



Universiteit
Leiden

The Netherlands

Elementary: the chemical fingerprints of massive galaxy formation over cosmic time

Cheng, C.M.T.

Citation

Cheng, C. M. T. (2026, June 25). *Elementary: the chemical fingerprints of massive galaxy formation over cosmic time*. Retrieved from <https://hdl.handle.net/1887/4307012>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4307012>

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

3 | AGES AND METALLICITIES OF QUIESCENT GALAXIES: CONFRONTING BROAD-BAND (UVJ) COLOURS WITH STELLAR ABSORPTION LINES

ABSTRACT

For decades, studying quiescent galaxies beyond $z \sim 1$ has been challenging due to the reliance on photometric spectral energy distributions, which are highly susceptible to degeneracies between age, metallicity, dust, and star-formation history. Only recently has deep, rest-frame, optical spectroscopy made robust metallicity and age measurements possible, allowing us to empirically assess their effects on continuum shapes. To this end, we measure ages and metallicities of ~ 700 massive ($10.2 \lesssim \log(M_*/M_\odot) \lesssim 11.8$) quiescent galaxies at $0.6 \lesssim z \lesssim 1.0$ from the Large Early Galaxy Astrophysics Census (LEGA-C) via continuum-normalized, absorption-line spectra, and compare with independent rest-frame $U - V$ and $V - J$ colours. Age increases along the quiescent sequence as both colours redden, consistent with stellar population synthesis (SPS) model predictions. Metallicity increases perpendicularly to the age trend, with higher metallicities at redder $U - V$ and bluer $V - J$ colours. Thus, age and metallicity behave differently in the UVJ diagram. Moreover, this trend conflicts with SPS model expectations of increasing metallicity approximately along the quiescent sequence. Independent dynamical mass-to-light ratio trends also differ dramatically from SPS model predictions. These results demonstrate that relying on model fits to continuum shapes alone may lead to systematic biases in ages, metallicities, and stellar masses. The cause of these data-model disparities may stem from non-solar abundance patterns in quiescent galaxies or the treatment of evolved stellar phases in the models. Resolving these discrepancies is crucial, as photometric data remain central even with *JWST*.

Chloe M. Cheng, Mariska Kriek, Aliza G. Beverage, Martje Slob, Rachel Bezanson, Marijn Franx, Joel Leja, Pavel E. Mancera Piña, Katherine A. Suess, Arjen van der Wel, Jesse van de Sande, and Pieter G. van Dokkum

2025, *Monthly Notices of the Royal Astronomical Society*, **540**, 1527

3.1 Introduction

Stellar populations of massive quiescent galaxies provide insight into the physical processes driving galaxy formation and evolution. In particular, tracing these populations over cosmic time is crucial for distinguishing different galaxy assembly and quenching mechanisms (e.g., Matteucci 1994; Trager et al. 2000; Conroy et al. 2014; Choi et al. 2014; Peng et al. 2015; Kriek et al. 2016; Kriek et al. 2019; Spitoni et al. 2017; Maiolino & Mannucci 2019; Trussler et al. 2020; Beverage et al. 2021, 2024, 2025; Cheng et al. 2024). These investigations subsequently provide constraints on cosmological simulations and theoretical models for galaxy formation and evolution (Maraston et al. 2006; Maiolino & Mannucci 2019).

A common approach to understanding quiescent stellar populations is to fit broad-band spectral energy distribution (SED) shapes with stellar population synthesis (SPS) models (Tinsley 1972; Searle et al. 1973; Larson & Tinsley 1978). Broad-band photometric data are relatively inexpensive to obtain and primarily reveal information about stellar mass-to-light (M/L) ratios, specific star-formation rates, and dust attenuation (e.g., Sawicki & Yee 1998; Bell & de Jong 2001; Papovich et al. 2001; Bell et al. 2003; Gallazzi & Bell 2009; Wuyts et al. 2009; Zibetti et al. 2009; Reddy et al. 2010, 2012; Leja et al. 2019b; Abdurro'uf et al. 2021; Johnson et al. 2021; see Conroy 2013 as well). They can also indicate ages and metallicities (e.g., Bell & de Jong 2000; MacArthur et al. 2004; Whitaker et al. 2010; Pacifici et al. 2016; Nersesian et al. 2024, 2025); however, these measurements are challenging due to well-known degeneracies between various stellar properties, including age, metallicity, star-formation history (SFH), and dust (Worthey 1994; Bell & de Jong 2001; Papovich et al. 2001; Bruzual & Charlot 2003; Gallazzi et al. 2005; Leja et al. 2017, 2019a,b).

While the degeneracy between age and dust can be broken with longer wavelength data (i.e. UVJ colours; Burgarella et al. 2005; Labbé et al. 2005; Wuyts et al. 2007; Leja et al. 2017), the age–metallicity degeneracy cannot reliably be broken with photometric data alone (Worthey 1994; Lee et al. 2007; Eminian et al. 2008; Conroy 2013; Nersesian et al. 2024, 2025). In this case, spectroscopic data can be used to disentangle and infer these quantities robustly. By fitting Lick indices (Burstein et al. 1984; Worthey et al. 1994) or the full spectrum (Conroy & van Dokkum 2012a; Conroy et al. 2018) with SPS models, one can directly quantify ages, metallicities, and elemental abundances. However, high-quality spectra are required to make these measurements. Thus, while this has been done extensively in local quiescent galaxies (e.g., Gonzalez et al. 1993; Caldwell & Rose 1998; Trager et al. 2000; Vazdekis 2001; Bruzual & Charlot 2003; Gallazzi et al. 2005; Thomas et al. 2005; Graves & Schiavon 2008; Choi et al. 2014; Conroy et al. 2014; Gu et al. 2018; Bernardi et al. 2023), these spectra were historically difficult to obtain beyond $z \sim 1$ as faint absorption features needed to measure ages and metallicities are shifted into the near-infrared. With developments allowing for ultra-deep ground- and space-based spectra, these analyses are now also being extended to high redshifts (up to $z \sim 3$), where massive quiescent galaxies are found to be metal-poor and α -enhanced compared to their low- z counterparts (e.g., Kriek et al. 2016; Kriek et al. 2019; Jafariyazani et al. 2020, 2025; Beverage

et al. 2021, 2023, 2024, 2025; Carnall et al. 2022; Zhuang et al. 2023).

Ideally, to maximize the information being considered when modelling stellar populations, one can fit photometric and spectroscopic data simultaneously (see e.g. Kriek et al. 2008b; Belli et al. 2019; Tacchella et al. 2022; Akhshik et al. 2023; Cappellari 2023; Kaushal et al. 2024; Slob et al. 2024; Nersesian et al. 2025). While powerful constraints on SFHs and, therefore, quenching mechanisms can be obtained by leveraging the complementary strengths of spectroscopy and photometry to break the degeneracies discussed above (Tacchella et al. 2022), models often struggle to reproduce both types of data simultaneously (e.g., Fumagalli et al. 2016; Carnall et al. 2024; Kriek et al. 2024; Slob et al. 2024). Additionally, different available SPS models can lead to different and inconsistent stellar population properties (e.g., Lee et al. 2007; Eminian et al. 2008; Muzzin et al. 2009; Conroy & Gunn 2010; Walcher et al. 2011; Pacifici et al. 2023; Whitler et al. 2023), in particular when including redder wavelength data for quiescent galaxies (van de Sande et al. 2015). These discrepancies may be due to the fact that there are a variety of uncertainties associated with SPS models, as different models make varying assumptions about stellar evolution and implement different spectral libraries, as well as to the choices that the user can make for the models and priors within the available parameter spaces (Walcher et al. 2011; Conroy 2013, see also Coelho et al. 2020; Byrne & Stanway 2023; Nersesian et al. 2024; Slob et al. 2024; Bellstedt & Robotham 2025). Thus, resulting galaxy properties may be uncertain and systematically biased. In combination with the modelling degeneracies discussed above, possible template mismatches are therefore concerning.

To overcome some of these issues and better understand the inconsistencies between both continuum shapes and absorption lines and different SPS models, we require a large sample of distant quiescent galaxies with high-quality (i) broadband photometry and (ii) age and metallicity measurements from deep, absorption-line spectra that are independent of the shape of the continuum. This is now possible with the Large Early Galaxy Astrophysics Census (LEGA-C; van der Wel et al. 2016, 2021; Straatman et al. 2018), a deep spectroscopic survey of galaxies at $0.6 \lesssim z \lesssim 1.0$ with UltraVISTA photometry (McCracken et al. 2012). In this work, we model the spectra of ~ 700 quiescent galaxies from LEGA-C with the ABSORPTION LINE FITTER (ALF, which continuum-normalizes spectra prior to fitting, Conroy & van Dokkum 2012a; Conroy et al. 2018). We combine rest-frame UVJ colours with independent, SSP-equivalent ages and metallicities to understand the relationships between these quantities as well as how our data compare to popular SPS models.

This paper is organized as follows: in Section 3.2 we summarize the LEGA-C data set and detail the full-spectrum fitting employed to derive galaxy ages and metallicities. In Section 3.3 we present our results. In particular, we discuss relationships between age and colour and metallicity and colour in Section 3.3.1. We also compare our results to predictions from two popular SPS models, the Flexible Stellar Population Synthesis models (FSPS; Conroy et al. 2009; Conroy & Gunn 2010) and the Bruzual & Charlot (2003) models, in Section 3.3.2. We compare our results with previous work, discuss the implications for galaxy evolution studies and SPS models, and outline some caveats of our work in Section 3.4. Finally, we

summarize and conclude in Section 3.5.

Throughout this work, we assume a flat Lambda cold dark matter (Λ CDM) cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. All magnitudes are given in the AB-magnitude system (Oke & Gunn 1983). A Kroupa (2001) initial mass function is assumed throughout.

3.2 Data and methods

In this work, we use data from LEGA-C (van der Wel et al. 2021). We provide a brief summary of the survey and sample selection in Section 3.2.1 and direct the reader to van der Wel et al. (2021) and Cheng et al. (2024) (C24 hereafter) for more details. In Section 3.2.2, we summarize the methods used to derive stellar population ages and elemental abundances for each data set.

3.2.1 LEGA-C

We use spectroscopic data from the third data release of LEGA-C, a European Southern Observatory (ESO) Public Spectroscopic survey of 3600 galaxies at $0.6 \lesssim z \lesssim 1.0$. Galaxies were selected from the UltraVISTA (McCracken et al. 2012) K -band catalogue by Muzzin et al. (2013b) and are located in the Cosmic Evolution Survey (COSMOS) field (Scoville et al. 2007b). The data were collected using the Visible MultiObject Spectrograph (VIMOS) on the ESO *Very Large Telescope* (VLT) over 128 nights, providing deep (20-h integration), $R \sim 3500$ spectra with signal-to-noise (S/N) $\sim 20 \text{ \AA}^{-1}$ on average. The raw 2D and reduced 1D spectra are available on the ESO Science Archive Facility^{1,2}. See van der Wel et al. (2016, 2021), and Straatman et al. (2018) for details.

