The Stellar Metallicities of Massive Quiescent Galaxies at $1.0 < z < 1.3$ from KMOS + VANDELS

Adam C. Carnall[1], Ross J. McLure[1], James S. Dunlop[1], Massissilia Hamadouche[1], Fergus Cullen[1], Derek J. McLeod[1], Ryan Begley[1], Ricardo Amorin[2,3], Micol Bolzonella[4], Marco Castellano[5], Andrea Cimatti[6,7], Fabio Fontanot[8,9], Adriana Gargiulo[10], Bianca Garilli[10], Filippo Mannucci[7], Laura Pentericci[5], Margherita Talia[4,6], Giovani Zamorani[4], Antonello Calabro[5], Giovanni Cresci[7], and Nimish P. Hathi[11]

[1] SUPA, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK; adamc@roe.ac.uk
[2] Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile
[3] Departamento de Física y Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile
[4] INAF—OAS Bologna, Via P. Gobetti 93/3, I-40129, Bologna, Italy
[5] INAF—Osservatorio Astronomico di Roma, Via Frascati 33, I-00078 Monteporzio Catone, Italy
[6] University of Bologna, Department of Physics and Astronomy (DIFA), Via Gobetti 93/2, I-40129 Bologna, Italy
[7] INAF—Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy
[8] INAF—Astronomical Observatory of Trieste, via G.B. Tiepolo 11, I-34143 Trieste, Italy
[9] IFPU—Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy
[10] INAF—IASF Milano, Via A. Corti 12, I-20133 Milano, Italy
[11] Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
Received 2021 October 7; revised 2022 February 18; accepted 2022 March 4; published 2022 April 20

Abstract

We present a rest-frame UV–optical ($\lambda = 2500$–6400 Å) stacked spectrum representative of massive quiescent galaxies at $1.0 < z < 1.3$ with $\log(M_*/M_\odot) > 10.8$. The stack is constructed using VANDELS survey data, combined with new KMOS observations. We apply two independent full-spectral-fitting approaches, measuring a total metallicity [Z/H] $= -0.13 \pm 0.08$ with BAGPIPES and [Z/H] $= 0.04 \pm 0.14$ with ALF, a fall of ~ 0.2–0.3 dex compared with the local universe. We also measure an iron abundance [Fe/H] $= -0.18 \pm 0.08$, a fall of ~ 0.15 dex compared with the local universe. We measure the alpha enhancement via the magnesium abundance, obtaining [Mg/Fe] $= 0.23 \pm 0.12$, consistent with galaxies of similar mass in the local universe, indicating no evolution in the average alpha enhancement of $\log(M_*/M_\odot) \sim 11$ quiescent galaxies over the last ~ 8 Gyr. This suggests the very high alpha enhancements recently reported for several bright $z \sim 1$–2 quiescent galaxies are due to their extreme masses, $\log(M_*/M_\odot) \gtrsim 11.5$, in accordance with the well-known downsizing trend, rather than being typical of the $z \gtrsim 1$ population. The metallicity evolution we observe with redshift (falling [Z/H], [Fe/H], constant [Mg/Fe]) is consistent with recent studies. We recover a mean stellar age of $2.5^{+0.6}_{-0.4}$ Gyr, corresponding to a formation redshift $z_{\rm form} = 2.4^{+0.6}_{-0.3}$. Recent studies have obtained varying average formation redshifts for $z \gtrsim 1$ massive quiescent galaxies, and, as these studies report consistent metallicities, we identify models with different star formation histories as the most likely cause. Larger spectroscopic samples from upcoming ground-based instruments will provide precise constraints on ages and metallicities at $z \gtrsim 1$. Combining these with precise stellar mass functions for $z > 2$ quiescent galaxies from the James Webb Space Telescope will provide an independent test of formation redshifts derived from spectral fitting.

Unified Astronomy Thesaurus concepts: Metallicity (1031); Chemical abundances (224); High-redshift galaxies (734); Quenched galaxies (2016); Galaxy spectroscopy (2171)

1. Introduction

In the local, present-day universe, the massive galaxy population ($\log_{10}(M_*/M_\odot) \gtrsim 10.5$) is dominated by quiescent galaxies, which have shut down (quenched) their star formation activity (e.g., McLeod et al. 2021). The formation and quenching processes leading to the rise of this dominant population, across at least the last 12 Gyr since the first known quiescent galaxies at redshift $z \simeq 4$, are therefore of central importance to our understanding of galaxy evolution.

Studying the massive quiescent galaxy population in the local universe presents several key challenges. First, as the rate of change in stellar population spectra is roughly logarithmic with age (e.g., Ocvirk et al. 2006), constraints on ages via

spectral fitting become steadily less precise for older stellar populations. This means that formation redshifts for local quiescent galaxies, with stellar population ages $\gtrsim 10$ Gyr, are highly uncertain. In addition, massive galaxies gradually accrete new stellar populations via merger events, which have the potential to change their physical sizes, as well as the average ages and metallicities of their stellar populations. This obscures the signatures of the dominant physical processes that acted on these galaxies during their main epoch of formation.

To understand the rise of massive quiescent galaxies, it is therefore necessary to conduct detailed observational studies across the whole history of the universe, from the first 1–2 Gyr to the present day. By measuring how the distributions of key physical parameters evolve as a function of redshift, it should be possible to disentangle the degenerate effects of different processes acting at different times, providing strong constraints on the key physical ingredients required to produce the local massive galaxy population.

As a practical first step toward achieving this goal, much attention is currently focused on measuring the number densities of quiescent galaxies, as well as the distributions of their physical sizes, mean stellar ages, and metallicities, as a function of stellar mass and redshift. Armed with this information across a large fraction of cosmic history, we may then aspire to build a self-consistent model for the assembly of the quiescent population, including the physical processes that influence massive galaxies during their main epoch of formation and quenching, the subsequent growth of individual quiescent galaxies through merger events, and the growth of the quiescent population as a whole via new galaxies quenching their star formation activity.

The largest single factor currently limiting progress in measuring the evolution of these four key properties (number density, size, stellar age, and metallicity) across cosmic time is the availability of high-quality observational data for faint, red, massive quiescent galaxies in the high-redshift universe. The best-constrained parameters for the quiescent population currently are number densities and physical sizes. These are now widely studied out to $z \sim 3$, because they can be reliably constrained via high-quality photometric data (e.g., McLure et al. 2013; Straatman et al. 2014, 2016; van der Wel et al. 2014; Cecchi et al. 2019; Girelli et al. 2019; Merlin et al. 2019; Mowla et al. 2019a, 2019b; Suess et al. 2019a, 2019b; Carnall et al. 2020; Marsan et al. 2022; Sherman et al. 2020; Santini et al. 2021; Hamadouche et al. 2022).

Stellar ages (more generally star formation histories; SFHs) have also been studied using photometric data (e.g., Pacifici et al. 2016; Carnall et al. 2018). However, the age–metallicity– dust degeneracy in galaxy spectral shapes results in relatively weak constraints, meaning the applied priors significantly impact the results obtained (Carnall et al. 2019a; Leja et al. 2019a). Recently, the increasing availability of spectroscopic data with medium to high signal-to-noise ratio (S/N) for quiescent galaxies at intermediate redshifts (e.g., van der Wel et al. 2016; McLure et al. 2018—hereafter M18), combined with sophisticated full-spectral-fitting approaches, has produced the first strong constraints on the SFHs of representative samples out to $z \gtrsim 1$ (e.g., Wu et al. 2018a, 2018b; Belli et al. 2019; Carnall et al. 2019b; Estrada-Carpenter et al. 2019, 2020; Wild et al. 2020; Tacchella et al. 2022).

However, systematic differences still exist in the results of these studies, likely due to a combination of different assumed SFH models, and the fact that stellar metallicities (which are strongly degenerate with ages) are relatively weakly constrained by intermediate-S/N data, meaning the applied priors still play a significant role.

Measuring the stellar metallicities of quiescent galaxies represents another step-change in observational difficulty, with strong constraints (~ 0.1 dex) only available at continuum S/N $\gtrsim 10$–15 $\mathrm{\AA}^{-1}$ in the rest-frame optical (e.g., Gallazzi et al. 2005; Pacifici et al. 2012). This is compounded at $z \gtrsim 1$ by the key features being shifted into the near-IR, where stronger atmospheric absorption and emission make continuum observations from the ground far more challenging.

Despite these challenges, stellar metallicity measurements are highly valuable, as they are strongly constraining on models of galaxy formation, being intimately linked to the physics of star formation and gas recycling (e.g., Maiolino & Mannucci 2019). While substantial progress has been made out to $z \sim 3$ with the brighter, bluer continua of star-forming

galaxies (e.g., Steidel et al. 2016; Cullen et al. 2019, 2021), to date, studies of the stellar metallicities of quiescent galaxies at $z > 1$ are rare, and often restricted to either individual bright objects or low-resolution grism spectra (e.g., Whitaker et al. 2013; Lonoce et al. 2015; Onodera et al. 2015; Kriek et al. 2016).

While the upcoming James Webb Space Telescope (JWST) will produce exceptionally high-quality data for limited numbers of objects at $z > 1$, truly statistical studies are still several years away, awaiting the advent of the next generation of ground-based multiobject spectrographs (e.g., Cirasuolo et al. 2020).

In this work, we present the first determination of the stellar metallicities of a mass-selected sample of quiescent galaxies at $z > 1$. We combine rest-frame near-UV data from the VANDELS survey (M18; Pentericci et al. 2018; Garilli et al. 2021) with new rest-frame optical KMOS YJ-band spectroscopy, to produce a stacked spectrum covering rest-frame 2500–6400 Å for UVJ-selected galaxies at $1.0 < z < 1.3$ with $\log_{10}(M_*/M_\odot) > 10.8$. We fit our stack using both BAGPIPES (Carnall et al. 2018) and ALF (Conroy & van Dokkum 2012; Conroy et al. 2018), obtaining consistent ages and metallicities.

