The SFR-M Correlation Extends to Low Mass at High Redshift

THE SFR-M$_*$ CORRELATION EXTENDS TO LOW MASS AT HIGH REDSHIFT

Kartheik Iyer,[1] Eric Gawiser,[1,2] Romeel Davé,[3,4] Philip Davis,[5] Steven L. Finkelstein,[6] Dritan Kodra,[7] Anton M. Koekemoer,[8] Peter Kurczynski,[1] Jeffery A. Newman,[7] Camilla Pacifici,[8] and Rachel S. Somerville[1,2]

—

[1]*Department of Physics and Astronomy, Rutgers, The State University of New Jersey, 136 Frelinghuysen Road, Piscataway, NJ 08854-8019 USA*

[2]*Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, NY 10010, USA*

[3]*Institute for Astronomy, Royal Observatory, Edinburgh EH9 3HJ, UK*

[4]*Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town 7535, South Africa*

[5]*Rutgers Discovery Informatics Institute, Rutgers, The State University of New Jersey, 96 Frelinghuysen Road, Piscataway, NJ 08854-8019 USA*

[6]*Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA*

[7]*Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA*

[8]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

ABSTRACT

To achieve a fuller understanding of galaxy evolution, SED fitting can be used to recover quantities beyond stellar masses (M$_*$) and star formation rates (SFRs). We use Star Formation Histories (SFHs) reconstructed via the Dense Basis method of Iyer & Gawiser (2017) for a sample of $17,873$ galaxies at $0.5 < z < 6$ in the CANDELS GOODS-S field to study the nature and evolution of the SFR-M$_*$ correlation. The reconstructed SFHs represent trajectories in SFR-M$_*$ space, enabling us to study galaxies at epochs earlier than observed by propagating them backwards in time along these trajectories. We study the SFR-M$_*$ correlation at $z = 1, 2, 3, 4, 5, 6$ using both direct fits to galaxies observed at those epochs and SFR-M$_*$ trajectories of galaxies observed at lower redshifts. The SFR-M$_*$ correlations obtained using the two approaches are found to be consistent with each other through a KS test. Validation tests using SFHs from semi-analytic models and cosmological hydrodynamical simulations confirm the sensitivity of the method to changes in the slope, normalization and shape of the SFR-M$_*$ correlation. This technique allows us to further probe the low-mass regime of the correlation at high-z by ~ 1 dex and over an effective volume of $\sim 10\times$ larger than possible with just direct fits. We find that the SFR-M$_*$ correlation is consistent with being linear down to M$_* \sim 10^7 M_\odot$ at $z > 4$. The evolution of the correlation is well described by $\log SFR = (0.80 \pm 0.029 - 0.017 \pm 0.010 \times t_{univ}) \log M_* - (6.487 \pm 0.282 - 0.039 \pm 0.008 \times t_{univ})$, where t_{univ} is the age of the universe in Gyr.

Keywords: galaxies: star formation — galaxies: evolution — techniques: photometric

Corresponding author: Kartheik Iyer
kgi1@physics.rutgers.edu

1. INTRODUCTION

The SFR-M$_*$ correlation couples a galaxy's Star Formation Rate (SFR), an effectively instantaneous quantity, to its stellar mass (M$_*$), accumulated over its lifetime (Noeske et al. 2007; Daddi et al. 2007; Elbaz et al. 2007; Salim et al. 2007). The correlation persists across a wide range of stellar masses and SFRs and over a range of redshifts (Whitaker et al. 2012; Speagle et al. 2014; Salmon et al. 2015; Tasca et al. 2015; Schreiber et al. 2015; Kurczynski et al. 2016; Johnston et al. 2015; Santini et al. 2017). This has led to speculations about its origin, with theories suggesting this is controlled by halo mass accretion (Dutton et al. 2010; Forbes et al. 2014; Rodríguez-Puebla et al. 2015), or the regulation of gas infall and feedback (Tacchella et al. 2016; Mitra et al. 2016), or that the observed correlation is simply a cross-section of a more fundamental SFR-M$_*$-Z relation, with the evolution explained by increasing metallicities with cosmic time (Mannucci et al. 2010; Lilly et al. 2013; Torrey et al. 2017).

Galaxy formation models predict the evolution of individual star-forming galaxies comprising the SFR-M$_*$ correlation. Testing these predictions has been difficult, with observations up to now unable to reveal if individual galaxies evolve along the correlation, as assumed by the 'Main Sequence Integration' technique (Leitner 2012; Muñoz & Peeples 2015) or make significant excursions above and below it (Pacifici et al. 2012; Tacchella et al. 2016).

One of the most common ways of probing the SFR-M$_*$ relation is estimating the stellar mass and SFR through SED fitting, with modern techniques capable of handling large quantities of data and extracting high-fidelity information. A key improvement in obtaining the stellar masses and SFRs comes from relaxing the assumption that a galaxy's Star Formation History (SFH) be described by a single parametric form such as exponentially declining or constant star formation (Iyer & Gawiser 2017; Pacifici et al. 2012; Acquaviva et al. 2011; Lee et al. 2017; Ciesla et al. 2017). Although the stellar masses and SFRs obtained through SED fitting offer a probe of the SFR-M$_*$ correlation to extremely high redshifts, it is sensitive to the systematic assumptions inherent in SED fitting, as well as the decreasing S/N as we approach dimmer objects and higher redshifts. To probe the correlation in this regime, we thus need to probe beyond these traditionally estimated SED-fit quantities.

In this paper, we apply the Dense Basis SED fitting method developed in Iyer & Gawiser (2017), which reconstructs a galaxy's Star Formation History using the best-fit from among multiple families of smooth SFHs.

We use these SFHs to construct SFR-M$_*$ trajectories, along which individual galaxies are propagated backwards in time to reveal the SFR-M$_*$ diagram at higher redshifts. This method allows us to gain statistical power from the large number of low-redshift galaxies that can contribute to the SFR-M$_*$ correlation at higher redshifts. In contrast to direct fits, these trajectories possess the advantage of probing lower masses as we extend the method to higher redshifts. Currently, the lowest stellar masses the correlation has been probed at involve using galaxies observed in the Hubble Ultra Deep Field to go down to $10^7 M_\odot$ at $z \sim 1.5$ (Kurczynski et al. 2016), or using gravitationally lensed galaxies in the HST Frontier Fields to go down to $10^{8.8} M_\odot$ at $z \sim 6$ (Santini et al. 2017). In this paper, we show that our technique of reconstructing SFR-M$_*$ trajectories allows us to recover the SFR-M$_*$ correlation down to $10^6 M_\odot$ at $1 < z < 6$.

The paper is structured as follows; in §2, we specify the choice of dataset, followed by the details of our analysis in §3. In §4 we present our main results, describe the validation tests we performed in §5, and discuss the implications and caveats in §6. Throughout this paper magnitudes are in the AB system; we use a standard ΛCDM cosmology, with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70$ km Mpc^{-1} s^{-1}.

2. DATASET

We fit 17-band photometric SEDs[1] spanning the rest-frame UV through near-IR (IRAC) from the Cosmic Assembly Near-Infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. (2011); Koekemoer et al. (2011); Nonino et al. (2009); Retzlaff et al. (2010); Fontana et al. (2014); Ashby et al. (2015)). The catalog selects objects in the GOODS-S field via SExtractor in dual-image mode using F160w as the detection band. The dual image mode (Galametz et al. 2013) is optimized to detect both faint, small galaxies in 'hot' mode without over de-blending large, resolved galaxies in 'cold' mode. The HST (ACS and WFC3) bands were point spread function (PSF) matched to measure photometry, and TFIT (Laidler et al. 2007) was used to measure the photometry of ground based and IRAC bands using the HST WFC3 imaging as a template. We consider galaxies in the GOODS-S field (Guo et al. 2013) at redshifts $0.5 < z < 6$. The three different depths (wide, deep, HUDF) allow us to probe a population of

---

[1] The photometric bands used are: U(CTIO), U(VIMOS), HST/ACS F435w, F606w, F775w, F814w, F850lp, HST/WFC3 F098w, F105w, F125w, F160w, VLT/HAWK-I Ks, VLT/Isaac Ks, and Spitzer/IRAC 3.6, 4.5, 5.8, 8μm. (Guo et al. 2013)
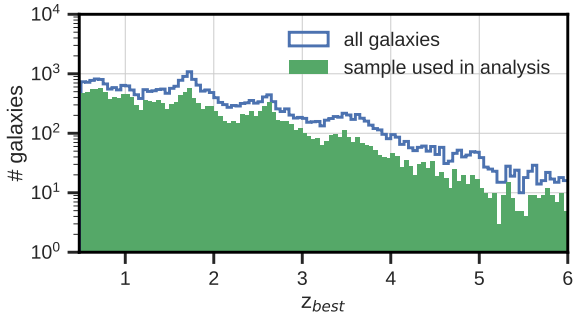
Figure 1. Redshift distribution of the CANDELS/GOODS-S galaxies at $0.5 < z < 6.0$ (blue line) and for the sample we use in our analysis (solid green histogram) excluding bad fits and galaxies with F160w magnitudes < 25.9, 26.6 and 28.1 in the wide, deep and HUDF regions, respectively.

galaxies across a wide range of masses to construct SFR-M$_*$ trajectories. To reduce the effects of incompleteness for the sample while accounting for the different depths, we require our sample to be brighter than a F160w magnitude limit of 25.9, 26.6 and 28.1 in the wide, deep and HUDF regions, respectively (Guo et al. 2013).

In performing our fits, we exclude objects that are marked as stars or X-ray detected AGN (Hsu et al. 2014) in the Santini et al. (2015) mass catalog (487 objects), as well as poor fits in our SED fitting routine ($\chi^2 > 50$, 595 objects). After these cuts, our analysis includes $\sim 17,873$ galaxies (94.7% of our parent sample). To perform our fits, we use an updated photometric redshift catalog by Kodra et al (in prep.) containing an increased number of spec-z measurements as well as photometric redshifts with Bayesian combined uncertainties estimated by comparing the redshift probability distributions of four different SED fitting methods. Figure 1 shows the redshift distribution of our analysis sample. We perform our fits using their $z_{\rm best}$ binned to the resolution of our pre-grid, with $\delta z = 0.01$. The $z_{\rm best}$ in our redshift range includes ~ 1917 spectroscopic and ~ 384 grism redshifts, in addition to photometric redshifts.

3. METHODOLOGY

3.1. *SED fitting*

The Dense Basis method described in Iyer & Gawiser (2017) uses a physically motivated basis of Star Formation Histories to generate an atlas of template SEDs. The best-fit SFH, dust and metallicity values for each observed galaxy SED are computed using standard χ^2 minimization over the entire atlas. To generate spectra corresponding to a galaxy with a given basis SFH, we use the Flexible Stellar Population Synthesis (FSPS) model (Conroy et al. 2009; Conroy & Gunn 2010; Foreman-Mackey et al. 2014). We use a Chabrier IMF (Chabrier