We use the LEGA-C sample selected in C24. We summarize the sample selection and stellar population parameter derivation here, but refer the reader to C24 for details. In particular, C24 determined rest-frame UVJ colours using EAZY (Brammer et al. 2008), via the Muzzin et al. (2013b) K -band catalogue³ with the spectroscopic redshifts fixed to those measured from the LEGA-C spectra (van der Wel et al. 2021), and selected quiescent galaxies by employing the classification from Muzzin et al. (2013a). C24 also removed galaxies with spectroscopic redshifts < 0.5 . Additionally, galaxies in LEGA-C’s mask 2 were discarded as their noise spectra are significantly underestimated. A median rest-frame S/N $\gtrsim 20 \text{ \AA}^{-1}$ was required, as well as a maximum wavelength of at least $4450 \text{ \AA}$ in order

¹http://archive.eso.org/eso/eso_archive_main.html.

²The reduced 1D spectra and catalogue have been released by ESO (<http://archive.eso.org/cms/eso-archive-news/Third-and-final-release-of-the-Large-Early-Galaxy-Census-LEGA-C-Spectroscopic-Public-Survey-published.html>). They are also available here: <https://users.ugent.be/~avdrwel/research.html\#legac>.

³There is a known systematic offset in the Muzzin et al. (2013b) photometry due to zero-point corrections (van der Wel et al. 2021). This offset is corrected in the UltraVISTA DR3 catalogue which does not cover the entire LEGA-C field, and thus we use the original Muzzin et al. (2013b) catalogue. We test our analysis for the common galaxies (~ 290) in the two catalogues. The redetermined UVJ colours are consistent between both sets of photometry, and the use of the DR3 photometry does not change our conclusions.

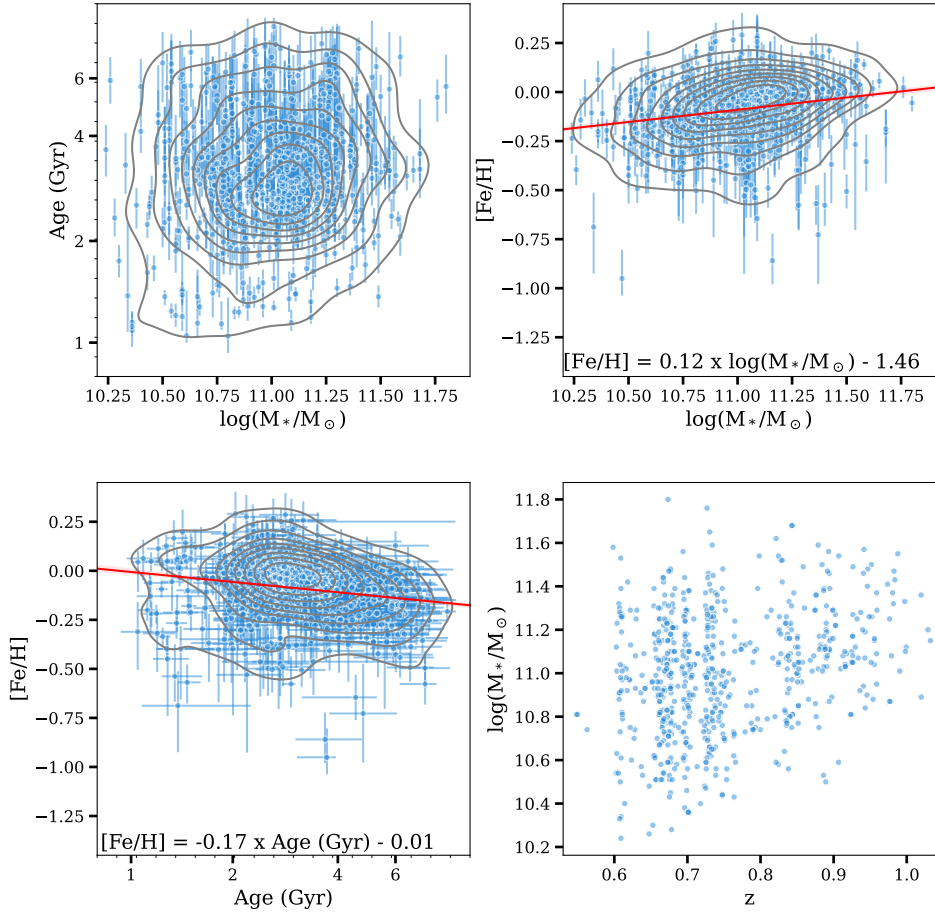


Figure 3.1: Properties of our final sample of massive quiescent galaxies from LEGA-C. In the top left panel, we show SSP-equivalent age as a function of stellar mass. In the top right panel, we show $[\text{Fe}/\text{H}]$ as a function of stellar mass. In the bottom left panel, we show $[\text{Fe}/\text{H}]$ as a function of age. In the bottom right panel, we show stellar mass as a function of redshift. We fit linear relations to the points in the top right and bottom left panels, respectively (solid lines with shaded uncertainties) and we state the corresponding best-fitting relations at the bottom of these panels. Stellar masses are measured by the LEGA-C team using MAGPHYS (da Cunha et al. 2008) and SSP-equivalent ages and $[\text{Fe}/\text{H}]$ abundances are derived from our ALF fits.

to sufficiently recover ages and metallicities. Note that the sample size in C24 was limited by the S/N of the spatially resolved spectra. However, this study is concerned with the full, integrated spectra. As such, we are able to add 244 more quiescent galaxies with $S/N \gtrsim 20 \text{ \AA}^{-1}$.

3.2.2 Spectral extraction and full-spectrum modelling

In this paper, we use the 1D spectra from C24, which were extracted from the 2D LEGA-C spectra as part of a spatially resolved analysis. In summary, an optimal extraction procedure was performed to obtain spatially resolved and integrated spectra, by fitting the spectral flux profiles with Moffat (1969) profiles. To extract the integrated spectra, these fits were used to perform weighted sums of all rows with significant flux (i.e. excluding (sub-)rows outside of the 3rd and 97th percentiles of the Moffat profiles to reduce noise, with the sums weighted by the Moffat profiles). For consistency, we perform the same extraction of the integrated spectra for the additional 244 galaxies in our sample. See C24 for details.

C24 derived stellar population parameters from the LEGA-C spectra using the ABSORPTION LINE FITTER (ALF⁴; Conroy & van Dokkum 2012a; Conroy et al. 2018). Empirical simple stellar populations (SSPs) constructed using the Mesa Isochrones and Stellar Tracks (MIST; Choi et al. 2016) and the Spectral Polynomial Interpolator (SPI; Villaume et al. 2017b)⁵ provide the foundation for the ALF models. ALF employs the Medium Resolution INT Library Of Empirical Spectra (MILES; Sánchez-Blázquez et al. 2006), the Extended Infrared Telescope Facility stellar library (E-IRTF; Villaume et al. 2017b), and a large sample of M-dwarf spectra (Mann et al. 2015) with SPI to generate stellar spectra as a function of effective temperature (T_{eff}), surface gravity, and metallicity from a data-driven model.

The empirical parameter space is set by the combination of the E-IRTF and Mann et al. (2015) samples. It spans $-2.0 \lesssim [\text{Fe}/\text{H}] \lesssim 0.5$ and $3.5 \lesssim \log(T_{\text{eff}}/\text{K}) \lesssim 3.9$. To ensure high-quality interpolation at the boundaries of the empirical parameter space, ALF also makes use of a theoretical stellar library (C3K; see Conroy & van Dokkum 2012a). By differentially including theoretical element response functions, ALF allows for variable abundance patterns. The elemental abundances of the base stellar population are estimated in Milone et al. (2011) and used in conjunction with the differential response functions to estimate the absolute abundances.

During the fitting procedure, ALF removes the continuum from the observations by fitting a high-order Chebyshev polynomial to the ratio of the data to the model. It then implements a Fortran version of the Markov chain Monte Carlo (MCMC) algorithm *emcee* (Foreman-Mackey et al. 2013) to sample the posteriors of 46 stellar population parameters. In particular, it allows for arbitrary variation in stellar population age and detailed elemental abundance patterns. To characterize observed errors, ALF also fits for systematic parameters. Note that ALF fits spectra between 3700 and 24000 Å and stellar populations older than 1 Gyr. For details, see Conroy & van Dokkum (2012a) and Conroy et al. (2018).

C24 fit each spectrum with 1024 walkers, 20000 burn-in steps, and a 1000-step production run. For each fit, 500 MCMC chains were examined. A Kroupa (2001) initial mass function (IMF) was assumed and a single stellar population age was fit. To initialize the age of each galaxy and to avoid the ALF walkers getting

⁴<https://github.com/cconroy20/alf> .

⁵https://github.com/AlexaVillaume/SPI_Utils .

trapped at an unrealistically high initial age, a random value was drawn from a uniform distribution centred at 3 Gyr. Additionally, the upper limit of the age prior was set to be the age of the Universe at the redshift of each galaxy, plus 2 Gyr to allow for uncertainties. The hot star component was not fit. Variation in all stellar parameters in ALF was allowed. Galaxies for which ALF fit an age < 1 Gyr were discarded (5 galaxies), as the abundances are extrapolated for ages younger than 1 Gyr and are unreliable (Conroy et al. 2018). Furthermore, the fits for galaxies where the posteriors ran up against the age, $[Z/H]$, and $[Fe/H]$ priors were discarded as these fits would result in unreliable ages and metallicities (~ 6 per cent of the sample). There are a variety of reasons why the parameter space may not be fully explored for these objects, for example low S/N, low levels of residual star formation, or non-stellar absorption. We fit the 244 galaxies that we add to this sample in exactly the same way as C24. Finally, we visually inspect all of the spectra and remove any objects near the edge of the UVJ box with strong $[O_{III}]$ emission (10 galaxies), which could potentially host active galactic nuclei or have low levels of star formation. The final LEGA-C sample examined in this study includes 683 quiescent galaxies.

We show the properties of our final sample in Fig. 3.1. In the top panels, we show, respectively, SSP-equivalent ages and $[Fe/H]$ abundances versus stellar masses (computed using MAGPHYS, da Cunha et al. 2008, and obtained via private communication with the LEGA-C team). In the bottom-left panel, we show our SSP-equivalent ages versus SSP-equivalent $[Fe/H]$ abundances. In the bottom right panel, we show stellar masses versus spectroscopic redshifts. We remove two galaxies with MAGPHYS masses $< 10^{10} M_{\odot}$ as the flux levels of their spectra are abnormally low. Overall, the galaxies in our sample are at $0.6 \lesssim z \lesssim 1.0$ with $10.2 \lesssim \log(M_*/M_{\odot}) \lesssim 11.8$ and $-1.3 \lesssim [Fe/H] \lesssim 0.3$, and are $1.0 - 8.6$ Gyr old.

