

How galaxies acquire their stellar mass at high redshift: High star formation efficiencies and the relative roles of dust and initial mass function

Hao Fu^{1,*}, Francesco Shankar², Fabio Fontanot³, Andrea Lapi⁴, Feng Yuan¹, Mohammadreza Ayromlou⁵, Daniel Roberts², Lumen Boco⁶, Nicola Menci⁷, Emiliano Merlin⁷, Laura Pentericci⁷, and Mengyuan Xiao⁸

¹ Center for Astronomy and Astrophysics and Department of Physics, Fudan University, Shanghai 200438, China

² School of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

³ INAF – Astronomical Observatory of Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy

⁴ SISSA, Via Bonomea 265, 34136 Trieste, Italy

⁵ Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany

⁶ Universität Heidelberg, Zentrum für Astronomie, Institut für theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany

⁷ INAF Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Rome, Italy

⁸ Department of Astronomy, University of Geneva, Versoix, Switzerland

Received 30 December 2025 / Accepted 21 May 2026

ABSTRACT

Aims. JWST has measured an unprecedented abundance of galaxies above redshift $z \gtrsim 4$ –5, whose formation and evolution are still difficult to reconcile within traditional galaxy evolution models in a Lambda Cold Dark Matter (Λ CDM) framework. Here we present a study on the star formation histories of these high-redshift galaxies between $z \approx 5$ –12 via a cutting-edge data-driven semi-empirical model that uses the observed ultra-violet (UV) luminosity functions (LFs) as input to retrieve star formation rates (SFRs), naturally bypassing any uncertain modelling of cooling, feedback and/or stochastic processes.

Methods. Galaxy stellar masses are progressively built in time by integrating their SFRs assigned along their progenitor haloes via the SFR-halo accretion rate relation, derived from abundance matching between the input observed UV LFs with the dark matter halo accretion rate distributions at each redshift. Our original method reverse-engineers empirical estimates of the high- z galactic SFRs directly from observations via abundance matching rather than fitting observed LFs with parametric star formation efficiencies (SFEs). This makes the SFEs a full prediction of the model rather than a tuned input, serving as a solid baseline to test burstiness, dust attenuation, or initial mass function variations.

Results. Our approach reproduces the total stellar mass function, the large-scale clustering, and the star-forming main sequence. We find that massive galaxies grew their stellar mass with a bursty star formation at $z \sim 9$ –10, broadly in agreement with the star formation histories inferred from spectral energy distribution fitting, with the SFE reaching high peaks of 0.8–0.9 at $z > 9$ and lowering to standard values of 0.2–0.3 below $z \lesssim 9$. We find that the presence of dust could enhance the predicted SFRs at $z \lesssim 8$, better reproducing the observed SFRs of massive dusty galaxies, and increase the SFEs to values close to or even above unity at $z \gtrsim 8$. Finally, switching to top-heavy initial mass functions reduces the SFEs by a factor of 2–3, highlighting the need for a variable initial mass function as an inevitable ingredient in the evolution of galaxies at high redshifts to avoid unphysical SFEs, especially in the presence of dust.

Key words. galaxies: abundances – galaxies: evolution – galaxies: star formation

1. Introduction

The formation and evolution of galaxies are still debated questions today. In a Lambda Cold Dark Matter (Λ CDM) Universe galaxies grow in stellar mass via both in situ and ex situ processes. The in situ growth is regulated by the star formation rate, i.e. the conversion of cold gas into stars, fuelled by the gas accretion of the host dark matter haloes. The ex situ growth is instead controlled by the rate of mergers with other galaxies, which in a Λ CDM Universe is the inevitable consequence of the mergers between the host dark matter haloes. Many studies have shown that at redshifts $z \lesssim 3$ less massive galaxies (with stellar mass $M_\star \lesssim 10^{11} M_\odot$ at $z \sim 0$) formed their stellar mass

mainly via star formation, whereas more massive galaxies had more contribution from mergers (e.g. van Dokkum et al. 2010; Shankar et al. 2015; Buchan & Shankar 2016; Pillepich et al. 2018a; Grylls et al. 2019, 2020; Fu et al. 2022, 2024; Eisert et al. 2023).

At progressively higher redshifts the new observations from the James Webb Space Telescope (JWST) have revolutionised our view of galaxy formation and evolution. Firstly, JWST has revealed a large number of star-forming, ultraviolet (UV)-bright galaxies up to $z \sim 17$ and possibly beyond, with abundances exceeding pre-JWST observations by up to an order of magnitude or more (e.g. Bouwens et al. 2023; Finkelstein et al. 2023, 2024; Adams et al. 2024; McLeod et al. 2024; Robertson et al. 2024; Pérez-González et al. 2025). Secondly, a large population of active galactic nuclei (AGN) powered by very massive

* Corresponding author: hao.fu@fudan.edu.cn

black holes have also been inferred from JWST observations (e.g. Matthee et al. 2024; Akins et al. 2025; Scholtz et al. 2025; Juodžbalis et al. 2026). Last but not least, JWST has also unveiled a population of extremely massive galaxies at $z \gtrsim 3$ that were completely missed in previous observations and are challenging to reconcile within current traditional models (e.g. Donnan et al. 2024; Matsumoto et al. 2024; McGaugh et al. 2024; Weibel et al. 2024; Shuntov et al. 2025a). For example, Glazebrook et al. (2024) identified a massive quiescent galaxy, ZF-UDS-7329, at redshift $z = 3.25$ with inferred stellar mass $M_\star \sim 10^{11.3} M_\odot$, speculating that it formed most of its stellar mass at $z \sim 11$ in a burst of star formation. They discuss that this scenario is in tension with current models of galaxy formation within the Λ CDM framework, pointing possibly towards a different stellar initial mass function (IMF), or changes in the underlying stellar population and feedback modelling. Indeed, evidence for a radially varying IMF in local massive galaxies is mounting (e.g. Cappellari et al. 2006; Conroy & van Dokkum 2012; La Barbera et al. 2016; Bernardi et al. 2019; Marsden et al. 2022; Lu et al. 2024).

More generally, the large number of UV-bright galaxies inferred from JWST observations has been explained in a number of different ways from the theoretical point of view. Some popular explanations rely on stochastic star formation to allow for epochs of larger-than-average galaxy growth rate (e.g. Shen et al. 2023; Sun et al. 2023; Gelli et al. 2024), or reducing the content of dust (e.g. Ferrara 2024), lowering stellar metallicity or assuming a top-heavy IMF to boost the UV emission for any given episode of star formation (e.g. Inayoshi et al. 2022; Trinca et al. 2024; Yung et al. 2024a), assuming density-modulated star formation efficiencies (e.g. Somerville et al. 2025), and/or a feedback-free scenario (e.g. Dekel et al. 2023). To explain the formation of massive galaxies at $z > 5$ in particular, some groups invoked the extreme value statistics (Lovell et al. 2023; Carnall et al. 2024; Jespersen et al. 2025; Enríquez-Vargas et al. 2026) to show that the existence of massive galaxies (with stellar mass $M_\star > 10^{11} M_\odot$) observed between $3 < z < 5$ may be accommodated within Λ CDM but only assuming very high star formation efficiency (SFE; $\epsilon_\star = 0.8-1$). Alternatively, a bottom-heavy IMF could also increase the stellar mass by boosting the fraction of lower mass stars (e.g. Lapi et al. 2024; Yung et al. 2024a; Mauerhofer et al. 2025; Fontanot et al. 2026). Alternative possibilities for boosting the formation of galaxies at early epochs may be found either in improving the modelling of the halo accretion histories within Λ CDM (e.g. ellipsoidal collapse) or in varying the underlying cosmological model, such as allowing for dynamical dark energy models (e.g. Boylan-Kolchin 2023; Menci et al. 2024; Sokoliuk 2025; Fakhry et al. 2026; Menci et al. 2026) or modified gravity models (e.g. McGaugh et al. 2024).

In the context of a Λ CDM Universe, some groups have attempted to model the galaxy star formation histories and implied UV luminosities via semi-empirical techniques. This approach bypasses the ab initio modelling of the baryonic physics behind star formation in galaxies, which is instead guided by observational data. For example, Kar et al. (2026) and Yung et al. (2025), along the lines of Moster et al. (2018) and Behroozi et al. (2019), assumed a parametric SFE of the type $\epsilon_\star(M_h, z)$ characterised by a number of free parameters defining its variation as a function of host halo mass and redshift. They concluded that the number density and spatial distribution of UV galaxies can be reproduced within Λ CDM with relatively standard SFEs of $\sim 20-30\%$, but steadily increasing at higher redshifts and without the need for very large scatters of up to ~ 1 dex

or more in the star formation rate-halo mass relation, as previously inferred by other more comprehensive theoretical models (e.g. Dekel et al. 2023; Mason et al. 2023; Nikopoulos & Dayal 2024).

In this work we made use, extending it to higher redshifts, of the semi-empirical approach recently devised by our group in two semi-empirical models, DECODE (Fu et al. 2025a,b) and TOPSEM (Boco et al. 2023), which have been successful in reproducing the stellar mass functions, star formation histories, and bulge-to-total distributions of galaxies at $z \lesssim 4-5$. Our method relies on extracting the star formation rate-halo accretion rate (SFR-HAR) relation from abundance matching between the SFR and HAR functions, where the former is directly inferred from observations and the latter is theoretically predicted by N-body simulations. There are several key advantages of using this data-driven technique. First, it does not require heavy parametrisations of the SFE, which is instead a natural prediction of the abundance matching procedure. Second, it allows us to extract the star formation histories by tracking the SFRs along the host halo progenitor histories. Third, the SFRs in input are directly inferred from observational data and as such they can be considered an effective balance between any feedback and any ex situ or in situ process (e.g. merger, disc instability, clumpy accretion) that may reduce or enhance the star formation in the galaxy. Finally, the SFRs in input do not require any modelling in terms of stochasticity or feedback, and can in turn be effectively adopted to test the degree of SFE, dust attenuation, or variations in IMF, as we carry out below.

The aim of this paper is twofold. On the one hand, we computed the star formation histories of galaxies along their progenitors to compare with direct observations. On the other hand, we integrated them to infer the implied stellar masses, in particular of massive galaxies, global stellar mass functions, and clustering to pin down the SFEs and impact of including dust attenuation or a variable IMF.

This paper is structured as follows. In Section 2, we describe the methodology followed in our model. In Section 3, we test the SFR-HAR assumption at high redshifts by applying it to the TNG simulation. In Section 4, we show our results on the galaxy abundances, stellar mass-halo mass connection and star formation histories. We also discuss the effect of choosing a different IMF and applying a dust correction to the SFRs. Finally, in Section 5 and 6, we discuss our results and draw our conclusions. In this work, we adopt the Λ CDM cosmology with best-fit parameters from Planck Collaboration VI (2020) (i.e. $(\Omega_m, \Omega_\Lambda, \Omega_b, h, n_s, \sigma_8) = (0.31, 0.69, 0.049, 0.68, 0.97, 0.81)$) and a Chabrier (2003) stellar IMF.

2. Methods

In this Section, we describe the implementation of our model DECODE¹ (Discrete statistical sEmi-empiriCal mODEl) and our approach to grow and merge galaxies following the dark matter halo assembly histories. We built up galaxies following their progenitors via input scaling relations between the galaxy SFR and HAR. In the first two papers of the series (Fu et al. 2025a,b) we accurately tested the SFR-HAR abundance matching at redshift $z \lesssim 5-6$. In this paper, we focus on the galaxy star formation rates at high redshifts and how they compare to the JWST spectral energy distribution (SED)-inferred star formation histories at redshift $z \gtrsim 5$.

¹ <https://github.com/haofuastro/DECODE2>

2.1. Dark matter halo merger tree

We based our analysis on a catalogue of $n = 5 \times 10^5$ central dark matter haloes with mass above $M_h > 10^{10.5} M_\odot$ at each redshift of interest. The choice of the volume and cut in halo mass represents a good balance among computational efficiency, statistics and mass resolution for the stellar mass range of interest. The cut in halo mass in particular ensures a completeness in stellar mass down to $M_\star \gtrsim 3 \times 10^9 M_\odot$. These haloes were randomly extracted from the Tinker et al. (2008) halo mass function at any redshift above the halo mass cut-off. We ignored the surviving satellite subhaloes whose abundance at these high redshifts is less relevant. The mass accretion history of each dark matter halo and their merger trees were computed via the SatGen analytical code from Jiang et al. (2021), based on the Parkinson et al. (2008) algorithm fitted to the outputs of the MultiDark Planck N-body simulation. Each of these parent dark matter haloes comes with a mass accretion history along with the infall redshift of the satellite subhaloes and their accretion history before infall. We also checked that these halo accretion tracks are on average well consistent with those from high-redshift dark matter-only simulations, such as GUREFT (Yung et al. 2024b). On top of this dark matter framework, we self-consistently grew central galaxies by assigning to them a SFR following the growth of the host haloes.

2.2. Abundance matching

The core of our model is based on the abundance matching between the galaxy SFR and host dark matter HAR. To compute the SFR-HAR relation, we performed the abundance matching between the observed SFR function and the theoretical HAR function (we used Equation 37 in Aversa et al. 2015, see also Fu et al. 2022, Boco et al. 2023, Fu et al. 2024 and Fu et al. 2025a,b). In the abundance matching we assumed a scatter of 0.4 dex in SFR at fixed HAR, as suggested by our analysis from the TNG simulation. We checked that by varying the value of the scatter within a reasonable range (0.3–0.5 dex) our results do not alter appreciably, as also discussed in Fu et al. (2025a) and Fu et al. (2025b), with stellar mass assemblies altering by less than $\lesssim 0.05$ dex.

In our abundance matching, we input the theoretical HAR function (from the results of the SatGen algorithm) and the observed SFR function (gathered from multiple datasets). For the former, we sampled the distributions from our mock halo catalogue at each redshift, while for the latter, we combined the most recently observed UV galaxy luminosity function (LF) at high redshifts (see Appendix A for details). We converted UV luminosities to SFRs via the calibration factor from Kennicutt (1998) which yields the following formula (e.g. Madau & Dickinson 2014)

$$\text{SFR}_{\text{UV}} = k_{\text{UV}} \cdot L_{\text{UV}}, \quad (1)$$

with $k_{\text{UV}} = 2.5 \times 10^{-10} M_\odot \text{yr}^{-1} L_\odot^{-1}$. We note that the Madau & Dickinson (2014) conversion factor is based on a fixed star formation history and solar metallicity, which might systematically overestimate the SFRs up to 1 dex (e.g. Wilkins et al. 2020). However, works on the mass-metallicity relation suggest a metallicity in the range of $Z \in [0.1 Z_\odot, Z_\odot]$ for galaxies with stellar mass $M_\star \gtrsim 10^9 M_\odot$ from both JWST emission-line analyses (e.g. Curti et al. 2023; Nakajima et al. 2023; Trump et al. 2023; Merlin et al. 2025) and cosmological simulations (e.g. Wilkins et al. 2023). We checked that in the most extreme case of lowest metallicity (i.e. $Z = 0.1 Z_\odot$), the UV-SFR calibration is expected to vary by up to $\sim 20\%$, as suggested by stellar pop-

ulation synthesis models (e.g. Bicker & Fritze-v. Alvensleben 2005; Wilkins et al. 2019). Accordingly, the resulting shift in SFR would be less than < 0.1 dex, corresponding to a ~ 0.1 dex shift in stellar mass. For the most massive galaxies, with $M_\star \gtrsim 10^{11} M_\odot$, both observations and models suggest a near-solar metallicity (e.g. Merlin et al. 2025), implying a negligible alteration in their stellar mass growth.

