR relationship (the *star-forming sequence*) across $0.2 < z < 3.0$ using the COSMOS-2015 and 3D-HST UV-IR photometric catalogs. We demonstrate that the chosen method of identifying star-forming galaxies introduces a systematic uncertainty in the inferred normalization and width of the star-forming sequence, peaking for massive galaxies at ∼0.5 and ∼0.2 dex, respectively. To avoid this systematic, we instead parameterize the density of the full galaxy population in the $\log M^*$–logSFR–redshift plane using a flexible neural network known as a normalizing flow. The resulting star-forming sequence has a low-mass slope near unity and a much flatter slope at higher masses, with a normalization 0.2–0.5 dex lower than typical inferences in the literature. We show this difference is due to the sophistication of the Prospector stellar populations modeling: the nonparametric star formation histories naturally produce higher masses while the combination of individualized metallicity, dust, and star formation history constraints produce lower star formation rates (SFRs) than typical UV+IR formulae. We introduce a simple formalism to understand the difference between SFRs inferred from SED fitting and standard template-based approaches such as UV+IR SFRs. Finally, we demonstrate the inferred star-forming sequence is consistent with predictions from theoretical models of galaxy formation, resolving a long-standing ~ 0.2–0.5 dex offset with observations at $0.5 < z < 3$. The fully trained normalizing flow including a nonparametric description of $\rho(\log M^*, \log\mathrm{SFR}, z)$ is available online[20] to facilitate straightforward comparisons with future work.

Unified Astronomy Thesaurus concepts: Galaxy formation (595); Galaxy photometry (611); Galaxy masses (607); Star formation (1569)

1. Introduction

The star-forming sequence is the relationship between galaxy star formation rate (SFR) and stellar mass (M^*) observed in star-forming galaxies.[21] The majority of galaxies form the majority of their mass either *on* (e.g., Leitner 2012) or *passing through* (e.g., Abramson et al. 2015) the star-forming sequence, making a robust characterization of the star-forming sequence crucial to understanding the process of galaxy evolution as a whole. As a result, the star-forming sequence has been well measured in the literature across a variety of rest-frame wavelengths and with a range of assumptions and techniques (Daddi et al. 2007; Noeske et al. 2007; Karim et al. 2011; Rodighiero et al. 2011; Whitaker et al. 2012, 2014; Speagle et al. 2014; Renzini & Peng 2015; Schreiber et al. 2015; Tomczak et al. 2016; Leslie et al. 2020).

Yet despite the importance of the star-forming sequence, there is a large dispersion across the literature in its inferred normalization and slope. Speagle et al. (2014) performed an exhaustive cross calibration of measurements of the star-forming sequence in the astronomical literature. They showed that there are two key sets of systematics that drive the dispersion in measurements of the star-forming sequence.

The first systematic is the method chosen to identify star-forming galaxies and—usually at the same time—to separate quiescent galaxies. These different methods include rest-frame color–color selection (e.g., *UVJ*, Williams et al. 2009;

---

[18] Banting & Dunlap Fellow.
[19] Hubble Fellow.
[20] https://github.com/jrleja/sfs_leja_trained_flow
[21] This is often called the *main sequence* of star-forming galaxies in the literature, although we do not use the term here to avoid confusion with the stellar main sequence.

sBzK, Daddi et al. 2004), fixed lines in the SFR–M_* plane (e.g., specific star formation rate (sSFR) cuts), lower limits on observed emission line luminosities or equivalent widths, sigma-clipping fits to the star-forming sequence, Lyman-break or luminous infrared galaxy selection, or no preselection at all. Speagle et al. (2014) demonstrate that the measured slope of the log(SFR)–log(M^*) relationship ranges between 0 and 1 in the literature, and this variance correlates strongly with the chosen method of identifying and classifying star-forming galaxies. This result has also been observed in simulations. For instance, Donnari et al. (2019) contrast rest-frame color-based selection with selection based on bimodality in the SFR–mass plane in the IllustrisTNG hydrodynamical simulation of galaxy formation. They find that this choice affects the normalization of the star-forming sequence at the ∼0.5 dex level, and while these selections generally produce identical fractions of quiescent galaxies at low stellar masses (to within 5%–10%), this measured fraction will vary between 20% and 40% at high masses ($M/M_\odot > 10^{10.5}$).

The second systematic challenge in observational measurements of the star-forming sequence is the method by which galaxy SFRs and (to a somewhat lesser extent) stellar masses are inferred. Speagle et al. (2014) find an interpublication 1σ scatter of ∼0.2 dex in the measured normalization of the star-forming main sequence, the great majority of which they attribute to systematics in the inference techniques for galaxy SFRs and stellar masses. Again, these challenges are straightforward to reproduce in simulated data: Katsianis et al. (2020) calculate UV, IR, and spectral energy distribution (SED) SFRs from observations of simulated galaxies including dusty radiative transfer effects and find even more dramatic offsets of 0.3–0.5 dex, with additional strong systematic trends with redshift.

By creating calibration offsets associated with each measurement methodology, Speagle et al. (2014) are able to bring these independent observational measurements of the star-forming sequence in the literature onto a common scale. This shows that these offsets are *precise* and *predictable*; however, it does not ensure that one or any of them are *accurate*.

More insight can be had by examining the star-forming sequence in simulations of galaxy formation. These simulations produce star-forming sequences that are systematically 0.2–0.5 dex lower than the observed star-forming sequences at $z = 0.5$–3, yet are still able to reproduce the evolution of the stellar mass function with reasonable accuracy. This offset has remained for nearly a decade and is present across a diverse range of galaxy formation models such as Illustris, both the original (Sparre et al. 2015) and TNG (Donnari et al. 2019); SIMBA (Davé et al. 2019); EAGLE (Furlong et al. 2015); the Santa Cruz semi-analytical model (Somerville & Davé 2015); and other analytical models of galaxy formation (Lilly et al. 2013; Dekel & Burkert 2014). Indeed, matching these observations seems to require a strong but temporary decoupling of baryonic growth from halo growth at $1 < z < 3$ (Mitchell et al. 2014). Indeed, direct calculations suggest the simulations may be attempting to match two incompatible sets of observations, as the total mass growth implied by the observed evolution of the mass function is lower than the total SFR density inferred from the observed star-forming sequence by ∼0.3 dex (Leja et al. 2015).

It is therefore crucial to develop better machinery to measure the star-forming sequence from observations, both to better characterize the star-forming sequence itself—which plays a central role in galaxy evolution—and to provide more accurate constraints on this sequence to simulations. One potential solution to the challenge of quiescent/star-forming separation is the adoption of data-driven definitions of the star-forming sequence which do not rely on any a priori separation between star-forming and quiescent galaxies. Recent solutions in the literature include defining the star-forming sequence as the *ridgeline* (i.e., the density peak in the SFR–mass plane) (Renzini & Peng 2015), or using flexible Gaussian mixture models with the number of Gaussian subcomponents determined by the data (Hahn et al. 2019). In the limit of wide-area surveys, these data-driven approaches can accurately measure higher-order moments of the star-forming sequence such as the width or the location of the so-called *green valley* (Sherman et al. 2021).

There are also methodologies emerging to infer stellar population parameters from observations more robustly, producing a promising way forward to mitigating or even eliminating the systematics in SFR and mass measurements. Often SFRs are inferred using direct flux-to-SFR conversions which are derived via fixed assumptions about the properties of galaxy stellar populations (e.g., Kennicutt & Evans 2012). This can result in significant measurement systematics when applied to a galaxy population with a rich diversity of properties (Wuyts et al. 2011a; Leja et al. 2019c). New generations of Bayesian galaxy SED models using highly flexible on-the-fly model generation (e.g., `Beagle`, Chevallard & Charlot 2016; `Prospector`, Leja et al. 2017; Johnson et al. 2021) or large pre-generated grids of models (e.g., `Bagpipes`, Carnall et al. 2018; `CIGALE`, Boquien et al. 2019) can now fit sophisticated models to all of the available data. These methods constrain stellar population properties such as star formation histories, dust attenuation, and metallicities on an object-by-object basis, as opposed to earlier works using fixed values or highly simplified models. The net effect is to produce more sophisticated (and, ideally, more accurate) conversions between the observed flux and the SFR and stellar masses.

In this work, we measure the star-forming sequence using stellar population properties inferred by the `Prospector-α` model built in the `Prospector` SED-fitting code. This work is motivated by the distinct differences in inferred masses and SFRs from previous analyses (Leja et al. 2019c). We additionally use a new machine-learning technique to measure the star-forming sequence: normalizing flows (Jimenez Rezende & Mohamed 2015). Normalizing flows allow the direct inference of the density in the SFR–M^* plane from individual observations. Crucially, this density inference can be modified to marginalize over measurement uncertainty as described later in this work—indeed, Curtis-Lake et al. (2021) have shown that proper modeling of the measurement uncertainties has a significant effect on the inferred slope and normalization of the star-forming sequence. This density estimate in turn permits measurement of the *ridge* of the star-forming sequence without any preselection of star-forming galaxies, as well as facilitating direct comparisons with simulations or other observations using *density* in SFR–M^*-redshift space. This paper is the second in a series; the first paper presented a Bayesian population model for the stellar mass function (Leja et al. 2020).

This paper is structured as follows. Section 2 describes the photometric data and the subsequent SED modeling. Section 3 uses these data to demonstrate the sensitivity of the star-forming sequence to a bimodal classification system. Section 4 introduces the normalizing flow method and describes our alterations to the standard methodology to include marginalization over the parameter uncertainties from SED modeling. Section 5 analyzes the resulting star-forming sequence and presents simple equations describing the evolution with stellar mass and redshift. Section 6 compares results in the literature, and provides a general framework in which to understand how SED-based SFRs are different than standard UV+IR formulae. Section 7 contains a broader discussion of the results and the conclusion is in Section 8. We use a Chabrier (2003) initial mass function and adopt a Wilkinson Microwave Anisotropy Probe 9 cosmology (Hinshaw et al. 2013) with $H_0 = 69.7$ km s^{-1} Mpc^{-1}, $\Omega_b = 0.0464$, and $\Omega_c = 0.235$. Reported SFRs are averaged over the most recent 100 Myr (rather than instantaneous values) unless explicitly stated otherwise.

2. Data and SED Modeling

Here, we describe how the inputs to the star-forming sequence (i.e., stellar masses, SFRs, and redshifts) are derived. The input photometry and SED-fitting methodology are identical to those in Paper I (Leja et al. 2020) and are briefly summarized here for convenience. The first two sections describe the publicly available surveys from which photometry and redshifts are taken. The subsequent section describes the SED modeling used to translate these data into stellar population properties.

2.1. 3D-HST Catalog

The 3D-HST photometric catalogs (Skelton et al. 2014) collate all publicly available photometry in five deep, well-studied extragalactic fields comprising an area of ~900 arcmin^2. In particular, this includes deep near-IR imaging from the CANDELS HST Treasury Program (Grogin et al. 2011; Koekemoer et al. 2011). This photometric catalog is supplemented with Spitzer/MIPS 24 μm photometry from Whitaker et al. (2014). In total, each galaxy is covered over 0.3–24 μm in the observed frame with between 17 and 44 bands of aperture-matched photometry.

The redshifts come from several sources. The 3D-HST grism survey (Momcheva et al. 2016) provides reliable space-based grism redshifts. The rest of the sample consists of photometric redshifts estimated with the publicly available EAZY (Brammer et al. 2008). Approximately 30% of the sample redshifts are from ground-based spectroscopy or reliable grism measurements, while the remainder is based on photometric estimates. The photometric redshifts are taken as the peak of the redshift probability distribution function. Thanks to the wide wavelength coverage and mixture of narrow-band and medium-band filters, they show a relatively low scatter of $0.0197(1+z)$ compared to the grism redshifts (Bezanson et al. 2016).

We select a stellar mass-complete sample of 26,971 galaxies at $0.5 < z < 3.0$ from this catalog. The lower redshift limit is set by the redshift at which the aperture-based photometry measurements start to become unreliable, and the upper redshift limit is where space-based (i.e., HST) rest-frame optical imaging is no longer available for target selection. Mass-complete limits are taken from Leja et al. (2020), with an

adjustment upwards by 0.1 dex in this work at $z = 0.65$ and $z = 2.1$ to remove an observed upturn in the star-forming sequence at the lowest masses. This is necessary because the stellar mass-complete limit is set at 90%, and a 10% incompleteness rate can have a substantive effect on $\langle \text{SFR} \rangle$, as red galaxies with low SFRs are more likely to drop out. The level of this adjustment is determined by examining the $\log M^* – \log$SFR distribution in small redshift slices at each redshift, and requiring that the distribution of galaxies in $\log$SFR at the stellar mass-complete limit is similar to that at slightly higher stellar masses. As none of the conclusions are sensitive to the SFR near the mass-complete limit, this adjustment has no effect on the conclusions of this work.

2.2. COSMOS-2015 Catalog

We additionally include photometry and redshifts from the COSMOS-2015 catalog (Laigle et al. 2016) to bolster the low-redshift volume. This includes 30 bands of photometry covering the observed frame of 0.2–24 μm, with a small fraction (<1%) of objects with Herschel far-IR imaging. We select only objects in that overlap with the UltraVISTA survey (McCracken et al. 2012) to include crucial near-IR photometric bands, producing an effective area of 1.38 deg^2.