A subset of 135 quiescent LEGA-C galaxies were fit using ALF in Beverage et al. (2023). They fit the 1D extracted spectra from the LEGA-C public release. We find that the stellar population parameters for the galaxies that we have in common (110 galaxies) are fully consistent between the two studies, despite subtle differences in the spectral extraction procedures (not shown). We also qualitatively recover the negative relation between $[Fe/H]$ and age (Beverage et al. 2021) and the slightly positive relation between $[Fe/H]$ and mass (Beverage et al. 2023), as shown by the red lines in Fig. 3.1. We state the linear fits in the corresponding panels in Fig. 3.1.

3.3 Results

In this paper we compare detailed stellar population properties derived from absorption lines to colours derived from broad SED shapes. In Section 3.3.1, we explore how age and metallicity vary on the UVJ diagram. In Section 3.3.2, we compare our results with predictions from popular SPS models.

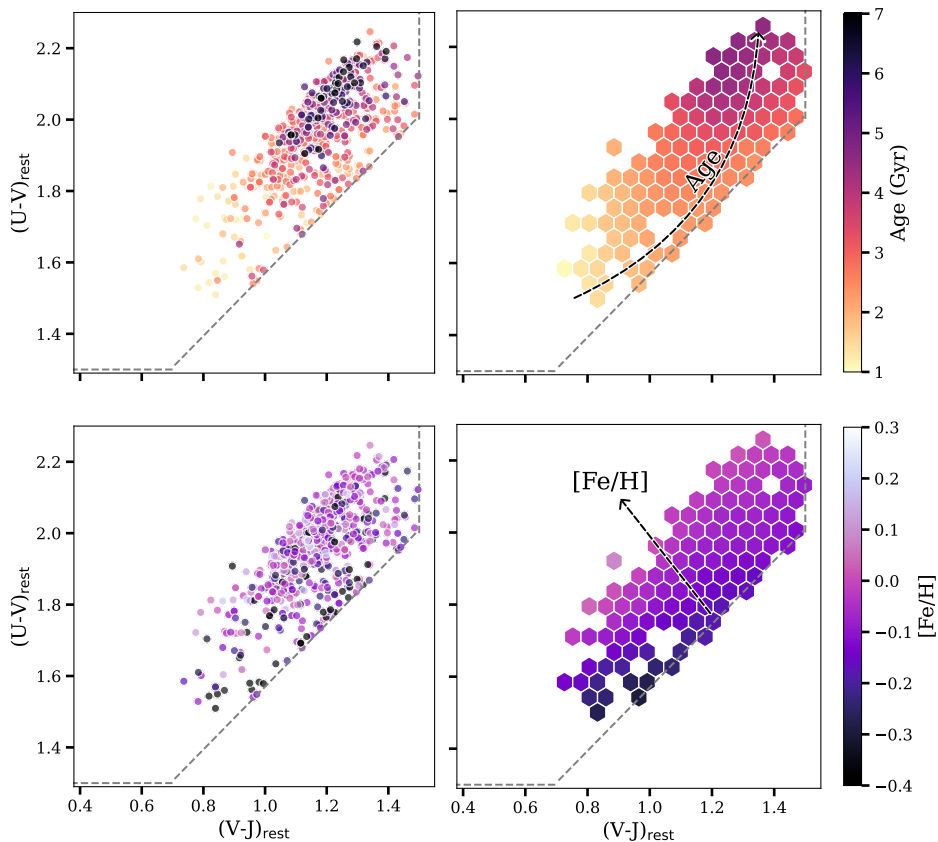


Figure 3.2: Rest-frame UVJ diagrams of quiescent galaxies at $0.6 \lesssim z \lesssim 1.0$ from the LEGA-C survey. *Left:* In these panels, we colour code the symbols by their measured ages (top) and their measured $[\text{Fe}/\text{H}]$ values (bottom), as derived by fitting absorption lines in continuum-normalized spectra with ALF. *Right:* In these panels, we LOESS-smooth the ages and metallicities from the left panels (Cappellari et al. 2013a; Cleveland & Devlin 1988). In each panel, we plot an arrow showing the general, qualitative direction of the dependence between age and colour (top) and metallicity and colour (bottom).

3.3.1 Trends with age and metallicity

In Fig. 3.2, we show several rest-frame UVJ diagrams colour coded by age and metallicity. In the left panels, we present the results for the individual galaxies in our LEGA-C sample. The symbols are colour coded by their measured SSP-equivalent ages (top) and $[\text{Fe}/\text{H}]$ values (bottom), both measured via absorption lines in continuum-normalized spectra (Section 3.2.2). In the right panels, we smooth the data using the Locally Weighted Regression (LOESS) 2D algorithm (Cleveland & Devlin 1988) via the LOESS Python package (Cappellari et al. 2013a).

There is a strong sequence in age in the top panels. In particular, age increases approximately parallel to the quiescent sequence, from ~ 1 Gyr in the bluest

regions of the UVJ diagram to $\sim 3 - 5$ Gyr in the reddest regions. There is some curvature to this trend, with the oldest galaxies populating the reddest end of the $U - V$ axis, but located slightly blueward on the $V - J$ axis. Interestingly, $[\text{Fe}/\text{H}]$ behaves differently, increasing approximately perpendicular to the quiescent sequence, from subsolar ($-0.3 \lesssim [\text{Fe}/\text{H}] \lesssim -0.2$) near the edge of the UVJ box to supersolar ($[\text{Fe}/\text{H}] \sim 0.08$) as one moves across the quiescent sequence. We indicate the approximate, qualitative trends that we describe here with arrows in the right panels.

This observed age trend has been suggested in previous work (e.g., Whitaker et al. 2012, 2013; Mendel et al. 2015; Belli et al. 2019; Díaz-García et al. 2019). However, in these studies, galaxy ages were derived either via the broad-band photometric SEDs alone, or by including the SEDs in simultaneous fits to spectra and photometry. In contrast, our ages are completely independent of continuum shape as they are derived directly from the continuum-normalized LEGA-C spectra.

Moreover, for the first time, we observe a trend in metallicity as derived from absorption lines. While ages and metallicities are known to be degenerate (Worthey 1994; Bruzual & Charlot 2003; Gallazzi et al. 2005), previous studies that examined the relationship between ages and broad-band colours (e.g., Whitaker et al. 2012, 2013; Mendel et al. 2015; Belli et al. 2019; Díaz-García et al. 2019) did not take metallicities into account⁶ as they were poorly constrained for distant quiescent galaxies. However, it is now possible to measure robust metallicities beyond the local Universe and disentangle them from ages using a combination of deep LEGA-C spectra and ALF (Conroy & van Dokkum 2012b; Choi et al. 2014; Beverage et al. 2021, 2023; Gu et al. 2022; Cheng et al. 2024). As our measured age and $[\text{Fe}/\text{H}]$ trends are distinct, this demonstrates that observed reddening along the quiescent sequence is indeed primarily due to an increase in age, as previously suggested (e.g., Whitaker et al. 2012, 2013; Mendel et al. 2015; Belli et al. 2019; Díaz-García et al. 2019). Furthermore, the differing trends may indicate that age and metallicity are not strongly degenerate in this UVJ space. However, this does not imply that we can simply derive metallicities and ages directly from UVJ colours as there are many other fitting degeneracies that must be considered. Thus, we do not attempt to quantify the strength of the age and metallicity trends here, but simply state that qualitative trends exist. We discuss this further in Section 3.4.1.

3.3.2 Comparison to stellar population models

Our current understanding of stellar populations in unresolved systems from both broad-band photometry and high-quality spectroscopy primarily relies on SPS models (Tinsley 1972; Searle et al. 1973; Larson & Tinsley 1978). Building these models involves combining isochrones, spectral libraries, and the IMF to construct simple stellar populations (SSPs). Every model implements these components in slightly different ways. Additionally, there are several aspects of population synthesis that still contain uncertainties (Renzini 2006; Walcher et al. 2011; Conroy

⁶Note that Díaz-García et al. (2019) examined trends between UVJ colours and metallicity, however, they derived total metallicity from photometry alone and used a different SPS model.

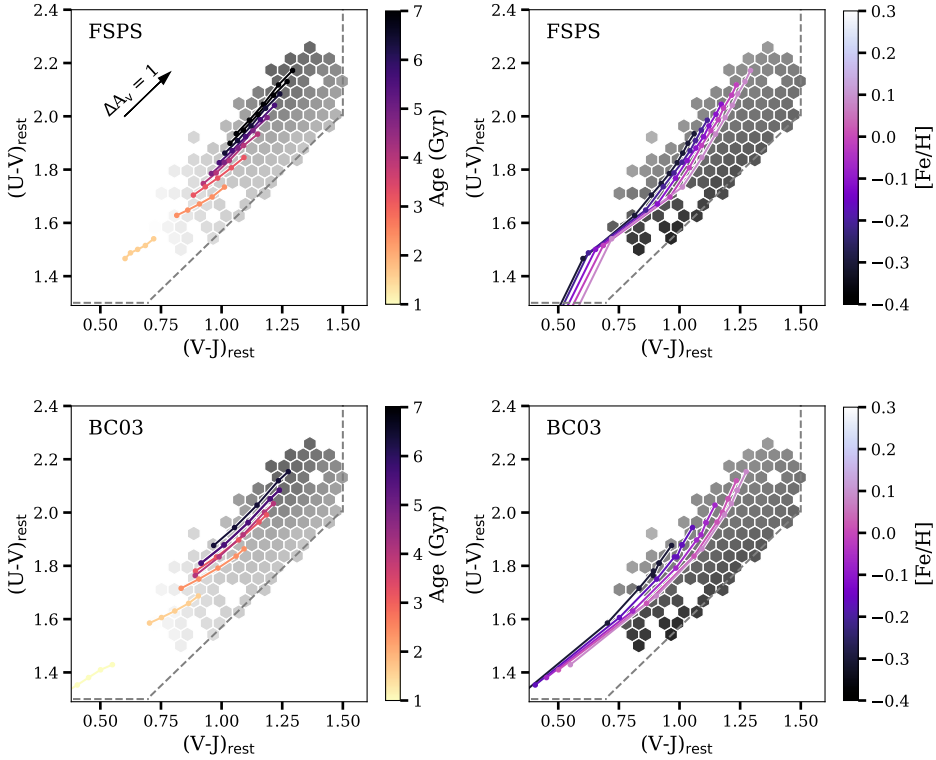


Figure 3.3: The same LOESS-smoothed rest-frame UVJ diagrams as in the right panels of Fig. 3.2 (grey) but with stellar evolution tracks overplotted. Tracks produced using PYTHON-FSPS (Johnson et al. 2023, see also Conroy et al. 2009; Conroy & Gunn 2010) are shown in the top panels and tracks produced using BAGPIPES (Carnall et al. 2018, which uses the Bruzual & Charlot 2003, or BC03, SPS models) are shown in the bottom panels. For each model, we produce tracks spanning the age ($\sim 1 - 8$ Gyr) and $[\text{Fe}/\text{H}]$ ($-0.31 < [\text{Fe}/\text{H}]/\text{dex} < 0.08$) ranges of our sample. The tracks in the left and the right panels are the same tracks for each model. On the left, we emphasize the age trend and the ages are indicated by colour-coded points, connected by lines to guide the eye. On the right, we emphasize the $[\text{Fe}/\text{H}]$ trend, with metallicities indicated by colour-coded points connected by lines. In general, we see the same kind of sequence in age in the stellar evolution tracks from both models that we see in our data (left panels). On the other hand, the metallicity sequence of the tracks seems to be perpendicular to that observed in the LEGA-C data (right panels).