Hereafter, we adopt the UV SFR as representative of the total SFR, along with a Chabrier (2003) IMF, as a reference model. We also explore the consequences of considering the dust correction in observed luminosities and adopting different IMFs in Sections 4.7 and 4.8.

2.3. Growing galaxies within DECODE

We grow our galaxies by assigning to each galaxy a SFR at each redshift via the SFR-HAR relation, following the mass accretion history of the host dark matter haloes. Starting our simulation at redshift $z \sim 12$, we then integrate the SFRs across cosmic time

$$M_\star(t) = \int_0^t \dot{M}_{\text{SFR}}(t')(1 - \mathcal{R})dt', \quad (2)$$

where $\mathcal{R}$ is the stellar mass loss fraction that returns into the interstellar medium. We note that the recycling factor $\mathcal{R}$ is dependent on the IMF which produces variations between ~ 0.3 – 0.6 (see, e.g. Segers et al. 2016; Yu & Wang 2016; Hopkins 2018). In what follows, we adopt $\mathcal{R} = 0.45$, which is a typical value for a Chabrier IMF (e.g. Reimers 1975; Renzini & Fusi Pecci 1988; Girardi et al. 2000; Pietrinferni et al. 2004; Romano et al. 2010; Nomoto et al. 2013; Vincenzo et al. 2016). We also clarify that we do not push our simulation beyond $z \gtrsim 12$ because the UV LF is not observationally well constrained at those redshifts, even though there are works that have provided preliminary estimates up to $z \sim 25$ (e.g. Donnan et al. 2024; Harikane et al. 2025; Pérez-González et al. 2025; Weibel et al. 2026).

2.4. Galaxy mergers

We add the contribution from mergers to the total galaxy stellar mass by following the merger histories of the host dark matter subhaloes. In particular, we assign satellite galaxies with a stellar mass at infall via the typical stellar mass corresponding to the halo mass of the host subhalo on the stellar mass-halo mass ($M_\star - M_h$) relation of the central counterpart, assuming that satellite galaxies align with the scaling relations of central galaxies at the time of infall. We assume that galaxies merge with a time delay with respect to the host subhaloes following the merger timescales described in Section 3.4 of Fu et al. (2022) corrected by a factor of ~ 1.5 at redshift $z \gtrsim 5$. These timescales are fitted to reproduce the number densities of surviving satellite subhaloes of the Millennium simulation up to redshift $z \sim 11$ – 12 .

For the purposes of this paper, we do not discuss in detail the contribution from mergers, even though we include them in the overall stellar mass assemblies of our galaxies for completeness. We have checked that mergers contribute to less than ~ 5 – 10% of the total stellar mass growth in all our mock galaxies at $z \gtrsim 5$ – 56 , resulting in a minimal impact on the stellar mass function. We note that this estimate is nicely consistent with the recent observational work by Calabrò et al. (2026), who claimed that mergers can contribute up to 5 – 10% at $5 < z < 14$ in a spectroscopic sample of 1233 JWST galaxies. We will dedicate a separate work on the prediction of merger rates of galaxies at high redshifts (Fu et al. in prep.).

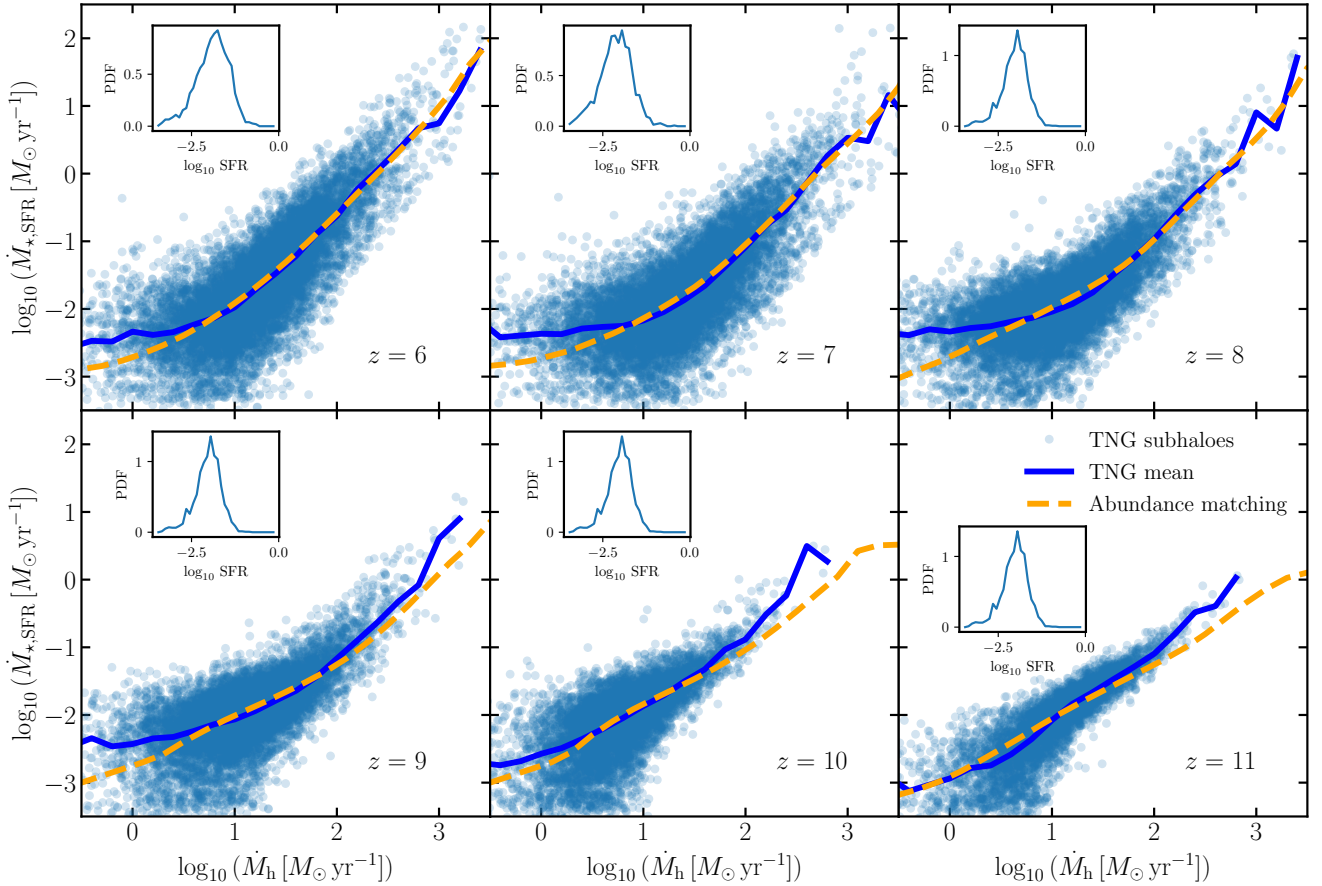


Fig. 1. Distribution of the TNG subhaloes on the star formation rate-halo accretion rate plane (blue dots), and median star formation rate-halo accretion rate relations from the TNG (blue lines) and computed using the TNG’s inputs via the abundance matching described in Section 2.2 (orange dashed lines) at redshifts $z = 6$ –11. The subplots in each panel show the distribution normalised to unity of the star formation rate in a halo accretion rate bin of 0.2 dex width centred at $\dot{M}_h \sim 10 M_\odot \text{ yr}^{-1}$, and the distributions in other bins follow a similar behaviour.

3. Verifying the assumption of a monotonic average relation between SFR and HAR at high redshift

Before applying our model to predict the star formation histories of high-redshift massive galaxies, following F25 we tested whether the assumption of a monotonic relation between the SFR and the HAR is still expected in cosmological galaxy evolution models even at $z \gtrsim 6$. For this purpose, we made use of The Next Generation (TNG henceforth) simulation to test whether a state-of-the-art cosmological hydrodynamic simulation is predicting a monotonic relation between SFR and HAR at $z > 6$ as it does at lower redshifts. We stress that this analysis is not aimed at calibrating our model on the TNG simulation which employs its own recipes for the baryonic physics and is calibrated on low-redshift data itself, but only to test the reliability of the SFR-HAR correlation hypothesis in a self-consistent model like the TNG. We note that a whole body of theoretical work supports a close connection between galactic SFR and host HAR (e.g. Dekel et al. 2009; Bournaud et al. 2015; Dou et al. 2025; Waterval et al. 2025).

3.1. The TNG simulation

We make use of the cosmological TNG100 simulation, a component of the IllustrisTNG project² (e.g. Nelson et al. 2019;

Pillepich et al. 2018b,a; Marinacci et al. 2018; Springel et al. 2018). The TNG simulations are performed via the moving-mesh AREPO code (Springel 2010), employing subgrid physics of galaxy formation such as gas cooling, star formation, stellar evolution, and AGN feedback (e.g. Pillepich et al. 2018b; Weinberger et al. 2017). The TNG simulations include different volume box sizes and mass relations designed to address specific physical processes. Below, we use the TNG100 simulation run on a 100 Mpc box on a side, which represents a good balance between volume size and mass resolution for the purposes of our test.

3.2. The SFR-HAR mapping

Figure 1 shows the SFR-HAR relation above redshift $z \gtrsim 6$, as extracted from the TNG simulation. We find that galaxy SFRs and host HARs are indeed connected via a monotonically increasing relation with a symmetric dispersion around the mean, which can be well approximated by a Gaussian scatter in our abundance matching procedure (as visualised in the inset of each panel in Figure 1). We also show that, by taking the TNG’s SFR function, HAR function, and scatter as input, our abundance matching (orange dashed lines) is able to reproduce the mean SFR-HAR relation of the simulation out to redshift $z \sim 11$ (blue solid lines), as shown by Figure 1. Having confirmed the plausibility of a mean SFR-HAR monotonic relation in a Λ CDM Universe, along with the validation of our abundance matching

² <https://www.tng-project.org/>

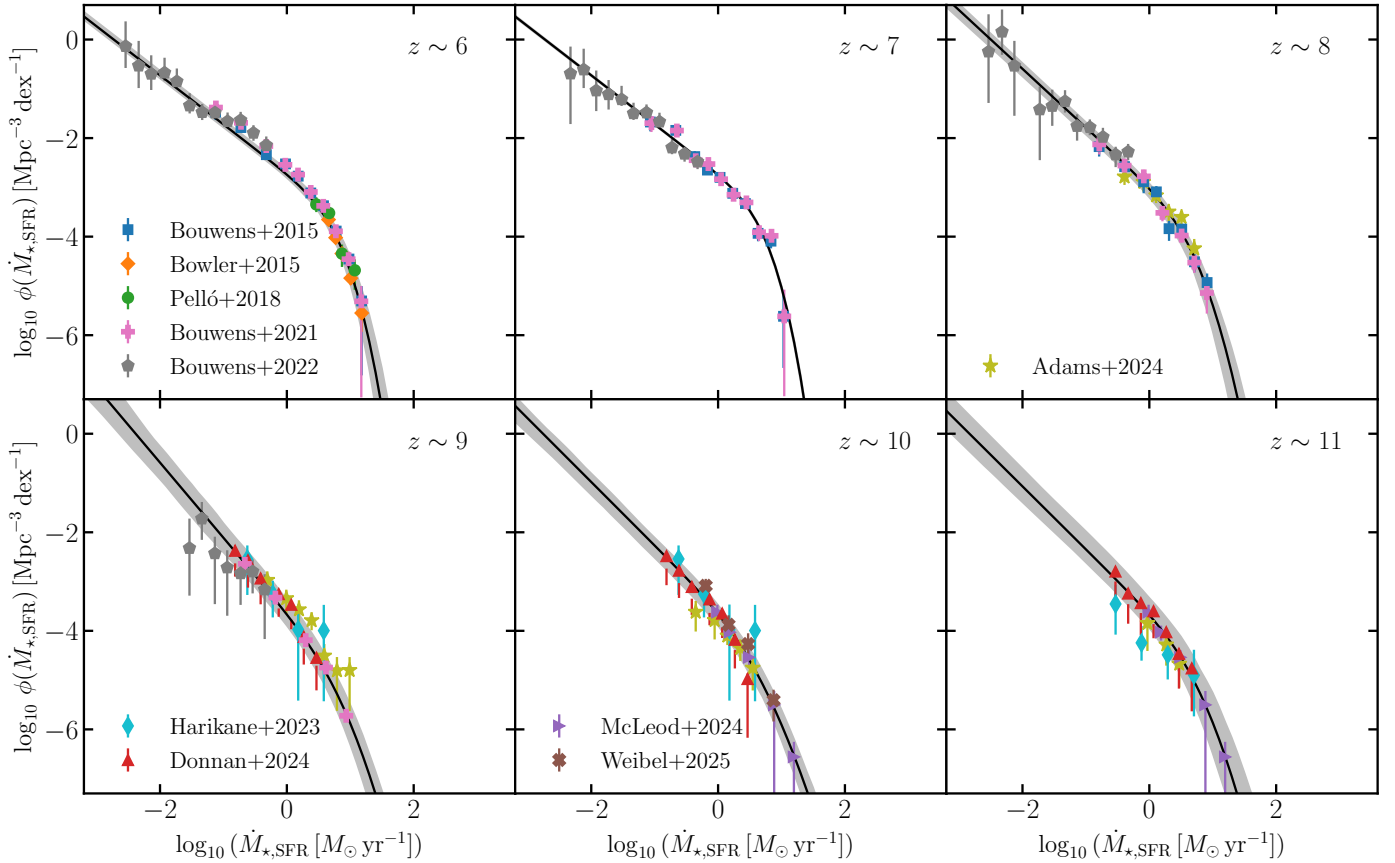


Fig. 2. Star formation rate function at from redshifts $z = 6$ to 11 . The data points with error bars show the data from HST (blue squares, pink pluses, and grey pentagons; Bouwens et al. 2015, 2021, 2022), UltraVISTA/COSMOS and UKIDSS (orange rhombuses; Bowler et al. 2015), WUDS (green dots; Pelló et al. 2018), PEARLS+JWST (yellow stars; Adams et al. 2024), JWST-PRIMER (red triangles; Donnan et al. 2024), JWST ERO+ERS (cyan rhombuses and purple triangles; Harikane et al. 2023; McLeod et al. 2024) and PANORAMIC (brown crosses; Weibel et al. 2026). The black solid lines and shaded areas show our fit to Equation (A.1) and 1σ uncertainty.

procedure to retrieve such a dependence, we are now best positioned to apply this method to the real Universe, with directly observed SFR functions and not derived from a simulation.