2003), Calzetti dust reddening (Calzetti 2001), and IGM absorption according to the Madau et al. (1996) prescription. Star Formation Histories are drawn from the Linexp (linear rise followed by exponential decline, sometimes called Delayed-τ models), Gaussian and Log-normal families of curves, excluding SFHs that are extremely similar in shape to reduce the size of the basis, with a slight modification of Iyer & Gawiser (2017), where we also considered Bessel function rise followed by exponential decline (Bessel-exp), Top-hat (Exp) and Constant Star Formation (CSF) histories. In this work, we do not consider these three families because Bessel-exp SFHs are extremely similar in shape to the SFHs we already consider in our basis, and Exponential and CSF were shown in Iyer & Gawiser (2017) to lead to biased estimates of galaxy properties. The parameter ranges for the various families are given in Table.1. Defining the slope as a tangent to the log SFR-log M$_*$ trajectory at a given point, the range of SFH shapes lead to trajectories that can have a wide range of slopes at low masses $\sim [0, 16]$ as well as flat and negative slopes $\sim [-34, 52]$ at high masses, as galaxies enter a quiescent phase. Figure.2 (a,b) shows examples of SFHs from each of the three families, as well as their corresponding trajectories in SFR-M$_*$ space. Insets in panels (c,d) of the same figure show examples of mock SFHs from the MU-FASA hydrodynamic simulation (panel c, Davé et al. (2016)) and a Semi-Analytic Model (panel d, Somerville et al. (2015)) and their reconstructed best-fit SFHs with uncertainties. The mock SEDs fitted to reconstruct the SFHs were generated using the same filters as the CANDELS/GOODS-S catalog, with realistic photometric noise and dust, as in Iyer & Gawiser (2017). While the method does not recover short stochastic episodes of star formation, it does well approximate the overall trend of the galaxy's SFH, thus allowing us to construct robust trajectories in SFR-M$_*$ space. Since we fit galaxies with a single episode of star formation, we find that our reconstructions can sometimes fail in cases where the true SFH of the galaxy contains multiple strong episodes of star formation. However, in Iyer & Gawiser (2017) we find that only about 15% of galaxies at z$\sim$1 support fits with two episodes of star formation. Fits to the galaxy SEDs at different redshifts provide us with both the Stellar Masses and effectively instantaneous Star Formation Rates at the epoch of observation, referred to as 'Direct Fits' for the rest of this work. In addition to this, the reconstructed Star Formation Histories are then used to construct SFR-M$_*$ trajectories, which allow us to infer the Stellar Masses and Star Formation Rates at higher redshifts of interest.
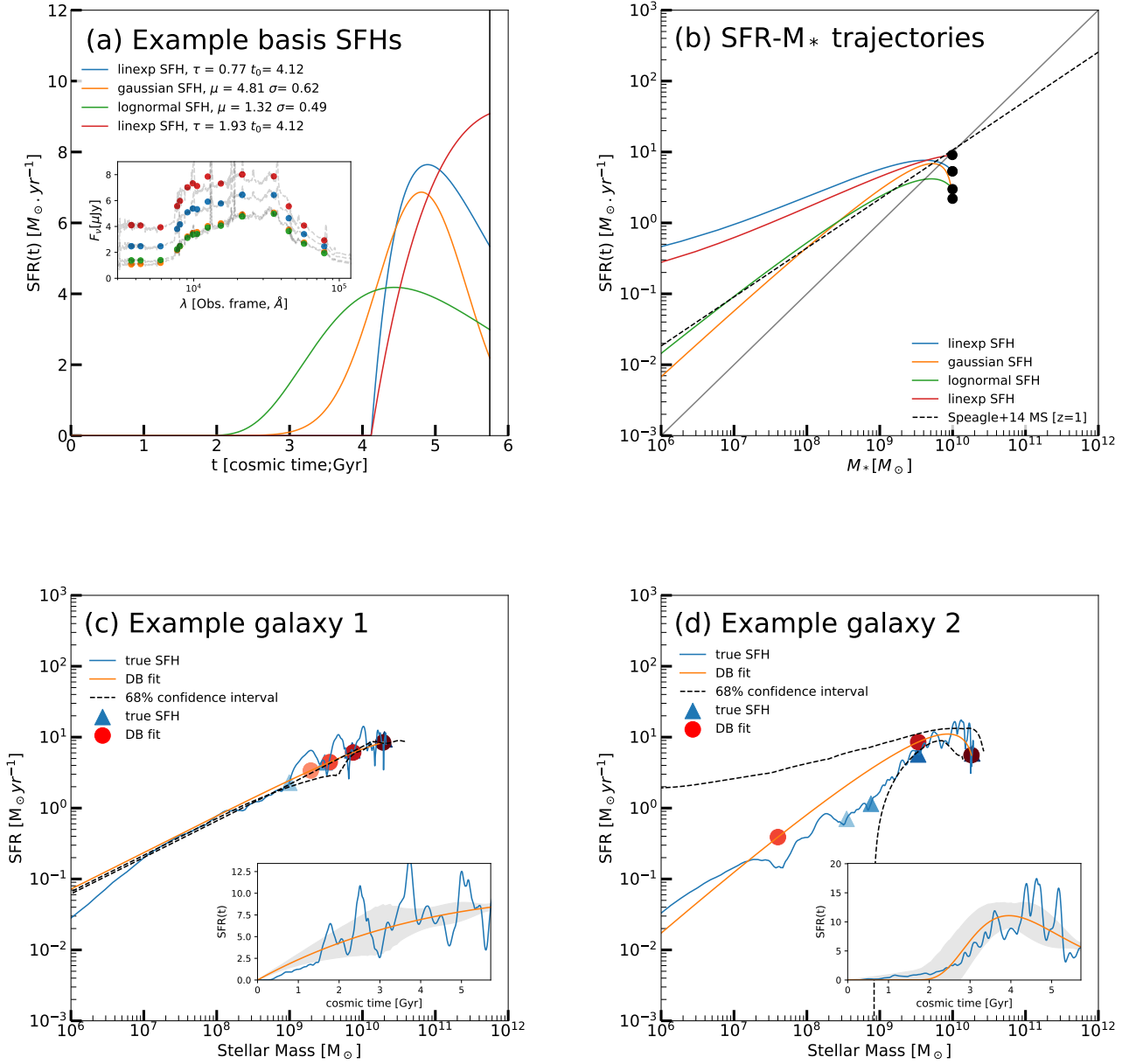
Figure 2. **(a)** Examples of basis Star Formation Histories belonging to the three functional families used for SED fitting in the paper, normalized to the same stellar mass of $10^{10} M_\odot$, with similar SFRs. Inset panel shows corresponding SEDs observed at z=1 with reference spectra plotted vs observed wavelength. **(b)** Trajectories in the SFR-M_* plane corresponding to the four SFHs shown above, with black circles illustrating their locations when lookback time equals zero. The dashed black line shows the Speagle et al. (2014) SFR-M_* relation at z=1 and the solid grey line shows $\log SFR = \log[M_*/10^9 M_\odot]$ for reference. **(c,d)** Examples of SFH reconstructions of individual $z = 1$ MUFASA (panel c) and SAM (panel d) galaxies with their uncertainties, extended to trajectories in SFR-M_* space. Inset figures show the simulated SFH (true SFH, blue) and the SFH reconstructed through SED fitting the noisy simulated photometry (DB fit, orange). The main panels show their corresponding trajectories in SFR-M_* space. Coloured circles and triangles show SFR, M_* estimates at z=1,2,3,4 (darker to lighter colors), and dashed lines show the 68% confidence interval, corresponding to the grey shaded region in the inset. Example galaxy 2 (panel d) can be reliably propagated back to $z \sim 2$, but not beyond that. In our analysis, we exclude such trajectories at redshifts where they have large uncertainties.

3.2. *SFH uncertainties:*

We compute uncertainties on the reconstructed SFH using the full χ^2 surface over the multidimensional space
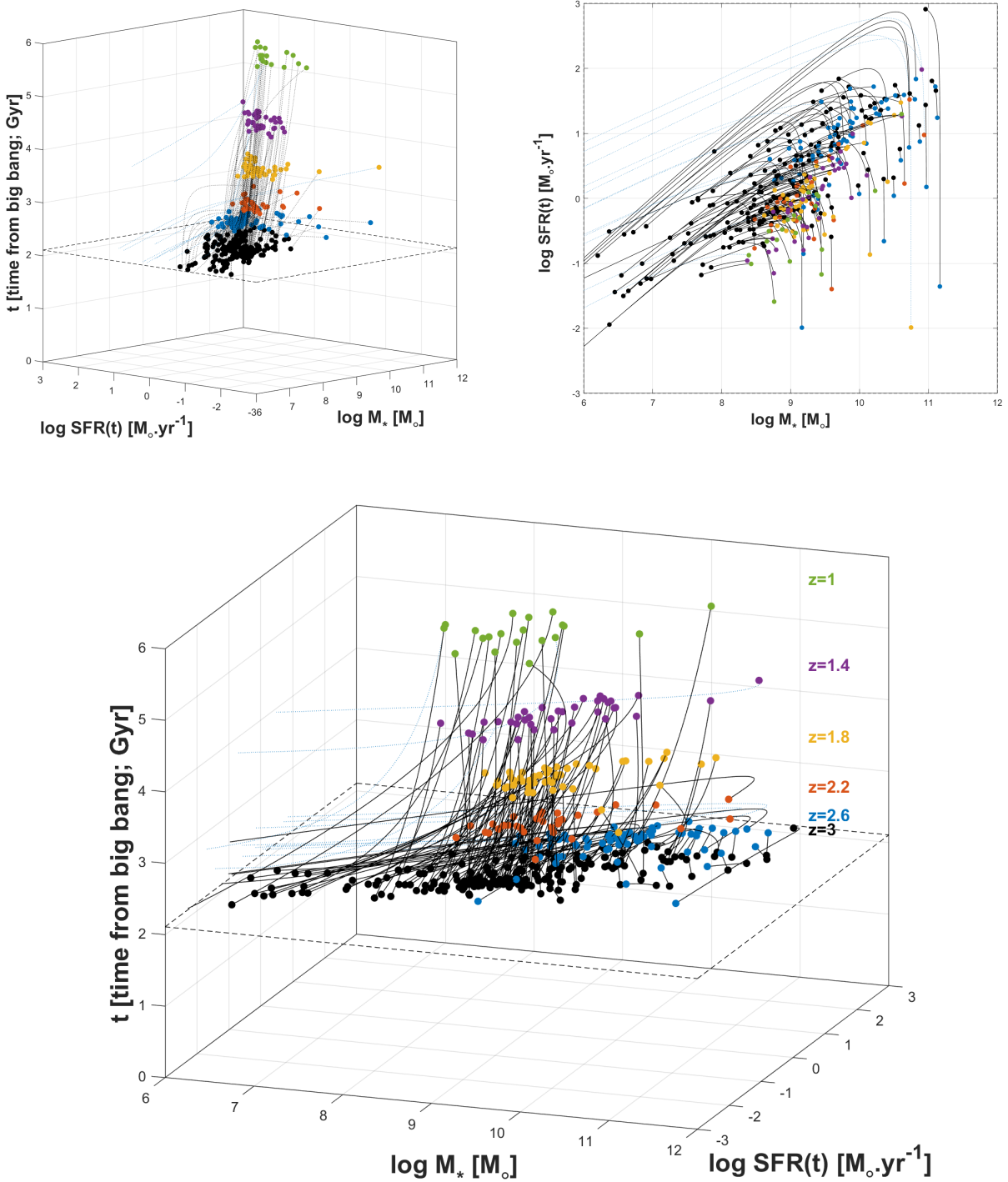
Figure 3. The SFR-M$_*$ correlation at $z = 3$ (black points) constructed by propagating a randomly chosen subset of galaxies at redshifts $z = [1.0, 1.4, 1.8, 2.2, 2.6]$ (colored points) backwards in time along their best-fit SFR-M$_*$ trajectories (black lines are galaxies whose trajectories go back to z=3, blue dotted lines are galaxies whose trajectories drop off the plot at lower redshifts). Trajectories that go beyond $z = 3$ are truncated at $z = 3$ for clarity. Three orientations of the figure are shown along different viewing angles. The plots allow us to see that the correlation evolves along a relatively narrow phase space, although individual galaxies can make significant excursions above and below it. This allows us to probe the correlation at high redshifts using higher S/N SED fits at lower redshifts. It also allows us to probe the correlation down to lower masses than previously possible.

Table 1. Parameter ranges for SFH families[a]

SFH	param 1	param 2
Linexp	$\tau \in [0.05, 10]$	$t_0 \in [0, t_{univ}]$
Gaussian	$\mu \in [0, t_{univ}]$	$\sigma \in [0.1, 10]$
Lognormal	$\mu \in [0, t_{univ}]$	$\sigma \in [0.1, 2]$

[a]Parameter ranges, in Gyr, adapted for $0.5 < z < 6$ from Iyer & Gawiser (2017). In addition to these, there is a normalization corresponding to the stellar mass, which can be considered a third free parameter in specifying the SFH.

spanning SFH, dust and metallicity, generated through the fitting procedure, following a method similar to Iyer & Gawiser (2017). Using the cumulative histogram of the χ^2 values, we select the 100 SFHs corresponding to the set of lowest χ^2 values among the set and compute the median, which we hereafter refer to as the median SFH. We find that changing this threshold of 100 SFHs does not substantially affect the computed uncertainties for this analysis. We then prune the set of good SFHs, removing those that are simply bad fits ($\chi^2 > 10 * min(\chi^2)$) or outliers in SFH space ($max(|SFH_i - SFH_{median}|)/\langle SFH_{median} \rangle > 5$), similar to any robust algorithm that is insensitive to outliers. We use the distribution of remaining SFHs to derive pointwise 68% confidence intervals in time for the reconstructed SFH, as well as a robust median SFH. We use the stellar masses and SFRs of the full set of good fits to derive confidence intervals for those quantities at any lookback time along the trajectory. This technique is also used to find the uncertainties on SFR and Stellar Mass at lookback times corresponding to the redshifts of interest in Sec. 4. To check that our uncertainties are robust at all lookback times as a function of rest-frame wavelength coverage (depending on z_{obs}) or SED S/N, we use mock SEDs to analyze possible biases in estimating SFR and M_* as a function of lookback time for different z_{obs}. This is detailed further in Appendix A, where we find that our median uncertainties are conservative and increase accordingly at lookback times where the Stellar Mass or SFR is poorly constrained.