At $1.0 < z < 1.3$, the VANDELS spectra span rest-frame wavelengths $\lambda \sim 2500$–4500 Å. This means they include the Balmer/4000 Å break region critical for precise age determination. However, generally, they do not include the key Fe and Mg absorption features at 4500–5500 Å that are most commonly used to measure stellar metallicities in the local universe (e.g., Gallazzi et al. 2005).

Stellar metallicity was fitted as a free parameter in the VANDELS full-spectral-fitting analysis of Carnall et al. (2019b). However, due to the lack of these key Fe and Mg features, combined with the fact that empirical stellar population models, generally still accepted as more accurate than those based on theoretical spectra (e.g., Coelho et al. 2020), are only available at $\gtrsim 3500$ Å, we chose not to report our derived stellar metallicities in that work. Instead, we obtained further observations with KMOS to gain access to these key rest-frame optical features, to ensure reliable results, and to facilitate direct comparisons with previous studies in the local universe.

The structure of this paper is as follows. In Sections 2 and 3, we introduce our VANDELS and KMOS spectroscopic data sets respectively. In Section 4, we describe the selection of our mass-complete sample, as well as the process of constructing and fitting our representative stacked spectrum with BAGPIPES and ALF. Our results are presented in Section 5 and discussed in Section 6. We present our conclusions in Section 7. All magnitudes are quoted in the AB system. For cosmological calculations, we adopt $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$. All times, t, are measured forwards from the beginning of the universe. We assume a Kroupa (2001) initial mass function. We also assume the solar abundances of Asplund et al. (2009), such that $Z_\odot = 0.0142$.

2. VANDELS Data and Sample Selection

VANDELS (M18; Pentericci et al. 2018; Garilli et al. 2021) is a large ESO public spectroscopic survey of the high-redshift universe, using the Visible Multi-Object Spectrograph (VIMOS) instrument on the Very Large Telescope. The primary aim of VANDELS is to detect continuum emission at high S/N for high-redshift galaxies, moving beyond redshift acquisition to

study the physical properties of galaxies within the first 6 billion years prior to $z = 1$.

2.1. The Parent Photometric Sample

The parent photometric sample for this study consists of 812 massive quiescent galaxies, selected by M18 as potential targets for the VANDELS survey. These objects were selected from four photometric catalogs: the CANDELS GOODS South and UDS catalogs of Guo et al. (2013) and Galametz et al. (2013), and two further ground-based photometric catalogs, purpose-built for VANDELS. These cover the regions immediately surrounding the CANDELS footprints. This approach was necessary as the field of view of the VIMOS spectrograph covers a larger area than the CANDELS Hubble Space Telescope (HST) imaging.

The parent sample was selected from these four photometric catalogs by the following process (described in full detail in Section 4 of M18). Objects were first required to meet the following criteria on apparent magnitude and photometric redshift, z_{phot}:

1. $H \leqslant 22.5$
2. $i \leqslant 25$
3. $1.0 \leqslant z_{phot} \leqslant 2.5$.

For the CANDELS catalogs, the photometric redshifts used were those published by the CANDELS team (Dahlen et al. 2013). For the ground-based catalogs, photometric redshifts were generated through a similar process by the VANDELS team, taking the median of results obtained using a variety of public codes. The H-band magnitude cut limits the sample to objects with $\log_{10}(M_*/M_\odot) \gtrsim 10$. The i-band cut was implemented to make sure the faintest objects would be detected in the VANDELS spectra, and is not relevant to this study, which is focused on a brighter, mass-selected subsample.

To select only quiescent galaxies, the following, permissive, rest-frame UVJ magnitude selection criteria were then applied:

1. $U - V > 0.88(V - J) + 0.49$
2. $V - J < 1.6$
3. $U - V > 1.2$.

From the 812 galaxies selected, 64% have $z_{phot} < 1.3$, with the sample being mass-complete down to $\log_{10}(M_*/M_\odot) = 10.3$ at $1.0 < z < 1.3$ (Carnall et al. 2019b). These 812 objects are referred to as the parent sample throughout the rest of this work.

2.2. VANDELS Spectroscopic Observations

From the parent sample described in the previous subsection, objects were assigned to slits at random to be observed as part of VANDELS. Spectra were obtained for 281 massive quiescent galaxies, roughly one third of the parent sample. These represent roughly 13% of the VANDELS survey, which includes ~2000 objects in total. The remaining 87% is composed of star-forming galaxies at $z > 2.4$. Objects were observed for 20, 40, or 80 hr, depending on their i-band magnitudes, with a mean integration of ~45 hr.

The VANDELS spectroscopic data (described in full in Pentericci et al. 2018) cover observed-frame wavelengths from 4800 to 10300 Å, with spectral resolving power, $R = \lambda/\Delta\lambda \sim 600$. The VANDELS data have high S/N for these redshifts. The spectra of quiescent galaxies within the mass-complete section of the parent sample discussed in Section 2.1 (those with $1.0 < z < 1.3$ and $\log_{10}(M_*/M_\odot) > 10.3$) have a median continuum S/N ~ 11 Å^{-1} at 7500 Å.

Spectroscopic redshifts were measured by the VANDELS team using the pandora.ez software (Garilli et al. 2010). Spectra were assigned redshift quality flags following Le Fèvre et al. (2013), with all but 12 of the 281 observed massive quiescent galaxies assigned quality flags 3 or 4, corresponding to 95% and 100% probabilities of correct identification respectively (in fact, Garilli et al. 2021 demonstrate ~99% reliability for the VANDELS redshifts with flags 3 and 4 combined). The 12 spectra with lower quality flags, as well as one low-redshift interloper, were excluded from this work, leaving 268 spectra. Three objects with spectroscopic redshifts in the range $0.96 < z < 1.0$ were retained. All VANDELS spectroscopic data used in this work come from the final public data release, DR4, described in Garilli et al. (2021).

3. KMOS Data and Sample Selection

We observed four KMOS pointings during ESO P104, from 2019 October to 2020 January, under program ID 0104.B −0885(A). Two pointings targeted each of the two VANDELS fields (UDS and GOODS South). Observations were in the YJ band, providing wavelength coverage from ~1 to 1.35 μm at $R \sim 3600$. During the period in which our data were taken, 23 KMOS arms were functional, meaning we obtained spectra for a total of 92 objects. Each pointing was observed for 8 hr on source, with 300 s exposures, and an ABAB nodding pattern between object and sky positions. The mean seeing for our observations was 0″.6.

3.1. KMOS Sample Selection

Targets for our KMOS observations were drawn almost exclusively from the parent sample of 812 objects described in Section 2.1. To maximize the utility of our YJ-band observations, objects were prioritized by their J-band magnitude. Objects were assigned to KMOS integral field units (IFUs) using the KMOS ArM Allocator (KARMA) tool, with priority classes being defined as follows:

1. Parent sample with $J < 21.5$
2. Parent sample with $J > 21.5$
3. Post-starburst galaxies from Wilkinson et al. (2021).

The priority 3 targets, obtained via private communication, are unrelated to this study, and were added only as fillers to make sure all IFUs were assigned. The positions of the four KMOS pointings were optimized to target the maximum possible number of priority class 1 objects, as well as to maximize the overlap with VANDELS spectroscopy. In total, we allocated IFUs to 62 objects with priority 1, 29 objects with priority 2, and one object with priority 3.

From the parent sample of 812 objects, a total of 273 have $J < 21.5$, meaning objects with priority class 1 were ~4 times more likely to be observed than those with priority class 2. From the 91 objects in the parent sample observed with KMOS, 51 objects were also observed by VANDELS, as described in Section 2.

3.2. KMOS Data Reduction

Our KMOS data were reduced using a combination of the standard ESOREX pipeline recipes and custom code, optimized

for these data. The KMOS pipeline v4.0.0 was used to produce flux-calibrated, sky-subtracted cubes for each individual pair of 300 s (object, sky) exposures. All our targets with $J \lesssim 22$ (approximately 75 out of 92 objects, including all those used in this paper) are clearly detected in each single-exposure, wavelength-collapsed cube.

Upon inspection of the wavelength-collapsed cubes, 1–2 spaxel ($0\rlap{.}''2$–$0\rlap{.}''4$) shifts in object centroids were noted with respect to the WCS coordinates of each object. New centroid positions were therefore measured for each frame by selecting the brightest pixel within a circular aperture of $1''$ diameter centered on the WCS position. These new centroids were verified by manual inspection of each cube.

Also noted were substantial sky-line residuals still present in the data cubes, along with varying systematic shifts in pixel values away from zero across the whole wavelength axis. To address these issues, we implement a further, custom, sky-subtraction step. We first mask all pixels within a circular aperture of $1''$ diameter centered on the new object centroid pixel, as well as all pixels bordering the edge of the detector. We then subtract the median of the remaining pixels from the cube slice at each wavelength. This approach is only possible in this specific instance, as all our targets, being $z > 1$ quiescent galaxies, are extremely compact, with effective radii $r_e \ll 1''$ (e.g., McLure et al. 2013; van der Wel et al. 2014).

Exposures were then aligned according to their updated centroid positions and median-stacked, with uncertainties calculated via the robust median absolute deviation (MAD) indicator. Finally, 1D spectra were extracted within circular apertures of $1''$ diameter, using the Horne (1986) optimal extraction algorithm.

Both the custom re-centroiding and additional sky-subtraction steps are critical to recovering the expected Fe, Mg, and Na absorption features in the spectra of individual objects. Using the default, pipeline-combined cubes produces visible continuum features only in the spectrum of our brightest target, with $J = 19.6$. However, with these additional steps, the strongest features (Mg I 5170 Å and Na I 5895 Å) are visible in almost all spectra for objects with $J < 21.5$.

