Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **487,** 2038–2060 (2019)
Advance Access publication 2019 May 21

doi:10.1093/mnras/stz1402

The VANDELS survey: the stellar metallicities of star-forming galaxies at $2.5 < z < 5.0$

F. Cullen,[1★] R. J. McLure,[1] J. S. Dunlop,[1] S. Khochfar,[1] R. Davé,[1,2,3] R. Amorín,[4,5] M. Bolzonella,[6] A. C. Carnall,[1] M. Castellano,[7] A. Cimatti,[8,9] M. Cirasuolo,[10] G. Cresci,[9] J. P. U. Fynbo,[11] F. Fontanot,[12] A. Gargiulo,[13] B. Garilli,[13] L. Guaita,[7,14] N. Hathi,[15] P. Hibon,[16] F. Mannucci,[9] F. Marchi,[7] D. J. McLeod,[1] L. Pentericci,[7] L. Pozzetti,[6] A. E. Shapley,[17] M. Talia[6,8] and G. Zamorani[6]

Affiliations are listed at the end of the paper

Accepted 2019 May 17. Received 2019 May 17; in original form 2019 March 28

ABSTRACT

We present the results of a study utilizing ultradeep, rest-frame UV, spectroscopy to quantify the relationship between stellar mass and stellar metallicity for 681 star-forming galaxies at $2.5 < z < 5.0$ ($\langle z \rangle = 3.5 \pm 0.6$) drawn from the VANDELS survey. Via a comparison with high-resolution stellar population synthesis models, we determine stellar metallicities (Z_*, here a proxy for the iron abundance) for a set of high signal-to-noise ratio composite spectra formed from subsamples selected by mass and redshift. Across the stellar mass range $8.5 < \log(\langle M_* \rangle / \mathrm{M}_\odot) < 10.2$, we find a strong correlation between stellar metallicity ($Z_*/\mathrm{Z}_\odot$) and stellar mass, with stellar metallicity monotonically increasing from $Z_*/\mathrm{Z}_\odot < 0.09$ at $\langle M_* \rangle = 3.2 \times 10^8\,\mathrm{M}_\odot$ to $Z_*/\mathrm{Z}_\odot = 0.27$ at $\langle M_* \rangle = 1.7 \times 10^{10}\,\mathrm{M}_\odot$. In contrast, at a given stellar mass, we find no evidence for significant metallicity evolution across the redshift range of our sample. However, comparing our results to the $z = 0$ stellar mass–metallicity relation for star-forming galaxies, we find that the $\langle z \rangle = 3.5$ relation is consistent with being shifted to lower metallicities by $\simeq$0.6 dex at all stellar masses. Contrasting our derived stellar metallicities with estimates of the gas-phase metallicities of galaxies at similar redshifts and stellar masses, we find evidence for enhanced O/Fe ratios in $z \gtrsim 2.5$ star-forming galaxies of the order $(\mathrm{O/Fe}) \gtrsim 1.8 \times (\mathrm{O/Fe})_\odot$. Finally, by comparing our results to the predictions of three cosmological simulations, we find that the $\langle z \rangle = 3.5$ stellar mass–metallicity relation is consistent with current predictions for how outflow strength scales with galaxy stellar mass. This conclusion is supported by an analysis of one-zone analytic chemical evolution models, and suggests that the mass-loading parameter ($\eta = \dot{M}_{\mathrm{outflow}}/M_*$) scales as $\eta \propto M_*^{\beta}$ with $\beta \simeq -0.4$.

Key words: galaxies: evolution – galaxies: high redshift.

1 INTRODUCTION

The relationship between the stellar mass and metallicity of galaxies as a function of cosmic epoch provides a fundamental constraint on models of galaxy formation (e.g. Maiolino & Mannucci 2019). The metallicity of a galaxy is determined by the integrated past history of star formation, the fraction of metal-enriched gas lost from the interstellar medium (ISM) through outflows, and the dilution of enriched ISM gas by pristine inflows from the intergalactic medium (IGM). Metallicity scaling relations therefore provide rich information regarding some of the key physical processes governing the evolution of galaxies.

In recent years, significant efforts have been made towards determining the gas-phase metal content of galaxies at $2 \lesssim z \lesssim 4$ from the ratios of strong optical nebular emission lines (e.g. Cullen et al. 2014; Maier et al. 2014; Steidel et al. 2014; Troncoso et al. 2014; Wuyts et al. 2014; Salim et al. 2015; Sanders et al. 2015; Onodera et al. 2016; Kashino et al. 2017; Sanders et al. 2018). Despite this, the substantial uncertainties inherent in converting nebular line ratios to abundances have frustrated efforts to reach a consensus. In addition to the known inconsistency between different metallicity calibrations in the local Universe (e.g. Kewley & Ellison 2008; Barrera-Ballesteros et al. 2017; Sánchez et al. 2019), the

★ E-mail: fc@roe.ac.uk

problem is compounded at higher redshifts by the evolution in the physical conditions in H II regions, potentially rendering local calibrations unusable (Steidel et al. 2014; Shapley et al. 2015; Cullen et al. 2016; Strom et al. 2017). Disconcertingly, it is still unclear whether the correlation between nebular line ratios and galaxy stellar mass in high-redshift galaxies is dictated primarily by the gas-phase metallicity or is purely driven by the ionizing properties of the massive stellar population (e.g. Steidel et al. 2014). To definitively address these concerns, direct estimates of gas-phase metallicity are needed at high redshift, requiring the detection of faint oxygen auroral lines such as [O III]λ4363 and O III]λ1661,1666, which are extremely challenging observations that have only been achieved a handful of times at $z \gtrsim 2$ (e.g. Sanders et al. 2016; Amorín et al. 2017).

An alternative method for probing the metal content of galaxies utilizes their stellar continuum emission directly. Abundances derived in this way are referred to as stellar (Z_*) as opposed to gas-phase (Z_g) metallicities. When available, stellar metallicities are a useful independent probe of the metal content of galaxies. In the local Universe, several authors have investigated the stellar mass–metallicity relationship for large statistical galaxy samples, primarily from the extensive Sloan Digital Sky Survey (SDSS) data set (e.g. Gallazzi et al. 2005; Panter et al. 2008; Zahid et al. 2017; Trussler et al. 2018), integral field spectroscopic surveys (e.g. CALIFA, MaNGA, SAMI; González Delgado et al. 2014; Scott et al. 2017; Lian et al. 2018b), and, at lower stellar masses, from stellar spectroscopy in local dwarf galaxies (e.g. Kirby et al. 2013). In the majority of these cases, metallicities have been derived from rest-frame optical continuum observations. Recently, Zahid et al. (2017) presented a compilation of Z_* estimates in the local Universe spanning ≃7 orders of magnitude in stellar mass (ranging from $M_* \simeq 10^4\,\mathrm{M}_\odot$ to $M_* \simeq 10^{11}\,\mathrm{M}_\odot$) finding evidence for a continuous relation that rises from $\log(Z_*/Z_\odot) \simeq -2.5$ at the lowest stellar masses, up to $\log(Z_*/Z_\odot) \simeq 0.0$, flattening at stellar masses above $M_* \simeq 10^{10}\,\mathrm{M}_\odot$. Determining whether or not a similar relation was already in place at earlier cosmic epochs is clearly of significant interest.

Unfortunately, estimates of Z_* for galaxies at $z \gtrsim 2$ are rare primarily because the measurement requires a high signal-to-noise (S/N) ratio detection of the stellar continuum, an expensive and challenging observation for faint high-redshift sources. Nevertheless, estimates have been made for small samples of lensed galaxies at $z \sim 2-3$ (Rix et al. 2004; Quider et al. 2009; Dessauges-Zavadsky et al. 2010), a number of unlensed sources at $z > 3$ (Sommariva et al. 2012) and a composite spectrum of star-forming galaxies at $z \simeq 2$ (Halliday et al. 2008). In all of these cases, the observations have been taken with ground-based optical spectrographs and therefore the stellar metallicities are based on rest-frame far-ultraviolet (FUV) spectra. The distinction is important because FUV-based metallicities are considered to be, to a first approximation, a measurement of the iron abundance in the photospheres of the young, massive, O- and B-type stars in the galaxy (e.g. Halliday et al. 2008). This is not necessarily the case for optical-based stellar metallicities, which trace more evolved stars and sample longer star formation time-scales, and is certainly not the case for estimates of Z_g derived from optical nebular emission lines, which trace the young stellar population, but mainly the oxygen and nitrogen abundance.

Most commonly, stellar metallicities at high-redshift have been determined using a calibration of FUV photospheric absorption indices developed by Rix et al. (2004), and later extended in Sommariva et al. (2012). However, the results from these studies have generally been inconclusive, especially with respect to the scaling relation between metallicity and stellar mass. This has been, in part, a consequence of the small sample sizes and subsequent lack of dynamic range, but is also a consequence of the fact that these indices are also prone to significant contamination by ISM absorption (e.g. Vidal-García et al. 2017). More recently, an alternative method for measuring Z_*, by fitting the full FUV spectrum at $\lambda \leq 2000$ Å, has been demonstrated by Steidel et al. (2016; S16) using the BPASSv2.0 stellar population synthesis models including massive stellar binaries (Eldridge et al. 2017). The method is a natural extension of the previous methods, making more complete and consistent use of the available data and should, in principle, place more robust constraints on Z_* than the standard indices approach (e.g. Walcher et al. 2011; Conroy et al. 2018).

Applying the full FUV spectral fitting method to a composite spectrum of 30 star-forming galaxies at $z = 2.4$, Steidel et al. (2016) found a low FUV-based stellar metallicity of $Z_*/Z_\odot \simeq 0.1$, roughly a factor of 5 lower than the gas-phase metallicity derived from rest-frame optical nebular emission lines ($Z_g/Z_\odot \simeq 0.5$), which they interpreted as evidence for supersolar oxygen-to-iron ratios (O/Fe) in high-redshift star-forming galaxies. This result was subsequently confirmed using a larger statistical sample of 150 star-forming galaxies at $z \sim 2-3$ in Strom et al. (2018) albeit via a slightly different method in which Z_* and Z_g were simultaneously estimated from fitting optical nebular emission lines. However, although Steidel et al. (2016) established that star-forming galaxies at $z \gtrsim 2$ have significantly subsolar stellar metallicities (iron abundances), they did not explore the existence of a scaling relation of Z_* with stellar mass. Therefore, determining the stellar mass–metallicity relation at $z \gtrsim 2$ remains a key open question in the study of galaxy evolution.

Fortunately, further progress can be made in this area using the deep spectroscopic data taken as part of the VANDELS survey (McLure et al. 2018b; Pentericci et al. 2018). VANDELS is an ESO public spectroscopic survey providing exceptionally deep integration times (of up to 80 h) for ∼2000 galaxies at $z \gtrsim 1$. At redshifts $2.5 < z < 5.0$, the VANDELS spectra cover the rest-frame FUV, allowing stellar metallicities to be estimated using the methods described above. Moreover, since all of the VANDELS targets are drawn from the CDFS and UDS survey fields, accurate stellar masses can be determined for all sources using the deep multiwavelength photometric catalogues available in these regions. In this paper, we present a study of the stellar mass–metallicity relationship at $2.5 < z < 5.0$ using deep rest-frame FUV spectra from the VANDELS survey.