3.3. SFR-M_* trajectories:

The best-fit reconstructed Star Formation Histories are curves of SFR(t) against time. At any instant in time, the Stellar Mass is given by

$$M_*(t) = \int_0^t SFR(t')f_{ret}(t' - t, Z)dt \qquad (1)$$

where $f_{ret}(t' - t, Z)$ is a metallicity dependent fraction of the mass of formed stars that is retained as stars

or stellar remnants at the time of observation obtained from FSPS (Conroy et al. 2009), which is typically between 0.6-1.0. Using this relation and the best-fit SFH, we can construct a parametrized curve corresponding to [SFR(t),M_*(t),t], which provides a trajectory in SFR-M_* space. Observing this trajectory at any redshift gives the [SFR(z),M_*(z)], allowing us to extend trajectories to higher redshifts and fill in the SFR-M_* correlation using previously inaccessible data from earlier periods in a galaxy's lifetime. Panel (b) of Figure 2 shows examples of trajectories corresponding to each of the basis SFHs shown in panel (a). Panels (c,d) of Figure.2 show a couple of examples of reconstructed SFR-M_* trajectories corresponding to a couple of SFHs at $z = 1$ from a hydrodynamical simulation (MUFASA, Davé et al. (2016)) and a Semi-Analytic Model (Somerville et al. 2015, 2008). As seen in the figure, the true trajectory is well approximated by the smooth reconstruction and its corresponding uncertainties, matching the observations not only at the epoch of observation ($z = 1$), but also at earlier epochs ($z = 2, 3, 4$). While the galaxy in panel (c) has a SFH that can be traced back to very high redshifts, this is not in general true for most observed galaxies. Galaxy trajectories may fail to contribute meaningfully at higher redshifts either because they drop off the plot, ie. they formed most of their mass at more recent epochs, or because their uncertainties grow extremely large. The latter case is illustrated through the example galaxy in panel (d), which can be reliably traced back to $z \sim 2$, but has large uncertainties beyond that. In our analysis, we exclude such trajectories at redshifts where they have large uncertainties.

Figure.3 shows a randomly selected sample galaxies at $1 < z < 3$, that are propagated backwards in time along their trajectories to infer the SFR-M_* correlation at $z \sim 3$. Black lines denote trajectories for galaxies that can be propagated backwards to $z = 3$, while blue dotted lines denote trajectories for galaxies whose trajectories do not reach back to $z = 3$. Since we choose a random subsample of galaxies to plot at each redshift, the F160w selection threshold results in the appearance of a rising lower limit in stellar mass as we go to higher redshifts - this doesn't imply that galaxies are more massive at $z \sim 2.6$, but that observationally selected galaxies, of which we pick a random sample, tend to be the more massive ones. The average amount of time a galaxy is propagated backwards in time along its trajectories shows a mild increase as we go to higher redshifts but remains much smaller than the amount of time between $z = 0.5$ and the redshift of interest, as shown in appendix A.

For the rest of this work, while considering a sample of galaxies propagated backwards in time along their SFR-M∗ trajectories, we restrict ourselves to the sub-sample of galaxies with low uncertainties ($\sqrt{\sigma_{SFR}^2 + \sigma_{M_*}^2} < 1$ dex) to minimize the effects of possible biases. This is explored in detail in Appendix A, where we fit mock SEDs corresponding to SFHs from simulations to assess the robustness of SFR and Stellar Mass as we propagate galaxies backwards in time along their trajectories. In doing so, we find that the uncertainties closely trace possible biases, incorporating effects due to factors like S/N and the rest-frame wavelength coverage during SED fitting. This allows us to isolate a subsample with minimal bias that we use for the analysis in this paper.

4. RESULTS: THE SFR-M∗ CORRELATION FROM DIRECT FITS AND TRAJECTORIES

In Figure 4, we present the SFR-M∗ correlation at z = [1,2,3,4,5,6], including estimates from galaxies observed at those epochs (henceforth direct fits), as well as from galaxies observed at later epochs propagated backwards in time along their trajectories (henceforth trajectories). The direct fits are shown as contours and as individual datapoints at high redshifts where the number of galaxies are small. The contribution from trajectories at each redshift is shown as a coloured heatmap.

We find that the locus of the direct fits and trajectories broadly agree with each other. To quantify this statistically, we perform a KS test comparing the two datasets at each redshift of interest, as shown in Table 2. Since we do not want to compare outlier distributions, due to starbursts or quenched galaxies, we impose a cutoff, excluding galaxies that are at a distance ≥ 0.4 dex from the best-fit SFR-M∗ correlation. In Appendix B we also compare distributions using a variable threshold based on the observed scatter and find that the results do not change. The number of galaxies that contribute to the SFR-M∗ correlation from direct fits and trajectories are given in the table. The p-values for all these comparisons are > α, indicating that the two distributions are not statistically different. The significance level for each test is $\alpha' = (0.05/6) \approx 0.0083$, where we apply a Bonferroni correction (Goeman & Solari 2014) to control for false positives since we are performing a family of tests to evaluate a single hypothesis. Since our results remain consistent across this broad range of tests, we can not reject the null hypothesis that the two samples are drawn from a common underlying distribution at > 95% confidence. While this does not completely rule out the possibility that the two distributions are different, this agreement justifies the usage of a combined sample to obtain our primary results. We further explore the comparison between the two distributions in Appendix B.

We plot the best-fit line to the combined dataset of direct fits and trajectories in Figure 4, determined using an iterative robust fitting routine that excludes outliers (Holland & Welsch 1977). To compare trajectories and direct fits on the same footing, the direct fits are analyzed at z=[1,2,3,4,5,6] in bins of $\Delta z = 0.1$. The uncertainties are determined using 1000 Monte Carlo realizations of the data perturbed within the M∗ and SFR error estimates obtained through the Dense Basis SED fitting routine. We find that the SFR-M∗ correlation extends to redshifts as high as $z \simeq 6$ (Steinhardt et al. 2014), and remains linear down to masses as low as $\log M_*/M_\odot \sim 7$, which is a factor of 10 below current estimates from direct fits. To test the linearity of the correlation, we fit the combined data at each redshift with polynomials of order 1 (linear) and 2 (quadratic) and see if the corresponding improvement in the goodness-of-fit is statistically significant using an F-test. At all redshifts, we find that the linear fit is preferred, at > 90% confidence, with values at individual redshifts given in Table 4. From the reconstructed SFHs, $\sim 92\%$ of the galaxies in the sample have $t_{10} \leq 3Gyr$ and $\sim 70\%$ of the galaxies in the sample have $age \leq 3Gyr$, where t_{10} is the lookback time at which they formed the first 10% of their observed stellar mass. This implies that most of the galaxies that contribute to the SFR-M∗ diagram form the majority of their stellar mass in ≤ 3 Gyr in the redshift range we consider, with 70% entering the observable SFR-M∗ range within that time. To account for the fact that the F-test need not guarantee that the correlation is indeed linear, we also perform non-parametric regression in Appendix D. Using this, we see that the nonparametric methods closely approximate the best-fit line as we go to low stellar masses at high redshifts.

To study the redshift evolution of the slope and normalization of the SFR-M∗ correlation, we analyze the direct fits and trajectories separately and present the results in Table 4 and graphically in Figure 5. The slope from both approaches are consistent within uncertainties, and roughly match the published meta-analysis of Speagle et al. (2014) in figure 5, shown as a solid purple line. This result is reassuring considering the Speagle et al. (2014) relation was calibrated at stellar masses above $10^9 M_\odot$. In comparison to the Speagle et al. (2014) relation, however, we find the slope to be consistent with little to no evolution with redshift. The normalization for the trajectories shows the same trend in redshift as the direct fits and the Speagle et al. (2014) relation, albeit being systematically lower at high redshifts. For better comparison with literature, we also
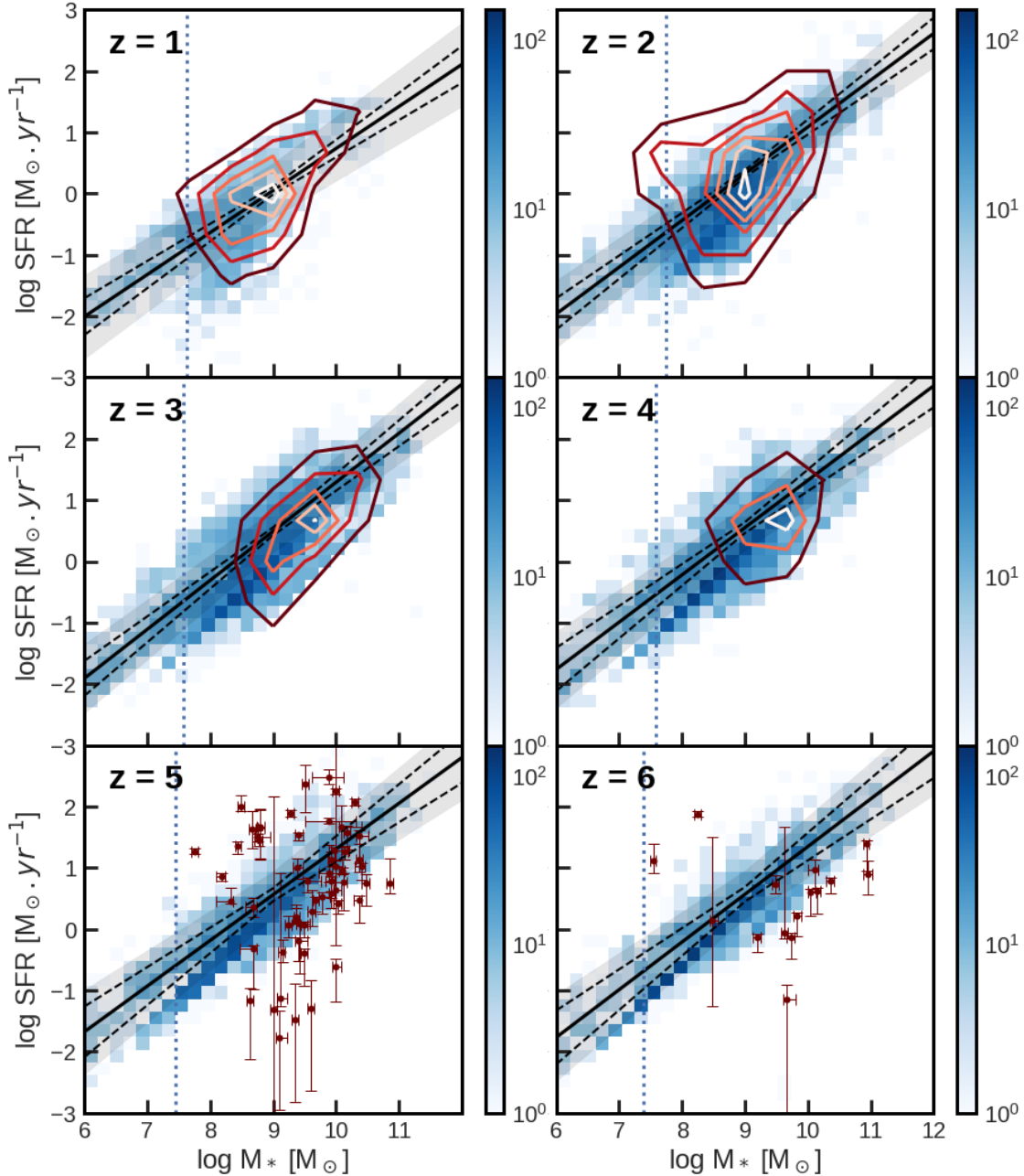
Figure 4. The SFR-M_* correlation at $z = [1, 2, 3, 4, 5, 6]$. Galaxies observed at the epochs of interest (direct fits) are shown as red contours for $z < 5$, and as red points with error bars for $z = 5, 6$ where there are insufficient points to yield representative contours. Galaxies observed at later epochs and propagated backwards in time along their SFR-M_* trajectories are shown as the colored heatmap, with the colorbars denoting the number of galaxies in a particular pixel. The black solid line shows our best-fit to the combined dataset, with uncertainties denoted by the dashed black lines.. The shaded black region shows the uncertainties + observed scatter around the best-fit. Dotted blue lines show the 10th percentile in stellar mass for trajectories. Additional non-parametric fits to the correlation are shown in Appendix D. We see that the SFR-M_* correlation is consistent with being linear out to very low masses and high redshifts.

considered a set of estimates for the slope and normalization where we use an additional UVJ selection criterion to select star forming galaxies for the direct fits dataset. Details of the UVJ selection can be found in

