






Uncertainties in high- z galaxy properties inferred from spectral energy distribution fittings using JWST NIRC*am* photometry

Jiyoung Choe¹, Taysun Kimm^{1,*} , Harley Katz^{2,3} , Maxime Rey¹ , Daniel Han¹ ,
Jaekyoung Jang¹ , and Joki Rosdahl⁴

¹ Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

² Department of Astronomy & Astrophysics, University of Chicago, 5640 S Ellis Avenue, Chicago, IL 60637, USA

³ Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

⁴ Centre de Recherche Astrophysique de Lyon UMR5574, Univ Lyon, Univ Lyon1, Ens de Lyon, F-69230 Saint-Genis-Laval, France

Received 1 September 2025 / Accepted 15 January 2026

ABSTRACT

Numerous high- z galaxies have recently been observed with the James Webb Space Telescope (JWST), providing new insights into early galaxy evolution. Their physical properties are typically derived through spectral energy distribution (SED) fitting, but the reliability of this approach remains uncertain owing to limited constraints on star formation histories (SFHs) and on the contribution from emission for such early systems. Applying BAGPIPES on simulated SEDs with $\text{SFR}_{10} > 0.3 M_{\odot} \text{yr}^{-1}$ at $z = 6$ from the SPHINX²⁰ cosmological simulation, we examine the uncertainties related to the recovery of stellar masses, star formation rates (SFR_{10}), and stellar metallicities from mock JWST/Near-Infrared Camera photometry, spanning F115W–F444W. Even without dust or emission lines, fitting the intrinsic stellar continuum overestimates the stellar mass by about 60%, on average (and by up to a factor of five for low-mass galaxies with recent starbursts). It also underestimates the SFR_{10} by a factor of 2, due to inaccurate SFHs and age–metallicity degeneracies. In full SED-fitting models that include dust attenuation and nebular emission, stellar mass estimates are primarily affected by age–metallicity degeneracy and emission lines. Short-term SFRs are most sensitive to dust attenuation and nebular emission, while long-term SFRs additionally depend on the assumed SFHs. Incorporating bands that are free of strong emission lines, such as F410M, helps mitigate stellar mass overestimation by disentangling line emission from older stellar populations. We also find that best fit or likelihood-weighted estimates are generally more accurate than median posterior values. Although stellar mass functions are reproduced reasonably well (particularly when the minimum- χ^2 estimates are used), the slope of the main sequence of star formation acutely depends on the adopted fitting model. Overall, these results underscore the importance of careful modelling when interpreting high- z photometry, particularly for galaxies with recent star formation burst and/or strong emission lines, to minimise systematic biases in derived physical properties.

Key words. galaxies: evolution – galaxies: high-redshift – galaxies: star formation

1. Introduction

Understanding how galaxies form and evolve at high redshift is vital for elucidating the earliest phases of cosmic history. After recombination, gas accumulates along the cosmic web of dark matter filaments, leading to the formation of the first stars and galaxies (e.g. Abel et al. 2002; Bromm et al. 2002). These primordial systems mark the onset of galaxy formation and they drive the reionisation of the intergalactic medium by emitting ionising photons (e.g. Dayal & Ferrara 2018; Robertson 2022). As they evolve hierarchically through mergers and baryonic processes, these high-redshift galaxies preserve crucial information regarding the interplay among structure formation, feedback, and the thermal and ionisation states of the Universe (e.g. Somerville & Davé 2015). Consequently, examining their properties is fundamental to building a coherent picture of galaxy evolution across cosmic time.

Several approaches are typically employed to investigate the evolution of galaxies in these early stages, including semi-analytic models of galaxy formation (e.g. Yung et al. 2019; Mauerhofer et al. 2025) and analytic methods (e.g. Dekel et al. 2023). In particular, recent advances in computational power and numerical methods have improved the fidelity of simula-

tions modelling the interstellar medium (ISM) in greater detail (Naab & Ostriker 2017). By incorporating strong stellar feedback, numerical simulations are able to successfully reproduce observed ultraviolet (UV) luminosity functions, stellar mass functions, and mass-metallicity relations at high redshift (e.g. Ma et al. 2016; Rosdahl et al. 2018; Finlator et al. 2018; Kannan et al. 2020). Such simulations provide theoretical frameworks for generating mock observations of high- z galaxies, enabling more direct comparisons with existing data and enhancing our understanding of early galaxy formation (Marshall et al. 2022; Katz et al. 2023).

In parallel, photometric and spectroscopic data continue to provide essential observational constraints on the nature of high redshift galaxies. For instance, oxygen abundances can be derived from strong-line diagnostics and the direct T_e (electron temperature) method (Curti et al. 2024; Nakajima et al. 2023; Chemerynska et al. 2024), while star formation rates (SFRs) are inferred from $H\alpha$ or [CII] line luminosities (De Looze et al. 2014; Smit et al. 2018; Béthermin et al. 2020; Rinaldi et al. 2023). While such spectroscopic observations provide direct and reliable insights into galaxy properties, they require substantial observation time and resources. Therefore, as a more feasible alternative for large samples, spectral energy distribution (SED) fittings based on photometric data are widely employed. For instance, Song et al. (2016) performed SED fitting of

* Corresponding author: tkimm@yonsei.ac.kr

ultraviolet (UV)-selected samples at $4 < z < 8$ and obtained UV luminosity functions and their stellar mass functions (SMFs). Building on this and using photometry spanning the rest-frame UV to near-infrared (NIR), Leung et al. (2025) investigated the observed ‘little red dots’ and demonstrated that including James Webb Space Telescope/Mid-Infrared Instrument (JWST/MIRI) bands and an active galactic nucleus (AGN) emission component substantially reduced stellar mass estimates, alleviating the apparent tension with respect to current cosmological models. Harvey et al. (2025) measured stellar masses for galaxies at $6.5 < z < 13.5$ using various stellar population synthesis (SPS) models and initial mass functions (IMFs); they proposed that the low-mass slope of the SMFs steepens toward higher redshift. Collectively, these empirical constraints are central to understanding galaxy assembly and serve as critical benchmarks for validating theoretical models (e.g. Mauerhofer et al. 2025).

Numerous publicly available SED fitting codes support this approach (Carnall et al. 2018; Chevallard & Charlot 2016; da Cunha et al. 2015; Leja et al. 2017), enabling the estimation of posterior distributions for physical parameters such as stellar mass by fitting observed fluxes across multiple bandpasses. In a recent comparative analysis, Pacifici et al. (2023) observed that while stellar mass estimates are generally robust, the SFR and dust attenuation (A_V) vary markedly with modelling assumptions, including the choice of star formation history (SFH), dust law, and nebular emission treatment. Among these factors, the assumed parametric SFH model proves particularly important for inferring the properties of quenched galaxies, as demonstrated using both mock and observed photometric datasets. In line with this approach, Carnall et al. (2018) reported that a double power-law SFH model yields an unbiased quenching timescale, whereas an exponentially declining SFH model underestimates both the stellar age and the quenching timescale. The reliability of the parametric SFH model was further tested through SED fitting of mock photometry derived from simulated galaxies (Haskell et al. 2024), revealing that such models typically underestimate the SFR of recently quenched systems by ~ 0.4 dex, but they do tend to overestimate the SFR of starburst galaxies. The choice of stellar population synthesis template likewise influences stellar mass estimates. For instance, using different combinations of stellar population and dust emission models for galaxies from the Cosmic Evolution Survey (Scoville et al. 2007), Jones et al. (2022) reported that stellar masses are, on average, 0.14 dex lower when the Binary Population and Spectral Synthesis (BPASS) model was employed, instead of the model introduced by Bruzual & Charlot (2003, 2016 version¹). In a related study, Meldorf et al. (2024) investigated whether a flexible dust attenuation law could recover the input dust parameters. They found that the correlation between the true slope of the dust attenuation curve and the residual is difficult to eliminate, even with the inclusion of IR bands. The importance of adopting non-parametric or flexible SFH models has also been increasingly emphasised over the past few years. Although they are computationally intensive, such models achieve greater accuracy in recovering the true SFH and associated physical parameters, especially for galaxies with stochastic SFHs or bursty young stellar populations (Lower et al. 2020; Jain et al. 2024).

In previous studies, SED fittings conducted on high- z galaxies were primarily confined to either highly luminous systems or those in lensed fields (McLure et al. 2011; Laporte et al. 2017; Jolly et al. 2021). However, with the successful launch of the James Webb Space Telescope (JWST), unprecedented volumes

of data are now available at much fainter magnitudes. Nevertheless, as JWST observations accumulate, new challenges appear, with the most notable being the apparent excess of luminous galaxies (Ferrara et al. 2023; Finkelstein et al. 2023; Carniani et al. 2024; Harikane et al. 2024; Sabti et al. 2024). Current theoretical frameworks suggest that the observed number of UV-bright galaxies is too high to align with the classical picture of galaxy formation models (Lovell et al. 2022). A plausible explanation is that star formation in these galaxies proceeds in a highly bursty manner (e.g. Sun et al. 2023; Shen et al. 2023). As spectroscopic confirmations of such galaxies increase (Carniani et al. 2024), the aforementioned tension with the canonical framework becomes more pronounced. Due to the high optical fluxes observed in these systems, several studies have either sought to attribute some of the emission to AGNs or refined the modelling of young stellar populations (Leung et al. 2025; Harvey et al. 2025). Nevertheless, quantifying the influence of AGNs and bursty star formation remains difficult, underscoring the importance of carefully assessing uncertainties in SED fitting. Recently, Cochrane et al. (2025) examined the impact of SED fitting on SMFs using mock JWST photometry of SPHINX²⁰ galaxies (Rosdahl et al. 2022; Katz et al. 2023). They concluded that stellar masses of low-mass galaxies are consistently overestimated because of inadequate modelling of strong emission lines.

While accurately inferring galaxy properties is essential, the accuracy of SED fitting and its associated uncertainties in high- z galaxies have received relatively little attention (c.f. Narayanan et al. 2024; Cochrane et al. 2025). Galaxies formed during the epoch of reionisation are believed to differ notably from those in the local Universe, frequently displaying active and bursty star formation (Caputi et al. 2017; Rinaldi et al. 2022; Dressler et al. 2023, 2024; Hu et al. 2023). Accordingly, the nebular emission is expected to contribute substantially to their total luminosity (Wilkins et al. 2020; Mármol-Queraltó & McLure 2015). Because these systems are at an early evolutionary stage, metallicity and dust extinction are generally low (Traina et al. 2024; Heintz et al. 2023), whereas the escape fraction of LyC radiation can be high (Hayes et al. 2011; Kimm et al. 2017). Determining whether SED fitting, which is validated primarily with local galaxies, can reliably recover the properties of high- z systems is therefore crucial. To this end, we use mock observations from the SEDs of simulated galaxies at $z = 6$ in the SPHINX²⁰ cosmological simulation, using JWST/NIRCam photometry. Furthermore, we applied SED fitting with BAGPIPES to evaluate the reliability of the derived physical quantities, such as stellar mass and SFR, and to assess how these affect the determination of the SMFs and the star-forming main sequence.

This paper is structured as follows. Section 2 briefly introduces the SPHINX²⁰ simulation, outlines the construction of mock SEDs for simulated galaxies, and describes the fitting procedure with BAGPIPES. Section 3 presents an evaluation of the accuracy of galaxy property recovery under different fitting configurations and an investigation of the sources of associated uncertainties and biases. Section 4 presents strategies for improving fitting accuracy and an assessment of discrepancies between the true and fitted statistical observables. Finally, Sect. 5 summarises our results and conclusions, while the appendix presents detailed fitting results.

2. Methods

To assess how emission lines influence observed photometry and introduce uncertainties in stellar mass and SFH estimations, we

¹ This updated version is available at: <https://www.bruzual.org/bc03>.

analysed mock spectra from the SPHINX²⁰ cosmological simulation. In this section, we outline the construction of mock SEDs for each galaxy and the generation of photometric data for the JWST Near-Infrared Camera (NIRCam) bands.

2.1. Simulation

The SPHINX suite comprises cosmological radiation-hydrodynamic simulations designed to investigate galaxy properties during the epoch of reionisation (Rosdahl et al. 2018, 2022). With a spatial resolution of 10.9 pc (physical) at $z = 6$ and a dark matter resolution of $2.5 \times 10^5 M_\odot$, the SPHINX²⁰ simulation captures the evolution of roughly 32 000 star-forming halos at $z = 6^2$ and resolves galaxies hosted by dark matter halos with masses down to the atomic cooling limit ($M_{\text{vir}} \sim 10^8 M_\odot$). Dark matter halos are identified using the ADAPTAHOP halo finder (Tweed et al. 2009) and star particles are assigned to the nearest (sub)halos. Galaxies are defined as the associations of gas cells and star particles within each (sub)halo. The large volume of the SPHINX²⁰ (20 cMpc)³ allows an analysis of the photometric and spectroscopic properties of galaxies spanning stellar masses up to $M_\star \approx 10^{10} M_\odot$ (Katz et al. 2023).

In SPHINX²⁰, gas cooling from hydrogen and helium is computed by solving non-equilibrium chemistry coupled with ionising radiation (Rosdahl et al. 2013), while metal and molecular cooling are incorporated, following the approach of Rosen & Bregman (1995). As described by Kimm et al. (2017), Rosdahl et al. (2018), stellar particles form when the gas collapses and becomes gravitationally unstable, with the star formation efficiency regulated by local turbulence. When a stellar particle with a mass as low as $m_\star = 400 M_\odot$ forms, ionising radiation is injected into its host cell by interpolating SEDs of different ages and metallicities from the BPASS model (Stanway & Eldridge 2018, version 2.2.1). This radiation then propagates through the interstellar medium (ISM), heating nearby gas and transferring momentum to the absorbing cells (Rosdahl et al. 2013). After 4–50 Myr of stellar evolution, supernovae occur and enrich both the ISM and the circumgalactic medium (Kimm & Cen 2014; Kimm et al. 2015). Additional details regarding the physical modelling in SPHINX²⁰ are provided by Rosdahl et al. (2018, 2022).

The SPHINX simulations reproduce the UV and Ly α luminosity functions at $z \geq 6$ (Garel et al. 2021). To further demonstrate that the simulations provide a realistic sample whose photometric properties resemble those of galaxies at $z \approx 6$, we compared the SPHINX²⁰ galaxies with NIRCam-selected sources in the CEERS and NGDEEP fields. As shown in Fig. 1, the SPHINX²⁰ DR galaxies occupy nearly the same region as the observed population in the F150W–F277W versus F277W–F444W colour–colour plane, indicating that their broad-band colours are representative of typical galaxies at this epoch.

2.2. Mock SEDs

Katz et al. (2023) post-processed 1,380 star-forming galaxies from SPHINX²⁰ with $\text{SFR}_{10} > 0.3 M_\odot \text{yr}^{-1}$ at $4.64 < z < 10$, where SFR_{10} is the star formation rate (SFR) averaged over the last 10 Myr, deriving dust-attenuated SEDs for ten viewing angles per galaxy. The sample was restricted to post-processed galaxies that are likely to be observable with JWST, adopting a limiting UV magnitude of $M_{\text{UV}} \approx -17$ (Katz et al. 2023;

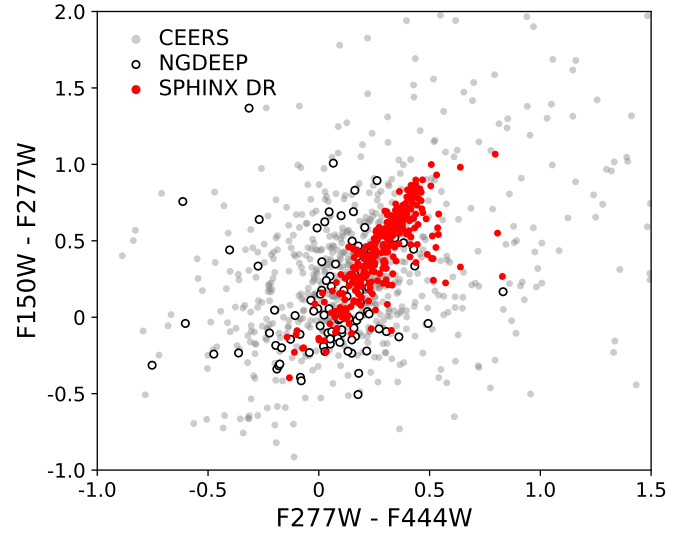


Fig. 1. Observed NIRCam colour–colour distribution of galaxies at $5.8 < z_{\text{phot}} < 6.2$ in the CEERS and NGDEEP fields (grey points and open circles), compared with the SPHINX²⁰ DR $z = 6$ galaxies used in our analysis (red points). The SPHINX²⁰ galaxies trace the same region in the F150W–F277W versus F277W–F444W plane as the observed population, indicating that their broad-band colours are consistent with those of typical star-forming galaxies at $z \approx 6$.

Rieke et al. 2023). In particular, using the same IMF parameters as in the simulation – a mass cutoff of 0.1–100 M_\odot with slopes of -1.3 at low mass and -2.35 at high mass – the authors first computed the intrinsic stellar continua with BPASS v2.2.1. They subsequently incorporated the contribution of 52 emission lines and the nebular continuum (free-free, free-bound, and two-photon) using CLOUDY (Ferland et al. 2017, v17.03). These emission lines include strong lines such as Ly α 1216 Å, [OIII] 4959/5007 Å, and H α 6563 Å, which strongly influence broad-band photometry (Zackrisson et al. 2008; Schaerer & de Barros 2009). For the H and He emission lines, recombinative and collisional rates were derived from non-equilibrium ionisation fractions directly extracted from the simulations when the local Stromgren sphere was resolved. Conversely, in under-resolved cells, the line luminosities were estimated from pre-computed CLOUDY model grids under the assumption of spherical geometry. The resulting intrinsic SEDs were then propagated with the Monte Carlo radiative transfer code RASCAS (Michel-Dansac et al. 2020), which self-consistently models resonant and non-resonant line transfer together with absorption and scattering by Small Magellanic Cloud (SMC) type dust. The dust content was assumed to scale with the neutral hydrogen density, following the model proposed by Laursen et al. (2009). Importantly, the attenuation curve of a galaxy does not necessarily resemble the SMC curve, as it is also determined by the spatial distribution of dust relative to the stars and by the total dust content (Narayanan et al. 2018; Trayford et al. 2019).

Figure 2 presents an SED of a SPHINX²⁰ dwarf galaxy with $M_\star = 10^{7.7} M_\odot$ (black line). The solid orange and yellow lines denote contributions from the intrinsic stellar and nebular continua, respectively, while the corresponding magnitudes in the JWST NIRCam filter system are plotted as filled circles with error bars. We find that dust attenuation notably reddens the UV slope relative to the intrinsic stellar continuum (e.g. Katz et al. 2023). The nebular continuum further modifies the SED (e.g. Katz et al. 2025; Narayanan et al. 2025), especially near the

² The SPHINX²⁰ simulation is run down to $z = 4.64$, where it contains approximately 32 500 star-forming halos.