2013). For example, all models must make approximations for processes such as convection, stellar rotation, mass loss, binary stars, and advanced stages of stellar evolution (e.g. thermally pulsating asymptotic giant branch, or TP-AGB, stars). Many codes treat these processes differently (compare the models of Bruzual & Charlot 2003 with those of Maraston 2005, for instance, and see Muzzin et al. 2009). Additionally, spectral libraries are often incomplete (Coelho et al. 2020;

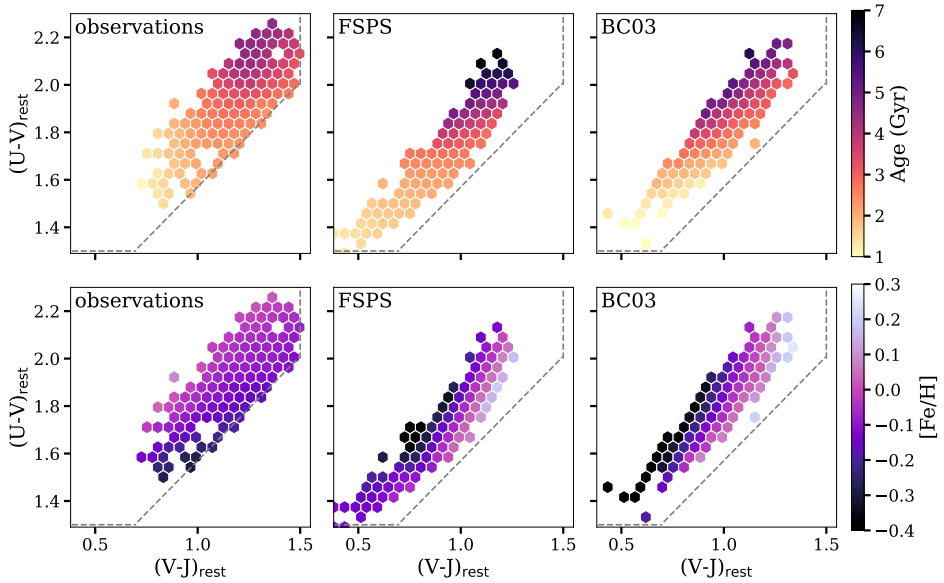


Figure 3.4: LOESS-smoothed rest-frame UVJ diagrams for the observations and the models. For the models, we input the ages and metallicities of each galaxy in our LEGA-C sample and output the predicted rest-frame UVJ colours, with some random scatter added to simulate observational uncertainties. The diagrams are colour coded by age in the top panels and by $[\text{Fe}/\text{H}]$ in the bottom panels. The same diagrams as in the right panels of Fig. 3.2 are shown for the data in the left panels, the diagrams for the FSPS models are shown in the middle panels, and the diagrams for the BC03 models are shown in the right panels. Age follows the quiescent sequence in all three panels. However, the trend in $[\text{Fe}/\text{H}]$ is inconsistent between the data and the models.

Byrne & Stanway 2023; Bellstedt & Robotham 2025). Thus, different SPS models can produce inconsistent results (see, e.g., Lee et al. 2007; Eminian et al. 2008; Muzzin et al. 2009; Conroy & Gunn 2010; Walcher et al. 2011; van de Sande et al. 2015; Coelho et al. 2020; Byrne & Stanway 2023). Moreover, they are often not able to reliably reproduce observations, either from photometry, spectra, or combined photometric and spectroscopic data, and in cases where models can reproduce the observations, the physics implied by each model can be conflicting (e.g., Lee et al. 2007; Fumagalli et al. 2016; Byrne & Stanway 2023; Cappellari 2023; Pacifici et al. 2023; Kaushal et al. 2024; Nersesian et al. 2024; Slob et al. 2024; Tortorelli et al. 2024; Bevacqua et al. 2025; Kim et al. 2025).

We note that several studies have fit the LEGA-C sample with different models (e.g., Chauke et al. 2018; Beverage et al. 2021, 2023; Barone et al. 2022; Cappellari 2023; Kaushal et al. 2024; Nersesian et al. 2024). However, in this work, we have access to both photometric colours from broad-band SEDs and completely independent measurements of ages and metallicities from continuum-normalized, absorption-line spectra. Thus, we are in a position to compare our results to predictions from popular SPS models to better understand the uncertainties and

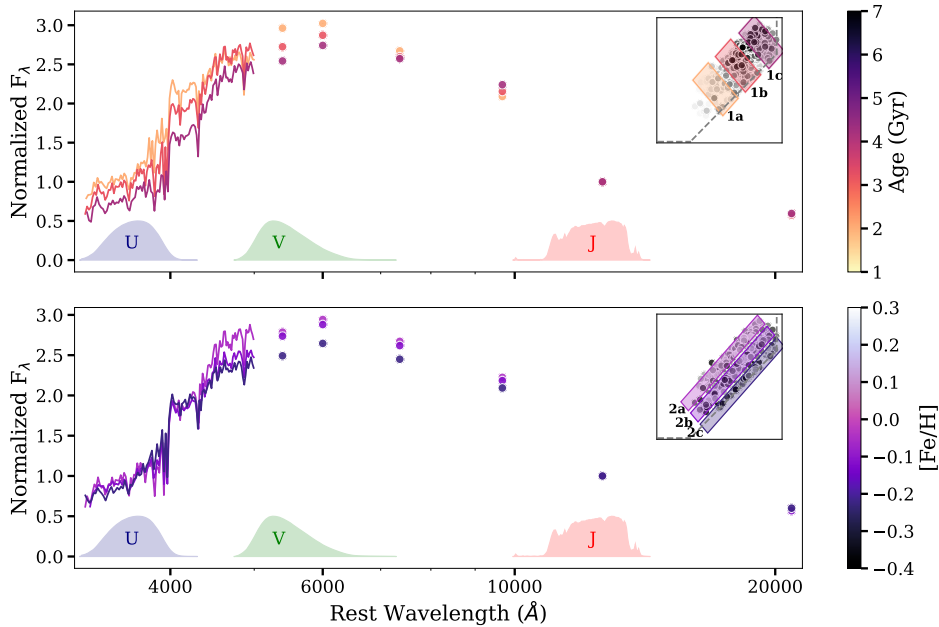


Figure 3.5: A comparison of stacked LEGA-C spectra. In the top panel, we select three bins along the quiescent sequence and approximately along our observed age sequence. In the bottom panel, we select three bins perpendicular to the quiescent sequence and approximately along our observed metallicity sequence. We normalize the spectra over $4400 - 4500 \text{ \AA}$ and mean-stack the spectra in each bin. We renormalize each stack by the photometric rest-frame J band. We also normalize and mean-stack the UltraVISTA photometry beyond $5000 \text{ \AA}$ for the same galaxies in each bin in the same way. We colour code the stacked spectra by the median age or metallicity in each bin and median-bin the spectra in bins of 50 pixels for visibility. We state each curve’s median age, $[Fe/H]$, sSFR averaged over the last 0.01 Gyr (measured via MAGPHYS by the LEGA-C team), and the number of galaxies in each stack in Table 3.1. The UVJ filter response curves are indicated in each panel. In the inset panels, we show the same rest-frame UVJ diagrams as in the left panels of Fig. 3.2 but greyscaled, with the bins indicated by the shaded regions. This figure serves to illustrate our results, as in the top panel the difference in ages in the three stacked spectra (along the observed age sequence) is greater than in the bottom panel (along the observed $[Fe/H]$ sequence), while the opposite is true for the difference in $[Fe/H]$.

discrepancies we discussed above. To facilitate this comparison, in Fig. 3.3 we overplot stellar evolution tracks generated using SPS models on the same LOESS-smoothed UVJ -diagrams as in Fig. 3.2 but with the hexbins greyscaled. In the top panels, we show tracks computed using PYTHON-FSPS (Johnson et al. 2023, and see also Conroy et al. 2009; Conroy & Gunn 2010), implemented with the MIST isochrones (Choi et al. 2016; Dotter 2016) and MILES spectral library (Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011). In the bottom panels, we similarly show tracks computed using BAGPIPES (Carnall et al. 2018), which makes use

Table 3.1: Median age, [Fe/H], sSFR over 0.01 Gyr, and number of galaxies in each stack for the stacked spectra in Fig. 3.5.

Box	Age (Gyr)	[Fe/H]	sSFR (0.01 Gyr) ($M_{\odot}/\text{yr}$)	Number of galaxies
1a	2.01	-0.14	-10.99	48
1b	3.19	-0.09	-11.37	297
1c	4.25	-0.07	-11.46	131
2a	3.66	-0.05	-11.58	297
2b	3.22	-0.12	-11.21	185
2c	2.67	-0.22	-10.73	62

of the 2016 version of the Bruzual & Charlot (2003) SPS models (BC03 hereafter), implemented with the Padova isochrones (Alongi et al. 1993; Bressan et al. 1993; Fagotto et al. 1994a,b; Girardi et al. 1996) and MILES spectral library. In both cases, we model a stellar population with a τ SFH truncated after 1 Gyr and no dust (i.e., $A_V = 0$, see Section 3.4.4 for a discussion of the impacts of our assumed SFH and A_V). We indicate with an arrow the direction in which dust would move the tracks according to the Calzetti et al. (2000) attenuation law. From the models, we compute UVJ colours for a grid of ages and metallicities, spanning $\sim 1 - 8$ Gyr (the age of the Universe at $z \sim 0.6$) and the metallicity range of our sample ($-0.31 < [\text{Fe}/\text{H}]/\text{dex} < 0.08$). We colour code the tracks by stellar population age in the left panels, where we connect points of the same age by lines to guide the eye. We colour code the tracks by metallicity in the right panels, where we assume $[\text{Fe}/\text{H}] \approx \log(Z/Z_{\odot})$ (see Beverage et al. 2025 and Section 3.4.3).

