Tracing galaxy evolution through the reconstruction of their star formation histories

by

Kristi Webb

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Supervisor:	Michael Balogh Associate Professor, Dept. of Physics and Astronomy, University of Waterloo
Committee Member:	Avery Broderick Associate Professor, Dept. of Physics and Astronomy, University of Waterloo
Committee Member:	Brian McNamara Professor, Dept. of Physics and Astronomy, University of Waterloo
Internal/External Member:	Marcel Nooijen Professor, Dept. of Chemistry, University of Waterloo
External Examiner:	Claudia Maraston Professor, Institute of Cosmology and Gravitation, University of Portsmouth

Author's Declaration

This thesis consists of material all of which I co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis is based on the following articles:

CHAPTER 3: **Webb K.**, Balogh M. L., Leja J., van der Burg R. F. J., Rudnick G., Muzzin A., Boak K., et al., 2020. The GOGREEN survey: post-infall environmental quenching fails to predict the observed age difference between quiescent field and cluster galaxies at $z > 1$. *Monthly Notices of the Royal Astronomical Society*, 498, 5317–5342. doi:10.1093/mnras/staa2752

CHAPTER 4: **Webb K.**, Villaume A., Laine S., Romanowsky A. J., Balogh M. L., van Dokkum P., Forbes D. A, et al., 2022. Still at Odds With Conventional Galaxy Evolution: Simultaneously Modelling Photometry and Spectroscopy of the Ultra-Diffuse Galaxy Dragonfly 44 to Measure its Star Formation History. *Monthly Notices of the Royal Astronomical Society*, 516, 3318-3341. doi:10.1093/mnras/stac2417

CHAPTER 5: Based on work currently in preparation of publication.

For each of these contributions, I did the majority of writing and analysis, including developing and running statistical models for the interpretation of the observations. This research was conducted under the supervision of Michael Balogh, with contributions and consultations from the GOGREEN collaborators, and Alexa Villaume. GOGREEN collaborators include Greg Rudnick, Adam Muzzin, Remco van der Berg, Benedetta Vulcani, in addition to the coauthors listed for Webb et al. (2020). Michael Balogh provided direction and helpful suggestions with respect to the analysis and interpretation of results of Webb et al. (2020, Chapter 3), and is the PI of the GOGREEN survey used in this work. I assisted with the Gemini Observatory observations as part of GOGREEN. Joel Leja and Ben Johnson built the software PROSPECTOR used throughout this thesis, and both provided assistance in its use. Alexa Villaume contributed to the development of Webb et al. (2022, Chapter 4), provided the reduced spectroscopy studied in this work, and helped in writing the introduction and discussion sections of this work. Seppo Laine provided the reduced photometry analysed in Chapter 4. Both Michael and Alexa have provided guidance in the analysis of Chapter 5, where Alexa originally suggested the application of deconvolution techniques for this work. Figure 2.2 is adapted from Bezanson et al. (2013), panel e) in Figure 2.1 is adapted from Schawinski et al. (2014), and Figure 4.7 is adapted from Behroozi et al. (2019). All other work in this thesis is my own.

Abstract

This thesis studies the star formation histories (SFHs) of galaxies in order to understand the nature of 'quenching.' The suppression of star formation, *i.e.*, quenching, over the history of the Universe results in a growing population of quiescent galaxies. The physical processes governing how and why galaxies are quenched remain unknown. Observations show that massive galaxies, and those in dense cluster environments, make up a great proportion of the quiescent population. A number of theories have been proposed to explain both the mass dependence and environmental dependence of this population. Through studying the ages and stellar properties of the quiescent galaxies, the predictions of these theories can be tested.

The first section of this thesis concentrates on the differences between galaxies in isolated ('field') or cluster environments as part of the Gemini Observations of Galaxies in Rich Early ENvironments (GOGREEN) survey. Leveraging multi-wavelength and spectroscopic observations of 331 quiescent galaxies at $1 < z < 1.5$, the data is fit to model spectral energy distributions (SEDs) to infer the SFHs and properties of the individual galaxies. In looking at the age trends between galaxies of different stellar masses, it is confirmed that more massive galaxies show evidence of earlier formation times, while lower mass galaxies exhibit more diverse SFHs. This result supports the paradigm of mass-dependent galaxy evolution. The novel result of this work was that any age difference between cluster and field galaxies was subtle; at fixed stellar mass cluster galaxies are < 0.5 Gyr older. Putting this result in the context of two simple quenching models rules out two proposed quenching scenarios: i) environmental quenching post-infall, and ii) a primordial quenched population among cluster galaxies. This is distinctly different from local clusters, for which the majority of the quiescent population is consistent with having been environmentally quenched upon infall. Our results suggest that the quiescent cluster population at $z > 1$ is driven by different physical processes than those at play at $z = 0$.

The second section of this thesis focuses instead on the detailed characterization of a single galaxy, the Ultra Diffuse Galaxy (UDG) Dragonfly 44 (DF44), whose curious set of properties is inconsistent with theoretical models of UDG formation. In fitting broad-band photometry with high signal-to-noise and high-resolution optical spectroscopy with SED models, the detailed stellar properties and SFH of this galaxy are investigated. The precision of the observations required a careful assessment of the SED models, where the conclusion was that DF44 formed between 7.2–12.9 Gyr ago. Regardless of whether DF44 is old or very old, the SFHs imply early formation and rapid quenching. This result in context with its large size, kinematics, stellar population properties, and its environment,

challenges conventional theories of galaxy evolution. The implication is that current theoretical models are missing the true diversity of galaxy formation and evolution.

The third section evaluates the assumptions of the SED fitting procedure. Modelling the SFHs of galaxies is an ill-defined problem, where the results are subject to a number of prior assumptions for what SFHs are more realistic. Modern SED fitting models make a number of these assumptions specific, which provides flexibility in studying diverse samples of galaxies. Without observational constraints to qualify such assumptions, however, the results can simply reflect the assumptions. A mock dataset of quiescent galaxies modelled after the GOGREEN sample is constructed in order to investigate the influence of the SFH prior to the results of that study. A statistical framework is used to infer the distribution of properties among a galaxy population, which in principle can mitigate unphysical assumptions made in SED-fitting. This work highlights the challenges involved in studying star formation timescales of old galaxies, and the nuances of SED-fitting procedures which can lead to spurious results.

Acknowledgements

This thesis is the product of a dedicated effort on behalf of a large number of people to support me and my dreams in a whole host of ways. I am extremely grateful for the shoulders that I stand on to have reached such heights. I found astronomy as an undergraduate and was enamoured with the interesting ways that the world could be understood with (relatively) simple mathematics and observations of light. I also found an incredibly wonderful community of people who share my interest, and who are responsible for a large portion of my education in their own right. Thank you to everyone who has ever acted as a mentor or a friend.

First and foremost I want to thank my PhD supervisor Dr. Michael Balogh for his endless encouragement0, patience, and guidance. I am grateful for all the opportunities you offered for collaboration, leading observations, and freedom in developing my own projects. I want to thank all of my collaborators, co-authors, and anonymous journal referees for the help that was provided in this thesis. My sincere gratitude to Alexa Villaume in particular for her mentorship and collaboration for a large component of this work. I also want to acknowledge the support from the University of Waterloo astrophysics group, and a whole legion of unofficial mentors.

Thank you to all my friends and family that have been incredibly supportive in so many ways. Most importantly, I want to thank my husband, Matthew. You do so much more than you know.

This work was supported by financial support provided through the Ontario graduate scholarship program, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the University of Waterloo. This work was dependent on the computing resources of the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute Canada. The majority of the work presented in this thesis was conducted at the University of Waterloo, which is located on the traditional territories of the Neutral, Anishinaabeg and Haudenosaunee peoples. The observations which formed the basis of this thesis were obtained from a variety of observatories across the globe, which were made possible by international collaborations and communities which support these telescopes.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ΛCDM Λ-Cold dark matter.

FSPS Flexible Stellar Population Synthesis library.

AGB Asymptotic giant branch.

AGN Active galactic nuclei.

bHB Blue horizontal branch.

CCD Charge-coupled device.

CFHT Canada-France-Hawaii Telescope.

CR Credible region.

CSFRD Cosmic SFR density.

CSP Composite stellar population.

dE Dwarf elliptical galaxy.

DF44 Dragonfly 44.

ETG Early-type galaxy.

EW Equivalent width.

FIR Far-infrared, $15 < \lambda/\mu\mathrm{m} < 1000$.

FUV Far-ultraviolet, $0.3 < \lambda/\mu\mathrm{m} < 0.4$.

FWHM Full-width half maximum.

GALEX Galaxy Evolution Explorer, a NASA orbiting space telescope.

GC Globular cluster.

HB Horizontal branch.

HBM Hierarchical Bayesian model.

HR diagram Hertzsprung–Russell diagram.

HST Hubble Space Telescope.

IMF Initial mass function.

IR Infrared.

IRAC Infrared Array Camera on the Spitzer space telescope.

ISM Interstellar medium.

KCWI Keck Cosmic Web Image.

LINER Low-ionization nuclear emission-line region.

LTG Late-type galaxy.

MAP Maximum a-posteriori.

MCMC Markov chain Monte Carlo.

MEH Metal enrichment history.

MIR Mid-infrared, $3 < \lambda/\mu\mathrm{m} < 15$.

MUV Mid-ultraviolet, $0.2 < \lambda/\mu\mathrm{m} < 0.3$.

MW Milky-Way.

MZR Stellar mass to stellar metallicity relation.

NIR Near-infrared, $0.7 < \lambda/\mu\mathrm{m} < 3$.

NUV Near-ultraviolet, $0.122 < \lambda/\mu\mathrm{m} < 0.2$.

PCA Principal component analysis.

RGB Red giant branch.

rHB Red horizontal branch.

S/N Signal-to-noise ratio.

SED Spectral energy distribution.

SFH Star formation history. The evolution of the SFR over time, $\mathrm{SFR}(t)$.

SFMS Star-forming main sequence.

SFR Star formation rate.

SMC Small Magellanic Cloud.

SMF Stellar mass function.

SMHM Stellar mass to halo mass.

SPS Stellar population synthesis.

sSFR Specific star formation rate. The SFR normalized by the total stellar mass, $\mathrm{sSFR} = \mathrm{SFR}(t)/M_* = (\mathrm{d}M_*/\mathrm{d}t)/M_*$.

SSP Simple stellar population.

TP-AGB Thermally pulsing asymptotic giant branch.

UDG Ultra diffuse galaxy.

UV Ultraviolet.

Chapter 1

Introduction

Galaxies are luminous tracers of the properties of our Universe, with galaxy evolution being tied to the evolving cosmological picture. Quantitative studies of galaxy demographics provide an empirical foundation for theoretical models, where the quality of a cosmological model is assessed by its ability to reproduce the basic demographics of observed galaxy populations. Our understanding of galaxy formation and evolution, and their usefulness as cosmological probes, however, is incomplete.

1.1 The current understanding of galaxy evolution

A galaxy is a collection of gas, dust, stars, and dark matter held together by gravity. Models of the large-scale structure of the Universe predict the formation and evolution of baryons and dark matter. Hierarchical cold dark matter models postulate that galaxies form through dissipative processes within the gravitational potential wells of dark matter halos (White & Rees, 1978; Fall & Efstathiou, 1980). Today's 'cosmic web' began as small fluctuations in the density field of the early Universe. As the Universe expanded and cooled, baryons condensed via gravity at the nodes and centres of dark matter halos, forming the first galaxies and galaxy clusters (Rees & Ostriker, 1977). Dark matter halos assemble hierarchically; small haloes merge to form larger structures. The assembly of *galaxies*, however, is not hierarchical. The growth of baryonic matter involves complex physical processes, acting on a variety of scales. Our understanding of galaxy evolution is challenged by the complexity of the processes controlling the cooling, heating, and feedback of baryonic material within galaxies. While the current cosmological paradigm (*e.g.*, Wechsler & Tinker, 2018) is neatly concordant with observations in its broad predictions,

the detailed predictions for galaxy evolution remain challenged in explaining the diverse properties of observed galaxies. This thesis investigates the process of galaxy evolution by studying the star formation histories of galaxies.

In collecting a census of galaxies as a function of stellar mass and halo mass, the relative mass distributions are distinct (*e.g.*, Bower et al., 2006). The observed evolution of the halo mass distribution out to $z \sim 10$ (*e.g.*, Springel, 2005; Driver et al., 2022) is consistent with predictions from hierarchical assembly, as demonstrated by a number of simulations (*e.g.*, Jenkins et al., 2001; Reed et al., 2003; Springel, 2005; Reed et al., 2007; Watson et al., 2013). In contrast, the stellar mass distribution of galaxies reveals that massive galaxies were already established at high-redshifts ($z > 4$), while the number of lower mass galaxies increases over time (*e.g.*, Pozzetti et al., 2003; Fontana et al., 2004; Drory et al., 2005; Bundy et al., 2005; Muzzin et al., 2013c; Behroozi et al., 2013). This apparent 'anti-hierarchical' nature of galaxy assembly implies that galaxy evolution is decoupled from that of dark matter. Moreover, if the gas within galaxies were to simply collapse under gravity, there would be a runaway cooling effect leading to efficient star formation until the supply of gas runs out. Given the distribution of gas in the Universe, the star forming epoch would be short, and all stars would now be extremely old (*i.e.*, the 'cooling crisis;' Cole 1991, White & Frenk 1991). The fact that galaxies continue to form stars at low redshifts implies that there are important internal physical processes regulating star formation.

Astronomers have long been collecting demographic data for galaxies in an effort to understand the physical mechanisms controlling their evolution. One of the first developed classification schemes for galaxies was based on their shapes. Hubble (1926) used a 'tuning fork' diagram to explain the various morphologies of galaxies as a sequence of 'early-type' (*i.e.*, elliptical) to 'late-type' (*i.e.*, spiral). Multi-wavelength observations later revealed the colours of galaxies, which are related to the properties of the stellar populations. Massive stars are luminous and short-lived, where their spectral energy distributions (SEDs) are dominated by ultraviolet (UV) emission such that they are 'blue.' Lower mass stars are comparatively faint, have longer lifetimes, and are 'red.' The integrated light of galaxies which have continued star formation is dominated by blue stars. Colour-magnitude diagrams (CMDs) are therefore a simple observational diagnostic where colour is a proxy for the luminosity-weighted mean stellar age (ignoring variable dust content), or on long timescales, the average specific star formation rate (sSFR). The magnitude is a proxy for the luminous matter content, *i.e.*, the stellar mass.

The nomenclature relating early-type galaxies (ETGs) with old stars (late-type, in the

sense of look back time), and vice-versa, is unfortunate.[1] It's not the case that late-type galaxies are devoid of old stellar populations, rather, the central cores of spiral galaxies are typically red. Instead, it's that ETGs are generally devoid of young stars. The suppression of star formation is typically called 'quenching.'[2] Since the 'early-type' designation is a derivative of morphological classifications, galaxies selected by colour (*e.g.*, along the 'red sequence') are more often labelled 'passive' or quiescent.

Many early studies established a number of empirical relations that described the strong correlation between the majority of global galaxy properties: *e.g.*, luminosity, colour, mass, star formation rate (SFR), size, morphology, mean stellar age (Freeman, 1970; Faber & Jackson, 1976; Tully & Fisher, 1977; Dressler et al., 1987). Since the advent of large-scale observational surveys in the 1980s (*e.g.*, Geller & Huchra, 1983; Folkes et al., 1999), multi-wavelength data has been collected for large numbers of galaxies over a wide range of redshifts. One of the first insights was that the bimodality of galaxy colours, and the relative fractions of red and blue galaxies, correlated with stellar mass. Since the early 2000s, surveys like the Sloan Digital Sky Survey (SDSS; York et al., 2000) helped to consolidate these fundamental relations, particularly for massive galaxies. The scaling relations were understood to be a one-dimensional sequence in stellar mass (*e.g.*, Kauffmann et al., 2003; Tremonti et al., 2004; Blanton et al., 2005; Gallazzi et al., 2005; Baldry et al., 2006), or perhaps more fundamentally, in terms of stellar surface density or central velocity dispersion (*e.g.*, Bell & de Jong, 2001; Bernardi et al., 2005; Kauffmann et al., 2006; Graves et al., 2009; Wake et al., 2012).

The bimodal distribution of galaxy properties persists in this picture. At fixed stellar mass, ETGs are generally redder, older, and lack ongoing star formation, compared to late-type galaxies (*e.g.*, Strateva et al., 2001; Blanton et al., 2003; Bell et al., 2004). The two different classes of galaxies therefore obey different sets of relations. Their evolutionary histories also appear distinct, where ETGs are, on average, more massive and more numerous in denser environments (*e.g.*, Dressler, 1980; Kauffmann et al., 2003; Blanton et al., 2005; Baldry et al., 2006).

The evolution of the colours (*e.g.*, Bell et al., 2003; Balogh et al., 2004) and stellar mass distributions of red/blue galaxies shows that the population continues to grow over time (*e.g.*, Bell et al., 2004; Faber et al., 2007). While the majority of galaxies are star forming at $z > 2.5$, a population of quiescent galaxies exists at $z \sim 4$, and later dominates the

[1] As noted in Tinsley (1980). An additional point of confusion is the selection of ETGs; colours and morphology and star formation rate proxies (*e.g.*, emission lines) are all used within the literature, but they are not necessarily equivalent. Not all ETGs are on the red sequence, not all red sequence galaxies have elliptical/early-type morphologies, and not all ETGs are absent of star formation.

[2] And we continue the tradition of using imperfect nomenclature.

massive end of the stellar mass function at $z = 0$ (*e.g.*, Ilbert et al., 2013; Muzzin et al., 2013c; Davidzon et al., 2017). The implication is that galaxies transition from star forming to quiescent, *i.e.*, 'quench,' and that this transition is related to stellar mass. Furthermore, galaxy clusters host a higher fraction of ETGs at lower redshifts (*e.g.*, Dressler, 1980; Butcher & Oemler, 1984) and lower star formation rates (SFRs; *e.g.*, Balogh et al. 1997; Kauffmann et al. 2004), relative to the field, suggestive of environmentally driven quenching processes (which may be independent of stellar mass; *e.g.*, Peng et al. 2010).

Based on the evolving star forming mass sequence and simple prescriptions for environmental-quenching and mass-independent self-quenching, Peng et al. (2010) developed a simple model to explain the $z = 0$ stellar mass distributions of star forming and quiescent galaxies. This 'grow-and-quench' paradigm explains the observed mass-dependent evolution, where the stellar populations of more massive galaxies are older, although their haloes continue to assemble (*e.g.*, Graves et al., 2009). However, the star forming main sequence is not necessarily a scaling law for individual galaxies, but the culmination of a diverse set of star formation histories (SFHs; *e.g.*, Abramson et al., 2016). This highlights a limitation of 'cross-sectional' studies of distinct galaxies compared over different epochs. In establishing a sequence, one also has to be careful about selection effects. For example, the population of quiescent galaxies is growing over time as star forming galaxies of lower stellar masses are increasingly undergoing quenching ('progenitor bias;' van Dokkum & Franx 1996). And given the hierarchical nature of galaxy assembly, galaxies at late times are an aggregate of many distinct stellar populations; there isn't a single 'progenitor' to connect to the 'descendant' galaxy at low redshifts.

When star formation quenched in galaxies was a prominent topic of the last several decades. Surveys like the SDSS were revolutionary in this venture, as the calibrated spectral continua provided a means to measure the properties of stellar populations with better precision, and the instrument design allows for the efficient observation of large numbers of galaxies. Through studying spectral features which reveal the galaxy's past SFHs for local quiescent galaxies (*e.g.*, Heavens et al., 2000; Panter et al., 2007; Thomas et al., 2010; Citro et al., 2016), the major epochs of star formation could be determined. Put another way, in contrast to cross-sectional methods, there is a 'fossil record' of different stellar populations encoded in the spectra of individual galaxies, revealing their integrated (*e.g.*, because of mergers) evolution over cosmic time. The star formation and quenching timescales can then be determined from the SFH. The amount of information needed to fully synthesize all the properties presents an observational challenge, which limits the sample sizes of surveys. Moreover, the interpretation of the properties from the integrated observations from stellar population synthesis (SPS) models relies on several careful considerations (see Section 2.3).

Spectral line-based studies showed that the SFHs of quiescent galaxies correlate with stellar mass (Thomas et al., 2005; Graves et al., 2009; Thomas et al., 2010). While mass-dependent galaxy evolution had been inferred by studies comparing galaxy scaling relations over time, this was confirmed with better statistical precision (Cowie et al., 1996; Brinchmann et al., 2004; Kodama et al., 2004). More recently, this has been complemented with observations at higher redshifts (*e.g.*, Drory et al., 2005; Moresco et al., 2010; Onodera et al., 2012, 2015; Jørgensen & Chiboucas, 2013; Choi et al., 2014; Gallazzi et al., 2014; Pacifici et al., 2016a; Carnall et al., 2019a; Leja et al., 2019b; Webb et al., 2020; Tacchella et al., 2022a), which provide better age precision given that the galaxies are systematically younger. Broadly, these studies conclude that massive quiescent galaxies ceased star formation early, $z > 3$, while less massive galaxies have more extended SFHs.

How star formation quenches remains an open question. Galaxy quenching involves mechanisms that disrupt the gas processes fuelling star formation, either by heating the gas or removing it. It's not necessarily that there is a dearth of gas in the galaxies, as rest-frame UV spectroscopy of diffuse gas and metals for a diverse range of galaxy types find gas present in galaxies out to $z \sim 2$–3 (*e.g.*, Rudie et al., 2012; Tumlinson et al., 2013; Peeples et al., 2014). A number of quenching mechanisms have been proposed, which can broadly be categorized as internal or external. Since the fraction of quenched central galaxies ('central' in the sense that they dominate their halo, as opposed to satellite galaxies) scales with stellar mass, internal (or 'self') quenching processes may be distinct at different mass scales (*e.g.*, stellar winds versus AGN feedback).

Galaxies in denser environments are likely influenced by their surroundings (*e.g.*, the inflow of gas being shut-off or gas being removed by tidal interactions or ram pressure). Through the comparison of galaxies in low- and high-density environments, the environmental-related quenching modes can be investigated. The rapid evolution of the low-mass quiescent population in clusters suggests that environmental quenching may be particularly important for less massive galaxies (*e.g.*, Nantais et al., 2016; Kawinwanichakij et al., 2017).

Quenching mechanisms can be delineated by their timescales; different physical processes act over timescales ranging from < 1 Myr to over a Hubble time (*e.g.*, Iyer et al., 2020). The population of massive and compact quiescent galaxies at high redshift suggests that quenching can happen on rapid timescales (*e.g.*, AGN-feedback; Barro et al. 2013). On the other hand, the number of local green valley galaxies (*i.e.*, galaxies which have colours between red and blue in CMDs; Schawinski et al., 2014) and the metallicity differences between local quiescent and star forming galaxies (*e.g.*, Peng et al., 2015), requires longer-timescale processes to explain (*e.g.*, gas exhaustion). Observations of post-starburst galaxies suggest that there are at least two dominant quenching timescales: rapid (< 200 Myr) and slow (*e.g.*, Belli et al., 2019; Wild et al., 2020). Through studying the

variability of SFR over different timescales, the cumulative effects of the distinct physical mechanisms controlling star formation can be quantified, and compared to predictions from simulations. While hydrodynamical simulations suggest that the distribution of quenching timescales is bimodal, and with different distributions between cluster and field galaxies, this cannot be confirmed until observations of slow quenchers can be studied in detail.

The dominant quenching process may differ with cosmic time. Therefore, while it is already appreciated that answering the question *'how do galaxies quench?'* requires a description of the dependencies on physical properties such as stellar mass, and how effective the processes are in different environments, observations at different redshifts can provide useful insights.

1.2 What this thesis addresses

This thesis chronicles several projects which attempt to understand galaxy evolution through their SFHs, as inferred from integrated observations (*i.e.*, individual stars are unresolved) of distant galaxies. The main topics involved are:

1. What suppresses (*i.e.*, 'quenches') star formation in galaxies?
2. How do quenching mechanisms differ at higher redshifts?
3. How is quenching related to stellar mass or local environment?

Each of the projects outlined in this work relies on inferring galaxy properties and SFHs from observations through fitting model spectral energy distributions (SEDs). Chapter 2 briefly introduces the history of understanding galaxy evolution with observations and provides an in-depth introduction to SED-fitting procedures.

Chapter 2 first provides a historical foundation for understanding the diverse properties of quiescent galaxies from observations. The SED-fitting framework used in Chapters 3–5 to infer the stellar properties from different sets of observables is then outlined.

Chapter 3 presents the first work of my Ph.D. (Webb et al., 2020) which measured the SFHs of quiescent galaxies observed from the GOGREEN survey ($1 < z < 1.5$; Balogh et al. 2020), and compared the SFHs of galaxies in cluster and field environments. The relative ages of the two populations are a constraint on timescales for different environmental quenching processes at $z > 1$.

Chapter 4 presents the second work of my Ph.D. (Webb et al., 2022), which studied in detail the SFH of the Ultra-Diffuse galaxy (UDG) Dragonfly44 (DF44). Leveraging very high signal-to-noise (S/N) rest-frame optical spectroscopy and broadband photometric coverage, timescales for DF44's formation and rapid quenching were constrained. Putting the SFH in the context of other anomalous properties, DF44 appears to be an outlier among canonical dwarf galaxies and is yet unexplained by current theories of galaxy evolution.

Chapter 5 then investigates the limitations of the SED-fitting framework with respect to modelling the diversity of properties among populations of galaxies with low-resolution or noisy observations. This provides guidance for interpreting results in the current literature based on different types of data or different SED models, and for how to design future surveys.

Chapter 6 then summarizes these projects and discusses ideas for future work. The published works presented in this thesis have been adapted to remove redundancy in describing the SED-fitting technique, and otherwise use consistent terminology.

Chapter 2

Inferring galaxy properties from observations

Galaxies are observable by the energy they emit; most commonly we observe galaxies by their photons.[1] The wavelength-dependent flux density is called the spectral energy distribution (SED). Stars have SEDs characteristic to their properties, such as age and chemical composition. Observations of the integrated light of a galaxy are the sum of the SEDs of its stellar populations, as well as the contribution of dust and gas.

The star formation history (SFH) of a galaxy describes the star formation rate (SFR) as a function of time, where the integrated value corresponds to the total stellar mass formed by the galaxy. The integrated stellar light within a galaxy is thus related to the distribution of birth stellar masses (*i.e.*, the initial mass function; IMF), the number of stars formed and when (the cumulative SFH), the properties of those stars, and mass lost during stellar evolution. If galaxies are in closed environments, successive stellar populations are formed of materials enriched by earlier populations, such that there is an overall increase in metallicity over time (*i.e.*, metal enrichment history; MEH). Otherwise, the inflow of pristine gas onto a galaxy, or the addition of materials from a merger with another galaxy, complicates this picture. Galaxies may also merge with each other, where the mass accretion history of a galaxy traces the merger history of its parent galaxies. Galaxies with complex mass assembly histories (*i.e.*, many mergers and/or interactions) can therefore host diverse stellar populations. The SFH tracks the overall SFR of the resulting stellar population.

[1] Although techniques for studying galaxies via gravitational waves or neutrinos do exist.

The stellar flux is obscured, scattered, and re-emitted by interstellar dust. Dust particles absorb high-frequency (short wavelength ultraviolet; UV) photons, and re-emit the energy at lower frequencies (*i.e.*, infra-red; IR). The SEDs of typical galaxies are dominated by dust emission beyond $\lambda \sim 10$ μm (*e.g.*, da Cunha et al., 2008).Young stars are very luminous in the UV, such that modelling the dust is particularly important for star forming galaxies. That said, characterizing the effect of dust-heating in older stellar populations is also important (*e.g.*, Utomo et al., 2014). For example, asymptotic giant branch (AGB) stars experience rapid mass-loss, leading to their light being obscured at optical wavelengths and re-radiated in the IR (Kelson & Holden, 2010).[2] The viewing angle can also affect the amount of dust along the line-of-sight of a galaxy; a spiral galaxy viewed edge-on will have a large dust component than if viewed face-on. Flexible dust models are necessary to describe the different attenuation and emission spectrums of a diverse set of galaxies, and simultaneously constraining dust and stellar properties, therefore, requires UV and IR observations.

The inference of galaxy properties through the comparison of model SEDs to observations largely depends on how the different properties affect the SEDs over the observed wavelength range. Typically, galaxy SEDs are only coarsely sampled by observations; the integrated flux over a set of photometric filters spanning a wide range in wavelength (photometry) or over a smaller wavelength range at higher resolution (spectroscopy). Photometric data are revealing of physical processes which have large-scale effects on the SED continuum. Due to the obscuring effects of the atmosphere, ground-based observations are limited to select wavelengths (*e.g.*, optical, radio). Until space-based telescopes were available, surveys were limited in their ability to distinguish parameters that are degenerate over optical wavelengths (*e.g.*, stellar metallicity, dust attenuation, and stellar age; see review by Conroy 2013).

Historically, empirically calibrated monochromatic or single luminosity-based scaling relations were used to interpret the integrated SED with only a few photometric data points (*e.g.*, Kennicutt, 1989; Kennicutt, 1998; Bell & de Jong, 2001). These scaling relations were often calibrated to a particular class of galaxy, limiting direct comparisons of diverse galaxy populations. Modern SED model comparison techniques instead provide a mapping between a larger suite of data and galaxy properties, allowing for better constraints. Through direct comparison of observations with physically motivated models of stellar populations, however, galaxy properties can be inferred uniformly.

The SFHs of quiescent galaxies are of particular interest as they trace the properties of

[2]Although the timescale that AGB stars dominate the IR flux varies between SPS models, *e.g.*, Maraston 2005 predicts the Thermally-pulsating AGB stars redden SEDs at a younger age than Bruzual & Charlot (2003).

the early Universe. While young stars are a product of the continually enriched ISM, old stellar populations are 'frozen-in' tracers of the early properties of the galaxy. Quiescent galaxies therefore provide a 'fossil record' of galaxy evolution, where 'fossil' evidence refers to the observable properties related to a galaxy's integrated past SFH (as opposed to active star formation). The diversity among quiescent galaxies is indicative of varying star formation processes over time. Through studying this diversity, the predictions of our understanding of galaxy evolution can be tested.

2.1 Interpreting galaxy properties from scaling relations

The properties of quiescent galaxies obey tight multidimensional correlations between their spectral and structural properties. One-dimensional relations include colour-magnitude (Faber, 1973; Sandage & Visvanathan, 1978; Bower et al., 1992), colour–line strengths (Faber, 1973), luminosity and velocity dispersion (*i.e.*, the Faber–Jackson relation; Faber & Jackson, 1976), galaxy mass-to-light ratio (M/L) with luminosity (Tinsley, 1981; Faber et al., 1987), and different combinations of velocity dispersion and luminosity with galaxy effective radius and effective surface brightness (Kormendy, 1985), or with galaxy core radius and central surface brightness (Lauer, 1985). Some of these relations are projections of two-dimensional correlations, namely the Fundamental Plane (Djorgovski & Davis, 1987; Dressler et al., 1987). With the development of large-scale galaxy surveys in the last few decades, nuances in the scaling relations have revealed diversity among the ages, metallicities, and abundances of the old stellar populations. Many studies have been devoted to connecting the features of the scaling relations to galaxy evolution models in order to understand the different physical processes causing this diversity.

The 'red sequence' characterizes the relationship between galaxy colour and velocity dispersion, σ, or luminosity, L (related to the dynamical mass of the galaxy, $M_{\rm dyn}$ as quiescent galaxies obey a tight mass–luminosity ratio, M/L). The two trends together culminate in a colour-magnitude trend, whereby colours redden with increasing σ (or L or M). Figure 2.1 demonstrates the connection between the evolution of stellar colours and SSP colours, where galaxies become redder as they age. Panel e) shows the observed colour–mass trend from SDSS, with sSFRs from *GALEX* (indicated by the colour map) from Schawinski et al. (2014), where galaxies are separated by morphology. ETG colours (and sSFRs) are correlated with stellar mass. The dispersion in the red sequence reveals age/metallicity trends, where the two parameters are degenerate (Faber, 1973; Worthey, 1994).

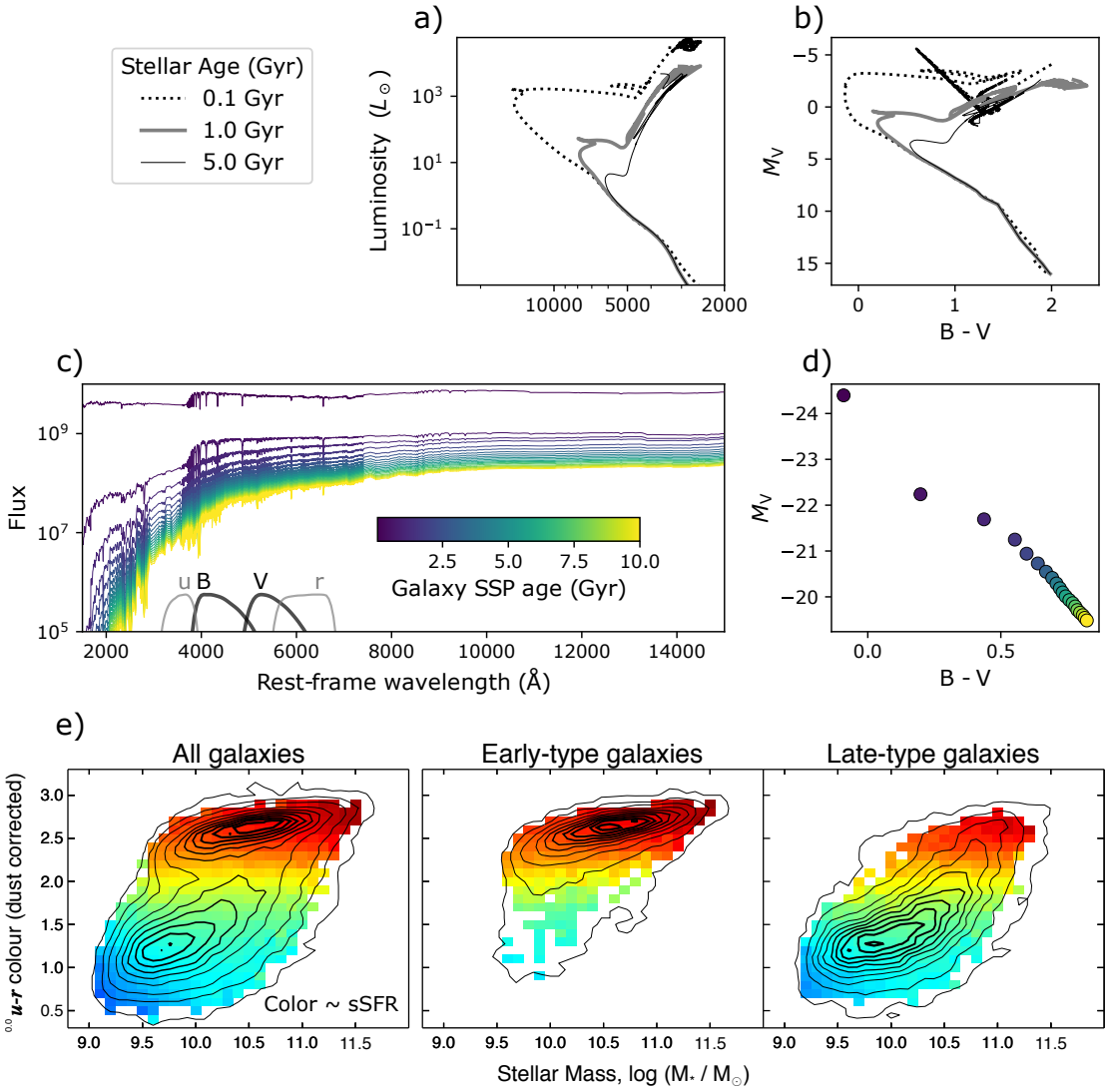

Figure 2.1: a) Stellar isochrones for coeval populations of a given age, and b) the corresponding V-band magnitude and $B-V$ colours. The colours of stars evolve over time. Isochrones from Marigo et al. (2017). c) SEDs for SSPs of varying age, and d) the corresponding V-band magnitude and $B-V$ colours as calculated by FSPSand Prospector. Galaxies, being populations of stars, also evolve in colour space over time. e) The $u-r$ colours and stellar masses as measured by SDSS and GALEX, for all galaxies and for ETG or LTG populations individually. The correlation between morphology and galaxy colour is apparent, where ETGs primarily follow a 'red sequence'. Colour is indicative of the measured sSFR. Figure adapted from Schawinski et al. (2014).

The tight correlation between the size, surface brightness, and velocity dispersion of ETGs is described as a Fundamental Plane. If ETGs are homologous, *i.e.*,with self-similar mass distributions, and have constant M/L ratios, then the virial theorem predicts a particular form for the Fundamental Plane. The finite thickness and tilt of the plane are not fully consistent with virial relations, however, suggesting a breakdown of these assumptions (Graves & Faber, 2010). Figure 2.2 summarizes the expected relationship between these three parameters and projections of the correlations, as well as the observed Fundamental Plane as a function of redshift as compiled by Bezanson et al. (2013). The degree of scatter perpendicular to the plane is sensitive to the dispersion in age and formation time, stellar metallicity, or IMF.

While interpreting galaxy ages directly from scaling relations is complicated by degeneracies (*e.g.*, Lee, 2007), the *evolution* of the scaling relations reveal the relative change in the ages of stellar populations (see review by Renzini, 2006). For example, the evolution of the size, morphology, and age of ETGs results in a vertical shift, a tilt, and an increase in the perpendicular scatter of the Fundamental Plane (see Figure 2.2). Changing other galaxy properties (*i.e.*, metallicity) affects the plane in orthogonal ways. The redshift evolution of the different scaling relations suggests that massive ellipticals are predominantly old, having formed at high redshift ($z > 3$) and been passively evolving ever since (Aragon-Salamanca et al., 1993; Bender et al., 1996; van Dokkum & Franx, 1996; Ziegler et al., 1999; Saglia et al., 2000).

2.2 Interpreting galaxy properties from spectral features

A complementary approach to studying the variance in ages among quiescent galaxies is by way of a set of specific spectral features (Lick/IDS indices) calibrated to stellar population synthesis (SPS) models. While age, metallicity, and abundances were degenerate in the scaling relations, the Lick indices were chosen such that they could be independently constrained (at least for elliptical galaxies; Burstein et al., 1984; Worthey, 1994).

Theoretical models of stellar evolution predict the signatures of different stellar types at a given age and chemical composition, which are combined in different ways to replicate observations (*e.g.*, Tinsley, 1968; Tinsley & Gunn, 1976; Bruzual, 1983). Modelling stellar properties is a complicated process, and building models of the integrated SEDs of galaxies relies on characterizing the properties and relative contributions of many stellar types,

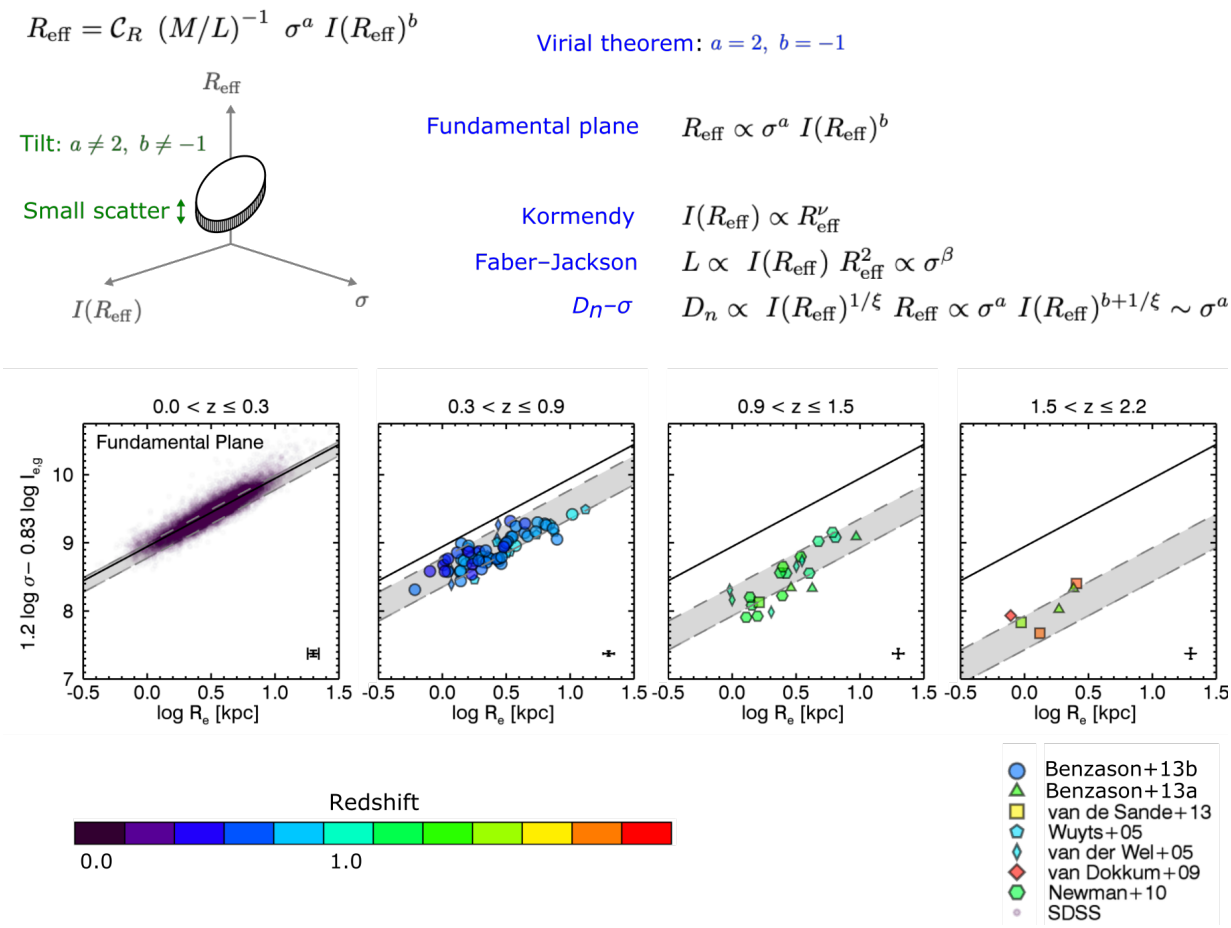

Figure 2.2: Elliptical galaxy scaling relations between the size (effective radius; $R_{\rm eff}$), surface brightness (within $R_{\rm eff}$; $I(R_{\rm eff})$), and velocity dispersion (σ). If elliptical galaxies are homologous, $\mathcal{C}_R$ is a constant. If elliptical galaxies follow a constant mass-to-light ratio, M/L is also constant. If elliptical galaxies are virialized structures, the power-law coefficients are $a = 2$ and $b = -1$. Deviations from these values, *i.e.*, that the Fundamental Plane is tilted and has small scatter, suggests a breakdown of these assumptions. Panels show the observed projection of the Fundamental Plane at several redshifts, as compiled by Bezanson et al. (2013) for massive elliptical galaxies. The evolution of the Fundamental Plane is expected given the evolution of galaxy sizes, morphologies, and stellar populations.

thus requiring complete stellar libraries. Empirical stellar libraries[3] can be incomplete (*e.g.*, Peletier, 1989; Thomas et al., 2003; Schiavon, 2007), as they rely on high-quality observations of stars across a large parameter space, presenting an observational challenge. Theoretical libraries, on the other hand, may not accurately characterize all stellar populations (*e.g.*, treatment of blue horizontal branch (HB) stars, AGB stars, stellar rotation and multiplicity). Moreover, different SPS models made different assumptions (*e.g.*, different assumptions of Thermally-pulsating AGB (TP-AGB) stars between Maraston 2005 and Bruzual & Charlot 2003 propagate to differences in predicted galaxy ages).

Galaxies, being collections of stars which have formed over many epochs, require a model for the star formation history. Early SFH models typically assumed simple stellar populations (SSPs), which describe the evolution of the SED of a single, coeval stellar population with uniform metallicity and abundance. SSPs assume a model for stellar evolution (in the form of stellar isochrones), stellar spectral libraries, and an IMF. Composite stellar populations (CSPs) are combinations of SSPs of different ages (as defined by their SFH) and metallicities, and contain dust. While the SFHs can be arbitrarily complex, simple forms are typically adopted for computational convenience. An exponentially declining SFR is a popular functional form used to describe local quiescent galaxies, which follows the predicted SFH from closed-box models (SFR evolves linearly with gas density; Schmidt 1959).

Galaxies are typically discussed in terms of the average age of their stellar content, where the age is a summary statistic of their star formation history. Ages reported in the literature vary between 'SSP-equivalent,' luminosity-weighted, or mass-weighted timescales. Since young stars are more luminous, weighting by luminosity or mass provides distinct metrics unless the M/L ratio is unity. Mass-weighted age is more difficult to measure, given that low mass galaxies which dominate the stellar mass fraction of the galaxy are intrinsically fainter, but is a better metric for the time of bulk star formation. SSP-equivalent or luminosity-weighted ages of galaxies with complex star formation histories (such as late bursts of star formation, where the old stars are outshone by small fractions of young/intermediate-age stellar populations) can therefore be very different from the mass-weighted age (*e.g.*, Trager et al., 2000a; Serra & Trager, 2007).

Until the importance of the distinction between different age metrics was apparent, and ETGs were recognized to have low levels of ongoing star formation (particularly for less luminous galaxies; *e.g.*, Schiavon, 2007; Smith et al., 2009; Trager & Somerville, 2009), there was a discord among results based on spectral indices or scaling relations (*e.g.*, Jørgensen,

[3]Spectral libraries convert the predictions of stellar evolution codes (*e.g.*, surface gravity and effective temperature) to observables, as a function of stellar metallicity.

1999; Poggianti et al., 2001). A second issue was that although Lick indices were selected to be independent constraints on different model parameters, the uncertainties among the indices are not strictly independent. The perverse circulation of observational errors between different indices required high S/N (> 100 Å^{-1}) spectra to overcome (Trager et al., 2000b); galaxy spectra were often averaged (*i.e.*, 'stacked') to meet this requirement.

In the late 2000s, improved SPS models could predict the full optical to near infrared (NIR) spectrum, providing a wealth of diagnostic features. Instead of relying on a small set of specific spectral features, the full spectra could be modelled and compared to observations. 'Full-spectrum' fitting could be used to capture complex SFHs, metal, and abundance patterns without as high of an S/N requirement (> 50 Å^{-1}; Ocvirk et al. 2006a, Tojeiro et al. 2007). As computational resources and statistical techniques have improved, it became feasible to increase the complexity of SED models further, and account for subtleties in the SFHs.

General concordance was eventually reached between studies based on scaling relations and spectral indices, once SED models began to incorporate more complex SFHs for ETGs (Renzini, 2006). By selecting galaxies in different regions of the scaling relations, and interpreting the average spectral features with SPS-calibrated Lick indices, the trends in age and metallicity or abundance could be separated. The different sequences in stellar population properties across the scaling relations suggest a diversity in the duration of star formation, at fixed velocity dispersion (σ). Different studies attribute between 25–40% of the colour dispersion in the red sequence to a diversity of galaxy ages (Trager et al., 2000c; Nelan et al., 2005; Gallazzi et al., 2006; Thomas et al., 2005; Graves et al., 2007; Gargiulo et al., 2009). Moreover, the high abundance ratios observed among elliptical galaxies (*e.g.*, Worthey et al., 1992) are inconsistent with extended star formation histories (*e.g.*, Thomas et al., 1999), which constrains the relative fraction of young stellar populations among these galaxies to be less than $\sim$10%.

In the next section, the features of modern SED models and fitting procedures are outlined. The capabilities of the Prospector code used in this work are discussed in detail.

2.3 Inferring galaxy properties via SED fitting

SED-fitting is the method of creating synthetic SEDs (SPS models) from the summed spectra of a population of stars, with weights according to the expected number distribution of each spectral type and spectral property (*e.g.*, stellar metallicity, chemical abundance).

The SPS models are compared to a set of observations, where a statistical model is used to assess a proposed weighting scheme. In comparison to 'recipes' which relate individual features between models and observations (*e.g.*, combinations of Lick indices), comparing the full SED (photometry and spectral continuum) consolidates all available information on the physical attributes of the galaxy related to stars, dust, and gas.

Out of computational necessity, many of the first SED-fitting routines simplified many of the components via parameterization. For example, SFHs were assumed to follow simple functional forms, fixed or discrete dust attenuation curves were assumed, and stellar metallicities were fixed or had limited variability (*e.g.*, Bolzonella et al., 2000; Heavens et al., 2000; Brammer et al., 2008; Kriek et al., 2009). Observed galaxies, however, are diverse; one set of assumptions does not represent all galaxies (*e.g.*, Tremonti et al., 2004; Reddy et al., 2015; Salmon et al., 2016). Importantly, because galaxy properties are correlated, restricting the parameter space of these variables can lead to strong biases in the SED fitting results (*e.g.*, Conroy, 2013; Mitchell et al., 2014). One of the most impactful assumptions is the form of the SFH, which influence both stellar mass and age estimates. Simplistic SSP or exponentially-declining 'τ' models fail to describe realistic SFHs (*e.g.*, Simha et al., 2014), underestimate ages of galaxies with continued star formation (*e.g.*, Trager et al., 2000c; Trager & Somerville, 2009; Lee et al., 2009, 2010; Maraston et al., 2010), and fail to reproduce the evolution of the mass function of galaxies (Wuyts et al., 2011). Another simplification common to SED-fitting routines is the use of χ^2-minimization metrics for model comparison, which cannot properly estimate the uncertainties, particularly for degenerate parameters (see, *e.g.*, Bundy et al. 2005, Taylor et al. 2011).

The challenge for highly-flexible SED models, on the other hand, is that they are subject to model misspecification. Even if the SED model itself can describe the true parameters, without high-quality observations, the truth cannot necessarily be distinguished from a number of other combinations of parameters which reasonably reproduce the data. This is sometimes phrased as SED-fitting being a 'poorly constrained' problem; a small observational uncertainty results in highly uncertain SFHs (Ocvirk et al., 2006a). In principle, this can be 'fixed' by specifying the model to prefer 'more-reasonable' solutions over others – which is in essence what simple-SFH models are doing. The specification of reasonable solutions is made explicit in models which adopt Bayesian statistical frameworks, which assume that each parameter follows a probability distribution defined by the prior knowledge. Moreover, the power of using flexible SED models with known 'priors' is that the posterior predictions have more realistic uncertainties (Leja et al., 2019a).

A large number of codes have been developed in the last two decades in an effort to produce more-flexible and more-realistic SEDs, and moreover include improved statistical routines to model the uncertainties: *e.g.*, MOPED, Heavens et al. 2000; STECMAP, Ocvirk

et al. 2006a; VESPA, Tojeiro et al. 2007; MAGPHYS, da Cunha et al. 2008; Zhang et al. 2012; Kelson et al. 2014; Dressler et al. 2016; AGNfitter, Calistro Rivera et al. (2016); BEAGLE, Chevallard & Charlot (2016); Pipe3D, Sánchez et al. (2016); Prospector, Leja et al. 2017, Johnson et al. 2021; Dense Basis, Iyer & Gawiser (2017), Iyer et al. (2019); Lightning, Eufrasio (2017); FIREFLY, Wilkinson et al. 2017; BAGPIPES, Carnall et al. 2018; MrMoose, Drouart & Falkendal 2018; FortesFit, Rosario 2019; PEGASE.3, Fioc & Rocca-Volmerange 2019; X-CIGALE, Yang et al. 2020; MCSED, Bowman et al. 2020; piXedfit, Abdurro'uf et al. 2021; ProSpect, Thorne et al. 2021; Starduster, Qiu & Kang 2022. The main issue at present is that the specification of the models is handled differently among these codes, and that the results can be strongly influenced by the choice of both models and priors (*e.g.*, Conroy, 2013; Leja et al., 2019a; Carnall et al., 2019b; Lower et al., 2020). Both broadband wavelength coverage and high S/N and high-resolution spectral data are needed to accurately specify such models. Not all parameters can be constrained even with high S/N data, however. For example, the choice of SFH model can have more influence on the inferred galaxy age than the S/N, as the earliest timescales of star formation have only subtle effects on the integrated SED (Leja et al., 2019a).

There are additional systematics between results from different SED codes related to their adopted SPS models. For example, Tojeiro et al. (2011) note that the SFRs over short timescales are particularly sensitive to the SPS models' treatment of AGB stars and convection processes. Many of the SED-fitting codes also rely on models with fixed solar abundance patterns; only stellar metallicity is allowed to vary. Pacifici et al. (2022) compared the results for a high-redshift sample of galaxies between 14 different SED-fitting codes, concluding that while stellar masses have relatively small systematic offsets, the SFRs and dust parameters are much more sensitive to the model specification, particularly where MIR constraints were unavailable.

This section describes the SED models and fitting routines used in this thesis. This introduction aims to provide a broad overview of the different model features used within the literature, which contextualizes the specific modelling choices used in Chapters 3 and 4. In Chapter 5, the limitations of SED-fitting techniques are discussed in detail, in particular given the choice of SFH model. More detailed reviews of SED-fitting techniques are provided in Walcher et al. (2011), Acquaviva et al. (2011), Gawiser (2009) and Conroy (2013). Section 2.3.1 describes the SPS models assumed in this work, and Section 2.3.2 introduces how galaxy properties are inferred from matching observations to such models. The various model components included in the SPS model are outlined in Sections 2.3.3–2.3.6, and additional components that are included to describe systematics in the observations in Sections 2.3.7–2.3.8. Lastly, the SFH models used in this work are discussed in detail

Table 2.1: Example SED model parameters. Notes: 1) Parameters of the SFH model, see Section 2.3.9. 2) Total stellar mass is the integral of the SFH, which includes the mass lost to outflows. To convert to stellar mass *remaining* at the time of observation, the spectral templates are regenerated in FSPS to calculate the mass loss. 3) The total stellar metallicity, fixed throughout the SFH. 4) Parameters of the dust attenuation model, see Section 2.3.4. 5) Parameters of the dust emission model, see Section 2.3.5. 6) Section 2.3.7. 7) Section 2.3.8.

Note	Parameter	Description
	$z_{\rm obs}$	Observed redshift
1	x	Parameters of the SFH model
2	$\log\left(M_*/\mathrm{M}_\odot\right)$	Total stellar mass formed
3	$\log\left(Z_*/\mathrm{Z}_\odot\right)$	Stellar metallicity
4	Dust attenuation	
	$\hat{\tau}_{\rm dust,\ diffuse}$	Diffuse dust optical depth
	$\hat{\tau}_{\rm dust,\ birth}/\hat{\tau}_{\rm dust,\ diffuse}$	Ratio of diffuse to birth-cloud dust optical depth
	$n_{\rm dust}$	Diffuse dust attenuation index
5	Dust emission	
	$Q_{\rm PAH}$	Per cent mass fraction of PAHs in dust
	$U_{\rm min,dust}$	Minimum starlight intensity to which the dust mass is exposed
	$\gamma_{\rm dust}$	Mass fraction of dust in high radiation intensity
6	Noise model	
	$j_{\rm spec}$	Multiplicative spectral noise inflation term
	$f_{\rm outlier,\ spec}$	Fraction of spectral pixels considered outliers
7	Spectrophotometric calibration	
	$\{c_i\}_{i=0}^{n}$	Chebyshev polynomial coefficients, degree= n

in Section 2.3.9. Figure 2.3 provides a schematic for the models and parameters included in the SED model. Table 2.1 lists several parameters that are typically adopted in SED models, and which are relevant to Chapters 3–5 of this thesis.

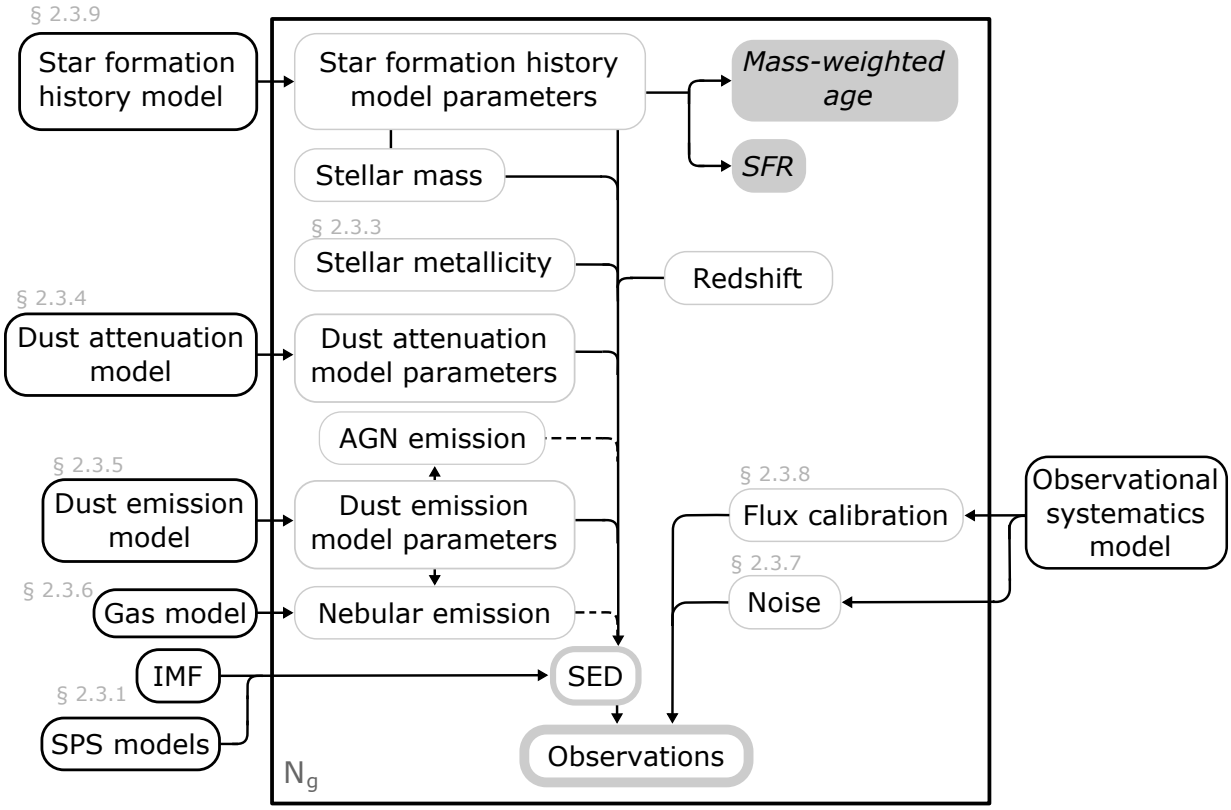

Figure 2.3: A graphical schematic for the models and parameters included in the SED model. The black rectangle, or 'plate,' denotes the set of parameters specified for a given galaxy, indexed by g. Black outlined boxes outside the plate refer to models which are assumed for every galaxy (*e.g.*, a fixed IMF and SPS models) which are connected to the relevant model parameters inside the plate. Boxes with thin grey outlines are explicit parameters of the SED model, where arrows denote how they are included in the SED model. Grey boxes denote latent parameters of the model. Dashed lines connect parameters which can be included in the SED model, but are not considered in this work.

2.3.1 SPS models

This thesis uses the code Prospector (v1.0.0 Leja et al., 2017; Johnson et al., 2020, 2021)[4] to build the physical SED models and perform a fully Bayesian inference of the stellar population properties. Prospector is a framework built on the Flexible Stellar Population Synthesis library (FSPS; Conroy et al., 2009; Conroy & Gunn, 2010a). FSPS is a highly-flexible model including a variety of metallicities, isochrones, and IMFs, and the assumptions about the morphology of the horizontal branch, the blue straggler population, the post-AGB phase, and the location in the HR diagram of the TP-AGB phase can be varied. The IMF and isochrones are not varied in this work. The MESA Isochrones and Stellar Tracks (MIST) evolutionary tracks and isochrones (Choi et al. 2016; Dotter 2016, based on the MESA stellar evolution code; Paxton et al. 2011, 2013, 2015, 2018), and MILES[5] spectral templates (Sánchez-Blázquez et al., 2006; Vazdekis et al., 2015a) are assumed. Notably, the MIST models include stellar rotation, which has the net effect of producing hotter, brighter, and longer-lived massive stars Choi et al. (2016), although the uncertainties related to different stellar rotation models are likely weaker than those for stellar binaries (Eldridge et al., 2017). A Chabrier (2003) IMF is assumed. Several simple prescriptions for dust attenuation are included in FSPS, where the emission from diffuse dust and AGN tori are calculated self (see Sections 2.3.4 and 2.3.5). Prospector wraps the FSPS code into a convenient framework which allows for efficient 'on the fly' model generation and sampling.

As noted in the previous section, a limitation of SPS models is the flexibility of the set of models in describing realistic stellar populations. The scope of SPS models is defined by the number of model components. Stellar libraries of SEDs are typically parameterized by stellar surface gravity, effective temperature, metallicity, α-abundances, IMF, etc. For the quiescent galaxies studied in this work, there are several uncertain components which can affect the age estimates. The unknown contribution of TP-AGB stars results in an uncertainty in the NIR emission of galaxies, particularly for galaxies of intermediate age ($\sim$2–3 Gyr; Maraston, 2005; Marigo et al., 2008; Conroy & Gunn, 2010b; Kriek et al., 2010; Capozzi et al., 2016). The UV-flux in old stellar populations is affected primarily by post-AGB and blue horizontal branch (HB) stars, and the flux at $\lambda < 1$ μm is sensitive to stellar rotation and binarity (Conroy, 2013, and references therin). In principle, FSPS allows the normalization of the TP-AGB stars or the fraction of blue HB stars to be varied, although it is currently not computationally feasible to marginalize over either within Prospector. As the NIR is particularly sensitive to both stellar masses and ages, this can result in large

[4]https://github.com/bd-j/prospector
[5]http://miles.iac.es/

systematic uncertainties in these parameters (Bruzual, 2007).

Different SPS models employ different prescriptions for stellar evolution, leading to systematic differences between SED-fitting results. Many studies have investigated the different aspects of this problem (*e.g.*, Pacifici et al., 2022). The impact of the choice of SPS models on the results is not quantified in this thesis. In Chapter 3, the relative differences in galaxy properties are investigated, with the expectation that any uncertainty in the absolute values cancels out. In Chapter 4 an unusually old (> 7 Gyr) object with a very-high S/N spectrum is studied, where several discrepancies between the models and the observations are discussed.

2.3.2 SED fitting

SPS models predict the SEDs of galaxies. The goal of SED-fitting procedures is to use these models to infer the properties of galaxies from their observed SEDs. Section 2 discusses a set of spectral indices calibrated to SPS models that were used for this purpose, which can be considered a special case of SED fitting (*e.g.*, Graves & Schiavon, 2008). The downside to this approach, however, is that not all the information from the observations is used. There is a wealth of information in galaxy spectra among the smaller spectral features. A second issue is that the measurement of Lick-indices can be complicated by small spectral features contaminating the sidebands in some galaxies more than others. An alternative approach is to describe a galaxy spectrum by a small set of continuous parameters which uniquely produce a given spectrum.

One way to do this is to decompose a spectrum into orthogonal components (*i.e.*, eigenvectors), which can be linearly combined. Principal Component Analysis (PCA) techniques do this by diagonalizing the covariance metric of the data points to identify the directions of largest variation, which reduces the problem of SED fitting to solving a linear equation (*e.g.*, Connolly et al., 1999; Wild et al., 2007, 2009). The main challenge of PCA is interpreting the physical properties from the empirically-determined principal components.

Alternatively, the full-spectrum can be fit, whereby a statistical framework infers the galaxy properties by inversion. As mentioned above, the SED of a galaxy can be represented as a sum of SEDs of SSPs, with weights according to the SFH (modulo dust and gas), which is a linear problem that can be solved by matrix inversion. That is, the observed SED is inverted onto a basis of independent components (SSPs, dust, etc.) drawn from the SED model.

First attempts at SED fitting were mainly template based, where model SEDs were generated based on SSPs with various properties such as age, dust extinction, stellar metallicity, and mass. To reduce the complex physics of SEDs, nearly all inversion methods make simplifying assumptions so that the SED can be described as a linear function of its parameters (*e.g.*, all stars have the same metallicity, which is also fixed over time, fixed dust attenuation). This approach follows a 'frequentist' logic, which attempts to quantify the likelihood of a galaxy's properties falling within a cell on the grid of parameters. A simple χ^2 metric was used to assess the goodness-of-fit of each model SED with the observations,

$$\chi^2 = \sum_{i=0}^{n} \left(\frac{f_i - \sum_{k=1}^{M} w_k \ \mathrm{SSP}_i(t_k, Z, \tau)}{\sigma_i} \right)^2, \tag{2.1}$$

where f_i is the observed spectrum in each of n wavelength bins i, and σ_i is the corresponding uncertainty. The weights, w_k, correspond to how each SSP of age t_k (with fixed metallicity Z and dust τ) contributes to the SFH. Typically, inversion codes search for the fewest number of SSPs that can describe the observations. Galaxies SFHs are perhaps better described as extended epochs of continued star formation, however. A common way to address this issue is to 'regularize' the linear combination of SSPs (*e.g.*, Cappellari & Emsellem, 2004; Ocvirk et al., 2006b). One of the downsides to traditional minimization techniques is that the uncertainties of the 'best-fit' model can be difficult to physically interpret. Simple bootstrap algorithms are not able to identify all sets of models which could realistically reproduce the observations. Regularization helps in the sense that it omits the 'unphysical' less-smooth solutions. However, the uncertainties may still be unrealistic (Walcher et al., 2011).

Linear-inversion techniques are more challenging when taking into account the non-linear effects of dust attenuation and emission, gas (*i.e.*, line) emission, AGN, etc., which are important when the observations span a broad range of wavelengths or resolutions. Bayesian-fitting methods are more flexible in that SEDs can be computed over a large parameter space, which are compared to the observed SEDs with some likelihood criterion. One of the advantages to the Bayesian approach is that the 'priors' that define the set of preferred solutions can be explicitly defined (*i.e.*, the size and shape of the parameter space).

Bayesian inference methodology

A 'Bayesian' approach poses the question: given the set of observations for a given galaxy, and the assumptions about the properties of said galaxy, what is the probability (or posterior probability distribution; PPD) for the properties of the galaxy? The advantage of this

Bayesian approach is that the prior knowledge about galaxy information can be directly applied in assessing the PPDs of galaxy properties (*i.e.*, the so-called 'priors' of the SED model). Importantly, within this framework the assumed *priors* are made explicit.

The physical model used to generate model SEDs is a function of the parameters $\boldsymbol{\theta}$, some of which are fixed (the SPS models, α-abundance, etc.) and some which are inferred from the observations. The prior information about the expected form of the model, gained from decades of galaxy studies, can be incorporated as a hypothesis, $\boldsymbol{\mathcal{H}}$. The 'prior probability distributions' are assigned for the parameters under this hypothesis, $p(\boldsymbol{\theta}|\boldsymbol{\mathcal{H}})$.

Following Bayes' equation, the probability of a set of properties, $\boldsymbol{\theta}_g$, in describing a galaxy g with observations, $\boldsymbol{\mathcal{O}}_g$, can be written,

$$p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}}) = \frac{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}})\ p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{H}})}{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})}, \tag{2.2}$$

where $p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}})$ describes the probability of the observations conditioned on $\boldsymbol{\theta}_g$ and $\boldsymbol{\mathcal{H}}$, which is called the 'likelihood.' The denominator describes the probability of the observations given the hypothesis, $p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})$, and is also called the 'evidence' or marginal likelihood. In practice, it can be safely assumed that the observations are independent of the choice of priors, such the evidence is a normalizing constant.

Where the observational errors are assumed to be Gaussian, the likelihood function is often expressed using the chi-squared metric, $p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}) \propto \exp(-\chi^2/2)$:

$$\ln(\boldsymbol{\mathcal{O}}_g|\boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}) = -\frac{1}{2}\sum_{i=1}^{N}\left[\ln\left(2\pi\boldsymbol{\sigma}_i^2\right) + \frac{\left(\boldsymbol{f}_{i,\ g} - \boldsymbol{f}_i^{\boldsymbol{\mathcal{H}}}(\boldsymbol{\theta}_g)\right)^2}{\boldsymbol{\sigma}_i^2}\right], \tag{2.3}$$

where at a given wavelength each observed flux, $\boldsymbol{f}_{i,\ g}$, (with an associated uncertainty, $\boldsymbol{\sigma}_i$) is compared to the model flux, $\boldsymbol{f}_i^{\boldsymbol{\mathcal{H}}}(\boldsymbol{\theta}_g)$. Where the χ^2 is high, there is a good probability that the model with parameters $\boldsymbol{\theta}_g$ is a good representation of the observations. In fitting both the photometry and spectroscopy, the log-likelihood of the model conditioned on the observations is taken to be the sum of the two individual likelihood functions:

$$\ln \mathcal{L}(\boldsymbol{\mathcal{O}}_g|\boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}) = \ln \mathcal{L}(\boldsymbol{\mathcal{O}}_{g,\ \mathrm{spec}}|\boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}) + \ln \mathcal{L}(\boldsymbol{\mathcal{O}}_{g,\ \mathrm{phot}}|\boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}), \tag{2.4}$$

where no relative weighting is used.

One of the challenges of SED fitting is that SED models are subject to misspecification; while a given model can produce a good match to the observations, its parameters may not accurately represent the properties of the object being studied. As an example, a

simple SFH which assumes that all the stars within a galaxy formed in a single burst (*i.e.*, an SSP) can produce a model SED matching the observed photometry over a narrow wavelength range (*e.g.*, optical colours). While an SSP model may accurately model *some* basic properties of the galaxy (*e.g.*, redshift), it can be strongly biased by recent star formation (young stars will dominate the integrated luminosity of the galaxy) such that the inferred ages of the stellar population may be underestimated. The limited functional form of the model (describing the SFH as a single epoch of star formation) fails to represent the complex, multi-episodic nature of star formation expected for real galaxies.

The priors are critical to this probability model as they define which solutions are possible, and most probable. Each parameter of the SED model (*e.g.*, stellar metallicity, the attenuation of the dust normalization curve, the parameters of the SFH model, etc.) is given an associated prior probability distribution.

Informative priors are useful when information exists as to what the solution to this probability equation should be (for example, based on results from independent experiments). Otherwise, it is preferable to use uninformative priors such that the probability equation will be solely informed by the likelihood function. Uninformative priors can be challenging to implement in practice, however, as many of the model parameters may be related. Recent studies have shown that the choice of SFH model (which is a prior in itself) and the priors assigned to the SFH model parameters, can significantly affect inferred galaxy properties.

2.3.3 Stellar metallicity

Galaxies are expected to have complex metal enrichment histories related to their merger and star formation histories. Modelling individual metal components is challenging, however, and requires high-S/N and high-resolution spectroscopy (Conroy et al., 2013). It is common to instead assume a fixed metallicity for the entire history of a galaxy. Notably, the FSPS models assume a fixed metallicity and scaled-solar abundance. This assumption is most certainly unphysical, as older stars would have formed out of more pristine gas compared to those at later epochs. Recently Bellstedt et al. (2020, 2021) demonstrated that the cosmic star formation history was better reproduced with SFHs derived with SED models which accounted for closed-box metal enrichment, rather than fixed metallicities. In Chapter 3 the data quality of observations is perhaps too poor to constrain the chemical composition of the galaxies, however, this is not the case for the observations analysed in Chapter 4 – a point which is discussed in detail in that chapter.

2.3.4 Dust attenuation model

FSPS includes several dust models, and each project in this work selects a model appropriate for the galaxies being studied. In this section, the fiducial two-component dust attenuation model from Charlot & Fall (2000) is introduced. The Charlot & Fall (2000) model separates the dust components associated with the birth-cloud and a uniform dust screen.[6] The birth-cloud dust only attenuates stellar emission for stars younger than 10 Myr, following the relation:

$$\tau_{\text{dust, birth}}(\lambda) = \hat{\tau}_{\text{dust, birth}} \left(\frac{\lambda}{5500 \text{ Å}} \right)^{-1}. \tag{2.5}$$

The uniform dust screen mimics diffuse-dust, and has a variable attenuation curve (Noll et al., 2009):

$$\tau_{\text{dust, diffuse}}(\lambda) = \frac{\hat{\tau}_{\text{dust, diffuse}}}{4.05} \left(k'(\lambda) + D(\lambda) \right) \left(\frac{\lambda}{5500 \text{ Å}} \right)^{n}, \tag{2.6}$$

where n is the diffuse dust attenuation index, $k'(\lambda)$ is the attenuation curve from Calzetti et al. (2000), and $D(\lambda)$ describes the UV bump based on Kriek & Conroy (2013). In principle, the parameters of this flexible model can be set to reproduce other various dust models in the literature. For example, the Calzetti et al. 2000 model corresponds to the Charlot & Fall (2000) model where $\hat{\tau}_{\text{dust, birth}} = 0$, $n = 0$, and $D(\lambda) = 0$.

The free parameters of the dust attenuation model are the diffuse dust normalization constant, $\hat{\tau}_{\text{dust, diffuse}}$, the ratio of the diffuse-to-birth-cloud dust, $\hat{\tau}_{\text{dust, birth}}/\hat{\tau}_{\text{dust, diffuse}}$, and the dust attenuation index, n. Using the ratio of dust constants helps to avoid degeneracies between the two dust parameters.

Salim & Narayanan (2020) recently reviewed a number of dust models and found that incorrect assumptions of the dust attenuation law propagated large uncertainties in the derived SFRs. The choice of dust model is only briefly explored in this work. In Chapter 3, a Milky-Way like extinction curve (Cardelli et al., 1989) was assumed in fitting a population of quiescent galaxies at $1 < z < 1.5$ with NUV–NIR photometry and rest-frame optical spectroscopy, where Appendix A.2 comments on this assumption in the context of the modelling results. In Chapters 4 and 5, however, the Charlot & Fall (2000) dust model is assumed, where each parameter was a free variable of the model. The results were then marginalized over all possible values of the dust parameters.

[6] Note that the diffuse dust optical depth is related to the dust extinction via $A_\lambda = 2.5 \ \ln(e) \ \tau_\lambda$.

2.3.5 Dust emission model

Dust emission is calculated by assuming energy balance, whereby all the energy attenuated by dust is re-emitted at IR wavelengths (da Cunha et al., 2008). Under this assumption, constraints on dust emission are also constraints on dust attenuation. For the IR luminosity to be inferred from the UV–MIR SED, the shape of the SED needs to be known. Prospector uses the templates from Draine et al. (2007), which include a silicate-graphite poly-aromatic hydrocarbon (PAH) model of interstellar dust (Mathis et al., 1977; Draine & Lee, 1984). The templates are parameterized by three parameters describing the shape of the SED. The shape and location of the thermal dust emission bump in the IR SED are described by $U_{\rm min}$ and γ_e, respectively. The fraction of the total dust mass attributed to PAHs is set by $q_{\rm PAH}$.

The inclusion of MIR photometry will constrain dust emission. In Chapter 3 the GOGREEN data typically covers the NUV–NIR, where only a handful of galaxies include MIR bands. For this reason, no dust emission model was included in the analysis, reducing the number of parameters of the SED model. Chapter 5 explores how the inclusion of MIR constrains the SED model parameters. Dust emission associated with AGN torus, and circumstellar dust around AGB stars are both models that can be included in FSPS. In practice, since AGN emission is strongest in the far-infrared (FIR), which is redder than any of the data included in this work, it is not incorporated in any of the SED models discussed in this work. Similarly, the contribution of dust around AGB stars is also not explored.

2.3.6 Nebular emission model

The standard approach to generating nebular emission in FSPS is to assume that the ionizing continuum from the model stellar population is fully absorbed by the gas and emitted as both line and continuum emission. The nebular emission can be easily generated according to a grid of pre-computed solutions (*e.g.*, from the program `cloudy`; Ferland et al. 2013). Nebular emission is tied to the gas properties through the parameters $Z_{\rm gas}$ and U of the dust emission models.

This work focuses on galaxies which are not actively forming stars, and hence without strong nebular emission. That said, [O II] emission is commonly found even within quiescent galaxies. Rather than fit the [O II] emission, in all cases in this work the spectral region about [O II] (∼3727 Å) is masked.

2.3.7 Spectroscopic Noise and outlier models

A noise model can be used to account for possible under- or over-estimates of the spectral uncertainties, where the noise is uniformly inflated (or deflated). Effectively, this modifies the spectral uncertainty by a multiplicative factor but is counterbalanced by a penalty in the likelihood calculation for larger uncertainties. This down-weights spectral points where the uncertainties are otherwise low, but there is a mismatch between the spectrum and the models. The mismatch may be because of a 'bad' spectral data point, or that the model itself cannot reproduce the feature.

To avoid being overly sensitive to outlier pixels in the spectrum, a mixture model is used to identify and mask pixels in the spectra which have large deviations from the model. This is again relevant where the spectroscopic uncertainties are smaller, and significant residuals can result from poor matches to the models where the model itself is inaccurate (due to differences in, for example, α-enhancement). Prospector uses the mixture model approach described in Hogg et al. (2010a).

The likelihood of the spectroscopic observations is modified (see Equation 2.4) to incorporate both the noise and outlier models,

$$\mathcal{L}(\boldsymbol{\mathcal{O}}_{g,\ \mathrm{spec}}|\boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}) = (1 - f_{\mathrm{out}})\ \mathcal{L}(\boldsymbol{\mathcal{O}}_{g,\ \mathrm{spec}}|\boldsymbol{\theta}_g, \sigma, \boldsymbol{\mathcal{H}}) + f_{\mathrm{out}}\ \mathcal{L}(\boldsymbol{\mathcal{O}}_{g,\ \mathrm{spec}}|\boldsymbol{\theta}_g, \sigma_{\mathrm{out}}, \boldsymbol{\mathcal{H}}), \quad (2.7)$$

where σ_{out} is multiplied by a large factor, *e.g.*, $\sigma_{\mathrm{out}} = j_{\mathrm{spec}} = 50\sigma$.

2.3.8 Spectro-photometric calibration model

Several parameters are included in the model to account for systematics in the spectra related to imprecise flux calibration. At each likelihood call, a Chebyshev polynomial of degree n is fit (via optimization) to the residual between the normalized model and the spectrum, which the spectrum is multiplied by before calculating the likelihood. The degree of the polynomial is chosen to allow the polynomial to be flexible enough to describe the spectral continuum without over-fitting, removing real features such as the break at $\sim$4000 Å ($\mathrm{D}_n(4000)$; Balogh et al., 1999). The effect of the calibration model is to remove the information from the spectroscopic continuum such that it does not constrain any of the model parameters. Instead, the broad shape of the continuum is set by the photometry.

2.3.9 Stellar mass and star formation history models

The SFH model describes the rate of star formation over time, SFR(t), where the integral is the total stellar mass formed,

$$M_{*,\ \mathrm{total}} = \int_0^{t_{\mathrm{univ}}(z_{\mathrm{obs}})} \mathrm{SFR}(t)\ \ \mathrm{d}t. \tag{2.8}$$

The remnant stellar mass at the time of observation, M_*, accounts for mass loss during the galaxy's evolution,

$$M_* = M_{*,\ \mathrm{total}}\,(1 - R)\,, \tag{2.9}$$

where R is the mass return fraction. FSPSfollows the stellar mass loss prescription from Renzini & Ciotti (1993). While $\log\left(M_{*,\ \mathrm{total}}/\mathrm{M}_\odot\right)$ is a free parameter in the Prospector $\log\left(M_*/\mathrm{M}_\odot\right)$ is more commonly discussed in the literature.

The historical approach to SED fitting has been to adopt simple functional forms to parameterize the SFH, where the restricted parameter space allows for fast computation of large numbers of models. The simplest assumption is that all star formation was concurrent, *i.e.*, the galaxy is described as a simple stellar population (SSP; *e.g.*, Bruzual & Charlot, 2003). In principle, this can be extrapolated by increasing the number of SSPs. Composite stellar populations describe linear combinations of SSPs, where the complexity of the model increases as more distinct stellar populations are included. 'Nonparametric' functions are a class of SFH models which parameterize CSPs without assuming a functional form for the SFR(t). 'Parametric' functions are those which simplify the dimensionality of the problem by parameterizing the SFR(t) such that it can be described by fewer parameters. This section introduces several SFH models used in this thesis. This is not a comprehensive list; both Carnall et al. (2019b) and Leja et al. (2019a) provide reviews on common parametric and nonparametric modelling approaches, respectively.

Parametric τ models

A model describing the SFH as a declining exponential is common within the literature, as it appeared well-suited to massive elliptical galaxies which formed the bulk of their stellar populations early, and ceased forming stars shortly thereafter. Historically, ETGs had been assumed to be well-represented by SSPs (*e.g.*, in modelling their Lick indices). In the early 2000s, however, several detailed spectroscopic (*e.g.*, Trager et al., 2000c) or near-UV (*e.g.*, Ferreras & Silk, 2000; Kaviraj et al., 2007) studies detected a population of hot stars indicative of low-level ongoing star formation. Declining exponential 'τ' models

were therefore proposed as a better description of the SFHs of these galaxies. The model follows the form:

$$\mathrm{SFR}(t) \propto \begin{cases} \exp\left(-\frac{t-t_0}{\tau}\right) & t > t_0 \\ 0 & t < t_0 \end{cases}, \tag{2.10}$$

where the star formation increases from zero to a maximum value at time t_0, after which it declines with the e-folding timescale τ. A notable limitation is that this model cannot describe rising SFRs, or SFHs with multiple star forming epochs. As a consequence, the model is ill-suited to describing star forming galaxies and galaxies at higher redshifts (*e.g.*, Maraston et al., 2010; Reddy et al., 2012). The permitted range of variable values (*i.e.*, the prior probability distributions) have been shown to be strongly influential on the SFH parameter estimates (Wuyts et al., 2011; Pforr et al., 2012), where poorly-calibrated priors produce biased stellar masses, mass-weighted ages, and SFRs when fit to mock observations based on simulated galaxies (Simha et al., 2014; Pacifici et al., 2015; Carnall et al., 2018).