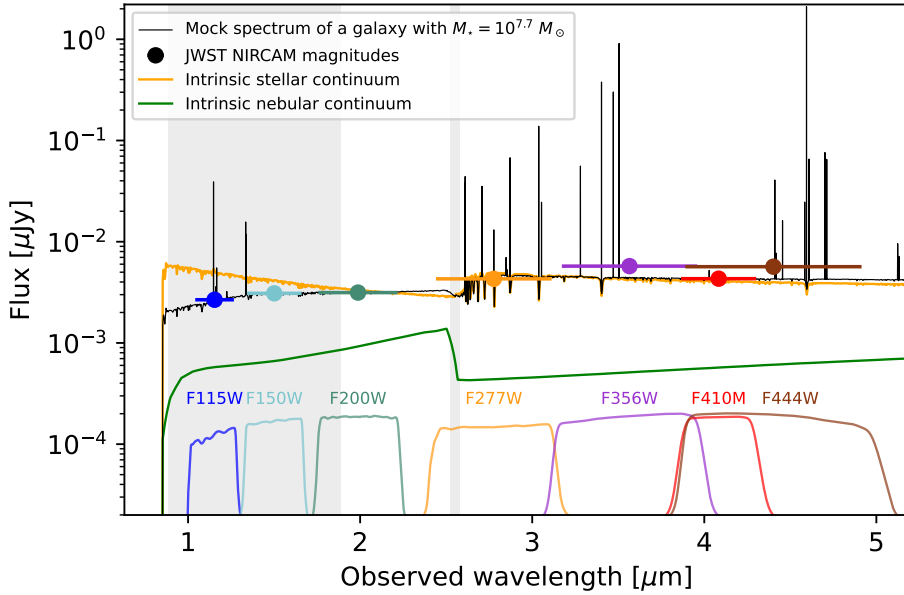


Fig. 2. Example SED of a SPHINX²⁰ galaxy at $z = 6$. The black line represents the attenuated spectrum along a random line of sight. Coloured points mark the pivot wavelengths and bandwidths of the seven JWST NIRCAM filters, with corresponding throughput curves shown below. The orange and green lines indicate the intrinsic stellar and nebular continua, respectively. The shaded regions indicate the wavelength range used to measure the UV slope (β) and the position of the Balmer break, respectively.

Balmer break, and can become bright enough to surpass the stellar continuum depending on the galaxy’s star formation history. Moreover, certain band magnitudes (e.g. F356W) are occasionally brighter than the continuum, suggesting that emission lines can significantly influence SED fitting.

To ensure both a sufficiently high redshift and an adequate sample size, we based our analysis on all 276 galaxies in the SPHINX²⁰ data release at $z = 6$ ³, each represented by ten spectra derived from ten distinct viewing angles. The stellar masses and mass-weighted stellar metallicities of these galaxies span $10^{6.6} - 10^{9.8} M_{\odot}$ and $0.002 - 0.309 Z_{\odot}$, respectively, where the solar metallicity is $Z_{\odot} = 0.02$. For the photometric coverage, we employed six JWST NIRCAM filters (F115W, F150W, F200W, F277W, F356W, and F444W), which have also been used in the Next Generation Deep Extragalactic Exploratory Public (NGDEEP) survey (Bagley et al. 2024). At $z = 6$, the filters cover the rest-frame far-UV to optical wavelength range of galaxies. The role of the medium-band filter is addressed in a later section.

2.3. SED fitting

We employed the widely used Bayesian inference code BAGPIPES (Carnall et al. 2018) to infer galaxy properties from SED fitting. BAGPIPES utilises the MultiNest sampling algorithm (Feroz & Hobson 2008; Feroz et al. 2009), which efficiently explores a wide range of parameter spaces. In the following, we outline the input models for SFH, nebular emission, and dust adopted in the SED modelling.

We adopted the BPASS v2.2.1 stellar population synthesis model coupled with the Kroupa & Boily (2002) IMF. This template is the same as that employed to generate the mock simulated SEDs in Katz et al. (2023)⁴. For the SFHs, we explored two

common parametric models (constant and double-power law), a flexible non-parametric model, and the normalised true SFHs. The constant SFH model prescribes a fixed SFR over a defined interval, with the SED shape governed by the initial and final stellar ages. Although straightforward and intuitive, the constant SFH cannot capture rising or declining trends, which are more effectively described by the double-power law,

$$\text{SFR}(t) \propto \left[\left(\frac{t}{\tau} \right)^{\alpha} + \left(\frac{t}{\tau} \right)^{-\beta} \right]^{-1}, \quad (1)$$

where τ denotes the time of peak star formation and $\alpha > 0$ and $\beta > 0$ represent free parameters defining the rising and declining phases, respectively. For the flexible model, we adopted the method of Leja et al. (2019), which prescribes a multi-component SFH with a constant SFR in each time bin. Rather than permitting fully unconstrained SFRs, we imposed a continuity prior defined as $x = \log(\text{SFR}_n / \text{SFR}_{n+1})$, which follows a Student’s-t distribution, expressed as

$$\text{PDF}(x, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{(x/\sigma)^2}{\nu} \right)^{-\frac{\nu+1}{2}}. \quad (2)$$

Here, Γ denotes the gamma function, σ represents the scale parameter, and ν is a parameter controlling the tail width of the distribution. We adopted $\sigma = 0.3$ and $\nu = 2$, following Leja et al. (2019). The SFH was discretised into seven age bins, with the first bin covering 0–10 Myr. The remaining bins are evenly spaced in log time from 10 Myr to $t_{\text{max}} - 100$ Myr, where t_{max} denotes the age of the universe. The normalised true SFH model employs the intrinsic SFH obtained directly from the simulated galaxy and normalised by the total stellar mass. For all SFH models, we adopted uniform priors for the stellar mass formed (M_{\star}) and stellar metallicity (Z_{\star}), spanning $[10, 10^{13}] M_{\odot}$ and $[0.001, 1] Z_{\odot}$, respectively. We can use BAGPIPES to model the nebular continuum and line emission with the precomputed grids provided by Byler et al. (2017). These authors employed the photoionisation code CLOUDY to compute the nebular emission from H II regions using input spectra with varying simple stellar population ages and ionisation parameter, U , under the assumption that all hydrogen-ionising photons from young stars end up absorbed. We assumed that stars younger than 10 Myr

³ We also examined SED-fitting uncertainties for 66 galaxies at $z \approx 9$ and confirmed that our conclusions remain unchanged (Sect. 4.1.1).

⁴ The IMF may deviate from the canonical Kroupa form in high-redshift, metal-poor galaxies (e.g. Cameron et al. 2024; Kroupa et al. 2026). For example, adopting a top-heavy IMF would lower the inferred stellar masses by approximately 0.5 dex (Harvey et al. 2025). While a qualitative exploration of IMF variations is beyond the scope of this study, a systematic assessment of their impact is warranted and is left to future work.

contribute to nebular emission. We confirm that treating this age threshold as a free parameter has a negligible impact on the fitting results. The prior range for log U was selected as $[-4, -1]$.

For the dust model, we adopted three different attenuation laws: SMC (Gordon et al. 2003), Calzetti (Calzetti et al. 2000), and Salim (Salim et al. 2018). Each law is characterised by A_λ/A_V , with A_λ denoting the attenuation at wavelength λ . The SMC and Calzetti laws can be widely applied to high- z galaxies, which are generally metal-poor or actively star-forming. In contrast to the SMC and Calzetti laws, which fix the slope (A_λ/A_V) for each λ , the Salim dust law introduces a tunable slope through the parameter δ , as

$$(A_\lambda/A_V)_{\text{Sal}} = \left(\frac{\lambda}{5500 \text{ \AA}} \right)^\delta (A_\lambda/A_V)_{\text{Cal}} + D_\lambda, \quad (3)$$

where $(A_\lambda/A_V)_{\text{Sal}}$ and $(A_\lambda/A_V)_{\text{Cal}}$ represent the slopes of the Salim and Calzetti laws, respectively. Furthermore, D_λ regulates the strength of the UV bump at 2175 Å, which is set to zero in this study. The Salim law simplifies to the Calzetti law when $\delta = 0$, while positive (negative) δ values yield a flatter (steeper) slope. For all dust models, we used a uniform prior of $[0, 2]$ for A_V . For the Salim law, we impose a Gaussian prior on δ with mean of $\mu = 0.35$ and standard deviation of $\sigma = 0.446$, motivated by the actual δ distribution of the simulated SPHINX²⁰ galaxies.

During the SED fitting, we fixed the redshift to the true values of the sample galaxies to minimise any uncertainties arising from the redshift estimation. We assumed a signal-to-noise ratio of $S/N=10$ for all filter magnitudes, consistent with the approach of Narayanan et al. (2024)⁵. Throughout this study, the fitted value of each parameter is considered as the median of its posterior distribution. Additional uncertainties arising from redshift determination and from the effects of the summary statistic are discussed separately in Appendix C and Sect. 4.

3. Results

This study quantifies how different physical assumptions in SED fitting—such as SFH, dust attenuation, and nebular emission—affect uncertainties in the recovered properties of high- z galaxies. We systematically evaluated the influence of each component on the accuracy and bias of SED-derived quantities, with a particular focus on the sparse photometric coverage typical of broadband surveys. We first examined how model SFHs and stellar metallicity affect the posterior distributions of stellar mass and SFR, disregarding nebular emission and dust attenuation. We then gradually added further model components to assess their impact on the fitting results. Finally, we incorporated the nebular emission and quantitatively compared the inferred values with the simulated ones. The fitting sets and their configurations are summarised in Table 1.

3.1. Impact of the model SFH and metallicity

A long-standing challenge in stellar population studies is the age-metallicity degeneracy (e.g. Worthey 1994), wherein older stellar ages and higher metallicities yield similar broad-band colours, making them difficult to disentangle using photometry

⁵ We verified that our inferred physical quantities are largely insensitive to this assumption: adopting the NGDEEP limiting magnitudes, as well as the NGDEEP limiting magnitudes combined with CEERS Poisson noise, yields consistent results.

alone. Previous studies have shown that this degeneracy can be partly alleviated by incorporating spectroscopic features such as absorption lines (e.g. Gallazzi et al. 2005; Conroy 2013). However, in the absence of such spectral information, quantifying how this degeneracy influences stellar properties inferred from SED fitting remains essential.

3.1.1. Overestimation of stellar masses in the presence of recent starbursts

We began by fitting photometric data derived from the intrinsic stellar continua of simulated galaxies at $z = 6$ to evaluate how SFHs and metallicities affect inferred galaxy properties. For each SFH model, the SED fitting was performed twice: once with $Z_{\star, \text{true}}$, the true mass-weighted stellar metallicity (`intS_Ztrue`) and once with metallicity treated as a free parameter (`intS_Z`). Both fits assumed a single metallicity for all stellar populations within each galaxy. Figure 3 illustrates the offsets between the fitted and true M_\star and Z_\star . When stellar metallicity is fixed (`intS_Ztrue`), the model with the normalised true SFH recovers M_\star almost perfectly⁶, whereas the constant, double power law, and flexible SFH models show deviations of 0.09, 0.07, and 0.08 dex, respectively. The magnitude of overestimation is closely tied to the contribution from young stellar populations. For instance, in SEDs with recent star formation bursts, the blue spectrum can be reproduced either by lowering the contribution of old stars or by boosting the numbers of both old stars and young stars together. Because the posterior distribution favours older ages owing to the larger number of possible combinations, fits to SEDs with recent SF bursts frequently overestimate the stellar mass. Under the adopted selection criterion in the SPHINX²⁰ data release (i.e. $\text{SFR}_{10} \geq 0.3 M_\odot/\text{yr}$), our low-mass galaxies are typically dominated by young populations, resulting in intrinsically blue SEDs with UV slopes as steep as -2.77 (without nebular emission). This can drive stellar mass overestimates of up to an order of magnitude. Conversely, in galaxies with substantial old stellar populations, the constant SFH model tends to underestimate M_\star because the fit is biased toward the luminosity-weighted age, which is lower than the mass-weighted age. These biases do not occur when using the normalised true SFH is applied, as the ratios of young to old stars are fixed across all posteriors. Accurately capturing recent star formation bursts is therefore crucial for recovering stellar masses in dwarf galaxies with diverse SFHs.

3.1.2. Underestimation of stellar metallicities in parametric fitting models

When stellar metallicity is instead treated as a free parameter (`intS_Z`), the model with the normalised true SFH still recovers M_\star almost perfectly and Z_\star with reasonable accuracy. The fitted Z_\star tends to exceed the true value at intermediate metallicities ($\log Z_\star/Z_\odot \sim -1.3$), although the offset is not severe. This upward bias reflects the posterior's preference for young, metal-enriched populations that dominate the UV fluxes. Because not all stars in the simulation share the same Z_\star , the fitting outcome depends jointly on the SFH and the stellar metallicity distribution.

For the parametric SFH models, the age-metallicity degeneracy significantly impacts the posterior distribution, since lowering

⁶ If an alternative input SED is employed in the fitting, such as Bruzual & Charlot (2003), the stellar mass is overestimated by 0.04 dex because the BPASS SED generates greater optical flux (see Appendix A).

Table 1. Description of the fitting models.

Fitting label	$Z_{\star, \text{fit}}$	Input SED	Model SED	Remark
ints_Z _{true}	$Z_{\star, \text{true}}$	intrinsic S	intrinsic S	
ints_Z	free param	intrinsic S	intrinsic S	
atts_Z _{true}	$Z_{\star, \text{true}}$	attenuated S	attenuated S	
atts_Z	free param	attenuated S	attenuated S	
attsNE_Z _{true}	$Z_{\star, \text{true}}$	attenuated S+N+E	attenuated S+N+E	
attsNE_Z	free param	attenuated S+N+E	attenuated S+N+E	
attsNE_Z_F410M	free param	attenuated S+N+E	attenuated S+N+E	F410M added
attsN_Z_SNE	free param	attenuated S+N	attenuated S+N+E	no E in input SED
attsNE_Z_SN	free param	attenuated S+N+E	attenuated S+N	no E in model SED

Notes. Summary of fitting models. From left to right, the columns specify the model name, stellar metallicity adopted in the fit, input SED obtained from SPHINX²⁰, and the model SED used in BAGPIPES. $Z_{\star, \text{fit}}$ is set to the mass-weighted true stellar metallicity ($Z_{\star, \text{true}}$) unless treated as a free parameter. S, N, and E indicate the stellar continuum, nebular continuum, and nebular emission lines, respectively. The attSNE_Z_F410M model additionally incorporates medium-band photometry from the F410M filter. The SMC dust law is adopted as the default, with the Salim and Calzetti laws considered as alternatives.

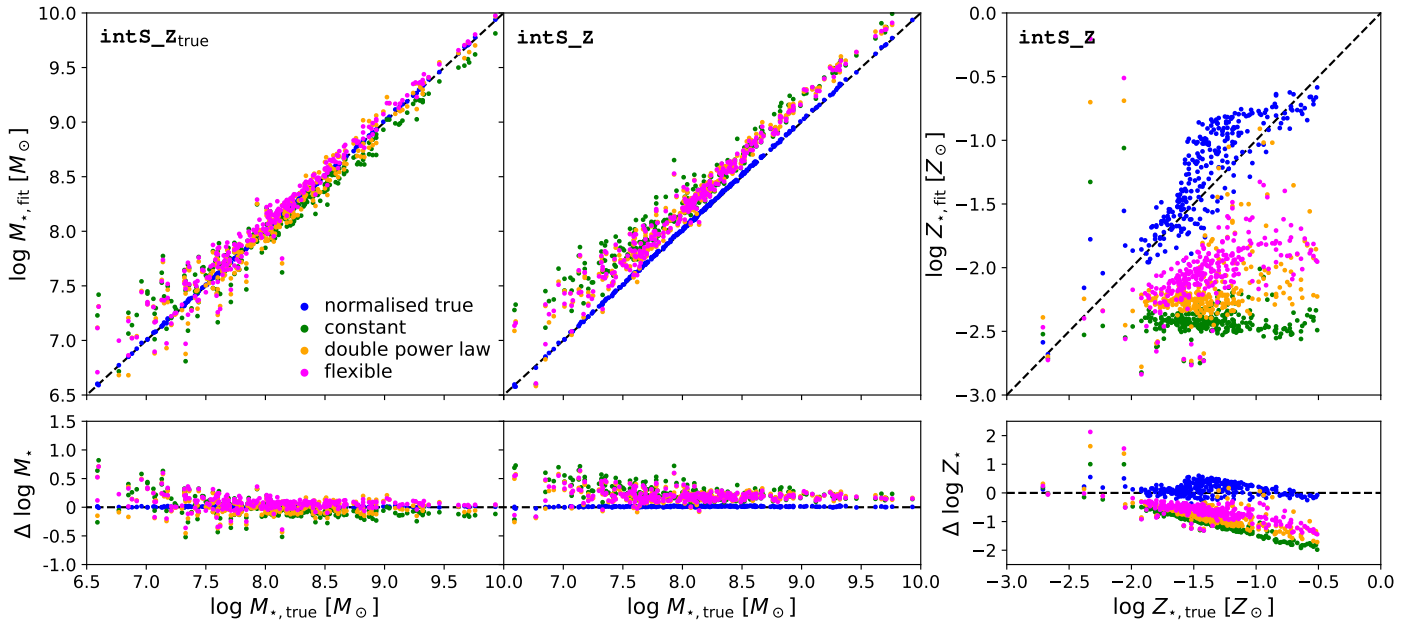


Fig. 3. Comparison of the true and derived stellar masses with (left; intS_Z_{true}) and without (right; intS_Z) fixing the stellar metallicity. Different SFH models are shown in different colours, as indicated in the legend. The right panel also displays the fitted Z_{\star} values for each SFH model, while the lower panels show the differences relative to the true quantities. When metallicity is treated as a free parameter, M_{\star} is overestimated by 0.2 dex for both parametric and non-parametric SFH models. This bias reflects the age–metallicity degeneracy, which tends to yield posteriors with underestimated metallicity and overly high fractions of old stars relative to the true values.

Z_{\star} or increasing the mass of young stars does end up yielding similarly blue SEDs. Our findings indicate that the photometric properties of the simulated galaxies are frequently reproduced by reducing Z_{\star} , while simultaneously enhancing the contribution from old populations, which, in turn, elevates M_{\star} . This outcome reflects the stronger degeneracy in SED modelling at low metallicities. Specifically, low- Z_{\star} SEDs permit a wider range of mass ratios between young and old stars, whereas high- Z_{\star} models necessitate the inclusion of young populations to offset redder colours. Consequently, the posterior distributions display a systematic bias towards low metallicities, with the constant SFH model often predicting values near the lower limit of the prior range (upper right panel of Fig. 3). Although this effect is weaker in the double-power law and non-parametric SFH models, high median offsets of 0.78 and 0.6 dex, respectively, are observed, indicating that the fitted metallicities are unreliable.