The majority of the redshifts are photometric, estimated using LePhare (Ilbert et al. 2006). These have a very low scatter of 0.007–0.009 $(1+z)$ in the targeted redshift range (Laigle et al. 2016), though notably this comparison is only performed for bright galaxies targeted by spectroscopic programs.

We select a stellar-mass-complete sample of 41,148 objects at $0.2 < z < 0.8$ from this catalog, where the lower limit avoids the saturation limit for bright galaxies (Davidzon et al. 2017) and the upper limit ensures overlap with the 3D-HST sample. The mass-complete limits are again taken from Leja et al. (2020). Following the same procedure described in Section 2.1, we also adjust the $z = 0.175$ ($z = 0.5$) mass-complete limit in COSMOS-2015 upward by 0.25 (0.1) dex.

2.3. SED Modeling

The photometry is fit using the galaxy SED-fitting code `Prospector` (Leja et al. 2017; Johnson et al. 2021), powered by the FSPS stellar population synthesis code (Conroy et al. 2009). Within FSPS, the MIST stellar isochrones are adopted (Choi et al. 2016; Dotter 2016).

The physical model is the `Prospector-α` model from Leja et al. (2019c), which is a modified version of the model from Leja et al. (2017). In brief, the model has 14 free parameters, consisting of a seven-component nonparametric star formation history using the *continuity* prior, which disfavors sharp changes in SFR(t) (Leja et al. 2019a), a two-component dust attenuation model with a flexible dust attenuation curve (Noll et al. 2009), free gas-phase and stellar metallicity, and a mid-IR active galactic nuclei (AGN) component with a free normalization and dust optical depth. `Prospector` includes dust emission powered by energy balance (da Cunha et al. 2008) with an SED and nebular emission self-consistently powered by the model stellar ionizing continuum (Byler et al. 2017).

Notably, in contrast to typical UV+IR SFRs, SFRs can be measured via this methodology for every object in the survey regardless of whether they are detected in Spitzer/MIPS

24 μm. Of course, objects without IR detections will naturally have larger parameter uncertainties.

3. Demonstrating the Sensitivity to Bimodal Classification

Here we use the SED-fitting posteriors from the previous section to demonstrate the sensitivity of the star-forming sequence to the method adopted for identifying star-forming galaxies. We select galaxies between $0.3 < z < 0.7$ from our data, separate them into star-forming and quiescent galaxies using four different methods described below, and fit a Gaussian with a mass-dependent mean and width to the distribution of log(SFR) in the identified star-forming galaxies. The mean and standard deviation of the Gaussian are allowed to vary quadratically with stellar mass. We take $N = 50$ posterior samples for each galaxy and marginalize the likelihood over the parameter uncertainties, following Leja et al. (2020). This effectively results in an error-deconvolved model of the star-forming sequence. We explore the posterior with the nested sampling code `dynesty` (Speagle 2020).

We examine four common methods to separate between star-forming and quiescent galaxies:

1. Hα equivalent width: The Hα equivalent width is estimated using the `Prospector` posterior predictions for each galaxy. These predictions are reasonably accurate, showing a 1σ scatter of $\sim$0.2 dex (Leja et al. 2017) when compared to spectroscopic observations. We use an equivalent width (EW) cut of 35 Å, which is approximately 5–10 Å below the EW mean of star-forming galaxies at this epoch (Fumagalli et al. 2012).

2. UVJ colors: Rest-frame *UVJ* colors are synthesized from the `Prospector` posteriors for each galaxy, and the galaxy is assigned to be star-forming or quiescent based on the *UVJ* quiescent region from Whitaker et al. (2012). When the object straddles the quiescent/star-forming boundary, the classification is assigned following the classification of the *majority* of the posterior weight.

3. Fixed sSFR: A fixed sSFR boundary at $\log(\mathrm{sSFR/yr}^{-1}) = -10.5$ separates star-forming and quiescent galaxies, with objects near the border again assigned based on the majority of their posterior weight.

4. Gaussian Mixture: Two Gaussians are fit to the log (SFR)–log(M^*) distribution, where here the second Gaussian is intended to describe the quiescent sequence. The means and widths vary quadratically with mass as before. To ensure robust separation between these two sequences, the mean of the second Gaussian is defined relative to the first one via

$$x_{\mathrm{qu}} = x_{\mathrm{sf}} - (\sigma_{\mathrm{SF}} + \sigma_{\mathrm{QU}}) - \Delta_{\mathrm{QU}}, \qquad (1)$$

where Δ_{QU} is the fit parameter. This parameter has a uniform prior taken between 0.0 and 2.0, σ_{SF} and σ_{qu} are the standard deviations of the star-forming and quiescent sequences, and x_{sf} and x_{qu} are their means. This formulation ensures that the two sequences have a minimum separation set by the sum of their standard deviations.

The results of these fits are shown in Figure 1. There is reasonable agreement between the techniques below $\log(M/M_\odot) \lesssim 10.5$, with agreement in the normalization of the star-forming sequence to within $\lesssim$0.1 dex and to within 0.05 dex for the 1σ width.

The striking differences occur at the massive end, where there is both a larger percentage of quiescent galaxies and a much less clear delineation between the two populations in the $\log M^*$–$\log$SFR plane (both of these effects can be seen directly in the Gaussian mixture panel of Figure 1). As a result, there is a 0.5 dex range in the inferred normalization at the massive end and a 0.2 dex range in the inferred 1σ widths. *UVJ* colors in particular produce a wide star-forming sequence at the massive end, a result that naturally follows from the flattening of the relationship between *UVJ* colors and specific star formation rate at $\log(\mathrm{sSFR/yr}^{-1}) < -10.5$ (Leja et al. 2019b).

The inferred quiescent fraction shows substantial variation, with a difference of $\sim$20% between techniques that shows no clear dependence on stellar mass (see Donnari et al. 2019 for similar conclusions for simulated galaxies). Interestingly, no one technique produces a consistently higher or lower quiescent fraction at all masses, demonstrating how the traditional observational signatures of quiescence are also affected by natural variation in underlying stellar populations properties (e.g., the level of time-variability in star formation histories will have a different effect on Hα versus *UVJ* quiescent indicators).

We also include the results from the analysis of the full population described later in this paper as gray dashed lines. These results are not quite directly comparable as the analysis is performed on the full population, but are useful as a point of comparison.

4. Inferring the Population Density with Normalizing Flows

In light of the challenges posed in cleanly separating star-forming and quiescent galaxies, we instead proceed by inferring the density $\rho(\log M^*, \log\mathrm{SFR}, z)$ in stellar mass, SFR, and redshift for the full galaxy population using a normalizing flow. From this, the star-forming sequence can be inferred directly by tracing the peak in the distribution of logSFR as a function of stellar mass and redshift, following Renzini & Peng (2015). We provide a brief description of normalizing flows below, along with some additional implementation details. It is not necessary to fully understand the normalizing flow treatment to follow the results of the paper, though it is important to note that the implementation below marginalizes over the physical parameter uncertainties for each galaxy (see Figure 2).

A normalizing flow is a type of neural network that tries to learn a nonparametric estimate of the continuous, underlying probability density function (PDF) $P_\theta(\boldsymbol{x})$ associated with a set of N observations $\{\boldsymbol{x}_1,...,\boldsymbol{x}_N\}$ given some parameters θ, which describe the transformation into the target distribution. For our purposes, $\boldsymbol{x} = \{\log M^*, \log\mathrm{SFR}, z\}$. The basic method by which a normalizing flow works is to determine a series of transformations f_{θ_k} from latent variables $\boldsymbol{z}$ (distinct from redshift, z) with a known simple distribution $P(\boldsymbol{z})$ to the observations $\boldsymbol{x}$ with unknown distribution $P_\theta(\boldsymbol{x})$. Here, θ is the set of parameters in the neural network used to transform to the target distribution. In most cases, $\boldsymbol{z}$ is assumed to follow a set of iid (independent and identically distributed) standard normal (i.e., unit Gaussian) distributions. In other words, the normalizing flow is trying to learn a series of transformations from three unit Gaussians to reproduce the distribution of the data in $\log M^*$, logSFR, and z.

Assuming that the data follow a point process and are independent and identically distributed (iid), a normalizing
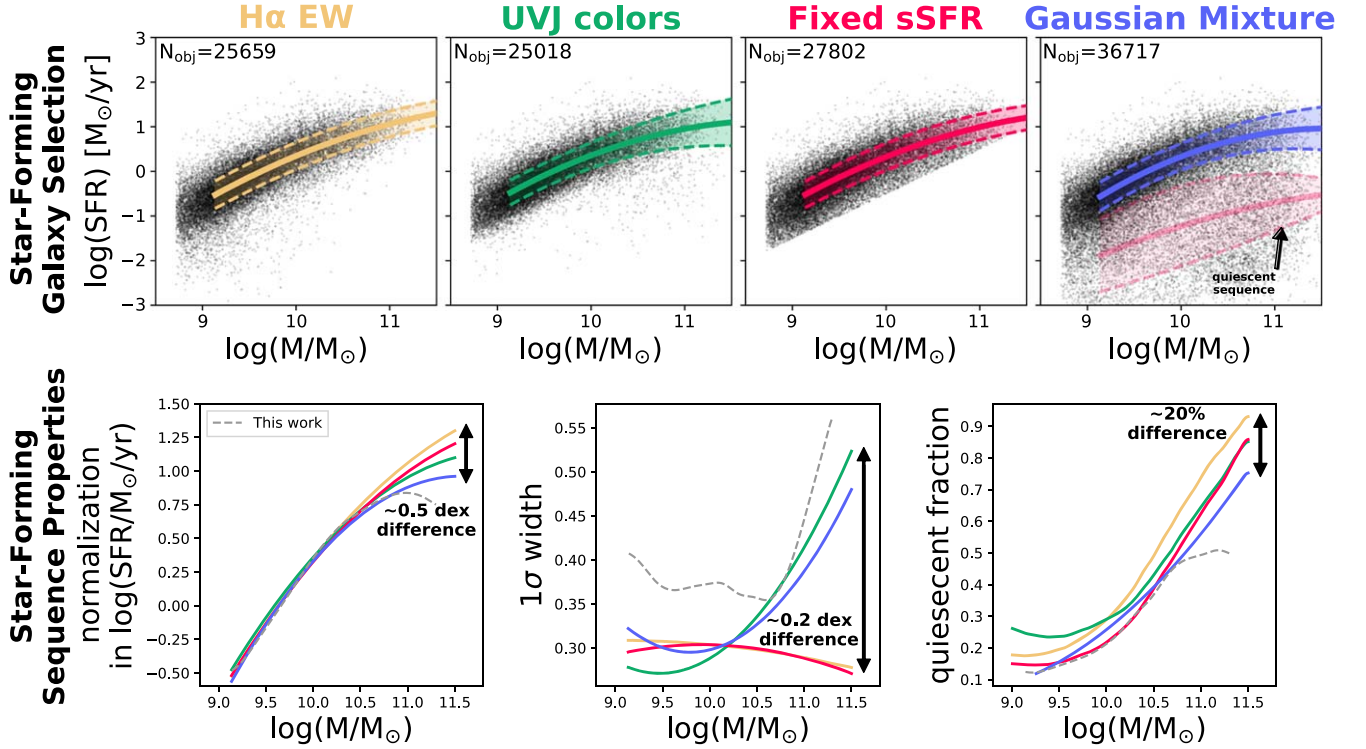
Figure 1. Four different methods to select star-forming galaxies, and the properties of the resulting star-forming sequences. The top panels show the SFR–M^* plane for galaxies at $0.3 < z < 0.7$ identified as star-forming via four different techniques, with the total number of objects included in the fit indicated in the upper left and the mean and 1σ scatter of the resulting star-forming sequence as solid and dashed lines, respectively. For the Gaussian mixture model, the quiescent sequence is also shown. The lower panels show the median and 1σ width of the resulting star-forming sequence and the fraction of quiescent galaxies inferred from each method. For reference, we additionally show the values derived later in this paper with a gray dashed line.

flow works by selecting a given $\theta_{\max}$ that maximizes the log-likelihood

$$\theta_{\max} = \mathrm{argmax}_\theta \left(\sum_{i=1}^{N} \log P_\theta(\boldsymbol{x}_i) \right). \qquad (2)$$

Taking advantage of the change of variables formula, for a series of k composite transformations

$$\boldsymbol{x} = f_\theta(\boldsymbol{z}) = f_{\theta_k}(\cdots f_{\theta_2}(f_{\theta_1}(\boldsymbol{z}))), \qquad (3)$$

where we have split $\theta = \{\theta_1,...,\theta_k\}$ into k portions, the PDF at any particular point depends on the initial PDF multiplied by the Jacobian:

$$P(\boldsymbol{x}) = P(f_\theta(\boldsymbol{z})) = P(\boldsymbol{z}) \times \left| \frac{\partial f_\theta(\boldsymbol{z})}{\partial \boldsymbol{z}} \right|^{-1}. \qquad (4)$$

Exploiting the chain rule and taking the logarithm, we can rewrite this expression as

$$\log P(\boldsymbol{x}) = \log P(f_\theta(\boldsymbol{z})) = \log P(\boldsymbol{z}) - \sum_{j=1}^{k} \log \left| \frac{\partial f_{\theta_j}(\boldsymbol{z}^{j-1})}{\partial \boldsymbol{z}^{j-1}} \right|, \qquad (5)$$

where $\boldsymbol{z}^j = f_{\theta_j}(\cdots f_{\theta_1}(\boldsymbol{z}))$ and $\boldsymbol{z}^0 = \boldsymbol{z}$. The core challenge for training normalizing flows is to ensure that the series of transformations f_{θ_j} have easy-to-compute Jacobians (to evaluate the log-PDF) and inverses $f_{\theta_j}^{-1}$ (to go from $\boldsymbol{x} \to \boldsymbol{z}$ as well as $\boldsymbol{z} \to \boldsymbol{x}$) while still retaining the flexibility to model complex distributions.