An extension of the τ model is a delayed declining-exponential (delayed-τ) model of the form:

$$\mathrm{SFR}(t) \propto \begin{cases} (t-t_0)\exp\left(-\frac{t-t_0}{\tau}\right) & t > t_0 \\ 0 & t < t_0 \end{cases}, \tag{2.11}$$

The extra term removes the discontinuity in the SFH at t_0 and allows for rising SFRs, and is generally preferred over the stricter τ-model (*e.g.*, Lee et al., 2010). Despite the improved flexibility compared to the τ-model, issues pertaining to how the specification of the model strongly influences the parameter estimates remain (*e.g.*, Lower et al., 2020). More flexible models with similar numbers of parameters are available in the literature: *e.g.*, log-normal (Gladders et al., 2013; Simha et al., 2014) and double-power-laws (Behroozi et al., 2013; Gladders et al., 2013; Carnall et al., 2018). Carnall et al. (2019b) recently tested whether various parametric models could accurately describe a diverse set of SFHs, finding that in all cases they impose strong priors on SFRs and mass-weighted ages, where the amount of bias depends on how well the functional form can match the true SFH.

Nonparametric models

Nonparametric models typically describe the SFH as the SFR in discrete time bins. Rather than describing the SFR as a functional form, the SFH is parameterized by how the SFR is distributed across the time bins, and how the time bins are fixed (or if the time bins

are themselves free parameters of the model; *e.g.*, Tojeiro et al. 2007). A second class of nonparametric model involves using mock galaxy SFHs from simulations directly (*e.g.*, Pacifici et al., 2012). In this method, complicated SFHs and MEHs can be incorporated, and the only parameters involved are the normalization of the SED and the parameters used to tune the simulation. While the SED-fitting can still be performed in the Bayesian sense, the drawback is that the implicit priors (imposed by the choice of simulation and the variety of mock SEDs) can add complexity to the interpretation of the results. A third class describes SFHs as a smooth function with varying numbers of parameters that describe the percentiles of the total stellar mass using Gaussian Processes (*e.g.*, Iyer & Gawiser, 2017; Iyer et al., 2019).

Due to the problem of model misspecification, historically nonparametric SFHs have typically been adopted when fitting high-S/N, high-resolution, full-spectrum data (*e.g.*, Ocvirk et al., 2006a; Tojeiro et al., 2007) or the combination of low-resolution spectra with broadband photometry (*e.g.*, Kelson et al., 2014; Dressler et al., 2016). This is because classical minimization routines struggle to sample the spectrum of plausible solutions due to the high dimensionality of the models, and the large number of correlated parameters. Over the last few decades, however, SED-fitting codes employing Bayesian statistics paired with advanced sampling algorithms have become available, which produce posterior probability distributions for each parameter with uncertainties reflecting the correlations. More recently, these codes adopt improved sampling algorithms which can recover complex multimodal posterior distributions (*e.g.*, Leja et al., 2017; Carnall et al., 2018; Johnson et al., 2021; Thorne et al., 2021).

This thesis primarily assumes that galaxy SFHs can be described with nonparametric models which parameterize SFHs as piece-wise constant functions of the SFR within N time bins. The time bins of nonparametric models are typically given widths according to the timescales of star formation expected to be resolved in the observations. For example, as massive stars are luminous and short-lived, their presence can typically be inferred on short timescales (a few Myr to 1 Gyr). As the SEDs of lower-mass stars evolve more slowly, the widths of the time bins typically increase as a function of lookback time. Models with fixed time bins assume the time that star formation begins. Within this class of models, two different sets of parameters which characterize the distribution of star formation across the time bins are explored: (1) a 'continuity' prior, which parameterizes the relative SFR between adjacent time bins, penalizing large transitions (*i.e.*, preferring smooth solutions), and (2) a Dirichlet prior, which parameterizes how star formation is dispersed across the time bins. Both of these SFH priors are described in detail in Leja et al. (2019a), where brief summaries are provided below.

Continuity model:

The continuity prior assumes a probability distribution for the logarithmic ratio of SFR between adjacent time bins, penalizing strong transitions. This is akin to regularization schemes adopted by earlier works when linearly combining SSPs (Ocvirk et al., 2006a; Tojeiro et al., 2007). The parameters of the model are the log-ratios of SFR between adjacent time bins, indexed by n: $\log(\mathrm{SFR}_n/\mathrm{SFR}_{n+1})$. A prior is adopted for the log-ratios which defines the expected strength of the transition. The fiducial assumption is that these follow a Student-T distribution with parameters μ, σ describing the mean and standard deviation, and two degrees of freedom (Leja et al., 2019a).

Dirichlet model:

The Dirichlet SFH model assumes a Dirichlet distribution for the concentration of star formation distributed across the time bins, with parameter α_{D}. The parameters include z_n, which are related to the fractional mass formed in each bin, m_n, via:

$$m_n = \frac{z_n \; t_n}{\sum_N \; z_n \; t_n}, \tag{2.12}$$

where t_n is the width of each time bin. Sampling from the Dirichlet prior is simplified by using the method described in Betancourt (2012). Larger values of α_{D} prefer solutions where star formation is more equally distributed between the time bins, while $\alpha_{\mathrm{D}} < 1$ prefers the star formation to occur within a few time bins. While there is no specific preference for any particular time bin in principle, the number and relative widths of the time bins will specify the preferred SFR(t).

Mass-weighted age

The mass-weighted stellar age, t_{mass}, is a summary statistic of the SFH which describes the average formation time of stars in a given galaxy (measured in units of lookback time), calculated from the SFH following:

$$t_{\mathrm{mass}} = \frac{\int_{t_{\mathrm{univ}}(z_{\mathrm{obs}})}^{0} t \;\; \mathrm{SFR}(t) \;\; \mathrm{d}t}{\int_{t_{\mathrm{univ}}(z_{\mathrm{obs}})}^{0} \mathrm{SFR}(t) \;\; \mathrm{d}t}, \tag{2.13}$$

where $t_{\mathrm{univ}}(z_{\mathrm{obs}})$ is the age of the Universe at the time of observation. When the timescale is discussed relative to the time the Universe began (as opposed to the lookback time), it's referred to as the mean-stellar *time* or formation time,

$$t_{\mathrm{form}} = t(z_{\mathrm{form}}) = t_{\mathrm{univ}}(z_{\mathrm{obs}}) - t_{\mathrm{mass}}, \tag{2.14}$$

where $z_{\rm form}$ is redshift corresponding to the formation time.

The mass-weighted age is a *latent* parameter of the parametric and nonparametric models. The prior is therefore not explicitly defined in the SED model; instead, there is an *implicit* prior, which is a function of the SFH model parameters and their priors (and time bins, for the nonparametric models). In the literature, an empirical distribution for the age prior is determined by sampling from the SED-model priors and deriving the corresponding ages. In fact, given that the SFH model parameters are independent of the other SED model parameters, samples can simply be drawn from the SFH model priors. The marginalized implicit prior for the mass-weighted stellar ages are defined by the surface integral,

$$p(\boldsymbol{t}_{\rm mass}|\boldsymbol{\mathcal{H}}) = \int_{\boldsymbol{t}_{\rm mass}=\text{constant}} p(\boldsymbol{\theta}_{\rm SFH}|\boldsymbol{\mathcal{H}}) \ \ {\rm d}S, \tag{2.15}$$

where $\boldsymbol{\theta}_{\rm SFH}$ is used to denote the parameters of the SFH model, and ${\rm d}S$ is the surface element in the SFH parameter space.

Each of the SFH models has different implicit prior distributions for the mass-weighted age. In Figure 2.4 the age-prior distributions are shown, obtained by drawing from three sets of priors for each of the delayed-τ, continuity, and Dirichlet model, where the corresponding ages were calculated following Equation 2.13. The priors are shown for an observed redshift of $z = 0$ and $z = 1.2$, which broadly correspond to the redshifts of galaxies studied in this thesis.

For the delayed-τ model, slight modifications to the prior on τ dramatically change the preference for old and young ages. For the Dirichlet model, the age prior distributions have widths corresponding to the assumed $\alpha_{\rm D}$ value. The 'lumpiness' in the age prior distributions from the nonparametric models is a result of the quantization of age at the centres and edges of the time bins.

The fiducial prior for the continuity model follows from Leja et al. (2019a), who set $\mu = 0$ and $\sigma = 0.3$ based on how well the predicted SFHs match those for simulated galaxies at $z = 0$. The centre of the prior falls around half the age of the Universe, which does not reflect the fact that observations of distant galaxies have higher SFHs than presently (*e.g.*, Madau & Dickinson, 2014). A more permissive version of this prior, with $\sigma = 0.5$, does not penalize strong changes in SFH between adjacent time bins as heavily and thus has a broader age prior distribution. The age prior distribution shown with a dashed line corresponds to a set of priors tuned to match the SFHs of quiescent galaxies at $z = 1.2$ from the Illustris simulation (Torrey et al., 2014; Vogelsberger et al., 2014), relevant to Chapter 5.

As discussed above, the SFH parameters are not equally constrained by the observations. The SFH at >1 Gyr is typically poorly constrained from broadband photometry (*e.g.*, Leja et al., 2017), and the SFH at large lookback times are indistinguishable without very high-S/N spectroscopy (*e.g.*, Ocvirk et al., 2006b; Tojeiro et al., 2007) due to the slowly evolving isochrones of old stellar populations. SFH models which connect the SFR at late timescales to those at early times (*i.e.*, all those introduced above, although to different degrees) impose secondary constraints on the mass-weighted ages. The result is that the age prior distributions shown in Figure 2.4 are incomplete descriptions of the influence of the prior in estimating galaxy ages. The effects of the SFH prior on galaxy ages are discussed in Chapters 3–5, where Chapter 5 delves into this issue in detail.

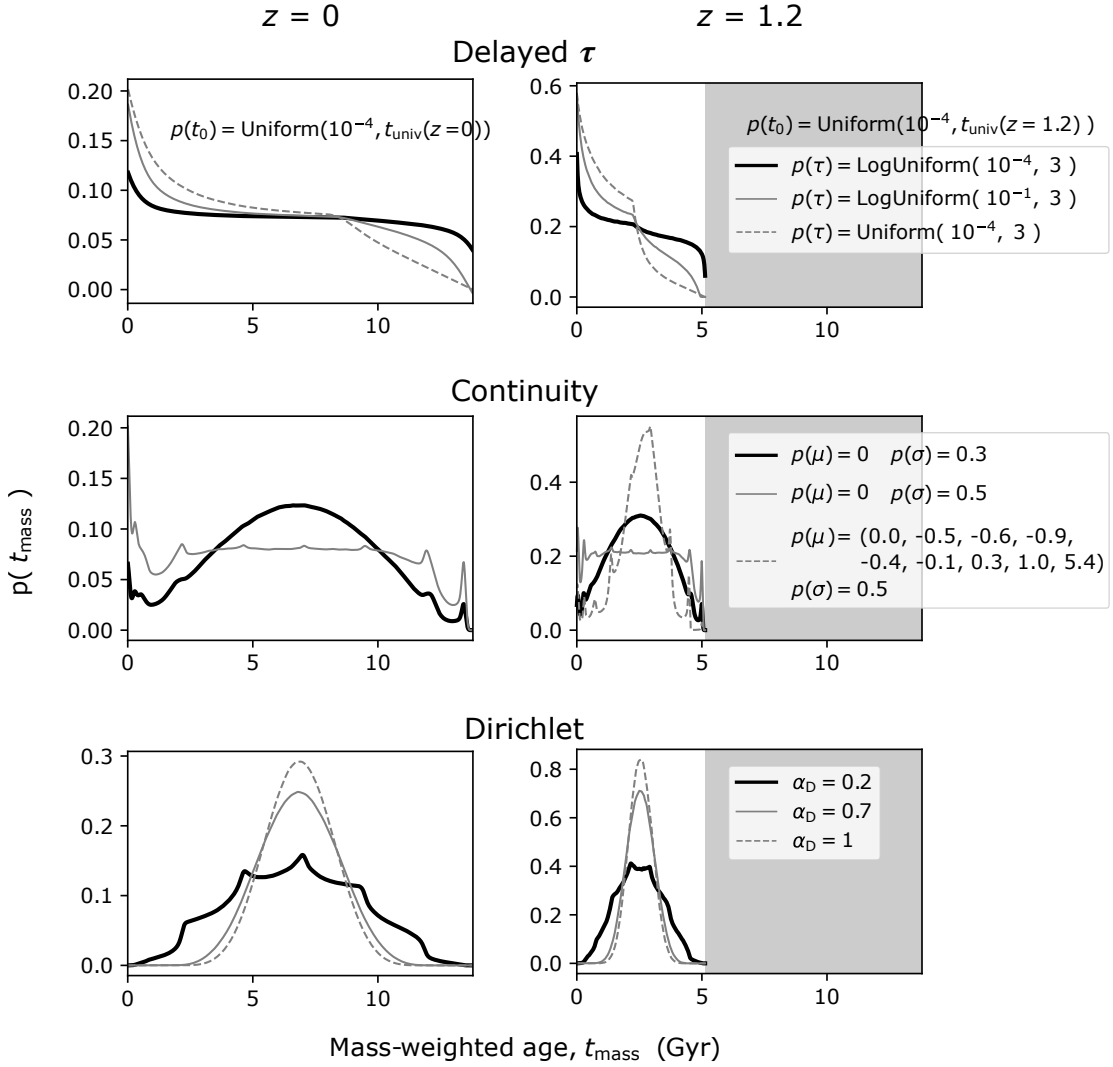

Figure 2.4: Implicit prior probability distributions for mass-weighted age, for three different SFH models, at $z = 0$ and 1.2. Each model is shown with three sets of priors on the parameters, and ages are calculated for $N = 10^6$ draws from each set of SFH model priors. For the nonparametric models, $N = 10$ time bins are assumed: 0–0.03, 0.03–0.1, 0.1–0.5, and 0.5–1 Gyr, $N - 5$ age bins spaced linearly between 1 Gyr–0.95× $t_{\rm univ}(z)$, and a final bin covering $0.95{\times}t_{\rm univ}(z)$–$t_{\rm univ}(z)$. For the Continuity model, the dashed line shows the age prior for the model tuned to the SFHs of the Illustris mock galaxies at $z = 1.2$ (see Chapter 5). The 'lumpiness' of the age prior distributions from the nonparametric models is a result of the 'quantization' of age given the choice of time bins, where the ages tend to fall at the centre of each bin.

Chapter 3

The GOGREEN survey: Post-infall environmental quenching fails to predict the observed age difference between quiescent field and cluster galaxies at $z > 1$

3.1 Introduction

Since $z \sim 2.5$, the galaxy population demonstrates a marked bimodality in star formation rates (SFRs, *e.g.*, Brinchmann et al., 2004; Brammer et al., 2011; Muzzin et al., 2012), and the quiescent component, representing galaxies with negligible current SFRs, has increased steadily (Faber et al., 2007; Muzzin et al., 2013b; Tomczak et al., 2014; Barro et al., 2017). This indicates that the relatively rapid suppression of star formation (quenching) is a fundamental aspect of galaxy evolution, and one that is largely responsible for the steep decline in cosmic SFR density (*e.g.*, Renzini, 2016). The rate of quenching, and indeed galaxy evolution in general, is observed to depend strongly on both stellar mass and environment. In particular, galaxies that are more massive or exist in denser environments are more likely to be quiescent (*e.g.*, Kauffmann et al., 2003, 2004; Brinchmann et al., 2004; Baldry et al., 2006; Weinmann et al., 2006; Kimm et al., 2009).

There have been many studies focused on identifying the main mechanisms that transform galaxies from star forming to quiescent. Simulations which include feedback from

active galactic nuclei (AGN) and feedback from star formation have successfully reproduced the SFR bimodality (*e.g.*, Croton et al., 2006; Bower et al., 2006; Hirschmann et al., 2016), if not quite replicating the observed quenched fractions (*e.g.*, Hirschmann et al., 2016; Xie et al., 2020). However, explaining the dependence of the quenched fraction on local environment appears to require additional processes related to the larger scale environmental densities of galaxies (*e.g.*, Baldry et al., 2006; Peng et al., 2012). Environmental quenching is commonly thought to take place as a galaxy accretes into the halo of a larger structure, either by the removal of its gas reservoir through tidal/ram pressure stripping or by preventing gas in the galaxy halo from accreting and forming new stars, sometimes called strangulation (*e.g.*, Gunn & Gott, 1972; Larson et al., 1980; Balogh et al., 2000). Evidence for the removal of gas can be seen by the lack of line emission from galaxies approaching larger haloes (*e.g.*, Odekon et al., 2016; Jaffé et al., 2016; Zhang et al., 2019). Denser environments could also favour tidal interactions, or harassment, between galaxies (*e.g.*, Merritt, 1983; Moore et al., 1996), which can lead to increased SFRs and accelerated gas consumption (Fujita, 2004). Given that the fraction of quiescent galaxies increases with the number density of surrounding galaxies (*i.e.*, rich galaxy clusters *vs* galaxy groups, *e.g.*, Kauffmann et al. 2004; Wilman et al. 2010; Peng et al. 2010; Darvish et al. 2016, or with distance from cluster cores, *e.g.*, Loh et al. 2008; Woo et al. 2013; Lin et al. 2014; Muzzin et al. 2014; Jian et al. 2017; Guglielmo et al. 2019; Pintos-Castro et al. 2019), the effectiveness of environmental quenching is thought to scale with environmental density.

A simple empirical model of environmental quenching is that, upon infall, the SFR of a galaxy rapidly declines, on an e-folding timescale called the 'fading time.' Motivated in part by the non-zero fraction of star forming galaxies in clusters, this quenching is thought to happen at some time after infall, called the delay time. Wetzel et al. (2013) used a sample of local galaxies in the Sloan Digital Sky Survey (SDSS; York et al., 2000) together with a cosmological N-body simulation in the context of this 'delayed-then-rapid' model, and found that typical delay times at $z = 0$ are 2–4 Gyr, and fading times $<$0.8 Gyr. Galaxy haloes grow hierarchically, however, and this infall-based quenching might happen upon the first infall of a galaxy into a larger halo, which might not be the final cluster halo. So-called 'pre-processing' within galaxy group environments may be an important preceding process (*e.g.*, Zabludoff & Mulchaey, 1998; Fujita, 2004; McGee et al., 2009; De Lucia et al., 2012; Pallero et al., 2019). Observations at higher redshifts have the potential to remove some degeneracies associated with this empirical picture, in part because the evolution in galaxy properties like SFRs and gas fractions is decoupled from the rate of dark matter halo mass growth (*e.g.*, McGee et al., 2014).

One direct way to trace galaxy evolution is to measure the stellar mass function (SMF) as a function of redshift and environment for passive and active galaxies (*e.g.*, Fontana

et al., 2004; Vulcani et al., 2011, 2013; Muzzin et al., 2013c; Tomczak et al., 2014; Nantais et al., 2016; van der Burg et al., 2013, 2020). Similarly, detailed studies of the stellar populations in galaxies compared across redshift epochs can reveal how the overall population of galaxies has evolved (*e.g.*, Poggianti et al., 1999; Trager et al., 2000a; Sánchez-Blázquez et al., 2009). This is only indirectly connected to changes in star formation, like quenching, and does not allow one to easily identify what subset of the population is undergoing changes at a given time. A complementary approach is to measure the star formation histories (SFHs) of individual galaxies and thus reconstruct the growth of populations (*e.g.*, Heavens et al., 2000, 2004; Panter et al., 2003). Comparing the SFHs of galaxies in isolated and dense environments has the potential to provide new information on the effect of environment-specific quenching processes.

Measuring the stellar ages of galaxies as a probe of the SFH is very challenging, however. For all but the nearest galaxies individual stars are not resolved; rather, observations measure the integrated luminosity of the stellar population, and thus it is necessary to disentangle the contribution of stars of various masses and ages. The galaxy spectral energy distributions (SEDs) also suffer from degeneracies between galaxy properties (*e.g.*, stellar age, metallicity, and dust) particularly at low resolution (*e.g.*, Worthey, 1994). Many studies rely on studying select spectral features, observed at high resolution, which are well calibrated against such degeneracies (*e.g.*, Vazdekis, 1999; Trager et al., 2000b) or more recently with full-spectrum fitting (*e.g.*, MacArthur et al., 2009; Sánchez-Blázquez et al., 2011a). Photometry alone cannot distinguish between such model parameters, and age estimates can be strongly influenced by priors (Carnall et al., 2019b; Leja et al., 2019a). The integrated luminosity is also dominated by bright young stars, 'outshining' evidence of older stellar populations (Papovich et al., 2001). Galaxies older than $\sim$5 Gyr have very similar SEDs, making it difficult to precisely estimate the stellar age of quiescent galaxies at low redshifts (*e.g.*, Gallazzi et al., 2005). Moreover, empirical models of stellar evolution are biased by systematic uncertainties and can significantly impact age estimates, particularly for galaxies dominated by intermediate age stars (*e.g.*, Maraston, 2005; Han & Han, 2018). While measuring the properties and SFHs of individual galaxies provides the clearest picture of galaxy evolution, this requires relatively large samples of galaxies with sufficiently high signal-to-noise ratio (S/N) continuum spectroscopy.

The consensus of observations at low to moderate redshifts, despite these challenges, is that there is a trend between the SFHs and stellar mass for quiescent galaxies. The SFRs of massive galaxies peaked at earlier times than lower mass systems (sometimes called 'downsizing;' *e.g.*, Cowie et al., 1996; Brinchmann et al., 2004; Kodama et al., 2004). Correspondingly, massive galaxies form their stellar mass earlier and are therefore older, on average ('archaeological downsizing;' *e.g.*, Nelan et al., 2005; Thomas et al., 2005, 2010;

Treu et al., 2005a,b; Cimatti et al., 2006; Gallazzi et al., 2014; Pacifici et al., 2016a; Carnall et al., 2018; Estrada-Carpenter et al., 2020; Saracco et al., 2020). These trends together are commonly referred to as 'mass dependent evolution.'

For massive galaxies, the majority of their stellar mass is formed within only 1–2 Gyr (Gallazzi et al., 2004, 2005; Glazebrook et al., 2004; Bell et al., 2005; Nelan et al., 2005; Thomas et al., 2005, 2010; Treu et al., 2005a; Toft et al., 2012; McDermid et al., 2015; Citro et al., 2016), and have quenched as early as $z \sim 3$–4 (*e.g.*, Straatman et al., 2014; Glazebrook et al., 2017; Schreiber et al., 2018b; Forrest et al., 2020a). Low-redshift observations of massive quiescent galaxies (typically early-type galaxies, ETGs) find that galaxies in less-dense environments are on average 1–2 Gyr younger than galaxies in massive clusters (*e.g.*, van Dokkum & Stanford, 2003; Thomas et al., 2005; Renzini, 2006, and ref's therein). Notably, age differences at low redshifts could be enhanced by environmental effects which come into play only at late times, such as 'rejuvenation' (Thomas et al., 2010) or 'frosting' (Trager et al., 2000b) of star formation via galaxy mergers or interactions – which occur more frequently in lower mass haloes (*e.g.*, Cooper et al., 2010). Paulino-Afonso et al. (2020) show that SF can be enhanced for low-to-moderate mass galaxies even at moderate, 'filament-like' overdensities. For these reasons, higher redshift observations can provide better leverage on the differences in galaxy properties related to how they evolved in different environments.

At $z \sim 1$, the average ages of ETGs in low-density environments are within 0.5 Gyr of comparable galaxies in galaxy clusters (*e.g.*, Gobat et al., 2008; Thomas et al., 2010; Rettura et al., 2010, 2011; Raichoor et al., 2011; Saracco et al., 2017; Woodrum et al., 2017). The lack of environmental influence at this epoch is supported by Fundamental Plane (FP) studies of ETGs which show that the mass-to-light (M/L) ratios evolve similarly for galaxies in field and cluster environments (*e.g.*, di Serego Alighieri et al., 2006a,b; van Dokkum & van der Marel, 2007; Saglia et al., 2010; Woodrum et al., 2017). Studies of SFRs between star forming galaxies in cluster and field environments show mixed results, either predicting little (*e.g.*, Peng et al., 2010; Muzzin et al., 2012) or modest (*e.g.*, Vulcani et al., 2010; Popesso et al., 2011; Koyama et al., 2013; Old et al., 2020, and McGee et al. 2011 for groups) trends between the star forming main sequence and environment.

Importantly, the present comparisons at $z>1$ between field and cluster galaxies are typically made for small samples and/or with limited stellar mass coverage, relying on the measurement of a few absorption line indices (*e.g.*, van der Wel et al., 2004; van Dokkum & van der Marel, 2007; Gobat et al., 2008; Woodrum et al., 2017; Saracco et al., 2020) or photometric SEDs (*e.g.*, Rettura et al., 2010, 2011; Raichoor et al., 2011; Saracco et al., 2017). While recent spectroscopic surveys have collected larger samples of quiescent galaxies at higher redshifts (*e.g.*, Pacifici et al., 2016a; Thomas et al., 2017; Carnall et al.,

2019a; Estrada-Carpenter et al., 2019), there has not yet been a systematic study of the SFHs with environment. We can significantly improve our understanding of the differences in SFHs of galaxies related to their environment with the Gemini Observations of Rich Early ENvironments survey (GOGREEN[1]; Balogh et al., 2017, 2020). The GOGREEN survey targeted galaxies in clusters and groups at $1<z<1.5$, and includes isolated 'field' galaxies along the line-of-sight of these structures. With galaxies at lower stellar masses, and at higher redshifts, than preceding surveys (*e.g.*, GCLASS, Muzzin et al. 2012; GEEC2, Balogh et al. 2014), GOGREEN is better suited to test the predictions of galaxy evolution models (*e.g.*, Bower et al., 2012; Weinmann et al., 2012).

Taking advantage of the well-sampled, homogeneously selected spectroscopy and broad photometric coverage for hundreds of galaxies observed as part of GOGREEN, we measure the SFHs and mass-weighted ages for quiescent galaxies in both average, 'field,' environments and in massive galaxy clusters. Comparing the star formation timescales between galaxies in clusters and field environments, we test simple quenching models which have been proposed to explain the difference in ages between the two populations. This work complements the comparison of the stellar mass distributions measured in (van der Burg et al., 2020), and the relation between stellar mass and star formation in star-forming galaxies (Old et al., 2020).

The paper is outlined as follows. In Section 3.2, we provide a brief description of the GOGREEN sample and the selection of quiescent galaxies used in our analysis. In Section 3.3, we describe the SFH fitting procedure. In Section 3.4, we show the SFHs and estimated average ages, and test the robustness of the results against our selection criteria for quiescent galaxies. Then, in Section 3.5, we discuss the SFHs and average ages as a function of stellar mass and environment in the context of the literature. We also discuss our results in the context of two toy models for environmental galaxy quenching scenarios: either galaxies quench upon being accreted into denser environments, or galaxies in denser environments simply formed earlier. Lastly, in Section 3.6 we summarize the results.

3.2 Data and sample selection

3.2.1 The GOGREEN survey

The GOGREEN survey includes 21 galaxy systems at $1<z<1.5$ selected to be representative of progenitors of local clusters and groups, described in detail in Balogh et al. (2017)

[1]http://gogreensurvey.ca/

Table 3.1: Description of the GOGREEN galaxy cluster targets. Notes: Coordinates and redshifts for each galaxy system in the GOGREEN sample. Spectroscopic redshifts are from Balogh et al. (2020). SpARCS1033 was excluded in this study because of the lack of K-band photometry. Notes: † indicates clusters also in the GCLASS survey.

Full name	BCG RA, Dec (J2000)	Redshift
SPT0205	02:05:48.19, -58:28:49.0	1.323
SPT0546	05:46:33.67, -53:45:40.6	1.068
SPT2016	21:06:04.59, -58:44:27.9	1.132
†SpARCS0035-3412	00:35:49.68, -43:12:23.8	1.335
SpARCS0219-0531	02:19:43.56, -05:31:29.6	1.328
SpARCS0335-2929	03:35:03.56, -29:28:55.8	1.368
SpARCS1034+5818	10:34:49.47, +58:18:33.1	1.388
†SpARCS1051+5818	10:51:11.23, +58:18:02.7	1.034
†SpARCS1616+545	16:16:41.32, +55:45:12.4	1.157
†SpARCS1634+4021	16:34:37.00, +40:21:49.3	1.177
†SpARCS1638+4038	16:38:51.64, +40:38:42.9	1.194

and Balogh et al. (2020). Groups and clusters with a wide range of halo masses were targeted, and within them galaxies with a wide range of stellar masses were targeted. For the present work we include eleven clusters from the GOGREEN survey which have complete spectroscopy and photometry as of 2020.

Table 3.1 lists the clusters with their coordinates and redshifts (Balogh et al. 2017, van der Burg et al. 2020, see Biviano et al. 2021 for halo masses and velocity dispersions). Three of these systems are from the South Pole Telescope survey (SPT, Brodwin et al., 2010; Foley et al., 2011; Stalder et al., 2013), nine are from the Spitzer Adaptation of the Red-Sequence Cluster survey (SpARCS, Wilson et al., 2009; Muzzin et al., 2009; Demarco et al., 2010). Five of the SpARCS clusters were also included in the Gemini cluster Astrophysics Spectroscopic Survey (GCLASS, Muzzin et al., 2012). We add to the number of low mass galaxies in the GCLASS sample, and increase the sampling at higher masses, particularly at z<1.3.

GOGREEN provides broadband photometry and Gemini Multi-Object Spectrograph (GMOS) spectroscopy for a selection of galaxies in each system. The survey strategy and magnitude limits (z'<24.25 and [3.6]<22.5) of the GOGREEN survey enables both a large sampling of bright galaxies and very deep spectroscopy of much fainter galaxies. The full survey is statistically complete for all galaxy types with stellar masses log $M_*/\mathrm{M}_\odot \gtrsim 10.3$

at $1<z<1.5$ (Balogh et al., 2020). Including the systematic offset between stellar mass estimates (see Section A.3) the mass completeness of the sample is log $M_*/\mathrm{M}_\odot \gtrsim 10.5$. Completeness here is characterized as a function of stellar mass and clustercentric distance, where above this limit our sample is representative of an unbiased sampling of the full galaxy population. We note that the lower mass selection used throughout this paper is below this mass completeness threshold, and the conclusions drawn from these galaxies are not necessarily statistically robust.

3.2.2 Spectroscopic sample

Spectroscopy for the GOGREEN galaxies was taken with the GMOS instruments using the R150 filter and three spectral dither positions (8300 Å, 8500 Å, and 8700 Å). Spectral dithers are done to fill in the gaps between the GMOS CCDs where spectral information is lost. This provides continuous wavelength coverage free of second order contamination over 6400–10200 Å. For the redshift range $1<z<1.5$, this corresponds to about 2500–5250 Å rest-frame.

The GMOS detector field of view is $5.5' \times 5.5'$, which roughly matches the size of our clusters ($\sim$2.8 Mpc at $z = 1.3$). With $1''$ slits, the spectra have an observed FWHM resolution of $\sim$20 Å, ($\mathrm{R} = 440 \pm 60$). We used the nod and shuffle mode to maximize the number of slits per exposure, particularly in the cluster centres, and to perform accurate sky subtraction. Specifics of the spectral data reduction can be found in Balogh et al. (2017) and Balogh et al. (2020). The basic steps follow the Gemini IRAF[2] reduction procedure, with additional corrections for scattered light and telluric absorption. Wavelength calibrations were established using CuAr lamp observations taken concurrent to the GMOS observations, with reference to night skylines to account for flexure shifts. The lack of features below 6400 Å results in unreliable calibrations at this end of the spectra. The 1D spectra were extracted and combined. Although a relative sensitivity correction was applied, based on standard star observations, the spectra were not absolute flux calibrated. This requires additional corrections in the fitting procedure discussed in Section 3.3.

In this study, we included only galaxies for which we could measure a spectroscopic redshift with confidence (quality flag 3 or 4) – this includes 970 galaxies. Spectroscopic redshifts were derived using the Manual and Automatic Redshifting Software (MARZ, Hinton et al., 2016), as described in Balogh et al. (2020).

[2]IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

3.2.3 Photometric coverage

GOGREEN has broad photometric coverage for each galaxy system. A full description of the photometry, as well as the calculation of stellar masses and rest-frame colours, is provided in van der Burg et al. (2020). The photometry includes Gemini GMOS (z'), Spitzer IRAC[3], VLT VIMOS[4] (U, B, V, R, I, z) and HAWK-I (Y, J, K_s), Subaru SuprimeCam (g, r, i) and HyperSuprimeCam (z, Y), Magellan Fourstar (J_1, J, K_s), CFHT WirCam (J, K_s) and MegaCam (U), and Blanco DECam (z). The one GOGREEN cluster not included in our sample (SpARCS1033) did not have K-band data as of the 2019A semester.

Rest-frame colours were derived from best fit templates to the observed photometry with EAZY (Brammer et al., 2008), assuming an exponentially declining SFR, Calzetti et al. (2000) dust law, Bruzual & Charlot (2003, BC03) stellar library, and solar metallicity. Templates were fixed to the spectroscopically determined redshift, and the redshift-corrected best-fit template was then convolved with U,V, and J filters (see Figure 3.1). Galaxies observed in the COSMOS fields have rest-frame colours as provided from the UltraVISTA v4.1 catalogue (Muzzin et al., 2013a).

3.2.4 Sample selection

The goal of this paper is to compare the ages and star formation histories of quiescent galaxies in cluster and field environments. Quiescent galaxies were selected based on their position in rest-frame U-V and V-J colour space, which has been shown to effectively separate star forming and quiescent galaxies (Labbé et al., 2005; Wuyts et al., 2007; Williams et al., 2009; Patel et al., 2012; Whitaker et al., 2012; Muzzin et al., 2013b) up to z<2.5 (Williams et al., 2010). Including the NIR colour allows quiescent galaxies to be more clearly distinguished from dusty star forming galaxies, since dust reddening scatters along the UVJ-colour selection vector. We consider alternative selections in Appendix A.1.

Figure 3.1 shows the rest-frame $U-V$ and $V-J$ colours of the GOGREEN spectroscopic sample, with the separation between star forming and quiescent galaxies,

$$(\mathrm{U}-\mathrm{V}) > 1.3 \quad \cap \quad (\mathrm{V}-\mathrm{J}) < 1.5 \quad \cap \quad (\mathrm{U}-\mathrm{V}) > 0.88\,(\mathrm{V}-\mathrm{J}) + 0.59 \tag{3.1}$$

as defined in Muzzin et al. (2013b) for 1<z<4, adapted from Williams et al. (2009). Of the 970 galaxies with spectra and robust redshift measurements, 338 quiescent galaxies were identified.

[3]Supplemented by archival data primarily from SERVS (Mauduit et al., 2012), S-COSMOS (Sanders et al., 2007), SPUDS (Galametz et al., 2013), and SWIRE (Lonsdale et al., 2003)

[4]Program ID: 097.A-0734.

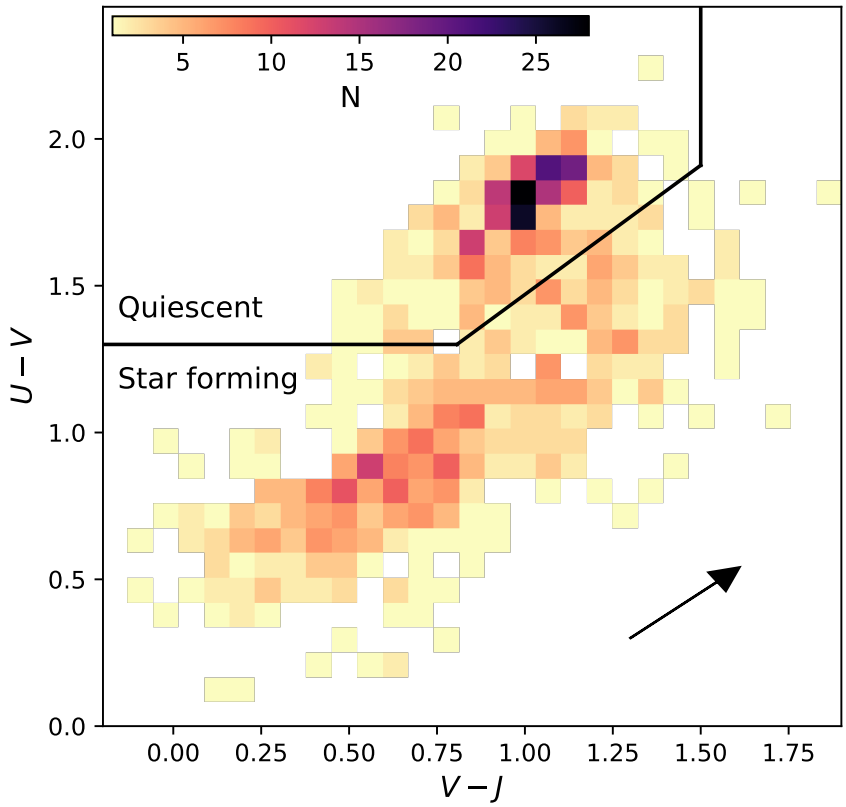

Figure 3.1: Rest-frame UVJ colours for the GOGREEN spectroscopic sample between $1<z<1.5$, shown as a 2D histogram for both cluster and field galaxies. Quiescent galaxies are selected above the shown boundaries, as defined by Muzzin et al. (2013b). The arrow indicates the impact of 1 mag of extinction in the V-band, using the Calzetti et al. (2000) dust law.

Galaxies were identified as cluster members or field based on their spectroscopic redshifts and projected phase space locations. A detailed description will be provided in Biviano et al. (2021). The field galaxy sample is taken as the galaxies along the line-of-sight of the clusters, not identified as members, and with spectroscopic redshifts within $1<z<1.5$. We also include galaxies in the five GOGREEN fields within COSMOS (Muzzin et al., 2013a). These pointings targeted group-mass systems that are otherwise not considered in this paper. We include galaxies that have a line-of-sight velocity more than 900 km s^{-1} from the targeted group redshift in our field sample. Our sample of quiescent galaxies includes 224 cluster members and 110 field galaxies.

Figure 3.2 shows the distribution of our quiescent sample as a function of stellar mass and redshift. Cluster galaxies are coloured orange and shown as hatched histograms, and field galaxies are blue with solid histograms. Stellar masses were determined from SED fits to the photometry and spectroscopy, discussed further in Section 3.3. The majority of the cluster galaxies are within $1.1<z<1.2$, while the field galaxies are more evenly spaced in redshift (Balogh et al., 2020).

3.3 Fitting star formation histories of quiescent galaxies

SFHs of the quiescent galaxies are constrained by fitting photometric and spectroscopic data with spectral energy templates using the Prospector inference code[5] (Leja et al., 2017; Johnson et al., 2020, v0.3.0). The physical models are generated from the flexible stellar population synthesis code FSPS (Conroy et al., 2009) with MIST stellar evolutionary tracks and isochrones (Choi et al., 2016; Dotter, 2016, based on the MESA stellar evolution code Paxton et al. 2011, 2013, 2015, 2018), and MILES[6] spectral templates (Vazdekis et al., 2015b). Biases related to metallicities are discussed further in Appendix A.2, where we conclude that any such systematics have a negligible impact on our results.

We assume a nonparametric[7] form for the SFHs with a continuity prior (described in Leja et al., 2019a, see also Section 2.3.9) and Milky Way extinction curve (Cardelli et al., 1989). We mask the only prominent emission line region within our spectral range ([O II]) rather than include a nebular line emission model. Table 3.2 lists the free parameters in the fitting procedure: redshift, total mass formed, dust optical depth, stellar metallicity,

[5]https://github.com/bd-j/prospector
[6]http://miles.iac.es/
[7]*Nonparametric* here means that the SFH has no specified functional form.

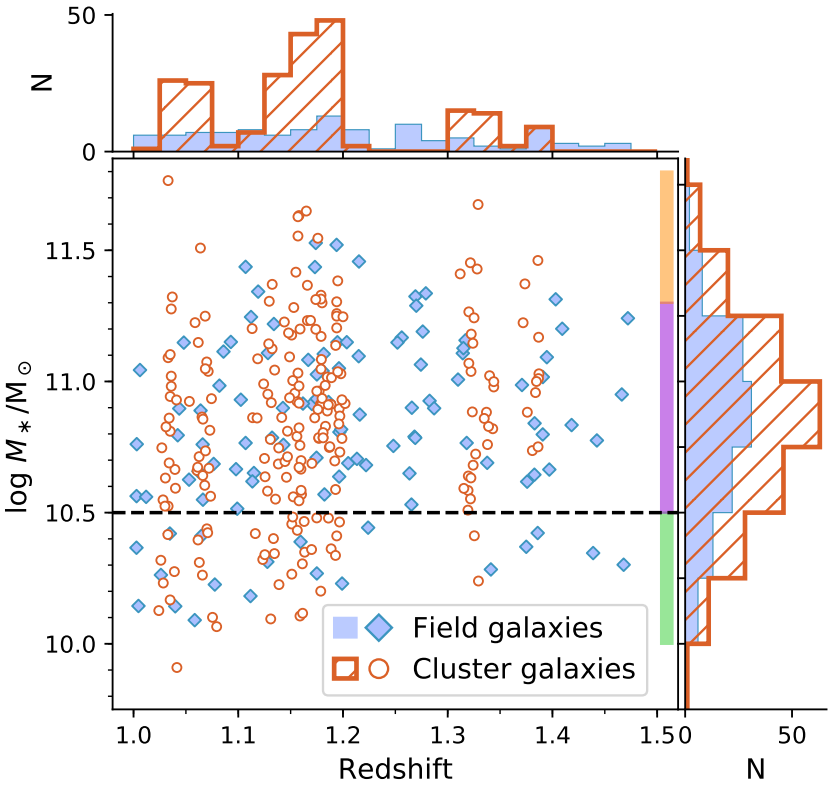

Figure 3.2: Stellar masses and redshifts for the *UVJ*-quiescent GOGREEN spectroscopic sample, with corresponding histograms on each axis. Field galaxies are shown as blue diamonds, and cluster member galaxies are red. Coloured blocks indicate the span of mass bins discussed throughout the analysis. A dashed line indicates the mass-completeness of our sample, where the lowest mass bin is below this threshold. Note that the stellar masses shown here are derived with nonparametric SFHs, and are systematically offset from those derived using FAST as reported in other GOGREEN papers, see Section 3.3 and Appendix A.3 for details.

relative[8] SFR ratios in ten time bins, spectral normalization and higher order spectral polynomial coefficients, spectral noise inflation, and spectral outlier fraction. The priors for each parameter are also provided in this table. The time bins are spaced so that the first four bins correspond to 30 Myr, 100 Myr, 500 Myr, and 1 Gyr in units of lookback time, and the final bin covers the first 5% of the age of the universe. The remaining time bins are spaced equally in time.[9] Note that galaxies observed at different redshifts will have different time binning in cosmic time (*i.e.*, time since the Big Bang).

Three of the free parameters help to identify systematics in the spectra. The spectral-noise inflation ($j_{\rm spec}$) effectively increases the uncertainties on all spectral points by a multiplicative factor. This is counter-balanced by a standard likelihood penalty term for larger uncertainties. This down-weighting of the spectra is rarely relevant unless the data has high S/N and is poorly matched by the models (*e.g.*, perhaps because the models do not include all realistic spectral features). We also include an outlier pixel model ($\rm f_{outlier,\ spec}$) which modifies the likelihood to be more permissive of large deviations from the model. Such large deviations can come from poor matches to the stellar models (due to differences in, for example, α-enhancement) and increases their errors by a factor of 50. The outlier fraction is less than 3% for the majority (95%) of our fits.

In fitting the spectroscopy and photometry together, we need to account for uncertainties in the spectral response calibration, and for the overall flux calibration due to slit losses. Several authors have demonstrated the challenge of simultaneously fitting spectral and photometric data, especially when the spectral continuum is not well characterized (Panter et al., 2007; van der Wel et al., 2016; Belli et al., 2019; Carnall et al., 2019a; Johnson et al., 2020). As described in Section 3.2.2, the spectra were not absolute flux calibrated. The flux calibration is uncertain due to slit losses, the lack of atmospheric dispersion correction, and uncertainties in the telluric absorption corrections. To accommodate for these effects, the spectral normalization (spec_norm) and a spectrophotometric calibration polynomial are calculated from the ratio of the observed and model spectrum, and applied to the template spectrum prior to assessing the goodness of fit. We use a third order Chebyshev polynomial, since a higher order polynomial could wash out real spectroscopic features.

The spectral fit was restricted to the wavelength range 3525-4400 Å rest-frame, covering the majority of useful spectral features (*e.g.*, CaH+K, $\rm D_n(4000)$, Hδ, G) while minimizing

[8]Relative with respect to adjacent bins. For N time bins, there are therefore N-1 free parameters. See Table 3.2.

[9]Convergence tests with some representative galaxies demonstrated that ten bins provide sufficient time resolution, while limiting the number of free parameters in the fitting procedure and the corresponding computational time.

Table 3.2: SFH parameters and priors. Notes: 1) Spectroscopic redshift. 2) Total mass is the sum of total stellar mass and mass lost to outflows. See note 3) for a comment on the prior. 3) We assume a Milky Way extinction curve (Cardelli et al., 1989). 4) We assume a prior on the stellar mass-metallicity relation (MZR) according to the local trend reported by Gallazzi et al. (2005), where we add the systematic offset between parametric and nonparametric stellar mass estimates (see Appendix A.3). 5) Ratio of the SFRs in adjacent bins of the ten-bin nonparametric SFH. The time bins are spaced in lookback time: 0, 30 Myr, 100 Myr, 500 Myr, and 1 Gyr, five equally spaced bins, and lastly $0.95\times$ the age of the universe at the observed redshift. For N time bins, there are N-1 free parameters. 6) The shape of the spectral continuum can be adjusted by a 3$^{\rm rd}$ degree Chevyshev polynomial to account for systematics in the relative flux calibration. The first coefficient controls the normalization of the spectra. 7) The uncertainty on the spectra can be increased by a given factor, with a likelihood penalty for factors giving reduced $\chi^2<1$. 8) An outlier pixel model can increase the errors for individual pixels by a factor of 50, to accommodate for poor matches between the data and spectral templates.

Note	Parameter	Description	Prior
1	$z_{\rm obs}$	Observed redshift	Uniform(min $= z_{\rm spec} - 0.01$, max $= z_{\rm spec} + 0.01$)
2	$\log(M_{*,\ \rm total}/{\rm M}_\odot)$	Total stellar mass formed	Uniform(min $= 8$, max $= 15$)
3	$\hat{\tau}_{\rm dust,\ diffuse}$	Diffuse dust optical depth	Uniform(min $= 0$, max $= 4$)
4	$\log(Z_*/{\rm Z}_\odot)$	Stellar metallicity	ClippedNormal($\mu = f(M_{*,\ \rm total})$, $\sigma = f(M_{*,\ \rm total})$, min $= -2$, max $= 0.19$)
5	$\log({\rm SFR}_n/{\rm SFR}_{n+1})$	Ratio of the SFR ratios in adjacent time bins	Student-t($\mu = 0$, $\sigma = 0.3$, $\nu = 2$)
6	c_n	Polynomial coefficients for spectrophotometric calibration model, $n = 4$	Uniform(min $= \dfrac{-0.1}{n+1}$, max $= \dfrac{0.1}{n+1}$) where $n > 0$
7	$j_{\rm spec}$	Spectral noise inflation term	Uniform(min $= 1$, max $= 3$)
8	$f_{\rm outlier,\ spec}$	Spectral outlier fraction	Uniform(min $= 10^{-5}$, max $= 0.5$)

sensitivity to the lowest and highest wavelength ranges where flux calibration is the most uncertain due to rapidly changing sensitivity. The lower bound is set by the different resolution of the MILES spectral templates at redder wavelengths. Beyond 4400 Å, a few of the spectra suffer systematic effects due to insufficiently well corrected telluric absorption. Due to the limitations of the template spectra, the metallicity was restricted to $-2 < \log\ Z/\mathrm{Z}_\odot < 0.19$ and the abundance patterns were fixed to solar. Prior to fitting the spectroscopy, the template spectra are smoothed to match the resolution of the observed spectrum. Lastly, we assumed a minimum photometric error of 5% as a conservative estimate of the calibration uncertainty in the photometry.

Prospector uses emcee (Foreman-Mackey et al., 2013) to create an ensemble of walkers which sample the parameter space following an affine invariant algorithm for a given number of steps. We used 64 walkers, iterative 'burn-in' in steps of 16, 34, 68, and 124, and a minimum of 1024 iterations thereafter. Each fit was visually confirmed as being converged (*i.e.*, the traces were stable), or the sampling was restarted from the previous maximum probability solution. We take only the last 500 iterations when building the posteriors. The SFHs were sometimes multimodal, particularly where the S/N was poor, which motivated us to use a weighted combination of a differential moves (80%) and snooker differential moves (20%) in the MCMC sampling[10].

An example of the output of this fitting procedure is shown in Figures 3.3 and 3.4. In Figure 3.3, the observed photometry (top, green circles) and spectrum (bottom, green line) are shown relative to the template with the highest combined likelihood and prior (maximum a posteriori, MAP; black line). The shaded green regions about the spectrum indicate the uncertainty, while masked regions in the fit are shown as faint green lines. The spectrum is shown relative to the MAP after the spectrophotometric polynomial was applied. A selection of the SFH parameters with their posteriors are shown in Figure 3.4 as a corner plot, and the range of SFRs as determined from the relative SFRs. The 50^{th} percentile value of each parameter is listed above its posterior, with uncertainties from the 68% confidence regions (CRs).

Throughout this work we report the uncertainties as 68% CRs (which corresponds to the 16^{th} to 84^{th} percent range) as the majority of the distributions are non-symmetric. The lower (16^{th}–50^{th}) and upper (50^{th}–84^{th}) reported are equivalent to $\pm 1\ \sigma$ for a Normal distribution.

From the SFH posteriors we calculate[11] the mass-weighted stellar age (t_{mass}, discussed

[10] As described in
https://emcee.readthedocs.io/en/stable/user/moves/#emcee.moves.DEMove

[11] FSPS calculates t_{mass} when `compute_light_ages` = True.

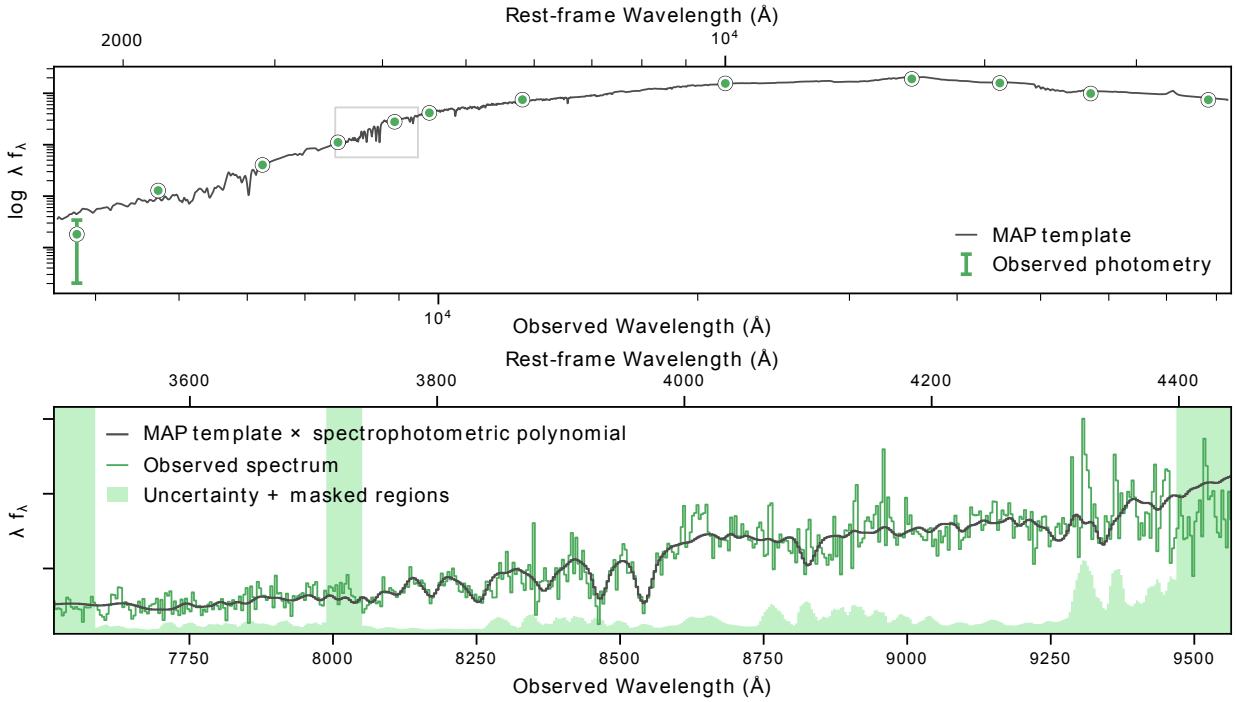

Figure 3.3: Example of typical photometric (green circles, top image) and spectroscopic (green line, bottom image) observations shown with the corresponding maximum a posteriori (MAP) template (black line), as a function of observed wavelength. The grey box indicates the wavelength region covered by the spectra relative to the photometry. The MAP template relative to the spectrum is shown with a polynomial 'correction' to account for systematics in the relative flux calibration. Green shaded regions indicate the uncertainty and masked regions of the spectrum for the SFH fitting (*e.g.*, the [O II] line at 3727 Å rest-frame).

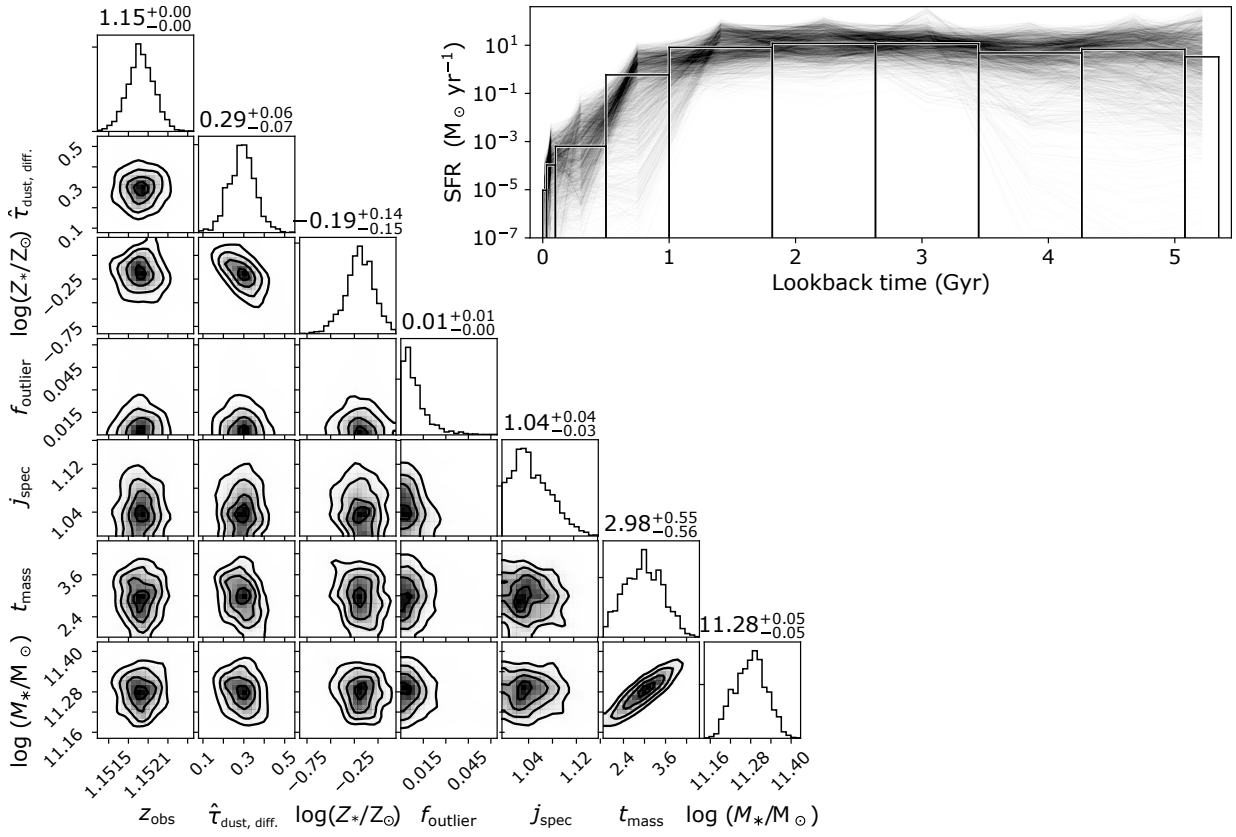

Figure 3.4: Selected posteriors for the fitting result shown in Figure 3.3, which is a typical galaxy from our sample of GOGREEN quiescent galaxies. Top: corner plot showing a selection of posterior distributions for SFH parameters: observed redshift, diffuse dust opacity, stellar metallicity, spectral outlier fraction, and spectral noise inflation, as well as two derived parameters: mass-weighted age and stellar mass (see Section 3.3). Posteriors are shown smoothed with a 1 σ Gaussian, and the 50^{th} percentiles are indicated on the top of each histogram with 68% CRs. Bottom: The posteriors for the SFRs are shown as a function of lookback time, where time bins are drawn with heights equal to the median in each bin.

in Section 3.4.1) and stellar mass. The latter is determined from the posterior of the total stellar mass formed and the corresponding fraction of surviving stellar mass for each sampling. We confirm that the stellar masses derived using nonparametric modelling are systematically offset from than those derived with parametric models, such as exponentially declining SFR models (*e.g.*, using FAST; Kriek et al. 2009). This comparison is discussed in Appendix A.3. We note that the stellar masses reported in other GOGREEN papers (*e.g.*, Balogh et al., 2017; Chan et al., 2019; Old et al., 2020; van der Burg et al., 2020) are derived using FAST, and therefore differ from the stellar masses in this paper by +0.2 dex. Since the focus of this paper is a differential comparison between galaxies in cluster and field environments, our results are less sensitive to the systematic effects related to model choices.

Only two of the fits clearly failed to reproduce the observations. For both the spectral continuum is dominated by telluric absorption that was not sufficiently corrected. The final sample includes 331 galaxies, 109 of which are field galaxies, and 222 are cluster galaxies.

3.4 Results

In this section, we present the results of the nonparametric SFH fitting applied to the sample of 331 quiescent GOGREEN galaxies. We explore differences related to stellar mass and density of local environment through comparing the SFHs and mass-weighted ages. We then test our result by refining our selection of quiescent galaxies. In Appendix A.4 we compare features in co-added spectra to the results of fitting the individual galaxies.

3.4.1 The dependence of star formation histories on mass and environment

Figure 3.5 shows the median sSFRs (star formation rates divided by the final stellar mass) for individual galaxies as a function of lookback time. Subplots separate galaxies according to environment and stellar mass. The overall median sSFRs for each selection of galaxies are shown as a bold lines, and the 68% CRs are shown as shaded regions. The right-hand column compares sSFRs for galaxies between the two environments, at fixed mass. The bottom row compares SFHs for galaxies between mass selections, at fixed environment. The hatched-shaded region in the right-hand column and bottom row shows the bootstrapped uncertainties on the medians.

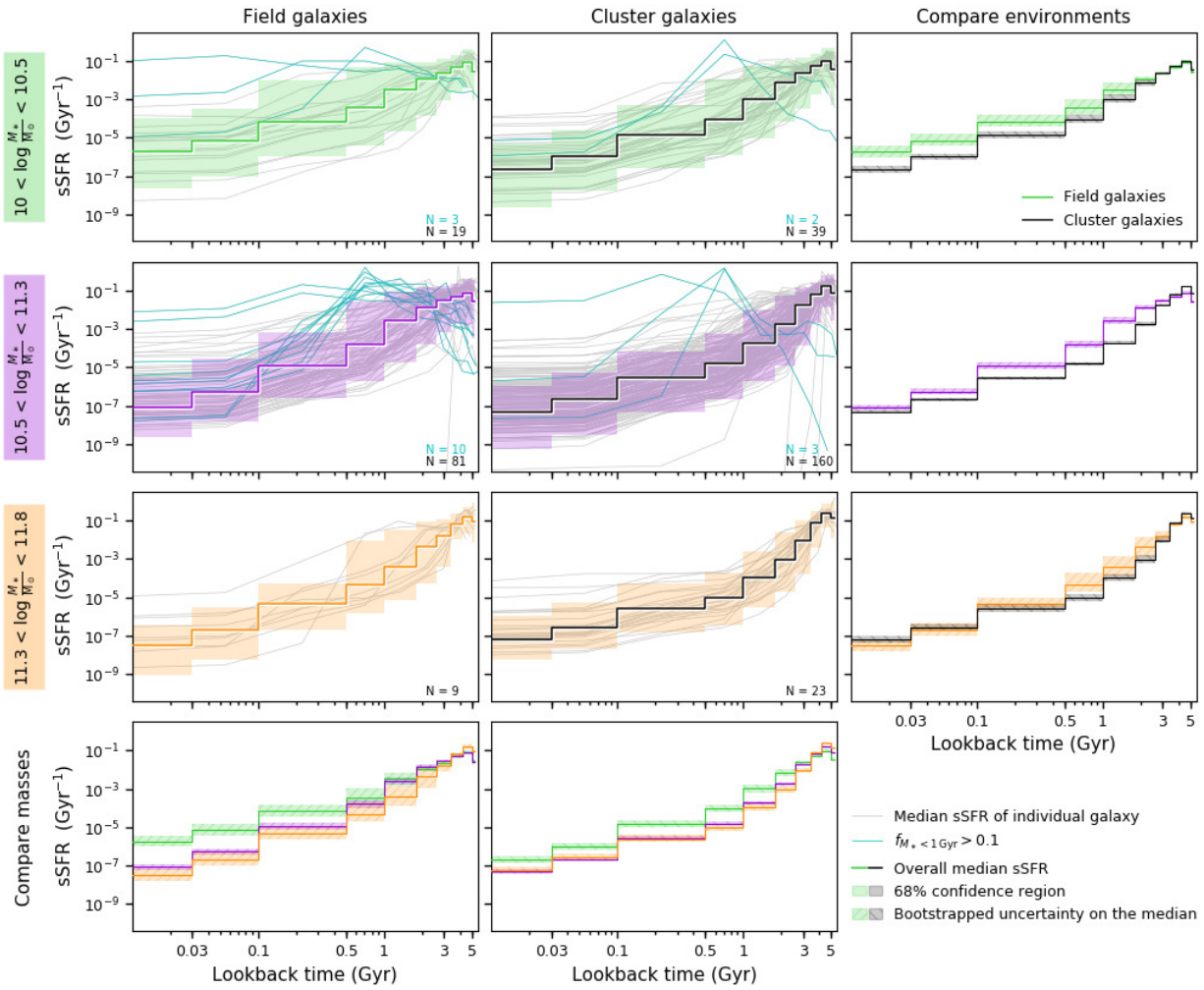

Figure 3.5: Specific SFR ($\mathrm{SFR}(t)/M_{*,z=z_{\mathrm{obs}}}$) as a function of lookback time for field and cluster galaxies, in three mass bins. Individual sSFRs are shown as grey lines, where galaxies in which more than 10% of their stellar mass has formed within the last 1 Gyr (see Section 3.4.3) are coloured cyan. The number of galaxies in each mass and environment selection is labelled at the bottom-right of the subplot. The overall median sSFR in each subsample is shown as a bold line, and is also shown in the right-hand column to compare between environments, and in the bottom row to compare between mass selections. The shaded region indicates the 68% CR of the combined sSFRs, while the hatched shaded regions show the bootstrapped uncertainty on the overall median. Two trends are apparent from the median SFHs: higher mass galaxies form their mass earlier (*i.e.*, mass-accelerated evolution), and cluster galaxies form their mass earlier.

The majority of galaxies follow a steady decline in sSFR, consistent with passive evolution. A few galaxies have more shallow declines or more recent star formation. We indicate galaxies which have more than 10% of their stellar mass formed within the last 1 Gyr, with cyan lines in Figure 3.5 (and list the number in each panel), and discuss them in Section 3.4.3. This population is not unexpected, as the UVJ colour selection can include younger galaxies, or those in transition. Four galaxies have extremely rapid declines in SFR, with negligible star formation within the last 1 Gyr.

Comparing galaxies at fixed environment (bottom row of Figure 3.5), we find that more massive galaxies have overall earlier star formation activity, and form their stars over shorter timescales. Lower mass galaxies, on average, have more extended SFHs. This trend is consistent with the 'mass-dependent evolution' scenario (*e.g.*, Nelan et al., 2005; Thomas et al., 2005), sometimes called 'archaeological downsizing' (Neistein et al., 2006). Interestingly, the galaxies in our moderate mass bin more closely resemble their higher mass counterparts, but have slightly longer star forming timescales.

Comparing galaxies at fixed mass (right-hand column of Figure 3.5), galaxies in clusters have overall earlier star formation activity, in the sense that the sSFRs decline more quickly. Below masses of $10^{11.3}$ $M_{\odot}$, the sSFRs of field galaxies are higher within the last $\sim$2 Gyr. In general, field galaxies in our lower mass sample have the flattest (most extended) SFHs.

Rettura et al. (2011) estimated the SFHs of massive ETGs in both clusters and the field at $z \sim 1.3$, based on photometric observations, and concluded that while the formation epochs are similar between environments, field galaxies take longer to assemble than cluster galaxies. Specifically, they found that after 1 Gyr of star formation, 75% of cluster galaxies had assembled at least 80% of their final stellar mass, compared with only 35% of field ETGs. We find a smaller difference, but also phrase it slightly differently given that we do not use parametric SFHs and do not constrain the onset of star formation: by $z \sim 5.4$ ($\sim$1 Gyr since the Big Bang), 75% of our higher mass cluster galaxies had formed at least 80% of their final stellar mass, compared to only 50% of field galaxies. Although we find a stronger difference between the SFHs of field and cluster galaxies at moderate stellar masses, the difference is smaller than found by Rettura et al. (2011) (75% *vs* 46%), but consistent within the uncertainties of the SFHs given the systematic differences in modelling (discussed in Raichoor et al., 2011, in the context of the Rettura et al. 2011 measurements). We discuss the SFHs in the context of mass-dependent evolution and the literature further in Section 3.5.1.

3.4.2 The dependence of age on mass and environment

From the SFHs we calculate the mass-weighted age ($t_{\rm mass}$; also known as the mean stellar age) which broadly describes the average formation time of stars in a given galaxy in units of lookback time (see Chapter 2 Equation 2.13). We also express the ages in units of cosmic time, $t_{\rm univ}(z_{\rm obs}) - t_{\rm mass}$ (sometimes called the formation time, $t_{\rm form}$), which is convenient when comparing galaxies observed across a range of redshifts. Trends between $t_{\rm mass}$ and UVJ colour are discussed in Appendix A.5.

Figure 3.6 shows the distribution of the stellar mass and mass-weighted ages, $t_{\rm mass}$, in units of cosmic time. Contours show the combined posteriors of the field (blue) and cluster (red) galaxies, where white points indicate the medians of the individual posteriors. The typical uncertainty for the mass-weighted ages is 0.52 Gyr, and for the stellar masses 0.043 dex. Diamonds indicate galaxies that have formed more than 10% of their stellar mass within the last 1 Gyr ($f_{M_*<1\ \rm Gyr} > 0.1$), discussed in Section 3.4.3. The right-hand column shows combined age histograms for field and cluster galaxies within three mass ranges. The galaxy sample is bootstrap sampled to determine the variances within the time bins. Medians and 68% CRs of the combined distributions are indicated with horizontal bars near the bottom axis.

The mass-weighted ages of our sample are distributed primarily between $2<z<8$, where there is a modest mass dependence in that galaxies in our higher mass selection have mass-weighted ages between $3<z<10$ while the lower mass galaxies fall within $2<z<6$. The majority (>50%) of the higher (lower) mass galaxies have formed at least half of their stellar mass by $z \sim 5.4$ ($z \sim$3.3). The shapes of the mass-weighted age distributions are also broader at lower stellar masses, as we saw from the SFHs shown in Figure 3.5 and discussed in the previous section. Specifically, at $z \sim 3.3$, the at least 90% of the higher mass galaxies have formed at least half their stellar mass, compared to only 50% of the lower mass galaxies.

For the lower and higher mass galaxies in our sample, the differences between the mass-weighted ages of galaxies between environments at fixed mass are smaller than the differences across our stellar mass range at fixed environment. This is apparent in the histograms of the mass-weighted ages shown in the right-hand column of Figure 3.6: the shapes of the distributions at fixed mass are more similar than between the higher and lower mass galaxies. We note, however, that the age-distributions for field galaxies are shifted towards younger ages, as well as broader. Interestingly, the distribution of mass-weighted ages for the moderate mass cluster galaxies more closely resemble that of their more massive counterparts, while the field galaxies are more similar to their lower mass counterparts. This is to say that the moderate mass galaxies in clusters are largely older,

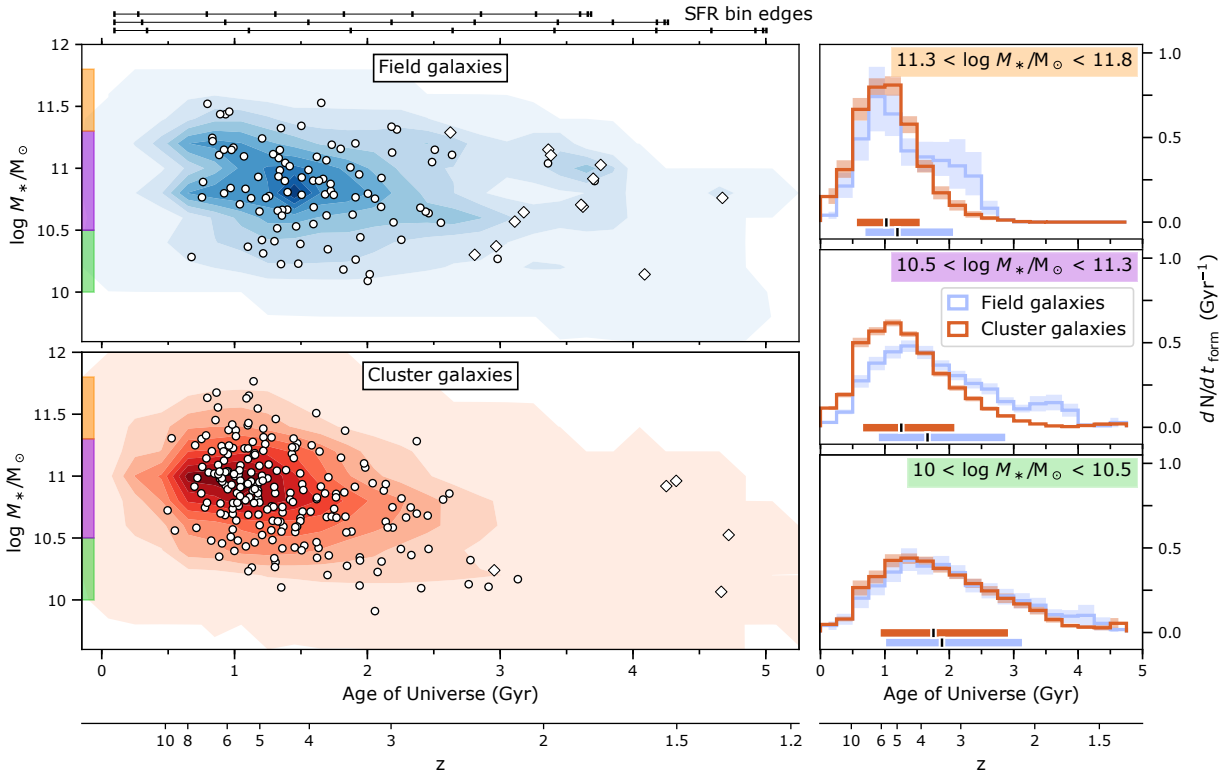

Figure 3.6: Comparison of stellar masses and mass-weighted ages between field (blue) and cluster (red) galaxies. Left: Combined posteriors of stellar masses and $t_{\rm mass}$ (in units of cosmic time), shown as contours. The medians of the individual posteriors are marked with white circles/diamonds. Diamonds indicate young galaxies, which have formed more than 10% of their stellar mass within the last 1 Gyr (discussed in Section 3.4.3). Horizontal bars at the top of the figure indicate the edges of the time bins for $z = 1.5$ (top), $z = 1.25$ (middle), and $z = 1$ (bottom). The bins were defined in units of lookback time, and therefore do not match up for galaxies observed at different redshifts. Right: Combined $t_{\rm mass}$ posteriors for field and cluster galaxies, shown in three mass bins. The medians (black mark) and 68% CRs (coloured bar) of each distribution are marked at the bottom of each subplot. The shaded regions show the bootstrapped uncertainty of each histogram. Although there are field galaxies that formed as early as the oldest cluster galaxies, and cluster galaxies that formed as late as the youngest field galaxies, *on average* field galaxies have more extended SFHs to reach the same final stellar mass.

while the field galaxies are both younger overall and have an extended tail towards younger ages.

Next, we attempt to compare the intrinsic distribution of ages between the field and cluster environments, accounting for the uncertainties on individual measurements. Comparing the rms uncertainties of individual posteriors to that of combined posteriors of similar mass (*i.e.*, $(\sigma_i^2 - \sigma_{\rm comb.}^2)^{-1/2}$, although neither are necessarily Normal), we find that there are significant intrinsic distributions of ages in both the cluster and field sample, with rms's of 0.74 Gyr and 0.73 Gyr, respectively. The distributions are consistent between environments, however.

In order to better quantify the difference in mass-weighted ages between field and cluster galaxies, we compare the combined age distributions in a cumulative sense. This allows us to compare the cosmic time at which the two populations reach a given fraction of their mass-weighted age distribution. Within small (0.1 dex) mass ranges we select field galaxies and cluster galaxies, calculate their respective combined age distributions, and interpolate the cumulative distributions to the same binning. Within a given mass bin, we include all portions of the posteriors that fall within the limits (*i.e.*, we are not selecting based on the median mass). We then measure the horizontal offset (*i.e.*, in units of time) between the distributions (field − cluster). An example of this procedure is shown in Figure 3.7. The mass-selected comparisons are then combined, weighted by the number of samplings from the respective posteriors, and the overall age offset is determined. We bootstrap our galaxy sample 500 times to capture the true variance.

Figure 3.7 shows the cumulative-age-distribution comparisons combined into broader mass selections (coloured histograms), and for the full mass range of our sample (black). The median age difference is shown for each mass selection, with error bars corresponding to the 68% CR. Across the mass range of our sample, $10 < \log\ M_*/{\rm M}_\odot < 11.8$, the median age difference between field and cluster galaxies is $0.31^{+0.51}_{-0.33}$ Gyr, in the sense that cluster galaxies are on average older than field galaxies. Interestingly, the age difference is slightly smaller for the lower and higher mass galaxies, and slightly larger for our moderate mass galaxies. Note that the lower mass selection is below the mass completeness limit of our sample, and is dominated by galaxies $z<1.2$. That is to say, the sample of galaxies below $10^{10.5}$ ${\rm M}_\odot$ is not a representative sample of the galaxy population, and the result is not as robust. Omitting the lower mass galaxies does not significantly change our result, however: the median age of the cluster galaxies is instead $0.35^{+0.51}_{-0.32}$ Gyr older than that of field galaxies.

We also consider the age comparison between galaxies at the lower end of our redshift selection, $1<z<1.2$, and find that the age difference is slightly smaller: $0.21^{+0.88}_{-0.39}$ Gyr, but

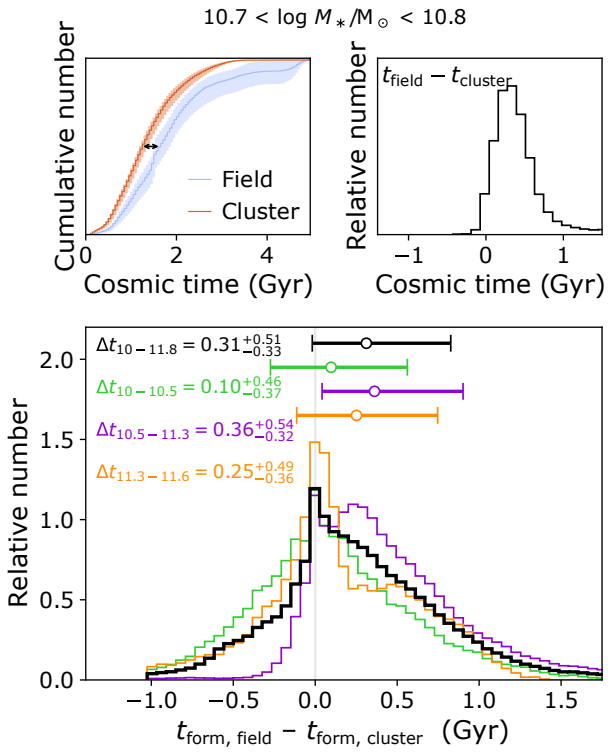

Figure 3.7: Distributions of offsets between cumulative age distributions of field and cluster galaxies (shown in Figure 3.6 as non-cumulative histograms), in units of cosmic time. The top row shows an example of this age comparison for galaxies with stellar masses $10.7 < \log\ M_*/M_\odot < 10.8$. The cumulative mass-weighted age distributions for the field (blue) and cluster (red) galaxies is shown on the top left, where the samples have been bootstrapped, and the variance is shown as a shaded region. The solid lines show the medians of the bootstrapped distributions. The corresponding offsets in the mass-weighted ages for interpolated bins spanning the cumulative distributions are shown in the top right plot. Galaxies are compared at fixed stellar mass (bins of 0.1 dex) and combined, weighted by the integrated mass within the bins. The combined distributions within the broader mass selections used throughout previous figures are included for reference: $10 < \log\ M_*/M_\odot < 10.5$, green; $10.5 < \log\ M_*/M_\odot < 11.3$, purple; $11.3 < \log\ M_*/M_\odot < 11.8$, orange. The full mass range combined distribution is shown in black. The median age difference for each mass selection is labelled in the figure, which shows the age difference is within $0.31^{+0.51}_{-0.33}$ Gyr. The median age difference is larger at lower stellar masses, and smaller for the highest mass galaxies. Error bars indicate the 68% CR.

still consistent with our main result. On the other hand, galaxies at the higher end of our redshift selection, $1.3<z<1.4$, have a slightly larger age difference: $0.39^{+0.49}_{-0.74}$ Gyr, but age consistent within the uncertainties. Figure 3.8 shows the mass-weighted age comparison for each mass and redshift selection of quiescent galaxies. We further test our result by identifying galaxies which are not necessarily passively evolving, discussed in the next Section.

3.4.3 Recent star formation

Our quiescent sample is selected based on *UVJ*-colours. We have seen in Section 3.4.1 that our *UVJ* colour selection does not yield exclusively old galaxies with exponentially declining SFRs (cyan coloured SFHs in Figure 3.5, marked with diamonds in Figure 3.6). While four galaxies have fairly flat SFHs, most of these galaxies are 'late-bloomers' with peaks in their sSFRs within the last 1 Gyr (similar to Dressler et al., 2018). These galaxies are not necessarily 'frosted' in the sense of Trager et al. 2000b, or 'rejuvenated' in the sense of Thomas et al. (2010) or Chauke et al. (2018), given that these recent peaks account for a substantial fraction of the stellar mass.

Given the breadth of the *UVJ*-colour selection of these quiescent galaxies, it is conceivable that these galaxies are still in transition (the *UVJ*-colour selection is discussed further below). In addition, some of our *UVJ*-selected galaxies show significant [O II] emission, which may be indicative of ongoing star formation. Both 'young' and [O II]-emitting galaxies are more frequent in our field sample (similar to studies at lower redshifts, *e.g.*, Treu et al., 1999, 2001; van Dokkum et al., 2001; van der Wel et al., 2004; Bernardi et al., 2006). We consider here if either population is the cause of the average mass-weighted age difference we find between field and cluster galaxies.

We identify galaxies which are not intrinsically old by the fraction of stellar mass formed within the last 1 Gyr,

$$f_{M_*<\ 1\ \mathrm{Gyr}} = \frac{\int_{t_{\mathrm{univ}}(z_{\mathrm{obs}})}^{t_{\mathrm{univ}}(z_{\mathrm{obs}})-1\ \mathrm{Gyr}} \mathrm{SFR}(t)\ \mathrm{d}t}{\int_{t_{\mathrm{univ}}(z_{\mathrm{obs}})}^{0} \mathrm{SFR}(t)\ \mathrm{d}t} \tag{3.2}$$

where we use $f_{M_*<1\ \mathrm{Gyr}} > 0.1$ as the criteria (*i.e.*, irrespective of [O II] emission). This selects 18 (5%) galaxies in our total sample, based on the median $f_{M_*<1\ \mathrm{Gyr}}$values. We note that four of these galaxies have $f_{M_*<1\ \mathrm{Gyr}} > 0.85$ and no [O II] emission, three of which are in clusters (one of which has particularly red *UVJ* colours). The spectra of these four 'young' galaxies are suggestive of recent star formation in the sense that they

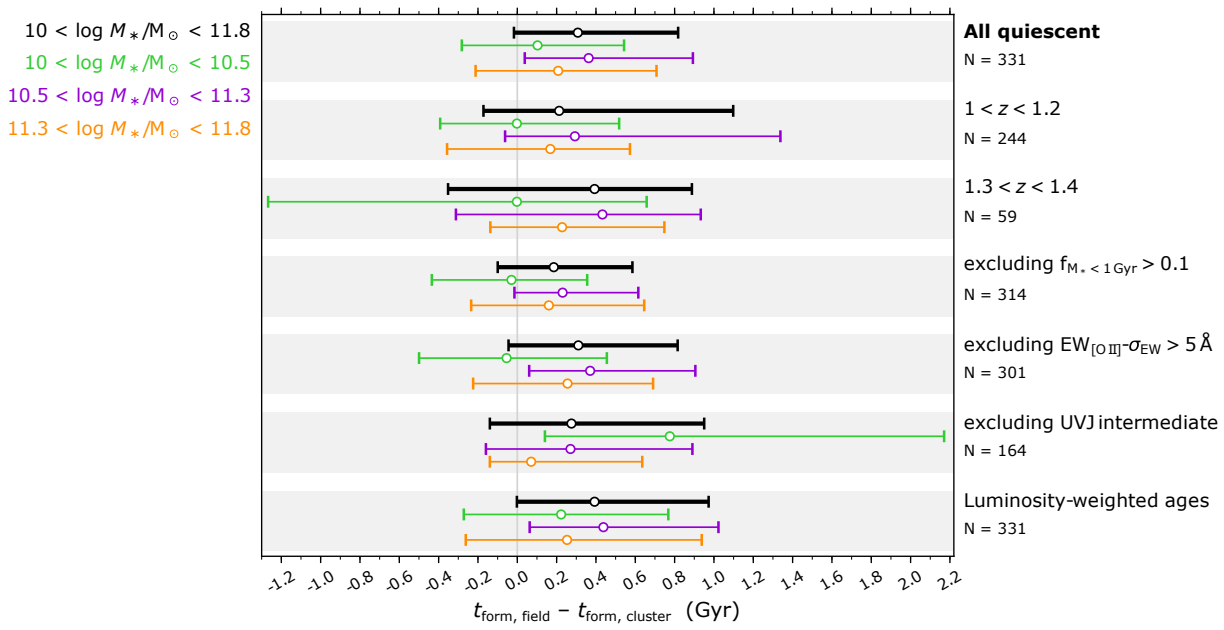

Figure 3.8: Differences in cumulative mass-weighted age distributions between field and cluster galaxies for different selections of our sample of quiescent galaxies, as described in the text, in units of cosmic time. Ages are first compared within 0.1 dex mass selections, and these comparisons are then combined, weighted by the number of posterior samplings in each selection (see Figure 3.7 for an example of this procedure). We show the comparisons in mass ranges of $10 < \log M_*/\mathrm{M}_\odot < 11.8$, black (*i.e.*, the full mass range); $10.0 < \log M_*/\mathrm{M}_\odot < 10.5$, green; $10.5 < \log M_*/\mathrm{M}_\odot < 11.3$, purple; $11.3 < \log M_*/\mathrm{M}_\odot < 11.8$, orange. Note that the lowest mass bin is below our completeness limit. The median age difference for each mass selection is marked as a circle with error bars indicating the 68% CR.

have relatively strong Balmer absorption lines, while two are particularly low S/N that their SFHs are not well constrained.