3.1.3. Underestimation of short-term SFRs and overestimation in long-term SFRs by recent starbursts

Our analysis also reveals notable biases in SFR_{100} ; for instance, the star formation rate averaged over the past 100 Myr, with its values overestimated when recent starbursts are present and underestimated when they are absent. Figure 4 presents the posterior SFHs for two representative galaxies, with stellar masses of $\log M_{\star}/M_{\odot} = 7.32$ (left panels) and 9.67 (right panels). In each panel, the blue, orange and red curves correspond to posterior samples from the intS_Z, attS_Z and attSNE_Z fits, respectively, selected to have SFR_{100} values near the median of their distributions. The constant SFH model (top panels) with intS_Z (blue lines) predicts the smallest stellar mass formed in the last 10 Myr and the lowest contribution from stars older than 100 Myr, concentrating most of the stellar mass in intermediate-age populations. Consequently, SFR_{100} is the

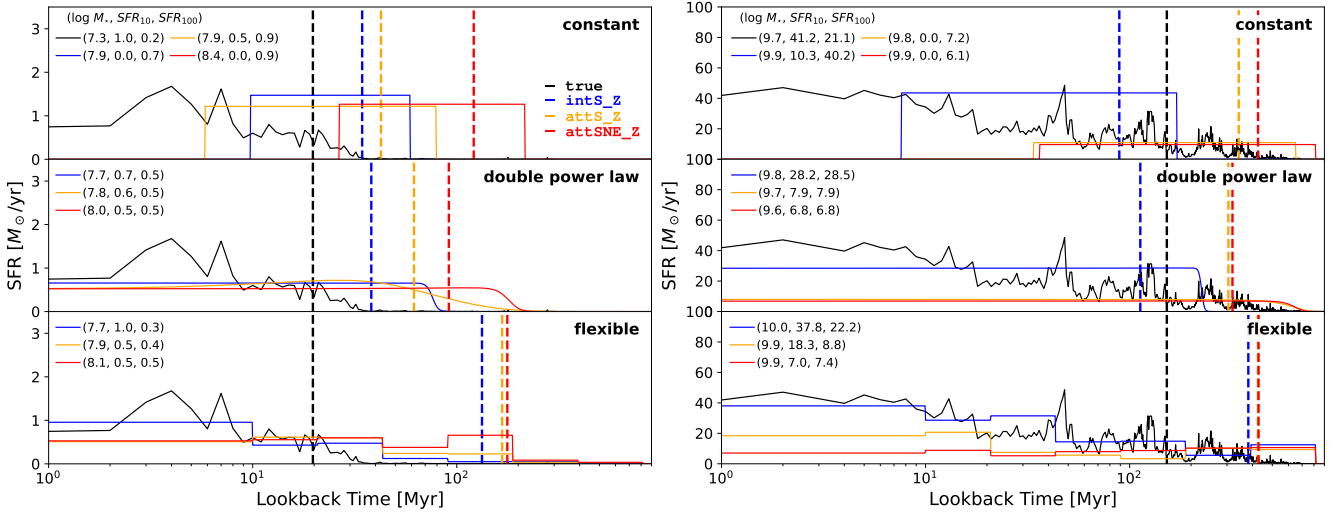


Fig. 4. Posterior SFHs corresponding to the median SFR_{100} values are displayed together with the true SFH (solid black line) for two representative galaxies: one low-mass ($\log M_*/M_\odot = 7.32$, left panels) and one high-mass ($\log M_*/M_\odot = 9.67$, right panels). Each row corresponds to a different SFH model used in the fitting, and each coloured line denotes a separate fitting configuration (blue: `intS_Z`, orange: `attS_Z`, red: `attSNE_Z`), with the SMC attenuation law applied when a dust model is included. Dashed vertical lines indicate the mass-weighted age. Under `intS_Z`, the constant SFH model generates the fewest young stars (<10 Myr) among all models and yields the lowest mass-weighted age. Consequently, its strongest underestimation is SFR_{10} and it also overestimates SFR_{100} , as most of the stellar mass is concentrated in intermediate-age populations. When dust and nebular components are added, all three SFH models predict larger old stellar populations, increasing the mass-weighted age and leading to an underestimation of SFR_{100} , particularly in high-mass galaxies.

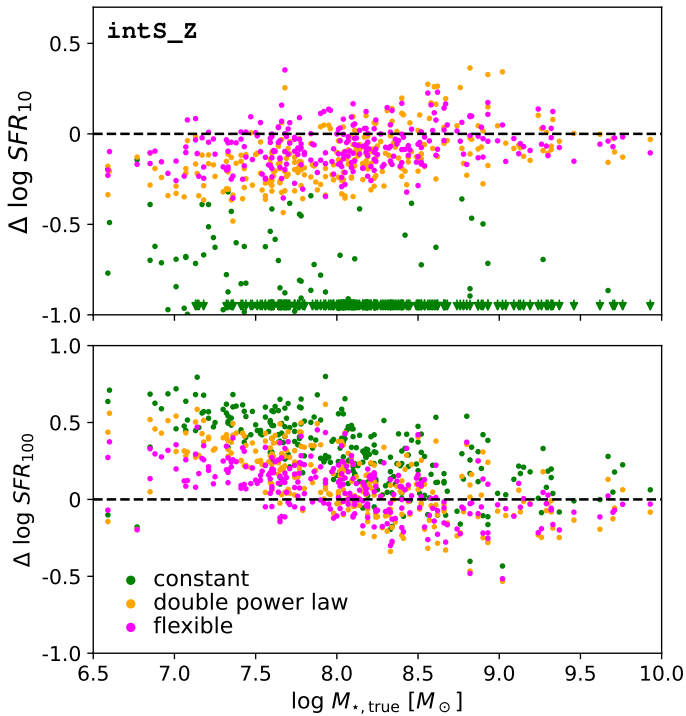


Fig. 5. Difference between the true and recovered SFRs from the intrinsic SEDs as a function of true M_* (`intS_Z`). The upper and lower panels present SFRs averaged over 10 Myr (SFR_{10}) and 100 Myr (SFR_{100}), respectively. In the fitting, stellar metallicity is treated as a free parameter. The constant SFH model substantially underestimates SFR_{10} but overestimates SFR_{100} , whereas the double power-law and flexible models recover SFRs to within ~ 0.5 dex.

highest among all SFH models. The double power-law and flexible SFH models similarly underestimate SFR_{10} , while overestimating SFR_{100} in this low-mass galaxy (and generally in the

low-mass galaxies in our sample; left panels). For the more massive galaxy (right panels), the posteriors contain larger fractions of old stellar populations, occasionally leading to an underestimation of SFR_{100} relative to the true value. This effect becomes more pronounced when dust and nebular components are included.

Figure 5 illustrates the differences between the true and recovered SFR_{10} and SFR_{100} across the full sample. SFR_{10} is consistently underestimated in all models, with the largest bias observed in the constant SFH model, where many galaxies are fitted with values of zero. This bias arises because the posteriors fail to reproduce the contribution from young stellar populations, while overproducing intermediate-to-old populations. In contrast, SFR_{100} exhibits a clear mass-dependent pattern: it is overestimated in low-mass galaxies, but this bias weakens with increasing mass and can even turn into underestimation in some high-mass systems. This outcome reflects the shift in the dominant stellar population age with increasing mass. In low-mass galaxies, where most stars form within the last 100 Myr, any overestimation of stellar mass directly translates into an overestimation of SFR_{100} . Conversely, in massive galaxies, the low fraction of young stars leads SED fitting to underpredict recent star formation, resulting in lower inferred values of SFR_{100} .

3.2. Impact of dust attenuation

The age-metallicity-dust degeneracy is well known to complicate reliable inference of galaxy properties from SED fitting (e.g. Papovich et al. 2001; Csizi et al. 2024). Although far-IR photometric data can help mitigate this effect by constraining dust content through the IR excess, considerable degeneracy in dust parameters may remain (e.g. Qin et al. 2022). To isolate the effect of dust attenuation on stellar property derivations, especially in the absence of IR data, we used the attenuated stellar continuum from SPHINX²⁰ and performed an SED fitting with three widely used dust laws: SMC (Gordon et al. 2003), Calzetti (Calzetti et al. 2000), and Salim (Salim et al. 2018). The

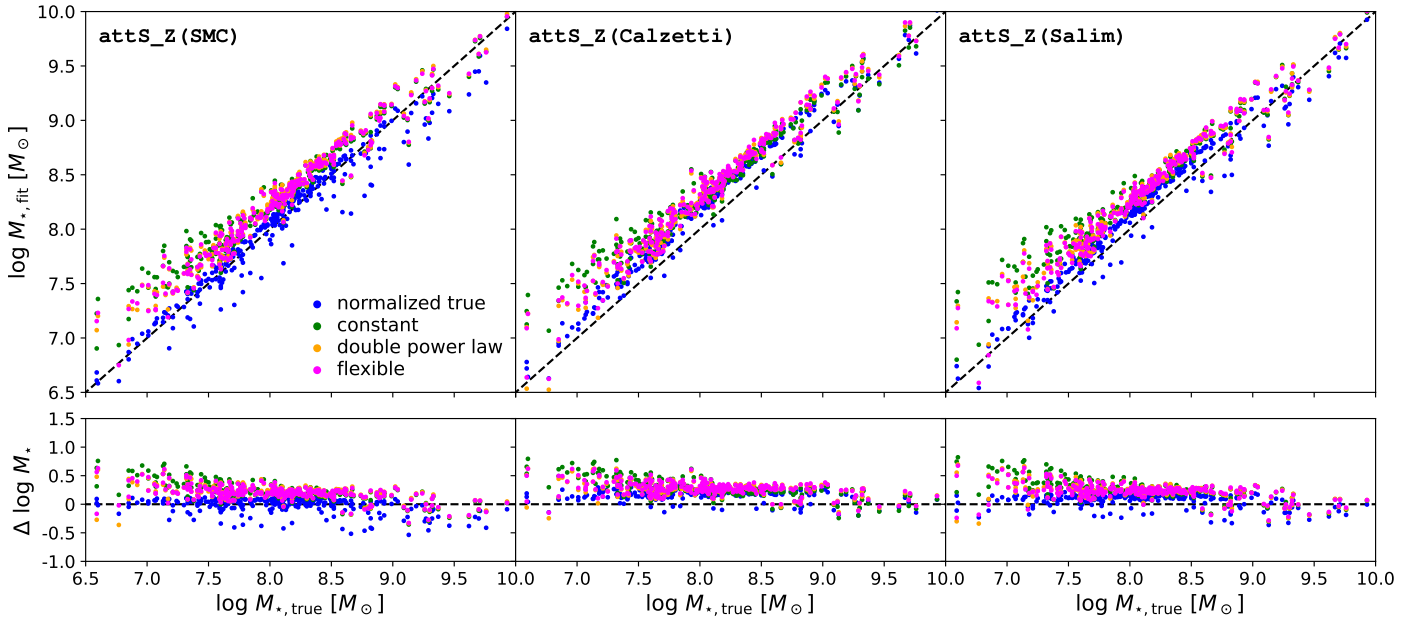


Fig. 6. Fitted M_\star and its offset from the true value for fits to the attenuated stellar continuum, with metallicity treated as a free parameter and three dust models applied: SMC (left), Calzetti (middle), and Salim (right). Considerable offsets in M_\star persist even when the normalised true SFH is used, driven mainly by slope mismatches between the true normalised attenuation curve and the assumed model curve. Parametric and non-parametric SFH models yield comparable results. Across all four SFH models, the SMC law yields the closest agreement in M_\star , whereas the Calzetti law produces the largest offset.

stellar metallicity is again treated as a free parameter during the fitting.

Figure 6 displays the fitted versus true values of M_\star . Offsets remain evident even when the normalised true SFH model is applied. The smallest median offset is obtained with the SMC law ($\Delta \log M_\star \equiv \log M_{\star, \text{fit}} - \log M_{\star, \text{true}} \approx 0.04$ dex), whereas larger offsets arise with the Salim (0.12 dex) and Calzetti laws (0.19 dex). A similar trend appears across the other three SFH models, yielding combined offsets of 0.20, 0.24, and 0.27 dex for the SMC, Salim, and Calzetti laws, respectively. In what follows, we first examine how the dust attenuation curve affects the SED fitting results and its influence in conjunction with different SFH models.

3.2.1. Biases in stellar mass and dust attenuation driven by attenuation-curve slope mismatches

To isolate the effect of the dust attenuation curve, we first examine the case with the ‘normalised true SFH’. The influence of alternative SFH models under different dust laws is addressed in the next sub-section.

Figure 6 (left-most panel, blue circles) illustrates that while most galaxies fitted with the SMC-type dust law exhibit a slight overestimation of M_\star , a sub-set of galaxies (particularly the more massive ones) display significant underestimations, with $\Delta \log M_\star \approx -0.6$ dex. This discrepancy mainly arises from differences in the attenuation curve slope between the simulated galaxies and the assumed SMC-type law during SED fitting, reflecting variations in dust geometry. This effect is further depicted in Fig. 7, which presents the ratio of fitted to true A_V and the stellar mass offset as functions of the true UV–optical slope, defined as $S \equiv A_{1500}/A_V$, where A_{1500} and A_V are the attenuations at 1500 Å and 5500 Å, respectively. Notably, S values greater than 4.75 indicate stronger UV attenuation than the SMC law, corresponding to a steeper attenuation curve. Our find-

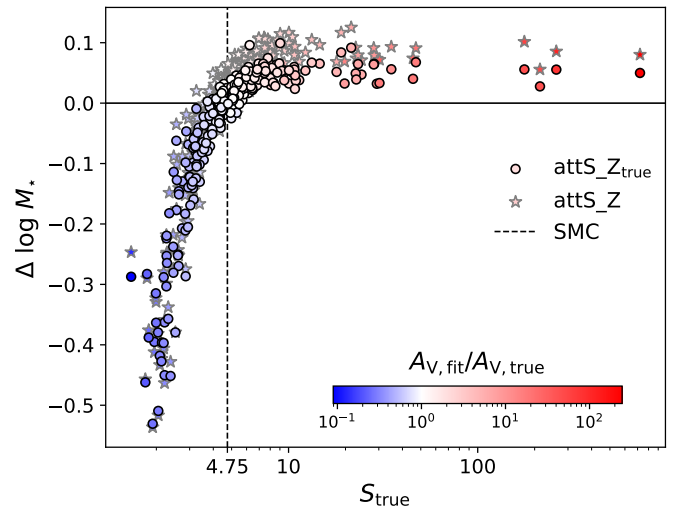


Fig. 7. Relation between the true UV–optical slope S ($\equiv A_{1500}/A_V$) and the stellar mass offset from the true value, derived by fitting the normalised true SFH with the SMC-type dust law under $\text{attS_Z}_{\text{true}}$ and attS_Z . The degree of over- or underestimation in M_\star and A_V depends on the steepness of the true normalised attenuation curve relative to the model curve. This behaviour persists even when metallicity is treated as a free parameter, although the fitted values of A_V and M_\star increase slightly.

ings indicate that when the attenuation curve is steeper than the SMC, both A_V and M_\star are overestimated, whereas a shallower curve leads to underestimation.

The physical origin of these correlations can be understood as follows. For the normalised SFH model fitted with the SMC-type dust law, the stellar mass offset (see Appendix B) can be

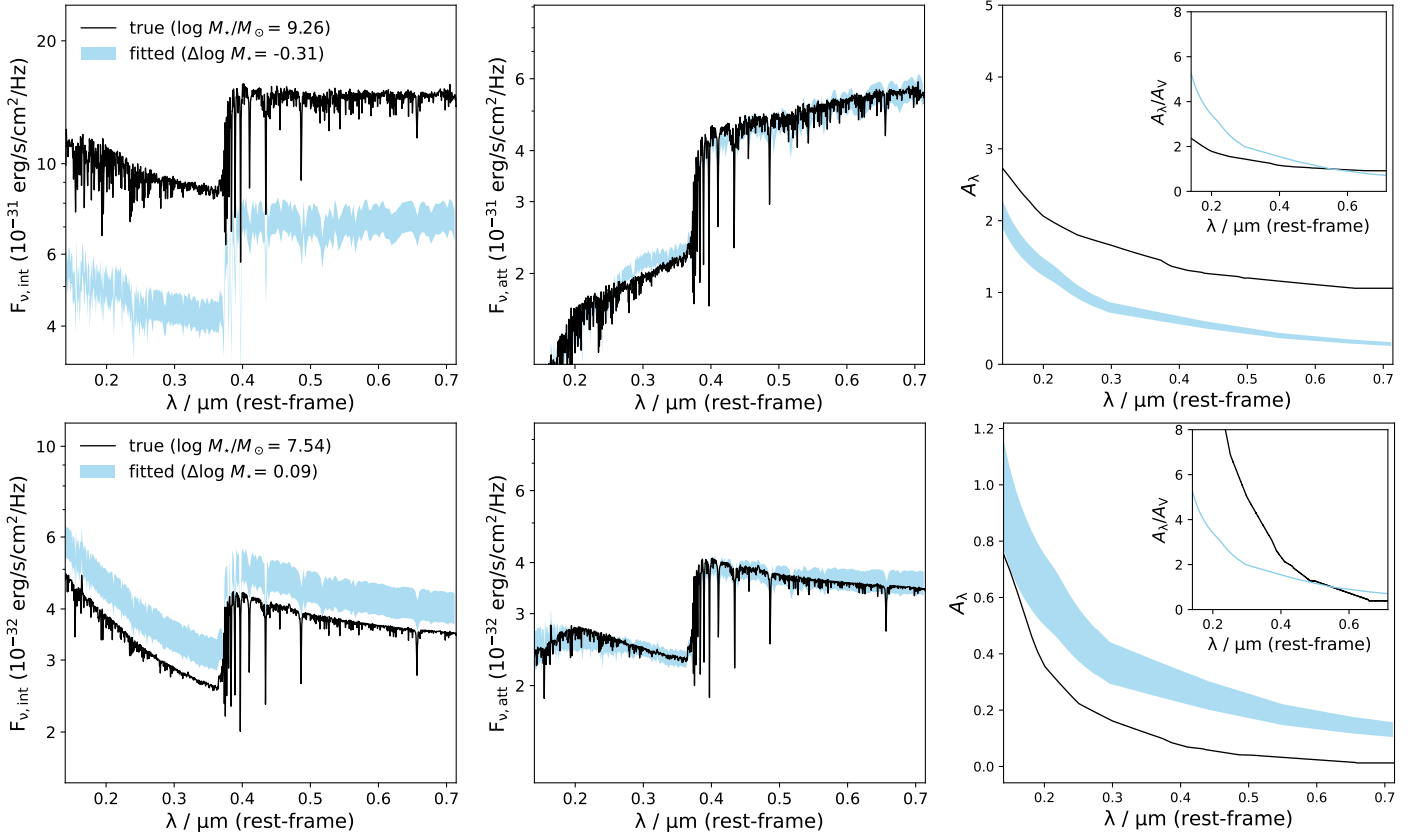


Fig. 8. 16th–84th percentile range of the posterior SEDs (blue shaded regions) is shown for two example galaxies, together with their true attenuated SEDs (black lines). Stellar metallicities are fixed to the true mass-weighted values. The left and middle panels display the intrinsic and attenuated stellar continua, respectively, and the right panels show the attenuation curves, with insets presenting their normalised forms (scaled by A_V). Here, A_λ is scaled to reproduce the attenuated SEDs while preserving the intrinsic SED shape ($Y(\lambda)$). The galaxy in the upper panel has a flatter dust slope than the model, yielding a fitted A_V lower than the true value, whereas the galaxy in the lower panel has a steeper dust slope, resulting in a fitted A_V higher than the true value.

expressed as

$$\Delta \log M_\star = \frac{A_V}{2.5} \left[\frac{A_{V,\text{fit}}}{A_V} s_{\text{SMC}}(\lambda) - s(\lambda) \right], \quad (4)$$

where s denotes the normalised dust attenuation curve (or slope), defined as $s(\lambda) \equiv A(\lambda)/A_V$. Because the stellar mass-to-light ratio $Y(\lambda)$ is known, the right-hand side becomes fixed for any λ once $\Delta \log M_\star$ is determined for a given posterior. At $\lambda = 1500 \text{ \AA}$, $s(\lambda = 1500 \text{ \AA}) \equiv S$ and Eq. (4) is simplified to

$$\Delta \log M_\star = \frac{A_V}{2.5} \left(\frac{A_{V,\text{fit}}}{A_V} S_{\text{SMC}} - S \right), \quad (5)$$

where S_{SMC} denotes the UV–optical slope derived from the SED attenuated by the SMC-type dust law. Because $\Delta \log M_\star$ is positively correlated with $A_{V,\text{fit}}/A_V$, Eq. (5) implies that if the normalised dust attenuation curve is steeper (flatter) than the model curve used in the SED fitting, the fitted A_V and M_\star will be overestimated (or underestimated, respectively). Representative examples of both cases are presented in Fig. 8, which compares the simulated stellar continua with the fitted results for two galaxies displaying different dust slopes. In the upper (lower) panel, the assumed dust slope (i.e. SMC, blue lines) is steeper (or shallower) than the actual extinction curve of the simulated galaxy (black lines), resulting in an underestimation (or overestimation) of the intrinsic fluxes and, consequently, the stellar mass.