Appendix C, where we find that the results don't vary much due the robust fitting algorithm we use. We do not consider such a criterion for trajectories since galaxies that are quiescent at the epoch of observation can
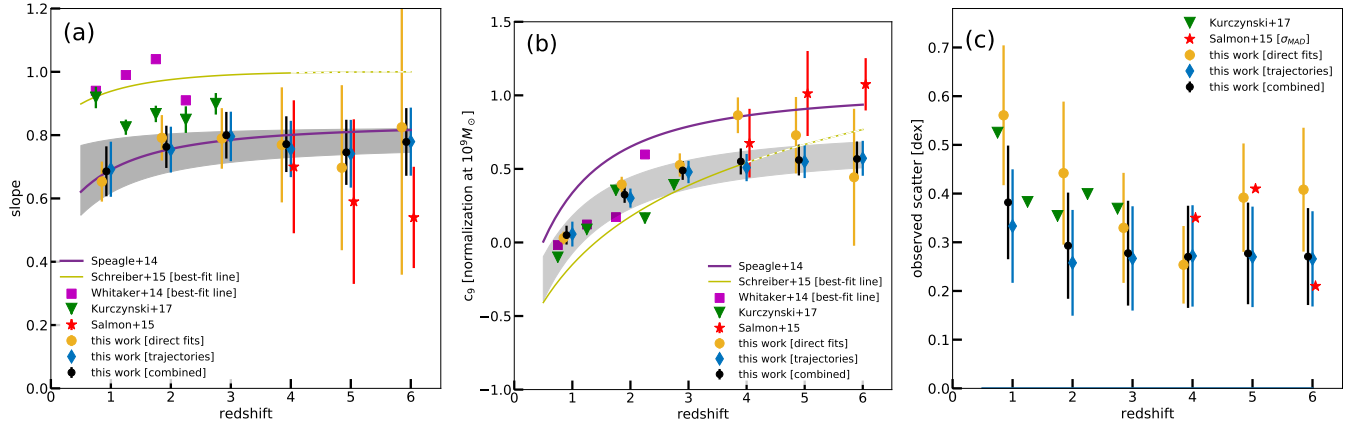
Figure 5. Evolution of the slope (m), normalization $(c_9$, the intercept at log $M_* = 9)$ and observed scatter of the SFR-M_* correlation. Black circles with errorbars show our results for the combined (direct fits + trajectories) dataset, and the grey shaded region is generated using the evolving relation defined in Eqn.2. We find that the slope from the direct fits (orange circles with error bars) and trajectories (blue diamonds with error bars) are roughly consistent within uncertainties with each other at all times, with some disagreement in normalization at $2 < z < 4$. Our estimates for slope agree well with Speagle et al. (2014) at low redshifts and Salmon et al. (2015) at high redshifts, while being consistently sub-linear, in comparison to Whitaker et al. (2014); Schreiber et al. (2015). Our normalization is closer to Schreiber et al. (2015) than Speagle et al. (2014) in value, but has a rate of evolution more consistent with the latter. Our measurement of observed scatter agrees well with the observed scatter reported in Kurczynski et al. (2016); Salmon et al. (2015). Measurements from Whitaker et al. (2014); Schreiber et al. (2015) are shown for comparison, using a local slope at $10^9 M_\odot$ for non-linear reported correlations.

Table 2. Comparing distributions from Direct fits vs Reconstructed Trajectories[a]

redshift	p-value	KS-statistic	cutoff [dex]	#traj	#direct
1.0 $(0.9 < z < 1.1)$	0.85	0.03	0.4	947	1667
2.0 $(1.9 < z < 2.1)$	0.10	0.06	0.4	2950	1003
3.0 $(2.9 < z < 3.1)$	0.02	0.10	0.4	3394	419
4.0 $(3.9 < z < 4.1)$	0.67	0.08	0.4	3019	162
5.0 $(4.5 < z < 5.0)$	0.20	0.13	0.4	2396	183
6.0 $(5.5 < z < 6.0)$	0.98	0.09	0.4	2101	81

[a]For both methods (galaxies observed at a particular epoch vs those observed at lower redshifts and propagated along their trajectories), we compute the distribution of distances of individual galaxies from the combined best-fit SFR-M_* correlation at z=1,2,3,4,5,6.
These distributions are compared using a KS test, to test the hypothesis that they are consistent with being drawn from the same distribution We reject this hypothesis if the p-value $> \alpha(= 0.0083)$. To arrive at this value we apply a Bonferroni correction (Goeman & Solari 2014) to control for false positives since we are performing a family of tests to evaluate a single hypothesis.
Since we do not want to include starburst and quiescent galaxies, we exclude galaxies that lie at a distance > 0.4 dex from the correlation. In Table 7 we also give p-values for the case where we exclude galaxies that lie farther than $1\times$ the observed scatter at each redshift. The last two columns show the number of galaxies from each dataset used in comparing the two distributions. The test shows that the direct fits and trajectories are consistent with being drawn from the same distribution at all redshifts as evinced by small values of the KS statistic, which measures the maximum distance between the CDF of the two distributions. The larger KS distance in the $z \sim 5$ redshift bin is due to the small number of points from direct fits.

be traced back to epochs when they were star forming and thus contribute to the SFR-M_* correlation at higher redshifts. Observed scatter is computed using the same procedure as Kurczynski et al. (2016), finding the standard deviation of the distribution of ΔSFR from the best-fit correlation excluding points beyond 1 dex. We

find that the scatter is close to 0.3 dex at all epochs with the possibility of being higher at low redshifts as seen for the direct fits. The observed scatter contains contributions from noise that needs to be deconvolved to estimate the intrinsic scatter (Kurczynski et al. 2016) and while our scatter for direct fits and trajectories are

consistent within uncertainties, it is possible that we underestimate the intrinsic scatter measured with trajectories since we do not consider short-timescale excursions from the smooth best-fit SFH for individual galaxies (Matthee & Schaye 2018). The published Kurczynski et al. (2016) and Salmon et al. (2015) values for slope, normalization and scatter are also shown as blue triangles at $0.5 < z < 2.5$ and purple stars at $3.5 < z < 6.5$. Whitaker et al. (2014) and Schreiber et al. (2015) report a nonlinear SFR-M_* correlation due to a turnover at high stellar mass ($M_* > 10^{10.5} M_\odot$). Since most of our galaxies fall below this mass range, we fit their reported correlation near $10^9 M_\odot$ with a line to find the effective low mass slope and normalization shown in Figure 5.

Fitting the slope and normalization of the best-fit SFR-M_* correlation as a function of cosmic time, we find that the relation is well described by a linear fit, given by

$$\log SFR = (0.80 \pm 0.029 - 0.017 \pm 0.010 \times t_{univ}) \log M_* \\ - (6.487 \pm 0.282 - 0.039 \pm 0.008 \times t_{univ}) \quad (2)$$

where t_{univ} is the age of the universe in Gyr at a given redshift. This relation is shown as the grey shaded region in Figure.5.

In estimating the slope and normalization of the SFR-M_* correlation, we use all galaxies that satisfy the selection criterion described in method (c) in Appendix A. In Table 3, we estimate the minimum well-sampled mass at each redshift, below which the statistics may be insufficient to confirm that the values for slope and normalization of the SFR-M_* correlation still apply. This is based on the distribution in Stellar Mass and SFR for direct fits and trajectories, conservatively quantified as the 10th percentile of the respective distributions. We find that we can probe the SFR-M_* correlation to ~ 1 dex lower than possible with just direct fits. This is possible since we are no longer limited by selection effects such as the F160w detection threshold, which does not allow us to detect the faint, low mass galaxies at high redshifts that we would see at lower redshifts. Using trajectories is thus a useful tool to go deeper in SFR and M_* at high redshifts.

Considering the trajectories allows us to effectively increase the survey volume obtained by propagating galaxies observed at later epochs backwards in time. We estimate this effective increase in the volume of the survey by comparing the ratio of the number of galaxies from just direct fits vs direct fits + trajectories in a comparable mass range. The comparable mass range is obtained by requiring that the median of M_* for the direct fits and trajectories in this mass range be separated by $< 0.1 dex$.

In theory, this can be applied to any survey that probes a wide variety of galaxy types to allow us to further extend our SED fitting results using SFR-M_* trajectories. Although it is beyond the scope of this work, it is important to include corrections to the effective volume on an individual galaxy basis (based on the amount of time they have been extrapolated backwards along their trajectory) while considering problems such as calculating luminosity functions or number densities using the combined trajectories + direct fits datasets.

While simulations don't yet make predictions for the slope and normalization of the correlation at the lowest masses, our results help put strong constraints on the models. When comparing M_*, SFR distributions obtained through trajectories with simulations, it is important to k is there a physical explanation for the discrepancy between the trajectory scatter and the direct fits scatterefore, to compare SFR-M_* correlations between our results and the simulations on the same footing, it is important that the correlation be compiled at any redshift using galaxies summed over all progenitors at the redshift of observation - for example, to compare accurately to reconstructed trajectories at $z \sim 6$, a simulation should be allowed to run to at least 3 Gyr in the future to about $z \sim 4$, then traced back to $z \sim 6$. This imposes resolution requirements at both the lower redshift, for the discovery of galaxies and at the redshift of interest, to be able to distinguish progenitors that contribute to the trajectories. The similarity of the two distributions when compared using the KS test indicate that this difference is not a major one for our observed sample of CANDELS/GOODS-S galaxies.

5. VALIDATION:

SED fitting allows us to estimate M_* and SFRs at the epoch of observation. In addition to this, we estimate the Stellar Masses and SFRs at previous epochs during which the galaxy was forming stars, by propagating galaxies backwards in time along their reconstructed SFR-M_* trajectories. In Appendix A we verify the robustness of our trajectories, and restrict our analysis to the subsample of galaxies whose trajectories have low uncertainties in SFR and M_* at a given redshift of interest. A reassuring check of the robustness of our method come from the similarity between the distributions around the SFR-M_* correlation from direct fits and trajectories in Table 2.

To further ensure that we do not get an artificially linear correlation due to our fitting method, we use a sample of mock SFHs from the MUFASA hydrodynamic simulations (Davé et al. 2016) to run a series of validation tests by changing the slope, normalization and

redshift	log M$_*$ (direct) (10th percentile)	log M$_*$ (traj.) (10th percentile)	log SFR (direct) (10th percentile)	log SFR (traj.) (10th percentile)	Eff. Volume (direct + traj.)
1.0	7.99	7.62	-1.19	-1.82	1.29x
2.0	7.91	7.74	-0.85	-1.16	3.25x
3.0	8.59	7.57	-0.50	-1.05	5.49x
4.0	8.59	7.59	-0.40	-1.03	11.74x
5.0	8.67	7.44	-1.13	-1.15	8.61x
6.0	8.34	7.38	-0.13	-1.2	13.74x

Table 3. 10th percentile of Stellar Mass and SFR probed using direct fits and trajectories at different redshifts. We see that using trajectories allows us to probe the SFR-M$_*$ correlation to nearly 1 dex deeper at high redshifts as compared to simply using direct fits. The last column estimates the the increase in the effective volume of the survey, using the increased number of galaxies at a particular redshift obtained by adding trajectories in a mass range where the direct fits and trajectories are comparable.

z	m_{direct}	m_{traj}	m_{total}	$c_{9,direct}$	$c_{9,traj}$	$c_{9,total}$	$\sigma_{obs,direct}$	$\sigma_{obs,traj}$	$\sigma_{obs,total}$	Conf. level
1	0.65±0.06	0.69±0.09	0.69±0.08	0.03±0.04	0.06±0.08	0.05±0.06	0.56±0.14	0.33±0.12	0.38±0.12	91.7%
2	0.79±0.07	0.75±0.07	0.76±0.07	0.39±0.05	0.30±0.06	0.32±0.06	0.44±0.15	0.26±0.11	0.29±0.11	91.2%
3	0.79±0.10	0.80±0.08	0.80±0.07	0.53±0.08	0.48±0.08	0.49±0.06	0.33±0.11	0.27±0.11	0.28±0.11	98.1%
4	0.77±0.18	0.76±0.09	0.77±0.09	0.86±0.12	0.51±0.09	0.55±0.09	0.25±0.08	0.27±0.10	0.27±0.10	97.9%
5	0.70±0.26	0.74±0.11	0.75±0.10	0.73±0.26	0.55±0.11	0.56±0.10	0.39±0.11	0.27±0.10	0.28±0.10	94.4%
6	0.82±0.47	0.78±0.11	0.78±0.11	0.44±0.47	0.57±0.12	0.57±0.12	0.41±0.13	0.27±0.10	0.27±0.10	97.6%

Table 4. Results: The slope, normalization at $10^9 M_\odot$ and observed scatter of the best-fit to the SFR-M$_*$ correlation at different redshifts as shown in Figure 5 for direct fits to galaxies observed at each epoch (**direct**; orange points in Figure 5), galaxies observed at later epochs propagated backwards along their SFR-M$_*$ trajectories (**traj**; blue points in Figure 5) and the combined sample (**total**; black points in Figure 5). Including the low mass data in our fits, we observe a milder evolution of the slope and normalization with time in comparison to Speagle et al. (2014), finding that the evolving correlation is best described by Eqn. 2: $\log SFR = (0.80 \pm 0.029 - 0.017 \pm 0.010 \times t_{univ}) \log M_* - (6.487 \pm 0.282 - 0.039 \pm 0.008 \times t_{univ})$, where t_{univ} is the age of the universe at a given redshift. The last column details the confidence levels (1−p-value) from the F-test to check the hypothesis that a linear fit to the SFR-M$_*$ is favoured over a quadratic fit.

shape of the simulated SFR-M$_*$ correlation to see if our fits can recover these changes. The results of are reported in Table 5. The first six columns report the results of a linear fit, to $\log SFR = m \log[M_*/10^9 M_\odot] + c_9$, reporting the slope ($m$), normalization ($c_9$) and the observed scatter (σ). The last two columns compare the goodness of fit of a first order (linear) and second order (quadratic) polynomial fit using an F-test, to determine if the improvement upon fitting with a second-order curve is statistically significant. This test allows us to test the hypothesis that the correlation is linear. For all tests, the simulated galaxies are fit at $z = 1$.