3.3. KMOS Redshift Measurement

As described above, 51 out of 91 objects from the parent sample observed with KMOS already have secure spectroscopic redshifts from VANDELS. For the remaining 40 objects, we followed the same process as described in Section 2.2, using Pandora.ez, to measure redshifts. In Table 1, we report 25 new spectroscopic redshifts, z_{KMOS}, measured from our KMOS data, along with their associated quality flags. The remaining 15 objects are all significantly fainter ($J > 21.8$), meaning that no reliable redshifts could be measured.

4. Stacking Analysis

The KMOS spectra we obtained for objects with $J < 21.5$ have a median $S/N \sim 4\,\text{Å}^{-1}$ at 12000 Å ($\sim$14 per resolution element at $R = 1000$). This is unfortunately not sufficient to constrain the stellar metallicities of individual objects. In this section, we therefore define a mass-complete sample, for which we can construct a representative stacked spectrum from the VANDELS + KMOS data described in Sections 2 and 3. We fit these data to constrain the average stellar metallicities of massive quiescent galaxies at $1.0 < z < 1.3$.

Table 1
Spectroscopic Redshifts Measured from Our KMOS Data (see Section 3.3)

ID	R.A.	Decl.	J	z_{KMOS}	Flag
CDFS-017418	53.15497	−27.76891	19.86	1.0956	4
UDS-196414	34.48731	−5.09687	20.23	1.0921	4
UDS-205452	34.50079	−5.05545	20.59	1.0945	4
UDS-190420	34.51167	−5.12379	20.60	1.0945	3
UDS-200825	34.52746	−5.07677	20.74	1.0369	3
UDS-201280	34.51318	−5.07575	20.78	1.4140	2
CDFS-004529	53.08042	−27.87204	20.78	1.0967	3
UDS-196179	34.50759	−5.09889	20.91	1.2710	3
CDFS-016336	53.22897	−27.77253	20.95	1.0386	3
CDFS-020067	53.15878	−27.74239	20.97	1.2210	3
CDFS-022694	53.14845	−27.71946	20.99	1.2221	3
CDFS-003952	53.07276	−27.87632	21.03	1.1014	4
CDFS-004376	53.07153	−27.87246	21.05	1.0976	3
UDS-197616	34.43864	−5.09219	21.07	1.6523	4
UDS-015126	34.31596	−5.19366	21.09	1.3611	2
UDS-207822	34.48340	−5.04458	21.18	1.0334	3
UDS-024934	34.52885	−5.12719	21.22	1.0954	3
UDS-009642	34.33053	−5.22374	21.26	1.0735	3
UDS-013785	34.32431	−5.20136	21.27	1.0915	4
UDS-013519	34.33534	−5.20168	21.40	1.5321	3
UDS-192952	34.48285	−5.11423	21.43	1.3295	3
UDS-020224	34.34611	−5.16692	21.44	1.0830	2
UDS-005970	34.31475	−5.24326	21.45	1.7127	2
UDS-010643	34.38065	−5.21789	21.46	1.2624	3
CDFS-014839	53.16516	−27.78587	21.53	1.3175	2

Note. From the 91 parent sample objects observed with KMOS, 51 already have robust VANDELS spectroscopic redshifts. The 25 objects in this table are those of the remaining 40 for which redshifts could be measured. The remaining 15 objects are all fainter than $J = 21.8$.

4.1. Sample Selection

We begin by refitting the photometric data described in Section 2.1 for the 812 objects in the M18 parent sample with BAGPIPES. We use the best available redshifts, z_{best}, defined by the following ranking:

1. VANDELS DR4 redshift if flag 3 or 4; else
2. KMOS redshift (Table 1) if flag 3 or 4; else
3. other spectroscopic redshift (see M18); else
4. photometric redshift, described in Section 2.1.

For consistency with other published work, we assume a Calzetti et al. (2000) dust attenuation curve and solar metallicity. We fit for five free parameters: V-band attenuation (A_V), total stellar mass formed, and the three shape parameters of a double-power-law SFH model (e.g., Carnall et al. 2018). The priors assumed for these five parameters are the same as shown in Table 2, which provides the full list of parameters and priors we use for our more sophisticated fits to our stacked spectrum plus photometry in Section 4.3.

BAGPIPES uses the 2016 updated version of the Bruzual & Charlot (2003) stellar population models,[12] using the MILES stellar spectral library (Falcón-Barroso et al. 2011) and updated stellar evolutionary tracks of Bressan et al. (2012) and Marigo et al. (2013).

From these fits, we obtain stellar masses, M_*, and rest-frame UVJ magnitudes. We now define a mass-complete sample for

---

[12] https://www.bruzual.org/~gbruzual/bc03/Updated_version_2016

Table 2
Parameters and Priors for the BAGPIPES Model We Fit to Our Stacked Spectrum and Photometry

Component	Parameter	Symbol/Unit	Range	Prior	Hyperparameters	
Global	Redshift	z	$(1.145, 1.155)$	Gaussian	$\mu = 1.15$	$\sigma = 0.001$
	Velocity dispersion/instrumental broadening	$\sigma_{vel}/\mathrm{km\ s^{-1}}$	$(100, 500)$	logarithmic		
SFH	Stellar mass formed	$M_*/M_\odot$	$(1, 10^{13})$	logarithmic		
	Metallicity	$Z/Z_\odot$	$(0.01, 2.5)$	logarithmic		
	Falling slope	α	$(0.1, 1000)$	logarithmic		
	Rising slope	β	$(0.1, 1000)$	logarithmic		
	Peak time	τ/Gyr	$(0.1, t_H)$	uniform		
Dust	V-band attenuation	A_V/mag	$(0, 4)$	uniform		
	Deviation from Calzetti et al. (2000) slope	δ	$(-0.3, 0.3)$	Gaussian	$\mu = 0$	$\sigma = 0.1$
	Strength of 2175 Å bump	B	$(0, 5)$	uniform		
Calibration	VANDELS zero order	P_{V0}	$(0.5, 1.5)$	Gaussian	$\mu = 1$	$\sigma = 0.25$
	VANDELS first order	P_{V1}	$(-0.5, 0.5)$	Gaussian	$\mu = 0$	$\sigma = 0.25$
	VANDELS second order	P_{V2}	$(-0.5, 0.5)$	Gaussian	$\mu = 0$	$\sigma = 0.25$
	KMOS zero order	P_{K0}	$(0.5, 1.5)$	Gaussian	$\mu = 1$	$\sigma = 0.25$
	KMOS first order	P_{K1}	$(-0.5, 0.5)$	Gaussian	$\mu = 0$	$\sigma = 0.25$
	KMOS second order	P_{K2}	$(-0.5, 0.5)$	Gaussian	$\mu = 0$	$\sigma = 0.25$
Noise	White noise scaling	a	$(0.1, 10)$	logarithmic		
	Correlated noise amplitude	b/f_{max}	$(0.0001, 1)$	logarithmic		
	Correlation length	$l/\Delta\lambda$	$(0.01, 1)$	logarithmic		

Note. The model is described in Section 4.3, and is adapted from the model presented in Section 4 of Carnall et al. (2019a). The first ten parameters describe our physical model, whereas the final nine describe our systematic uncertainties model. For Gaussian priors, μ is the mean and σ the standard deviation. Logarithmic priors are uniform in $\log_{10}$ of the parameter.

which we can construct a representative stacked spectrum from our combined VANDELS + KMOS spectroscopic data sets. We begin by imposing $z_{best} < 1.3$, leaving 512 objects. We then follow up on the permissive *UVJ* criteria of M18 (see Section 2.1) by requiring $U - V > 0.88 \times (V - J) + 0.69$. This is the diagonal *UVJ* cut proposed by Williams et al. (2009) for quiescent galaxy selection at $z < 0.5$. We use this criterion for our $z \gtrsim 1$ sample, because it has been shown by Carnall et al. (2018, 2019b) to consistently select objects with specific SFR $< 0.2t_H^{-1}$ across a wide redshift range (where t_H is the age of the universe as a function of redshift). This is a widely used criterion for separating star-forming and quiescent galaxies (e.g., Pacifici et al. 2016). This further reduces the sample to 409 objects.

As our KMOS observations targeted a more limited, brighter subsample than VANDELS, the selection criteria detailed in Section 3.1 are the most important for defining our mass-completeness limit. We define this limit as the lowest stellar mass for which 90% of more massive galaxies have $J < 21.5$, which is the criterion for inclusion as a priority 1 target in our KMOS observations. This lowest mass is approximately $\log_{10}(M_*/M_\odot) = 10.8$, and we therefore adopt this as our mass-completeness limit. Imposing this stellar-mass criterion returns 176 objects, with a median stellar mass of $\log_{10}(M_*/M_\odot) = 11.05$. VANDELS spectroscopy is available for 77 of these, whereas KMOS spectroscopy is available for 37 objects. A total of 23 objects have both VANDELS and KMOS spectra. In both cases, objects with spectroscopic data represent a random draw from our mass-complete sample of 176 objects. All VANDELS and KMOS objects have secure (flag 3 or 4) spectroscopic redshifts.