The structure of the paper is as follows. In Section 2, we describe the selection of our VANDELS star-forming galaxy sample along with the other relevant data sets used in this work. In Section 3, we begin by reviewing the metallicity information contained within rest-frame FUV galaxy spectra, and outline our method for determining FUV-based stellar metallicities. In Section 4, we present our stellar mass–metallicity relation and compare to the predictions from three independent cosmological simulations. In Section 5, we discuss some of the implications of our results, including an investigation into which physical parameters are potentially driving the observed stellar metallicities in our sample, and a discussion of O/Fe ratios in high-redshift star-forming galaxies. Finally, in Section 6 we summarize our results and conclusions. Throughout this paper, metallicities are quoted relative to the solar abundance taken from Asplund et al. (2009) that has a bulk composition by mass of $Z_* = 0.0142$, an iron mass fraction of $Z_{*,\mathrm{Fe}} = 0.0013$, and an oxygen mass fraction of $Z_{*,\mathrm{O}} = 0.0058$, and we assume the following cosmology: $\Omega_M = 0.3, \Omega_\Lambda = 0.7, H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

All magnitudes are quoted in the AB magnitude system (Oke & Gunn 1983).

2 DATA

In this section, we present the various observed and simulated data sets used in this work. The data set of primary interest, described in Section 2.1, is the rest-frame FUV spectra of star-forming galaxies at $2.5 < z < 5.0$ provided by the VANDELS survey. These high-redshift data are supplemented by a selection of FUV spectra of starburst galaxies and star-forming regions in the local Universe that are described in Section 2.2. Finally, in Section 2.3, we discuss the data extracted from two state-of-the-art cosmological simulations (FIBYand SIMBA; Paardekooper, Khochfar & Dalla Vecchia 2015; Davé et al. 2019) that are used to test our method and to aid in the analysis and interpretation of our results.

2.1 VANDELS

The primary spectroscopic data were obtained as part of the VANDELS ESO public spectroscopic survey (McLure et al. 2018b; Pentericci et al. 2018). The VANDELS survey is a deep optical spectroscopic survey of the CANDELS CDFS and UDS fields with the VIMOS spectrograph on ESO's Very Large Telescope (VLT) targeting massive passive galaxies at $1.0 \leq z \leq 2.5$, bright star-forming galaxies at $2.4 \leq z \leq 5.5$, and fainter star-forming galaxies at $3.0 \leq z \leq 7.0$. In this work, we focus exclusively on the star-forming galaxies at $z \leq 5$.

All galaxies were drawn from four independent *H*-band-selected catalogues. The CDFS and UDS regions are covered by the CANDELS survey, and benefit from extensive WFC3/IR imaging (CDFS-*HST*, UDS-*HST*; Grogin et al. 2011; Koekemoer et al. 2011). In these fields, photometry catalogues produced by the CANDELS team were used (Galametz et al. 2013; Guo et al. 2013). Within the wider field regions, two bespoke PSF-homogenized catalogues were produced, primarily from publicly available ground-based imaging (CDFS-GROUND, UDS-GROUND; McLure et al. 2018b). From these catalogues, objects were pre-selected as potential spectroscopic targets based on robust photometric redshift estimates. The bright star-forming galaxies were chosen to satisfy $2.4 \leq z_{\rm phot} \leq 5.0$ with an *i*-band magnitude of $i \leq 25$. This *i*-band constraint was chosen to ensure the final spectra had sufficient S/N to allow detailed analyses of individual objects. The faint star-forming galaxies relevant to our study (i.e. in the redshift interval $3.0 \leq z_{\rm phot} \leq 5.5$) were selected to have $25 \leq H \leq 27 \wedge i \leq 27.5$ in the *HST* regions.[1] As discussed in McLure et al. (2018b; illustrated in their fig. 3), this selection ensured that the sample of star-forming galaxies is consistent with being drawn from the main sequence of star-forming galaxies at all redshifts. An additional constraint, enforced by the observing strategy, was placed on the *i*-band magnitudes so that slits were efficiently allocated to objects requiring 20, 40, and 80 h of integration time in a roughly 1:2:1 ratio (McLure et al. 2018b). The final spectroscopic sample of faint star-forming galaxies was an unbiased random (approximately 1 in 4) subsample of the pre-selection catalogue.

The observations and reduction of the VIMOS spectra are described in detail in the first data release paper (Pentericci et al. 2018). Briefly, observations were obtained with the ESO-VLT VIMOS spectrograph using the medium resolution grism that covers the wavelength range $4800 < \lambda_{\rm obs} < 10000$ Å with a resolution of $R = 580$ and a dispersion of 2.5 Å pixel^{-1}. The median seeing across all observations was FWHM = 0.7 arcsec. The spectra were reduced using the EASY-LIFE pipeline (Garilli et al. 2012) that is an update of the algorithms originally described in Scodeggio et al. (2005). EASY-LIFE performs all standard data reduction procedures including the removal of bad CCD pixels and sky subtraction at the individual exposure level. The 2D spectrum for each individual exposure is extracted and resampled on to a common linear wavelength scale; these individual-exposure 2D spectra are then combined to create the final 2D spectrum for each object. The 1D spectra are extracted, flux calibrated using the spectrophotometric standard stars, and corrected for telluric absorption. A final correction for atmospheric and galactic extinction is applied and the 1D spectra are normalized to the *i*-band photometry.

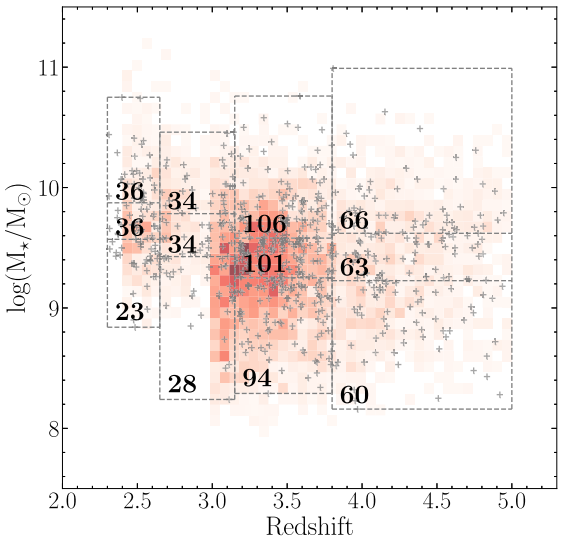

Figure 1. Stellar mass versus redshift for the VANDELS DR2 spectra at $2.3 < z < 5.0$ showing (via the grey dashed grid) the bins in stellar mass and redshift used to produce composite spectra. The number within each box indicates the number of individual galaxies in that bin, which are marked by the small grey '+' symbols. The background 2D histogram shows all the galaxies within the initial pre-selection catalogue with $2.4 \leq z_{\rm phot} \leq 5.0$, illustrating how the observed galaxies are effectively randomly drawn from this population.

We draw our sample from the second VANDELS data release[2] that comprises roughly 65 per cent (1362/2106) of the full survey. Redshifts for all of the spectra have been determined by members of the VANDELS team and assigned a redshift quality flag as described in Pentericci et al. (2018). We only consider galaxies at $2.3 \leq z_{\rm spec} \leq 5.0$ and with a redshift quality flag 3 or 4 (corresponding to a ≥95 per cent probability of being correct), leaving a total of 681 galaxies. Fig. 1 shows the distribution of our final sample in the mass–redshift plane compared to the underlying distribution of all $2.4 \leq z_{\rm phot} \leq 5.0$ star-forming galaxies in the pre-selection catalogue. We have extended our selection to $z_{\rm spec} \geq 2.3$ so that we include galaxies whose photometric redshift was slightly overestimated at the lower redshift boundary of the star-forming galaxy selection ($z_{\rm phot} = 2.4$). The upper redshift limit of $z = 5$ is

[1] In the wider field regions, the limiting *i*-band magnitude is slightly brighter at $i \leq 26.0$.

[2] Data available through the consortium data base at the link http://vandels.inaf.it/db.

enforced primarily because above this redshift the FUV wavelength coverage becomes prohibitively small. The average redshift of our final sample is $\langle z \rangle = 3.5 \pm 0.6$. Stellar masses for the full VANDELS sample were derived from SED fitting using Bruzual & Charlot (2003) templates with solar metallicity and assuming exponentially declining star formation histories (SFHs) as described in McLure et al. (2018b). We also derived stellar masses for our $2.5 < z < 5.0$ sample using FAST++ , a rewrite of FAST (Kriek et al. 2009) described in Schreiber et al. (2018). We adopted the same set of parameters as McLure et al. (2018b), but allowed for rising SFHs using a delayed exponentially declining model ($M_* \propto te^{t/\tau}$) and a range of metallicities values ($0.3 - 2.5 \times Z_\odot$). The stellar masses returned from these two procedures were fully consistent, but for this work we adopt the FAST++ values.

As discussed in McLure et al. (2018b), the pre-selection catalogue contains star-forming galaxies consistent with being drawn from the star-forming main sequence across the full redshift range. Therefore, since the final selection is a random subsample of galaxies from the pre-selection catalogue, the galaxies in our final sample can be considered typical star-forming galaxies for their stellar mass. The stellar mass range of the final sample is $8.2 < \log(M_*/\mathrm{M}_\odot) < 11.0$ with a median specific star formation rate (SFR) of $\mathrm{sSFR} = 4.4\,\mathrm{Gyr}^{-1}$.

2.1.1 *Composite FUV spectra*

The typical continuum S/N is too low to extract metallicity information therefore we focus our analysis on stacked spectra in bins of stellar mass and redshift. The mass and redshift bins are illustrated in Fig. 1 and detailed in Table 1. In total, we chose four redshift bins that were selected to encompass roughly equivalent intervals of cosmic time (≈400 Myr), with three stellar-mass bins per redshift. We also analyse seven composite spectra stacked in bins of stellar mass across the full redshift range, with the stellar mass ranges given in Table 1. The mass bins were chosen manually with the aim of keeping the bin widths as narrow as possible whilst also ensuring that they contained enough galaxies to have an acceptable S/N.