4. Results

In this Section, we present our results on the galaxy stellar mass growths at redshift $z \gtrsim 5$ by adopting as input the SFR-HAR relation from the UV LF abundance matching. In particular, we show the SFR function used as input in our abundance matching and the output SFR-HAR relation (Section 4.1). Then we show the predicted stellar mass function (SMF) in Section 4.2, and the corresponding $M_\star - M_h$ relation and the SFE in Section 4.3. We also show a case study on the star formation histories of the newly detected high-redshift galaxies from JWST (Section 4.5). Finally, we discuss the role of dust obscuration and different IMFs on the galaxy SFRs (Sections 4.7 and 4.8).

4.1. Star formation rate-halo accretion rate relation

We first show the SFR function used in our abundance matching, computed starting from the UV LF, as described in Section 2.2. Within this work we make use of UV data alone since the observational constraints on the infrared (IR) LF are available only up to $z < 6$ (e.g. Gruppioni et al. 2020; Wang et al. 2021; Traina et al. 2024, 2026). Figure 2 shows

the SFR number densities at redshifts $z = 6$ – 11 . In particular, we combine the latest observational determinations in the UV band from HST (Bouwens et al. 2015, 2021, 2022), UltraVISTA/COSMOS and UKIDSS (Bowler et al. 2015), WUDS (Pelló et al. 2018), PEARLS+JWST (Adams et al. 2024), JWST-PRIMER (Donnan et al. 2024), JWST ERO+ERS (Harikane et al. 2023; McLeod et al. 2024) and PANORAMIC (Weibel et al. 2026). We fit the above datasets to the Saunders et al. (1990) analytic form (Equation (A.1)) at each redshift via Markov chain Monte Carlo (MCMC) scans. The best-fitting form to the Saunders et al. (1990) formula is represented by the black lines and shaded areas in Figure 2, and the best-fitting parameters with their evolution in redshift are reported in Appendix A. We find that the observational SFR functions are well described by a power law, characterised by a steep drop towards the bright end ($\dot{M}_{\star,\text{SFR}} \gtrsim 3 M_\odot/\text{yr}$). In particular, galaxies with stellar mass $M_\star \sim 10^{11} M_\odot$ and SFR in the range $1 - 3 M_\odot/\text{yr}$ have a number density of $\sim 10^{-5} - 10^{-4} \text{ Mpc}^{-3} \text{ dex}^{-1}$ and extremely mild evolution in redshift from $z = 6$ to 11 .

We now turn our attention to the galaxy-halo connection. The upper panel of Figure 3 shows the SFR-HAR relation at different redshifts, which we stress is a direct result of the abundance matching between our empirical SFR functions and the HAR distributions extracted from merger trees as described in Section 2.1. The resulting SFR-HAR relation at $z \gtrsim 6$ can be well described by a single power law with slope ~ 1 and is

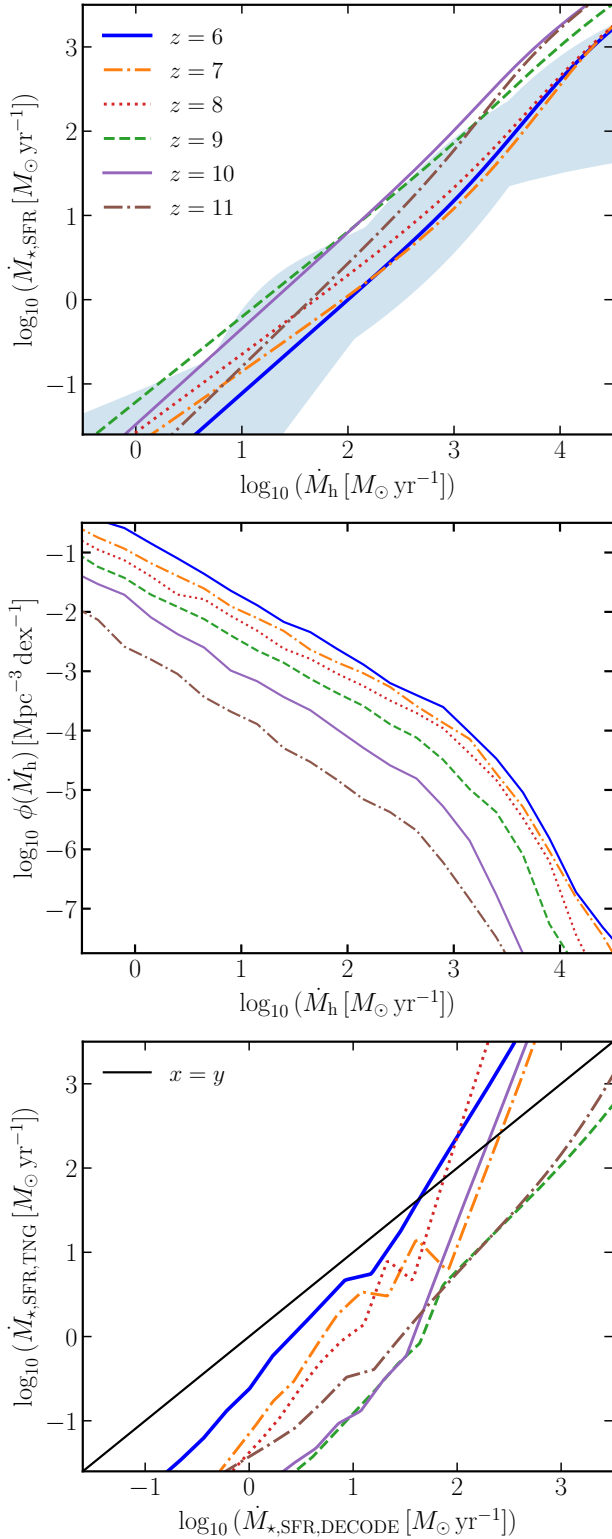


Fig. 3. Upper panel: Star formation rate-halo accretion rate relation from redshift $z = 6$ to 11 , from the abundance matching. The cyan coloured area shows the redshift evolution region of the relation from $z = 0$ to 6 , as presented in Fu et al. (2025a) and Fu et al. (2025b). Central panel: Halo accretion rate function in the same redshift bins. Lower panel: Star formation rate from the TNG simulation compared to that predicted by our abundance matching at the same halo accretion rates.

characterised by a mild time evolution, with increasing normalisation up to 0.5 dex from $z = 6$ to $z = 10$ – 11 . This varia-

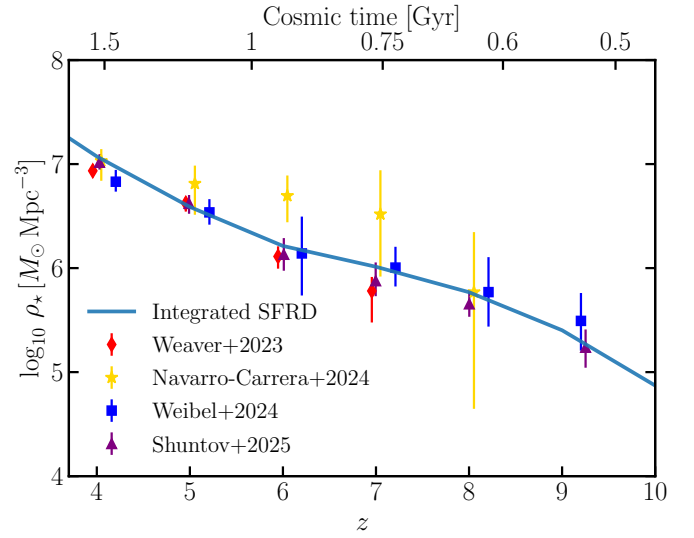


Fig. 4. Evolution of the cosmic stellar mass density. The blue solid line shows the cosmic star formation history computed by integrating the cosmic star formation rate density retrieved from our star formation rate functions shown in Figure 2. The data points show the observationally inferred stellar mass densities from COSMOS2020 (red rhombuses with error bars; Weaver et al. 2023), JWST HUDF and UDS fields (yellow stars with error bars; Navarro-Carrera et al. 2024), JWST NIRCам (blue squares with error bars; Weibel et al. 2024) and JWST COSMOS-Web (purple triangles with error bars; Shuntov et al. 2025a).

tion in redshift is a direct by-product of the quick HAR function’s evolution in normalisation, as shown in the central panel of Figure 3, whilst the evolution of the SFR function is weaker across those redshifts. As shown in the lower panel of Figure 3, we note that the TNG’s SFRs are systematically lower than ours at the same corresponding HARs. This is due to the fact that the TNG physics is calibrated to match the observations at low redshift ($z \lesssim 2$) which may not necessarily fit the observed LF at high redshifts, whereas our SFRs come directly from observational constraints. Indeed, at low redshift this difference between simulated and observed SFRs is no longer present. We checked that the origin of this mismatch does not arise from different timescale averages (e.g. Donnari et al. 2019b,a).

4.2. Galaxy stellar mass function

In this Section, we compare our SMFs, derived from the time integration of the star formation histories, with those independently measured from COSMOS2020 and JWST at high redshifts. Reproducing the observed SMFs is a crucial self-consistency test to confirm whether our star formation histories are truly representative of the underlying evolution of galaxies. We note that mergers (Section 2.4) and quenching play a minor role in shaping the SMF at $z > 6$. Indeed, although our SFR functions are by construction only representative of the star-forming population, the comparison with the total galaxy stellar mass functions is still valid as recent estimates of the quenched fractions at $z \gtrsim 5$ is limited to just $\lesssim 6\%$ for galaxies with $M_* \gtrsim 10^{10} M_{\odot}$ (e.g. Baker et al. 2025; Merlin et al. 2025; Russell et al. 2025; Yang et al. 2026). At redshifts below $z \lesssim 2$ – 3 the integral of the observed SFRs produces higher stellar masses with respect to the observed values by a factor of up to 0.5 – 0.6 dex, as also found by several works (e.g. Bernardi et al. 2010; Rodríguez-Puebla et al. 2017;

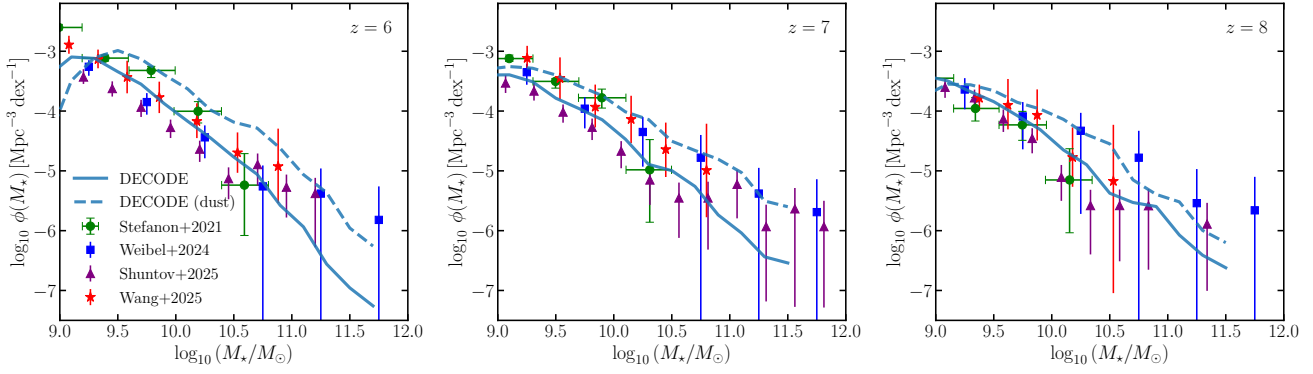


Fig. 5. Galaxy stellar mass function at redshifts $z = 6, 7$ and 8 . The blue solid lines show the prediction from DECODE’s reference model and the blue dashed lines show the scenario with dust correction (as discussed in Section 4.7). Our results are compared with the observational data from HST/CANDELS (green dots with error bars; [Stefanon et al. 2021](#)), JWST NIRCcam (blue squares with error bars; [Weibel et al. 2024](#)), JWST COSMOS-Web (purple triangles with error bars; [Shuntov et al. 2025a](#)), and JWST/MIRI (red stars with error bars; [Wang et al. 2025](#)).

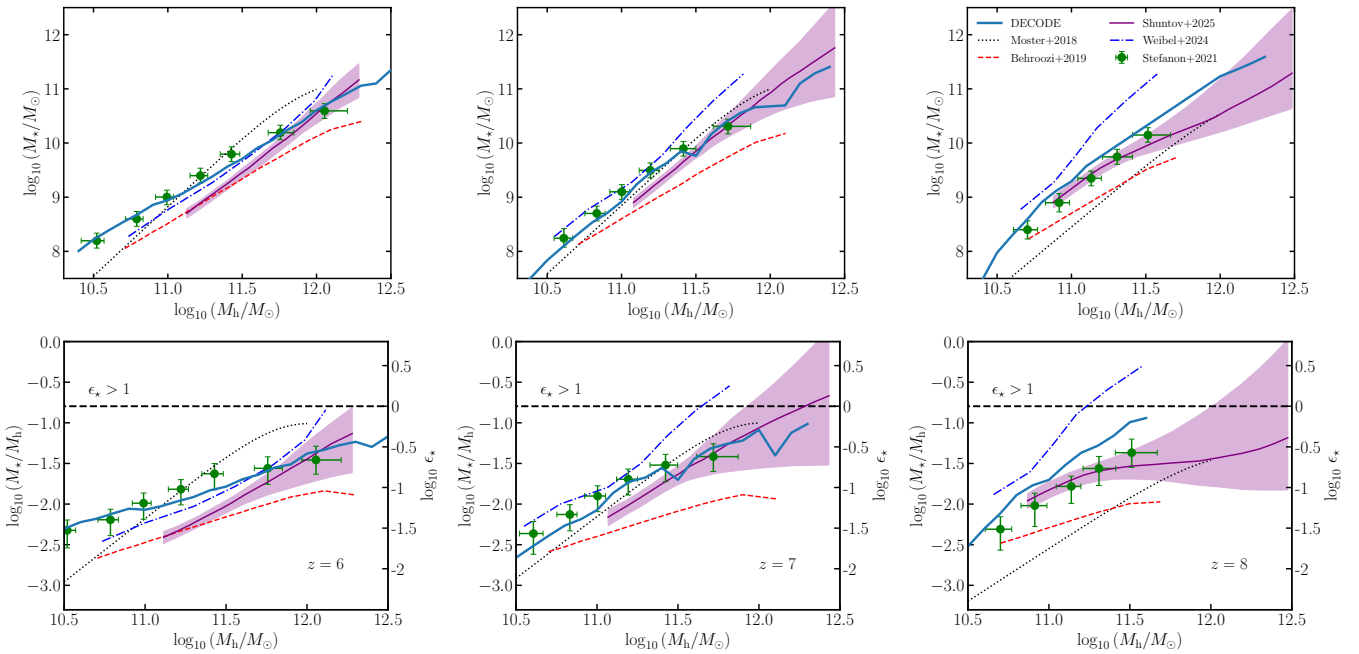


Fig. 6. Stellar mass-halo mass relation at redshifts $z = 6, 7$ and 8 . The upper and lower panels show the stellar mass and the stellar-to-halo mass ratio as a function of halo mass, respectively. The right labels in the lower panels represent the star formation efficiency, defined as $\epsilon_* = M_*/(M_h f_b)$. Our results (blue solid lines) are compared to the observational measurements from [Stefanon et al. \(2021\)](#), green dots with error bars), [Shuntov et al. \(2025a\)](#), purple solid lines with shaded areas), and to the empirical models of [Moster et al. \(2018\)](#), black dotted lines) and [Behroozi et al. \(2019\)](#), red dashed lines). The dark blue dash-dotted lines show the results from the abundance matching using the observed stellar mass function from [Weibel et al. \(2024\)](#). We note that the stellar mass-halo mass relation might be slightly biased towards higher values below $M_* \lesssim 10^9 M_\odot$, especially at $z \sim 6$, due to the incompleteness of the stellar mass function below that threshold.