Figure 1 shows the distribution of galaxies in *J*-band magnitude versus z_{best}. The M18 parent sample is shown with gray circles, whereas objects in our mass-complete sample are
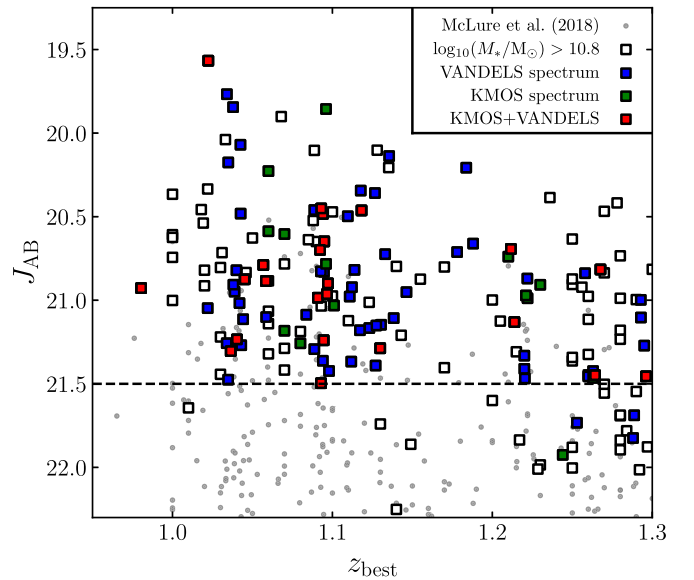

Figure 1. The distribution of the sample in *J*-band magnitude vs. redshift. Circles denote the parent sample from M18 (see Section 2.1). Objects included in our mass-complete photometric sample, selected as described in Section 4.1, are marked with squares. The colors denote the availability of spectroscopic data from VANDELS and KMOS for each object. The dashed horizontal line is the threshold for inclusion as a priority 1 target in our KMOS observations (see Section 3.1). Objects below it were assigned priority 2.

highlighted with open black squares. The availability of VANDELS and KMOS spectroscopic data is indicated by different colored fills of these open black squares, as indicated in the figure.

To summarize, our mass-complete sample is selected from the M18 parent sample, introduced in Section 2.1, as follows
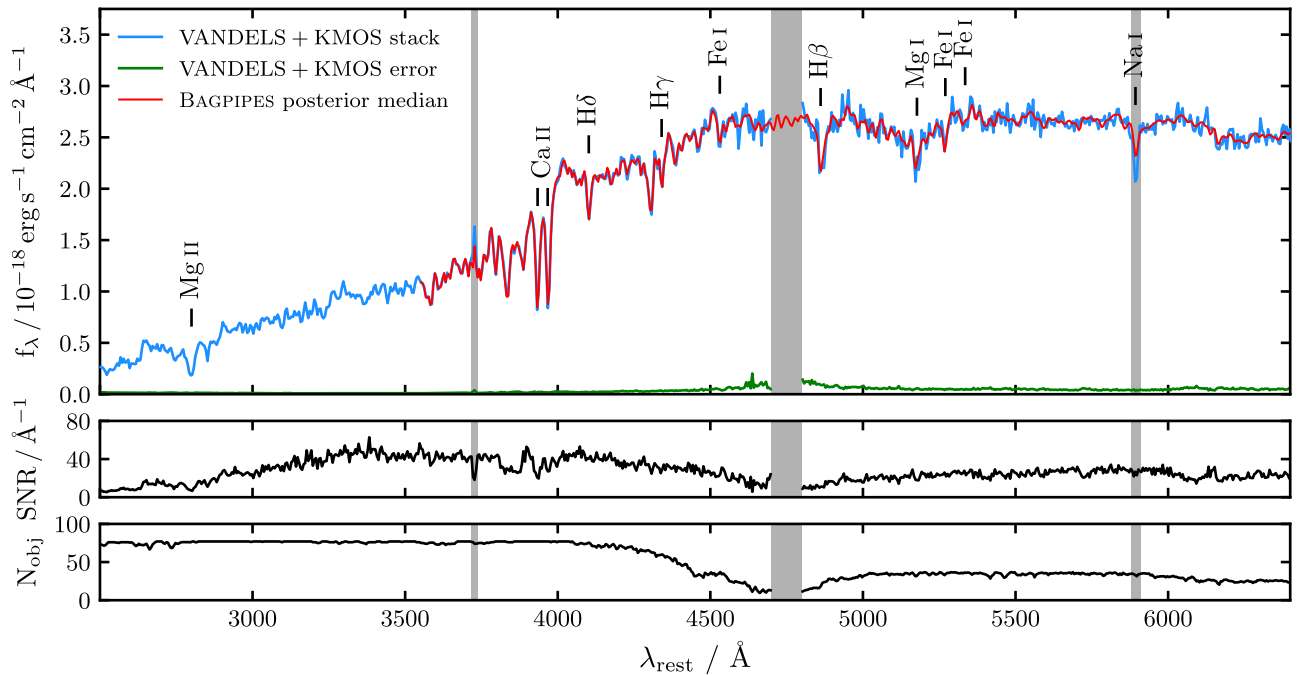
Figure 2. The stacked rest-frame UV–optical spectrum for quiescent galaxies at $1.0 < z < 1.3$ with $\log_{10}(M_*/M_\odot) > 10.8$. The stacked spectrum is shown in blue in the top panel, with the error spectrum shown in green. The S/N per Å and number of objects contributing to the stack are shown in the middle and lower panels respectively. Key age- and metallicity-sensitive features are marked in black. The red line overlaid on the spectrum shows the posterior median model fitted to the stack with BAGPIPES, as described in Section 4.3. This includes a calibration polynomial, which makes corrections of ~5%. We derive a stellar metallicity of [Z/ H] $= -0.13 \pm 0.08$ and a formation redshift of $z_{\rm form} = 2.4^{+0.6}_{-0.3}$. The gray vertical bands show regions masked from fitting (see Section 4).

1. $1.0 < z_{\rm best} < 1.3$
2. $\log_{10}(M_*/M_\odot) > 10.8$
3. $U - V > 0.88 \times (V - J) + 0.69$

with M_* and UVJ magnitudes determined by BAGPIPES fitting.

4.2. Stacking Procedure

We stack the 77 VANDELS and 37 KMOS spectra for our mass-complete sample to produce a single representative stacked spectrum for massive quiescent galaxies at $1.0 < z < 1.3$ with $\log_{10}(M_*/M_\odot) > 10.8$. The individual VANDELS and KMOS spectra were first shifted to the rest frame, then flux-normalized, using rest-frame wavelengths $\lambda = 2500$–4750 Å for VANDELS spectra and $\lambda = 5000$–6000 Å for KMOS spectra.

The spectra were then resampled to a common wavelength grid using SPECTRES (Carnall 2017). The VANDELS and KMOS spectra were then median-stacked separately to produce two stacks, with uncertainties calculated via the MAD estimator. Pixels with strong sky-line contamination were masked prior to stacking. The two stacked spectra were then multiplied by the median of the normalization factors applied to their input spectra. The two stacks, which do not overlap in wavelength, were then combined, and finally binned down to 5 Å sampling.

The stack covers rest-frame wavelengths from 2500 to 6400 Å, with the transition from VANDELS to KMOS data at $\simeq 4750$ Å. Fewer than 10 objects have wavelength coverage between 4700 and 4800 Å, and we therefore mask this region from the stack. We do not attempt to match the flux normalizations of the two stacks at this stage. Instead, during the fitting procedure described in Section 4.3, we fit a spectrophotometric calibration polynomial to both sections of

the stack separately, allowing their relative normalizations to be fitted.

The combined stacked spectrum is shown in blue in the top panel of Figure 2. The error spectrum is shown in green in the same panel. The S/N per Å and the number of objects contributing to the stack are shown in the lower two panels. The wavelengths of key age- and metallicity-sensitive absorption features are labeled in black.

We also generate stacked photometry for our mass-complete sample, by taking the posterior median model fitted to each of the 176 objects in Section 4.1, shifting this to the median redshift of our sample ($z = 1.15$), then calculating fluxes through a series of UV–IR filters. We use $UBVRiz$, HST F606W, F814W, F125W, and F160W, the HAWKI K band, and Spitzer–IRAC Channels 1 and 2. We then produce stacked photometry following the same normalization and median stacking process as detailed above for the spectroscopic data.

4.3. BAGPIPES Fitting of the Stacked Spectrum and Photometry

To constrain the average stellar metallicity of our mass-complete sample, we fit our stacked spectrum and photometry with BAGPIPES[13] (Carnall et al. 2018). We apply the fitting methodology developed in Carnall et al. (2019b), described in full detail in Section 4 of that work. We here provide a brief summary of the method, including a description of the minor changes that have been made to the fitted model for this work. A full list of the 19 free parameters of our model, along with their associated priors, is given in Table 2.

We fit a double-power-law SFH, this time allowing stellar metallicity to vary with a logarithmic prior from 0.01 to 2.5 $Z_\odot$.

---

[13] https://bagpipes.readthedocs.io

Dust attenuation is modeled using the form of Salim et al. (2018), which parameterizes dust curve shape with a power-law deviation, δ, from the Calzetti et al. (2000) model. Emission lines are included in the fit, using a method based on that of Byler et al. (2017) with the CLOUDY photoionization code (Ferland et al. 2017). The lifetime assumed for the stellar birth clouds giving rise to nebular emission is 10 Myr, and A_V is doubled for emission from stars younger than this, as well as nebular line and continuum emission.

Two separate second-order multiplicative Chebyshev polynomials are fitted to the VANDELS and KMOS portions of the stacked spectrum to model any imperfections in spectro-photometric calibration. We model the covariance matrix for our spectroscopic data as follows. The diagonal terms are given by the square of the green error spectrum plotted in Figure 2, multiplied by a factor, a^2, to allow for potential under-estimation of uncertainties. We fit a with a logarithmic prior from 0.1 to 10. The off-diagonal terms are modeled with a Gaussian process, using an exponential-squared kernel.

The stacked spectrum is shifted to the median redshift of our sample ($z = 1.15$) for fitting, and redshift is allowed to vary within a narrow range about this value. A Gaussian prior is applied, with a standard deviation of 0.001 and a maximum deviation of 0.005.