For the normalizing flow in this work, we build on the same basic architecture and training procedure as Ting & Weinberg (2022), using a batch size of 1024 and 500 epochs. This architecture consists of eight units of a *neural spline flow*, each of which consists of three layers of densely connected neural networks with 16 neurons, coupled with a "Conv1x1" operation (often referred to as "GLOW" in the machine learning literature; see, e.g., Kingma & Dhariwal 2018; Durkan et al. 2019). See Green & Ting (2020) and Ting & Weinberg (2022) for additional details, particularly Figure 3 in Ting & Weinberg (2022), which illustrates the efficacy of these transformations using complex analytical PDFs.

We elaborate on one feature of the above approach. Due to the coupling between neural network layers, the normalizing flow does not perform well with a small number of odd parameters (three in this work). As a result, we introduce a dummy variable ϵ drawn from a standard normal distribution and train our normalizing flow to infer $P_\theta(\log M^*, \log \mathrm{SFR}, z, \epsilon)$. We subsequently marginalize over ϵ when calculating the final PDF $P_\theta(\log M^*, \log \mathrm{SFR}, z)$.

We also make a modification to the above approach. This modification takes into account that the quantities inferred from the observations—$\log M^*$, $\log \mathrm{SFR}$, and z—are actually noisy and probabilistic estimates of their true underlying values $\log M^*_{\mathrm{true}}$, $\log \mathrm{SFR}_{\mathrm{true}}$, and z_{true}. Computing the log-probability for each observation for our normalizing flow therefore requires marginalizing over the corresponding uncertainty $P_\theta(\boldsymbol{x}_i | \boldsymbol{x}_{i,\mathrm{true}})$:

$$P_\theta(\boldsymbol{x}_i) = \int P_\theta(\boldsymbol{x}_{i,\mathrm{true}}) P(\boldsymbol{x}_i | \boldsymbol{x}_{i,\mathrm{true}}) \, d\boldsymbol{x}_{i,\mathrm{true}}. \qquad (6)$$
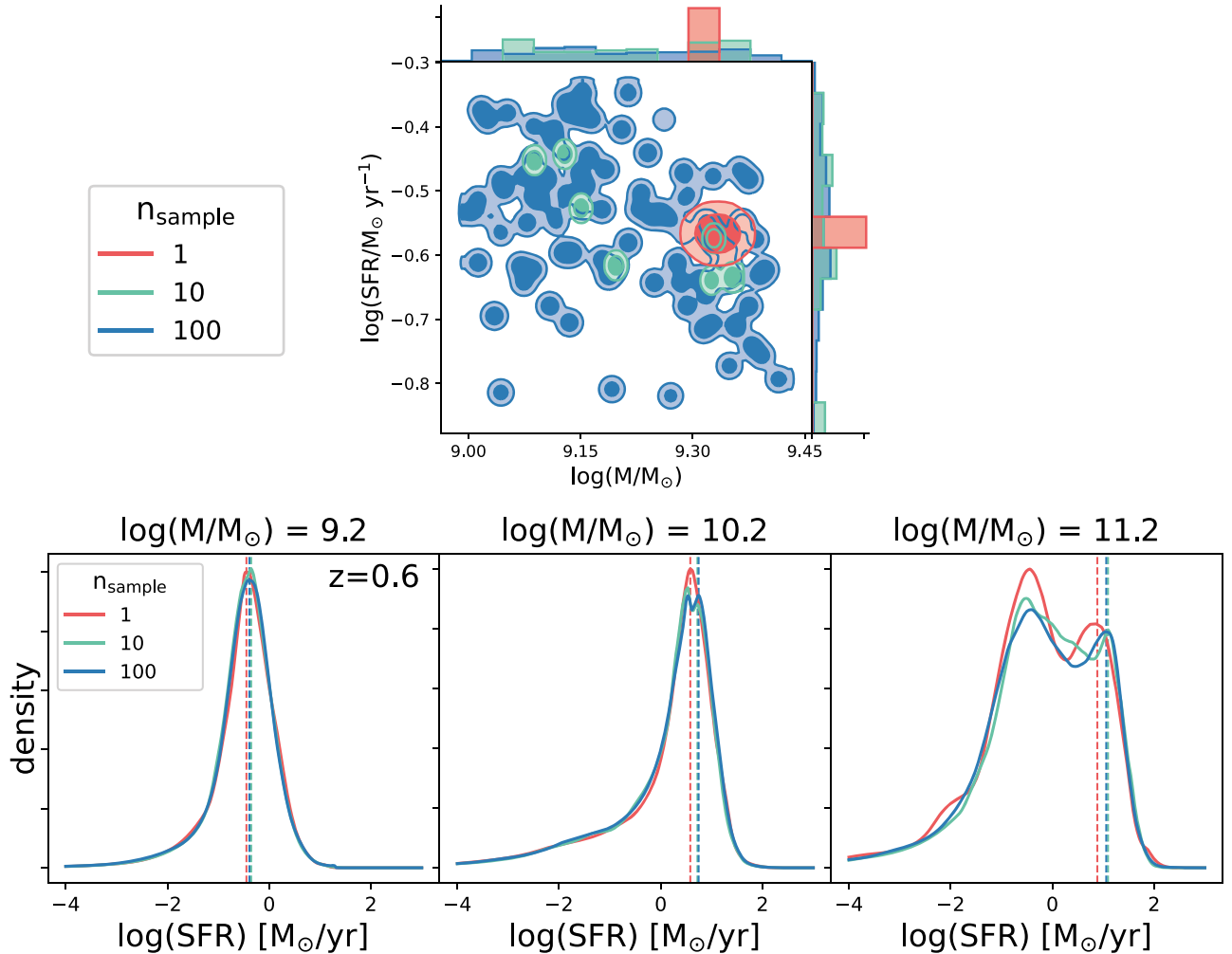
Figure 2. The effect of incorporating uncertainties in $\log M^*$ and $\log$SFR on the inferred density, $P(\log \text{SFR}|\log M^*, z)$. The top panel shows the joint $\log M^*$–$\log$SFR posterior for a single object, demonstrating that increasing the number of posterior samples for one object produces a better resolution in both the 1D uncertainties for mass and star formation rate, and of the correlation between them. The lower panels show the density profile across the star-forming sequence at fixed M^* and redshift. The colors in each panel denote how many posterior samples were included from each galaxy when inferring the density in ($\log M^*$, $\log$SFR, z) with the normalizing flow method. The ridge of the star-forming sequence is marked with a dashed line. Including uncertainties has a relatively strong effect on the profile of the sequence at high masses, shifting the median by ~0.2 dex. This sensitivity arises because the uncertainty on the average star formation rate is relatively high due to the larger number of low-sSFR systems, and the number of objects is relatively low. In contrast, measurement uncertainty has a relatively negligible effect on lower mass systems, likely due to the fact that they are far more numerous due to the shape of the galaxy stellar mass function.

Following the strategy in Leja et al. (2020), we approximate this integral using a weighted average based on $M = 100$ samples from the posterior distribution for each object:

$$\int P_\theta(\boldsymbol{x}_{i,\text{true}}) P(\boldsymbol{x}_i|\boldsymbol{x}_{i,\text{true}}) \, d\boldsymbol{x}_{i,\text{true}} \approx \frac{\sum_{j=1}^{M} w_{ij} P_\theta(\boldsymbol{x}_{ij,\text{true}})}{\sum_{j=1}^{M} w_{ij}}, \quad (7)$$

where $w_{i,j}$ are importance weights that mitigate the impact of the prior on $\log M^*$, $\log$SFR, and z. This effectively results in an uncertainty-deconvolved measurement of the SFR–M^* plane at a fixed redshift. We note that this only marginalizes over the uncertainties in free parameters in the Prospector-α physical model. It does not include more fundamental uncertainties (e.g., the initial mass function (IMF)).

Figure 2 shows the effect of explicitly accounting for the uncertainties this way, highlighting both the shape of the posterior in mass-SFR space for an individual galaxy and the vertical profile in the SFR–M plane at several slices. The error

deconvolution has the greatest effect where the uncertainties are highest, namely at low specific SFRs. Broadly speaking, the width of the star-forming sequence is not very sensitive to the error resolution except for this high-mass regime.

The subsequent analysis uses a fine sampling of the density field from the trained normalizing flow and derives the properties of the star-forming sequence from this sampled density field. We also smooth with a Gaussian kernel with $\delta z = 0.1$ in redshift to account for the effects of sampling variance (cosmic variance) caused by the finite areal coverage of the surveys. This is an imperfect solution; ideally, in future work it will be possible to train a normalizing flow that marginalizes over density fluctuations using an explicit model for cosmic sampling variance (e.g., Leja et al. 2020).

For convenience, in the rest of this work we will drop the θ subscript and instead refer to the actual 3D density estimated by the flow

$$\rho(\log M^*, \log \text{SFR}, z) = N \times P_\theta(\log M^*, \log \text{SFR}, z), \quad (8)$$
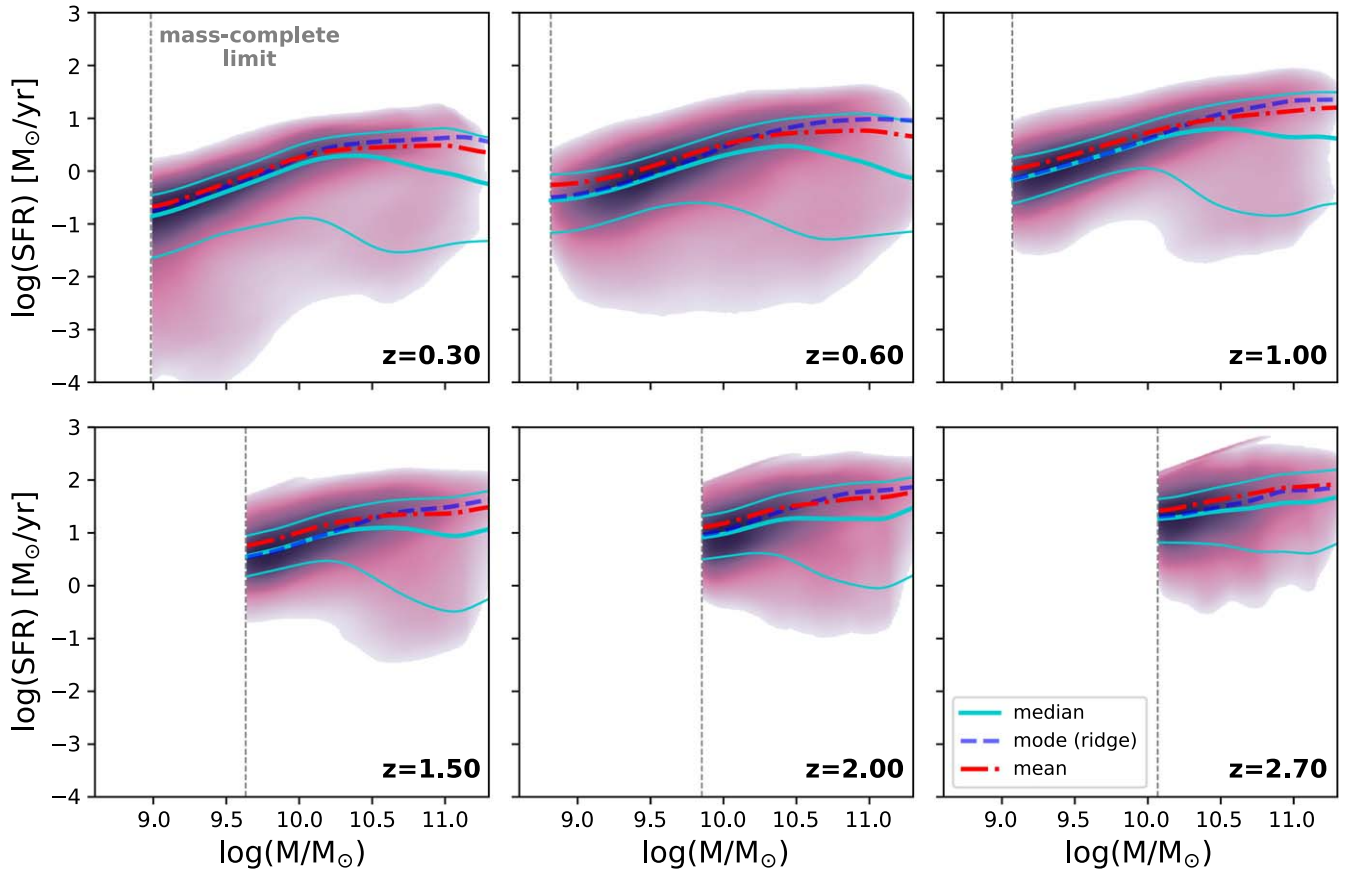
Figure 3. The log M^*–logSFR relationship at six redshifts. The color shading indicates the log-density log ρ(log M^*, logSFR, z) estimated from our normalizing flow. The thick teal line shows the running median, with the thin teal lines showing the 16th/84th percentiles. The dashed blue line shows the mode (i.e., the ridgeline), while the dashed–dotted red line shows the mean SFR (as opposed to the mean logSFR). Below $z = 0.5$, the mass-complete limit is from COSMOS-2015, while above $z = 0.5$ it is from the deeper 3D-HST survey.

where N is the number of objects (galaxies) in our data set.