Figure 3.9 shows $f_{M_*<1\ \mathrm{Gyr}}$as a function of stellar mass, separating cluster and field galaxies in colour, and galaxies which also have [O II] emission are circled. Coloured boxes indicate the ranges of the three mass bins used throughout the paper. The number of galaxies which are 'young' by this definition are labelled in Figure 3.5 for each mass and environment subsample; 13 of these galaxies are in the field population, accounting for 16% (12%) of the lower (moderate) mass sample. Comparatively, the four 'young' galaxies in our cluster sample account for 5% (2%) of the lower (moderate) mass samples. Although the relative fractions of these galaxies are higher in the field population, the overall fractions are still quite low. Indeed, the overall median SFHs shown in Figure 3.5 are unchanged within the bootstrapped uncertainty when the 'young' galaxies are excluded.

The fraction of field galaxies in our sample with significant [O II] emission (EW([O II])-$\sigma_{\mathrm{EW}} > 5$ Å, c.f. Appendix A.1), 17% (19/109), is similarly larger than the 5% (11/222) of cluster galaxies. Moreover, as apparent in the co-added spectra discussed in Appendix A.4, the strength of [O II] emission is higher for field galaxies. Similar to our results, Rudnick et al. (2017) find that for a selection of intrinsically old galaxies the prevalence of [O II] emission was higher for field galaxies, which they attributed to clusters (and groups) being sites where gas accretion onto massive galaxies (above $10^{10.4}$ $\mathrm{M}_\odot$) was shut off. Indeed, [O II] emission can result from processes other than star formation (AGN and/or LINER, *e.g.*, Heckman, 1980; Yan et al., 2006; Singh et al., 2013), and has complex dependence on ISM properties (Hogg et al., 1998). Interestingly, the sites of [O II] emission in our sample have different mass ranges between environments: for field galaxies the [O II] emitting galaxies have masses $<10^{10.9}$ $\mathrm{M}_\odot$ for all but three galaxies, while the cluster galaxies have masses $>10^{10.9}$ $\mathrm{M}_\odot$ for all but three galaxies. We also note that only four of the $f_{M_*<1\ \mathrm{Gyr}}> 0.1$ galaxies also have [O II] emission.

Figure 3.10 shows our quiescent sample in UVJ colour space, where diamonds indicate $f_{M_*<1\ \mathrm{Gyr}}> 0.1$ galaxies, and galaxies with EW([O II])-$\sigma_{\mathrm{EW}} > 5$ Å are circled. Interestingly, and perhaps as expected, the 'young' galaxies occupy the bluer end of the UVJ colours (except one galaxy), and both the 'young' and [O II] emitting galaxies preferentially occupy the colour space closer to the boundary of the quiescent selection. This region is below the dashed line in Figure 3.10 where the U-V delimiter was increased by 0.3 dex.

We now repeat our measurement of the mass-weighted age difference between field and cluster galaxies, now excluding galaxies which are not intrinsically old. Figure 3.8 summarizes the age comparisons for these various selections of quiescent galaxies, relative to the full sample. Our result does not significantly change when excluding 'young'

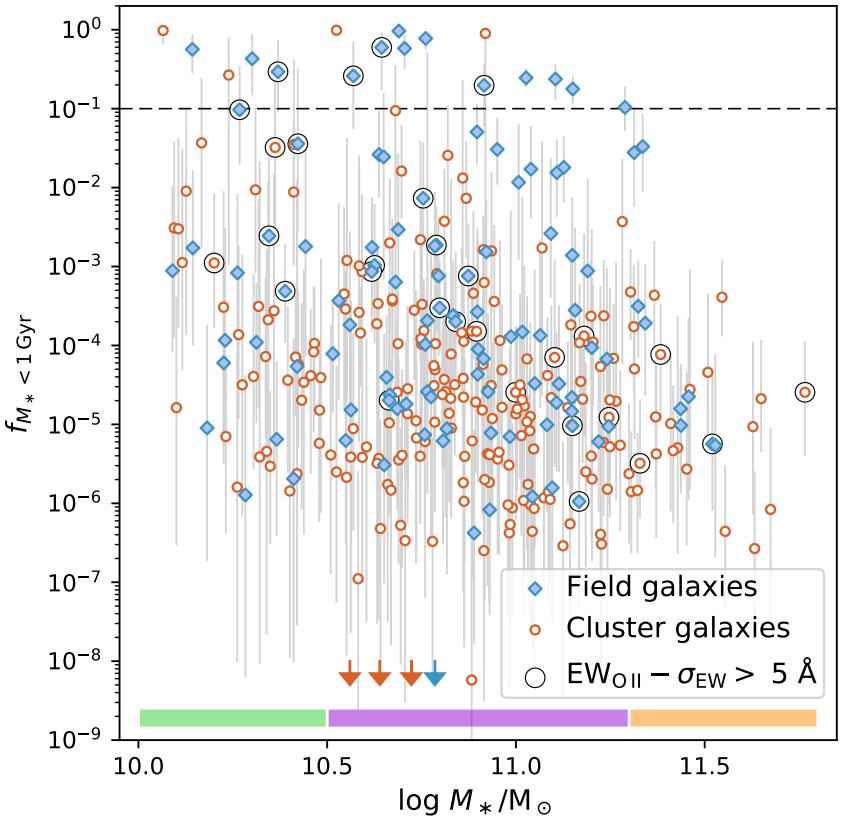

Figure 3.9: The fraction of stellar mass formed within the last 1 Gyr as a function of total stellar mass, for our UVJ-selected sample of quiescent galaxies. Galaxies with EW([O II])-$\sigma_{\mathrm{EW}} > 5$ Å are circled. Field galaxies are shown as blue diamonds, and cluster galaxies as orange circles. Arrows indicate points below the shown scale. 'young' $f_{M_*<1\ \mathrm{Gyr}} > 0.1$ galaxies are more common among field galaxies, and at stellar masses $<10^{11.3}$ $\mathrm{M}_\odot$. There is no correlation between the presence of [O II] emission and $f_{M_*<1\ \mathrm{Gyr}} > 0.1$, however.

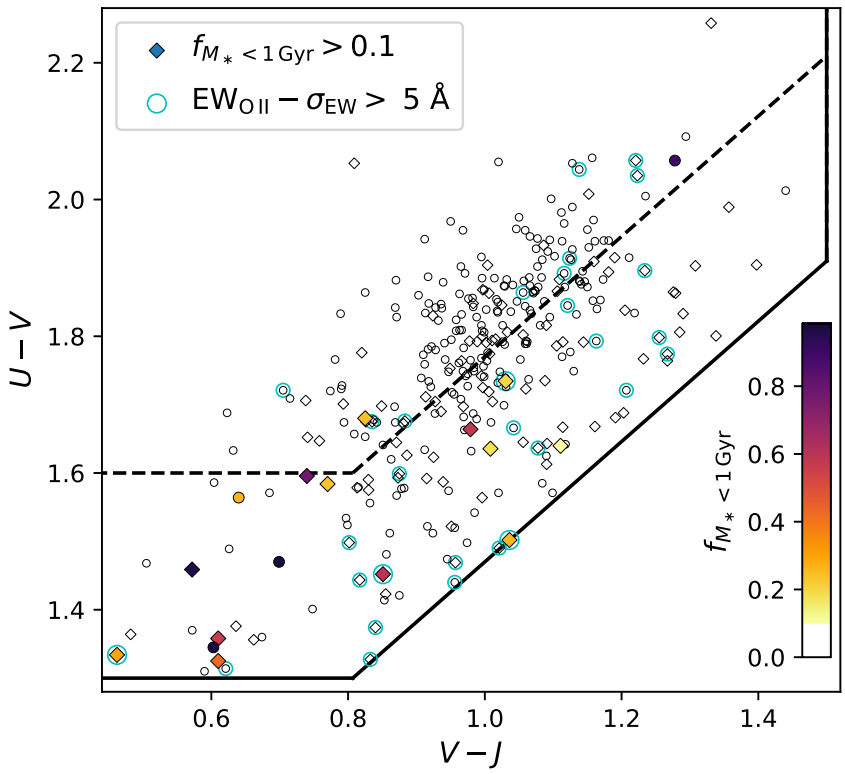

Figure 3.10: Rest-frame UVJ colours of the GOGREEN quiescent galaxies (plus marks), where $f_{M_*<1\ \mathrm{Gyr}} > 0.1$ galaxies are shown with colours according to the fraction of stellar mass formed within the last 1 Gyr, $f_{M_*<1\ \mathrm{Gyr}}$. Cluster galaxies are shown as circles, field galaxies as diamonds. Galaxies with significant [O II] emission are circled. Only a few of the $f_{M_*<1\ \mathrm{Gyr}} > 0.1$ or [O II] emitting galaxies occupy the densest region, *i.e.*, the 'red clump.' We therefore test our age comparison for galaxies in this clump, which we select by increasing the $U-V$ selection by 0.3 dex, shown as a dashed line.

($f_{M_*<1\ \mathrm{Gyr}} > 0.1$) galaxies, galaxies with [O II] emission, or galaxies near the UVJ-colour quiescent selection boundary (labelled UVJ intermediate). The latter selection would reasonably exclude galaxies transitioning between star-forming and quiescence, or which have complex dust properties obscuring star forming populations. Saying that, the largest change comes from excluding the $f_{M_*<1\ \mathrm{Gyr}} > 0.1$ galaxies, particularly at lower stellar masses. On the other hand, excluding the UVJ intermediate primarily increases the age difference between low mass galaxies, although the error bars are larger due to smaller numbers of galaxies. The exclusion of [O II] emitting galaxies does not visibly affect the age difference at all except for the lower mass galaxies, decreasing the age difference. Overall, the age difference between field and cluster galaxies is insensitive to recent star formation, unlike the low-redshift galaxies studied in Thomas et al. (2010).

The age comparison between luminosity-weighted ages is also shown, discussed in Appendix A.6, which predicts a larger (by 0.1 Gyr) age difference than mass-weighted ages, except for the highest mass galaxies. The luminosity-weighted age is more sensitive to recent star formation, so it is not unexpected that there is a mass dependence between t_{mass} and t_{light}related to the mass-dependent SFHs.

3.5 Discussion

The main goal of this work is to compare quiescent galaxies in average density (field) and high density (galaxy cluster) environments, while accounting for any differences related to their stellar mass. We now discuss our result that the age difference is within $0.31^{+0.51}_{-0.33}$ Gyr in the context of the literature. In Section 3.5.1, we discuss that our SFHs are consistent with 'mass-dependent evolution,' and the environmental dependence of the SFHs. In Section 3.5.2, we compare the mass-weighted age measurements to similar results from the literature, and discuss the difference between mass-weighted ages as a function of environment. We then place the measured age difference in the context of two simple quenching models in Section 3.5.3, where environmental quenching is purely related to the time since infall, or where there is no environmental quenching but the field population forms later than cluster galaxies.

3.5.1 Mass-dependent evolution of quiescent galaxies

Decades of work has shown that the bulk of star formation in massive ETGs occurred at high redshifts, and these galaxies have been passively evolving since. Studies connecting intermediate-redshift and local observations of the colour-magnitude relations (*e.g.*,

Dressler, 1980; Ellis et al., 1997; Stanford et al., 1998), the evolution of the luminosity function (*e.g.*, De Propris et al., 1999; Toft et al., 2004), the Fundamental Plane (*e.g.*, van Dokkum et al., 1998; Kelson et al., 2000; Cimatti et al., 2006; di Serego Alighieri et al., 2006a,b; Jørgensen et al., 2006, 2007; Beifiori et al., 2017; Woodrum et al., 2017; Saracco et al., 2020), and absorption lines (e.g, Bender et al., 1996; Kelson et al., 2001; Sánchez-Blázquez et al., 2009) suggest that ETGs have been evolving passively since $z \sim 2$–3 (see Renzini 2006 for a review). As large scale surveys became available (*e.g.*, SDSS; York et al., 2000), trends between the star formation histories and galaxy properties have increasingly been explored. A robust finding is that more massive galaxies form their stellar mass earlier and over shorter time scales than lower mass galaxies (*e.g.*, Gallazzi et al., 2014; Heavens et al., 2004; Jimenez et al., 2005; Nelan et al., 2005; Thomas et al., 2005, 2010, 2017; Sánchez-Blázquez et al., 2009), *i.e.*, 'mass-dependent evolution.' This is similar to the concept of 'downsizing' in the sense that there is mass-dependent decline in the SFRs of galaxies with time (*e.g.*, Cowie et al., 1996; Bell et al., 2005; Juneau et al., 2005), or in the growth of the stellar mass function (*e.g.*, Cimatti et al., 2006; Leitner, 2012). This downsizing trend can be explained by the fact that more low mass galaxies are continuously (over time) being added to the quiescent population (Brammer et al., 2011; Muzzin et al., 2013b; Tomczak et al., 2014). On the other hand, merger rates are mass dependent (Khochfar & Silk, 2009; Emsellem et al., 2011), and late time rejuvenated star formation is more common in low mass galaxies (Poggianti et al., 2008, 2009; Thomas et al., 2010; Belli et al., 2015).

The mass-dependence of SFHs has been confirmed at higher redshifts, where age indicators are more sensitive to older stellar populations. However, observations beyond $z \sim 1$ are challenging. As a result, studies have been mainly limited to surveys of massive galaxies with small samples (*e.g.*, van Dokkum & Brammer, 2010; Toft et al., 2012; van de Sande et al., 2013; Kriek et al., 2016; Belli et al., 2015, 2019; Estrada-Carpenter et al., 2019; Saracco et al., 2020), and rely on averaging photometric SFHs (*e.g.*, Rettura et al., 2011; Snyder et al., 2012; Strazzullo et al., 2013; Pacifici et al., 2016a; Iyer & Gawiser, 2017) or combining spectra (*e.g.*, Gobat et al., 2008; Tanaka et al., 2013; Whitaker et al., 2013; Choi et al., 2014; Onodera et al., 2015). Only recently have large, high-redshift spectroscopic surveys been completed which allow more precise age estimates of individual galaxies. Notably, Chauke et al. (2018) combine high resolution spectroscopy and photometry for more than 600 galaxies at $0.6<z<1$ from LEGA-C (van der Wel et al., 2016; Straatman et al., 2018) to show that galaxies with higher stellar velocity dispersions formed both earlier and faster, and that the majority of quiescent galaxies evolve passively since their main star forming epoch. Carnall et al. (2019a) likewise use the VANDALS survey (McLure et al., 2018; Pentericci et al., 2018) to determine the SFHs for 75 massive quiescent galaxies at

$1<z<1.3$, finding a trend between the average formation times of galaxies and their stellar mass of $1.48^{+0.34}_{-0.39}$ Gyr per dex for $M_*<10^{11}$ $M_\odot$.

The mass-dependent evolution in the GOGREEN quiescent galaxies is apparent in Figure 3.5, where we find the more massive galaxies to have sSFRs which are higher at earlier times, and decline at earlier times, than the lower mass galaxies, at fixed environment. The median mass-weighted ages are shown in Figure 3.6 relative to stellar mass, where contours show the combined posteriors. While we see that the ages of lower mass galaxies are younger *on average*, this appears to be driven by the fact that there is a broader distribution of ages among the lower mass galaxies. Indeed, our trend between age and mass in our field sample is both flatter and offset towards older ages than found by Carnall et al. (2019a, and references therein).

The $\sim$1 Gyr offset in ages could be a result of differences in fitting procedure, where Leja et al. (2019b) report that Prospector-α[12] predicts older ages and higher stellar masses than standard parametric modelling. Carnall et al. (2019a) use a double-power-law form for their SFHs, however, which is more flexible than fiducial declining-exponential models, so the ages should be more similar than those reported by Leja et al. (2019b). Along the same lines, Forrest et al. (2020b) reconstruct the SFHs of ultra massive ($>10^{11}$ $M_\odot$) galaxies in protoclusters at $z>3$ and find that the bulk of star formation occurred between $4<z<5$ (only $\sim$0.5 Gyr later than the median mass-weighted ages we measure), and the galaxies had quenched by $z\sim3$. Besides the difference in parameterization of the SFHs, an alternative explanation for the older ages we find is that it is a result of a lower metallicity in the best-fit model, since metallicities are strongly degenerate with ages. In fact, as discussed in Appendix A.2, our metallicities are systematically lower than other studies at intermediate redshifts (*e.g.*, Choi et al., 2014; Estrada-Carpenter et al., 2019; Morishita et al., 2019). Indeed, Forrest et al. (2020b) assume solar metallicity. An increase in metallicity by a factor of three (i.e, +0.5 dex) would decrease the mass-weighted age by $\sim$0.5 Gyr, which would account for most of the age difference.

There is a stronger age difference between lower and higher mass galaxies at fixed environment, than between environments at fixed mass – despite the fact that we find a flatter mass-dependence of the SFHs than other studies. For both the cluster and the field populations, the median difference in mass-weighted ages is $\sim0.7^{+0.3}_{-0.6}$ Gyr between galaxies of mass 10^{10}–$10^{10.5}$ $M_\odot$ and $10^{11.3}$–$10^{11.8}$ $M_\odot$, while the age differences between environments are $<$0.4 Gyr (see Figure 3.8), and are discussed further in the next Section. This result is consistent with the results of Saglia et al. (2010) and Woodrum et al. (2017), where both

[12]Prospector-α uses the Prospector framework, but includes additional parameters (such as dust emission, nebular emission, AGN emission).

measured the evolution of the M/L ratio between cluster and field galaxies at $z \lesssim 0.9$ and $z<1.2$, respectively, and found stronger differences between galaxies of different stellar mass than between environments. Similarly, Raichoor et al. (2011) compared ETGs at $z \sim 1.3$ to conclude that the age difference between galaxies in cluster and field environments was less significant than between galaxies of different mass.

3.5.2 Environment-dependent evolution of quiescent galaxies

A number of recent studies find that field galaxies form over longer timescales than cluster galaxies, however, the exact timescales have been challenging to robustly quantify. Line strength studies of early type galaxies (ETGs) at low redshifts find that star formation in low density environments is delayed by 1–2 Gyr (*e.g.*, Bernardi et al., 1998; Balogh et al., 1999; Thomas et al., 2005; Clemens et al., 2006; Sánchez-Blázquez et al., 2006, see also the review by Renzini 2006). Using the Fundamental Plane, the evolution of the M/L between galaxies at $z \lesssim 1.2$ has shown that the slopes are steeper for galaxies in cluster environments, indicating that they formed at slightly higher redshifts than field galaxies (*e.g.*, van Dokkum & van der Marel, 2007; Saglia et al., 2010; Woodrum et al., 2017). The M/L evolution can be interpreted as SFHs with models of simple stellar populations (SSPs), taking into account the structural evolution in the size of galaxies (and progenitor bias). van Dokkum & van der Marel (2007) infer that massive galaxies in clusters are ∼0.4 Gyr older than field galaxies, Saglia et al. (2010) estimate a ∼1.6 Gyr age difference, while Woodrum et al. (2017) estimate $\lesssim$ 0.3 Gyr difference for galaxies with low-velocity dispersions but $\lesssim$1 Gyr for high-velocity dispersions. Compared to luminosity-weighted ages derived from Balmer absorption lines, Saglia et al. (2010) find consistent age estimates within their large uncertainties, while Woodrum et al. (2017) find a larger age difference of 1–3 Gyr.

Measuring age differences at low redshifts does not necessarily reflect differences in star formation histories at early times, however. Late-stage environmental effects on galaxy evolution (*e.g.*, Thomas et al., 2010), or progenitor effects, can obscure estimates of the ages of the oldest stellar populations; recent star formation can 'outshine' older stars, making age estimates from the integrated light difficult (Papovich et al., 2001). Moreover, the population of ETGs has been in place since $z \sim 2$ (*e.g.*, Bernardi et al., 1998; van Dokkum et al., 2010). The old ages of these galaxies presents a challenge because of the nonlinear evolution of stellar spectra; over time, the ages of galaxies are increasingly difficult to distinguish. In order to explore whether environmental factors affected galaxy formation during the period where the galaxies assembled the majority of their mass requires higher redshift observations.

At $z \sim 1.2$, Gobat et al. (2008) measured the ages of ETGs in a massive cluster relative to galaxies in the GOODS/CDF-S survey via SED fitting photometry and coadded spectroscopy, finding that cluster galaxies formed ∼0.5 Gyr before field galaxies (particularly at $<10^{11}$ $M_\odot$). On the other hand, for the same cluster Rettura et al. (2010) independently compare the massive ETGs with equivalent galaxies in the GOODS survey, measuring ages from fitting photometry to SEDs (without spectroscopy), and conclude that there is no significant delay in formation epochs between the two environments within the typical uncertainty of ∼0.5 Gyr. Two additional clusters are included in the comparison by Rettura et al. (2011), where again no difference was found in formation times within their average uncertainty, 0.5 Gyr. At $z \sim 1.3$, Saracco et al. (2017) compare the median luminosity-weighted ages of elliptical galaxies in three clusters relative to the GOODS, COSMOS, and CANDELS fields. While they find that the structural properties of galaxies in cluster and field environments are consistent at fixed mass, and $<10^{11}$ $M_\odot$, massive galaxies either assemble ∼0.3 Gyr earlier or assemble more efficiently in clusters.

Our results are fully consistent with these studies. We find that cluster galaxies are *on average* $0.31^{+0.51}_{-0.33}$ Gyr older than field galaxies, at fixed stellar mass. While the age difference is largest for galaxies of masses $10^{10.5}$–$10^{11.3}$ $M_\odot$, the age difference is positive (although sometimes consistent with zero) for all mass ranges. This result is robust when carefully removing galaxies which show recent star formation, [O II] emission, or *UVJ* colours outside the red clump (see Section 3.4.3).

Muzzin et al. (2012) compare $D_n(4000)$ values, as a proxy for stellar age, for quiescent[13] galaxies in the GCLASS survey; a subset of these clusters, and galaxy spectra, are included in GOGREEN. At fixed stellar mass, they find that $D_n(4000)$ is independent of environment except perhaps for their lowest mass galaxies $<10^{10}$ $M_\odot$. We compare the $D_n(4000)$ of our spectra relative to Muzzin et al. (2012) in Appendix A.4, where we find modestly larger differences between environments, consistent with our result of a small positive age difference. The GCLASS sample is dominated by $z \sim 0.8$ clusters, however, particularly at low stellar masses. Thus, the small difference we observe may be a result of evolution.

An important consideration when comparing to results from the literature is how the lower density sample is defined. Some studies separate galaxies in the cores and outskirts of clusters, or in higher- and lower-density regions within their sample, or carefully select for galaxies in clusters, groups, or in isolation. Our field sample is selected from the distant fore- and background of our clusters, and is therefore expected to be representative of an average patch of the Universe. Comparing galaxies in clusters with those truly isolated in

[13] Muzzin et al. (2012) select quiescent galaxies based on the lack of [O II] emission, rather than *UVJ* colours. See Appendix A.1 for a comparison of these selections.

cosmic voids, or exclusively galaxies central to their halo, will likely have a larger contrast in properties than our results. Importantly, the 'field' environment may be different at different mass scales; for example, more massive galaxies could be more likely to exist in cosmic overdensities (*e.g.*, groups) than lower mass galaxies. Therefore, the comparison between galaxies of lower stellar masses could reflect different physical factors than between galaxies of higher mass. We leave a comparison of galaxies between different local densities to a future paper.

A second consideration is the selection of quiescent galaxies: several studies classify quiescent galaxies based on morphology, or other star formation tracers than UVJ colours. We do not expect this to significantly impact the relative age measurements, however, as long as the selection is consistent between environments. Saracco et al. (2017) find that at $z \sim 1.3$ elliptical galaxies have consistent structure and properties between field and cluster environments, however there are fewer large and massive elliptical galaxies in the field relative to clusters. Such differences between galaxy properties and environment could be important to the quiescent-selection in detail.

3.5.3 Toy models of cluster galaxy evolution

It is well established that at low redshifts the fraction of quiescent galaxies is higher in denser environments (*e.g.*, Baldry et al., 2006). Several studies also find a higher fraction of low-mass quiescent galaxies in denser environments (*e.g.*, Muzzin et al., 2012; Woo et al., 2013). Peng et al. (2010) suggested that these two observations are consistent if galaxies in dense environments are subject to extra 'environmental-quenching' which is independent of stellar mass, in addition to mass-dependent 'self-quenching.'

At $z \gtrsim 1$ the situation is very different. While there is still an excess of quiescent galaxies in dense environments, the SMFs of quiescent galaxies are consistent between low- and high-density environments (Nantais et al., 2016; van der Burg et al., 2020). Moreover, the shapes of the SMFs for *star forming* galaxies are also the same between cluster and field. We add to this picture the fact that there is a small, positive age difference between quiescent cluster and field galaxies. This is difficult to reconcile with the hypothesis that the higher fractions of quenched galaxies in galaxy clusters at this epoch result from the transformation of recently accreted, star-forming galaxies.

We first consider whether a simple infall-based quenching model can be simultaneously compatible with both our measured average age difference, and the quenched fractions in cluster and field environments measured by van der Burg et al. (2020). We then consider an

alternative model where cluster galaxies formed earlier than field galaxies, and infall-driven quenching is negligible.

In order to determine the mass-weighted age evolution, we need a prediction of the average SFH of star forming galaxies; we assume the SFRs evolve as defined in Schreiber et al. (2015), and that the SF is truncated when the galaxy is 'quenched.' We compare galaxies with final stellar masses between $10^{9.5}$-$10^{11.5}$ $M_{\odot}$. As we are only interested in modelling the global properties of 'average' galaxies, we ignore any mass dependence in the data. Therefore, we model the self-quenching efficiency using the same form as proposed by Peng et al. (2010) (*i.e.*, $\eta \propto \mathrm{SFR}/M^*$) using the SFR for an $M^* = 10^{10.8}$ $M_{\odot}$ galaxy. Given that van der Burg et al. (2020) find the SMFs between star forming cluster and field populations to have the same shape, we require that the star forming SMFs in our model similarly do not evolve. Our toy model consists of tracking the number of star forming and quiescent galaxies from $z = 10$ (when cluster galaxies are assumed to form) to $z = 1.2$.[14]

Post-infall environmental quenching and pre-processing

For the infall-based quenching model we assume that all galaxies are subject to self-quenching, while in addition star forming galaxies that join clusters quench at a given time after infall (t_{delay}). The infall rate we assume follows the predictions of McGee et al. (2009) for timescales of galaxies becoming satellites of larger haloes ($>10^{13}$ $M_{\odot}$), based on the Millennium simulation (Springel, 2005) with additional prescriptions for halo assembly via merger trees (Helly et al., 2003; Harker et al., 2006) and using the semi-analytic models of Bower et al. (2006)[15]. This predicts that the rate at which galaxies join larger haloes is effectively constant in time. There are then two parameters in this model which determine the relative populations of star forming/quiescent and field/cluster galaxies: the normalization of the self-quenching efficiency, and t_{delay}. Both of these parameters are constrained by observations of the quenched fractions at our fiducial stellar mass, measured for the GOGREEN sample to be $f_{\mathrm{Q,\ field}}(z = 1.2) = 0.3$ and $f_{\mathrm{Q,\ cluster}}(z = 1.2) = 0.65$ (van der Burg et al., 2020). The self-quenching efficiency drives the quenched fraction in the field, while the delay time determines the additional quenching in clusters. We find that a delay time of $t_{\mathrm{delay}} \sim 2.4$ Gyr is required to match the observed quenched fractions. This is somewhat longer than expected from dynamical timescales at this redshift (*e.g.*, Balogh et al., 2016); we caution that our toy model is merely illustrative (we ignore mass

[14]This toy model is qualitatively different from the mass-quenching model proposed by Peng et al. (2010), or as implemented by van der Burg et al. (2020). Furthermore, we neglect mergers. Including mergers, however, would only enhance the different galaxy properties between cluster and field environments.

[15]With updated modelling of strangulation, as per Font et al. (2008).

dependence and the mass-quenching rate is somewhat arbitrary), and this discrepancy does not significantly affect our conclusions here.

An important consequence of post-infall environmental quenching models is that quiescent galaxies in the field would be on average *older* than quiescent galaxies in clusters at fixed mass (by $1.5^{+1.3}_{-0.2}$ Gyr given $t_{\rm delay}$and the quenched fractions listed above). This is because the rate that recently-quenched galaxies are added to the quiescent population is higher at later times in the cluster, such that the overall population is younger. As we have constructed our model, environmental-quenching is stronger at later times (*e.g.*, Nantais et al., 2017), while self-quenching dominates at early times. Muzzin et al. (2012) come to a similar conclusion, modelling the evolution of $D_n(4000)$ for early-self-quenching late-environmental-quenching dominated efficiencies. Given that we find a small, but significant, average age difference between field and cluster galaxies in the opposite sense, we can exclude this model even for delay times approaching the age of the universe.

One important simplification of this infall-based quenching model is that we have neglected the role of pre-possessing in the field population. That is, galaxies which quenched in locally overdense clumps (*i.e.*, groups or filaments) prior to joining clusters (*e.g.*, Dressler, 1980; Fujita, 2004; Moran et al., 2007). The infall rate we use predicts the number of galaxies which become satellites of haloes with masses $>10^{13}$ $M_\odot$ within a given time, and we have considered all such structures 'clusters' when realistically some fraction makes up the 'field.' Secondly, some fraction of these pre-processed groups will eventually accrete onto clusters. McGee et al. (2009)'s model predicts that at $z = 1.5$ around 20% of galaxies were in haloes of mass 10^{13}–10^{14} $M_\odot$ h^{-1} prior to becoming a member of their final halo. Along the same lines, Poggianti et al. (2006) use the fraction of cluster galaxies with [O II] emission to constrain the fraction of galaxies which were 'primordially quenched' at high redshifts, or experienced environmental quenching in haloes above 10^{14} $M_\odot$. They find that $z \sim 1.5$ marks a turnover between these two populations, where only galaxies in haloes with high velocity dispersions have appreciable numbers of 'quenched' galaxies. De Lucia et al. (2012) build on the model of McGee et al. (2009) to show that the accretion history of satellites onto clusters is stellar mass dependent, where lower stellar mass galaxies are more likely to be satellites of a smaller structure when joining a cluster. Moreover, if the groups which accrete onto clusters represent a biased sample (*e.g.*, the oldest groups) this would make the cluster quenched population older on average, and the field younger. In this scenario, it is no longer clear that the field quenched population is necessarily older than quiescent cluster galaxies. The exact age differences are difficult to predict, however, as they depend on the distribution of galaxies in groups between field and cluster environments. Lastly, even considering the fact that some galaxies may be part of smaller substructures prior to joining clusters, whether they are quenched in such environments

likely depends on the halo mass and how long they have been satellites. If the dominant environmental-quenching processes are only relevant over long timescales, the effect of pre-processing at high redshifts may not be significant. We leave a more complete analysis to a future paper.

Delayed formation of field galaxies

Motivated by this challenge for the simplistic post-infall quenching model to explain our results, we now turn to a model where the self-quenching of cluster galaxies gets a head start relative to the average field. Figure 3.11 illustrates this toy model of delayed formation times between cluster and field galaxies. Here the only quenching is self-quenching, which is shown in the top row for cluster galaxies (thin black line, starting at $z = 10$) normalized such that $f_{\mathrm{Q,\ cluster}}(z = 1.2) = 0.65$, and for the field galaxies (coloured dashed lines). Galaxies in the two environments quench through the same processes; however, field galaxies form and quench later, starting at a time offset from the cluster, labelled as t_Δ[16]. For the four delay times, (t_Δ= 0.25 Gyr, 0.5 Gyr, 1 Gyr, and 2 Gyr), the difference in quenched fractions between cluster and field galaxies, Δf_{Q}, and the median cumulative mass-weighted age difference, Δt_{mass}, are calculated and shown in the second and third plots, respectively. The error bars on Δt_{mass} correspond to the 68% CR of the age comparison, and the grey region indicates the measured age difference discussed in Section 3.4.2.

Figure 3.11 shows that any delay time $\gtrsim$1 Gyr would result in a mass-weighted age difference that is excluded by our data. To match the observed $\Delta f_{\mathrm{Q}} \sim 0.35$, however, would require $t_\Delta \sim$1.75 Gyr in our simple model. This is larger than the delay in formation time predicted to match the quenched fractions in the toy model of van der Burg et al. (2020) of $\sim$1 Gyr, which likely is due to different assumptions of the growth of the SMF and mass-dependent self-quenching. In either model, such a long formation delay time would result in a mass-weighted aged difference of >1 Gyr, strongly excluded by our observations.

Neither the simple post-infall environment quenching model nor the delayed-formation model can fully explain the difference in galaxy properties between high and low density environments at $z = 1$. In principle, a combination of the two models can, even without pre-processing. For example, with a delay between infall and environmental quenching of $\sim$2.8 Gyr (which is still long), and delaying the formation of field galaxies by 1 Gyr relative

[16]We note that in the simple model of Peng et al. (2010) (their Sec 6) the formation of field galaxies is delayed $\sim$1 Gyr ($z_{\mathrm{form}} = 4$)

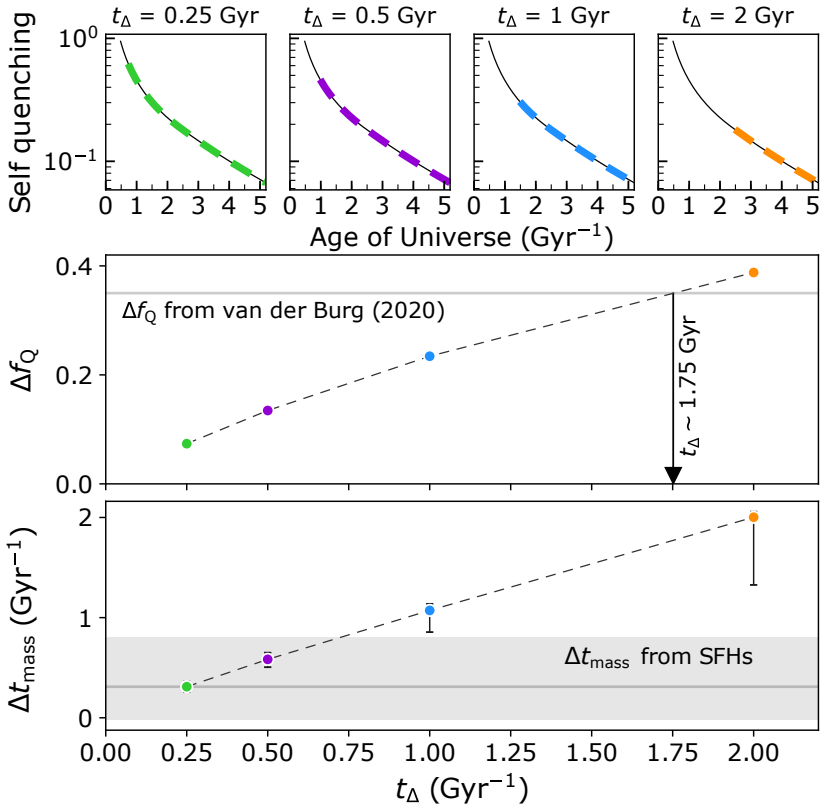

Figure 3.11: Toy model of the expected difference in quenched fractions and $t_{\rm mass}$ given an offset in the formation of the field population of t_Δ. Top: quenching rate for four values of t_Δ: 0.25 Gyr, 0.5 Gyr, 1 Gyr, and 2 Gyr. The cluster quenching rate is shown as a black line starting at $z_{\rm form} = 10$, and field quenching rate as a coloured (according to t_Δ) dashed line. Middle: The difference in quenched fractions, $\Delta f_{\rm Q}$, for fixed stellar mass at $z = 1.2$, for the four $t_\Delta$models shown. Larger offsets in the formation of field galaxies corresponds to larger $\Delta f_{\rm Q}$. A horizontal line indicates the measured difference in quenched fractions from van der Burg et al. (2020). Bottom: Average difference in cumulative $t_{\rm mass}$ distributions, $\Delta t_{\rm mass}$, between field and cluster galaxies, with error bars indicating the 68% CR. The grey shaded region indicates the measured average age difference, see Section 3.4.2. Larger offsets in the formation of field galaxies corresponds to larger $\Delta t_{\rm mass}$. In the context of this simple model, t_Δ<0.75 Gyr is consistent with our observations, but is inconsistent with the time derived by the difference in quenched fractions.

to cluster galaxies, this simple model can simultaneously match both the observed quenched fraction and age difference. Before concluding that such a hybrid model is successful, however, it will be important to test the stellar mass and halo-mass dependence of the predictions. We leave this work to a future paper.

3.6 Conclusions

In this work, we determined the SFHs for 331 quiescent galaxies in 11 GOGREEN clusters and field galaxies at $1<z<1.5$ based on rest-frame optical spectroscopy and multi-band photometry fit to SED templates with the Bayesian inference code Prospector. The following summarizes our comparison of the quiescent field and cluster galaxies:

1. Comparing SFHs between galaxies of different mass, we found that more massive galaxies form earlier, and over shorter timescales, than lower mass galaxies (see Section 3.4.1 and the bottom row of Figure 3.5). This picture is consistent with the 'mass-dependent evolution' scenario. Comparing SFHs between galaxies in cluster and field environments, we conclude that below $10^{11.3}$ $M_\odot$ the SFRs declined earlier and more rapidly for galaxies in denser environments, at fixed mass (see the right-hand column of Figure 3.5).

2. From the SFHs, we calculate posteriors for mass-weighted ages for each galaxy, shown in Figure 3.6 relative to stellar mass. Overall, 90% of all galaxies have formed half their stellar mass by $z \sim 2.2$. The majority (>50%) of galaxies with masses $10^{11.3}$–$10^{11.8}$ $M_\odot$ have formed at least half of their stellar mass by $z \sim 5.4$, while the same is true for galaxies with masses 10^{10}–$10^{10.5}$ $M_\odot$ at $z \sim 3.4$. The formation times we estimate are older than similar ages in the literature; this may be a consequence of the age-metallicity degeneracy and the fact that our fits prefer somewhat lower metallicities than other studies (see discussion in Sec 3.5.2). Future telescopes, such as the James Webb Space Telescope (JWST), can directly observe galaxies at these redshifts and will be able to confirm our predictions.

3. We compare the mass-weighted ages for galaxies of similar stellar mass between the two environments (see Figure 3.7). The distribution of ages for field galaxies is broader than for cluster galaxies, where the field population has a higher relative fraction of young galaxies. As a result, the mass-weighted age difference between field and cluster galaxies with stellar masses between 10^{10}–$10^{11.8}$ $M_\odot$ is within $0.31^{+0.51}_{-0.33}$ Gyr, in the sense that cluster galaxies are older on average. This result holds when we

exclude galaxies which have formed more than 10% of the stellar mass within the last 1 Gyr, have significant [O II] emission, or have UVJ colours closer to the star formation population (see discussion in Sec 3.4.3).

4. We consider two simple models consistent with the higher fraction of quenched galaxies in clusters, and show neither one is consistent with our age measurements. If the environmentally-quenched population is built up entirely through post-infall quenching processes (without pre-processing), we predict field galaxies would be *older* than cluster galaxies – in contrast to our results. On the other hand, if quenching in cluster environments gets a head start, this needs to be >1 Gyr to explain the difference in quenched fraction, which results in an average mass-weighted age difference that is much larger than we observe.

This work builds on previous evidence (Balogh et al., 2016; van der Burg et al., 2020) that the substantial quenched population in galaxy clusters at z>1 has been built up in a fundamentally different way from clusters at $z = 0$. In particular, the infall-based environmental quenching models of Peng et al. (2010), Wetzel et al. (2012), and others, that are so successful at matching local observations, are not able to account for the properties of the GOGREEN cluster sample. The data seem to require that a substantial population of protocluster galaxies are quenched at early times, z>3, via a process which is accelerated but otherwise indistinguishable from the mass-quenching that affects all galaxies. Evidence of quenched populations of massive galaxies at this epoch is growing (*e.g.*, Glazebrook et al., 2017; Schreiber et al., 2018a; Tanaka et al., 2019; Forrest et al., 2020a; Forrest et al., 2020b; Tanaka et al., 2020; Valentino et al., 2020). Environmental quenching must still play a role, but it may only become dominant at z<1 (although McConachie et al. 2022 find an excess of quenched galaxies in protoclusters relative to the field at $z \sim 3.3$, which they interpret as evidence of environmental quenching). In future work, we will use the stellar-mass and halo-mass dependence of these observations to further constrain these toy models; comparison with simulations and semianalytic models will be important to help identify the physical origin of the quenching mechanisms postulated here. Finally, these data indicate that much of the quenching activity responsible for building up galaxy clusters occurred in the protocluster environment at z>3; data from JWST will be crucial for understanding the nature and cause of this phenomenon.

Data Availability

The data underlying this article are available in Zenodo, at https://dx.doi.org/10.5281/zenodo.3942064.

Chapter 4

Still at Odds with Conventional Galaxy Evolution: The Star Formation History of Ultra-Diffuse Galaxy Dragonfly 44

4.1 Introduction

Matching predictions to observations of how, and when, galaxies assemble serves as an important test for our greater understanding of cosmology and baryonic physics. Modern theories that suggest galaxy evolution is determined by the growth of their dark matter haloes, as well as the regulation of their gas processes (*i.e.*, infall and star formation histories; *e.g.*, White & Frenk 1991; Schaye et al. 2010; Davé et al. 2012; Wechsler & Tinker 2018), have successfully replicated some observed relations between galaxy properties – for example, the tight connection between stellar mass and halo mass (*i.e.*, the SMHM relation; Moster et al. 2010). A number of outstanding issues remain, however. A particularly challenging problem is explaining the increasing number of galaxies that cease forming stars (*i.e.*, 'quench') over time (Renzini, 2006; Faber et al., 2007). While simulations correctly predict scaling relations for massive galaxies (*e.g.*, the mass–metallicity relation; MZR, and star formation main sequence), there are still fundamental discrepancies at lower stellar masses.

In the low mass regime, observations have shown that quenched galaxies associated with massive host haloes are rare (Geha et al., 2012), such that quenching at $z < 1$ is

thought to predominantly be a result of environmental effects (*e.g.*, Boselli & Gavazzi, 2006; Fillingham et al., 2018; Mao et al., 2021). Rather than remain quenched, recent studies instead suggest that isolated quiescent dwarfs may in fact oscillate between 'star forming' and 'quenched' states (*e.g.*, Polzin et al., 2021). Yet cosmological simulations typically over-predict the abundance of quiescent field dwarfs (*e.g.*, Dickey et al., 2021).

The recently discovered ultra-diffuse galaxies (UDGs) potentially exemplify our limited understanding of the true diversity of galaxy evolution and quenching. UDGs were initially noted for their surprisingly large sizes given their low surface brightnesses ($R_{\rm eff} \geq 1.5$ kpc and $\mu_0(g) \geq 24$ mag arcsec^{-2}; van Dokkum et al. 2015) which, along with their red colours, distinguished them from classical low surface brightness (LSB) galaxies (*e.g.*, Dalcanton et al., 1997).

Several current cosmological model predictions suggest that conventional processes can explain the UDG population, thus maintaining standard dark matter halo occupancy relations (*e.g.*, Tremmel et al., 2020). Such models typically focus on the mechanisms which increase the size of otherwise canonical dwarf galaxies to make them 'ultra-diffuse' (for a summary of UDG origins, see Jiang et al. 2019a). Simulations have shown that unusual star formation or galaxy evolution processes can 'puff up' canonical dwarfs (*e.g.*, high-spin scenarios, Amorisco & Loeb 2016; Rong et al. 2017; energetic star formation feedback, Di Cintio et al. 2017; Chan et al. 2018; Jackson et al. 2021) or dynamically redistribute their stellar populations (*e.g.*, tidal heating and/or stripping; Jiang et al. 2019a; Liao et al. 2019; Carleton et al. 2019; Sales et al. 2020). Alternatively, UDGs may represent the tail of galaxy evolution processes, such that only minor differences in their evolution (*e.g.*, when they infall or have major mergers) distinguish their final properties from normal dwarfs (*e.g.*, Tremmel et al., 2020; Wright et al., 2021).

Despite these differences, nearly all models rely on environmental processes to explain the lack of star formation in the subset of UDGs that are quiescent (*e.g.*, via ram pressure stripping; Yozin & Bekki, 2015; Rong et al., 2017; Chan et al., 2018; Tremmel et al., 2020). Accordingly, all the scenarios follow a dichotomy related to when UDGs infall into a cluster environment: whether the proto-UDGs surpassed the size-threshold prior-to or post infall, is tied to whether they infall 'late' or 'early.' While UDGs are found both in the field and clusters, those that are quiescent are usually located in clusters (the few exceptions may be on backsplash orbits; *e.g.*, Papastergis et al., 2017; Benavides et al., 2021). Explaining the origin of UDGs and the diversity of their properties in the context of their environments remains a key question in understanding galaxy formation and evolution.

Testing the predicted UDG properties (*e.g.*, kinematics, Amorisco & Loeb 2016; stellar populations, Rong et al. 2017; Ferré-Mateu et al. 2018; globular cluster (GC) properties,

Carleton et al. 2021; infall versus quenching times, Gannon et al. 2022) from these scenarios against the observed properties, however, has revealed a number of discrepancies. And while some UDGs are found with very large sizes ($R_{\rm eff} > 4.5$ kpc), these exotic objects are beyond the predictions of most models (Di Cintio et al., 2017; Carleton et al., 2019). Along the same lines, models which accurately predict the distribution of UDG sizes fail to reproduce the distribution of sizes among normal dwarfs (*e.g.*, Rong et al., 2017; Jiang et al., 2019a; Tremmel et al., 2020).

On the other hand, van Dokkum et al. (2015) proposed that some UDGs originate similar to today's massive galaxies (and have sizes reflecting their massive haloes), but lost their gas early in their histories. As a result of their early quenching, these 'failed' galaxies did not build up the stellar mass expected for their haloes. This scenario deviates from the expected galaxy–halo connection, in that either these failed galaxies do not follow the SMHM relation or at least have a larger scatter than the standard relation.

A particularly interesting UDG is Dragonfly 44 (DF44) which is the largest galaxy in the original van Dokkum et al. (2015) sample, with $R_{\rm eff} = 4.7 \pm 0.2$ kpc (van Dokkum et al., 2017). High S/N spectroscopy has revealed an extremely old and metal-poor stellar population ($\sim$2.3σ below the canonical dwarf MZR; Villaume et al., 2022), implying that DF44 quenched very early and over a short timescale. Moreover, while DF44 appears to have very low rotation (van Dokkum et al., 2019) characteristic of dwarf spheroidal galaxies, the stellar population gradients are 'inverted' compared to the gradients typical of dwarf spheroidals (Villaume et al., 2022). Regardless of whether DF44 has an over-massive halo or not (van Dokkum et al., 2017; Wasserman et al., 2019; Bogdán, 2020; Lee et al., 2020; Saifollahi et al., 2021), this UDG is inconsistent with the majority of UDG formation models.

Late-quenching (after infall into a dense environment) scenarios can be ruled out for DF44 given its old age (*e.g.*, Rong et al., 2017; Chan et al., 2018; Liao et al., 2019; Jiang et al., 2019a; Jackson et al., 2021). Moreover, DF44's low rotation conflicts with high-spin scenarios (*e.g.*, Rong et al. 2017; although the rotation could increase at larger radii, Grishin et al. 2021). Yet, given the uncertainty in establishing the cluster infall time for an individual galaxy, we cannot preclude early-infall scenarios (*e.g.*, Yozin & Bekki, 2015; Liao et al., 2019; Carleton et al., 2019, 2021; Tremmel et al., 2020). While some evidence (*e.g.*, Alabi et al., 2018; van Dokkum et al., 2019) suggests that DF44 is on its first infall into Coma, this is difficult to prove.

There is more to be learned, however, as UDG formation scenarios can be tested via their inferred star formation histories (SFHs). The time-scales of star formation reveal important epochs (*e.g.*, mergers, infall, and/or quenching), which can be compared against

observations. A number of studies have investigated the ages and mass assembly histories of UDGs, relying either only on broadband colours, or low to moderate S/N spectroscopy (*e.g.*, Kadowaki et al., 2017; Ferré-Mateu et al., 2018; Gu et al., 2018b; Pandya et al., 2018; Ruiz-Lara et al., 2018; Martín-Navarro et al., 2019; Buzzo et al., 2022). While these studies provide important first steps, comparisons with predictions are not necessarily straightforward. This is primarily because constraining the detailed shape of a galaxy's SFH is a complex problem.

Several galaxy properties can conspire to alter the spectral energy distribution (SED) in similar ways (*e.g.*, stellar age, metallicity, and dust), which are particularly difficult to disentangle with low spectral resolution data (*e.g.*, with photometry alone; Bell & de Jong, 2001). Recovering the SFHs for old stellar populations is particularly difficult – the integrated spectrum evolves non-linearly with age (Serra & Trager, 2007) such that old populations appear relatively similar (for a complete discussion see the review by Conroy, 2013). Moreover, a late burst of star formation can 'outshine' a (dominant) older population (*e.g.*, Papovich et al., 2001; Allanson et al., 2009). While broad wavelength coverage is needed to precisely determine the dust absorption (and emission, with mid-infrared coverage), high resolution data of select spectral features are needed to precisely constrain the stellar metallicity and age. Both observations are necessary to break the degeneracy between these parameters (*e.g.*, Vazdekis, 1999; Trager et al., 2000b). Using spectra that span a relatively wide wavelength range, full-spectrum fitting has proven to be effective in this respect (*e.g.*, MacArthur et al., 2009; Sánchez-Blázquez et al., 2011b). However, this technique requires a well-calibrated spectral continuum. Simultaneously fitting photometry and spectra can bypass this issue, as the photometry provides a means to fit the continuum[1] and increases the wavelength coverage.

In fitting the data it is necessary to impose 'prior knowledge,'[2] such as the flexibility of the SFH. The choice of a prior for the shape of the SFH can significantly impact age estimates, particularly for older stellar populations, and for low resolution and/or low S/N data (as shown in, *e.g.*, Maraston, 2005; Leja et al., 2017, 2019a; Han & Han, 2018; Carnall et al., 2019b). In order to draw connections between the predicted and observed properties of UDGs, it is necessary to give due attention to the choice of a prior. While it is advantageous to use flexible models together with physically motivated priors, a 'good prior' is not necessarily known a priori. Therefore, results should be discussed in the

[1] In practice, it is generally easier to calibrate photometry to standard filters than to calibrate a spectrum.

[2] 'Prior' here is used in the Bayesian sense, where the probability of a model given the data (*i.e.*, the 'posterior') is proportional to both the likelihood of the data (given the model) and the prior knowledge about the model.

context of the prior used (which may not be as 'uninformative' as intended; *e.g.*, Leja et al. 2019a).

In this work we simultaneously fit near-ultraviolet (NUV) to near-infrared (NIR) photometry (nine bands) with high S/N ($\sim$96 Å^{-1}) rest-frame optical spectroscopy (from KCWI, the Keck Cosmic Web Imager). The same data set was used in van Dokkum et al. (2019) and Villaume et al. (2022) to study the stellar kinematics and populations of DF44. We adopt flexible SFHs in our fiducial model which do not assume a certain shape with time. Moreover, we compare the results between SFH priors of different degrees of 'smoothness' in order to identify which results are fully constrained by the observations. We address the unique stellar population properties of this UDG, and its epoch of formation and quenching, in order to test models of UDG formation.

The data are described in Section 4.2, and Section 4.3 details how we fit the data with an advanced physical model. In Section 4.4 we discuss the results, and put the results in the context of the literature. What our results imply about the origins of DF44 in the context of theoretical models is discussed in Section 4.5. A summary of the key results is provided in Section 4.6. The SFHs of DF44 determined by this work are listed in full in Appendix B.1. We provide additional details on the above discussion in the Appendix, touching on systematic biases in measuring SFHs in Appendix B.2, and degeneracies between dust extinction and the flux from old stellar populations in Appendix B.3.

The magnitudes reported follow the AB magnitude system. We use a Chabrier (2003) initial mass function (IMF), and adopt a flat Λ cold dark matter (ΛCDM) cosmology with $\Omega_m = 0.3$ and $H_0 = 70$ km s^{-1} Mpc^{-1}.

4.2 Data

Our data for DF44 include both rest-frame optical spectroscopy and NUV to NIR photometry, shown in Figure 4.1, and described in more detail below. We assume the spectroscopic redshift measured by van Dokkum et al. (2017): $z = 0.02132 \pm 0.00002$.

4.2.1 Spectroscopy

The spectroscopy is described in detail in van Dokkum et al. (2019) and analysed further in Villaume et al. (2022); we summarize the relevant details here.

Of particular note is the sky subtraction, as the sky is much brighter than the UDG. Sky exposures were obtained $1'.5$ away from DF44 intermittently between DF44 observations. The wavelength-dependent time variation in the sky spectrum was obtained from the spatially collapsed individual sky spectra, as parameterized by principal component analysis (PCA). The sky in each science cube was determined from a linear combination of templates, where the bestfit sky spectrum for the given exposure was subtracted from each spatial pixel. Additional details are provided in van Dokkum et al. (2019).

KCWI integral field spectroscopy was obtained for DF44, and spectra were extracted in nine elliptical apertures after masking the ten brightest point sources. The apertures were sized $9'' \times 6''$, to match the UV photometry; see the following section. The integrated spectrum was determined through bootstrapping the individual spectra, where we used the 50^{th} percentile of the bootstrapped flux distribution and the average of the 16^{th} and 84^{th} percentile as the uncertainty. With 17 hours of exposure on-target, the integrated spectrum reaches a S/N ~ 96 Å^{-1} (see the third panel in Figure 4.1).

The KCWI Medium slicer with BM grating was used, yielding a spectral resolution of R ~ 4000. After masking and interpolating over regions badly affected by sky transmission, the spectrum was smoothed to a resolution of 110 km s^{-1}, for the purpose of later comparing with templates at this resolution. The final spectrum is shown in Figure 4.1 (the unsmoothed spectrum shown with grey lines), covering 4578–5337 Å rest-frame, with notable absorption features labelled. Also shown is the S/N of the spectrum as a function of wavelength.

Given the challenge of precisely flux-calibrating the spectrum (*e.g.*, due to residuals from the spectral extraction), we instead rely on the calibration of the photometry to provide constraints on the SED continuum when fitting the galaxy properties and SFH (see Section 4.3). For this reason, we do not flux calibrate the spectrum, and the continuum shape therefore reflects primarily the instrument response function and not the galaxy SED. We then effectively flatten the continuum by dividing through by a polynomial fit. In the fitting routine, we therefore need to marginalize over the shape of the spectral continuum in comparing the models to the observations (see Section 4.3.3).

Lastly, we chose to mask the spectrum between 4700–4750 Å rest-frame where there is a broad dip in the spectrum that does not appear in the models. We note that the blue end of the spectrum ($\lesssim$4800 Å) was not fitted by either van Dokkum et al. (2019) or Villaume et al. (2022). Our results are not impacted by masking this region of the spectrum, although the χ^2 values are slightly higher without masking.

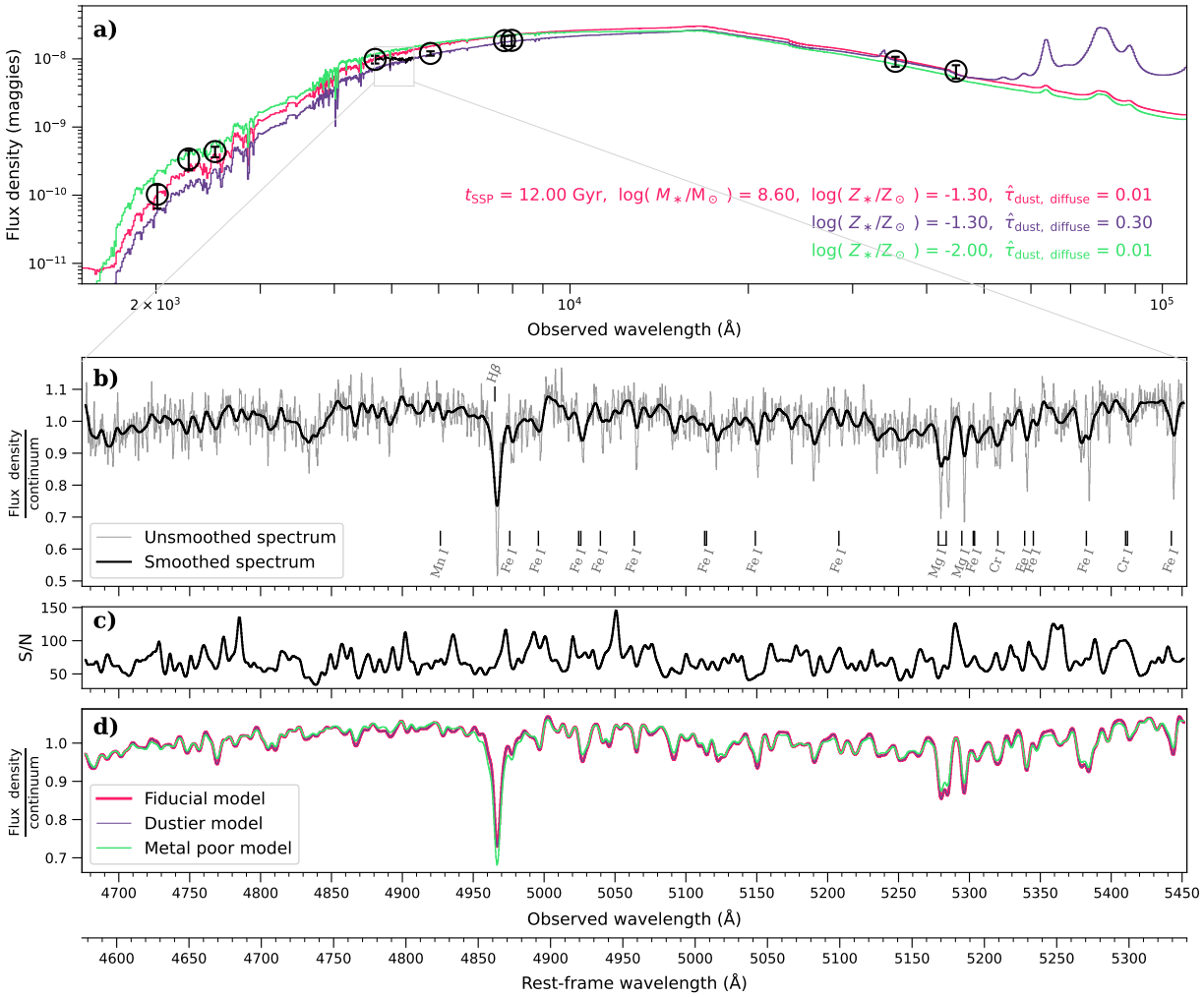

Figure 4.1: Observations of DF44 and three simple stellar populations (SSPs) shown to briefly illustrate the effect of the dust attenuation and metallicity on the shape of the SED. *(a)* The photometry with error bars (black markers) and model SEDs (coloured lines) with listed parameter values. *(b)* The original (continuum removed) KCWI spectrum (thin grey line) and the smoothed spectrum (thick line) for fitting purposes. Significant spectral features are labelled for reference. *(c)* The S/N of the smoothed spectrum. *(d)* The (continuum removed) model spectra, shown with the same resolution as the smoothed KCWI spectrum. While changing the dust attenuation affects the shape of the overall continuum, changing the metallicity affects both the continuum shape and the absorption features.

4.2.2 Photometry

Photometry in all the broadband images was performed by measuring fluxes within a $9'' \times 6''$ elliptical aperture, with a position angle of 65 degrees, to be consistent with the UV photometry reported by Lee et al. (2020). As this is significantly larger than the image resolution in all filters, no point spread function homogenization was applied, though appropriate aperture corrections are made to the *Spitzer* and *GALEX* images to account for light lost outside the aperture due to the point spread function. Details on the reduction and analysis of each image is described in more detail, below. The photometric measurements in each broad-band filter were corrected for foreground extinction in the Milky Way in the direction of the Coma Cluster using the website `http://argonaut.skymaps.info/usage` and Table 6 of Schlafly & Finkbeiner (2011) with $R_V = 3.1$.

Spitzer-IRAC Near-Infrared (NIR) Imaging

Spitzer-IRAC (Fazio et al., 2004; Werner et al., 2004) observations of DF44 were taken on 2017 May 12 starting at 07:19 (UT). Both 3.6 and 4.5 μm (channels 1 and 2, respectively) observations were taken. 50 medium-scale (median dither separation 53 pixels) cycling dither pattern 100-second frames were taken in each channel. The total exposure time was $93.6 \times 50 = 4680$ s in channel 1 and $96.8 \times 50 = 4840$ s in channel 2.

We removed the 'first-frame correction' (to address imperfect bias subtraction; see Section 5.1.10 of the IRAC Instrument Handbook). The rectification of each individual data frame for history effects in the IRAC arrays was performed in two steps that are explained in detail in Pandya et al. (2018). In short, we first performed a per pixel correction that was based on IRAC idling time characteristics in the IRAC skydarks, matched to those that took place before our observations. The typical magnitude of the per pixel correction was about 4 kJy sr^{-1} in channel 1 and 1 kJy sr^{-1} in channel 2. The typical corrections are much smaller than the read noise error, and we do not add any systematic magnitude uncertainties due to these first-frame corrections.

In the second step, a mean background is calculated for each frame, and a function fitted to these means is subtracted. The typical function consisted of a constant term plus terms that are declining exponentially with time. The uncertainties in these first-frame effect corrections are negligible compared to other sources of systematic error. We also formed a median image after doing a 3σ clipping from all the frames on the source in each channel and subtracted that median image separately in each frame. Such a median image will subtract the residual images that have been formed on the detector from previous

observations. We determined that the uncertainty in the final magnitudes added by this step is less than 0.01 mag.

The DF44 frames include a point source on top of the faint galaxy. We used *Spitzer* Science Center provided software MOPEX, specifically the APEX and APEX-QA modules, to subtract this point source using point response function (PRF) fitting. The estimated uncertainty due to this step is about 0.5 μJy in both channels.

We used the contributed *Spitzer*/IRAC software IMCLEAN (Jhora99, 2021) to remove leftover column pulldown artefacts from the CBCD frames. We then used the *Spitzer* custom software package MOPEX to create mosaics of the 50 frames in each channel, using the default parameters and the North up, East left orientation. Before mosaicking, we ran the overlap correction module to adjust for background offsets among the CBCD frames (one number per frame). We used only the multiframe outlier rejection scheme in MOPEX to reject outlier pixels in the input frames.

Next, we manually created masks of other sources (including point-like sources on the galaxy) in both channels with the custom software GIPSY (van der Hulst et al., 1992). We then measured the 'sky background' in five empty areas of sky close to DF44 in channels 1 and 2, and from the results we estimated an average sky background (0.00408 and 0.00415 MJy sr^{-1} in channels 1 and 2, respectively) to be subtracted at the position of DF44, applying the mask and using Astropy Python library commands in a $9'' \times 6''$ (P.A. $+65°$) aperture centred on the coordinates given by van Dokkum et al. (2015): R.A. = $13^{\rm h}00^{\rm m}58^{\rm s}.0$, Dec. = $26°58'35''$. We corrected the results with the appropriate aperture corrections from the IRAC Instrument Handbook.

The uncertainty in aperture photometry was estimated by performing aperture photometry on several positions in empty sky and taking the rms scatter in these measurements. This gave 0.05 and 0.10 mag in IRAC channels 1 and 2, respectively. We estimated the uncertainty due to masking by replacing the pixel values under the masks by the average pixel values within the unmasked aperture, and performed the photometry again, and took the difference between this measurement and the measurement using the masks as the uncertainty. The channel 1 masking uncertainty is thus 0.14 mag, and 0.18 mag for channel 2.

The sky background subtraction uncertainty is estimated by taking the maximum difference in the sky background measurements in three areas of empty sky around DF44 in the images and adding this difference to all the pixels within the photometry aperture and summing them up. This method gives 0.01 mag and 0.11 mag as the sky uncertainty in channels 1 and 2, respectively.

The calibration uncertainty was estimated to be 2% in IRAC channels 1 and 2, amounting to 0.02 mag in systematic uncertainty. There is an additional uncertainty of 9% in channel 1 and 2% fractional flux in channel 2 due to the uncertainty in integrated aperture flux correction factor (limiting case is infinite aperture). These convert to 0.09 and 0.02 mag in channels 1 and 2. In addition, there is the point source subtracting uncertainty of 0.01 mag.

We list the final AB magnitudes for channels 1 and 2 and their respective uncertainties from the quadrature sum of the magnitude uncertainties in Table 4.1.

Gemini GMOS g- and i-Band Imaging

DF44 was observed on 2017 May 12 with the Gemini Multi-Object Spectrometer (GMOS) for a total of 3000 s in both the g- and i-bands. The observations have been described by van Dokkum et al. (2016). We flux-calibrated the images with SDSS, accounting for a $g-i$ colour term and using four SDSS catalogued stars in our images. The data were obtained in photometric conditions, and we adopt an absolute calibration magnitude uncertainty to be 3%, amounting to 0.03 mag in the g- and i-bands, based on https://www.gemini.edu/instrumentation/gmos/calibrations. The sky background uncertainty was calculated as above for the IRAC channels, and amounted to 0.03 mag in the g-band and 0.09 mag in the i-band. Aperture photometry was performed using the coordinates from van Dokkum et al. (2015) and the Astropy Python library commands.

We list the final AB magnitudes for the g- and i-bands and the respective uncertainties in Table 4.1.

HST/WFC3/UVIS F606W and F814W imaging

Additional visual images of DF44 were taken on 2017 April 23 with the *Hubble Space Telescope* using the WFC3 camera and its UVIS detector and broadband filters F606W and F814W. van Dokkum et al. (2017) reported 5σ AB depths of F606W= 28.4 and F814W= 26.8 for DF44. A total of 2430 s and 2420 s were spent on the source in F606W and F814W filters. In both filters we calculated the sky mode in five different 'empty' regions of the sky and took an average and subtracted those values from the images. We also manually masked out point sources in the images. We used the image headers to calculate the conversion from electrons/s to AB magnitudes and performed elliptical aperture photometry within the same apertures as mentioned above for IRAC.

Table 4.1: DF44 Photometry.

Filter	m_0 (AB)	$\lambda_{\rm eff}$ (Å)
UVOT UV1	23.40 ± 0.19	2516.7
UVOT UV2	24.97 ± 0.41	2010.4
GALEX NUV	23.67 ± 0.35	2271.1
GMOS g_G0301	20.02 ± 0.14	4687.6
GMOS i_G0302	19.33 ± 0.18	7751.6
WFC3 F606W	19.80 ± 0.08	5813.0
WFC3 F814W	19.32 ± 0.19	7972.9
IRAC1	20.09 ± 0.18	35439.4
IRAC2	20.45 ± 0.24	44840.9

The uncertainties were estimated in the following way: we estimate a photometric calibration offset uncertainty of 0.03 mag, and the uncertainty due to background subtraction (estimated as above) is 0.05 mag in F606W and 0.13 mag in F814W. The uncertainty due to masked point sources within the aperture is estimated to be 0.03 and 0.01 mag in F606W and F814W. The uncertainty in performing aperture photometry was estimated as above and results in an additional 0.05 and 0.14 mag in F606W and F814W.

We list the final AB magnitudes for F606W and F814W and the respective uncertainties in Table 4.1.

Ultraviolet

The UV data reduction and analysis was presented in Lee et al. (2020). This consists of two filters observed with *Swift* UVOT (UV1 at 2600 Å and UV2 at 1928 Å), and *GALEX* NUV images. The UVOT data include a correction for red leakage and scattered light, where the correction (14%) was comparable to the flux uncertainty. Again, we list the final results in Table 4.1.

4.3 Stellar population modelling and fitting

In this section, we describe how we fit the DF44 observations using the fully Bayesian inference code Prospector, as introduced in Section 2.3. The photometry and spectroscopy are fit simultaneously, incorporating the information on the stellar properties and SFH

from both data sets. A summary of the parameters and priors of our physical model is shown in Table 4.2. Section 4.3.4 briefly describes the sampling method.

4.3.1 The physical model

The dust is modelled with the two-component dust attenuation model from Charlot & Fall (2000), which separates the dust components between those associated with the birth-cloud, and a uniform dust screen. While we expect DF44 to have an old stellar population with very little dust content, we prefer to include a flexible dust model and marginalize over the parameters rather than assume a simplistic model. This avoids the assumption that dust attenuation in DF44 is the same as dust attenuation in the local Universe. The birth-cloud dust acts to only attenuate stellar emission for stars younger than 10 Myr, while the diffuse-dust acts as a uniform screen with a variable attenuation curve (Noll et al., 2009).

The diffuse dust constant is given an uninformative prior, $\hat{\tau}_{\text{dust, diffuse}} \sim \text{Uniform}(\text{min} = 0,\ \text{max} = 1.5)$, and the dust ratio prior $\hat{\tau}_{\text{young}}/\hat{\tau}_{\text{dust, diffuse}} \sim \text{Normal}(\mu = 1, \sigma = 0.3, \text{min} = 0,\ \text{max} = 1.5)$ which broadly follows the results among the literature for massive galaxies while allowing for some variation. The prior on the diffuse dust attenuation index is assumed to be $n \sim \text{Uniform}(\text{min} = -2,\ \text{max} = 0.5)$.

Dust emission is calculated assuming energy conservation, *i.e.*, all the energy attenuated by dust is re-emitted at infrared wavelengths (da Cunha et al., 2008). As our photometry is limited to < 4.4 μm (rest-frame) there is no significant information in the SED constraining dust emission. We chose to include the full dust model and marginalize over the unconstrained parameters, rather than a more simplistic model, in order to avoid biasing the result. We assumed a permissive set of priors for the dust emission parameters, as listed in Table 3.2.

The stellar metallicity is a free parameter; however, we assume a constant metallicity for all the stars and for the entire history of the galaxy. This single metallicity has a uniform prior in $\log(Z_*/\text{Z}_\odot) \sim \text{Uniform}(\text{min} = -2, \text{max} = 0.19)$, where $\text{Z}_\odot = 0.0142$ (Asplund et al., 2009). In addition, scaled-Solar abundances are assumed, which is a current limitation of the FSPS models.

Nonparametric SFH

To characterize the SFH we use a nonparametric Dirichlet model of the form of a piece-wise constant function with $N = 12$ time bins. The time bins are defined in lookback time,

Table 4.2: SFH parameters and priors. Notes: 1) Fraction of SFR in a given time bin, where the SFH is a piece-wise constant function with N parameters. The prior is a Dirichlet function, see Section 4.3.1. 2) Redshift, with a tight prior about the measured spectroscopic redshift, $z_{\rm spec}$. 3) Total stellar mass is the integral of the SFH, which includes the mass lost to outflows. 4) The total stellar metallicity where scaled-Solar α-abundance is assumed. 5) Parameters for the two-component Charlot & Fall (2000) dust absorption model, with an adjustable attenuation curve slope from Noll et al. (2009) with a UV bump based on Kriek & Conroy (2013). 6) Parameters for the Draine et al. (2007) dust emission model. 7) The uncertainty on the spectra can be increased by a given factor, with a likelihood penalty for factors giving reduced χ^2<1. 8) An outlier pixel model can increase the errors for individual pixels by a factor of 50, to accommodate for poor matches between the data and spectral templates. 9) A fourth degree Chebyshev polynomial is fit (via optimization) to the residual of the normalized ratio between the observed spectrum and the proposed model spectrum and multiplied out prior to each likelihood calculation. This effectively accounts for the lack of flux-calibration in the spectrum.

Note	Parameter	Description	Prior
	SFH		
1	f_n	sSFR fraction.	Dirichlet($\alpha_{\rm D}$)
2	$z_{\rm obs}$	Redshift	Uniform(min $= z_{\rm spec} - 0.01$, max $= z_{\rm spec} + 0.01$)
3	$\log(M_*/{\rm M}_\odot)$	Total stellar mass formed	Uniform(min = 8, max = 12)
4	$\log(Z/{\rm Z}_\odot)$	Stellar metallicity	Uniform(min = -2, max = 0.19)
	Dust attenuation		
5	$\hat{\tau}_{\rm dust,\ diffuse}$	Diffuse dust optical depth (eq. 2.6)	Uniform(min = 0, max = 1.5)
	$\frac{\hat{\tau}_{\rm young}}{\hat{\tau}_{\rm dust,\ diffuse}}$	Ratio of diffuse to birth-cloud dust optical depth (eq. 2.5)	Clipped Normal($\mu = 1$, $\sigma = 0.3$, min = 0, max = 1.5)
	$n_{\rm dust}$	Diffuse dust attenuation index	Uniform(min = -2, max = 0.5)
	Dust emission		
6	$Q_{\rm PAH}$	Percent mass fraction of PAHs in dust	Uniform(min = 0.5, max = 7)
	$U_{\rm min,dust}$	Minimum starlight intensity to which the dust mass is exposed	Uniform(min = 0.1, max = 25)
	$\gamma_{\rm dust}$	Mass fraction of dust in high radiation intensity	LogUniform(min = 0.001, max = 0.15)
	Noise model		
7	$j_{\rm spec}$	Spectral noise inflation term	Uniform(min = 1, max = 3)
8	$f_{\rm outlier,\ spec}$	Fraction of outlier spectral pixels	Uniform(min = 10^{-5}, max = 0.5)
	Spectrophotometric calibration		
9	c_n	Polynomial coefficients, $n = 4$	

spaced so that the first seven bins correspond to 0–30 Myr, 30–100 Myr, 100–500 Myr, 500 Myr–1 Gyr, 1.0–2.0 Gyr, 2.0–3.0 Gyr, and 3.0–4.0 Gyr. There are four bins spaced logarithmically between 4 Gyr to $0.95\times$ $t_{\rm univ}$ (4.0–5.4 Gyr, 5.4–7.2 Gyr, 7.2–9.6 Gyr, and 9.6–12.6 Gyr), and the last bin covers $0.95\times t_{\rm univ}$–$t_{\rm univ}$, where $t_{\rm univ}$ is the age of the Universe at the time of observation. Defining the time bins this way reflects the non-linear evolution in the SEDs: the narrower time bins at recent lookback times allow a sufficient precision in capturing recent star formation, while the wider bins at later lookback times reflect the modest evolution of older stellar populations. The last time bin is included to permit a maximally old population.

Fitting SEDs to recover SFHs is an ill-defined problem, and prone to overfitting (*e.g.*, Moultaka & Pelat, 2000; Moultaka et al., 2004; Ocvirk et al., 2006a). In order to recover a physically plausible SFH it is common to invoke 'regularization.' There are a number of ways that this can be done, which differ in technical detail. One approach is to impose Gaussian-like priors on the SFH and/or the age-metallicity relation (*e.g.*, as in the commonly used code STECKMAP; Ocvirk et al., 2006a; Ocvirk et al., 2006b), and another is to penalize sharp transitions in the SFH (*e.g.*, the continuity prior; Leja et al., 2019a). In this work, we use a third method, which is to control the degree of concentration of fractional specific SFR (sSFR) between the time bins of the nonparametric function. While these approaches differ in detail, they all attempt to avoid non-physical solutions by imposing constraints on the variability of the SFH over time.

We adopt a Dirichlet prior which includes a concentration parameter, $\alpha_{\rm D}$, that controls the preference to distribute the fractional sSFR in one bin ($\alpha_{\rm D} < 1$) or evenly between all bins ($\alpha_{\rm D} \geq 1$), respectively. A detailed description of this prior is provided in Leja et al. (2017). Without direct physical motivation to inform a choice of $\alpha_{\rm D}$, we consider both $\alpha_{\rm D} = 1$ and $\alpha_{\rm D} = 0.2$ as valid options, labelling them as 'extended' and 'concentrated' versions of the SFH prior. In comparing the results produced from these two choices of SFH prior, we explore the dependence of the results on the degree of regularization.

Fig. 4.2 shows random draws (thin lines) for priors with $\alpha_{\rm D} = 1$ (extended) and $\alpha_{\rm D} = 0.2$ (concentrated), with the time bins as defined above. The median and 68% credible regions (CRs) of the priors are shown with a thick line and shaded regions, respectively. The corresponding implicit prior on the mass-weighted age is shown in the bottom panel for reference. The mass-weighted stellar age, $t_{\rm mass}$, broadly describes the average formation time of stars in a given galaxy in units of lookback time) is calculated from the SFH (see Chapter 2 Equation 2.13). The implicit age prior for an extended SFH is centred at half the age of the Universe with a 99.9% CR between 3.08–9.98 Gyr, and thus is a strong prior against both very old and very young ages. In comparison, the concentrated SFH also peaks around half the age of the Universe (although offset given the varying widths of the

time bins), but the prior is not as tight (99.9% CR between 0.83–12.17 Gyr) such that old ages are less strongly disfavoured.

4.3.2 Noise and outlier models

A noise model is used to account for possible under- or over-estimates of the spectral uncertainties, where the noise is uniformly inflated (or deflated). This effectively modifies the spectral uncertainty by a multiplicative factor, but is counterbalanced by a penalty in the likelihood calculation for larger uncertainties. This down-weights spectra where the uncertainties are otherwise low, but there is a mismatch between the spectrum and the models. The minimum uncertainty in the photometry is 7%, and as we expect this to be large enough to account for deviations with the template SEDs, we do not include a noise model for the photometry in our model.

A mixture model is used to identify and mask pixels in the spectra which have large deviations from the model. The purpose here is to avoid being overly sensitive to outlier pixels in the spectrum. This is again relevant where the S/N is large and significant residuals can result from poor matches to the models where the model itself is inaccurate (due to differences in, for example, α-enhancement). Prospector uses the mixture model approach described in Hogg et al. (2010a).

The spectral outlier model finds that less than 1% of the pixels are inconsistent with the model templates beyond the specified uncertainty. Note that the spectral noise inflation model prefers to inflate the uncertainties by ∼1–3%, which is not unexpected given that the S/N of the spectrum is high, 40–140 (median 96) and that the models are not flexible enough to precisely match the metallicity- and α-abundance sensitive spectral features (i.e, the Mg triplet).

4.3.3 Spectrophotometric calibration

We rely on the calibration of the photometry to constrain the shape of the SED continuum. The DF44 spectrum is not flux-calibrated, such that neither the normalization nor the shape of the spectral continuum provides information about the stellar properties. In fact, the spectrum was flattened prior to fitting (see Section 4.2.1). For this reason we ignore the shape of the spectrum when computing the likelihood of the SED model (relative to the spectrum). We do this by following the routine provided through Prospector which fits (via optimization) a polynomial to the residual between the spectrum and the model, which is then multiplied to the model. We use an $n = (\lambda_{\rm max} - \lambda_{\rm min})/100$ Å ∼8 order

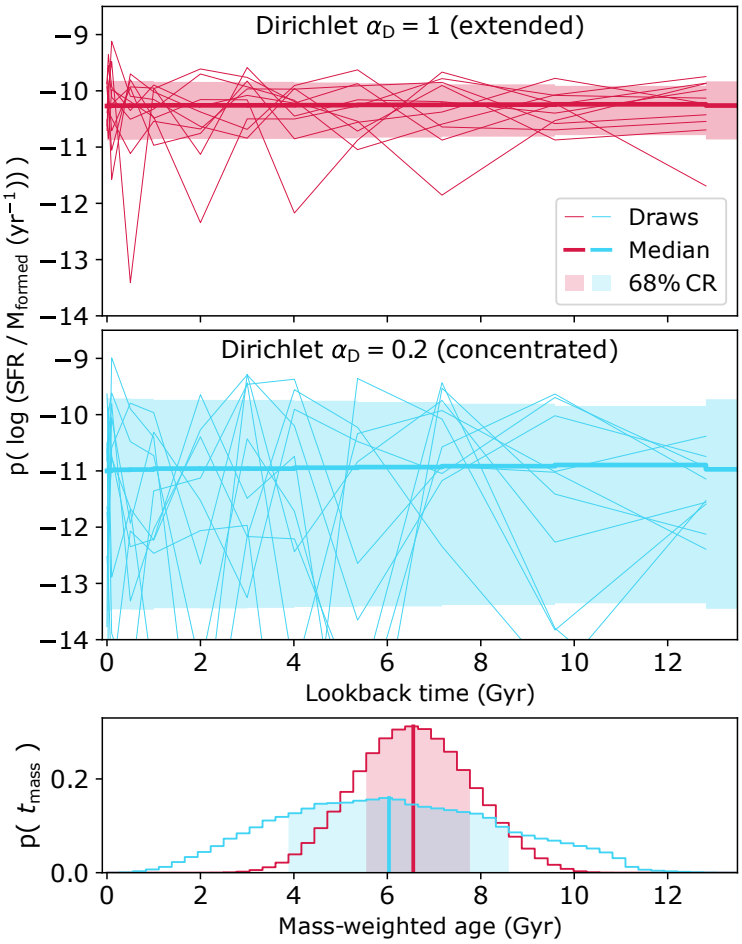

Figure 4.2: Comparison of SFH priors between specific SFR (SFR per unit mass; first and second panels) and corresponding mass-weighted age (third panel). The $\alpha_D = 1$ SFH prior prefers solutions where the star formation is equally weighted between the time bins, hence we label it 'extended.' In comparison, the $\alpha_D = 0.2$ SFH prior prefers solutions where the star formation is unequally distributed between time bins, which we label as 'concentrated.' The sSFR is shown as a function of lookback time for ten random draws (thin lines) from the Dirichlet SFH priors with $\alpha_D = 1$ (extended; red) and $\alpha_D = 0.2$ (concentrated; blue). The medians and 68% CRs of the prior are indicated with thick lines and shaded regions, respectively. The implicit mass-weighted age priors are shown in the lower panel, with vertical lines indicating the median, and shaded regions indicating the 68% CRs.