The final stacked composite spectra were formed by first shifting each contributing spectrum into the rest-frame using z_{spec}. To correct for redshift differences, the flux of each spectrum was scaled to the flux that would be observed at the mean redshift of the stack. For the stellar mass–redshift stacks, these flux-correction factors are all $\lesssim$5 per cent; for the stellar-mass-only stacks the maximum average correction is 15 ± 7 per cent that occurs in the highest redshift bin. We have confirmed that our results do not change significantly if, instead of this flux-correction method, the spectra are simply normalized and then combined. The primary benefit of preserving the absolute flux values is that an error spectrum can be robustly estimated. The individual spectra were then resampled on to a common wavelength grid, which varied slightly depending on the redshift range (see Table 1), with a dispersion of 1Å per pixel. The final flux at each dispersion point was taken as the median of all the individual flux values after rejecting 3σ outliers, and the 1σ error was calculated from bootstrap re-sampling of the individual flux values. An example of one composite spectrum is shown in Fig. 2. The effective spectral resolution element of the composites is 3.0 Å.

2.2 FOS–GHRS local sample

The main high-redshift sample was supplemented by a sample of local galaxies for which published rest-frame FUV spectra are available. The motivation for including a local reference sample was that the majority of published stellar metallicities in the local Universe are derived from rest-frame optical spectra (e.g. Gallazzi et al. 2005; Panter et al. 2008; González Delgado et al. 2014; Zahid et al. 2017), and it is not clear that optical and FUV light will originate from the same population of stars or trace the same abundance type. Even in the cases where rest-frame UV spectra are analysed, the focus is primarily on deriving gas-phase metallicities (not stellar metallicities) from the depths of ISM absorption features (e.g. Leitherer et al. 2011; Zetterlund et al. 2015; Faisst et al. 2016). To our knowledge, there are no local examples of stellar metallicities derived solely from global FUV continuum fitting. For this reason, we have constructed a comparison sample of local starbursts and star-forming galaxies from the FUV spectroscopic atlas presented in Leitherer et al. (2011), which is comprised of spectra taken with the faint object spectrograph (FOS) and Goddard high resolution spectrograph (GHRS).

Our sample is drawn from the 46 rest-frame UV spectra observed with the FOS and the GHRS on-board the *HST* presented in Leitherer et al. (2011). These 46 spectra are drawn from 28 individual galaxies (i.e. some spectra are just different regions of the same galaxy). The spectral resolution of the individual spectra spans the range 0.5–3 Å depending on the instrument, aperture size, and physical extent of the object being observed. For consistency with the VANDELS data, we smooth all spectra to a common 3 Å resolution. We also require that the spectra have wavelength coverage in the interval $1410 \leq \lambda \leq 1450$ Å to enable an analysis of normalized composite spectra, and finally that the corresponding galaxy has a measurement of absolute K-band magnitude that we can use to estimate the stellar mass using the relation from McGaugh & Schombert (2014). This selection leaves 26 of 46 of the original spectra from 18 of 28 of the original local starbursts and star-forming galaxies presented in Leitherer et al. (2011). A list of the individual spectra used in this work is given in the Appendix (Table A1). Composite spectra were constructed in three bins of stellar mass as outlined in Table A1. To create the composites, all individual spectra were median combined on the same wavelength grid used for the VANDELS galaxies. In this case, an error spectrum was estimated by propagating the error spectra of each of the individual spectra (bootstrapping could not be applied in this case since the individual spectra were normalized).

2.3 Simulation data

To compare the observations to simulation predictions, we have extracted data from two state-of-the-art cosmological hydrodynamical simulations: FIBY (e.g. Johnson, Dalla Vecchia & Khochfar 2013; Paardekooper et al. 2015; Cullen et al. 2017) and SIMBA (Davé et al. 2019). In Section 3.4, we use the simulation data to assess the accuracy of our adopted method for measuring stellar metallicities, and in Section 4 we compare the observed stellar mass–metallicity relationship to the simulation predictions. Having two independent simulations allows us to account for systematic variations between their methodologies and theoretical predictions. Here, we give a brief overview of the simulation details but refer the reader to the references above for further information.

2.3.1 *FIBY*

The FIBY simulation suite is a set of high-resolution cosmological hydrodynamical simulations using a modified version of

Table 1. Details of the VANDELS composite spectra stacked in bins of stellar mass and redshift.

Composite ID	Bin range $\log(M_*/M_\odot)$	Median $\log(M_*/M_\odot)$	$\langle z \rangle$	Wavelength coverage (Å)	N_{gal}	Median S/N per pixel	$\log(Z_*)$
$2.30 \leq z \leq 5.00$:							
VANDELS-m1	8.16–8.70	8.51	3.82	1000–2000	38	6	<-2.89
VANDELS-m2	8.70–9.20	9.00	3.67	1000–2000	131	14	-2.78 ± 0.03
VANDELS-m3	9.20–9.50	9.36	3.47	1000–2000	153	19	-2.78 ± 0.03
VANDELS-m4	9.50–9.65	9.56	3.50	1000–2000	111	15	-2.70 ± 0.03
VANDELS-m5	9.65–9.80	9.71	3.48	1000–2000	95	15	-2.71 ± 0.03
VANDELS-m6	9.80–10.00	9.89	3.34	1000–2000	74	14	-2.56 ± 0.06
VANDELS-m7	10.00–11.00	10.24	3.24	1000–2000	79	12	-2.42 ± 0.06
$2.30 \leq z < 2.65$:							
VANDELS-z1-m1	8.84–9.47	9.30	2.54	1450–2000	23	13	-2.64 ± 0.06
VANDELS-z1-m2	9.49–9.81	9.74	2.50	1450–2000	36	14	-2.68 ± 0.05
VANDELS-z1-m3	9.82–10.75	9.82	2.50	1450–2000	36	13	-2.50 ± 0.08
$2.65 \leq z < 3.15$:							
VANDELS-z2-m1	8.24–9.39	9.08	2.98	1200–2000	28	8	-2.64 ± 0.11
VANDELS-z2-m2	9.41–9.74	9.57	2.92	1200–2000	34	9	-2.63 ± 0.10
VANDELS-z2-m3	9.78–10.46	10.02	2.97	1200–2000	34	9	-2.44 ± 0.10
$3.15 \leq z < 3.80$:							
VANDELS-z3-m1	8.29–9.22	8.94	3.50	1100–2000	94	12	-2.90 ± 0.03
VANDELS-z3-m2	9.23–9.57	9.43	3.45	1100–2000	101	16	-2.82 ± 0.03
VANDELS-z3-m3	9.58–10.76	9.99	3.50	1100–2000	106	15	-2.68 ± 0.04
$3.80 \leq z \leq 5.00$:							
VANDELS-z4-m1	8.16–9.20	8.85	4.31	900–2000	60	8	-2.89 ± 0.05
VANDELS-z4-m2	9.21–9.60	9.41	4.26	900–2000	63	9	-2.71 ± 0.11
VANDELS-z4-m3	9.61–11.00	9.87	4.26	900–2000	66	10	-2.59 ± 0.08

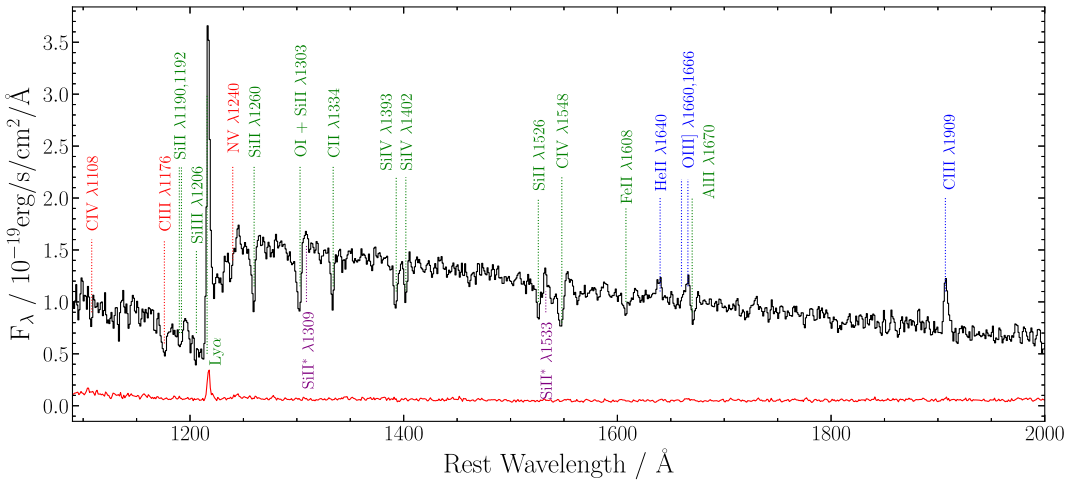

Figure 2. An example composite spectrum (black curve) built from the spectra of the 131 galaxies in our sample within the mass range $8.7 < \log(M_*/M_\odot) < 9.2$ (VANDELS-m2; see Table 1 for details) along with the error spectrum (red line) estimated via a bootstrap re-sampling technique. The galaxies contributing to the composite span the full redshift range of our sample ($2.3 < z < 5.0$) with a mean redshift of $z = 3.67$. To account for redshift differences, the flux of each individual spectrum was scaled to the flux that would be observed at the mean redshift. The labels indicate prominent emission/absorption features colour-coded by their physical origin: interstellar absorption (green), stellar absorption (red), nebular/stellar emission (blue), and fine structure emission (purple). The remainder of the FUV spectrum is dominated by absorption due to heavy photospheric line-blanketing in the atmosphere of massive O- and B-type stars; these are the regions that were used to constrain the stellar metallicity (Section 3.3 for details).

the GADGET code used in the Overwhelmingly Large Simulations (OWLS) project (Schaye et al. 2010). The code tracks metal pollution for 11 elements: H, He, C, N, O, Ne, Mg, Si, S, Ca, and Fe and calculates the cooling of gas based on line-cooling in photoionization equilibrium for these elements (Wiersma, Schaye & Smith 2009) using tables pre-calculated with CLOUDY v07.02 (Ferland et al. 1998). Furthermore, the simulation incorporates full non-equilibrium primordial chemistry networks (Abel et al. 1997; Galli & Palla 1998; Yoshida et al. 2006) including molecular cooling functions for H_2 and HD. Star formation is modelled using the pressure law implementation of Schaye & Dalla Vecchia (2008), which yields results consistent with the Schmidt–Kennicutt law (Schmidt 1959; Kennicutt 1998). The simulations include feedback from stars by injecting thermal energy into the neighbouring

particles (Dalla Vecchia & Schaye 2012). Element yields from Type Ia and core-collapse supernovae (CCSNe) are implemented following the prescription of Wiersma et al. (2009). In this work, we will focus on the FIBY_XL simulation that has individual gas and star particle masses of $\log(M_*/\mathrm{M}_\odot) = 5.68$ and covers a co-moving volume of $(32\,\mathrm{Mpc})^3$. The lowest redshift simulated in the FIBY simulation is $z = 4$, and this is the closest redshift to the mean redshift of our sample ($\langle z \rangle = 3.5$). Therefore, for our FIBY comparison sample, we extracted the 591 galaxies in FIBY_XL with $\log(M_*/\mathrm{M}_\odot) > 8.0$ at $z = 4$. The maximum galaxy stellar mass is $M_* = 3.8 \times 10^{10}\,\mathrm{M}_\odot$, with a median value of $4.0 \times 10^8\,\mathrm{M}_\odot$.