5. Results from the Analysis of the logM*–logSFR-z Density Field

Here, we explore the density field ρ(log M^*, logSFR, z) inferred by our normalizing flow. We calculate and parameterize the ridge and mean of the star-forming sequence in Section 5.1 and fit the profile of the star-forming sequence to derive the width and quiescent fraction in Section 5.2.

5.1. Star-forming Sequence

The density ρ(log M^*, logSFR, z) at several redshifts is shown in Figure 3. The star-forming sequence is clearly visible as a peak in the density field. As stellar mass increases, the peak of the density increases and the distribution in logSFR grows wider. The overall increase in star formation activity with increasing redshift can be seen clearly by examining the galaxy density. Additionally, the growth of the quiescent sequence can be seen, first appearing at high stellar masses then at a later time extending to lower stellar masses as well.

We measure the evolution of the ridge (i.e., the number density peak) and the mean SFR as a function of mass and redshift. No star-forming galaxy selection is performed, i.e., this calculation is applied to the full density field. The ridge is measured as the peak of the conditional density field ρ(logSFR | log M^*, z). In practice, the ridge is measured by starting at the stellar mass-complete limit and stepping upward

in increments of ≈ 0.01 dex, requiring |d(logSFR)/d(log M^*)| < 3 at each step to ensure local continuity. This local continuity requirement prevents the ridge measurement from being affected by noise in the density field and also helps to avoid erroneously tracing the quiescent sequence where it dominates the density field. Starting at the low-mass end where the star-forming sequence is clearly defined, rather than the massive end where the density peak is lower and the shot noise is relatively high, helps to ensure the robustness of the measurement. While in principle there's no a priori reason that the density will peak in the star-forming galaxy regime, particularly if quiescent galaxies dominate at the given mass and redshift, as we will demonstrate shortly the local continuity requirement ensures that this measurement is robust a posteriori.

The goal of the nonparametric analysis is to avoid the bias created by the assumption of a parametric form; however, it is also useful to create a compact form of the results to compare with the literature. Accordingly, the evolution of these relationships is parameterized using a double power law in stellar mass (Whitaker et al. 2014; Leja et al. 2015):

$$\log(\text{SFR}) = \begin{cases} a(M - M_t) + c & M > M_t \\ b(M - M_t) + c & M \leqslant M_t \end{cases} \tag{9}$$

where $M \equiv \log M^*/M_\odot$, a is the slope at high masses, b is the slope at low masses, c is the y-axis intercept, and M_t is the
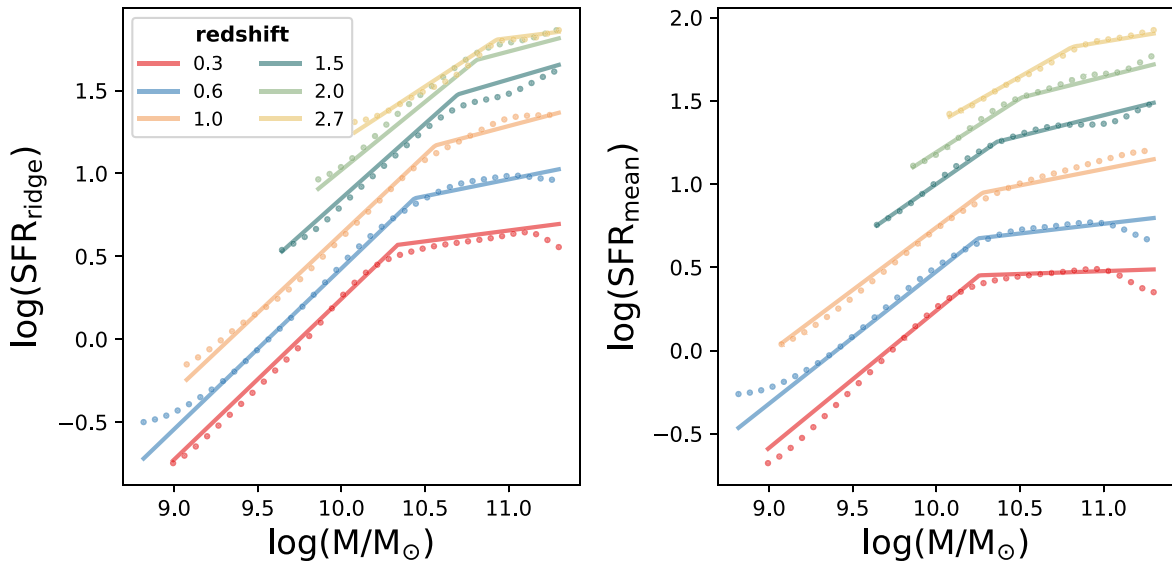
Figure 4. The left panel shows the evolution of the ridge of the star-forming sequence while the right panel shows the evolution of the mean SFR. The mean SFR includes contributions from both star-forming and quiescent galaxies. The points are measurements from the density field evaluated using the normalizing flow. The lines show the broken power-law fits described in Equations (9) and (10).

Table 1
Coefficients to the Broken Power-law Parameterization (Equation (9)) of the Ridge and Mean of the Star-forming Sequence

Parameter	X_0	X_1	X_2
Ridge (mode)			
a	0.03746 ± 0.01739	0.3448 ± 0.0297	-0.1156 ± 0.0107
b	0.9605 ± 0.0100	0.04990 ± 0.01518	-0.05984 ± 0.00482
c	0.2516 ± 0.0082	1.118 ± 0.013	-0.2006 ± 0.0039
$\log M_t$	10.22 ± 0.01	0.3826 ± 0.0188	-0.04491 ± 0.00613
Mean			
a	-0.06707 ± 0.00821	0.3684 ± 0.0128	-0.1047 ± 0.0044
b	0.8552 ± 0.0068	-0.1010 ± 0.0107	-0.001816 ± 0.003290
c	0.2148 ± 0.0045	0.8137 ± 0.0069	-0.08052 ± 0.00219
$\log M_t$	10.29 ± 0.01	-0.1284 ± 0.0135	0.1203 ± 0.0044

Note. The uncertainties are calculated from Poisson uncertainties in the underlying density field.

logarithm of the stellar mass where the slope transitions from the low-mass to the high-mass component.

The redshift evolution of each of these parameters is in turn parameterized by a quadratic in redshift. Linear and third-order polynomial fits were both explored, but the linear fit was found to have very poor predictive power and the third-order fit was found to be overly flexible without significantly improving the agreement with the density field. The equations describing the coefficients over $0.2 < z < 3.0$ follow the form

$$y_i = X_{i,0} + X_{i,1}z + X_{i,2}z^2, \tag{10}$$

where y_i represents one of the four parameters in Equation (9). The equations describing the evolution in $\log(\mathrm{SFR_{mean}})$ are defined in an analogous fashion. The mean and ridge are smoothed before fitting with Gaussian kernels with width $\delta z = 0.1$, $\delta \log M^* = 0.1$ to remove small inhomogeneities in the density field. The coefficients are shown in Table 1. The uncertainties are from counting (Poisson) uncertainties, i.e., underlying uncertainties in the density from the neural network. While the neural network produces only the maximum

likelihood model as the output, it is conditioned using the uncertainties on stellar mass and SFR from SED fitting, so these are also taken into account in this step. We note that the largest uncertainties are likely systematic errors endemic to SED fitting. These have been estimated in a number of cross-code and simulation-based studies to be approximately 0.1 dex for stellar mass and 0.3 dex in SFR (Wuyts et al. 2011a, 2011b; Pforr et al. 2012; Mobasher et al. 2015; Leja et al. 2019a; Carnall et al. 2019; Lower et al. 2020).

The redshift evolution of the mean and ridge are shown in Figure 4. Both the parametric fit and the direct measurement from the density field are shown. The mean and the ridge increase steadily with increasing redshift up to $z \sim 2$; above this, the mean continues to increase—perhaps driven by continuing decrease in the fraction of passive galaxies—whereas the evolution of the ridge begins to saturate. This is consistent with the observed flattening of sSFR evolution above $z \gtrsim 2.5$ (e.g., Santini et al. 2017; Leslie et al. 2020). Despite the fact that passive galaxies decrease the normalization of the mean SFR, the mean is still generally higher than
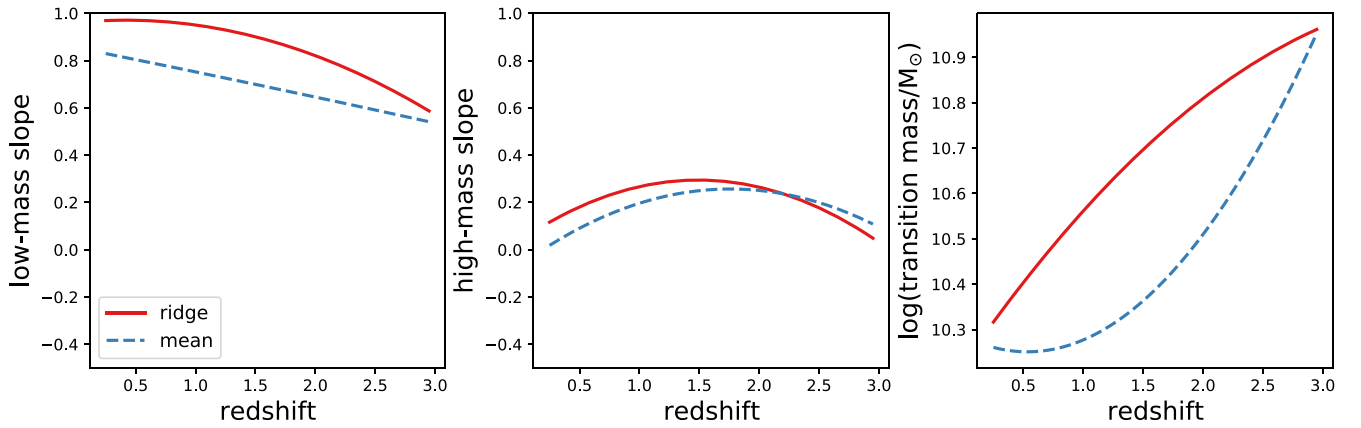
Figure 5. The redshift evolution of the low- and high-mass slope and the stellar mass at which the slope transitions between the two. The high- and low-mass slopes are distinct at all redshifts. However, the need for multiple slopes in the ridge of the star-forming sequence is most clear at low redshifts. This is because at higher redshifts the transition mass increases to higher values of stellar mass, at which there are few ($N \sim 5$–10, see, Figure 1 in Leja et al. 2020) observable galaxies in these fields, meaning both the mean and ridgeline SFRs are more weakly constrained.

the ridge (i.e., the peak of the density field), which is a standard feature in log-normal distributions such as the star-forming sequence.

Figure 5 shows the redshift evolution of the low-mass slope, high-mass slope, and transition mass. The low-mass slope of the ridge of the star-forming sequence stays near unity across most of the observed redshift range. Conversely, the high-mass slope stays relatively constant between 0 and 0.2, in contrast to high-mass slopes of 0.5–0.7 found at $1 < z < 2.5$ in previous work but in agreement at $0.5 < z < 1$ (Whitaker et al. 2014). The transition mass increases with redshift, from $\log M_t \sim 10.2$ at $z = 0.2$ to $\log M_t = 11$ at $z = 3$, at which point the high-mass slope is poorly constrained and even the existence of a two-slope star-forming sequence is not clear, due to the small number of objects at high masses and redshifts. The quadratic fit may underestimate the extent to which the high-mass slope is flat or even negative at high masses and low redshifts, as the density field in Figure 3 illustrates. This cannot be due to inaccuracies in the separation of star-forming/quiescent galaxies, since none is performed; instead, it suggests that there is little or no relationship between SFR and mass for galaxies with $\log(M/M_\odot) \gtrsim 10.5$ between $0.2 < z \lesssim 1.5$.

The low-mass slope of the mean behaves similarly to the ridge, though the slope is generally shallower at all redshifts. This is likely because the increase of quiescent galaxies with increasing stellar mass will naturally flatten out the average relationship between SFR and mass. The high-mass slope shows similar behavior. The transition mass for the mean is lower than for the ridge by $\sim$0.1–0.5 dex, likely reflecting the effect of an increasing quiescent fraction flattening out the slope at intermediate masses.

At the lowest masses there are small upturns in the SFR–M^* relationship at some redshifts. This likely represents imperfections in the low-mass 90% completeness limit, specifically a combination of up-scattering of bright galaxies with high sSFRs below the limit and an incomplete census of low-sSFR galaxies above the limit. The broken power-law fits are relatively robust to these effects and so no adjustment is made for them.

