

This is a self-archived – parallel-published version of an original article.

This article has been accepted for publication in Monthly notices of the royal astronomical society ©: 2023 .Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

AUTHOR Saracco P, La Barbera F, De Propris R, Bevacqua D, Marchesini D, De Lucia G, Fontanot F, Hirschmann M, Nonino M, Pasquali A, Spiniello C, Tortora C

TITLE The star formation history and the nature of the mass–metallicity relation of passive galaxies at $1.0 < z < 1.4$ from VANDELS

YEAR 2023

DOI <https://doi.org/10.1093/mnras/stad241>

VERSION Publisher’s PDF

CITATION P Saracco, F La Barbera, R De Propris, D Bevacqua, D Marchesini, G De Lucia, F Fontanot, M Hirschmann, M Nonino, A Pasquali, C Spiniello, C Tortora, The star formation history and the nature of the mass–metallicity relation of passive galaxies at $1.0 < z < 1.4$ from VANDELS, *Monthly Notices of the Royal Astronomical Society*, Volume 520, Issue 2, April 2023, Pages 3027–3048, <https://doi.org/10.1093/mnras/stad241>

The star formation history and the nature of the mass–metallicity relation of passive galaxies at $1.0 < z < 1.4$ from VANDELS

P. Saracco¹,^{1★} F. La Barbera,^{2★} R. De Propris^{1,3,4★} D. Bevacqua^{1,5} D. Marchesini,⁶ G. De Lucia^{1,7},
F. Fontanot^{1,7}, M. Hirschmann,^{7,8} M. Nonino,⁷ A. Pasquali^{1,9}, C. Spiniello^{1,2,10} and C. Tortora^{1,2}

¹INAF – Osservatorio Astronomico di Brera, via Brera 28, I-20121 Milano, Italy

²INAF – Osservatorio Astronomico di Capodimonte, sal. Moiariello 16, I-80131 Napoli, Italy

³FINCA, University of Turku, Vesilinnantie 5, FI-20014 Turku, Finland

⁴Department of Physics and Astronomy, Botswana International University of Science and Technology, Private Bag 16, Palapye, Plot 10071, Botswana

⁵Università degli studi dell’Insubria, via Valleggio 11, I-22100 Como, Italy

⁶Physics and Astronomy Department, Tufts University, 574 Boston Ave, Medford, MA 02155, USA

⁷INAF – Osservatorio Astronomico di Trieste, via G.B. Tiepolo 11, I-34143 Trieste, Italy

⁸Institute of Physics, Laboratory for Galaxy Evolution, Ecole Polytechnique Federale de Lausanne, Observatoire de Sauverny, Chemin Pegasi 51, CH-1290 Versoix, Switzerland

⁹Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstrasse 12 - 14, D-69120 Heidelberg, Germany

¹⁰Sub-Department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

Accepted 2023 January 18. Received 2023 January 11; in original form 2022 November 4

ABSTRACT

We derived stellar ages and metallicities $[Z/H]$ for ~ 70 passive early-type galaxies (ETGs) selected from VANDELS survey over the redshift range $1.0 < z < 1.4$ and stellar mass range $10 < \log(M_*/M_\odot) < 11.6$. We find significant systematics in their estimates depending on models and wavelength ranges considered. Using the full-spectrum fitting technique, we find that both $[Z/H]$ and age increase with mass as for local ETGs. Age and metallicity sensitive spectral indices independently confirm these trends. According to EMILES models, for 67 per cent of the galaxies we find $[Z/H] > 0.0$, a percentage which rises to ~ 90 per cent for $\log(M_*/M_\odot) > 11$ where the mean metallicity is $[Z/H] = 0.17 \pm 0.1$. A comparison with homogeneous measurements at similar and lower redshift does not show any metallicity evolution over the redshift range $0.0 < z < 1.4$. The derived star formation (SF) histories show that the stellar mass fraction formed at early epoch increases with the mass of the galaxy. Galaxies with $\log(M_*/M_\odot) > 11.0$ host stellar populations with $[Z/H] > 0.05$, formed over short time-scales ($\Delta t_{50} < 1$ Gyr) at early epochs ($t_{\text{form}} < 2$ Gyr), implying high star formation rates ($\text{SFR} > 100 M_\odot \text{yr}^{-1}$) in high-mass density regions ($\log(\Sigma_{1\text{kpc}}) > 10 M_\odot/\text{kpc}^2$). This sharp picture tends to blur at lower masses: $\log(M_*/M_\odot) \sim 10.6$ galaxies can host either old stars with $[Z/H] < 0.0$ or younger stars with $[Z/H] > 0.0$, depending on the duration (Δt_{50}) of the SF. The relations between galaxy mass, age, and metallicities are therefore largely set up *ab initio* as part of the galaxy formation process. Mass, SFR, and SF time-scale all contribute to shape up the stellar mass–metallicity relation with the mass that modulates metals retention.

Key words: galaxies: elliptical and lenticular, cD – galaxies: evolution – galaxies: formation – galaxies: high-redshift.

1 INTRODUCTION

Early-type galaxies (hereafter ETGs) are known to obey a series of ‘scaling relations’ with small intrinsic scatter, most of them being a projection of a thin distribution in the multidimensional fundamental plane (FP; e.g. Djorgovski & Davis 1987; Dressler et al. 1987; Jørgensen, Franx & Kjaergaard 1996). In particular, the colour–magnitude (CM) relation (Sandage & Visvanathan 1978; Bower, Lucey & Ellis 1992) is observed to persist, with largely unchanged slope and scatter, except for the effects of stellar aging, even at $z = 2$ (Blakeslee et al. 2003; Mei et al. 2006a, b, 2012; Newman et al. 2014; Lemaux et al. 2019; Willis et al. 2020). This is conventionally

interpreted as a relation between mass and metallicity, with the intrinsic scatter being provided by a spread in age (Kodama & Arimoto 1997; Gallazzi et al. 2006). Spectroscopy confirms that the CM relation is driven by a mass–metallicity trend out to $z = 1.2$ at least (e.g. Jørgensen et al. 2017; Lemaux et al. 2019; Saracco et al. 2019). However, interpretation of galaxy colours and their evolution in terms of age and metallicity is hampered by the well-known degeneracies between these two quantities (e.g. Worthey 1994).

The simplest interpretation is one where the CM relation is established by an intense burst of star formation (SF) at high redshift (e.g. Pipino & Matteucci 2004; Dekel et al. 2009) where most of the stellar mass is formed *in situ* and the depth of the potential well determines the epoch at which stellar winds and supernova explosions manage to eject the remaining gas from the forming protogalaxy. However, it is not necessary to form stars *in situ* to have older stars in more massive galaxies. In fact, older stars can also

* E-mail: paolo.saracco@inaf.it (PS); francesco.labarbera@inaf.it (FLB); rodepr@utu.fi (RDP)

be found in more massive galaxies when later assembled through merging (e.g. De Lucia, Kauffmann & White 2004). Moreover, integrated star formation histories (SFHs) for stars belonging to massive galaxies can also produce the correct scaling relation in models where mergers are significant in the stellar mass assembly (see e.g. fig. 7 in Fontanot et al. 2017).

The age– and stellar metallicity–mass relations are well studied in the local Universe (e.g. Gallazzi et al. 2005; Thomas et al. 2005, 2010; Choi et al. 2014; McDermid et al. 2015). These studies established that most massive ETGs host, on average, older and more metal-rich stellar populations than lower mass galaxies, a result which is very counterintuitive, even in the hypothesis of mergers, as time is needed to produce metals.

The metal content of ETGs and its evolution across time provides information about their past SF activity, their quenching phase, and their evolution. The total metallicity is the result of the duration of the SF and the gas exchange with the inter/circumgalactic medium. For instance, we expect that metallicities will be low and show no evolution with redshift if gas is quickly removed by an outflow [e.g. active galactic nucleus (AGN)-driven]. Indeed, if star formation is interrupted at early times due to the sudden removal of the gas instead of being smoothly quenched at later times, the total metallicity will be lower (e.g. De Lucia, Fontanot & Hirschmann 2017; Okamoto et al. 2017; Trussler et al. 2020). On the contrary, if the external gas supply (inflow) is stopped or suppressed by feedback mechanisms (e.g. AGN or SF feedback), or gas infall is negligible with respect to the rate at which gas is converted into stars (star formation rate, SFR), the system resembles a closed box and metallicity increases rapidly to its maximum value (e.g. Vazdekis et al. 1997; Peng et al. 2015).

Since ETGs are seen to be quiescent, it is expected that after the formation of the bulk of the galaxy stellar mass, their stellar population properties do not change significantly because of *in situ* SF. If one then observes evolution in the metallicity of passive ETGs with redshift, it must be due to progenitor bias (van Dokkum et al. 2008; Carollo et al. 2013) or mergers adding stellar populations to the mixture. Minor mergers are expected to lower the metallicity over time (if the mass accreted is significant) since low-mass galaxies have lower metallicities; major mergers to leave the metallicity unchanged, since equal mass galaxies are expected to have similar metallicity. However, in the simulations, mergers (see e.g. Mo, van den Bosch & White 2010) require some fine-tuning to match the scatter observed in the scaling relations (e.g. Nipoti et al. 2009; Skelton, Bell & Somerville 2012). In general, if galaxies undergo numerous mergers, the scatter in any pre-existing relation between mass and metallicity is likely to increase, if it is even able to survive, as most mergers will take place between random objects. In case of progenitor bias, the evolution (if any) depends on the formation and quenching mechanisms mentioned above.

A better understanding of the star formation history (SFH) and the mass assembly history of ETGs can be achieved by studying the stellar population properties (age and metallicity) and their relationship with mass and other properties of galaxies at increasingly higher redshift. Gallazzi et al. (2014) find relationships of increasing age and metallicity with the galaxy mass at $z \sim 0.7$ similar to those for ETGs in the local Universe, consistently with the results of Choi et al. (2014). Similar non-evolving trends between stellar metallicity and mass up to $z \sim 1.0$ are found by Ferreras et al. (2009). On the other hand, Beverage et al. (2021), using LEGA-C data (van der Wel et al. 2016, 2021), find that galaxies at $z \sim 0.7$ have metallicities 0.2 dex lower than their local counterparts, and that older galaxies have lower metallicities than younger ones.

Stellar metallicity measurements at $z \geq 1$ have been carried out for few massive or stacked galaxies (e.g. Lonoce et al. 2014; Onodera et al. 2015; Kriek et al. 2019; Saracco et al. 2019) and a handful of galaxies at even higher redshifts ($z \sim 2.1$, Kriek et al. 2016; $z \sim 3.35$, Saracco et al. 2020b). Recently, Carnall et al. (2019, 2022) used VANDELS data (McLure et al. 2018; Pentericci et al. 2018) to study the metallicity of passive galaxies at $z \sim 1.2$ and its evolution. We will discuss and compare their results in Sections 4 and 8.

Here, we present the study of a sample of field early-type and passive galaxies at $1.0 < z < 1.4$ selected from the VANDELS survey data. We describe the data set in Section 2, where we also give details on our selection procedures and adopted stellar population models. In Section 3, we describe the method used to derive the stellar age and metallicity of galaxies, and we study the dependence on models and fitting assumptions. In Section 4, we derive the stellar mass–metallicity relation (MZR) at $z \sim 1.2$, study its evolution down to $z \sim 0$ and we compare our results with the literature. In Section 5, we derive the SFH of VANDELS galaxies. In Section 6, we study the relationships between SFH and metallicity. Section 7 summarizes the results which are then discussed in Section 8 where we present also our conclusions. In Appendix C, we describe the procedure used to measure absorption line spectral indices in the rest-frame wavelength range [2600–4350] Å, and report them for the whole sample. Moreover, we use indices to test the results in a nearly model-independent way and mid-UV indices to constrain chemical abundances.

Throughout this paper, we use a cosmology with $H_0 = 70$ km s^{-1} Mpc $^{-1}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$ and assume a Chabrier (2003) initial stellar mass function (IMF). Magnitudes are in the AB system, unless otherwise specified.

2 DATA AND MODELS

2.1 VANDELS observations and data

A complete and detailed description of the VANDELS survey, spectroscopic observations, data reduction, and quality of the data can be found in McLure et al. (2018), Pentericci et al. (2018), and Garilli et al. (2021). Fully reduced spectra have been made publicly available by the VANDELS team (see Data Availability statement). Here, we provide only a brief summary of the relevant points.

VANDELS is a VIMOS/VLT deep public spectroscopic survey in the wavelength range [4800–9800] Å, of galaxies selected in the Hubble Space Telescope CANDELS and UDS fields. The survey aims to detect star-forming galaxies at high-redshift and passive galaxies at $1.0 < z < 2.5$ with spectra of sufficient quality and resolution ($R \sim 600$, FWHM ~ 15 Å at 9000 Å, 1 arcsec of slit width) to determine not only the redshift but also some stellar population parameters.

Spectra have an average dispersion of 2.5 Å pixel $^{-1}$. The pixel scale of the images is 0.205 arcsec pixel $^{-1}$. The seeing, as measured on the science images, was below 1 arcsec in ~ 90 per cent of the observations with a median value of ~ 0.7 arcsec (Garilli et al. 2021), corresponding to a spatial scale of ~ 6 kpc at $z \sim 1.2$. Therefore, any possible radial variation of stellar population properties on angular scale lower than 0.7–1 arcsec, i.e. lower than 6–8 kpc at $z = 1.2$, cannot be seen. It is worth noting that this angular scale is larger than the angular diameter ($2R_e$) of 95 per cent of our selected passive galaxies (see below) whose median value is $2R_e \sim 3.5$ kpc.

1D spectra were extracted applying the Horne optimal extraction algorithm (Horne 1986) which delivers the maximum possible signal-to-noise ratio for each spectrum. This implies that spectra have been extracted within a variable aperture. However, being the observations

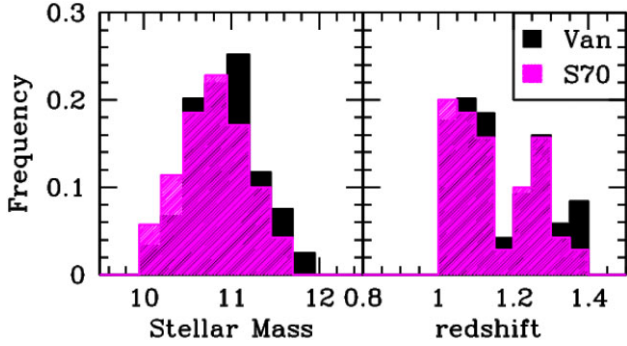


Figure 1. Stellar mass and redshift distributions of the 70 galaxies of our sample (shaded magenta histogram) are compared to the distributions of the parent VANDELS sample of passive galaxies (black histogram) in the redshift range $1.0 < z < 1.4$.

seeing limited, for the reasons discussed above, the optimal extraction algorithm is not expected to introduce any systematic effect in the estimate of stellar population properties.

2.2 Sample selection

The 70 passive galaxies studied here were extracted from the sample of 268 UVJ-selected passive galaxies at $1.0 < z_{\text{phot}} < 2.5$ targeted by the VANDELS survey (see McLure et al. 2018; Pentericci et al. 2018, for a detailed description of the sample selection and spectroscopic observations) according to the following criteria. We first selected all the passive galaxies (119) with spectroscopic redshift in the range $1.0 < z_{\text{spec}} < 1.4$ and reliability redshift flag ≥ 3 (i.e. probability of z_{spec} to be correct > 80 per cent). The redshift selection allows us to sample the rest-frame wavelength range $[2600\text{--}4200]$ Å for all the selected galaxies, and $[2600\text{--}4350]$ Å for those at $z_{\text{spec}} < 1.3$. In this wavelength range, the main mid-UV indices (e.g. Mg II($\lambda 2800$), Mg I($\lambda 2852$), Fe I($\lambda 3000$)) and a number of optical (e.g. CN3883, Ca IIH&K, D4000, H δ , Ca4227, *G* band) spectral features lie. Finally, on the basis of previous experience (Saracco et al. 2019), we selected galaxies whose spectrum has an S/N > 6 per Å over the rest-frame range $[3400\text{--}3600]$ Å to assure an average accuracy on spectral indices of ~ 15 per cent and reliable stellar population properties, and no truncation (due to technical problems) over the range $[3350\text{--}4350]$ Å to allow a reliable fitting of the spectrum. After this cleaning we remained with a sample of 70 galaxies in the mass range $10.0 < \log(M_*/M_\odot) < 11.7$. Fig. 1 compares the stellar mass and the redshift distributions of the 70 passive galaxies (magenta shaded histogram) with the distributions of the parent VANDELS sample of passive galaxies in the same redshift range (black histogram).

We morphologically classified galaxies as early-type (ETG), and late-type (LTG, spiral S and irregular I) by inspecting the *HST* images for the 50 galaxies covered by *HST* observations. The classification results into 18 LTGs (six irregulars and 12 spirals), 32 ETGs and 20 unclassified (for which we thus expect ~ 13 ETGs and 7 LTGs). Among the 18 LTGs, six galaxies were found to be a superposition or a merger of two galaxies. These six galaxies were not included in the analysis (even if the fitting to their spectra was performed, see below), resulting in a final sample of 64 passive galaxies, of which ~ 70 per cent are ETGs, as seen also in other samples covering similar mass range (see e.g. Tamburri et al. 2014). We do not find a dependence of the fraction of LTGs on the mass of the galaxies. Hereafter, we refer to the sample of 64 galaxies as passive galaxies.

2.3 Stellar population models

In this analysis, we adopt as reference the UV-extended EMILES simple stellar population (SSP) models (Vazdekis et al. 2015, 2016), based on BaSTI isochrones (Pietrinferni et al. 2004) and Chabrier (2003) stellar IMF (see e.g. Ge et al. 2019, for a comparison among models and different IMFs in full spectral fitting). We considered ages in the range $[0.1; 5.0]$ Gyr, and metallicity $[Z/H]$ in the range $[-2.27; 0.26]$.¹ These models have a full width at half-maximum (FWHM) spectral resolution of 3 Å at $\lambda < 3540$ Å and 2.5 Å in the optical domain (Vazdekis et al. 2016), higher than the rest-frame resolution ($\sim 6\text{--}7$ Å) of the VANDELS spectra. We adopted, as reference, the set of EMILES ‘base’ models for which it is assumed that $[\text{Fe}/\text{H}] = [Z/H]$, although this is only true for $[Z/H] \geq 0.0$. For low metallicities the input stars are α -enhanced. Therefore, when a model with total metallicity $[Z/H]$ is selected, its $[\text{Fe}/\text{H}]$ is lower (see Vazdekis et al. 2016, for a detailed description).