2.3.2 SIMBA

The SIMBA simulation suite is based on the GIZMO cosmological gravity plus hydrodynamics solver (Hopkins 2015, 2017) and is described in detail in Davé et al. (2019). The code implements a H_2-based SFR, with H_2 fractions calculated using the subgrid model of Krumholz & Gnedin (2011). The chemical enrichment model tracks the same 11 elements as in FIBY, with radiative cooling and photoionization heating modelled using the GRACKLE-3.1 library (Smith et al. 2017), and element yields calculated for CCSNe, Type Ia SNe, and asymptotic giant branch (AGB) stars following the prescriptions of Nomoto et al. (2006), Iwamoto et al. (1999), and Oppenheimer & Davé (2008), respectively. SIMBA employs a mass outflow rate scaling with galaxy stellar mass motivated by the results derived from the extremely high resolution FIRE zoom simulations (Muratov et al. 2015; Anglés-Alcázar et al. 2017). In this work, we focus on the data from the m50n1024 simulation that has a co-moving box length of $50\,h^{-1}$Mpc and individual gas/star element resolution of $2.28 \times 10^6\,\mathrm{M}_\odot$. For consistency with FIBY, we focus on the $z = 4$ snapshot in the simulation. From this, we extract a sample of 1749 galaxies down to a minimum stellar mass of $4.5 \times 10^8\,\mathrm{M}_\odot$, with a maximum stellar mass of $1.7 \times 10^{11}\,\mathrm{M}_\odot$ and a median value of $8.2 \times 10^8\,\mathrm{M}_\odot$.

3 MEASURING STELLAR METALLICITIES

We now describe our method for estimating stellar metallicities from rest-frame FUV spectra. In Section 3.1, we describe the metallicity information content of the rest-frame FUV spectra of star-forming galaxies. The stellar population synthesis models we use to compare to the data are discussed in Section 3.2. Then, in Section 3.3, we describe the statistical method we employ to robustly constrain the metallicity. Finally, in Section 3.4 we discuss various tests of our method.

3.1 Stellar metallicity indicators in the UV

To begin with, it is worthwhile reviewing what metallicity information is contained within the FUV spectrum of a star-forming galaxy. Figs 2 and 3 serve as a useful references for this discussion. From simple photon energy arguments, we know that the rest-frame FUV is dominated by the light from young, massive, O- and B-type stars. These stars emit a continuum that contains absorption (and potentially emission) features due to elements within the stellar photosphere and the expanding stellar wind. The strength of these features is naturally a strong function of the total photospheric metallicity (e.g. Leitherer et al. 2010).

The most prominent absorption features are typically the strong stellar wind lines produced from the stellar atmospheres, namely

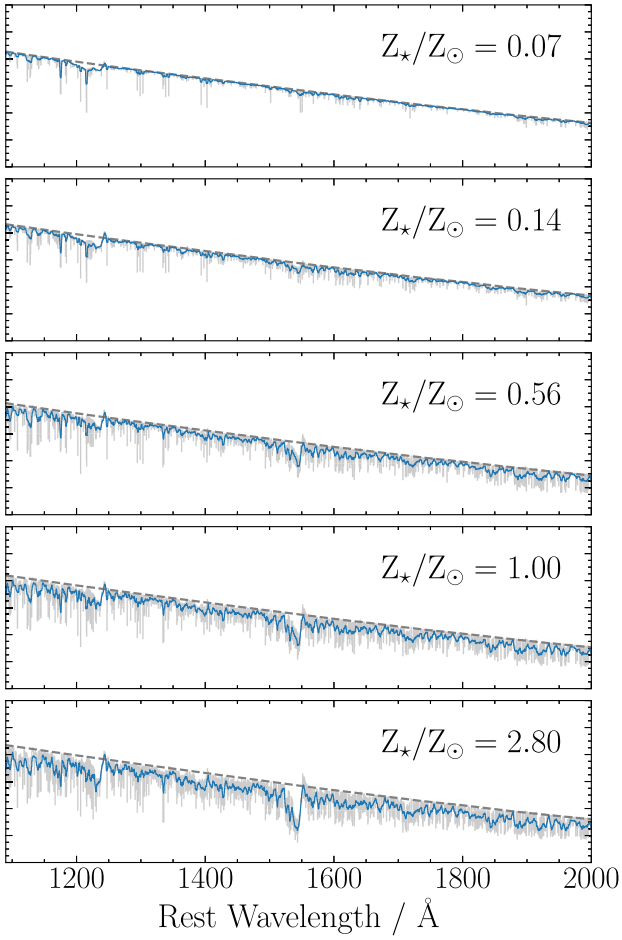

Figure 3. The Starburst99 (SB99) high-resolution WM-Basic stellar population models used in this work. All models assume a constant star formation rate over 100 Myr. Each panel shows the model FUV spectra at one of the five default metallicities provided by the Starburst99 Geneva tacks ($Z_* = 0.001$, 0.002, 0.008, 0.014, and 0.040) running from the lowest metallicity (top panel) to the highest (bottom panel). Values of $Z_*/Z_\odot$ are indicated in the legend. The background grey spectrum shows the full resolution SB99 models (0.4 Å), and the blue spectrum shows the models at the resolution of the VANDELS data (3 Å). The grey dashed line shows the intrinsic stellar emission (i.e. before passing through the stellar photosphere). The figure illustrates how increasing the photospheric metallicity results in heavier FUV line-blanketing that, at the resolution of our data, results in a depression of the FUV continuum at all wavelengths.

N Vλ1240, Si IVλ1400, and C IVλ1548. These lines are often extremely broad, blueshifted by $\sim 1000\,\mathrm{km\,s^{-1}}$ and exhibit P-Cygni line profiles. They are useful metallicity diagnostics since the strength of the stellar winds, and hence the line profiles, has a strong metallicity dependence that is linked to the metallicity-dependent mass-loss rates (Kudritzki & Puls 2000; Puls, Vink & Najarro 2008). Naturally, however, these lines are also sensitive to the initial mass function (IMF) and SFH.

Outside these strong wind features, the rest-frame FUV shortward of $\lambda = 2000$ Å contains an abundance of stellar photospheric absorption features, primarily due to transitions of highly ionized iron (Dean & Bruhweiler 1985; Brandt et al. 1998). The individual absorption features have low equivalent widths that require high spectral resolution and S/N to be seen individually, but the resulting strong photospheric line blanketing can be seen in unresolved O- and B-star populations in local starburst galaxies, star-forming regions,

and even high-redshift galaxies (Leitherer et al. 2011; Halliday et al. 2008).

These photospheric absorption features have been utilized in the past as metallicity indicators, for example, Rix et al. (2004) developed the '1978 index' at 1935–2050 Å, a wavelength regime that is believed to be dominated by Fe III transitions in early B-type stars. The 1978 index, which is thought to be a direct probe of the iron abundance ([Fe/H]), was used to derive the stellar metallicity of a composite spectrum of 75 star-forming galaxies at $z \approx 2$ by Halliday et al. (2008). Additional wavelength regions (e.g. Fe Vλ1363, Si IIIλ1417, C III$\lambda\lambda$1426/28, S Vλ1502) have been investigated by other authors and shown to be similarly sensitive to the photospheric abundance (e.g. Sommariva et al. 2012). These individual absorption features are analogous to the Lick index system developed for constraining abundances from rest-frame optical spectra (Faber et al. 1985; Thomas, Maraston & Bender 2003). However, Vidal-García et al. (2017) have recently demonstrated that this approach is limited by the fact that the many of these indices are significantly contaminated by ISM absorption lines.

An alternative approach, demonstrated most recently by Steidel et al. (2016), is to simply fit the full FUV spectra directly, thereby constraining the FUV abundance using all of the metallicity-sensitive regions simultaneously. Indeed, similar approaches have proven successful for constraining stellar metallicities at optical wavelengths (e.g. Panter et al. 2008; González Delgado et al. 2014). The obvious merit of this approach is that potentially relevant information is not being thrown away (which may be the case when focusing on specific, narrow, wavelength regions; see e.g. Conroy et al. 2018). On the other hand, the fact that the absorption features across the full FUV spectrum are a result of a number of different element species means there is some ambiguity as to what abundance is actually being measured. Steidel et al. (2016; among others) have pointed out that, because the vast majority of the FUV line-blanketing is a result of transitions in highly ionized iron, a metallicity derived by fitting the FUV spectrum is, to a first approximation, a measure of the iron abundance in the stellar photosphere. The point is demonstrated in table 1 of Leitherer et al. (2011), which provides a comprehensive list of the photospheric spectral lines identified in star-forming galaxies both locally and from lensed high-redshift sources (Pettini et al. 2000). In this work, we will adopt a full spectra fitting approach similar to Steidel et al. (2016), which is outlined in detail in Section 3.3. Throughout this paper, we refer to the stellar metallicity derived from rest-frame FUV observations as $Z_*/Z_\odot$ (or $\log(Z_*/Z_\odot)$, but emphasize that this should be understood as a proxy for the iron abundance.

A final important point to note is that, while these FUV absorption features are strongly dependent on metallicity, the low-order shape of the FUV continuum is not (in contrast to the case at optical wavelengths). By far the dominant factor in determining the FUV continuum shape is the wavelength-dependent dust attenuation law. Interestingly, the attenuation law derived from local starbursts by Calzetti et al. (2000) still appears to be a good approximation for the average shape of the attenuation law for star-forming galaxies at high-redshift with $\log(M_*/\mathrm{M}_\odot) \gtrsim 9.5$ (e.g. Cullen et al. 2017, 2018; McLure et al. 2018a), although this is still a matter of debate (e.g. Reddy et al. 2018), and object by object variation, and even some stellar mass dependence, is expected (e.g. Kriek & Conroy 2013). Crucially, however, the lack of metallicity dependence on the global continuum shape (i.e the overall curvature of the FUV spectrum) negates any strong degeneracies between the metallicity and the wavelength-dependence/normalization of the dust attenuation.

3.2 Stellar population synthesis models

To derive metallicities from the observed FUV spectra, we require models to compare to the data. A number of different stellar population synthesis models exist (e.g. Bruzual & Charlot 2003; Mollá, García-Vargas & Bressan 2009; Leitherer et al. 2010; Eldridge et al. 2017), but in this work we have opted to use the Starburst99 (SB99) high-resolution WM-Basic theoretical stellar library described in Leitherer et al. (2010). Our choice was motivated partly for clarity, and also by the fact that the WM-Basic SB99 spectra are the highest resolution FUV spectral models available, and have been extensively tested against individual spectra of hot stars in the Galaxy and Magellanic Clouds. We argue that the WM-Basic SB99 models still represent the most robust stellar population models available for comparing to spectroscopic data in the FUV. However, the SB99 models do not account for some physical effects that are known to be important in massive star evolution, such as the prevalence of binary stars and the phenomenon of 'quasi-homogeneous evolution' at low metallicities (e.g. Yoon & Langer 2005; Eldridge & Stanway 2012). In Appendix C, we have demonstrated that using stellar populations models that do account for these phenomena (e.g. BPASSv2.1; Eldridge et al. 2017) would not affect our main results.