Another important feature of Eq. (4) concerns the dust properties of a galaxy. For any pair of wavelengths λ_1 and λ_2 , we have

$$\begin{aligned} A_V[s(\lambda_2) - s(\lambda_1)] &= A_{V,\text{fit}}[s_{\text{SMC}}(\lambda_2) - s_{\text{SMC}}(\lambda_1)], \\ E(\lambda_2 - \lambda_1) &= E_{\text{fit}}(\lambda_2 - \lambda_1), \end{aligned} \quad (6)$$

where $E(\lambda_2 - \lambda_1)$ and $E_{\text{fit}}(\lambda_2 - \lambda_1)$ denote the colour excesses between λ_1 and λ_2 for the input attenuated SED of the simulated galaxy and the corresponding model fit, respectively. This relation indicates that SED fitting primarily recovers the colour excess rather than the absolute dust content (A_V) or the slope of the attenuation curve (S). The data in Fig. 8 (right panels) support this interpretation: the fitted A_λ curves closely follow those of the simulated galaxy, even when the corresponding stellar masses may diverge from the true values.

The trends obtained with `attS_Ztrue`, depicted in Fig. 7, persist even when metallicity is allowed to vary (`attS_Z`). In this case, both the fitted A_V and M_\star increase slightly relative to those using `attS_Ztrue`, by +0.04 dex and +0.03 dex, respectively. These increases result from the underestimation of Z_\star (−0.11 dex), which renders the intrinsic model stellar continuum bluer. To match the observed colour of the input attenuated stellar continuum, the fit requires a steeper attenuation curve or equivalently a larger colour excess. Because the dust law in the SED fitting is fixed to the SMC curve, this steepening can only be achieved by raising A_V . This, in turn, enhances the attenua-

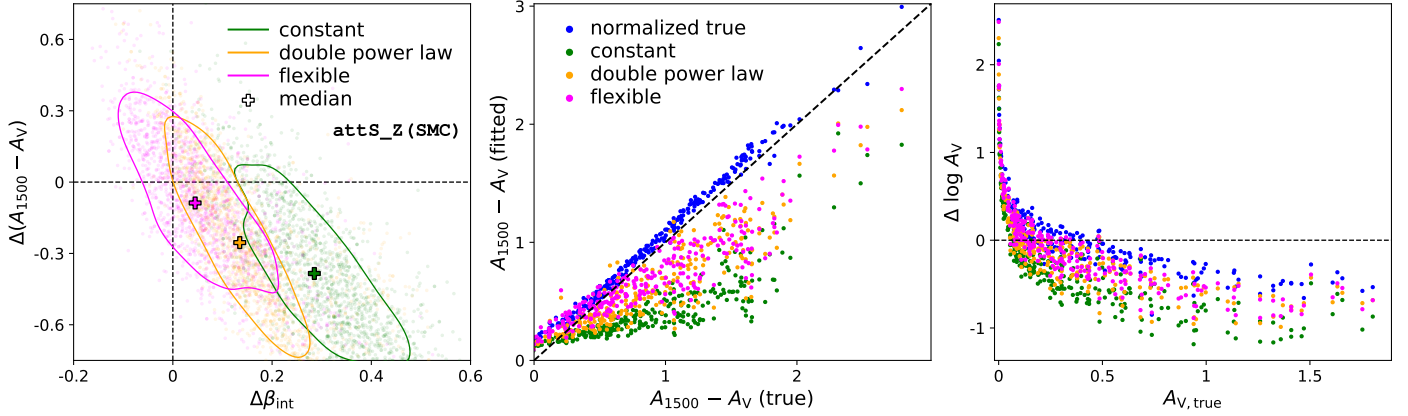


Fig. 9. Distribution of the difference between the fitted and true values of β_{int} and the colour excess for a randomly selected galaxy, shown in the left panel. Crosses mark the medians, and contours enclose the 1σ confidence regions of the posterior distributions for each SFH model. A negative correlation is evident between the two quantities. Among the models, the constant SFH model yields the largest overestimation of β_{int} , followed by the double power-law and flexible models. This trend appears in most galaxies in our sample, independent of the dust model adopted. The middle panel presents the resulting distribution of the recovered colour excess, while the right panel displays the distribution of the A_V offset, which is directly related to the colour excess attributed to their positive correlation.

tion across all wavelengths, resulting in an increase in intrinsic model fluxes and, thus, in the fitted M_\star .

The correlation between S_{true} and the stellar mass offset (Fig. 7) remains when alternative dust attenuation laws are used in the SED fitting, although the absolute values differ slightly. For instance, applying the Calzetti law yields larger deviations in M_\star ($\Delta \log M_\star \approx 0.19$) and A_V ($\Delta A_V \approx 0.42$ dex) than applying SMC law ($\Delta \log M_\star \approx 0.07$ dex and $\Delta A_V \approx 0.20$ dex). This occurs because the normalised attenuation curves of simulated galaxies in SPHINX²⁰ are generally steeper than the Calzetti law, but more closely resemble the SMC law. Consequently, reproducing the true colour excess requires a higher A_V when the Calzetti law is adopted. The stellar mass, M_\star , is then overestimated to compensate for the stronger attenuation at a fixed observed flux. By contrast, the Salim law model allows the slope to vary through an additional parameter, δ . Through simultaneous adjustment of δ and A_V , the Salim law can more flexibly recover the colour excess, reducing the systematic biases in A_V ($\Delta A_V \approx 0.26$ dex) and in M_\star ($\Delta \log M_\star \approx 0.12$ dex) relative to the Calzetti law; however, it still does not surpass the SMC law despite its greater flexibility.

In summary, the slope of the dust attenuation curve has a significant impact on the inferred stellar mass and A_V in the SED fitting, with shallower (steeper) true attenuation curves than the model leading to underestimation (overestimation). This bias arises because SED fitting primarily retrieves the colour excess rather than the absolute dust content or the attenuation slope. Although flexible models such as the Salim law alleviate this bias more effectively than the Calzetti law, the SMC law continues to provide the most accurate fits for galaxies with steep attenuation curves, such as those in the SPHINX²⁰ simulation.

3.2.2. Impact of intrinsic UV slope errors (SFH model) on colour excess, A_V , and SFR

Because reconstructed SFHs rarely reproduce the true stellar mass-to-light ratio (Υ) precisely, Eqs. (4) to (6) tend to not be satisfied overall. In such cases, the colour excess can be accurately recovered only if the shape of the fitted intrinsic stellar continuum closely matches the true shape. Otherwise, the fitting behaves as when metallicity is treated as a free parameter with the normalised true SFH: if the fitted intrinsic SED

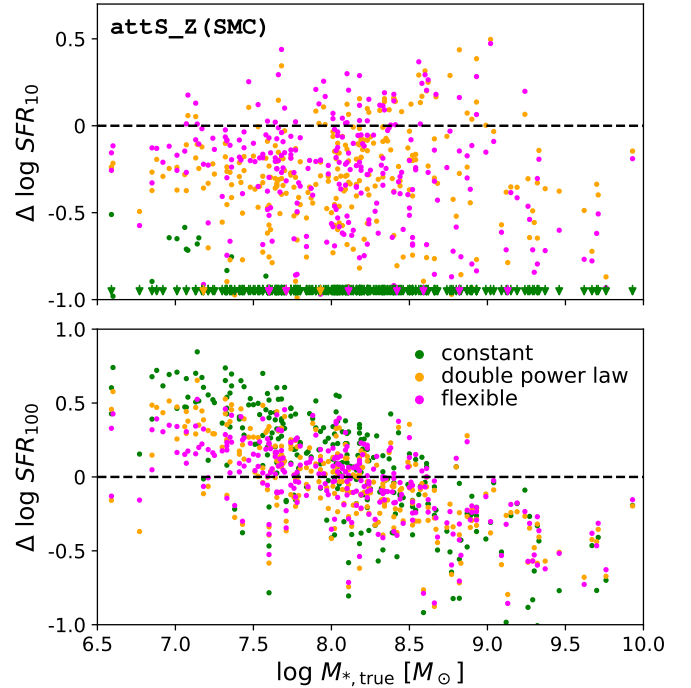


Fig. 10. Same as Fig. 5, but for the attS_Z model adopting an SMC-type dust attenuation curve. Compared to Fig. 5, the underestimation of SFR_{10} is apparent across the entire stellar-mass range, while the underestimation of SFR_{100} becomes progressively more severe toward the high-mass end.

becomes redder (bluer), the fitted attenuation curve correspondingly becomes shallower (steeper) to reproduce the colour of the input attenuated SED. This behaviour is supported by Fig. 9 (left panel), which displays the posterior distribution of the differences between the fitted and true colour excess ($\Delta(A_{1500} - A_{5500})$) and between the fitted and true UV slope of the intrinsic stellar continuum ($\Delta\beta_{\text{int}}$) for a randomly selected galaxy. Here, β_{int} is measured from the intrinsic stellar continuum across 1268 to 2700 Å (e.g. Calzetti et al. 1994). Our findings reveal that the difference in colour excess is consistently and negatively correlated with β_{int} , irrespective of the SFH model used.

For most galaxies in our sample, the fitted β_{int} is also overestimated across all three SFH models. The constant SFH model exhibits the largest deviation ($\Delta\beta_{\text{int}} \approx +0.31$), followed by the double power-law ($\Delta\beta_{\text{int}} \approx +0.08$) and then the flexible model ($\Delta\beta_{\text{int}} \approx +0.04$). This trend is directly connected to the distribution of the fitted colour excess. As illustrated in the middle panel of Fig. 9, the constant SFH model substantially underestimates the colour excess (≈ 0.32 dex) because of the overestimation of β_{int} . The normalised true SFH model yields the smallest offset in colour excess (0.04 dex), followed by the flexible (0.12 dex) and double power-law (0.15 dex) models. Because the colour excess is determined solely by A_V under both the SMC and Calzetti laws, the fitted A_V tends to be smallest when the constant SFH model is applied. This explains the fitted A_V distribution in the right panel of Fig. 9, where the constant SFH model produces the lowest A_V across all attenuation laws, including Salim.

A shallower attenuation curve, such as the Calzetti law, requires a larger colour excess to reproduce the same level of reddening between two wavelengths (e.g. 1500 Å and 5500 Å), particularly when fitting the full shape of the observed SED. This is analogous to fitting a curve through multiple points: a shallower attenuation curve constrained by fixed endpoints (i.e. the same colour excess) will inevitably fail to reproduce the intermediate fluxes accurately. Consequently, the fitting procedure favours a larger colour excess to better match the overall spectral shape, which in turn yields a steeper UV slope owing to the increased SFR. Figure 10 shows the differences between the true and fitted star formation rates as a function of stellar mass for different SFH models. As a direct consequence of the behaviour described above, the median ΔSFR_{100} increases systematically from SMC to the Salim to the Calzetti attenuation curve, with values of 0.02, 0.09, and 0.19, respectively, while the overall mass-dependent trends remain qualitatively similar. Because shallower attenuation curves (e.g. Calzetti) require large values of A_V to reproduce a given colour excess, the inferred stellar mass increases progressively from SMC to Salim to Calzetti (0.20, 0.24, and 0.27, respectively; see Fig. 6).

In summary, because reconstructed SFHs are rarely sufficient to recover the true intrinsic continuum, errors in the fitted UV slope translate directly into dust biases: redder inferred slopes force shallower attenuation curves and yield underestimated colour excesses and A_V , especially for the constant SFH model. These systematics are further amplified, with shallower curves (e.g. Calzetti) requiring larger colour excesses and higher A_V , thereby boosting the recovered SFR and stellar mass.

3.3. Impact of nebular emission

Because high-redshift galaxies are actively star-forming, nebular emission from ionised gas, such as H β 4861, [O III] 4959/5007, H α 6563, and [N II] 6548/6584, can contribute substantially to broad-band fluxes in the UV-optical range (e.g. Anders & Fritze-v. Alvensleben 2003; Zackrisson et al. 2008; Wilkins et al. 2013). The intrinsic Ly α emission at 1216 Å is also very strong, but it is heavily attenuated by the neutral IGM at $z = 6$ in SPHINX²⁰ (Garel et al. 2021). The presence of strong emission lines reddens the UV-optical colours, causing stellar masses to be overestimated when nebular emission is neglected (Schaerer & de Barros 2009; Salmon et al. 2015; Yuan et al. 2019; Miranda et al. 2025).

For galaxies in our sample, including nebular emission indeed makes the input SED appear redder. To illustrate its impact on galaxy colour, we compared the filter magnitudes with and without nebular emission for massive ($8.5 < \log M_*/M_\odot < 10$) and

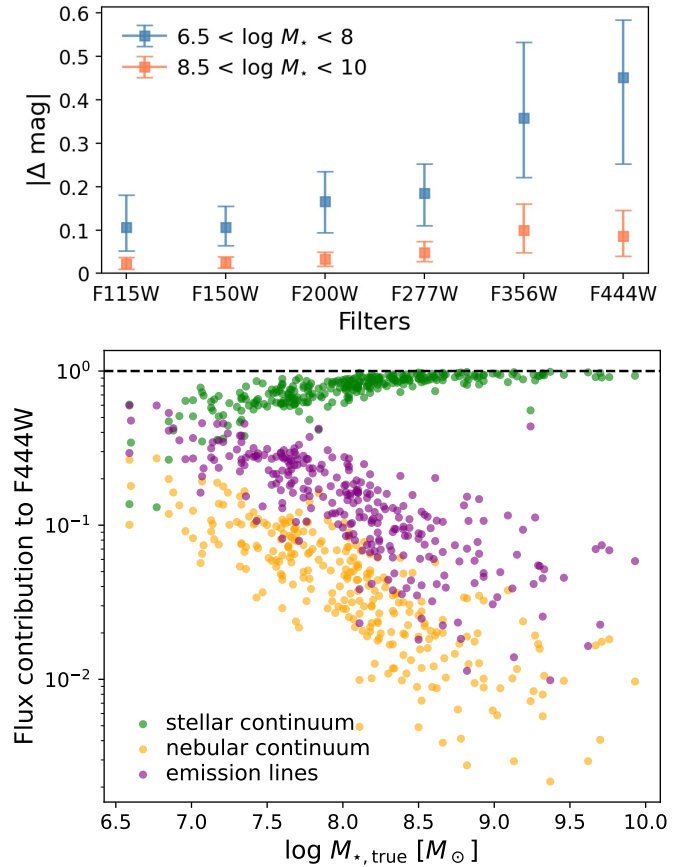


Fig. 11. Change in filter magnitudes when nebular emission is added to the stellar continuum for low- and high-mass galaxies, shown in the upper panel. Each point marks the median, with error bars indicating the 16th–84th percentile range. Both galaxy populations appear redder, mainly due to strong optical emission lines in the rest frame optical range, with the effect being more pronounced in low-mass systems. The lower panel presents the flux contributions of the stellar continuum, nebular continuum, and nebular emission lines to the F444W magnitude for each galaxy. In galaxies with $\log M_*/M_\odot \lesssim 7.3$, the emission-line contribution is comparable to that of the stellar continuum.

less massive ($6.5 < \log M_*/M_\odot < 8$) sub-samples. As depicted in the upper panel of Fig. 11, the inclusion of nebular emission results in substantial flux increases in the F356W and F444W filters. This enhancement is especially pronounced in low-mass galaxies, leading to a higher inferred M_* , as discussed further below.

The lower panel of Fig. 11 displays the relative contributions of the stellar continuum, nebular continuum, and emission lines to the F444W filter. In most galaxies, the stellar continuum dominates the total flux, whereas emission lines contribute comparably in low-mass, actively star-forming systems. Within the F444W bandpass, the dominant emission lines are H α λ 6563 and [NII] λ 6583, whereas in the F356W bandpass, H β 4861, [O III] 4959/5007 dominate. The nebular continuum is generally approximately three times fainter than the line emission and therefore remains a sub-dominant component of the input SED.

To evaluate the impact of nebular emission on the inferred galaxy properties, we included the attenuated nebular component (both line and continuum) together with the attenuated stellar continuum in the input spectrum. The combined spectrum is then fitted using BAGPIPES, with the nebular component enabled (attSNE_Z). Figure 12 illustrates that when nebular emission is included under the normalised true SFH, stellar masses can still

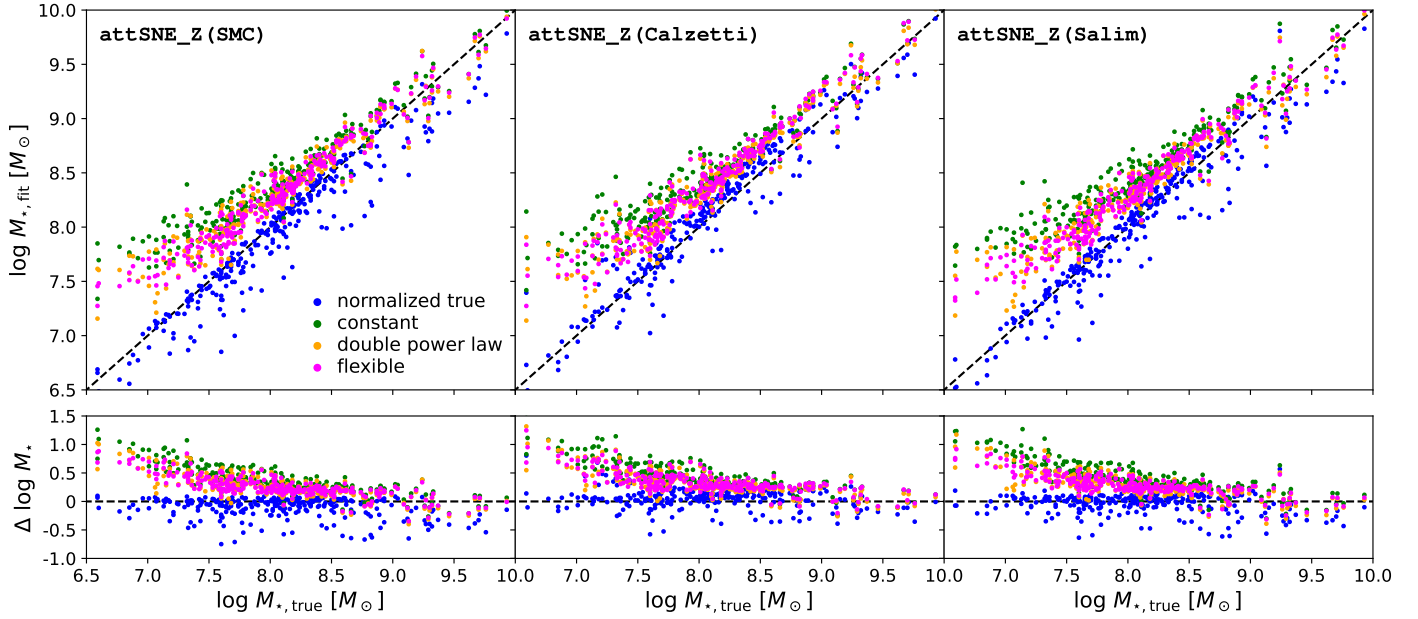


Fig. 12. Same as Fig. 6, but with nebular emission included (i.e. the attSNE_Z model). Across all dust attenuation laws and SFH models, the overestimation of M_* exceeds that in the attS_Z case, especially in low-mass galaxies, where it reaches ~ 1.5 dex. Consistent with the attS_Z results, the bias is smallest for the SMC law, largest for the Calzetti law, and intermediate for the Salim law.

be recovered with a reasonable level of accuracy. The median differences in $\log M_*$ are -0.05 , 0.06 , and 0.01 for the SMC, Calzetti, and Salim dust models, respectively; these values are smaller than the corresponding offsets without nebular emission (0.04 , 0.19 , and 0.12 ; see Fig. 6). In contrast, estimates deviate more strongly when using the other three SFH models, with offsets of 0.36 , 0.26 , and 0.26 dex for the constant, double power-law, and flexible models, respectively. This discrepancy is most evident in low-mass galaxies, where strong optical lines redden the SED and bias the posterior towards older stellar populations, producing mass overestimations of up to 1.31 dex. Conversely, in massive galaxies, the effect is weak, as the contribution of nebular emission to the total flux is negligible (see the lower panel of Fig. 11). Galaxies with $M_{*,\text{true}} > 10^{8.5} M_\odot$ yield fitted masses consistent with those from attS_Z. Although the nebular flux contribution is generally small, excluding it may not be justified, as it can still bias M_* estimates.