Chebyshev polynomial, which is flexible enough to remove the broad continuum shape without over-fitting absorption features (*e.g.*, Conroy et al., 2018). We test our results using several orders of the polynomial, and find that we are generally insensitive to the choice of n as long as $n > 4$ (otherwise the dust attenuation pdf is skewed).

4.3.4 Sampling

The complete model includes 19 free parameters (11 of which describe the shape of the SFH), which are summarized in Table 4.2. We follow the sampling procedure outlined in Johnson et al. (2021) (see also Tacchella et al., 2022b), using the dynamic nested sampling (Skilling, 2004; Higson et al., 2019) algorithm DYNESTY[3] (Speagle, 2020) to efficiently sample the high-dimensional parameter space of the model and build posterior pdfs. This approach provides full posterior distributions of the model parameters together with their degeneracies. A useful primer on Bayesian methods can be found in van de Schoot et al. (2021).

Throughout this work we report the uncertainties as 68% CRs (which corresponds to the 16^{th} to 84^{th} percent range) of the posterior pdfs as the majority of the distributions are non-symmetric.

4.3.5 Simultaneously fitting the photometry and spectroscopy

In fitting both the photometry and spectroscopy we consider the log-likelihood of the model, conditioned on the observation, to be the sum of the two individual likelihood functions:

$$\ln \mathcal{L}(d_s, d_p|\theta, \phi, \alpha) = \ln \mathcal{L}(d_s|\theta, \phi, \alpha) + \ln \mathcal{L}(d_p|\theta) \tag{4.1}$$

where d_s is the spectroscopic data, d_p is the photometric data, the parameters θ describe the physical model used in PROSPECTOR, the parameters α describe the spectroscopic noise model (Section 4.3.2), and the parameters ϕ include the spectro-photometric calibration (Section 4.3.3). The parameters of the physical model are summarised in Table 4.2. We apply no relative weighting between fitting the spectroscopy and photometry in assessing the match between the observations and SEDs.

The basic likelihood calculation is effectively a χ^2 calculation for both the spectral and the photometric data. We alter the likelihood calculation for the spectroscopy to include the noise model and outlier model described in Section 4.3.2, following the procedure outlined in Appendix D of Johnson et al. (2021).

[3]https://dynesty.readthedocs.io/en/latest/

4.4 Results

Given the sensitivity of modelling ages of old stellar populations, and their dependence on both the flexibility of the assumed SFH and the choice of SFH prior (*e.g.*, Leja et al., 2017, 2019a), we present the results for two 'extremes' of the SFH prior: i) an 'extended' SFH, preferring equal distribution of fractional sSFR between the time bins ($\alpha_{\rm D} = 1$), and ii) a so-called 'concentrated' SFH, preferring an unequal distribution of fractional sSFR between time bins ($\alpha_{\rm D} = 0.2$). The difference between these priors is discussed in Section 4.3.1.

In assuming the SFH is extended, there is a preference for ages of half the age of the Universe and against old ages ($\gtrsim 10$ Gyr; see Figure 4.2). However, the results of Villaume et al. (2022) suggest that DF44 formed its stellar population early, and shortly thereafter rapidly quenched, as determined from its inverted stellar population gradients and low iron metallicity for its mass. The 'concentrated' SFH prior has a higher likelihood for such an SFH. Moreover, the concentrated SFH prior has an overall broader implicit prior on the mass-weighted age as there is no preference for where the fractional sSFR is concentrated in the time bins.

The results from the full-spectral modelling of DF44 are shown in Figures 4.3–4.5, where the fits for the extended SFH prior are shown in red, and for the concentrated SFH prior in blue. Table 4.3 summarizes the posteriors for select parameters of the SED models, as well as latent parameters. In Figure 4.3 the observations are shown with the 'bestfit' models (the maximum a-posteriori model; *i.e.*, with the highest probability of the set of samplings) and the 68% CR of 500 random draws from the posteriors.

Overall, the fits to the photometry are similar between the two SFH priors; the extended SFH model has marginally smaller residuals at NUV wavelengths. Similarly, the bestfit model spectra (multiplied by the spectrophotometric calibration polynomial) compared to the spectroscopy are nearly identical, with differences only at the $< 1\%$ level. Given the degeneracy between age, dust, and metallicity, the subtle differences in these features lead to the differences in the predicted stellar population parameters.

4.4.1 Star formation history and stellar population parameters at z=0

Figure 4.4 shows the median (solid line) and 68% CR (shaded) of the posteriors for the sSFR, and corresponding SFR and mass-assembly history. Similarly, the median (dashed

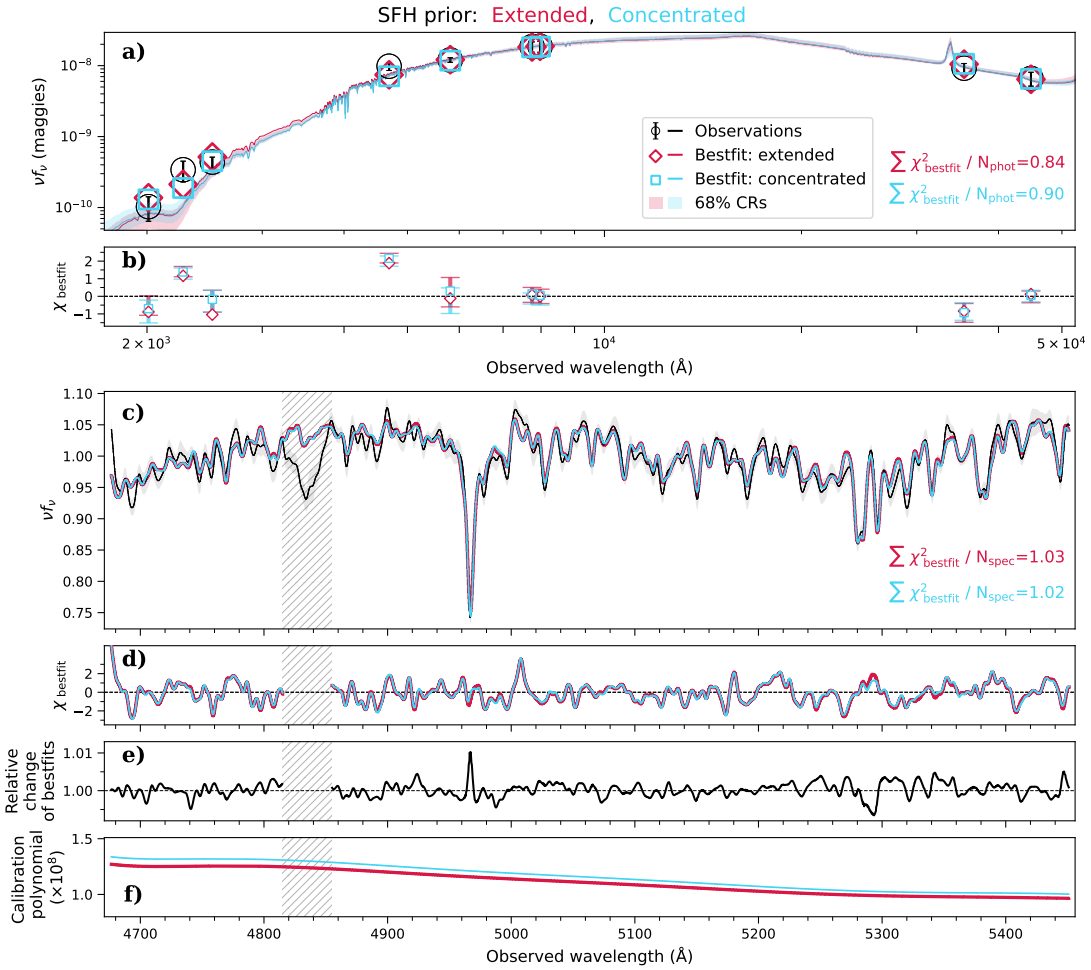

Figure 4.3: Summary of the fitting results for DF44. The observed data (black) is compared to the bestfit models (red, extended SFH prior; blue, concentrated SFH prior) and the 68% CR of 500 randomly drawn models from the posteriors (red/blue shaded). The corresponding posteriors are shown in Figs. 4.5 and 4.4. *(a)* The observed (circles) and bestfit (diamonds and squares) photometric points, where the $\chi^2/N_{\rm data}$ of the bestfit SED is listed and *(b)* shows χ ([data − model]/σ) of the bestfit points. *(c)* The observed (uncertainties shown in grey) spectrum and bestfit spectra (multiplied by the spectrophotometric calibration polynomial). The light grey region indicates the spectral region masked throughout the fitting process. *(d)* The χ of the bestfit spectra as a function of wavelength. *(e)* The relative change of the bestfit models, *i.e.*, the ratio of the two bestfit spectra. *(f)* The spectrophotometric calibration polynomials.

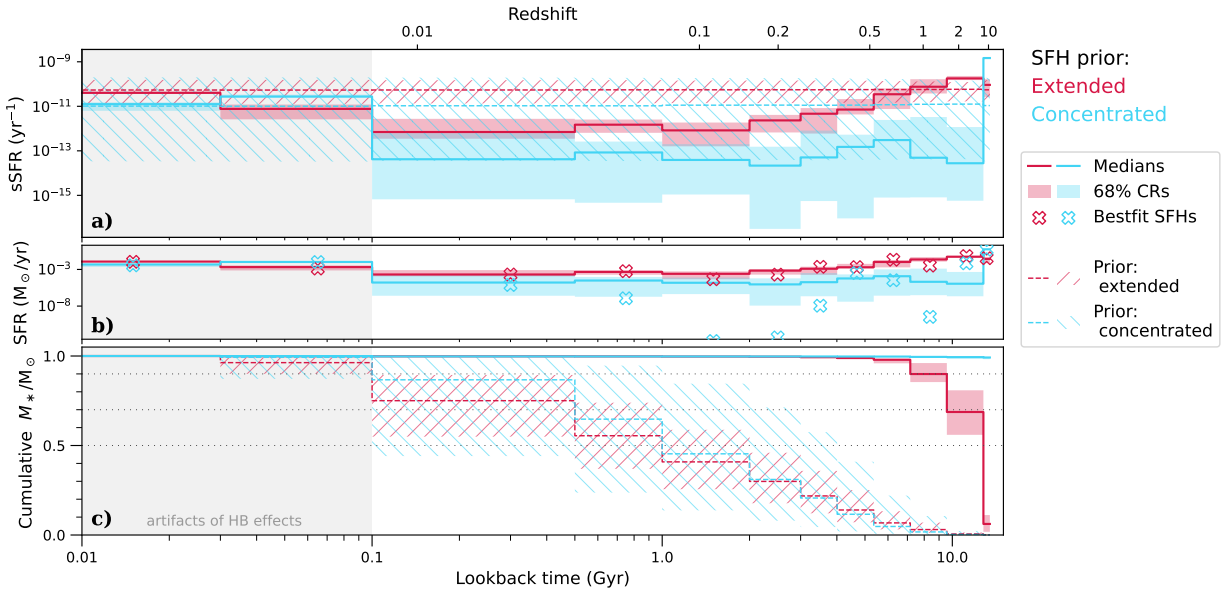

Figure 4.4: Posteriors for the (*a*) sSFR, (*b*) derived SFR, and (*c*) derived cumulative stellar mass for the fits shown in Figure 4.3. The results and priors shown in red (blue) colours correspond to the fit with an extended (concentrated) SFH prior. Solid lines indicate the median, and the shaded regions the 68% CR. These values are listed in Table B.1 in Appendix B.1. The SFHs of the bestfit models are indicated with open crosses. The median (dashed line) and 68% CRs (hatched) of the priors are shown for reference. Note that the cumulative mass and mass-weighted age priors are implicit, as they are derived from the sSFR prior. Dotted lines are drawn at 50, 70, and 90% of the cumulative mass for reference. The last 100 Myr are shaded grey to indicate that the SFH is affected by artefacts such as HB stars (see text).

Table 4.3: Summary of posteriors for DF44's properties and SFH timescales. Parameters with asterisks are directly included in the SED model, while those without are derived from the SED model parameters. While the mass-weighted age is in units of lookback time, the timescales t_x are in units of Gyr since the Big Bang. Time-scales t_x correspond to the time at which x percent of stellar mass had formed. Given that the SFH is a step function, the SFH is interpolated to estimate t_x. The 16^{th}, 50^{th}, and 84^{th} percentiles of the posterior (*i.e.*, the 68% CR) are shown. For the SFH timescales, this is likely an underestimate of the uncertainty given the width of the time bins in our SFH. Given the observed redshift of DF44 and the adopted cosmology, the age of the universe is 13.47 Gyr. Timescales in units of lookback time are therefore $t_{\text{Lookback}} = 13.47 \text{ Gyr} - t_x$. The fractional SFH within the time bins of the nonparametric model (*i.e.*, not interpolated) are listed in Table B.1 in Appendix B.1.

Parameter	SFH prior					
	Extended			Concentrated		
	16	50	84	16	50	84
$^*\log(Z_*/\text{Z}_\odot)$	-1.19	-1.18	-1.17	-1.29	-1.27	-1.24
$^*\hat{\tau}_{\text{dust, diffuse}}$	0.19	0.24	0.27	0.17	0.20	0.24
A_V	0.38	0.51	0.62	0.45	0.37	0.55
$\log(M_*/\text{M}_\odot)$	8.17	8.23	8.25	8.30	8.33	8.36
t_{mass} (Gyr)	9.72	10.20	10.54	13.02	13.06	13.08
t_{10} (Gyr)	0.60	0.87	1.12	0.068	0.068	0.069
t_{20} (Gyr)	1.18	1.39	1.64	0.135	0.136	0.137
t_{30} (Gyr)	1.63	1.79	2.33	0.203	0.204	0.206
t_{40} (Gyr)	2.07	2.33	2.92	0.271	0.272	0.274
t_{50} (Gyr)	2.55	2.85	3.53	0.339	0.340	0.343
t_{60} (Gyr)	2.93	3.41	4.17	0.406	0.408	0.411
t_{70} (Gyr)	3.37	4.07	4.92	0.474	0.475	0.480
t_{80} (Gyr)	3.86	5.39	5.70	0.542	0.543	0.549
t_{90} (Gyr)	5.73	6.32	7.01	0.610	0.611	0.617
t_{95} (Gyr)	6.25	7.57	7.93	0.643	0.645	0.652
$t_{50} - t_{90}$ (Gyr)	2.43	3.30	4.21	0.27	0.27	0.27

line) and 68% CRs (hatched) for the explicit and implicit priors are shown (see also Figure 4.2). The SFHs of the bestfit models (shown in Figure 4.3) are indicated with open crosses. Dotted lines are drawn at the $50^{\rm th}$, $70^{\rm th}$, and $90^{\rm th}$ percent levels of the cumulative stellar mass for reference. In Table 4.3 we summarize the two SFH results by listing the times at which different percentiles of the final stellar mass were in place. The SFH within the time bins of the nonparametric model are provided in Appendix B.1.

The SFHs determined from both priors suggest that DF44 formed early, having 90% of its stellar mass in place at least ∼7.2 Gyr ago ($z \sim 0.9$). Using the extended SFH prior, we find that it took ∼3.5 Gyr for DF44 to assemble between 50% to 90% of its mass, suggesting a relatively fast transition between star forming and quiescent states. The SFH determined with the concentrated prior is extreme in that more than 90% of the mass formed within the first time bin, *i.e.*, ∼12.8 Gyr ago ($z \sim 8$). During the last 5 Gyr the two results are otherwise similar, with low levels of star formation until the last 100 Myr.[4]

A curious feature of both SFHs is the rise in SFR within the last 100 Myr (corresponding to the first two time bins and shaded grey in Figure 4.4; by 1.8–2.4 dex). Although residual star formation appears to be common for massive early type galaxies, where ∼0.5 per cent of their mass formed within the last 2 Gyr, the fraction decreases at lower stellar masses, consistent with galaxy 'downsizing' (*e.g.*, Salvador-Rusiñol et al., 2020). The recent rise in DF44's SFH accounts for $\lesssim 1\%$ of the total stellar mass, assuming either SFH prior. While DF44 shows no indication of recent star formation from the photometry, and similarly lacks emission lines in the spectrum, it is possible that Hα emission (perhaps related to star formation ignited by a late infall in to the Coma cluster) recently stopped. This is perhaps unlikely, however, given the lack of blue regions within the galaxy. Lee et al. (2020) concluded, based on the difference in NUV and UVW2 bands, that the light traces older stars (on $\sim$ Gyr time-scales, as opposed to young stars which evolve on the order of $\sim$ Myr time-scales). The 'recent burst' is not a consequence of an artifact in the KCWI spectrum; the same feature is apparent when fitting the MaNGA data from Gu et al. (2018b). Rather, we expect this recent star formation to be an artifact of the stellar models not being flexible to the contribution of blue horizontal branch (HB) stars (discussed in Appendix B.2.1) or non-solar Mg-abundances.

We nonetheless test the sensitivity of the models to the presence of a very young stellar population by re-defining the time bins of our SFH, only allowing for star formation older

[4] Additional testing of the prior-sensitivity of the SFH showed that using $\alpha_{\rm D} = 0.5$ (mildly concentrated) produced parameter values between the results from $\alpha_{\rm D} = 0.2$ and $\alpha_{\rm D} = 1$, as expected. The mass-weighted age was found to be 11.9 Gyr, which indicates that the very old age is not overly sensitive to the choice of the $\alpha_{\rm D}$ value.

than 1 Gyr. This places a strong prior against recent star formation (SF) to counteract the inability of the SPS models to correctly model the influence of the blue HB stars. In excluding star formation younger than 1 Gyr, the models are better able to recover the shape of the SED, particularly in the NUV, but are marginally worse in matching the spectrum. With this revised model we recover SFHs equivalent to that of our primary results (at times > 1 Gyr), with statistically consistent but less dust and metallicity, and slightly higher stellar mass. Interestingly, with the extended SFH prior, the revised age estimate is $\sim$2.4 Gyr older. With the concentrated SFH prior, the revised age estimate is unchanged from that of our main result. As such, we conclude that the presence of the 'recent burst' of SF does not affect our conclusion that DF44 formed and quenched very early in the history of the Universe.

Figure 4.5 shows the posteriors for the normalization of the diffuse dust attenuation curve, stellar metallicity, stellar mass, and mass-weighted age. The parameters marked with an asterisk are not directly fit in our physical model, but derived from the posterior distributions. We calculate the dust extinction following equations (2.5) and (2.6) in the V-band, where we use $\lambda = 5500$ Å. We note that 'total stellar mass formed' is a free parameter in our model, which we convert to 'stellar mass' by subtracting the mass lost throughout the SFH, as calculated by FSPS (see Chapter 2 Section 2.3.9).

Contrary to our expectation that old (*e.g.*, Péroux & Howk, 2020) and metal-poor (*e.g.*, Galliano et al., 2018) populations are devoid of dust (see also Barbosa et al., 2020), DF44 appears to have a non-negligible amount: the normalization of the diffuse-dust attenuation curve is $\hat{\tau}_{\rm dust,\ diffuse} \gtrsim 0.2$ and $A_{\rm V} \gtrsim 0.5$. The origin of such dust is not clear, however, Buzzo et al. (2022) recently measured similar extinction values from optical to mid-infrared photometry for a sample of quiescent UDGs. The overall shape of the SED constrains the dust content, however there are degeneracies with both metallicity and age. If we instead fix $\hat{\tau}_{\rm dust,\ diffuse} = 0$ and refit DF44 (with an extended SFH prior), the posteriors are statistically consistent with that of our main result, although we note that the age increases (as expected) by $\sim$0.23 Gyr. In Appendix B.2.2 we discuss the fit to just the photometry, which prefers an even dustier solution ($\hat{\tau}_{\rm dust,\ diffuse} \sim 0.36$ and $A_{\rm V} \sim 0.8$, although the photometry provides no direct constraint for the metallicity, and little constraint for the age). While the spectroscopy breaks the degeneracy between dust and metallicity, the degeneracy with age remains; adding either more dust or a stellar population older than $\sim$3 Gyr lowers the flux at wavelengths < 5000 Å (see Appendix B.3). Additional observations in the mid-infrared would provide better constraints on the dust content, as age and dust affect the flux in opposite directions at this wavelength range.

Other than $\hat{\tau}_{\rm dust,\ diffuse}$, the posteriors of the dust model parameters largely reflect their

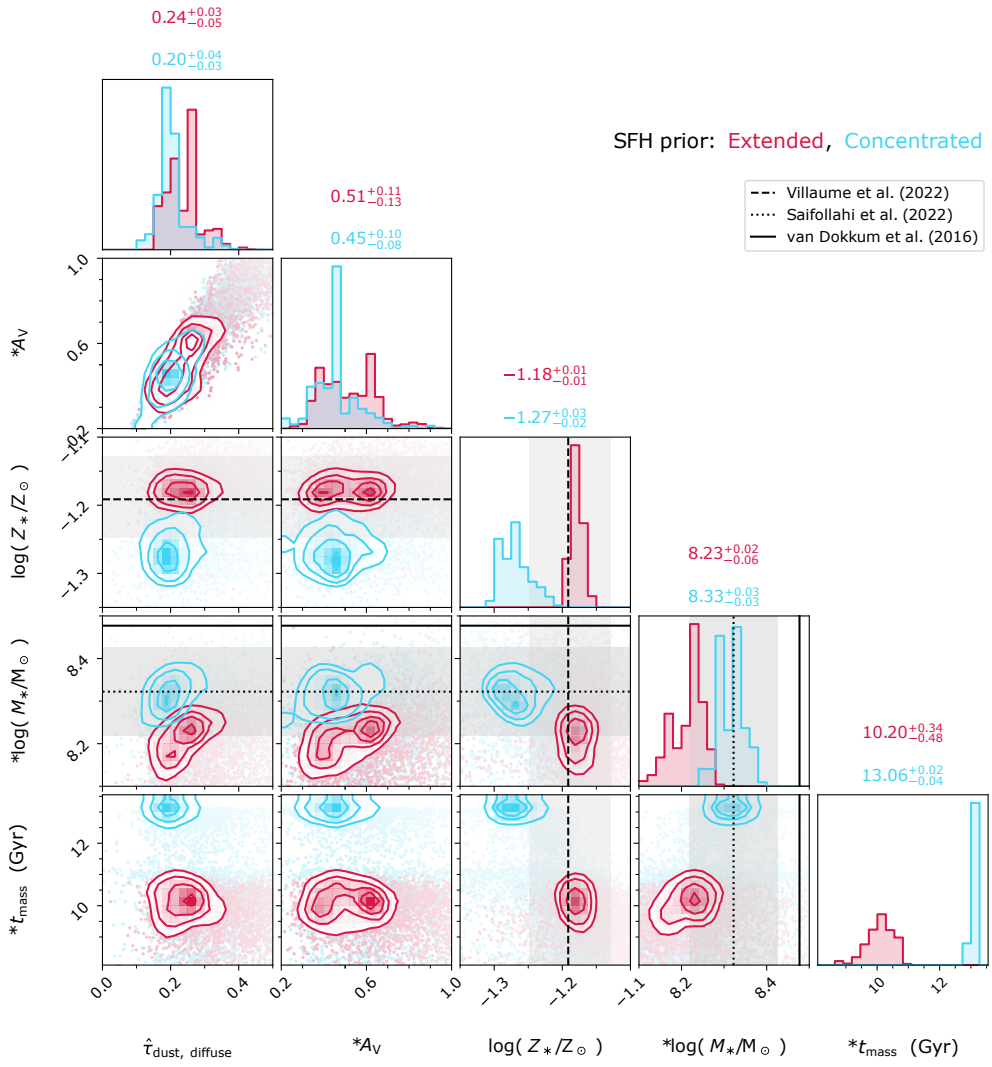

Figure 4.5: Posteriors of selected fitted and derived parameters (indicated with an asterisk) for the fits shown in Figure 4.3. Contours are shown smoothed with a $n = 1$ Gaussian kernel, where red (blue) contours show the fits with an extended (concentrated) SFH prior. Black lines denote the expected results given the analysis by Villaume et al. (2022) for $\log(Z_*/Z_\odot)$, and stellar mass from van Dokkum et al. (2016) and Saifollahi et al. (2022). Grey shaded regions indicate the uncertainties on these values. The median and uncertainties from the 68% CR for our results are listed along the top of the one-dimensional histograms.

priors – which is to be expected given the lack of constraining data. Nonetheless, to check that our results do not depend on the particular dust model, we also fit the data with the dust model of Gordon et al. (2003) based on the SMC Bar (thought to have similar dust properties to dwarf elliptical galaxies, *i.e.*, without a UV bump in the extinction curve), and find no change to our result. A degeneracy between the dust normalization and stellar mass can be seen in the joint posterior in Figure 4.5, where an increase in dust suggests a higher stellar mass. As a point of comparison, a solid black line indicates the estimated stellar mass from van Dokkum et al. (2016), and a dotted black line indicates that measured by Saifollahi et al. (2022) with uncertainties reflecting the systematics of the model fitting. Both of our fits produce stellar masses lower than (and statistically inconsistent within their 68% CRs) with the van Dokkum et al. (2016) value, but consistent with Saifollahi et al. (2022). Given that the photometry included in our fits is measured within an aperture, and thus does not include all the light of the galaxy, it is not unexpected that the stellar mass we recover underestimates that from the literature.

There is a $\sim$0.1 dex difference in $\log(Z_*/\mathrm{Z}_\odot)$ between the fits with an extended or concentrated SFH prior, where the sense of the metallicity difference is consistent with that of the age difference ($\sim$2.9 Gyr) with respect to the age–metallicity degeneracy. This indicates that we are not able to fully break the age–metallicity degeneracy with the data at hand. While in Figure 4.5 we show the stellar 'isochrone' metallicity measured by Villaume et al. (2022) as a black dashed line for comparison, there are several caveats to their comparison which are discussed in the following section.

At this point, the dichotomy of DF44 being 'old' or 'very-old' is subject to the choice of SFH prior. We remind the reader that the extended SFH prior behaves analogously to regularization methods used throughout the literature. While the concentrated SFH prior provides more flexibility to better recover the short and early star formation expected for DF44, it is not necessarily a 'good' prior; we provide no physical information for the shape of the SFH. We simply tune the prior such that it prefers to distribute the SF within fewer time bins (see Section 4.3.1).

This prior-dependency problem is exacerbated with less complete or lower S/N data sets. As a brief example, in Figure 4.6 we compare the stellar metallicities and ages determined through fitting both the spectrum and photometry (diamond), with that fitted to only the photometry (circle) for the extended SFH prior (points marked with an 'E'). While the NUV–NIR photometry provides information on the dust in DF44 (see Appendix B.2), the age estimate is more heavily weighted by the SFH prior than are the full spectrum fitting results. Accordingly, the photometry-only fit gives a median age $\sim$3.4 Gyr younger

than the fit to the spectrum and photometry together.[5]

4.4.2 Which SFH prior is preferred?

There is little statistical evidence to decide whether the results from either SFH prior better reflects the 'true' properties (or SFH) of DF44.[6] The distributions of SED models shown in Figure 4.3 are similar between the fits with each prior, and the models have similar residuals.

There are subtle differences, however, particularly around the Hβ and Mg II features where the concentrated SFH gives a (statistically) lower χ^2. The Hβ line is sensitive to recent star formation (and to HB stars, as discussed in Section 4.4.1), while Mg II is sensitive to the α-abundance of the stellar population. The FSPS models that we use are currently limited to fixed solar α-abundance. However, Villaume et al. (2022) found that DF44 has [Mg/Fe]$= 0.11^{+0.06}_{-0.04}$ through fitting the same spectrum of DF44 as this work with the full-spectrum fitting code `alf` (Conroy et al., 2018), which includes response functions to measure the non-solar chemical abundance variations. Given the relationship between both features and the age of the stellar population, this points to the need to include more complex stellar populations variables, *e.g.*, α-abundance, in models in order to break this degeneracy.[7]

Figure 4.6 compares the stellar metallicities and ages measured for DF44 by this work,

[5] If instead of the nonparametric model, we assume the SFH follows a delayed exponential form (a common parametric model adopted within the literature) we find similar results. With a log-uniform prior on the e-folding time, τ, and linearly uniform prior for the delay time, $t_{\rm mass}$, the implicit age prior has a complex form with $16^{\rm th}$, $50^{\rm th}$, and $84^{\rm th}$percentiles of 1 Gyr, 3.8 Gyr, and 8.4 Gyr respectively – preferring younger ages than the extended SFH prior results. The implicit age skews even younger if instead τ is linearly sampled. Fitting the photometry of DF44 suggests the age is ∼8.2 Gyr, and slightly less dusty than using the extended SFH model. Fitting both the photometry and spectroscopy suggests the age is ∼13.6 Gyr, and slightly less dusty and more metal poor than our main result. We note that the photometry-only results with the delayed parametric model appear particularly sensitive to the S/N – if we inflate the photometric uncertainties by a factor of two, the age posterior decreases by ∼2 Gyr. The same is not true when using the nonparametric models.

[6] Comparing the Bayesian evidence of the two fits (as derived from the nested sampling described in Section 4.3.4) we find a strong preference (according to the Jefferys scale, see for example Kass & Raftery 1995) for the concentrated SFH prior ($\ln Z_{\rm concentrated} = 62590$ is much larger than $\ln Z_{\rm extended} = 62542$), where here Z is the Bayesian evidence. However, this likely reflects the fact that the old age of DF44 is more disfavoured by the extended SFH prior (see Section 4.3.1) more than a preference of the data itself.

[7] While FSPS does include an option to set the fraction of blue HB stars, for technical reasons we cannot include it as a free parameter in our models.

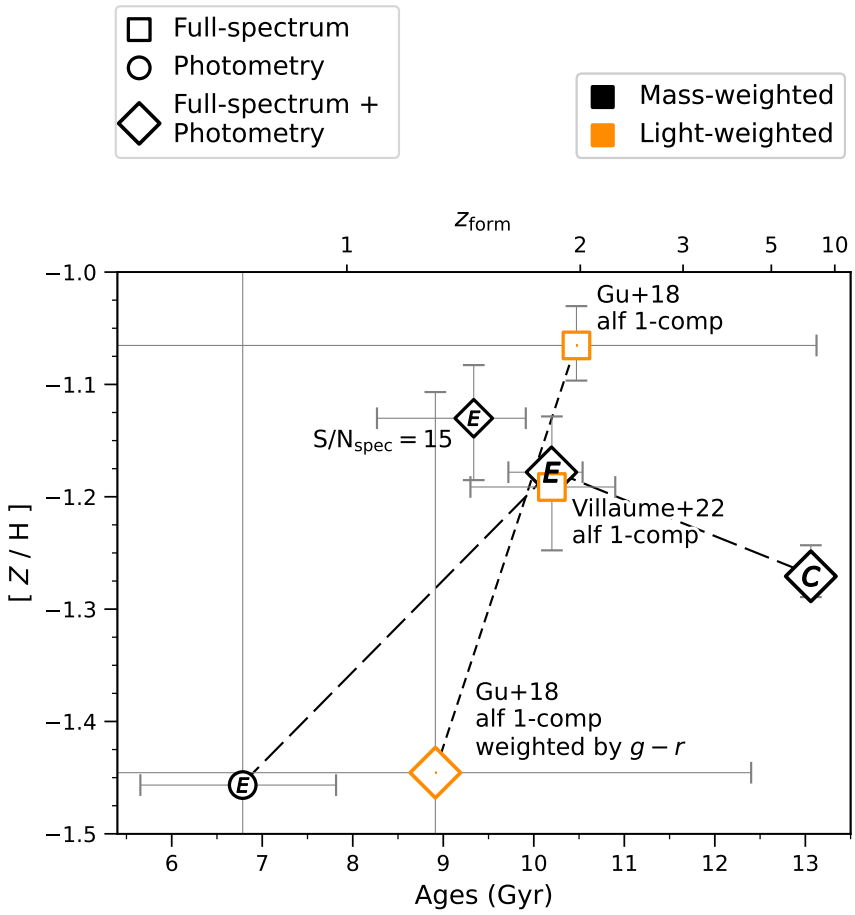

Figure 4.6: Comparison of stellar metallicity and age for DF44 from this work, Gu et al. (2018b), and Villaume et al. (2022, using the same spectroscopic data set as this work). Both results from the literature derived values using alf, and thus are not directly comparable to our results using Prospector (see text). Black coloured points show mass-weighted ages, while orange points show luminosity-weighted ages. Marker shapes indicate the data used in fitting the stellar properties. Dashed lines connect results obtained from the same study. We mark the results from this work derived with an extended SFH with an 'E,' or with the concentrated SFH with a 'C.'

Table 4.4: Summary of age and metallicity measurements from the literature and this work (as listed in Table 4.3), as shown in Figure 4.6. Both Gu et al. (2018a) and Villaume et al. (2020) fit their spectroscopic data with alf, which produces metallicities and ages which are not directly comparable to those derived with Prospector. That said, given the old age of DF44 the two approaches are expected to provide fairly consistent results.

	SFH model/prior	Data	[Z/H]	[Fe/H]	$t_{\rm mass}$ (Gyr)	$t_{\rm light}$ (Gyr)
G18	1 SSP	MaNGA	$-1.07^{+0.03}_{-0.03}$	$-1.25^{+0.33}_{-0.39}$		$10.47^{+2.65}_{-5.79}$
		MaNGA + $g-r$	$-1.45^{+0.34}_{-0.20}$	$-1.25^{+0.35}_{-0.41}$		$8.91^{+3.49}_{-4.10}$
V20	1 SSP	KCWI	$-1.19^{+0.06}_{-0.06}$	$-1.33^{+0.05}_{-0.04}$		$10.20^{+0.70}_{-0.90}$
	2 SSP	KCWI		$-1.29^{+0.03}_{-0.02}$	$9.70^{+1.10}_{-0.90}$	$9.80^{+0.90}_{-0.70}$
This work	Nonpar., extended	KCWI + UV–MIR	$-1.18^{+0.01}_{-0.01}$		$10.20^{+0.34}_{-0.48}$	
		KCWI (S/N = 15) + UV–MIR	$-1.13^{+0.05}_{-0.06}$		$9.34^{+0.58}_{-1.07}$	
		UV–MIR	$-1.46^{+0.47}_{-0.33}$		$6.78^{+1.03}_{-1.13}$	
	Nonpar., concentrated	KCWI + UV–MIR	$-1.27^{+0.03}_{-0.02}$		$13.06^{+0.02}_{-0.04}$	

Villaume et al. (2022), and Gu et al. (2018b).[8] The same data is shown in Table 4.4. Both previous studies fitted rest-frame optical spectra of DF44 with the full-spectrum fitting code alf. We caution that there are fundamental differences between alf and Prospector which make their results only broadly comparable: *e.g.*, the inclusion of non-solar abundance patterns (as mentioned above), and alf fits a single-age stellar component (with a uniform prior with minimum age of 1 Gyr) rather than an SFH. That said, the luminosity- and mass-weighted ages should be comparable, given that DF44 is old.

Villaume et al. (2022) fitted the same KCWI spectrum as this work, while Gu et al. (2018b) fitted a MaNGA spectrum which covers a broader wavelength range (including several additional age diagnostics: Hδ, Hγ, Ca II H and K, and G-band). The MaNGA spectrum has S/N ~ 8 Å^{-1}, however, which is only $\sim$12% the S/N of the KCWI spectrum. Despite differences in data, the two studies both found the age of DF44 to be $\sim$10.5 Gyr,

[8]The stellar 'isochrone' metallicity (distinct from that which includes the response function for individual elements) [Z/H] values from Villaume et al. (2022) and Gu et al. (2018b) were provided via private communication.

although the stellar metallicities are formally discrepant.[9] Notably, Gu et al. (2018b) also considered the $g-r$ colour of DF44 from Dragonfly imaging, and re-weighted their posteriors, which considerably lowers their metallicity value (and is then consistent with Villaume et al. 2022 owing to its large uncertainty). In degrading the S/N of the KCWI spectrum and refitting the data, we obtain a similar age as Gu et al. (2018a) (see Appendix B.2.3 for a discussion of the S/N dependence of the results).

Considering that we fit DF44 in a completely independent way compared to these studies, it is at least encouraging that the results are fairly similar. Significant variations among age and metallicity measurements for the same object, measured between different studies, is not unique to DF44. In Appendix B.2.4 we outline two additional examples and discuss the reasons behind their differences.

The comparison shown in Figure 4.6 demonstrates the difficulty in measuring the stellar properties of old stellar populations, related both to limitations of data and modelling. As discussed in the previous section, a solution is within reach as the inclusion of a variable α-abundance or the addition of mid-IR photometry would help to break degeneracies between the stellar population properties.

We conclude that DF44 has an age of $\sim$10–13 Gyr. Without clear statistical evidence to favour one SFH model over the other, throughout the remainder of this work we present both sets of results. In the next section, we discuss the implications of such a large sized galaxy having formed the bulk of its stellar mass very early.

4.5 Discussion

In this work, we sought to measure the detailed SFH of DF44 as a means to distinguish between UDG formation scenarios, which predict a variety of quenching times (*i.e.*, SFHs). The consistent narrative among theoretical simulations is that UDGs are contiguous with the canonical dwarf population. However, Villaume et al. (2022) established that DF44 is dissimilar to canonical dwarf galaxies with respect to the stellar population gradients, stellar metallicity, and kinematics. In measuring the SFH of DF44, we can further test this scenario.

Previous analyses of DF44 found that its stellar population is old, having an age of $\sim$10 Gyr (see Figure 4.6; Gu et al., 2018b; Villaume et al., 2022). In this work, we have

[9] Villaume et al. (2022) considered the presence of a second young population (aged 1–3 Gyr), which lowers their age estimate by 0.6 Gyr but is consistent with their fiducial fit.

shown that DF44 formed the majority of its mass early, where we consider the galaxy 'quenched' after it forms ∼90% of its mass. In using an extended SFH prior, we obtain a lower limit of the quenching epoch of $z \sim 0.9$ (∼6.3 Gyr after the Big Bang). Alternatively, in using a concentrated SFH prior (motivated by the results of Villaume et al. 2022), we recover an extremely early quenching epoch of $z \sim 8$ (∼0.6 Gyr after the Big Bang). In either case, we find that DF44 is old, the distinction being that a concentrated SFH prior suggests that it is *very* old. Without clear statistical evidence to favour one prior over the other (see Section 4.4.2) we instead focus on providing a qualitative comparison of the implications of the two results.

For either of our two results, the bulk formation of DF44 occurs during an epoch where the evolution of galaxies in dwarf-scale dark matter haloes ($\lesssim 10^{11}$ $M_\odot$) significantly differs from that of galaxies in more massive haloes. The mass assembly histories expected for average galaxies with dark matter halo masses between 10^{11}–10^{13} $M_\odot$ are shown in Figure 4.7, from the empirical model of Behroozi et al. (2019).[10] The mass assembly history of DF44 (as shown in Figure 4.4) is shown for comparison.

While the current stellar mass of DF44 falls within the range expected for the $z = 0$ canonical central dwarf population, and its halo mass is in the neighbourhood of $\sim 10^{11}$ $M_\odot$ (*e.g.*, van Dokkum et al. 2016, 2019; Wasserman et al. 2019; see also Bogdán 2020), its mass assembly history is not necessarily compatible with this population. In fact, the mass of DF44 at $z \sim 8$ was typical for galaxies destined to become the brightest cluster galaxies (BCGs) – however, the mass growth was halted. This provides our first evidence that DF44 may not originate among the canonical field dwarf population.

We now further investigate our results in the context of the predictions of UDG formation scenarios from theoretical work.

4.5.1 DF44 in tension with UDG formation scenarios

There have been many scenarios that have come out of cosmological simulations and SAMs which satisfy the size and surface brightness constraints of the observed UDG population. As these scenarios all work under the same constraints of conventional galaxy evolution physics, they follow the same dichotomy: whether the significant size growth necessary to

[10]The population averages likely overestimate the star formation time-scales of galaxies in dense environments (*e.g.*, Thomas et al., 2010). Indeed, dEs and UDGs in clusters are found to be old (*e.g.*, Weisz et al., 2011; Ferré-Mateu et al., 2018; Ruiz-Lara et al., 2018), and satellite dwarfs in simulations appear to form much earlier than central dwarfs (Digby et al., 2019; Garrison-Kimmel et al., 2019; Joshi et al., 2021). However, DF44 is likely only on its first infall into the Coma cluster (discussed further below).

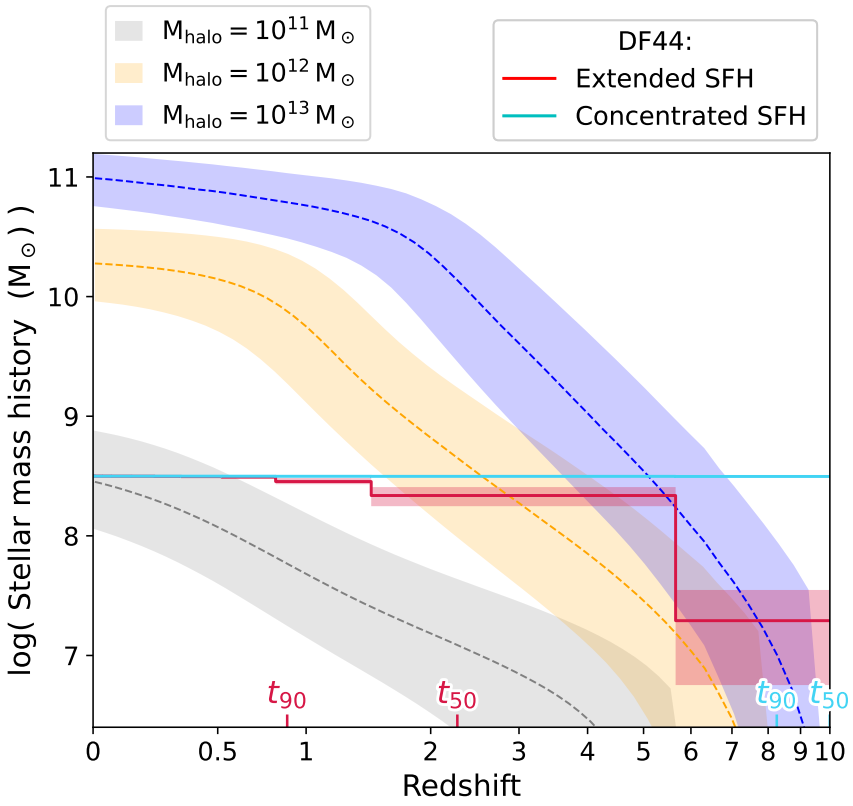

Figure 4.7: DF44's mass assembly history compared to total stellar mass histories of the main progenitor for haloes of several masses from the empirical model of Behroozi et al. (2019). Lines indicate the median values, and shaded regions correspond to the 68% CRs. The halo mass bins have a width of ±0.25 dex. Time-scales of DF44's SFH are labelled on the plot: when 50% and 90% of the final mass had been formed. At the time that DF44 quenched ($z > 0.9$) it was already more massive than expected for most normal dwarf galaxies.

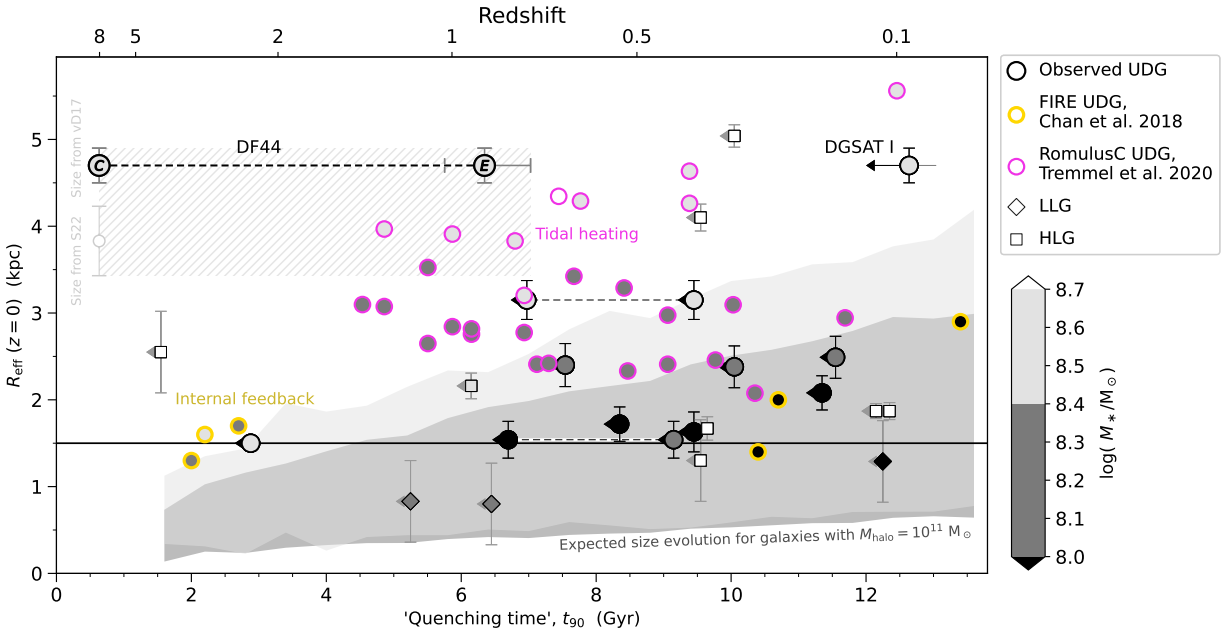

Figure 4.8: Comparison of effective radius and quenching time of UDGs (circles) between observations (black outlines) and predictions from two UDG formation scenarios (coloured outlines). Quenching times are shown in units of age of the Universe, such that $t_{90} = 0$ is the Big Bang. We show the quenching time results for DF44 as points with a 'C' (concentrated SFH prior) and with an 'E' (extended SFH prior). A hatched region fills the parameter space covered by the range in DF44's size and quenching time measurements. The distribution of dwarf galaxies in two mass ranges is shown with grey shaded regions. A horizontal line shows the size threshold for UDGs, $R_{\rm eff} > 1.5$ kpc. We show UDGs from two simulations: FIRE (yellow outlines; Chan et al. 2018) and RomulusC (magenta outlines; Tremmel et al. 2020), which are described in the main text of Section 4.5. *The large size and early quenching of DF44 are inconsistent with the internal feedback model of FIRE, but is not inconsistent with the tidal-heating model of* RomulusC (see text). Observations from the literature are shown with arrows, indicating that they may in fact be upper limits, given the potential bias in their measured SFH time-scales, see Section 4.5.2. Most of these UDGs fall within the expected size range of normal or tidally-enlarged dwarfs. We also include low luminosity galaxies (LLGs; diamonds), and high luminosity galaxies (HLGs; squares) from the literature. Dashed lines connect values obtained for the same object. Points are coloured according to their stellar mass.

transform a canonical dwarf galaxy into a UDG occurs pre- or post-infall (which is to say, whether the cluster environment is necessary to the process) is related to whether they are late or early infallers into that environment. Nearly[11] all of these models require the infall into the dense, hot environment of a cluster to quench the UDGs, such that the quenching time is directly linked to the infall time.

This dichotomy is demonstrated in Figure 4.8 where we show the relation between the effective radius and the quenching time for three different scenarios: i) what is typically expected under 'normal' conditions (grey shaded regions)[12], ii) the predictions for the 'internal feedback' scenario from the FIRE simulations (gold outlined circles; Chan et al., 2018), and iii) the predictions from the RomulusC simulations (magenta outlined circles; Tremmel et al., 2020)[13]. The symbols are additionally colour-coded by stellar mass.

We show our quenching time results for DF44 in this figure, with the size measured by van Dokkum et al. (2017) based on the deepest imaging available ($R_{\rm eff} = 4.7 \pm 0.2$ kpc; with black outlines). However, Saifollahi et al. (2022) obtained a smaller size ($R_{\rm eff} = 3.83 \pm 0.4$ kpc; grey point) based on the same data. A hatched region covers the parameter space of the various quenching times from our results, and the two size measurements.

The FIRE UDGs largely follow the expected size growth trend, where the distinction is that their internal feedback causes bursts of SF (which puff up the sizes of the galaxies, prior to quenching) which places them at the top end of the size distribution. While Chan et al. (2018) predicted that there are objects with quenching times as early as measured for DF44, these objects barely reach the nominal size of UDGs ($R_{\rm eff} < 2$ kpc). Indeed, they could not reach the size of DF44 without significantly more time to form stars, which would then violate the stellar mass/surface brightness constraint.

A formation scenario that can explain both large and early-quenched UDGs is tidal heating, where the expected size–quenching time trend is the exact opposite of the FIRE scenario, *i.e.*, the earliest infallers/quenchers will be the largest because they have spent the longest time expanding due to the cluster environment. We see this effect demonstrated in the RomulusC points.[14] While Tremmel et al. (2020) show that some objects reach the

[11]The exception is Wright et al. (2021) in which a small fraction of isolated UDGs which experience major mergers are quenched at $z = 0$. This scenario is discussed further below.

[12]The expected size growth was determined from the stellar mass assembly histories of Behroozi et al. (2019) and the size–mass relation of Sales et al. (2020).

[13]These values were taken from Figures 2 and 11 of Tremmel et al. (2020).

[14]We note that other simulations and SAMs which invoke tidal heating (*e.g.*, Carleton et al., 2019; Jiang et al., 2019a; Liao et al., 2019; Sales et al., 2020) with slightly different prescriptions (*i.e.*, cuspy vs. cored dark matter haloes, how early satellite infall begins) could change the exact predictions of the sizes of UDGs. We note that RomulusC appears to over-predict the sizes of cluster dwarfs by a factor of ∼2

nominal sizes of UDGs prior to infall into a cluster, in order to reach the large end of UDG sizes requires the additional effect of tidal heating from the cluster environment. With the extended SFH prior, the quenching time of DF44 is reasonably consistent with the tidally heated RomulusC UDGs.

Certainly tidal heating is happening on some level to some galaxies in clusters. Evidence of such has been observed among proto-UDGs in clusters (*e.g.*, Grishin et al., 2021). Carleton et al. (2019) interpreted the radial alignment of UDGs in Coma (Yagi et al. 2016, which includes DF44) as evidence that these galaxies have been tidally influenced. While we cannot discount the tidal heating scenario in explaining the size and quenching times of DF44, this scenario conflicts with other properties of DF44.

Measurements of the kinematics and dynamics of DF44 indicate that it has not been in the cluster environment long enough to be impacted by tidal effects. Its position in phase-space points to a late infall into the Coma cluster (< 2 Gyr ago; Alabi et al., 2018). Moreover, DF44 appears to be part of a dynamically cold group that would have surely been disrupted if tidal heating had taken place (van Dokkum et al., 2019), and there is no distortion in its ellipticity that would be a marker of tidal heating (Mowla et al., 2017).

Together with the above points, the SFH provides evidence that DF44 certainly quenched prior to cluster infall. This would suggest that its progenitor was larger than a dwarf galaxy, or that a process unrelated to environment caused an expansion. This interpretation is consistent with the conclusion of Saifollahi et al. (2022), who find that the elevated GC populations at a given stellar mass ($N_{\rm GC}/M_*$) of large UDGs (including DF44) are inconsistent with scenarios which explain the sizes of UDGs via redistributing the stars to larger radii (*i.e.*, tidal interactions, stellar feedback, or high-spin). Villaume et al. (2022) similarly ruled out such scenarios given DF44's 'inside-out' stellar population gradients. Therefore, *how* DF44 quenched is the crucial question to answer to understand its origins.

From simulations, only Wright et al. (2021, based on Romulus25; Tremmel et al. 2020) have proposed a scenario, 'early major mergers'[15], in which UDGs can form and quench[16] without relying on environmental quenching mechanisms. The UDGs in Romulus25 had their star forming gas and star formation moved outwards from the central cores of the galaxies to larger radii by major mergers ∼8–11 Gyr ago. For most of the simulated UDGs, star formation continued in the galaxy outskirts, while the central core passively dimmed, leading to negative radial age gradients.

(compared to observations from Eigenthaler et al. 2018).

[15]We note that Saifollahi et al. (2022) refer to this scenario as 'lack of late mergers.'

[16]Less than 5% of the simulated UDGs with masses $M_* > 10^8$ $M_\odot$ are quenched, in the sense that they are gas poor. This population is dominated by galaxies that have had an interaction with a more massive halo and/or AGN activity.

Considering that DF44 quenched $\gtrsim 7$ Gyr ago, this may suggest that a major merger is responsible for (or at least concurrent with) its quenching – and that there would be a flat age gradient. The central (< 0.5 kpc) SFH predicted for Romulus25 UDGs is broadly consistent with DF44's SFH when assuming an extended SFH prior, although not when assuming a concentrated prior (which quenches much earlier). Villaume et al. (2022) measured a flat-to-negative [Mg/Fe] gradient out to $\sim$2.5 kpc, which taken as a proxy for an age gradient is not strictly inconsistent with this scenario.[17]

Further work is needed in order to establish whether DF44 is the product of an early major merger. For instance, the mechanism that quenches $\lesssim 5\%$ of the Romulus25 UDGs is not fully described, providing no point of comparison with DF44's SFH or stellar population gradients. Moreover, when this quenching occurs, or whether the galaxies remain quenched, is unclear. While Wright et al. (2021) and Van Nest et al. (2022) explored the predictions of 'early major mergers' in differentiating average UDGs and non-UDGs, the fact that DF44 is a rare case warrants more detailed comparisons.

The results of this work show that DF44 has been shaped by some rare galaxy evolution process, no matter whether the 'true' SFH resembles our result with an extended or concentrated SFH prior, or falls somewhere in between. As was shown in Figure 4.7, the early SFR of DF44 is more typical of normal (MW-like) star forming galaxies at $z > 3$ (Rinaldi et al., 2021). The implication is that it is not the early, extreme SFH that makes DF44 unusual among $z = 0$ galaxies, but rather its sudden quenching. Given the lack of galaxies like DF44 in cosmological simulations, this would imply that galaxy evolution models are not capturing the true diversity of quenching mechanisms.

In fact, cosmological simulations already struggle to reconcile the opposing stellar mass–effective radius constraints for objects like DF44 in the context of the broader galaxy population. A common problem among cosmological simulations is that they do not accurately reproduce the population of normal-sized dwarfs (*e.g.*, Chan et al. 2018; El-Badry et al. 2016; Lupi et al. 2017; Tremmel et al. 2020; Benavides et al. 2021; see also Jiang et al. 2019a). Since this points to issues in the implementation of star formation and related feedback, the evidence from this work and Villaume et al. (2022) that there are objects like DF44 that require even more intense star formation feedback exacerbates this problem.

Analytic and semi-analytic models can avoid such issues to some degree. With respect to size, several UDG formation scenarios apply empirical distributions (*e.g.*, Carleton et al., 2019; Sales et al., 2020) but they are then subject to the likely bias of 'getting out what

[17]While Villaume et al. (2022) measured a flat age gradient, they note that given the limitations of modelling granular differences in old stellar populations, the [Mg/Fe] gradient is more sensitive to age variations.

they put in' (see Jiang et al., 2019b). With respect to star formation and feedback, Danieli et al. (2021) analysed the large number of GC candidates hosted by NGC 5846_UDG1 (Forbes et al., 2021) with a model that connects the evolution of a galaxy with its dark matter halo and GC populations (Trujillo-Gomez et al., 2019) to show that it is plausible that clustered supernova feedback could significantly increase the mass loading factor of gas outflows. However, these models miss an important component of galaxy evolution – the impact of the different environments a galaxy moves through over its lifetime. DF44's very early quenching and relatively late infall into the Coma cluster invokes the question of what has it been doing for the last ∼10 billion years? Given the potential 'pre-processing' by group environments or filaments that can affect everything from the size of a galaxy's dark matter halo, to its SFH and present-day GC population, makes it vital to understand this aspect of galaxy evolution in general.

4.5.2 DF44 in context

The prior-dependence of the SFH for old stellar populations, even with high-S/N data, means that further work is needed to understand what 'good' SFH priors are for these systems. The problem is amplified at lower S/N, where the prior will have a stronger influence on the posteriors (see Appendix B.2.3 for an example). Consequently, it is not straightforward to compare results between studies in the literature. With this caveat in mind, we also show in Figure 4.8 the quenching times and sizes of UDGs from three studies (Ferré-Mateu et al., 2018; Ruiz-Lara et al., 2018; Martín-Navarro et al., 2019), and for comparison high- and low-luminosity dwarfs in Coma (squares and diamonds, respectively; Ferré-Mateu et al., 2018). Arrows attached to these points indicate that they are perhaps upper limits, given potential biases from the use of regularized SFHs (akin to the extended SFH prior used in this work; see the discussion in Appendix B.2.4). We note that the UDGs from the literature are shown with effective radii from the catalogue of Alabi et al. (2020) when possible, where DF44 was found to have a size of 3.74 ± 0.23 kpc in the Subaru/Suprime-Cam R-band.

Regardless of potential biases in the SFHs, there are still interesting conclusions to draw from this data set. DF44 stands out as an outlier among the largest observed UDGs with an early quenching time, for any of the discussed quenching times or sizes. On the other hand, the UDG DGSAT I stands out with both the largest size and latest quenching time among the literature values shown in Figure 4.8, and it is also the only non-cluster member. Unlike the rest of the UDGs, DGSAT I is similar to a subset of the RomulusC UDGs, which follow a trend in size–quenching time in distinct disagreement with the standard expectations of tidal heating. Its size is also well outside what is plausible for the

concentrated SFH scenario, or normal expectations of size growth given its late quenching time.

While it is outside the scope of this work to examine DGSAT I in detail, it is relevant to this discussion in that it further provides evidence that multiple observed objects, all of which are 'UDGs,' in fact have distinct formation pathways.

That DF44 attained a similar stellar mass and size as the other large galaxies, but much earlier, supports the idea that it is either the product of unconventional galaxy evolution processes, or it was interrupted from becoming a much more massive galaxy by some catastrophic quenching event. Speculation of the latter has also been drawn on the basis of the wide range of GC counts among UDGs, and the range of implied dark matter halo masses (with some having little to no dark matter). This is the first time this diversity has been shown in the SFHs of the galaxies' field star populations.

4.6 Summary

In this work, we simultaneously fit NUV to NIR photometry and high S/N rest-frame optical spectroscopy of the UDG DF44 with an advanced physical model. Our model includes nonparametric SFHs, a flexible dust attenuation law, a spectral noise model, and an outlier model, which we fit to the observations in a fully Bayesian framework with Prospector.

We find that DF44 formed the majority of its stellar mass ($> 90\%$) early, although how early is sensitive to the choice of the SFH prior and degeneracies between stellar population parameters. Using an extended SFH prior akin to similar studies in the literature (which strongly favours ages of half the age of the Universe, and therefore disfavours very old ages) we find that DF44 formed by $z \gtrsim 0.9$. If we instead adopt prior knowledge from DF44's stellar population gradients that the DF44 formed early and rapidly quenched (Villaume et al., 2022), such that its SFH is concentrated at short timescales, we find that DF44 assembled as early as $z \sim 8$. Neither of these priors encode physical information of the shape of the SFH based on a priori knowledge, and thus neither are necessarily 'good' priors. Further work is needed to understand what 'good' SFH priors are for such old galaxies from a theoretical standpoint. Even with the high-S/N spectral data used in this work ($\sim$96 Å^{-1}) the data showed no statistical preference for either result. Improved age constraints are possible with the inclusion of observations in the mid-infrared, in that this would pin down the dust attenuation, which in the NUV is degenerate with the contribution of old stellar populations. Improvements in the models (*e.g.*, including variable α-abundance) to replicate old and complex stellar populations are also needed.

DF44's early and short SFH determined from this work, together with previous results that DF44 is very metal poor for its mass, and that stellar population gradients indicate 'inside-out' formation (unlike kinematically- and morphologically-similar dwarfs; Villaume et al., 2022), points towards an unusual origin, likely distinct from the canonical dwarf population. UDG formation scenarios outlined in simulations only predict the SFH and size of DF44 through invoking prolonged environmental effects, yet we conclude that DF44 quenched prior to accretion into the Coma cluster. While analysis of the ROMULUS25 simulation by Wright et al. (2021) proposes early major mergers as a means to produce UDGs in the field, it is not yet clear if the properties of DF44 are fully consistent with this scenario. Instead, DF44 may be a 'failed galaxy' with its initial size, or whatever processes that expanded it, being unrelated to its environment. In Summary, early quenching and late infall taken together rules out most UDG formation scenarios except for the failed-galaxy and early-major-mergers (with the caveats above). Additional work is needed to explain the old quiescent UDGs from a theoretical standpoint, while reproducing the observed stellar properties beyond general size–mass trends.

Chapter 5

SFH prior dependence limits the uniform interpretation of galaxy properties: implications for population models and scaling relations

5.1 Introduction

Our understanding of galaxy evolution relies on our ability to constrain the properties of distant galaxies from multi-wavelength observations. Where data is limited, single colours or luminosities have been empirically calibrated to physical parameters (*e.g.*, Kennicutt, 1989; Kennicutt, 1998; Bell & de Jong, 2001), where often the calibrations are only valid for specific galaxy types. With more data, a better mapping between observations and galaxy properties can be determined by comparison to physically-motivated model spectral energy distributions (SEDs). The SED models are constructed by evolving models of stellar populations over an assumed star formation history (SFH) with a given initial mass function (IMF), and metal enrichment history (MEH), and producing a composite spectrum, which is then attenuated by an assumed dust attenuation model and further influenced by dust emission, gas properties, and active galactic nuclei (AGN; see Chapter 1 Section 2.3, and reviews by Walcher et al., 2011; Conroy, 2013). Historically, sparsely sampled SEDs of luminous local galaxies could reasonably be matched to simple SED

models (*e.g.*, Spinrad & Taylor, 1969; Bruzual, 1983; Guiderdoni & Rocca-Volmerange, 1987; Worthey, 1994; Bruzual A. & Charlot, 1993; Leitherer et al., 1996; Maraston, 2005; Vazdekis, 1999). Interpreting broadband and high-S/N observations for a diverse set of galaxies at a variety of redshifts requires more complexity in SED models. The advantage of SED-fitting approaches is they are flexible: all the contributing elements can be customized.

The accuracy and precision of the SED-model comparison are subject to both i) the ability of the model to describe the true properties of the galaxy, and ii) the ability of the observations to inform the model. The first point is related to the fact that SED models involve many assumptions, including model stellar isochrones, spectral libraries, IMF, SFHs, and dust physics. Modern SED-fitting codes pair complex, self-consistent SED models with efficient sampling algorithms such that many of these variables can be 'free' instead of 'fixed.' The model can adapt to best suit the data being fit, or otherwise follow a probability distribution based on realistic expectations. SED-fitting codes which adopt Bayesian statistics make these 'prior' assumptions explicit. Given that each of these assumptions introduces uncertainties on the galaxy properties (*e.g.*, Wuyts et al., 2009; Acquaviva et al., 2015; Simha et al., 2014; Salim & Narayanan, 2020) there has been a significant effort in the last decade to improve SED models or validate the model assumptions.

The choice of priors is critical and should be subject to careful consideration. A good prior is either well-calibrated to the truth or is strictly 'uninformative' in that it *does not* prefer any particular solution. Uninformative priors are difficult to implement in practice, however, as subtle degeneracies between parameters can lead to unintentionally complex correlations.[1] Parameters which are only weakly constrained can therefore be biased in unexpected ways.

The SFH model is a prior in the sense that it defines the parameter space of permitted SFR(t) forms. Several studies have identified that the choice of SFH model is particularly influential on the results (*e.g.*, Wuyts et al., 2009; **?**; Sobral et al., 2014; Simha et al., 2014; Acquaviva et al., 2015); in some cases more so than the observational uncertainties (*e.g.*, Pacifici et al., 2015; Iyer & Gawiser, 2017; Carnall et al., 2019b; Leja et al., 2019a; Lower et al., 2020). As discussed in Chapter 1, while a τ SFH model is perhaps a fair description of massive quiescent galaxies observed at low redshifts, it is a poor model for star forming galaxies, or galaxies at high redshifts. Simple SFH models can impose strong priors on other parameters, such that the relationships between parameters are biased;

[1]For example, there is a well-known degeneracy between age, dust, and metallicity among UV and optical colours, where the age–metallicity degeneracy is challenging for stellar populations older than 5 Gyr even with spectral constraints (Worthey, 1994).

fundamental cosmic relations inferred from the biased data are therefore wrong.[2] Given that mass-weighted age is a latent parameter of the SFH model and priors, it follows that our understanding for when galaxies form their stars is significantly influenced by these assumptions (*e.g.*, Carnall et al., 2019b; Leja et al., 2019a, 2020).

Some works attempt to circumvent this issue by adopting SFH models and prior tuned to reproduce the predicted galaxy properties from simulations (*e.g.*, Leja et al., 2019b) or empirical scaling relations that scale with stellar mass (*e.g.*, Suess et al., 2022). There is not necessarily a consensus among simulations as to what the dominant modes of energetic feedback are at high redshifts, and thus they can predict SFHs with a variety of timescales. For example, while the quenching models of Illustris TNG and Simba are broadly similar (Torrey et al., 2014; Vogelsberger et al., 2014; Davé et al., 2020),[3] Iyer et al. (2020) measure less power on short timescales (*i.e.*, smoother SFHs) from Illustris. Moreover, this practice limits the ability of the observations to provide independent constraints for theoretical models.

Testing the ability of SED models to accurately report galaxy properties with different choices of priors has been a focus of numerous studies over the last few years. Typically, observations are constructed for mock galaxies, which are compared to the SED-fitting results; *i.e.*, 'ground truth' comparisons (*e.g.*, Simha et al., 2014; Carnall et al., 2018, 2019b; Leja et al., 2019a; Lower et al., 2020; Pacifici et al., 2016b; Suess et al., 2022; Tacchella et al., 2022a; Ji & Giavalisco, 2022). The general conclusion of these works is that flexible SFH models are capable of providing accurate results *when constrained by good quality observations*, *i.e.*, spanning a broad range of wavelengths (UV–NIR) and the redshift and stellar metallicity are reasonably known. While some works aim to identify the 'best' set of priors for a given class of galaxy (*e.g.*, quiescent, star forming, or post-starburst), the issue of how to define the boundaries of where the prior is well-suited is often unaddressed. And given that the influence of the priors depends on the S/N of the observations, if there is a data-quality threshold beyond which the results are reasonably prior-independent.

With the problem of prior dependence in mind, rather than fit individual noisy observations, some studies opt to fit averaged observations which have lower uncertainties. It is generally acknowledged that interpreting averaged spectra can be misleading, particularly

[2]For examples, the cosmic star formation rate density, CSFRD, and stellar mass functions (Ciesla et al., 2017; Carnall et al., 2019a; Leja et al., 2019b). Gladders et al. (2013) demonstrated that the CSFRD could be recovered where more flexible SFH models were used.

[3]Each simulation adopts kinetic AGN feedback as the dominant mode regulating star formation, but the subtleties in their implementation lead to differences in the cold gas content of star forming galaxies at high-redshifts, in particular (Davé et al., 2020).

where there are non-linear correlations in the spectral flux over a range of parameter values. SEDs evolve non-linearly with time (Serra & Trager, 2007), for example, such that the spectral variance across a sample of galaxies with a uniform distribution of ages is non-Gaussian. So while the properties of the average spectrum do not directly correspond to the average of the properties of the contributing galaxies, they are often assumed at least demonstrative (*e.g.*, Khullar et al., 2022). The selection of which galaxies to combine presents an additional challenge, as ideally the selection would be such that there is a gain in S/N without averaging out the variations in the dataset.

In an attempt to bypass the issues of SED model specification altogether, recently a number of works have instead developed machine learning algorithms to construct a mapping between galaxy observables and their underlying properties. For example, neural networks can be trained with a sample of galaxies with known properties (*e.g.*, Davidzon et al., 2019; Lovell et al., 2019; Surana et al., 2020; Gilda et al., 2021). A training set based on high-quality galaxy observations can offer more accuracy than SSP templates. However, a diverse set of galaxies are required to build a fully representative sample. This poses challenges for studying high-redshift galaxies, or otherwise exotic galaxy populations not predicted by current theoretical models (*e.g.*, Dragonfly 44; see Chapter 4). There are other practical complications related to how to train the algorithm, or adapt the algorithm for incomplete datasets (*e.g.*, Gilda et al., 2021).

Instead of avoiding the issue of prior-dependent results, this work attempts to tackle the issue head on by constructing physically motivated and calibrated prior distributions through leveraging the combined information from high-quality observations. Moreover, if the priors are sufficiently characterized, the influence of a poorly-calibrated prior can be mitigated through 'importance sampling.' Hierarchical Bayesian models (HBM), or 'population models,' can be used to model the combined posteriors for large samples of objects (*e.g.*, Kelly, 2012). Some recent examples from the astronomical literature using population models include: Hogg et al. (2010b) construct a model to describe the eccentricities of binary-star and exoplanets, Foreman-Mackey et al. (2014) model the eccentricity distribution of exoplanets, Wolfgang & Lopez (2015) model the composition of planet candidates, Leja et al. (2020) model the stellar mass function of galaxies, Galliano (2018) and Nagaraj et al. (2021) model the dust properties of galaxies, Li & Leja (2022) model the optical colour to stellar mass relation, Sandles et al. (2022) model the star forming main sequence, and Whitler et al. (2023b) model the ages of high-redshift ($z \sim 6.8$) UV-luminous galaxies. This last example is particularly intriguing given the connection between galaxy ages and quenching timescales described in Chapter 3.

Besides providing a statistical framework for describing the probability distribution of a

given property, the inferred population distributions can be used to validate or improve the choice of priors assumed when SED-fitting. While it remains computationally infeasible to simultaneously fit SED to galaxies in a strictly hierarchical model (where the population distribution is constructed in tandem with the individual galaxy properties), a pseudo-hierarchical framework is achievable. The population distribution constructed based on a 'training set' could be used to update the choice of priors, which are then used to fit a second set of galaxies, and the process is repeated until all the galaxies in the set are fit. The final population distribution leverages the information from the full set of galaxies, without being as sensitive to the initial choice of priors, which may not be well-calibrated. The crux of this procedure is that the training set is representative of the population and that their likelihoods are well-calibrated to the true properties.

Following the procedure outlined in several works in the literature, which demonstrate the success of Bayesian hierarchical modelling in describing the distribution of properties as estimated via SED-fitting with Prospector (*e.g.*, Leja et al., 2020; Nagaraj et al., 2021; Whitler et al., 2023b), this work aims to model the diversity of ages among high-redshift quiescent galaxies. To validate this model, a set of mock observations are constructed from simulated galaxies, modelled after the GOGREEN quiescent galaxies studied in Chapter 3. While several studies characterize the biases imposed on galaxy ages, given different SFH models at $z = 0$ (Wuyts et al., 2009; Carnall et al., 2018; Leja et al., 2019a; Lower et al., 2020) or at $z \sim 1$ (*e.g.*, Carnall et al., 2019a; Suess et al., 2022; Tacchella et al., 2022a), they are limited to samples with higher S/N than GOGREEN, or different types of galaxies (*e.g.*, post-starbursts). The 'ground-truth' assessment of the ages inferred from a given SFH model in this work therefore complements such studies, and will be useful for upcoming surveys which will provide large numbers of galaxies with lower S/N. That said, we only focus on quiescent galaxies at $z \sim 1$. The challenge of characterizing SFH-model and prior-dependent age bias related to the more diverse population of galaxies is beyond the scope of this work. Even within a relatively uniform population of galaxies, however, we confront a number of issues.

The mock population constructed to validate this model is introduced in Section 5.2. The SED-fitting framework is described briefly in Section 5.3, where in Section 5.3.2 we describe the different SFH models explored in this work. The results of SED-fitting the mock observations are presented in Section 5.4. The Bayesian inference framework for the population model is then laid out in Section 5.5. In Section 5.6 the results of the population model are presented, and Section 5.7 the implications of this work.

Table 5.1: Properties of the mock quiescent galaxies. Notes: 1) Redshifts are assumed to be known, and are fixed in the model. 2) Metallicities are sampled from the local mass-metallicity relation (see text). 3) Photometric filters listed in Table 5.2. 4) The spectral continuum was not included in the fitting procedure, see Section 5.3.3).

Notes	Property	Values
Properties from Illustris TNG300		
1	Redshift	1.2
	sSFR($t < 100$ Myr)	$\lesssim 10^{11.2}$ yr^{-1}
	Mass-weighted age, $t_{\rm mass}$	1.6 – 4 Gyr
	$\log(M_{*,\ \rm total}/\rm M_\odot)$	10.5 – 11.3
	Sample size	100
Galaxy properties set from empirical relations		
2	$\log Z_*/\rm Z_\odot$	-2 – 0.2
	$\hat{\tau}_{\rm dust,\ diffuse}$	10^{-3} – 0.7
Observables		
	Photometric S/N	10
3	Photometric wavelength coverage	NUV–NIR
	Spectroscopic S/N	$\sim$1, 5, or 10 Å^{-1}
	Spectroscopic resolution	$\sim$8.2 Å
4	Spectroscopic wavelength coverage	3525–4400 Å, rest-frame (continuum not fit)

5.2 Constructing mock quiescent galaxies at $z = 1.2$

The SFHs of the mock galaxies are based on the predicted SFHs from the cosmological, hydrodynamical simulation Illustris TNG (Marinacci et al., 2018; Naiman et al., 2018; Nelson et al., 2018; Pillepich et al., 2018; Springel et al., 2018). The mock observations are constructed such that they reflect the same qualities of the real observations we study in Chapter 1 with respect to i) photometric filters, ii) spectroscopic wavelength range and resolution, and iii) redshift, with $z = 1.2$ being the median redshift of the GOGREEN galaxies. The details of how we build this mock population is provided in Section 5.2, and summarized in Table 5.1.

Stellar masses and SFRs are drawn from Illustris TNG300-1 (TNG300 hereafter), which combines moderate resolution with large volume to identify a large number of mock quiescent galaxies. We model this choice based on the GOGREEN study comparing the environmental trends between observations and simulations (Kukstas et al., 2023). The

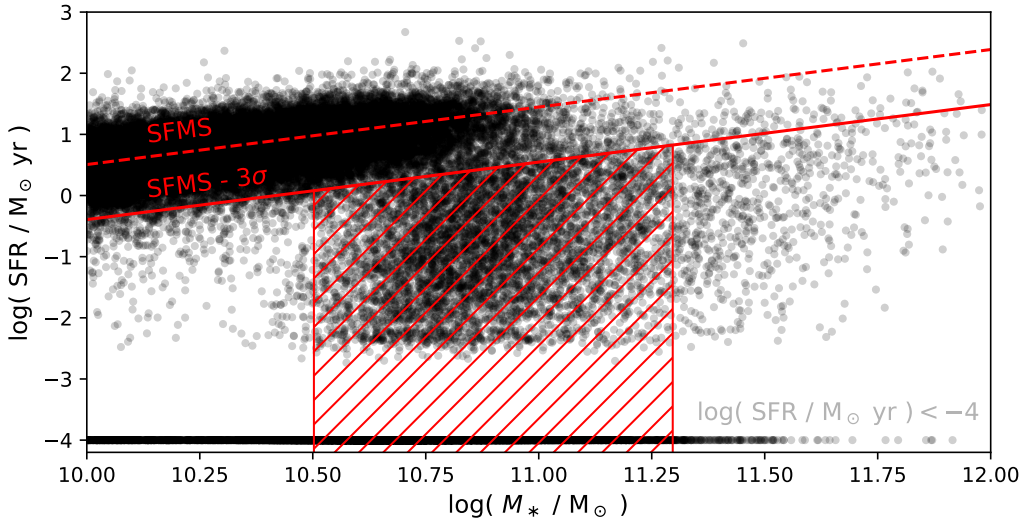

Figure 5.1: Selection of quiescent central galaxies from Illustris TNG300 at $z = 1.2$. The mock galaxies in our sample are selected to be below a 3σ threshold below the star forming main sequence (linear fit) and within the stellar mass range $10.5 < \log(M_*/\mathrm{M}_\odot) < 11.3$.

baryonic mass resolution of this box is 4.8×10^8 $\mathrm{M}_\odot$ h^{-1}, and the box size is 303 cMpc^3.

We identify all the subhalos in the $z = 1.2$ snapshot with `SubhaloFlag` > 0, and tag galaxies that are satellites of clusters, or are central subhalos in the field. The subhalos are then selected based on a stellar mass and SFR cut. Matching the mass range of GOGREEN, we identify subhalos in range $10 < \log(M_*/\mathrm{M}_\odot) < 11.8$. The star forming main sequence (SFMS) is then approximated by fitting a linear polynomial to the SFR of all the subhalos as a function of stellar mass, and quiescent galaxies are selected to be $< 3\ \sigma$ below the SFMS, and confirm that this selection reproduces the quenched fractions presented in Kukstas et al. (2023). Figure 5.1 shows the selection criteria. This selection identified 11634 massive quiescent subhalos in the field, and 264 satellite subhalos. Given that SFHs were qualitatively different at the lower and upper ends of the mass range, we then refined the mass selection to $10.5 < \log(M_*/\mathrm{M}_\odot) < 11.3$, and given the larger sample of subhalos in the field (8254 subhalos, compared to 111 satellites), considered this to be the source for our mock galaxies. Despite the different quiescent fractions between field and cluster subhalos, the SFHs of the selected subhalos were similar at masses $> 10^{10.5}$ $\mathrm{M}_\odot$, such that this decision will not affect our results.

For each subhalo, we constructed the star formation histories by tracking the star particle ages in 500 time bins to $t_{\mathrm{univ}}(z = 1.2)$. Mass-weighted ages were then calculated

following Equation 2.13. Given the computational expense of fitting galaxies with flexible SFH models, we selected 100 of these galaxies to be a diverse sampling of this population. Figure 5.2 shows the distribution of ages among the quiescent field galaxies and the final set, their SFHs, and quenching timescales related to how the sSFRs transition through the green valley. The ages span 1.6–4 Gyr, with formation times between 0.3–1 Gyr after the Big Bang, and specific star formation rates dipping below 0.1 $M_\odot$ yr^{-1} between 0.02–3.3 Gyr ago.

Stellar metallicities were assigned based on the empirical stellar-mass stellar-metallicity relation (MZR) from Gallazzi et al. (2005) calibrated to $z = 0$ galaxies. We double the width of this distribution to account for possible redshift evolution, although recent observations of passive galaxies at $1 < z < 1.4$ appear to be consistent with this relation (*e.g.*, Saracco et al., 2023). Metallicities are selected for the mock galaxies by sample from this MZR, according to the galaxy's stellar mass. We assume that the mock galaxies are generally free of dust, where dust extinctions are assigned at random by drawing from a Normal distribution with $\mu = 0$ and $\sigma = 0.25$, bounded by zero. This distribution is consistent with the diffuse dust measured for the GOGREEN quiescent galaxies in Chapter 3, and the expectation that quiescent galaxies are not particularly dusty.

SEDs for the mock galaxies are generated with the FSPS population synthesis codes, which is the same model used in Prospector when fitting the observations. We therefore overlook any potential issues related to how accurately the SPS templates reflect real galaxies. We confirm that the SEDs for the mock galaxies have rest-frame UVJ colours within the 'quiescent region' as defined by Muzzin et al. (2013b) for the redshift range of our mock observations (ignoring the horizontal and vertical delimiters, as suggested by Belli et al. 2019).

From the SEDs, mock observations were constructed for fourteen filters from the UV to MIR; the filters are listed in Table 5.2. The cSFR set of observations covers the NUV–NIR, although we also explore the added parameter constraints with the inclusion of longer wavelengths (MIR) or the exclusion of the UV. The choice of filters and spectroscopic resolution is based on the GOGREEN sample studied in Chapter 3.

The observed spectroscopy is based on the GMOS R150 filter setup, spanning 6398 Å to 10199 Å (observed-frame; ∼2900-4650 Å rest-frame at $z \sim 1.2$) with a pixel scale of 3.9 Å/pix. The model spectra are convolved to match the resolution of ∼8.2 Å, or ∼19 FWHM.[4] Only the wavelength region 3525-4400 Å (rest-frame), which includes useful

[4]The model spectra are smoothed to the same resolution as the mock spectra prior to calculating the model likelihood.