The first row of Table 5 evaluates that the SFR-M$_*$ correlation is robustly recovered for a randomly selected sample of galaxies at the epoch of observation ($z = 1$, direct fits) with no modifications to their SFHs. This represents the control case for our validation tests. The

next two rows change the slope of the correlation to see if our method can recover the artificially high or low slope. This is done using a gaussian envelope to modify the SFH in a way that the overall galaxy mass remains constant. We find that the recovered slope and normalization match the truth within uncertainties, as seen in columns 2-5 of Table.5. In the fourth row we check that the recovered correlation is sensitive to the linearity of the underlying correlation by curving the correlation such that it is better fit by a quadratic curve. This is achieved by adding a single gaussian component to a randomly drawn MUFASA SFH while varying lookback time at which the SFH peaks. This results in a curved SFR-M$_*$ correlation. Although there is always an improvement to the fit with an additional degree of freedom, we use an F-test with a threshold p-value of 0.1 to see if the quadratic fit provides a statistically sig-

Validation test	m_{orig}	m_{fit}	$c_{9,orig}$	$c_{9,fit}$	σ_{true}	σ_{fit}	α_{orig}	α_{fit}
Unchanged	0.886	0.896± 0.067	-0.212	-0.180± 0.041	0.238	0.231	1	1
Increased slope (z=1)	1.208	1.189± 0.038	-0.135	-0.091± 0.040	0.198	0.213	1	1
Decreased slope (z=1)	0.478	0.461± 0.072	-0.302	-0.288± 0.038	0.231	0.217	1	1
Changed shape (z=1)[†]	-0.075	-0.086± 0.093	-0.215	-0.228± 0.142	0.607	0.629	2	2
Increased slope (z=2)	1.049	1.028± 0.015	-0.158	0.082± 0.054	0.249	0.403	1	1
Decreased slope (z=2)	0.663	0.532± 0.055	0.165	0.268± 0.027	0.059	0.181	1	1
Changed shape (z=2)[†]	0.459	0.495± 0.049	-0.421	-0.139± 0.067	0.640	0.389	2	2

Table 5. Validation: estimating the sensitivity of the fits to slope, normalization and shape of the SFR-M_* correlation. m is the slope of the linear correlation, c is the normalization, and α denotes the degree of the polynomial that the correlation is best fit with. The m_{orig} and c_{orig} are the linear coefficients obtained from the best-fit to the SFR-M_* correlation generated directly using SED fitting with the modified SFHs for each test case.
[†]Although the slope and normalization are reported, the correlation in this case is better fit with a quadratic, and is not well described by the linear coefficients. For the $z = 1$ case, the true correlation is described by log SFR $= -0.364(logM_* - 9)^2 + 0.515(logM_* - 9) + 0.053$. For the $z = 2$ case, the coefficients are log SFR $= -0.199(logM_* - 9)^2 + 0.342(logM_* - 9) + 0.271$.

nificant improvement in describing the variance of the data and find that the order of our recovered correlation matches the input, shown in columns 8-9 of Table.5.

To check the robustness of the SFR-M_* correlation recovered from galaxies propagated backwards in time along their trajectories, we also perform the same tests using a randomly selected sample of galaxies that are fit at $z = 1$ and analyzed at $z = 2$. We choose $z = 2$ to test the trajectories, since we find that most galaxies contributing to the SFR-M_* correlation through trajectories at different redshifts form the bulk of their stellar mass at $\leq 3Gyr$ from the epoch of interest. We repeat the same tests for the trajectories as we did for the direct fits, while attempting to keep the z=1 correlation the same. Since achieving such changes through modifications to the MUFASA SFHs was difficult, we use a Monte Carlo method generating random gaussian contributions to the SFH until the resulting SFH satisfies a 'normal' MS slope at $z = 1$ and a modified slope at $z = 2$ significantly higher or lower than $m \sim 0.89$. We then generate SEDs corresponding to these SFHs and fit them, finding that the fitting method is reasonably sensitive to changes in the slope of the correlation at higher redshifts. A small number of quiescent galaxies at $z = 1$ also contribute to the correlation at higher redshifts, which helps increase the sensitivity of our approach. Additionally, we repeated the test where the shape of the correlation is changed to being better described by a quadratic rather than a linear curve, finding that this too is robustly recovered by the fits.

6. DISCUSSION

The tight correlation between the SFRs and Stellar Mass of star forming galaxies has been extensively stud-

ied, with simulations matching observations at z∼0 and high redshifts (Sparre et al. 2015; Salmon et al. 2015), but with some tension at intermediate redshifts around $z \sim 2$ (Sparre et al. 2015). The Dense Basis method allows us to extend the dynamic range across which we fit the SFR-M_* correlation to estimate the slope and normalization, helping provide more robust estimates of these quantities. We find sub-linear slopes at all redshifts consistent with mild evolution, similar to Speagle et al. (2014); Salmon et al. (2015) and Kurczynski et al. (2016) at $z > 1$. This is in contrast (about 2.5 σ) to Schreiber et al. (2015) and Whitaker et al. (2012), who find a slope closer to 1. Speagle et al. (2014) attributes the slope to a steady, environment driven mode of star formation, where a slope slightly below unity occurs due to feedback, leading to the growth of hot halos around higher mass galaxies and slows down gas accretion (Finlator et al. 2006; Davé 2008). Salmon et al. (2015) argues that gas accretion onto dark-matter halos at high-z is smooth over large timescales (Cattaneo et al. 2011; Finlator et al. 2011) assuming a power-law form of the SFH, which controls the scaling of both the SFR and stellar mass (Stark et al. 2009; González et al. 2011; Papovich et al. 2011). Results from our more versatile SFHs appear to extend this interpretation across a wider range of redshifts. A sub-linear slope to the SFR-M_* correlation is also relevant in the context of Abramson et al. (2016), which considers how the growth of a bulge adds M_* but not SFR. It would be an interesting analysis to further study how the scatter around the SFR-M_* correlation correlates with explicit SFH parameters like t10 (the lookback time when the galaxy forms the first 10% of its stellar mass) and morphological quantities like the bulge/disk ratio (Abramson et al. 2014). The

evolution of the normalization, which could be related to changing cosmological gas accretion rates with redshift (Dutton et al. 2010) generally agrees with the literature, while being $\sim 0.2 - 0.4$ dex lower than the Speagle et al. (2014) meta-analysis at all redshifts and ~ 0.5 dex lower than Salmon et al. (2015) at $z \sim 6$, albeit with larger uncertainties.

SFR-M$_*$ trajectories obtained through SED fitting provide a valuable tool to extract information about where galaxies lie on the Stellar Mass - SFR plane at multiple epochs, allowing us to probe the low-mass portion of the SFR-M$_*$ correlation as we go to higher redshifts, with greater numbers than previously available. However, in interpreting the results we obtain, it is important to keep in mind the limitations of the observational data, as well as the current implementation of the Dense Basis method. Since our method estimates the smooth overall trend of star formation in a galaxy's past, it does not recover stochastic 'short timescale' star formation events that contribute to the intrinsic scatter of the SFR-M$_*$ correlation. However, we find that only $\sim$5% of a sample of galaxies from SAMs (Somerville et al. 2015) and $\sim$7% of galaxies from MUFASA (Davé et al. 2016) show a short burst where SFR$_{10Myr}$ / SFR$_{life} > 10$, where SFR$_{10Myr}$ is the SFR averaged over the last 10 Myr lookback time, and SFR$_{life}$ is the SFR of a galaxy averaged over its lifetime. This is not significant enough to bias estimations of the slope/normalization, or to alter the results from our validation tests.

In hierarchical cosmology, the process of galaxy evolution includes major and minor mergers along with gradual mass growth due to infall. When SED fitting yields information about an observed galaxy's star formation history, even a summary statistic such as the age of its stellar population, this information is about the sum of all stars in that galaxys progenitors. Hence the trajectories that result from our SFH reconstruction represent trajectories of the summed stellar masses and star formation rates of each observed galaxy's progenitors, rather than those of its most massive progenitor at each epoch. However, observable samples of galaxies at e.g., $z = 4$, will contain the most massive progenitors of observable galaxies at e.g, $z = 2$, along with a few additional progenitors that are massive enough to be detected. Looking backwards in time, a minor merger with 10:1 or 3:1 mass ratio causes only a 0.04 or 0.12 dex offset, respectively, in mass between the sum-of-progenitors and the most massive progenitor, with the maximum offset of 0.3 dex coming from a 1:1 major merger. Such major mergers are predicted to be rare (Kaviraj et al. 2015; Rodriguez-Gomez et al. 2015; Ventou et al. 2017), even at high redshift, for galaxies

massive enough to be detected at $z \sim 1$. Nonetheless, the KS test described in §4 found the distribution of trajectory values about the inferred SFR-M$_*$ correlation to match that of the observed values from direct fits well enough that the hypothesis of these being drawn from a single underlying population is not ruled out at higher than 99% confidence for any of the samples.

Recent studies (Hsieh et al. 2017) indicate that a strong correlation exists between the Star Formation Rate Density (Σ_{SFR}) and the Surface Mass Density (Σ_{M_*}) in star forming galaxies at kpc scales. This indicates that the SFR-M$_*$ correlation may extend to much lower scales than currently measured, with the method described in this work providing a unique bridge to intermediate scales.

7. CONCLUSIONS

SFH reconstruction through SED fitting yields the trajectories of galaxies that evolve through SFR-M$_*$ space and is thus uniquely suited to probe the low-mass end of the SFR-M$_*$ correlation as we go to higher redshifts. In this paper, we used the Dense Basis method (Iyer & Gawiser 2017) to fit a sample of $\sim$17,800 galaxies in the CANDELS GOODS-S field at redshifts $0.5 < z < 6.0$. We used the reconstructed SFHs to obtain the stellar masses and star formation rates of galaxies at $z = 1, 2, 3, 4, 5, 6$. Using the combined dataset from galaxies observed at the epochs of interest (direct fits) and galaxies observed at lower epochs propagated backwards in time along their SFR-M$_*$ trajectories, we find that the SFR-M$_*$ correlation is linear to $\sim 10^6 M_\odot$ at high redshifts.

This allows us to study the nature and evolution of the SFR-M$_*$ correlation in greater detail than allowed by previous approaches like Main Sequence Integration (Leitner 2012) which assumes that star forming galaxies stay on the correlation throughout their lifetimes. We find that the overall trend of the evolution of the slope of the SFR-M$_*$ correlation with redshift is roughly consistent with the evolving Speagle et al. (2014) relation, while the normalization seems systematically lower by a factor of ~ 0.2 dex.

Thus new approach provides a probe of the correlation at much lower masses than previously possible (Salmon et al. 2015; Kurczynski et al. 2016) since stellar masses decrease as we propagate galaxies backwards along their SFR-M$_*$ trajectories. This is more important in view of the selection effects in the direction of increasing stellar mass as we go to higher redshift in galaxy surveys like CANDELS. It also allows for a closer comparison between observations and simulations, by providing constraints for simulations through the com-

parison of the predicted SFR-M$_*$ distributions to the reconstructed ones down to much lower masses.