Finally, we note that there are deviations at the $\lesssim$0.1 dex level between the broken power-law parameterizations and the true evolution of the mean and ridgeline of the galaxy SFR distribution. Such deviations are inevitable when fitting smooth

functions to data with complex behaviors. For work with higher accuracy requirements, it is recommended to work directly on the trained density field, which is available online.

5.2. Profile of the Star-forming Sequence

A Gaussian distribution is fit to the density, $P(\log \mathrm{SFR}$—$\log M^*, z)$, to determine the width of the star-forming sequence and the fraction of galaxies on this sequence. This is intended to produce a qualitative description of the evolving properties of the star-forming sequence population; detailed quantitative comparisons with other results in the literature are best performed using the ridge as inferred in the previous section, or using the density field directly.

Specifically, a Gaussian is fit to the distribution of $\rho(\log \mathrm{SFR} \mid \log M^*, z)$, with the center of the Gaussian fixed to the ridgeline derived in Section 5.1. A key challenge is that the distribution of the star-forming sequence looks increasingly less like a Gaussian distribution as the fraction of quiescent galaxies increases. This is addressed by performing an asymmetric sigma-clipping fit, with the clipping occurring *only* below the star-forming sequence.

The point where the sigma-clipping process begins is chosen carefully to both accurately describe the density profile above the star-forming sequence while also excluding the long tail of quiescent galaxies below it. For each small slice in mass and redshift, an iterative sigma-clipping fit is performed to the density in $\log \mathrm{SFR}$ with $\sigma_{\mathrm{clip}} = -3.0$. Here, the negative sign indicates that clipping is performed to remove galaxies *below* the star-forming sequence. To ensure a high quality of fit, the fraction of galaxies existing *above* the ridge of the star-forming sequence is estimated from the fitted Gaussian. This is then compared to this same quantity estimated directly using the density from the trained normalizing flow. If these two estimates differ by more than 10%, σ_{clip} is adjusted upwards by 0.5 in order to exclude more quiescent galaxies and the fit is performed again. We allow σ_{clip} to vary between $-3.0 \leqslant \sigma_{\mathrm{clip}} \leqslant -1.0$. Examples of the fit to $\rho(\log \mathrm{SFR} \mid \log M^*, z)$ are shown in Figure 6.

The redshift evolution of all parameters inferred from the density field is shown in Figure 7. While the evolution of the ridge, mean, and slopes are parameterized explicitly in Equation (9), here we show the raw values as measured from the data. The width of the star-forming sequence comes from
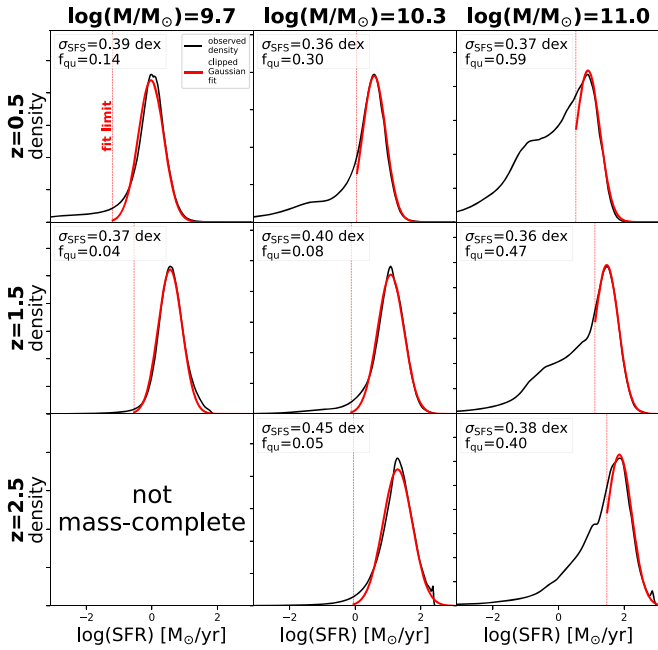
Figure 6. Slices of the log SFR distribution at fixed values of logM^* and redshift from the density estimate are in black, while Gaussian fits to the star-forming sequence shown in red. The fit is performed iteratively using sigma-clipping with a variable lower bound as described in the text. The fit limit in each panel is indicated with a dashed line. The stellar mass and redshift of each panel are indicated in the upper left, along with the inferred width of the star-forming sequence. There is a fundamental trade-off between the fidelity of the Gaussian fit to the high-SFR end of the distribution and the exclusion of the long tail of quenching or quiescent galaxies below the star-forming sequence.

the fitted Gaussian, while the quiescent fraction is defined as (1—fractional area covered by the fitted Gaussian). The parameter maps are smoothed with Gaussian kernels with width $\delta z = 0.1$, $\delta \log M^* = 0.1$ for presentation purposes.

The maps of the slope and normalization (both mean and mode) of the star-forming sequence show evolution consistent with the parameterized fits in Figure 5. The quiescent fraction shows nearly monotonic increases with decreasing redshift and with increasing stellar mass, as expected. The width of the star-forming sequence from the log-normal fit is between 0.3 and 0.4 dex at most masses and redshifts, increasing to ∼0.5 dex at higher masses where the distinction between star-forming and quiescent galaxies becomes blurred. Intriguingly, the width of the star-forming sequence also increases at high redshift, approaching $\sigma \sim 0.5$ dex at $z > 2$ for intermediate-mass galaxies.

Notably, the width of the star-forming sequence and the quiescent fraction are correlated with one another in the Gaussian profile and thus partially degenerate. Without strong preselection of star-forming galaxies and in the presence of a significant quiescent population, the width of the star-forming sequence is generally not very well constrained (e.g., Figure 1). The simple sigma-clipping procedure which produces the quiescent fractions and the width of the star-forming sequence in Figure 7 is only precise to within ∼0.1 dex (in width) or ∼20% (in quiescent fraction), and we recommend using the density from the trained flow directly for operations requiring higher precision than this. Given these caveats, the primary trend that we observe in the width of the star-forming sequence is that it is approximately constant between 0.3 and 0.4 dex at $z < 2$ and then becomes 0.4–0.5 dex at $2 < z < 3$. Indeed,

Figure 1 cleanly illustrates how the width of the star-forming sequence, $\sigma_{m\mathrm{SFS}}$, follows from the definition of a star-forming galaxy.

In contrast, the width of the full sequence (i.e., σ_{all} = half of the 84th–16th percentiles of the distribution) is independent of any distinction between star-forming and quiescent galaxies. This makes it a relatively robust property of the distribution, albeit at the *cost* of now being sensitive to the SFRs of non-star-forming galaxies. At low masses and/or high redshifts where quiescent galaxies are rare, it effectively collapses to σ_{SFS}.

6. Comparison to Previous Results and Techniques

This section compares the results from this work with previous results in the literature. Section 6.1 compares the star-forming sequence derived in this work to previous observational results, showing particularly that the normalization is considerably lower than previous work. Section 6.2 explores the origin of this difference and introduces a general framework to interpret the difference between SED SFRs and standard UV+IR SFRs.

6.1. Comparison with Other Star-forming Sequences from the Literature

Figure 8 compares the star-forming sequence derived in this work to results in the literature (Speagle et al. 2014; Whitaker et al. 2014; Lee et al. 2015; Schreiber et al. 2015; Tomczak et al. 2016; Santini et al. 2017; Boogaard et al. 2018; Iyer et al. 2018; Pearson et al. 2018; Leslie et al. 2020; Thorne et al. 2021). The literature results are corrected to a Chabrier IMF to match this work, and all studies are mass complete in this regime. The Santini et al. (2017) results are clipped at high masses where their assumption of linearity in the SFR–M relationship breaks down.

The star-forming sequence in this work is systematically lower than the measurements in the literature by ∼0.2–0.5 dex, with the offset decreasing at lower redshifts. This is not due to differences in definitions of the star-forming sequence, e.g., via preselection of star-forming galaxies. Such differences mostly affect the massive end (Figure 1), where the separation between star-forming and quiescent systems is the most confused (Figure 6). Instead, this offset is seen across the full range of stellar masses. This suggests the offset is instead due to differences in the inferred galaxy masses and SFRs. These differences are explained in detail in the next section.

There are two notable exceptions to this systematic offset in the literature, specifically the studies of Boogaard et al. (2018) and Iyer et al. (2018). What these two share in common is that they do not attempt to measure SFRs directly from broadband photometry: Boogaard et al. (2018) use Hα and Hβ emission lines from the MUSE Hubble Ultra Deep Field, permitting an internal dust correction, and Iyer et al. (2018) infer the star-forming sequence slope and normalization indirectly, by constraining them with star formation histories derived from galaxies at lower redshifts. It is a positive sign that these relatively independent methodologies agree with the Pro-spector-α modeling of the broadband photometry.

Of the studies based on broadband photometry, the Pearson et al. (2018) and Thorne et al. (2021) results are most similar to the Prospector-α inferences. This is likely because they are also measuring both SFRs and stellar masses at the same time
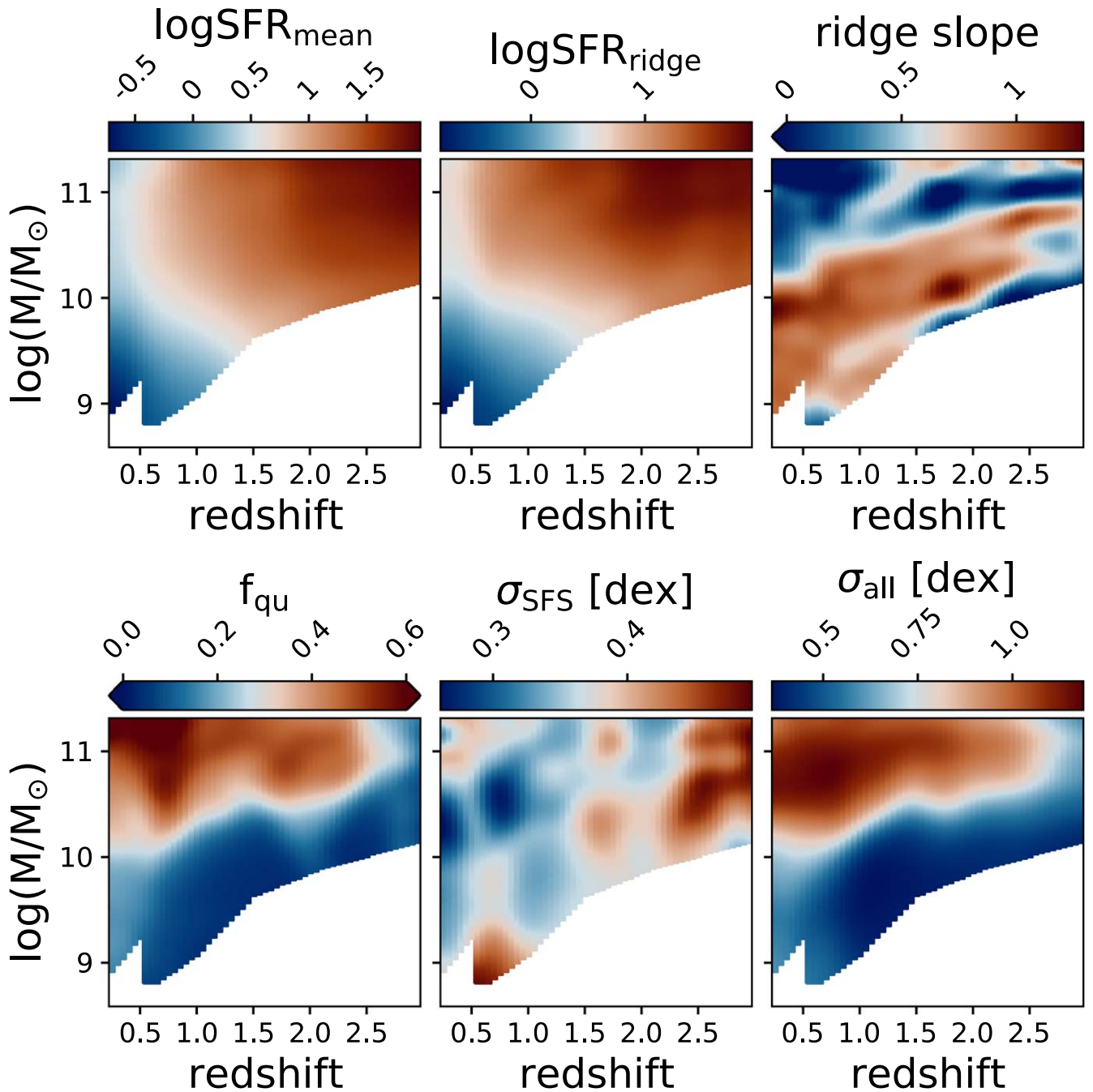
Figure 7. The evolution of parameters derived from $\rho(\log SFR \mid \log M^*, z)$. In order of left to right, the first row shows the mean SFR, the ridge of the star-forming sequence, and the slope of the ridge, while the second row shows the quiescent fraction, the 1σ width of the star-forming sequence in logSFR, and the 1σ width in logSFR of the full population. The slope of the ridge is clipped at 0 to avoid having the color scale dominated by the down-turn at high masses and low redshifts. The parameter maps are smoothed with Gaussian kernels with width $\delta z = 0.1$, $\delta \log M^* = 0.1$ for presentation purposes.

using SED fitting. Solving for the stellar populations on a galaxy-by-galaxy basis naturally lowers the SFR estimates as compared to standard UV+IR formulae, as explained in the next section. Additionally, Thorne et al. (2021) also find more massive galaxies than in previous work; they similarly have chosen a star formation history model, which also returns older ages and thus higher stellar masses (Robotham et al. 2020). Yet the `Prospector`-α results still typically lie $\sim$0.1–0.3 dex below these inferences, with the offset increasing at low masses.