3 STELLAR AGE AND METALLICITY ESTIMATES

Stellar metallicities and ages for the whole sample of passive galaxies at $1.0 < z < 1.4$ were derived through non-parametric full-spectrum fitting (npFSF) performed over the rest-frame wavelength range $[3350\text{--}4350]$ Å. We limited the fit to 3350 Å to exclude the UV regime where observations for local galaxies are not available (as this wavelength range is not accessible from the ground). This allows us to perform a consistent comparison down to $z \sim 0$ (see Section 4.2). We used the STARLIGHT code (Cid Fernandes et al. 2005, 2007) and EMILES models. This code performs a fit by linearly combining SSPs with different ages Age_i and metallicities Z_i , each one contributing with a different weight to the light and to the stellar mass. Light-weighted (L) and mass-weighted (M) age $\text{Age}_{L,M}$ and metallicity $[Z/H]_{L,M}$ are thus defined according to the relations (e.g. Asari et al. 2007)

$$\text{Age}_{L,M} = \sum_i w_i(L, M) \text{Age}_i \quad (1)$$

and

$$[Z/H]_{L,M} = \log \sum_i w_i(L, M) Z_i / Z_\odot, \quad (2)$$

where $w_i(L, M)$ are the light- and mass-weights. An advantage of this fitting approach compared to the parametric FSF is that no a priori assumption on the SFH is done, assumption which might significantly affect the resulting duration of the star-formation, as well as the inferred age and metallicity.

Before studying the relations among the stellar properties, we first assess their dependence on the stellar population models and on the main assumptions considered in the fitting, and the difference between luminosity-weighted and mass-weighted quantities. This analysis (presented in Sections 3.1 and 3.2) was performed on stacked spectra having higher signal-to-noise than individual spectra to minimize the uncertainties.

¹Models with $[Z/H] = 0.4$, although available, were not used in this analysis given the lower quality than the other models (see Vazdekis et al. 2015, for details).

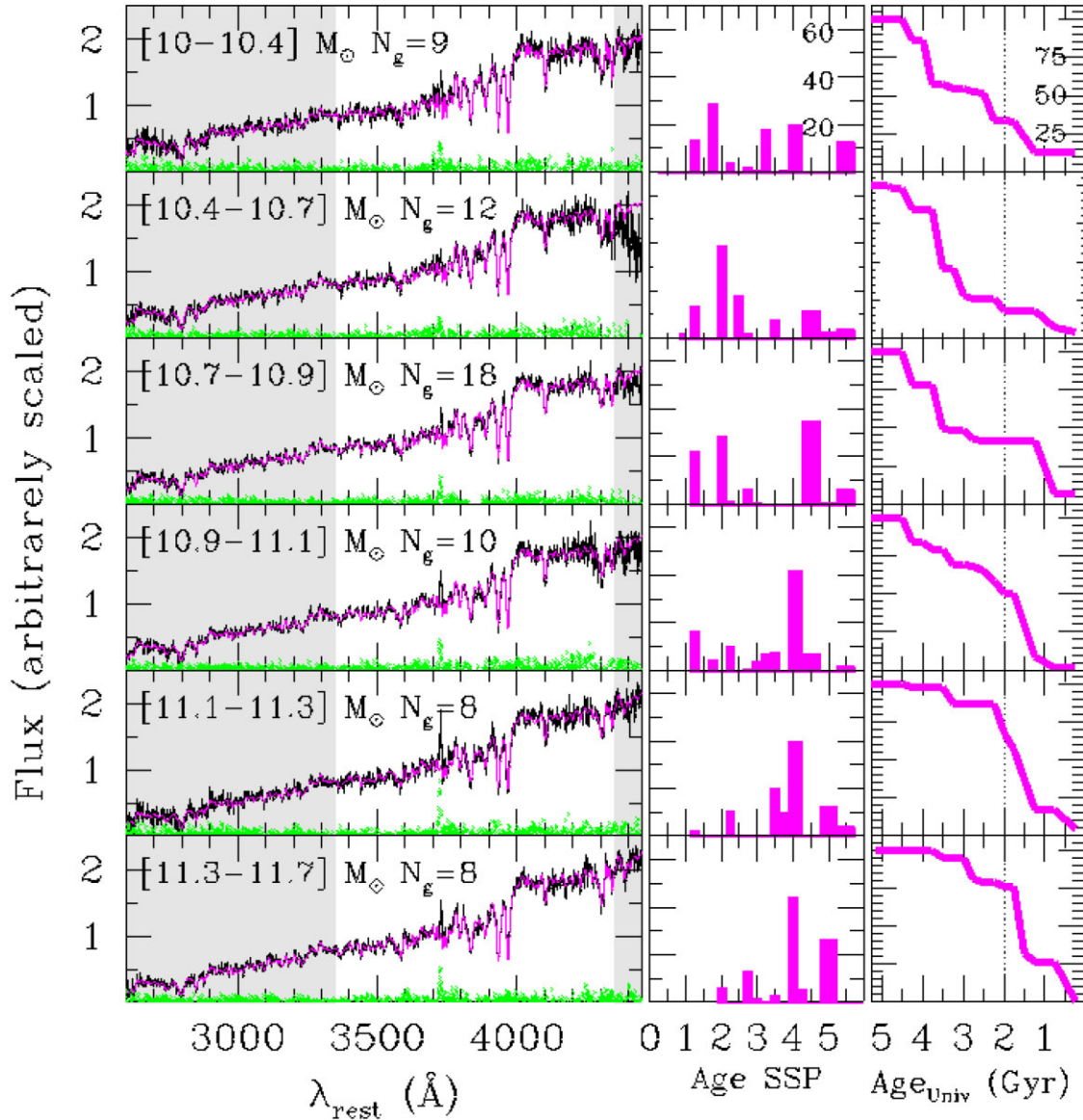


Figure 2. Left-hand panels: Stacked spectra (black curves) of galaxies in six stellar mass ranges. Magenta curve is the best-fitting composite model resulting from full-spectrum fitting performed with STARLIGHT over the wavelength range 3350–4350 Å with EMILES models (see Section 3). Residuals are shown in green. In each panel, the mass range (in logarithmic form) and the number of stacked spectra are shown. The grey shaded regions mark the wavelength ranges masked in the fit: the UV wavelength range inaccessible from the ground for galaxies at $z < 0.4$ and $\lambda > 4350$ Å sampled only for few galaxies. Central panels: Star formation history of galaxies. Fraction of stellar mass associated with each SSP contributing to the best-fitting composite model as a function of age. The numbers inside the upper panel mark the fractions 20 per cent, 40 per cent, and 60 per cent respectively. Right-hand panels: Cumulative SFH showing the growth of stellar mass as a function of the age of the cosmic time. To go from the ages of the SSPs to the epoch of formation t_{form} (see Section 5.1) we considered $\langle z \rangle = 1.2$ ($\text{Age}_{\text{Univ}}(z = 1.2) \simeq 5.05$ Gyr) as mean redshift of the stacks. The numbers inside the upper panel mark 25 per cent, 50 per cent, and 75 per cent of the total mass, respectively. The dotted line marks 2 Gyr ($z \simeq 3.0$) just as reference.

3.1 Stacked spectra

We divided the sample of 64 VANDELS passive galaxies in six mass ranges between 10^{10} and $10^{11.7} M_{\odot}$ to derive a single stacked spectrum representative of galaxies in each mass range. Mass ranges were chosen to have a minimum of eight galaxies in each of them. For the stacking, each spectrum was first shifted to the rest-frame, then normalized to the mean flux measured in the rest-frame wavelength range 3350–3550 Å, flat and free from significant features, and finally resampled to a common dispersion of 1 \AA pixel^{-1} . The spectra were then median stacked and the uncertainties were calculated using the median absolute deviation (MAD) estimator. Fig. 2 shows the

stacked spectra corresponding to the six mass bins. The resulting signal-to-noise ratios are in the range $\text{SNR} \simeq 18\text{--}25 \text{ \AA}^{-1}$.

The stacked spectra of galaxies with mass $\log(M_*/M_{\odot}) > 11.0$ show weak [O II]($\lambda 3727$) emission lines, not detectable in the individual spectra (see also Maseda et al. 2021, for O II emission in LEGA-C passive galaxies). The measured flux² of the strongest [O II] emission is $F(\text{O II}) = 6.1 \pm 0.6 \times 10^{-18} \text{ erg cm}^{-2} \text{ s}^{-1}$ associated

²Flux was estimated by fitting a Gaussian function to the line after having removed the underlying continuum evaluated through a polynomial fitting of the regions adjacent to the line.

with the stack of galaxies in the mass range $[11.1-11.3] M_{\odot}$. This flux corresponds to a luminosity $L(\text{O II}) = 5.0 \pm 0.5 \times 10^{40} \text{ erg s}^{-1}$ at the mean redshift $z = 1.2$. Assuming that this emission is due only to SF, and that the local relation $\text{SFR}(M_{\odot} \text{ yr}^{-1}) = (1.4 \pm 0.4) \times 10^{-41} L(\text{O II})$ (Kennicutt 1998) is valid also at higher redshift, we derive an $\text{SFR} = 0.7 \pm 0.3 M_{\odot} \text{ yr}^{-1}$. Therefore, the current residual star formation for galaxies with $\log(M_*/M_{\odot}) > 11.0$ of our sample is, on average, lower than $\text{SFR} \sim 1.0 M_{\odot} \text{ yr}^{-1}$ and decreases towards lower masses. We note that no differences are obtained in the full spectral fitting by masking or not the spectral regions with the above emission lines.

3.2 Dependence on models, fitting, and definitions

We performed npFSF to the stacked spectra shown in the left-hand panels of Fig. 2 using three different sets of models and two fitting codes to assess and quantify the dependence of the results on these basic fitting assumptions. We underline the fact that the following analysis does not want to be an exhaustive comparison between the SSP models in the literature nor the spectral fitting codes, but is simply aimed at verifying if the results can depend on their choice. Besides the EMILES models, we considered the 2016 updated version of the Bruzual & Charlot (2003) models (CB16 hereafter) with a Chabrier IMF, ages in the range $[0.1; 5.0]$ Gyr and metallicity in the range $[-2.3; 0.4]$, and the Maraston & Strömbäck (2011) models (M11 hereafter) with Chabrier IMF, ages in the range $[0.1; 5.0]$ Gyr and metallicity in the range $[-1.3; 0.3]$. These three sets of models are based on the same MILES spectral stellar library in the optical \AA (i.e. $\lambda > 3500 \text{\AA}$; Falcón-Barroso et al. 2011). The exclusion of the UV spectral range ($\lambda < 3350 \text{\AA}$) from the fit allows us, among other things, a fair comparison between these different SSP models.

As an alternative to the STARLIGHT code, we also considered the PPXF code (Cappellari 2017). Both codes perform the npFSF by linearly combining SSPs extracted from the same base of spectral templates. The best-fitting composite model is found by χ^2 minimization. The main differences between the two codes are that PPXF performs a regularization of the solutions (see Cappellari 2017) allowing to explore their degeneracy while STARLIGHT does not, and that this latter explores the composite models using a Markov Chain Monte Carlo algorithm (see Ge et al. 2018, for a comparison of the performances of the two codes).

In Fig. 3, the stellar metallicity and age resulting from the fitting of the stacked spectra are shown as a function of stellar mass for the two fitting codes and the different models. For E-MILES models, the two fitting codes provide consistent values both for metallicity and ages. We notice, however, that luminosity-weighted ages from PPXF are systematically younger than those from STARLIGHT. The resulting values are summarized in Table 1.

On the contrary, significant systematics (of the order of $\sim 0.2-0.4$ dex) are present among the values obtained with different models. In particular, the metallicity $[Z/H]$ (Fig. 3, upper panel) obtained for M11 models is systematically lower than the metallicity obtained with EMILES and CB16 models, being on average, $[Z/H]_{L, \text{EMILES}} = [Z/H]_{L, \text{M11}} + 0.17 \pm 0.08$ for the luminosity-weighted values, offset that reduces to $+0.11 \pm 0.08$ considering the different solar metallicity of models.³ The discrepancy between CB16 and M11

³Note that solar metallicity of EMILES models based on BASTI isochrones (Pietrinferni et al. 2004) differs by $+0.06$ dex with respect to models based on PADOVA isochrones (Girardi et al. 2000), such as CB16 and M11 (see Vazdekis et al. 2015).

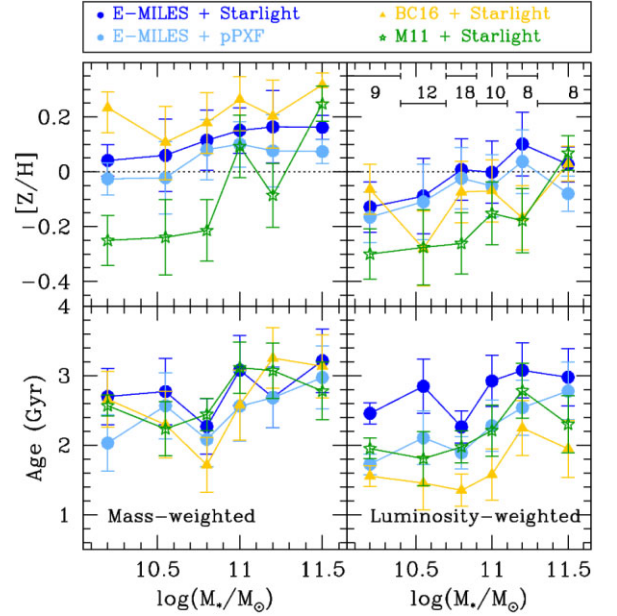


Figure 3. Metallicity (top) and age (bottom) estimates as a function of stellar mass. Left-hand and right-hand panels show mass- and luminosity-weighted quantities, respectively, resulting from full-spectrum fitting of stacked spectra of passive galaxies in different mass ranges (Fig. 2) for different stellar population models and fitting codes. Blue and light blue filled circles are the results obtained with EMILES SSPs using STARLIGHT and PPXF fitting codes, respectively; yellow triangles and green stars represent STARLIGHT fitting with CB16 and M11 models, respectively (see note 3 for the ± 0.06 dex offset between models). The number of galaxies contributing to the stack in each bin is shown on top of right-hand panel together with the width of the bins.

is even higher ($\Delta[Z/H] \sim 0.3$) when mass-weighted metallicity is considered. For the age, the differences among the mass-weighted values obtained with the different models are not significant. However, a significant systematic is found for CB16 models when the luminosity-weighted stellar age is considered, being on average $\text{Age}_{L, \text{EMILES}} = 1.5(\pm 0.08) \times \text{Age}_{L, \text{CB16}}$. Therefore, except for the mass-weighted age, which all models return consistent values, luminosity-weighted age and metallicity values depends on the models assumed in the analysis.

A well known, but important systematic difference exists between luminosity- and mass-weighted values independently of the models considered, as shown in Fig. 3 (see also Barone et al. 2020). As for the metallicity, the difference is particularly significant being, on average, $[Z/H]_{M*} \simeq [Z/H]_L + 0.2(\pm 0.05)$, while it is less important for the age, $\text{Age}_{M*} = 1.05(\pm 0.08) \times \text{Age}_L$. The systematic difference is due to the different M/L values of SSPs seen at different age and metallicity. Therefore, as expected, considering luminosity-weighted or mass-weighted values implies significant systematic differences.

Finally, the spectral range considered in the fitting can affect the resulting stellar population properties. This is shown in Fig. 4 where full-spectrum fitting of stacked spectra was performed with STARLIGHT and EMILES models over two wavelength ranges, $2600 \text{\AA} < \lambda_{\text{rest}} < 4350 \text{\AA}$ and $3350 \text{\AA} < \lambda_{\text{rest}} < 4350 \text{\AA}$. Also in this case, the largest systematic is seen for the metallicity. The inclusion in the fitting of the UV wavelength range $2600-3350 \text{\AA}$, typically missed for ground-based observations of galaxies at $z < 0.4$, results in metallicities systematically lower (by ~ 0.15 dex) than those obtained by fitting the range $\lambda_{\text{rest}} > 3350 \text{\AA}$. This is true both for luminosity-

Table 1. Mean stellar age and metallicity of galaxies in different ranges of stellar mass as resulting from the full-spectrum fitting^a of the stacked spectra performed with STARLIGHT (S) and PPXF (p).

$\text{Log}(M_*)$ (M_\odot)	Age_L (Gyr)	$[\text{Z}/\text{H}]_L$	Age_{M^*} (Gyr)	$[\text{Z}/\text{H}]_{M^*}$	A_V^b (mag)	Code
10.20	2.4 ± 0.2	-0.13 ± 0.09	2.7 ± 0.4	0.05 ± 0.06	0.00	S
"	1.7 ± 0.2	-0.17 ± 0.09	2.0 ± 0.4	-0.04 ± 0.06	p
10.55	2.8 ± 0.4	-0.09 ± 0.10	2.7 ± 0.5	0.07 ± 0.10	0.16	S
"	2.1 ± 0.4	-0.11 ± 0.10	2.6 ± 0.5	-0.03 ± 0.10	p
10.80	2.3 ± 0.2	0.01 ± 0.10	2.3 ± 0.4	0.12 ± 0.10	0.00	S
"	1.9 ± 0.2	-0.02 ± 0.10	2.1 ± 0.4	0.07 ± 0.10	p
11.00	2.9 ± 0.4	0.00 ± 0.10	3.1 ± 0.5	0.16 ± 0.08	0.00	S
"	2.3 ± 0.4	-0.05 ± 0.10	2.6 ± 0.5	0.09 ± 0.08	p
11.20	3.1 ± 0.4	0.10 ± 0.10	2.7 ± 0.4	0.17 ± 0.10	0.24	S
"	2.5 ± 0.4	0.04 ± 0.10	2.7 ± 0.4	0.07 ± 0.10	p
11.50	3.0 ± 0.4	0.03 ± 0.06	3.2 ± 0.5	0.17 ± 0.04	0.33	S
"	2.8 ± 0.4	-0.08 ± 0.06	3.0 ± 0.5	0.06 ± 0.04	p

Note. ^aThe fitting was performed with EMILES models. Luminosity-weighted and mass-weighted values are marked with L and M^* , respectively. ^bInternal reddening for different extinction laws is allowed by STARLIGHT, while PPXF uses multiplicative polynomials to correct low frequency continuum variations.