We considered constant SFR models from the latest version of SB99 (Leitherer et al. 2014), assuming the weaker wind Geneva tracks with stellar rotation and single-star evolution. We considered the Geneva evolution tracks at $Z_* = (0.001, 0.002, 0.008, 0.014, 0.040)$, and an $\eta = 2.3$ IMF slope in the mass range $0.5 < M_*/M_\odot < 100$ corresponding to the default Kroupa (2001) IMF. The high-resolution WM-basic spectra provided by Starbust99 cover the wavelength region 900–3000 Å at a dispersion of 0.4 Å pixel^{-1}, so we smoothed and resampled the models to match the resolution ($\approx$3 Å) and sampling (1Å pixel^{-1}) of the VANDELS composites (Fig. 3). We considered constant star formation models over timescales of 100, 300, and 500 Myr but adopt the 100 Myr models as our fiducial set (we discuss the motivation for this choice next).

Nebular continuum emission was added to all models using CLOUDY v17.00 (Ferland et al. 2017) following the method described in Cullen et al. (2017), assuming H II region parameters consistent with the current best estimates at $z \simeq 2 - 3$ (Strom et al. 2018). When fitting the observed data we considered models with both maximal nebular contribution ($f_{\rm esc} = 0$ per cent) and zero nebular contribution ($f_{\rm esc} = 100$ per cent). However, since the average escape fraction is measured to be relatively low in high-redshift star-forming galaxies ($\lesssim$20 per cent; Grazian et al. 2017; Fletcher et al. 2018; Steidel et al. 2018), we adopt the models including maximal nebular continuum as our fiducial set. We also note that, since the nebular continuum only acts to alter the shape of the continuum it is degenerate with the dust prescription, not with the stellar metallicity. Assuming $f_{\rm esc} = 100$ per cent does not change the results presented here.

Finally, in Section 3.4, we will discuss how we tested our FUV-fitting method using synthetic FUV spectra derived from FIBY and SIMBA simulation data. For generating these synthetic spectra, we also used WM-Basic SB99 models, but in this case considered the instantaneous burst models generated at 1000 time-steps split logarithmically between 1 Myr and 1.5 Gyr. These instantaneous burst models are required so that the synthetic spectra can be constructed to match the known SFH and chemical-abundance histories of the simulations. The instantaneous burst models assume a total stellar mass within the burst of $10^6\,\mathrm{M}_\odot$ with all other parameters (e.g. IMF, metallicity range) being identical to the

constant SFR models. As will be described in more detail next, the synthetic spectra were constructed from these instantaneous burst models using the average star formation and chemical-abundance histories of the simulated galaxies following the method outlined in Cullen et al. (2017).

3.3 A statistical estimate of the stellar metallicity

To fit the SB99 models to our data, we adopted a Bayesian forward-modelling approach. The method follows the familiar approach of model fitting using Bayes' theorem with

$$P(\theta|D) \propto P(D|\theta)P(\theta), \tag{1}$$

where $P(\theta|D)$ is the posterior probability on the input parameters θ given the data (D), $P(D|\theta)$ is the likelihood function (L), and $P(\theta)$ is the prior. Assuming the error bars are Gaussian and independent, the logarithm of the likelihood function is given by

$$\ln(L) = -\frac{1}{2}\sum_i \left[\frac{(f_i - f(\theta)_i)^2}{\sigma_i^2} + \ln\left(2\pi\sigma_i^2\right)\right], \tag{2}$$

where f is the observed flux, $f(\theta)$ is the model flux for a given set of parameters θ, and σ is the error on the observed flux. The summation is over all wavelength pixels included in the fit. For our model, we adopted four free parameters: the logarithm of the stellar metallicity ($\log(Z_*/Z_\odot)$ and three parameters used to fit the overall continuum shape, based on a physically motivated parametrization of the dust attenuation law.

The dust law parametrization, taken from Salim, Boquien & Lee (2018), is a modification of the Calzetti et al. (2000) attenuation law for starburst galaxies that allows for the slope of the curve to be modified and includes a prescription for the UV bump at 2175 Å. It is given by

$$A_{\lambda,\mathrm{mod}} = \frac{A_V}{R_{V,\mathrm{mod}}}\left[k_{\lambda,\mathrm{Calz}}\frac{R_{V,\mathrm{mod}}}{R_{V,\mathrm{Calz}}}\left(\frac{\lambda}{5500\text{Å}}\right)^{\delta} + D_\lambda\right], \tag{3}$$

where A_V is the absolute V-band attenuation (i.e. the normalization of the attenuation law), $k_{\lambda,\mathrm{Calz}}$ is the total-to-selective attenuation curve for the Calzetti et al. (2000) law and δ is the power-law exponent used to modify the slope of the Calzetti et al. (2000) law. $R_{V,\mathrm{mod}}$ is the modified total-to-selective attenuation ratio that is simply a function of δ (see equation 4 in Salim et al. 2018) and $R_{V,\mathrm{Calz}} = 4.05$. Finally, $D(\lambda)$ is the Drude profile that is commonly used as the functional form of the UV bump at 2175 Å and is given by

$$D_\lambda(B) = \frac{B\lambda^2\Delta^2}{\left[\lambda^2 - \lambda_c^2\right]^2 - \lambda^2\Delta^2}, \tag{4}$$

where λ_c is the central wavelength of the feature (2175 Å), Δ is the width (350 Å) that are both held fixed, and B is the amplitude, which we allowed to vary.[3] This dust attenuation prescription therefore has three free parameters: A_V (normalization), δ (slope), and B (UV bump strength). When $\delta = 0$ and $B = 0$, the parametrization simply becomes the Calzetti et al. (2000) law. We note that the role of the three dust parameters is to fit the shape of the continuum, which, as discussed above, has little dependence on the stellar metallicity (Fig. 3).

To perform the fitting, we used the nested sampling code MULTINEST (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009),[4] which is an implementation of the nested sampling algorithm described in Skilling (2006). Nested sampling is an alternative to the traditional Markov Chain Monte Carlo method of sampling the posterior distribution for a given Bayesian inference problem, which enables the extraction of 1D posterior distributions for any given parameter in the model by marginalizing over all other free parameters. We adopted simple flat priors on all the model parameters. The prior in $\log(Z_*/Z_\odot)$ is imposed by the SB99 models to be $-1.15 < \log(Z_*/Z_\odot) < 1.45$. As the models are only provided at five fixed metallicity values, we linearly interpolated the logarithmic flux values between the models to generate a model at any metallicity within the prescribed range. We experimented with other interpolation schemes (e.g. interpolating in flux linear flux values) but found that this did not strongly affect our results, and performed simple tests (e.g. recovering the known metallicity of the SB99 templates) to ensure the interpolation scheme was not strongly biasing the recovered metallicities. For A_V and B, we considered values in the range 0–5 and for δ we considered $-1.0 < \delta < 1.0$.

A final point to note is that not all wavelength pixels were included in the fitting. As discussed above (and illustrated in Fig. 2), a significant portion of the rest-frame FUV spectra is contaminated by features not related to stellar emission (e.g. interstellar absorption lines and nebular emission lines). We excluded these wavelength regions, using only the spectral windows sensitive to photospheric absorption and stellar-wind features outlined in table 3 of Steidel et al. (2016). Finally, we also only considered wavelengths redward of Ly α at 1216 Å to avoid the additional uncertainties related to the IGM H I absorption and dust correction at shorter wavelengths. An example of a fit to one spectrum (VANDELS-m5) is shown in Fig. 4. We list the best-fitting $\log(Z_*)$ values and errors in the last column of Table 1. These quoted errors are the statistical uncertainties only, a discussion of the various systematic uncertainties, which we find to be at the $\simeq$10 per cent level, is given in the appendix.

3.4 Accuracy of the estimate and systematic uncertainties

Before discussing the stellar metallicities derived from the VANDELS data, it is worth considering how reliable we expect our method to be. We performed two tests to explore any potential systematic uncertainties related to our adoption of a simplified star formation and chemical-abundance history. First, we used data from the FIBY and SIMBA simulations to construct synthetic FUV spectra based on realistic star formation and chemical-abundance histories, and compared our recovered $\log(Z_*/Z_\odot)$ values to the true FUV-weighted stellar metallicities. Secondly, we derived stellar metallicities from the FUV spectra of local star-forming regions and starburst galaxies, which we compared to the published gas-phase metallicities of the same objects. The results of both of these independent tests are discussed next.

3.4.1 Synthetic spectra based on simulation data

In previous analyses of the FUV spectra of star-forming galaxies at high redshift, it has been common to assume models with constant SFHs (Rix et al. 2004; Sommariva et al. 2012; Steidel et al. 2016;

[3] Although the peak of the UV bump at 2175 Å is outside the range of our fitting, it can still potentially affect the shape of the UV continuum due to its width (350 Å).

[4] We accessed MULTINEST via the python interface PYMULTINEST (Buchner et al. 2014).

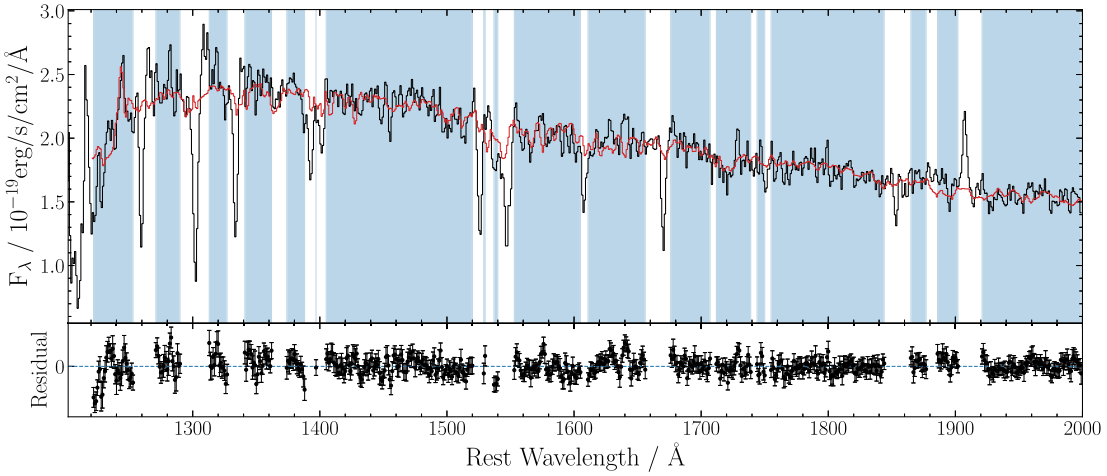

Figure 4. An example of a full spectral fit to the VANDELS-m5 spectrum (see Table 1 for details). The upper panel shows the observed composite spectrum in black with the wavelength regions used in the fitting shaded in blue. The best-fitting Starburst99 WM-Basic spectrum is overplotted in red. The lower panel shows the fit residuals; the reduced chi-squared value for this particular fit is $\chi^2_r = 1.03$.