When performing full spectral fitting on spectroscopic observations of galaxies, it is critical to model broadening of spectral features as a result of stellar velocity dispersion within the galaxy, as well as instrumental broadening due to the finite spectral resolution of the optical system. In BAGPIPES, these effects are treated as a nuisance parameter, and jointly modeled by convolving the spectral model with a Gaussian kernel in velocity space, with the standard deviation allowed to vary.

Because the VANDELS and KMOS spectra have different spectral resolutions, it was initially unclear whether it would be appropriate to fit the whole stack using a single Gaussian kernel or whether separate kernels for both sections of the stack would be more appropriate. To investigate this, we performed separate fits to the two sections of the stack, obtaining consistent values for the standard deviation. We therefore proceeded with our final fit using a single Gaussian kernel, permitting the standard deviation to vary from 100 to 500 km s^{-1} with a logarithmic prior. The resulting values are not used in our analysis, and we make no attempt to correct these values for instrumental effects to obtain the true stellar velocity dispersion.

We exclude rest-frame wavelengths <3550 Å from the fit, as the MILES library does not provide coverage bluer than this, with the Bruzual & Charlot (2003) models instead employing a combination of theoretical stellar spectral libraries. We mask the [O II] line at 3727 Å, as it is currently unclear whether [O II] emission in quiescent galaxies originates from star-forming regions, which is the only source of line emission in our BAGPIPES model. Finally, we also mask the Na D absorption feature at 5895 Å, as this has a potential strong component from the interstellar medium (however, see Conroy et al. 2014).

To sample the posterior distribution for a model, BAGPIPES uses the MULTINEST nested sampling algorithm (Skilling 2006; Feroz & Hobson 2008; Feroz et al. 2009, 2019), via the PYMULTINEST interface (Buchner et al. 2014). For fitting our stacked data, MULTINEST was run with 1000 live points, requiring ∼500 CPU hours.

4.4. ALF Fitting of the Stacked Spectrum

In addition to fitting our stacked data with BAGPIPES, we also carried out an independent analysis with the ALF code[14] (Conroy & van Dokkum 2012; Conroy et al. 2018). This decision was made, first to provide a cross-check on our results via a more established method, and second because the Bruzual & Charlot (2003) models assume scaled-solar abundances. By contrast, ALF allows individual element abundances to vary separately, allowing us to constrain the level of alpha enhancement in our target population, a valuable indicator of their formation timescales (e.g., Thomas et al. 2003).

The ALF code is designed for fitting optical to near-IR continuum spectroscopy for old ($\gtrsim$1 Gyr) stellar populations. Originally designed for constraining the initial mass function, recent applications have focused on stellar metallicities, with the code having been developed to fit individual abundances for up to 19 elements.

The code also makes use of the MILES stellar spectral library, and includes empirical stellar population models spanning 0.37–2.4 μm. We therefore fit our stacked spectrum across a wavelength range similar to that described in Section 4.3, this time also omitting rest-frame wavelengths from 3550 to 3700 Å. We investigate the potential impact of this difference by rerunning our BAGPIPES fit while also excluding this wavelength range, and find this has no effect on our results.

ALF continuum-normalizes input spectra using a high-order polynomial, with one order per 100 Å of rest-frame spectral coverage, prior to fitting. The code does not currently include the capability to fit photometric data, so our stacked photometry was not used in this analysis. The code uses the Markov Chain Monte Carlo sampler EMCEE (Foreman-Mackey et al. 2013). We run ALF in simple mode, which includes 13 free parameters: redshift, velocity dispersion, stellar age (a single burst SFH is assumed), total stellar metallicity, [Z/H], and abundances for nine individual elements, including Fe and Mg. As an additional check, we have also fitted our data using full mode in ALF, which results in nearly identical metallicity values.

5. Results

The posterior median model fitted to our stacked spectrum with BAGPIPES is shown in red in the top panel of Figure 2. We measure a stellar metallicity of [Z/H] = -0.13 ± 0.08. We also measure a mean stellar age of $2.5^{+0.6}_{-0.4}$ Gyr, which, at $z = 1.15$, corresponds to a mean formation time $t_{\mathrm{form}} = 2.7^{+0.4}_{-0.6}$ Gyr after the Big Bang, or a formation redshift of $z_{\mathrm{form}} = 2.4^{+0.6}_{-0.3}$. We measure a dust curve slope consistent with Calzetti et al. (2000), and $A_V = 0.48 \pm 0.08$ mag. This is in good agreement with Belli et al. (2019) and Carnall et al. (2019b) (but see Appendix B of van der Wel et al. 2021).

With ALF, we obtain element abundances [Fe/H] = -0.18 ± 0.08 and [Mg/H] = 0.07 ± 0.09. The enhancement of Mg relative to Fe can be used as a proxy for alpha enhancement, and we obtain [Mg/Fe] = 0.23 ± 0.12. We convert from Fe and Mg abundances to total metallicity using [Z/H] = [Fe/H] + 0.94[Mg/Fe] (e.g., Thomas et al. 2003), for consistency with other recent work (e.g., Kriek et al. 2019). This yields a total metallicity of [Z/H] = 0.04 ± 0.14. Our
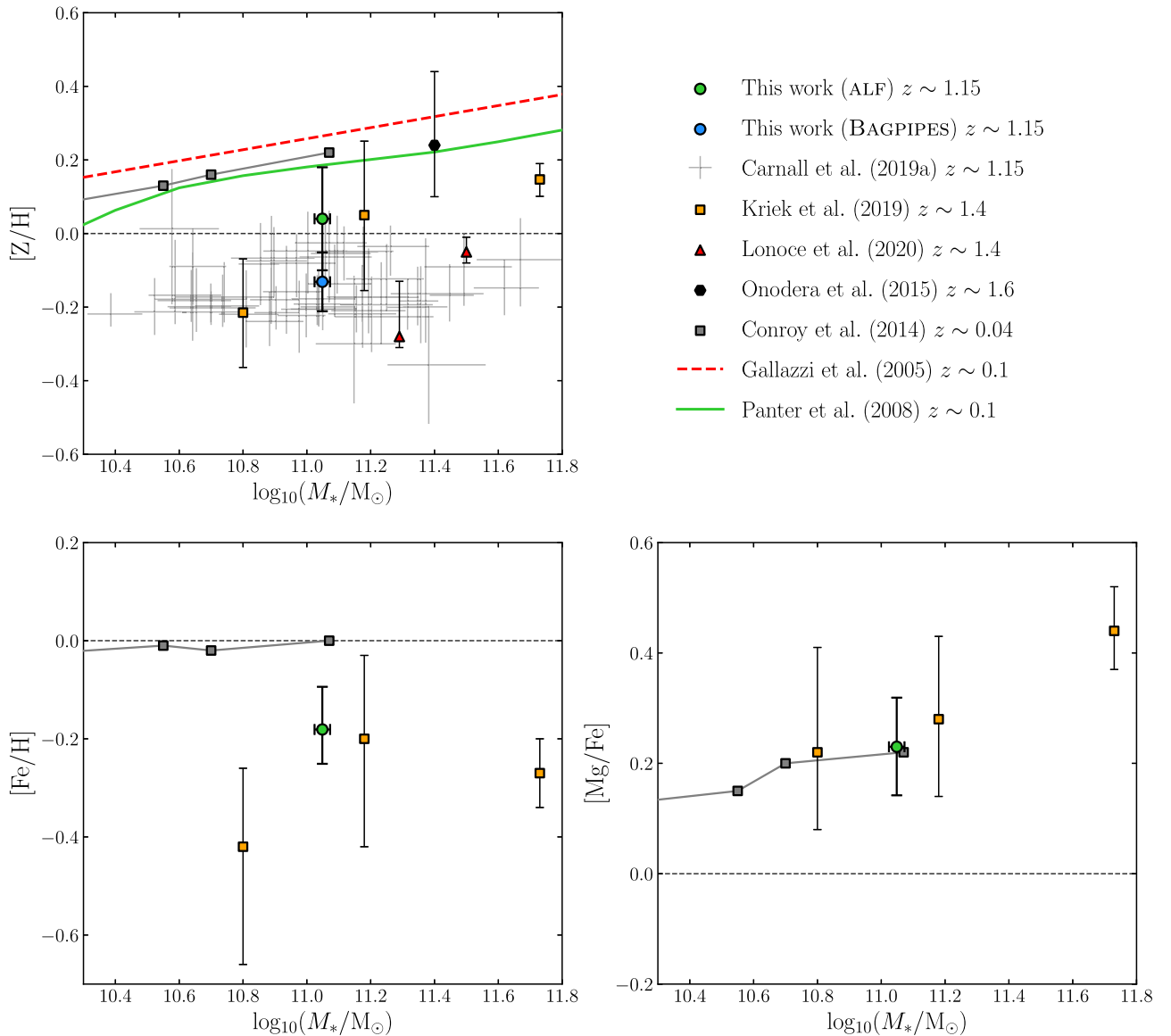
---

[14] https://www.github.com/cconroy20/alf

Figure 3. Top: stellar metallicity measurements for massive quiescent galaxies at $z \gtrsim 1$, compared with the local universe. Our results, which are broadly consistent with previous work at similar redshifts, suggest that [Z/H] was ~0.2–0.3 dex lower at $z \gtrsim 1$ compared with the present day. Lower left: Fe abundances for the subset of studies, including ours, that use ALF. We find [Fe/H] ~ 0.15 dex lower than the local universe. Lower right: alpha enhancement, traced by [Mg/Fe]. We find ~0.2 dex enhancement, suggesting no evolution across the last ~8 Gyr since $z \gtrsim 1$. The dashed horizontal lines show the solar value.