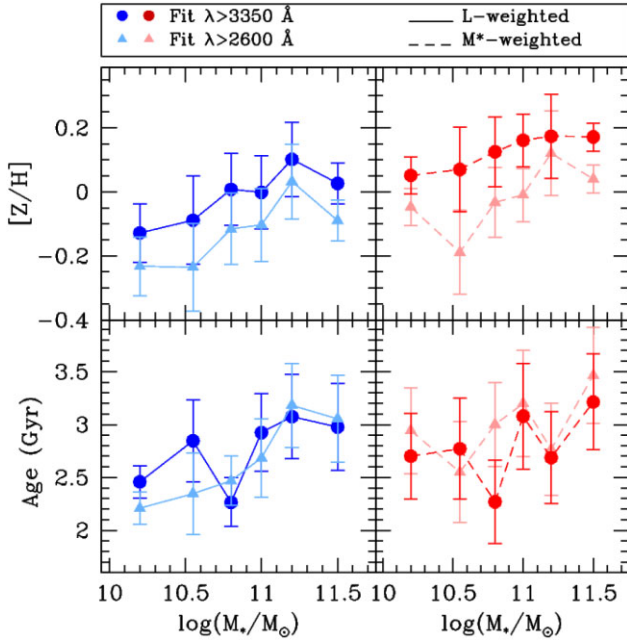


Figure 4. Stellar age (lower panels) and metallicity (upper panels) resulting from full-spectrum fitting of stacked spectra (Fig. 2) in different wavelength ranges are shown as a function of stellar mass for EMILES SSP models. Filled circles and triangles are values resulting from fitting the spectra at $\lambda_{\text{rest}} > 3350 \text{ \AA}$ and $\lambda_{\text{rest}} > 2600 \text{ \AA}$, respectively. Solid and dashed lines mark luminosity-weighted and mass-weighted values, respectively.

weighted and mass-weighted values. It is worth noting that this systematic is not dependent on the models used. Indeed, we obtained the same result with BC16 models, which extend to UV: by fitting the spectral range $[2600\text{--}4350] \text{ \AA}$, we derived metallicities, on average, ~ 0.15 dex lower than fitting the range $[3350\text{--}4350] \text{ \AA}$.

These metallicity offsets could be due to the combined effect of the different stellar populations sampled by the two different wavelength ranges, and by the poorer knowledge and implementation of the UV properties in the stellar population synthesis models (see e.g. Maraston et al. 2009; Le Cras et al. 2016; Vazdekis et al. 2016;

Lonoce et al. 2020). Simulations cannot help in disentangling these effects.⁴ It is important to note that, since the observed spectral range at rest shifts toward shorter wavelengths with increasing redshift, the effect above would result in lowering the metallicity values at high redshift with respect to those at lower redshift, mimicking an evolution.

Any comparison of the stellar population properties of different samples of galaxies cannot neglect the dependencies and systematics derived above. This aspect becomes decisive in detecting evolution of these properties across cosmic time through the comparison of measurements at different redshift. The analysis shows that such kind of measurements must be homogeneous to be comparable and that the variation (if any) should be considered relative since values cannot be considered absolute, being dependent on models, methodology, and spectral range.

In the next sections, we first determine the main relations among stellar population properties of passive galaxies at $z \sim 1.2$, then we probe their evolution with redshift using homogeneous measurements and method for different galaxy samples.

4 THE STELLAR MASS–METALLICITY RELATION

4.1 The MZR of passive galaxies at $z \sim 1.2$

Fig. 3 shows that the stellar metallicity of passive galaxies at $z \sim 1.2$ increases with stellar mass independently of the models assumed in the fitting and of the values considered, i.e. luminosity- or mass-weighted. This is also confirmed by the best-fitting linear relations obtained for the different models reported in Table 2. The best-fitting parameters reflect the large differences among the spectral fitting obtained for different models, with the large variation of the zero-

⁴By simulating a spectrum with SSP models and fit it with SSP models, we cannot test how the poor knowledge of the UV spectral range may affect the results since different spectral regions would be, by construction, self-consistent. However, in Appendix A, we simulate a galaxy with mixed stellar population to show qualitatively how the inclusion in the fit of the UV spectral range may affect the metallicity estimate.

Table 2. Best-fitting linear relations to the luminosity- and mass-weighted metallicity values derived from STARLIGHT FSF to the stacked spectra obtained in the different mass ranges, for different models $[Z/H] = a_{L,M} \log(M_{11}) + b_{L,M}$, where $M_{11} = M_*/(10^{11} M_\odot)$.

a_L	b_L	a_M	b_M	Models
0.15 ± 0.06	0.01 ± 0.02	0.11 ± 0.06	0.14 ± 0.01	EMILES
0.09 ± 0.09	-0.09 ± 0.05	0.09 ± 0.06	0.23 ± 0.03	CB16
0.26 ± 0.07	-0.15 ± 0.03	0.39 ± 0.10	-0.07 ± 0.05	M11

point b showing the large systematic in the $[Z/H]$ values discussed in the previous section.

The stellar MZR of passive galaxies at $z \sim 1.2$ in the mass range $10 \leq \log(M_*/M_\odot) < 11.6$ resulting from npFSF with STARLIGHT and E-MILES models to stacked spectra are

$$[Z/H]_L = 0.15 \pm 0.06 \log(M_{11}) + 0.01 \pm 0.02$$

$$Z/H$$

$$M_* = 0.11 \pm 0.06 \log(M_{11}) + 0.14 \pm 0.01 \quad (3)$$

for luminosity-weighted and mass-weighted values, respectively (first line of Table 2), where $M_{11} = M_*/(10^{11} M_\odot)$.

Fig. 5 shows the mass-weighted metallicity as a function of mass for the individual VANDELS passive galaxies (magenta symbols)⁵ superimposed to the stacked values (purple filled circles). The purple solid line is the mass-weighted relation reported in equation (3). The npFSF was performed with STARLIGHT and E-MILES SSPs over the wavelength range $\lambda_{\text{rest}} > 3350 \text{ \AA}$. Internal reddening in the range $A_V = 0-2$ mag was allowed in the fitting by considering the Calzetti et al. (2000) extinction law. In Fig. 6, the distributions of the values of extinction and of the χ^2 values resulting from the fitting are shown.

The stellar metallicity of ~ 95 per cent of the sample falls in the range $-0.35 < [Z/H] < 0.25$, as shown by the distribution in the right-hand panel of Fig. 5. In particular, 67 (58) per cent of the galaxies at $z \sim 1.2$ have mass-weighted (luminosity-weighted) metallicity higher than solar ($[Z/H] > 0$). In Fig. 5, we also show, for comparison, the metallicity of massive ($\log(M_*/M_\odot) > 11$) ETGs at $z \sim 1.4$ derived with the same procedure adopted in this work from the VLT-FORS2 spectra studied by Gargiulo et al. (2016, Ga16; blue triangles), and for passive galaxies in cluster XLSS0223 at $z \sim 1.2$ from LBT-MODS spectra (Saracco et al. 2019, Sa19; cyan filled squares). The agreement among the different data at comparable redshift is very good.

We confirm that the stellar metallicity of passive galaxies is positively correlated with their stellar mass at $z \sim 1.2$, as tentatively previously found by Kriek et al. (2019) and Saracco et al. (2019) on much smaller samples. This trend is independent of the models assumed,⁶ as shown in Fig. 3, even if different models provide

⁵The typical uncertainty on single measurement has been derived according to the following procedure. We considered the best-fitting composite model of one of the VANDELS spectra with an average $S/N \sim 7-8 \text{ \AA}^{-1}$, representative of the selected spectra. We obtained a number of realizations by summing to this template the shuffled residuals of the fit itself. For each realization, we perform the fit and estimate the metallicity. As typical error, we considered the standard deviation from the realizations.

⁶The Spearman rank test performed on the metallicity values derived from the stacked spectra for the different models and codes (Fig. 3) provided correlation coefficients $\rho_s > 0.92$ (i.e. probabilities $p_s \sim 0.005$ that the data are not correlated) in all the cases, with the exception of CB16 that provided $\rho_s \simeq 0.54$ and $p_s \simeq 0.26$. The test performed on single VANDELS values provided $\rho_s \simeq 0.27$ with an associated probability $p_s \simeq 0.04$.

systematic differences in the derived stellar populations properties. The scatter in the metallicity of individual VANDELS galaxies at $\log(M_*/M_\odot) > 11$ is about ± 0.1 dex, while it is larger at lower masses, where galaxies span a wider range of metallicity. It is worth noting that the positive trend between metallicity and mass is mainly due to the lack of low-metallicity galaxies with high mass rather than to a higher metallicity of high-mass galaxies, as also seen in the local Universe (e.g. Gallazzi et al. 2005; Asari et al. 2009; Choi et al. 2014; McDermid et al. 2015), and this produces also the flattening of the relation for masses $\log(M_*/M_\odot) > 11.2$.

4.2 The evolution of the stellar MZR

In Fig. 7, we compare the stellar metallicity of passive galaxies over the redshift range $0 < z < 3.35$ derived homogeneously according to the procedure adopted in this work, i.e. same fitting code (STARLIGHT), simple stellar population models (EMILES) and fitting wavelength range ($[3350-4350] \text{ \AA}$). At redshifts lower than VANDELS data, $z < 1.2$, we derived stellar metallicity from stacked spectra ($S/N > 15 \text{ \AA}^{-1}$) of passive galaxies selected from the LEGA-C survey (van der Wel et al. 2016, 2021) in the mass range $10.2 < \log(M_*/M_\odot) < 11.6$ and in the redshift range $0.6 < z < 0.7$ (Bevacqua et al, in preparation). The metallicity values at $z \sim 0.65$ derived at the different masses agree with those at $z \sim 1.2$ derived from VANDELS stacked spectra. Also the LEGA-C data show a positive correlation between metallicity and mass.

At $z \sim 0.38$ we derived the metallicity of massive ETGs from the high S/N (~ 100) stacked spectra of Salvador-Rusiñol et al. (2020). They selected ETGs in the narrow mass range $11.2 < \log(M_*/M_\odot) < 11.45$ from BOSS and stacked the spectra according to their velocity dispersion ($220 \text{ km s}^{-1} < \sigma_v < 340 \text{ km s}^{-1}$) in bin of 100 km s^{-1} considering, for each stack, the median stellar mass (see Salvador-Rusiñol et al. 2020, for details). The metallicity values we derived lie on the relation described by the LEGA-C data and are consistent with those from VANDELS data at $z \sim 1.2$.

At even lower redshift, $z \sim 0.05$, we used the high S/N (~ 100) stacked spectra of ETGs selected from the SPIDER sample (La Barbera et al. 2013) in the mass range $10.6 < \log(M_*/M_\odot) < 11.2$. The stacking was made according to their velocity dispersion as in Salvador-Rusiñol et al. (2020). Also in this case, the metallicity values agree with those at higher redshift.

At redshift higher than VANDELS data, we considered the mean metallicity value (blue triangle) of the five massive ($11 < \log(M_*/M_\odot) < 11.6$) ETGs at $z \sim 1.4$ studied by Gargiulo et al. (2016) and shown individually in Fig. 5, and the metallicity derived for the massive ($\log(M_*/M_\odot) \sim 11.3$) ETG C1-23152 at $z \sim 3.35$, whose estimate was obtained following the same procedure adopted in this work (Saracco et al. 2020b). The metallicity of these galaxies is consistent with the values derived for VANDELS galaxies and for massive ETGs at ~ 1.4 at similar mass, as well as with those derived at lower redshift for LEGA-C, BOSS, and SDSS massive ETGs.

Therefore, our analysis does not show evidence of metallicity evolution in the redshift range probed. In particular, we do not detect change of the stellar metallicity of passive galaxies with mass $\log(M_*/M_\odot) > 11$ at least in the redshift range $0.1 < z < 1.4$. At higher redshift, we note that the metallicity of C1-23152, the only massive passive galaxy for which an estimate of the stellar metallicity has been obtained to date at $z > 3.0$ suggests a lack of variation even up to these redshifts. In fact, the metallicity values derived from the high S/N stacked spectra in the above mass range, do not show any trend with redshift, and the metallicity values of individual massive galaxies are consistent with each other up to the highest redshift

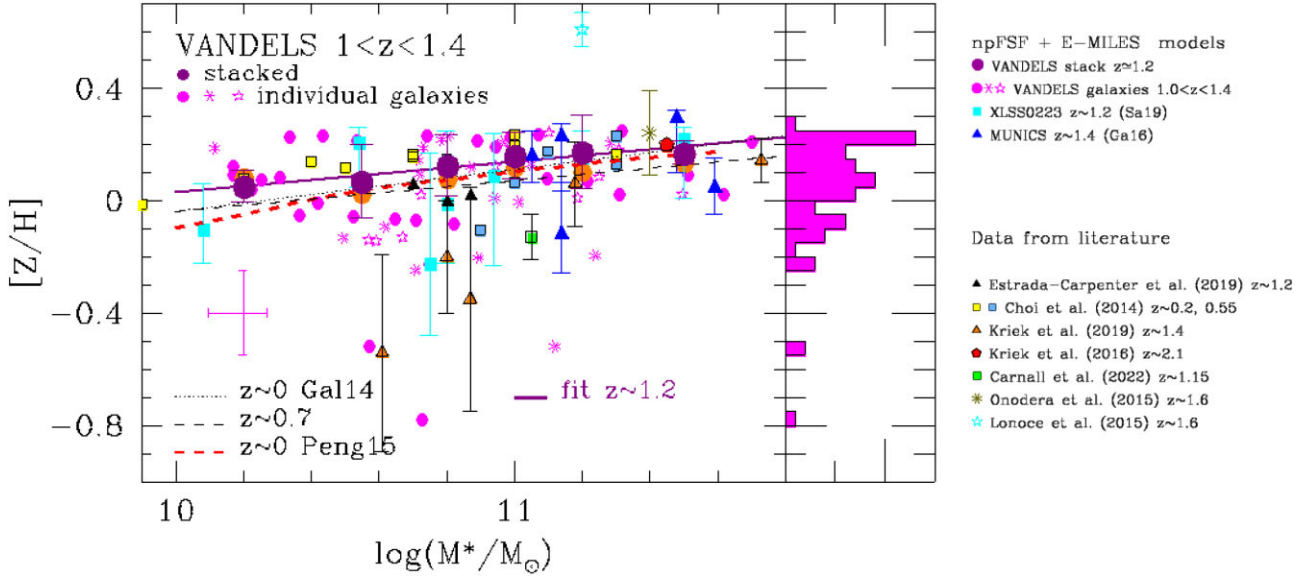


Figure 5. Mass-weighted metallicity derived through non-parametric full-spectrum fitting (npFSF) with STARLIGHT and E-MILES models as a function of mass for quiescent galaxies at comparable redshift $z \sim 1.2$. Magenta symbols (filled circles, stars, and skeletons) are VANDELS ETGs, LTGs, and unclassified galaxies, respectively, at $1.0 < z < 1.4$; big orange filled circles are the median values of VANDELS galaxies in the same mass bins of stacked spectra; big purple filled circles are VANDELS stacked spectra (errorbars are the standard deviations of the values in each bin); cyan filled squares are individual passive galaxies in cluster XLSS0223 at $z \sim 1.2$ from Saracco et al. (2019, Sa19); blue triangles are individual massive ETGs at $z \sim 1.4$ from Gargiulo et al. (2016, Ga16). The purple continuous line is the best-fitting linear relation to the stacked values of VANDELS (see equation 3). The black dotted and dashed lines are the MZR of quiescent galaxies at $z = 0.1$ and $z = 0.7$, respectively, from Gallazzi et al. (2014, Gal14). The red dashed line is the observed stellar MZR of local passive galaxies derived by Peng, Maiolino & Cochrane (2015) from SDSS data. Data from the literature, as derived by Estrada-Carpenter et al. (2019), Choi et al. (2014), Kriek et al. (2016, 2019), Carnall et al. (2022), Onodera et al. (2015), and Lonoce et al. (2015), are superimposed to npFSF + EMILES values. Symbols are as in the legend. Cross in the lower left of the left-hand panel represents the typical error at 1σ of individual measurements for VANDELS galaxies.

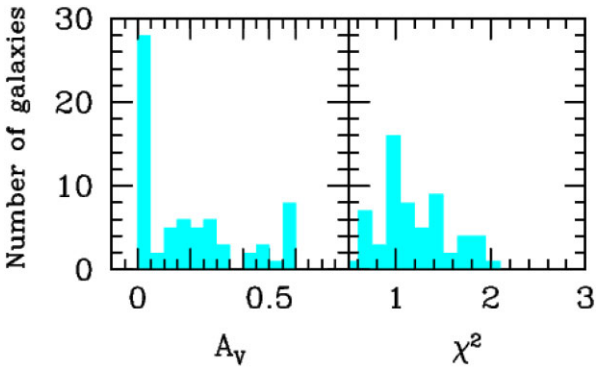


Figure 6. Distributions of extinction values (left) and of χ^2 values (right) resulting from the npFSF of VANDELS galaxies with STARLIGHT and E-MILES models.

probed by our data. The stellar MZR defined over the mass range $10 < \log(M_*/M_\odot) < 11.6$ does not show evolution between $z \sim 1.2$ and $z \sim 0.65$, either in the mean value or in the slope.

It is worth noting that some hydrodynamic simulations and semi-analytical models predict an evolution of the stellar metallicity of galaxies with redshift, even if typically less than 0.1 dex (see e.g. fig. 11 in Guo et al. 2016). However, the direction of the metallicity evolution (increasing or decreasing with redshift) is not always coincident among them. In Fig. 7, the mass-metallicity for passive galaxies as resulting from the predictions of the GALaxy Evolution and Assembly (GAEA) semi-analytical model (Fontanot et al. 2021) is also shown as an example. Models show the relation for mass-weighted values expected at $z \sim 1.2$, $z \sim 0.7$, and $z \sim 0$ for

passive galaxies defined according to their specific star formation rate $\log(\text{sSFR}) < -0.15 \times \log(M_*) - 9.66$ (Gallazzi et al. 2021).⁷ The agreement between the slope of the predicted (0.14 ± 0.05) and observed relation (see equation 3) is remarkable, as well as the flattening of the relation at large masses, as shown in Fig. 7 where the GAEA predicted MZR at $z \sim 1.2$ is rescaled up by $[Z/H] = +0.23$ to match the median $[Z/H]$ of VANDELS stacks. The predicted systematic increase of the metallicity with decreasing redshift is consistent with no evolution especially at $\log(M_*/M_\odot) > 11$, where the evolution is much smaller than the scatter in the data. We note that the possible tension between the normalization of the observed and predicted relations could be not significant. Indeed, as shown in Section 3.2, different SSP models can provide metallicity values differing even by 0.3 dex (see Fig. 3), a difference that alone could justify the apparent discrepancy. Therefore, we believe that there are not the conditions to support a discrepancy between the GAEA predicted stellar metallicities and those observed. We believe that the low-mass regime of the relation, still difficult to probe with the current ground-based observing facilities, would deserve to be investigated since, at low masses, a larger evolution is predicted.

4.3 Comparison with previous estimates of metallicity in the literature

For completeness, we also report metallicity estimates from the literature, even if based on different models and methods. Whenever possible, we distinguish mass-weighted from luminosity-weighted

⁷We note that the use of different criteria to select passive galaxies (e.g. $\text{sSFR} < 0.3/t_{\text{Hubble}}(z)$) does not introduce significant differences in the result.