In the left panels of Fig. 3.3, we see that both models have age trends that are qualitatively similar to those seen in our observations. In particular, ages generally increase along the quiescent sequence, with the trend steepening at the reddest tip of the quiescent sequence. This comparison indicates that SPS models predict similar UVJ colours as a function of age to what we observe for the LEGA-C galaxies.

However, the same cannot be said when we examine the metallicity panels (right panels of Fig. 3.3). Here, both models do not follow the trends that we see in our observations, and perhaps even show a sequence in the perpendicular direction (i.e. while our metallicities increase approximately perpendicular to the quiescent sequence, the metallicities of the models seem to increase more along the quiescent sequence). This finding suggests that models like FSPS and BC03 may not accurately predict broad-band colours for varying metallicities.

We note that the model tracks in Fig. 3.3 only span a small range in the UVJ quiescent box and that the real galaxies (grey hexagons) show a much broader distribution. This is likely due to photometric uncertainties, differences between the real SFH and the assumed τ -model SFH, differences between the real dust properties and the assumed dust extinction ($A_V = 0$), or template mismatches. To assess the significance of our results and evaluate whether we expect to be able to observe the metallicity trend that we see in the UVJ diagram, we generate

the predicted UVJ colours for each galaxy in our sample assuming the FSPS and BC03 models. In particular, we input the measured SSP-equivalent ages and metallicities of our sample and output rest-frame UVJ colours for each galaxy. Uncertainties on the rest-frame fluxes are estimated using the median S/N of the observed photometric data points (between 1 and 7 bands, depending on the filter) which are close in rest-frame wavelength. For the U , V , and J bands, we find typical uncertainties of ~ 3.5 , ~ 1.0 , and ~ 1.0 per cent, respectively. We then add random scatter to the model fluxes and use these to calculate realistic U , V , and J colours by randomly sampling from a normal distribution multiplied by each of these typical observed relative errors.

We plot the LOESS-smoothed UVJ diagrams colour coded by age and metallicity for the observations, the FSPS models, and the BC03 models in Fig. 3.4. In the top panels, the age trend is qualitatively consistent between the observations and the models. In particular, ages always increase along the quiescent sequence, with some curvature near the top right corner of the UVJ box. In the bottom panels, however, the trend in $[\text{Fe}/\text{H}]$ is not consistent between the observations and the models, with the trend being nearly perpendicular between the two, consistent with what we see in Fig. 3.3. This test indicates that we can realistically expect to detect the trend with metallicity in our data. Thus, our result is significant. Additionally, even if we are underestimating the typical relative observed errors on the UVJ colours as described above, we find that we are still able to recover the trends with typical uncertainties of up to 5 per cent for each of the U , V , and J bands (this percentage allows the model panels to approximately display the same spread in UVJ colours as the observations). We discuss possible additional sources of scatter in Section 3.4.4.

Finally, in Fig. 3.5, we visualize how the observed SEDs and spectra of galaxies change along and perpendicular to the quiescent sequence, in roughly similar directions to those in which age and metallicity increase, respectively. In the top panel, we select three bins along the quiescent sequence (i.e. approximately along the age sequence). In the bottom panel, we select three bins perpendicular to the quiescent sequence (i.e. approximately along the metallicity sequence). These bins are shown by the shaded rectangles on the rest-frame UVJ diagrams in the inset panels. We normalize the spectra over the region from 4400 to 4500 Å and mean-stack the spectra in each bin. We then normalize each stack by the UltraVISTA photometry in the rest-frame J band. We also normalize and mean-stack the UltraVISTA photometry beyond 5100 Å for the same galaxies in each bin in the same way. In particular, as the galaxies in our sample are at similar redshifts, we mean-stack the photometry in the observed frame and de-redshift the median wavelength of each band when plotting the points in Fig. 3.5. In the top panel, we colour code the stacked spectra and photometry by the median age in each bin and in the bottom panel we colour code by the median metallicity in each bin. We median-bin the spectra in bins of 50 pixels for ease of visualization. The UVJ filter response curves are also indicated in each panel. We state the median age, $[\text{Fe}/\text{H}]$, specific star-formation rate (sSFR) over 0.01 Gyr (computed using MAGPHYS, da Cunha et al. 2008, and obtained via private communication with the LEGA-C team), and the number of galaxies in each stack in Table 3.1.

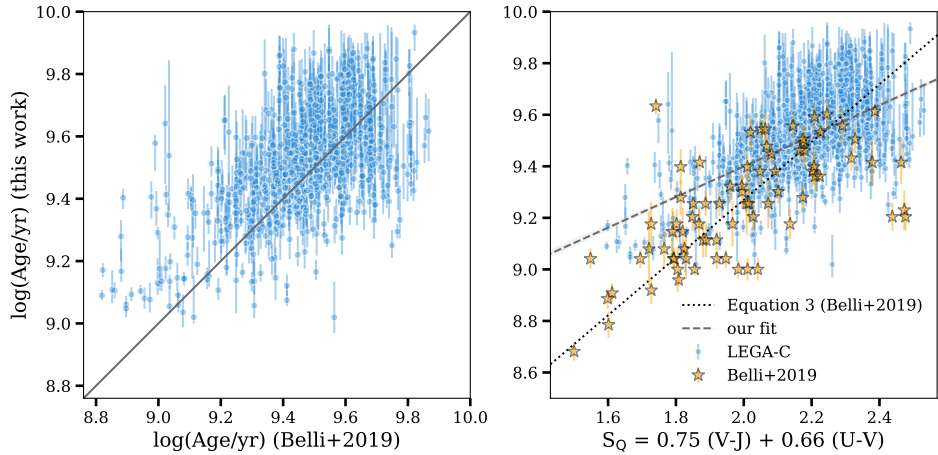


Figure 3.6: *Left:* A comparison between the ages that we measure by using ALF to fit absorption lines in continuum-normalized spectra (y -axis) to those derived using equation (3) in B19 (i.e. using only the UVJ colours, x -axis). The one-to-one relation is represented by the solid line. In general, we find older ages compared to those predicted from the B19 age-colour relation. However, overall, the ages of the LEGA-C galaxies are relatively consistent with the predicted ages from B19. *Right:* SSP-equivalent ages plotted as a function of the S_Q axis defined in equation (2) in B19. We also reproduce the points from fig. (8) in B19 (orange stars). The dotted line is the age-colour relation from B19. The dashed line shows our illustrative linear fit to the points in this plot, with shaded regions indicating the 1σ uncertainties. Our fit differs qualitatively from the age-colour relation derived in B19. Furthermore, there is a lot of scatter that is not captured by a linear relation on these axes.

This figure serves as an illustration of our results. In particular, the difference in ages in the three stacked spectra in the top panel (~ 1 Gyr on average) is greater than in the bottom panel (~ 0.6 Gyr on average). On the other hand, the difference in $[\text{Fe}/\text{H}]$ values in the top panel (~ 0.05 dex on average) is smaller than in the bottom panel (~ 0.1 dex on average). Thus, the selected bins in the top panel are tracing the age sequence that we observe with very little dependence on $[\text{Fe}/\text{H}]$. When we examine the combined spectra and SEDs, we see that age is primarily shifting the spectra to redder wavelengths and making the $4000 \text{ \AA}$ break more pronounced. Meanwhile, the selected bins in the bottom panel are tracing the observed $[\text{Fe}/\text{H}]$ sequence with a smaller dependence on age compared to the top panel. Metallicity is not shifting the spectra in the same way as age in the top panel (i.e. the spectra all have approximately the same peak wavelength in different metallicity bins), but it acts to make the galaxies redder in $U - V$. However, as a caveat we note that this is not a perfect comparison as metallicity and age are changing in both panels, since the observed age and metallicity trends are not exactly along and perpendicular to the quiescent sequence, respectively. The observed shifts in the spectra here may also be affected by other degenerate effects, such as dust, SFH, or sSFR. See Section 3.4.4 for details.

3.4 Discussion

In Section 3.3, we examined trends between spectroscopic ages and metallicities and independent UVJ colours. We found that observed reddening along the quiescent sequence in the UVJ diagram is primarily due to age, while metallicity increases approximately perpendicular to the quiescent sequence (Section 3.3.1). We also investigated whether these trends can be reproduced by popular SPS models. While these models predict similar UVJ colours as a function of age compared to what we see in the data, this is not the case for metallicity (Section 3.3.2).

Here, we discuss the implications of our results. In Section 3.4.1, we compare our results to previous studies, in particular Belli et al. (2019). We discuss the implications for galaxy evolution studies in general, and for derived stellar M/L ratios and masses specifically, in Section 3.4.2. We examine the implications of our results for SPS models in Section 3.4.3. Finally, in Section 3.4.4, we discuss caveats of our work that should be taken into account.

3.4.1 Comparison to previous work

In Section 3.3.1, we used independent measurements of age and UVJ colours to demonstrate that reddening along the quiescent sequence is primarily due to age, as previously suggested by observations (e.g., Whitaker et al. 2012, 2013; Mendel et al. 2015; Belli et al. 2019; Díaz-García et al. 2019) and cosmological simulations (e.g., Akins et al. 2022). In Belli et al. (2019) (B19 hereafter), this trend was used to derive a relationship between ages in quiescent galaxies at $1.0 < z < 2.5$ and UVJ colours (see their equation (3))⁷. This relation has since been used to predict quiescent galaxy ages using only UVJ colours (e.g., Suess et al. 2019b; Clausen et al. 2025). However, in B19 and other previous studies, the broad-band photometric SEDs were included when deriving galaxy ages. In contrast, our ages are completely independent of continuum shape as they are derived directly from the continuum-normalized LEGA-C spectra, by only fitting absorption lines using ALF. Thus, this study provides an independent check of the age trend along the quiescent sequence in the UVJ diagram.

In the left panel of Fig. 3.6, we compare our ages measured from the LEGA-C spectra using ALF to those derived using equation (3) in B19 (i.e. predicted from UVJ colours). In general, we find older ages compared to those predicted from the B19 age–colour relation. This difference is somewhat surprising as we measure SSP-equivalent ages while B19 considers mass-weighted ages, and SSP-equivalent ages are typically found to be younger than mass-weighted ages (Trager & Somerville 2009). However, the difference is small, and the overall ages that we find for the LEGA-C galaxies are relatively consistent with the B19 predicted ages, even though we examine a different redshift range. Thus, the B19 relation is able to predict approximate quiescent galaxy ages down to $z \sim 0.6$.

⁷Díaz-García et al. (2019) derived similar relationships between stellar population parameters (including ages and metallicities) and UVJ colours. However, they examined photometric data only, in bands offset from U , V , and J . Thus, we examine only the relationship from Belli et al. (2019) here.