BAGPIPES and ALF measurements of [Z/H] are therefore in broad agreement, with the ALF posterior median value 0.17 dex higher than that from BAGPIPES (formally the two posterior distributions are in tension with ~1.1σ confidence). The ALF fitted spectrum is virtually indistinguishable from the BAGPIPES fitted spectrum shown in Figure 2, and is therefore not shown. Our stellar metallicity results are summarized in Figure 3, and compared to results from the literature in Section 6.1.

The stellar population ages are also in broad agreement, with ALF returning an age of 2.4 ± 0.3 Gyr. The fact ALF returns a ~1σ higher metallicity and a slightly lower age than BAGPIPES is consistent with our expectations, based on the different SFH models used by the two codes. When run in simple mode, ALF uses a single burst SFH, a simplifying assumption known to result in lower ages and higher metallicities, closer to the light-weighted values (e.g., Conroy 2013). These issues will be further discussed in Section 6.

6. Discussion

6.1. Stellar Metallicities

6.1.1. Comparisons with Other Results at z ≳ 1

While this work presents the first measurement of the average stellar metallicity of a mass-selected sample of massive quiescent galaxies at $z > 1$, several other studies have analyzed individual bright objects or magnitude-selected samples at similar redshifts. In Figure 3, we compare our results for [Z/H], [Fe/H], and [Mg/Fe] with other recent work at $1.0 < z \lesssim 1.5$, as well as results from the local universe, which will be discussed in Section 6.1.2.

Kriek et al. (2019) report stellar metallicities for three massive quiescent galaxies at $z \sim 1.4$ via full spectral fitting of Keck-LRIS and MOSFIRE spectroscopy, also with the ALF code. These results are shown in Figure 3 with orange squares, and are in good agreement with the measurements we derive

from our stacked spectrum. Their most massive object is more Mg-enhanced, leading to a higher total metallicity, consistent with the positive correlation between mass and metallicity observed in the local universe.

Lonoce et al. (2015, 2020) also report stellar metallicities for two galaxies at $z \sim 1.4$ via full spectral fitting, which are shown in the top left panel of Figure 3 with red triangles. These results are also consistent with our findings, though it should be noted that Lonoce et al. (2020) obtain strongly contrasting results via an alternative, spectral-index-fitting analysis: $[Z/H] \sim -0.7$ and $+0.6$ for their lower- and higher-mass galaxies, respectively.

Onodera et al. (2015) report a stellar metallicity of $[Z/H] = 0.24^{+0.20}_{-0.14}$ via spectral index fitting, for a stacked spectrum constructed from a magnitude-selected sample ($K < 21.5$) of 24 quiescent galaxies at $1.2 \lesssim z \lesssim 2.0$. This is considerably higher than our result at $1.0 < z < 1.3$; however, their relatively large reported uncertainties mean the two results are not strongly in tension. It should also be noted that the average mass of galaxies in the sample of Onodera et al. (2015) is 0.35 dex higher than that in our sample.

Finally, we show, with gray error bars, the stellar metallicities we obtained for 53 individual massive quiescent galaxies in Carnall et al. (2019b), via BAGPIPES full spectral fitting of intermediate-S/N ($\sim 10\,Å^{-1}$) VANDELS rest-frame near-UV spectroscopy. The results of Carnall et al. (2019b) are in very good agreement with the new analysis we present in this work, which differs from our previous work in three key respects: the use of stacking, the inclusion of additional rest-frame optical KMOS data, and the exclusion of wavelengths $\lambda < 3550\,Å$, which were fitted with theoretical stellar templates in Carnall et al. (2019b). Our new result validates the metallicities measured by our previous analysis, and hence the ages/SFHs that were the main result of that work.

The individual-object results of Carnall et al. (2019b) shown in Figure 3 do not suggest a strong stellar mass–metallicity relation at $z \gtrsim 1$ above $\log(M_*/M_\odot) = 10.5$, and are consistent with a flat relationship. However, a relatively weak mass–metallicity relation, as seen in the local universe, is by no means ruled out. We have investigated the possibility of splitting the mass-complete sample in this work into two mass bins to further probe the mass–metallicity relation at $z \gtrsim 1$; however, our KMOS data set does not provide sufficient S/N to obtain meaningful constraints when splitting our sample into two or more bins.

We conclude that our new results are generally consistent with previous work at $z \gtrsim 1$. In particular, multiple full spectral fitting analyses agree on average stellar metallicities, $[Z/H] \sim -0.1$ for massive quiescent galaxies with $\log_{10}(M_*/M_\odot) \sim 11$ at $z \gtrsim 1$. Our measurement of $[Fe/H] \sim -0.2$ is in good agreement with Kriek et al. (2019), as is our measurement of ~ 0.2 dex of Mg enhancement over Fe. Kriek et al. (2016, 2019) show that this level of alpha enhancement suggests formation timescales of 0.2–1.0 Gyr, again in good agreement with the SFHs reported by Carnall et al. (2019b).

6.1.2. Comparisons with the Local Universe

In the top left panel of Figure 3, we also show three determinations of the stellar mass–metallicity relation in the local universe ($z \lesssim 0.1$). These three sets of results are all in reasonable agreement, but there are key differences in methodology and sample selection that should be considered

when making comparisons. All three studies use data from the Sloan Digital Sky Survey (SDSS; York et al. 2000).

Gallazzi et al. (2005) and Panter et al. (2008) report mean stellar metallicity as a function of stellar mass for all galaxies, including both star-forming and quiescent objects. However, the dominance of quiescent galaxies at high masses in the local universe means these results are still reasonably comparable with the $z \gtrsim 1$ results shown. In addition, Gallazzi et al. (2005) report r-band light-weighted metallicities, rather than the mass-weighted quantities reported by the other studies. This may explain the higher metallicities they report, as their results are weighted more heavily toward younger stars.

It should also be noted, however, that even among studies reporting mass-weighted quantities, the SFH model adopted has the potential to influence the stellar metallicities obtained (e.g., Carnall et al. 2019a; Leja et al. 2019a). This is of particular note for Conroy et al. (2014), who adopt a simplified two-burst SFH model that would be expected to return metallicities closer to the light-weighted values. This may explain the slightly higher metallicities found by Conroy et al. (2014) with respect to Panter et al. (2008).

The results of Conroy et al. (2014) are the most comparable with our results at $z \gtrsim 1$, in terms of both methodology and sample selection. They select quiescent galaxies by requiring no Hα or [O II] emission, and require that objects lie on the Fundamental Plane, using the central slice defined by Graves et al. (2010). At $\log(M_*/M_\odot) \sim 11$, they report $[Z/H] = 0.21$. Our result therefore implies, at fixed stellar mass, ~ 0.2–0.3 dex evolution in the average $[Z/H]$ for massive quiescent galaxies across the ~ 8 Gyr from $z \sim 1.15$ to the present. From the lower right panel of Figure 3, it can be seen that, surprisingly, this 0.2–0.3 dex evolution in $[Z/H]$ is not accompanied by any change in alpha (Mg) enhancement, with $[Mg/Fe]$ remaining at ~ 0.2.

Alpha elements are primarily produced by core-collapse supernovae (CCSNe), which are the end-point in the evolution of massive stars. Enrichment with alpha elements therefore rapidly follows the onset of star formation, within a few million years. By contrast, Fe-peak elements are produced by both CCSNe and Type Ia supernovae, which occur on much longer ($\sim$Gyr) timescales. This means that both Fe abundance and alpha enhancement in stars are linked to the formation timescale of the galaxy, with lower-Fe, more highly alpha-enhanced stellar populations expected in galaxies that formed on short timescales (e.g., Weinberg et al. 2017; Kobayashi et al. 2020).

As the number density of massive quiescent galaxies rises steadily over cosmic time, those that exist at $z \gtrsim 1$ are expected to be a biased subsample of the local quiescent galaxy population: the earliest and fastest formed. This is often referred to as progenitor bias. We would therefore expect to see, on average, both lower Fe abundance and greater alpha enhancement for $z \gtrsim 1$ quiescent galaxies, compared with the local population.

Beverage et al. (2021) have recently reported similar evolution of the quiescent galaxy population from the local universe to $z \sim 0.7$: falling $[Z/H]$ and $[Fe/H]$, but constant $[Mg/Fe]$ with increasing redshift (see also Choi et al. 2014; Leethochawalit et al. 2019). They suggest the expected increase in $[Mg/Fe]$ might begin to become apparent at higher redshifts, pointing to several studies of individual objects that find very high $[Mg/Fe]$ for individual ultramassive galaxies at $z \sim 2$ (Kriek et al. 2016; Jafariyazani et al. 2020).

Our results demonstrate the expected increase in [Mg/Fe] is still not apparent at $z \gtrsim 1$ for typical $\log(M_*/M_\odot) \sim 11$ quiescent galaxies. This is in agreement with Kriek et al. (2019) at $z \sim 1.4$ (bottom right panel of Figure 3), who report higher alpha enhancement than the local universe for only a single, extremely massive galaxy with $\log(M_*/M_\odot) \sim 11.7$. Our results suggest the very high [Mg/Fe] values (implying ultrashort, ~ 100 Myr, formation timescales) recently reported by Kriek et al. (2016, 2019) and Jafariyazani et al. (2020) for ultramassive galaxies at $z \sim 1$–2 are more closely associated with their exceptionally high masses, in accordance with the well-known downsizing trend, rather than their high redshifts.

The expected change in [Mg/Fe] across the ~ 0.15 dex interval between our [Fe/H] result and those of Conroy et al. (2014) can be estimated from Figure 12 of Kobayashi et al. (2020) to be ~ 0.1 dex. This is of the same order of magnitude as the uncertainty we measure on our [Mg/Fe] value. This means that, while we do not observe any change in [Mg/Fe] between the local universe and our results at $z \gtrsim 1$, the level of [Mg/Fe] evolution predicted by Kobayashi et al. (2020) is not strongly excluded by our measurement.