Strom et al. 2018). Generally, a star formation time-scale of 100 Myr is assumed. It is further assumed that the metallicity of the stars does not vary strongly across the SFH. These assumptions make modelling the FUV spectra of high-redshift galaxies relatively simple compared to longer wavelength regimes and/or lower redshift galaxies, where the time-scales involved are longer, so that more complex star formation and chemical evolution histories must be considered (e.g. Carnall et al. 2018a; Leja et al. 2019).

Nevertheless, it is worth testing the validity of these assumptions against predictions from simulations. To do this, we have used the FIBY and SIMBA simulation data described in Section 2.3. Because our composite spectra are averages across a population, we focused on the average star formation and chemical abundances histories of galaxies in the FIBY and SIMBA simulations. We considered two stellar mass bins that encompass the observed data, a low-mass bin with $8.5 < \log(M_*/\mathrm{M}_\odot) < 9.5$ and a high-mass bin with $\log(M_*/\mathrm{M}_\odot) > 9.5$. As a result of their different volumes, the maximum mass is different in the two simulations ($10^{10.5}\mathrm{M}_\odot$ in FIBY and $10^{11.2}\mathrm{M}_\odot$ in SIMBA). The average star formation and chemical evolution histories for these simulated galaxies are shown in Fig. 5.

In accordance with common assumptions, it can be seen that, although the average SFRs are smoothly rising over the full formation history, all are relatively constant within the past 100 Myr. Even in the case of the most extreme variation (the high-mass bin of SIMBA galaxies) the 100 Myr averaged SFR is $29 \pm 4\,\mathrm{M}_\odot\mathrm{yr}^{-1}$ (ranging from 23–35 $\mathrm{M}_\odot\mathrm{yr}^{-1}$). In fact, for the SFRs, one could argue that there is relatively little variation over roughly 300–500 Myr time-scales in all cases. This is in good agreement with previous analyses of the average SFHs of simulated galaxies at similar redshifts (Finlator, Oppenheimer & Davé 2011).

An invariance on time-scales of $\simeq$100 Myr is also true for the chemical abundances, which overall follow a very similar trend to the SFRs. There is, in general, a slightly steeper rise over the longer time-scales, although even in the most extreme case (again the high-mass bin of SIMBA galaxies) the increase is only a factor $\simeq$2 over 500 Myr. Interestingly, despite the exact details of the global histories varying quite significantly, all share a common theme: rapid evolution over the first Gyr (i.e. from formation up to $z \simeq 6$) followed by a more gradual evolution over the next 500 Myr (i.e. $z = 4-6$). Based on the simulation data, it would seem that assuming a simple constant SFR and chemical abundance model over 100 Myr time-scales is a reasonable approximation for the composite galaxy spectra in our sample, since, across 100 Myr, *both* the average SFRs and chemical abundances of the galaxy population are relatively invariant in the two stellar mass bins.

However, next we will also demonstrate explicitly that our results should not be strongly affected by assuming longer time-scales. This is mainly due to the fact that, assuming constant star formation, the model FUV spectra reach an equilibrium (i.e. become time invariant) after $\simeq$50 Myr (e.g. Leitherer et al. 2010), and therefore any time-scale $\gtrsim$50 Myr should yield similar results. We also note that while these assumptions hold for galaxies on average, the same might not be true for individual galaxies. On an individual basis, the SFHs will not necessarily be smooth, and bursts of star formation will likely play a more significant role (e.g. Hopkins 2015).

As an explicit test, we constructed four synthetic FUV spectra based on the four star formation and chemical-abundance histories illustrated in Fig. 5, using the SB99 WM-Basic instantaneous burst models scaled appropriately to match the integrated stellar mass. To account for the effect of dust, we also attenuated each individual burst model according to its metallicity. Under our simple prescription, the absolute attenuation at 1500 Å is related to the metallicity via $A_{1500} = -2.24 \times \log(Z_*) - 2.16$, with a random scatter of $\sigma_{A_{1500}} = 0.5$ magnitudes (in this case Z_* is the total metallicity not the iron abundance). This scaling relation is motivated by results from the FiBY simulation (Khochfar et al., in preparation), and implies $\simeq$2 magnitudes of FUV attenuation for stars formed within solar metallicity environments. We note that, despite its simplicity, the relation yields global (i.e. galaxy averaged) A_{1500} values in reasonable agreement with the attenuation versus stellar mass relation presented in McLure et al. (2018a). Finally, for simplicity, we used a Calzetti attenuation curve to convert A_{1500} into A_λ across the full FUV spectral range. The synthetic spectra were smoothed to the resolution of the VANDELS data and Gaussian noise was added assuming an S/N per pixel of 15 (comparable to the observed stacks).

One unfortunate aspect of this procedure is that, for a portion of all the SFHs, the metallicity is below the minimum imposed by the SB99 models ($\mathrm{Z}_*/\mathrm{Z}_\odot = 0.07$; $\log(\mathrm{Z}_*/\mathrm{Z}_\odot) = -1.15$). In these cases, we are forced to simply use the lowest metallicity SB99 spectrum.

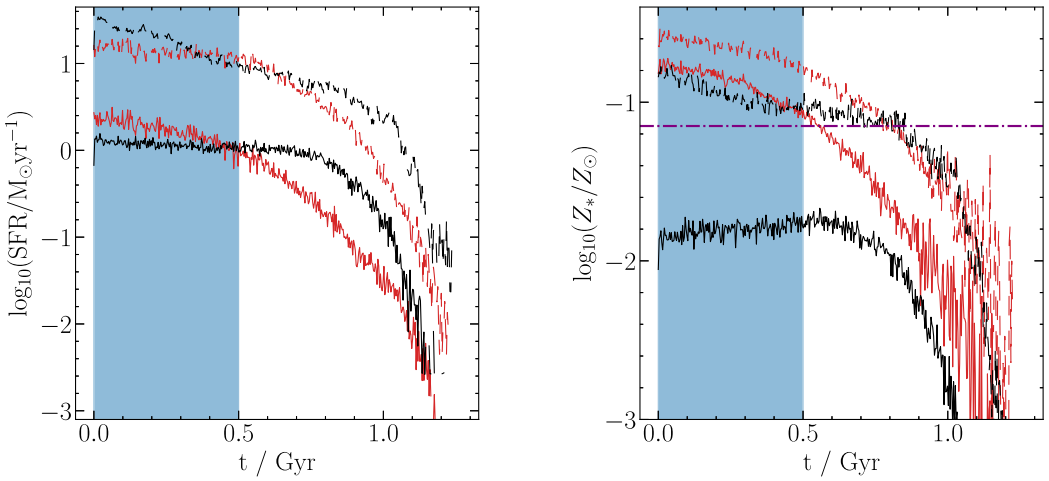

Figure 5. The star formation history (left-hand panel) and chemical-abundance history (in our case iron abundance; right-hand panel) of galaxies from the FIBY (the red curves) and SIMBA (the black curves) simulations at $z = 4$. The values on the x-axis are lookback times relative to $z = 4$. In each panel, the solid curves show the average histories for galaxies with $8.5 < \log(M_*/\mathrm{M}_\odot) < 9.5$ and the dashed curves show galaxies with $\log(M_*/\mathrm{M}_\odot) > 9.5$, roughly representative of the range of stellar masses in our sample. Despite the obvious differences, all curves show a rapid increase in star formation rate and metallicity over the first $\sim$ Gyr followed by a more gradual evolution over the final $\sim$500 Myr (the blue shaded region). The purple dot–dashed line in the right-hand panel indicates the lower metallicity limit of the SB99 WM-Basic models ($Z_*/Z_\odot = 0.07$).

This is particularly a problem for the low metallicities predicted in the low-mass bin of the SIMBA simulation (the black solid curve in the right-hand panel of Fig. 5). For the other three cases, the issue is somewhat mitigated since the FUV light is dominated by stars with metallicities within the SB99 parameter space. Nevertheless, we note that this effect could still introduce some small bias into the recovered $\log(Z_*/Z_\odot)$ values.

The results of the test are shown in Fig. 6, and generally confirm the conclusions drawn from Fig. 5. For the three cases in which the FUV-weighted metallicity is within the model parameter space the true value is well recovered, with derived (true) $\log(Z_*/Z_\odot)$ values of -0.78 ± 0.06 (0.78), -0.63 ± 0.07 (0.62), and -0.80 ± 0.06 (0.83). As mentioned above, the low-mass galaxies in the SIMBA simulation have an average FUV-weighted metallicity of $\log(Z_*/Z_\odot) = -1.88$, below the lower limit of the models. In this case, obviously somewhat by design, the solution bumps up against the edge of the parameter space. For a similar case within the observed sample, we would consider this an upper limit on $\log(Z_*/Z_\odot)$. It can also be seen that adopting constant SFR models with longer time-scales (300 and 500 Myr) does not have a strong systematic effect on the recovered metallicity, although, in general, the 100 Myr models perform best.

Finally, we note that, although it is encouraging, this test cannot provide a definitive justification of our method since we are building and fitting the FUV spectra with, fundamentally, the same set of SB99 models. However, it does serve to demonstrate that our simplified assumptions regarding the star formation and chemical-abundance histories should not significantly bias the recovered metallicity values.