6.2. Origin of the Differences in the Star-forming Sequence

Here, we explore the shift in inferred stellar masses and, specifically, SFRs. We adopt the results from previous analysis of the 3D-HST survey as a reference point: stellar masses inferred using the FAST SED-fitting code (Kriek et al. 2009) from Skelton et al. (2014), and UV+IR SFRs from Whitaker et al. (2014). Figure 9 shows the median shift in inferred values of $\log M^*$ and logSFR between the `Prospector`-α values and the aforementioned works. The ridge of the star-forming sequence is included for reference.
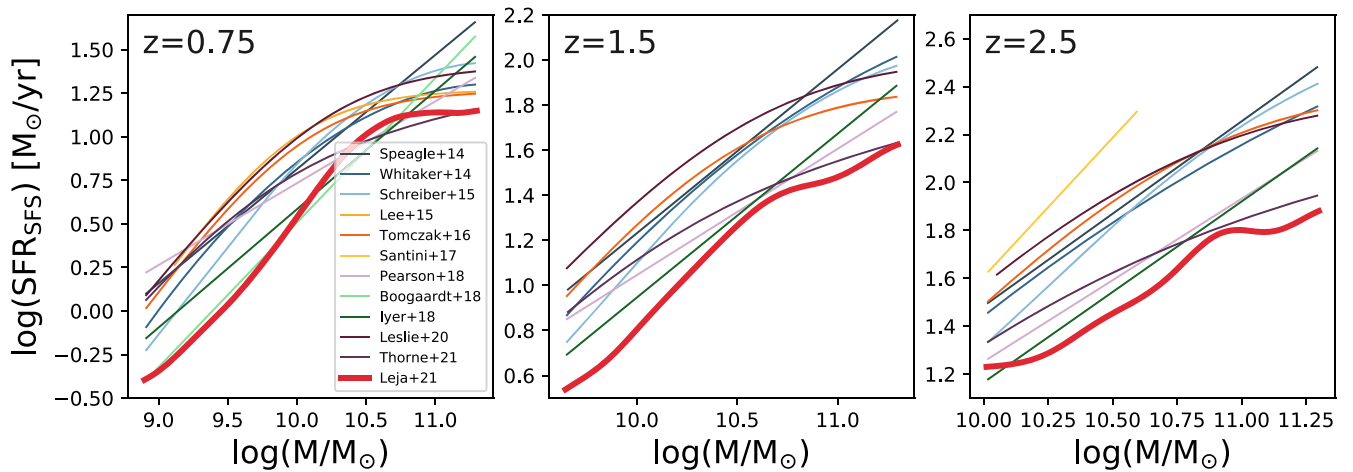
Figure 8. Comparing the star-forming sequence derived in this work with other results from the literature. In general, the star-forming sequence inferred here has a normalization well below previous results. This is due to the methodology by which the stellar masses and SFRs are inferred—see Section 6.2 for further discussion.

The systematic increase in stellar masses is at this point well understood to come from the use of nonparametric star formation histories, which find significantly older stellar ages (and thus higher stellar masses) than parametric methods. In brief, this is because they assume a larger fraction of old stars for a given observed SED; see Carnall et al. (2019) and Leja et al. (2019a) for more details.

There exists some evidence that these larger stellar masses are more accurate. Simple recovery tests using a wide range of input star formation histories show that common parametric star formation histories systematically underestimate stellar ages (Leja et al. 2019a; Carnall et al. 2019). Leja et al. (2019c) showed that the more extended star formation histories inferred from nonparametric techniques are much more consistent with the observed evolution of the stellar mass function over $0.5 < z < 3$. Lower et al. (2020) test these techniques on hydrodynamical simulations; they find that nonparametric techniques lead to significantly improved stellar masses in SED fitting, with an average bias of 0.4 dex with a delayed-τ model compared to a bias of less than 0.1 dex for a nonparametric model.

The difference in inferred SFRs has a more complex origin. Leja et al. (2019c) showed that the largest contribution to this shift is the effect of dust heating from *old* (>100 Myr) stars, which is incorrectly interpreted as star formation in standard UV+IR estimates. This effect is strongest for galaxies with low sSFR, and can be seen in Figure 9 as a decrease in inferred SFRs for massive objects, with a magnitude increasing with decreasing SFR. However, this interpretation alone is incomplete: for example, it does not explain the increase in SFR for low-mass galaxies above the star-forming sequence. To address this more clearly, we generalize the UV+IR SFR equation to explain the physical origins of the difference between the `Prospector`-α SFRs and UV+IR SFRs.

6.2.1. Difference between UV+IR and SED SFRs

The simple assumption underlying UV+IR SFRs is that the SFR is directly proportional to the luminosity of young stars, which follows from energy conservation:

$$\text{SFR} = \alpha L_{\text{bolometric,young stars}},$$

where α is a constant parameterizing the total energy output of young stars. This direct equivalence is why UV+IR SFRs are generally considered to be a reliable SFR estimator (e.g., Kennicutt 1998; Kennicutt & Evans 2012). One challenge in estimating $L_{\text{bolometric,young stars}}$ is that young stars are often enshrouded in dust that attenuates much of their light and reemits it in the infrared. The above equation then becomes

$$\text{SFR} = \alpha(L_{\text{unattenuated}} + L_{\text{attenuated}}).$$

Typically, these two components are estimated from observations by defining specific bandpasses in the UV and the IR and inferring the total energy emitted in these bands. This luminosity is then converted into an SFR by comparing to simple models of galaxy stellar populations. We use the templates from Bell et al. (2005) as an example. This work assumes $L_{\text{unattenuated}} \propto L_{\text{UV}}$ (1216–3000 Å) and $L_{\text{attenuated}} = L_{\text{IR}}(8–1000\,\mu\text{m})$. They postulate a relation of the form

$$\text{SFR } [M_\odot/\text{yr}] = \alpha(\delta L_{\text{IR}} + \gamma L_{\text{UV}}). \tag{11}$$

Bell et al. (2005) compute the coefficients using a theoretical stellar population with a constant SFR over 100 Myr, finding coefficients (adjusted to a Chabrier 2003 IMF) $\alpha_0 = 1.09 \times 10^{-10}$ [$M_\odot$ yr^{-1} $L_\odot^{-1}$], $\delta_0 = 1$, and $\gamma_0 = 2.2$. These values correspond to the parameters in the above equation, but are given subscripts in order to differentiate them from `Prospector`-inferred values for α, δ, and γ introduced later in this section.

While the general form of Equation (11) follows directly from energy conservation, the *coefficients* must be derived using stellar evolutionary theory and simplified models of the stellar populations in galaxies and can thus be subject to significant uncertainty. Standard UV+IR SFRs will assume simple model stellar population's properties to derive these coefficients (e.g., Bell et al. 2005), or infer them via empirical methods (e.g., Hao et al. 2011). Conversely, SFRs from SED models such as `Prospector`-α will (implicitly) infer these coefficients on an object-by-object basis using models for the stellar populations in each galaxy. Notably, Equation (11) itself holds true independent of the standard assumption of energy
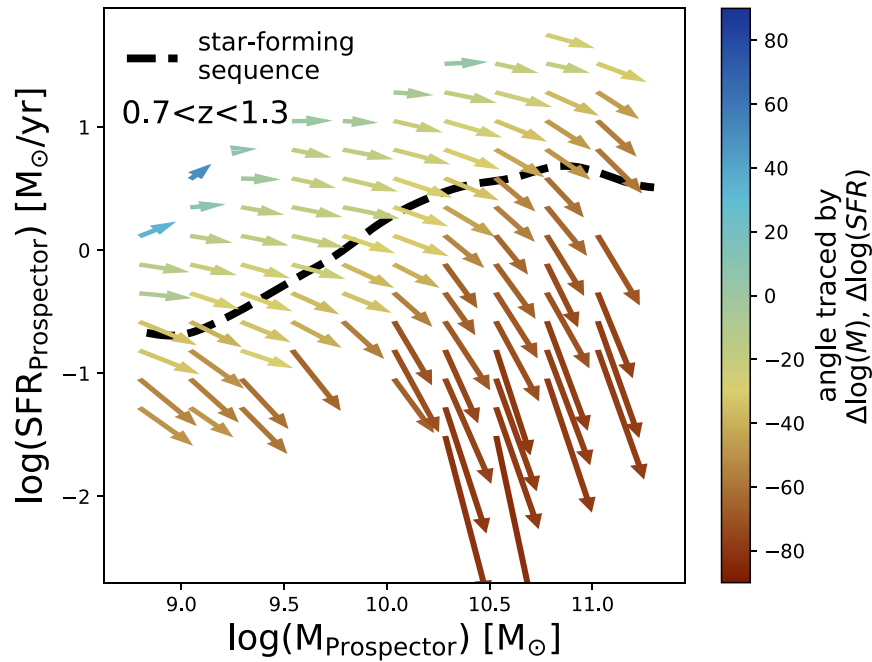
Figure 9. The median shift in inferred values of logM^* and logSFR between the `Prospector`-α model and FAST stellar masses/UV+IR SFRs for galaxies at $0.7 < z < 1.3$. Only cells with $N > 20$ galaxies are shown. The masses increase by ∼0.2 dex due to the use of nonparametric star formation histories. The inferred SFRs decrease by 0.1–1+ dex largely due to dust heating from old stars, though other effects can contribute substantially in certain regimes (see Section 6.2 for an in-depth discussion). The ridgeline of the star-forming sequence as measured in this work is shown as a black dashed line. The color bar traces the angle of the shift in log$M$–logSFR space, where $\theta = 0$ is a horizontal arrow pointing to the right.

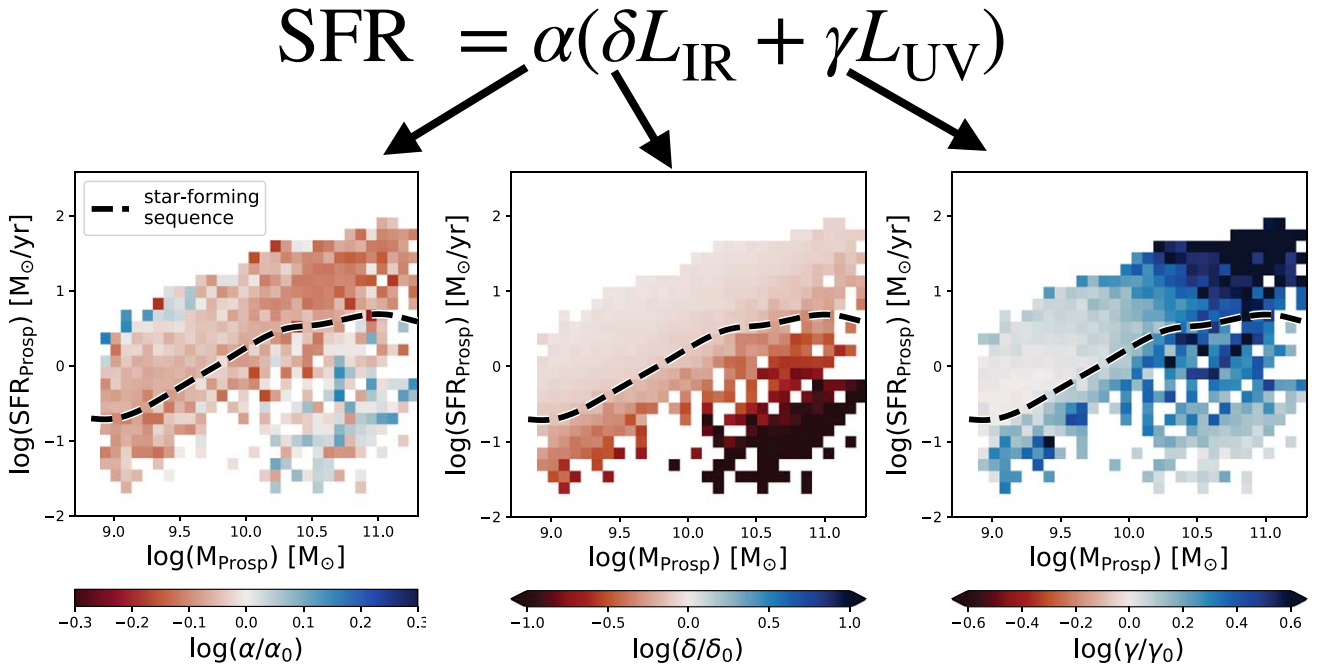$$\text{SFR} = \alpha(\delta L_{\text{IR}} + \gamma L_{\text{UV}})$$

Figure 10. The median difference in the coefficients to the UV+IR SFR in Equation (11) between `Prospector`-α model and the fixed template-based coefficients from Bell et al. (2005) is shown for galaxies at $0.7 < z < 1.3$. α parameterizes the total energy output of young stars, δ parameterizes the fraction of IR luminosity which is powered by energy from young stars, and γ parameterizes the conversion from the observed UV luminosity to the bolometric luminosity of the young stars *after* dust attenuation. Arrows on the color bar indicate where the data is clipped to enhance the information content of the map. The y-axis shows the SFR averaged over the most recent 100 Myr.

balance (i.e., $L_{\text{attenuated}} = L_{\text{emitted,IR}}$, da Cunha et al. 2008). Rather, energy balance is used in order to derive estimates of the coefficients α, δ, and γ.