[Donnari et al. 2019b](#); [Leja et al. 2022](#); [Fu et al. 2024](#); [Bosi et al. 2025](#)), but we do not show and discuss this any further as it lies beyond the purposes of this work. We also note that satellites contribute $\lesssim 1\%$ to the overall galaxies of $M_* \gtrsim 10^9 M_\odot$ and $z > 5$ ([Fu et al. in prep.](#)), further validating our assumption of focusing on central galaxies only in this work.

We start by showing in Figure 4 our integrated stellar mass density (blue solid line) compared against a variety of measurements at $z \gtrsim 4$. We find a very good agreement between our predictions and the integrated measurements, thus providing a first validation of our approach. Having compared with integrated quantities, the next step is to generate the full SMF from our model by integrating the SFR of each galaxy along its main progenitor branch. Figure 5 shows the galaxy SMF at redshifts

$z = 6, 7$, and 8 , as computed from DECODE (blue solid lines). We compare our results to the observations from multiple datasets, such as HST/CANDELS ([Stefanon et al. 2021](#)), JWST NIRCcam ([Weibel et al. 2024](#)), JWST/COSMOS-Web ([Shuntov et al. 2025a](#)), and JWST/MIRI ([Wang et al. 2025](#)).

Our model can reproduce the bulk of the SMF at all redshifts. This agreement both in normalisation and shape highlights the power of the SFR-HAR connection in tracing galaxy evolution without any further assumption, for example, on the complex and multiple physical processes regulating star formation in galaxies, especially at high redshifts, since the latter are implicitly included in our input SFRs, which should be regarded as a balance among the effects of AGN feedback, stellar feedback, and gas cooling.

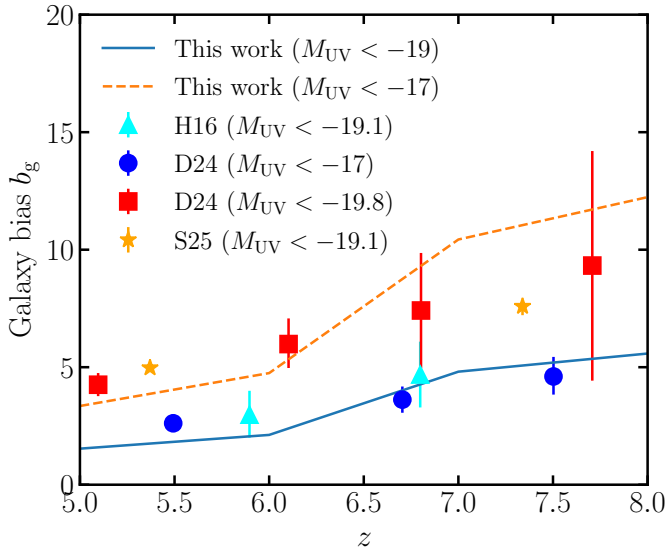


Fig. 7. Galaxy bias as a function of redshift for UV magnitude $M_{UV} \lesssim -19$ (blue solid line) and $M_{UV} \lesssim -17$ (orange dashed line). We compare our predictions to the observed galaxy bias from Subaru/Hyper Suprime-Cam (cyan triangles with error bars; Harikane et al. 2016), JWST JADES (blue dots with error bars; Dalmasso et al. 2024a), HST CANDELS (red squares with error bars; Dalmasso et al. 2024b) and JWST NIR-Cam/grism (orange stars with error bars; Shuntov et al. 2025b).

4.3. Stellar-to-halo connection

We now turn our attention to the connection between the galaxy stellar mass and host dark matter halo mass. As discussed by several groups, the shape and evolution of the $M_\star - M_h$ relation are directly linked to those of the galaxy SMF (e.g. Rodríguez-Puebla et al. 2017; Moster et al. 2018; Grylls et al. 2019; Fu et al. 2022, 2024).

Figure 6 displays the $M_\star - M_h$ relation at redshifts $z = 6, 7$, and 8 compared to the measurements from Spitzer/IRAC (Stefanon et al. 2021) and COSMOS-Web (Shuntov et al. 2025a) data. For completeness, we also include the outputs of the empirical models EMERGE (Moster et al. 2018) and UniverseMachine (Behroozi et al. 2019). The upper panels show the stellar mass as a function of halo mass, and the lower panels report the stellar-to-halo mass ratio and the SFE, defined as $\epsilon_\star = M_\star / (M_h f_b)$, where $f_b = 0.16$ is the cosmic baryon fraction (Planck Collaboration VI 2020). We find a strong correlation between SFE and halo mass, with a SFE of $\sim 70\%$ the baryonic value for the highest masses, significantly above common expectations of $\sim 10 - 20\%$ as calibrated at lower redshifts (e.g. Hopkins et al. 2011; Schaye et al. 2015; Pillepich et al. 2018a; Davé et al. 2019). We find that our predicted SFEs are well aligned with independent estimates derived from abundance matching of the SMFs and halo mass functions, with a slight tendency for our results to increase by a factor of $\sim 2-3$ at $z \geq 8$. Our predictions both in stellar mass and in SFE are also higher in normalisation with respect to the results of previous semi-empirical estimates from, e.g. Moster et al. (2018) and Behroozi et al. (2019), due to the fact that their models were calibrated to match the SMFs and quenched fractions mostly up to $z = 4$. Our current predictions are particularly noteworthy as they point to very high baryon conversion efficiencies, particularly in massive haloes at $z \geq 7$, in ways independent of any specific theoretical model but simply relying on observed UV data and halo abundances.

4.4. Clustering

Prior to predicting galaxy star formation histories, we computed the large-scale galaxy clustering, as a further test of our approach. Indeed, a robust model of galaxy evolution should be able to reproduce also the clustering in addition to galaxy abundances (e.g. Mirocha 2020; Muñoz et al. 2023; Gelli et al. 2024; Shuntov et al. 2025b; Kar et al. 2026). For this purpose, we calculate the bias as a tracer of the large-scale spatial distribution of galaxies, following the formalism put forward by Muñoz et al. (2023)

$$b_g = \frac{1}{\phi(M_{UV})} \int b(M_h) P(M_h | M_{UV}) \frac{dn}{dM_h} dM_h, \quad (3)$$

where $\phi(M_{UV})$ is the UV LF, $b(M_h)$ is the halo bias (Tinker et al. 2010), $P(M_h | M_{UV})$ is the probability distribution of halo mass at fixed UV magnitude, and dn/dM_h is the halo mass number density.

Figure 7 shows the galaxy bias evolution as predicted by our model, for $M_{UV} \lesssim -19$ and $M_{UV} \lesssim -17$, compared to a variety of observational measurements (Harikane et al. 2016; Dalmasso et al. 2024b,a; Shuntov et al. 2025b). Our model predicts an increasing trend for the galaxy bias as a function of redshift and as a function of luminosity at fixed epoch, in good agreement with the data and theoretical expectations of structure formation.

4.5. Star formation histories

As anticipated above, our methodology allows to track the star formation of galaxies along the progenitors to build full star formation histories and integrated stellar masses. In this Section, we explore the performance of our semi-empirical model in reproducing very massive galaxies of the order of $M_\star > 10^{11} M_\odot$ at $z \approx 5 - 6$, as observed by Xiao et al. (2024, 2026).

Figure 8 shows the star formation histories from DECODE in different stellar mass bins compared to two massive star-forming candidates detected by JWST (Xiao et al. 2024; Wang et al. 2024b; de Graaff et al. 2025). In particular, we show the stellar mass (upper panel), SFR (second panel), specific SFR (third panel), and SFE (lower panel) as a function of redshift. Our model naturally predicts that massive galaxies comparable to those observed by Xiao et al. (2024) underwent a high star formation phase at redshift $z > 9$, with a peak of SFR at $z \sim 10$. In particular, these galaxies formed half of their stellar mass in less than 100 Myr with a bursty star formation. The first peak in SFR at redshift $z \sim 10$ is characterised by values of SFR between $100 - 200 M_\odot/\text{yr}$ and SFE close to unity. After 100 – 150 Myr of intense star formation the galaxies entered a slightly milder phase with lower values of (specific) SFR and SFE, reaching standard values of $\sim 30\%$ below $z \lesssim 7$. As shown in the lower panel of Figure 8, massive galaxies experience more frequent major mergers with increasing redshift (with $\Psi_{\text{merger}} \gtrsim 2/\text{Gyr}$ at $z \gtrsim 8$), suggesting that part of the high SFR before the peak may be ex situ triggered from major mergers. Such a correspondence between high major merger rates and SFRs, expected from the SFR-HAR relation, suggests that major mergers may have contributed to boost early and high SFRs, even if their direct contribution of ex situ stellar mass may have remained relatively minor.

We now compare our predictions for the SFR evolution to those inferred from SEDs. Figure 9 compares the mean star formation history from our catalogue to those inferred for massive candidates from JWST (Glazebrook et al. 2024; Wang et al.

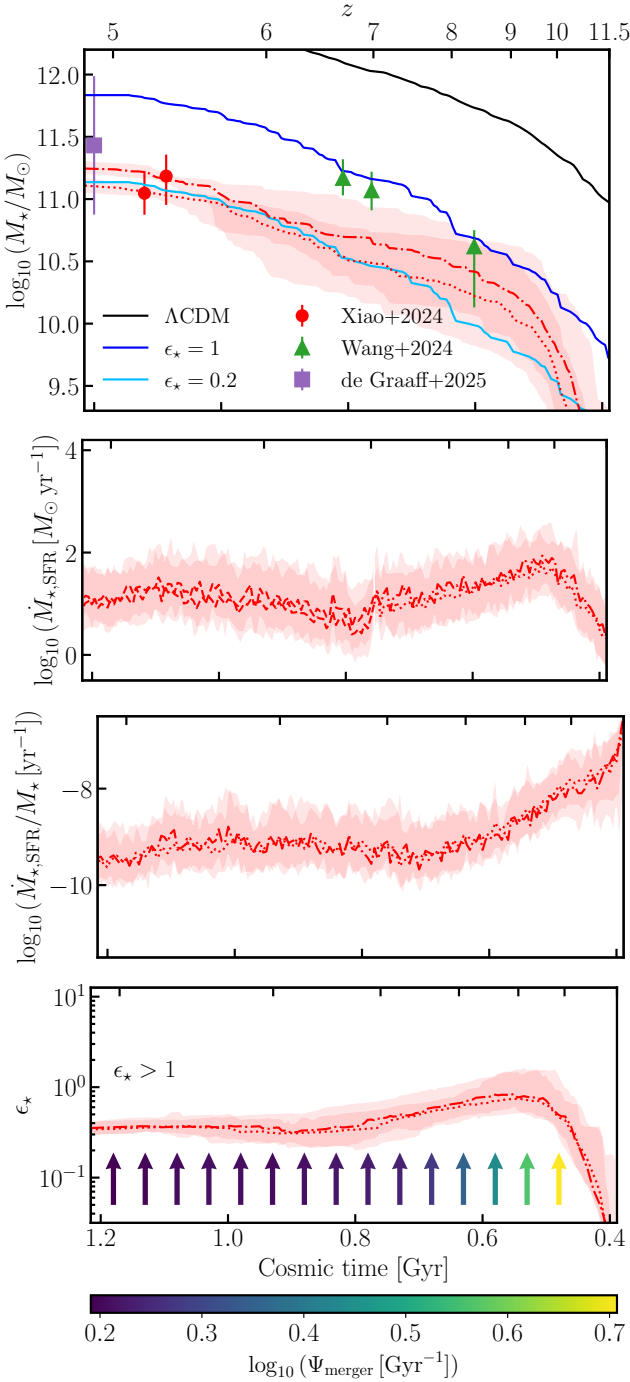


Fig. 8. Upper panel: Stellar mass growth of galaxies in different mass bins as a function of redshift. The red dots with error bars show three massive galaxies in the JWST sample from Xiao et al. (2024). The massive candidates detected by de Graaff et al. (2025) and Wang et al. (2024b) are also plotted as a reference (purple square and green triangles with error bars). The red lines and shaded areas show the mean stellar mass and 1σ dispersion in the same mass bins from DECODE. The cyan and blue lines show the baryon mass limit computed from the dark matter halo growth with star formation efficiency of $\epsilon_* = 0.2$ and 1, respectively. The black solid line shows the Λ CDM limit with $\epsilon_* = 1 + 3\sigma$. Second, third and fourth panels: Star formation rate, specific star formation rate and star formation efficiency for the same stellar mass bins as in the upper panel. The grey shaded area in the lower panel shows the region where the star formation efficiency exceeds 100%. The colour-coded arrows in the lower panel represent the merger rate, plotted in terms of number of major mergers per unit time, as a function of redshift.