ACKNOWLEDGEMENTS

REFERENCES

Abramson, L. E., Gladders, M. D., Dressler, A., et al. 2016, The Astrophysical Journal, 832, 7

Abramson, L. E., Kelson, D. D., Dressler, A., et al. 2014, The Astrophysical Journal Letters, 785, L36

Acquaviva, V., Gawiser, E., & Guaita, L. 2011, Proceedings of the International Astronomical Union, 7, 42

Ashby, M., Willner, S., Fazio, G., et al. 2015, The Astrophysical Journal Supplement Series, 218, 33

Calzetti, D. 2001, Publications of the Astronomical Society of the Pacific, 113, 1449

Cattaneo, A., Mamon, G. A., Warnick, K., & Knebe, A. 2011, Astronomy & Astrophysics, 533, A5

Chabrier, G. 2003, Publications of the Astronomical Society of the Pacific, 115, 763

Ciesla, L., Elbaz, D., & Fensch, J. 2017, arXiv preprint arXiv:1706.08531

Cleveland, W. S. 1979, Journal of the American statistical association, 74, 829

Conroy, C., & Gunn, J. E. 2010, The Astrophysical Journal, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, The Astrophysical Journal, 699, 486

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, The Astrophysical Journal, 670, 156

Davé, R. 2008, Monthly Notices of the Royal Astronomical Society, 385, 147

Davé, R., Thompson, R., & Hopkins, P. F. 2016, Monthly Notices of the Royal Astronomical Society, 462, 3265

Dutton, A. A., van den Bosch, F. C., & Dekel, A. 2010, Monthly Notices of the Royal Astronomical Society, 405, 1690

Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, Astronomy & Astrophysics, 468, 33

Finlator, K., Davé, R., Papovich, C., & Hernquist, L. 2006, The Astrophysical Journal, 639, 672

Finlator, K., Oppenheimer, B. D., & Davé, R. 2011, Monthly Notices of the Royal Astronomical Society, 410, 1703

Fontana, A., Dunlop, J., Paris, D., et al. 2014, Astronomy & Astrophysics, 570, A11

Forbes, J. C., Krumholz, M. R., Burkert, A., & Dekel, A. 2014, Monthly Notices of the Royal Astronomical Society, 443, 168

Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS (v0.1.1), doi:10.5281/zenodo.12157

Galametz, A., Grazian, A., Fontana, A., et al. 2013, The Astrophysical Journal Supplement Series, 206, 10

Goeman, J. J., & Solari, A. 2014, Statistics in medicine, 33, 1946

González, V., Labbé, I., Bouwens, R. J., et al. 2011, The Astrophysical Journal Letters, 735, L34

Grogin, N. A., Kocevski, D. D., Faber, S., et al. 2011, The Astrophysical Journal Supplement Series, 197, 35

Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, The Astrophysical Journal Supplement Series, 207, 24

Holland, P. W., & Welsch, R. E. 1977, Communications in Statistics-theory and Methods, 6, 813

Hsieh, B., Lin, L., Lin, J., et al. 2017, The Astrophysical Journal Letters, 851, L24

Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, The Astrophysical Journal, 796, 60

Iyer, K., & Gawiser, E. 2017, The Astrophysical Journal, 838, 127

Johnston, R., Vaccari, M., Jarvis, M., et al. 2015, Monthly Notices of the Royal Astronomical Society, 453, 2540

Kaviraj, S., Devriendt, J., Dubois, Y., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 2845

Koekemoer, A. M., Faber, S., Ferguson, H. C., et al. 2011, The Astrophysical Journal Supplement Series, 197, 36

Kurczynski, P., Gawiser, E., Acquaviva, V., et al. 2016, The Astrophysical Journal Letters, 820, L1

Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, Publications of the Astronomical Society of the Pacific, 119, 1325

Lee, B., Giavalisco, M., Whitaker, K., et al. 2017, arXiv preprint arXiv:1706.02311

Leitner, S. N. 2012, The Astrophysical Journal, 745, 149

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, The Astrophysical Journal, 772, 119

Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, Monthly Notices of the Royal Astronomical Society, 283, 1388

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, Monthly Notices of the Royal Astronomical Society, 408, 2115

Matthee, J., & Schaye, J. 2018, arXiv preprint arXiv:1805.05956

Mitra, S., Davé, R., Simha, V., & Finlator, K. 2016, Monthly Notices of the Royal Astronomical Society, 464, 2766

Muñoz, J. A., & Peeples, M. S. 2015, Monthly Notices of the Royal Astronomical Society, 448, 1430

Noeske, K., Weiner, B., Faber, S., et al. 2007, The Astrophysical Journal Letters, 660, L43

Nonino, M., Dickinson, M., Rosati, P., et al. 2009, The Astrophysical Journal Supplement Series, 183, 244

Pacifici, C., Kassin, S. A., Weiner, B., Charlot, S., & Gardner, J. P. 2012, The Astrophysical Journal Letters, 762, L15

Pacifici, C., Oh, S., Oh, K., Lee, J., & Yi, S. K. 2016, arXiv preprint arXiv:1604.02460

Papovich, C., Finkelstein, S. L., Ferguson, H. C., Lotz, J. M., & Giavalisco, M. 2011, Monthly Notices of the Royal Astronomical Society, 412, 1123

Rasmussen, C. E., & Williams, C. K. 2006, The MIT Press, Cambridge, MA, USA, 38, 715

Retzlaff, J., Rosati, P., Dickinson, M., et al. 2010, Astronomy & Astrophysics, 511, A50

Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, Monthly Notices of the Royal Astronomical Society, 449, 49

Rodríguez-Puebla, A., Primack, J. R., Behroozi, P., & Faber, S. 2015, Monthly Notices of the Royal Astronomical Society, 455, 2592

Salim, S., Rich, R. M., Charlot, S., et al. 2007, The Astrophysical Journal Supplement Series, 173, 267

Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, The Astrophysical Journal, 799, 183

Santini, P., Ferguson, H., Fontana, A., et al. 2015, The Astrophysical Journal, 801, 97

Santini, P., Fontana, A., Castellano, M., et al. 2017, The Astrophysical Journal, 847, 76

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

Šidák, Z. 1967, Journal of the American Statistical Association, 62, 626

Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, Monthly Notices of the Royal Astronomical Society, 391, 481

Somerville, R. S., Popping, G., & Trager, S. C. 2015, Monthly Notices of the Royal Astronomical Society, 453, 4337

Sparre, M., Hayward, C. C., Springel, V., et al. 2015, Monthly Notices of the Royal Astronomical Society, 447, 3548

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, The Astrophysical Journal Supplement Series, 214, 15

Stark, D. P., Ellis, R. S., Bunker, A., et al. 2009, The Astrophysical Journal, 697, 1493

Steinhardt, C. L., Speagle, J. S., Capak, P., et al. 2014, The Astrophysical journal letters, 791, L25

Tacchella, S., Dekel, A., Carollo, C. M., et al. 2016, Monthly Notices of the Royal Astronomical Society, 457, 2790

Tasca, L., Le Fèvre, O., Hathi, N., et al. 2015, Astronomy & Astrophysics, 581, A54

Torrey, P., Vogelsberger, M., Hernquist, L., et al. 2017, arXiv preprint arXiv:1711.11039

Ventou, E., Contini, T., Bouché, N., et al. 2017, Astronomy & Astrophysics, 608, A9

Whitaker, K. E., Van Dokkum, P. G., Brammer, G., & Franx, M. 2012, The Astrophysical Journal Letters, 754, L29

Whitaker, K. E., Franx, M., Leja, J., et al. 2014, The Astrophysical Journal, 795, 104

Williams, R. J., Quadri, R. F., Franx, M., Van Dokkum, P., & Labbé, I. 2009, The Astrophysical Journal, 691, 1879

APPENDIX

A. VALIDATION OF TRAJECTORY ROBUSTNESS

In Sec. 5, we showed that our SED fitting technique is robustly able to recover the SFR-M_* correlation corresponding to a mock dataset from the MUFASA simulation. This is true both for the case of direct fits at $z \sim 1$, and for stellar masses and star formation rates recovered from the SFR-M_* trajectories at $z \sim 2$.

Here we check for biases in the reconstruction of individual trajectories in SFR-M_* space at different redshifts. If the estimated uncertainty from our SED fit is bigger than the bias, this should not affect our analysis since the corresponding uncertainties will re-weight these points when we fit for the slope and normalization. We also examine possible ways where we can reduce any possible ensemble bias using tracers that are sensitive to factors like the rest-frame wavelength coverage, and SED S/N.

Using a sample of mock galaxies at different redshifts and stellar masses, we span a range in S/N and wavelength coverage similar to what we see in the CANDELS data. For all of these galaxies we reconstruct the SFHs through SED fitting, along with uncertainties on log SFR(t) and log M_*(t) at each point in lookback time. We find that these estimated uncertainties at each point in lookback time trace of the bias in both log M_* and log SFR, and allow us to restrict ourselves to a minimally biased subset of galaxies for fits at any redshift. This is seen in Figure 6.

We consider a few different options for reconstructing the high-z SFR-M_* relation based on trajectory data from galaxies observed at later epochs (lower redshifts).

- Method (a): Use all available galaxies at lower redshifts / later epochs.

- Method (b): Use a subset of available galaxies with bounds on how far back in lookback time the galaxies are propagated, using SFHs from simulations to find the bounds.

- Method (c): Using the simulations, we find that bias in estimating quantities like SFR, M_* are traced closely by the pointwise uncertainties on these quantities estimated during SED fitting. These uncertainties depend on factors such as the S/N of the SED, rest frame wavelength coverage, and degeneracies of the SFH with dust and metallicity. Since this criterion is based on the uncertainties in log SFR and log M_* and not their actual values, this does not correspond to a selection in log sSFR. We can thus reduce the bias by limiting our analysis to a 'high confidence subsample that avoids significant biases. To ensure that the bias on M_* and SFR are less than 0.2 and 0.3 dex respectively, we use a threshold of 1 dex on the combined uncertainties ($\sqrt{\sigma_{logM_*(t)}^2 + \sigma_{logSFR(t)}^2}$) to select galaxies for analysis at different epochs. We find that varying the threshold in the range of 0.5 to 2 dex does not significantly impact the analysis. This selection is possible for both the direct fits and trajectories since the SED fitting procedure estimates uncertainties on M_* and SFR at each point in lookback time. For example, at $z = 1$ we consider the uncertainties for all galaxies that are directly fit at $z \sim 1$ and exclude those with $\sqrt{\sigma_{logM_*}^2 + \sigma_{logSFR}^2} > 1 dex$. Similarly, for the trajectories we consider all the galaxies at $z < 1$, and compute the uncertainties on their M_* and SFR after propagating them backwards in time to $z = 1$ before applying our selection criterion.

These three approaches are highlighted in Figure 6, where we show the error in estimating Stellar Mass and SFR for a sample from the simulations at $z \sim 1$ as a function of lookback time (i.e., trajectory run-time). In Figure 7 and Figure 8 we show this test run at a range of redshifts. We find that the dependence of the reconstruction on SED coverage or S/N is captured by the uncertainties, and thus using them as a tracer of the bias in our fits is an effective method that accomplishes both objectives: propagating robust SFR-M_* trajectories farther back in time while avoiding samples with large uncertainties due to bad SED coverage, low S/N, or bad fits. In Table 6 we show the median amounts of time that galaxies in our actual analysis are propagated backwards along their trajectories to be included in the analysis at a given redshift of interest. While this increases as we go to higher redshifts, most galaxies are propagated backwards by only about $15 - 30\%$ of their full SFHs at any epoch.