We directly calculate the UV+IR coefficients for each galaxy from the `Prospector`-α physical models by

perturbing the SFRs from their measured values and measuring the response in L_{IR} and L_{UV}. Figure 10 shows the median difference between the `Prospector`-α coefficients and those from the Bell et al. (2005) templates in the SFR–M plane. These differences are discussed in turn below.

The first coefficient, α, represents the total luminosity of young stars (*young* here defined as age <100 Myr). At these redshifts, Prospector-α finds median values in the SFR–M plane of $-0.2 \lesssim \log(\alpha_{\mathrm{Prosp}}/\alpha_0) \lesssim 0.15$, though the scatter can vary up to ± 0.4 for individual galaxies.

In the Prospector-α model, variations in α are primarily determined by the SFR(0–30 Myr)/SFR(30–100 Myr) ratio (Spearman-r coefficient of 0.81). This is a straightforward consequence of the fact that younger stars (0–30 Myr) are relatively more luminous than older stars (30–100 Myr). Thus, galaxies with rising star formation histories have lower values of α and galaxies with falling star formation histories have higher values of α. Since galaxies on and above the star-forming sequence tend to have rising star formation histories while galaxies below have falling star formation histories (Leja et al. 2019c), this naturally produces an increase in α with decreasing SFR at a fixed stellar mass. This relationship has considerable scatter though, particularly at the low-mass end, suggesting that some galaxies are being observed in the process of moving back toward their equilibrium values (i.e., around the ridgeline of the star-forming sequence).

The second coefficient, δ, represents the fraction of the emitted IR luminosity which is powered by young stars. The primary other source of energy is old stellar populations, although AGN can be significant in a minority of cases. In the Bell et al. (2005) UV+IR formulation, $\delta = 1$ because there is no other source of energy; however, it has long been known that δ *should* be <1 (e.g., Fumagalli et al. 2014; Utomo et al. 2014; Hayward & Smith 2015). In empirical calibrations, δ is inferred to be ∼0.6 for starburst galaxies (e.g., Meurer et al. 1999), or 0.46 for normal star-forming galaxies (Hao et al. 2011).

Prospector-α infers $0.01 \lesssim \delta \leqslant 1$. Variations in δ in the Prospector-α model are almost entirely driven by heating from old stars: Leja et al. (2019c) find a strong relationship between sSFR and the fraction of dust heating performed by old stars, with galaxies at $\log(\mathrm{sSFR\ yr^{-1}}) = -11$ having ∼80% of their dust emission powered by old stars. This relationship is in excellent qualitative agreement with detailed radiative transfer simulations of nearby galaxies from Dustpedia (Nersesian et al. 2020). Notably, this effect gets *stronger* at higher redshifts, where the *old* stars are relatively more luminous than at later times. This relationship is evident in Figure 10 as a decrease in δ with decreasing SFR at a fixed stellar mass, ranging from $\delta \approx 1$ for highly star-forming galaxies to $\delta \ll 1$ for quiescent galaxies.

The final coefficient, γ, represents the conversion from the observed UV luminosity at 1216–3000 Å to the bolometric luminosity of young stars. This measurement is performed *after* dust attenuation. Thus it represents the complement to the IR luminosity, ensuring that the full energy budget of young stars is included in the estimate.

Variations of γ from Prospector-α span a range of $-0.05 \lesssim \log(\gamma/\gamma_0) \leqslant 1.1$. The γ coefficient correlates with the shape of the SED of young stars after the effects of dust attenuation. The largest determinant of γ in Prospector-α is perhaps unsurprisingly $\tau_{\mathrm{diffuse\ dust}}$ (Spearman-r coefficient of 0.88). This is the normalization of the dust attenuation curve (closely related to A_V) and represents the most important determinant of the total amount of reddening in the observed SED. Stellar metallicity is an important secondary effect as decreasing stellar metallicity will make the SED of young stars

bluer (Spearman-r coefficient of 0.4). The γ map in Figure 10 generally increases with both mass and SFR, consistent with trends in the observed dust attenuation (Whitaker et al. 2017).

The sum of these effects can now explain the full transformation of the star-forming sequence in Figure 9. Galaxies on and below the star-forming sequence have lower inferred SFRs due to the dust-heating effect from old stars. Galaxies above the star-forming sequence show little effect on their star formation rates; here, the standard assumption that L_{IR} is entirely powered by star formation appears to hold. Conversely, low-mass galaxies above the star-forming sequence show an increase in their SFRs due to a combination of the effects of dust on the interpretation of their UV luminosities and bursty recent star formation histories.

An important point is that the IR energy dominates the UV+IR equation above $\log(M/M_\odot) \gtrsim 9.5$ (Whitaker et al. 2017), while UV energy dominates for galaxies with $\log(M/M_\odot) \lesssim 9.5$; this means that δ will have a larger influence at high masses, and γ a larger influence at low masses.

We emphasize that while measuring these coefficients with a model such as Prospector-α is a more robust approach than using fixed coefficients, these inferences are still based on simplified models of galaxy stellar populations and should be viewed with scrutiny. One method to investigate their accuracy is via posterior predictive checks, where one set of observations is fit by a model which is in turn used to predict a second independent set of observations. The SFRs from Prospector-α perform well here, finding a scatter of 0.2 dex and negligible offset when fitting UV-IR broadband photometry and predicting measurements of Hα (Leja et al. 2017) and Br-γ (Pasha et al. 2020) emission line luminosities. Br-γ predictions are particularly informative: Br-γ is a hydrogen emission line at ∼2.16 μm, which is nearly insensitive to dust attenuation, thus serving as a dust-independent measure of SFRs. However, these tests are complicated by the systematic ∼0.2 dex sensitivity of emission line luminosities to the chosen stellar isochrones and their treatment of rotation or binary effects (Steidel et al. 2016; Choi et al. 2017; Pasha et al. 2020). Additional posterior predictive checks using a wide range of observables across a wide range of galaxy types and redshifts will be beneficial in further calibrating SED model outputs.

7. Discussion

Here, the broader implications of the results are discussed. Section 7.1 compares the star-forming sequence in this work with those from hydrodynamical simulations of galaxy formation, and Section 7.2 discusses the prospects for robust measurements of the star-forming sequence in the absence of a clear star-forming/quiescent galaxy separation.

7.1. A New Agreement with Theoretical Models of Galaxy Formation

There has been a long-standing systematic offset of ∼0.2–0.5 dex in the normalization of the $0.5 \lesssim z \lesssim 3$ star-forming sequence between observations and predictions from galaxy formation simulations (Mitchell et al. 2014; Furlong et al. 2015; Leja et al. 2015; Tomczak et al. 2016; Davé et al. 2019; Donnari et al. 2019; Katsianis et al. 2020). This offset is such that simulations under-predict the observations. This offset is persistent across most of the major theoretical models of galaxy formation. The universality comes from the fact that this
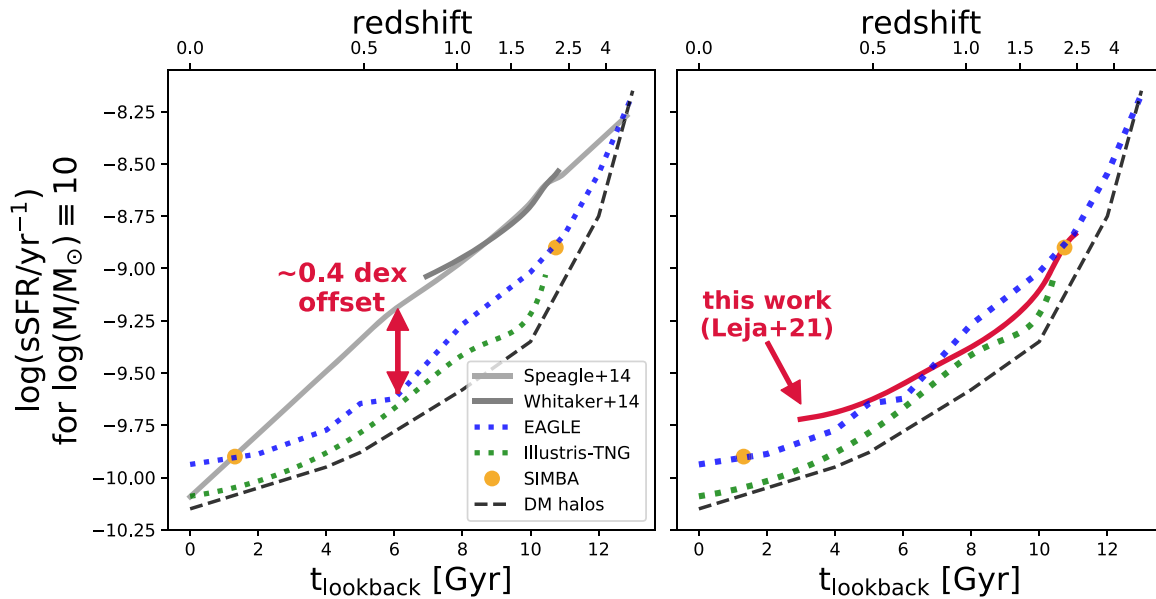
Figure 11. This figure compares the redshift evolution of sSFR for a $\log(M/M_\odot) = 10$ galaxy on the star-forming sequence between observations and galaxy formation simulations. The specific dark matter accretion rate from the Millennium simulation is also shown for reference (Fakhouri et al. 2010). The left panel shows previous observational star-forming sequences, which have a ∼0.4 dex offset with the simulations. The right panel shows results from this work, which resolve these systematic differences.

problem is challenging to solve with baryonic feedback prescriptions; galaxy star formation rates typically track the gas accretion rate, which is in turn set by the behavior of dark matter at large scales (Mitchell et al. 2014).

We show that this tension is resolved by the star-forming sequence presented in this work (Figure 11), confirming the preliminary work in Nelson et al. (2021). This figure shows the redshift evolution of the specific star formation rate for a $\log(M/M_\odot) = 10$ galaxy from the observations (Speagle et al. 2014; Whitaker et al. 2014) and from this work. Predictions for the galaxy star-forming sequence are shown from several major cosmological hydrodynamical simulations of galaxy formation: IllustrisTNG (Donnari et al. 2019), the EAGLE simulations Furlong et al. (2015), and SIMBA (Davé et al. 2019). The specific dark matter accretion rates are taken from the Millennium N-body simulations (Fakhouri et al. 2010). Whereas before there was a systematic offset of ∼0.2−0.5 dex with the observations, there is now broad agreement between the observations, the hydrodynamical simulations, and the dark matter accretion rates.

This new agreement is important to the field because it is now possible to use the observed star-forming sequence across a wide range of redshifts as a constraint on hydrodynamical simulations of galaxy formation. Hahn et al. (2019) measure the star-forming sequence across multiple simulations using a consistent nonparametric definition and show that the simulated star-forming sequences differ by up to 0.7 dex in normalization and have significantly different slopes, in spite of the fact that they all approximately match the $z = 0$ galaxy stellar mass function (Somerville & Davé 2015). Having the new capacity to adjust their physics and feedback models to match the $0.5 < z < 3$ star-forming sequence will help to bring consensus in models of galaxy formation.

Constraints on the star-forming sequence will in turn provide a strong constraint on the feedback models in simulations, which currently have substantial qualitative and quantitative differences (see the review in Somerville & Davé 2015 for a

discussion). We note that the release of tools like the trained normalizing flow used in this work means that these comparisons can now be done directly in the space of the SFR–M^*-redshift density, $\rho(\log SFR, \log M^*, z)$. This bypasses classic challenges in comparing observations and simulations, such as preselection of star-forming galaxies or definitions of the star-forming sequence.

7.2. Is It Meaningful to Enforce a Separation between Star-forming and Quiescent Galaxies?

It has long been known that galaxies have clear bimodal distributions in rest-frame color diagrams such as the red sequence (Bell et al. 2004) or the *UVJ* diagram (Williams et al. 2009). This naturally leads to the idea of two separate classes of galaxies: star-forming and quiescent. In such a framework, it is straightforward to use colors or an SFR indicator such as Hα to distinguish between the two classes. It is also natural to focus more interest on measuring the star formation rates in star-forming galaxies (i.e., the star-forming sequence), as growth by star formation in quiescent galaxies is thought to be negligible.

However, the distribution of galaxies in the SFR–M plane inferred by this analysis—importantly, with no preselection of star-forming galaxies—instead resembles a unimodal distribution with a long skew or tail toward low star formation rates (e.g., Figure 6). Feldmann (2017) observes a similar distribution, and argues that the star-forming/quiescent galaxy distinction as defined by a log-normal star-forming sequence is inconsistent with the idea of star formation as a discrete and burst-driven event. Detailed studies of the local universe such as the Herschel Reference Survey show no separation at all between star-forming and quiescent galaxies (Eales et al. 2017). In this new schema, using SFR indicators to distinguish between star-forming and quiescent galaxies is not a robust methodology due to the strong overlap between the two classes. This danger is compounded by the fact that SFR indicators at low sSFRs are (a) much noisier due to a lower intrinsic brightness, and (b) and more prone to be contaminated

with emission from processes unrelated to star formation (e.g., LINER emission, AGN, or dust heating by old stars).