While we confirm that the B19 relation can be used in a general sense, we find that it may be too simplistic to derive precise ages for *individual* galaxies. In the right panel of Fig. 3.6, we plot our SSP-equivalent ages as a function of the $S_Q = 0.75(V - J) + 0.66(U - V)$ axis defined in B19 (see their equation (2)). We overplot the age-colour relation derived by B19 (dotted black line, see their equation (3)) as well as their individual data points (orange stars). As an illustrative comparison, we fit a similar linear relation to our galaxies on these axes, shown by the dashed grey line, with shaded regions indicating the 1σ uncertainty on our fit. Our fit differs from the age-colour relation in B19. In particular, we find a qualitatively shallower slope. Moreover, there is a significant amount of scatter that is not captured well by either fit, likely due in part to the fact that B19 have younger galaxies (ALF cannot fit galaxies younger than 1 Gyr, see Section 3.2.2), and because the relationship between UVJ colours and galaxy ages is not actually linear (see the curvature in the age trend in the top right panel of Fig. 3.2). Thus, we do not give the functional form of our linear fit as we do not believe that it accurately represents a quantitative relationship between UVJ colours and galaxy ages. We also apply multiple linear and polynomial regression algorithms to attempt to derive quantitative relationships between UVJ colours and our metallicities and ages (not shown). Specifically, we attempt to find the best linear or polynomial combination of $U - V$ and $V - J$ colours that minimizes the residuals in age and metallicity, similar to the S_Q axis defined in Belli et al. (2019). However, we find that no simple functional form fits the data well and captures the relationships between UVJ colours and ages and metallicities.

Accordingly, we suggest that the UVJ diagram can, as found previously, be used to identify whether a quiescent galaxy is relatively older or younger, but it may not be possible to specify a precise value of age based on an individual galaxy's UVJ colours. Additionally, we show that the UVJ diagram can also be used to estimate whether a quiescent galaxy is relatively more metal-rich or metal-poor, though the scatter in metallicity is larger than in age (and again, it is difficult to specify an exact metallicity based on UVJ colours alone, see also Leja et al. 2019c). Other fitting degeneracies discussed in Section 3.1 could be complicating these relationships, for example degeneracies between age, dust, and SFH (see Section 3.4.3 for further discussion). Looking forward, more advanced fitting techniques, such as random forest algorithms or neural networks, in addition to combining photometry and spectra and developing SPS models further, may be able to provide a quantitative translation between UVJ colours and ages and metallicities for individual galaxies.

3.4.2 Implications for galaxy evolution studies

In Section 3.3.2, we examined how trends between our spectroscopically measured ages and metallicities and independent photometric UVJ colours compare to predictions from popular SPS models. We found that, while the age trend is generally reproduced well by both FSPS and BC03, these models do not recover the metallicity trend. Here, we discuss the implications of our results for galaxy evolution studies in general.

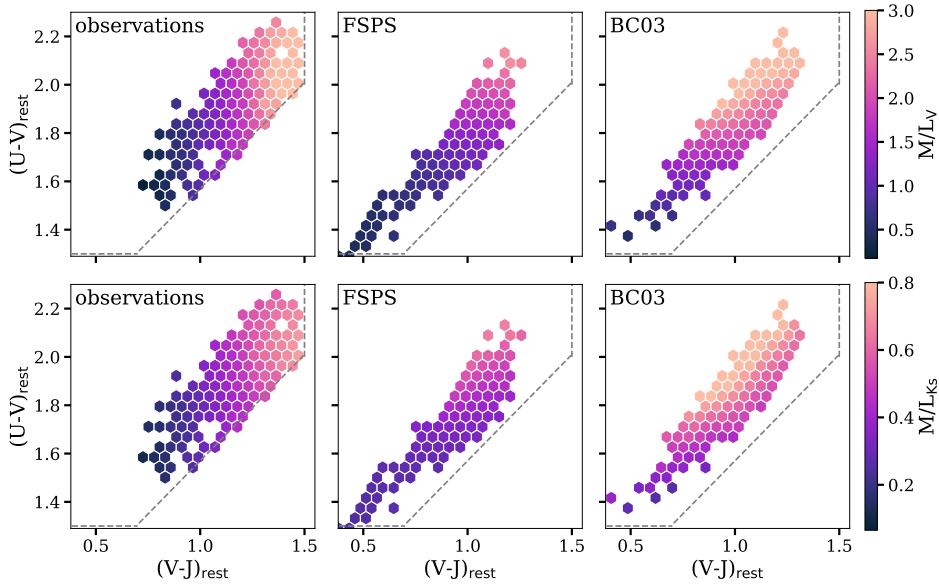


Figure 3.7: LOESS-smoothed rest-frame UVJ diagrams for the data and the models, where the diagrams are colour coded by the M/L ratio. The LEGA-C UVJ diagrams are shown in the left column, the FSPS diagrams are shown in the middle column, and the BC03 diagrams are shown in the right column. This is similar to Fig. 3.4, where we input the ages and metallicities of each galaxy in our LEGA-C sample into the models, and output the predicted rest-frame UVJ colours (from which we compute luminosities) and stellar masses. For the LEGA-C sample, we compute M/L ratios using independently measured, dynamically scaled virial masses from van der Wel et al. (2022). The top row is colour coded by M/L ratios in the V band (M/L_V) and the bottom row is colour coded by M/L ratios in the K_s band (M/L_{K_s}). While the M/L ratio increases approximately along the quiescent sequence in most cases, the trend curves in opposite directions for the data and the models at the reddest tip of the quiescent sequence. Additionally, the trends in the M/L ratios predicted by the two models are not consistent with each other.

SPS models have been used for decades to fit photometric and spectroscopic data of quiescent galaxies and interpret their formation and evolution. However, if these models cannot fully reproduce observations as we suggest here, we may need to re-evaluate certain aspects of our understanding of galaxy evolution. For example, while the age trends that we find in this work are qualitatively consistent with those predicted by the models, individual age measurements may not be reliable due to degeneracies with other fitting parameters (Worthey 1994; Bell & de Jong 2001; Papovich et al. 2001; Bruzual & Charlot 2003; Gallazzi et al. 2005; Leja et al. 2017, 2019a,b). In particular, an incorrect metallicity will be compensated by an incorrect estimation of age, dust, or SFH. Accordingly, derived metallicities, SFHs, and dust components may be systematically biased (Leja et al. 2017; Carnall et al. 2019; Leja et al. 2019a,b; Pacifici et al. 2023; Bellstedt & Robotham 2025). Thus, our work may have significant implications for our understanding of the

SFHs of massive galaxies in the early Universe as derived from SED modelling (e.g., Glazebrook et al. 2017, 2024; Carnall et al. 2023b, 2024; de Graaff et al. 2025; Bevacqua et al. 2025; Nanayakkara et al. 2025), or for why massive quiescent galaxies are still quite dusty (e.g., Gobat et al. 2018; Lee et al. 2024; Setton et al. 2024; Siegel et al. 2025).

Another set of key quantities that could be affected is the M/L ratios and therefore stellar masses. To explore this issue, we compare trends between M/L ratios in our data and those predicted by the models. It has been found previously that M/L ratios, similar to ages, have a tight relationship with rest-frame colours (e.g., Bell & de Jong 2001; van de Sande et al. 2015). Thus, in Fig. 3.7 we show a similar comparison between our observations and SPS models as in Fig. 3.4, but now colour coding the rest-frame UVJ diagrams with M/L ratios.

In the top row we show the V -band M/L ratios (M/L_V) and in the bottom row we show the K_s -band M/L ratios (M/L_{K_s}). We show the LEGA-C galaxies in the left panels, where we compute M/L ratios using dynamical-model-calibrated virial masses derived in van der Wel et al. (2022). We derive luminosities from the EAZY V - and K_s -band colours (redetermined as in Section 3.2, using the LEGA-C spectroscopic redshifts). The virial masses were scaled to dynamical masses obtained via Jeans modelling in van Houdt et al. (2021), and are defined as

$$M_{\text{vir}} = K(n)K(q)\frac{\sigma_{*,\text{int}}'^2 R_{\text{sma}}}{G} \quad (3.1)$$

(equation (9) in van der Wel et al. 2022), where $K(n) = 8.87 - 0.831n + 0.0241n^2$ and where n is the Sérsic index (Cappellari et al. 2013b), $K(q) = (0.87 + 0.38e^{-3.78(1-q)})^2$ and where q is the projected axis ratio $q = b/a$, $\sigma_{*,\text{int}}'^2$ is the spatially integrated stellar velocity second moment from LEGA-C, and R_{sma} is the semimajor axis of the ellipse that contains 50 per cent of the projected Sérsic light model from the best-fitting Jeans model.

We use dynamical masses instead of stellar masses as they are independent of SPS models (see also van de Sande et al. 2015), similar to our age and metallicity measurements in the left panels of Fig. 3.4. As a caveat, we note that they include both baryonic and dark matter contributions⁸. The dark matter component may systematically bias the M/L ratios to higher values; however, the qualitative trends along the UVJ diagram would be the same assuming a similar dark matter fraction within one R_e .

In the middle and right panels of Fig. 3.7, we show the stellar M/L ratios for the FSPS and BC03 models, respectively, derived for each galaxy using the assumed age and metallicity (with added random scatter for the UVJ colours as in Section 3.3.2) from which we compute luminosities. As an additional caveat, we note that our treatment of dust could affect the observed luminosities while the assumed IMF will affect the model M/L ratios. We discuss this further in Section 3.4.4.

Similar to the metallicity panels in Fig. 3.4, the qualitative trends between M/L ratios in the data and in the models are not consistent. In particular, while the

⁸As these are quiescent galaxies, we neglect the gas mass.

M/L ratio increases roughly along the quiescent sequence in most panels, the data and models curve in opposite directions at the reddest tip of the quiescent sequence. Moreover, in the bottom-right panel, the M/L_{K_s} ratio increases approximately perpendicular to the quiescent sequence in the BC03 models, while this is not the case for the data. Thus, the data-model inconsistencies that we find in Section 3.3.2 also apply to stellar mass measurements.

Our results are consistent with those of van de Sande et al. (2015), who compared M/L -colour relations to those predicted by SPS models for massive quiescent galaxies out to $z \sim 2$. They found that the BC03, FSPS, and Maraston & Strömbäck (2011) SPS models were not able to simultaneously reproduce the relationships that they found between dynamical g -band M/L ratios and $(g - z)$ rest-frame colours and dynamical K -band M/L ratios and $(g - K)$ rest-frame colour. Furthermore, Bellstedt & Robotham (2025) explored the impacts of different stellar libraries and model assumptions on SPS-model-derived stellar masses and found that the primary impact on stellar masses was in assuming a fixed, solar metallicity, followed by stellar library differences (see also Jones et al. 2022) and SFH parameterizations (see also Carnall et al. 2019; Leja et al. 2019a; Lower et al. 2020; Suess et al. 2022; Haskell et al. 2024).