Because emission lines increase the total flux and redden the SED, the stellar population is inferred to be older and/or more dust-rich, which, in turn, raises the estimated stellar mass. To evaluate this effect, we repeated the fitting with input spectra that exclude emission lines (attSN_Z_SNE) and compared the results with the fiducial case (attSNE_Z), thereby isolating the influence of intrinsic emission on mass inference. The upper panel of Fig. 13 indicates that the recovered mass remains consistent under the normalised true SFH model is used. In contrast, the other three models yield systematically lower masses, as the omission of emission lines produces a bluer SED that favours younger populations, as illustrated in the lower panel of Fig. 13. This shift in stellar population leads to a pronounced reduction in the inferred stellar mass, particularly for low-mass galaxies.

Conversely, when emission is included in the input SED, but not modelled during the fitting (attSNE_Z_SN), as in the study of Schaerer & de Barros (2009), the excess optical flux is attributed to older stellar populations under the double power-law and flexible SFH models (Fig. 14). The normalised true SFH model also returns higher M_* values, but in this case, the increase is driven by a larger fitted A_V , as the stellar population

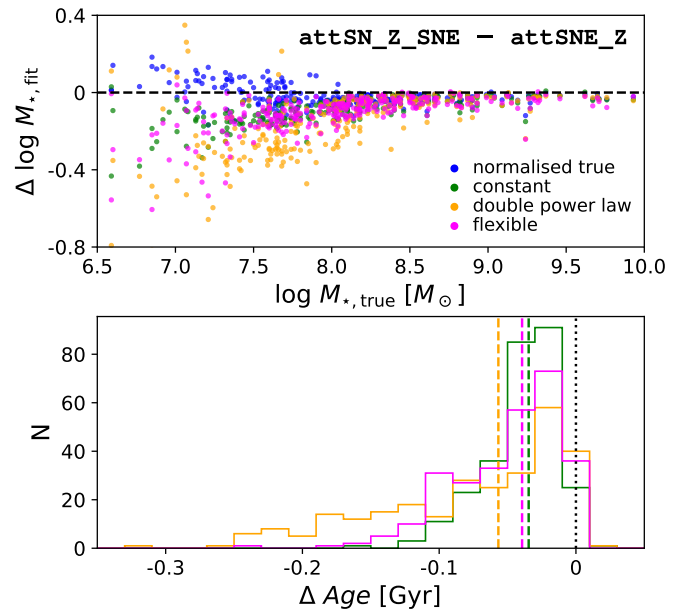


Fig. 13. Impact of emission lines in the input spectrum on inferred galaxy properties. The upper panel presents the difference in the fitted M_* between attSN_Z_SNE (which excludes emission lines from the input spectra, but includes them in the fitting) and attSNE_Z. Emission lines have the strongest effect in low-mass galaxies. The reduction in fitted M_* arises because attSN_Z_SNE infers younger stellar populations, producing a bluer continuum to match the observed colour in the absence of emission lines. The lower panel illustrates this change in stellar populations by illustrating the distribution of differences in mass-weighted age between attSN_Z_SNE and attSNE_Z with the median marked by a dashed line.

is fixed. By comparison, the constant SFH model exhibits negligible variation in any fitted parameter, as it primarily reproduces the broadband colours through the stellar continuum alone, which limits the influence of emission lines on the inferred age

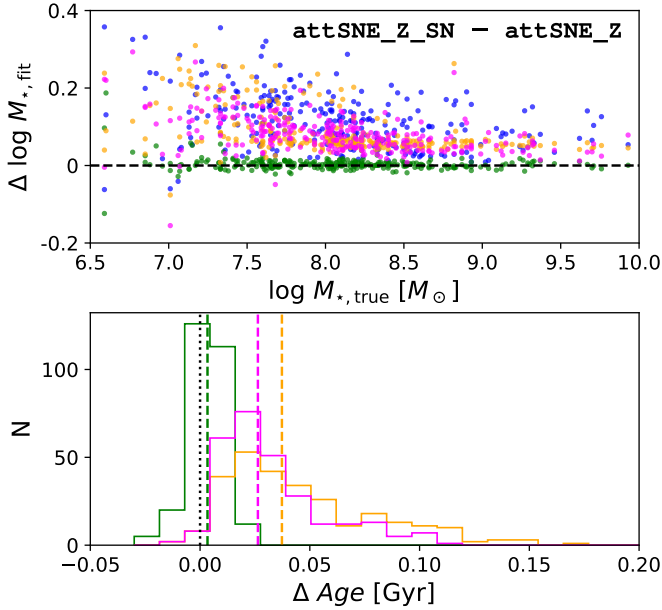


Fig. 14. Same as Fig. 13 but for the attSNE_Z_SN model, which includes emission lines in the input spectra but neglects them in the fitting process. Here, the increase in the fitted M_{\star} mainly reflects changes in the inferred stellar populations, except for the normalised true SFH model, which increases A_V to redden the model spectrum instead.

or mass. Although including emission lines lowers the χ^2 values, this improvement only accounts for only a small fraction of the posterior distributions.

In summary, strong emission lines substantially enhance the rest-frame optical flux, leading to an overestimation of the fitted M_{\star} in the attSNE_Z model compared to that in attS_Z. This effect is most pronounced in low-mass galaxies, with biases reaching up to 0.8 dex. Neglecting emission-line modelling in SED fitting further exacerbates the bias, introducing an additional overestimation of up to 0.4 dex because emission-line optical flux is misattributed to the stellar continuum of older populations. By contrast, the constant SFH model yields few or no stars younger than 10 Myr irrespective of emission, indicating that it is the least reliable of the tested models for recovering recent SFRs.

3.4. Summary of the impact of modelling assumptions

Figure 15 summarises the impact of various modelling assumptions adopted in the SED fitting. We first quantified the uncertainty associated with the choice of SFH by taking the mean logarithmic offsets in the derived physical quantities obtained using the intS_Z_{true} configuration. We then illustrate how the inferred values deviate from the true ones as additional modelling ingredients are introduced sequentially, as indicated along the x-axis. We emphasise that this figure is intended solely to provide a qualitative indication of the assumptions that influence the recovered galaxy properties most strongly. A detailed quantitative breakdown is presented in Appendix D.

For less massive galaxies ($M_{\star} < 10^8 M_{\odot}$, top panel), we find that the stellar mass estimates are most strongly affected by the age–metallicity degeneracy (~ 0.2 dex) and by emission line modelling (~ 0.2 dex). The short-term SFR is likewise most sensitive to the inclusion of emission lines, while dust modelling contributes comparably to systematic underestimation. In con-

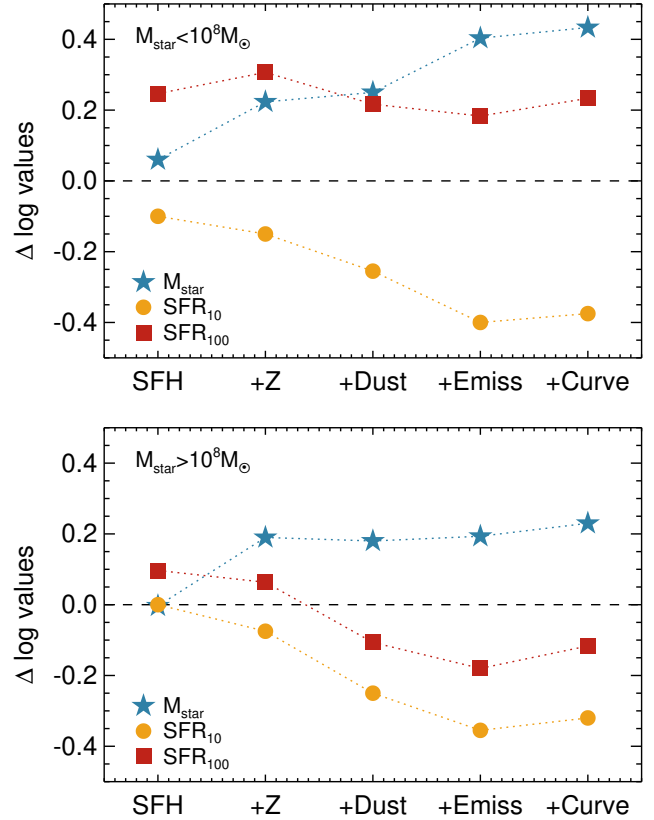


Fig. 15. Summary of how different modelling assumptions influence the inferred physical quantities. Top and bottom panels show the logarithmic offsets for low-mass ($M_{\star} < 10^8 M_{\odot}$) and high-mass ($M_{\star} > 10^8 M_{\odot}$) galaxies, respectively. Each point represents the mean offsets in the inferred physical properties obtained with the intS_Z_{true}, intS_Z, and attS_Z configurations; the two right-most points use the same attSNE_Z configuration. The left-most point isolates the effects of the assumed SFH, while the points to the right (marked by plus signs on the x-axis) illustrate the cumulative impact of successively adding each modelling ingredient.

trast, the long-term SFR is most strongly affected by the assumed SFHs.

For more massive galaxies ($M_{\star} > 10^8 M_{\odot}$, top panel), the stellar mass estimates remain primarily driven by the age–metallicity degeneracy, whereas the impact of the other ingredients is negligible. For both the short-term and long-term SFR estimates, dust modelling emerges as the dominant source of uncertainty, followed by emission line modelling.

4. Discussion

In the previous section, we detail how SFH, metallicity, dust, and emission influence SED fitting. Building on this, we go on to evaluate how choices in the observational setup and statistical treatment affect the accuracy and robustness of SED fitting, with particular attention to the role of adding a medium-band filter in recovering galaxy properties. We further consider how uncertainties in the inferred properties can propagate into the derived quantities, such as the SMF and the star formation main sequence.

4.1. Improving fitting accuracy

4.1.1. Impact of adding a medium band on fitting results

Figure 12 illustrates that stellar mass is considerably overestimated when strong emission lines are present in the input photometry (attSNE_Z) compared to the case without emission (attS_Z, Fig. 6). As noted earlier, this bias arises because the optical flux in the stellar continuum is boosted by demanding a larger contribution from older stellar populations. The inclusion of a medium-band filter without strong emission lines can therefore potentially mitigate this bias in the SED fitting (e.g. Roberts-Borsani et al. 2021).

To test this, we recompute stellar masses after adding the F410M band, which is largely unaffected by strong emission lines at $z = 6$, to the input photometry (attSNE_Z_F410M). The upper panel of Fig. 16 displays an example of the posterior distribution for a low-mass galaxy with $M_{\star} = 10^{7.2} M_{\odot}$. We find that the addition of the extra band reduces the stellar mass offset from 0.7 dex in attSNE_Z to nearly zero. Furthermore, the true SFR_{10} ($0.78 M_{\odot} \text{ yr}^{-1}$) is more accurately recovered ($0.23 \rightarrow 0.5 M_{\odot} \text{ yr}^{-1}$), as the contribution from older stellar populations is diminished.

A statistical comparison of the stellar mass offset is presented in the lower panel of Fig. 16. The maximum offset in low-mass galaxies under attSNE_Z is 1.26 dex (Fig. 12), but it decreases to 0.57 dex when the F410M band is included. On average, $\Delta \log M_{\star}$ in low-mass galaxies with $M_{\star, \text{true}} \lesssim 10^8 M_{\odot}$ decreases from 0.53 to 0.39, 0.40 to 0.27, and 0.39 to 0.24 dex for the constant, double power-law, and flexible SFH models, respectively, relative to attSNE_Z. By contrast, galaxies more massive than $\sim 10^8 M_{\odot}$ are only marginally influenced by the addition of the medium band, again because the contribution from emission lines is negligible. Nevertheless, the medium band can still be useful when bright galaxies display strong emission lines.

This exercise highlights the importance of emission-line-free photometric bands in the optical range, in line with the findings of Roberts-Borsani et al. (2021). In practice, however, such medium bands may not be accessible for galaxies at higher redshift. For instance, the F410M band is likely contaminated by the [OII] $\lambda 3727$, [OIII] $\lambda 5007$ and Balmer lines in galaxies at $z \gtrsim 7$. Moreover, galaxies at higher redshift are typically more actively star-forming and therefore more strongly influenced by emission lines. To evaluate whether these effects alter our conclusions, we repeated the same analysis for a sample of 66 galaxies at $z = 9$. Although we do not show them in this work, we did find that the trends observed at $z = 6$ persist at $z = 9$, indicating that our conclusions remain robust across these epochs. At $z = 9$, however, the F410M filter is no longer free of strong emission lines, and thus less useful than at $z = 6$. Consequently, the stellar mass of low-mass galaxies is overestimated by ≈ 0.1 dex more than that at $z = 6$. We also observe that the degree of mass overestimation from SED fitting decreases with increasing true stellar mass, confirming that the overall conclusions remain valid even at higher redshift ($z \sim 9$).

4.1.2. Choice of summary statistics

In Bayesian SED fitting, consensus on the appropriate statistical criterion for defining the ‘best-fit’ value is still lacking. The minimum χ^2 solution is widely used, yet it does not necessarily represent the most probable outcome, as posterior distributions are often non-Gaussian and asymmetric. The posterior median is another common choice, but, as illustrated in the case of stellar metallicity in Fig. 3, it can be biased by distribution shape

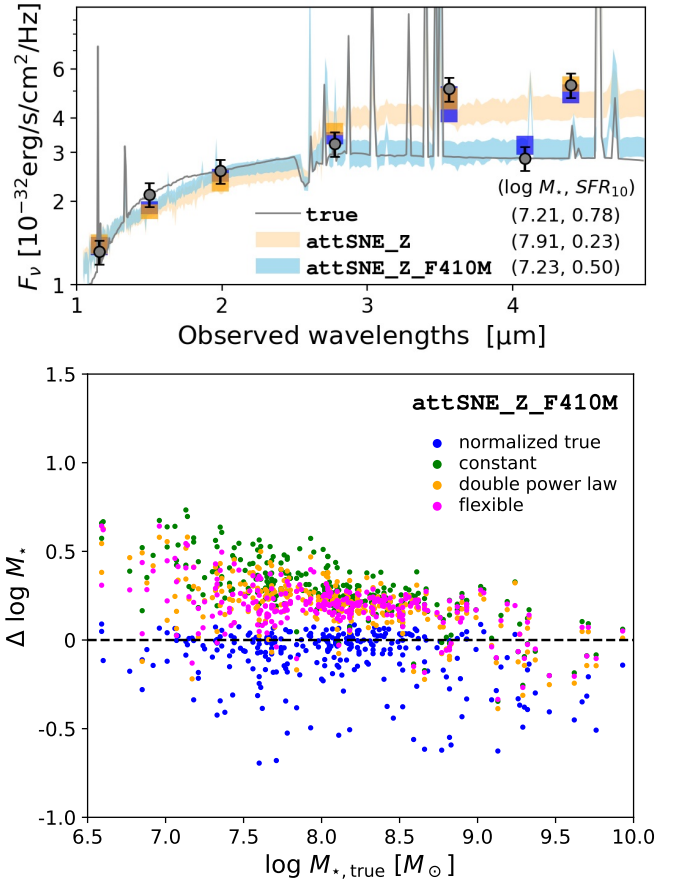


Fig. 16. Upper panel: 16th–84th percentile distribution of the fitted SEDs is presented for an example galaxy using the attSNE_Z (orange) and attSNE_Z_F410M (sky blue) models with the double power-law SFH. Comparing each shaded region with the true SED (gray) indicates that including F410M photometry provides tighter constraints on the SED fitting and consequently on M_{\star} and SFR_{10} . Lower panel: Difference between the true and fitted M_{\star} is shown when the SMC law is applied with attSNE_Z_F410M. The inclusion of F410M photometry substantially reduces the overestimation of M_{\star} , particularly in low-mass galaxies.

and may not reliably trace the true value. Therefore, assessing which of the three widely used statistics (minimum χ^2 , posterior median, or likelihood (\mathcal{L})-weighted mean) best recovers the underlying physical parameters is particularly instructive.

Figure 17 presents the differences between the true and fitted values of M_{\star} , Z_{\star} , SFR_{10} , and SFR_{100} in the attSNE_Z_F410M model across the three statistics. For the normalised true SFH case, M_{\star} and the SFRs display no notable dependence on the choice of statistic. By contrast, Z_{\star} is better reproduced by the posterior median than by the \mathcal{L} -weighted mean. The minimum χ^2 method often performs well, but also produces the largest offsets.

In the constant SFH model, M_{\star} is more biased when the posterior median or the \mathcal{L} -weighted mean is used, whereas the minimum χ^2 estimate is less biased but exhibits greater scatter. A similar pattern emerges for Z_{\star} , except that the \mathcal{L} -weighted mean appears least biased. This apparent accuracy largely stems from the restricted prior range, $Z_{\star}/Z_{\odot} = [0.001, 1]$. As noted in Sect. 3.1, the fitted Z_{\star} values cluster near the lower bound of this range, generating systematic bias when the posterior median is adopted. The effect is strongest for the constant SFH model and also influences the \mathcal{L} -weighted mean. Extending the prior range to $[10^{-5}, 1] Z_{\odot}$ causes both the posterior median and the

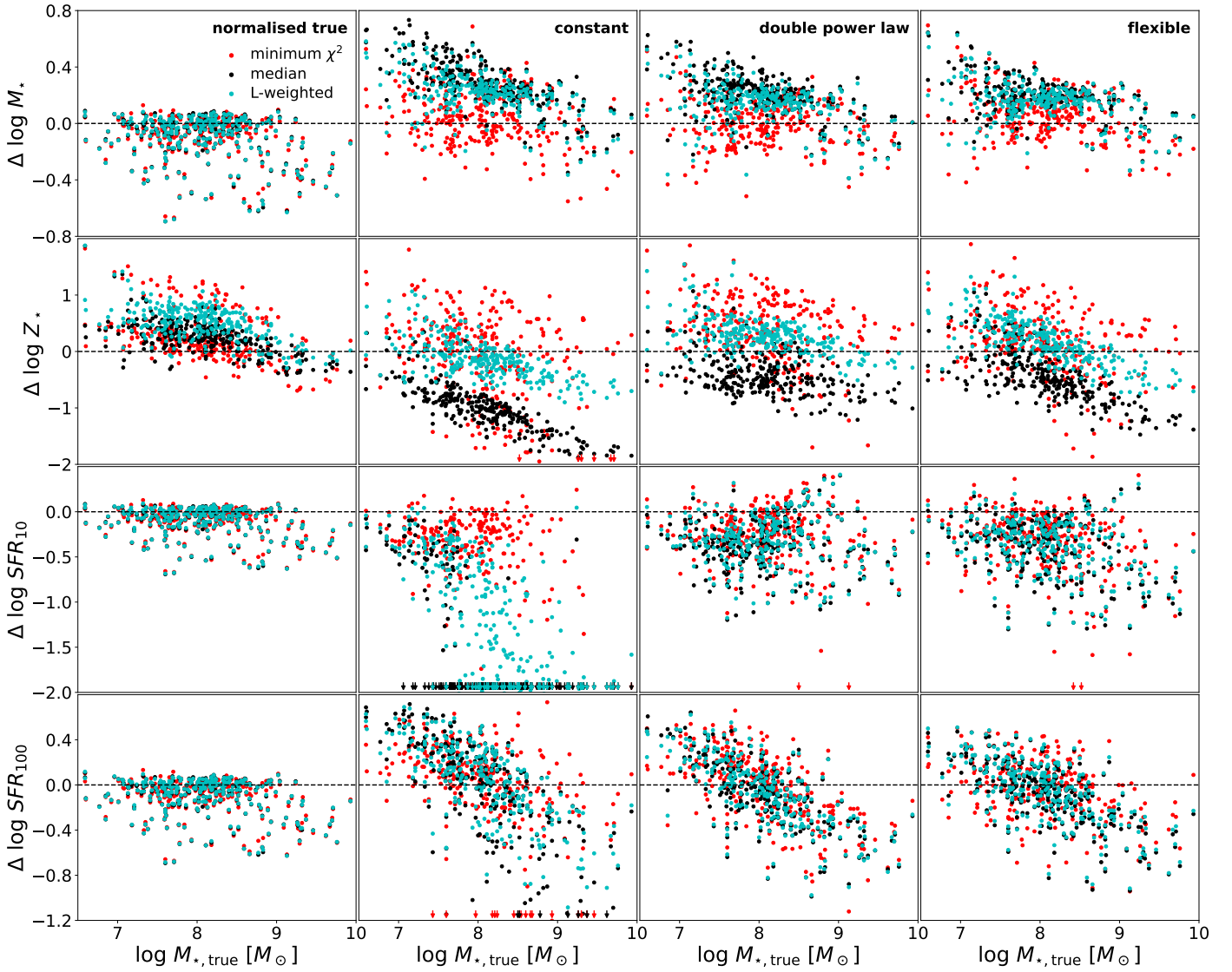


Fig. 17. Impact of summary statistics on the recovered physical quantities in the attSNE_Z_F410M model. Each row displays the difference between the true and fitted values of M_* , Z_* , SFR_{10} , and SFR_{100} , while each column corresponds to a different SFH model. The fitted values are derived from the posterior distribution using one of the three statistics: minimum χ^2 (red), median posterior (black), or likelihood-weighted mean (cyan). The minimum χ^2 posterior yields the most accurate recovery of M_* , whereas the likelihood-weighted mean performs best for Z_* . The recovered SFRs are generally insensitive to the choice of the summary statistic, except in the constant SFH model.