It may not yet be possible to quantitatively distinguish between these two schemata, at least not using these data. This is because colors can be bimodal even if the sSFR is not, as photometric colors saturate at low sSFRs (see, e.g., Leja et al. 2020 and references therein). Put another way, the key challenge is that the observed SFRs become highly uncertain at low sSFR. Figure 2 provides a clear demonstration of this. The sensitivity of the star-forming sequence at low SFRs to the uncertainty resolution suggests that the inferred shape of the star-forming sequence at low sSFRs is not a robust measurement, but instead, largely traces the shape of the uncertainties in SFRs.

Regardless of whether there is truly a bimodal sequence that is unobservable because it is smeared out by currently-irreducible observational error, or whether there is indeed no bimodality at all, there are clear practical ramifications to this question. Figure 1 demonstrates that the normalization and width of the star-forming sequence and the fraction of quiescent galaxies both depend strongly on the separation of star-forming and quiescent galaxies at the $\sim$0.2–0.3 dex level, even when using the same star formation rate indicator. Given this, it is unsurprising that the observed width of the star-forming sequence varies between $\sim$0.25 and 0.5 dex in the literature (Speagle et al. 2014).

In light of these uncertainties, we argue that it is more meaningful to compare measurements of the density in the SFR–M plane than it is to parameterize the star-forming sequence itself using a bimodal galaxy classification system. The ridgeline of the star-forming sequence is largely invariant to bimodal classification (Renzini & Peng 2015), and the 1σ scatter of the *full* galaxy population in SFR–M is also meaningful and straightforward to measure. In this work, we additionally package the inferred density of the star-forming sequence into a trained normalizing flow and make it publicly available. In this way, galaxy formation models and future observational studies can compare to the full set of observations rather than attempting to emulate fragile observational selection criteria. This is an important step on the path toward likelihood-free inference. This approach has the additional advantage of including the full structure in the SFR–M^* plane, including such features as star-bursting galaxies above the star-forming sequence and the distribution of green valley and quiescent galaxies below the star-forming sequence.

8. Conclusion

In this paper, we infer the galaxy star-forming sequence across $0.2 < z < 3$ using a combination of machine-learning techniques and nonparametric methods applied to state-of-the-art UV-IR galaxy SED modeling from Prospector-α. Specifically, we train a normalizing flow to reproduce the distribution $\rho(\log M^*, \log\mathrm{SFR}, z)$ of galaxies in $\log M^*$, log SFR, and redshift, with alterations included to marginalize over the full $\log M^*$ and log SFR uncertainties derived from our previous Prospector-α fits.

Our conclusions are summarized here:

1. We use the Prospector-α fits to show that preselection of star-forming galaxies adds substantial systematic uncertainty to the inferred properties of the star-forming sequence. Applied to the same data, four standard but different preselection methods produce differences of $\sim$0.3 dex in the normalization, a $\sim$0.2 dex in the width, and $\sim$20% differences in the fraction of quiescent galaxies.

2. We instead present mass- and redshift-dependent fits to the ridgeline (number density peak) of the star-forming sequence and the mean SFR for all galaxies, avoiding fragile observational selection criteria for star-forming galaxies.

3. We show that the star-forming sequence has a mass-dependent slope below $z \sim 2$, with a flatter relationship at high masses and a steeper relationship at low masses with a slope near unity. Above $z \sim 2$ a single slope is able to fit the data.

4. The profile of the star-forming sequence is relatively narrow ($\sigma \sim 0.3$–0.35 dex) across a wide range of masses and redshifts. It is, however, significantly wider ($\sigma \sim 0.4$–0.5 dex) at high redshift ($z \gtrsim 2$) and high masses ($M^* > 10^{11}\ M_\odot$).

5. The star-forming sequence inferred in this work is 0.2–0.5 dex below the great majority of previous results in the literature. We show that this lower normalization is specifically caused by the nonparametric SFHs in Prospector-α which produce older ages and higher stellar masses and the overall lower inferred SFRs.

6. A general framework is developed to compare SED-based SFRs with UV+IR SFRs. This framework demonstrates that inferring star formation histories, dust attenuation, and metallicities on an object-by-object basis via SED-fitting produces more sophisticated (and, likely, more accurate) conversions between the observed flux and the SFRs. It is shown that the net effect of these changes in Prospector-α produces lower SFRs than standard UV+IR formulae, particularly at higher redshifts.

7. Our new star-forming sequence follows the specific dark matter accretion rate much more closely than previous work, removing a long-standing 0.2–0.5 dex offset at $0.5 < z < 3$ in the star-forming sequence between observations and predictions from models of galaxy formation.

8. Our trained normalizing flow describing the full density $\rho(\log M^*, \log\mathrm{SFR}, z)$ of galaxies in $\log M^*$, logSFR, and redshift is made available online.[22] This permits direct comparison between this work and future theoretical and observational works, removing the need to define a star-forming sequence or separate star-forming and quiescent galaxies.

In this work, we have used state-of-the-art tools in machine learning, statistics, and computing to solve long-standing problems in observational galaxy evolution. In the future, we look forward to employing more sophisticated models and better tools to further address extant issues in this work. Some avenues for future improvement include better accounting for the effects of cosmic variance (sampling variance), more sophisticated modeling of the selection function near the stellar mass-complete limit, and a significant reduction of the computational resources required to fit high-dimensional SED models.

Finally, the galaxy star-forming sequence is only as accurate as the models used to produce it. Due to the complex nature of

---

[22] https://github.com/jrleja/sfs_leja_trained_flow

stellar populations in galaxies, these physical models greatly benefit from cross-calibration tests to ensure their accuracy. Such tests for Prospector-α have produced promising results in spectroscopic observations of nearby galaxies (Leja et al. 2017; Pasha et al. 2020), in hydrodynamical simulations (Lower et al. 2020), and in comparisons between low-redshift star formation histories and high-redshift observations (Leja et al. 2019c). However, existing posterior predictive tests typically only cover small, well-studied samples or simulated objects (*models fitting models*). Going forward, there is a sore need for tests covering a variety of stellar population parameters across a wide range of both stellar masses and cosmic times, particularly in the early universe, to ensure robust parameter inference.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), dynesty (Speagle 2020) FSPS (Conroy et al. 2009), ipython (Pérez & Granger 2007), matplotlib (Caswell et al. 2018), numpy (van der Walt et al. 2011), Prospector (Johnson et al. 2021), python-fsps (Foreman-Mackey et al. 2014), PyTorch (Paszke et al. 2019), scipy (Virtanen et al. 2020).

ORCID iDs

Joel Leja ⬤ https://orcid.org/0000-0001-6755-1315
Joshua S. Speagle (沈佳士)
 ⬤ https://orcid.org/0000-0003-2573-9832
Yuan-Sen Ting (丁源森)
 ⬤ https://orcid.org/0000-0001-5082-9536
Benjamin D. Johnson ⬤ https://orcid.org/0000-0002-9280-7594
Charlie Conroy ⬤ https://orcid.org/0000-0002-1590-8551
Katherine E. Whitaker ⬤ https://orcid.org/0000-0001-7160-3632
Erica J. Nelson ⬤ https://orcid.org/0000-0002-7524-374X
Pieter van Dokkum ⬤ https://orcid.org/0000-0002-8282-9888
Marijn Franx ⬤ https://orcid.org/0000-0002-8871-3026

References

Abramson, L. E., Gladders, M. D., Dressler, A., et al. 2015, ApJL, 801, L12
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E., et al. 2013, A&A, 558, A33
Bell, E. F., Papovich, C., Wolf, C., et al. 2005, ApJ, 625, 23
Bell, E. F., Wolf, C., Meisenheimer, K., et al. 2004, ApJ, 608, 752
Bezanson, R., Wake, D. A., Brammer, G. B., et al. 2016, ApJ, 822, 30
Boogaard, L. A., Brinchmann, J., Bouché, N., et al. 2018, A&A, 619, A27
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019, ApJ, 873, 44
Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, MNRAS, 480, 4379
Caswell, T. A., Droettboom, M., Hunter, J., et al. 2018, matplotlib/matplotlib v3.0.0, Zenodo, doi:10.5281/zenodo.1420605
Chabrier, G. 2003, PASP, 115, 763
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
Choi, J., Conroy, C., & Byler, N. 2017, ApJ, 838, 159
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Curtis-Lake, E., Chevallard, J., Charlot, S., & Sandles, L. 2021, MNRAS, 503, 4855
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746
Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
Dekel, A., & Burkert, A. 2014, MNRAS, 438, 1870
Donnari, M., Pillepich, A., Nelson, D., et al. 2019, MNRAS, 485, 4817
Dotter, A. 2016, ApJS, 222, 8
Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. 2019, arXiv:1906.04032
Eales, S., de Vis, P., Smith, M. W. L., et al. 2017, MNRAS, 465, 3125
Fakhouri, O., Ma, C.-P., & Boylan-Kolchin, M. 2010, MNRAS, 406, 2267
Feldmann, R. 2017, MNRAS, 470, L59
Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS v0.1.1, Zenodo, doi:10.5281/zenodo.12157
Fumagalli, M., Labbé, I., Patel, S. G., et al. 2014, ApJ, 796, 35
Fumagalli, M., Patel, S. G., Franx, M., et al. 2012, ApJL, 757, L22
Furlong, M., Bower, R. G., Theuns, T., et al. 2015, MNRAS, 450, 4486
Green, G. M., & Ting, Y.-S. 2020, arXiv:2011.04673
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Hahn, C., Starkenburg, T. K., Choi, E., et al. 2019, ApJ, 872, 160
Hao, C.-N., Kennicutt, R. C., Johnson, B. D., et al. 2011, ApJ, 741, 124
Hayward, C. C., & Smith, D. J. B. 2015, MNRAS, 446, 1512
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Iyer, K., Gawiser, E., Davé, R., et al. 2018, ApJ, 866, 120
Jimenez Rezende, D., & Mohamed, S. 2015, arXiv:1505.05770
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. 2011, ApJ, 730, 61
Katsianis, A., Gonzalez, V., Barrientos, D., et al. 2020, MNRAS, 492, 5592
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kennicutt, R. C., Jr. 1998, ARA&A, 36, 189
Kingma, D. P., & Dhariwal, P. 2018, arXiv:1807.03039
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
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Lee, N., Sanders, D. B., Casey, C. M., et al. 2015, ApJ, 801, 80
Leitner, S. N. 2012, ApJ, 745, 149
Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019a, ApJ, 876, 3
Leja, J., Johnson, B. D., Conroy, C., et al. 2019c, ApJ, 877, 140
Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, ApJ, 837, 170
Leja, J., Speagle, J. S., Johnson, B. D., et al. 2020, ApJ, 893, 111
Leja, J., Tacchella, S., & Conroy, C. 2019b, ApJL, 880, L9
Leja, J., van Dokkum, P. G., Franx, M., & Whitaker, K. E. 2015, ApJ, 798, 115
Leslie, S. K., Schinnerer, E., Liu, D., et al. 2020, ApJ, 899, 58
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Lower, S., Narayanan, D., Leja, J., et al. 2020, ApJ, 904, 33
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Mitchell, P. D., Lacey, C. G., Cole, S., & Baugh, C. M. 2014, MNRAS, 444, 2637
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
Nelson, E. J., Tacchella, S., Diemer, B., et al. 2021, MNRAS, 508, 219
Nersesian, A., Verstocken, S., Viaene, S., et al. 2020, A&A, 637, A25
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJL, 660, L43
Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793
Pasha, I., Leja, J., van Dokkum, P. G., Conroy, C., & Johnson, B. D. 2020, ApJ, 898, 165
Paszke, A., Gross, S., Massa, F., et al. 2019, arXiv:1912.01703
Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, A&A, 615, A146
Pérez, F., & Granger, B. E. 2007, CSE, 9, 21
Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285
Renzini, A., & Peng, Y.-j 2015, ApJL, 801, L29

Robotham, A. S. G., Bellstedt, S., Lagos, C. d. P., et al. 2020, MNRAS, 495, 905

Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJL, 739, L40

Santini, P., Fontana, A., Castellano, M., et al. 2017, ApJ, 847, 76

Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74

Sherman, S., Jogee, S., Florez, J., et al. 2021, MNRAS, 505, 947

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

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Sparre, M., Hayward, C. C., Springel, V., et al. 2015, MNRAS, 447, 3548

Speagle, J. S. 2020, MNRAS, 493, 3132

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15

Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159

Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., et al. 2021, MNRAS, 505, 540

Ting, Y.-S., & Weinberg, D. H. 2022, ApJ, 927, 209

Tomczak, A. R., Quadri, R. F., Tran, K.-V. H., et al. 2016, ApJ, 817, 118

Utomo, D., Kriek, M., Labbé, I., Conroy, C., & Fumagalli, M. 2014, ApJL, 783, L30

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261

Whitaker, K. E., Franx, M., Leja, J., et al. 2014, ApJ, 795, 104

Whitaker, K. E., Pope, A., Cybulski, R., et al. 2017, ApJ, 850, 208

Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, ApJL, 754, L29

Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879

Wuyts, S., Förster Schreiber, N. M., Lutz, D., et al. 2011a, ApJ, 738, 106

Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011b, ApJ, 742, 96