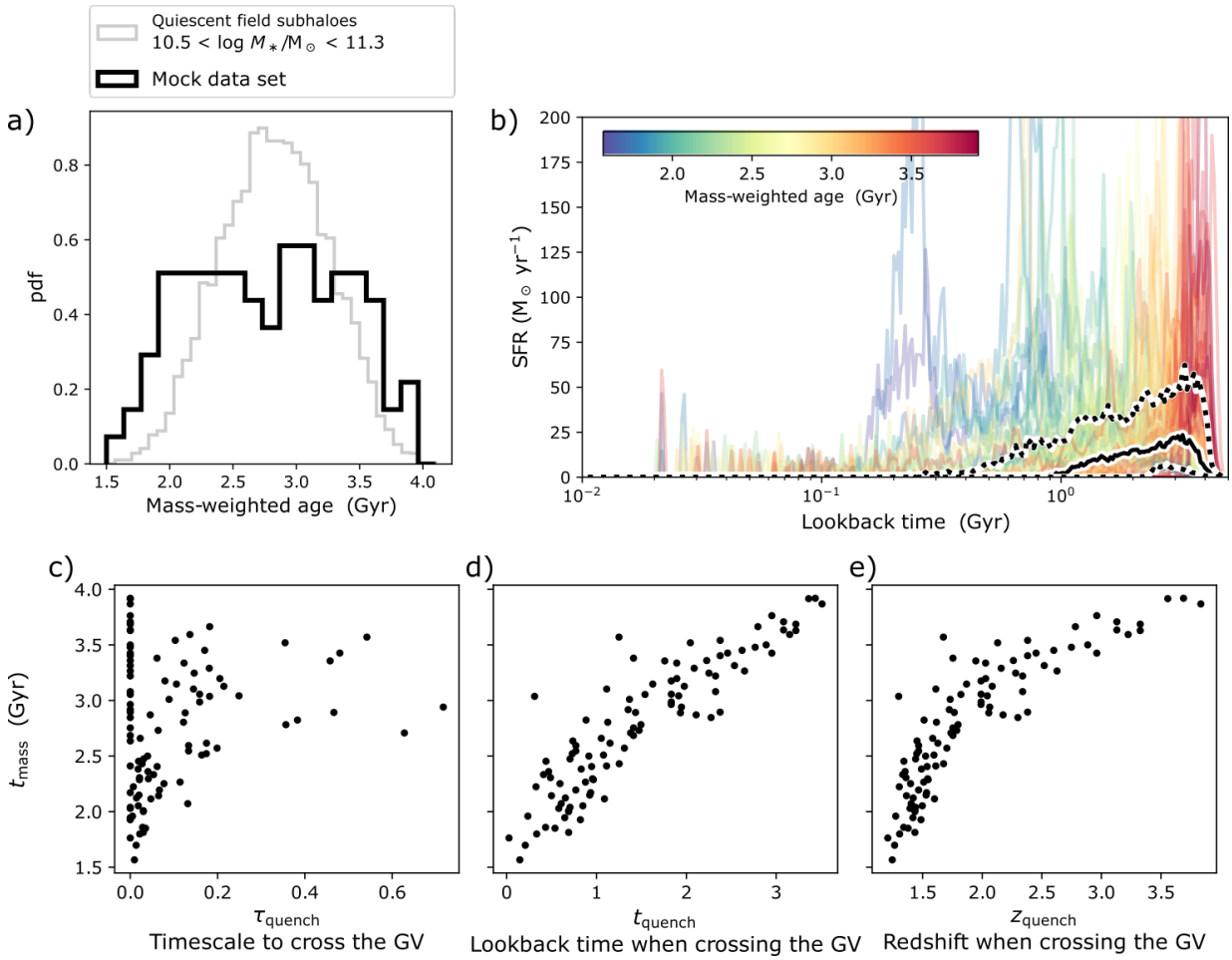

Figure 5.2: SFHs and ages of Illustris TNG300 quiescent galaxies. a) The distribution of mass-weighted ages of the full sample of quiescent field subhalos, within the desired mass range, and the final sample selected to be a diverse sampling of this group. b) Star formation rates as a function of look back time for the final sample, with the median shown as a thick line, and 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles shown as dashed lines. c,d,e) The mass-weighted age as a function relative to the time that each SFH takes to cross the green valley (GV) region, degined as $\frac{1}{20\ t_{\rm univ}(z)} < \text{sSFR} < \frac{1}{3\ t_{\rm univ}(z)}$ (*e.g.*, Tacchella et al., 2022a), the median lookback time when the galaxy crosses the GV, and the corresponding redshift.

Table 5.2: Photometric filters the included in mock data set. The rest-frame wavelength ranges at the redshift of the mock galaxies ($z = 1.2$) are also listed. The filters listed with an asterisk are not included in the fiducial set of observations.

Filter	Instrument	$\lambda_{\rm eff}$ (Å)	$\lambda_{z=1.2}$
U	VIMOS	3741.25	UV
V		4300.51	
B		5460.20	
R		6443.79	
I		8166.90	
Z		8996.55	Optical
J	FOURSTAR	12361.81	
$J1$		10538.94	
K		21414.84	NIR
IRAC1	IRAC	35439.39	
IRAC2		44840.88	
IRAC3		57164.36	
*IRAC4		78285.09	MIR
*MIPS 24 μm	Spitzer MIPS	234336.49	

spectral features are found (*e.g.*, CaH+K, $D_n(4000)$, Hδ, G), is considered when comparing to model SEDs. The lower bound of 3525 Å is based on the limit of the MILES spectral templates, where the resolution of the templates is lower at bluer wavelengths. The upper bound corresponds to the wavelength region where telluric contamination is typically problematic. We also do not fit the spectral continuum, instead assuming that the flux calibration is uncertain beyond the S/N of each pixel. A spectrophotometric calibration model is used to marginalize over the spectral continuum prior to calculating the model likelihood (see Section 5.3.3).

The mock observations were assumed to follow the true SED model, with uncertainties scaled to the desired S/N. If instead Gaussian noise was included in the mock observations, additional random noise would be apparent in the fitting results, which could obscure comparisons. Mock observations including Gaussian noise are discussed further in Section 5.7.2. The mock data were fit with the following noise-levels: the uncertainty of the photometry is fixed at 10% (*i.e.*, $\mathrm{S/N_{phot}} = 10$), and the spectroscopy $\mathrm{S/N_{spec}} = 5$, 10, or 40 pix^{-1}($\sim$1.3, 5.3, or 10.2 Å^{-1}). While higher-quality observations exist within the literature at this redshift, the intent of this work is to explore how to obtain accurate results from SED-fitting data with low S/N.

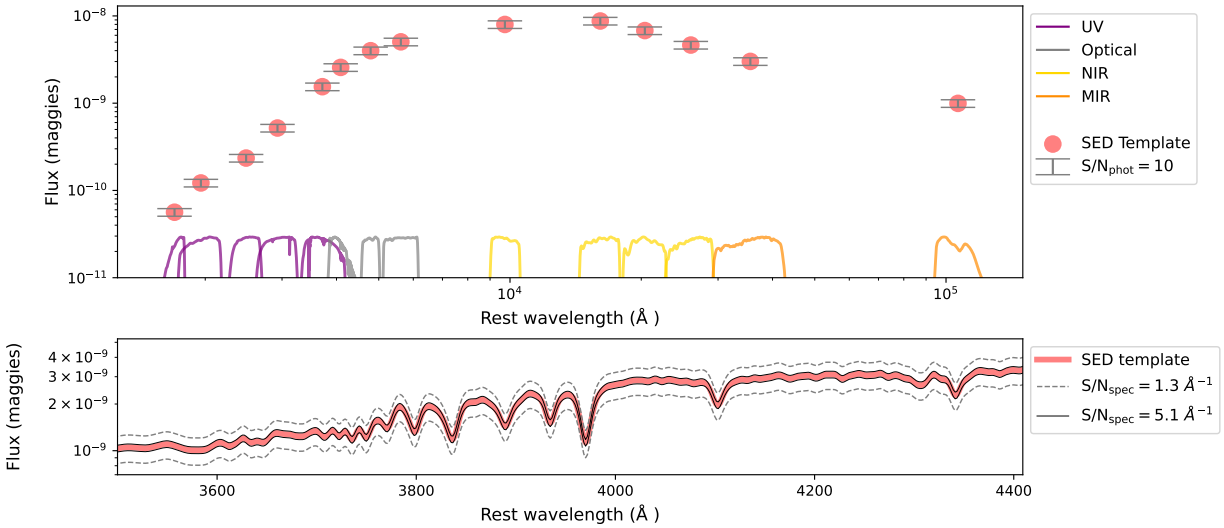

Figure 5.3: Example mock observation of a galaxy at $z = 1.2$. The template SED is shown in red, with error bars or black lines indicating various noise-levels assigned to the mock observations. The filter curves (see Table 5.2) are coloured according to their different wavelength regions. While the template spectrum is shown with its continuum in this figure, the shape of the continuum is not fit in the SED-model.

Figure 5.3 shows an example mock SED (red) and mock observations with different levels of noise (black error bars or lines). Coloured regions indicate different wavelength regions, where we compare results when fitting the photometry in only a subset of these regions (with or without the UV or MIR).

5.3 SED model and fitting procedure

The Prospector modelling and fitting framework is used to estimate the galaxy properties, as described in Chapter 1 Section 2.3. The specific model components are briefly described in this section. Table 5.3 summarizes the parameters included in the physical model, and their associated priors.

Table 5.3: Parameters of the SED-model, and their priors. (*) Only one SFH model is used at a time when fitting any given set of observations. Notes: 1) Redshift, fixed to $z = 1.2$. 2) Total stellar mass is the integral of the SFH, which includes the mass lost to outflows. 3) The total stellar metallicity, where scaled-Solar α-abundance is assumed. 4) Parameters of the delayed-τ SFH model. The FSPS manual lists the variable t_0 as `tage`, which we relabel to avoid confusion with the mass-weighted age. See Equation 2.11. 5) Logarithm of the relative SFR between adjacent time bins, where the SFH is a piece-wise constant function with N time bins. The cSFR prior assumes a flat, smooth SFH with $\mu = 0$ and $\sigma = 0.3$. The xSFR prior uses $\mu = [0., 0 - 0.5, -0.6, -0.9, -0.4, -0.1, 0.3, 1.0, 5.4]$ and $\sigma = 0.5$ (see text). 6) The Dirichlet parameters are related to the fractional sSFR in a given time bin via Equation 2.12, where the SFH is a piece-wise constant function with N time bins). The prior is a Dirichlet function, controlled by the parameter $\alpha_{\rm D}$. 7) Parameters for the dust attenuation model. 8) Parameters for the dust emission model.

Note	Parameter	Description	Prior
1	$z_{\rm obs}$	Redshift	
2	$\log(M_{*,\ \rm total}/{\rm M}_\odot)$	Total stellar mass formed	Uniform(min= 9, max= 12)
3	$\log(Z_*/{\rm Z}_\odot)$	Stellar metallicity	Uniform(min= -2, max= 0.19)
4	*Delayed-τ SFH model:		
	τ	e-folding time	LogUniform(min= 10^{-4}, max= 3)
	t_0	Time of initial star formation	Uniform(min= 10^{-4}, max= $t_{\rm univ}(z)$)
5	*Continuity SFH model:		
	$\log({\rm SFR}_n/{\rm SFR}_{n+1})$	Logarithm of SFR ratios between adjacent time bins	Student-T(μ, σ, $\nu = 2$)
6	*Dirichlet SFH model:		
	z_n	Dirichlet parameters describing the fractional sSFR in each time bin	Dirichlet($\alpha_{\rm D}$)
7	Dust attenuation:		
	$\hat{\tau}_{\rm dust,\ diffuse}$	Diffuse dust optical depth (eq. 2.6)	Uniform(min= 0, max= 2)
	$\hat{\tau}_{\rm young}/\hat{\tau}_{\rm dust,\ diffuse}$	Ratio of diffuse to birth-cloud dust optical depth (eq. 2.5)	Clipped Normal($\mu = 1$, $\sigma = 0.3$, min= 0, max= 1.5)
	$n_{\rm dust}$	Diffuse dust attenuation index	Uniform(min= -2, max= 0.5)
8	Dust emission:		
	$Q_{\rm PAH}$	Percent mass fraction of PAHs in dust	Uniform(min= 0.5, max= 7)
	$U_{\rm min,dust}$	Minimum starlight intensity to which the dust mass is exposed	Uniform(min= 0.1, max= 25)
	$\gamma_{\rm dust}$	Mass fraction of dust in high radiation intensity	LogUniform(min= 0.001, max= 0.15)
	Noise model:		
	$j_{\rm spec}$	Spectral noise inflation term	Uniform(min= 1, max= 3)
	$f_{\rm outlier,\ spec}$	Outlier fraction of spectral pixels	Uniform(min= 10^{-5}, max= 0.5)
	Spectrophotometric calibration model:		
	c_n	Polynomial coefficients, $n = 4$	

5.3.1 The physical model

The free parameters describing the stellar properties in our model include: total stellar mass formed (which includes the mass components lost through evolution, *e.g.*, through stellar feedback and winds) and stellar metallicity. A limitation of the FSPS models is that the stellar metallicity is fixed to the same value for the entire stellar population, and over time (*i.e.*, there is no metallicity enrichment history), and scaled-solar abundance are assumed. Neither of these limitations are discussed in detail in this work as the dataset cannot constrain either aspect. Both the total stellar mass (hereafter referred to as simply stellar mass) and stellar metallicities are given log uniform priors, as listed in Table 5.3.

The dust is modelled with the two-component dust attenuation model (see Section 2.3.4), which separates the dust components between those associated with the birth-cloud and a uniform dust screen. The birth-cloud dust acts to only attenuate stellar emission for stars younger than 10 Myr, while the diffuse-dust acts as a uniform screen with a variable attenuation curve (Noll et al., 2009). The free parameters of the model include the diffuse dust normalization constant, $\hat{\tau}_{\text{dust, diffuse}}$, the birth colour normalization constant, $\hat{\tau}_{\text{young}}$, and the diffuse dust attenuation index, n. The diffuse-dust constant and the ratio of the diffuse-to-birth-cloud dust are variables in this model, $\hat{\tau}_{\text{young}}/\hat{\tau}_{\text{dust, diffuse}}$, which avoids degeneracies between the parameters. The diffuse dust constant is given an uninformative prior, $\hat{\tau}_{\text{dust, diffuse}} \sim \text{Uniform}(\text{min} = 0,\ \text{max} = 1.5)$, and the dust ratio prior, $\hat{\tau}_{\text{young}}/\hat{\tau}_{\text{dust, diffuse}} \sim \text{Normal}(\mu = 1, \sigma = 0.3, \text{min} = 0,\ \text{max} = 1.5)$ which broadly follows the results among the literature for massive galaxies while allowing for some variation. The prior on the diffuse dust attenuation index is assumed to be $n \sim \text{Uniform}(\text{min} = -2,\ \text{max} = 0.5)$. Dust emission is calculated assuming energy conservation, where a permissive set of priors for the dust emission parameters were assumed, as listed in Table 5.3.

In this work, the focus is limited to studying the recovery of SFHs and ages from SED-fitting techniques for quiescent galaxies. Contributions from AGN and nebular emission in the SED model are therefore neglected. In fitting spectroscopy, however, the [O II] spectral line is masked similarly as it would be for real data, given that [O II] emission is not uncommon among quiescent galaxies (*e.g.*, see Chapter 3).

5.3.2 Star formation history models

The SFH models in Prospector cannot be completely agnostic to those of the galaxies being fit. It is either assumed that the SFH follows a functional form, or can be reasonably described as a piece-wise step function (*i.e.*, a nonparametric model). With a large enough

number of parameters or time bins, the two approaches become equivalent. In practice, nonparametric models are generally more flexible with fewer parameters. In either case, the parameters of the models are assigned priors which favour what are thought to be realistic SFHs. Mass-weighted age is a parameter derived from the SFH model, where its implicit prior is expressed by both the form of the SFH model and the priors for the SFH parameters.

In this work, SFH models widely used in the literature are studied, both parametric and nonparametric. The bias imposed on the galaxy ages given the associated priors of the models is explored in detail. That said, this work does not perform a comprehensive investigation of SFH models (a number of which are discussed in detail in Leja et al. 2019a and Carnall et al. 2019b; see also Lower et al. 2020) as the intention is not to select the best SFH model suited for a particular set of galaxies. Rather, the SFH models were chosen on the basis of having distinct implicit priors on the stellar ages (as discussed in Section 2.3.9), and attempt to characterize the biases, as needed to build unbiased population models of the ages (see Section 5.5). The SFH models included: a delayed-τ parametric model, a continuity nonparametric model, and a Dirichlet nonparmetric model.

Our discussion of the delayed-τ parametric SFH model is mainly to demonstrate its failures in accurately modelling quiescent galaxy SFHs and ages. Several works in the literature have already shown that the star formation timescales derived from delayed-τ models are biased, compared to more flexible models (*e.g.*, Carnall et al., 2018, 2019a; Lower et al., 2020; Tacchella et al., 2022a) yet this model is still common among recent galaxy studies (*e.g.*, Khullar et al., 2022; Annunziatella et al., 2023). As discussed in Section 2.3.9, the delayed-τ characterizes the SFH as starting at a time t_0, and thereafter evolving with an e-folding timescale, τ.

For the Continuity nonparametric model, two sets of priors are assumed. The 'cSFR' prior follows from Leja et al. (2019a), where the preference is for smooth SFHs with a probability model for the transitions in SFR between time bins matched to $\sim$30,000 simulated SFHs with stellar masses of 10^9 $M_\odot$ from Illustris at $z = 0$ (Vogelsberger et al., 2014; Diemer et al., 2017) – a sample dominated by star forming galaxies. A Student-T distribution was used to tune the probability of logarithmic transitions in SFR between adjacent time bins has the parameters $\mu = 0$ (preference for a constant SFR) and $\sigma = 0.3$ (*i.e.*, the strength of this preference; *i.e.*, controls the smoothness of the transitions in log SFR). The preference for smooth SFHs imposes similar constraints as does regularization schemes in models which fit linear combinations of SSPs (Ocvirk et al., 2006a; Tojeiro et al., 2007), which are still used in studies of high-redshift galaxies (*e.g.*, Belli et al., 2019; Morishita et al., 2019). While not explicitly tuned to reproduce the predicted SFHs of high-redshift or quiescent galaxies, Leja et al. (2019a) demonstrate that it can reasonably reproduce a

variety of different SFHs. That said, they also note that the ages of SFHs which have sharp transitions in SFR (*e.g.*, the 'sudden quench' scenario) will likely be overestimated under this model. Several works have since adopted its use when studying quiescent at a number of redshifts (Leja et al., 2019b; Webb et al., 2020; Cutler et al., 2022; Sukay et al., 2022; Whitler et al., 2023a), or a slightly modified version with $\sigma = 1$ (Tacchella et al., 2022a; Whitler et al., 2023b,a).

The alternative prior assumed for the Continuity model, 'xSFR,' is explicitly tuned to the mean SFHs of Illustris TNG300 galaxies, from which the mock galaxies were selected. The simulated SFHs are binned according to the time bins defined for our SED model, and the median $\log(\mathrm{SFR}_n/\mathrm{SFR}_{n+1})$ for each time bin, n, was measured. This sets the Student-T parameter $\mu = \{0.0, -0.55, -0.60, -0.85, -0.37, -0.11, 0.26, 1.02, 5.37\}$. Given the diversity in SFHs among this sample, a more permissive standard deviation $\sigma = 0.5$ was adopted. The marginalized prior for the mass-weighted age for this xSFR prior was shown in Chapter 1 Figure 2.4 (dashed line) relative to the cSFR prior (thick line). The idea behind this 'tuning' is similar to schemes used in the literature (*e.g.*, Suess et al., 2022; Setton et al., 2022) which tune the Continuity model based on the predicted SFHs from the UniverseMachine (Behroozi et al., 2019). Whether this explicit tuning actually leads to accurate ages for a diverse sampling of galaxies among the 'tuning set' is tested in this work; although fundamental issues with this practice are discussed in the introduction of this chapter.

Dirichlet nonparametric models are also common in the literature (*e.g.*, Leja et al., 2017; Lower et al., 2020; Ji & Giavalisco, 2022; Webb et al., 2022; Qiu et al., 2023), in part because they provide a more direct conceptual mapping between the model parameters and the distribution of star formation over time. In Chapter 4 (*i.e.*, Webb et al., 2022) it was argued that a Dirichlet prior was a more natural choice when fitting a galaxy that appeared to have a distinctly non-smooth SFH based. Lower et al. (2020) suggest that a Dirichlet prior with $\alpha_\mathrm{D} = 0.7$ best reproduces the stellar masses, dust properties, and mass-weighted ages for a set of mock star forming and quiescent galaxies based on the Simba simulation at $z = 0$, while Ji & Giavalisco (2022) argue that $\alpha_\mathrm{D} = 1$ is a good prior based on a set of mock quiescent galaxies from TNG100 at $z = 2$. This work sets $\alpha_\mathrm{D} = 0.2$, which has a strong preference for an unequal distribution of star formation among the time bins, *i.e.*, 'burstier' star formation (with timescales according to the width of the time bins). There is no physical motivation for the choice of $\alpha_\mathrm{D} = 0.2$, although it appears a reasonable choice for describing SFHs dominated by early star formation (*i.e.*, fewer than half the time bins), and it provides a strong contrast to the Continuity models in the sense that it does not prefer smooth SFHs (more so than α_D= 0.5, for example). As such, the constraints on the recent star formation are largely-decoupled from the early SFH (although the degree of the

coupling will depend on the number and relative widths of the time bins).

For this work, fixed time bins are defined in units of lookback time, spaced so that the first four bins correspond to 0–0.03, 0.03–0.1, 0.1–0.5, and 0.5–1 Gyr, there are $N - 5$ age bins spaced linearly between 1 Gyr–$0.95\times$ $t_{\rm univ}(z)$, and the final bin covers $0.95\times t_{\rm univ}(z)$–$t_{\rm univ}(z)$, where $t_{\rm univ}(z)$ is the age of the Universe at the observed redshift of the galaxy. Defining the bins this way allows sufficient precision in capturing recent star formation, as well as a maximally old population. Note that because the time bins are fixed in lookback time, galaxies observed at different redshifts will have different age bins (because the age of the Universe will differ). For our preliminary analysis with mock galaxy observations, which have fixed $z_{\rm obs} = 1.2$, this is not of concern.

For the nonparametric models, the time that SF begins is fixed at the age of the Universe. Since the SFR at the earliest timescales (largest lookback times) is the most difficult to discern (*e.g.*, Conroy, 2013), the SFH at early times is most sensitive to the choice of SFH prior. For example, Whitler et al. (2023b) discuss the dependence of SED-fitting derived stellar masses for a sample of UV-luminous galaxies at $z \sim 7$ on the choice of SFH model, and demonstrate that the results differ depending on when star formation is allowed to begin (see their appendix A). While star formation is allowed to begin at the Big Bang, this assumption cannot be tested with the quality of observations typically available. Assuming a later formation time (*e.g.*, $z \sim 10$) would produce younger ages, without significant penalty to the overall likelihood of the model. In comparison, most parametric models allow the formation time to be a free parameter, including the delayed-τ model considered in this work.

5.3.3 Observational systematic models

While the mock observations were built from simulated galaxies using the same SED templates as used in fitting the observations, a noise model and a spectral outlier pixel model are included in the SED model, as would be used for real observations (see Section 2.3). The noise model is used to uniformly increase the spectral uncertainties by a multiplicative factor (with a prior against large values) in the case of model mismatch. Similarly, the outlier model can mask spectral pixels which have large deviations from the SED model.

The spectral continuum of the mock observations is assumed to not be flux calibrated. The spectral continuum is removed from the mock observations prior to fitting, and thus the continuum of the SED model also needs to be marginalized over prior to calculating the likelihood. Instead, the normalization and shape of the SED model should be informed by the photometry, while the spectroscopy provides information of small scale features

(which in the wavelength range considered includes Hδ, CaK+H, $D_n(4000)$, etc.). At each likelihood call, a Chebyshev polynomial is fit (via optimization) to the residual between the normalized model and the spectrum, which the spectrum is multiplied by before calculating the likelihood. A fourth order polynomial is used, which is flexible enough to describe the spectral continuum without over-fitting, and removing real features such as $D_n(4000)$. Subtle regularization terms are adopted in order to avoid the polynomial diverging from the spectrum-ratio at the edges of the wavelength range.

With these additional systematic models, the likelihood equation follows Equation 2.4.

5.3.4 Sampling

The complete model includes 12 parameters, exclusive of the SFH model parameters (which include 2–9 additional parameters; see Table 5.3). A dynamic nested sampling (Higson et al., 2019) algorithm `dynesty` (Speagle, 2020)[5] is used to sample the parameter space of the model and build posterior distributions. Given the motivation of this work was to explore the effect of the SFH prior on the posteriors, a high effective sample size (ESS) was needed to fully describe the shape and extent of the posteriors. The stopping criteria was set to ESS=10,000.

Throughout this work, the uncertainties are reported as the 68% confidence regions (CRs; which corresponds to the 16^{th} to 84^{th} percentile range) of the posterior probability distributions.

5.4 SED-fitting mock galaxy observations

The mock galaxy observations were constructed to have low-S/N such that the influence of the priors on the results could be explored in detail. Four different SFH priors are used, as described in Section 5.3.2, where each is tuned to prefer a different SFR(t). In this section, the influence of the different tuning schemes on the recovery of the direct and latent model parameters is explored. The objective is not to find the 'best' prior, but to characterize any bias imposed on the parameters so that it can be 'corrected' in the population model outlined in the following section. Section 5.4.1 begins with a comparison of the posteriors to the true values for stellar properties which are directly parameterized in the SED model. The accuracy of the mass-weighted ages are then assessed in Section 5.4.2, and the degeneracies between parameters are discussed in Section 5.4.3.

[5] https://dynesty.readthedocs.io/en/latest/

5.4.1 SFH prior dependence of direct model parameters

Despite that priors are explicitly set for the direct model parameters, some parameters are also related to the SFH, and thus its prior. For example, the prior-dependence of the mass-weighted age can propagate to degenerate parameters (*e.g.*, stellar metallicities, dust). Stellar mass is also linked to the SFH through the M/L ratio, where several authors have noted how simple parametric functions lead to discrepant stellar masses (*e.g.*, Maraston et al., 2010; Pforr et al., 2012). Figure 5.4 compares the posterior medians and true values for these parameters, recovered for observations of varying spectral S/N, fit with the four SFH models. Contours are drawn about the medians of the posteriors, and coloured according to the spectral S/N of the observations. The wavelength is fixed to NUV–NIR and photometric S/N $= 10$ (as is the case for all the observations discussed here, unless otherwise stated).

The contribution of the prior is expected to be larger where the observations are less informative (*i.e.*, have lower S/N). For each of these parameters, the explicit prior in the model is 'uninformative:' either uniform, or uniform in logarithmic space. Systematic offsets between the true values and median posterior values are apparent among the majority of low-S/N results. The fact that the posteriors for the lesser S/N data based on different SFH priors, are somewhat different suggests that there is a mild influence of the SFH prior. The biases are small, however: total stellar masses are within ∼0.1 dex, diffuse dust within ∼0.15 dex, and stellar metallicities within ∼0.25 dex. In Section 5.4.2 the correlation between these biases and that for the mass-weighted ages are discussed in more detail.

For the results from the cSFR Continuity prior, the stellar and dust properties were generally accurate where the observations had S/N $\gtrsim 10$ Å^{-1}. Otherwise, the offsets were larger than the uncertainties in the posteriors. This prior tends to overestimate the total stellar mass and diffuse dust, while underestimating the metallicity. In comparison, the posteriors from the Dirichlet($\alpha_{\rm D} = 0.2$) or xSFR Continuity prior were noisy, but not biased. SFH priors which prefer (or do not penalize) the true properties of a galaxy are labelled *commensurate*.[6] The strength of the bias imposed by an incommensurate prior is related to how strongly the true solution is disfavoured. The demand on the observations to constrain the true properties is therefore greater when using an incommensurate prior (*i.e.*, higher S/N observations are necessary). The delayed-τ prior, on the other hand, strictly disallows some solutions (*i.e.*, SFHs which cannot be described by the functional form of this model), such that increasing the S/N does not always produce more accurate results.

[6]Note that this does not imply that the Dirichlet($\alpha_{\rm D} = 0.2$) or xSFR Continuity prior are uniformly commensurate with all the model parameters, or with all the galaxies in the sample.

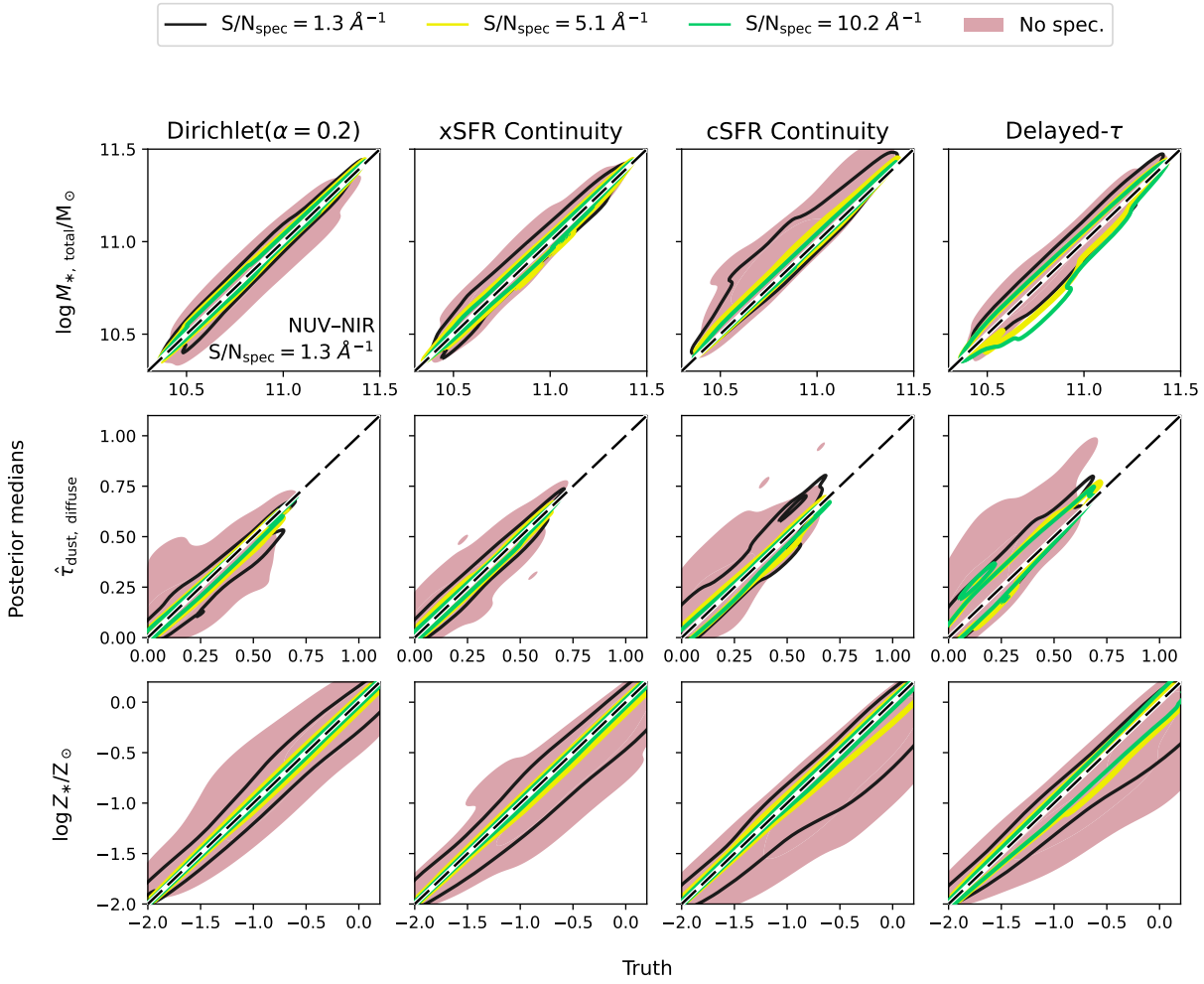

Figure 5.4: Stellar mass, diffuse dust normalization constant, and stellar metallicity posteriors, based on fitting data with different spectral S/N (contour colours; see legend) and different SFH models, compared to the true values. Contours are drawn around the posterior medians, and a dashed line shows the 1:1 relation. The accuracy of the posteriors improves with increasing spectral S/N for all but delayed-τ model results.

This issue is discussed further in Section 5.6.2, as the consequences for the mass-weighted age posteriors are more apparent.

5.4.2 SFH prior dependence of mass-weighted ages

Mass-weighted age is a latent parameter of the SFH model; its prior depends on the parameter space of SFHs set by the form of the SFR(t), and the preferred SFHs set by priors for the SFH parameters. The procedure for determining the *implicit* interim prior distributions for Prospector models is described in several works in the literature (*e.g.*, for sSFR; Leja et al., 2020; Nagaraj et al., 2021; Li & Leja, 2022; Alsing et al., 2023), and was introduced in Chapter 1, Section 2.3.9. The implicit priors for the latent parameters can be described by marginalizing over the prior space, where Equation 2.13 describes the functional relationship between the mass-weighted age and the SFH parameters, and Equation 2.15 describes the surface integral over draws from the priors for the direct parameters. From the posteriors for the SFH parameters, the posteriors for mass-weighted age were calculated following Equation 2.13. In this section, the accuracy of the mass-weighted age posteriors are compared across a set of results based on different spectral S/N, and different SFH models. The relationship between the age posteriors and the marginalized age prior distributions is discussed in the Section 5.6.

Figure 5.5 compares the mass-weighted age posteriors to the true values for the mock quiescent galaxies. There are clear differences among the age posteriors between the SFH prior; as expected, the bias imposed by the different priors are stronger where the observations have lower S/N. The Dirichlet($\alpha_{\rm D} = 0.2$) and xSFR Continuity priors produce similar mass-weighted age posteriors. Where the posteriors are most prior-dependent, there is a preference for ages of $\sim$2.5–3 Gyr (*i.e.*, the ages of the younger galaxies are overestimated, and those of the older are underestimated). A mild bias persists until the spectroscopic S/N $\gtrsim$5 Å^{-1}. On the other hand, the posteriors under the cSFR Continuity typically overestimate the ages of all the galaxies. The bias is larger than the uncertainties in most cases. Lastly, the delayed-τ model recovers the ages of the oldest galaxies even at low S/N, although the ages of the youngest galaxies are consistently underestimated.

The biases among the mass-weighted ages can be understood through comparing the true and posterior SFHs. Figure 5.6 compares a set of median SFHs inferred from the different SFH priors, based on the same observations. The preferred SFH for each prior is shown in cyan, as determined from the median SFH of 10^5 draws, where thin lines indicate the strength of this preference. The prior for the Continuity models directly

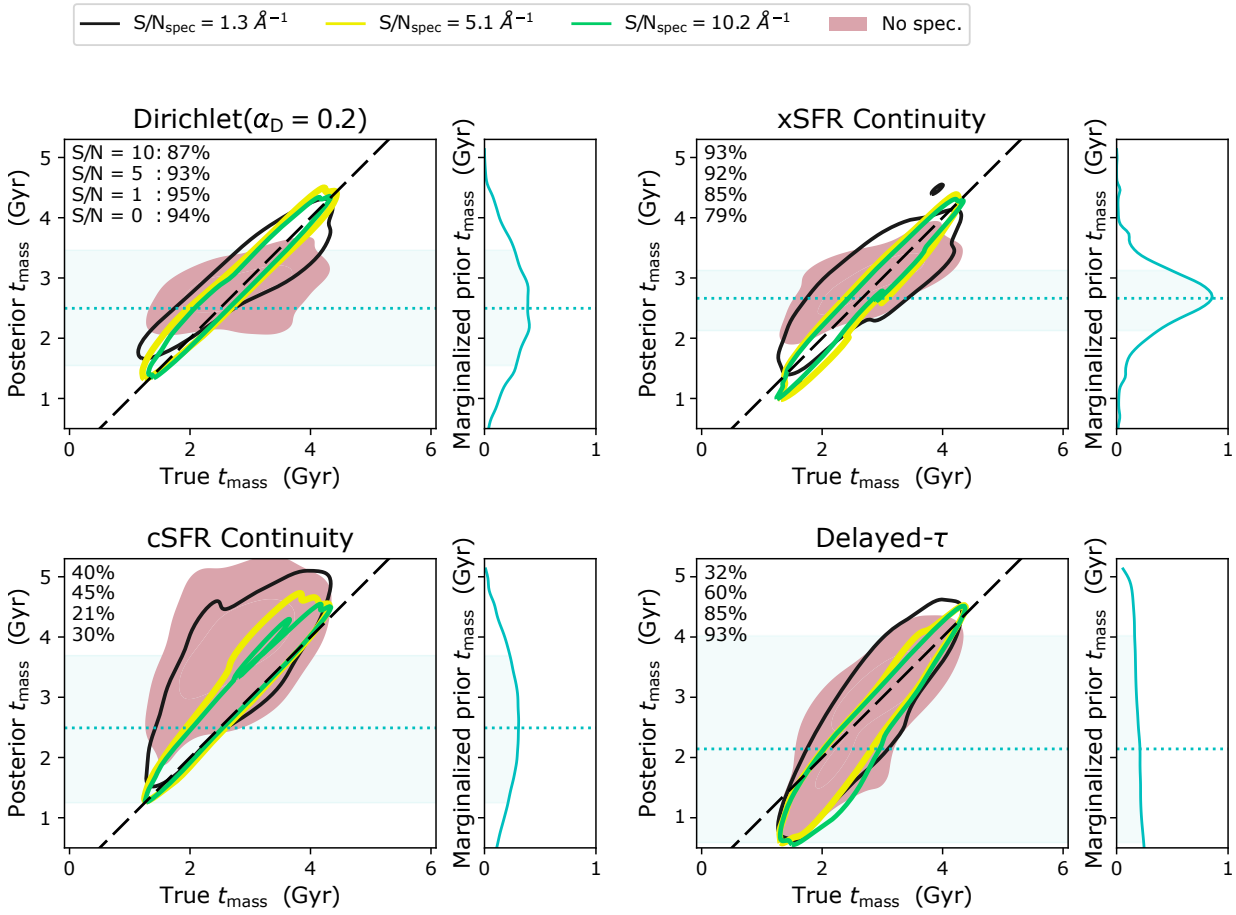

Figure 5.5: Mass-weighted age posteriors, based on fitting data with different spectral S/N (contour colours; see legend) and different SFH models, compared to the true values. The accuracy of the posteriors improves with increasing spectral S/N and wavelength coverage. Where the posteriors differ between the SFH models, the posteriors are prior-dependent. Contours are drawn around the posterior medians, and the percentage of galaxies consistent with the true values within the 68% CRs are listed in the top right corner. A dashed line shows the 1:1 relation. The marginalized prior age distributions are shown in the right-hand panels (cyan), where the median (dotted line) and 68% CRs (shaded region) are shown (see Section 5.6.2).

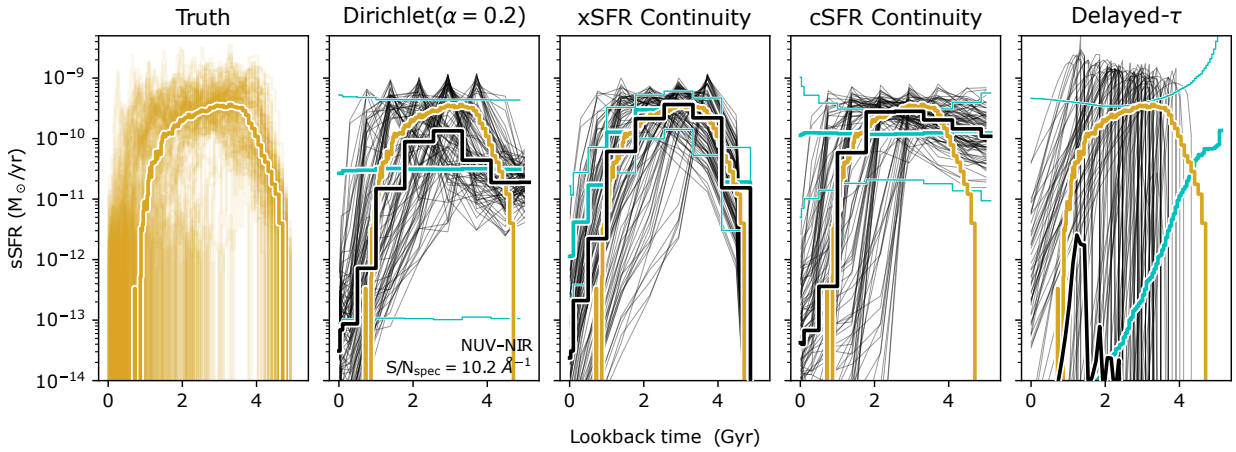

Figure 5.6: Comparison of true SFHs (yellow lines) of the mock galaxies, with the median SFHs for posteriors from the different SFH models (black lines). Median SFHs are shown with thick lines. Age posteriors were obtained by fitting mock galaxies with NUV–NIR photometry (S/N $= 10$) and rest-frame optical spectroscopy (S/N ~ 10.2 Å^{-1}). Median and 68% percentiles of sSFRs drawn from each SFH prior are shown in cyan.

informs the shape of the SFH.[7] Comparing the posterior, prior-preferred, and true SFHs, it is evident that while the declining SFRs at recent times is generally well recovered, early SFH remain prior dependent. For example, the cSFR Continuity model was tuned to preferred constant SFHs, which overestimates the SFR at early times, leading to the ages being overestimated. In contrast, the xSFR Continuity prior was tuned to prefer rising and falling SFHs (tracing the median SFH of the mock quiescent galaxies from Illustris) and does not act to consistently overestimate the ages, instead preferring a narrow distribution of ages around the peak of the preferred SFH. The Dirichlet($\alpha_{\rm D} = 0.2$) acts similarly, where a low recent SFR, and a preference for concentrated distribution of star formation and a low early SFH ($\sim$1 dex below that of the cSFR Continuity prior), centres the preferred distribution of ages around half the age of the Universe.

Figure 5.5 showed that the mass-weighted ages from the delayed-τ model were increasingly underestimated for younger galaxies. The prior of this model is that the SFHs of the mock galaxies follow a delayed-τ form, which is not the case for the youngest galaxies:

[7] This representation is not necessarily meaningful for the Dirichlet($\alpha_{\rm D} = 0.2$) prior, which prefers unequal distributions of the SFR between the time bins, without a preference for any particular bin; this averages out to a flat SFH. Similarly, the median SFH of the delayed-τ prior does not follow delayed-τ form.

typically they have early formation times followed by an extended period of star formation, and later, a sharp decline in SFR. Where the observations correctly constrain the recent SFR, the model assumes too late of a formation time, thus underestimating the mass-weighted age.

Fitting combined observations (*i.e.*, stacking) does not always provide a better understanding of galaxy diversity, despite overcoming this prior-dependence issue. While averaged observations will have higher S/N than the individual observations, such that the posteriors will be less prior-dependent, the posteriors are not always meaningful in the same way as the full set of individual posteriors. That is, the age of the averaged-galaxy is not the same as the average age of the galaxies, and the information of the variance among the galaxies properties can be lost. A demonstration of this point is discussed in Appendix C.1, where we compare the individual posteriors to that from averaged observations for galaxies of different ages.

5.4.3 Degeneracies between age, dust, and metallicity

The fact that the mass-weighted age posteriors were noisier where the observations had lower S/N is related, in part, to the degeneracy between the ages, stellar metallicity, and dust. Figure 5.7 shows the accuracy of the age posteriors compared to that for stellar metallicity and the diffuse dust normalization constant, for the same fits as shown in Figure 5.5. There is a clear correlation between the three parameters where the observations have spectral S/N $\lesssim 5$ Å^{-1}, while the posteriors remain degenerate at higher S/Ns for the cSFR Continuity and delayed-τ priors. This degeneracy explains how the SFH prior can affect the posteriors for metallicity and dust, as observed in Section 5.4.2.

The degeneracy between age and dust can be broken with multi-wavelength observations. The cSFR set of observations cover the NUV–NIR which provides constraints for the presence of UV luminous (young) stellar populations and the effect of dust attenuation, as well as the degree to which the heated dust re-emits that light in the IR. Observations at redder wavelengths increase the constraints on dust emission, which under the assumption of energy conservation propagates to constraints for the dust attenuation and young stellar component. The observations are refit with NUV–MIR photometry to test whether the inclusion of MIR data provides sufficient information to overcome the dependence on the SFH prior. The observations are also refit without UV data (V–NIR) to demonstrate the age constraints from such wavelengths.

Figure 5.8 compares the mass-weighted age posteriors based on photometric data, coloured according to the wavelength region included in the observations: V–NIR (grey),

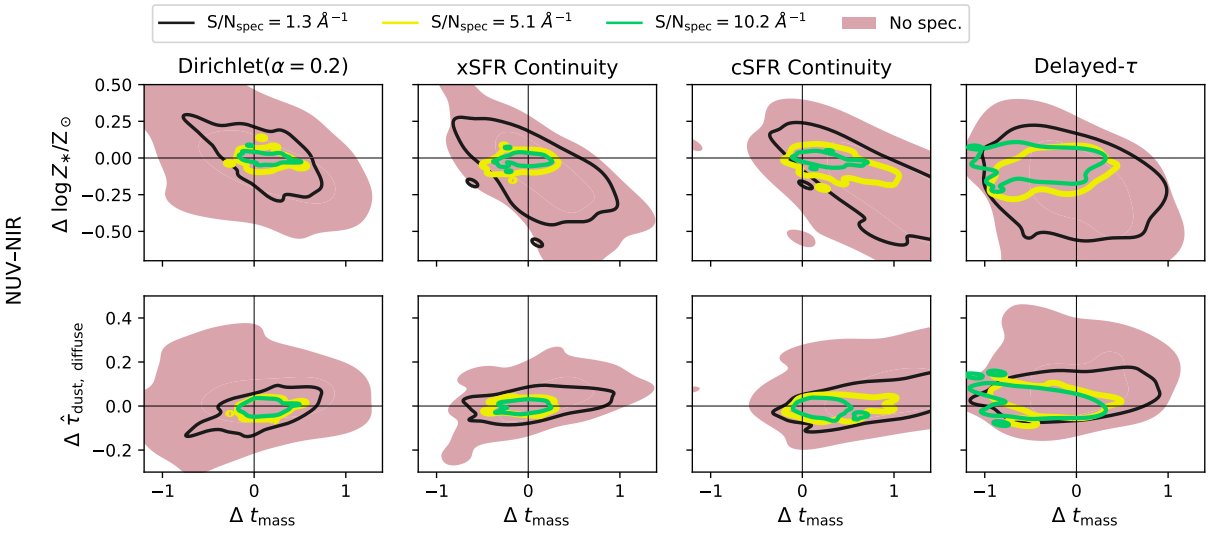

Figure 5.7: Comparison of the accuracy of stellar metallicity, and the diffuse dust normalization constant, relative to mass-weighted age. Contours are drawn around the posterior medians minus the true values, where the contours are coloured according to the spectral S/N of the observations (see legend).

NUV–NIR (red outlined), and NUV–MIR (solid yellow). The median posterior values are compared against the true ages, where black lines indicate the running median. Violin markers shown in the right-hand panel summarize these comparisons, with markers indicating the overall age difference, and that for the first and last bins of the running median. For the Dirichlet(α_D = 0.2) prior, which was shown to be fairly commensurate with the mock galaxies, the inclusion of UV–MIR data provides the best constraints for the ages of the oldest galaxies. This follows, since the lack of flux in the UV (given the lack of recent star formation) could be misinterpreted as stronger dust attenuation.

Figure 5.9 summarizes the difference between the posterior medians and true values for stellar mass, diffuse dust, stellar metallicity, and mass-weighted ages for all sets of observations discussed in this work. Dash markers indicate the overall median offset (Δ), and triangle markers indicate the Δ for the first and last bins of the running median (where the data is grouped by their true values into four bins of equal width; see Figure 5.8). The breadth of the violin markers indicate the level of noise *and* relative bias within the set of posteriors. Where the three markers converge, the posteriors show no *relative* bias – although there may still be a systematic bias. Posteriors based on different SFH models are shown across the different columns, and posteriors constrained by observations of different

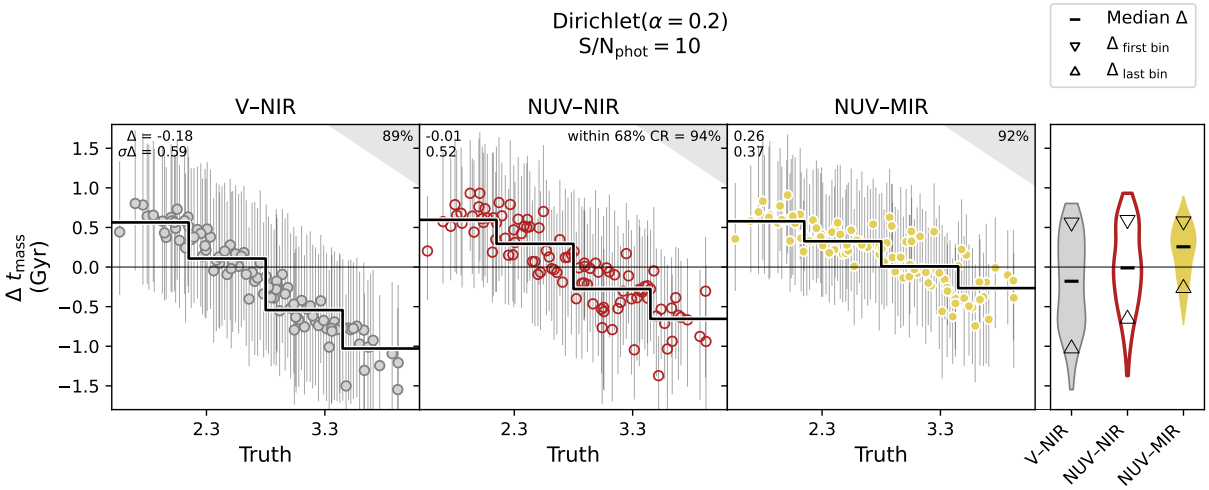

Figure 5.8: Accuracy of the mass-weighted age posteriors as a function of the wavelength coverage of the observations. Photometric observations, with S/N = 10, were fit with the Dirichlet($\alpha_{\rm D} = 0.2$) prior. Points indicate the posterior median values, and error bars the 68% CRs, and coloured according to the discrepancy with the true values. The median and standard deviation of the offsets are labelled in each panel, as well as the percent of posteriors which are consistent with the true values within their 68% CRs. Thick black lines indicate the running median. Grey regions indicate unphysical values. Violin markers summarize the distribution of $\Delta t_{\rm mass}$, where dashes indicate the overall offset, and triangles indicate the median offset in the first and last bin of the running median.

spectral S/N are grouped along the horizontal axes.

Comparing the results based on NUV–NIR or NUV–MIR data, there is a marked improvement in the accuracy of the dust posteriors where the observations have low spectral S/N. The inclusion of the MIR bands also improves the accuracy of the metallicity posteriors, although not to a degree where the metallicities are always unbiased. Any improvement in the accuracy of the ages is subtle. The MIR bands do not provide better constraints on the dust–age or dust–metallicity degeneracies than having low S/N spectroscopy. Note that the spectral continuum was excluded during the SED fitting procedure, such that the information from the spectra is only from the small set of rest-frame optical spectral features.

In comparison, having UV constraints appears to be crucial for recovering accurate dust parameters; otherwise, the posteriors prefer too much dust, and too young of ages to compensate. Even with higher S/N spectra, the dust posteriors based on V–NIR observations alone are offset from the true values. That said, the other parameters are less sensitive to the wavelength coverage when the observations include spectroscopy. Of the four galaxy parameters shown in Figure 5.4, stellar mass is the least sensitive to the wavelength coverage of the observations (with systematic offsets $\lesssim 0.1$ dex).

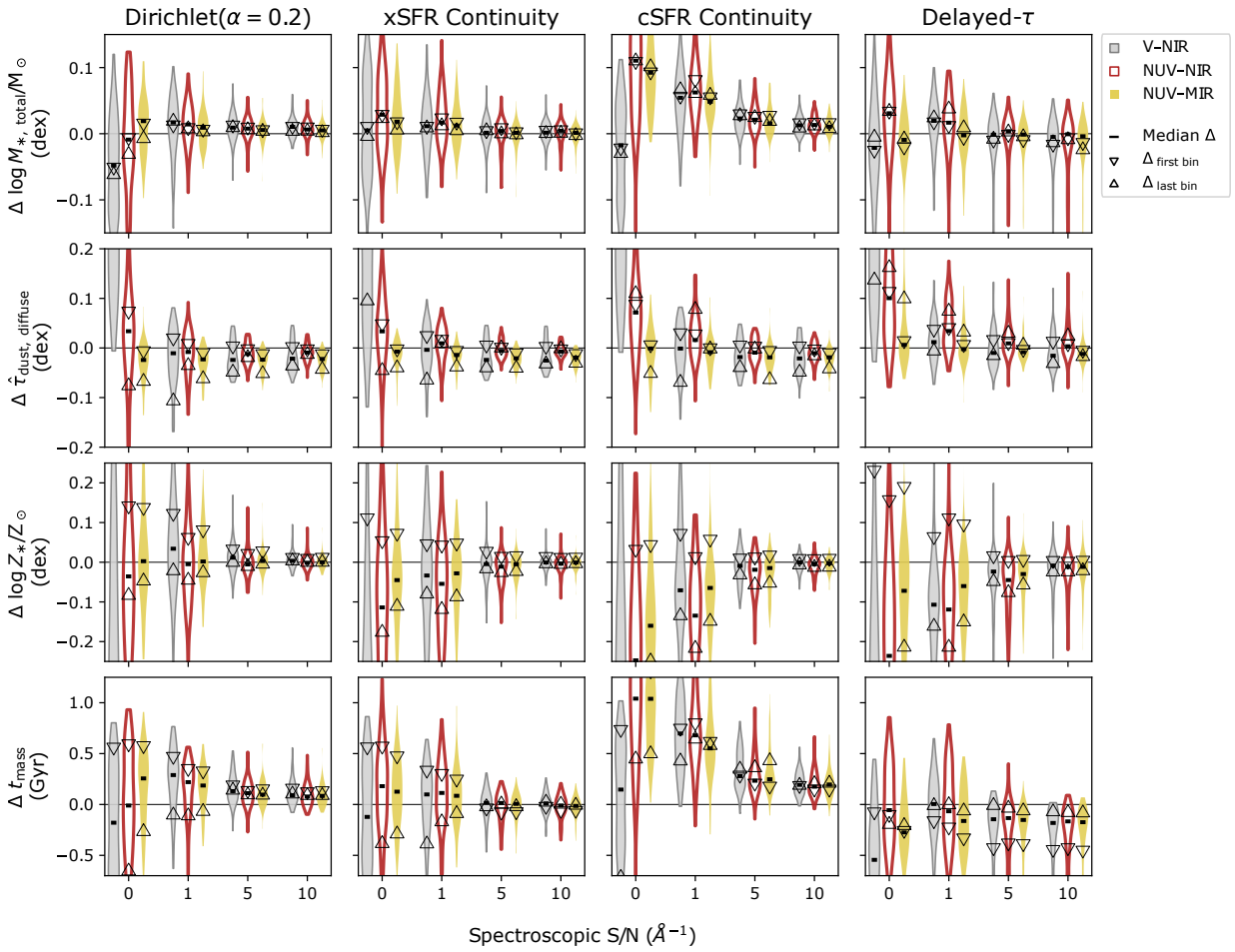

Figure 5.9: Accuracy of posterior median values for fits based on observations of different spectral S/N (horizontal axis) or wavelength coverage (V–NIR, NUV–NIR, or NUV–MIR; see legend), and with different SFH models. Comparisons are shown for total stellar mass formed, the diffuse dust normalization constants, stellar metallicity, and mass-weighted age. Violin markers show the distribution of posterior medians minus the true values, Δ, where the median offset is indicated with a dash marker. Triangle markers indicate the median offset of the first and last bins, where the data is grouped by their true values into four bins of equal width (see Figure 5.8). Where the markers converge, there is no relative bias across the population. The accuracy of the posteriors improves with increasing spectral S/N and wavelength coverage, however low S/N spectra are more informative (even when not fitting the spectral continuum) than marginally increasing the wavelength coverage.

5.5 Interpreting the diversity of galaxy parameters through population modelling

Population models provide a statistical basis for inferring the distribution of galaxy properties within a population. These models are conceptually similar to HBM, but instead condition the population distribution on the inferred *properties* of the galaxies (constrained by the individual observations), rather than from the observations directly. While in principle the SED-fitting framework could be made hierarchical, such that the individual galaxy properties are inferred simultaneously with the population distribution, this remains computationally challenging (although, see Leistedt et al. 2023). It is instead convenient to model the properties of individual galaxies by SED-fitting first (as described in Section 5.3, and model the distribution of those properties after-the-fact, as described in this section. A number of works in the astronomical literature take this approach, where Hogg et al. (2010b) outline the statistical basis for the pseudo-hierarchical model for the eccentricities of exoplanets. Section 5.5.1 expands on their work, making the model explicit to a subset of parameters from the SED model.

Population models allow for the self-consistent propagation of measurement uncertainties and correlations. The full shape and extent of the distribution of data, and co-varying uncertainties, can be taken into account. This provides an advantage over traditional minimization techniques that assume Gaussian-like uncertainties. Moreover, by leveraging the constraints from the full sample, this effectively deconvolves the noisy and heteroscedastic data (*e.g.*, Hogg et al., 2010b). Sandles et al. (2022) demonstrate the advantage of this technique when inferring the star forming main sequence for datasets with correlated measurements of stellar mass and redshift. This approach is naturally suited to studying the ages of galaxies, given the complex degeneracies between ages and other galaxy properties (*e.g.*, see Section 5.4.3).

The crux of population models is that they cannot intrinsically handle biased data. In modelling the properties of individual galaxies from SED-fitting, it is therefore important that the SED model is constructed such that the posteriors are unbiased. The priors for the SED model should therefore be uninformative. In practice, however, it can be challenging to implement uninformative priors. For example, mass-weighted age is a latent parameter of the SFH model, such that it's prior is subject to the flexibility of the model and the assumed priors for the model. In such a case, the priors associated with the SED-model needs to be 'taken out' of the posteriors. Put another way, the population model infers the distribution of properties for a population of galaxies from the (noisy, but unbiased) *likelihood* distributions for individual galaxies. Where the priors are uninformative, the

posteriors are equivalent to the likelihoods (see Chapter 1 Section 2.3.2). Otherwise, the likelihoods can be recovered by 'importance sampling' the posteriors, weighting the samples by their inverse prior probability.

This section outlines the statistical framework for a general population model conditioned on the posteriors for individual galaxies, as obtained with SED-fitting methods (as described in Section 5.3; the results of which are shown in Section 5.4). The validity of this model is then tested for the mock galaxy dataset in Section 5.6.

5.5.1 A population model for galaxy parameters

Consider a set of galaxies $\boldsymbol{g}$ such that a galaxy g is one of N_g galaxies[8]. These galaxies form and evolve according to a cosmological model which defines the distribution of galaxy properties, $p(\boldsymbol{\Theta}|\ \boldsymbol{g})$, where $\boldsymbol{\Theta}$ are the parameters of this model. Each galaxy has a set of properties, $\boldsymbol{\theta}_g = (z_g,\ Z_{*,g},\ t_{\mathrm{mass},g},\ M_{*,g}\dots)$, drawn from this model: $\boldsymbol{\theta}_g \sim p(\boldsymbol{\Theta}|\ \boldsymbol{g})$. In observing each galaxy we collect data, $\boldsymbol{\mathcal{O}}_g$, from which we can infer noisy estimates of the galaxy properties by comparing the observations against theoretical models (*e.g.*, SED-fitting). The population model defines a probability distribution for the parameters of the population distribution, Θ given a set of galaxy observations, $\{\boldsymbol{\mathcal{O}}_g\}_{N_g}$.

In practice, we are not always interested in constructing a population model for all the parameters. Instead, we focus on a subset of $\boldsymbol{\theta}$, which we label $\boldsymbol{\phi}$ (*e.g.*, age, or metallicity), and marginalize over all other parameters. To make this clear, we label the population model parameters describing the subset of physical parameters $\boldsymbol{\Phi}$. It follows that the probability of a given parameter can be described by $p(\boldsymbol{\phi}|\boldsymbol{\Phi},\boldsymbol{g})$. We construct a model which infers the probability of the parameters $\boldsymbol{\Phi}$ conditioned on the observations and a hypothesis for the form of the distribution (which defines the expected distribution of the model, $p(\boldsymbol{\Phi})$).

Ideally we would be able to write the probability of the population model in terms of the set of observations directly,

$$p\left(\boldsymbol{\Phi}\ \middle|\ \{\boldsymbol{\mathcal{O}}_g\}_{N_g}\right) = \frac{p\left(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}\ \middle|\ \boldsymbol{\Phi}\right)\ p\left(\boldsymbol{\Phi}\right)}{p\left(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}\right)} \tag{5.1}$$

In most cases, however, it is computationally intractable to simultaneously model observations for multiple galaxies with modern SED-fitting codes. Instead, we can use hierarchical

[8]In this notation, a bold-face indicates a vector of parameters.

inference, leveraging the results found in fitting each set of observations individually. We can re-write the probability equation in terms of the posteriors obtained from fitting each set of galaxy observations, $p(\boldsymbol{\theta}_g|\boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}})$ since $\boldsymbol{\phi_g}$ is a subset of $\boldsymbol{\theta}_g$. Being explicit about which parameters are relevant to the population model, $\boldsymbol{\phi}$, we separate $\boldsymbol{\theta} = (\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v})$, where

1. $\boldsymbol{y}$ include all parameters fit in the SED-model which *are not* included in the population model,
2. $\boldsymbol{u}$ include all parameters fit in the SED-model which *are directly* included in the population model, and
3. $\boldsymbol{v}$ include all parameters fit in the SED-model which *are indirectly* included in the population model, *i.e.*, $\boldsymbol{t} = f(\boldsymbol{v})$, where f represents some deterministic function.

The parameters relevant to the population model are therefore $\boldsymbol{\phi} = \{\boldsymbol{u}, \boldsymbol{t}\}$.[9].

Following this notation, we re-write Equation 2.2 which describes the posterior probability distributions obtained from SED-fitting galaxy observations,

$$p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}}) = p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}}) = \frac{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g, \boldsymbol{\mathcal{H}})\ p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{H}})}{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})} \tag{5.2}$$

We now want to leverage this information in building an inference model for the parameters of the population model, as laid out in Equation 5.1. For simplicity, we will continue to use the label $\boldsymbol{\theta}_g$ in the following equations until it is necessary to distinguish between $\boldsymbol{y}_g$, $\boldsymbol{u}_g$ and $\boldsymbol{v}_g$.

5.5.2 Hierarchical inference

We start by making the reasonable assumption that the observations of the individual galaxies are independent of each other such that their likelihoods are separable,

$$p\left(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}\Big|\ \boldsymbol{\Phi}\right) = \prod_g^{N_g} p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi}) \tag{5.3}$$

[9]We label the variables following Nagaraj et al. (2021), despite the potential for confusion between t and $t_{\rm mass}$. This notation is only relevant for explaining the components of the population model, and is not used elsewhere in this chapter.

We can therefore consider the marginal likelihood for each galaxy individually,

$$p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi}, \boldsymbol{\mathcal{H}}) = \int p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}})\ p(\boldsymbol{\theta}_g|\ \boldsymbol{\Phi})\ \ \mathrm{d}\boldsymbol{\theta}_g \tag{5.4}$$

The first term in the integral is the likelihood function of the observations, given a set of properties, as shown in Equation 5.2. Rearranging Equation 5.2,

$$p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\theta}_g, \boldsymbol{\mathcal{H}}) = \frac{p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}})\ p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})}{p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{H}})} \tag{5.5}$$

and substituting it into Equation 5.4, we get

$$p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi}, \boldsymbol{\mathcal{H}}) = \int \frac{p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}})\ p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})}{p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{H}})}\ p(\boldsymbol{\theta}_g|\ \boldsymbol{\Phi})\ \ \mathrm{d}\boldsymbol{\theta}_g \tag{5.6}$$

We can rearrange further by factoring out $p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})$ given that it is independent of the variables of integration, and make explicit the distinction of variables contributing to the population model,

$$\begin{aligned} \frac{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})} &= \int \frac{p(\boldsymbol{\theta}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{H}})}\ p(\boldsymbol{\theta}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}})\ \ \mathrm{d}\boldsymbol{\theta}_g \\ &= \int \frac{p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{H}})}\ p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{H}})\ \ \mathrm{d}\boldsymbol{y}_g\ \mathrm{d}\boldsymbol{u}_g\ \mathrm{d}\boldsymbol{v}_g \end{aligned} \tag{5.7}$$

We can factor the first term in the integral to separate the parameters which are ($\boldsymbol{u}_g$, $\boldsymbol{v}_g$) and are not ($\boldsymbol{y}_g$) included in the population model, so long that both the parameters and their priors are independent. Additionally, we assume that $\boldsymbol{y}_g$ is unrelated to the population model such that their priors are also irrelevant: $p(\boldsymbol{y}_g|\ \boldsymbol{\Phi}) = p(\boldsymbol{y}_g|\ \boldsymbol{\mathcal{H}})$,

$$\frac{p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{H}})} = \frac{p(\boldsymbol{y}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{y}_g|\ \boldsymbol{\mathcal{H}})}\ \frac{p(\boldsymbol{u}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{u}_g|\ \boldsymbol{\mathcal{H}})}\ \frac{p(\boldsymbol{v}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{v}_g|\ \boldsymbol{\mathcal{H}})} = \frac{p(\boldsymbol{u}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{u}_g|\ \boldsymbol{\mathcal{H}})}\ \frac{p(\boldsymbol{v}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{v}_g|\ \boldsymbol{\mathcal{H}})} \tag{5.8}$$

Furthermore, provided that the deterministic relationships between $\boldsymbol{v}_g$ and $\boldsymbol{t}_g$ are independent of both $\boldsymbol{\Phi}$ and $\boldsymbol{\mathcal{H}}$, we can write,

$$\frac{p(\boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{H}})} = \frac{p(\boldsymbol{u}_g, \boldsymbol{t}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{u}_g, \boldsymbol{t}_g|\ \boldsymbol{\mathcal{H}})} \tag{5.9}$$

We now simplify Equation 5.7,

$$\frac{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})} = \int \frac{p(\boldsymbol{u}_g, \boldsymbol{t}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{u}_g, \boldsymbol{t}_g|\ \boldsymbol{\mathcal{H}})}\ p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{O}_G}, \boldsymbol{\mathcal{H}})\ \ \mathrm{d}\boldsymbol{y}_g\ \mathrm{d}\boldsymbol{u}_g\ \mathrm{d}\boldsymbol{v}_g \tag{5.10}$$

From Equation 2.2 we know $p(\boldsymbol{y}_g, \boldsymbol{u}_g, \boldsymbol{v}_g|\ \boldsymbol{\mathcal{O}}_g, \boldsymbol{\mathcal{H}})$ are posteriors output from the SED-fitting procedure. Here $p(\boldsymbol{u}_g, \boldsymbol{t}_g|\ \boldsymbol{\mathcal{H}})$ refers to the joint prior probability distribution for the parameters $\boldsymbol{u}_g$ and $\boldsymbol{t}_g$ given the explicit prior probabilities assigned to the SFH model parameters, and the function relating $\boldsymbol{t}_g = f(\boldsymbol{v}_g)$.

The integral in Equation 5.10 can be approximated by a Monte Carlo integral approximation (*i.e.*, using the 'importance sampling approximation') such that

$$\frac{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})} \approx \frac{1}{M} \sum_{m=1}^{M} \frac{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)}|\ \boldsymbol{\Phi}\right)}{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)}|\ \boldsymbol{\mathcal{H}}\right)} \tag{5.11}$$

where M is the number of independent samples drawn from the distribution. In the limit of infinite samples, the approximation becomes exact. This is efficient to compute where it is easy to sample from $p(\boldsymbol{u}_g, \boldsymbol{t}_g|\ \boldsymbol{\Phi})$; in our case it is, because we can sample $\boldsymbol{u}_g$ and $\boldsymbol{t}_g$ by drawing from posteriors output from the SED-fitting procedure. The index m label a given draw from the distributions.

Combining all the individual galaxy information following Equation 5.3, we can now write the likelihood of the population model as

$$\frac{p(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}|\ \boldsymbol{\Phi})}{p(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}|\ \boldsymbol{\mathcal{H}})} = \prod_{g=1}^{N_g} \frac{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\Phi})}{p(\boldsymbol{\mathcal{O}}_g|\ \boldsymbol{\mathcal{H}})} \approx \prod_{g=1}^{N_g} \frac{1}{M} \sum_{m=1}^{M} \frac{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)}|\ \boldsymbol{\Phi}\right)}{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)}|\ \boldsymbol{\mathcal{H}}\right)} \tag{5.12}$$

For our purposes the term $p\left(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}\big|\ \boldsymbol{\mathcal{H}}\right)$ is a constant that we can ignore since it is independent of Φ. Taking the logarithm,

$$\ln p\left(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}\big|\ \boldsymbol{\Phi}\right) = \sum_{g=1}^{N_g}\ \ln\left(\sum_{m=1}^{M} \frac{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)}|\ \boldsymbol{\Phi}\right)}{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)}|\ \boldsymbol{\mathcal{H}}\right)}\right) \tag{5.13}$$

To re-iterate, the two parts to this likelihood equation are 1) a likelihood function for the population model where the probability of $\boldsymbol{\theta}_g$ (drawn from the posteriors from SED-fitting) are conditioned on a set of parameters describing the population model for galaxy

ages, $\boldsymbol{\Phi}$, and 2) the associated probabilities of $\boldsymbol{\theta}_g$ given the interim priors assumed in the SED-model.

Having now developed a likelihood function for a given set of parameters, we can write the probability model following Equation 5.1,

$$\ln p\left(\boldsymbol{\Phi} \mid \{\boldsymbol{\mathcal{O}}_g\}_{N_g}\right) = \sum_{g=1}^{N_g} \ln\left(\sum_{m=1}^{M} \frac{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)} \mid \boldsymbol{\Phi}\right)}{p\left(\boldsymbol{u}_g^{(m)}, \boldsymbol{t}_g^{(m)} \mid \boldsymbol{\mathcal{H}}\right)}\right) + \ln p\left(\boldsymbol{\Phi}\right) - \ln p\left(\{\boldsymbol{\mathcal{O}}_g\}_{N_g}\right) \tag{5.14}$$

If we are interested in building a population model for the mass-weighted ages of a sample of galaxies, we would write $\boldsymbol{\phi} = \{\boldsymbol{t}_{\rm mass}\}$ such that $\boldsymbol{t}_g = \boldsymbol{t}_{{\rm mass},g}$ and $\boldsymbol{u}_g$ is an empty set.

5.5.3 The form of the population model

For the population models described in this work, we assume that the population distributions can be described by a piece-wise constant function with K steps:

$$p(\boldsymbol{\phi}_g^{(m)}|\boldsymbol{\Phi}) = \sum_{k=1}^{K} s\left(\boldsymbol{\phi}_g^{(m)}; \Delta_k,\ \Delta_{k+1}\right) \tag{5.15}$$

where Δ describes the edges of the bins, and

$$s\left(\boldsymbol{\phi}_g^{(m)}; \Delta_k,\ \Delta_{k+1}\right) \equiv \begin{cases} 0 & \text{for } x < \Delta_k \\ (\Delta_{k+1} - \Delta_k)^{-1} & \text{for } \Delta_k \leq x \leq \Delta_{k+1} \\ 0 & \text{for } \Delta_{k+1} < x \end{cases}$$

This provides a convenient framework that can easily be extrapolated to multiple dimensions (*e.g.*, if fitting a population model of two or more parameters) without specifying an exact functional form. The width and number of the bins are easily adapted to suit the variance among different datasets, or computational limitations. Sampling from this model is efficient, where we follow the approach outlined by Speagle et. al (in prep.) for `frankenz`.[10] We assume an uninformative (conjugate) prior for this model: Dirichlet($\alpha_{\rm D} = 1$).

Hierarchical shrinkage

One of the benefits of constructing a population model is that we can leverage the information to better constrain the individual properties for each galaxy. Applying the population

[10] https://github.com/joshspeagle/frankenz

information as a *prior* in this sense is known as 'hierarchical shrinkage,'

$$p(\boldsymbol{\phi}_g|\boldsymbol{\mathcal{O}}_g, \boldsymbol{\Phi}) = p(\boldsymbol{\phi}_g|\boldsymbol{\mathcal{O}}_g)\ p(\boldsymbol{\Phi}) \tag{5.16}$$

Given that we have described the population model as a step function, we can apply the overall probability of a given parameter under this model to each posterior by drawing from a Multinomial distribution,

$$p(\boldsymbol{\phi}_g|\boldsymbol{\mathcal{O}}_g, \boldsymbol{\Phi}) = \text{Mult}\left(\frac{p(\boldsymbol{\phi}_g|\boldsymbol{\mathcal{O}}_g)\ \odot\ p(\boldsymbol{\Phi})}{p\boldsymbol{\phi}_g|\boldsymbol{\mathcal{O}}_g)\ \cdot\ p(\boldsymbol{\Phi})}\right) \tag{5.17}$$

where $\odot$ is element-wise multiplication, and $\cdot$ is the dot product.

5.6 SFH prior dependence limits the uniform interpretation of galaxy properties: implications for population models

In this section, the SED-fit results for a population of mock galaxies are tested against the requirements of the population model introduced in Section 5.5. The population model requires that (noisy, but unbiased) likelihood distributions can be recovered by importance sampling the posteriors by the inverse prior probabilities assumed in the SED model. As described in Section 5.2, the mock quiescent galaxies are based on a set of SFHs predicted from TNG300 at $z = 1.2$, with observations modelled after the GOGREEN sample from Chapter 3. The mock observations were fit with four different SFH priors, as described in Section 5.3.2. As a reminder, the main details of the SFH models are as follows:

- The Dirichlet($\alpha_{\rm D} = 0.2$) prior favours SFHs with star formation in only a few of the time bins, as set by $\alpha_{\rm D} = 0.2$.
- The Continuity prior favours SFHs which have relative SFRs between time bins as set by the parameter μ, where the cSFR prior sets $\mu = 0$ giving a preference for constant SFHs, and the xSFR prior sets μ based on the measured SFHs from the population of mock quiescent galaxies at $z = 1.2$ in Illustris.
- The delayed-τ prior uses two parameters, t_0 and τ, to specify the time that star formation begins, and the timescale of the exponential decline thereafter.

The influence of these different SFH priors on the posteriors for stellar mass, diffuse dust, stellar metallicity, and mass-weighted age, were discussed in Section 5.4. These results are now discussed in the context of the expected behaviour of the SFH priors. Based on the SED-fitting results discussed in Section 5.4, two important points are highlighted:

1. Decreasing the quality of the observations does not just increase the noise in the posteriors; the posteriors are increasingly prior-dependent.

2. If the prior is incommensurate with the true properties of the galaxy, the posteriors will be biased. The extent of the bias is related to how probable the true set of galaxy properties are under the assumed prior. Additional observational constraints are required to overcome an incommensurate prior.

For mass-weighted age, a commensurate prior is one that broadly traces the correct SFR over time, SFR(t). Despite the relatively narrow distribution of properties among the mock galaxy population, none of the SFH priors that were explored in this work were uniformly commensurate with this population. Nor are the SFH priors 'uninformative.'

In order to uniformly interpret the properties of a given stellar population, the posteriors need to be 'importance sampled,' weighting by the inverse prior probabilities, to correct for any bias imposed by an incommensurate prior. The importance sampling requires that the prior probabilities are well-characterized. The implications for modelling the distribution of stellar mass are discussed in Section 5.6.1. In Sections 5.6.2 and 5.6.3, the implicit priors for mass-weighted age are discussed in detail.

5.6.1 A population model for total stellar mass

In the SED model used in this work, the 'uninformative' priors were assumed for total stellar mass, stellar metallicity, and the diffuse dust normalization constant (uniform, or uniform in logarithmic space; see Table 5.3). As demonstrated in Section 5.4.1, however, the SFH prior can also influence stellar properties which are related to the SFH. In the literature of HBM studies, however, the contribution of the SFH prior is often overlooked when importance sampling (*e.g.*, Leja et al., 2020). Rather, the prior is taken to be only that which is explicitly defined in the model. While not strictly proper, this is only an issue if the posteriors are biased by the SFH prior.

To demonstrate the population model, modulo the importance sampling approximation,[11] the posteriors for total stellar mass determined from the Dirichlet($\alpha_{\rm D} = 0.2$), which

[11] And ignoring several other simplifications of the mock test results, *e.g.*, model misspecification.

are unbiased, are used. Figure 5.10 shows the success of the population model under such conditions. Four sets of results are shown (black PDFs), for posteriors with different amounts of noise related to the spectral S/N of the observations. Where the observations are based only on the photometry,[12] the posteriors have standard deviations of 0.06–0.16 dex, which decreases to < 0.1 dex (0.08, 0.06 dex) with $\mathrm{S/N_{spec}} \sim 1$ Å^{-1} (5, 10 Å^{-1}). The posteriors are sorted into ten bins of 0.12 dex width, and shown relative to the binned true mass distribution (yellow). Red violin markers indicate the probability assigned to each mass bin by the population model, where dashes indicate the median of the probability distribution. The population model recovers the true mass distribution better than simply combining the posteriors.

Figure 5.10 shows that as the noise in the posteriors decreases, the population model provides less of an advantage over simply combining the posteriors, which is in part because the bin widths are larger than the variance. As mentioned in Section 5.5.3, the population model distribution can be used to improve the individual galaxy posteriors through hierarchical shrinkage. The 'shrinkage' among the posteriors depends on the original variance relative to the detailed information in the population model (*i.e.*, bin widths). While it is straightforward to decrease the bin widths, the caveat is that increasing the number of bins also increases the shot noise in the model, although this can be compensated for by increasing the number of galaxies. Marginalizing over multiple draws from the population model in calculating the shrinkage (Equation 5.17) can lessen the impact of shot noise.

In principle, it is easy to adapt the population model to higher dimensions, such that multiple parameters can be modelled simultaneously, accounting for correlations between the parameters. An example of a population model for mass and metallicity is discussed in Appendix C.2.

If the posteriors for total stellar mass from the cSFR Continuity model were instead used here, the effect of the SFH prior would have been to bias the results as a function of the S/N of the observations. The population distribution inferred from lower quality data would be offset from that obtained with better data. The same problem would apply if the SFH prior is commensurate with one population of galaxies more than another. Both issues are ignored in the example presented above. This is not a failure of the population model, but in meeting the criteria of sampling unbiased likelihoods. In the following section, the action of the SFH prior on the mass-weighted age estimates is explored, for which this issue is more significant.

[12]Note that it is assumed that the redshifts are known.

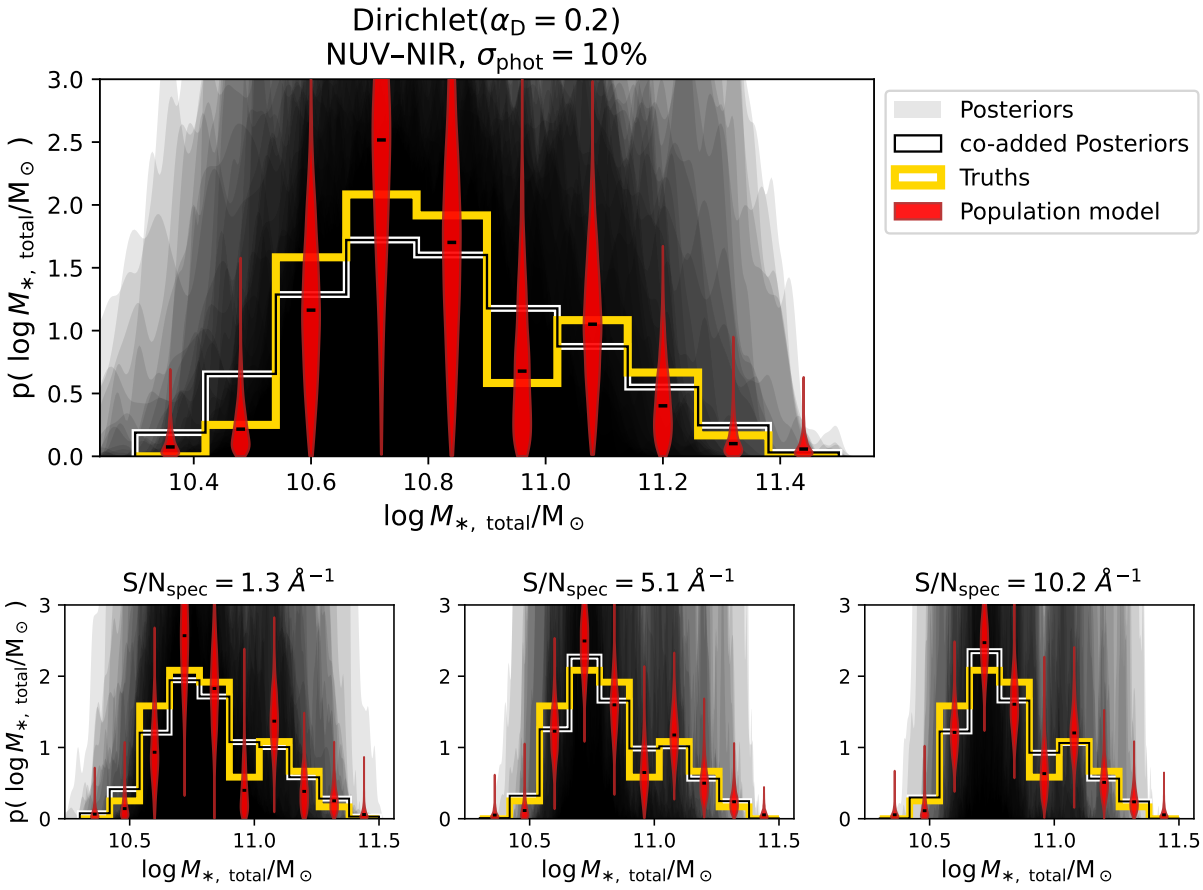

Figure 5.10: A population model for stellar mass. Since the stellar mass posteriors under the Dirichlet($\alpha_D = 0.2$) are noisy but unbiased, importance sampling is not included in this example. The individual posteriors are constrained with NUV–NIR photometry with S/N$_{phot}$ = 10, where the four sets of panels show the results based on observations of spectral S/N. The individual posteriors are shown as grey distributions, smoothed with a 0.1 dex kernel. For the population model, ten bins were chosen to be equally spaced between $10.3 < \log M_{*,\ total}/M_\odot < 11.5$, and show the co-added and binned posteriors with a black histogram. The binned true stellar masses are shown in yellow. The probabilities assigned to the bins are shown with red violin markers, with dashes indicating the median values. The population model successfully deconvolves the noisy posteriors to recover the true mass distribution. As the noise level decreases in the posteriors (*i.e.*, the observations increase in S/N) the advantage of the population model over combining the posteriors lessens.