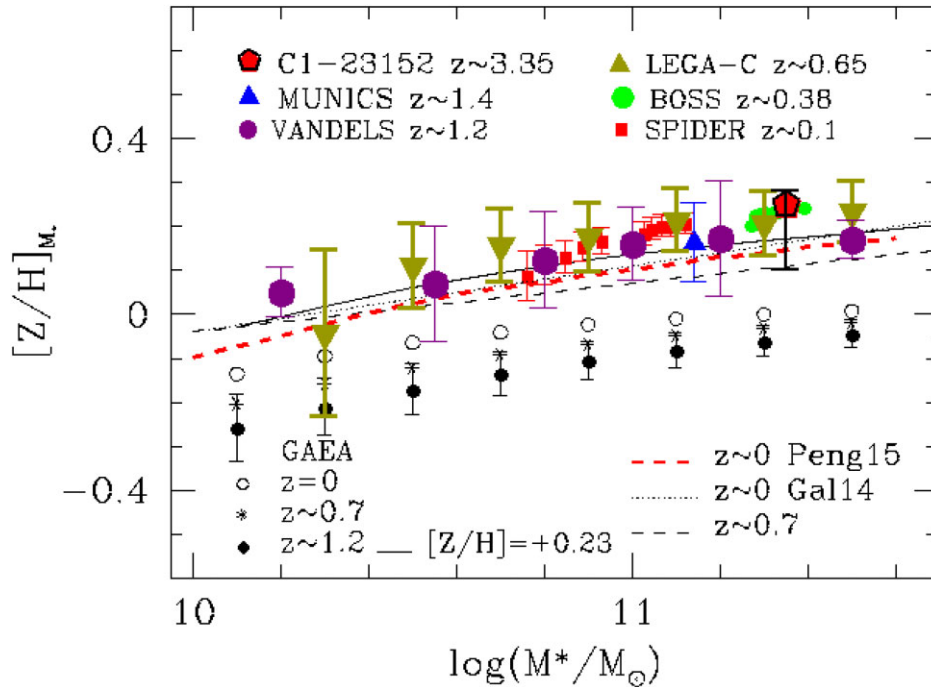


Figure 7. Mass-weighted metallicity derived through non-parametric full-spectrum fitting (npFSF) to stacked spectra of quiescent galaxies at different redshifts as a function of mass. The fitting was performed with STARLIGHT code and EMILES models over the same wavelength range ([3350–4350] Å) for all the spectra. Purple circles are VANDELS stacked spectra, military triangles are LEGA-C stacked spectra at $0.6 < z < 0.7$; green filled circles are stacked BOSS spectra of massive ETGs at ($z \sim 0.38$ from Salvador-Rusiñol et al. (2020); red squares are stacked SPIDER spectra of massive ETGs at $z \sim 0.05$ (La Barbera et al. 2013); blue triangle is the median value of the 5 massive ETGs at $z \sim 1.4$ from Gargiulo et al. (2016); red diamond is the massive ETG C1-23152 at $z \sim 3.35$ studied by Saracco et al. (2020b). Superimposed to the data are also predictions from GAEA models (Fontanot et al. 2021, black open and filled symbols) for different redshifts as in the legend. The black thin curve is the GAEA prediction at $z \sim 1.2$ rescaled up by $+0.23$ in $[Z/H]$ to match the median value of VANDELS data. The mass–metallicity relations by Gallazzi et al. (2014, black dotted and dashed lines) and Peng et al. (2015, red dashed curve) are also shown.

estimates given the difference existing between these two quantities. We remind the reader that a systematic difference of 0.06 dex in metallicity exists between EMILES and some other models (see note 3). When relevant for the comparison, we will explicitly take this offset into account.

At low redshift, we show in Fig. 5 the SSP-equivalent metallicity derived by Choi et al. (2014)⁸ from stacked spectra of passive galaxies selected from the AGES survey (Kochanek et al. 2012). We considered their estimates in the redshift range $0.1 < z < 0.3$ and $0.4 < z < 0.7$ ($z \sim 0.2$ and $z \sim 0.55$ in the legend, respectively). They used the fitting code and models developed by Conroy & van Dokkum (2012). Their data follow the MZR found by Gallazzi et al. (2006) for local ($z \sim 0.1$) ETGs (dotted line) and agree well with the values we derived both at $z \sim 1.2$ from VANDELS data and at $z \sim 0.65$ from LEGA-C data even when an offset by 0.06 dex is considered.

Besides these data, we show the metallicity value obtained by (Carnall et al. 2022,; C22 sample hereafter) for stacked spectra of 91 massive passive galaxies in the redshift interval $1.0 < z < 1.3$. They selected galaxies brighter than $J = 21.5$ from the VANDELS photometric sample (McLure et al. 2018) applying a slightly different UVJ colour selection criterion to define passive galaxies with respect to VANDELS, and recomputed stellar masses. The resulting sample is complete down to $\log(M^*/M_\odot) = 10.8$ and

composed of 77 galaxies with VANDELS spectra and 14 galaxies with KMOS spectra (see Carnall et al. 2022). We expect no more than 30 galaxies in common between our and C22 sample,⁹ i.e. the galaxies more massive than $\log(M^*/M_\odot) = 10.8$ in our sample. Their metallicity estimate is based on parametric FSF performed with BAGPIPES code (Carnall et al. 2018) and CB16 models (Bruzual & Charlot 2003) over the wavelength range [3550–6400] Å thanks to KMOS observations. The mean mass-weighted metallicity they obtained, $[Z/H] = -0.13 \pm 0.08$ ($[Z/H] = -0.07$ when corrected by $+0.06$ dex), is $\sim 2\sigma$ lower than our estimate for the same stellar mass (see Table 1). According to the comparison among models shown in Fig. 3, the offset between the metallicity values found in our and in their work cannot be due to the different models used (E-MILES vs CB16): by adopting CB16 models we obtain a metallicity even higher than the one obtained with E-MILES models, as shown in the upper left panel of Fig. 3.

A possible reason for the discrepancy could be the different wavelength range considered in our and in their fitting procedure. Indeed, the range [4400–6400] Å is not present in our spectra because of the lack of near-IR observations. We verified whether the inclusion of this range of wavelengths in the fitting affects the metallicity estimate with respect to the estimate based on the range [3350–4350] Å. By extending the fitting of high S/N BOSS stacked spectra of massive ETGs at $z \sim 0.38$ (Salvador-Rusiñol et al. 2020) up to λ

⁸Choi et al. (2014) derive the abundances of metal elements. We obtained the metallicity using the relation $[Z/H] = [Fe/H] + 0.94[Mg/Fe]$ (Thomas, Maraston & Bender 2003).

⁹The actual overlap between our sample and C22 sample, as well as the overlap with the Carnall et al. (2019) sample, cannot be quantified as both samples are not disclosed.

$\sim 6000 \text{ \AA}$, we did not detect significant offset in the metallicity (see Appendix B, Fig. B1).

A more direct comparison can be made with the results obtained by Carnall et al. (2019), based on the same spectra used in our analysis. In this case they selected 75 passive galaxies from the VANDELS spectroscopic sample in the redshift range $1.0 < z < 1.3$ at $\log(M_*/M_\odot) > 10.3$. Considering that there are five galaxies in our sample with $z > 1.3$ and five galaxies with $\log(M_*/M_\odot) < 10.3$, we expect a sample of about 54 galaxies in common (60 galaxies considering also the six galaxies we removed being superposition or merger of two galaxies, see Section 2.1). A systematic difference between our and their $[Z/H]$ estimates exists among the metallicity values for individual passive galaxies that they show in fig. 3 of Carnall et al. (2022): none of the passive galaxies of their sample has metallicity higher than solar, contrary to a fraction of ~ 60 per cent with supersolar metallicity in our sample (see above). The spectral fitting in Carnall et al. (2019) was performed in the wavelength range $[2600\text{--}4400] \text{ \AA}$. We have already shown that the inclusion of the UV part of the spectrum in the fitting, not observed in local and low redshift ($z < 0.4$) galaxies, leads to metallicity values systematically lower by 0.15 dex (see Section 3.2). However, from Fig. 4, it can be seen that the inclusion of the UV wavelength range in our fitting, leads to a mean metallicity $[Z/H] \geq 0$ for galaxies more massive than $\log(M_*/M_\odot) > 11.0$, hence higher than the estimates in Carnall et al. (2019). Moreover, as noticed above, by adopting CB16 models we would obtain metallicities even higher than those obtained with E-MILES models. Finally, they do not find a positive trend of the metallicity with the stellar mass. Therefore, we hypothesize that the difference between our and their results is mainly due to the significant differences between the methods used to fit the data: on one hand, a npFSF method based on a linear combination of SSPs to fit the spectrum; on the other hand, a parametric FSF based on constrained SFHs to fit the spectrum and the photometric data.

At redshift comparable to VANDELS data, we show in Fig. 5 the mass-weighted metallicity values derived by Kriek et al. (2019) for five massive ($\log(M_*/M_\odot) > 10.6$) galaxies at $z \sim 1.4$, the metallicity derived by Onodera et al. (2015) from the stacked spectra of 16 passive galaxies at $z \sim 1.6$ and the metallicity estimate by Lonoce et al. (2015) for a massive galaxy at $z \sim 1.4$. The measurements by Kriek et al. (2019), based on the absorption lines fitting code and models developed by Conroy & van Dokkum (2012) (see also Conroy, Graves & van Dokkum 2014; Choi et al. 2014), agree with the positive correlation of the metallicity with mass. The measurement by Lonoce et al. (2015) is based on absorption lines fitting with M11 models (Maraston & Strömbäck 2011) and, taken together with the Kriek et al. (2019) estimates, confirm the large scatter we also observe in the stellar metallicity of individual galaxies (see also Lonoce et al. 2020). It is worth to mention that large differences are found by Spiniello et al. (2012) between predictions of some line indices of Conroy & van Dokkum (2012) models with respect to Vazdekis et al. (2015) models. The metallicity derived by Onodera et al. (2015), based on the comparison of absorption line indices with the models' predictions by Thomas, Maraston & Johansson (2011), is consistent with the metallicity values we derived at comparable mass.

At higher redshift, we show the mass-weighted metallicity derived by Kriek et al. (2016) for a massive quiescent galaxy at $z = 2.1^{10}$

¹⁰As far as we know, there are no other metallicity estimates for quiescent galaxies at $z > 2.5$ besides the estimate at $z \simeq 3.35$ by Saracco et al.

according to the same method used in Kriek et al. (2019). The metallicity agrees with those derived for galaxies of similar mass both at lower and higher redshift.

Therefore, the different estimates from the literature, with the exception of those by Carnall et al. (2019, 2022), agree with our estimates: the average stellar metallicity of massive ($\log(M_*/M_\odot) > 11$) passive galaxies is supersolar, higher than for lower mass galaxies, and it did not change across cosmic time, at least over the last 9 Gyr. This lack of a significant evolution in the mean metallicity value seems to apply to the whole population of passive galaxies in the mass range $10 < \log(M_*/M_\odot) < 11.6$, as suggested by the agreement among the different estimates in the redshift range $0 < z < 1.4$. The detection at $z \sim 1.2$ of a positive correlation between metallicity and mass, consistent with the one observed in the local Universe, shows that the observed trend was established at earlier epochs as result of the formation process rather than their evolution.

5 THE STAR FORMATION HISTORY

5.1 The formation epoch of stellar mass

Figs 3 and 4 (lower panels) show the stellar age resulting from the fitting to stacked spectra as a function of their mass. A mild increase of the age with the mass can be seen, regardless of the choice of the models. However, galaxies that contribute to the stacked spectra in each mass interval are spread over the redshift range $1.0 < z < 1.4$, corresponding to an interval of time $\Delta t \sim 1.3$ Gyr which could affect the real trend. To properly compare the mean stellar age of galaxies seen at different redshift, we considered the mean formation epoch of the stellar mass defined as $t_{\text{form}} = \text{Age}_U(z) - \text{Age}_{M_*}(z)$, where $\text{Age}_U(z)$ is the age of the Universe at the redshift of the galaxy and $\text{Age}_{M_*}(z)$ is its mass-weighted age.

Fig. 8 shows the mean formation epoch t_{form} of the stellar mass as a function of mass for the VANDELS sample and the other passive galaxies at similar and higher redshift as well as those derived for stacked massive ETGs at lower redshift. Large purple filled circles are the median t_{form} values of VANDELS passive galaxies in the different mass intervals. The well known general trend between formation epoch and stellar mass is clearly visible from the figure: the higher the stellar mass of a galaxy the earlier formed (the older are) its stars, in agreement with previous studies of stellar population properties in local galaxies (e.g. Cowie et al. 1996; Kauffmann et al. 2003; Gallazzi et al. 2005; Thomas et al. 2005, 2010; Conroy et al. 2014; McDermid et al. 2015) and of scaling relations for local and higher redshift galaxies (e.g. McDermid et al. 2015; Barone et al. 2018; Saracco et al. 2020a). For comparison, Fig. 8 shows the $t_{\text{form}} - M_*$ relation found by Carnall et al. (2019, dashed line) on VANDELS passive galaxies, and the relation derived by Saracco et al. (2020a, thin solid curve) from the study of the Fundamental Plane of cluster ETGs at $z \sim 1.2$. The relations follow the same trend with similar slopes even if they are offset by about 1.0 Gyr. By fitting the median t_{form} values of VANDELS galaxies as a function of stellar mass we derived the relation:

$$\log(t_f/\text{Gyr}) = (-0.29 \pm 0.04) \times \log(M_*/M_\odot) + (3.5 \pm 0.4) \quad (4)$$

(2020b). Recently, Cullen et al. (2019) and Calabrò et al. (2021) derived stellar metallicities for VANDELS star-forming galaxies at $z > 2.5$ by comparing the far UV ($\lambda_{\text{rest}} < 2000 \text{ \AA}$) spectral features of stacked spectra with the theoretical stellar library of massive stars by Leitherer et al. (2010).

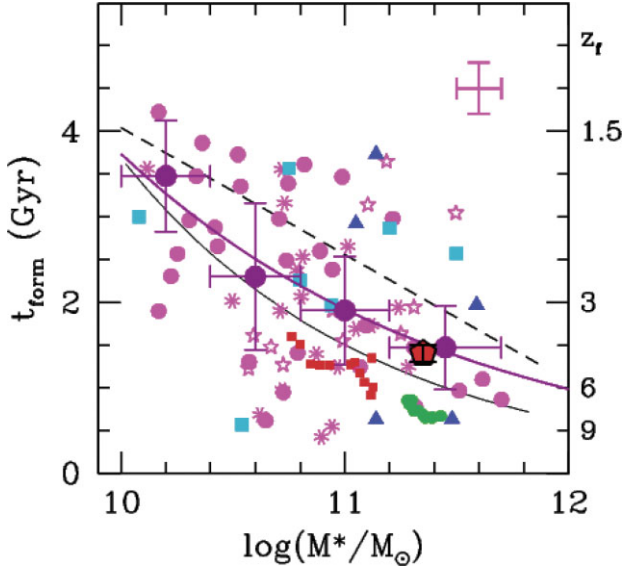


Figure 8. Formation epoch/redshift (left/right axis) as a function of stellar mass. Symbols are as in Fig. 7. The purple filled circles are the median t_{form} values of VANDELS galaxies in the different mass intervals as marked by horizontal error bars. Vertical error bars represent the standard deviation of the values in each mass bin. The cross in the upper right corner represents the typical error at 1σ of individual values. The purple thick solid curve is the best-fitting $t_{\text{form}}-M_*$ relation derived for the median t_{form} values of VANDELS galaxies (see text, equation 4). The black dashed line is the best-fitting $t_{\text{form}}-M_*$ relation found by Carnall et al. (2019). The black thin solid curve is the $t_{\text{form}}-M_{\text{dyn}}$ relation derived by Saracco et al. (2020a, equation 8) from the FP scaled according to the mean value $\langle M_{\text{dyn}}/M_* \rangle \simeq 1.6$ they find for their sample.

represented by the thick purple solid curve which lies in between the two previous relations.

5.2 Star formation and stellar mass

The $t_{\text{form}}-M_*$ relation expressed by equation (4) summarizes the relationship between SFH and mass. From the fit to VANDELS stacked spectra (Fig. 2), the SFH seems to be smoother and longer (i.e. star formation episodes distributed over a larger interval of time) in lower mass galaxies, while is sharper and shorter (i.e. characterized by one or two major episodes accounting for more than 50 per cent of the mass) in higher mass ones. In these latter, more than $10^{11} M_{\odot}$ formed within ~ 1 Gyr.

Fig. 9 (left) shows the mean SFH of VANDELS galaxies in different mass bins, i.e. the mean fraction of stellar mass as a function of time as resulting from the fitting to individual galaxies, instead of considering stacked spectra as in Fig. 2. The right-hand panel shows the mean integrated SFH, the cumulative fraction of mass. The average SFH has been obtained according to the following procedure. For each galaxy, we considered the fraction of stellar mass associated with each SSP contributing to the composite best-fitting model and, for each of them, we derived the corresponding t_{form} from the age of the SSP. We then re-sampled and summed the fractions for all the galaxies within intervals of $\log(dt[\text{Gyr}]) = 0.05$ (see also Asari et al. 2007). Finally, we run a moving mean (MM) over $5 \log(dt)$ to smooth the data. We verified that the results do not depend on the binning assumptions by repeating the procedure for different binning. We note that quite different SFHs exist among galaxies with similar mass (see e.g. Carnall et al. 2019; Tacchella et al. 2022).

Fig. 9 shows that at increasing stellar mass the fraction of stars that formed at early epochs increases: the SFH tends to be more skewed towards early epochs. Less massive galaxies host younger stars whose formation started later, according to an SFH peaked at more recent epochs than for more massive galaxies. Moreover, Fig. 9 also suggests the presence of a double peaked SFH. The relative intensity of the peaks seems to correlate with the mass: the higher the mass the more pronounced the first peak at early epochs. Whether the double peak is real or simply the result of the discrete nature of the SFH derived here (sum of the SSPs), the formation of stellar mass begins earlier and earlier as the mass of the galaxy increases more and more. Indeed, the right-hand panel shows that, on average, the stellar mass in higher mass galaxies is formed earlier than in lower mass ones, as shown by the epoch at which 50 per cent of all stars in a galaxy is already formed (see e.g. Estrada-Carpenter et al. 2020). More than 50 per cent of the stellar mass in massive $\log(M_*/M_{\odot}) > 11.3$ passive galaxies is formed at $z > 5$, and almost 80 per cent within the first 2 Gyr of the cosmic time, i.e. by $z \sim 3$. This agrees with the results shown in Fig. 2 derived from the fit to the stacked spectra.

Significant star formation took place at early epochs. Indeed, galaxies with masses $\log(M_*/M_{\odot}) > 10.5$ experience an epoch of significant star formation at the earliest cosmic times, as shown by the growth of stellar mass within the first 1.5 Gyr. However, at $z \sim 1.2$, these galaxies show quite different properties, in terms of age and metallicity. These differences are mainly driven by the different SFHs and by the likely different rate at which star formation decline. We take up this aspect in more detail in the next section.

These results are qualitatively similar to those derived for local early-type galaxies (e.g. Thomas et al. 2005, 2010; McDermid et al. 2015). We underline the fact that the results obtained here are independent of any assumption. The power of the non-parametric full spectral fitting, as the one adopted in this analysis, is just to detect differences in the SFH resulting from differences in the spectral properties. All the galaxies can indeed be modelled by linearly combining SSPs extracted from the same base of models without any prior, apart from the discretization of the models (the same for all the galaxies). Therefore, the fact that stellar populations in galaxies of greater mass are, on average, described by different SFHs than stellar populations in galaxies of lower mass reflects precisely the presence of systematic differences in their spectral and therefore stellar properties.