Additionally, the middle and right panels in Fig. 3.7 demonstrate that there are significant differences in the M/L ratio trends between the models themselves. This was also indicated by van de Sande et al. (2015), who found that introducing a variable IMF improved the match between observed and predicted M/L -colour relations for certain models, but not all (see also Bellstedt & Robotham 2025). Similarly, Muzzin et al. (2009) found that different SPS models result in best-fitting stellar masses that differ by a factor of up to 1.5. Finally, Lee et al. (2025) found systematic differences in M/L ratios measured using four different SPS models (Vazdekis et al. 2016, BC03, Plat et al. 2019, and FSPS).

Given these findings, it would be informative to understand whether ages, metallicities, or other quantities are primarily driving these M/L ratio differences. To this end, we examine relationships between our measured ages and metallicities and M/L ratios (not shown), and find that there may be a correlation between M/L and age in our observations, albeit with large scatter. There is no clear relation between M/L and metallicity in our observed LEGA-C galaxies. Additionally, we examine the ratio between the observed and modelled M/L ratios (not shown). In this case, there is also no clear relation with age or metallicity, indicating that we do not observe any specific biases for specific properties. While this is unexpected given the inconsistencies between the models and the data in UVJ space, it is not surprising that this analysis does not yield any further insights as the observed M/L ratios derived from dynamical masses rely on many assumptions, have significant uncertainties, and include dark matter (van Houdt et al. 2021; van der Wel et al. 2022). Thus, understanding the driver of M/L ratio differences between the data and the models will benefit from a more in-depth investigation.

3.4.3 Implications for SPS models

In Section 3.3.2, we showed that popular SPS models such as FSPS and BC03 can broadly reproduce the age trend in the UVJ diagram, but cannot reproduce the $[\text{Fe}/\text{H}]$ trend. We also showed that these models fail to predict the observed M/L ratio trends in Section 3.4.2. In this section, we discuss possible explanations for these data-model inconsistencies.

First, the observed discrepancies could be attributed to differences in the models used to measure stellar population parameters from absorption lines (i.e., ALF) and the models used to produce the UVJ colours (i.e., FSPS, BC03). While both FSPS and ALF use the same MIST isochrones (Choi et al. 2016) and MILES stellar libraries (Sánchez-Blázquez et al. 2006), the major difference between the SPS models considered here and ALF is that ALF allows for a variable abundance pattern while FSPS and BC03 assume solar-scaled abundance patterns. If models with α -enhancements or variable abundance patterns predict different continuum shapes compared to models with solar-scaled abundance patterns, this could explain the differences between the observations and model predictions in Figs 3.4 and 3.7. For example, Choi et al. (2019) found that the variable abundance pattern implemented in ALF was necessary to simultaneously reproduce spectra and broad-band, rest-frame, optical colours (ugr). Vazdekis et al. (2015) found that α -enhancement and a bottom-heavy IMF are required to reproduce observed photometric colours. Several other groups have also introduced α -enhanced models (e.g., Coelho et al. 2007; Lee et al. 2009; Percival et al. 2009; Knowles et al. 2021, 2023; Byrne et al. 2025; Park et al. 2025). Thus, implementing more widespread use of non-solar abundance patterns, especially when studying α -enhanced stellar populations such as those in distant massive quiescent galaxies, may be a promising way forward in explaining the discrepancies in UVJ space. In fact, Park et al. (2025) recently⁹ presented α -MC, a new α -enhanced implementation of the FSPS models. In the future, it will be interesting to generate the models in Section 3.3.2 with α -MC to examine how α -enhancement affects the observed trends in stellar properties.

In this context, it is interesting to note that these different approaches (solar-scaled versus variable abundance patterns) lead to different fitting results. For example, Nersesian et al. (2024) computed ages and metallicities for the LEGA-C galaxies using PROSPECTOR (Johnson et al. 2021, which makes use of the FSPS models, Conroy et al. 2009; Conroy & Gunn 2010). We find that there is good agreement between the PROSPECTOR and ALF ages (albeit with a large amount of scatter), but poor agreement between the metallicities (not shown, A. Nersesian, private communication). This was also found in Beverage et al. (2025), who fit the spectra of massive quiescent galaxies at $1 < z < 3$ with a PYTHON version of ALF (ALF α , Beverage 2024; Beverage et al. 2025) and compared their ages and metallicities to PROSPECTOR results from Slob et al. (2024). Thus, implementing non-solar abundance patterns in standard fitting codes may improve the consistency in age and metallicity measurements between the different codes as well.

⁹At the time of writing, the α -enhanced FSPS models from Park et al. (2025) were not available as part of PYTHON-FSPS for us to test here.

Another potential reason for the discrepancies between the models and the data (from a modelling perspective) is the contribution of evolved stellar stages. Specifically, the TP-AGB phase is prominent at low metallicities for young galaxies (< 2 Gyr) and can contribute significantly to the M/L ratio (Walcher et al. 2011). The TP-AGB phase is implemented differently in every SPS model, and its exact contribution is debated (compare Maraston 2005, Maraston et al. 2006, and Lu et al. 2025, who found that a prominent TP-AGB phase reproduced observations the best, with Kriek et al. 2010 who found that a prominent TP-AGB phase overpredicted near-infrared luminosities, and see Bevacqua et al. 2025 as well). These differences can have varying impacts on SED fits. In particular, Muzzin et al. (2009) fit the near-infrared SEDs of young galaxies, where the effects of the TP-AGB should be most evident, with different SPS models, and found that the best-fitting stellar population parameters differed by a factor of up to 3. Thus, agreement on the implementation of TP-AGB stars and other evolved stellar stages is needed.

Other aspects of SPS models could help alleviate the discrepancy with observations. For instance, it will also be important to reconcile different spectral libraries, as they have been found to produce different results. In particular, Byrne & Stanway (2023) compared multiple spectral libraries using the same SPS model and found that the discrepancies with spectroscopic quantities when using different libraries were worse for metallicities than for ages, as the 4000 Å break was relatively insensitive to library differences. Similar results were found by Bellstedt & Robotham (2025). We note that ALF makes use of the C3K stellar library, which Byrne & Stanway (2023) found to be the most robust out of their tested libraries. Finally, different SPS models do not always agree on the definition of metallicity. For example, total metallicity is often computed using a combination of $[Mg/Fe]$ and $[Fe/H]$ (e.g. Thomas et al. 2003) when Mg and Fe are both measured individually. However, we note that the opacities are primarily driven by iron. Moreover, as discussed above, both FSPS and BC03 assume solar-scaled abundances, while massive quiescent galaxies are α -enhanced. In particular, Beverage et al. (2025) showed that the best-fitting FSPS metallicities follow $[Fe/H]$, and not $[Z/H]$, for galaxies that are α -enhanced (like the ones in this sample). See also Choi et al. (2014) and Leja et al. (2019a).

In summary, our results demonstrate that standard SPS models do not predict the correct UVJ colours for quiescent galaxies when given stellar metallicities or ages (even when taking into account dust and extended SFHs, see Section 3.4.4). Thus, we emphasize the need for caution when interpreting or assuming metallicities or stellar masses predicted from photometry, or derived via solar-abundance-scaled codes that rely on the shape of the continuum (see also Nersesian et al. 2024; Bellstedt & Robotham 2025; Beverage et al. 2025). More work is needed to implement non-solar abundance patterns (see e.g. Park et al. 2025), understand the contribution of continuum shapes and advanced stellar stages, and reconcile the ingredients in different SPS models.

3.4.4 Caveats

Here, we discuss the main caveats that should be taken into consideration when interpreting our results, including whether our observed trends may be driven by other physical properties or assumptions that we have made, and our use of ALF.

Throughout this work, we assume that there is no dust in the LEGA-C galaxies and we do not implement a dust law in our generation of FSPS and BC03 models. However, it is likely that at least some of the LEGA-C galaxies are affected by dust. In particular, the location of our data in the UVJ diagram in Figs 3.3, 3.4, and 3.7 is slightly redder than expected from the models, especially when we compare with FSPS, and dust could be contributing to this offset. The presence of dust would also introduce additional scatter in the SPS model panels (middle and right) in Figs 3.4 and 3.7. We test these possibilities by generating the same models as shown in Figs 3.4 and 3.7, but assuming a Calzetti et al. (2000) dust law with (i) FAST (Kriek et al. 2009, 2018) A_V values, measured by and obtained via private communication with the LEGA-C team for each galaxy (although these values will be heavily affected by the degeneracies and model inaccuracies that we discuss throughout the paper), and (ii) A_V values that scale with age (as dust approximately follows the age trend in the UVJ diagram, see the arrow in Fig. 3.3), such that $A_V = 0.5$ at 1 Gyr, decreasing down to an age of 3 Gyr, beyond which $A_V = 0$. We correct the model FSPS and BC03 luminosities for the respective dust extinction in each case (not shown). In all cases, dust moves the distribution along the arrow shown in Fig. 3.3, but does not provide a better match between the observations and the models. However, we note that implementing the FAST A_V values for each galaxy broadens the distribution of the models in UVJ space in the middle and right panels of Figs 3.4 and 3.7 to better match the observations. Dust also acts to increase the M/L ratios; however, it does not change their distribution in UVJ space, preserving the discrepancy between the observations and the models. While we find that dust does not drive the disagreement between the observations and the models, we note that correlations between galaxy properties could alter the distribution of galaxies in UVJ space. Thus, it is still possible that better A_V constraints could bring the observations and models into agreement. Further exploration of methods to disentangle dust from other physical properties is required.

Additionally, in our FSPS and BC03 model generation, we assume a relatively simple SFH with a short, truncated burst. We also assume a single age in our ALF fits. However, if multiple stellar populations are present (see e.g. Carnall et al. 2024; Nanayakkara et al. 2025), this could also play a role in the offset that we see between the data and the models. We test the effects of several different SFHs, by extending the burst, varying the τ value, and implementing a delayed- τ model. We find that altering the assumed SFH may help to explain some of the scatter that we see in the right-most panels of Figs 3.4 and 3.7, but does not account for all of it. Crucially, implementing different SFHs does not generate the trends that we observe in age, metallicity, or M/L ratios, with these parameters behaving the same way in these test models as with our assumed simple SFH.