3.4.2 The FOS–GHRS local Universe sample

Given that the test described above, using synthetic spectra, is not truly independent, it is desirable to perform an independent test using real data. To this end, we have used the above method to fit both the individual and composite spectra from the local FOS–GHRS galaxy sample described in Section 2.2. The idea here is to

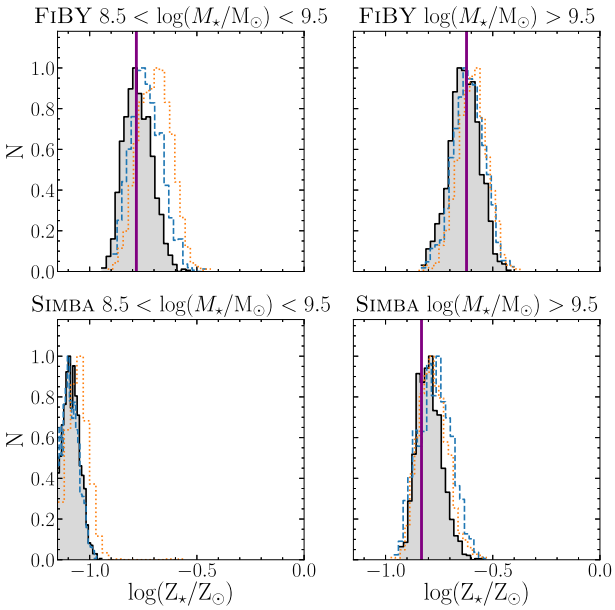

Figure 6. The 1D posterior distribution for the UV luminosity weighted $\log(Z_*/Z_\odot)$ recovered for four simulated galaxies. The simulated galaxies were built using the four average star formation and chemical-abundance histories shown in Fig. 5 as described in the text. The simulation (FIBY/SIMBA) and stellar mass regime corresponding to each panel are given in the title. The grey filled histograms are the 1D posterior distributions obtained using our fiducial Starburst99 models assuming 100 Myr of constant star formation. The blue dashed and orange dotted curves are constant star formation models with longer time-scales of 300 and 500 Myr. The purple vertical lines show the true UV-weighted metallicity of each simulated galaxy. For the lowest mass galaxies in SIMBA (lower left-hand panel), the true UV-weighted metallicity is below the lower limit enforced by the Starburst99 models ($\log(Z_*/Z_\odot) = -1.15$) and therefore not shown. In this case, the fitted metallicity is, as expected, only an upper limit.

compare the measured UV stellar metallicities to the published gas-phase metallicities of the same galaxies, under the assumption that the two parameters should correlate. This should be a reasonable assumption since the nebular emission lines used to determine the gas-phase metallicity are expected to originate from the hot, ionized, gas surrounding the same young, massive O- and B-type stars that dominate the FUV stellar emission. Indeed, such a correlation is commonly observed in the H II regions of Local Group galaxies, and is consistent with a 1:1 relation (e.g. Toribio San Cipriano et al. 2017). Moreover, for the integrated emission of local galaxies (out to $z \simeq 0.3$), a 1:1 correlation is observed for the young (<2 Gyr old) stellar population (i.e. the population traced by rest-frame FUV observations; González Delgado et al. 2014).

As discussed in Section 2, we produced three composite FOS–GHRS spectra in bins of absolute K-band magnitude (M_K, a proxy for the stellar mass) as outlined in Table A1. Spectral fits to the M_K composite spectra are shown in Fig. 7. We find that the stellar metallicities of the two brightest composites ($M_K = -23.1$ and -25.1; or $\log(M_*/\mathrm{M}_\odot) = 10.33$ and 11.13) are similar with $\log(Z_*/Z_\odot) \simeq 0.2$ (or $Z/Z_\odot \simeq 1.6$) in excellent agreement with their median gas-phase metallicities. For the faint, low-mass, composite with $M_K = -18.5$ ($\log(M_*/\mathrm{M}_\odot) = 8.5$), we find a stellar metallicity of $\log(Z_*/Z_\odot) \simeq -0.58$ (or $Z/Z_\odot \simeq 0.26$), roughly a factor 6 lower) and, again, the stellar metallicity is in good agreement with the median gas-phase metallicity. We note that the fits to the low-mass and high-mass composites, while generally good, do not appear to match the observed spectra in the range $\simeq$1200–1400 Å despite having statistically acceptable reduced χ-squared values ($\chi^2_r \simeq 0.7$). These fits would likely be improved with a more complex dust model, as the current discrepancy could plausibly be a result of strong variation in the dust law at shorter wavelengths due to the relatively low number of objects in each composite. However, as the current fits are statistically acceptable, we did not consider a more complex dust model in this analysis.

We find a similar result for the individual galaxies, which are fitted in the same way (although often using only a restricted portion of the full wavelength range), albeit with a much larger scatter. In fact, the results in Fig. 7 are consistent with the stellar and gas-phase metallicities scattering about a 1:1 relation. The Spearman rank correlation coefficient between the two quantities is 0.70, with the probability of obtaining this value by chance formally zero ($p = 3.1 \times 10^{-4}$). We note that although a correlation between stellar and gas-phase abundances is to be expected, a direct 1:1 correspondence is only expected if the O/Fe ratios (i.e. chemical abundance patterns) in these galaxies are close to the solar value; we will return to this issue in the case of high-redshift galaxies in Section 5.3.

In summary, we have performed two independent tests of our adopted method. First, we have shown that the assumption of a constant SFH with a time-scale of 100 Myr, at a single metallicity, is sufficient to recover the FUV-weighted stellar metallicities of synthetic galaxy spectra built from more realistic star formation and chemical-abundance histories. Furthermore, by applying our method to composite FUV spectra of local galaxies we have shown that we are capable of reproducing the expected trend between stellar and gas-phase abundances. One further potential source of bias that we do not discuss in detail here, relates to the effect of redshift uncertainties on the metallicity-sensitive continuum features when constructing composite spectra. However, we do not expect this effect to significantly affect our derived metallicities, and a brief discussion is provided in Appendix B.

4 THE STELLAR METALLICITIES OF STAR-FORMING GALAXIES AT 2.5 < z < 5.0

In this section, we discuss stellar mass–metallicity relationship at $2.5 < z < 5.0$ derived from the VANDELS composite spectra. Before discussing the observational results, however, it is important to consider that fact that the stellar metallicities derived using our method are FUV weighted, and it is therefore worth briefly discussing how FUV-based metallicities (tracing the recent SFH) might be biased with respect to the mass-weighted metallicities (which trace the global SFH) at these redshifts.

4.1 FUV-weighted versus stellar-mass-weighted metallicities

From the FIBY and SIMBA simulation data, we were able to derive both the mass-weighted and FUV-weighted stellar metallicities for each simulated galaxy and we present the results of the comparison in Fig. 8, which shows the median mass-weighted and FUV-weighted stellar mass–metallicity relations for both simulations. The mass-weighted metallicities of each simulated galaxy were derived simply using the mass and metallicity information of each star particle. For the FUV-weighted metallicities, we used the FUV luminosity of each star particle at 1500 Å taken from the SB99 instantaneous burst models. In both cases, the FUV-weighted relation is offset to lower metallicity, although crucially the shape of the relation is unaffected. The offset is due to the fact that the youngest stars, which dominate the FUV flux, are generally more metal enriched and therefore according to our attenuation prescription, are more heavily attenuated and do not contribute as much to the FUV. For example, it can be seen from Fig. 8 that the intrinsic FUV-weighted relation (i.e. assuming no dust attenuation) for the SIMBA galaxies (the red dotted line), closely follows the mass-weighted relation. Therefore, any preferential attenuation of the youngest, most metal-enriched, stars will result in a bias to lower metallicities when using FUV spectra. There is clearly some systematic uncertainty here related to our method of attenuating the individual star-particle spectra, however, based on reasonable dust assumptions, we expect the FUV-weighted metallicities of high-redshift star-forming galaxies to be biased low with respect to the mass-weighted values by $\simeq$0.05−0.1 dex, but predict that the shape of the mass–metallicity relation to be similar in either case. Finally, we note that throughout this section we will also compare our results to the relations extracted from the FIRE simulation presented in Ma et al. (2016).[5] Since these FIRE relations are for mass-weighted metallicities, we have scaled them down by 0.05 dex.

4.2 The 2.5 < z < 5.0 stellar mass–metallicity relation

The $2.5 < z < 5.0$ stellar mass–metallicity relation derived from the VANDELS sample is shown in Fig. 9. The vertical error bars show the 16th and 84th percentiles of the posterior distribution on $\log(Z_*/Z_\odot)$ and the horizontal error bars show the range of stellar mass in each bin. The first thing to note is that a clear monotonic increase in $\log(Z_*/Z_\odot)$ with $\log(M_*/\mathrm{M}_\odot)$ is observed for our sample. Focusing on the purely stellar mass stacks, $\log(Z_*/Z_\odot)$ evolves from a lower limit of <-1.04 (68 per cent) at $\log(M_*/\mathrm{M}_\odot) \simeq 8.5$ to -0.57 ± 0.06 at $\log(M_*/\mathrm{M}_\odot) \simeq 10.2$, an increase by a factor $\geq$3 in the average metallicity across roughly two decades of stellar mass. In Fig. 10, we show how this metallicity evolution affects the form of

[5] Ma et al. (2016) assume a solar metallicity of 0.02 and therefore we have converted their relations to our assumed value of 0.0142.

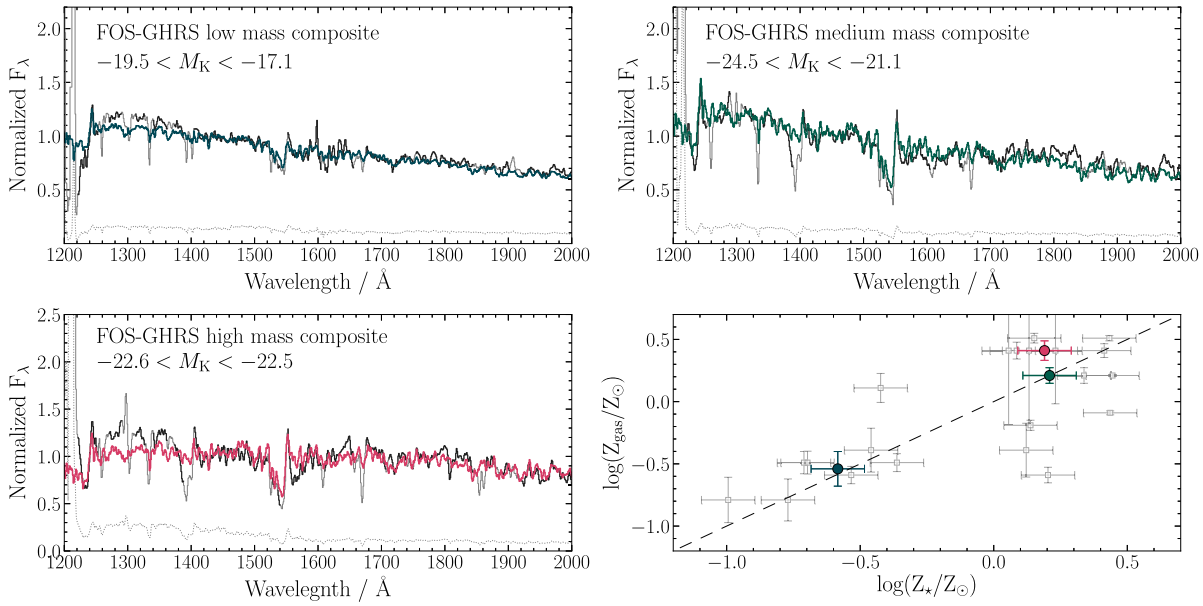

Figure 7. Composite FOS–GHRS spectra in bins of absolute *K*-band magnitude (M_K) with the best-fitting Starburst99 stellar population synthesis models overplotted in colour (top two panels and lower left-hand panel). The lower right-hand panel shows the relationship between our derived stellar metallicities and published gas-phase metallicities for the FOS–GHRS galaxies, both for the individual galaxies (the grey points) and the composites (the coloured points), illustrating the expected correlation between these quantities (the Spearman rank correlation coefficient is 0.70).