B. ROBUSTNESS OF TRAJECTORY - DIRECT FIT COMPARISON TO SAMPLE SELECTION

We use a KS test to compare the distributions of the distances from the best-fit SFR-M_* correlation that we get from the direct fits to the distances we get from galaxies observed at later epochs and propagated backwards in time
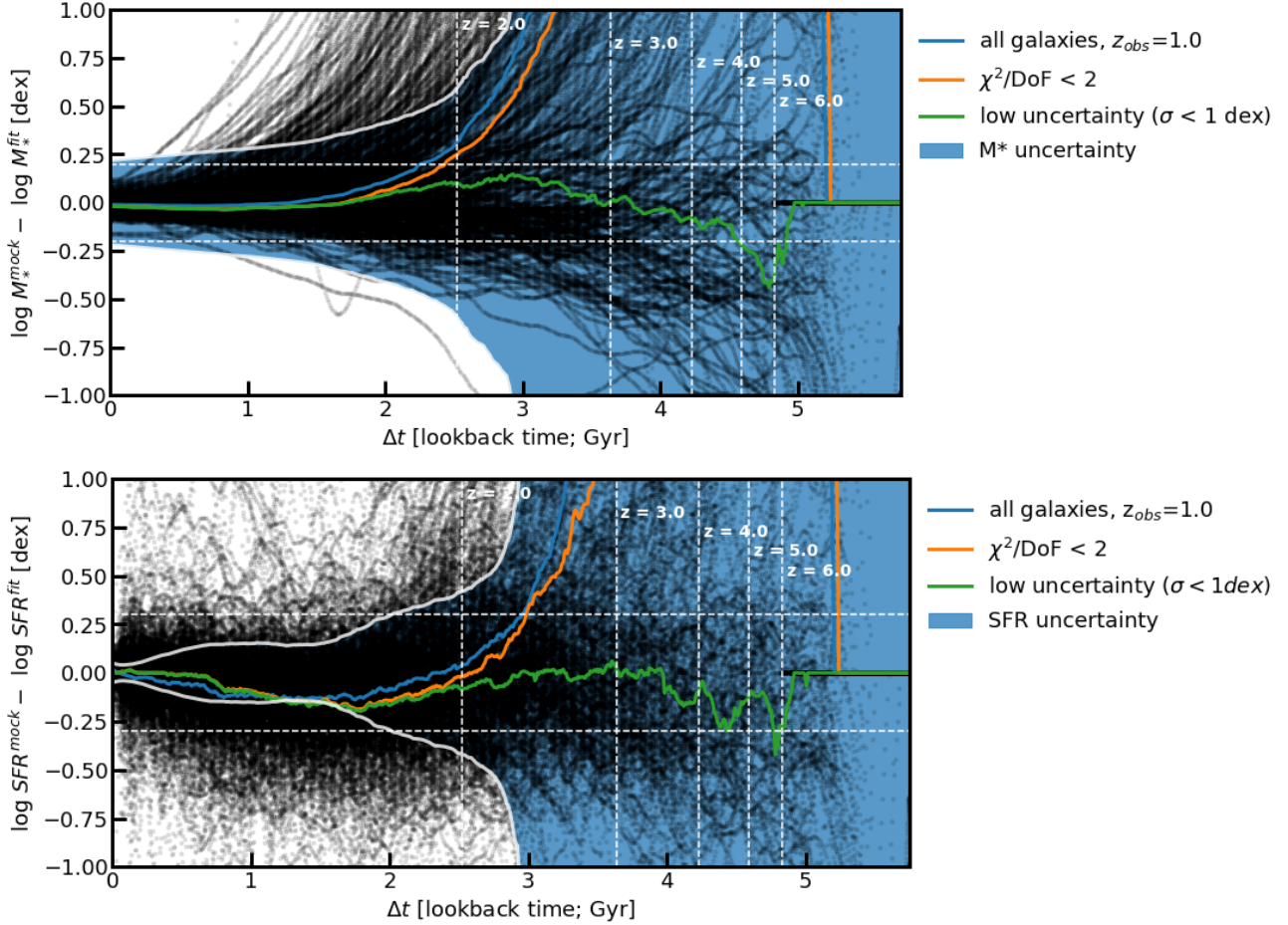
Figure 6. Reconstructed SFHs for an ensemble of mock galaxies at $z \sim 1$ using the Dense Basis SED fitting method. Top panel shows the estimated bias (model - SED fitting estimate) and uncertainties (from SED fitting) in stellar mass as a function of lookback time, and the bottom panel shows the same for SFR.

In each plot, the **blue solid line** is pointwise median for all galaxies in the sample. This generally grows with time since our smooth SFHs can go to 0 while the mock SFHs generally go to some small nonzero value, which leads to a one-sided bias. However, since this bias would affect both SFR and M$_*$ identically, it simply shifts points along the diagonal, and shouldn't affect our estimates of slope and normalization. The **orange line** is pointwise median for the subsample of 'good fits' (method a) and the **green line** is pointwise median excluding contributions from galaxies which have large uncertainties ($\sigma \log(SFR(t), M_*(t)) > 1\ dex$). (method c, using the uncertainties to avoid SFHs with biases).

The white horizontal line shows the average uncertainties on SFR ($\sim$0.3 dex) and M$_*$ ($\sim$0.2 dex), for comparison. The **white vertical lines** indicate redshifts instead of lookback times. (method b would be truncating our trajectories to where the orange line hits our tolerance bias of the white horizontal lines, here around $\Delta t \sim 2.5 Gyr$.) **The blue shaded region + white solid line shows the uncertainty on SFR-M$_*$ estimates, showing that the uncertainties closely follow the bias. Additionally, we see that for the full ensemble of galaxies (sample (a)), the uncertainties are larger than the bias and thus should not affect a statistical analysis that takes the uncertainties into account.**

along their trajectories. The similarity of the two distributions in addition to our previous validation suggests that the reconstructions are robust and that the effects due to mergers do not significantly affect our analysis of quantities like the slope and normalization of the SFR-M$_*$ correlation. To ensure that the KS test is not affected by possible systematics arising from sample selection, we perform the test on a few different samples at each redshift:

- The full distribution of distances from both direct fits and trajectories, out to a distance of 0.4 dex from the best-fit correlation.

- The distribution of distances corresponding to the sample of galaxies with uncertainties < 1 dex, out to a distance of 0.4 dex from the best-fit correlation.
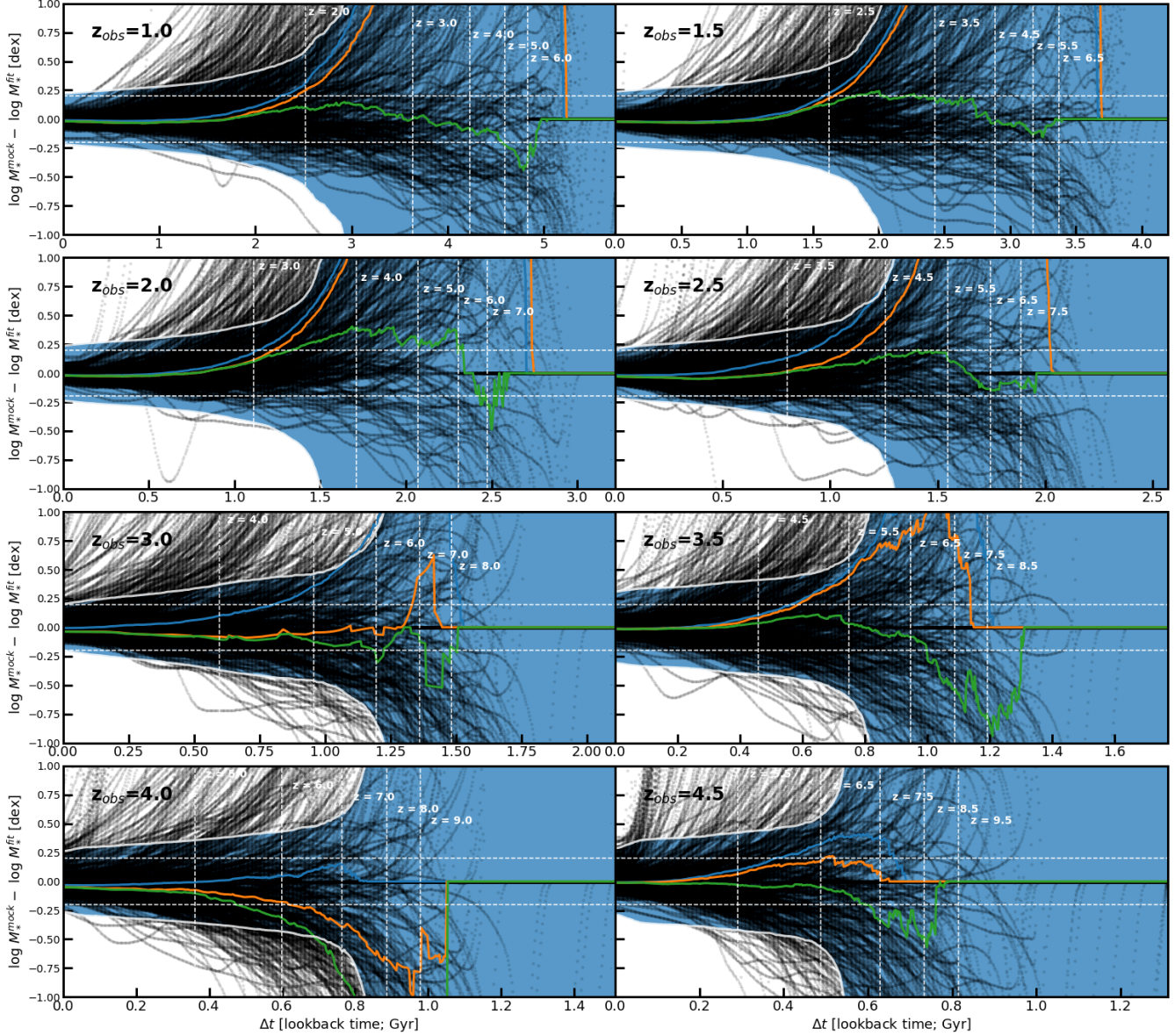
Figure 7. Same as figure 6 for Stellar Mass, repeated across a range of redshifts.

- The distribution of distances corresponding to the sample of galaxies with uncertainties < 1 dex, out to $1 \times$ the observed scatter from the best-fit correlation.

- The distribution of distances corresponding to the sample of galaxies with uncertainties < 1 dex restricted to the mass range where the direct fits have good statistics, out to $1 \times$ the observed scatter from the best-fit correlation. To find the lower mass threshold, we find the minimum mass at for which the $median(M_*^{directfits}) - median(m_*^{trajectories}) < 0.1$ dex .

The results of our KS test are consistent at all redshifts (p-value $> \alpha$), as summarized in table 7. The significance level for each test is $\alpha = 0.05$. However, since we are performing a family of comparisons to test a single hypothesis, we need to control for the increased probability of false positives. To this end, we use adjusted significance levels of $\alpha' = \alpha/N = 0.05/6 = 0.0083$ using a Bonferroni correction (Goeman & Solari 2014), or $\alpha' = 1 - (1 - \alpha)^{1/N} = 1 - (1 - 0.05)^{1/6} \approx 0.0085$ using the more conservative Sidak correction (Šidák 1967). This choice of correction does not affect our results since our lowest p-value is 0.02. Since our results remain consistent across this broad range of tests, we can not reject the null hypothesis that the two distributions are the same. While this does not completely
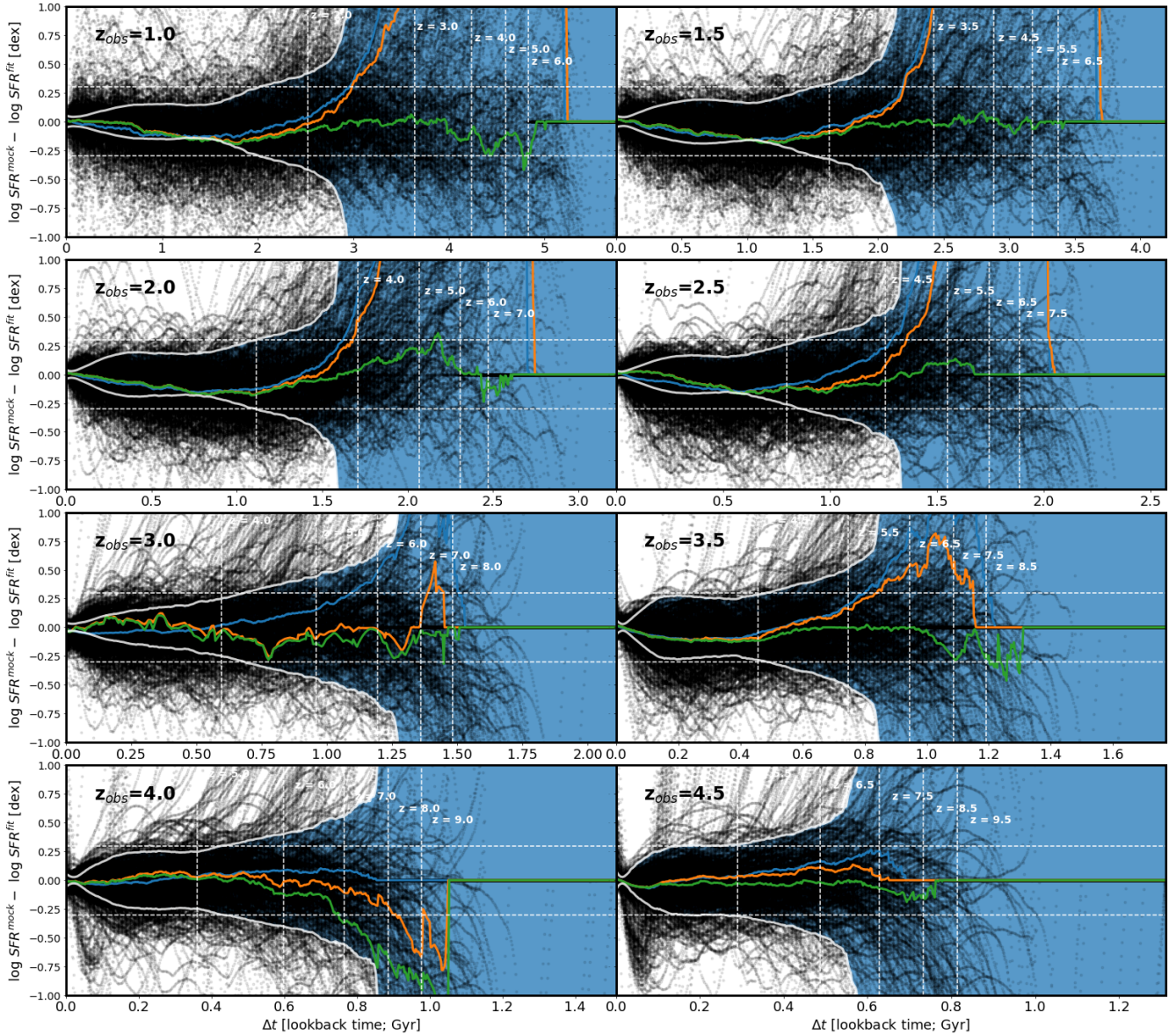
Figure 8. Same as figure 6 for Star Formation Rate, repeated across a range of redshifts.

rule out the possibility that the two distributions are different, the probability of this being the case is lower than $\alpha = 5\%$.