6 STELLAR METALLICITY AND SFH

Fig. 10 shows how the stellar metallicity, the time-scale to form half of the stellar mass and the mean SFR (see below) are distributed among galaxies on the $t_{\text{form}}-M_*$ plane. The upper panel shows the distribution of the mass-weighted metallicity $[Z/H]$. The colourscale shows $[Z/H]$ smoothed using the locally weighted regression algorithm LOESS (Cleveland & Devlin 1988; Cappellari et al. 2013) to highlight the presence of possible trends. The figure joins the relationships between metallicity and age with the mass of galaxies, correlations individually seen in Figs 5 and 8, respectively. Even if the statistic is low, this 3D view adds an information that was not evident from the two individual figures: contrary to high-mass galaxies ($\log(M_*/M_{\odot}) > 11$), that all have supersolar metallicity and whose stellar mass formed earlier (old stellar ages), the stellar metallicity of lower mass galaxies (e.g. $\log(M_*/M_{\odot}) \sim 10.6$) is subsolar if formed at earlier epochs (old ages), supersolar if formed at later epochs (young ages). Invoking the age-metallicity degeneracy to justify this trend is of no help: high-mass galaxies, all formed at

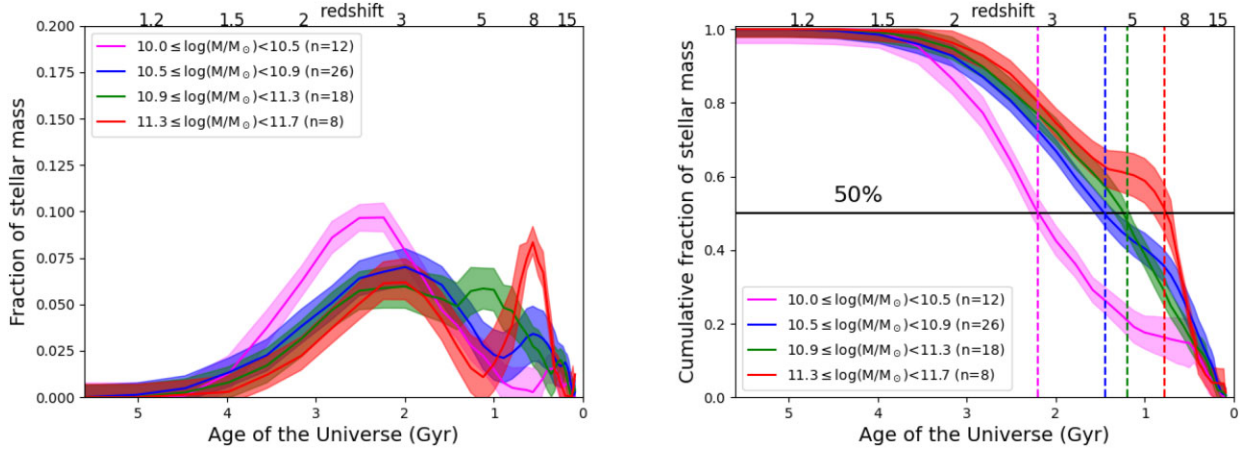


Figure 9. Left: Average SFH of VANDELS passive galaxies in different ranges of stellar mass as a function of cosmic time. Continuous curves are the mean mass fraction as resulting from the best-fitting composite model of individual galaxies for different range of stellar masses. The dashed regions indicate the dispersion of the values divided by $\sqrt{N_{\text{gal}}}$ at each time interval. Right: Average integrated SFH showing the cumulative mass fraction existing at a given time. The vertical dashed lines marks t_{50} (z_{50}), the age of the Universe (redshift, top x-axis) at which 50 per cent of the stellar mass of galaxies for the different mass ranges was already formed.

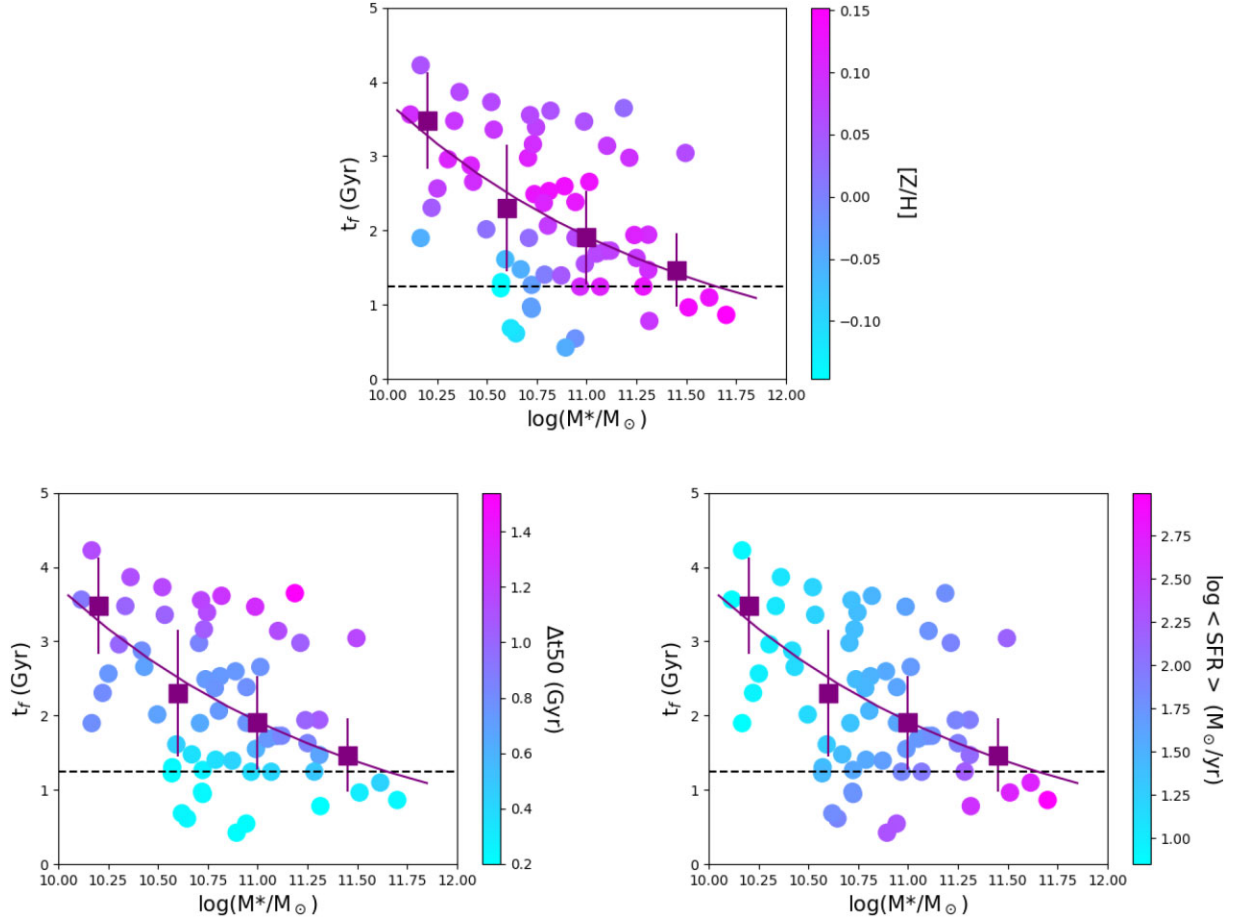


Figure 10. Formation epoch t_{form} as a function of stellar mass for the VANDELS galaxies (filled circles). Colourscale represent LOESS-smoothed mass-weighted metallicity $[Z/H]$ (upper panel), time interval Δt_{50} within which 50 per cent of the stellar mass is formed (lower left; see text) and SFR averaged over the time required to form 90 per cent of the stellar mass (lower right). The median t_{form} values of VANDELS galaxies in the different mass intervals (purple squares) and the best-fitting t_{form} -mass relation (purple curve; equation 4) are also shown. The dashed line defines Maximally Old Galaxies (MOGs, see Section 6), those hosting stellar populations formed at $z > 5$, within the first ~ 1.2 Gyr of the cosmic time.

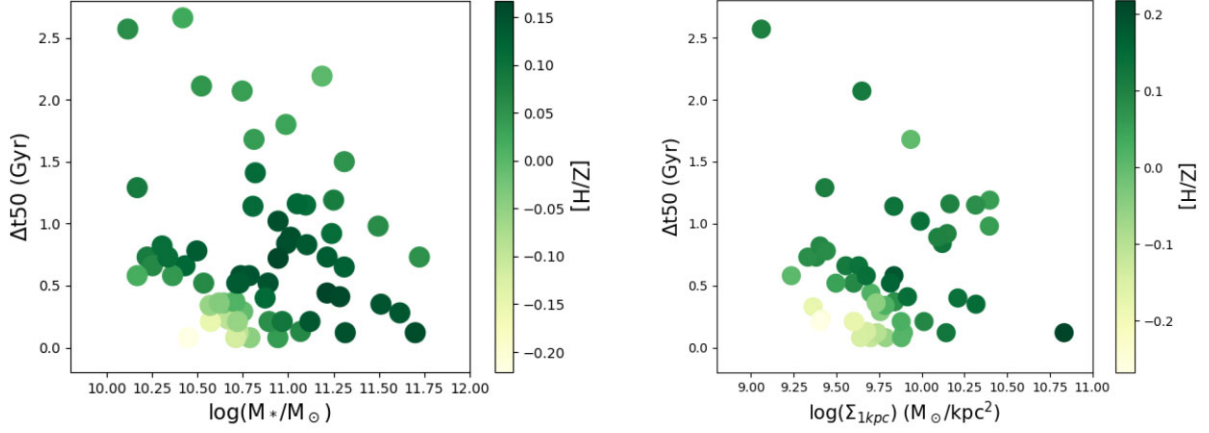


Figure 11. Time Δt_{50} required to form 50 per cent of the mass as a function of mass (left) and central stellar mass density $\Sigma_{1\text{kpc}}$ (right). Colourscale represents LOESS-smoothed mass-weighted metallicity $[Z/H]$. Note that right-hand panel shows 51 passive galaxies out of the 64: 25 galaxies for which an effective radius (R_e) was available from the literature (van der Wel et al. 2012, 2014) and 26 for which R_e has been derived from *HST-F160W* archive images.

earliest epochs, i.e. hosting oldest stars, should be all metal poor, contrary to what it is found.

Focusing on galaxies hosting the oldest stellar population, that is those with $t_{\text{form}} < 1.2$ Gyr ($z_{\text{form}} > 5$) that we call maximally old (MOGs)¹¹ hereafter, the positive trend between $[Z/H]$ and mass is extremely clean. By selection, the stellar populations in these galaxies formed nearly within the same (short) time. However, their metallicity is significantly different and increases systematically with the mass of the galaxy. Conversely, focusing on low-mass galaxies, at fixed mass say, e.g. $\log(M_*/M_\odot) \sim 10.6$, the metallicity is not constant for all of them but increases systematically with t_{form} and reaches supersolar values as high as for the most massive galaxies. Therefore, the trend between metallicity and mass cannot be the only causal relationship but there must be other processes that overlap and modulate this relationship.

In the lower left panel, the colourscale shows the quantity Δt_{50} (Gyr) defined as the time to form 50 per cent of the stellar mass of the galaxy, starting when the star formation begins. It is the time needed to form half of the mass (a proxy for the duration of the SF) derived from the SFH of galaxies described in the previous section. We remark that Δt_{50} differs from t_{50} shown in Fig. 9 since the latter marks the cosmic epoch at which 50 per cent of the stellar mass formed while the former indicates the time required to do it. Correctly, MOGs are characterized by similar Δt_{50} . On the contrary, the systematic increase of the metallicity with t_{form} at fixed (low) mass, is accompanied by increasing values of Δt_{50} : in low-mass galaxies, shorter SF produces lower stellar metallicity, longer SF higher metallicity. We notice that while low-mass galaxies can be described by both long and short values of Δt_{50} , high-mass galaxies are described only by short values. In fact, we find that Δt_{50} is anticorrelated with the mass (and mass density) as shown in Fig. 11, i.e. the time-scale of star formation depends on the mass (and mass density) of galaxy, as found for local ETGs (e.g. Thomas et al. 2010; McDermid et al. 2015; but see Beverage et al. 2021 for a different result).

In the lower right panel of Fig. 10, the colourscale shows the average $\langle SFR \rangle$ ($M_\odot \text{ yr}^{-1}$) defined as the ratio between 90 per cent of

the mass and the time interval to form it. As expected, $\langle SFR \rangle$ increases with the mass, showing that the rate at which gas is converted into stars increases faster with mass than the duration of the SF. At fixed mass, $\langle SFR \rangle$ increases with decreasing t_{form} since decreases also Δt_{50} . Focusing on MOGs, given that they all have approximately the same Δt_{50} , $\langle SFR \rangle$ scales simply according to the mass, from $\sim 30 M_\odot \text{ yr}^{-1}$ at $\log(M_*/M_\odot) \sim 10.6$ to more than $300 M_\odot \text{ yr}^{-1}$ at $\log(M_*/M_\odot) \sim 11.6$.

Fig. 11 shows Δt_{50} as a function of mass and central stellar mass density $\Sigma_{1\text{kpc}} = M_{1\text{kpc}}/(\pi R_{1\text{kpc}}^2)$, defined as the mass density within 1 kpc radius (Saracco, Gargiulo & Longhetti 2012; Saracco et al. 2017).¹² The colourscale shows the metallicity $[Z/H]$. The figure shows a trend between Δt_{50} and $\Sigma_{1\text{kpc}}$ recalling the one in the upper panel of Fig. 10: highest densities are associated with shortest duration and highest metallicity; lower densities are associated both with short and long duration with the metallicity that increases as Δt_{50} increases.

Taken together, these relations suggest that the duration of the SF (the time-scale to form half the mass, Δt_{50}), and the rate at which the gas is converted into stars ($\langle SFR \rangle$) play a role in the definition of the stellar metallicity depending on the mass and/or on the mass density of the galaxy. We discuss further these results in the Discussion.

In Appendix C, we derive spectral indices (and make them publicly available) and use them to qualitatively test, in a model independent way, the relationship between SFH, age, metallicity and mass, as derived with npFSF. We also use Mg and Fe UV lines in an attempt to constrain $[\alpha/\text{Fe}]$ abundance ratio. We remind that the relative

¹²The stellar mass interior to 1 kpc is given by

$$\mathcal{M}_{1\text{kpc}} = L_{1\text{kpc}} \times \left(\frac{\mathcal{M}_*}{L_{\text{tot}}} \right)_{\text{gal}} = \frac{\gamma(2n, x)}{\Gamma(2n)} \times \mathcal{M}_* \quad (5)$$

where

$$L_{1\text{kpc}} = 2\pi I_e R_e^2 n \frac{e^{b_n}}{b_n^{2n}} \gamma(2n, x) \quad (6)$$

is the luminosity within the central region of 1 kpc, L_{tot} is the total luminosity obtained by replacing in equation (6) $\gamma(2n, x)$ with the complete gamma function $\Gamma(2n)$ (Ciotti 1991), \mathcal{M}_* is the stellar mass of the galaxy, n is the Sérsic index, $x = b_n (R_1/R_e)^{1/n}$, and $\gamma(2n, x)$ is the incomplete gamma function. We assumed the analytic expression $b_n = 1.9992n - 0.3271$ (Capaccioli & Corwin 1989) to approximate the value of b_n .

¹¹The definition of MOGs is arbitrary. The selection $t_{\text{form}} < 1.2$ Gyr follows the choice of considering the ~ 1 Gyr of SF between the first star-forming events possibly taken place at $z \sim 15-20$ (cosmic time ~ 0.2 Gyr), and the end of the re-ionization ($z \sim 5-6$).

abundances of α elements and iron would provide crucial information in terms of star formation time-scales (e.g. Tinsley 1979; Matteucci 1994; see also Yan et al. 2021), and insights on possible variations of the IMF (e.g. La Barbera et al. 2013). Unfortunately, our analysis cannot be conclusive due to the still poor characterization of features in the UV spectral range. High S/N near-IR observations enabling to extend the spectral coverage up to $\lambda_{\text{rest}} \sim 5800 \text{ \AA}$ would be needed (see e.g. Bevacqua et al., in preparation).

7 SUMMARY OF THE RESULTS

In this work, we studied the stellar population properties of early-type and passive galaxies in the mass range $10 < \log(M_*/M_\odot) < 11.6$ at $1.0 < z < 1.4$ using VANDELS data. We first investigated the dependence of the age and metallicity estimates on the methods and assumptions usually adopted in this kind of analysis. This part of analysis showed that:

(i) Significant systematics exist among estimates from different stellar population models, different spectral ranges as often happens for samples at different redshifts, and when luminosity-weighted or mass-weighted quantities are considered. These systematics highlight the need for a uniform and homogeneous methodology at all redshifts.

On the basis of these results, the stellar population properties of galaxies have been derived using homogeneous measurements and method over the whole redshift range considered, i.e. same fitting code (STARLIGHT), simple stellar population models (EMILES) and fitting wavelength range ([3350–4350] \AA). We thus defined the age–mass and the metallicity–mass relations at these redshifts and studied the evolution of the stellar metallicity down to $z \sim 0$. We then reconstructed the SFH to study its relation with stellar population and physical properties of ETGs, and to trace the origin of the stellar MZR and its evolution. The main results are the following:

(i) The stellar metallicity and age of VANDELS passive galaxies at $\langle z \rangle \sim 1.2$ are positively correlated with their stellar mass, as for galaxies in the local Universe: higher mass galaxies host stars formed earlier and are more metal rich than most of the lower mass galaxies (see Section 4.1). These trends are, in a relative sense, independent of models or any particular assumption, as also confirmed by the observed trends of age and metallicity sensitive absorption spectral indices with mass (see Appendix). However, different models might change the absolute values one infers, as shown in Section 3. It is worth to remind the reader that these correlations are mainly due to the increasing lack of young and low-metallicity galaxies as the mass increases, an effect clearly present also in local (e.g. Gallazzi et al. 2005; McDermid et al. 2015; Barone et al. 2018) and intermediate redshift (e.g. Barone et al. 2022, Bevacqua et al., in preparation) samples.

(ii) The stellar metallicity of VANDELS passive galaxies with mass $10 < \log(M_*/M_\odot) < 11.6$ falls in the range $-0.35 < [Z/H] < 0.26$,¹³ with 67 per cent of them having $[Z/H] > 0.0$. This percentage rises to 90 per cent (19/21) at masses $\log(M_*/M_\odot) > 11$. These metallicity values agree with those estimated for field ETGs with comparable mass at $z \sim 1.4$ from Gargiulo et al. (2016), with those in the cluster XLSS0223 at $z \sim 1.2$ from Saracco et al. (2019) and with the few estimates at similar redshift from the literature (Onodera et al.