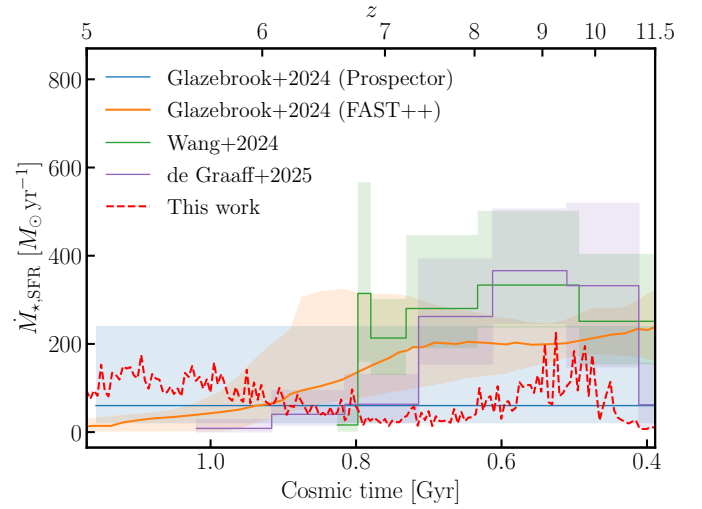


Fig. 9. Average star formation rate evolution for galaxies selected in our catalogue with stellar mass $M_* \sim 10^{11} M_\odot$ at redshift $z \approx 5$ (red dashed line). We compare our results to the star formation histories inferred for the JWST ZF-UDS-7329 galaxy from Prospector and FAST++ ($M_* = 10^{11} M_\odot$ at $z = 3.2$; Glazebrook et al. 2024; blue and orange lines with shaded areas), JWST RUBIES-EGS-49140 ($M_* = 1.5 \times 10^{11} M_\odot$ at $z = 6.7$; Wang et al. 2024b; green line with shaded area) and RUBIES-EGS-QG-1 ($M_* = 10^{11} M_\odot$ at $z = 4.9$; de Graaff et al. 2025; purple line with shaded area) candidates.

2024b; de Graaff et al. 2025). We selected galaxies of stellar mass $M_* \sim 10^{11} M_\odot$ at $z = 5$ to be as consistent as possible with the mass and spectroscopic redshift of the samples from the literature. Our results are broadly in agreement with the star formation histories from SED fitting, yielding a high burst of star formation $\sim 400 \text{ Myr}$ after the Big Bang ($z \sim 11$), with peaks reaching $150\text{--}200 M_\odot/\text{yr}$. Interestingly, such SFR quickly drops after $100\text{--}150 \text{ Myr}$, bringing the galaxy to a quiescent mode and after $\sim 200 \text{ Myr}$ further resuming to appreciably higher values of the SFR. Such a breathing alternate phases of SFR and quenching have been recently put forward by Merlin et al. (2025) to explain the puzzling non-linear evolution of the relative number densities of quenched and star-forming galaxies at high redshift from ASTRODEEP-JWST. It is interesting to note that we do not include any quenching mechanism in our model galaxies and their apparent periodic behaviour in SFR with the low-SFR phase is uniquely driven by the evolution of the underlying HAR, on the assumption that the host halo (characterised by masses well below the quenching scale of $M_h \sim 10^{12}\text{--}10^{13} M_\odot$, e.g. Dekel & Birnboim 2006) continues carrying cold gas to fuel the star formation in the central galaxy.

4.6. Star-forming main sequence

In this Section, we turn our attention to the star-forming main sequence of galaxies as predicted by our model. The three panels of Figure 10 show the SFR-stellar mass relation at redshifts $z = 6, 7$ and 8 , as labelled, compared to the observed relations from Spitzer+Herschel (Popesso et al. 2023), JWST JADES/CEERS (Clarke et al. 2024), and COSMOS/SMUVS, JWST JADES/GOODS-S, and MIDIS/XDF (Rinaldi et al. 2025). The overall relation is consistent in slope with the Clarke et al. (2024) and Rinaldi et al. (2025) determinations at all redshifts. We also find a mildly increasing normalisation of the SFR as a function of redshift, in agreement

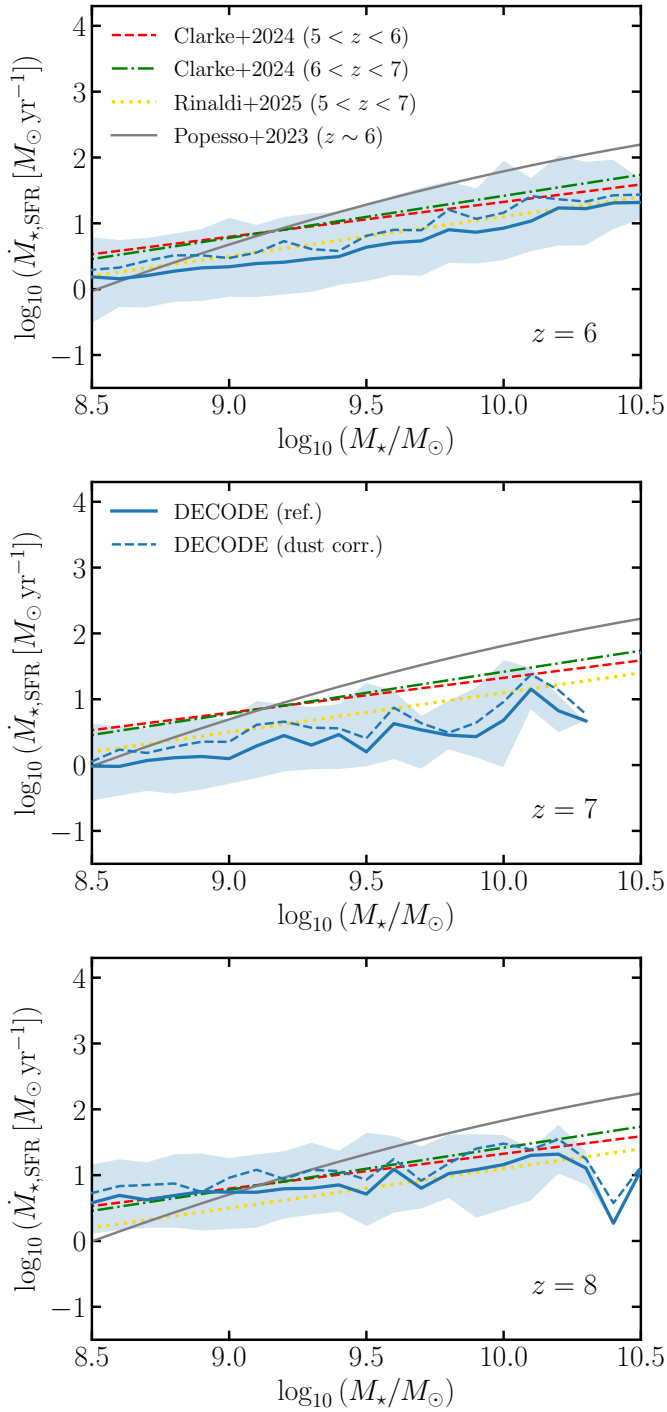


Fig. 10. Star formation rate-stellar mass relation at redshifts $z = 6$, 7 , and 8 . The predictions from DECODE (blue solid lines and shaded areas) are compared to the observational determinations from Spitzer+Herschel (grey solid lines; Popesso et al. 2023), JWST JADES/CEERS (red dashed and green dash-dotted lines; Clarke et al. 2024), and COSMOS/SMUVS, JWST JADES/GOODS-S and MIDIS/XDF (yellow dotted lines; Rinaldi et al. 2025). The blue dashed lines show the model where dust correction is included in the star formation rates.

with the increase in accretion rate of the host dark matter halo. The good agreement with the main sequence further strengthens the validity and overall self-consistency of our semi-empirical approach.

4.7. The role of dust obscuration

As described above, our SFRs are empirically based on direct observation of the UV luminosities, under the assumption that the dust content at high redshift is low (e.g. Pettini et al. 1997; Nath et al. 2023; Shchekinov & Nath 2025). However, the SFRs may still be underestimated due to the effect of dust obscuration that could damp the intrinsic UV luminosity arising from a galaxy (e.g. Payoor Vijayan et al. 2026), and mounting evidence suggests that the dust may still be present even in very high-redshift galaxies (e.g. Ferrara 2024; Somerville et al. 2025; Rodighiero et al. 2026). Here, we tested the effect of dust obscuration by assuming different attenuation recipes from the literature. In particular, we explored the attenuation laws from Charlot & Fall (2000), Calzetti et al. (2000) and Salim et al. (2018), which we describe in detail in Appendix C. In the same Appendix we also show that the dust-corrected UV LF would match the IR LF at $z \lesssim 6$, where IR measurements are available, providing independent support to our dust-correction recipes.

We find that within the UV luminosity range explored here, all the recipes quoted above produce similar results, yielding SFRs higher by a factor of ~ 2 – 3 with respect to those predicted by our reference model. This shift would in turn enhance our predicted SFRs for the JWST massive candidates from ~ 150 – $200 M_\odot/\text{yr}$ to ~ 300 – $600 M_\odot/\text{yr}$, in closer agreement with what measured by Xiao et al. (2024) at redshift $z \sim 5$ – 6 , although still lower than their quoted value of $\sim 700 M_\odot/\text{yr}$. When including dust corrections, the galaxy SMF, shown in Figure 5, is shifted towards higher stellar masses, though still broadly aligned with current data, except possibly at $z \sim 6$ and at higher stellar masses. Also the star-forming main sequence, reported in Figure 10, increases by 0.2 – 0.3 dex after dust correction. Intriguingly, the dust correction would enhance the SFRs in Figure 8 at redshifts $z \sim 8$ – 10 pushing to SFEs above unity, as we further discuss below in Figure 12. These findings would imply 1) that the dust content may still be low at these high redshifts and high stellar masses and/or 2) that a variable IMF is an inevitable ingredient in the evolution of galaxies at high redshifts. We explore the impact of a variable IMF in the next Section.

4.8. The dependence on the initial mass function

We now discuss the implications of varying the input IMF on the stellar masses and SFRs of the galaxies analysed in this work. The shape of the IMF and its dependence on the galaxy SFR continue to be a matter of intense scrutiny and with the advent of JWST data the topic has become even more relevant. Several works have pointed out the dependence of the stellar population and the derived total stellar mass on the assumed IMF, with the new observed massive candidates pointing towards more exotic, e.g. top-heavy, IMFs (e.g. Steinhart et al. 2023; Carnall et al. 2024; Lapi et al. 2024; van Dokkum & Conroy 2024; Wang et al. 2024a). Below, we employ two analytical recipes for correcting the galaxy SFR for different assumed IMFs. Our aim is to test whether a variable IMF can relax the pressing need for extreme SFE values in forming massive galaxies. For example, a top-heavy IMF would generate more massive stars increasing the UV luminosity for a given burst of star formation, reducing the need for a high efficiency of star formation.

First, following Lapi et al. (2024), we parametrise the IMF via the Larson (1998) analytic formula

$$\phi(m_\star) \propto m_\star^\xi \cdot e^{-m_\star/c/m_\star}, \quad (4)$$

where ξ is the slope at high star masses and $m_{\star,c}$ is the characteristic mass below which the IMF flattens or bends. In order to

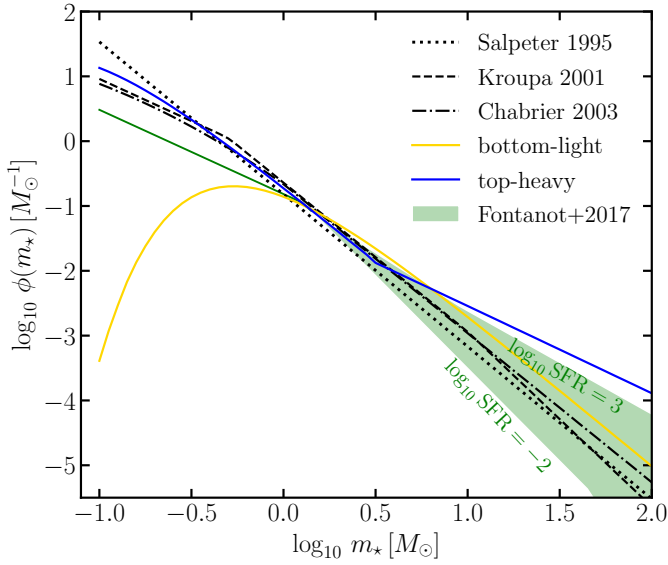


Fig. 11. Initial mass function from [Salpeter \(1955\)](#), black dotted line), [Kroupa \(2001\)](#), black dashed line) and [Chabrier \(2003\)](#), black dash-dotted line), compared to the bottom-light (gold solid line) and top-heavy (red solid line) toy model IMFs parametrised in terms of the [Larson \(1998\)](#) formula. The green shaded area represents the range of variation of the integrated galaxy-wide initial mass function for star formation rates between 10^{-2} and $10^3 M_{\odot} \text{yr}^{-1}$ from [F17](#). The different IMFs are normalised to the value of the [Chabrier \(2003\)](#) IMF at $m_{\star} = 1 M_{\odot}$.

compute the k_{UV} conversion factor from UV luminosity to SFR, we make use of the results from [Lapi et al. \(2024\)](#), see Section 2.2 therein) obtained using the PARSEC code for stellar evolution ([Bressan et al. 2012](#); [Goswami et al. 2022](#)).

We also take advantage of the predictions of the Galaxy Evolution and Assembly (GAEA) semi-analytical model implementing variable IMF prescriptions. In particular, we consider the GAEA realisation discussed in [Fontanot et al. \(2017, F17 hereafter\)](#), see also [Fontanot et al. 2024](#) including the integrated galaxy-wide IMF (IGIMF) model proposed by [Weidner & Kroupa \(2005\)](#). The IGIMF framework provides an estimate for the IMF shape for any given reference SFR. In particular, we use the same broken power law parametrisation for the IMF shape adopted in [F17](#), and the corresponding best-fit parameters from their Table 1. The resulting estimates for the k_{UV} factors in the different scenarios allow us to explore a broader, but still physically plausible, range of conversions from UV LF to the SFR function, and probe their impact on our results compared to a Milky Way-like IMF. We show in Appendix B the resulting SFR functions.