There may be other effects creating the trends in stellar population parameters

that we see that are not inherent to the SPS models. For example, the age and metallicity measurements from which we derive our main results are based solely on our fits using ALF, which may not be as accurate for galaxies near the edge of the UVJ box; these galaxies could potentially be dusty post-starburst galaxies (Zick et al. 2018) or have low levels of star formation (Akins et al. 2022). Since our conclusion of a metallicity trend depends on this region of the UVJ diagram, we test shifting the UVJ box by 0.1 magnitudes toward redder $U - V$ colours to examine the effect of excluding potential contaminant galaxies near the edge of the UVJ box. We find that both the age and metallicity trends are still present and follow the same qualitative direction as in Fig. 3.2, although the age trend is slightly tighter. Thus, the inclusion of the galaxies towards the edge of the UVJ box does not affect our conclusions.

To quantitatively examine the potential biases of using ALF, we additionally test whether ALF could be predicting lower $[\text{Fe}/\text{H}]$ abundances for a quiescent population which had a recent minor starburst. In this case, the equivalent widths of absorption lines would be decreased by an additional, mostly featureless, continuum contribution. To test this scenario, we select 33 quiescent galaxies from LEGA-C that lie along the quiescent sequence, but away from the edge of the UVJ box. For each selected LEGA-C galaxy, we generate a young (0.1 Gyr) SED with FSPS that is 6 per cent of the luminosity of the corresponding LEGA-C galaxy in the V band, and add the young SED to the LEGA-C spectrum to make 33 composite spectra that still look visually quiescent. We determine this luminosity percentage by estimating the UVJ colours of the composite spectra such that the composite galaxies lie near the edge of the UVJ box. Note that we give the young SEDs the same metallicity as derived from the fiducial fit to the corresponding LEGA-C spectrum. We also test giving the young SEDs higher and lower metallicities and find that this does not affect our conclusions. We fit the composite spectra with ALF in the same way as the corresponding LEGA-C spectra. We find that including a young population on top of an older quiescent one results in $[\text{Fe}/\text{H}]$ values that are consistent with the $[\text{Fe}/\text{H}]$ values that we derive from the fiducial fits within uncertainties. However, we note that in many cases the fits are driven to slightly lower median $[\text{Fe}/\text{H}]$ values compared to the fiducial case. In particular, we find that an average decrease of ~ 0.24 in $U - V$ colour and ~ 0.04 in $V - J$ colour results in an average decrease in $[\text{Fe}/\text{H}]$ of ~ 0.05 dex. When we compare this to our expectations from a similar decrease in $U - V$ and $V - J$ in our observations in the lower right panel of Fig. 3.2, we see that the average observed decrease in $[\text{Fe}/\text{H}]$ is ~ 0.11 dex. Therefore, low levels of star formation may contribute to the observed metallicity trend, and in general we caution that composite stellar populations near the edge of the UVJ box may lead to lower metallicities. However, based on the test presented here, these low levels of star formation cannot fully explain the observed metallicity trend, and we note that such a recent burst would generate emission lines, which we do not observe in our spectra.

In the future, it will still be important to confirm our ALF fitting results via other means. Encouragingly, we note that our observed metallicity trend is consistent with the results of Nersesian et al. (2025), who simultaneously fit photometry

and spectroscopy in a large sample of star-forming and quiescent LEGA-C galaxies. Another possibility is to estimate ages and metallicities via traditional Lick indices (Worthey et al. 1994), which could potentially provide us with insight into whether our results are robust. However, these measurements would not have the same independence as our ALF measurements, as SPS models like FSPS or BC03 are required to interpret Lick indices. Additionally, this method will likely not be able to provide age and metallicity estimates to the same level of accuracy as ALF as it only makes use of a few individual features while ALF considers the full spectrum (see e.g. Kriek et al. 2016). Other methods of constraining ages and metallicities are beyond the scope of this work, but should be considered in the future.

Finally, we test whether the metallicity trend could be caused by distinct types of galaxies populating different parts of the UVJ diagram. For example, if a specific population with low metallicity is preferentially found towards the edge of the UVJ box, this could drive the observed trend. For quiescent galaxies, properties that are typically correlated with lower metallicity include lower velocity dispersion, lower mass, or higher redshift (e.g., Smith et al. 2009; Leethochawalit et al. 2019; Beverage et al. 2023, 2024, 2025). We do not find any preference of lower velocity dispersion or lower mass populations towards the edge of the UVJ box. There is a slight preference of lower metallicities for higher redshift galaxies, which are indeed closer to the edge of the UVJ box. It is still unclear why these galaxies fall in this region of the UVJ diagram; however, we note that our sample also contains fewer galaxies at these redshifts. Larger samples of high redshift galaxies will be required going forward to examine whether there is a true dependence on redshift. The only additional property that is preferentially found towards the edge of the box is slightly higher sSFR. However, at the same mass, we do not expect quiescent galaxies with higher sSFR to have lower metallicity. In fact, we would expect the opposite trend, as galaxies that quench at later times and have longer star-formation time-scales will have higher $[\text{Fe}/\text{H}]$ values (e.g., Beverage et al. 2021, 2024, 2025).

3.5 Summary and conclusions

The age–metallicity degeneracy (Worthey 1994; Bruzual & Charlot 2003; Gallazzi et al. 2005) poses a major challenge to interpreting photometric data of quiescent galaxies. Measuring metallicities has historically been challenging, especially beyond the local Universe, where low S/N data make it difficult to constrain weak metallicity-sensitive features. As a result, metallicity has typically been left unconstrained or fixed to solar values, leading to (systematic) uncertainties in stellar population studies. Consequently, the red colours of quiescent galaxies, for example in the well-known UVJ diagram, were fully attributed to age (e.g., Whitaker et al. 2012, 2013; Mendel et al. 2015; Belli et al. 2019; Díaz-García et al. 2019), and our knowledge of how metallicity affected the shapes of galaxy continua was completely dependent on solar-scaled models, which become increasingly uncertain at earlier times.

In this work we compared, for the first time, spectroscopically derived ages and metallicities (from continuum-normalized, absorption-line spectra) to independently measured broad-band colours of quiescent galaxies beyond the low-redshift Universe. In particular, we examined a sample of ~ 700 massive quiescent galaxies at $0.6 \lesssim z \lesssim 1.0$ from the LEGA-C survey (van der Wel et al. 2021). We fit the continuum-normalized, absorption-line spectra using ALF (Conroy & van Dokkum 2012a; Conroy et al. 2018) and recovered SSP-equivalent ages and $[\text{Fe}/\text{H}]$ values. We compared these measurements to UltraVISTA (McCracken et al. 2012) UVJ colours (Muzzin et al. 2013b) by colour coding the UVJ diagrams of these galaxies by age and metallicity. We also compared our observed trends between spectroscopic ages and metallicities and broad-band UVJ colours to predictions from popular SPS models including FSPS (Conroy et al. 2009; Conroy & Gunn 2010) and BC03 (Bruzual & Charlot 2003). Our findings can be summarized as follows:

- Galaxy age generally increases along the quiescent sequence in the UVJ diagram. This was implied in Whitaker et al. (2012, 2013), Mendel et al. (2015), Belli et al. (2019), and Díaz-García et al. (2019); however, for the first time, we detected this trend using age measurements that are independent of UVJ colours, while also accounting for stellar metallicity. We also found that there is a slight curvature to the trend, with the oldest galaxies populating the reddest end of the $U - V$ axis, but located slightly blueward on the $V - J$ axis. Thus, we concluded that a simple linear relation between UVJ colours and ages does not represent the age trend very well at these redshifts, in contrast to the findings of Belli et al. (2019) at $1.0 < z < 2.5$.
- Stellar metallicity increases approximately perpendicular to the quiescent sequence. *Thus, the sequence in metallicity does not follow the sequence in age*, in contrast to predictions from stellar population models (Bruzual & Charlot 2003; Conroy et al. 2009; Conroy & Gunn 2010). This finding indicates that the interpretation of the quiescent sequence as an age sequence (Whitaker et al. 2012, 2013; Mendel et al. 2015; Belli et al. 2019; Díaz-García et al. 2019) is robust.
- In general, SPS models are able to qualitatively reproduce our observed age trend, though the observed galaxies are on average redder than model predictions. This difference could be explained by some dust attenuation.
- However, the models predict that metallicity increases nearly perpendicular to the trend that we observe in the data. In other words, SPS models predict very different UVJ colours for a given metallicity than the colours that we observe.
- These data-model inconsistencies likely impact our understanding of stellar population properties of quiescent galaxies derived from photometric data, as an incorrect metallicity will be compensated by age, dust, or SFH, which may subsequently bias M/L ratios. To explore this further, we examined the trends between M/L ratios and the UVJ diagram measured in our observed

LEGA-C galaxies and predicted by FSPS and BC03. For our observations, we made use of independent M/L ratios derived from dynamically scaled virial masses. Similar to the metallicity trend, we found that the M/L ratio trends predicted by SPS models behave very differently from those which we observe. This difference indicates that stellar masses inferred from M/L ratios measured via SPS models may also suffer from model inaccuracies.

Our work suggests that, while state-of-the-art SPS models are powerful tools in deriving stellar population properties, further development is required to accurately recover the stellar metallicities, ages, and M/L ratios of quiescent galaxies. In particular, the implementation of non-solar abundance patterns and a better understanding of the contribution from evolved stellar stages may improve the mismatch between observations and SPS model predictions, as well as the discrepancies between different SPS models. The need to improve model predictions will only become more urgent as we continue to push the frontier of quiescent galaxy studies to higher redshifts with *JWST*.

Acknowledgments

We thank the anonymous referee for taking the time to give us useful feedback that improved this manuscript. We thank the LEGA-C team for making their data set public. We also thank Brian Lorenz, Angelos Nersesian, and Colin Yip for useful conversations. This work was performed using the compute resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. This work also used the Dutch national e-infrastructure with the support of the Samenwerkende Universitaire Rekenfaciliteiten (SURF) Cooperative using grant no. EINF-6344 and grant no. EINF-10017 which are financed by the Dutch Research Council (NWO). MK acknowledges funding from the NWO through the award of the Vici grant VI.C.222.047 and NSF AAG grant AST-1909942. PEMP acknowledges the support from the Dutch Research Council (NWO) through the Veni grant VI.Veni.222.364.

Data Availability

This study makes use of data from the LEGA-C survey. The 2D and reduced 1D spectra can be obtained from the ESO Science Archive Facility (http://archive.eso.org/eso/eso_archive_main.html). The reduced spectra and catalogue have been released by ESO (<http://archive.eso.org/cms/eso-archive-news/Third-and-final-release-of-the-Large-Early-Galaxy-Census-LEGA-C-Spectroscopic-Public-Survey-published.html>) and are also available here: <https://users.ugent.be/~avdrwel/research.html#legac>. For details, see van der Wel et al. (2016, 2021) and Straatman et al. (2018).

The re-determined UVJ colours as well as ages, metallicities, and $[Mg/Fe]$ values are available in an online catalogue. Other data products generated in the course of this work will be made available upon reasonable request.