2015; Estrada-Carpenter et al. 2019; Kriek et al. 2019). On the contrary, for galaxies with mass $\log(M_*/M_\odot) \sim 11.2$ we estimate a mean stellar metallicity ($[Z/H] = 0.17 \pm 0.1$) $\sim 2\sigma$ higher than the metallicity estimated by Carnall et al. (2019) for VANDELS passive galaxies with the same redshift and mass ($[Z/H] = -0.07 \pm 0.08$ when corrected by the offset of $+0.06$ between models, see note 3). We verified that this significant difference is not due to any of the possible systematics introduced by different SSP models, spectral range or quantities considered. We believe that the difference is due to the very different methods adopted by the used fitting codes (see Section 4.3).

(iii) We do not detect any cosmic evolution of the metallicity–mass relation, either in the slope or in the normalization down to $z \sim 0$, as confirmed by the comparison with the relation derived from LEGA-C passive galaxies at $z \sim 0.65$ and with the local relations from the literature (Choi et al. 2014; Gallazzi et al. 2014; Peng et al. 2015). This confirms that the stellar metallicity is mainly defined during the formation of the dominant (in mass) stellar population. According to EMILES models, massive galaxies ($\log(M_*/M_\odot) > 11$) have supersolar metallicity and subsequent evolutionary processes (merging and/or SF, those processes able to modify the macroscopic properties of a galaxy) do not modify it (see Section 4.2).

(iv) The comparison of Mg and Fe UV spectral indices of VANDELS stacked spectra with those of stacked massive ETGs at $z \sim 0.38$ does not allow us to reach a firm conclusion about the possible evolution of the $[\alpha/\text{Fe}]$ ratio with redshift (see Appendix).

(v) The cumulative SFHs of VANDELS passive galaxies show that the fraction of stellar mass formed at early epochs increases with the mass of the galaxy, in agreement with the positive age–mass relation. On average, about 80 per cent of the stellar mass of very massive ($\log(M_*/M_\odot) > 11.3$) galaxies formed within the first 2 Gyr of cosmic time ($z > 3$), and 50 per cent within the first Gyr (by $z \sim 5$), results qualitatively in agreement with those derived for local ETGs (e.g. Thomas et al. 2005, 2010; McDermid et al. 2015; see Section 5).

(vi) Massive galaxies ($\log(M_*/M_\odot) > 11.0$) host old stellar populations ($t_{\text{form}} < 2$ Gyr) characterized by supersolar metallicity ($[Z/H] > 0.05$). These stars have been formed in short time ($\Delta t_{50} < 1$ Gyr) implying high star formation rates ($\text{SFR} > 100 M_\odot \text{ yr}^{-1}$) originating in high-mass density regions, $\log(\Sigma_{1\text{kpc}}) > 10 M_\odot/\text{kpc}^2$. This sharp picture tends to blur with decreasing mass: galaxies with intermediate mass, e.g. $\log(M_*/M_\odot) \sim 10.6$ can host either stars with sub-solar metallicity as old as those in massive galaxies, or younger stars with supersolar metallicity, depending on the duration of the star formation, shorter or longer respectively (see Section 6;), in agreement with other studies at intermediate redshift (Bevacqua et al., in preparation).

8 DISCUSSION AND CONCLUSIONS

To study the evolution of the stellar populations properties of galaxies, in particular of metallicity, it is essential to compare estimates obtained in a homogeneous way at the different redshifts. Our analysis showed, indeed, that such kind of estimates must be homogeneous to be comparable and that a metallicity variation (if any) should be considered relative since values, being dependent on the considered models, methodology, and spectral range, are not absolute.

The fact that positive correlations between stellar age and metallicity with the mass of the galaxies were already in place at redshift $z > 1.2$ implies that they were defined during the first 4 Gyr of cosmic time, as a result of the formation processes of stellar mass and galaxies. The lack of evolution in stellar metallicity of galaxies with $\log(M_*/M_\odot) > 10.6$ in the last 9–10 Gyr ($0 < z < 1.4$) and, perhaps,

¹³We stress that $[Z/H] = 0.26$ is the upper limit of the metallicity of the EMILES models considered here. Therefore, we cannot rule out metallicity values even higher than this one for some massive galaxies.

~ 12 Gyr ($z < 3.35$) confirms on the one hand that evolutionary processes (merging and/or later SF) do not significantly influence the stellar metallicity of galaxies, at least of the massive ones, and that the metallicity of galaxies is determined at early epochs, during the main massive star formation event.

First, let us consider the constraints on the evolution imposed by the lack of metallicity evolution over the last ~ 9 Gyr. Once a given stellar mass is seen assembled in a galaxy (progenitor) with a given age and metallicity $[Z/H]$, to generate a descendant with a metallicity, e.g. 0.1 dex lower (higher), a similar mass with metallicity, at least, half (twice) of the progenitor must be added. The lower the accreted mass fraction, the higher its metallicity offset with respect to the progenitor and vice versa. The same reasoning applies to the mean metallicity of the population of galaxies: to change it by $\Delta[Z/H] = \pm 0.1$ a similar amount of galaxies (with similar mass distribution) with metallicity half or twice should be added to the population. Therefore, a lack of metallicity evolution especially for masses $\sim 10^{11} M_{\odot}$ or higher is not surprising. New star formation of this magnitude is not seen at $z < 1.4$ (e.g. Madau & Dickinson 2014, and references therein), the mass growth of massive galaxies in this redshift range, if any, is negligible (e.g. Muzzin et al. 2013) and, most importantly, the stellar populations in local massive ETGs are, on average, old as expected for passive evolution (e.g. Gallazzi et al. 2005; Thomas et al. 2005). Major merging (mass ratio $\sim 1:1$) would leave the metallicity unchanged since it is expected that similar mass galaxies have similar stellar metallicity. Minor mergers would not affect the mean stellar population properties of a massive progenitor given the low mass (expected metal-poor) accreted, even if they could efficiently affect the structural properties of a galaxy (e.g. Ciotti, Lanzoni & Volonteri 2007; Bezanson et al. 2009; Hopkins et al. 2009; Naab, Johansson & Ostriker 2009).¹⁴ For instance, Hirschmann et al. (2015) showed that minor mergers can steepen the stellar metallicity gradients due to the accretion of metal-poor stars at the outskirts of massive galaxies, leaving unchanged their mean metallicity.

Newly formed massive galaxies can play a role in the evolution of the mean properties of the population of early-type and passive galaxies at different redshifts (progenitor bias, e.g. Fagioli et al. 2016). However, considering the arguments above, their stellar population properties cannot be substantially different from galaxies already assembled, apart from their size and mass density. For these reasons, and independently of our result, it is unclear how the fall by ~ 0.3 dex in the mean stellar metallicity of galaxies with $\log(M_*/M_{\odot}) > 11.2$ found by Carnall et al. (2022) at $z \sim 1.2$ with respect to the local Universe can be justified, unless we assume that local estimates are all biased (for some reason) towards metallicity values ~ 2 times higher than the value they find at that redshift. A comparison among estimates homogeneously obtained with their methodology on different samples down to the local Universe (as we did with the method adopted in this work) would clarify whether the strong evolution of the metallicity they found is due to the different methodology used in the comparison estimates.

Much more indicative than the lack of metallicity evolution is the nature of the stellar MZR. Our analysis shows an increasing lack of young and low-metallicity galaxies as the stellar mass increases: there are no massive galaxies as young as some of the low-mass galaxies, and no massive galaxies with metallicity as low as some of the

low-mass galaxies. On the contrary, low-mass galaxies can display either supersolar metallicity, as high as massive ones, associated with young ages, or sub-solar metallicity associated with old ages. This is evident in our analysis, for galaxies at $z \sim 1.4$, in those at intermediate redshift (see e.g. Bevacqua et al., in preparation) and in the local Universe (see e.g. Gallazzi et al. 2005; Asari et al. 2009; Peng et al. 2015).

Given the redshift of our galaxies, these differences must be the results of different SFHs and/or initial conditions. The fact that the most massive galaxies host old stellar populations characterized by supersolar metallicity, formed within short time and high SFR associated with high central stellar mass densities are all properties expected for stellar systems that form in the early Universe (e.g. Wellons et al. 2015; see also Eappen et al. 2022). The higher mean density of the early Universe favours higher molecular gas densities and, according to the Kennicutt–Schmidt law (Kennicutt 1998), higher SFR densities. The high SFR determined by the short time and the high mass, efficiently enriches the interstellar medium (e.g. Matteucci 1994; Calura et al. 2009). This enhances the stellar metallicity more efficiently than in lower mass galaxies even in case of outflows (at first guess, proportional to the SFR; e.g. Calura et al. 2009; Spitoni et al. 2010; Spitoni, Vincenzo & Matteucci 2017). Indeed, low-mass galaxies are expected to have a lower ability than high-mass ones in retaining the metals because of their shallower potential well (e.g. Dekel & Silk 1986; De Lucia et al. 2004; Tremonti et al. 2004; Kobayashi, Springel & White 2007; Finlator & Davé 2008). This is confirmed by the observed correlation between velocity dispersion, a direct measure of the potential well, and stellar metallicity (e.g. Jørgensen 1999; Trager et al. 2000; Harrison et al. 2011; McDermid et al. 2015).

For short SF time-scales, such as those of massive galaxies, low-mass galaxies show much lower stellar metallicities associated with old ages. The SFR is low given the low mass and the low density, the enrichment of the ISM is less efficient than in high-mass galaxies and requires much longer time to enhance the stellar metallicity. Indeed, for low-mass galaxies, metallicity increases as the duration of the SF increases and it is, therefore, associated with younger stellar population. Calura et al. (2009) show that a variable star formation efficiency from low- to high-mass galaxies (higher in massive galaxies) can produce the observed stellar MZR. Spitoni et al. (2010) conclude that a plausible scenario is variable star formation efficiency coupled to galactic winds becoming more important in low-mass galaxies (see also Spitoni et al. 2017). We cannot exclude that SF efficiency changes systematically with mass, even if it is not clear what physical mechanism can change the SF efficiency. In this respect, it is worth to remind that GAEA correctly reproduces the shape of the stellar MZR for passive galaxies at $z < 1.4$ as well as the lack of evolution at large stellar masses without assuming any scaling of the SF efficiency with stellar mass.

To conclude, our results show that some properties characterizing the SFH change systematically among galaxies and that these variations underlie the MZR. In particular, we find evidence that the SFR and the duration of the SF are the properties playing a major role in defining the stellar MZR and age–mass relations of ETGs with the fundamental complicity of the underlying mass which modulates the retention of metals as the mass increases.

ACKNOWLEDGEMENTS

Based on public data products from observations made with ESO Telescopes at the La Silla or Paranal Observatories under programs ID 194.A-2003(E-K) and ID 194.A-2005(A-F). This work

¹⁴Major and minor mergers can play a role in the evolution of galaxies even if in a limited way because of the large scatter that they would introduce in the scaling relations (see e.g. Nipoti et al. 2009). We do not discuss further this topic since it is out of the scope of this analysis.

is partially based also on observations carried out with the ESO Very Large Telescope under programme ID 085.A-0135, with the Large Binocular Telescope (LBT) under programs ID 2015-2016-28 and ID 2017B-C2743-3. PS, FLB, RDP, DB, FF, AP, CS, and CT acknowledge support by the grant PRIN-INAF-2019 1.05.01.85 (#11). CS is supported by an ‘Hintze Fellow’ at the Oxford Centre for Astrophysical Surveys, which is funded through generous support from the Hintze Family Charitable Foundation. PS would like to thank V. Lynd for the useful discussions. PS thanks Anna Gallazzi and Stefano Zibetti for their careful reading and comments on the manuscript.

DATA AVAILABILITY

The VANDELS data underlying this article are available at <http://vandel.inaf.it/> or via the ESO Archive (<https://www.eso.org/qi/>). The codes STARLIGHT and PPXF are publicly available. The spectral indices generated in this research are published in electronic form (Tables B1 and B2). The remaining derived data will be shared on reasonable request to the corresponding author.

REFERENCES

- Asari N. V., Cid Fernandes R., Stasińska G., Torres-Papaqui J. P., Mateus A., Sodr e L., Schoenell W., Gomes J. M., 2007, *MNRAS*, 381, 263
- Asari N. V., Stasińska G., Cid Fernandes R., Gomes J. M., Schlickmann M., Mateus A., Schoenell W., 2009, *MNRAS*, 396, L71
- Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, *ApJ*, 527, 54
- Barone T. M. et al., 2018, *ApJ*, 856, 64
- Barone T. M., D’Eugenio F., Colless M., Scott N., 2020, *ApJ*, 898, 62
- Barone T. M. et al., 2022, *MNRAS*, 512, 3828
- Beverage A. G., Kriek M., Conroy C., Bezanson R., Franx M., van der Wel A., 2021, *ApJ*, 917, L1
- Bezanson R., van Dokkum P. G., Tal T., Marchesini D., Kriek M., Franx M., Coppi P., 2009, *ApJ*, 697, 1290
- Blakeslee J. P. et al., 2003, *ApJ*, 596, L143
- Bower R. G., Lucey J. R., Ellis R. S., 1992, *MNRAS*, 254, 601
- Bruzual A. G., 1983, *ApJS*, 53, 497
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Calabr o A. et al., 2021, *A&A*, 646, A39
- Calura F., Pipino A., Chiappini C., Matteucci F., Maiolino R., 2009, *A&A*, 504, 373
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682
- Capaccioli M., Corwin Jr. G. H., 1989, *Advanced Series in Astrophysics and Cosmology*, Vol. 4, World Scientific Europe, London, UK
- Cappellari M., 2017, *MNRAS*, 466, 798
- Cappellari M. et al., 2013, *MNRAS*, 432, 1862
- Carnall A. C., McLure R. J., Dunlop J. S., Dav e R., 2018, *MNRAS*, 480, 4379
- Carnall A. C. et al., 2019, *MNRAS*, 490, 417
- Carnall A. C. et al., 2022, *ApJ*, 929, 131
- Carollo C. M. et al., 2013, *ApJ*, 773, 112
- Chabrier G., 2003, *PASP*, 115, 763
- Chavez M., Bertone E., Buzzoni A., Franchini M., Malagnini M. L., Morossi C., Rodriguez-Merino L. H., 2007, *ApJ*, 657, 1046
- 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
- Cid Fernandes R., Mateus A., Sodr e L., Stasińska G., Gomes J. M., 2005, *MNRAS*, 358, 363
- Cid Fernandes R., Asari N. V., Sodr e L., Stasińska G., Mateus A., Torres-Papaqui J. P., Schoenell W., 2007, *MNRAS*, 375, L16
- Ciotti L., 1991, *A&A*, 249, 99
- Ciotti L., Lanzoni B., Volonteri M., 2007, *ApJ*, 658, 65
- Cleveland W. S., Devlin S. J., 1988, *J. Am. Stat. Assoc.*, 83, 596
- Conroy C., van Dokkum P. G., 2012, *ApJ*, 760, 71
- Conroy C., Graves G. J., van Dokkum P. G., 2014, *ApJ*, 780, 33
- Cowie L. L., Songaila A., Hu E. M., Cohen J. G., 1996, *AJ*, 112, 839
- Cullen F. et al., 2019, *MNRAS*, 487, 2038
- Davidge T. J., Clark C. C., 1994, *AJ*, 107, 946
- De Lucia G., Kauffmann G., White S. D. M., 2004, *MNRAS*, 349, 1101
- De Lucia G., Fontanot F., Hirschmann M., 2017, *MNRAS*, 466, L88
- Dekel A., Silk J., 1986, *ApJ*, 303, 39
- Dekel A. et al., 2009, *Nature*, 457, 451
- Djorgovski S., Davis M., 1987, *ApJ*, 313, 59
- Dressler A., Lynden-Bell D., Burstein D., Davies R. L., Faber S. M., Terlevich R., Wegner G., 1987, *ApJ*, 313, 42
- Eappen R., Kroupa P., Wittenburg N., Haslbauer M., Famaey B., 2022, *MNRAS*, 516, 1081
- Estrada-Carpenter V. et al., 2019, *ApJ*, 870, 133
- Estrada-Carpenter V. et al., 2020, *ApJ*, 898, 171
- Fagioli M., Carollo C. M., Renzini A., Lilly S. J., Onodera M., Tacchella S., 2016, *ApJ*, 831, 173
- Falc n-Barroso J., S nchez-Bl azquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, *A&A*, 532, A95
- Fanelli M. N., O’Connell R. W., Burstein D., Wu C.-C., 1990, *ApJ*, 364, 272
- Ferreras I. et al., 2009, *ApJ*, 706, 158
- Finlator K., Dav e R., 2008, *MNRAS*, 385, 2181
- Fontanot F., De Lucia G., Hirschmann M., 2017, *MNRAS*, 464, 3812
- Fontanot F. et al., 2021, *MNRAS*, 504, 4481
- 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
- Gargiulo A., Saracco P., Tamburri S., Lonoce I., Ciocca F., 2016, *A&A*, 592, A132
- Gallazzi A. R., Pasquali A., Zibetti S., Barbera F. L., 2021, *MNRAS*, 502, 4457
- Garilli B. et al., 2021, *A&A*, 647, A150
- Ge J., Yan R., Cappellari M., Mao S., Li H., Lu Y., 2018, *MNRAS*, 478, 2633
- Ge J., Mao S., Lu Y., Cappellari M., Yan R., 2019, *MNRAS*, 485, 1675
- Girardi L., Bressan A., Bertelli G., Chiosi C., 2000, *A&AS*, 141, 371
- Gorgas J., Cardiel N., Pedraz S., Gonz alez J. J., 1999, *A&AS*, 139, 29
- Guo Q. et al., 2016, *MNRAS*, 461, 3457
- Harrison C. D., Colless M., Kuntschner H., Couch W. J., de Propris R., Pracy M. B., 2011, *MNRAS*, 413, 1036
- Hirschmann M., Naab T., Ostriker J. P., Forbes D. A., Duc P.-A., Dav e R., Oser L., Karabal E., 2015, *MNRAS*, 449, 528
- Hopkins P. F., Bundy K., Murray N., Quataert E., Lauer T. R., Ma C.-P., 2009, *MNRAS*, 398, 898
- Horne K., 1986, *PASP*, 98, 609
- J rgensen I., 1999, *MNRAS*, 306, 607
- J rgensen I., Franx M., Kjaergaard P., 1996, *MNRAS*, 280, 167
- J rgensen I., Chiboucas K., Berkson E., Smith O., Takamiya M., Villaume A., 2017, *AJ*, 154, 251
- Kauffmann G. et al., 2003, *MNRAS*, 341, 54
- Kennicutt R. C. Jr, 1998, *ARA&A*, 36, 189
- Kobayashi C., Springel V., White S. D. M., 2007, *MNRAS*, 376, 1465
- Kochanek C. S. et al., 2012, *ApJS*, 200, 8
- Kodama T., Arimoto N., 1997, *A&A*, 320, 41
- Kriek M. et al., 2016, *Nature*, 540, 248
- Kriek M. et al., 2019, *ApJ*, 880, L31
- La Barbera F., Ferreras I., Vazdekis A., de la Rosa I. G., de Carvalho R. R., Trevisan M., Falc n-Barroso J., Ricciardelli E., 2013, *MNRAS*, 433, 3017
- Le Cras C., Maraston C., Thomas D., York D. G., 2016, *MNRAS*, 461, 766
- Leitherer C., Ortiz Ot alvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, *ApJS*, 189, 309
- Lemaux B. C. et al., 2019, *MNRAS*, 490, 1231