5.6.2 SFH prior dependence of the age posteriors

The procedure for determining the *implicit* interim prior distributions for Prospector models is described in several works in the literature (*e.g.*, Leja et al., 2020; Nagaraj et al., 2021; Li & Leja, 2022; Alsing et al., 2023), and was introduced in Chapter 1, Section 2.3.9. Effectively, the implicit priors can be described by marginalizing over the prior space, where Equation 2.15 described the surface integral over draws from the direct parameter priors. Marginalized mass-weighted age distributions are often presented in the literature as a characterization of the SFH prior (*e.g.*, Chapter 4 Figure 4.2, Carnall et al., 2019b; Leja et al., 2019a; Lower et al., 2020; Tacchella et al., 2022a). The purpose of this section is to compare the bias imposed on the mass-weighted age posteriors by an incommensurate prior, to the expected behaviour of the prior given the marginalized prior distributions.

Figure 5.5 compares the mass-weighted age posteriors to the marginalized age prior distributions (cyan distributions). The expectation is that the mass-weighted age *likelihoods* are noisy but otherwise well-calibrated, where any discrepancy between the posterior and the truth is a product of a poorly calibrated prior. The influences of the SFHs priors are apparent in comparing the mass-weighted age posteriors between results based on observations of different S/N (where the contribution of the prior is larger when the data is less informative). Where the SFH priors are incommensurate with the true galaxy properties, the ages are biased from their true values. Comparing the direction of the bias to the peak of the marginalized prior distributions, however, shows that the action of the SFH prior is inconsistent with our expectations. For example, given the peak in the marginalized age prior about ∼2.5 Gyr for the cSFR Continuity prior, it was unexpected that the prior-dependent mass-weighted age posteriors do not show a preference for this particular age, but are instead uniformly overestimated.

As a further example of this point, Figure 5.11 compares the true and posterior mass-weighted ages, with the marginalized mass-weighted age prior distributions, for ten of the mock galaxies.[13] The posteriors shown correspond to observations with $S/N_{spec} \sim 5$ Å^{-1}, such that the age–metallicity degeneracy is broken, or subtle. The difference between the true and posteriors ages (yellow and black lines, respectively) is not consistent with the marginalized age prior distributions (cyan; as shown in Figure 5.5). In other words, the marginalized age prior distribution cannot inform us *how* the posteriors are (possibly) inaccurate.

[13]The age posteriors shown in Figure 5.11 are not typically Gaussian; which is true for the majority of the age posteriors discussed in this work. While typically only the posterior medians are discussed, this overlooks the fact that this is not necessarily meaningful when the posteriors are strongly skewed or multimodal.

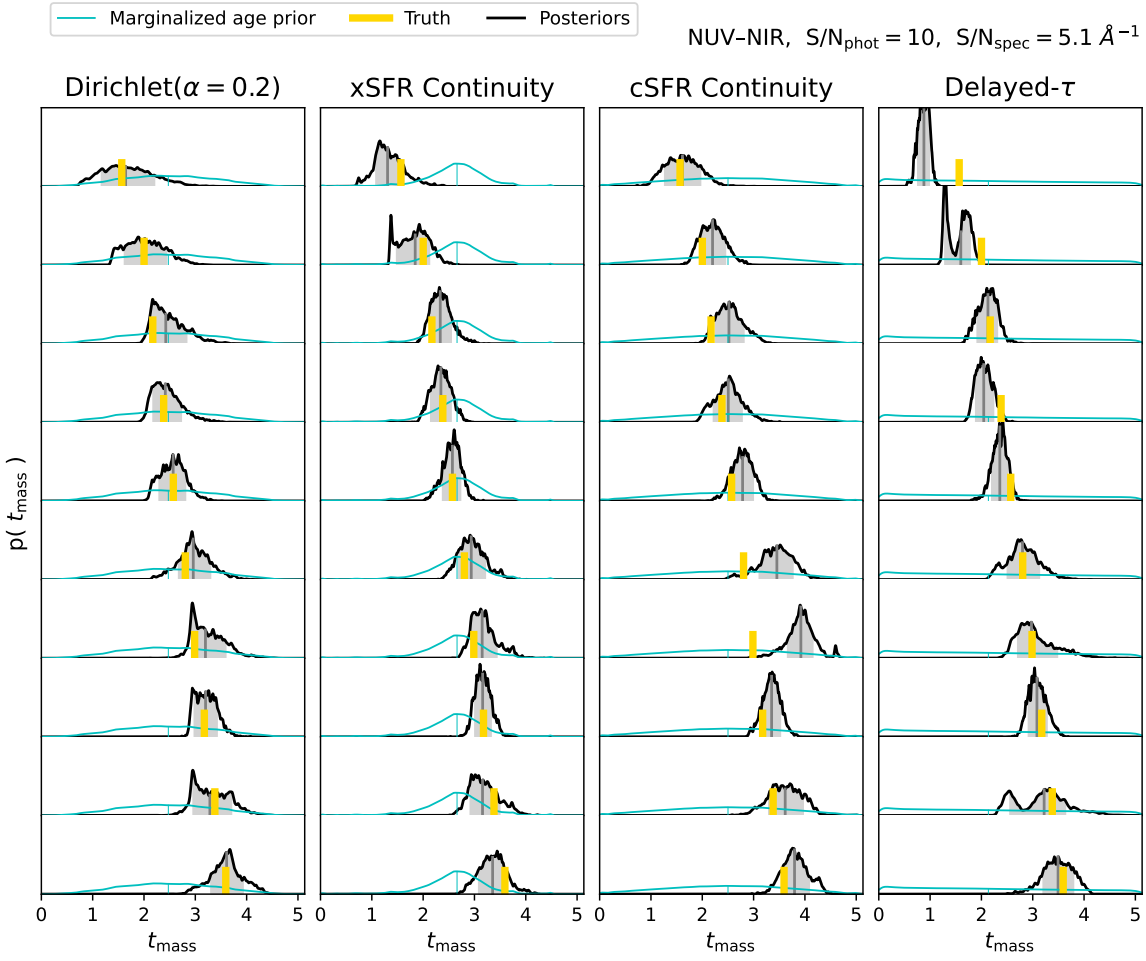

Figure 5.11: Comparison of true mass-weighted ages (yellow line) with age posteriors for a subset of the mock galaxies (black), and the marginalized age prior distributions (cyan; see Section 5.6.2). Age posteriors were obtained by fitting mock galaxies with NUV–NIR photometry (S/N $= 10$) and rest-frame optical spectroscopy (S/N ~ 5 Å^{-1}), where Figure 5.7 suggests that degeneracy between stellar metallicity and age is insignificant for all but the cSFR Continuity results. For a given galaxy (*i.e.*, each row) the posteriors based on the different SFH models differ, suggesting that the posteriors are prior-dependent. Where the age posteriors are inaccurate, the bias is not consistent with the influence of the marginalized age prior. This suggests that the marginalized age prior distribution *does not* characterize the age prior. Vertical lines indicate the median of each distribution, and the shaded region corresponds to the 68% CR of the posteriors. Posteriors are shown smoothed with a 0.1 Gyr kernel. The posteriors under these models exhibit subtle substructure about the centres and edges of the time bins (see Section 5.6.2).

One of the motivations for using the delayed-τ prior was that the marginalized age prior suggested that this prior would be less informative compared to the nonparametric models (where their prior distributions are more peaked). As was discussed in earlier sections, however, this prior is more constraining given the fact that it cannot reproduce complex SFHs. Moreover, the age posteriors under this model are often multimodal, indicative of complex degeneracies between model parameters.

The marginalized mass-weighted age prior describes the preferred distribution of ages where the data provides no constraints on the model. However, for any set of mock observations in this work, the SFH is not completely unconstrained by the data. Photometric observations alone can typically constrain the recent SFR (*e.g.*, Conroy, 2013). Recall that the mass-weighted age is a latent parameter in our model, and is related to the shape of the SFH. The SFR at different timescales are not equally constrained by the observations; the recent SFR is comparatively better constrained by the observations than the early SFH, given the significant luminosity of young stars and the nonlinear evolution of SEDs over time (*e.g.*, Iyer et al., 2019). Consequently, the observations are least constraining of the ages of the oldest stellar populations, such that the early SFH is more prior dependent. This was observed in the results for the mock galaxies discussed in Section 5.4.2. The mass-weighted age is therefore sensitive to how the SFH model propagates constraints between the late- and early SFH.

To demonstrate the connection between the mass-weighted age and SFR, in Figure 5.12 the marginalized age distributions are shown for each of the SFH models, relative to the marginalized ages at fixed sSFR averaged over the last 100 Myr (sSFR_{100}; cyan dashed lines) or 500 Myr (sSFR_{500}; magenta dotted lines), as derived from the SFHs. Timescales of 100 Myr are sensitive to the presence of O and B-type stars which strongly contribute to the UV-luminosity of the integrated SED, with lifetimes of ∼30–100 Myr. Where the recent SFRs are constrained to be low, the action of the Dirichlet($\alpha_{\mathrm{D}} = 0.2$) and xSFR Continuity priors is to strongly prefer ages of ∼3 Gyr, which explains the 'pile up' of the age posteriors at this time among the prior-dependent results. On the other hand, the action of the cSFR Continuity prior is to pair low SFRs with old ages, hence the systematic overestimation of the ages under this prior. Johnson et al. (2021) notes that the preference for 'smooth' SFHs acts to disperse the mass into more time bins, where there is little penalty in the likelihood model for doing so at large lookback times. As a consequence of the 'smoothness' prior, and the fact that it was assumed that star formation could begin as early as the Big Bang, both the ages and stellar masses were overestimated.[14] Ji & Giavalisco (2022) also recently

[14] Changing the location of the first time bin (*i.e.*, oldest in lookback time) will affect how the ages are biased (*e.g.*, Whitler et al., 2023b). This issue will be exacerbated at lower redshifts as older stellar

reported this effect in their 'ground truth' study of mock quiescent galaxies from TNG100 at $z = 2$. Although 'smoothness' is still favoured in the xSFR Continuity prior, the shape of the SFH is tuned to rise and decline, rather than be constant. So while the implicit age prior for the cSFR Continuity prior may naively appear more permissive (because the marginalized distribution has the largest dispersion of the nonparametric models), the age posteriors under this prior are generally more biased than under the other nonparametric model priors discussed in this work.

While the conditional dependence of the mass-weighted age distribution on the recent sSFR is informative, it is also an incomplete description of the SFH prior. Importance sampling the posteriors by the joint prior of mass-weighted age and sSFR_{100} (or any timescale) does not explain all the discrepancies among the age posteriors. Neural Network models were used to construct the multidimensional marginalized prior distributions to explore the behaviour of the mass-weighted age as related to the recent SFR. This is in part because the posteriors for sSFR_{100} are typically discrepant from the true values. Low SFRs ($< 10^{-1}$ $\mathrm{M}_{\odot}$ yr^{-1}) are notoriously challenging for SED-fitting codes to correctly measure (*e.g.*, Pacifici et al., 2022), which is particularly problematic when fitting quiescent galaxies. The sSFR posteriors typically span a large range of values, regardless of the S/N of the observations. Since low sSFRs have a very low prior probability in all the SFH priors described here, if such points were used in the importance sampling approximation, they would strongly outweigh the other solutions.

To fully characterize the impact of the mass-weighted age prior probably requires a high-dimensional model which captures the correlation between SFH parameters and age. Reducing the behaviour of the prior to one-dimension misses these complex relationships. In other words, *the marginalized age prior is an incomplete description of the influence of the SFH prior*. Developing a model which is capable of describing the nature of these SFH priors, however, is beyond the scope of this project. Without an understanding of how the SFH prior influences parameter estimates for galaxies of different intrinsic properties, mass-weighted ages inferred from observations without high-quality constraints of age-sensitive spectral features should be interpreted with caution. In Section 5.7 the conclusions of Chapter 3 are revisited with a newfound appreciation for the SFH prior dependence.

Lacking a proper characterization of the SFH prior precludes the general application of population models for mass-weighted ages. Without importance sampling, the technique is only robust in modelling the ages of galaxies where the posteriors are prior-independent,

populations are increasingly difficult to resolve. As noted in Section 5.2, the mock galaxies begin star formation 0.3–1 Gyr after the Big Bang, where the first time bin covers 5% of the age of the Universe ($\sim$0.26 Gyr at $z = 1.2$).

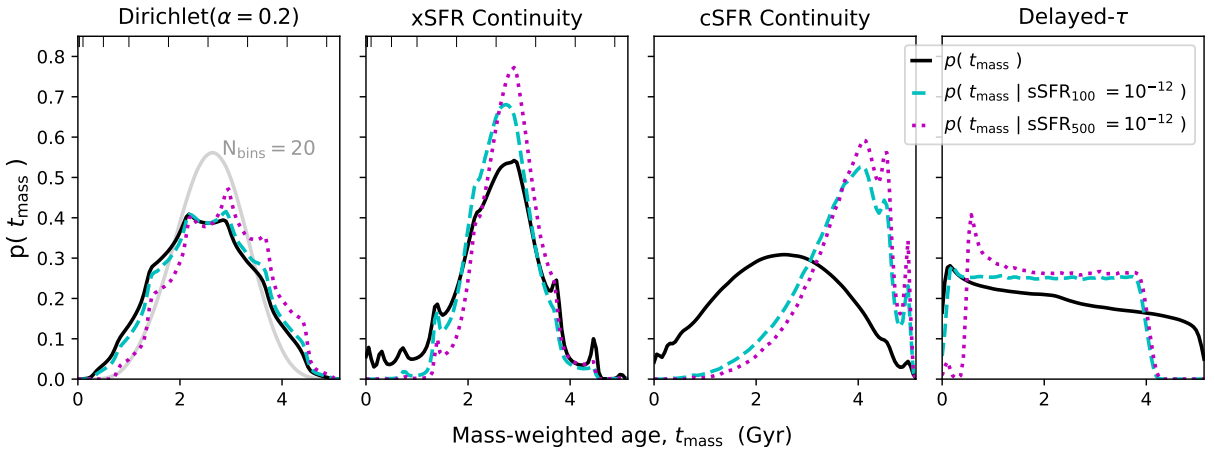

Figure 5.12: Marginalized prior distributions for mass-weighted age for the SFH models. Age distributions are also shown at fixed sSFR_{100} (*i.e.*, $\mathrm{sSFR}(t < 100$ Myr); cyan dashed line) and sSFR_{500} (dotted magenta line) to indicate that a constraint on the recent SFR changes the preferred mass-weighted age. For the cSFR Continuity model, the assumption of a smooth, constant SFH produces a prior preferring very old ages, given a low recent SFR. In contrast, the Dirichlet model age prior does not noticeably change given the constraint of a low recent SFR (which accounts for only one of $N = 10$ time bins). For the nonparametric models, the edges of the time bins are shown as ticks along the top of each panel. The 'quantization' of ages at the centres of the time bins is apparent, particularly for the Dirichlet($\alpha_{\rm D} = 0.2$) model; see Section 5.6.3). A grey line shows the marginalized age distribution for the Dirichlet($\alpha_{\rm D} = 0.2$) prior if the number of time bins was increased from $N = 10$ to $N = 20$, which lessens the 'quantization' of ages about the edges and centres of the time bins. For the delayed-τ model, the truncation in the distributions at old ages is related to the upper limit on the prior for τ.

which is to say having high-S/N observations and a commensurate SFH prior. In this work, it was demonstrated that NUV–NIR photometry with S/N = 10, and rest-frame optical spectroscopy with S/N > 10 is required to obtain reasonably prior-independent results for mock quiescent galaxies at $z = 1.2$, where the incommensurate cSFR Continuity prior was assumed. This result should not be interpreted as a baseline for real observations of galaxies, however, as a number of simplifications were assumed for this experiment.

One of the advantages to the population model outlined in Section 5.5 was that the inferred distribution of properties could be used as a new prior, taking advantage of hierarchical 'shrinkage.' The implicit assumption of the hierarchical model is that the objects are drawn from a given distribution (*i.e.*, the population distribution), and the 'observations' (*i.e.*, the posteriors from SED-fitting) are noisy representations of that population distribution (Loredo & Hendry, 2019). Using the population distribution as a new prior can be used to 'shrink' the noise of the individual posteriors. Alternatively, the population distribution can be used to better inform the choice of SFH prior, or be used as an explicit prior in the SED model itself.

In practice, a meaningful population distribution for the mock galaxy ages could not be constructed for the nonparametric models explored in this work. Even with high-quality data such that the mass-weighted age posteriors are accurate, there can be additional challenges for population modelling related to the SFH *model* itself; *i.e.*, the functional form of the SFH, this is discussed in the following section.

Empirically determined marginalized prior distributions, as calculated here, have been used in the literature to importance sample posteriors from SED-fitting to build population models describing the stellar mass function, (Leja et al., 2020), star forming main sequence (Leja et al., 2022), dust properties (Nagaraj et al., 2021), and others. The critical difference may be that the parameters of these models are not as sensitive to the choice of SFH model as the mass-weighted ages. In Section 5.6.1 it was noted that stellar masses can be influenced by the SFH prior (and has been noted previously in the literature; *e.g.*, Leja et al. 2019b, Lower et al. 2020), however, this effect may be insignificant compared to the realistic uncertainties of mass estimates. And while SFRs are latent parameters of the SFH model, observations can provide more direct constraints compared to mass-weighted age.

Whitler et al. (2023b) describe a population model based on ages for UV-luminous galaxies at $z \sim 6.8$, where their SFR model is parameterized by a constant SFR truncated at a given time. They had assumed that since mass-weighted age was a direct parameter of the SFH model, the weights used in importance sampling the posteriors simply followed the explicit prior. However, they note the limitation of their SFH model in describing the complex SFHs in their sample. That is, the prior of the SFH model itself is poorly

calibrated, which was not taken into account in their importance sampling.

5.6.3 SFH model dependence of the age posteriors

The choice of SFH model is a prior in itself. For example, that the delayed-τ model assumes a functional form for the SFR(t): following the onset of star formation, the SFR exponentially declines. Simply marginalizing over the prior space (as shown in Figure 5.12) misses this aspect of this prior assumption. The delayed-τ model is only appropriate for galaxies with SFHs that can reasonably be described with this form (*i.e.*, the oldest mock galaxies in our sample). That said, where the observations lack sufficient constraints for the shape of the SFH, the mass-weighted ages were least biased under this model (see Figure 5.6), although not to a degree where the true age distribution could be recovered with the population model.

While the nonparametric models can describe a broader spectrum of SFHs, there is the implicit assumption that the choice of time bins matches the resolution of distinct stellar populations constrained by the observations. Leja et al. (2019a) caution that using 'more bins than the data warrant' can lead to 'underfitting' whereby the uncertainties in the posteriors are overestimated, but note that underfitting can be mitigated by the choice of physically motivated SFR(t). Several studies have also demonstrated that using 'more time bins than the data warrants' does not necessarily improve the accuracy of SED-fitting results (*e.g.*, Lower et al., 2020; Tacchella et al., 2022a). The downside to using fewer bins is that the mass-weighted age tend to be 'quantized' at the centres and edges of the time bins, particularly for SFHs without regularization (*e.g.*, Leja et al., 2019a).

Quantized ages are apparent in the mass-weighted age posteriors shown in Figure 5.11 and the marginalized prior distributions in Figure 5.12. The effect is stronger where the SFHs have stronger transitions in SFR between time bins, such that the Continuity priors are less affected than the Dirichlet($\alpha_{\rm D} = 0.2$) prior. The grey line in Figure 5.12 for the Dirichlet($\alpha_{\rm D} = 0.2$) prior shows how increasing the number of time bins can resolve this issue. For the purposes of modelling the population of galaxy ages, this structure is problematic because the coincident peaks between posteriors will be compounded. Even where the posteriors have sufficient S/N to overcome SFH-prior dependence, the mass-weighted age posteriors exhibit this quantization such that the posterior distributions cannot be used in our population model. A more flexible approach to modelling the shape of the SFH may be required (*e.g.*, Iyer & Gawiser, 2017; Iyer et al., 2019).

5.7 SFH prior dependence limits the uniform interpretation of galaxy properties: implications for understanding galaxy evolution

To uniformly interpret the diverse properties of galaxies requires unbiased estimates of those properties for all galaxies. Noisy posteriors are not the issue; deconvolution techniques such as those used in the population model described in Section 5.5 can mitigate noise. The population model cannot correct for bias, however. Based on a 'ground truth' comparison of 'true' ages of mock quiescent galaxies to those recovered with SED-fitting, this work has demonstrated that the requirements for unbiased mass-weighted ages are high-S/N observations and an appropriate SFH model. While in principle importance sampling approximations can be used to mitigate prior-imposed bias, this requires that the prior is well described. However, in Section 5.6 it was demonstrated that marginalized prior distributions, as adopted in the literature, are not sufficient characterizations of the SFH prior. The complex behaviour of the SFH prior requires high-dimensional modelling of the correlations between the SFH parameters, which is beyond the scope of this work. Without a full understanding of the influence of the SFH prior on mass-weighted ages, population modelling cannot be generally applied to prior-dependent datasets. The implications of interpreting age trends from prior-dependent results are discussed in this section.

The prior-dependence of SED-fitting results is subject to the balance between model constraints from the observations or the prior. That is, the demands on the observations to constrain the true galaxy properties is higher where the SFH prior is incommensurate. Leja et al. (2019a) noted that the choice of SFH prior can be more important than the S/N of the observations, based on a 'ground truth' comparison of SED-fitting results from different nonparametric SFH priors with a $z = 0$ photometric dataset. Adding to this point, the nature of *how* the results are biased by an incommensurate prior will depend on how the prior is incommensurate. For example, star forming galaxies may be biased differently than quiescent galaxies, or post-starburst galaxies. This work focused on demonstrating that the bias can be different even among a fairly uniform population of high-redshift quiescent (mock) galaxies.

Ground-truth studies in the literature are complementary to this work in exploring the accuracy of SED-fitting results for a broader diversity of galaxy, datasets, and redshifts (*e.g.*, Wuyts et al., 2009; Hayward & Smith, 2015; Carnall et al., 2018, 2019b,a; Leja et al., 2019a; Lower et al., 2020; Suess et al., 2022). Many of these works attempt to identify SFH models which are 'best-calibrated' to a particular set of galaxies (or observational dataset);

however, this approach neglects the fact that some galaxies within the sample are likely disproportionately influenced by the favoured priors. In some cases, the observations are constructed to be sufficiently informative of the relevant parameters. Because the prior is increasingly influential as the observational constraints lessen, the results of such studies cannot be easily extrapolated. This work highlights that the most discrepant results are those for galaxies with SFHs most dissimilar to the assumed SFH priors – the inferred ages are not just noisy, they are *biased*. This is an important distinction given our interest in characterizing the diversity of galaxy ages within a population.

The requirement for high-quality observational constraints is more challenging for lower redshift observations, where it is more difficult to resolve the oldest stellar populations. The mock galaxies discussed in this work were observed at $z = 1.2$ such that the maximum age of the stellar populations was $\sim$5 Gyr. In Section 5.4 it was determined that rest-frame optical spectroscopy with S/N $\gtrsim$ 10 Å^{-1} was required to overcome the bias imposed by the incommensurate cSFR Continuity prior within the uncertainties of the posteriors. Mass-weighted ages inferred from photometric data alone (with S/N = 10), were prior-dependent such that age *trends* were inaccurate.[15] In Appendix C.2 it is demonstrated that a mildly informative metallicity prior, as adopted in Chapter 3 and in the literature (*e.g.*, for 3D-HST; Leja et al., 2019a), was not enough to overcome the prior-dependence of the low-S/N based results.

An additional concern is that the same SFH prior may not be commensurate with a given class of galaxy at all redshifts. Massive galaxies at low redshifts are likely passively evolving, while those at higher redshifts may still have declining star formation rates (*e.g.*, Madau & Dickinson, 2014). In a similar vein, the same SFH prior may not be commensurate with galaxies at all mass scales, given the apparent correlation between SFR and stellar mass at all redshifts (*e.g.*, Whitaker et al., 2012; Schreiber et al., 2015). The implication is that interpreting mass-weighted ages from datasets with mixed data quality, over a broad range of wavelengths, can be distorted by prior-dependence. Section 5.7.1 explores how SFH prior-dependence can conspire to look like physical trends reported in the literature. In Section 5.7.2, the results of Chapter 3 are revisited.

Studies which try to draw direct comparisons across a diverse set of galaxies are confronted with the challenge of constructing an SFH model that is flexible enough, and an SFH prior that is permissive enough, with respect to the true galaxy properties, to not bias the SED-fitting results. With this issue in mind, Pacifici et al. (2022) recommend adopting

[15] Increasing the photometric S/N = 30 for the mock observations provided comparable results as having S/N = 10 photometry and S/N $\gtrsim$ 1 Å^{-1} spectroscopy; that is, the correct age trends were recovered only where a commensurate prior was assumed.

a set of priors that permit all possible combinations of parameters. How to construct such a prior remains a challenge; none of the priors discussed in this work achieve this goal.

An alternative approach is to tune SFH priors to specific galaxies. For example, Suess et al. (2022) use the predicted SFHs from the UniverseMachine for galaxies of a given mass to tune a version of the Continuity model. This is similar to how the Continuity model was tuned in this work based on the true SFHs from the Illustris mock galaxies (the 'xSFR Continuity' prior), although the prior was not scaled with mass given the lack of age–mass trend among the mock galaxies. In Section 5.4 it was demonstrated that even this 'reasonable' SFH prior led to biased results, where the true SFHs were slightly different from the prior and the observations were not sufficiently constraining. Consequently, naive tuning schemes do not solve the problem of determining unbiased ages for samples of low S/N data. A second caveat of the approach of tuning SFH priors to simulations is that it assumes the validity of the simulations. SFH prior-dependent results are therefore only as meaningful as simulations can predict realistic SFHs. Moreover, if the simulations do not reproduce galaxy properties or underestimate the diversity of the properties, the tuning scheme may conceal an interesting population of unusual galaxies.

Without knowing the true galaxy properties *a priori*, it can be difficult to choose a commensurate SFH prior or be confident that the quality of the observations is sufficient to overcome the influence of an incommensurate prior. Pacifici et al. (2022) suggest using caution if the results resemble the shape and extent of the prior distribution. However, implicit in this procedure is knowing the behaviour of the prior, which this work has shown to be challenging for mass-weighted ages. While the comparison of results and marginalized prior distributions are used to gauge whether the results are purely prior-dependent (*e.g.*, Tacchella et al., 2022a), this practice can be misleading. Conversely, the fact that a posterior traces the prior distribution does not imply that the results are prior-dependent, it may simply be well-calibrated to the truth. Instead, it is suggested that the results could be considered robust if they are consistent between distinct SFH priors. The caveat to this approach is that it can be challenging to choose distinct priors.[16]

5.7.1 SFH prior contrived relations between star formation timescales

All timescales inferred from the SFH are sensitive to the SFH prior. The implication is that trends observed for star formation timescales from prior-dependent results will be

[16] For example, while the bias imposed on old quiescent galaxies is distinct between the Dirichlet($\alpha_{\rm D} = 0.2$) and cSFR Continuity priors, for young galaxies the bias is coincident. The fact that the two priors give the same age for a young galaxy does not imply the result is robust, or prior-independent.

unphysical if the SFH prior is unphysical. Besides the mass-weighted age, which provides a sense of when the main epoch of star formation occurred, there are a number of timescales discussed in the literature of galaxy SFHs.

Metrics for the overall duration of star formation are often taken from the percentiles of the mass assembly history, (*e.g.*, $\tau_{\rm SF} = t(p_1) - t(p_2)$), where the choice of p's are somewhat arbitrary. In Chapter 4 the SFH of the UDG DF44 was discussed in terms of $t_{50} - t_{90}$, while other works adopt different values (*e.g.*, $t_{20} - t_{80}$; Tacchella et al. 2022a). Ji & Giavalisco (2022) instead propose quantifying the duration of star formation by the doubled square root of the mass-weighted second central moment of stellar ages. For any definition, the fact that the early SFH is particularly SFH prior dependent can skew these statistics. As an example, Figure 5.13 compares the true and prior-dependent trends between the formation redshift, $z_{\rm form} = z(t_{\rm form}) = z\left[t_{\rm univ}(z_{\rm obs}) - t_{\rm mass}\right]$, $t_{50} - t_{90}$, and $t_{20} - t_{80}$. The formation redshift, being another representation of the mass-weighted age, is biased by the SFH prior as discussed in Section 5.4.2. The star formation timescales are also biased, where the nature of the bias is very different between the different SFH priors shown. In this figure, two versions of the Dirichlet prior are compared: with $\alpha_{\rm D} = 0.2$ and $\alpha_{\rm D} = 1$, given that the latter is commonly discussed among recent studies (*e.g.*, Tacchella et al., 2022a; Ji & Giavalisco, 2022). Neither the Dirichlet($\alpha_{\rm D} = 1$) nor cSFR Continuity priors are commensurate with the star formation timescales of the mock quiescent galaxies discussed in this work, such that any interpretation of the prior-dependent results would be misleading. The Dirichlet($\alpha_{\rm D} = 1$) mass-weighted ages are strongly skewed, predicting no trend between either Δt and $z_{\rm form}$, while the cSFR Continuity prior overestimates the ages of younger galaxies in particular, thus predicting too strong of a Δt and $z_{\rm form}$ relation.

Timescales associated with the last epoch of star formation are less sensitive to the shape of the SFH, and consequently less sensitive to the SFH prior. Figure 5.13 compares the true and prior-dependent trends between the formation redshift and the redshift at which the SFR crosses the green valley (GV; $1/[20\ t_{\rm univ}(z_{\rm obs})] < {\rm sSFR} < 1/[3\ t_{\rm univ}(z_{\rm obs})]$) $z(t_{\rm cross\ GV})$, and $z(t_{90})$. The correct trend between formation redshift and 'quenching' redshift is recovered reasonably well for all SFH priors with modest S/N spectroscopic observations. While the cSFR Continuity prior recovers a strong trend between these timescales where the observations have lower S/N, this is a contrivance of the prior, as shown in Figure 5.13 .

Tacchella et al. (2022a) compared SFH timescales for 161 massive galaxies at $0.4 < z < 1.2$, including star forming, post-starburst, and quiescent galaxies. Their dataset included 17–44 photometric bands covering 0.3–8 μm (UV–NIR rest-frame), and optical spec-

troscopy with S/N between 5–62 Å^{-1} (average 17.4 Å^{-1}) covering $\lambda_{\text{obs}} = 7000$–$9200$ Å^{-1}.[17] Their dataset, therefore, includes a variety of different age-sensitive features, with variations in observational quality, and for a diverse sample of galaxies. In fitting the SFHs of these galaxies with Prospector, assuming the cSFR Continuity prior, they recover a trend between z_{form} and $t_{20} - t_{80}$ which broadly traces the prior-dependent results shown here (see also their appendix D where they note the coincidence of the results and draws from the prior). In performing a 'ground truth' test with mock observations matching their dataset for galaxies from TNG100, they found that their timescales were overestimated, suggesting that despite the 'exquisite' spectroscopic and photometric data, the main results presented in their work are somewhat prior-dependent. Importantly, they also fit their data with the Dirichlet($\alpha_{\text{D}} = 1$) prior, which predicts no trend between z_{form} and $t_{20} - t_{80}$ (see Figure 5.13). They justify their recovery of a z_{form} and $t_{20} - t_{80}$ trend on the basis that a similar trend is recovered between the two SFH priors. Given the results of this chapter, the trend recovered with the Dirichlet($\alpha_{\text{D}} = 1$) prior for the quiescent galaxies in their sample is likely less prior-dependent. Moreover, one of their conclusions is that TNG100 predicts too narrow of a distribution of $z(t_{\text{cross GV}})$. The results of this chapter suggest that the large distribution of quenching times reported by Tacchella et al. (2022a) may be an artefact of the cSFR Continuity prior.

Li & Leja (2022) also recently compared the SFHs for a sample of 361 massive quiescent galaxies at $1.2 < z < 4$ (average $z \sim 2$) with ∼40 photometric bands covering the UV–NIR (S/N > 10), fit with Prospector. They are careful in their choice of SFH prior, ultimately selecting the Dirichlet($\alpha_{\text{D}} = 1$) prior based on the flexibility of this model reported in Leja et al. (2019a) and a 'ground truth' comparison with $z = 2$ mock galaxies from TNG100 (see their appendix A). While the Dirichlet($\alpha_{\text{D}} = 1$) prior may be a good choice of prior for $z > 2$ quiescent galaxies, results from this chapter demonstrate that the Dirichlet($\alpha_{\text{D}} = 1$) prior is not commensurate with mock $z = 1.2$ quiescent galaxies (the assumption that the SFR is distributed equally among the nonparametric time bins, as set by $\alpha_{\text{D}} = 1$, is likely a poor descriptor of low-redshift quiescent galaxies). The degree to which the Li & Leja (2022) sample is influenced by the choice of SFH prior is likely different over the redshift range because of the evolving SFHs (as well as the redshift dependence of the S/N requirements). That said, they demonstrate that their results are prior-dependent in the sense that there are systematic offsets between their fiducial results and those determined assuming the cSFR Continuity prior (see their appendix B), and are careful to interpret the conclusions on the basis of the uncertainties related to the SFH prior dependence.

[17] $\lambda_{\text{rest}} \sim 5000$–$6600$ Å^{-1} at $z = 0.4$, and 3200–4200 Å^{-1} at $z = 1.2$.

In exploring more diverse galaxy samples, there are additional challenges to consider. For example, Suess et al. (2022) demonstrate that 'flexible time bin' nonparametric models are better at capturing the recent SFR variability of post-starburst galaxies. That said, their 'ground truth' mock tests demonstrate that both fixed and flexible time bin models underestimate the mass-weighted ages of galaxies with recent bursts of star formation, an artefact of the 'outshining' effect (*e.g.*, Papovich et al., 2001), despite tuning the SFH prior based on the predicted SFHs from the UniverseMachine (akin to the xSFR Continuity tuned prior in this work). Their results highlight the challenge of uniformly interpreting the SFHs of quiescent and post-starburst galaxies under the same priors – the results are biased in different ways.

Together, these works highlight the complexity of determining robust SFH timescales from observations given the varying influence of the SFH prior. This severely limits the interpretation of trends between galaxies of different properties or redshifts. For any progress to be made in understanding the true diversity of SFHs, any inferences drawn from diverse datasets must consider the validity of the choice of SFH prior, and justify the conclusions on the basis of the assumed prior.

5.7.2 Revising the age comparison for GOGREEN quiescent galaxies in Chapter 3

The mock galaxies constructed in this work are modelled after the observational dataset for the GOGREEN spectroscopic sample of quiescent galaxies at $1 < z < 1.5$ studied in Chapter 3. The data set includes NUV–NIR photometry and rest-frame optical spectroscopy, as was originally fit assuming the cSFR Continuity prior. While in Chapter 3 it was expected that the absolute ages inferred from the observations would be prior-dependent, it was assumed that the relative age trend would be less prior-dependent, such that any difference in the results between field and cluster galaxies would reflect the true age difference of these populations. Given the results of Chapter 5, however, it is appreciated that this assumption is dependent on the field and cluster galaxies being equivalently commensurate with a given SFH prior. Moreover, the cSFR Continuity prior was shown to be incommensurate with the mock quiescent galaxies. In this section, the comparison of mass-weighted ages between UVJ-quiescent GOGREEN field and clusters galaxies is revisited with a new appreciation for the behaviour of the assumed SFH priors.

Given that the prior-dependence of the results is related to the S/N of the observations, it is important to establish that the data quality is similar between the field and cluster samples. The GOGREEN observations include photometry with varying S/N (minimum

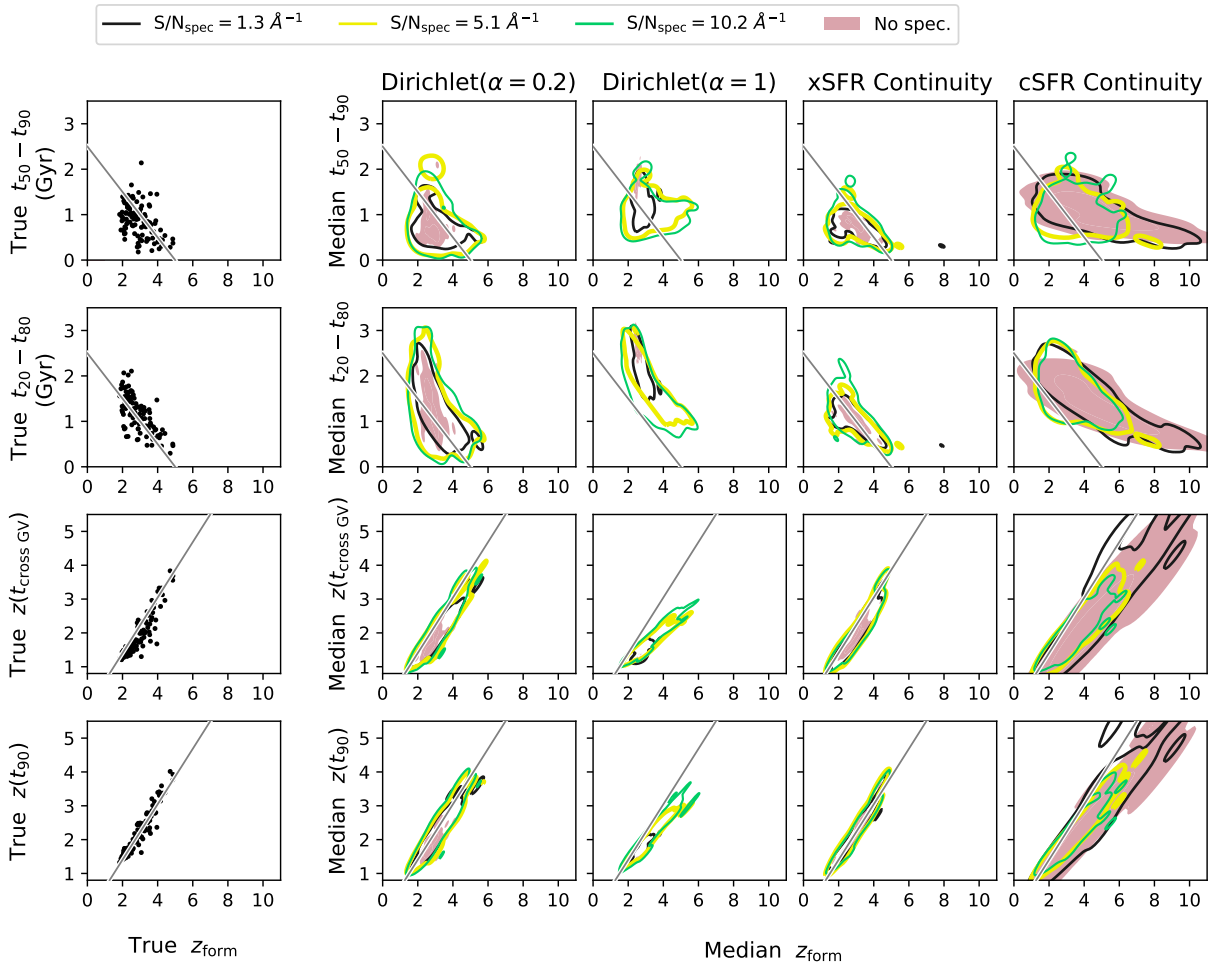

Figure 5.13: Prior-driven scaling relations between the formation redshift and (from the top): i) timescale to form between 50–90% of the stellar mass, or ii) between 20–80%, iii) the redshift at which the SFH crosses the GV, or iv) the redshift when 90% of the total stellar mass had formed. The true values for the mock quiescent galaxies are shown in left-hand column, and those recovered by fitting (S/N = 10) NUV–NIR photometry (filled contours) and with S/N $\sim$ 5 Å^{-1} spectroscopy (yellow contours) in the right-hand columns, for four different nonparametric SFH priors. Contours are smoothed with a 0.25σ kernel. Grey lines and contours are shown to guide visual comparison between plots, and are otherwise arbitrary. The grey lines are the same for all panels in a given row.

5% uncertainties assumed) and low-S/N rest-frame optical spectroscopy, such that the mass-weighted ages are likely prior-dependent. Figure 5.14 compares the S/N for the galaxies as a function of redshift and total stellar mass. The data quality is similar for the galaxies between the two environments, where the lowest S/N data is for the lowest stellar mass and higher redshift galaxies. The results for the mock quiescent galaxies presented in Section 5.4 suggest that age trends can be recovered assuming the Dirichlet($\alpha_{\rm D} = 0.2$) prior where the observations have $\rm S/N_{phot} \gtrsim 10$ and $\rm S/N_{spec} \gtrsim 1$. It should be noted, however, that the mock galaxy observations are unrealistic in that they consist of the true SED templates with error bars according to the set S/N. Figure 5.15 compares the scatter between the results presented in Section 5.4 to those determined from observations which include Gaussian noise.[18] The inclusion of Gaussian noise adds scatter to the posteriors, but not to a degree which demands a signifiant increase in S/N to overcome. Based on these results, any age trends among the GOGREEN galaxies are expected to be recovered where $\rm S/N_{spec} \gtrsim 1$ where the Dirichlet($\alpha_{\rm D} = 0.2$) prior is assumed, and assuming concordance between the mock and GOGREEN galaxies. In comparison, the formation times (mass-weighted ages) are expected to have a larger scatter among the earliest (oldest) reported values, such that the age trend may be overestimated.

The UVJ-quiescent GOGREEN galaxies are fit with the cSFR Continuity prior (as assumed in Chapter 3) and the Dirichlet($\alpha_{\rm D} = 0.2$) prior, where the latter is expected to be more commensurate with this dataset given the results from the mock quiescent galaxy tests. There are a number of subtle differences among these fits than those performed in Chapter 3. No joint mass-metallicity prior is assumed, where the results from Appendix C.2 suggest that this prior has negligible influence on the recovery of the mass-weighted ages. A more flexible dust model is assumed, instead of a fixed Milky-Way dust model, and a dust emission model is included. Lastly, the SED models were sampled with nested sampling (`dynesty`) instead of MCMC (`emcee`), where the stopping criteria was set to an effective sample size ESS = 10000, such that the extent of posteriors would be well sampled.

Figure 5.16 compares the median formation times ($t_{\rm form} = t_{\rm univ}(z) - t_{\rm mass}$) and total stellar masses for the results for both SFH priors. Results which are expected to be more strongly influenced by either prior (with $\rm S/N_{spec} < 1$ or $\rm S/N_{phot} < 10$) are shown as black points. The galaxies above this S/N threshold are typically those at $z < 1.2$ and with $\log\left(M_{*,\ \rm total}/\rm M_{\odot}\right) > 10.6$. A horizontal grey band indicates the middle mass bin specified in Chapter 3, which is offset by 0.2 dex to account for the typical difference between stellar mass and total stellar mass.[19] A grey line is drawn to help qualitatively compare the

[18] Where a data point x with uncertainty $\sigma_x = x/(\rm S/N)$ is drawn from $x \sim \rm Normal(x_{true}, \sigma_x)$.

[19] The variable fit in the SED models is the total stellar mass formed, not accounting for mass loss. In order to convert to the stellar mass remaining at the time of observation, the SED models need to be

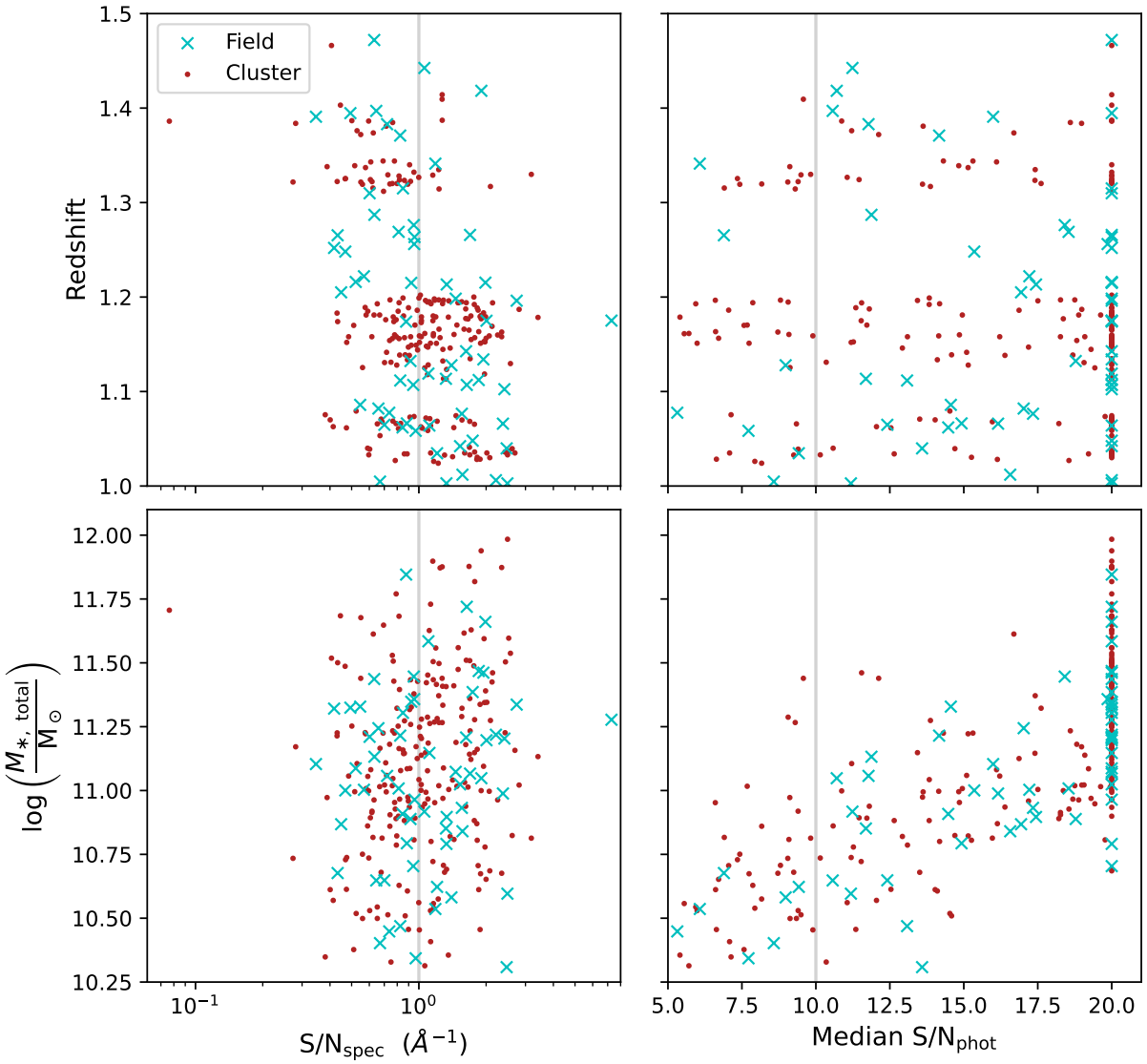

Figure 5.14: The distribution of spectral S/N (left) and photometric S/N (right; assuming minimum 5% uncertainties) for the GOGREEN quiescent spectroscopic sample, as a function of observed redshift (top) and total stellar mass (bottom). Markers and colours indicate galaxy environments: satellite galaxies in clusters are red (circles), and those in the field are blue (crosses). Overall, the data quality is similar for the entire sample of galaxies, suggesting that the SED-fitting results are prior-dependent to a similar degree, if the galaxies are otherwise equivalent. Vertical lines denote the S/N threshold used to select a subset of higher-S/N observations.

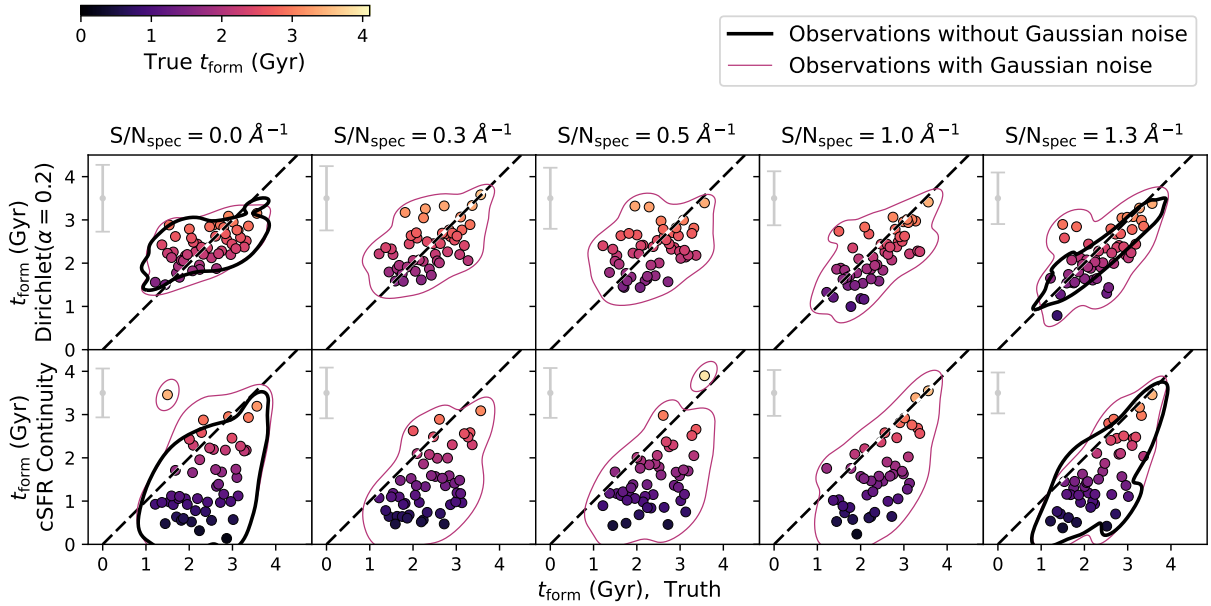

Figure 5.15: Comparison of true and posterior median values for mock quiescent galaxies fit with the Dirichlet($\alpha_{\rm D} = 0.2$) and cSFR Continuity priors, based on observations with (coloured points and red contours) and without (black contours; as shown in Figure 5.5) Gaussian noise. Grey points indicate the median uncertainty of the individual posteriors. Points are coloured by the true formation time of the mock galaxies, where half of the sample is shown.

results between the different panels, and is otherwise arbitrary. Consistent with the mock quiescent galaxies results shown in Section 5.4.2, the results for the GOGREEN galaxies from the cSFR Continuity prior suggest earlier formation times for the oldest galaxies than from the Dirichlet($\alpha_{\rm D} = 0.2$) prior. The ages of the youngest galaxies (at later formation times) are consistent between priors, consistent with the fact that these priors are not distinct for younger galaxies. Consequently, there is a larger scatter among the formation times inferred from the cSFR Continuity prior.

The main results of Chapter 3 were that the GOGREEN galaxies i) supported the picture of mass-dependent evolution, where more massive galaxies form earlier, and ii) that on average cluster galaxies form earlier than field galaxies. Without a full characterization of the influence of the SFH prior on the mass-weighted age posteriors, a quantitative comparison is not warranted. Qualitatively, the results from the Dirichlet($\alpha_{\rm D} = 0.2$) prior are consistent with mass-dependent evolution. Massive galaxies in both field and cluster environments have early formation times: $\log(M_{*,\ \rm total}/{\rm M}_{\odot}) > 11.5$ galaxies form < 2 Gyr after the Big Bang. While some of the lower stellar mass galaxies also have early formation times, on average the formation times are later for galaxies of decreasing stellar mass. Both field and cluster populations have a few young galaxies; in Chapter 3 these galaxies had 'rejuvenated' SFHs.

The evidence presented in Chapter 3 for cluster galaxies being slightly older than for field galaxies appears to be a conspiracy of the age bias imposed by the cSFR Continuity prior, particularly among the results from lower S/N observations. In making a careful S/N selection, and using a more commensurate SFH prior, the evidence for an age difference is less apparent. In as much as can be distinguished from the uncertainties of the age posteriors, it cannot be concluded whether the cluster galaxies are older than those in the field, at fixed stellar mass. That said, the results from the mock galaxies show that the Dirichlet($\alpha_{\rm D} = 0.2$) prior would act to underestimate any age differences. Higher-S/N observations are required to investigate the age difference in a meaningful way.

This revised analysis does not change the broad conclusions of Chapter 3 regarding feasible quenching models at $z > 1$. A pure infall-induced quenching model would predict that field galaxies are older than cluster galaxies, which we find no evidence for.

---

regenerated for each draw from the posterior to calculate the mass loss. The added computational expense of this process was not undertaken in this project.

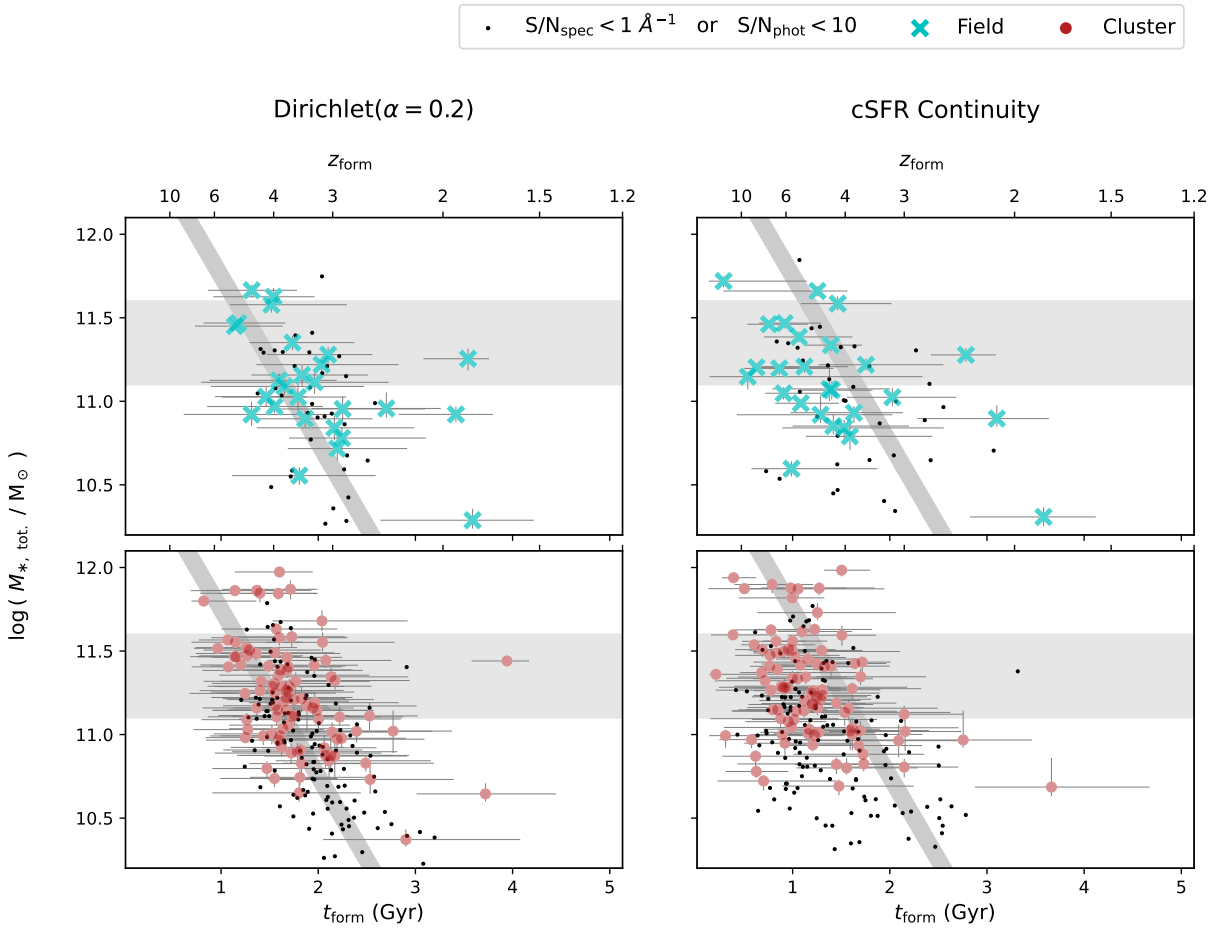

Figure 5.16: Comparison of median formation times (age of the Universe minus mass-weighted ages) as a function of total stellar mass, from the Dirichlet($\alpha_D = 0.2$) (left) and cSFR Continuity (right) priors, for the UVJ-quiescent GOGREEN spectroscopic sample. Marker shapes indicate the galaxy's environment, where black points are below the S/N threshold. A horizontal band denotes the middle mass bin, selected in Chapter 3, shifted by 0.2 dex to account for the typical offset between stellar mass and total stellar mass. A grey line is shown to guide the comparison between plots, and is otherwise arbitrary.

5.8 Summary and Conclusions

Recognizing the issue of SFH prior-dependence on the interpretation of star formation timescales, this work attempted to construct a statistical model which could leverage the information from individual galaxies and construct physically motivated SFH priors. The crux of this procedure is that the priors must be fully characterized, where any prior-dependence in the SED-fitting results could be mitigated with importance sampling. The framework of this population model follows several examples from the literature. In this work, it was demonstrated that the marginalized prior distributions often used to describe the latent behaviour of the SFH priors are incomplete characterizations. As such, the criteria for the population model are not satisfied, and it is not suitable for a general study of the mass-weighted ages of galaxies. Issues remain even where observations have high enough quality to produce prior-independent results; the assumed nonparametric form of the SFH can imprint subtle substructures in the distribution of mass-weighted ages. In constructing a population model from mass-weighted ages with 'quantized ages,' the substructure is compounded, and dominates the inferred population distribution. Without a method to correct for the substructure (which requires a characterization of the SFH priors), more flexible SFH models are needed.

The results of this work have broader implications than qualifying the use of population models. Through performing a ground truth comparison of the properties of mock quiescent galaxies at $z = 1.2$ with the recovered properties from SED fitting, assuming different SFH models, the prior-driven trends between star formation timescales are investigated. Of the various SFH models and priors explored in this work, none were uninformative with respect to mass-weighted ages.

Given that mass-weighted ages are sensitive to the shape of the SFHs, the unequal distribution of constraints on the late and early SFRs can lead to biased estimates. For the delayed-τ parametric models, the inflexibility of the functional form was such that the presence of a younger stellar population restricted the presence of an older stellar population, such that the ages of the young quiescent mock galaxies were underestimated. While the nonparametric models discussed in this work are more flexible in this respect, the fraction of the older stellar population is sensitive to the SFH prior. For example, in assuming a prior for constant and smooth SFHs (*i.e.*, the cSFR Continuity prior discussed in this work) the SFRs at early times are overestimated, and thus without observational constraints to correct this poor assumption the mass-weighted ages of quiescent galaxies are overestimated. Other SFH prior tuning schemes explored in this work affect the inferred mass-weighted ages in other unique ways; this SFH-prior dependence on mass-weighted ages (and other SFH timescales) acts to bias the values depending on how commensurate

the SFH prior is with the true galaxy properties.

Given the diverse nature of galaxy properties, any one SFH prior is likely not equally commensurate with a population of galaxies. In this work, this point was demonstrated for a relatively uniform population of mock quiescent galaxies with mass-weighted ages spanning $\sim$2.5 Gyr. In Section 5.7 the implications for studies which include a broader diversity of galaxies, observational datasets, or observed redshifts are addressed. Based on the results of this chapter, caution is suggested in interpreting trends of star formation timescales of such surveys. Some of the reported trends may be contrivances of the assumed SFH priors.

The results of Chapter 3 are also revisited with a newfound critical viewpoint for the validity of field and cluster galaxy population comparison. A revised analysis with a second SFH prior (thought to be more commensurate with quiescent galaxies based on the results of this chapter) washes out the original evidence for cluster galaxies being definitively older than field galaxies, at fixed mass. That said, the observed mass-dependent evolution and conclusions regarding feasible quenching models at $z > 1$ were unchanged.

Chapter 6

Summary, and looking forward

The physical properties of galaxies can be understood by interpreting their observed spectral energy distributions (SEDs) with theoretical models. Through investigating the direct observables (luminosities and colours) or derived physical parameters (stellar mass, metallicity, ages) the trends between these properties reveal a diversity in galaxy formation and evolution. Chapter 1 provided a brief review of the history of studying the 'fossil' record of quiescent galaxies from scaling relations. Advances in modern galaxy surveys, which provided multi-wavelength and spectroscopic information for large samples of galaxies, motivated the development of flexible SED models to provide a mapping between observables and physical parameters. The method of comparing model SEDs to observations was introduced in Section 2, where Section 2.3 then discussed a number of components which can be included in the models which make them highly flexible in describing a diverse set of galaxy properties.

Chapters 3 and 4 of this thesis study galaxies through reconstructing their star formation histories (SFHs). Chapter 3 focuses on a large sample of quiescent galaxies at $1 < z < 1.5$ found in low-density 'field' or high-density cluster environments. Differences in the SFHs for galaxies in either environment provides a constraint of the environmental-related processes which affect galaxy formation and evolution at high redshifts. The wide range of stellar masses in the sample also allowed for an exploration of the mass-dependent evolution of galaxies. The results of this work are summarized in Section 6.1. Chapter 5 instead focuses on a detailed study of one particular galaxy, Dragonfly 44 (DF44), whose properties are at odds with predictions of galaxy evolution. The results are summarized in Section 6.2. The precision afforded by the high-quality observations of DF44 highlighted limitations of the SED modelling framework, related both to the stellar population synthesis (SPS) and SFH models.

Motivated by the need for well-calibrated SFH models and priors to properly characterize diverse samples of galaxies from modern surveys, Chapter 6 then explored how SED-fitting results could be leveraged to construct physically motivated priors. The premise of this chapter was challenged by the complexity of the SFH model and priors, however. This work instead highlights the influence of SFH priors on the inferred star formation timescales, and investigates how priors can conspire to resemble physical trends.

Ideas for future projects that build on the content of this thesis are then outlined in Section 6.4.

6.1 Summary of Chapter 3

Chapter 3 studied a population of quiescent galaxies at $1 < z < 1.5$ in different local environments, and attempted to characterize the difference in ages and SFHs as related to environment. This work was based on observations from the GOGREEN survey, including UV–NIR photometry and low S/N rest-frame optical spectroscopy. Following several examples in the literature, the observations were fit with an SED model with a nonparametric SFH model which assumed a prior for a smooth SFR(t) (*i.e.*, a Continuity prior). It was assumed at the time that the relative SFHs among the GOGREEN data would not be substantially biased by the choice of SFH priors. By comparing the combined posteriors of the mass-weighted age distributions for galaxies of equivalent stellar mass, it was found that massive clusters lacked the population of young rejuvenated galaxies that was apparent in the field sample, while field galaxies had a lower fraction of old galaxies. A simple quantitative comparison of the combined posteriors suggested the median ages of the two populations were offset by < 0.5 Gyr, where the largest difference was among $10 < \log(M_{*,\,\mathrm{total}}/\mathrm{M}_{\odot}) < 11.3$ galaxies at $z > 1.3$. This result was tested under different selections of the galaxy sample: excluding 'rejuvenated' galaxies, those with modest [O II] flux, and those with UVJ colours closer to the green valley populations. In all cases, there was a positive or negligible age difference between galaxies in clusters or the field. The only exception was the subset of low mass galaxies without [O II] emission, although the uncertainties were large (as they are in all the comparison statistics) and only marginally less than zero. $\mathrm{D}_n(4000)$ values measured from the stacked spectroscopy were also consistent with the fact that galaxies in clusters are older than those in the field, and larger than observed among samples of galaxies from GCLASS at lower redshifts (see Appendix A.4).

The environmental age difference was put into context of two distinct galaxy formation models in Chapter 3. Unlike at $z = 0$, where the overabundance of lower mass quiescent

galaxies in clusters can be explained by a continued supply of newly quenched galaxies with mass-independent environmental quenching process, van der Burg et al. (2020) demonstrated that at $1 < z < 1.4$ mass-independent quenching could not explain the homogeneous shapes of the field and cluster quiescent mass functions. Such a scenario was also discrepant with the measured age difference between galaxies in clusters and the field. Using an efficiency of self-quenching which traces that cosmic star formation efficiency, and expected cluster infall rate from the literature, we constructed a toy model that predicted the ages of galaxies in either environment. Given the fact that at high redshifts the cluster infall rate outpaced the efficiency of self-quenching, in this scenario the population of quiescent cluster galaxies was constantly being supplemented by newly-quenched galaxies. The predicted average age of the quiescent galaxies in clusters was therefore lower than in the field, in contrast to our results. This is distinctly different from local clusters, for which the majority of the quiescent population is consistent with having been environmentally quenched upon infall. Our results suggest that the quiescent cluster population at $z > 1$ is driven by different physical processes than those at play at $z = 0$.

Alternative to post-infall based quenching models, van der Burg et al. (2020) demonstrated that the mass-independent overabundance of quiescent cluster galaxies could be explained by primordial enhanced quenching in proto-cluster environments. That is, galaxies which eventually end up in clusters had a head start in formation. To explain the quenched fractions observed among the GOGREEN field and cluster sample requires $\gtrsim 1$ Gyr of a head start (van der Burg et al., 2020). While we cannot rule this out based on the measured ages of the same galaxies, the results favour only a small (or negligible) age difference. Whether galaxies are 'pre-processed' in group-scale environments before assembling in clusters remains an open question.

6.1.1 Revisiting the age comparison

In Chapter 5 Section 5.7.2, the age comparison between field and cluster GOGREEN galaxies was revisited with a more careful consideration of the SFH prior-dependence of the result. The GOGREEN data varied somewhat in quality, and was largely of lower quality than identified in Chapter 5 as necessary to obtain accurate ages independent of the choice of SFH prior. The fact that the ages reported in Chapter 3 were older ages than reported in the literature, based on higher S/N observations, was consistent with the prior-dependent bias for the cSFR Continuity prior assumed in that work. In refitting the data with a Dirichlet($\alpha_{\rm D} = 0.2$) SFH prior, which was shown to be more commensurate with high-redshift quiescent galaxies in Chapter 5, the ages were less discrepant with the literature. While the age distributions from the Dirichlet($\alpha_{\rm D} = 0.2$) prior were somewhat different

from the original values, the conclusions of Chapter 3 were not significantly different. The results supported the picture of mass-dependent evolution, where on average, more massive galaxies form earlier than less massive galaxies; this result was not a contrivance of the SFH prior. Without a thorough characterization of the SFH prior, a quantitative comparison between cluster and field galaxies was not warranted. Considering the uncertainties of the individual posteriors a qualitative comparison showed no evidence for the cluster galaxies being older, or younger, than the field sample. The influence of the prior would be to underestimate age trends, however, such that higher quality observations could confirm the nature of the difference between populations.

6.2 Summary of Chapter 4

Ultra Diffuse Galaxies (UDGs) have sizes like giant galaxies, but luminosities like dwarf galaxies. Theoretical models typically attempt to explain UDGs as enlarged dwarfs, invoking only conventional galaxy evolution and environmental quenching to match the sizes and surface brightnesses of UDGs. This implies a connection between UDGs and dwarfs properties. However, UDGs have been observed to have a diverse set of properties, some of which are discrepant with model predictions. Chapter 4 studied the SFH of the quiescent UDG DF44, which is an oddity among both UDGs and dwarf galaxies. DF44's collective properties are inconsistent with the canonical dwarf population, and as yet unexplained by theories of UDG formation.

Previous studies of DF44 have shown it to be one of the largest UDGs in the Coma cluster (van Dokkum et al., 2017), where its location in the cluster outskirts suggests that it is on its first infall into the cluster (< 2 Gyr ago; Alabi et al. 2018, van Dokkum et al. 2019) and hence has not yet been environmentally processed. Moreover, it appears to be a part of a cold group (van Dokkum et al., 2019), and has no tidal distortion features (Mowla et al., 2017). DF44's stellar populations are old and quiescent (Gu et al., 2018a; Villaume et al., 2022), and therefore unrelated to recent infall into Coma. This is in tension with nearly all theoretical models, since they require the infall into the dense, hot environment of a cluster to quench UDGs.

High-resolution and high-S/N KCWI observations also showed that the stellar population properties and gradients of DF44 are dissimilar to canonical dwarf galaxies. Villaume et al. (2022) found that its metallicity is $\sim$2.3 σ below the expected mass-metallicity relation, pointing to a very short star formation timescale. In the same work, DF44's stellar population gradients were shown to be inconsistent with both observations of dwarfs, and

predictions from simulations of canonical dwarfs. That is, its gradients suggest that it quenched early and catastrophically.

Complementing the KCWI spectroscopy with UV–NIR photometry, in Chapter 4 Prospector was used to constrain the SFH of DF44. In contrast to the higher-redshift and lower-S/N observations studied in Chapter 3, DF44's observations provided a wealth of spectroscopic information related to its metallicities, abundances, and SFH. In fitting the SFH, it was concluded that DF44 experienced a catastrophic quenching event early shortly after its formation. The precise time that the quenching occurred, and the duration of the quenching, were dependent on the assumed SFH prior.

The SED-fitting results were extensively tested. The impact of two distinct SFH priors[1] were discussed at length, where the prior preferring extended SFHs provided a median mass-weighted age of ∼10.2 Gyr, compared to a prior preferring concentrated SFHs favoured ∼13.1 Gyr. Even with such high-quality data, the early SFH was prior-dependent, and these results were interpreted as limits on the true mass-weighted age. In an effort to explore the dependence on the SFH prior, the various subsets of the data were fit, and with varying degradations of S/N (see Appendix B.2) which motivated exploring the influence of the SFH prior in Chapter 5.

For either set of results, DF44's SFH was distinct from that of canonical dwarfs. Comparing its SFH to the predictions for average galaxies in halos of different mass, DF44 formed the bulk of its stellar mass during an epoch where massive galaxies were forming; galaxies in dwarf-scale haloes had significantly different stellar mass histories at that time. Since the publication of Chapter 3, Joshi et al. (2021) used models from Illustris TNG50 (Nelson et al., 2019; Pillepich et al., 2019) to predict the stellar mass assembly of dwarf galaxies in different environments. DF44's SFH resembles that of the satellite dwarfs which have early infall times and large stellar-to-halo mass ratios, which is inconsistent with its recent accretion into Coma.

Putting its SFH in context with its other properties, DF44's progenitor was likely larger than a dwarf galaxy, or some process unrelated to its environment caused its size to expand. Complementary evidence for this interpretation was provided by Saifollahi et al. (2022) who measured an excess globular cluster (GC) population, relative to stellar mass, for DF44 (including other UDGs). The GC population is inconsistent with UDG formation scenarios that explain the sizes of UDGs by redistributing the stars to larger radii, and Villaume et al. (2022) similarly excluded such scenarios given DF44's 'inside-out' stellar population gradients. *How* DF44 quenched remains a critical question needed

[1]Dirichlet with $\alpha_D = 1$, preferring uniform distributions of stellar mass across the time bins of the nonparametric model, or $\alpha_D = 0.2$, which preferred concentrated distributions

to understand its origins. Nonetheless, current theoretical models need to be revised to explain the diverse set of UDG formation pathways, and exotic galaxies like DF44 in particular.

6.3 Summary of Chapter 5

One of the critical assumptions required in SED-fitting is the nature of the SFH. Both the flexibility and preferred form of the SFH model place strong priors on the mass-weighted age. Chapter 5 addresses the nature of the SFH prior-dependence of galaxy properties head on by attempting to build physically motivated prior distributions using population modelling techniques. A statistical framework was constructed to leverage the information of individual galaxy properties based on an initial set of fiducial fits assuming uncalibrated SFH priors, and recover the true distribution of properties within the population. Any bias imposed by a poorly-calibrated initial prior could be mitigated by importance sampling the posteriors, weighting by the inverse prior probability. The crux of this procedure is that the prior must be sufficiently characterized. While several examples in the literature demonstrate how to calculate the prior distributions for latent model parameters (such as mass-weighted ages, which is derived from the SFH model), this was found to be an incomplete description of the behaviour of the SFH prior. In effect, the lack of a proper characterization of the influence of the SFH prior on the inferred mass-weighted ages precludes the general application of population models for understanding mass-weighted ages. The broader implication is that results presented in the literature may be prior-dependent in unexpected ways.

Given the diverse nature of galaxy properties, any one SFH prior is likely not equally commensurate with a population of galaxies. This point was demonstrated in Chapter 5 for a relatively uniform population of mock quiescent galaxies with mass-weighted ages spanning $\sim$2.5 Gyr. A 'ground truth' experiment compared the properties of mock quiescent galaxies at $z = 1.2$ (modelled after the sample studied in Chapter 3) to the recovered properties from SED-fitting techniques, assuming several common SFH priors. None of the SFH priors were uniformly commensurate with the mock galaxy population. Where a incommensurate prior was assumed, better observational constraints were needed to obtain accurate results – otherwise the results were biased. The nature of the bias was related to how the prior was incommensurate with the true galaxies properties, and thus the bias can be different for different types of galaxies.

Based on the results of Chapter 5, caution is suggested in interpreting trends of star formation timescales from surveys which include a diversity of galaxies, with varying ob-

servational constraints, and span a wide redshift, but assume a fixed SFH prior. Some reported trends may be contrivances of the assumed SFH prior.

6.4 Future directions

6.4.1 Further study of GOGREEN galaxies

Since the publication of Chapter 3, McNab et al. (2021) studied a population of GOGREEN galaxies that are transitioning between star forming and quiescent states, and found an overabundance of such galaxies in clusters, particularly at lower stellar masses. They conclude that the green valley (GV) galaxies are dominated by those that are not transitioning, but those that are quench on fast timescales ($<$ 1 Gyr, *e.g.*, post-starbursts or 'blue-quesicent' galaxies; Belli et al. 2019). This implies that for low mass galaxies, ($< 10^{10.5}$ $M_\odot$) most of the cluster-driven transformation is driven by post-infall quenching mechanisms. There still may be a long delay between infall and the onset of quenching (*e.g.*, Wetzel et al., 2012; De Lucia et al., 2012; Mok et al., 2014; McGee et al., 2014; Foltz et al., 2018), where the suppressed SFRs among star forming galaxies in clusters, relative to the field, suggests an environmental effect independent of galaxy colour (Old et al., 2020). Applying a similar toy model as described in Chapter 3, McNab et al. (2021) interpreted the measured excess in transition galaxies relative to the quiescent galaxy mass functions, and concluded that larger transition timescales ($\sim$1 Gyr) than expected were necessary to simultaneously explain the two observations.

Studying the SFHs of the transition population would be complementary to this work. This thesis has focused on the SFHs of quiescent galaxies, but GOGREEN includes photometry and spectroscopy for a diverse sample of galaxies. The detailed study of galaxies with active star formation requires modelling the nebular emission lines, which has only recently been included in the Prospector framework (Johnson et al., 2021). Another limitation was that the nonparametric models discussed in this work assumed fixed time bins. The widths of the time bins were set according to the expected resolution in ages of stellar populations expected to be recovered by the observations, given the nonlinear evolution of galaxy SEDs. However, this can limit the measurement of rapid fluctuations in SFR. Suess et al. (2022) recently demonstrated the need for more flexible approaches to constrain the quenching timescales of recently-quenched post-starburst galaxies. In Chapter 3, a subset of the UVJ-quiescent galaxies were found to be 'rejuvenated,' exhibiting late bursts of star formation in their late SFHs. Investigating 'rejuvenated,' post-star burst, and GV galaxies among the GOGREEN sample could help constrain the timescale

of the fast-quenching mode proposed by McNab et al. (2021). Whether the low S/N of the GOGREEN observations are a barrier to measuring quenching times remains to be determined.

6.4.2 Consideration of pre-processing for environmental quenching

Reeves et al. (2021) recently compared the relative excess of quiescent galaxies in clusters to that in galaxy groups (halo masses $< 10^{14}$ $M_{\odot}$) using the GOGREEN survey data, finding that the population of massive quiescent galaxies are in place in group environments. A complementary approach was taken by Werner et al. (2022), who investigated the relative importance of environmental quenching versus pre-processing by studying galaxies from the GOGREEN and GCLASS surveys, over $0.8 < z < 1.4$. They compared the quiescent fraction of galaxies between isolated environments, the infall region surrounding clusters, and within clusters. Half of the lower mass galaxies quenched after crossing the virial cluster radius, suggesting that environmentally-driven processes are relevant for this population. In contrast, massive ($> 10^{11}$ $M_{\odot}$) galaxies were either self-quenched or pre-processed prior to infall into clusters.

Put another way, pre-processing in group environments at high redshifts is an important aspect of galaxy evolution for massive quiescent galaxies, but perhaps not for low-mass galaxies. While pre-processing was not included in the toy models explored in Chapter 3, mass-dependent evolution is still difficult to explain with the observed stellar mass functions from van der Burg et al. (2020). This finding also needs to be reconciled with the fact that the mass-weighted ages of massive field and cluster populations are similar, as measured in Chapter 3 (and Chapter 5 Section 5.7.2). Observations of the progenitors of massive cluster galaxies, in proto-cluster environments, will help to reconcile the relevance of pre-processing.

6.4.3 Improved SED and SPS models

Despite the complexity of the SED models discussed in this work, a number of simplifications limit the detailed study of galaxy properties. This issue was particularly relevant in Chapter 4 where high-S/N and high-resolution spectroscopy highlighted several discrepancies with the FSPS SED models.

The SED models used in this work assumed a fixed stellar metallicity. However, the metallicity history of galaxies is complementary to both the stellar mass and star formation

histories. Galaxies are observed to follow a sequence between stellar mass and stellar metallicity (MZR; *e.g.*, Lequeux et al. 1979; Gallazzi et al. 2005) and between stellar mass, gas-phase metallicity, and SFR (*i.e.*, the 'fundamental mass-metallicity relation'; FMR, *e.g.*, Mannucci et al. 2010). Gas-phase metallicities trace the enrichment of the interstellar medium (ISM) by star formation and feedback, or dilution due to cosmological gas inflows or ejections due via stellar winds. Stellar metallicities are sensitive to galaxy growth on timescales on which the enriched ISM is used as fuel for successive generations of stars. Evolution along the FMR has been interpreted to indicate that SFR is self-regulated by stellar feedback or baryonic cycling, where scatter across the FMR is suggestive of the strengths of these processes. Chapter 3 and 5 of this thesis explore using joint prior for stellar mass and metallicity based on the empirical relationship of local quiescent galaxies. If a metallicity enrichment history (MEH) were included it the SED model, it would be natural to link the SFH and MEH (*e.g.*, Tremonti et al., 2004). Including the MEH in the FSPS models appears to be a computational issue; however, other recent codes have had success in including a gas-phase MEH (*e.g.*, Bellstedt et al., 2020). Simultaneous study of the SFH and MEH could help to reveal the key processes which regulate galaxy evolution.

The SED models used in this work also assumed a fixed scaled-solar abundances. This assumption was not suited to the detailed study of DF44, where complementary analysis with abundance-fitting SED models (Villaume et al., 2022) demonstrated distinctly non-solar abundance patterns. Given the connection between star formation feedback from supernova and abundance patterns, the inclusion of abundance pattern response functions in the SED model would be complementary to the inference of the SFH. While full-spectrum abundance matching codes exist in the literature (*e.g.*, alf; Conroy et al., 2018), such techniques have not yet been incorporated into Prospector.

This work assumed a set of SPS models and a fixed initial mass function (IMF); however, there is a rich literature demonstrating the uncertainties associated with both of these components (Conroy, 2013). Comparing results between different SPS models can help mitigate for poor assumptions made about stellar evolution, as discussed in Chapter 1 Section 2.3.1. Pacifici et al. (2022) recently demonstrated that the uncertainties from SED-fitting codes underestimate the differences between results from codes assuming different SPS models or IMF. Different SPS codes vary the contribution of particularly phases of stellar evolution (*e.g.*, TP-AGB stars), which can propagate to uncertainties in the inferred SFHs (*e.g.*, Maraston, 2005). Beyond the galaxy samples discussed in this thesis, detailed observations of galaxies at higher redshifts (such as offered with the James Webb Space Telescope; JWST) may be more sensitive to some aspects of stellar evolution (*e.g.*, stellar rotation or multiplicity) than local galaxies, requiring more flexible SPS models.

The assumption of a given initial mass function (IMF) should also be challenged. Davé

(2008) point out a fact of two uncertainty in between the integrated cosmic SFH and stellar mass functions may be related to the assumption of a Salpeter IMF, where either a Chabrier or Kroupa IMF help resolve the tension. On the other hand, Leja et al. (2017) not that this tension is likely driven by underestimated stellar mass estimates derived with parametric functions.

6.4.4 Machine learning methods

'Inference free' machine learning (ML) algorithms can be constructed to map galaxy properties to observables directly. The application of ML methods for studying galaxy redshifts (*e.g.*, Wadadekar, 2005; Gerdes et al., 2010; Collister & Lahav, 2004; Leistedt & Hogg, 2017) and physical properties (*e.g.*, Masters et al., 2015; Lovell et al., 2019; Stensbo-Smidt et al., 2017; Surana et al., 2020; Simet et al., 2021; Gilda et al., 2021) have been investigated by a number of recent studies. Baron (2019) recently provided a review of ML applications in astronomy. Some shortcomings of SED-fitting approaches related to uncertain SFH or dust models can be bypassed by training neural networks to trace the nonlinear mapping between galaxy fluxes and their properties. Moreover, the properties of an entire population of galaxies can be conveniently constrained simultaneously, thus mapping the population distribution of properties. Gilda et al. (2021) demonstrated the success of ML methods for recovering the properties of mock galaxies over traditional SED-fitting approaches, although they note several caveats and practical limitations of their approach.

One of the caveats is the same as noted in Chapter 5; the ability of ML algorithms to constrain realistic galaxy properties is only as good as the training set is representative of realistic galaxy properties. Training sets are either taken from simulations or empirical libraries. Neither may be fully representative of the true diversity of galaxy formation and evolution, which limits the discovery of novel galaxy populations. While the same issue is also relevant for SED models used for SED-fitting, ML approaches may be less sensitive to outliers (*e.g.*, Gilda et al., 2021) such that interesting populations of exotic galaxies may be overlooked. Overcoming such challenges will take a consolidated effort from observers, modellers, and theorists alike.

6.4.5 Machine learning methods for characterizing the prior space

Chapter 5 attempted to characterize the nature of the SFH prior in order to reconstruct unbiased likelihood distributions from the posteriors output from SED-fitting codes. Simply marginalizing the prior space to construct empirical prior distributions for latent model

parameters, such as mass-weighted age, failed to replicate the observed bias among prior-dependent results. One aspect of this problem was the non-uniform constraints on the shape of the SFH provided by the observations. Mass-weighted ages were sensitive to how the SFH prior propagated information on the recent SFR to the entirety of the SFH. The connection of the mass-weighted age and recent SFR was explored with Neural Network models, but a two-dimensional model also failed to fully characterize the SFH prior imposed bias on the measured ages. Higher-dimensional ML methods would be better suited to this problem. Ideally, such models could also incorporate the redshift dependence of the SFH prior.