A detailed understanding of these issues will require a data set of similar quality to SDSS at $z \gtrsim 1$. Several planned surveys with upcoming instruments, such as the Multi-Object Optical and Near-infrared Spectrograph (MOONS) GTO survey MOONRISE (Cirasuolo et al. 2020; Maiolino et al. 2020), will provide much larger numbers of high-S/N spectra at $z \gtrsim 1$. This will allow individual metallicity determinations for large numbers of bright targets, and high-S/N stacking experiments to be performed with fine resolution in stellar mass and other key parameters, such as physical size.

6.2. Stellar Ages

As discussed in Section 6.1.1, the stellar metallicity we derive for our stacked spectrum is in good agreement with those we obtained for individual galaxies in Carnall et al. (2019b). In that work, we derived the following relationship between formation time, $t_{\rm form}$ (which we define as the time corresponding to the mean stellar age), and stellar mass for quiescent galaxies at $z \gtrsim 1$:

$$\frac{t_{\rm form}}{\rm Gyr} = 2.56^{+0.12}_{-0.10} - 1.48^{+0.34}_{-0.39} \log_{10}\left(\frac{M_*}{10^{11} M_\odot}\right). \qquad (1)$$

At the median stellar mass of $\log_{10}(M_*/M_\odot) = 11.05$ for the mass-complete sample from which we construct our stack, this relationship predicts a formation time of 2.5 ± 0.1 Gyr after the Big Bang. We derive consistent results from both the BAGPIPES and ALF analyses of our stack: $t_{\rm form} = 2.7^{+0.4}_{-0.6}$ Gyr and 2.4 ± 0.3 Gyr respectively.

Several other recent studies have reported ages for massive quiescent galaxies at $z \gtrsim 1$ using similar methodologies (Belli et al. 2019; Estrada-Carpenter et al. 2020; Tacchella et al. 2022). These results are summarized in Figure 10 of Tacchella et al. (2022), demonstrating that the formation times they derive for their $z \sim 0.8$ sample are earlier on average than those we derive at $z \gtrsim 1$. This is the opposite to the effect that would be expected due to progenitor bias.

Due to the strong age–metallicity degeneracy in galaxy spectral shapes, discrepant metallicities is an obvious potential cause for the unexpected differences in derived ages between these studies. However, as can be seen from Figure 19 of

Tacchella et al. (2022), the average metallicity they derive is just below solar,[15] as would be expected considering the comparison between metallicities in the local universe and at $z \gtrsim 1$ shown in the top panel of Figure 3. The metallicities we derive in this work and Carnall et al. (2019b) are therefore fully consistent with Tacchella et al. (2022), and are ruled out as the cause of the differences in derived ages.

The remaining likely cause for these differences lies in the SFH models used by the respective studies. In this work and Carnall et al. (2019b), we have used the double-power-law SFH model introduced by Carnall et al. (2018). This is a parametric model, and similar to the models used by Belli et al. (2019), who report results consistent with Carnall et al. (2019b). By contrast, Tacchella et al. (2022) use the nonparametric, continuity SFH prior introduced by Leja et al. (2019a), which is known to return older stellar populations than more traditional parametric models (Leja et al. 2019b). Interestingly, Estrada-Carpenter et al. (2020) also use the nonparametric continuity prior, finding results consistent with Tacchella et al. (2022).

It seems likely, therefore, that the underlying cause of the discrepant results in the recent literature for the formation redshifts of massive quiescent galaxies at $z \gtrsim 1$ is the use of different SFH models, rather than conflicting derived metallicities. These SFH models constitute a set of prior beliefs about when and how galaxies form, introducing an unavoidable subjectivity into results obtained in the absence of strongly constraining data (e.g., Ocvirk et al. 2006; Carnall et al. 2019a; Johnson et al. 2021).

To evaluate the success of these subjective prior choices, one should perform cross-validation checks, testing whether results obtained via spectral fitting are consistent with other, better established results in the literature (e.g., Wuyts et al. 2011; Carnall et al. 2019a). In this case, we would wish to use the SFHs derived for $z \gtrsim 1$ massive quiescent galaxies to predict the stellar mass function for quiescent galaxies at $z \sim 2$–4, then to compare these predictions with direct observations of $z > 2$ mass functions and/or number densities.

To perform this challenging experiment, we are currently missing two vital components. First, the number of massive quiescent galaxies at $z \gtrsim 1$ with high-S/N rest-frame UV–optical spectroscopy is currently no more than a few hundred. This, in combination with the significant uncertainties on individual recovered SFHs, leads to huge statistical uncertainties in predicted mass functions at $z > 2$. To make useful predictions, vastly larger samples will be required, numbering tens to hundreds of thousands of galaxies. A number of upcoming surveys, such as MOONRISE (Maiolino et al. 2020), will provide the necessary data sets over the next ~ 5 yr.

The second missing component is a precise determination of the observed stellar mass function for massive quiescent galaxies at $z > 2$, with current studies suffering from considerable uncertainties in sample selection (e.g., Schreiber et al. 2018; Carnall et al. 2020). Large imaging surveys with JWST (e.g., PRIMER; Dunlop et al. 2021) will provide the extremely deep, high-resolution infrared imaging necessary to select mass-complete samples of quiescent galaxies with confidence out to the highest redshifts, resulting in precise constraints on the stellar mass functions of quiescent galaxies at $z > 2$.

---

[15] The average metallicity derived for the *UVJ*-quiescent sample in Tacchella et al. (2022) is [Z/H] $= -0.13 \pm 0.02$ (private communication).

Once these data sets are in place, it should be possible to clearly distinguish between the parametric and nonparametric SFH models that currently return substantially different formation redshifts for massive quiescent galaxies at $z \gtrsim 1$.

7. Conclusion

In this work, we have combined data from the VANDELS survey with new KMOS observations to construct a representative stacked spectrum for quiescent galaxies at $1.0 < z < 1.3$ with $\log_{10}(M_*/M_\odot) > 10.8$, covering rest-frame wavelengths $\lambda = 2500$–6400 Å. The stacked spectrum is shown in Figure 2. We also report 25 new spectroscopic redshifts (Table 1) from our KMOS data.

We fit our stacked spectrum with BAGPIPES, obtaining a stellar metallicity $[Z/H] = -0.13 \pm 0.08$. We also obtain a formation time $t_{\mathrm{form}} = 2.7^{+0.4}_{-0.6}$ Gyr after the Big Bang, corresponding to a formation redshift of $z_{\mathrm{form}} = 2.4^{+0.6}_{-0.3}$. Both of these results are consistent with the results we presented in Carnall et al. (2019b), which were obtained by fitting only VANDELS data and working with individual spectra rather than stacking. We also fit our stacked spectrum with the ALF code, obtaining a consistent result for $[Z/H]$, as well as an iron abundance $[Fe/H] = -0.18 \pm 0.08$ and alpha enhancement $[Mg/Fe] = 0.23 \pm 0.12$ (see Figure 3).

By comparing our results at $z \gtrsim 1$ with results from the local universe ($z \lesssim 0.1$), we demonstrate that the average $[Z/H]$ for $\log_{10}(M_*/M_\odot) \sim 11$ quiescent galaxies has risen by ~ 0.3 dex across the ~ 8 Gyr since $z \gtrsim 1$, whereas $[Fe/H]$ has risen by ~ 0.15 dex. However, the alpha enhancement, $[Mg/Fe]$, we measure is the same as found by Conroy et al. (2014) in the local universe, implying no evolution at fixed stellar mass across at least the last ~ 8 Gyr.

Given that $z \gtrsim 1$ massive quiescent galaxies are a biased subsample of the local quiescent population (those that formed fastest and earliest), we would expect them to have lower $[Fe/H]$, as observed, but we would also expect greater alpha enhancement. Our finding of no redshift evolution in $[Mg/Fe]$ at $\log_{10}(M_*/M_\odot) \sim 11$ from $z \gtrsim 1$ to the present is in agreement with the results of Kriek et al. (2019) at $z \sim 1.4$ and Beverage et al. (2021) at $z \sim 0.7$.

This result suggests that the highly alpha-enhanced, ultra-massive ($\log_{10}(M_*/M_\odot) \gtrsim 11.5$) galaxies recently reported at $z \sim 1$–2 by some authors (e.g., Kriek et al. 2016, 2019; Leethochawalit et al. 2019) are highly alpha-enhanced due to their extreme masses, in accordance with the well-known downsizing trend, rather than being typical of the $z \gtrsim 1$ quiescent population.

The model of Kobayashi et al. (2020) predicts a relatively modest change in $[Mg/Fe]$ of ~ 0.1 dex across the ~ 0.15 dex $[Fe/H]$ interval separating our results from the local-universe measurements of Conroy et al. (2014). This level of $[Mg/Fe]$ evolution is not strongly ruled out by our measurement, meaning that stronger constraints from higher-quality data at $z \gtrsim 1$ will be necessary to confidently determine whether the level of $[Mg/Fe]$ evolution observed between the local universe and $z \gtrsim 1$ is in agreement with theoretical predictions.

Recently, Tacchella et al. (2022) have highlighted differences in the average formation redshifts measured for $z \gtrsim 1$ massive quiescent galaxies by several recent studies. In particular, they, in agreement with Estrada-Carpenter et al. (2020), find earlier formation than Belli et al. (2019) and Carnall et al. (2019b). Given the latter two studies are at higher

redshifts, this is the opposite to the effect that would be expected due to progenitor bias.