Figure 11 reports the various IMFs that we investigate in this Section. In particular, in addition to the canonical [Chabrier \(2003\)](#) IMFs (see also [Salpeter 1955](#); [Kroupa 2001](#)), we consider more exotic ones such as bottom-light and top-heavy, which have been constructed by tweaking the characteristic mass of the [Larson \(1998\)](#) formula to $m_{\star,c} = 0.1 M_{\odot}$ for the former, and the slope at high star masses to $\xi = -1.5$ for the latter. From the IGIMF library considered in [F17](#), we focus on the IMF corresponding to a $\log_{10} \text{SFR} = 3$, as the corresponding top-heavy shape provides the largest deviations from the canonical Chabrier IMF. We also note that, in principle, the k_{UV} conversion factor should vary with stellar age and metallicity, but we find that it has a weak dependence on these parameters in the ranges of interest to this work, stellar ages of 10–200 Myr and metallicity

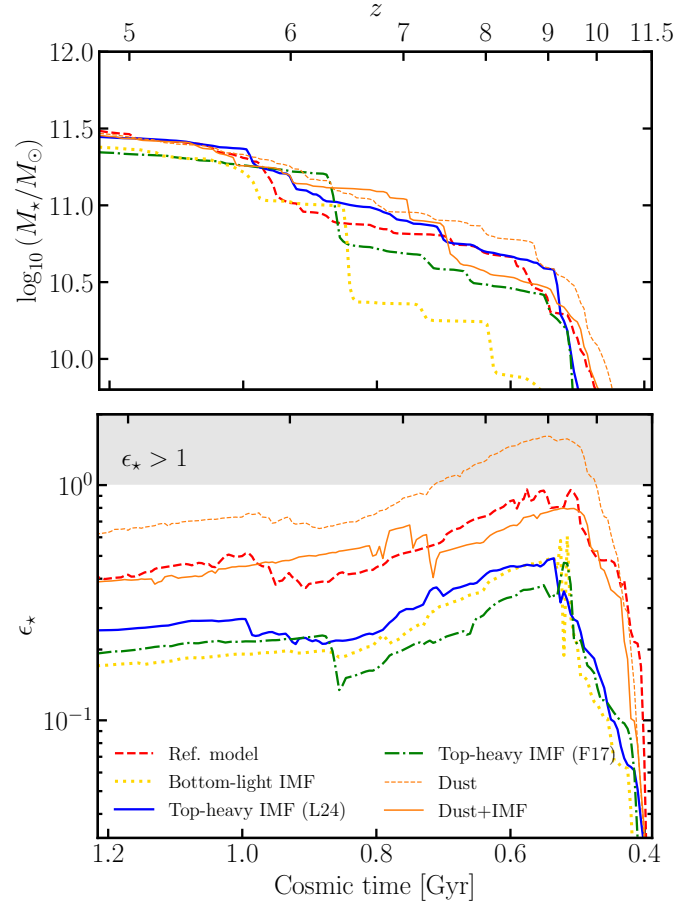


Fig. 12. Upper panel: Stellar mass growth of galaxies in the same highest stellar mass bins as in Figure 8, for our reference model IMF (red dashed line), the bottom-light (gold dotted line), and the [Lapi et al. \(2024\)](#) and [F17](#) top-heavy IMFs (blue solid and green dash-dotted lines). The orange dashed and solid lines show the cases with dust, and dust along with a top-heavy [Lapi et al. \(2024\)](#) IMF. Lower panel: Same as the upper panel, but for the star formation efficiency.

ties $0.001-1 Z_{\odot}$ (e.g. [Carnall et al. 2023](#); [Curtis-Lake et al. 2023](#); [Labbé et al. 2023](#); [Tacchella et al. 2023](#); [Topping et al. 2024](#)).

For the above-mentioned IMFs and input assumptions, we retrieve the following values of the UV luminosity-to-SFR conversion factors:

$$k_{\text{UV},\text{Kroupa}} = 2.1 \times 10^{-10} M_{\odot} \text{yr}^{-1} L_{\odot}^{-1} L_{\text{UV}}, \quad (5)$$

$$k_{\text{UV},\text{bottom-light}}^{\text{L24}} = 1.1 \times 10^{-10} M_{\odot} \text{yr}^{-1} L_{\odot}^{-1} L_{\text{UV}}, \quad (6)$$

$$k_{\text{UV},\text{top-heavy}}^{\text{L24}} = 1.7 \times 10^{-10} M_{\odot} \text{yr}^{-1} L_{\odot}^{-1} L_{\text{UV}}, \quad (7)$$

$$k_{\text{UV},\text{top-heavy}}^{\text{F17}} = 1 \times 10^{-10} M_{\odot} \text{yr}^{-1} L_{\odot}^{-1} L_{\text{UV}}. \quad (8)$$

The abundance matching procedure is then repeated by using the new conversion factors which lead to different SFR functions (see also Appendix B).

Interestingly, when assuming an input bottom-light IMF, the output SFRs would be a factor of ~ 2 lower than that predicted by the fiducial [Chabrier \(2003\)](#). Conversely, when adopting the top-heavy IMF from the [Lapi et al. \(2024\)](#) formalism, the SFRs would be lowered by the same factor, and by up to $\sim 50\%$ when adopting the top-heavy IMF from the [F17](#) formalism. Figure 12

shows the stellar mass assembly (upper panel) and SFE (lower panel) in the most massive bin as in Figure 8, for the variable bottom-light and top-heavy IMFs. Both a bottom-light and a top-heavy IMF could alleviate the need for extreme values of the SFEs and better accommodate the observed massive galaxies within Λ CDM, though still requiring peak values of the SFEs of around 30 – 40%. However, if we reinsert dust corrections in our non-canonical IMF models, we would still recover very high SFEs of $\epsilon_\star \gtrsim 0.7$ at $z \gtrsim 7$ (orange solid lines in Figure 12). Instead, as mentioned above, the reference model with only dust correction would yield unphysical SFEs greater than unity at $z \gtrsim 7$ (orange dashed lines in Figure 12). Thus, a non-canonical IMF appears to be a necessary ingredient to contain the SFEs of high redshifts, in particular in massive and dusty galaxies.

5. Discussion

In this work, we laid out a self-consistent and minimal data-driven approach to assign empirical SFRs to haloes and build the implied galaxy star formation histories. Instead of using observed UV LFs as outputs to test the model, we used them as input to infer SFRs mapped onto host dark matter haloes via abundance matching techniques. With the least possible assumptions and theoretical inputs, our methodology transparently allows us 1) to infer the efficiencies of star formation in haloes as emerging directly from the observed galaxy luminosities, and 2) to test the consequences of adding dust obscuration and/or IMF variations on empirical SFRs. In general, we find that high SFEs appear to be unavoidable within the framework of a Λ CDM Universe, independently of any stochasticity in the star formation histories or dust obscuration, although the inclusion of a variable IMF could reduce the SFEs by a factor of ~ 2 . Below we further discuss some implications of our results and how they compare to other works.

We showed in Figures 8 and 9 that our method naturally predicts bursts of star formation characterised by extremely high SFEs following the sustained halo accretion histories characterising host haloes at early epochs. This finding aligns with what has been recently found in cosmological simulations by Dekel et al. (2025), who reported feedback-free starbursts with massive galaxies at cosmic dawn. Similar results are also reported by the work of Ferrara et al. (2025a), who found feedback-free SFRs in the pre-supernova phase with SFE $\epsilon_\star > 0.4$. We also found that these galaxies are characterised by a number density of $\sim 1.8 \times 10^{-5} \text{ Mpc}^{-3}$ at redshift $z \sim 5\text{--}6$ for $M_\star \gtrsim 10^{11} M_\odot$, which lies far below the hard upper limit imposed by the dark matter halo mass function of $\sim 10^{-3} \text{ Mpc}^{-3}$ at the same redshifts (as shown in Boylan-Kolchin 2023), showing no tension between these objects and Λ CDM.

Another important caveat of our modelling is the lack of IR restframe data at redshift $z \gtrsim 6$. We have shown that assuming an extinction law to account for the dust obscuration, may generate some overestimation of the measured SMF and further increase the already extreme SFEs above unity (dashed orange lines in Figure 12), although it could help in boosting the star formation histories of the very massive galaxies. We speculate that dust correction is a relevant factor in massive galaxies at the late phases of cosmic dawn, but less impactful at higher redshift and lower stellar masses. This hypothesis is also supported by the findings of several other works, that support a dust-free star formation scenario above $z \gtrsim 10$ with a gradual transition period at redshift $z \sim 8\text{--}9$ (e.g. Ferrara et al. 2022, 2023, 2025b; Algera et al. 2023; Cullen et al. 2024).

Finally, we studied how the galaxy SFE, for a given observed luminosity, can vary by assuming different IMFs. In particular, we found that a more exotic IMF, e.g. characterised by more massive stars, is expected to improve the above tension between the JWST high-luminosity objects and Λ CDM. This is due to the fact that more massive stars generally emit a higher amount of light per unit mass, producing higher luminosities at fixed galactic stellar masses. Indeed, as a result of this, we found that a top-heavy or bottom-light IMF yields SFEs a factor of $\sim 10\text{--}50\%$ lower with respect to the extreme values of SFEs $\epsilon_\star \sim 1$ suggested by the fiducial model, in good agreement with what reported in Ziegler et al. (2025) and Fontanot et al. (2026). We found that our reference model with dust and/or the inclusion of a non-canonical IMF, still yields a clustering strength at large scales within $\pm 10\text{--}20\%$. Therefore, at the level of precision of the current data, galaxy clustering may not yet be able to efficiently break degeneracies among different models, but it can still be used as a consistency check for the validation of the mean correlation between stellar mass and host halo mass.

All in all, our findings suggest that the observed UV LFs necessarily translate into high SFEs when mapped against the halo accretion histories in a Λ CDM early Universe at $z \gtrsim 7$, in line with a variety of previous studies (e.g. Dekel et al. 2023; Ferrara et al. 2023; Lapi et al. 2024; Menci et al. 2024; Somerville et al. 2025; Prada et al. 2026). In turn, such high SFEs, when integrated in time, can generate massive galaxies close in stellar mass to those observed by JWST. Even in the presence of a non-canonical stellar IMF, the SFEs remain considerably high and close to unity if dust corrections are included. Beyond- Λ CDM cosmological models, making use of, for example, dynamical dark energy, could release some of the tension, although they would still require high SFEs (e.g. Menci et al. 2026; Comini et al. 2026).

6. Conclusions

In this work, we have presented a flexible yet powerful data-driven method to estimate the star formation histories and efficiencies of the high-redshift galaxies detected by JWST. Instead of fitting the observed UV LFs through parametric SFEs, we have used these data as input in our framework to extract the implied SFRs, which we have mapped onto host dark matter haloes via abundance matching techniques with the halo accretion rates. Our SFRs are thus by design directly rooted in observations and as such can be considered an effective balance among any cooling, feedback and accretion, without requiring explicit modelling of any of these processes. By assigning SFRs to galaxies via our abundance matching relations along the main progenitors, we have built star formation histories, star formation efficiencies, and integrated stellar masses. In our transparent and minimalistic method based on abundance matching, SFEs simply emerge from the data rather than being a tuned input of the model, thus serving as a natural baseline to test burstiness, dust attenuation, or IMF variations.

Our main results can be summarised as follows:

- The assumption of a monotonic relation between the galaxy SFR and dark matter halo accretion rate can robustly predict and match the bulk of the observed galaxy stellar mass function, supporting the view that the vast majority of galaxies above $z \gtrsim 5$ are mostly star-forming (Figures 4 and 5).
- The star formation efficiency implied by our data-driven formalism is contained within 30% up to $z \sim 6$ but rapidly increasing to $\epsilon_\star \gtrsim 0.8$ at $z > 9$ for $M_\star \gtrsim 5 \times 10^{10} M_\odot$ (Figure 6).

- We find that massive galaxies formed the majority of their stellar mass via a bursty star formation roughly 400 Myr after the Big Bang with extreme values of baryon conversion efficiencies (Figures 8 and 9).
- Our predicted star formation histories along the main progenitors generate a main sequence consistent with current measurements at these high redshifts (Figure 10).
- Dust obscuration may play an important role in massive galaxies and can increase the total SFR by a factor of up to ~ 2 on average at redshift $5 \lesssim z \lesssim 8$, with respect to the reference case without dust, implying SFEs greater than unity at $z \gtrsim 7$.
- Non-canonical IMF scenarios, like a top-heavy one (Figure 11), produce lower SFRs and SFEs at fixed luminosity, alleviating the pressing need for an extreme baryon conversion efficiency at $z > 9$ (Figure 12).

In conclusion, all our results point to high SFEs, close to unity at $z \gtrsim 7$, even when accounting for non-standard IMFs, especially when correcting for dust extinction. Semi-empirical models represent a powerful data-driven approach for addressing key issues of galaxy evolution at high redshifts. Semi-empirical models can accurately predict several galactic properties, such as the star formation histories, galaxy abundances, and stellar mass-halo mass relation in line with observations, with only a minimal number of observed parameters as input. With the upcoming high-quality data from observatories, such as Euclid (e.g. [Euclid Collaboration: Mellier et al. 2025](#)) and Rubin-LSST (e.g. [Ivezić et al. 2019](#)), the role played by data-driven models like ours will become even more relevant to extract key constraints on galaxy formation and evolution quantities in a self-consistent and flexible way (e.g. [Lapi et al. 2026](#)).

Acknowledgements. We warmly thank Y. Harikane for useful discussion on the UV luminosity function. We cordially thank R. Somerville, K. Glazebrook and L. Graziani for insightful discussions on the evolution of high-redshift galaxies. HF acknowledges support at Fudan University from the Shanghai Super Post-doctoral Excellence Program grant No. 2024008 and from the National Science Foundation of China (NSFC) grant No. W2533002. FS acknowledges partial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860744 for the BiD4BEST project. AL acknowledges support from the Italian Research Center on High Performance Computing Big Data and Quantum Computing (ICSC), project funded by European Union – NextGenerationEU – and National Recovery and Resilience Plan (NRRP) – Mission 4 Component 2 within the activities of Spoke 3 (Astrophysics and Cosmos Observations). FY is supported in part by NSFC (grant Nos. 12133008, 12192220, and 12192223). MA is supported at the Argelander Institute für Astronomie through the Argelander Fellowship. LB acknowledges financial support from the German Excellence Strategy via the Heidelberg Cluster of Excellence (EXC 2181 – 390900948) STRUCTURES.