Lonoce I., Longhetti M., Saracco P., Gargiulo A., Tamburri S., 2014, *MNRAS*, 444, 2048

Lonoce I. et al., 2015, *MNRAS*, 454, 3912

Lonoce I., Maraston C., Thomas D., Longhetti M., Parikh T., Guarnieri P., Comparat J., 2020, *MNRAS*, 492, 326

Madau P., Dickinson M., 2014, *ARA&A*, 52, 415

Maraston C., Strömbäck G., 2011, *MNRAS*, 418, 2785

Maraston C., Nieves Colmenáez L., Bender R., Thomas D., 2009, *A&A*, 493, 425

Maseda M. V. et al., 2021, *ApJ*, 923, 18

Matteucci F., 1994, *A&A*, 288, 57

McDermid R. M. et al., 2015, *MNRAS*, 448, 3484

McLure R. J. et al., 2018, *MNRAS*, 479, 25

Mei S. et al., 2006a, *ApJ*, 639, 81

Mei S. et al., 2006b, *ApJ*, 644, 759

Mei S. et al., 2012, *ApJ*, 754, 141

Mo H., van den Bosch F. C., White S., 2010, *Galaxy Formation and Evolution*. Cambridge Univ. Press, Cambridge

Muzzin A. et al., 2013, *ApJS*, 206, 8

Naab T., Johansson P. H., Ostriker J. P., 2009, *ApJ*, 699, L178

Newman A. B., Ellis R. S., Andreon S., Treu T., Raichoor A., Trinchieri G., 2014, *ApJ*, 788, 51

Nipoti C., Treu T., Auger M. W., Bolton A. S., 2009, *ApJ*, 706, L86

Okamoto T., Nagashima M., Lacey C. G., Frenk C. S., 2017, *MNRAS*, 464, 4866

Onodera M. et al., 2015, *ApJ*, 808, 161

Peng Y., Maiolino R., Cochrane R., 2015, *Nature*, 521, 192

Pentericci L. et al., 2018, *A&A*, 616, A174

Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2004, *ApJ*, 612, 168

Pipino A., Matteucci F., 2004, *MNRAS*, 347, 968

Poggianti B. M., Barbaro G., 1997, *A&A*, 325, 1025

Salvador-Rusiñol N., Vazdekis A., La Barbera F., Beasley M. A., Ferreras I., Negri A., Dalla Vecchia C., 2020, *Nat. Astron.*, 4, 252

Sandage A., Visvanathan N., 1978, *ApJ*, 225, 742

Saracco P., Gargiulo A., Longhetti M., 2012, *MNRAS*, 422, 3107

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., 2020a, *MNRAS*, 491, 1777

Saracco P. et al., 2020b, *ApJ*, 905, 40

Serven J., Worthey G., Briley M. M., 2005, *ApJ*, 627, 754

Skelton R. E., Bell E. F., Somerville R. S., 2012, *ApJ*, 753, 44

Spiniello C., Trager S. C., Koopmans L. V. E., Chen Y. P., 2012, *ApJ*, 753, L32

Spitoni E., Calura F., Matteucci F., Recchi S., 2010, *A&A*, 514, A73

Spitoni E., Vincenzo F., Matteucci F., 2017, *A&A*, 599, A6

Tacchella S. et al., 2022, *ApJ*, 926, 134

Tamburri S., Saracco P., Longhetti M., Gargiulo A., Lonoce I., Ciocca F., 2014, *A&A*, 570, A102

Thomas D., Maraston C., Bender R., 2003, *MNRAS*, 343, 279

Thomas D., Maraston C., Bender R., Mendes de Oliveira C., 2005, *ApJ*, 621, 673

Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, *MNRAS*, 404, 1775

Thomas D., Maraston C., Johansson J., 2011, *MNRAS*, 412, 2183

Tinsley B. M., 1979, *ApJ*, 229, 1046

Trager S. C., Worthey G., Faber S. M., Burstein D., González J. J., 1998, *ApJS*, 116, 1

Trager S. C., Faber S. M., Worthey G., González J. J., 2000, *AJ*, 120, 165

Tremonti C. A. et al., 2004, *ApJ*, 613, 898

Trussler J., Maiolino R., Maraston C., Peng Y., Thomas D., Goddard D., Lian J., 2020, *MNRAS*, 491, 5406

van der Wel A. et al., 2012, *ApJS*, 203, 24

van der Wel A. et al., 2014, *ApJ*, 788, 28

van der Wel A. et al., 2016, *ApJS*, 223, 29

van der Wel A. et al., 2021, *ApJS*, 256, 44

van Dokkum P. G. et al., 2008, *ApJ*, 677, L5

Vazdekis A., Peletier R. F., Beckman J. E., Casuso E., 1997, *ApJS*, 111, 203

Vazdekis A. et al., 2015, *MNRAS*, 449, 1177

Vazdekis A., Koleva M., Ricciardelli E., Röck B., Falcón-Barroso J., 2016, *MNRAS*, 463, 3409

Wellons S. et al., 2015, *MNRAS*, 449, 361

Willis J. P. et al., 2020, *Nature*, 577, 39

Worthey G., 1994, *ApJS*, 95, 107

Worthey G., Ottaviani D. L., 1997, *ApJS*, 111, 377

Yan Z., Yerabkova T., Kroupa P., 2019, *A&A*, 632, 110

SUPPORTING INFORMATION

Supplementary data are available at [MNRAS](https://www.mnras.org) online.

Table C1. Absorption line spectral indices.

Table C2. Errors (based on photon statistics) on measurements of absorption line indices.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: SIMULATING MIXED STELLAR POPULATIONS

We checked whether and how the spectral range [2600–3350] Å affects the results of the full spectral fitting performed over a range extending up to $\lambda_{\text{rest}} \leq 4350$ Å, in case of multiple stellar populations. To this end, we simulated a galaxy (a template) with a mixed stellar population starting from EMILES SSP models. The simulated galaxy has the bulk (weight = 1) of the stellar mass represented by a SSP 3 Gyr old. To this stellar population we added 10 per cent of stellar mass (weight = 0.1) represented by a SSP 0.2 Gyr old. Both the SSPs have metallicity $[Z/H] = 0.15$. Therefore, the resulting stellar population is 2.75 Gyr old and has a metallicity $[Z/H] = 0.15$.

This template has been normalized to the VANDELS stacked spectrum with $\log(M^*/M_{\odot}) = 11$ in the range 4100–4250 Å. Then, the residuals of the full spectral fitting to the stacked spectrum have been randomly reshuffled 30 times in λ and summed to the template to reproduce similar S/N and mimic the true uncertainties. In Fig. A1,

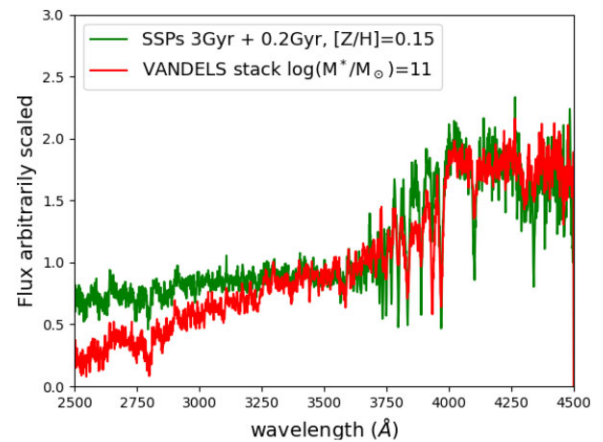


Figure A1. Simulated galaxy with mixed stellar populations. The simulated galaxy (green curve) is the weighted sum of an SSP 3 Gyr old weighted 1, and an SSP 0.2 Gyr old weighted 0.1, both with metallicity $[Z/H] = 0.15$. The VANDELS stacked spectrum with $\log(M^*/M_{\odot}) = 11$ is shown for comparison (red curve). The two spectra are normalized to the mean flux in the range [4100–4250] Å.

a realization of the template with summed residuals is shown as example (green curve).

Finally, we run STARLIGHT on these simulated spectra with different noise, for the two spectral ranges [3350–4350] Å and [2600–4350] Å. The range [3350–4350] Å provided a mean age of 3.25 ± 0.2 Gyr and a mean metallicity $[Z/H] = 0.11 \pm 0.03$; the range [2600–4350] Å provided a mean age of 3.30 ± 0.2 Gyr and a mean metallicity $[Z/H] = 0.08 \pm 0.02$. Therefore, the presence of a young stellar component, even if accounting just for about ten per cent of the mass, can induce metallicity estimates systematically lower (at 1σ) when the fit extends to UV range with respect to the metallicity values derived from optical spectral range.

We remind that this exercise is not meant to probe the nature of the systematic in the metallicity estimates shown in Section 3.2, since it is based on the assumption that spectral features in the UV are well characterized and modelled. This exercise simply shows that a mixed stellar population could, in principle, affect in a systematic way the estimate of the mean stellar metallicity when the UV spectral range, dominated by younger stellar components and affected by dust extinction, is taken into account.

We used these simulations also to probe the internal age–metallicity degeneracy of STARLIGHT full spectral fitting. This to assess whether and how much the stellar population properties we derived for VANDELs galaxies can be affected. To this end, we searched for a correlation between the age and the metallicity values derived for the 30 renditions of the simulated galaxy. We did not detect a correlations among the scattered values, meaning that the possible systematic is much smaller than the uncertainties due to the S/N of our spectra.

APPENDIX B: EXTENDING THE SPECTRAL FITTING TO >5000 Å

In this appendix, we check for the possible different stellar population properties resulting from full spectral fitting performed over different rest-frame wavelength ranges. In particular, a fit performed over [3350–4350] Å, as in our analysis, and a fit extending to ~ 6000 Å, as in Carnall et al. (2022) to verify whether this is the reason of the discrepancy between our and their stellar metallicity estimates. We used the high S/N BOSS stacked spectra of massive ETGs at

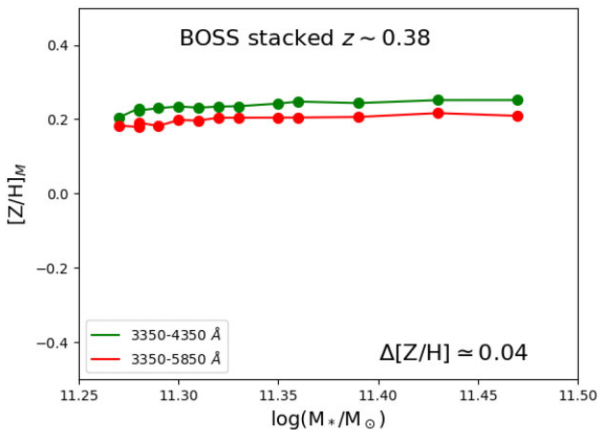


Figure B1. Mass-weighted metallicity of massive ETGs at $z \sim 0.38$ as a function of mass as resulting from the fitting of BOSS stacked spectra (Salvador-Rusiñol et al. 2020) over two different wavelength ranges, 3350–4350 Å (green curve) and 3350–5850 Å (red curve). The fitting was performed with STARLIGHT and EMILES models.

$z \sim 0.38$ (Salvador-Rusiñol et al. 2020) whose spectral coverage allow us to perform this comparison. The resulting stellar metallicity values resulting from the STARLIGHT full spectral fitting are shown in Fig. B1. The extension of the fitting to ~ 6000 Å affects the metallicity by $\Delta[Z/H] = -0.04$ with respect to the fit performed in the range [3350–4350] Å. This metallicity offset is less than half of the error associated with the metallicity estimate of VANDELs stacked spectra and is one seventh of the difference ($\Delta[Z/H] \sim 0.3$) between our stellar metallicity estimate for galaxies with $\log(M_*/M_\odot) > 11$ and the estimate by Carnall et al. (2022). Therefore, the extension of the fitting to wavelength $\lambda > 4350$ Å does not justify the discrepancy between our and their estimate.

APPENDIX C: ABSORPTION LINE SPECTRAL INDICES

Absorption line indices in the rest-frame wavelength range 2600–4200(4350)Å were measured for the whole sample of passive galaxies (for those at $z < 1.3$) and for the stacked galaxies making use of the software LECTOR.¹⁵ The list of the measured indices and their definition are reported in Table C3. The strength of the 4000 Å break was measured according to the D4000 definition by Bruzual (1983; see also Gorgas et al. 1999) and the D_n definition by Balogh et al. (1999). The measured indices were corrected for the broadening of the spectra due to the low instrumental resolution ($R \sim 600$). The corrections were obtained by comparing the indices measured on an SSP 3 Gyr old (the median age of the sample) smoothed to the instrumental resolution ($R \simeq 600$), and those of the same model at the nominal resolution of the EMILES spectral library. We verified that considering ages older or younger than 3 Gyr in the range 1–5 Gyr would produce completely negligible variations in the corrections. The corrected measured indices with their errors are summarized in Tables C1 and C2, respectively.

Fig. C1 shows the metallicity sensitive features Mg_W , BL3584, CN3883, and FeI of ETGs, LTGs, and unclassified galaxies as a function of their stellar mass. The black line represent the best-fitting linear relation. An increasing trend with the mass for the indices considered is visible even if the correlation is not statistically significant given the large scatter.

C1 Age and metallicity versus stellar mass and SFH

In this section, we use the indices to qualitatively verify, in a model independent way, the relationships between SFH, age, metallicity, and mass of galaxies resulting from the analysis in Sections 4, 5, and 6. In the upper panel of Fig. C2, the age sensitive indices $H\delta_A$, $H\gamma_F$, and D4000, and metallicity sensitive indices Fe($\lambda 2604$), Mg_W , and CN3883 measured on stacked spectra are shown as a function of the stellar mass. In the lower panel, the same indices predicted by EMILES models are shown as a function of the age of the stellar population for different metallicity values. It is important to note that all the indices are sensitive to both age and metallicity, and in particular UV spectral indices are strongly sensitive to even tiny fractions of young stars in a stellar population (see Salvador-Rusiñol et al. 2020) Therefore, individually, they cannot be used to directly derive the age and the metallicity of the underlying stellar population. However, taken together, they provide information on the SFH and the metallicity independently of models.

¹⁵http://www.iac.es/galeria/vazdekis/vazdekis_software.html

Table C1. Absorption line spectral indices (full table available in electronic form).

ID	Mg II (Å)	Mg I (Å)	MgW (Å)	Fe I (Å)	CN3883 (mag)	Ca II(H&K) (Å)	H δ_A (Å)	H δ_F (Å)	CN1 (mag)	CN2 (mag)	Ca4227 (Å)	G-b (Å)	D4000	D _n
UDS021385	26.2	5.5	100.0	15.6	0.166	22.1	1.9	1.8	0.034	0.066	1.4	3.6	1.48	1.24
UDS197769	15.8	8.4	26.1	0.4	0.096	12.2	5.4	4.4	-0.081	-0.043	2.0	3.0	1.38	1.15
CDFS126089	5.5	2.9	29.3	3.4	0.167	18.5	3.6	1.9	-0.107	-0.039	1.0	3.1	1.39	1.10

Notes. The indices are corrected for the broadening of the spectra due to the low instrumental resolution (see Section 2). Values equal to 99.99 indicate failed or unreliable measurements because of problems in the spectrum (e.g. uncovered wavelength range, strong emission/absorption sky line).

Table C2. Errors (based on photon statistics) on measurements of absorption line indices (full table available in electronic form).

ID	eMg II	eMg I	eMgW	eFe I	eCN3883	eCa IIHK	eH δ_A	eH δ_F	eCN1	eCN2	eCa4227	eG-b	eD4000	eD _n
UDS021385	1.6	0.8	100.0	1.6	0.03	1.3	0.3	0.2	0.02	0.03	0.1	0.4	0.04	0.03
UDS197769	0.6	0.4	1.3	1.0	0.03	1.2	0.2	0.1	0.02	0.03	0.1	0.3	0.03	0.03
CDFS126089	1.2	0.7	2.1	1.5	0.03	1.4	0.3	0.2	0.02	0.03	0.2	0.4	0.03	0.03

Table C3. Definition of absorption line spectral indices.

Index	Blue cont.	Feature	Red cont.	A/M ^a	Ref
Fe2609	2562–2588	2596–2622	2647–2673	A	1,2,3,4
BL2740	2647–2673	2736–2762	2762–2782	A	1,2,3,4
Mg II	2762–2782	2784–2814	2818–2838	A	1,2,3,4
Mg I	2818–2838	2839–2865	2906–2936	A	1,2,3,4
Mg _{wide}	2470–2670	2670–2870	2930–3130	A	1,2,3,4
Fe I	2906–2936	2965–3025	3031–3051	A	1,2,3,4
D4000	3750–3950		4050–4250		9,10
D _n	3850–3950		4000–4100		11
Mg3334	3310–3320	3328–3340	3342–3355	A	4, 6
BL3584	3540–3569	3570–3600	3601–3630	A	...
CN3883	3760–3780	3780–3900	3900–3915	M	5
CaII(H&K)	3900–3915	3915–4000	4000–4020	A	6
H δ_A	4041–4079	4083–4122	4128–4161	A	7,8
H δ_F	4057–4088	4091–4112	4114–4137	A	7,8
CN1	4080–4117	4142–4177	4244–4284	M	7,8
CN2	4083–4096	4142–4177	4244–4284	M	7,8
Ca4227	4211–4219	4222–4234	4241–4251	A	7,8
G band	4266–4282	4281–4316	4318–4335	A	7,8

Note. ^aA = EW(Å), M = magnitude; 1 - Fanelli et al. (1990); 2 - Chavez et al. (2007), 3 - Maraston et al. (2009), 4 - Vazdekis et al. (2016), 5 - Davidge & Clark (1994), 6 - Serven, Worthey & Briley (2005), 7 - Worthey & Ottaviani (1997), 8 - Trager et al. (1998), 9 - Bruzual (1983), Gorgas et al. (1999), 11 - Balogh et al. (1999).

The Balmer lines H δ_A and H γ_F are prominent in hot, high-mass (1.5–2 M_⊙) rapidly evolving stars. Their main-sequence lifetimes are less than 1 Gyr (e.g. Poggianti & Barbaro 1997). Therefore, the strength of these indices are related to the time since the last episode of star formation, the weaker the feature the longer the time elapsed since last burst. This is shown in the lower panel of Fig. C2 where the expected H δ_A and H γ_F indices for EMILES SSPs are shown. These two indices measured on stacked spectra (upper panel) are anticorrelated with the mass, suggesting that higher mass galaxies experienced their last burst at earlier epochs than lower mass ones, in agreement with the SFH resulting from the FSF (see Figs 2 and 9).