6.4.6 Future surveys

Our understanding of galaxy formation and evolution was revolutionized in the early 2000s by the introduction of large surveys like SDSS (Eisenstein et al., 2003), which provided observations for $\sim 10^6$ objects. Upcoming massive surveys will provide yet larger samples, and more rich spectro-photometric datasets (*e.g.*, LSST, Ivezic et al. 2008; MSE, The MSE Science Team et al. 2019). Moreover, detailed studies are now possible with JWST (Gardner et al., 2006), which can provide rest-frame ultraviolet and rest-frame optical spectroscopy of high-redshift galaxies, as well as far-infrared constraints on dust properties for local galaxies. By mapping the diversity of galaxy properties over a wide range of redshifts, these surveys will provide a comprehensive picture of galaxy evolution and lead to new insights in the coming decade. That said, it is anticipated that the best use of these datasets will require improved models and robust statistical tools.

References

Abdurro'uf Lin Y.-T., Wu P.-F., Akiyama M., 2021, ApJS, 254, 15

Abramson L. E., Gladders M. D., Dressler A., Oemler A., Poggianti B., Vulcani B., 2016, ApJ, 832, 7

Acquaviva V., Gawiser E., Guaita L., 2011, ApJ, 737, 47

Acquaviva V., Raichoor A., Gawiser E., 2015, ApJ, 804, 8

Alabi A., et al., 2018, MNRAS, 479, 3308

Alabi A. B., Romanowsky A. J., Forbes D. A., Brodie J. P., Okabe N., 2020, MNRAS, 496, 3182

Allanson S. P., Hudson M. J., Smith R. J., Lucey J. R., 2009, ApJ, 702, 1275

Alsing J., Peiris H., Mortlock D., Leja J., Leistedt B., 2023, ApJS, 264, 29

Amorisco N. C., Loeb A., 2016, MNRAS, 459, L51

Annunziatella M., Pérez-González P. G., García Argumánez Á., Barro G., Alcalde Pampliega B., Costantin L., Koekemoer A. M., Mérida R. M., 2023, MNRAS, 519, 1476

Aragon-Salamanca A., Ellis R. S., Couch W. J., Carter D., 1993, MNRAS, 262, 764

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469

Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1997, ApJ, 488, L75

Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, ApJ, 527, 54

Balogh M. L., Navarro J. F., Morris S. L., 2000, ApJ, 540, 113

Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, ApJ, 615, L101

Balogh M. L., et al., 2014, MNRAS, 443, 2679

Balogh M. L., et al., 2016, MNRAS, 456, 4364

Balogh M. L., et al., 2017, MNRAS, 470, 4168

Balogh M. L., et al., 2020, MNRAS, 500, 358

Barbosa C. E., et al., 2020, ApJS, 247, 46

Baron D., 2019, arXiv e-prints, p. arXiv:1904.07248

Barro G., et al., 2013, ApJ, 765, 104

Barro G., et al., 2017, ApJ, 840, 47

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Beifiori A., et al., 2017, ApJ, 846, 120

Bell E. F., de Jong R. S., 2001, ApJ, 550, 212

Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289

Bell E. F., et al., 2004, ApJ, 608, 752

Bell E. F., et al., 2005, ApJ, 625, 23

Belli S., Newman A. B., Ellis R. S., 2015, ApJ, 799, 206

Belli S., Newman A. B., Ellis R. S., 2019, ApJ, 874, 17

Bellstedt S., et al., 2020, MNRAS, 498, 5581

Bellstedt S., et al., 2021, MNRAS, 503, 3309

Benavides J. A., et al., 2021, Nature Astronomy, pp 1–6

Bender R., Ziegler B., Bruzual G., 1996, ApJ, 463, L51

Bernardi M., Renzini A., da Costa L. N., Wegner G., Alonso M. V., Pellegrini P. S., Rité C., Willmer C. N. A., 1998, ApJ, 508, L143

Bernardi M., Sheth R. K., Nichol R. C., Schneider D. P., Brinkmann J., 2005, AJ, 129, 61

Bernardi M., Nichol R. C., Sheth R. K., Miller C. J., Brinkmann J., 2006, AJ, 131, 1288

Betancourt M., 2012. eprint: arXiv:1010.3436, pp 157–164

Bezanson R., van Dokkum P., van de Sande J., Franx M., Kriek M., 2013, ApJ, 764, L8

Biviano A., et al., 2021, A&A, 650, A105

Blanton M. R., et al., 2003, ApJ, 594, 186

Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005, ApJ, 629, 143

Bogdán Á., 2020, ApJ, 901, L30

Bolzonella M., Miralles J. M., Pelló R., 2000, A&A, 363, 476

Boselli A., Gavazzi G., 2006, Publications of the Astronomical Society of the Pacific, 118, 517

Bower R. G., Lucey J. R., Ellis R. S., 1992, MNRAS, 254, 601

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Bower R. G., Benson A. J., Crain R. A., 2012, MNRAS, 422, 2816

Bowman W. P., et al., 2020, ApJ, 899, 7

Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503

Brammer G. B., et al., 2011, ApJ, 739, 24

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Brodwin M., et al., 2010, ApJ, 721, 90

Bruzual A. G., 1983, ApJ, 273, 105

Bruzual A. G., 2007, in Vazdekis A., Peletier R., eds, Vol. 241, Stellar Populations as Building Blocks of Galaxies. pp 125–132 (arXiv:astro-ph/0703052), doi:10.1017/S1743921307007624

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Bruzual A. G., Charlot S., 1993, ApJ, 405, 538

Bundy K., Ellis R. S., Conselice C. J., 2005, ApJ, 625, 621

Burstein D., Faber S. M., Gaskell C. M., Krumm N., 1984, ApJ, 287, 586

Butcher H., Oemler A. J., 1984, ApJ, 285, 426

Buzzo M. L., et al., 2022, MNRAS, 517, 2231

Cabrera-Ziri I., Conroy C., 2022, MNRAS

Calistro Rivera G., Lusso E., Hennawi J. F., Hogg D. W., 2016, ApJ, 833, 98

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Capozzi D., Maraston C., Daddi E., Renzini A., Strazzullo V., Gobat R., 2016, Monthly Notices of the Royal Astronomical Society, 456, 790

Cappellari M., Emsellem E., 2004, PASP, 116, 138

Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

Carleton T., Errani R., Cooper M., Kaplinghat M., Peñarrubia J., Guo Y., 2019, MNRAS, 485, 382

Carleton T., Guo Y., Munshi F., Tremmel M., Wright A., 2021, MNRAS, 502, 398

Carnall A. C., McLure R. J., Dunlop J. S., Davé R., 2018, MNRAS, 480, 4379

Carnall A. C., et al., 2019a, MNRAS, 490, 417

Carnall A. C., Leja J., Johnson B. D., McLure R. J., Dunlop J. S., Conroy C., 2019b, ApJ, 873, 44

Chabrier G., 2003, PASP, 115, 763

Chan T. K., Kereš D., Wetzel A., Hopkins P. F., Faucher-Giguère C. A., El-Badry K., Garrison-Kimmel S., Boylan-Kolchin M., 2018, MNRAS, 478, 906

Chan J. C. C., et al., 2019, ApJ, 880, 119

Charlot S., Fall S. M., 2000, ApJ, 539, 718

Chauke P., et al., 2018, ApJ, 861, 13

Chevallard J., Charlot S., 2016, MNRAS, 462, 1415

Choi J., Conroy C., Moustakas J., Graves G. J., Holden B. P., Brodwin M., Brown M. J. I., van Dokkum P. G., 2014, ApJ, 792, 95

Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, ApJ, 823, 102

Ciesla L., Elbaz D., Fensch J., 2017, A&A, 608, A41

Cimatti A., Daddi E., Renzini A., 2006, A&A, 453, L29

Citro A., Pozzetti L., Moresco M., Cimatti A., 2016, A&A, 592, A19

Clemens M. S., Bressan A., Nikolic B., Alexander P., Annibali F., Rampazzo R., 2006, MNRAS, 370, 702

Cole S., 1991, ApJ, 367, 45

Collister A. A., Lahav O., 2004, PASP, 116, 345

Connolly A. J., Budavári T., Szalay A. S., Csabai I., Brunner R. J., 1999, in Weymann R., Storrie-Lombardi L., Sawicki M., Brunner R., eds, Astronomical Society of the Pacific Conference Series Vol. 191, Photometric Redshifts and the Detection of High Redshift Galaxies. p. 13

Conroy C., 2013, ARA&A, 51, 393

Conroy C., Gunn J. E., 2010a, p. ascl:1010.043

Conroy C., Gunn J. E., 2010b, ApJ, 712, 833

Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

Conroy C., Graves G. J., van Dokkum P. G., 2013, ApJ, 780, 33

Conroy C., Villaume A., van Dokkum P. G., Lind K., 2018, ApJ, 854, 139

Cooper M. C., Gallazzi A., Newman J. A., Yan R., 2010, MNRAS, 402, 1942

Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, AJ, 112, 839

Croton D. J., et al., 2006, MNRAS, 365, 11

Cutler S. E., Giavalisco M., Ji Z., Cheng Y., 2022, arXiv e-prints, p. arXiv:2208.01653

Dalcanton J. J., Spergel D. N., Summers F. J., 1997, ApJ, 482, 659

Danieli S., et al., 2021, arXiv e-prints, p. arXiv:2111.14851

Darvish B., Mobasher B., Sobral D., Rettura A., Scoville N., Faisst A., Capak P., 2016, ApJ, 825, 113

Davé R., 2008, MNRAS, 385, 147

Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98

Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, MNRAS, 497, 146

Davidzon I., et al., 2017, A&A, 605, A70

Davidzon I., et al., 2019, MNRAS, 489, 4817

De Lucia G., Weinmann S., Poggianti B. M., Aragón-Salamanca A., Zaritsky D., 2012, MNRAS, 423, 1277

De Propris R., Stanford S. A., Eisenhardt P. R., Dickinson M., Elston R., 1999, AJ, 118, 719

Demarco R., et al., 2010, ApJ, 711, 1185

Di Cintio A., Brook C. B., Dutton A. A., Macciò A. V., Obreja A., Dekel A., 2017, MNRAS, 466, L1

Dickey C. M., et al., 2021, ApJ, 915, 53

Diemer B., Sparre M., Abramson L. E., Torrey P., 2017, ApJ, 839, 26

Digby R., et al., 2019, MNRAS, 485, 5423

Djorgovski S., Davis M., 1987, ApJ, 313, 59

Dotter A., 2016, ApJS, 222, 8

Draine B. T., Lee H. M., 1984, ApJ, 285, 89

Draine B. T., et al., 2007, ApJ, 663, 866

Dressler A., 1980, ApJ, 236, 351

Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, ApJ, 313, 42

Dressler A., et al., 2016, ApJ, 833, 251

Dressler A., Kelson D. D., Abramson L. E., 2018, ApJ, 869, 152

Driver S. P., et al., 2022, MNRAS, 513, 439

Drory N., Salvato M., Gabasch A., Bender R., Hopp U., Feulner G., Pannella M., 2005, ApJ, 619, L131

Drouart G., Falkendal T., 2018, MNRAS, 477, 4981

Eigenthaler P., et al., 2018, ApJ, 855, 142

Eisenstein D. J., et al., 2003, ApJ, 585, 694

El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, ApJ, 820, 131

Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, Publ. Astron. Soc. Australia, 34, e058

Ellis R. S., Smail I., Dressler A., Couch W. J., Oemler Jr. A., Butcher H., Sharples R. M., 1997, ApJ, 483, 582

Emsellem E., et al., 2011, MNRAS, 414, 888

Estrada-Carpenter V., et al., 2019, ApJ, 870, 133

Estrada-Carpenter V., et al., 2020, ApJ, 898, 171

Eufrasio R. T., 2017, Lightning: SED Fitting Package, Astrophysics Source Code Library, record ascl:1711.009 (ascl:1711.009)

Faber S. M., 1973, ApJ, 179, 731

Faber S. M., Jackson R. E., 1976, ApJ, 204, 668

Faber S. M., Dressler A., Davies R. L., Burstein D., Lynden Bell D., Terlevich R., Wegner G., 1987, in Faber S. M., ed., Nearly Normal Galaxies. From the Planck Time to the Present. p. 175

Faber S. M., et al., 2007, ApJ, 665, 265

Fall S. M., Efstathiou G., 1980, MNRAS, 193, 189

Fazio G. G., et al., 2004, ApJS, 154, 10

Ferland G. J., et al., 2013, Rev. Mex. Astron. Astrofis., 49, 137

Ferré-Mateu A., et al., 2018, MNRAS, 479, 4891

Ferreras I., Silk J., 2000, ApJ, 541, L37

Ferreras I., et al., 2019, MNRAS, 486, 1358

Fillingham S. P., Cooper M. C., Boylan-Kolchin M., Bullock J. S., Garrison-Kimmel S., Wheeler C., 2018, MNRAS, 477, 4491

Fioc M., Rocca-Volmerange B., 2019, A&A, 623, A143

Foley R. J., et al., 2011, ApJ, 731, 86

Folkes S., et al., 1999, MNRAS, 308, 459

Foltz R., et al., 2018, ApJ, 866, 136

Font A. S., et al., 2008, MNRAS, 389, 1619

Fontana A., et al., 2004, A&A, 424, 23

Forbes D. A., Gannon J. S., Romanowsky A. J., Alabi A., Brodie J. P., Couch W. J., Ferré-Mateu A., 2021, MNRAS, 500, 1279

Foreman-Mackey D., et al., 2013, p. ascl:1303.002

Foreman-Mackey D., Hogg D. W., Morton T. D., 2014, ApJ, 795, 64

Forrest B., et al., 2020a, ApJ, 890, L1

Forrest B., et al., 2020b, ApJ, 903, 47

Freeman K. C., 1970, ApJ, 160, 811

Fujita Y., 2004, PASJ, 56, 29

Galametz A., et al., 2013, ApJS, 206, 10

Gallazzi A., Charlot S., White S. D. M., Brinchmann J., 2004, IAU, 2004

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., 2006, MNRAS, 370, 1106

Gallazzi A., Bell E. F., Zibetti S., Brinchmann J., Kelson D. D., 2014, ApJ, 788, 72

Galliano F., 2018, MNRAS, 476, 1445

Galliano F., Galametz M., Jones A. P., 2018, ARA&A, 56, 673

Gannon J. S., et al., 2022, MNRAS, 510, 946

Gardner J. P., Mather J. C., Clampin M., et al., 2006, Space Sci. Rev., 123, 485

Gargiulo A., et al., 2009, MNRAS, 397, 75

Garrison-Kimmel S., et al., 2019, MNRAS, 487, 1380

Gawiser E., 2009, New Astron. Rev., 53, 50

Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, ApJ, 757, 85

Geller M. J., Huchra J. P., 1983, ApJS, 52, 61

Gerdes D. W., Sypniewski A. J., McKay T. A., Hao J., Weis M. R., Wechsler R. H., Busha M. T., 2010, ApJ, 715, 823

Gilda S., Lower S., Narayanan D., 2021, ApJ, 916, 43

Gladders M. D., Oemler A., Dressler A., Poggianti B., Vulcani B., Abramson L., 2013, ApJ, 770, 64

Glazebrook K., et al., 2004, Nature, 430, 181

Glazebrook K., et al., 2017, Nature, 544, 71

Gobat R., Rosati P., Strazzullo V., Rettura A., Demarco R., Nonino M., 2008, A&A, 488, 853

Gordon K. D., Clayton G. C., Misselt K. A., Landolt A. U., Wolff M. J., 2003, ApJ, 594, 279

Graves G. J., Faber S. M., 2010, The Astrophysical Journal, 717, 803

Graves G. J., Schiavon R. P., 2008, ApJS, 177, 446

Graves G. J., Faber S. M., Schiavon R. P., Yan R., 2007, ApJ, 671, 243

Graves G. J., Faber S. M., Schiavon R. P., 2009, The Astrophysical Journal, 693, 486

Greco J. P., Goulding A. D., Greene J. E., Strauss M. A., Huang S., Kim J. H., Komiyama Y., 2018, ApJ, 866, 112

Grishin K., Chilingarian I., Afanasiev A., Fabricant D., Katkov I., Moran S., Yagi M., 2021, Nature Astronomy

Gu Y., Fang G., Yuan Q., Cai Z., Wang T., 2018a, ApJ, 855, 10

Gu M., et al., 2018b, ApJ, 859, 37

Guglielmo V., et al., 2019, A&A, 625, A112

Guiderdoni B., Rocca-Volmerange B., 1987, A&A, 186, 1

Gunn J. E., Gott III J. R., 1972, ApJ, 176, 1

Han Y., Han Z., 2018, ApJS, 240, 3

Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039

Hayward C. C., Smith D. J. B., 2015, MNRAS, 446, 1512

Heavens A. F., Jimenez R., Lahav O., 2000, MNRAS, 317, 965

Heavens A., Panter B., Jimenez R., Dunlop J., 2004, Nature, 428, 625

Heckman T. M., 1980, A&A, 500, 187

Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., 2003, MNRAS, 338, 903

Higson E., Handley W., Hobson M., Lasenby A., 2019, Statistics and Computing, Vol. 29, No. 5, pp. 891-913, 29, 891

Hinton S. R., Davis T. M., Lidman C., Glazebrook K., Lewis G. F., 2016, Astronomy and Computing, 15, 61

Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760

Hogg D. W., Cohen J. G., Blandford R., Pahre M. A., 1998, ApJ, 504, 622

Hogg D. W., Bovy J., Lang D., 2010a, arXiv, p. arXiv:1008.4686

Hogg D. W., Myers A. D., Bovy J., 2010b, ApJ, 725, 2166

Hubble E. P., 1926, ApJ, 64, 321

Ilbert O., et al., 2013, A&A, 556, A55

Ivezic Z., et al., 2008, Serbian Astronomical Journal, 176, 1

Iyer K., Gawiser E., 2017, ApJ, 838, 127

Iyer K. G., Gawiser E., Faber S. M., Ferguson H. C., Kartaltepe J., Koekemoer A. M., Pacifici C., Somerville R. S., 2019, ApJ, 879, 116

Iyer K. G., et al., 2020, MNRAS, 498, 430

Jackson R. A., et al., 2021, MNRAS, 502, 4262

Jaffé Y. L., et al., 2016, MNRAS, 461, 1202

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Jhora99 2021, jhora99/imclean: imclean 20210323, Zenodo, doi:10.5281/zenodo.4850526

Ji Z., Giavalisco M., 2022, ApJ, 935, 120

Jian H.-Y., et al., 2017, ApJ, 845, 74

Jiang F., Dekel A., Freundlich J., Romanowsky A. J., Dutton A. A., Macciò A. V., Di Cintio A., 2019a, MNRAS, 487, 5272

Jiang F., et al., 2019b, MNRAS, 488, 4801

Jimenez R., Panter B., Heavens A. F., Verde L., 2005, MNRAS, 356, 495

Johnson B. D., Leja J., Conroy C., Speagle J. S., 2020, bd-j/prospector: prospector v1.0.0, Zenodo, doi:10.5281/zenodo.4586953, https://ui.adsabs.harvard.edu/abs/2020zndo...4586953J

Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021, ApJS, 254, 22

Jørgensen I., 1999, MNRAS, 306, 607

Jørgensen I., Chiboucas K., 2013, AJ, 145, 77

Jørgensen I., Chiboucas K., Flint K., Bergmann M., Barr J., Davies R., 2006, ApJ, 639, L9

Jørgensen I., Chiboucas K., Flint K., Bergmann M., Barr J., Davies R., 2007, ApJ, 654, L179

Jørgensen I., Chiboucas K., Berkson E., Smith O., Takamiya M., Villaume A., 2017, AJ, 154, 251

Jørgensen I., Chiboucas K., Webb K., Woodrum C., 2018, AJ, 156, 224

Joshi G. D., Pillepich A., Nelson D., Zinger E., Marinacci F., Springel V., Vogelsberger M., Hernquist L., 2021, MNRAS, 508, 1652

Juneau S., et al., 2005, ApJ, 619, L135

Kadowaki J., Zaritsky D., Donnerstein R. L., 2017, ApJ, 838, L21

Kass R. E., Raftery A. E., 1995, Journal of the American Statistical Association, 90, 773

Kauffmann G., et al., 2003, MNRAS, 341, 54

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713

Kauffmann G., Heckman T. M., De Lucia G., Brinchmann J., Charlot S., Tremonti C., White S. D. M., Brinkmann J., 2006, MNRAS, 367, 1394

Kaviraj S., et al., 2007, ApJS, 173, 619

Kawinwanichakij L., et al., 2017, ApJ, 847, 134

Kelly B. C., 2012, in Statistical Challenges in Modern Astronomy V. pp 147–162 (arXiv:1112.1745), doi:10.1007/978-1-4614-3520-4˙13

Kelson D. D., Holden B. P., 2010, ApJ, 713, L28

Kelson D. D., Illingworth G. D., van Dokkum P. G., Franx M., 2000, ApJ, 531, 184

Kelson D. D., Illingworth G. D., Franx M., van Dokkum P. G., 2001, ApJ, 552, L17

Kelson D. D., et al., 2014, ApJ, 783, 110

Kennicutt Robert C. J., 1989, ApJ, 344, 685

Kennicutt R. C., 1998, ARA&A, 36, 189

Khochfar S., Silk J., 2009, MNRAS, 397, 506

Khullar G., et al., 2022, ApJ, 934, 177

Kimm T., et al., 2009, MNRAS, 394, 1131

Kodama T., et al., 2004, MNRAS, 350, 1005

Kormendy J., 1985, ApJ, 295, 73

Koyama Y., et al., 2013, MNRAS, 434, 423

Kriek M., Conroy C., 2013, ApJ, 775, L16

Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221

Kriek M., et al., 2010, ApJ, 722, L64

Kriek M., et al., 2016, Nature, 540, 248

Kriek M., et al., 2019, ApJ, 880, L31

Kukstas E., et al., 2023, MNRAS, 518, 4782

Labbé I., et al., 2005, ApJ, 624, L81

Larson R. B., Tinsley B. M., Caldwell C. N., 1980, ApJ, 237, 692

Lauer T. R., 1985, ApJ, 292, 104

Lee S.-K., 2007, in American Astronomical Society Meeting Abstracts #210. p. 121.07

Lee-Brown D. B., et al., 2017, ApJ, 844, 43

Lee S.-K., Idzi R., Ferguson H. C., Somerville R. S., Wiklind T., Giavalisco M., 2009, ApJS, 184, 100

Lee S.-K., Ferguson H. C., Somerville R. S., Wiklind T., Giavalisco M., 2010, ApJ, 725, 1644

Lee C. H., Hodges-Kluck E., Gallo E., 2020, MNRAS, 497, 2759

Leethochawalit N., Kirby E. N., Moran S. M., Ellis R. S., Treu T., 2018, ApJ, 856, 15

Leistedt B., Hogg D. W., 2017, ApJ, 838, 5

Leistedt B., Alsing J., Peiris H., Mortlock D., Leja J., 2023, ApJS, 264, 23

Leitherer C., Alloin D., Fritze-v. Alvensleben U., et al., 1996, PASP, 108, 996

Leitner S. N., 2012, ApJ, 745, 149

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., 2019a, ApJ, 876, 3

Leja J., et al., 2019b, ApJ, 877, 140

Leja J., Speagle J. S., Johnson B. D., Conroy C., van Dokkum P., Franx M., 2020, ApJ, 893, 111

Leja J., et al., 2022, ApJ, 936, 165

Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, A&A, 80, 155

Li Y., Leja J., 2022, ApJ, 940, 88

Liao S., et al., 2019, MNRAS, 490, 5182

Lin L., et al., 2014, ApJ, 782, 33

Loh Y., Ellingson E., Yee H. K. C., Gilbank D. G., Gladders M. D., Barrientos L. F., 2008, ApJ, 680, 214

Lonsdale C. J., et al., 2003, PASP, 115, 897

Loredo T. J., Hendry M. A., 2019, arXiv e-prints, p. arXiv:1911.12337

Lovell C. C., Acquaviva V., Thomas P. A., Iyer K. G., Gawiser E., Wilkins S. M., 2019, Monthly Notices of the Royal Astronomical Society, 490, 5503

Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, ApJ, 904, 33

Lupi A., Volonteri M., Silk J., 2017, MNRAS, 470, 1673

MacArthur L. A., González J. J., Courteau S., 2009, MNRAS, 395, 28

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Maier C., et al., 2016, A&A, 590, A108

Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115

Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2021, ApJ, 907, 85

Maraston C., 2005, MNRAS, 362, 799

Maraston C., Thomas D., 2000, ApJ, 541, 126

Maraston C., Pforr J., Renzini A., Daddi E., Dickinson M., Cimatti A., Tonini C., 2010, MNRAS, 407, 830

Marigo P., Girardi L., Bressan A., Groenewegen M. A. T., Silva L., Granato G. L., 2008, A&A, 482, 883

Marigo P., et al., 2017, ApJ, 835, 77

Marinacci F., et al., 2018, MNRAS, 480, 5113

Martín-Navarro I., et al., 2019, MNRAS, 484, 3425

Masters D., et al., 2015, ApJ, 813, 53

Mathis J. S., Rumpl W., Nordsieck K. H., 1977, ApJ, 217, 425

Mauduit J.-C., et al., 2012, PASP, 124, 714

McConachie I., et al., 2022, ApJ, 926, 37

McDermid R. M., et al., 2015, MNRAS, 448, 3484

McGee S. L., Balogh M. L., Bower R. G., Font A. S., McCarthy I. G., 2009, MNRAS, 400, 937

McGee S. L., Balogh M. L., Wilman D. J., Bower R. G., Mulchaey J. S., Parker L. C., Oemler A., 2011, MNRAS, 413, 996

McGee S. L., Bower R. G., Balogh M. L., 2014, MNRAS, 442, L105

McLure R. J., et al., 2018, MNRAS

McNab K., et al., 2021, MNRAS, 508, 157

Mehta V., et al., 2018, ApJS, 235, 36

Merritt D., 1983, ApJ, 264, 24

Mitchell P. D., Lacey C. G., Cole S., Baugh C. M., 2014, MNRAS, 444, 2637

Mok A., et al., 2014, MNRAS, 438, 3070

Monaco L., Bellazzini M., Ferraro F. R., Pancino E., 2003, ApJ, 597, L25

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613

Moran S. M., Ellis R. S., Treu T., Smith G. P., Rich R. M., Smail I., 2007, ApJ, 671, 1503

Moresco M., et al., 2010, A&A, 524, A67

Morishita T., et al., 2018, ApJ, 856, L4

Morishita T., et al., 2019, ApJ, 877, 141

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

Moultaka J., Pelat D., 2000, MNRAS, 314, 409

Moultaka J., Boisson C., Joly M., Pelat D., 2004, A&A, 420, 459

Mowla L., van Dokkum P., Merritt A., Abraham R., Yagi M., Koda J., 2017, ApJ, 851, 27

Muzzin A., Marchesini D., van Dokkum P. G., Labbé I., Kriek M., Franx M., 2009, ApJ, 701, 1839

Muzzin A., et al., 2012, ApJ, 746, 188

Muzzin A., et al., 2013a, ApJS, 206, 8

Muzzin A., Wilson G., Demarco R., Lidman C., Nantais J., Hoekstra H., Yee H. K. C., Rettura A., 2013b, ApJ, 767, 39

Muzzin A., et al., 2013c, ApJ, 777, 18

Muzzin A., et al., 2014, ApJ, 796, 65

Nagaraj G., Ciardullo R., Bowman W. P., Gronwall C., 2021, ApJ, 913, 34

Naiman J. P., et al., 2018, MNRAS, 477, 1206

Nantais J. B., et al., 2016, A&A, 592, A161

Nantais J. B., et al., 2017, MNRAS, 465, L104

Neistein E., van den Bosch F. C., Dekel A., 2006, MNRAS, 372, 933

Nelan J. E., Smith R. J., Hudson M. J., Wegner G. A., Lucey J. R., Moore S. A. W., Quinney S. J., Suntzeff N. B., 2005, ApJ, 632, 137

Nelson D., et al., 2018, MNRAS, 475, 624

Nelson D., et al., 2019, MNRAS, 490, 3234

Newman A. B., Ellis R. S., Andreon S., Treu T., Raichoor A., Trinchieri G., 2014, ApJ, 788, 51

Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793

Ocvirk P., 2010, ApJ, 709, 88

Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006a, MNRAS, 365, 46

Ocvirk P., Pichon C., Lançon A., Thiébaut E., 2006b, MNRAS, 365, 74

Odekon M. C., et al., 2016, ApJ, 824, 110

Old L. J., et al., 2020, MNRAS, 493, 5987

Onodera M., et al., 2012, ApJ, 755, 26

Onodera M., et al., 2015, ApJ, 808, 161

Pacifici C., Charlot S., Blaizot J., Brinchmann J., 2012, MNRAS, 421, 2002

Pacifici C., et al., 2015, MNRAS, 447, 786

Pacifici C., et al., 2016a, ApJ, 832, 79

Pacifici C., et al., 2016b, ApJ, 832, 79

Pacifici C., Iyer K. G., Mobasher B., et al., 2022, arXiv e-prints, p. arXiv:2212.01915

Pallero D., Gómez F. A., Padilla N. D., Torres-Flores S., Demarco R., Cerulo P., Olave-Rojas D., 2019, MNRAS, 488, 847

Pandya V., et al., 2018, ApJ, 858, 29

Panter B., Heavens A. F., Jimenez R., 2003, MNRAS, 343, 1145

Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, MNRAS, 378, 1550

Panter B., Jimenez R., Heavens A. F., Charlot S., 2008, MNRAS, 391, 1117

Papastergis E., Adams E. a. K., Romanowsky A. J., 2017, A&A, 601, L10

Papovich C., Dickinson M., Ferguson H. C., 2001, ApJ, 559, 620

Patel S. G., Holden B. P., Kelson D. D., Franx M., van der Wel A., Illingworth G. D., 2012, ApJ, 748, L27

Paulino-Afonso A., Sobral D., Darvish B., Ribeiro B., Smail I., Best P., Stroe A., Cairns J., 2020, A&A, 633, A70

Paxton B., Bildsten L., Dotter A., Herwig F., Lesaffre P., Timmes F., 2011, ApJS, 192, 3

Paxton B., et al., 2013, ApJS, 208, 4

Paxton B., et al., 2015, ApJS, 220, 15

Paxton B., et al., 2018, ApJS, 234, 34

Peeples M. S., Werk J. K., Tumlinson J., Oppenheimer B. D., Prochaska J. X., Katz N., Weinberg D. H., 2014, ApJ, 786, 54

Peletier R. F., 1989, PhD thesis, -

Peng Y.-j., et al., 2010, ApJ, 721, 193

Peng Y.-j., Lilly S. J., Renzini A., Carollo M., 2012, ApJ, 757, 4

Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192

Pentericci L., McLure R. J., Garilli B., et al., 2018, A&A, 616, A174

Péroux C., Howk J. C., 2020, ARA&A, 58, annurev

Pforr J., Maraston C., Tonini C., 2012, MNRAS, 422, 3285

Pillepich A., et al., 2018, MNRAS, 475, 648

Pillepich A., et al., 2019, MNRAS, 490, 3196

Pintos-Castro I., Yee H. K. C., Muzzin A., Old L., Wilson G., 2019, ApJ, 876, 40

Poggianti B. M., Smail I., Dressler A., Couch W. J., Barger A. J., Butcher H., Ellis R. S., Oemler Jr. A., 1999, ApJ, 518, 576

Poggianti B. M., et al., 2001, ApJ, 562, 689

Poggianti B. M., et al., 2006, ApJ, 642, 188

Poggianti B. M., et al., 2008, ApJ, 684, 888

Poggianti B. M., et al., 2009, ApJ, 693, 112

Polzin A., van Dokkum P., Danieli S., Greco J. P., Romanowsky A. J., 2021, ApJ, 914, L23

Popesso P., et al., 2011, A&A, 532, A145

Pozzetti L., et al., 2003, A&A, 402, 837

Qiu Y., Kang X., 2022, ApJ, 930, 66

Qiu Y., Kang X., Luo Y., 2023, MNRAS, 519, 2268

Raichoor A., et al., 2011, ApJ, 732, 12

Reddy N. A., Pettini M., Steidel C. C., Shapley A. E., Erb D. K., Law D. R., 2012, ApJ, 754, 25

Reddy N. A., et al., 2015, ApJ, 806, 259

Reed D., Gardner J., Quinn T., Stadel J., Fardal M., Lake G., Governato F., 2003, MNRAS, 346, 565

Reed D. S., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, MNRAS, 374, 2

Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541

Reeves A. M. M., et al., 2021, MNRAS, 506, 3364

Renzini A., 2006, ARA&A, 44, 141

Renzini A., 2016, MNRAS, 460, L45

Renzini A., Ciotti L., 1993, The Astrophysical Journal, 416, L49

Rettura A., et al., 2010, ApJ, 709, 512

Rettura A., et al., 2011, ApJ, 732, 94

Rinaldi P., Caputi K. I., van Mierlo S., Ashby M. L. N., Caminha G. B., Iani E., 2021, arXiv:2112.03935 [astro-ph]

Rong Y., Guo Q., Gao L., Liao S., Xie L., Puzia T. H., Sun S., Pan J., 2017, MNRAS, 470, 4231

Rosario D. J., 2019, FortesFit: Flexible spectral energy distribution modelling with a Bayesian backbone, Astrophysics Source Code Library, record ascl:1904.011 (ascl:1904.011)

Rudie G. C., et al., 2012, ApJ, 750, 67

Rudnick G., et al., 2017, ApJ, 850, 181

Ruiz-Lara T., et al., 2015, A&A, 583, A60

Ruiz-Lara T., et al., 2018, MNRAS, 478, 2034

Saglia R. P., Maraston C., Greggio L., Bender R., Ziegler B., 2000, A&A, 360, 911

Saglia R. P., et al., 2010, A&A, 524, A6

Saifollahi T., Trujillo I., Beasley M. A., Peletier R. F., Knapen J. H., 2021, MNRAS, 502, 5921

Saifollahi T., Zaritsky D., Trujillo I., Peletier R. F., Knapen J. H., Amorisco N., Beasley M. A., Donnerstein R., 2022, arXiv:2201.11750 [astro-ph]

Sales L. V., Navarro J. F., Peñafiel L., Peng E. W., Lim S., Hernquist L., 2020, MNRAS, 494, 1848

Salim S., Narayanan D., 2020, ARA&A, 58, 529

Salmon B., et al., 2016, ApJ, 827, 20

Salvador-Rusiñol N., Vazdekis A., La Barbera F., Beasley M. A., Ferreras I., Negri A., Dalla Vecchia C., 2020, Nature Astronomy, 4, 252

Sánchez-Blázquez P., Gorgas J., Cardiel N., González J. J., 2006, A&A, 457, 787

Sánchez-Blázquez P., et al., 2009, A&A, 499, 47

Sánchez-Blázquez P., Ocvirk P., Gibson B. K., Pérez I., Peletier R. F., 2011a, MNRAS, 415, 709

Sánchez-Blázquez P., Ocvirk P., Gibson B. K., Pérez I., Peletier R. F., 2011b, MNRAS, 415, 709

Sánchez S. F., et al., 2016, Rev. Mex. Astron. Astrofis., 52, 21

Sandage A., Visvanathan N., 1978, ApJ, 223, 707

Sanders D. B., et al., 2007, ApJS, 172, 86

Sandles L., Curtis-Lake E., Charlot S., Chevallard J., Maiolino R., 2022, MNRAS, 515, 2951

Saracco P., Gargiulo A., Ciocca F., Marchesini D., 2017, A&A, 597, A122

Saracco P., La Barbera F., Gargiulo A., Mannucci F., Marchesini D., Nonino M., Ciliegi P., 2019, MNRAS, 484, 2281

Saracco P., Gargiulo A., La Barbera F., Annunziatella M., Marchesini D., 2020, MNRAS, 491, 1777

Saracco P., et al., 2023, MNRAS

Schawinski K., et al., 2014, MNRAS, 440, 889

Schaye J., et al., 2010, MNRAS, 402, 1536

Schiavon R. P., 2007, ApJS, 171, 146

Schiavon R. P., Rose J. A., Courteau S., MacArthur L. A., 2004, ApJ, 608, L33

Schiavon R. P., Rose J. A., Courteau S., MacArthur L. A., 2005, ApJS, 160, 163

Schiavon R. P., et al., 2006, ApJ, 651, L93

Schlafly E. F., Finkbeiner D. P., 2011, ApJ, 737, 103

Schmidt M., 1959, ApJ, 129, 243

Schreiber C., et al., 2015, A&A, 575, A74

Schreiber C., et al., 2018a, A&A, 611, A22

Schreiber C., et al., 2018b, A&A, 618, A85

Serra P., Trager S. C., 2007, MNRAS, 374, 769

Setton D. J., et al., 2022, arXiv e-prints, p. arXiv:2212.05070

Simet M., Chartab N., Lu Y., Mobasher B., 2021, ApJ, 908, 47

Simha V., Weinberg D. H., Conroy C., Dave R., Fardal M., Katz N., Oppenheimer B. D., 2014, arXiv e-prints, p. arXiv:1404.0402

Singh R., et al., 2013, A&A, 558, A43

Skilling J., 2004,] 10/d9p42j, 735, 395

Smith R. J., Lucey J. R., Hudson M. J., Allanson S. P., Bridges T. J., Hornschemeier A. E., Marzke R. O., Miller N. A., 2009, MNRAS, 392, 1265

Snyder G. F., et al., 2012, ApJ, 756, 114

Sobral D., Best P. N., Smail I., Mobasher B., Stott J., Nisbet D., 2014, MNRAS, 437, 3516

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

Spinrad H., Taylor B. J., 1969, ApJ, 157, 1279

Springel V., 2005, MNRAS, 364, 1105

Springel V., et al., 2018, MNRAS, 475, 676

Stalder B., et al., 2013, ApJ, 763, 93

Stanford S. A., Eisenhardt P. R., Dickinson M., 1998, ApJ, 492, 461

Stensbo-Smidt K., Gieseke F., Igel C., Zirm A., Steenstrup Pedersen K., 2017, MNRAS, 464, 2577

Straatman C. M. S., et al., 2014, ApJ, 783, L14

Straatman C. M. S., et al., 2018, ApJS, 239, 27

Strateva I., et al., 2001, AJ, 122, 1861

Strazzullo V., et al., 2013, ApJ, 772, 118

Suess K. A., et al., 2022, ApJ, 935, 146

Sukay E., et al., 2022, ApJ, 940, 42

Surana S., Wadadekar Y., Bait O., Bhosale H., 2020, MNRAS, 493, 4808

Tacchella S., et al., 2022b, ApJ, 926, 134

Tacchella S., et al., 2022a, ApJ, 926, 134

Tanaka M., et al., 2013, ApJ, 772, 113

Tanaka M., et al., 2019, ApJ, 885, L34

Tanaka M., et al., 2020, ApJ, 894, L13

Taylor E. N., et al., 2011, MNRAS, 418, 1587

The MSE Science Team Babusiaux C., Bergemann M., et al., 2019, arXiv e-prints, p. arXiv:1904.04907

Thomas D., Greggio L., Bender R., 1999, MNRAS, 302, 537

Thomas D., Maraston C., Bender R., 2003, MNRAS, 339, 897

Thomas D., Maraston C., Bender R., de Oliveira C. M., 2005, ApJ, 621, 673

Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, MNRAS

Thomas R., et al., 2017, A&A, 602, A35

Thorne J. E., et al., 2021, MNRAS, 505, 540

Tinsley B. M., 1968, ApJ, 151, 547

Tinsley B. M., 1980, Fundamentals Cosmic Phys., 5, 287

Tinsley B. M., 1981, MNRAS, 194, 63

Tinsley B. M., Gunn J. E., 1976, ApJ, 203, 52

Toft S., Mainieri V., Rosati P., Lidman C., Demarco R., Nonino M., Stanford S. A., 2004, A&A, 422, 29

Toft S., Gallazzi A., Zirm A., Wold M., Zibetti S., Grillo C., Man A., 2012, ApJ, 754, 3

Tojeiro R., Heavens A. F., Jimenez R., Panter B., 2007, MNRAS, 381, 1252

Tojeiro R., Percival W. J., Heavens A. F., Jimenez R., 2011, MNRAS, 413, 434

Tomczak A. R., et al., 2014, ApJ, 783, 85

Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, MNRAS, 438, 1985

Trager S. C., Somerville R. S., 2009, MNRAS, 395, 608

Trager S. C., Faber S. M., Worthey G., González J. J., 2000a, AJ, 119, 1645

Trager S. C., Faber S. M., Worthey G., González J. J., 2000b, AJ, 120, 165

Trager S. C., Faber S. M., Worthey G., González J. J., 2000c, AJ, 120, 165

Tremmel M., Wright A. C., Brooks A. M., Munshi F., Nagai D., Quinn T. R., 2020, MNRAS, 497, 2786

Tremonti C. A., et al., 2004, ApJ, 613, 898

Treu T., Stiavelli M., Casertano S., Moller P., Bertin G., 1999, MNRAS, 308, 1037

Treu T., Stiavelli M., Bertin G., Casertano S., Møller P., 2001, MNRAS, 326, 237

Treu T., Ellis R. S., Liao T. X., van Dokkum P. G., 2005a, ApJ, 622, L5

Treu T., et al., 2005b, ApJ, 633, 174

Trujillo-Gomez S., Reina-Campos M., Kruijssen J. M. D., 2019, MNRAS, 488, 3972

Tully R. B., Fisher J. R., 1977, A&A, 54, 661

Tumlinson J., et al., 2013, ApJ, 777, 59

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

Valentino F., et al., 2020, ApJ, 889, 93

Van Nest J. D., Munshi F., Wright A. C., Tremmel M., Brooks A. M., Nagai D., Quinn T., 2022, ApJ, 926, 92

Vazdekis A., 1999, ApJ, 513, 224

Vazdekis A., et al., 2015a, MNRAS, 449, 1177

Vazdekis A., et al., 2015b, MNRAS, 449, 1177

Villaume A., Foreman-Mackey D., Romanowsky A. J., Brodie J., Strader J., 2020, ApJ, 900, 95

Villaume A., et al., 2022, ApJ, 924, 32

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Vulcani B., Poggianti B. M., Finn R. A., Rudnick G., Desai V., Bamford S., 2010, ApJ, 710, L1

Vulcani B., et al., 2011, MNRAS, 412, 246

Vulcani B., et al., 2013, A&A, 550, A58

Wadadekar Y., 2005, PASP, 117, 79

Wake D. A., van Dokkum P. G., Franx M., 2012, ApJ, 751, L44

Walcher J., Groves B., Budavári T., Dale D., 2011, Ap&SS, 331, 1

Wasserman A., et al., 2019, ApJ, 885, 155

Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230

Webb K., et al., 2020, MNRAS, 498, 5317

Webb K. A., et al., 2022, MNRAS, 516, 3318

Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2

Weinmann S. M., Pasquali A., Oppenheimer B. D., Finlator K., Mendel J. T., Crain R. A., Macciò A. V., 2012, MNRAS, 426, 2797

Weisz D. R., et al., 2011, ApJ, 739, 5

Werner M. W., et al., 2004, ApJS, 154, 1

Werner S. V., Hatch N. A., Muzzin A., van der Burg R. F. J., Balogh M. L., Rudnick G., Wilson G., 2022, MNRAS, 510, 674

Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232

Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336

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

Whitaker K. E., et al., 2013, ApJ, 770, L39

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Whitler L., Endsley R., Stark D. P., Topping M., Chen Z., Charlot S., 2023a, MNRAS, 519, 157

Whitler L., Stark D. P., Endsley R., Leja J., Charlot S., Chevallard J., 2023b, MNRAS, 519, 5859

Wild V., Kauffmann G., Heckman T., Charlot S., Lemson G., Brinchmann J., Reichard T., Pasquali A., 2007, MNRAS, 381, 543

Wild V., Walcher C. J., Johansson P. H., Tresse L., Charlot S., Pollo A., Le Fèvre O., de Ravel L., 2009, MNRAS, 395, 144

Wild V., et al., 2020, MNRAS, 494, 529

Wilkinson D. M., Maraston C., Goddard D., Thomas D., Parikh T., 2017, MNRAS, 472, 4297

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

Williams R. J., Quadri R. F., Franx M., van Dokkum P., Toft S., Kriek M., Labbé I., 2010, ApJ, 713, 738

Wilman D. J., Zibetti S., Budavári T., 2010, MNRAS

Wilson G., et al., 2009, ApJ, 698, 1943

Wolfgang A., Lopez E., 2015, ApJ, 806, 183

Woo J., et al., 2013, MNRAS, 428, 3306

Woodrum C., Jørgensen I., Fisher R. S., Oberhelman L., Demarco R., Contreras T., Bieker J., 2017, ApJ, 847, 20

Worthey G., 1994, ApJS, 95, 107

Worthey G., Faber S. M., Gonzalez J. J., 1992, ApJ, 398, 69

Wright A. C., Tremmel M., Brooks A. M., Munshi F., Nagai D., Sharma R. S., Quinn T. R., 2021, MNRAS, 502, 5370

Wuyts S., et al., 2007, ApJ, 655, 51

Wuyts S., Franx M., Cox T. J., Hernquist L., Hopkins P. F., Robertson B. E., van Dokkum P. G., 2009, ApJ, 696, 348

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

Xie L., De Lucia G., Hirschmann M., Fontanot F., 2020, MNRAS, 498, 4327

Yagi M., Koda J., Komiyama Y., Yamanoi H., 2016, ApJS, 225, 11

Yan R., Newman J. A., Faber S. M., Konidaris N., Koo D., Davis M., 2006, ApJ, 648, 281

Yang G., et al., 2020, MNRAS, 491, 740

York D. G., Adelman J., Anderson Jr. J. E., et al., 2000, AJ, 120, 1579

Yozin C., Bekki K., 2015, MNRAS, 452, 937

Zabludoff A. I., Mulchaey J. S., 1998, ApJ, 496, 39

Zhang H.-X., Hunter D. A., Elmegreen B. G., Gao Y., Schruba A., 2012, AJ, 143, 47

Zhang H., Zaritsky D., Behroozi P., Werk J., 2019, ApJ, 880, 28

Ziegler B. L., Saglia R. P., Bender R., Belloni P., Greggio L., Seitz S., 1999, A&A, 346, 13

da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595

di Serego Alighieri S., Lanzoni B., Jørgensen I., 2006a, ApJ, 647, L99

di Serego Alighieri S., Lanzoni B., Jørgensen I., 2006b, ApJ, 652, L145

van Dokkum P. G., Brammer G., 2010, ApJ, 718, L73

van Dokkum P. G., Franx M., 1996, MNRAS, 281, 985

van Dokkum P. G., Stanford S. A., 2003, ApJ, 585, 78

van Dokkum P. G., van der Marel R. P., 2007, ApJ, 655, 30

van Dokkum P. G., Franx M., Kelson D. D., Illingworth G. D., 1998, ApJ, 504, L17

van Dokkum P. G., Franx M., Kelson D. D., Illingworth G. D., 2001, ApJ, 553, L39

van Dokkum P. G., et al., 2010, ApJ, 709, 1018

van Dokkum P. G., Abraham R., Merritt A., Zhang J., Geha M., Conroy C., 2015, ApJ, 798, L45

van Dokkum P., et al., 2016, ApJ, 828, L6

van Dokkum P., et al., 2017, ApJ, 844, L11

van Dokkum P., et al., 2019, ApJ, 880, 91

van de Sande J., et al., 2013, ApJ, 771, 85

van de Schoot R., et al., 2021, Nature Reviews Methods Primers, 1, 1

van der Burg R. F. J., et al., 2013, A&A, 557, A15

van der Burg R. F. J., et al., 2020, A&A, 638, A112

van der Hulst J. M., Terlouw J. P., Begeman K. G., Zwitser W., Roelfsema P. R., 1992, in Worrall D. M., Biemesderfer C., Barnes J., eds, Astronomical Society of the Pacific Conference Series Vol. 25, Astronomical Data Analysis Software and Systems I. p. 131

van der Wel A., Franx M., van Dokkum P. G., Rix H.-W., 2004, ApJ, 601, L5

van der Wel A., et al., 2016, The Messenger, 164, 36

APPENDICES

Appendix A

Appendices from Chapter 3

A.1 Quiescent indicators

In this work, we selected quiescent galaxies by their position in rest-frame UVJ colour space. However, there are several other tracers of SFR that could have been used instead. The $D_n(4000)$ has been used as a proxy for the age of a stellar population (Balogh et al., 1999; Kauffmann et al., 2003; Muzzin et al., 2012) as the strength of the break increases with the fraction of old stars (but also with metallicity). The flux of the [O II] emission line is sensitive to recent excitations in the ISM from young stars – although indirectly, and is also dependent on the metallicity of the gas. Galaxies selected by each tracer as quiescent are shown in the UVJ plane in Figure A.1. The first columns show the 2D histograms of the GOGREEN spectroscopic sample in UVJ colour space, with galaxies in clusters shown in the first row and galaxies in the field in the second row. The separation of quiescent and star forming galaxies is shown as a black line.

The positions of galaxies in UVJ colour space are then shown for galaxies which satisfy alternative indicators of passive evolution: $D_n(4000)>1.4$ in the middle column, and EW([O II])$+\sigma_{\rm EW} < 5$ Å in the right-hand column. Among the cluster galaxies, the highest density of galaxies selected by $D_n(4000)$ or [O II] is predominantly in the UVJ-quiescent region. A much larger fraction of the 'quenched' galaxies in the field are UVJ-star forming.

The distribution of UVJ-quiescent (red) or UVJ-star forming (blue) according to $D_n(4000)$ (EW([O II])) is shown in the inset histogram in the second (third) columns. The selection of quiescent galaxies from $D_n(4000)$ or EW([O II]) is determined by the break in the UVJ-quiescent and UVJ-star forming distributions, corresponding to $D_n(4000)\sim1.4$

and EW([O II])+$\sigma_{\rm EW}$ < 5 Å. Both $D_n(4000)$ and [O II] emission select the majority of the UVJ-selected quiescent sample – 79% and 63%, respectively. While only 32% (24%) are UVJ-quiescent among the $D_n(4000)$-quiescent ([O II]-quiescent), the contamination of UVJ-star forming galaxies is only 13% (3%). This brief comparison shows that these tracers are broadly consistent, and using $D_n(4000)$ or EW([O II]) instead to select quiescent galaxies would not qualitatively change our conclusions.

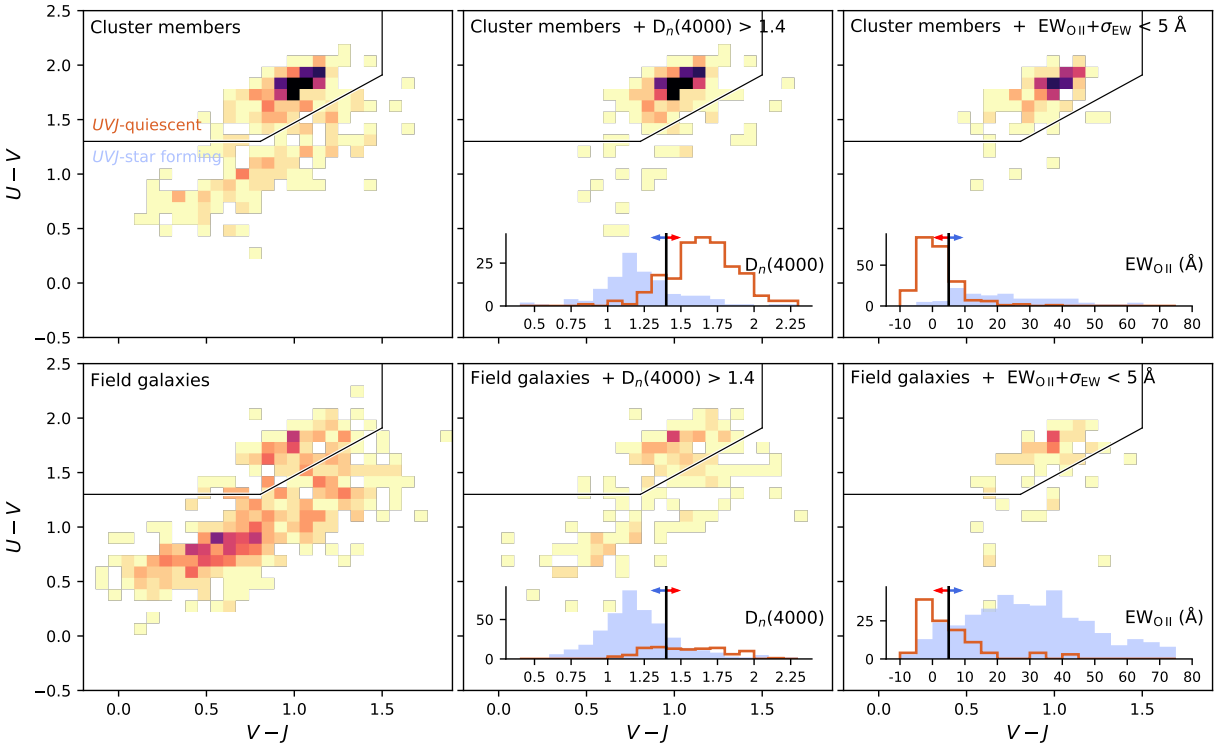

Figure A.1: 2D histograms of the spectroscopic sample in UVJ colour space. The selection criteria of UVJ-quiescent galaxies is shown as a black line with arrows indicating the region of quiescent (red) or star forming (blue) galaxies. In the top (bottom) row, the cluster (field) galaxies are shown. The middle column compares the UVJ selection against $D_n(4000)$>1.4, where the $D_n(4000)$ threshold was chosen based on the bimodality of the UVJ selection relative to $D_n(4000)$ shown in the inset histogram. The right-hand column compares the UVJ selection against EW([O II])+$\sigma_{\rm EW}$ $<$ 5 Å, where the threshold was chosen based on the bimodality of the UVJ selection relative to EW([O II]) shown in the inset histogram. UVJ colours are broadly consistent with both $D_n(4000)$ and EW([O II]) tracers for quiescent galaxies.

A.2 Mass-metallicity relation

Stellar mass, dust, and metallicity are correlated throughout a galaxy's evolution, and the relation between the two has been well studied in the local universe (Gallazzi et al., 2005, 2014; Tremonti et al., 2004; Panter et al., 2008; Choi et al., 2014). Observables used to estimate the ages of stellar populations, such as colours and spectral lines, can be strongly degenerate with dust and metallicity. Understanding such degeneracies at $z>1$ is challenging, especially given that most studies are limited to small numbers of massive galaxies (Onodera et al., 2012, 2015; Newman et al., 2014; Kriek et al., 2016; Lee-Brown et al., 2017; Morishita et al., 2018, 2019; Estrada-Carpenter et al., 2019). Moreover, without high-resolution spectroscopy, it is difficult to accurately model the complex behaviour of these parameters. Given the limited wavelength coverage in our spectra, and typically low SNR, we do not tightly constrain metallicity in our fits – however, it is important to consider the average metallicity we fit, as a function of mass and environment, because of its degeneracy with age. For instance, we find that a difference in metallicity of a factor of three ($\sim$0.5 dex) can change the mass-weighted age estimate by $\sim$0.5 Gyr.

The MIST isochrones cover an extended range of metallicities (-4<[Z/H]<0.5), while the MILES templates are limited to [Z/H]<0.19. We also impose an additional limit of [Z/H]>-2 to avoid extrapolating the templates to less well constrained parameter space. Although updated isochrones libraries include variation of α-abundances, the current version of FSPS includes only scaled-solar abundances. Studies of high SNR spectra of passive galaxies show that [α/Fe]scales with galaxy properties (*e.g.*, velocity dispersion, stellar mass), and a number of old massive galaxies with super-solar α-abundances have been discovered (Thomas et al., 2005; Choi et al., 2014; Conroy et al., 2013; Onodera et al., 2015; Kriek et al., 2016, 2019; Jørgensen et al., 2017, 2018). Underestimating α-abundance affects the slope of the UV-NIR continuum, where Vazdekis et al. (2015b) show differences of 10% in optical colours, or 40% in flux within a bandpass, between solar [α/Fe]and +0.4 albeit for galaxies much older than included in our study.

We explored the sensitivity of the metallicity measurements in our fits through the stellar mass–metallicity relation (MZR) and relative to the diffuse dust optical depth. Figure A.2 shows the posteriors of metallicity and dust (left) and stellar masses (right) for the galaxies in our sample, with circles showing the medians of individual posteriors. The GOGREEN measurements are shown relative to the local relation for quiescent SDSS galaxies from Gallazzi et al. (2005), marked as cyan lines, corresponding to the 16$^{\text{th}}$ and 84$^{\text{th}}$ percentiles of the reported trend. Note that this relation was used as a prior in our fitting procedure. We also include the $z\sim$0.7 MZR for quiescent galaxies from Gallazzi et al. (2014) as a blue region. The MZR for quiescent galaxies reported by Choi et al. (2014) at

$0.4<z<0.55$ is shown as pink points, and $0.55<z<0.7$ as black points with pink error bars. Lastly, we show the 1 σ region of individual measurements of $1.6<z<2.5$ massive galaxies from Morishita et al. (2019) as green boxes. The Gallazzi et al. (2005), Gallazzi et al. (2014), and Choi et al. (2014) data are shown corrected for differences in stellar mass estimates (*i.e.*, +0.2 dex, see Appendix A.3), but not corrected for differences in definitions of solar metallicity or α-abundance. Choi et al. (2014) incorporated α-abundance corrections in their continuum-normalized spectral fitting. Morishita et al. (2019) used a higher limit on metallicity, as they use the updated MIST isochrones which extend to [Z/H]<0.5.

We note that these studies all use different methodologies: Gallazzi et al. (2005) and Gallazzi et al. (2014) relied on line-indices, Choi et al. (2014) use full-spectrum SPS modelling for continuum-corrected co-added spectra, while Morishita et al. (2019) use full-spectrum SPS modelling of spectroscopy and photometry, more similar to our own procedure. Although not shown in Figure A.2, Leethochawalit et al. (2018) study the MZR with respect to [Fe/H] for quiescent galaxies at $z\sim0.4$ using spectral modelling, and recover values consistent with the highest density (purple) region in our plot (see their figure 7). Interestingly, Kriek et al. (2019) measure the metallicity of three massive quiescent galaxies at $z\sim1.4$, using high-resolution spectroscopy to measure absorption lines, and find that the [Fe/H] values are $\sim$0.2 dex lower than the $z<0.7$ relation. Similarly, Jørgensen et al. (2007) find evidence of evolution of cluster galaxies since $z\sim1$. On the other hand, Onodera et al. (2015) find the [Z/H] of 24 massive quiescent galaxies at $z\sim1.6$ to be well in line with the local relation, based on a similar line index analysis.

While our metallicities are lower the reported by similar studies, as long as the mass-metallicity relation does not have a strong environmental dependence, the *relative* comparison of cluster and field galaxy ages will not be sensitive to our model metallicities. Indeed, we find no difference in the MZR between field and cluster galaxies from our fits. Peng et al. (2015) compared the stellar metallicities of galaxies in SDSS, and found no significant difference between satellite and central galaxies above 10^{10} $M_{\odot}$. Tangentially, Maier et al. (2016) measured enhanced *gas-phase* metallicities of accreted star-forming cluster galaxies relative to comparable field galaxies at $z\sim0.4$ for $<10^{10.5}$ $M_{\odot}$, but no significant difference at higher masses.

As mentioned above, there is a degeneracy between age, metallicity, and dust. For completeness, we show the combined posteriors of metallicity and the diffuse dust optical depth in the left plot of Figure A.2. The majority of galaxies have very little dust, $\tau_{\lambda,2}<0.5$, even the galaxies with very low metallicities. This perhaps suggests that the dust model we have assumed (*i.e.*, Milky Way extinction curve; Cardelli et al. 1989) is insufficient.

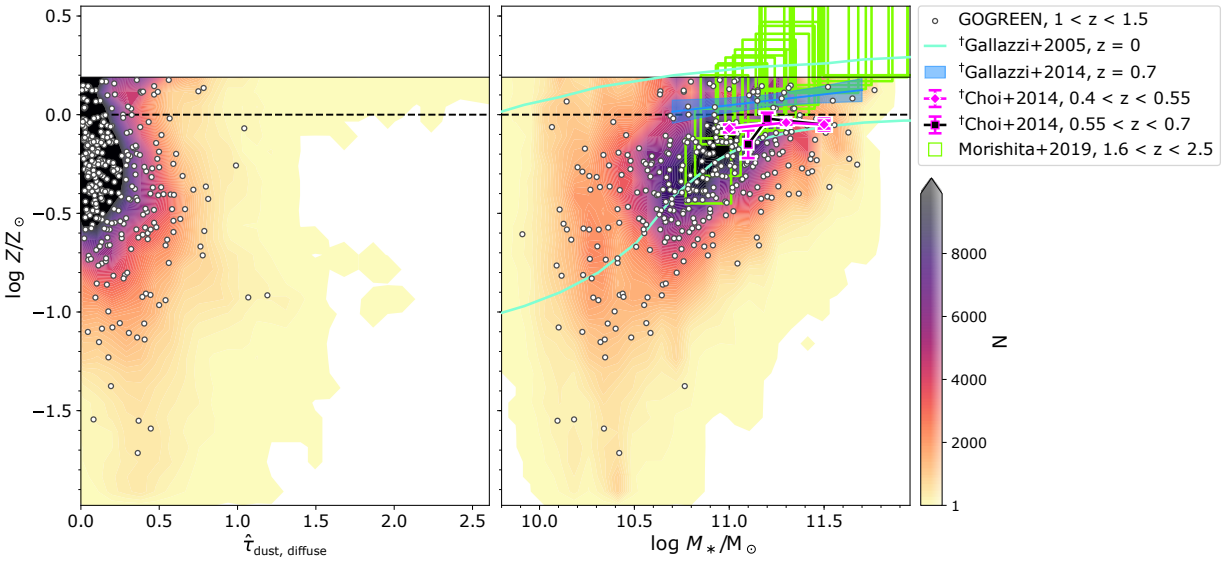

Figure A.2: Metallicity as a function of diffuse dust optical depth (left) and stellar mass (right) for the GOGREEN UVJ-quiescent sample. A dashed line indicates solar metallicity, and a solid line indicates the maximum metallicity allowed by the MILES spectral templates. The local mass-metallicity relation (MZR) for early type galaxies from Gallazzi et al. (2004) at $z\lesssim0.22$ is shown with two cyan lines indicating the lower and upper limits of the reported 68% CR. This relation was used as a prior in our SFH fitting procedure. The MZR for $z\sim0.7$ quiescent galaxies is shown as a blue line with a shaded region indicating the uncertainty region, from Gallazzi et al. (2014). A selection of moderate-redshift quiescent galaxies at $0.4<z<0.55$ and $0.55<z<0.7$ from full continuum-normalized spectral fits from Choi et al. (2014) are shown, without correction for differences in α-abundance. A high-redshift sample of massive quiescent galaxies from Morishita et al. (2019) are also included, shown in green. Daggers denote where data have been adapted from the relevant study to compensate for difference in stellar mass estimates. The colour scale shows the density of the combined posteriors in the GOGREEN data, with white circles indicating the median values of the individual posteriors.

A.3 Prospector nonparametric vs FAST parametric models

We confirm the systematic offset between derived stellar masses using a parametric SFH model (FAST, Kriek et al. 2009) versus a nonparametric SFH model with a continuity prior (as assumed in this work, using Prospector) reported by other studies (*e.g.*, Leja et al., 2019b). Our comparison is shown in Figure A.3, where nonparametric SFH masses are on average 1.6× (*i.e.*,0.2 dex) higher. Stellar masses were derived with FAST for the same SFH as was used to measure the rest-frame colours with EZGAL (see Section 3.2.3 – a declining exponential SFR). The SXDF galaxies are marked as yellow diamonds in Figure A.3 as their fiducial masses were not derived from FAST but from similar template fitting with Bruzual & Charlot (2003) models, as described in Mehta et al. (2018). The stellar masses used in Old et al. (2020) and van der Burg et al. (2020), as well as in the upcoming data release (Balogh et al., 2020), are based on FAST masses, and therefore will differ from the stellar masses in this paper.

Differences between the FAST and Prospectorfits include the assumed SPS libraries, Bruzual & Charlot (2003) or FSPS, respectively. Leja et al. (2019a) explored the significance of this difference when fitting galaxies from 3D-HST and demonstrated that it accounted for $\lesssim 0.05$ dex of the offset in stellar masses. The more important difference is the assumed SFH model and prior. Figure A.4 compares the discrepancy between stellar masses with that for stellar metallicity (where FAST assumes a fixed solar metallicity), V-band dust optical depth, and mass-weighted ages. The clear correlation between the difference in masses and mass-weighted ages, where because mass-weighted age is a summary statistic of the SFH, this suggests the SFH model is the primary driver of the mass offset. We find that the offset is largest for the young galaxies, according to FAST. This is consistent with the expectation that a declining exponential model is a poor description of young, star forming galaxies. A further exploration of the SFH-model dependence of the mass and ages of galaxies is left to future work (see Chapter 5).

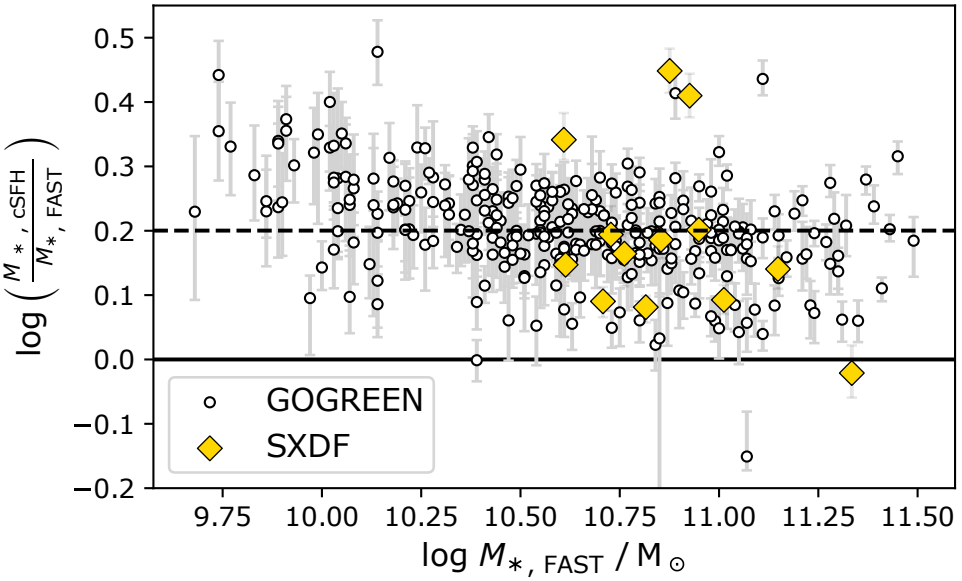

Figure A.3: Comparison of FAST (parametric SFH) and Prospector (nonparametric SFH with a continuity prior; cSFH) derived stellar masses. Nonparametric SFHs yields larger masses, by ∼0.2 dex (shown as a dashed line), with a mild mass dependence. Yellow diamonds indicate SXDF galaxies which have parametric stellar masses from Mehta et al. (2018).

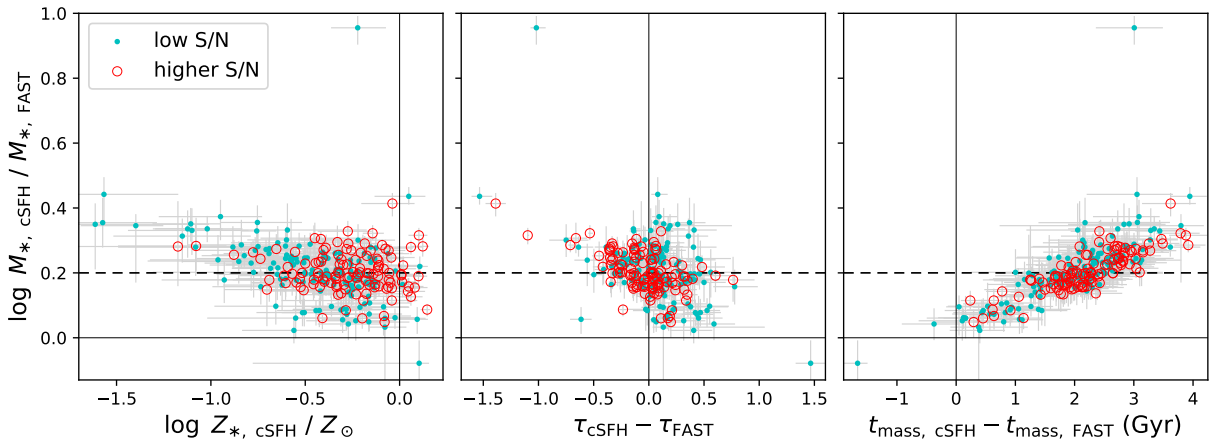

Figure A.4: Comparison of FAST (parametric SFH) and Prospector (nonparametric SFH with a continuity prior; cSFH) derived stellar masses, stellar metallicities, V-band dust optical depth, and mass-weighted ages. Solar metallicity is assumed in FAST, such that the comparison is simply the stellar metallicity from Prospector. Dust extinctions from FAST are converted to opacity. The parametric SFH parameters fit with FAST are used to derive the mass-weighted ages. Points are coloured according to the S/N of the spectroscopic (> 1 Å^{-1}) and photometric (> 10) data. The galaxies with most discrepant values typically have lower S/N data. The discrepancy between stellar mass values is most correlated with that for the ages.

A.4 Average spectral characteristics

In this paper, we have measured galaxy properties on individual galaxies, and then considered the statistics of those measurements. A common alternative in the literature is to combine the *data* to create an average spectrum/SED, and measure physical parameters from that. As the parameters are non-linearly related to SED shape, these two approaches do not necessarily give the same result.

Figure A.5 shows co-added spectra of cluster galaxies and field galaxies in our sample, each separated into three stellar mass subsamples. Before stacking, the spectra were redshift corrected, binned to a common wavelength sampling, and flux normalized at 4120 Å. Spectra within a given stellar mass and environment subsample were then averaged and bootstrap sampled to determine the uncertainty. Combined galaxies within clusters are shown in orange, and within the field in blue, where the number of contributing galaxies to each spectrum is labelled on the left.

The average cluster and field spectra appear very similar overall, with only a few apparent differences. The field population has more prominent [O II] emission at lower masses, while the cluster galaxies have stronger [O II] emission at higher masses (although much weaker than in the field). This is likely related to the fact that [O II] is not strictly related to recent star formation (*e.g.*, from AGN and/or LINER; Heckman 1980, Yan et al. 2006, Singh et al. 2013). On the other hand, absorption lines from Hδ appear stronger for cluster galaxies (except at the lowest stellar masses) suggesting that the cluster galaxies experienced, on average, more recent star formation.

$D_n(4000)$ is commonly used as an age indicator (*e.g.*, Balogh et al., 1999; Kauffmann et al., 2003; Muzzin et al., 2012), because it is insensitive to dust and, as a relatively wide feature, can be measured at high SNR relative to other indices. $D_n(4000)$ is not sensitive to the SFH, however; a galaxy that quenched rapidly and one that quenched slowly can have the same $D_n(4000)$, depending on the relative timing of the quenching. Less apparent from the co-added spectra (and only statistically significant for the moderate mass galaxies) is that the field spectra have smaller $D_n(4000)$ than cluster galaxies. A comparison is shown in Figure A.6 *for different mass selections* than for the co-added spectra shown in Figure A.5, relative to values measured for galaxies in the GCLASS survey (Muzzin et al., 2012, averaged over radial bins – see their table 5). We note that the GCLASS sample between $1<z<1.5$ is included in our GOGREEN sample. We increase the reported GCLASS masses and mass selections by 0.2 dex to account for differences in how the stellar masses were estimated; see the discussion in Appendix A.3: $\log M_*/M_\odot \in$ [9.45, 10.15), [10.15,10.85), and [10.85, 12.15). Black error bars indicate the uncertainties of

the $D_n(4000)$ measurements from the combined spectra, while cyan error bars indicate the systematic uncertainty due to how the spectra are combined (*i.e.*, inverse-weighted averaged, or median combined).

While the GCLASS sample shows small differences in $D_n(4000)$ between environments, on average we find larger differences in the $D_n(4000)$ of the average spectrum of cluster galaxies than field galaxies for the GOGREEN sample. This is consistent with the sense of the age difference we measure from fitting the SFHs of individual galaxies. The GCLASS sample is dominated by galaxies at $z\sim0.8$, particularly at low stellar masses. That we find larger age differences than in GCLASS could hint that the age difference evolves between $z\sim0.8$ and $z\sim1.2$.

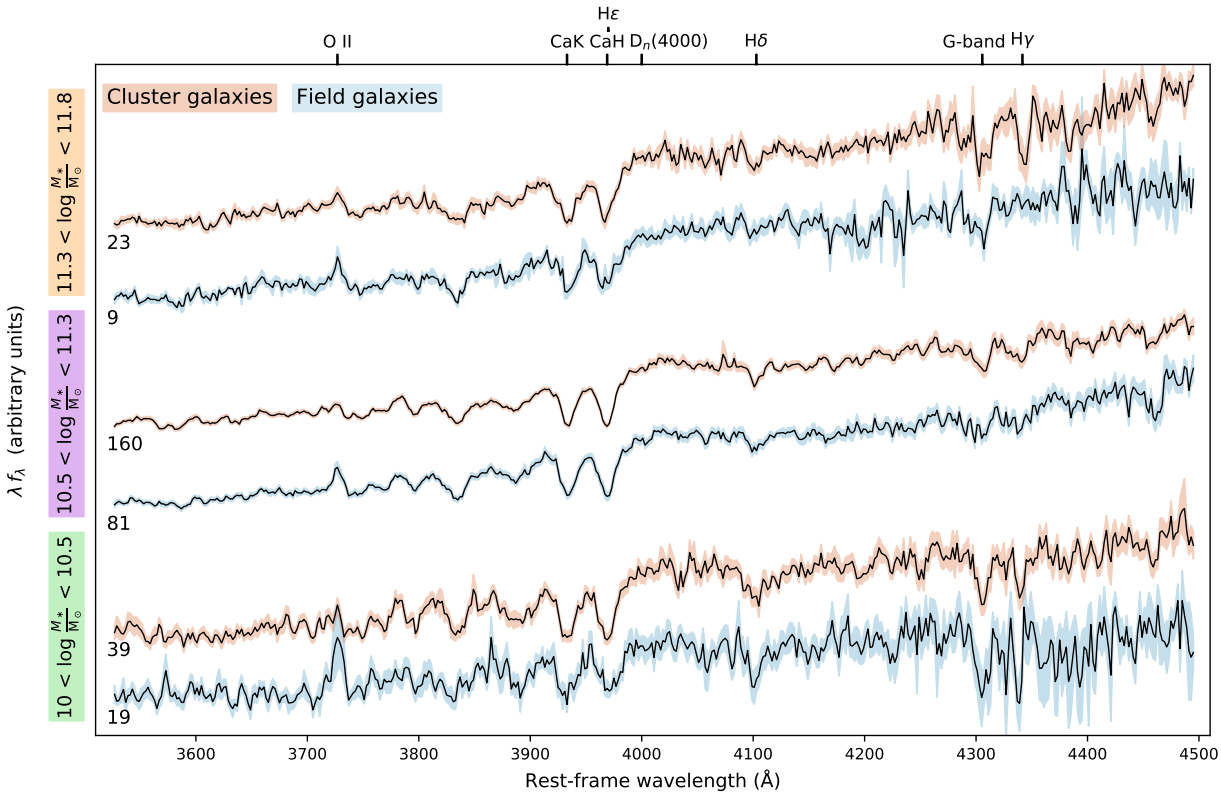

Figure A.5: Combined spectra of quiescent galaxies within mass and environment selections, shown within the wavelength region included in the SFH fitting procedure. The spectra in each subsample were de-redshifted, re-binned to a common wavelength sampling, flux normalized about 4120 Å, and then averaged. The uncertainty in the co-added spectra was determined from bootstrapping. Prominent spectral features are labelled on the top axis, and number of galaxies in each co-add are indicated on the left.

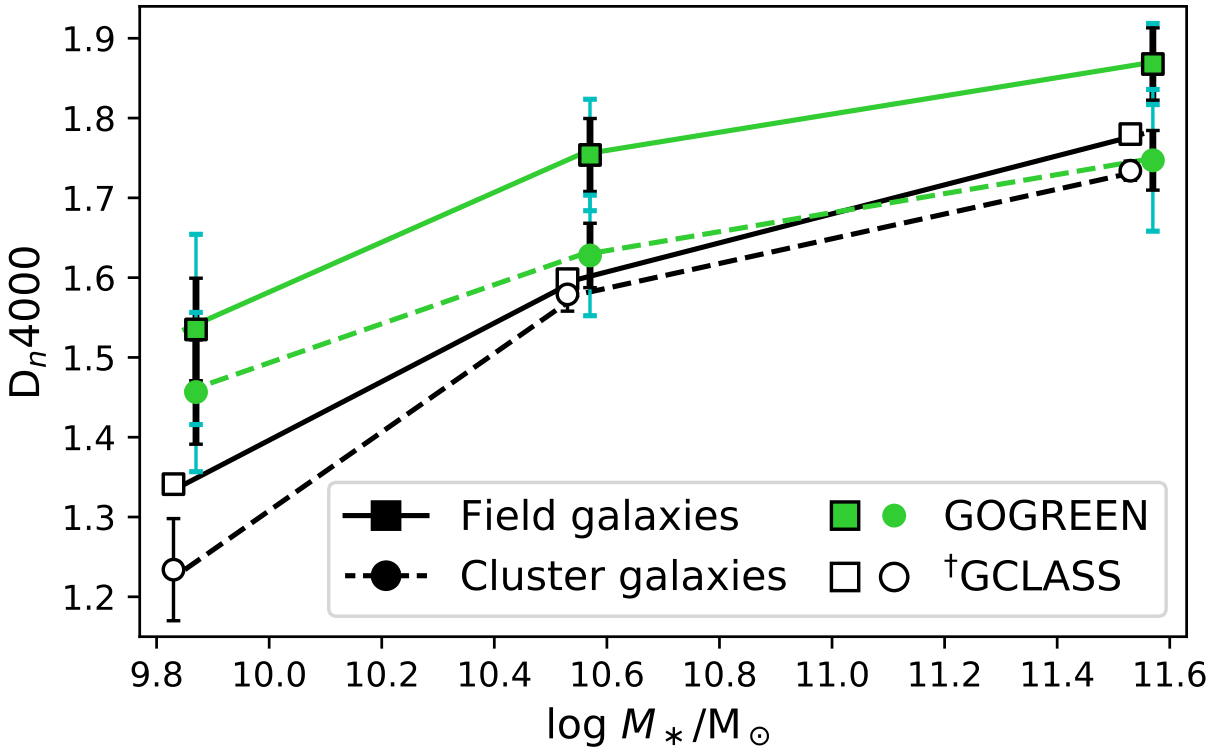

Figure A.6: $D_n(4000)$ of *averaged* spectra as a function of stellar mass, relative to equivalent results from GCLASS Muzzin et al. (2012) – *not* the same mass binning as used throughout the paper or the co-added spectra shown in Figure A.5. Masses selected within bins of log $M_*/M_\odot \in$[9.45,10.15), [10.15,10.85), and [10.85,12.15) where a 0.2 dex offset was applied to the selection of Muzzin et al. (2012) based on the difference in mass measurement techniques (see Appendix A.3). Points are shown slightly offset for clarity. Cluster galaxy values are marked with circles, field galaxy values with squares. Green colours mark measurements with GOGREEN, with black error bars corresponding the uncertainty in averaged $D_n(4000)$ values, and cyan error bars showing the systematic error between methods of combining the values. Black outlined points show the measurements from Muzzin et al. (2012) (taken from their table 5, averaged over radial bins).

A.5 Age as a function of *UVJ* colour

Mass-weighted ages, $t_{\rm mass}$, are shown in UVJ colour space in Figure A.7. The sample is divided into five regions in UVJ-colour space, delineated by dotted lines, and the median age (and 68% CRs) are labelled for each. As expected, there is a positive trend between $t_{\rm mass}$ and rest-frame $U-V$ and $V-J$ colours, where the oldest galaxies are clustered towards the upper right of the quiescent region. We find good consistency between our UVJ-ages trend and trends in the literature (*e.g.*, Belli et al., 2019; Estrada-Carpenter et al., 2019; Ferreras et al., 2019), despite systematic or procedural differences between studies, for example: SFR parameterization, SED-fitting procedures, how the ages were measured (luminosity weighted, mass-weighted, median, etc.), and the mass or redshift range of the samples. The overall age gradient in UVJ-colour space is flatter than predicted by Belli et al. (2019), which could be attributed to the aforementioned systematics. However, Carnall et al. (2019a) report their sample of $1<z<1.3$ quiescent galaxies to have $t_{\rm mass}$ in good agreement with the Belli et al. (2019) relationship despite having similar methodological differences. Although the systematics related to our fitting procedure are important when comparing to the literature, they are less important for the purposes of this study – the differential comparison of cluster and field populations. Our age estimates are discussed further in Section 3.4.2.