References

- Adams, N. J., Conselice, C. J., Austin, D., et al. 2024, *ApJ*, **965**, 169
- Akins, H. B., Casey, C. M., Lambrides, E., et al. 2025, *ApJ*, **991**, 37
- Algera, H. S. B., Inami, H., Oesch, P. A., et al. 2023, *MNRAS*, **518**, 6142
- Aversa, R., Lapi, A., de Zotti, G., Shankar, F., & Danese, L. 2015, *ApJ*, **810**, 74
- Baker, W. M., Valentino, F., Lagos, C. d. P., et al. 2025, *A&A*, **702**, A270
- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, *MNRAS*, **488**, 3143
- Bernardi, M., Shankar, F., Hyde, J. B., et al. 2010, *MNRAS*, **404**, 2087
- Bernardi, M., Domínguez Sánchez, H., Brownstein, J. R., Drory, N., & Sheth, R. K. 2019, *MNRAS*, **489**, 5633
- Bicker, J., & Fritze-v. Alvensleben, U. 2005, *A&A*, **443**, L19
- Boco, L., Lapi, A., Shankar, F., et al. 2023, *ApJ*, **954**, 97
- Bosi, M., Lapi, A., Boco, L., et al. 2025, *ApJ*, **984**, 117
- Bournaud, F., Daddi, E., Weiß, A., et al. 2015, *A&A*, **575**, A56
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, **803**, 34
- Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, *AJ*, **162**, 47
- Bouwens, R. J., Illingworth, G., Ellis, R. S., Oesch, P., & Stefanon, M. 2022, *ApJ*, **940**, 55
- Bouwens, R., Illingworth, G., Oesch, P., et al. 2023, *MNRAS*, **523**, 1009
- Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, *MNRAS*, **452**, 1817
- Boylan-Kolchin, M. 2023, *Nat. Astron.*, **7**, 731
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Buchan, S., & Shankar, F. 2016, *MNRAS*, **462**, 2001
- Calabrò, A., Pentericci, L., Llerena, M., et al. 2026, *A&A*, submitted, [arXiv:2602.18068]
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
- Cappellari, M., Bacon, R., Bureau, M., et al. 2006, *MNRAS*, **366**, 1126
- Carnall, A. C., Begley, R., McLeod, D. J., et al. 2023, *MNRAS*, **518**, L45
- Carnall, A. C., Cullen, F., McLure, R. J., et al. 2024, *MNRAS*, **534**, 325
- Chabrier, G. 2003, *PASP*, **115**, 763
- Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718
- Clarke, L., Shapley, A. E., Sanders, R. L., et al. 2024, *ApJ*, **977**, 133
- Comini, L., Vagnozzi, S., & Loeb, A. 2026, *J. High Energy Astrophys.*, **53**, 100626
- Conroy, C., & van Dokkum, P. G. 2012, *ApJ*, **760**, 71
- Cullen, F., McLeod, D. J., McLure, R. J., et al. 2024, *MNRAS*, **531**, 997
- Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, *MNRAS*, **518**, 425
- Curtis-Lake, E., Carniani, S., Cameron, A., et al. 2023, *Nat. Astron.*, **7**, 622
- Dalmaso, N., Leethochawalit, N., Trenti, M., & Boyett, K. 2024a, *MNRAS*, **533**, 2391
- Dalmaso, N., Trenti, M., & Leethochawalit, N. 2024b, *MNRAS*, **528**, 898
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, *MNRAS*, **486**, 2827
- de Graaff, A., Setton, D. J., Brammer, G., et al. 2025, *Nat. Astron.*, **9**, 280
- Dekel, A., & Birnboim, Y. 2006, *MNRAS*, **368**, 2
- Dekel, A., Birnboim, Y., Engel, G., et al. 2009, *Nature*, **457**, 451
- Dekel, A., Sarkar, K. C., Birnboim, Y., Mandelker, N., & Li, Z. 2023, *MNRAS*, **523**, 3201
- Dekel, A., Mandelker, N., Li, Z., et al. 2025, *MNRAS*, **544**, 160
- Donnan, C. T., McLure, R. J., Dunlop, J. S., et al. 2024, *MNRAS*, **533**, 3222
- Donnari, M., Pillepich, A., Nelson, D., et al. 2019a, *MNRAS*, **489**, 3036
- Donnari, M., Pillepich, A., Nelson, D., et al. 2019b, *MNRAS*, **485**, 4817
- Dou, J., Peng, Y., Gu, Q., et al. 2025, *ApJ*, **982**, L12
- Eisert, L., Pillepich, A., Nelson, D., et al. 2023, *MNRAS*, **519**, 2199
- Enriquez-Vargas, M., Rodríguez-Puebla, A., Manuwal, A., et al. 2026, arXiv e-prints [arXiv:2605.08353]
- Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, **697**, A1
- Fakhry, S., Shiravand, M., & Del Popolo, A. 2026, *ApJ*, **998**, 178
- Ferrara, A. 2024, *A&A*, **684**, A207
- Ferrara, A., Sommovigo, L., Dayal, P., et al. 2022, *MNRAS*, **512**, 58
- Ferrara, A., Pallottini, A., & Dayal, P. 2023, *MNRAS*, **522**, 3986
- Ferrara, A., Manzoni, D., & Ntormousi, E. 2025a, *Open J. Astrophys.*, **8**, 140
- Ferrara, A., Pallottini, A., & Sommovigo, L. 2025b, *A&A*, **694**, A286
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJ*, **946**, L13
- Finkelstein, S. L., Leung, G. C. K., Bagley, M. B., et al. 2024, *ApJ*, **969**, L2
- Fitzpatrick, E. L., & Massa, D. 1986, *ApJ*, **307**, 286
- Fontanot, F., De Lucia, G., Hirschmann, M., et al. 2017, *MNRAS*, **464**, 3812
- Fontanot, F., La Barbera, F., De Lucia, G., et al. 2024, *A&A*, **686**, A302
- Fontanot, F., De Lucia, G., Xie, L., et al. 2026, *A&A*, submitted, [arXiv:2603.22405]
- Fu, H., Shankar, F., Ayromlou, M., et al. 2022, *MNRAS*, **516**, 3206
- Fu, H., Shankar, F., Ayromlou, M., et al. 2024, *MNRAS*, **532**, 177
- Fu, H., Boco, L., Shankar, F., et al. 2025a, *A&A*, **695**, A252
- Fu, H., Shankar, F., Yuan, F., et al. 2025b, *A&A*, **704**, A244
- Gelli, V., Mason, C., & Hayward, C. C. 2024, *ApJ*, **975**, 192
- Girardi, L., Bressan, A., Bertelli, G., & Chiosi, C. 2000, *A&AS*, **141**, 371
- Glazebrook, K., Nanayakkara, T., Schreiber, C., et al. 2024, *Nature*, **628**, 277
- Goswami, S., Silva, L., Bressan, A., et al. 2022, *A&A*, **663**, A1
- Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, *A&A*, **643**, A8
- Grylls, P. J., Shankar, F., Zanisi, L., & Bernardi, M. 2019, *MNRAS*, **483**, 2506
- Grylls, P. J., Shankar, F., Leja, J., et al. 2020, *MNRAS*, **491**, 634
- Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, *ApJ*, **821**, 123
- Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, *ApJS*, **265**, 5
- Harikane, Y., Inoue, A. K., Ellis, R. S., et al. 2025, *ApJ*, **980**, 138
- Hopkins, A. M. 2018, *PASA*, **35**, e039a
- Hopkins, P. F., Quataert, E., & Murray, N. 2011, *MNRAS*, **417**, 950
- Inayoshi, K., Harikane, Y., Inoue, A. K., Li, W., & Ho, L. C. 2022, *ApJ*, **938**, L10
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, **873**, 111
- Jespersen, C. K., Carnall, A. C., & Lovell, C. C. 2025, *ApJ*, **988**, L19
- Jiang, F., Dekel, A., Freundlich, J., et al. 2021, *MNRAS*, **502**, 621
- Juodžbalis, I., Maiolino, R., Baker, W. M., et al. 2026, *MNRAS*, **546**, stag086
- Kar, A., Alam, S., & Silk, J. 2026, *ApJ*, **996**, 103
- Kennicutt, R. C., Jr 1998, *ARA&A*, **36**, 189
- Kroupa, P. 2001, *MNRAS*, **322**, 231
- La Barbera, F., Vazdekis, A., Ferreras, I., et al. 2016, *MNRAS*, **457**, 1468
- Labbé, I., van Dokkum, P., Nelson, E., et al. 2023, *Nature*, **616**, 266

- Lapi, A., Gandolfi, G., Boco, L., et al. 2024, *Universe*, **10**, 141
- Lapi, A., Boco, L., & Shankar, F. 2026, *Encycl. Astrophys.*, **4**, 554
- Larson, R. B. 1998, *MNRAS*, **301**, 569
- Leja, J., Speagle, J. S., Ting, Y.-S., et al. 2022, *ApJ*, **936**, 165
- Lovell, C. C., Harrison, I., Harikane, Y., Tacchella, S., & Wilkins, S. M. 2023, *MNRAS*, **518**, 2511
- Lu, S., Zhu, K., Cappellari, M., et al. 2024, *MNRAS*, **530**, 4474
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, *MNRAS*, **480**, 5113
- Marsden, C., Shankar, F., Bernardi, M., et al. 2022, *MNRAS*, **510**, 5639
- Mason, C. A., Trenti, M., & Treu, T. 2023, *MNRAS*, **521**, 497
- Matsumoto, T., Harikane, Y., Maeda, K., & Ioka, K. 2024, *ApJ*, **976**, L16
- Matthee, J., Naidu, R. P., Brammer, G., et al. 2024, *ApJ*, **963**, 129
- Mauerhofer, V., Dayal, P., Haehnelt, M. G., et al. 2025, *A&A*, **696**, A157
- McGaugh, S. S., Schombert, J. M., Lelli, F., & Franck, J. 2024, *ApJ*, **976**, 13
- McLeod, D. J., Donnan, C. T., McLure, R. J., et al. 2024, *MNRAS*, **527**, 5004
- Menci, N., Sen, A. A., & Castellano, M. 2024, *ApJ*, **976**, 227
- Menci, N., Castellano, M., Mukherjee, P., et al. 2026, *A&A*, **707**, A300
- Merlin, E., Fortuni, F., Calabró, A., et al. 2025, *Open J. Astrophys.*, **8**, E170
- Mirocha, J. 2020, *MNRAS*, **499**, 4534
- Moster, B. P., Naab, T., & White, S. D. M. 2018, *MNRAS*, **477**, 1822
- Muñoz, J. B., Mirocha, J., Furlanetto, S., & Sabti, N. 2023, *MNRAS*, **526**, L47
- Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, *ApJS*, **269**, 33
- Nath, B. B., Vasiliev, E. O., Drozdov, S. A., & Shchekinov, Y. A. 2023, *MNRAS*, **521**, 662
- Navarro-Carrera, R., Rinaldi, P., Caputi, K. I., et al. 2024, *ApJ*, **961**, 207
- Nelson, D., Springel, V., Pillepich, A., et al. 2019, *Comput. Astrophys. Cosmol.*, **6**, 2
- Nikopoulos, G. P., & Dayal, P. 2024, arXiv e-prints [arXiv:2409.10613]
- Muñoz, J. B., Kobayashi, C., & Tominaga, N. 2013, *ARA&A*, **51**, 457
- Parkinson, H., Cole, S., & Helly, J. 2008, *MNRAS*, **383**, 557
- Payoor Vijayan, A., Yates, R. M., Lovell, C. C., et al. 2026, *Open J. Astrophys.*, **9**, 57554
- Pelló, R., Hudelot, P., Laporte, N., et al. 2018, *A&A*, **620**, A51
- Pérez-González, P. G., Östlin, G., Costantin, L., et al. 2025, *ApJ*, **991**, 179
- Pettini, M., King, D. L., Smith, L. J., & Hunstead, R. W. 1997, *ApJ*, **478**, 536
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castellì, F. 2004, *ApJ*, **612**, 168
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018a, *MNRAS*, **475**, 648
- Pillepich, A., Springel, V., Nelson, D., et al. 2018b, *MNRAS*, **473**, 4077
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Popesso, P., Concas, A., Cresci, G., et al. 2023, *MNRAS*, **519**, 1526
- Prada, F., Behroozi, P., Ishiyama, T., et al. 2026, arXiv e-prints [arXiv:2604.18683]
- Reimers, D. 1975, *Memoires of the Societe Royale des Sciences de Liege*, **8**, 369
- Renzini, A., & Fusi Pecci, F. 1988, *ARA&A*, **26**, 199
- Rinaldi, P., Navarro-Carrera, R., Caputi, K. I., et al. 2025, *ApJ*, **981**, 161
- Robertson, B., Johnson, B. D., Tacchella, S., et al. 2024, *ApJ*, **970**, 31
- Rodighiero, G., Ferrara, A., Catone, M., et al. 2026, *A&A*, submitted, [arXiv:2603.15841]
- Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V., & Faber, S. M. 2017, *MNRAS*, **470**, 651
- Romano, D., Karakas, A. I., Tosi, M., & Matteucci, F. 2010, *A&A*, **522**, A32
- Russell, T. A., Dobric, N., Adams, N. J., et al. 2025, *MNRAS*, **544**, 4482
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, **859**, 11
- Salpeter, E. E. 1955, *ApJ*, **121**, 161
- Saunders, W., Rowan-Robinson, M., Lawrence, A., et al. 1990, *MNRAS*, **242**, 318
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Scholtz, J., Maiolino, R., D'Eugenio, F., et al. 2025, *A&A*, **697**, A175
- Segers, M. C., Crain, R. A., Schaye, J., et al. 2016, *MNRAS*, **456**, 1235
- Shankar, F., Buchan, S., Rettura, A., et al. 2015, *ApJ*, **802**, 73
- Shchekinov, Y. A., & Nath, B. B. 2025, *Galaxies*, **13**, 64
- Shen, X., Vogelsberger, M., Boylan-Kolchin, M., Tacchella, S., & Kannan, R. 2023, *MNRAS*, **525**, 3254
- Shuntov, M., Ilbert, O., Toft, S., et al. 2025a, *A&A*, **695**, A20
- Shuntov, M., Oesch, P. A., Toft, S., et al. 2025b, *A&A*, **699**, A231
- Sokoliuk, O. 2025, *A&A*, **699**, A59
- Somerville, R. S., Yung, L. Y. A., Lancaster, L., et al. 2025, *MNRAS*, **544**, 3774
- Springel, V. 2010, *MNRAS*, **401**, 791
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, **475**, 676
- Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2021, *ApJ*, **922**, 29
- Steinhardt, C. L., Kokorev, V., Rusakov, V., Garcia, E., & Sneppen, A. 2023, *ApJ*, **951**, L40
- Sun, G., Faucher-Giguère, C.-A., Hayward, C. C., et al. 2023, *ApJ*, **955**, L35
- Tacchella, S., Eisenstein, D. J., Hainline, K., et al. 2023, *ApJ*, **952**, 74
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, **688**, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, **724**, 878
- Topping, M. W., Stark, D. P., Senchyna, P., et al. 2024, *MNRAS*, **529**, 3301
- Traina, A., Gruppioni, C., Delvecchio, I., et al. 2024, *A&A*, **681**, A118
- Traina, A., Gruppioni, C., Delvecchio, I., et al. 2026, *A&A*, **705**, A255
- Trinca, A., Schneider, R., Valiante, R., et al. 2024, *MNRAS*, **529**, 3563
- Trump, J. R., Arrabal Haro, P., Simons, R. C., et al. 2023, *ApJ*, **945**, 35
- van Dokkum, P., & Conroy, C. 2024, *ApJ*, **973**, L32
- van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, *ApJ*, **709**, 1018
- Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, *MNRAS*, **455**, 4183
- Wang, L., Gao, F., Best, P. N., et al. 2021, *A&A*, **648**, A8
- Wang, B., Leja, J., Atek, H., et al. 2024a, *ApJ*, **963**, 74
- Wang, B., Leja, J., de Graaff, A., et al. 2024b, *ApJ*, **969**, L13
- Wang, T., Sun, H., Zhou, L., et al. 2025, *ApJ*, **988**, L35
- Waterval, S., Cannarozzo, C., & Macciò, A. V. 2025, *MNRAS*, **537**, 2726
- Weaver, J. R., Davidzon, I., Toft, S., et al. 2023, *A&A*, **677**, A184
- Weibel, A., Oesch, P. A., Barrufet, L., et al. 2024, *MNRAS*, **533**, 1808
- Weibel, A., Oesch, P. A., Williams, C. C., et al. 2026, *ApJ*, **1002**, 136
- Weidner, C., & Kroupa, P. 2005, *ApJ*, **625**, 754
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, *MNRAS*, **465**, 3291
- Wilkins, S. M., Lovell, C. C., & Stanway, E. R. 2019, *MNRAS*, **490**, 5359
- Wilkins, S. M., Lovell, C. C., Fairhurst, C., et al. 2020, *MNRAS*, **493**, 6079
- Wilkins, S. M., Vijayan, A. P., Lovell, C. C., et al. 2023, *MNRAS*, **518**, 3935
- Xiao, M., Oesch, P. A., Elbaz, D., et al. 2024, *Nature*, **635**, 311
- Xiao, M., Bing, L., Brammer, G., et al. 2026, *A&A*, submitted, [arXiv:2602.03030]
- Yang, T., Wang, T., Xu, K., et al. 2026, *ApJ*, **1000**, L42
- Yu, H., & Wang, F. Y. 2016, *ApJ*, **820**, 114
- Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Wilkins, S. M., & Gardner, J. P. 2024a, *MNRAS*, **527**, 5929
- Yung, L. Y. A., Somerville, R. S., Nguyen, T., et al. 2024b, *MNRAS*, **530**, 4868
- Yung, L. Y. A., Somerville, R. S., & Iyer, K. G. 2025, *MNRAS*, **543**, 3802
- Ziegler, J. J., Freese, K., Lozano, J., & Montefalcone, G. 2025, arXiv e-prints [arXiv:2507.21409]