C. EFFECTS OF UVJ PRE-SELECTION

The Star Formation Rates of actively star forming galaxies are tightly correlated with their Stellar Masses across a range of redshifts, with $\geq 68\%$ of such galaxies found within a narrow range of a single best-fit line. However, when galaxies enter periods of quiescence or undergo starbursts, they make excursions from this correlation. Improperly taking these galaxies into account (both by failing to exclude them or by being too rigorous in excluding them, thereby excluding some star forming galaxies as well) could lead to biases in our estimates of slope and normalization for the SFR-M$_*$ correlation. To mitigate this issue, we used an iterative robust fitting routine that excludes outliers (Holland & Welsch 1977), which effectively re-calibrates to the data at each redshift slice to identify which points could be outliers. This avoids the use of pre-determined conditions for when galaxies are quiescent, since we find that our star-formation histories allow us to robustly distinguish between galaxies with low SFRs throughout their lifetime and galaxies that are experiencing a rapid fall in their SFR.

$z_{analysis}$	Median Δt that galaxies are propagated [Gyr]	Median z_{obs} [$t(z_{analysis})$- Median(Δt)]	Δt between $z = 0.5$ and $z_{analysis}$ [Gyr]
1	0.8	0.82	2.7
2	0.8	1.58	5.1
3	1.3	1.89	6.3
4	1.4	2.21	6.9
5	2.1	1.98	7.3
6	2.2	2.07	7.5

Table 6. Median amounts of time that sub-samples of low-uncertainty galaxies (observed at z_{obs}) are propagated along their redshifts to reach desired redshifts at which we have performed our analysis ($z_{analysis}$). We see that the amount of time grows as we go to higher redshifts, but is much less than the time interval between the lowest redshift and the redshifts of interest (the maximum amount of time a galaxy can be propagated in our current analysis). This indicates that most galaxies analyzed at a given redshift come from the vicinity of that redshift, rather than being propagated backwards all the way from $z \sim 1$.

redshift of analysis	p-values (full)	p-values (low uncert.)	p-values (low uncert., 1 × scatter)	p-values (low uncert., 1 × scatter, M$_*$ threshold)
1	0.14	0.85	0.91	0.18
2	0.04	0.10	0.11	0.15
3	0.07	0.02	0.15	0.80
4	0.88	0.67	0.56	0.30
5	0.56	0.20	0.50	0.03
6	0.52	0.98	0.82	0.56

Table 7. P-values corresponding to the KS test comparing the distributions of distances from the best-fit SFR=M$_*$ correlation for the direct fits and trajectories. The first column (full) shows the p-values comparing the distribution across all M_* within 0.4 dex of the best-fit line. The second column (low uncert.) repeats this analysis for the subsample of galaxies used for trajectories that have low uncertainties at the redshift of interest. The third column (low uncert., 1 × scatter) uses a threshold of 1 × the observed scatter at each redshift instead of a fixed threshold of 0.4 dex. The fourth column (low uncert., 1 × scatter, M$_*$ threshold) performs the KS test on a further reduced dataset restricted to the mass range where the direct fits have good statistics so that stellar mass effects don't enter into our comparison. The p-values for all these comparisons are $> \alpha$, indicating that the two distributions are not different to a statistically significant level.

However, in order to better compare our results to literature that uses a pre-selection step to select star forming galaxies (see for example Schreiber et al. (2015), where they use an optical selection criterion since even quiescent galaxies could still show residual IR emission due to a warm ISM), we adopt the UVJ selection criterion from Williams et al. (2009). We use the rest-frame U-V and V-J colors derived by Pacifici et al. (2016). In brief, Pacifici et al. (2016) use a large library of model SEDs to fit all available photometric data and derive median values and uncertainties of the rest-frame colors for each galaxy in the sample. The library is generated by combining the output of a semi-analytical model of galaxy formation with models of the stellar and gas emission and the attenuation by dust (see Pacifici et al. (2012) for more details). The $z \sim 1$ sample with the selection criteria is shown in Figure 9.

Refitting the correlation to determine the slope and normalization does not significantly change our results, since most of the points excluded by the selection criterion would be classified as outliers by our algorithm. The fractional changes to slope and normalization at different redshifts are: -1.48 %, -0.05 %, 0.1 % , 0.78 %, 0.4 % , and 0.02 %, while the fractional changes to normalization are: 0.01 , 0.011, 0.025, 0.033, 0.011, and 0 dex at z=1,2,3,4,5,6 respectively.
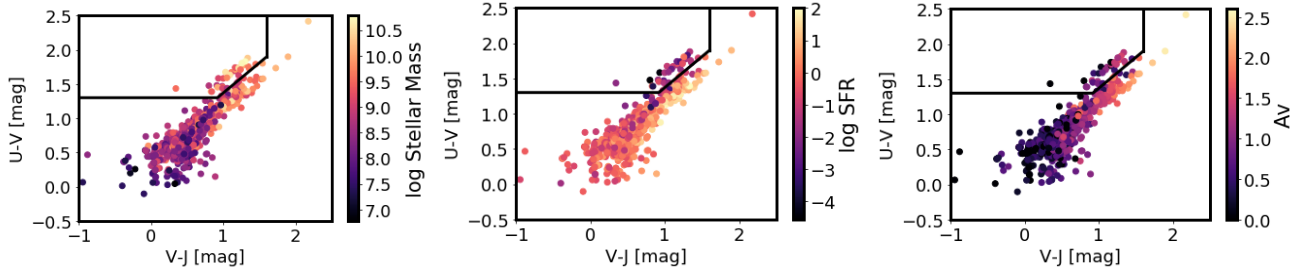
Figure 9. UVJ diagrams for the sample at $z \sim 1$. Rest frame U-V and V-J colors are computed through SED fitting, using the best-fit model template as a prior. The diagram shows the expected correlations with SFR and dust.

D. USING NONPARAMETERIC REGRESSION METHODS TO QUANTIFY THE SFR-M_* CORRELATION

In our analysis, we assume a linear relation between log SFR and log Stellar Mass and fit for its slope and normalization. However, it is not necessary that the correlation be linear. Indeed, Schreiber et al. (2015) and Whitaker et al. (2014) find that the relation flattens out at the high mass end. Using an F-test, we checked to see if a quadratic fit is statistically preferred over a linear one, and found this not to be the case.

However, this is not enough to prove the linearity of the relationship. While some studies quantify the correlation between SFR and Stellar Mass by finding the effective SFR in bins of Stellar Mass, or even by binning perpendicular to the correlation, methods involving binning potentially suffer from effects due to bin size and the locations of bin centers. Here, we use LOWESS (Cleveland 1979) a nonparametric regression technique that uses local weighting to create a smooth nonparametric estimate of the correlation. We also create a similar set of plots using Gaussian Process Regression (GPR) (Rasmussen & Williams 2006) to estimate the correlation at each point in M_*. From Figure 10, we can see that the relation is indeed linear at the low mass end, closely matching our best-fit. While this alone isn't enough to state that the correlation is linear, it is certainly consistent within uncertainties with our best-fit linear relation.

From Figure 10 we also see that the direct fits and trajectories agree extremely well in the high-mass regime at redshifts where both sets have significant statistics. However, the high mass end has a slightly higher slope than the low-mass end at $z \sim 1 - 3$. From this, we conclude that when we fit the direct fits and trajectories, possible discrepancies may arise due to the different mass ranges over which they are being fit.
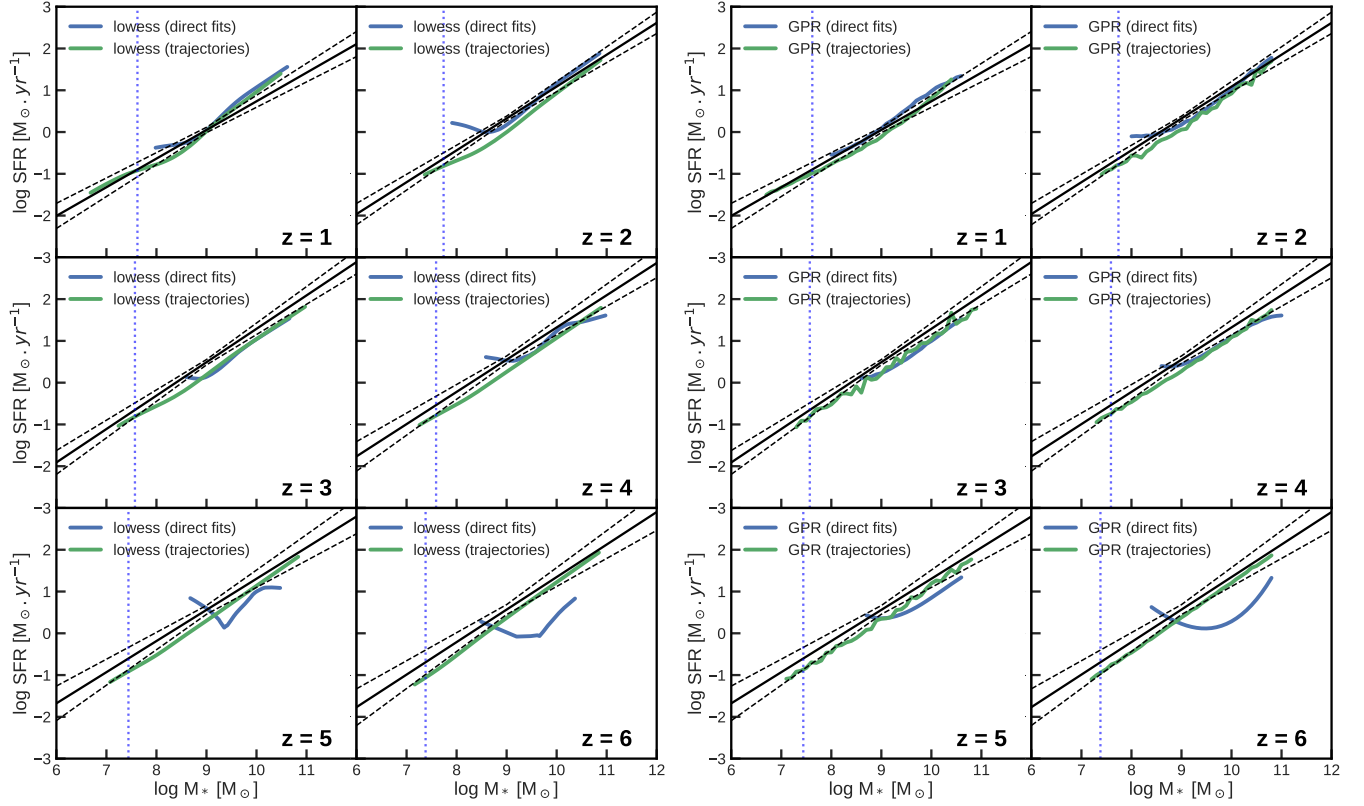
Figure 10. Nonparametric regression using Locally Weighted Scatterplot Smoothing (LOWESS) and Gaussian Process Regression (GPR) performed on the direct fits (blue lines) and trajectories (green lines) datasets at different redshifts. The curves are limited to the 5th to 95th percentile in Stellar Mass for each dataset with GPR, and from the 1st to 95th percentile for LOWESS. The curves corresponding to direct fits at $z > 4$ are not reliable due to the small number of points available at those redshifts (see Figure 4. Dotted blue lines show the 10th percentile of the M_* distributions at different epochs as reported in Table. 3.)