\mathcal{L} -weighted mean of Z_* to decline sharply (by nearly a factor of 50). For SFR_{10} , the minimum χ^2 method yields closer agreement with the true value than the other two, although SFR_{10} remains generally underestimated by a (relatively low) factor. This underestimation indicates that most posteriors lack recent star formation, even when solutions with non-zero SFR in the past 10 Myr provide lower χ^2 values. In contrast, SFR_{100} is less sensitive to the chosen statistic, although the minimum χ^2 method still produces the largest scatter.

The double power law and flexible SFH models display similar behaviour. The minimum χ^2 statistic more effectively recovers M_* than the other two measures, although the scatter is comparable to that of the constant SFH model. For Z_* , the \mathcal{L} -weighted mean provides the most reasonable estimates, even when the prior range is extended. While the posterior median of Z_* is strongly influenced by the adopted prior, both the minimum χ^2 and \mathcal{L} -weighted mean are less sensitive to this effect. For SFR_{10} , the minimum χ^2 achieves slightly better agreement

with the true value, whereas SFR_{100} remains largely unaffected by the choice of summary statistics.

We attribute these performance differences to two main factors. Firstly, BAGPIPES is a Bayesian inference code that explores the full prior space. Therefore, parameters that are weakly constrained (either because the model SED is relatively insensitive to them or because they suffer from strong degeneracies) can yield posterior distributions that are heavily shaped by the assumed priors. In such cases, summary statistics such as the median or \mathcal{L} -weighted mean become less reliable, whereas the minimum χ^2 estimate often tracks the true value more closely. Secondly, for strongly constrained parameters, where small changes lead to large variations in the model SED, the minimum χ^2 estimate is more susceptible to outliers. In these cases, the \mathcal{L} -weighted mean, which considers the entire posterior distribution, often provides a more reliable estimate.

In summary, the minimum χ^2 statistic more accurately recovers M_* than the other two, albeit with substantial scatter. In

Table 2. Offsets in the inferred physical properties for different summary statistics.

Variable	SFH model	Median		Minimum- χ^2		L-weighted	
		low($\leq 10^8 M_\odot$)	high($> 10^8 M_\odot$)	low	high	low	high
$\Delta \log M_\star$	Normalised true	$-0.04^{+0.07}_{-0.15}$	$-0.04^{+0.07}_{-0.25}$	$-0.06^{+0.08}_{-0.15}$	$-0.05^{+0.08}_{-0.24}$	$-0.04^{+0.07}_{-0.14}$	$-0.04^{+0.07}_{-0.25}$
	Constant	$0.37^{+0.15}_{-0.11}$	$0.23^{+0.07}_{-0.12}$	$0.05^{+0.18}_{-0.17}$	$0.03^{+0.20}_{-0.16}$	$0.27^{+0.14}_{-0.13}$	$0.20^{+0.06}_{-0.13}$
	Double power	$0.27^{+0.13}_{-0.11}$	$0.17^{+0.08}_{-0.12}$	$-0.01^{+0.17}_{-0.16}$	$0.02^{+0.14}_{-0.15}$	$0.18^{+0.15}_{-0.14}$	$0.15^{+0.07}_{-0.12}$
	Flexible	$0.21^{+0.16}_{-0.09}$	$0.19^{+0.05}_{-0.12}$	$0.08^{+0.18}_{-0.15}$	$0.05^{+0.12}_{-0.15}$	$0.17^{+0.16}_{-0.11}$	$0.17^{+0.05}_{-0.12}$
$\Delta \log \text{SFR}_{10}$	Normalised true	$-0.05^{+0.06}_{-0.15}$	$-0.05^{+0.07}_{-0.25}$	$-0.06^{+0.07}_{-0.15}$	$-0.06^{+0.08}_{-0.24}$	$-0.04^{+0.07}_{-0.14}$	$-0.04^{+0.07}_{-0.25}$
	Double power	$-0.37^{+0.18}_{-0.21}$	$-0.29^{+0.30}_{-0.33}$	$-0.23^{+0.18}_{-0.18}$	$-0.21^{+0.23}_{-0.36}$	$-0.32^{+0.20}_{-0.21}$	$-0.28^{+0.29}_{-0.31}$
	Flexible	$-0.28^{+0.16}_{-0.26}$	$-0.41^{+0.27}_{-0.40}$	$-0.22^{+0.23}_{-0.26}$	$-0.30^{+0.20}_{-0.43}$	$-0.24^{+0.16}_{-0.27}$	$-0.38^{+0.26}_{-0.37}$
$\Delta \log \text{SFR}_{100}$	Normalised true	$-0.04^{+0.07}_{-0.15}$	$-0.05^{+0.08}_{-0.25}$	$-0.07^{+0.09}_{-0.14}$	$-0.06^{+0.08}_{-0.24}$	$-0.04^{+0.07}_{-0.16}$	$-0.04^{+0.07}_{-0.26}$
	Constant	$0.30^{+0.23}_{-0.26}$	$-0.14^{+0.29}_{-0.46}$	$0.16^{+0.21}_{-0.18}$	$-0.03^{+0.22}_{-0.38}$	$0.28^{+0.19}_{-0.24}$	$-0.10^{+0.29}_{-0.38}$
	Double power	$0.12^{+0.23}_{-0.23}$	$-0.21^{+0.22}_{-0.24}$	$0.17^{+0.22}_{-0.18}$	$-0.15^{+0.29}_{-0.27}$	$0.14^{+0.21}_{-0.22}$	$-0.19^{+0.22}_{-0.25}$
	Flexible	$0.08^{+0.19}_{-0.19}$	$-0.16^{+0.22}_{-0.28}$	$0.03^{+0.20}_{-0.23}$	$-0.02^{+0.17}_{-0.32}$	$0.10^{+0.17}_{-0.20}$	$-0.14^{+0.22}_{-0.27}$

Notes. The table shows the 16th, 50th and 84th percentiles for the fitting model of attSNE_Z_F410M. ΔSFR_{10} for the constant SFH model is not shown, as $\text{SFR}_{10,\text{fit}} = 0$ in many galaxies. For each SFH model, we highlight the value that has the minimum offset.

contrast, the \mathcal{L} -weighted mean yields a more accurate estimate of Z_\star , although it remains sensitive to the prior range, particularly under the constant SFH model. SFR_{10} is more reliably recovered with the minimum χ^2 , especially when the constant SFH model is applied, whereas SFR_{100} appears largely insensitive to the choice of the summary statistic. Taken together, these results support the use of flexible or double power-law SFH models in combination with the minimum χ^2 statistic for estimating stellar mass, while the \mathcal{L} -weighted mean might be preferable for recovering other parameters (see Table 2 for detailed values).

4.2. Impact on derived galaxy population statistics

Stellar mass and SFR are widely used to characterise galaxy populations across cosmic time and an accurate recovery of these properties from SED fitting is essential for constraining galaxy formation models. In this sub-section, we examine how the inferred stellar mass and SFR influence the SMF and star formation main sequence (SFMS). Here, $M_{\star,\text{curr}}$ refers to the total stellar mass of a galaxy at $z = 6$; namely, the mass that remains in stars after stellar mass loss has been accounted for, rather than the total mass ever formed in stars.

4.2.1. Stellar mass functions

Figure 18 displays the SMFs of star-forming galaxies at $z = 6$, derived from the SPHINX²⁰ data (dashed black lines; Katz et al. 2023). The upper panels illustrate the SMFs obtained with the attSNE_Z model and an SMC-type dust law, while the lower panels show results from attSNE_Z_F410M. The left and right panels present results based on the median posterior and the minimum χ^2 estimate, respectively. For attSNE_Z with the median posterior, the SMF shows an offset of ~ 0.3 dex at $M_{\star,\text{curr}} \sim 10^{8.5} M_\odot$ relative to the true SMF, arising from stellar mass overestimation in low-mass galaxies that shifts the function rightward. The SMFs from the attSNE_Z_F410M model are nearly identical, indicating that inclusion of the medium band has little influence on the overall shape. By contrast, the offset is notably reduced with the minimum χ^2 estimate, implying that the SMF can be more reliably recovered using NGDEEP-like broad-band photometry.

Having established a baseline for uncertainties in stellar mass estimates, we can then compare the SMFs from SPHINX²⁰ with empirical determinations. Figure 18 displays SMFs from the studies of Song et al. (2016), Stefanon et al. (2021), and Navarro-Carrera et al. (2024), each derived via SED fitting of galaxy samples at $z \sim 6$. The first two are based on HST data, whereas the latter relies on JWST observations. As reported by Katz et al. (2023), SPHINX²⁰ galaxies generally overproduce stars compared to Song et al. (2016) and Stefanon et al. (2021). Because uncertainties in SED fitting typically bias stellar masses upward, this offset implies that the true discrepancy could be even larger. By contrast, the SMF obtained by Navarro-Carrera et al. (2024) agrees more closely with SPHINX²⁰, suggesting possible consistency between the simulation and JWST-based measurements. Nevertheless, given the limited survey area of that study, wider-field observations will be crucial for establishing whether the apparent agreement is genuine or merely coincidental.

4.2.2. Star formation main sequence

We also examined the SFMS to gain insight into ongoing star formation across the galaxy population. Figure 19 illustrates the SFMSs derived from the attSNE_Z_F410M model and emphasises how parameter biases affect the inferred relation. The upper and lower panels correspond to SFR_{10} and SFR_{100} , respectively, while the left and right panels contrast the median posterior with the minimum χ^2 . The SFMS from the attSNE_Z model is not included, as its behaviour largely mirrors that of attSNE_Z_F410M, apart from the constant SFH case. This model consistently yields the least accurate SFR, irrespective of the addition of F410M photometry.

The normalised true SFH model closely matches the true SFMS relation (solid black line), which represents the best-fit trend of the SPHINX²⁰ simulation at $z = 6$. This results from the fact that, under a true SFH, variations in stellar mass formed scale directly with the SFR, limiting deviations from the intrinsic relation. In contrast, the constant SFH model markedly underestimates SFR_{10} , making it the least reliable model for recovering the SFMS, regardless of the posterior statistic applied. Accordingly, the best-fit line for the constant SFH model is

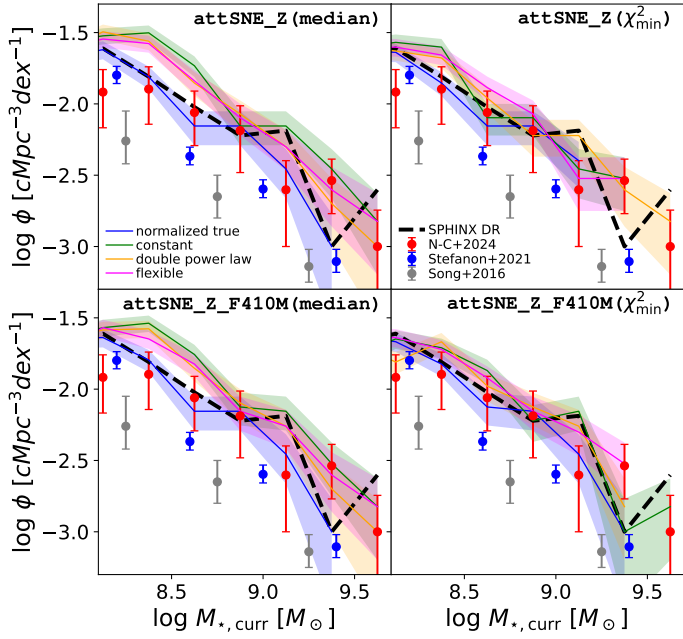


Fig. 18. SMFs of star-forming galaxies at $z = 6$. The dashed gray line denotes the true SMF from the SPHINX²⁰ simulation, while the solid coloured lines show the SMFs derived from the attSNE_Z (upper panel) and attSNE_Z_F410M (lower panel) models assuming the SMC dust law. Shaded regions mark Poisson uncertainties. The left panels use the posterior median, and the right panels use the minimum χ^2 posterior. Adding the F410M photometry yields a modest improvement in agreement with the true SMF when the median posterior is adopted, whereas the minimum χ^2 posterior produces a more accurate SMF, even without the extra band.

not displayed in the upper panel of Fig. 19. The limitation is alleviated in the $M_{*,\text{curr}}\text{-SFR}_{100}$ plane and is further suppressed when the minimum χ^2 posterior is employed. The double power-law and flexible SFH models produce broadly consistent results. Using the median posterior, galaxies are inferred to form stars less actively on 10 Myr timescales than they actually do in SPHINX²⁰. This bias stems from the joint effect of M_* overestimation and SFR_{10} underestimation. Adopting the minimum χ^2 posterior mitigates M_* overestimation and thereby reduces the deviation from the true SFMS, although an offset remains for galaxies with $M_{*,\text{curr}} \lesssim 10^{8.5} M_\odot$.

Although the $M_{*,\text{curr}}\text{-SFR}_{10}$ relation is sensitive to both the SFH model and the choice of summary statistic, the SFMS constructed from SFR_{100} reproduces the true relation more reliably. In the lower panel of Fig. 19, low-mass galaxies generally lie above the SFMS of the SPHINX²⁰ data release (solid black line), whereas intermediate- and high-mass galaxies lie below it. This behaviour is a result of the systematic overestimation of SFR_{100} in low-mass galaxies and its underestimation in high-mass galaxies, as illustrated in the bottom row of Fig. 17. A comparable trend is evident in Figs. 4 and 5: as M_* is overestimated owing to the enhanced contribution of older stellar populations, SFR_{100} is correspondingly underestimated. This effect is absent in low-mass galaxies, as most of their stars have formed within the past 100 Myr. Consequently, the reconstructed SFMS slopes are systematically shallower than the true relation, even when the minimum χ^2 is adopted. Notably, recent observations by Heintz et al. (2023) and Rinaldi et al. (2022) report a slope consistent with that of our reconstructed SFMSs, although the SPHINX²⁰ galaxies generally are inferred to form stars at a lower rate at fixed

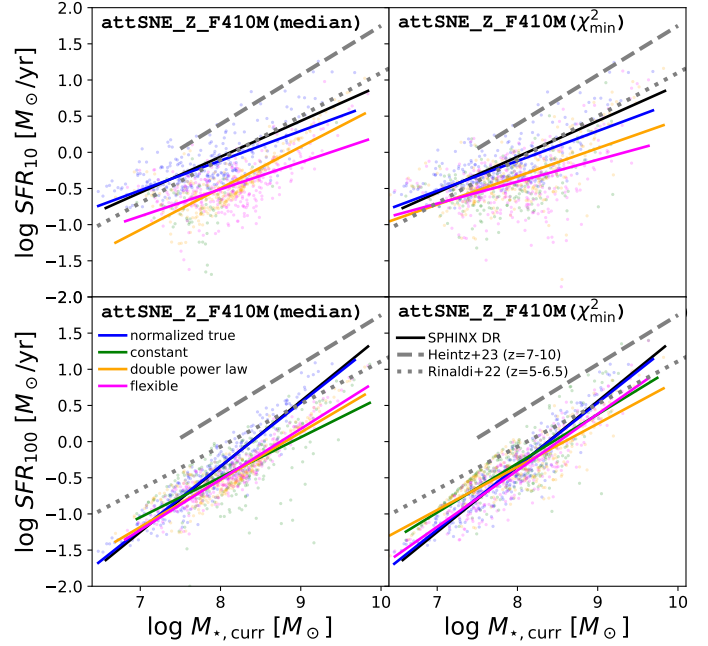


Fig. 19. SFMSs of the sample galaxies at $z = 6$. The upper and lower panels present the SFMSs constructed with SFR_{10} and SFR_{100} , respectively, while the left and right panels show results obtained from the median posterior and the posterior with the minimum χ^2 , respectively. The solid black line marks the best fit to the true SPHINX²⁰ data, and the solid coloured lines indicate best-fit relations derived with different SFH models in the attSNE_Z_F410M model. The best fit for the constant SFH model is excluded from the upper panels, as many galaxies return SFR_{10} values of zero. The minimum χ^2 posterior more accurately reproduces the true SFMS from the simulation more closely, although it lies slightly below current high- z observations (Heintz et al. 2023; Rinaldi et al. 2022).

stellar mass. This offset again likely reflects the excessive star formation in the simulation, leading to a horizontal shift in the relation. More importantly, the shallower slope inferred from SED fitting underscores the need for accurate and flexible SFH models to reliably recover the SFMS at high redshift.