Table A.1. Best-fitting parameters to Equation (A.1) for the SFR function at redshifts $z \geq 6$.

z	$\log_{10} \phi^*$	$\log_{10} \dot{M}_{\star, \text{SFR}}^*$	α	σ
6	$-3.425^{+0.027}_{-0.027}$	$0.712^{+0.027}_{-0.027}$	$1.995^{+0.002}_{-0.002}$	$0.220^{+0.012}_{-0.012}$
7	$-3.425^{+0.007}_{-0.007}$	$0.712^{+0.004}_{-0.004}$	$1.995^{+0.003}_{-0.003}$	$0.19^{+0.001}_{-0.001}$
8	$-3.414^{+0.051}_{-0.052}$	$0.367^{+0.038}_{-0.037}$	$2.191^{+0.007}_{-0.007}$	$0.305^{+0.014}_{-0.014}$
9	$-4.148^{+0.104}_{-0.111}$	$0.336^{+0.058}_{-0.055}$	$2.525^{+0.054}_{-0.054}$	0.412
10	$-3.406^{+0.102}_{-0.099}$	$-0.116^{+0.066}_{-0.069}$	$2.288^{+0.026}_{-0.028}$	$0.519^{+0.019}_{-0.019}$
11	$-3.917^{+0.125}_{-0.128}$	$0.229^{+0.082}_{-0.083}$	$2.278^{+0.036}_{-0.039}$	$0.399^{+0.023}_{-0.024}$

Appendix A: High-redshift star formation rate function

We fit the collection of data for the SFR function, as described in Section 4.1 (Figure 2), with a modified Schechter function following Saunders et al. (1990)

$$\phi(\dot{M}_{\star, \text{SFR}}) d \log \dot{M}_{\star, \text{SFR}} = \dot{M}_{\star, \text{SFR}}^* \left(\frac{\dot{M}_{\star, \text{SFR}}}{\dot{M}_{\star, \text{SFR}}^*} \right)^{1-\alpha} \exp \left[-\frac{1}{2\sigma^2} \log_{10}^2 \left(1 + \frac{\dot{M}_{\star, \text{SFR}}}{\dot{M}_{\star, \text{SFR}}^*} \right) \right] d \log \dot{M}_{\star, \text{SFR}} \quad (\text{A.1})$$

The best-fitting parameters of the SFR function to Equation (A.1) at each redshift are shown in Table A.1. Figure A.1 shows, as an example of robustness of our fit, the posterior distributions of the parameters fitted at redshift $z = 10$, and we obtained also similar corner plots for the fits at any other redshifts considered in this work. We also fit the redshift evolution of each parameter of the Saunders formula, using minimum likelihood method, finding the following analytic redshift-dependent fits

$$\log_{10} \phi^*(z) = -3.43 + 1.59 \cdot \log_{10}(1 + \tilde{z}) - 6.39 \cdot \log_{10}^2(1 + \tilde{z}) + 4.83 \cdot \log_{10}^3(1 + \tilde{z}), \quad (\text{A.2})$$

$$\log_{10} \dot{M}_{\star, \text{SFR}}^*(z) = 0.71 + 2.62 \cdot \log_{10}(1 + \tilde{z}) - 11.06 \cdot \log_{10}^2(1 + \tilde{z}) + 8.69 \cdot \log_{10}^3(1 + \tilde{z}), \quad (\text{A.3})$$

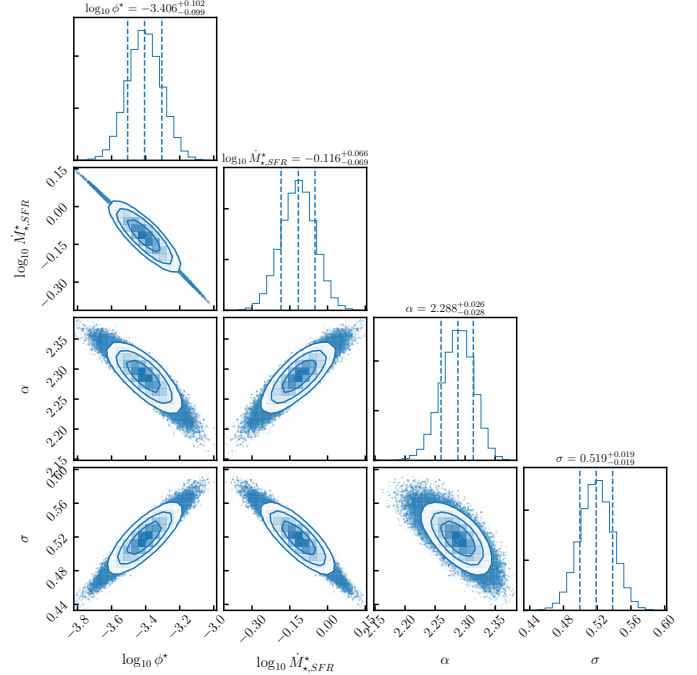
$$\alpha(z) = 2 - 2.48 \cdot \log_{10}(1 + \tilde{z}) + 10.8 \cdot \log_{10}^2(1 + \tilde{z}) - 9.23 \cdot \log_{10}^3(1 + \tilde{z}), \quad (\text{A.4})$$

$$\sigma(z) = 0.22 - 1.52 \cdot \log_{10}(1 + \tilde{z}) + 5.88 \cdot \log_{10}^2(1 + \tilde{z}) - 4.59 \cdot \log_{10}^3(1 + \tilde{z}), \quad (\text{A.5})$$

where $\tilde{z} = z - z_0$ and $z_0 = 6$.

Appendix B: Star formation rate function for different initial mass functions

Figure B.1 shows the SFR function for the different IMFs we explore in Section 4.8. For each input IMF, we repeat the abundance matching by building a new SFR function depending on the L_{UV} -to-SFR conversion corresponding to the assumed IMF.

**Fig. A.1.** Poster distribution of the fitting parameters to Equation (A.1) for the star formation rate function at redshift $z = 10$.

Appendix C: Dust attenuation laws

We first employ the Charlot & Fall (2000) attenuation law, according to which the intrinsic luminosity is correlated to the observed luminosity with the following power law

$$L_{\text{int}} = L_{\text{obs}} \times e^{\tau_{\lambda}}, \quad \tau_{\lambda} = \tau_{\lambda}^{\text{ISM}} \cdot \left(\frac{\lambda}{5500 \text{ \AA}} \right)^{-n} + \tau_{\lambda}^{\text{BC}} \cdot \left(\frac{\lambda}{5500 \text{ \AA}} \right)^{-n}. \quad (\text{C.1})$$

where λ is the observation wavelength and $(\tau_{\lambda}^{\text{BC}}, \tau_{\lambda}^{\text{ISM}}, n) = (1, 0.5, 0.7)$.

We secondly explore the Calzetti et al. (2000) attenuation law

$$L_{\text{int}} = L_{\text{obs}} \times 10^{0.4A_{\lambda}}, \quad (\text{C.2})$$

with $A_{\lambda} = k(\lambda) \cdot E(B - V)$. Here, $E(B - V)$ is the colour excess and $k(\lambda) = 2.659(-1.857 + 1.040/\lambda) + R_V$, with $R_V = 4.05$.

Lastly, we also include the modified Calzetti law as presented in Salim et al. (2018)

$$k(\lambda) = k_{\lambda, \text{Cal}} \frac{R_{V, \text{mod}}}{R_{V, \text{Cal}}} \left(\frac{\lambda}{5500 \text{ \AA}} \right)^{\delta} + D_{\lambda}, \quad R_{V, \text{mod}} = \frac{R_{V, \text{Cal}}}{(R_{V, \text{Cal}} + 1)(4400/5500)^{\delta} - R_{V, \text{Cal}}}, \quad \delta = -0.45 + 0.19(\log_{10}(M_{\star}/M_{\odot}) - 10), \quad (\text{C.3})$$

where $R_{V, \text{Cal}} = 4.05$ and D_{λ} is the Drude profile UV bump (Fitzpatrick & Massa 1986).

We note that the above correction for dust depends on the specific observation wavelength of the sources. We checked that by applying the correction from the above recipes within the wavelengths of interest the bright end of the galaxy LF matches the observed one in the IR band, but naturally overestimates the faint end where the dust content should be significantly lower (Figure C.1).

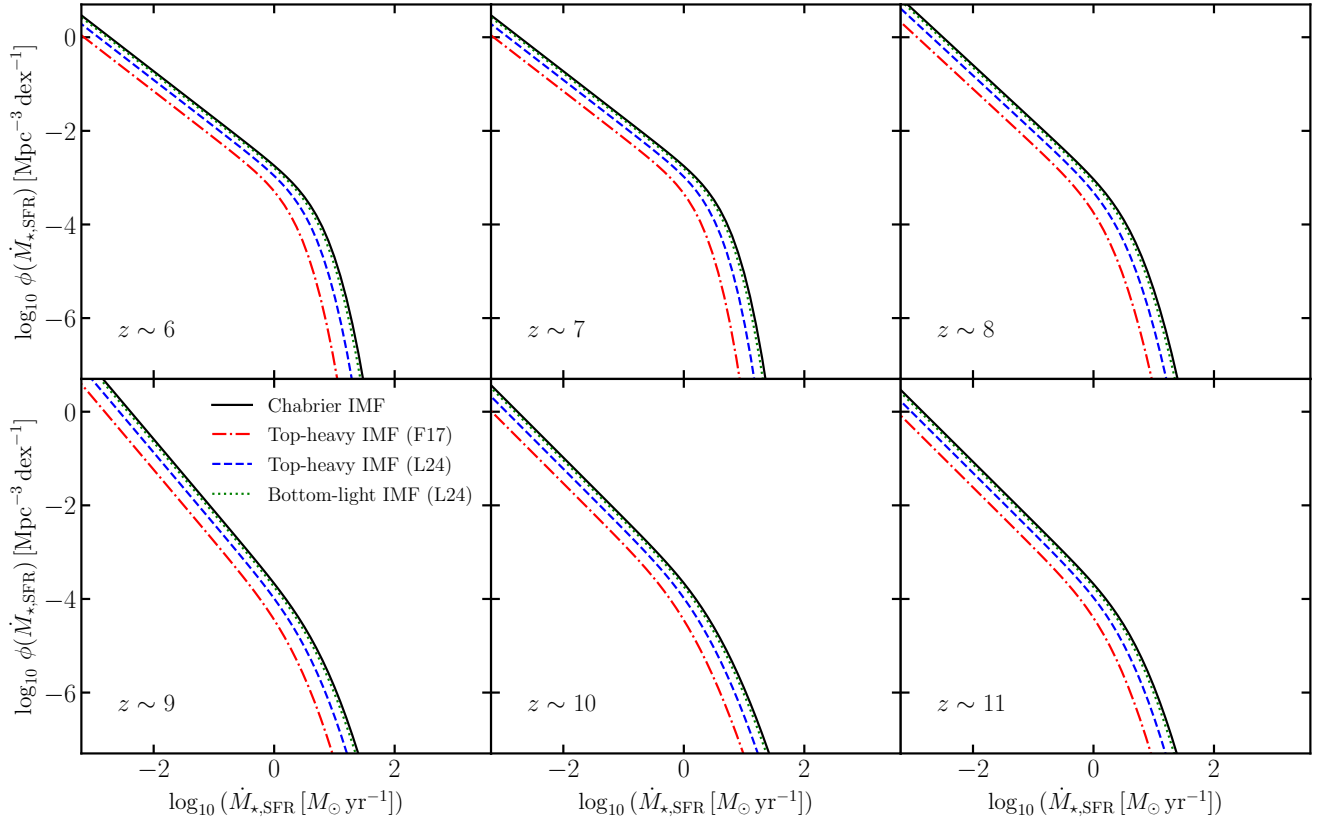


Fig. B.1. Star formation rate function from redshifts $z = 6$ to 11 for different assumed initial mass functions: canonical (black solid lines; Chabrier 2003), top-heavy F17 (red dash-dotted lines; F17), top-heavy L24 (blue dashed lines; Lapi et al. 2024), and bottom-light (green dotted lines; Lapi et al. 2024). The red dash-dotted lines refer to the F17 IGIMF for $\text{SFR} = 10^3 M_{\odot}/\text{yr}$.

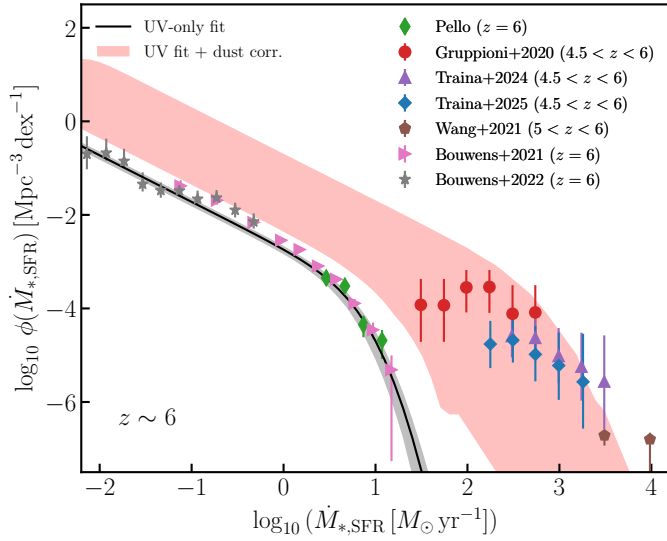


Fig. C.1. Star formation rate function at redshift $z = 6$ from the UV-only fit and by including the dust correction, compared to several observations both in the UV and IR bands.