We demonstrate that the metallicities recovered by Tacchella et al. (2022) at $z \sim 0.8$ are consistent with the results of this work and Carnall et al. (2019b). This means discrepancies in recovered metallicities are ruled out as the cause of the differences in recovered ages. We therefore conclude that the differences in formation redshifts obtained by different authors are likely due to the use of different models of star formation history, with Estrada-Carpenter et al. (2020) and Tacchella et al. (2022) using the continuity nonparametric SFH model of Leja et al. (2019a), and Belli et al. (2019) and Carnall et al. (2019b) using more traditional parametric models.

To determine which of these models more accurately represents the SFHs of $z \gtrsim 1$ massive quiescent galaxies, we will require much larger statistical samples at $z \gtrsim 1$ with high-S/N rest-frame UV–optical spectroscopy, as will be provided by upcoming instruments such as MOONS (Cirasuolo et al. 2020). This will allow us to make firm predictions for the stellar mass functions of $z > 2$ quiescent galaxies via spectral fitting with different SFH models, which can be compared with the precise $z > 2$ stellar mass functions that will be provided by data from the upcoming JWST.

Data Availability

The VANDELS survey is a European Southern Observatory Public Spectroscopic Survey. The full spectroscopic data set, together with photometric catalogs and derived quantities, is available from http://vandels.inaf.it, as well as from the ESO archive https://www.eso.org/qi. The KMOS data used in this work are available from the ESO archive under program ID 0104.B-0885(A). Reduced KMOS data products, our stacked spectrum, and model posteriors may be made available upon reasonable request.

ORCID iDs

Adam C. Carnall ⬤ https://orcid.org/0000-0002-1482-5818
Massissilia Hamadouche ⬤ https://orcid.org/0000-0001-6763-5551
Fergus Cullen ⬤ https://orcid.org/0000-0002-3736-476X
Derek J. McLeod ⬤ https://orcid.org/0000-0003-4368-3326
Ricardo Amorin ⬤ https://orcid.org/0000-0001-5758-1000
Micol Bolzonella ⬤ https://orcid.org/0000-0003-3278-4607
Marco Castellano ⬤ https://orcid.org/0000-0001-9875-8263
Andrea Cimatti ⬤ https://orcid.org/0000-0002-4409-5633
Fabio Fontanot ⬤ https://orcid.org/0000-0003-4744-0188
Adriana Gargiulo ⬤ https://orcid.org/0000-0002-3351-1216

Bianca Garilli ⓘ https://orcid.org/0000-0001-7455-8750
Filippo Mannucci ⓘ https://orcid.org/0000-0002-4803-2381
Laura Pentericci ⓘ https://orcid.org/0000-0001-8940-6768
Margherita Talia ⓘ https://orcid.org/0000-0003-4352-2063
Giovani Zamorani ⓘ https://orcid.org/0000-0002-2318-301X
Antonello Calabro ⓘ https://orcid.org/0000-0003-2536-1614
Giovanni Cresci ⓘ https://orcid.org/0000-0002-5281-1417
Nimish P. Hathi ⓘ https://orcid.org/0000-0001-6145-5090

References

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Belli, S., Newman, A. B., & Ellis, R. S. 2019, ApJ, 874, 17
Beverage, A. G., Kriek, M., Conroy, C., et al. 2021, ApJL, 917, L1
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Carnall, A. C. 2017, arXiv:1705.05165
Carnall, A. C., Leja, J., Johnson, B. D., et al. 2019a, ApJ, 873, 44
Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, MNRAS, 480, 4379
Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2019b, MNRAS, 490, 417
Carnall, A. C., Walker, S., McLure, R. J., et al. 2020, MNRAS, 496, 695
Cecchi, R., Bolzonella, M., Cimatti, A., & Girelli, G. 2019, ApJL, 880, L14
Choi, J., Conroy, C., Moustakas, J., et al. 2014, ApJ, 792, 95
Cirasuolo, M., Fairley, A., Rees, P., et al. 2020, Msngr, 180, 10
Coelho, P. R. T., Bruzual, G., & Charlot, S. 2020, MNRAS, 491, 2025
Conroy, C. 2013, ARA&A, 51, 393
Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, ApJ, 780, 33
Conroy, C., & van Dokkum, P. G. 2012, ApJ, 760, 71
Conroy, C., Villaume, A., van Dokkum, P. G., & Lind, K. 2018, ApJ, 854, 139
Cullen, F., McLure, R. J., Dunlop, J. S., et al. 2019, MNRAS, 487, 2038
Cullen, F., Shapley, A. E., McLure, R. J., et al. 2021, MNRAS, 505, 903
Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93
Dunlop, J. S., Abraham, R. G., Ashby, M. L. N., et al. 2021, PRIMER: Public Release IMaging for Extragalactic Research, JWST Proposal. Cycle 1, ID. #1837
Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2019, ApJ, 870, 133
Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2020, ApJ, 898, 171
Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2019, OJAp, 2, 10
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Galametz, A., Grazian, A., Fontana, A., et al. 2013, ApJS, 206, 10
Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41
Garilli, B., Fumana, M., Franzetti, P., et al. 2010, PASP, 122, 827
Garilli, B., McLure, R., Pentericci, L., et al. 2021, A&A, 647, A150
Girelli, G., Bolzonella, M., & Cimatti, A. 2019, A&A, 632, A80
Graves, G. J., Faber, S. M., & Schiavon, R. P. 2010, ApJ, 721, 278
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Hamadouche, M. L., Carnall, A. C., McLure, R. J., et al. 2022, MNRAS, 512, 1262

Horne, K. 1986, PASP, 98, 609
Jafariyazani, M., Newman, A. B., Mobasher, B., et al. 2020, ApJL, 897, L42
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, ApJ, 900, 179
Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, Natur, 540, 248
Kriek, M., Price, S. H., Conroy, C., et al. 2019, ApJL, 880, L31
Kroupa, P. 2001, MNRAS, 322, 231
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Leethochawalit, N., Kirby, E. N., Ellis, R. S., Moran, S. M., & Treu, T. 2019, ApJ, 885, 100
Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. 2019a, ApJ, 876, 3
Leja, J., Johnson, B. D., Conroy, C., et al. 2019b, ApJ, 877, 140
Lonoce, I., Longhetti, M., Maraston, C., et al. 2015, MNRAS, 454, 3912
Lonoce, I., Maraston, C., Thomas, D., et al. 2020, MNRAS, 492, 326
Maiolino, R., Cirasuolo, M., Afonso, J., et al. 2020, Msngr, 180, 24
Maiolino, R., & Mannucci, F. 2019, A&ARv, 27, 3
Marigo, P., Bressan, A., Nanni, A., Girardi, L., & Pumo, M. L. 2013, MNRAS, 434, 488
Marsan, Z. C., Muzzin, A., Marchesini, D., et al. 2022, ApJ, 924, 25
McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2021, MNRAS, 503, 4413
McLure, R. J., Pearce, H. J., Dunlop, J. S., et al. 2013, MNRAS, 428, 1088
McLure, R. J., Pentericci, L., Cimatti, A., et al. 2018, MNRAS, 479, 25
Merlin, E., Fortuni, F., Torelli, M., et al. 2019, MNRAS, 490, 3309
Mowla, L., van der Wel, A., van Dokkum, P., & Miller, T. B. 2019a, ApJL, 872, L13
Mowla, L. A., van Dokkum, P., Brammer, G. B., et al. 2019b, ApJ, 880, 57
Ocvirk, P., Pichon, C., Lançon, A., & Thiébaut, E. 2006, MNRAS, 365, 46
Onodera, M., Carollo, C. M., Renzini, A., et al. 2015, ApJ, 808, 161
Pacifici, C., Charlot, S., Blaizot, J., & Brinchmann, J. 2012, MNRAS, 421, 2002
Pacifici, C., Kassin, S. A., Weiner, B. J., et al. 2016, ApJ, 832, 79
Panter, B., Jimenez, R., Heavens, A. F., & Charlot, S. 2008, MNRAS, 391, 1117
Pentericci, L., McLure, R. J., Garilli, B., et al. 2018, A&A, 616, A174
Salim, S., Boquien, M., & Lee, J. C. 2018, ApJ, 859, 11
Santini, P., Castellano, M., Merlin, E., et al. 2021, A&A, 652, A30
Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018, A&A, 618, A85
Sherman, S., Jogee, S., Florez, J., et al. 2020, MNRAS, 491, 3318
Skilling, J. 2006, BayAn, 1, 833
Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159
Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, ApJL, 783, L14
Straatman, C. M. S., Spitler, L. R., Quadri, R. F., et al. 2016, ApJ, 830, 51
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2019a, ApJ, 877, 103
Suess, K. A., Kriek, M., Price, S. H., & Barro, G. 2019b, ApJL, 885, L22
Tacchella, S., Conroy, C., Faber, S., et al. 2022, ApJ, 926, 134
Thomas, D., Maraston, C., & Bender, R. 2003, MNRAS, 339, 897
van der Wel, A., Bezanson, R., D'Eugenio, F., et al. 2021, ApJS, 256, 44
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
van der Wel, A., Noeske, K., Bezanson, R., et al. 2016, ApJS, 223, 29
Weinberg, D. H., Andrews, B. H., & Freudenburg, J. 2017, ApJ, 837, 183
Whitaker, K. E., van Dokkum, P. G., Brammer, G., et al. 2013, ApJL, 770, L39
Wild, V., Taj Aldeen, L., Carnall, A., et al. 2020, MNRAS, 494, 529
Wilkinson, A., Almaini, O., Wild, V., et al. 2021, MNRAS, 504, 4533
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Wu, P.-F., van der Wel, A., Bezanson, R., et al. 2018a, ApJ, 868, 37
Wu, P.-F., van der Wel, A., Gallazzi, A., et al. 2018b, ApJ, 855, 85
Wuyts, S., Förster Schreiber, N. M., Lutz, D., et al. 2011, ApJ, 738, 106
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579