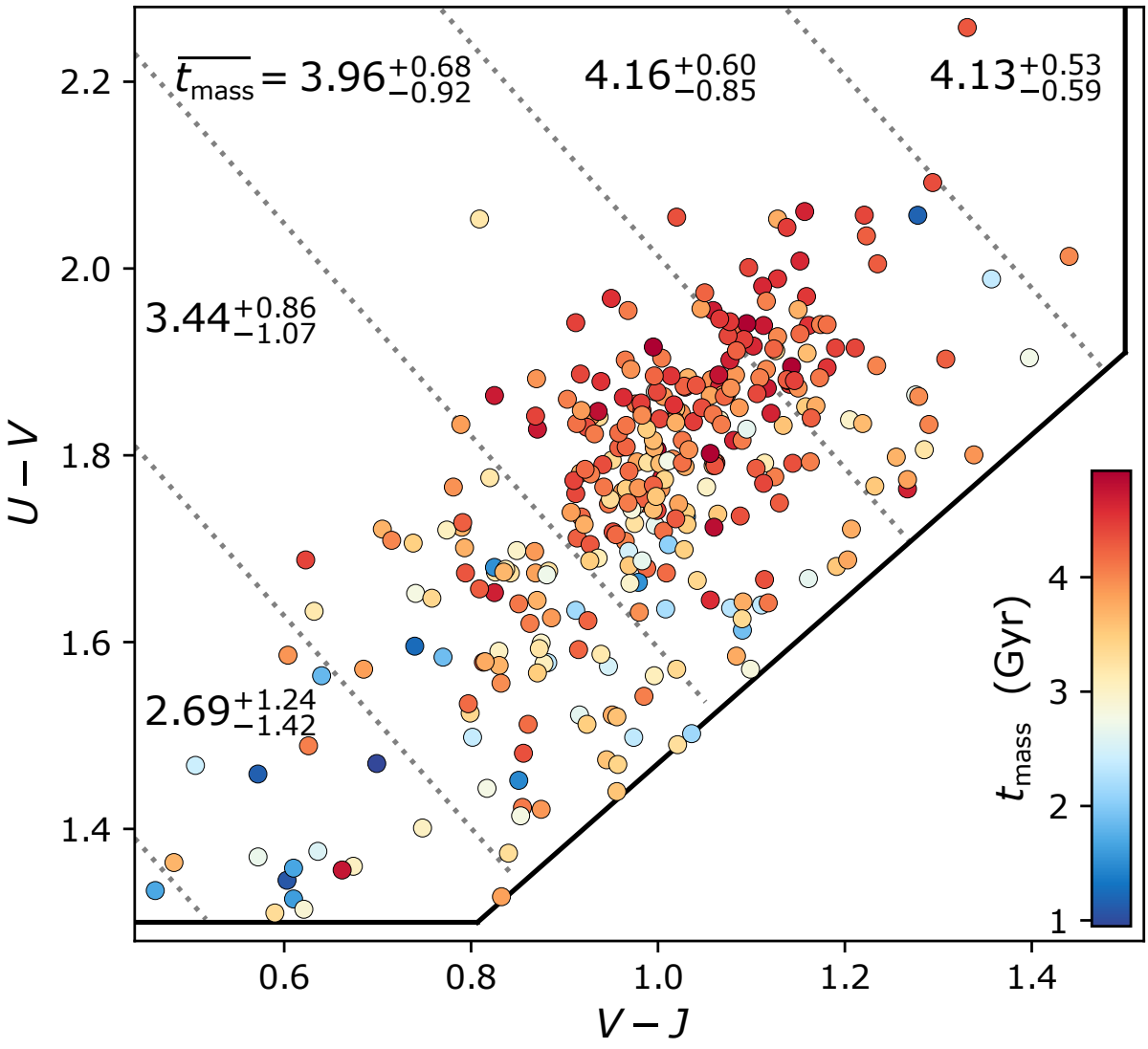

Figure A.7: Mass-weighted ages in rest-frame UVJ colour space. The sample is divided into five regions, where the median $t_{\rm mass}$ and 68% CRs for the galaxies in each bin are labelled. The majority of galaxies in the 'red clump' are the oldest galaxies in our sample, but otherwise there is not a smooth distribution of $t_{\rm mass}$ relative to UVJ colours.

A.6 Luminosity weighted ages

The luminosity-weighted age is more sensitive to recent star formation, as younger stars dominate the integrated luminosity. For passively evolving galaxies, which formed all their stars a long time ago, the mass-weighted age and luminosity-weighted ages should be equivalent. We calculate the luminosity-weighted age from the SFH posteriors,

$$t_{\rm light} = \frac{\int_{t_{\rm univ}(z_{\rm obs})}^{0} t\ {\rm SFR}(t)\ L(t)\ {\rm d}t}{\int_{t_{\rm univ}(z_{\rm obs})}^{0} {\rm SFR}(t)\ L(t)\ \ {\rm d}t} \tag{A.1}$$

where L is the g-band luminosity.

Figure A.8 shows the distribution of the stellar mass and luminosity-weighted ages, in units of cosmic time (similar to Fig 3.6 for mass-weighted ages). Contours show the combined posteriors of the field (blue) and cluster (red) galaxies, where white points indicate the medians of the individual posteriors. Diamonds mark galaxies which have formed more than 10% of their stellar mass within the last 1 Gyr, $f_{M_*<1\ {\rm Gyr}}>0.1$, discussed in Sec 3.4.3. Compared to the mass-weighted ages, the luminosity-weighted ages are younger on average, but not uniformly younger. As a result the age distributions are broadened.

Following the same mass-matched cumulative age comparison as for $t_{\rm mass}$, we find that cluster galaxies are on average $0.39^{+0.58}_{-0.40}$ Gyr older than field galaxies, a $\sim$0.1 Gyr larger difference. Figure 3.8 compares this age comparison to that with mass-weighted ages.

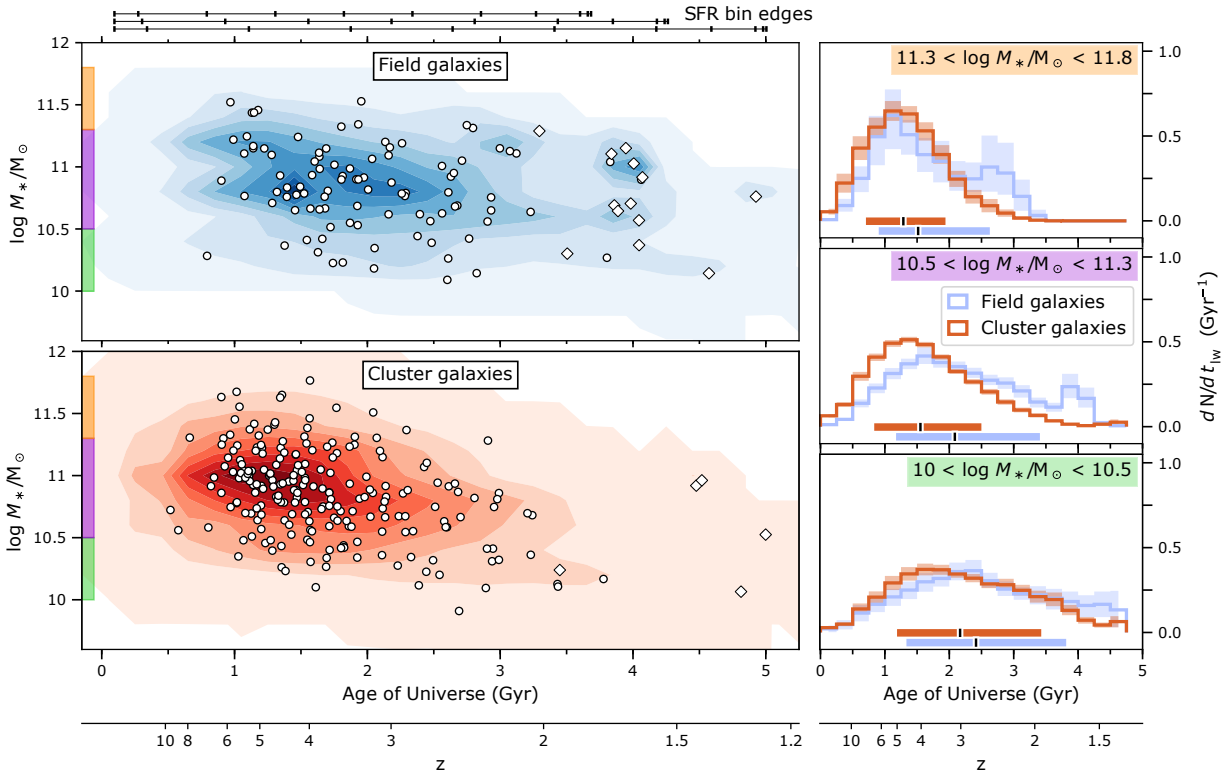

Figure A.8: Comparison of stellar masses and luminosity-weighted ages between field (blue) and cluster (red) galaxies. Left: Combined posteriors of stellar masses and $t_{\rm light}$ (in units of cosmic time), shown as contours. The medians of the individual posteriors are marked with white circles/diamonds. Diamonds indicate $f_{M_*<1\ \rm Gyr}$>0.1 galaxies (formed more than 10% of their stellar mass within the last 1 Gyr). Horizontal bars at the top of the figure indicate the edges of the age bins for $z = 1.5$ (top), $z = 1.25$ (middle), and $z = 1$ (bottom). The bins were defined in units of lookback time, and therefore do not match up for galaxies observed at different redshifts. Right: Combined $t_{\rm light}$ posteriors for field and cluster galaxies, shown in three mass bins. The medians (black mark) and 68% CRs (coloured bar) of each distribution are marked at the bottom of each subplot. The shaded regions show the bootstrapped uncertainty of each histogram. Although there are field galaxies that formed as early as the oldest cluster galaxies, and cluster galaxies that formed as late as the youngest field galaxies, *on average* field galaxies formed at later times.

Appendix B

Appendices from Chapter 4

B.1 The SFH of DF44

For comparison with future works, in Table B.1 we provide the fraction of SF, and cumulative fraction of stellar mass formed, within the time bins of the nonparametric models. We list the 16^{th}, 50^{th}, and 84^{th} percentiles of the distributions, where we note that the 50^{th} percentiles of the fractional SFHs do not necessarily sum to unity.

Table B.1: Summary of SFH results. The fraction of SF and the cumulative fraction of stellar mass formed are listed for each time bin of the nonparametric SFH model. The 16^{th}, 50^{th}, and 84^{th} percentiles of the posterior (*i.e.*, the 68% CR) are listed. We note that the 50^{th} percentiles of the fractional SFH do not necessarily sum to unity. The SF time-scales listed in Table 4.3 are interpolated from these step functions.

Time bin (Gyr)	Extended SFH prior						concentrated SFH prior					
	SF Fraction			Cumulative fraction of M_*			SF Fraction			Cumulative fraction of M_*		
	16	50	84	16	50	84	16	50	84	16	50	84
10^{-9} – 0.03	0.0703	0.0795	0.0967	1.0000	1.0000	1.0000	0.0042	0.0082	0.0099	1.0000	1.0000	1.0000
0.03 – 0.10	0.0050	0.0152	0.0240	0.9986	0.9988	0.9989	0.0160	0.0181	0.0221	0.9995	0.9996	0.9998
0.10 – 0.50	0.0007	0.0014	0.0061	0.9980	0.9983	0.9986	0.0000	0.0000	0.0004	0.9974	0.9977	0.9980
0.50 – 1.00	0.0014	0.0031	0.0049	0.9972	0.9979	0.9983	0.0000	0.0001	0.0002	0.9974	0.9975	0.9979
1.00 – 2.00	0.0003	0.0015	0.0039	0.9964	0.9970	0.9975	0.0000	0.0000	0.0001	0.9973	0.9974	0.9978
2.00 – 3.00	0.0014	0.0046	0.0086	0.9949	0.9962	0.9970	0.0000	0.0000	0.0001	0.9971	0.9974	0.9978
3.00 – 4.01	0.0011	0.0089	0.0158	0.9917	0.9932	0.9955	0.0000	0.0000	0.0005	0.9969	0.9973	0.9977
4.01 – 5.36	0.0091	0.0145	0.0435	0.9833	0.9886	0.9925	0.0000	0.0001	0.0004	0.9964	0.9971	0.9973
5.36 – 7.16	0.0166	0.0702	0.1351	0.9591	0.9790	0.9847	0.0000	0.0002	0.0016	0.9943	0.9969	0.9971
7.16 – 9.57	0.0722	0.1548	0.3310	0.8543	0.8993	0.9606	0.0000	0.0000	0.0022	0.9917	0.9950	0.9969
9.57 – 12.80	0.3036	0.3726	0.5208	0.5596	0.6872	0.8085	0.0000	0.0000	0.0008	0.9851	0.9938	0.9958
12.80 – 13.47	0.0583	0.1918	0.3111	0.0179	0.0618	0.1116	0.9678	0.9714	0.9737	0.9822	0.9916	0.9947

B.2 Systematic biases in measuring SFHs

B.2.1 SFH biased by blue horizontal branch stars

The use of integrated light to reconstruct stellar populations has the caveat that multiple types of stars can share spectral signatures. This is the case for young, massive main-sequence stars and old, metal poor stars on the blue side of the horizontal branch (HB); both act to amplify the equivalent width of the Balmer lines. A population of blue HB stars produces a flux shortward of 3000 Å which increases with decreasing metallicity due to a hotter main-sequence turn-off. Neglecting to include a blue HB population in models can lead to predictions of unrealistically young ages (*e.g.*, Worthey, 1994; Schiavon et al., 2004; Thomas et al., 2005; Schiavon, 2007). The difficulty of distinguishing between these two stellar populations has been noted in GCs and dwarf galaxies (*e.g.*, Monaco et al., 2003; Schiavon et al., 2004; Conroy et al., 2018; Cabrera-Ziri & Conroy, 2022), as well as in elliptical galaxies (Maraston & Thomas, 2000).

In the SFH fit to DF44 (for which the primary age indicator is the Hβ absorption line) we see a rise in the SFR in the two most recent time bins corresponding to the last 100 Myrs (by $1.8^{+0.2}_{-0.4}$ dex for the extended SFH, $2.5^{+1.5}_{-1.1}$ dex for the concentrated SFH) – yet there are no corresponding emission lines to suggest the presence of a young stellar population. Given the low metallicity of this UDG ($\log(Z_*/Z_\odot) \sim -1.2$), the presence of

a blue HB population would not be unexpected.

While the bias between the age and the blue HB stars is well known when fitting simple stellar populations (SSPs; Conroy et al., 2018), it is not yet well studied for nonparametric SFHs. Ocvirk (2010) provided a first look at the impact of blue HB stars on linear combinations of SSP models, finding that the presence of blue HB stars can be inferred as a recent burst of star formation at ∼100 Myr, contributing less than around 10% of the total stellar mass. While this provides a promising explanation for the apparent star formation bursts we observe in the SFH of DF44, we follow a similar test using the nonparametric SFH described in Section 4.3.1.

In order to investigate if our SFHs are affected by the presence of blue HB stars which mimic a burst of SF within the last 100 Myrs, we fit the SFHs of two Galactic GCs, one with a known blue HB and the other without. We select the GCs from Schiavon et al. (2005), with metallicities similar to DF44: NGC 2808 has [Fe/H] $= -1.29$ and $(B-R)/(B+V+R) = -0.49$ (bluer HB), and NGC 6218 (M12) has [Fe/H] $= -1.32$ and $(B-R)/(B+V+R) = 0.97$ (redder HB). Given that GCs are reasonable approximations of SSPs, we expect an early single burst of star formation only. Fits to the spectra[1] of NGC 2808 and NGC 6218, over the same wavelength range as DF44, following the procedure described in Section 4.3, are shown in Fig. B.1. The top panels summarize the comparison between the observations (black) lines, and models (coloured lines). Similar to Fig. 4.4, the bottom panels show the sSFR, SFR, and mass assembly histories. An extended SFH was assumed, and the total stellar mass was fixed to 10^8 $M_\odot$.

In both cases we find an increase in the SFR within the last 100 Myr, although to a larger extent for the GC with the blue HB stars (by $1.5^{+0.5}_{-0.3}$ dex for NGC 6218, and by $2.6^{+0.4}_{-0.4}$ dex for NGC 2808). In addition, we see that both SFHs are early and short-lived, although there are modest levels of star formation at > 2 Gyr, which likely results from the models being unable to precisely match the high S/N spectra (∼180 and ∼480, respectively). We conclude from this comparison that some component of the recent SF burst we measure for DF44 could plausibly be related to a population of blue HB stars.

[1]Downloaded from http://www.noao.edu/ggclib.

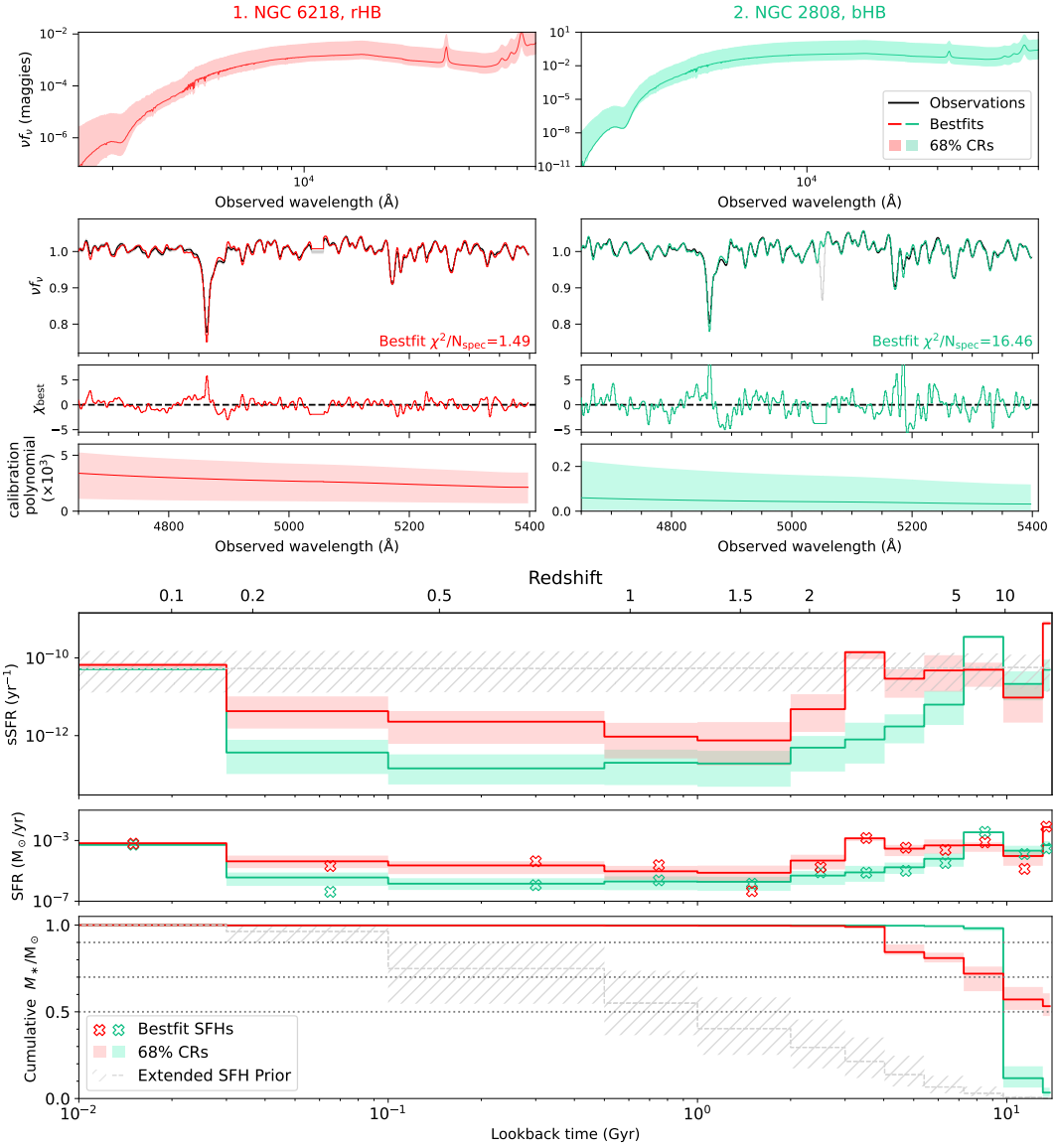

Figure B.1: Summary of the fitting results for two Milky Way globular clusters selected to have a similar metallicity as DF44: NGC 6362 (redder horizontal branch) and NGC 2808 (bluer horizontal branch). *Top:* Fits to observations, similar to Fig. 4.3 for DF44. *Bottom:* Posterior distributions for star formation and mass growth, similar to Fig. 4.4 for DF44. Both clusters are fitted with a fixed mass $\log(M_*/M_\odot) = 8$.

B.2.2 Fitting the spectroscopy and photometry together vs separately

Figs. B.2 and B.3 show the results of fitting the models to observations of DF44, where we include the following input: i) using only the photometry (yellow), ii) using only the spectroscopy (green), and iii) using both the photometry and spectroscopy (red), and assuming an extended SFH prior. We note that the stellar mass is fixed (to the value reported by van Dokkum et al., 2016) for the spectrum-only fit as the continuum was subtracted from the spectrum.

Similar to Fig. 4.3 discussed in Section 4.4, in Fig. B.2 the observations (black lines and markers) are shown relative to the bestfit models (coloured lines and markers, where the colours denote which observations were fit). Shaded coloured regions indicate the 68% CRs from sampling the posteriors, where the grey shaded region indicates the uncertainties in the spectrum.

Both bestfit SED models match the photometry with reasonable $\chi_{\rm bestfit}$. In comparison, the UV flux is significantly overestimated when fitting only the spectroscopy. Since the UV provides information about recent star formation, and the UV to optical colours constrain the dust attenuation, we do not expect to constrain these properties from the spectrum alone.

A comparison of the observed spectrum with the bestfit models is also shown in Fig. B.2, with the $\chi_{\rm bestfit}$ as a function of wavelength, and the spectrophotometric calibration polynomial (see Section 4.3.3). The ratio of the two bestfit models, shown flattened by dividing through by a polynomial, shows that the fits are similar at the 2% level. The only notable differences between the two bestfit models are around the Hβ line and Mg II features at $\sim$5285 Å – 5305 Å (observed-frame). The positive ratio of the Hβ line between the spectrum-and-photometry fit over the spectrum-only fit is consistent with the UV flux being constrained for the former, such that the absorption line is preferentially shallower. The difference in the Mg II lines reflects the difference in metallicities predicted for each fit, as well as the inability of the (fixed scaled-solar abundance) models to be flexible to such features.

Fig. B.3 compares the basic stellar properties (normalization of the diffuse dust attenuation curve, V-band extinction, stellar metallicity, stellar mass, and mass-weighted age) for the fits to the three sets of observations. This figure is akin to Fig. 4.5, discussed in Section 4.4. For comparison, black lines indicate values measured in the literature: dashed lines indicate the stellar isochrone metallicity measured by Villaume et al. (2022), while dotted and solid lines indicate the stellar mass measured by van Dokkum et al. (2016) and

Saifollahi et al. (2021), respectively. For reference, the prior on the age (which is implicit, as age is determined by the time bin widths and SFH) is shown as a black histogram.

The broadband NUV to NIR photometry (yellow) and continuum normalized spectroscopy (green) carry different information about the galaxy properties. The broad (yet coarse) photometry provides a tighter constraint on the dust attenuation, while the spectroscopy constrains the metallicity. The dust attenuation cannot be determined from the spectroscopy alone because of the lack of continuum information; the spectrophotometric calibration marginalizes over the continuum shape, and is degenerate with both the stellar mass and dust attenuation. On the other hand, the metallicity is tightly constrained by the spectroscopy, as there is detailed information among the numerous absorption lines.

Despite the formal consistency of the dust and metallicity parameters between these two fits (given the large uncertainties), the age posteriors are significantly different. The age posterior from the photometry largely traces the (implicit) prior. A tighter posterior for the stellar metallicity provides a more precise estimate of the age, as expected given the degeneracy between these two parameters.

Simultaneously fitting the photometry and spectroscopy (shown in red) constrains the full set of parameters. In the particular case of DF44, the results are largely informed by the spectroscopy, which covers a broad range of metallicity and age features – the inclusion of the photometry only modestly affects the posteriors. The stellar mass derived from the combined data sets is consistent with that of Saifollahi et al. (2022), while the photometry-derived posterior is skewed lower by $\sim$0.23 dex, which is likely also related to the lower estimate for the stellar metallicity. The combined result shows DF44 to be very old, metal poor, and perhaps with some small amount of dust.

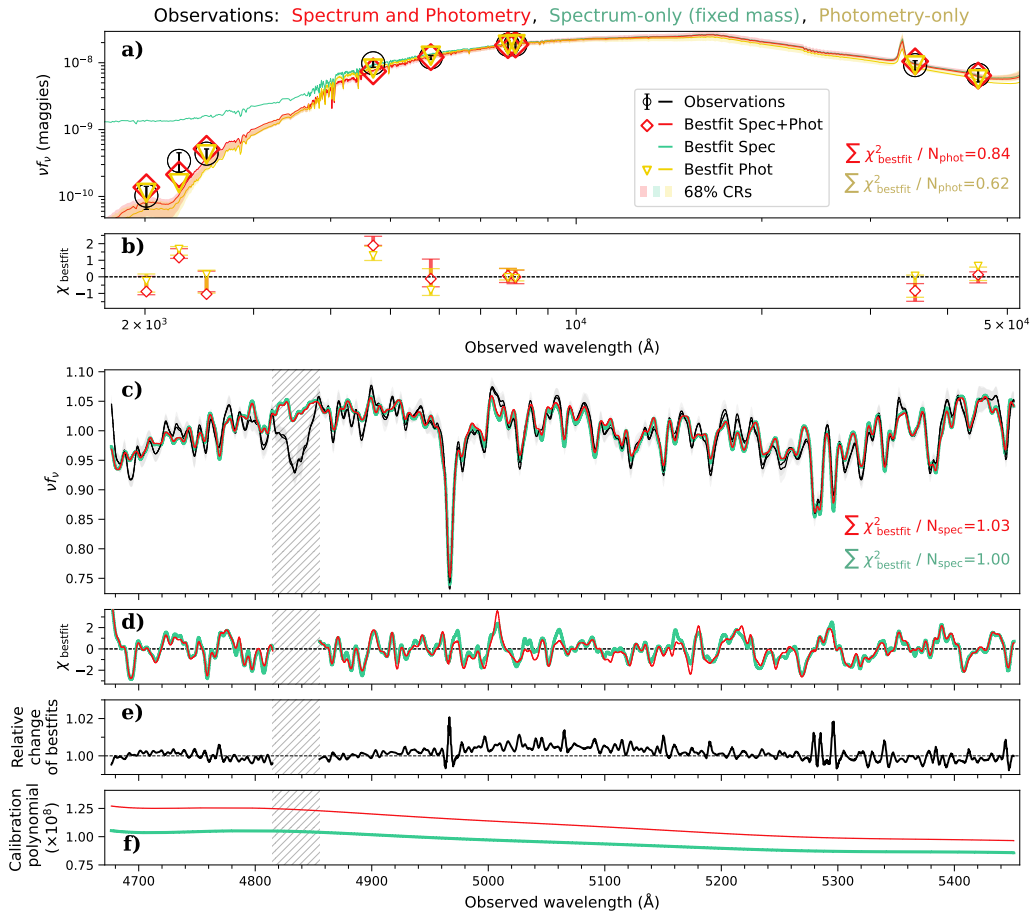

Figure B.2: Comparison of fits with the spectrum and photometry (red), spectrum only (green, with mass fixed to the value from van Dokkum et al., 2016), and photometry only (yellow), assuming an extended SFH prior. The observed data (black) is compared to the bestfit models (coloured lines) and the 68% CR of 500 randomly drawn models from the posteriors (shaded coloured regions). The corresponding posteriors are shown in Fig. B.3. *(a)* The observed (circles) and bestfit (diamonds and triangles) photometric points, where the reduced $\chi^2/N_{\rm data}$ of the bestfit SED are listed. *(b)* The χ ([data − model]/σ) of the bestfit photometric points. *(c)* The observed spectrum (uncertainties shown in grey) and bestfit spectra (multiplied by the spectrophotometric calibration polynomial). The hatched grey region indicates the spectral region masked throughout the fitting process. *(d)* The χ of the bestfit spectra as a function of wavelength. *(e)* The relative change of the bestfit models, *i.e.*, the ratio of the two bestfit spectra. *(f)* The spectrophotometric calibration models, with 68% CRs shown as shaded regions.

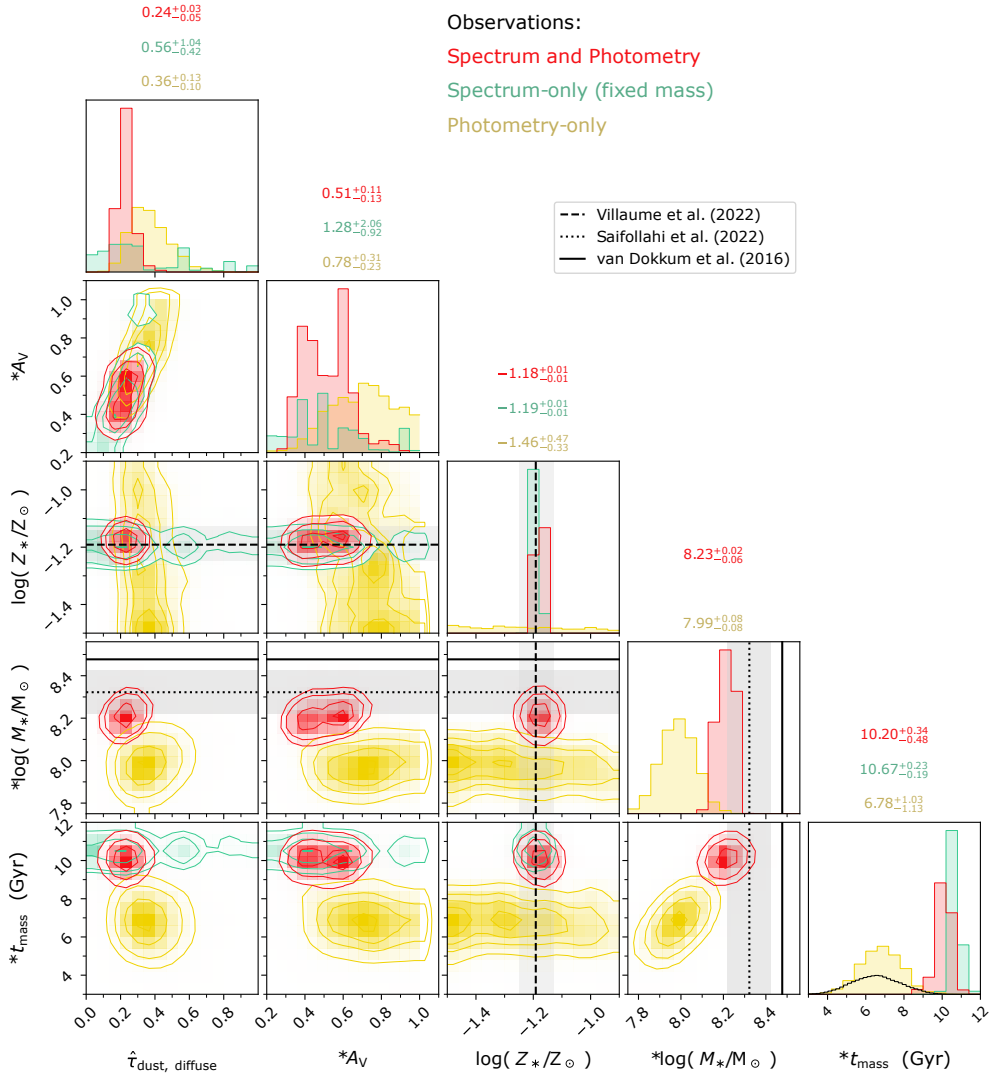

Figure B.3: Comparison of posteriors derived with the photometry-only (yellow), spectrum-only (green, with mass fixed to the value from van Dokkum et al., 2016), and both spectrum and photometry (red), assuming an extended SFH prior. Fits are shown in Fig. B.2. Posteriors of selected fitted and derived (marked with asterisks) parameters are shown. Contours are shown smoothed with a $n = 1$ Gaussian kernel. Black lines denote the expected results from the literature: stellar metallicity from Villaume et al. (2022), and stellar mass from van Dokkum et al. (2016) and Saifollahi et al. (2022). The median and uncertainties from the 68% CR are listed along the top of the one-dimensional histograms. The implicit age prior is shown as a black histogram for reference.

B.2.3 SFH biased by choice of prior

Fig. B.4 demonstrates the S/N dependence of the bias imposed by the choice of SFH prior, which in this case is an extended SFH in describing a very old stellar population. We refit the KCWI spectrum of DF44 with the extended SFH prior ($\alpha_{\rm D} = 1$), successively increasing the uncertainties of the spectrum such that the $\rm S/N_{spec} = 5$, 10, 15, and 20. The medians of the recovered posteriors are shown for the mass-weighted age (in lookback time), t_{50} and t_{90} (in time since the Big Bang), $\log(Z_*/\rm Z_\odot)$, and diffuse dust, with error bars corresponding to the 68% (thick and wide) and 95% CRs (thin and narrow). Points mark the results from fitting the spectrum and photometry simultaneously (diamonds), the spectrum alone (squares, offset vertically for clarity), and the photometry alone (circles). The prior distributions are shown in the top panels. Note that because the implicit priors for the SFH time-scales depend on the widths of the SFH time bins (a step function), the distributions are not necessarily smooth.

The SFH time-scales are more heavily weighted by the SFH prior at low S/N. This is particularly true for t_{90}, which we use as a proxy of the quenching time. In contrast, neither the stellar metallicity nor the dust is significantly biased, or at least the offsets are well within the (large) uncertainties. While having a complete set of observations informs many of the galaxy properties, the choice of a 'good' SFH prior is important.

B.2.4 Comparing results between studies – prior and data dependence

Fig. B.5 shows a comparison of the star formation time-scales of UDGs (circles) and dwarfs (squares and diamonds) for observations from the literature (for Coma galaxies in almost all cases). We compare the time at which we consider the galaxy quenched, t_{90}, with how extended the SFH is, $t_{50} - t_{90}$. The grey shaded region denotes the parameter space where ages (t_{50}) are older than the Universe (*e.g.*, OGS1 from Ruiz-Lara et al. 2018). We show the results from the literature as upper limits given the possible biases in SFH time-scales discussed above related to the S/N, and choice of SFH priors.

Except for DF44, all the literature values were measured using the full-spectrum fitting code STECKMAP. Notably, STECKMAP smooths the SFHs via (tuneable) regularization akin to Gaussian priors on the SFH and age–metallicity relations (see the discussion in Section 4.3.1). The details of the regularization differ between all studies, where, for example, Ruiz-Lara et al. (2015) present the outcome of averaging several results with various smoothing parameters. Martín-Navarro et al. (2019) show in their appendix A the

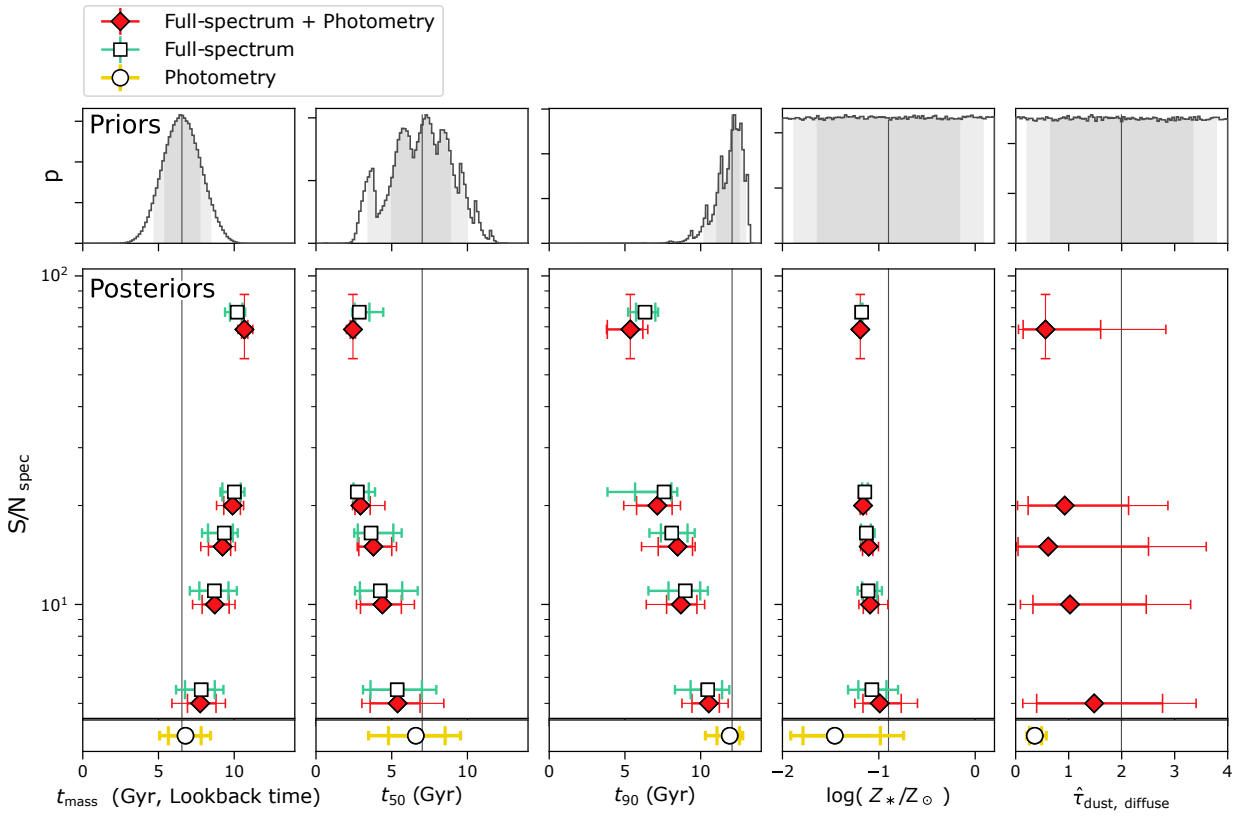

Figure B.4: Using the extended SFH prior, we compare the priors (top panels) and posterior parameter estimates (bottom panels) as a function of S/N. Values determined from full-spectrum fits with (diamonds) and without (squares; vertically offset for clarity) including the photometry are shown, as well as from photometry alone (circles). Points are shown with error bars corresponding to the 68% (thick and wide) and 95% (thin and narrow) percentiles of the posteriors. Similarly, for the prior we show the median with a grey line, and 68% and 95% CRs with shaded regions. While mass-weighted ages are shown in units of lookback time, the time-scales t_{50} and t_{90} have units of time since the Big Bang ($t = 0$ is the Big Bang).

difference in their regularized and un-regularized results to be $\sim$1 Gyr in t_{50} and $\lesssim$ 0.4 Gyr in t_{90}.

Ferré-Mateu et al. (2018) compared their SFH time-scales derived from STECKMAP with those from an alternative fitting code, STARLIGHT, which does not impose regularization but does require relative-flux calibrated spectra. Between the two fitting approaches, Ferré-Mateu et al. (2018) found consistent results in that the SFHs are extended and had similar quenching times. That said, STARLIGHT preferred starting star formation $\sim$2 Gyr later, such that the ages were younger and star forming time-scales were shorter. In contrast, the 'burstier' prior used in this work produced earlier star formation and quenching.

Because of the difficulties in determining the ages of old stellar populations, even subtle differences in data or analysis can impact results beyond the expected uncertainties. As an example, we can compare measurements for two UDGs, DF26/Yagi93 and Yagi418, both studied by Ferré-Mateu et al. (2018) and Ruiz-Lara et al. (2018); the values are connected with dashed lines in Fig. B.5. Each author used rest-frame optical spectroscopy (where Ruiz-Lara et al. 2018 reported higher S/N and had a wider wavelength coverage) and they used the same code (STECKMAP). However, the median mass-weighted ages differ by $\sim$1 Gyr (uncertainties were not reported, but the luminosity weighted ages are formally consistent). In both cases, the higher S/N data provided a solution shifted in the expected direction (*i.e.*, towards older and less-extended SFHs).

While DF44 appears to have (one of) the shortest SFHs and earliest quenching times, we caution that a detailed comparison should consider priors and the S/N. A poorly chosen SFH prior will have a stronger bias at a low S/N. For example, in using an extended SFH prior with the DF44 KCWI spectrum degraded to S/N = 20, we recover $t_{50} \sim 2.9 \pm 0.5$ Gyr and $t_{90} \sim 7.1 \pm 1.2$ Gyr (see Fig. B.4 in Appendix B.2.3), which overlaps with the lower end of UDGs in Fig. B.5. This suggests that some of these objects could be older, and have less-extended SFHs.

Along the same lines, we do not include photometry-derived results in Fig. B.5 as the comparison can be misleading given the different choices (and relative contributions) of SFH priors. In the preceding sections, we have shown that the photometry-derived ages are younger than the spectroscopy- or combined-derived ages. There is a similar difference between the results of Pandya et al. (2018, with optical to NIR photometry; not shown in Fig. 4.8) and Martín-Navarro et al. (2019, with rest-frame optical spectroscopy, S/N$\sim$ 10 Å^{-1}). Both studied the UDG DGSAT I, although using different fitting methods and assuming different SFHs. Pandya et al. (2018) fitted their photometry (via MCMC) to a delayed-exponential model, while Martín-Navarro et al. (2019) fitted their spectroscopy with STECKMAP. We note that in this example, the priors are considerably different.

For a delayed exponential model with linearly uniform priors with $\tau = 0.1$–10 Gyr and $t_0 = 0.1$–14 Gyr, the implicit prior on the mass-weighted age has a median of 3.2 Gyr. In comparison, a constant SFH has a median age of half the age of the Universe, $\sim$6.8 Gyr (see also the discussion in Johnson et al. 2021). While the luminosity-weighted ages are similar ($\sim$3 Gyr), their mass-weighted ages are discrepant by > 1 Gyr (t_0 in the delayed-exponential model is the onset of star formation, where for a $\tau > 3$ this corresponds to ages considerably younger than t_0). The metallicities are also discrepant by > 1 dex, although Martín-Navarro et al. (2019) found that DGSAT I is unusually α-enhanced. Several other studies have studied UDGs from photometry alone (*e.g.*, Greco et al., 2018; Barbosa et al., 2020), and have similarly noted younger ages than spectroscopy-derived results.

We additionally note that Martín-Navarro et al. (2019) uses a set of SSP models different from those used in both this work, and the other UDGs studies discussed here. Neither the choice of SSP models nor the application of regularization would explain the significant offset between the SFHs of DGSAT I and the other UDGs, however.

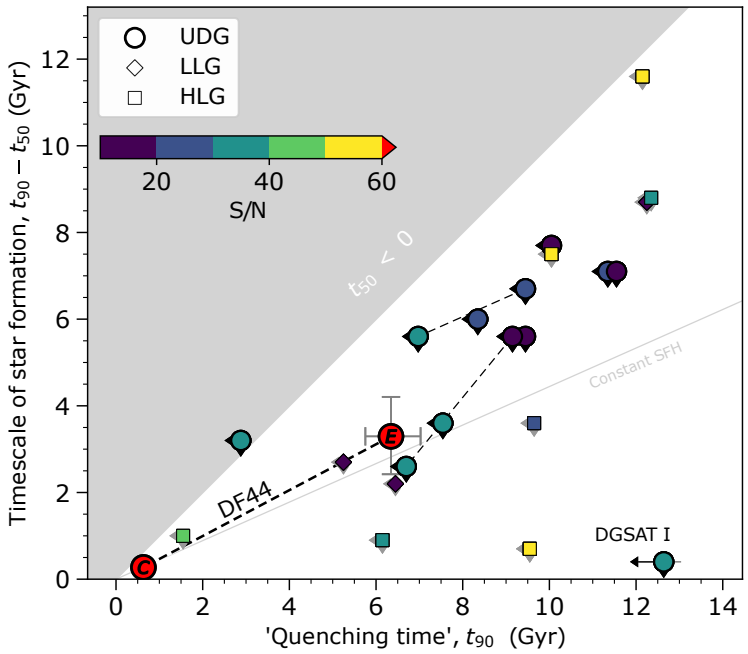

Figure B.5: Star formation time-scales of UDGs (circles) and dwarfs (low luminosity and high luminosity galaxies; squares and diamonds, respectively) for observations from the literature. We approximate the quenching time as when 90% of the stellar mass is in place (t_{90}), while the timescale $t_{90} - t_{50}$ gives a sense of the duration of star formation, *i.e.*, how concentrated/extended the SFH is. Other than DF44, we show the points from observations with arrows indicating that they are upper limits (see text). Points are coloured according to their S/N, where DF44 has a mean S/N of 96 Å^{-1}. Dashed lines connect points measured for the same object, but from different studies. Sources: Ferré-Mateu et al. (2018), Ruiz-Lara et al. (2018), and Martín-Navarro et al. (2019). The points from Ruiz-Lara et al. (2018) are shown with S/N= 32 Å^{-1}, the median of the reported range in values.

B.3 Degeneracy between dust attenuation and flux from old stellar populations in the NUV

The normalization of the dust attenuation curve ($\hat{\tau}_{\text{dust, diffuse}}$) and the fraction of old stars, both parameters of our physical model, are degenerate at optical and UV wavelengths. As a brief example of this degeneracy, Fig. B.6 shows the photometry for DF44 (black points) relative to three model SEDs with simple stellar populations (*i.e.*, not the results of fitting the physical model described in Section 4.3). Taking the grey model as the 'fiducial' model, slight variations in age and dust are shown by the purple and cyan models, respectively. While the 2.8 Gyr age increase or 0.2 dex increase in diffuse dust produces an equivalent effect in the NUV, they have opposing effects at wavelengths > 1 μm. Coloured markers show the expected photometry in two *JWST* filters in the mid-infrared, with S/N ~ 5 to reflect the average uncertainty of the IRAC data. In this example, the 'old' and 'dusty' models are slightly distinguishable in F560W ($\Delta m_{\text{AB}} \sim 0.6\sigma_m$) but very different in F770W ($\Delta m_{\text{AB}} \sim 3\sigma_m$). The inclusion of mid-infrared data to our data set would allow us to assess whether DF44 is as dusty as our results suggest or a product of the complex degeneracies between physical parameters (see Section 4.4.1).

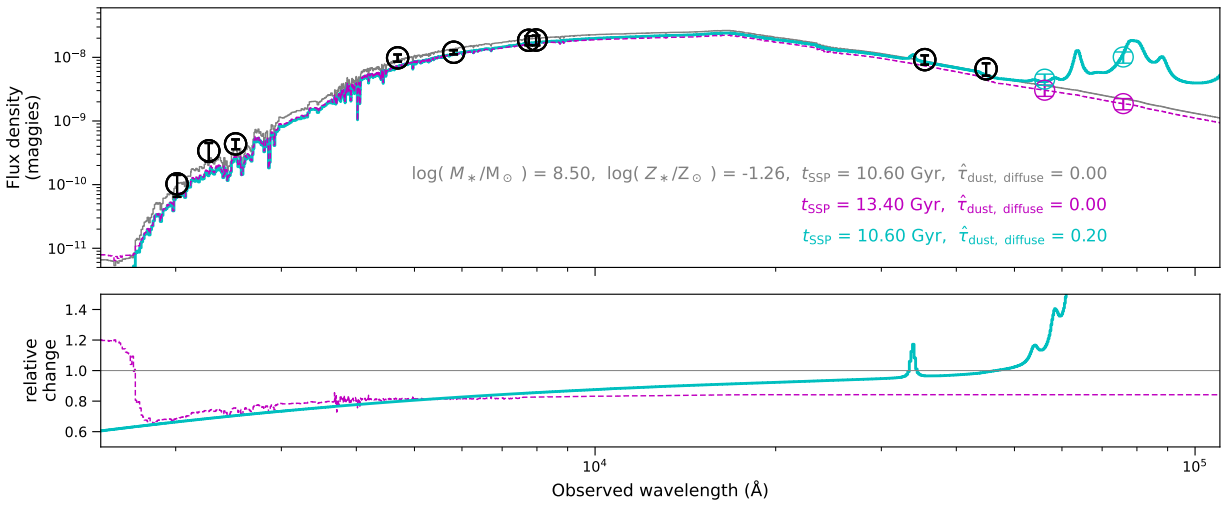

Figure B.6: A brief demonstration of the degeneracy between dust attenuation and age on the shape of SEDs. *Top:* Photometry of DF44 (black markers) and models (coloured lines) for three SSP populations. Photometric points corresponding to the 'old' (purple dashed) and 'dusty' (solid cyan) models are shown as measured by the *JWST* F560W and F770W filters (coloured markers), with S/N $\sim$ 5. *Bottom:* The relative change between the fiducial (grey) and older or dustier models. While the effect of either increasing the age or dust acts similarly at wavelengths < 1 μm, the effect acts in the opposite sense in the mid-infrared.

Appendix C

Appendices from Chapter 5

C.1 Issues with averaging observations

In this section, we argue that the nature of averaging observations to obtain higher S/N composite observations hinders, more than helps, the study of galaxy populations. While many studies in the literature ((*e.g.*, Eisenstein et al., 2003; Schiavon et al., 2006; Gallazzi et al., 2005; Onodera et al., 2015; Kriek et al., 2019; Saracco et al., 2019, 2023)) employ stacking techniques looking to relative differences between composite spectra related to differences in the properties of the galaxies contributing to each composite, the absolute values or absolute relative values obtained via composite spectra are subject to a number of complicated biases.

As an example of the non-equivalence of an 'average observation' and an 'observation of the average,' in Figure C.1 we show three SSPs are shown with ages of 2, 3, and 4 Gyr (broadly spanning the range of ages in our mock population) and otherwise identical properties. The ratio of the 'average spectrum' and the 'spectrum with the average age' is shown in the bottom panel. The ratio is also sensitive to the relative normalization of each observation, particularly if the observations cover a broad wavelength range. While the differences are within 2% for this particular example, if the uncertainty of the average spectrum is underestimated,[1] the average age obtained by SED-fitting the composite spectrum will be inaccurate. Given that there is larger evolution in the SED for younger ages, this can lead to a stronger bias for younger composite spectra. Figure C.2 compares

[1] Measuring a non-Gaussian distribution via statistics which assume Gaussian distributions will lead to biased results.

the average spectra of two sets of SEDs, each with a uniform distribution of mass-weighted ages and exponentially declining SFHs, to the spectrum with the average age of each group (panels e). Where the ages of the SFHs are between 2–4 Gyr, the average spectrum is biased towards older ages (as interpreted from the CaK+H and Balmer absorption features), while the opposite is true when the ages span 3–5 Gyr. The relative ages inferred from the average spectra will also be inaccurate. While in practice non-Gaussian uncertainties would better capture the variance across the observations, SED-fitting codes typically rely on the fast computational advantages of Gaussian statistics.

The 'variance-induced bias' effect can of course be minimized by only averaging sets of observations which are alike, where galaxies are grouped according to their colours or SED shapes(*e.g.*, Eisenstein et al., 2003). The complication of this approach, however, is having a large enough sample to populate narrow subgroupings to balance the gain in S/N in stacking the data with the information of the variance within the population. While young (1–2 Gyr) and old ($>$ 3 Gyr) galaxies are easily distinguishable, it is more challenging to isolate galaxies that are $\sim$5 Gyr from those that are older. Moreover, if galaxies at a given age have a spread in other properties (*e.g.*, metallicity or dust), this may warrant further sub-categorization.

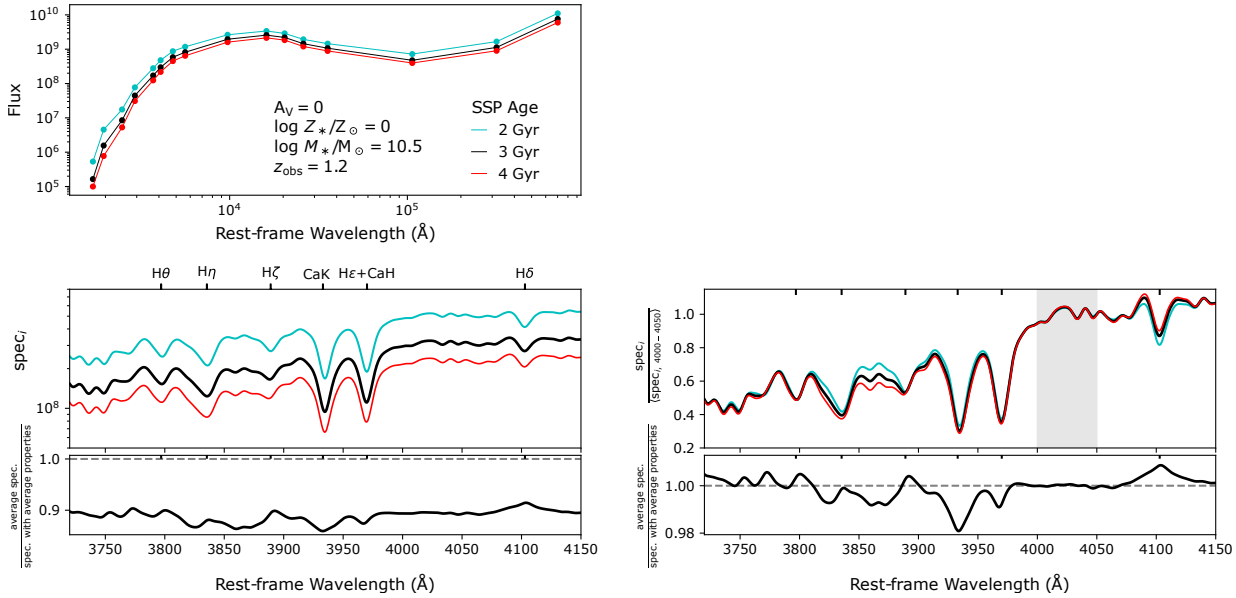

Figure C.1: Comparison of SEDs for three SSPs of varying ages, and the average SED. The average spectrum is compared to the spectrum with the average properties ($t_{\rm mass} = 3$ Gyr) where no flux normalization is applied (left), or the spectra are normalized over the average flux between 4000–4050 Å (right). The average spectrum does not correspond to the spectrum with the average properties. In this example, there is relatively minor SED evolution over the age range, the average spectrum is unphysical at the 1–2% level.

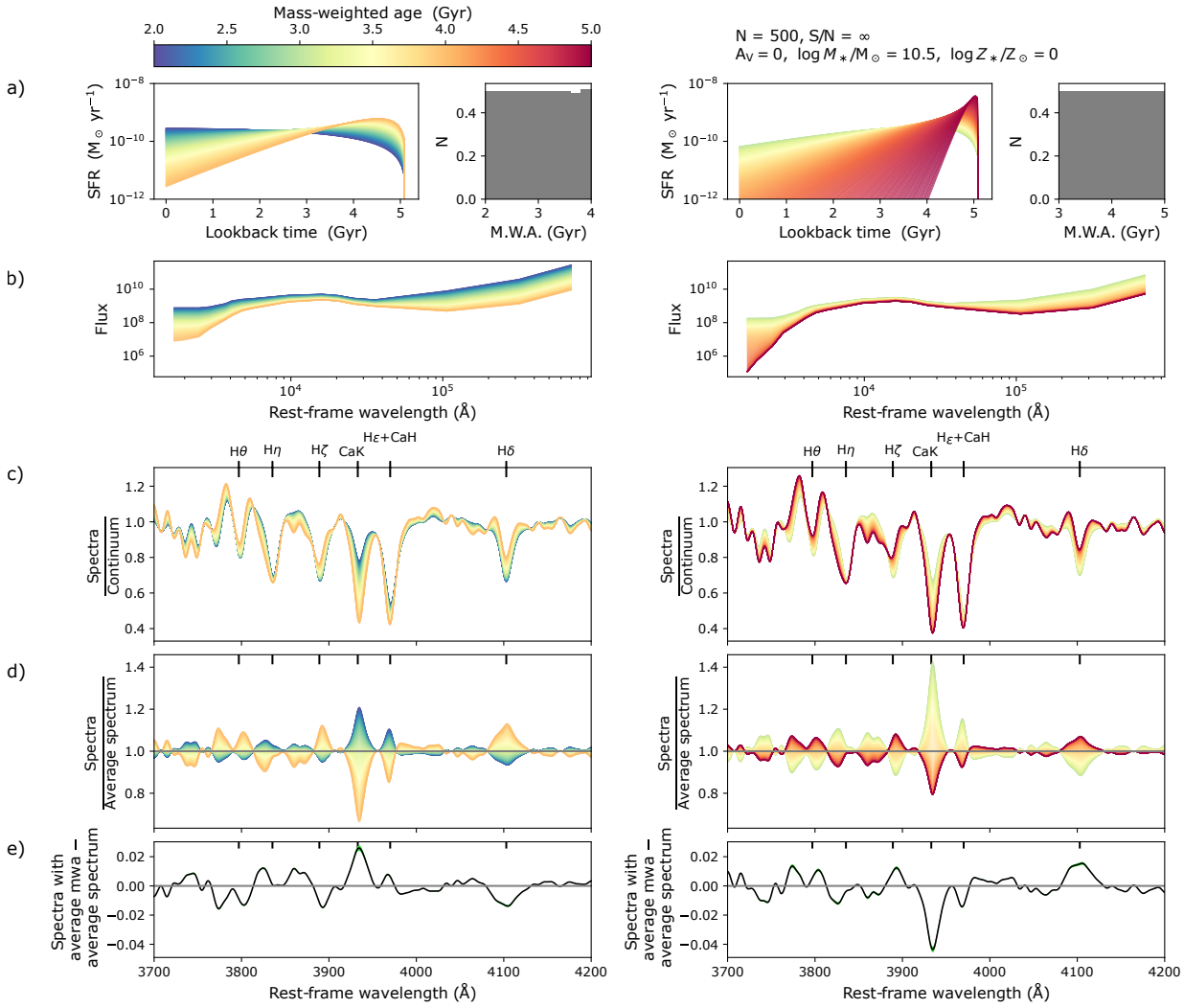

Figure C.2: SEDs for exponentially declining SFHs with varying e-folding time and mass-weighted ages, and otherwise fixed properties. a) SFHs, coloured by the mass-weighted age, and b) the corresponding SEDs. SEDs with ages of 2–4 Gyr (left; with a uniform distribution of ages, as shown in the histograms) and 3–5 Gyr (right) are averaged, where the spectra were first continuum normalized (as shown in panels c). The average spectra are compared to the individual spectra (panels d), and to the spectra with the average age (e), over the rest-frame optical wavelength range. The average spectra are biased from the spectrum with the average age of each selection differently.

C.1.1 Stacking mock galaxies of different ages

We demonstrate the failure of stacking approaches in measuring the relative ages of populations by comparing the results of i) fitting galaxies individually, and ii) fits of the averaged observations. Three sets of 100 mock galaxies were selected from the TNG300 sample of quiescent galaxies, with ages spanning a narrow range of values: 1.9, 2.8, and 3.7 Gyr. For each age selection, a composite observation was built by taking the median of the photometry (covering the NUV–NIR, each with S/N = 10), and the median of the spectroscopy (S/N = 1.3 Å^{-1}). Given that the mock galaxies are all observed at the same redshift, it was not necessary to sample the photometry or spectra to coincident filters or wavelength bins. For each observation, the photometry was normalized based on the flux about 4125 Å(rest-frame) and the spectral continuum was removed (by fitting an eighth-degree Chebyshev polynomial to the spectrum and dividing it through) before stacking. The uncertainties of the composite observations were determined from the median absolute deviation (MAD). The MAD typically reports uncertainties slightly below the standard error on the average, such that the median S/N of the spectra are ∼11 rather than $\sqrt{N} \times \text{S/N}_i \sim 13$.

Posteriors determined from fitting the individual and composite observations are shown in Figure C.3, compared to the true properties, where we fit the data with either the cSFR Continuity or Dirichlet($\alpha_\text{D} = 0.2$) SFH priors. The median and 68% CRs of each distribution are indicated by ticks and horizontal bars, respectively, at the top of each panel. The posteriors based on the composite observations recover the relative ages of the three sets of galaxies but overestimate the age difference. The stellar metallicity and diffuse dust posteriors also do not reflect the true average, nor the range of values. In comparison, the combined posteriors based on fitting the individual observations are noisy, but better trace the true properties of the mock galaxies.

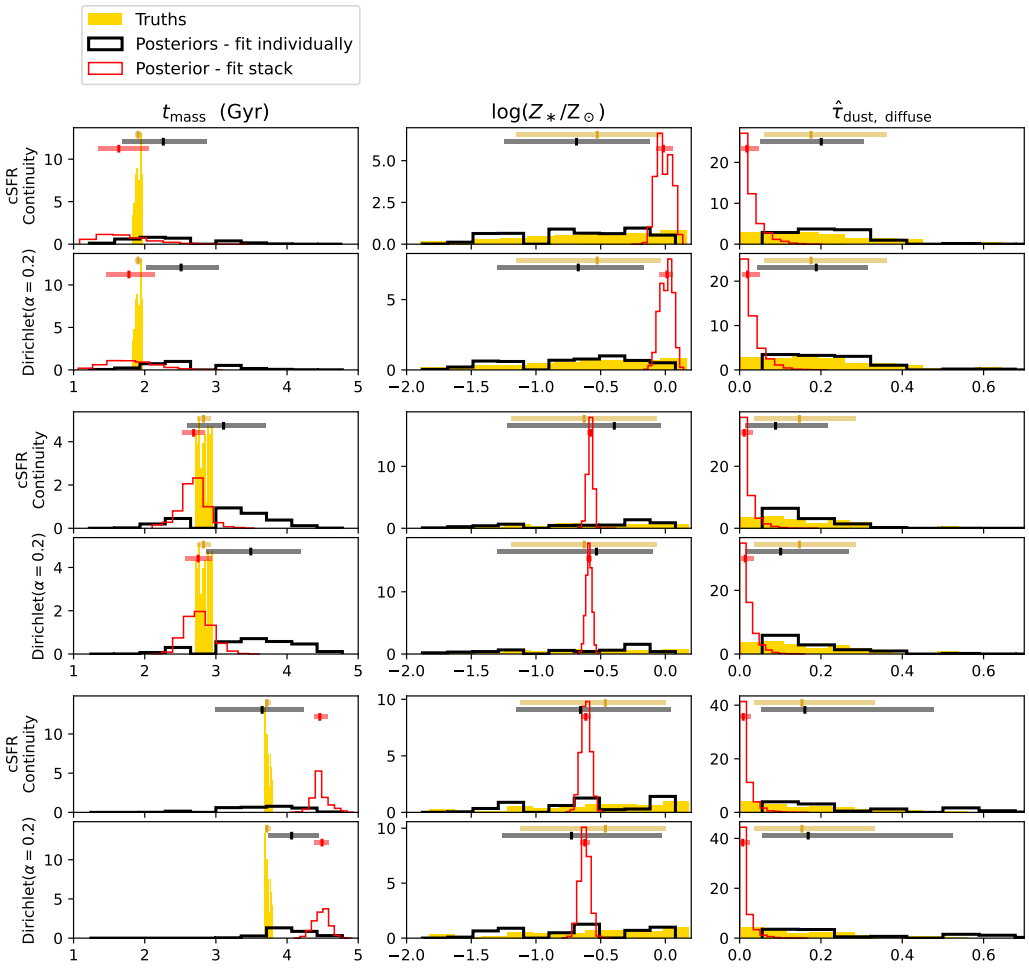

Figure C.3: Comparison of posteriors from fitting individual (black) or composite (red) observations, relative to the true distribution of properties (yellow solid histograms). The median and 68% CRs of each distribution are indicated by ticks and horizontal bars along the top of each panel. 100 mock galaxies are selected to have true ages of ∼1.8, 2.8, or 3.7 Gyr. The results from fitting the individual observations with the cSFR Continuity (black) or Dirichlet($\alpha_{\rm D} = 0.2$) (blue) priors are compared to the results from fitting the composite observations. While the relative ages of the three populations is apparent in each set of results, the age posteriors from the composite observations overestimate the age difference. The range of metallicity and dust properties, as inferred from the composite observations, are also underestimated. In comparison, the individual posteriors are noisy, but better trace the true properties of the mock galaxies.

C.2 A population model for stellar mass and metallicity

The stellar masses and metallicity posteriors shown in Figure 5.4 are generally unbiased from the true values. The exceptions are the posteriors based on low S/N observations fit with the cSFR Continuity model, which are systematically underestimated, and the delayed-τ model, which are increasingly biased with increasing $\log(Z_*/\mathrm{Z}_\odot)$. Similar to Section 5.6.1, we now focus on a set of posteriors which are noisy but unbiased to demonstrate the success of the population model when simultaneously modelling stellar mass and metallicity to determine a mass-metallicity relation (MZR). We note that the demand for the sample size increases with the dimensionality of the population model.

The population model is fit to the joint distribution of stellar mass and metallicity using the results of the NUV–NIR photometry, with spectral S/N $\sim$ 5 $\mathrm{\AA}^{-1}$, with the Dirichlet($\alpha_\mathrm{D} = 0.2$) model, for 100 (left) or 528 (right) mock galaxies. While for the mock quiescent galaxies, the Dirichlet($\alpha_\mathrm{D} = 0.2$) provides unbiased posteriors for these parameters, this may not be the case for all types of galaxies. Figure C.4 shows the probability of each mass and metallicity bin according to the population model with violin markers, relative to the true values (yellow), the co-added posteriors (black line), and the true MZR assumed when assigning metallicity values to the mock galaxies. As was the case when modelling the stellar mass alone, the population model generally offers a better representation of the underlying distribution, compared to simply adding the posteriors. This is particularly clear in the top panels of Figure C.4, where despite there being no galaxies with true masses within this mass-selection, the summed-posteriors suggest a peak about $\log(Z_*/\mathrm{Z}_\odot) \sim -0.5$, while the population model finds the distribution of metallicities to be uniform (none of the bins have zero probability). Comparing the results, the effect of increasing the number of galaxy posteriors is to tighten the probability distributions in each bin. That is, the shot noise is reduced as the sample size is increased.

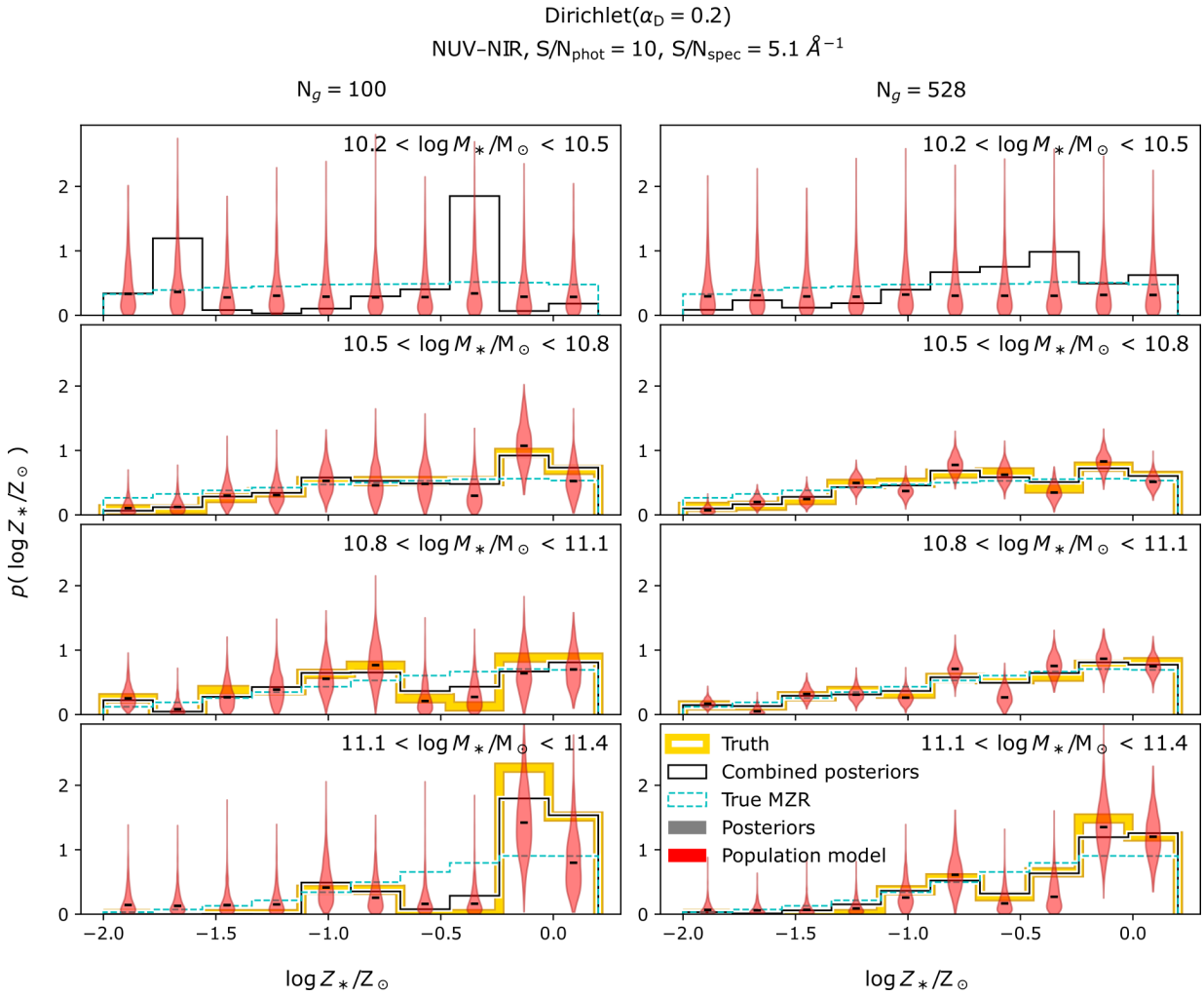

Figure C.4: Population model for stellar mass and metallicity (red) relative to the true values (yellow), co-added posteriors (black), and the true MZR (cyan). Posteriors were determined from fitting NUV–NIR photometry, S/N$_{\rm spec}$ $\sim$ 5 Å^{-1} spectroscopy with a Dirichlet($\alpha_{\rm D} = 0.2$) SFH model, which were unbiased; see Figure 5.4. Red violin markers indicate the probability attributed to each bin of metallicity and mass, with dashes marking the median of the distribution. On the left 100 galaxies were included in the model, while on the right there are 528 galaxies.

C.2.1 An informed mass-metallicity prior with population modelling

In Section 5.4.2 we noted that the stellar metallicities and mass-weighted ages were degenerate where the observations had low S/N. Without direct observational constraints on the metallicity, it still may be possible to break this degeneracy with the use of a well-calibrated prior. If ages could be more accurately recovered from poorer quality observations, perhaps the ages would be less sensitive to the choice of SFH model. In this section, we construct an informed mass-metallicity prior from the population model as described in Section C.2 which is then used in the SED model to re-derive the ages of the mock galaxies.

The crux of this procedure is building the mass-metallicity population model based on well-calibrated data, which requires high S/N observations. One therefore needs high quality observations for a representative sample of the population. We fit the population model to the joint distribution of stellar mass and metallicity using the results of the NUV–NIR photometry, with spectral S/N ~ 5 Å^{-1}, with the Dirichlet($\alpha_{\rm D} = 0.2$) model, which is well-calibrated to the true values (see Figure 5.4). We have implicitly assumed in applying the population model that the small sample of galaxies is a fair representation of the true diversity of properties. However, this is not the case. The true MZR of the mock galaxies, as described in Section 5.2, is shown with dashed cyan lines in Figure C.4. The subset of 100 mock galaxies were not selected to be representative of the mass-metallicity parameter space, and therefore the population model is not expected to be wholly consistent with the true MZR. We therefore supplemented our mock galaxy sample with 428 additional fits randomly drawn from the mock galaxy population. Given the breadth of the true MZR prior, we suggest that more than 528 data points are necessary to provide a fair representation, despite the narrow mass range we explore in this work.

Despite having an imperfect characterization of the true MZR, we nonetheless are interested in testing whether assuming this MZR as a *prior* when SED fitting improves the accuracy of the age posteriors. We construct an MZR prior based using a two-dimensional Heavyside function, with weights determined by the summed probability model (*i.e.*, essentially tracing the median probability in each bin). Ten bins cover $9 < \log M_*/{\rm M}_\odot < 12$ (despite the true masses being between $\log M_*/{\rm M}_\odot = 10.5$–$11.3$), and ten bins cover $-2 < \log Z_*/{\rm Z}_\odot < 0.2$. The cSFR set of 100 mock galaxies are then refit with this MZR prior.

We also test whether using the true MZR as a prior would improve the age posteriors. In this case, the same MZR distribution used to assign metallicities to the mock galaxies based on their stellar mass, as described in Section 5.2, is used as a prior. This follows the same procedure as used in Leja et al. (2019a) to fit a large number of observations from

3D-HST, and Chapter 3 in fitting the GOGREEN quiescent galaxies from which we have modelled the mock galaxies in this work.

Figure C.5 compares the posteriors based on fitting $S/N_{spec} \sim 1$ Å^{-1} observations with the Dirichlet($\alpha_D = 0.2$) model: i) determined with an uninformative prior for stellar mass and metallicity, ii) an informed MZR prior following the population-model fit to high-quality posteriors of a non-representative sample of galaxies, and iii) using an informed MZR prior from the true MZR. The last set of results, which include MIR constraints, are discussed in Section 5.4.3. The points are coloured by the discrepancy of the age posteriors with the true galaxy ages, and error bars indicate the 68% CRs. In each panel, the median and standard deviation of the difference between the posteriors medians and true values are listed, along with the percentage of galaxies consistent with the true values within the 68% CRs. Violin markers in the last column summarize the Δ distributions, with markers indicating the median Δ, and that for the first and last 25% of the population (see legend). In comparing these statistics between the different sets of posteriors, we see that using the poorly-calibrated MZR prior increased both the overall age discrepancy and scatter (although marginally; from $\Delta = 0.22$ Gyr and $\sigma_\Delta = 0.25$ Gyr, to $\Delta = 0.23$ Gyr and $\sigma_\Delta = 0.27$ Gyr). On the other hand, a perfectly-calibrated prior did improve the age estimates, although again only marginally (to $\Delta = 0.17$ Gyr, and $\sigma_\Delta = 0.25$ Gyr scatter). The dust parameters were not significantly affected by the use of either MZR prior, while the stellar metallicities were marginally less accurate where a MZR prior was assumed. More important to our desire to characterize the diversity of galaxy ages, neither MZR prior fixed the issue of the relative age bias – the ages of young ages are still overestimated, and old galaxies underestimated, such that the total diversity of the population is underestimated.

The results of fitting the mock galaxies with the true MZR prior where we have instead assumed the cSFR Continuity prior are shown in Figure C.6. This prior was assumed by Leja et al. (2019a) and in Chapter 3 when assuming the same MZR prior. The thought here is that the ages under this prior are more discrepant from the true values, such that a MZR constraint would be more helpful. However, as we saw with the results for the Dirichlet($\alpha_D = 0.2$) prior, any improvement in the accuracy of metallicities or ages is marginal.

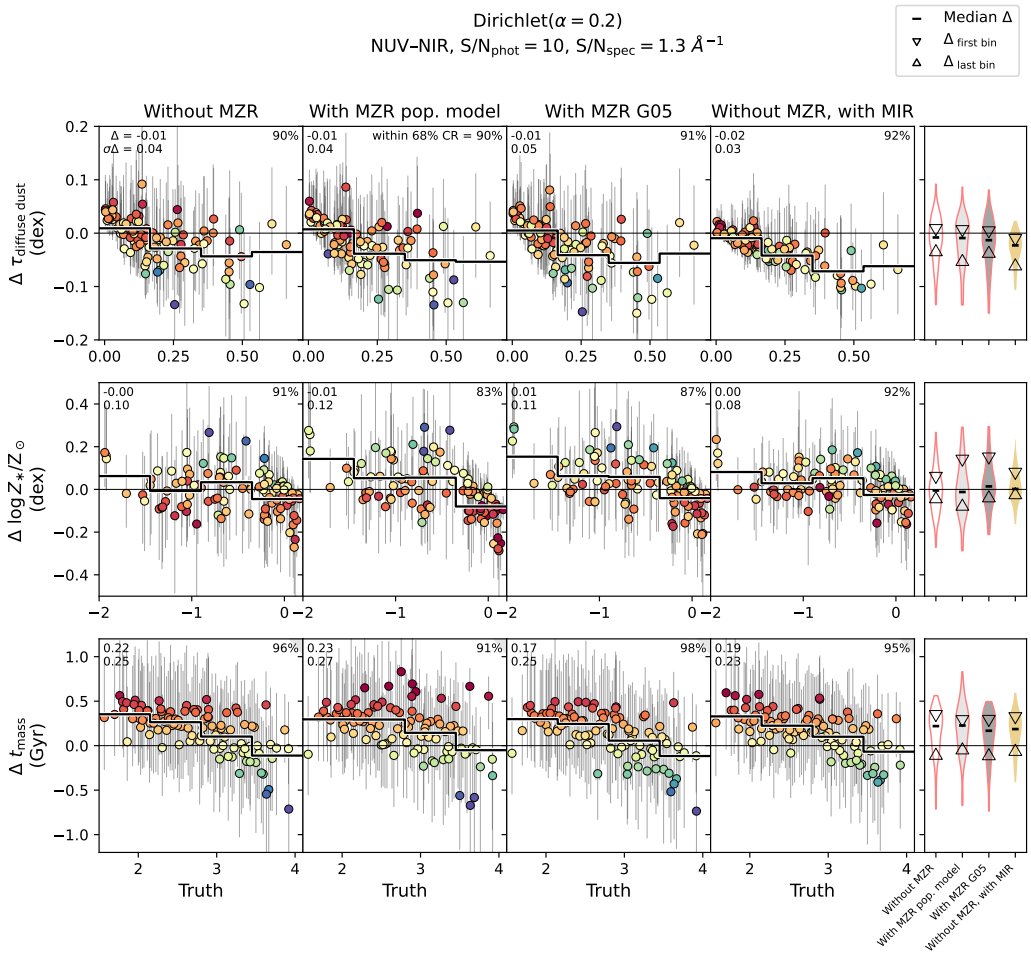

Figure C.5: Posterior median values from the Dirichlet($\alpha_D = 0.2$) SFH prior are compared to the true values for fits assuming: i) no correlation between stellar mass and metallicity, ii) an MZR from the population model fit to the same 100 mock galaxies as shown in Figure C.4, iii) the 'true' MZR prior from Gallazzi et al. (2005) used to assign metallicities to the mock galaxies, and iv) no MZR prior, with MIR constraints. The comparison is shown for the diffuse dust normalization constant, stellar metallicity, and mass-weighted age. Points are coloured according to the accuracy of the recovered ages. In each panel, the median difference between the posterior medians and true values are listed, along with the standard deviation. Error bars indicate the 68% CR of each posterior, where the percentage of points consistent within the true values within this interval are listed at the top right of each panel. Thick lines indicate the running median. Violin markers summarize the Δ distribution in the last column, where markers indicate the median Δ (dash), and first and last Δ of the running median (triangles). There is no *relative* bias across the sample where the three markers converge. The inclusion of the MZR prior, or MIR data, does not uniformly improve the posteriors based on low S/N observations.

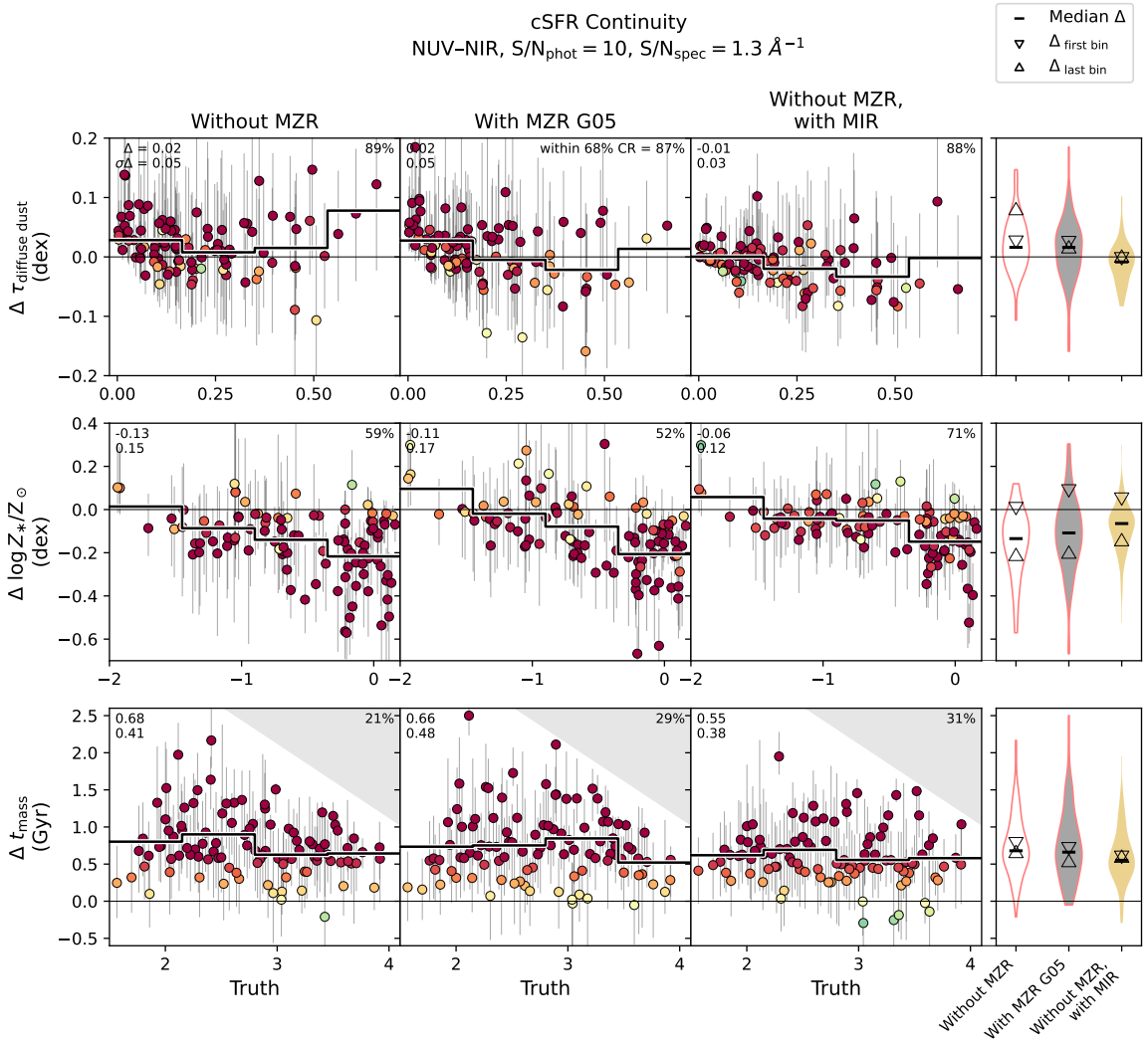

Figure C.6: As shown in Figure C.5, with results from the cSFR Continuity prior. Results from the population model based MZR prior are not shown.