4.2.3. Overabundance of bright galaxies

Finally, we discuss the implications of our results for the overabundance of bright galaxies observed at high redshift (Lovell et al. 2022; Finkelstein et al. 2023; Harikane et al. 2024), a phenomenon often regarded as a challenge to the Λ CDM framework. Figure 20 demonstrates that stellar masses inferred for actively star-forming galaxies can easily deviate by as much as a factor of two when the median posterior is adopted. This offset increases with specific SFR (sSFR), reaching ~ 0.6 dex (approximately a factor of four) at $\text{sSFR} \sim 10^{-7} \text{ yr}^{-1}$. If star formation in high- z galaxies is more bursty than in SPHINX²⁰ (e.g. Mauerhofer et al. 2025), the SED fitting might considerably overestimate stellar masses because of short-lived starbursts and strong emission lines. However, this does not mean that the observed tension can be explained solely by SED-fitting uncertainties, as the comparison is performed directly in the UV rather than in stellar mass (e.g. Yung et al. 2024; Mauerhofer et al. 2025). Instead, our findings reinforce previous arguments that the prevalence of UV-bright galaxies does not necessarily correspond to an overabundance of massive galaxies (Katz et al. 2023; Yung et al. 2024). By contrast,

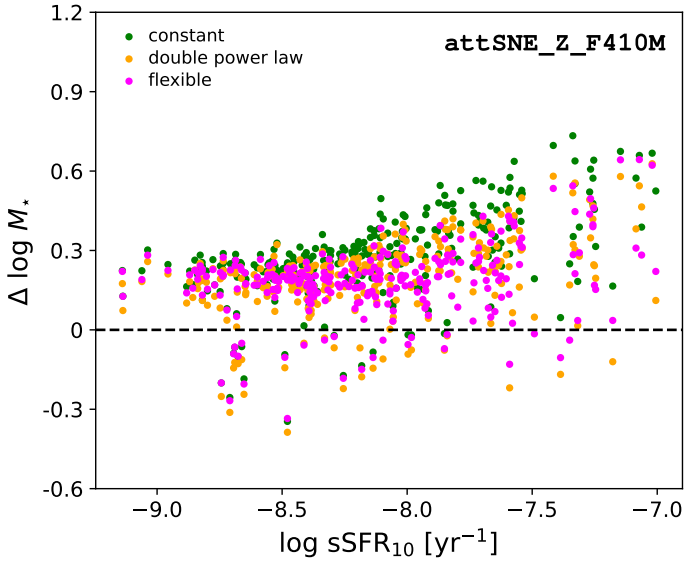


Fig. 20. Uncertainties in the inferred stellar mass as a function of specific SFR for the attSNE_Z_F410M model (median posterior). Different colours indicate results obtained with different SFH models, as shown in the legend. Stellar masses are increasingly overestimated at higher specific SFRs.

Narayanan et al. (2024) contended that massive systems ($M_{\star} \sim 10^9 M_{\odot}$) are largely dominated in mass by old stars, while outshining from young stellar populations leads to a systematic underestimation of M_{\star} . Because the test simulations, such as SIMBA (Davé et al. 2019) and SMUGGLE (Marinacci et al. 2019), and the SED-fitting methods (PROSPECTOR, Leja et al. 2017) differ, the origin of this discrepancy remains unresolved. Nevertheless, both studies emphasise the need for caution when estimating stellar masses and their uncertainties, before appealing to more exotic explanations for the excess of bright galaxies, such as non-standard cosmology.

5. Summary and conclusions

In this study, we investigated how assumptions made in the process of performing an SED fitting, such as the SFH, dust, and nebular emission, affect the recovery of key galaxy physical parameters of galaxies when using JWST/NIRCam photometry. Using a suite of synthetic SEDs of star-forming galaxies at $z = 6$ taken from the SPHINX²⁰ simulation (Katz et al. 2023) and applying a Bayesian fitting with BAGPIPES (Carnall et al. 2018), we systematically analysed the impact of each component through a series of model variations. Our primary results are summarised below.

- In the absence of dust attenuation and nebular emission (i.e. the intrinsic stellar continuum), SED fitting with JWST/NIRCam photometry (covering rest-frame UV to optical bands) systematically overestimates stellar masses (Fig. 3). Although the mean deviation is modest (0.20 dex), the stellar masses of low-mass galaxies ($M_{\star} \lesssim 10^{7.5} M_{\odot}$) end up overestimated by up to 0.73 dex. This bias arises primarily from metallicity and age effects, as the fits tend to favour older stellar populations. Consequently, SFR_{10} is underestimated by an average of 0.11 dex for the double power-law and flexible SFH models, whereas the constant SFH model yields $\text{SFR}_{10} = 0$ in most cases (Fig. 5). This trend worsens when the dust attenuation and nebular emission are included,

highlighting how these components exacerbate degeneracies among stellar age, metallicity, and mass-to-light ratio.

- By comparing the slope of the assumed model attenuation curve with that of the true curve from the simulated galaxy, we find that the stellar mass is overestimated when the model curve is shallower and underestimated when it is steeper (Fig. 7). This bias persists even when the normalised true SFH is used because mismatches in the attenuation slope force the model to adjust A_V to reproduce the observed colours, thereby yielding an incorrect total attenuation.
- The choice of the SFH model influences the distribution of the intrinsic UV slope. The constant SFH model leads to the largest overestimation, followed by the double power-law and flexible SFH models. This, in turn, alters the distribution of the inferred colour excess and A_V by causing both to be underestimated, which introduces systematic biases in stellar mass and SFR estimates (Fig. 9).
- Strong emission lines at optical wavelengths redden the observed colours, particularly in low-mass galaxies ($\lesssim 10^8 M_{\odot}$) where nebular contributions are most pronounced (Fig. 11). To reproduce these redder colours, the fitting procedure tends to favour older stellar populations, resulting in stellar masses being overestimated by more than 1 dex. In contrast, for galaxies with stellar masses above $10^{8.5} M_{\odot}$, where the emission lines contribute only marginally to the total flux, this effect remains minimal.
- By fitting SEDs without emission lines in either the model or the input spectrum, we were able to isolate their effect. Stellar mass estimates can vary by up to 0.8 dex solely because of emission lines in the input spectrum (Fig. 13), whereas neglecting them in SED fitting introduces biases of up to 0.3 dex (Fig. 14), driven mainly by changes in the inferred stellar population.
- Including photometric data largely free from strong emission-line contamination, such as the F410M filter, improves the recovery of young stellar populations by constraining the contribution of emission lines relative to the stellar continuum. The overestimation of stellar mass, which reached 1.26 dex for galaxies with $M_{\star} \lesssim 10^8 M_{\odot}$, was reduced to within 0.7 dex, while the effect was less pronounced in more massive systems (Fig. 16).
- Furthermore, adopting either minimum χ^2 values or likelihood-weighted means, rather than posterior medians, yields closer agreement with the true bulk galaxy properties. These approaches more reliably recovered key observables, including the SMF and the star-forming main sequence, even under complex modelling assumptions (Figs. 18 and 19).

Taken together, our analysis underscores the cumulative influence of model assumptions in SED fittings, particularly in regimes where bursty star formation, nebular emission, and dust attenuation are non-negligible. Future studies should account for these factors when interpreting photometric SEDs to better constrain galaxy formation at high redshift, especially for early galaxies, which are increasingly thought to have bursty SFHs. A promising avenue for improvement is machine-learning-based SED fitting trained on successful cosmological simulations, which could yield more accurate stellar mass estimates (e.g. Gilda et al. 2021).

Acknowledgements. We thank Sukyoung Yi and Hyunmi Song for detailed comments on the manuscript. J.C. and T.K. are supported by the National Research Foundation of Korea (RS-2022-NR070872 and RS-2025-00516961) and also by the Yonsei Fellowship, funded by Lee Youn Jae. T.K. acted as the cor-

responding author. The large data transfer was supported by KREONET, which is managed and operated by KISTI.

References

- Abel, T., Bryan, G. L., & Norman, M. L. 2002, *Science*, 295, 93
- Anders, P., & Fritze-v. Alvensleben, U. 2003, *A&A*, 401, 1063
- Bagley, M. B., Pirzkal, N., Finkelstein, S. L., et al. 2024, *ApJ*, 965, L6
- Béthermin, M., Dessauges-Zavadsky, M., Faisst, A. L., et al. 2020, *The Messenger*, 180, 31
- Bromm, V., Coppi, P. S., & Larson, R. B. 2002, *ApJ*, 564, 23
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
- Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, *ApJ*, 840, 44
- Calzetti, D., Kinney, A. L., & Storchi-Bergmann, T. 1994, *ApJ*, 429, 582
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
- Cameron, A. J., Katz, H., Witten, C., et al. 2024, *MNRAS*, 534, 523
- Caputi, K. I., Deshmukh, S., Ashby, M. L. N., et al. 2017, *ApJ*, 849, 45
- Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, *MNRAS*, 480, 4379
- Carniani, S., Hainline, K., D'Eugenio, F., et al. 2024, *Nature*, 633, 318
- Chemerynska, I., Atek, H., Dayal, P., et al. 2024, *ApJ*, 976, L15
- Chevallard, J., & Charlot, S. 2016, *MNRAS*, 462, 1415
- Cochrane, R. K., Katz, H., Begley, R., Hayward, C. C., & Best, P. N. 2025, *ApJ*, 978, L42
- Conroy, C. 2013, *ARA&A*, 51, 393
- Csizi, B., Tortorelli, L., Siudek, M., et al. 2024, *A&A*, 689, A37
- Curti, M., Maiolino, R., Curtis-Lake, E., et al. 2024, *A&A*, 684, A75
- da Cunha, E., Walter, F., Smail, I. R., et al. 2015, *ApJ*, 806, 110
- Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, *MNRAS*, 486, 2827
- Dayal, P., & Ferrara, A. 2018, *Phys. Rep.*, 780, 1
- De Looze, I., Cormier, D., Leboutteiller, V., et al. 2014, *A&A*, 568, A62
- Dekel, A., Sarkar, K. C., Birnboim, Y., Mandelker, N., & Li, Z. 2023, *MNRAS*, 523, 3201
- Dressler, A., Vulcani, B., Treu, T., et al. 2023, *ApJ*, 947, L27
- Dressler, A., Rieke, M., Eisenstein, D., et al. 2024, *ApJ*, 964, 150
- Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, *Rev. Mex. Astron. Astrofis.*, 53, 385
- Feroz, F., & Hobson, M. P. 2008, *MNRAS*, 384, 449
- Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, 398, 1601
- Ferrara, A., Pallottini, A., & Dayal, P. 2023, *MNRAS*, 522, 3986
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJ*, 946, L13
- Finlator, K., Keating, L., Oppenheimer, B. D., Davé, R., & Zackrisson, E. 2018, *MNRAS*, 480, 2628
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
- Garel, T., Blaizot, J., Rosdahl, J., et al. 2021, *MNRAS*, 504, 1902
- Gilda, S., Lower, S., & Narayanan, D. 2021, *ApJ*, 916, 43
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, *ApJ*, 594, 279
- Harikane, Y., Nakajima, K., Ouchi, M., et al. 2024, *ApJ*, 960, 56
- Harvey, T., Conselice, C. J., Adams, N. J., et al. 2025, *ApJ*, 978, 89
- Haskell, P., Das, S., Smith, D. J. B., et al. 2024, *MNRAS*, 530, L7
- Hayes, M., Schaefer, D., Östlin, G., et al. 2011, *ApJ*, 730, 8
- Heintz, K. E., De Cia, A., Thöne, C. C., et al. 2023, *A&A*, 679, A91
- Hu, C.-Y., Smith, M. C., Teyssier, R., et al. 2023, *ApJ*, 950, 132
- Jain, S., Tacchella, S., & Mosleh, M. 2024, *MNRAS*, 527, 3291
- Jolly, J.-B., Knudsen, K., Laporte, N., et al. 2021, *A&A*, 652, A128
- Jones, G. T., Stanway, E. R., & Carnall, A. C. 2022, *MNRAS*, 514, 5706
- Kannan, R., Marinacci, F., Vogelsberger, M., et al. 2020, *MNRAS*, 499, 5732
- Katz, H., Rosdahl, J., Kimm, T., et al. 2023, *Open J. Astrophys.*, 6, 44
- Katz, H., Cameron, A. J., Saxena, A., et al. 2025, *Open J. Astrophys.*, 8, 104
- Kimm, T., & Cen, R. 2014, *ApJ*, 788, 121
- Kimm, T., Cen, R., Devriendt, J., Dubois, Y., & Slyz, A. 2015, *MNRAS*, 451, 2900
- Kimm, T., Katz, H., Haehnelt, M., et al. 2017, *MNRAS*, 466, 4826
- Kroupa, P., & Boily, C. M. 2002, *MNRAS*, 336, 1188
- Kroupa, P., Gjergo, E., Jerabkova, T., & Yan, Z. 2026, *Encycl. Astrophys.*, 2, 173
- Laporte, N., Ellis, R. S., Boone, F., et al. 2017, *ApJ*, 837, L21
- Laursen, P., Sommer-Larsen, J., & Andersen, A. C. 2009, *ApJ*, 704, 1640
- Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, *ApJ*, 837, 170
- Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019, *ApJ*, 876, 3
- Leung, G. C. K., Finkelstein, S. L., Pérez-González, P. G., et al. 2025, *ApJ*, 992, 26
- Lovell, C. C., Harrison, I., Harikane, Y., Tacchella, S., & Wilkins, S. M. 2022, *MNRAS*, 518, 2511
- Lower, S., Narayanan, D., Leja, J., et al. 2020, *ApJ*, 904, 33
- Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, *MNRAS*, 456, 2140
- Marinacci, F., Sales, L. V., Vogelsberger, M., Torrey, P., & Springel, V. 2019, *MNRAS*, 489, 4233
- Mármol-Queraltó, E., McLure, R., & Cullen, F. 2015, in *Highlights of Spanish Astrophysics VIII*, eds. A. J. Cenarro, F. Figueras, C. Hernández-Monteagudo, J. Trujillo Bueno, & L. Valdivielso, 263
- Marshall, M. A., Watts, K., Wilkins, S., et al. 2022, *MNRAS*, 516, 1047
- Mauerhofer, V., Dayal, P., Haehnelt, M. G., et al. 2025, *A&A*, 696, A157
- McLure, R. J., Dunlop, J. S., de Ravel, L., et al. 2011, *MNRAS*, 418, 2074
- Meldorf, C., Palmese, A., & Salim, S. 2024, *MNRAS*, 531, 3242
- Michel-Dansac, L., Blaizot, J., Garel, T., et al. 2020, *A&A*, 635, A154
- Miranda, H., Pappalardo, C., Afonso, J., et al. 2025, *A&A*, 694, A102
- Naab, T., & Ostriker, J. P. 2017, *ARA&A*, 55, 59
- Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, *ApJS*, 269, 33
- Narayanan, D., Conroy, C., Davé, R., Johnson, B. D., & Popping, G. 2018, *ApJ*, 869, 70
- Narayanan, D., Lower, S., Torrey, P., et al. 2024, *ApJ*, 961, 73
- Narayanan, D., Stark, D. P., Finkelstein, S. L., et al. 2025, *ApJ*, 982, 7
- Navarro-Carrera, R., Rinaldi, P., Caputi, K. I., et al. 2024, *ApJ*, 961, 207
- Pacifici, C., Iyer, K. G., Mobasher, B., et al. 2023, *ApJ*, 944, 141
- Papovich, C., Dickinson, M., & Ferguson, H. C. 2001, *ApJ*, 559, 620
- Qin, J., Zheng, X. Z., Fang, M., et al. 2022, *MNRAS*, 511, 765
- Rieke, M. J., Robertson, B., Tacchella, S., et al. 2023, *ApJS*, 269, 16
- Rinaldi, P., Caputi, K. I., van Mierlo, S. E., et al. 2022, *ApJ*, 930, 128
- Rinaldi, P., Caputi, K. I., Costantin, L., et al. 2023, *ApJ*, 952, 143
- Roberts-Borsani, G., Treu, T., Mason, C., et al. 2021, *ApJ*, 910, 86
- Robertson, B. E. 2022, *ARA&A*, 60, 121
- Rosdahl, J., Blaizot, J., Aubert, D., Stranex, T., & Teyssier, R. 2013, *MNRAS*, 436, 2188
- Rosdahl, J., Katz, H., Blaizot, J., et al. 2018, *MNRAS*, 479, 994
- Rosdahl, J., Blaizot, J., Katz, H., et al. 2022, *MNRAS*, 515, 2386
- Rosen, A., & Bregman, J. N. 1995, *ApJ*, 440, 634
- Sabti, N., Muñoz, J. B., & Kamionkowski, M. 2024, *Phys. Rev. Lett.*, 132, 061002
- Salim, S., Boquien, M., & Lee, J. C. 2018, *ApJ*, 859, 11
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, 799, 183
- Schaerer, D., & de Barros, S. 2009, *A&A*, 502, 423
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Shen, X., Vogelsberger, M., Boylan-Kolchin, M., Tacchella, S., & Kannan, R. 2023, *MNRAS*, 525, 3254
- Smit, R., Bouwens, R. J., Carniani, S., et al. 2018, *Nature*, 553, 178
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, 53, 51
- Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, *ApJ*, 825, 5
- Stanway, E. R., & Eldridge, J. J. 2018, *MNRAS*, 479, 75
- Stefanon, M., Bouwens, R. J., Labbé, I., et al. 2021, *ApJ*, 922, 29
- Sun, G., Faucher-Giguère, C.-A., Hayward, C. C., et al. 2023, *ApJ*, 955, L35
- Traina, A., Magnelli, B., Gruppioni, C., et al. 2024, *A&A*, 690, A84
- Trayford, J. W., Lagos, C. d. P., Robotham, A. S. G., & Obreschkow, D. 2019, *MNRAS*, 491, 3937
- Tweed, D., Devriendt, J., Blaizot, J., Colombi, S., & Slyz, A. 2009, *A&A*, 506, 647
- Wilkins, S. M., Coulton, W., Caruana, J., et al. 2013, *MNRAS*, 435, 2885
- Wilkins, S. M., Lovell, C. C., Fairhurst, C., et al. 2020, *MNRAS*, 493, 6079
- Worthey, G. 1994, *ApJS*, 95, 107
- Yuan, F.-T., Burgarella, D., Corre, D., et al. 2019, *A&A*, 631, A123
- Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Popping, G., & Davé, R. 2019, *MNRAS*, 483, 2983
- Yung, L. Y. A., Somerville, R. S., Finkelstein, S. L., Wilkins, S. M., & Gardner, J. P. 2024, *MNRAS*, 527, 5929
- Zackrisson, E., Bergvall, N., & Leitet, E. 2008, *ApJ*, 676, L9

Appendix A: Impact of SED template and SFH model choice on fitting

As indicated in the main text, we adopt the BPASS SPS model in BAGPIPES, as the input SEDs from SPHINX²⁰ are likewise generated with this model. In contrast, [Cochrane et al. \(2025\)](#) employed the [Bruzual & Charlot \(2003, BC03\)](#) model to evaluate the accuracy of recovering stellar masses and SFRs from SPHINX²⁰ SEDs. Therefore, examining how the choice of SED template influences the inferred galaxy properties is necessary.

Figures [A.1](#) and [A.2](#) illustrate the impact of the SPS template on SED fitting results, highlighting the comparison between BPASS and BC03. We perform intrinsic stellar continuum fitting with several SFH models and compare the resulting M_\star values with those obtained using BPASS (i.e. the $\text{intS_Z}_{\text{true}}$ model in [Table 1](#)). [Figure A.1](#) displays the difference between the true and fitted M_\star values when each SPS template is applied with the true normalised SFH model, and the inset depicts an example of the intrinsic stellar continuum spanning rest-frame NUV to optical wavelengths. We find that the recovered M_\star with BC03 is slightly larger (by $\sim 10\%$) than that with BPASS. This difference arises from the lower flux at rest-frame optical wavelengths in the stellar continuum when BC03 is applied, as can be seen in the inset of [Fig. A.1](#). As shown by [Stanway & Eldridge \(2018\)](#), BPASS yields slightly more luminous and redder spectra at intermediate ages, driven by differences in the treatment of AGB stars, stellar atmosphere models, and binary evolution. Consequently, when Z_\star is allowed to vary, the fitted Z_\star with BC03 becomes larger to match the redder input colour of the SPHINX²⁰ galaxy, which is based on the BPASS.