The D4000 index (Bruzual 1983) is a measure of the discontinuity produced by the opacity of the stellar atmospheres that depends on the ionized metals hence, on the stellar temperature. Hot (high-mass) stars do not contribute to the amplitude of D4000 being their opacity low because their elements are ionized. Therefore, the D4000 index is dominated by low-mass stars, its amplitude is related to the age

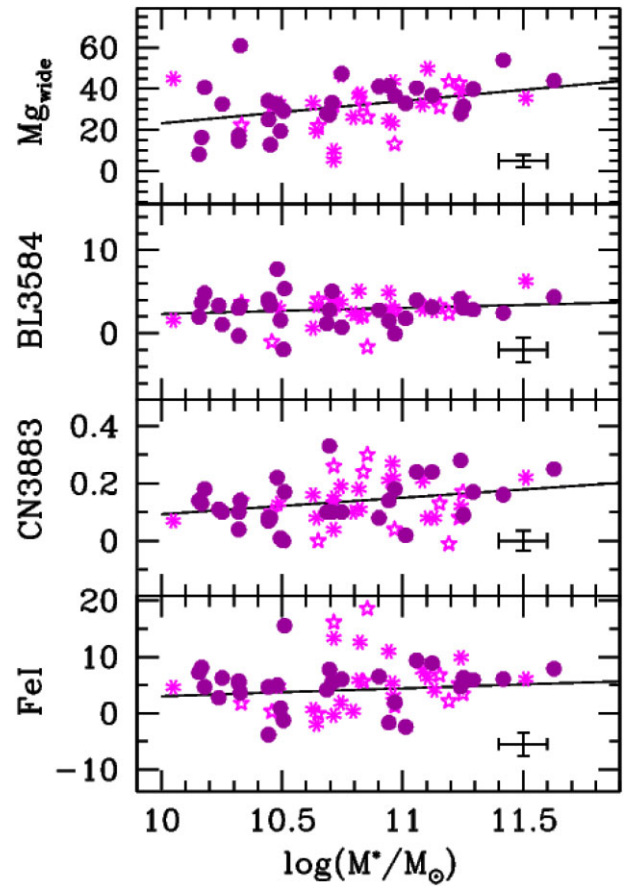


Figure C1. The metallicity sensitive indices Mg_{wide}, BL3584, CN3883, and Fe I (see Table C3) are shown as a function of the stellar mass of ETGs (purple filled dots), unclassified galaxies (light purple skeletal symbols), and S galaxies (stars). The black line in each panel is the best-fitting line to the data (orthogonal fit). The black crosses on the bottom-right corners represent the typical formal error (based on photon statistic) on the data.

(and metallicity) of the bulk of the stellar population: it gets larger for increasing age (and metallicity) as shown in the lower panel of Fig. C2. The D4000 index measured on stacked spectra shows a clear positive trend with the mass indicating that the age of the stellar population increases with the mass of the galaxy, in agreement with the results derived from the FSF analysis summarized in Fig. 8 and in

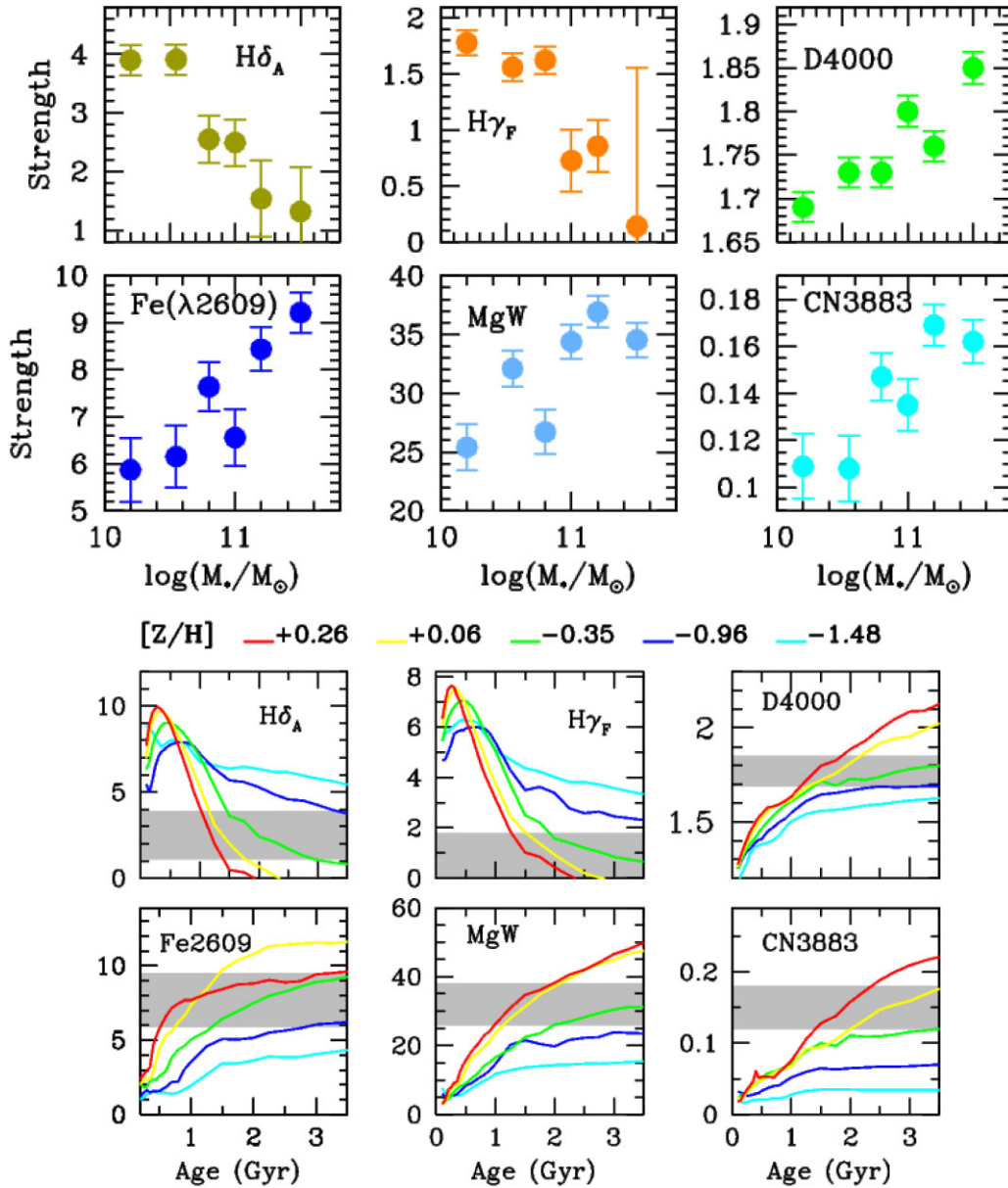


Figure C2. Upper panel: The age sensitive indices $H\delta_A$, $H\gamma_F$ and D4000, and metallicity sensitive indices Fe($\lambda 2604$), MgW and CN3883 (see Table C3) measured on stacked spectra (large filled symbols) are shown as a function of the stellar mass. Lower panel: The different curves are the same indices above as predicted by EMILES models for different metallicity values shown as a function of age of the stellar population. The grey regions mark the range of the values of the indices measured on stacked spectra.

equation (4). Therefore, the behaviour of the Balmer lines, combined with the D4000 index, suggest that galaxies with larger stellar masses formed their stars in a shorter SF episode that occurred earlier on in cosmic time. This also implies that they have higher SFR. This is in agreement with what shown in the lower panel of Fig. 10.

The strength of the indices Fe($\lambda 2604$), MgW and CN3883 is sensitive to the abundance of those elements in the stellar atmospheres. The strength of the three indices (upper panel of Fig. C2) increases systematically with the mass, suggesting a corresponding increase of the metallicity.¹⁶ It is interesting to note that the indices predicted

¹⁶However, notice that CN3883 is mostly sensitive to C and N abundance ratios, with both elements giving only a minor contribution to the total metallicity of a stellar metallicity.

by models (lower panel) for metallicity lower than solar (green, blue, and cyan curves) do not increase significantly with age in the range 1–3.5 Gyr, and do not match the observed values, especially MgW and CN3883. The lower indices values measured on the stacked spectra (whose range is represented by the grey regions in the figure) associated with lower mass galaxies, seem to rule out metallicity values much lower than $[Z/H] \sim -0.35$. At the same time, the upper limits associated with high-mass galaxies require metallicity higher than solar to be matched. Analogous conclusions can be reached looking at the age sensitive indices $H\delta_A$, $H\gamma_F$ and D4000, whose range of values provide a similar lower limit to metallicity when compared to the indices predicted by models.

In conclusion, the metallicity-sensitive indices that we measured all suggest a positive trend of the stellar metallicity with the stellar

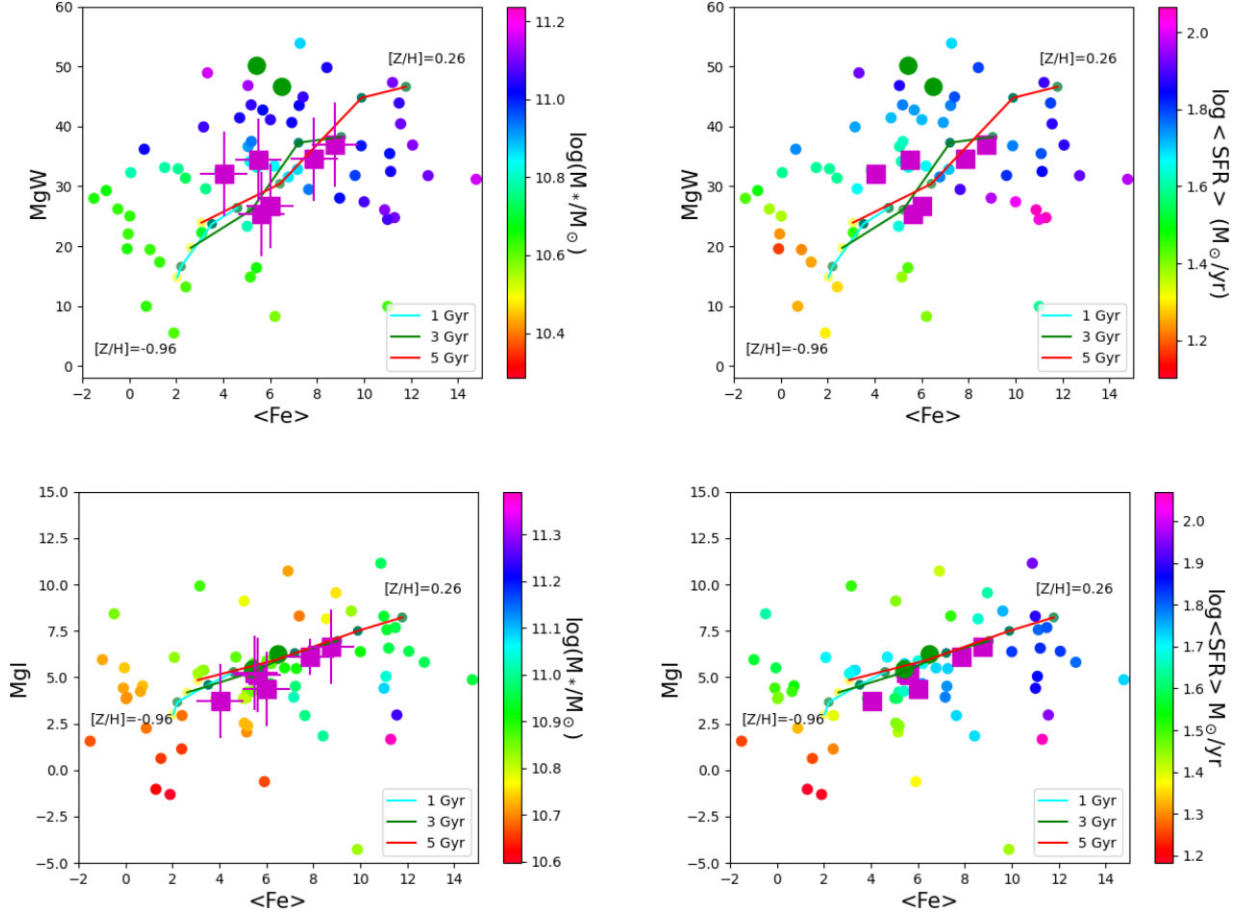


Figure C3. Magnesium versus Iron line indices for individual VANDELS galaxies (small coloured filled circles), VANDELS stacks (magenta big squares) and BOSS stacks of massive ETGs (big green filled circles; Salvador-Rusiñol et al. 2020). These latter are shown as two points representing the mean values obtained from the individual stacks in the two velocity dispersion intervals [160–220] and [220–280] km s^{-1} respectively. In the upper panels the Magnesium index MgW is plotted versus the index $\langle \text{Fe} \rangle$ defined as $\langle \text{Fe} \rangle = (\text{Fe}2609 + \text{FeI})/2$. In the lower panels the Magnesium index MgI is plotted versus $\langle \text{Fe} \rangle$ (see Table C3 for the indices definition). In the left panels, individual VANDELS galaxies are coloured according to their stellar mass, in the right panels according to their $\langle \text{SFR} \rangle$ (see Section 6). Errorbars on VANDELS stacks are the MAD of the values of individual galaxies. Lines show the predictions of EMILES models for 3 fixed ages (as in the legend) for metallicity in the range $-0.96 \leq [\text{Z}/\text{H}] \leq 0.26$ and $[\alpha/\text{Fe}] \sim 0.0$ (see text).

mass of the galaxy, in agreement with the results obtained in Section 4. Hence, the metallicity follows the same trend with the SFH of galaxies shown in Fig. 10. These results seem to confirm the existence of the stellar MZR of passive galaxies at $z \leq 1.4$ (cosmic time <4.5 Gyr) independently of models and of spectral fitting. This, in turn, implies that the stellar metallicity and the resulting relation have been established well in advance, during the formation process. It is worth noting that, taking as reference EMILES models at solar abundance ratio, the indices associated with higher mass galaxies are consistent with supersolar metallicity, while those associated with lower mass galaxies are consistent with lower metallicities.

C2 Constraints on $[\alpha/\text{Fe}]$ ratio and its evolution

The $\text{Mgb}[\lambda 5177]$ index and Fe lines at $\lambda > 5000 \text{ \AA}$ (Fe5015, Fe5270, and Fe5335) are considered the best lines to constrain the $[\text{Mg}/\text{Fe}]$ abundance ratio (e.g. Trager et al. 1998; Thomas et al. 2003). However, these features fall outside the rest-frame wavelength range covered by VIMOS spectra of VANDELS galaxies at $1.0 < z < 1.4$. Therefore, we considered the UV spectral features MgW, MgI, MgII, Fe2609, and FeI (see Table C3 for the indices definition).

UV spectral indices (see e.g. Maraston et al. 2009; Vazdekis et al. 2016, for a description) are more affected by the presence of young populations than optical indices. Moreover, the effect of elemental abundance ratios on these indices is not yet known and they are still affected by higher uncertainties in the stellar population synthesis models. For instance, they can be affected by the presence of the UV-upturn, particularly Mg II, as discussed by, e.g. Le Cras et al. (2016) and Lonoce et al. (2020).

Fig. C3 shows Magnesium indices versus Iron index for VANDELS galaxies at $\langle z \rangle \sim 1.2$ and for massive $\log(M_*/M_\odot) > 11$ BOSS ETGs at $\langle z \rangle \sim 0.38$ (Salvador-Rusiñol et al. 2020). As Magnesium tracers we considered MgW and MgI and, for Iron, we defined the quantity $\langle \text{Fe} \rangle = (\text{Fe}2609 + \text{FeI})/2$. It is worth to remind that the expected effect of enhancing $[\alpha/\text{Fe}]$ on Mg–Fe diagrams is to make Mg higher and Fe lower, i.e. points should be offset along an oblique direction compared to models predictions for varying age and metallicity. In the left-hand panels, individual VANDELS galaxies are coloured according to their stellar mass. The positive trend with the mass of both Magnesium and Iron confirms, once again independently of models, the mass-metallicity trend found in Section 4. In the right-hand panels galaxies are coloured according to their $\langle \text{SFR} \rangle$ (see Section 6). Metallicity depends also on $\langle \text{SFR} \rangle$,

even if the plot $\text{MgI}-\langle\text{Fe}\rangle$ shows a more complex dependence, suggesting that $\langle\text{SFR}\rangle$ may affect mainly Fe. In fact, Fig. 10 shows that also the duration of the SF plays an increasingly important role in the metallicity of a galaxy as the mass decreases (see Discussion).

We probed the possible evolution of $[\alpha/\text{Fe}]$ ratio in massive ETGs by comparing the same indices measured for VANDELS galaxies with those measured on the stacked spectra of massive ETGs at $\langle z \rangle \sim 0.38$. In the $\text{MgW}-\langle\text{Fe}\rangle$ plane. These massive ETGs are offset with respect to VANDELS stacks and are located among individual VANDELS galaxies with highest MgW, suggesting a possible higher $[\alpha/\text{Fe}]$ abundance. However, this is not the only possible explanation for the offset. Indeed, Salvador-Rusiñol et al. (2020) show that it is Fe2609 in their spectra to be too low compared to SSP model predictions because of the presence of young stellar populations. Therefore, it is not necessarily MgW to be too high, but indeed the offset could be explained because of Fe being too low. In the $\text{MgI}-\langle\text{Fe}\rangle$ plane, low redshift massive ETGs are not offset, they are exactly superimposed to VANDELS stacks. Therefore, the comparison between UV indices of VANDELS galaxies at $\langle z \rangle \sim 1.2$ and massive ETGs at $\langle z \rangle \sim 0.38$ shown in Fig. C3 does not

provide a clear answer about a possible redshift evolution of $[\alpha/\text{Fe}]$ in massive galaxies.

To constrain the $[\alpha/\text{Fe}]$ value, in Fig. C3 the EMILES models for 3 fixed ages (as in the legend) and for metallicity in the range $-0.96 \leq [Z/H] \leq 0.26$ are shown. These models assume $[Z/H] = [\text{Fe}/\text{H}]$ and are based on the empirical stellar spectra following the Milky Way (MW) abundance pattern as a function of metallicity. Therefore, in these models, $[\alpha/\text{Fe}] \sim 0.0$ for metallicity solar or higher, while at lower metallicity this is not true (see Vazdekis et al. 2016, for a detailed description). VANDELS galaxies perfectly agree with EMILES models independently of the Magnesium tracer considered, while massive ETGs at $\langle z \rangle \sim 0.38$ are offset when the MgW index is considered. On the basis of these diagnostic plots, and considering the poor knowledge of the effects of element abundance ratios on UV indices, we cannot draw a firm conclusion about the evolution of the $[\alpha/\text{Fe}]$ of massive galaxies from UV indices.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.