A higher Z_\star generally produces lower fluxes across all wavelengths, which can drive an increase in the fitted M_\star to match the input flux. This preference for higher Z_\star in BC03-based SED fitting persists across different SFH models and remains evident even with more complex approaches, such as `attSNE_Z_F410M`, although the trend weakens as model complexity increases. This mitigation is illustrated in [Fig. A.2](#), which presents the median posterior for `attSNE_Z_F410M` (cf. [Fig. 17](#)). Because BPASS generates more UV flux than BC03, the associated emission-line luminosity is brighter for the same SFH, which yields a redder colour (cf., [Fig. 11](#)). As the F410M photometry disentangles emission-line contributions from the optical continuum, a higher SFR_{10} is required to reproduce the emission-line luminosity when BC03 is employed. At the same time, because the stellar continuum is redder in BPASS (inset of [Fig. A.1](#)), a higher metallicity is still necessary. While the first effect tends to reduce M_\star , the second drives it upward. Moreover, an increase in SFR_{10} steepens the UV slope, necessitating a higher colour excess to reproduce the input UV colour. Because the SMC-type dust law raises the colour excess only through a higher A_V , the fitted M_\star must also increase to compensate for the added dust extinction. Consequently, when BC03 is adopted, the increase in SFR_{10} partly counterbalances the higher Z_\star and A_V , yielding a modest overestimation of the fitted M_\star . We find that the increases in these three parameters are broadly consistent across other complex SFH models ([Fig. A.2](#)). Nevertheless, using BC03 introduces only minor changes in the derived stellar mass, and we confirm that these differences are not insufficient to affect our main conclusions.

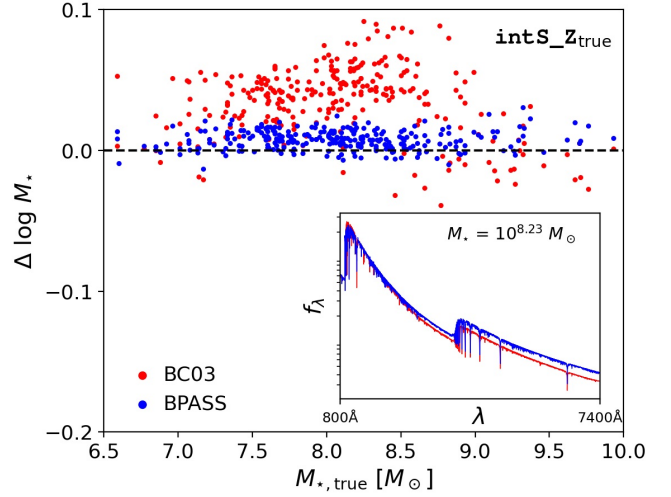


Fig. A.1. Difference between the true and fitted M_\star for SPHINX²⁰ galaxies at $z = 6$ when the normalised true SFH model is fitted with $\text{intS_Z}_{\text{true}}$. Using BC03 template is used (red) yields systematically larger M_\star values than BPASS (blue). This offset arises because BC03 predicts lower rest-frame optical fluxes than BPASS, even under the same SFH is used. The inset presents an example SED fit with both templates, with the x-axis showing to observed wavelengths from 0.9 to 5.0 microns.

Appendix B: Mass offset for the true normalise SFH model

In [Sect. 3.2](#), we demonstrate a tight correlation between the UV-optical slope S and $\Delta \log M_\star$ under the assumption of the true normalised SFH model. Here, we explain our derivation of [Eq. 4](#) in more detail.

We consider a galaxy for which the intrinsic and attenuated stellar continuum fluxes are denoted by $F(\lambda)$ and $f(\lambda)$, respectively. These two quantities are related by dust attenuation, $A(\lambda)$, as follows:

$$A(\lambda) = -2.5 \log \frac{f(\lambda)}{F(\lambda)} = A_V s(\lambda), \quad (\text{B.1})$$

where $s(\lambda)$ denotes the normalised dust attenuation curve (or slope), and $A_V = A(5500 \text{ \AA})$ represents the attenuation at 5500 Å. As the shape of the intrinsic stellar continuum is determined by the stellar population, we can express $F(\lambda)$ as $F(\lambda) = M_\star \Upsilon(\lambda)$, where $\Upsilon(\lambda)$ denotes the stellar mass-to-light ratio. Equation [B.1](#) can then be re-expressed as

$$\log M_\star + \log \Upsilon(\lambda) - \frac{A_V s(\lambda)}{2.5} = \log f(\lambda). \quad (\text{B.2})$$

Now consider a fitted model galaxy with the same stellar population but subject to SMC-type dust attenuation, as

$$\log M_{\star, \text{fit}} + \log \Upsilon(\lambda) - \frac{A_{V, \text{fit}} s_{\text{SMC}}(\lambda)}{2.5} = \log f_{\text{fit}}(\lambda), \quad (\text{B.3})$$

where $M_{\star, \text{fit}}$, $A_{V, \text{fit}}$ and $f_{\text{fit}}(\lambda)$ denote the fitted stellar mass, attenuation, and attenuated flux, respectively; and $s_{\text{SMC}}(\lambda)$ is the SMC attenuation curve. If the fitted model perfectly reproduces the attenuated SED of the galaxy, then $f(\lambda) = f_{\text{fit}}(\lambda)$ is true for all wavelengths. Combining [Eqs. B.2](#) and [B.3](#), we obtain

$$\begin{aligned} \log M_{\star, \text{fit}} - \log M_\star &= \frac{A_{V, \text{fit}} s_{\text{SMC}}(\lambda)}{2.5} - \frac{A_V s(\lambda)}{2.5}, \\ &= \frac{A_V}{2.5} \left[\frac{A_{V, \text{fit}}}{A_V} s_{\text{SMC}}(\lambda) - s(\lambda) \right]. \end{aligned} \quad (\text{B.4})$$

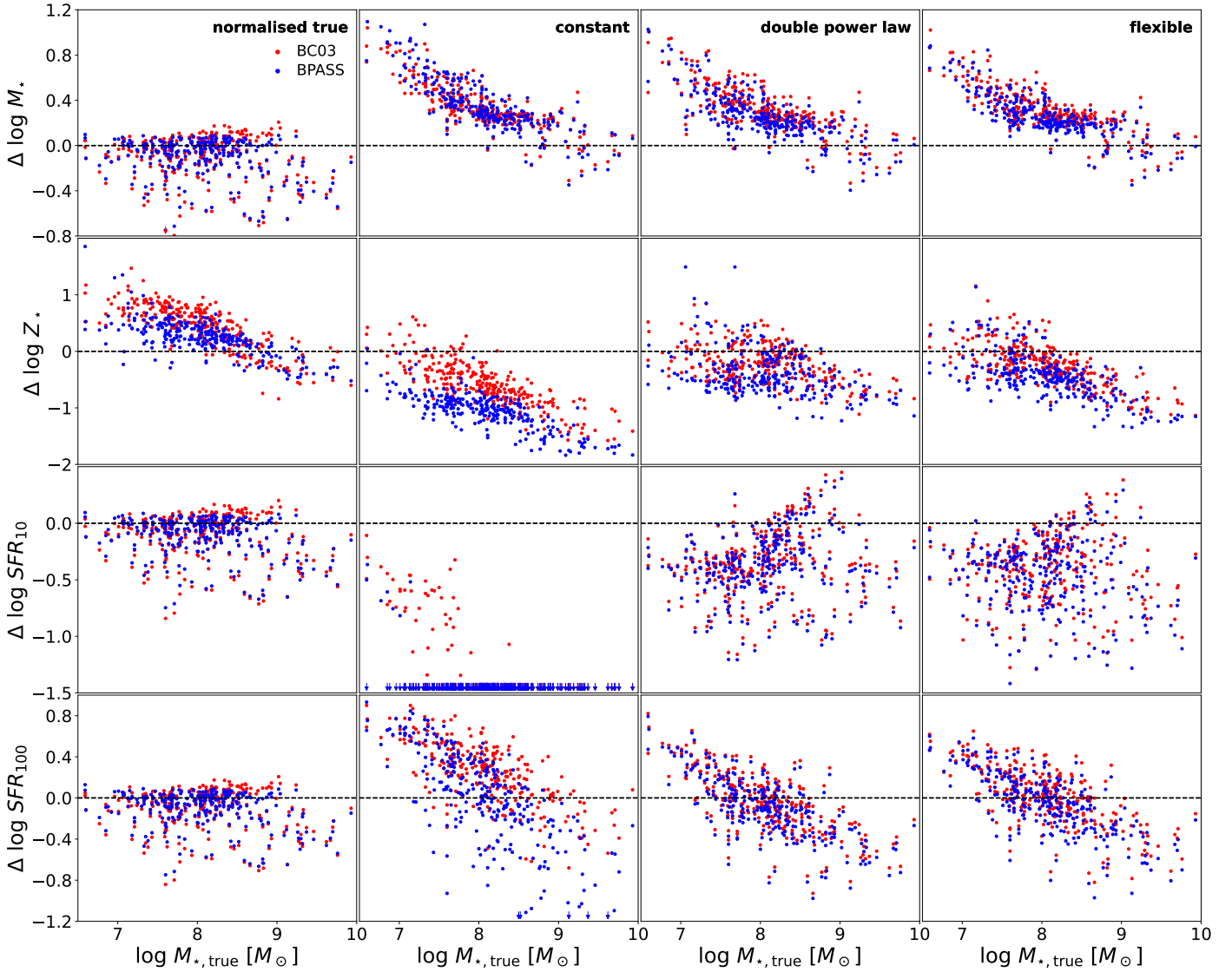


Fig. A.2. Same as Fig. 17, but with the BC03 template used for SED fitting. Only median posteriors are displayed. On average, the fitted M_* and SFRs differ by ~ 0.04 dex between templates, whereas the inferred metallicities differ by up to a factor of three. Notably, SFR_{10} remains severely underestimated when the constant SFH is fitted with BC03.

At $\lambda = 5500 \text{ \AA}$, Eq. B.4 becomes

$$\Delta \log M_{*,\text{fit}} = \frac{A_V}{2.5} \left(\frac{A_{V,\text{fit}}}{A_V} - 1 \right). \quad (\text{B.5})$$

Thus, when $A_{V,\text{fit}}$ is larger (smaller) than the true A_V , M_* is overestimated (underestimated, respectively).

Appendix C: Fitting results with redshift as a free parameter

As detailed in the main text, the redshift is fixed to its true value to focus on uncertainties in galactic properties, including SFHs. However, the redshift of observed galaxies is typically not known a priori. To assess the effect of redshift uncertainties, we repeat the analysis with the redshift treated as a free parameter, adopting a prior range of 0–10 and using the `attSNE_Z_F410M` prescriptions.

Figure C.1 presents the fitting results for stellar mass (left) and for redshift (right). The recovered stellar masses remain broadly consistent with those from the `attSNE_Z_F410M` model

(lower panel of Fig. 16), when the recovered redshift deviates by less than unity from the true value (i.e. $5 \lesssim z \lesssim 7$). Redshift recovery is most reliable with the double power-law (97%), followed by the normalised true (94%), flexible (85%) and constant SFH models (59%). Galaxies wherein the redshift is not recovered typically yield $z \sim 1.5$, as the Balmer break in the fitted SED resembles the Gunn–Peterson trough in the true SED. Our simulated samples are prone to this redshift degeneracy because the filter does not cover the UV region of the SED where IGM attenuation is strong. This degeneracy can be mitigated by employing a filter that covers shorter wavelengths, such as F070W. Alternatively, adopting the minimum- χ^2 posterior can mitigate this issue, because posteriors with $z \sim 1.5$ generally yield larger χ^2 values than those with $z \sim 6$.

Appendix D: Uncertainties in derived physical properties from SED fitting

We summarise the uncertainties in various physical properties derived from SED fitting in Table 2–D.2.

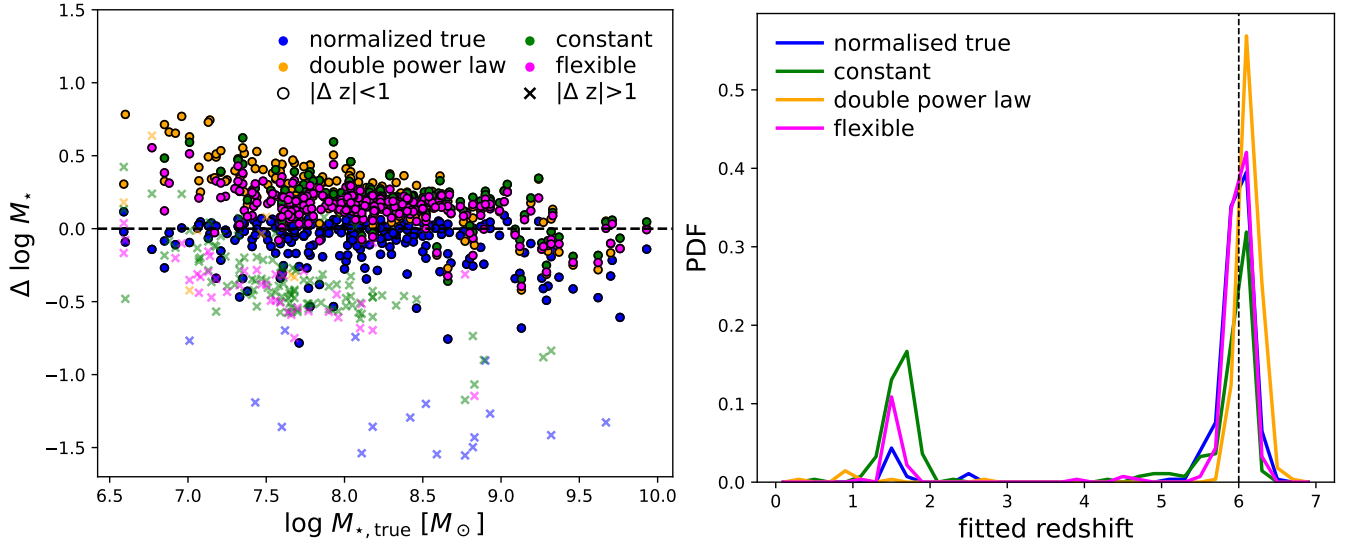


Fig. C.1. Impact of redshift determination on the inferred stellar mass offset (left). Different colours denote the SFH models listed in the legend. Stellar masses are recovered reasonably well when the estimated redshift lies within $|\Delta z| < 1$ (circles), whereas greater scatter arises when $|\Delta z| > 1$ (crosses). The corresponding redshift distributions are displayed in the right panel, with the true redshift ($z = 6$) indicated by the dashed line.

Table D.1. Offsets in the inferred physical properties from different SED fitting models for low-mass ($< 10^8 M_\odot$) galaxies.

Variable	SFH model	intS_Z _{true}	intS_Z	attS_Z _{true}	attS_Z	attSNE_Z _{true}	attSNE_Z	attSNE_Z_F410M
$\Delta \log M_*$	Normalised true	0.01 ± 0.01	0.01 ± 0.01	0.02 ± 0.10	0.05 ± 0.11	-0.07 ± 0.15	-0.05 ± 0.16	-0.04 ± 0.15
	Constant	0.07 ± 0.21	0.27 ± 0.16	0.12 ± 0.17	0.29 ± 0.15	0.34 ± 0.23	0.50 ± 0.22	0.37 ± 0.13
	Double power law	0.03 ± 0.18	0.20 ± 0.15	0.09 ± 0.13	0.24 ± 0.14	0.26 ± 0.22	0.36 ± 0.19	0.27 ± 0.14
	Flexible	0.09 ± 0.17	0.21 ± 0.15	0.12 ± 0.13	0.23 ± 0.13	0.25 ± 0.21	0.35 ± 0.17	0.21 ± 0.14
$\Delta \log \text{SFR}_{10}$	Normalised true	0.00 ± 0.01	0.01 ± 0.01	0.01 ± 0.10	0.04 ± 0.11	-0.07 ± 0.15	-0.05 ± 0.16	-0.05 ± 0.15
	Double power law	-0.14 ± 0.12	-0.20 ± 0.11	-0.27 ± 0.27	-0.31 ± 0.27	-0.41 ± 0.26	-0.42 ± 0.25	-0.37 ± 0.24
	Flexible	-0.07 ± 0.16	-0.10 ± 0.12	-0.15 ± 0.28	-0.20 ± 0.30	-0.35 ± 0.27	-0.38 ± 0.26	-0.27 ± 0.24
$\Delta \log \text{SFR}_{100}$	Normalised true	0.01 ± 0.01	0.01 ± 0.01	0.02 ± 0.10	0.05 ± 0.11	-0.07 ± 0.15	-0.05 ± 0.16	-0.04 ± 0.15
	Constant	0.36 ± 0.20	0.47 ± 0.16	0.37 ± 0.21	0.38 ± 0.28	0.38 ± 0.26	0.34 ± 0.34	0.31 ± 0.31
	Double power law	0.25 ± 0.17	0.28 ± 0.17	0.18 ± 0.21	0.15 ± 0.24	0.09 ± 0.26	0.08 ± 0.25	0.12 ± 0.24
	Flexible	0.13 ± 0.14	0.17 ± 0.14	0.09 ± 0.18	0.13 ± 0.19	0.12 ± 0.22	0.13 ± 0.23	0.08 ± 0.20

Notes. The table shows the 16th, 50th, and 84th percentiles. ΔSFR_{10} for the constant SFH model is not shown, as $\text{SFR}_{10,\text{fit}} = 0$ in many galaxies. Values with the minimum offset in attSNE_Z and attSNE_Z_F410M are shown in bold.

Table D.2. Same as Table D.1 but for high-mass ($> 10^8 M_\odot$) galaxies.

Variable	SFH model	intS_Z _{true}	intS_Z	attS_Z _{true}	attS_Z	attSNE_Z _{true}	attSNE_Z	attSNE_Z_F410M
$\Delta \log M_*$	Normalised true	0.01 ± 0.01	0.02 ± 0.01	-0.01 ± 0.14	0.02 ± 0.15	-0.08 ± 0.18	-0.06 ± 0.19	-0.04 ± 0.17
	Constant	-0.07 ± 0.08	0.19 ± 0.07	-0.05 ± 0.13	0.17 ± 0.12	0.01 ± 0.16	0.23 ± 0.15	0.22 ± 0.14
	Double power law	-0.00 ± 0.09	0.20 ± 0.06	0.05 ± 0.12	0.19 ± 0.12	0.08 ± 0.15	0.17 ± 0.15	0.17 ± 0.14
	Flexible	0.06 ± 0.06	0.18 ± 0.06	0.05 ± 0.11	0.18 ± 0.12	0.05 ± 0.15	0.18 ± 0.13	0.19 ± 0.12
$\Delta \log \text{SFR}_{10}$	Normalised true	0.00 ± 0.01	0.01 ± 0.01	-0.01 ± 0.14	0.02 ± 0.15	-0.08 ± 0.18	-0.06 ± 0.19	-0.05 ± 0.17
	Double power law	-0.02 ± 0.14	-0.09 ± 0.14	-0.20 ± 0.31	-0.23 ± 0.31	-0.29 ± 0.31	-0.30 ± 0.31	-0.29 ± 0.31
	Flexible	0.02 ± 0.12	-0.06 ± 0.11	-0.14 ± 0.33	-0.27 ± 0.35	-0.28 ± 0.32	-0.41 ± 0.34	-0.41 ± 0.34
$\Delta \log \text{SFR}_{100}$	Normalised true	0.01 ± 0.01	0.01 ± 0.01	-0.01 ± 0.14	0.02 ± 0.15	-0.08 ± 0.18	-0.06 ± 0.19	-0.05 ± 0.17
	Constant	0.24 ± 0.16	0.20 ± 0.19	0.11 ± 0.28	-0.10 ± 0.39	0.07 ± 0.27	-0.15 ± 0.41	-0.13 ± 0.42
	Double power law	0.05 ± 0.17	0.00 ± 0.18	-0.11 ± 0.25	-0.12 ± 0.24	-0.20 ± 0.24	-0.21 ± 0.25	-0.21 ± 0.25
	Flexible	-0.00 ± 0.14	-0.01 ± 0.15	-0.06 ± 0.24	-0.10 ± 0.26	-0.13 ± 0.24	-0.17 ± 0.26	-0.16 ± 0.26