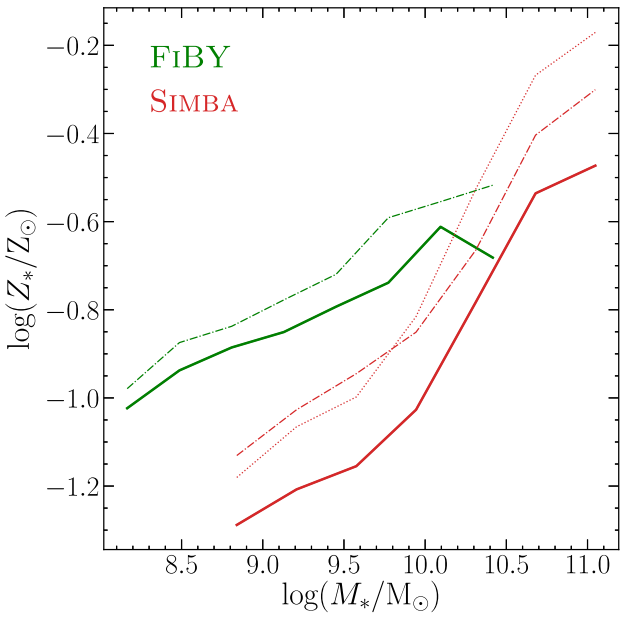

Figure 8. The mass-weighted and FUV-weighted stellar mass–metallicity relations derived from the FIBY (green) and SIMBA (red) simulations at $z = 4$. In each case, the dot–dashed line shows the mass-weighted relation and the solid line shows the FUV-weighted relation, which is derived as described in the text (including a correction for dust). The FUV metallicities are generally offset to lower values by $\sim$0.05$-$0.1 dex due to the preferential extinction of the youngest, most metal-enriched stars in the galaxy. For reference, the dotted red line shows the intrinsic FUV-weighted relation (i.e. assuming no dust) for the SIMBA galaxies.

the FUV spectrum by showing the best-fitting SB99 models overlaid on the composite spectra (the effect is subtle, but as the average mass of the composite spectra increases, the continuum becomes less smooth and the undulations resulting from metal line blanketing become more pronounced). At all masses, the stellar metallicities (to first order a proxy for the iron abundance) are significantly subsolar ($\lesssim$25 per cent). From an analysis of the redshift and mass tacks, we do not find strong evidence for a significant trend with redshift at a given stellar mass in the VANDELS sample. It can be seen from Fig. 9 that the redshift-mass stacks are consistent with scattering around the stellar-mass-only relation.

With respect to the normalization of the stellar mass–metallicity relation, our results are consistent with recent results from the independent KBSS-MOSFIRE survey (Steidel et al. 2016; Strom et al. 2018). The orange data point in Fig. 9 shows the result derived from a stack of 30 star-forming galaxies at $z = 2.40 \pm 0.11$ presented in Steidel et al. (2016) who estimated $Z_*/Z_\odot \simeq 0.14$ using a similar method to our own, by fitting SB99 WM-Basic models to the composite FUV spectrum, with the additional requirement that the best-fitting stellar model should also account for the observed nebular optical emission lines (a constraint we currently do not have for our sample). Steidel et al. (2016) find a lower best-fitting metallicity model using the alternative BPASSv2 stellar models, which is consistent with what we report in Appendix C. Despite the slight difference in method, average redshift, and the stellar models employed, the consistency is encouraging, and provides further evidence that the stellar metallicities (or iron abundances) of galaxies at high redshift are significantly subsolar ($Z_*/Z_\odot \lesssim 0.25$).

4.3 Redshift evolution of the stellar mass–metallicity relation

Extending the comparison to lower redshifts, we show the evolution of the mass–metallicity relation between $2.3 < z < 5.0$ ($\langle z \rangle = 3.5$) and $z = 0$ in Fig. 11. For our main local Universe comparison sample, we have used a recent determination of the stellar mass–metallicity relation for $\sim$200 000 star-forming galaxies in the SDSS by Zahid et al. (2017; Z17). The solid line (connecting the triangular data points) in Fig. 11 shows the $z = 0$ relation and the dashed line shows an arbitrary downward shift of this relation by 0.6 dex. This

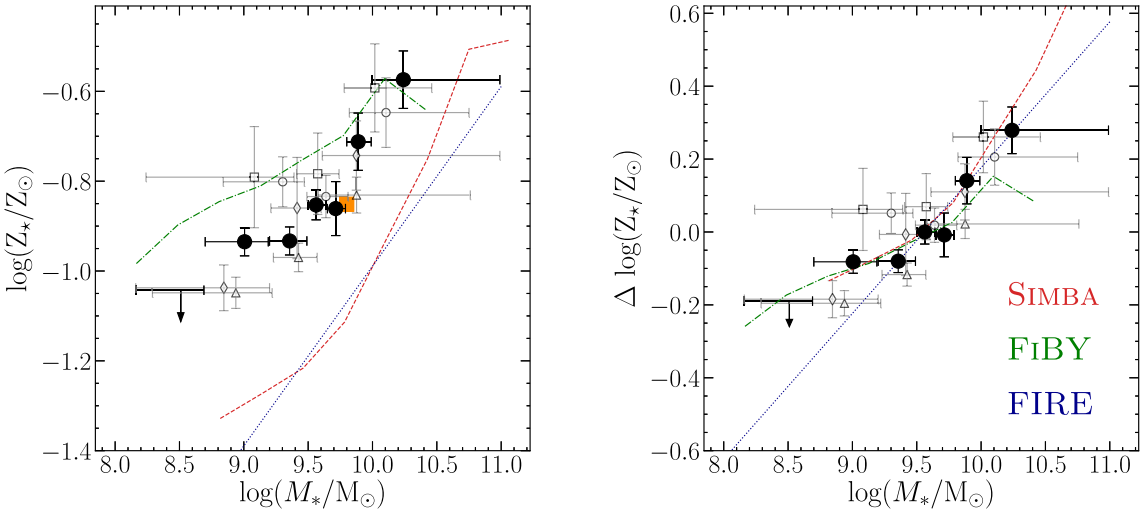

Figure 9. The stellar mass–metallicity relationship for the VANDELS galaxies at $2.3 < z < 5.0$ compared to predictions from hydrodynamical simulations. The left-hand panel shows the absolute values of $\log(Z_*/Z_\odot)$ plotted against $\log(M_*/M_\odot)$ with the VANDELS data shown as the filled black circles (the stellar-mass-only composites) and the open grey symbols (the composites binned in redshift and stellar mass). The redshift ranges corresponding to the open symbols are as follows: $2.3 \leq z < 2.65$ (the circles), $2.65 \leq z < 3.15$ (the squares), $3.15 \leq z < 3.80$ (the triangles), and $3.80 \leq z < 5.0c$ (the diamonds). The vertical error bars show the 16th and 84th percentiles of the posterior distribution on $\log(Z_*/Z_\odot)$ and the horizontal error bars show the range of stellar mass in each bin. The orange square data point is a measurement of $\log(Z_*/Z_\odot)$ for a sample of 30 star-forming galaxies at $z = 2.4$ taken from Steidel et al. (2016). The various lines are predictions for stellar metallicity (weighted by the 1500 Å luminosity) versus stellar mass from three cosmological simulations: FIBY (green, dot–dashed), SIMBA (red, dashed), and FIRE (blue, dotted). The absolute values in all three simulations are inconsistent with the observed data. However, the shape of the relation is well recovered by the simulations as illustrated in the right-hand panel where, instead of absolute $\log(Z_*/Z_\odot)$, the values relative to $\log(M_*/M_\odot) \simeq 9.56$ are shown.

simple shift matches the VANDELS data remarkably well, implying an increase in metallicity of a factor ∼4 from $z \sim 3.5$ to the present day without any obvious dependence on stellar mass.

However, it is important to note that there will be some biases associated with this simple comparison. For example, whereas our metallicities are determined from the rest-frame FUV, tracing the massive OB stellar population, and are sensitive to the iron abundance, Z17 derive metallicities by fitting to rest-frame optical spectra in the wavelength range ≃4000–7000 Å. It is not immediately obvious that the methods trace either the same stellar population or abundance type. We can partially address this issue via a comparison with the metallicities derived from the composite FOS–GHRS spectra (the blue data points in Fig. 11). In this case, we know that the method, at least, is not biasing the comparison since the metallicities have been derived in identical ways. Although the mass sampling is not ideal, from the three stellar-mass bins available we find the derived stellar metallicities are in approximate agreement with the $z = 0$ relation, especially when taking into account that, when measured for individual galaxies, the scatter in the local stellar mass–metallicity relation is of the order $\gtrsim$0.1 dex (Panter et al. 2008). There are again a number of caveats, primarily the fact that, for the high-mass galaxies, the FOS–GHRS observations are often taken for individual H II regions within the galaxy rather than the galaxy globally. However, one can make the argument that these regions are likely to dominate the global FUV emission of these galaxies. Taken as a whole, the data presented in Fig. 11 provide strong evidence for an evolution in stellar metallicities of roughly a factor 4 between $2.3 < z < 5.0$ and $z = 0$.

Another potential complication worth acknowledging is the fact that, although comparing the stellar metallicities of star-forming galaxies at these redshifts is clearly of interest, it is likely that the two data sets do not form an evolutionary sequence since, depending on their stellar mass, star-forming galaxies at high redshift will likely be the progenitors of present-day quiescent galaxies (e.g. Carnall et al. 2018b). Nevertheless, the stellar metallicities of early-type galaxies with $M_* > 10^{10}\,M_\odot$ estimated from rest-frame optical absorption features also fall predominantly within the range $-0.2 < \log(Z_*/Z_\odot) < 0.2$ (Gallazzi et al. 2006), similar to the values for the star-forming population, and therefore a similar conclusion would be drawn from this comparison.

In summary, we find clear evidence for a monotonic increase in stellar metallicity with galaxy stellar mass across the redshift range $2.3 < z < 5.0$. Within this narrow redshift window, we find no strong evidence for redshift evolution at fixed stellar mass. However, a direct comparison to the stellar metallicities of star-forming galaxies in the local Universe implies that metallicities increase by a factor of 4, independent of mass, between $z \sim 3.5$ and $z = 0$.

5 DISCUSSION

Here, we discuss some of the implications of our results. First, we compare the observed data to the predictions of three cosmological simulations. Then, using a simple one-zone chemical evolution model, we attempt to identify the key physical parameters driving both the significantly subsolar stellar metallicities, as well as the observed mass dependence in our sample. Finally, we compare the stellar metallicities derived here with published gas-phase metallicities at similar redshifts in order to investigate suggestions of enhanced O/Fe ratios in star-forming galaxies at $z \gtrsim 2.5$.

5.1 Comparison to cosmological simulations

The left-hand panel of Fig. 9 compares the absolute $\log(Z_*/Z_\odot)$ values to simulation predictions for the FUV-weighted stellar mass–metallicity relation. In general, the absolute metallicity values predicted in the simulations are either systematically too large

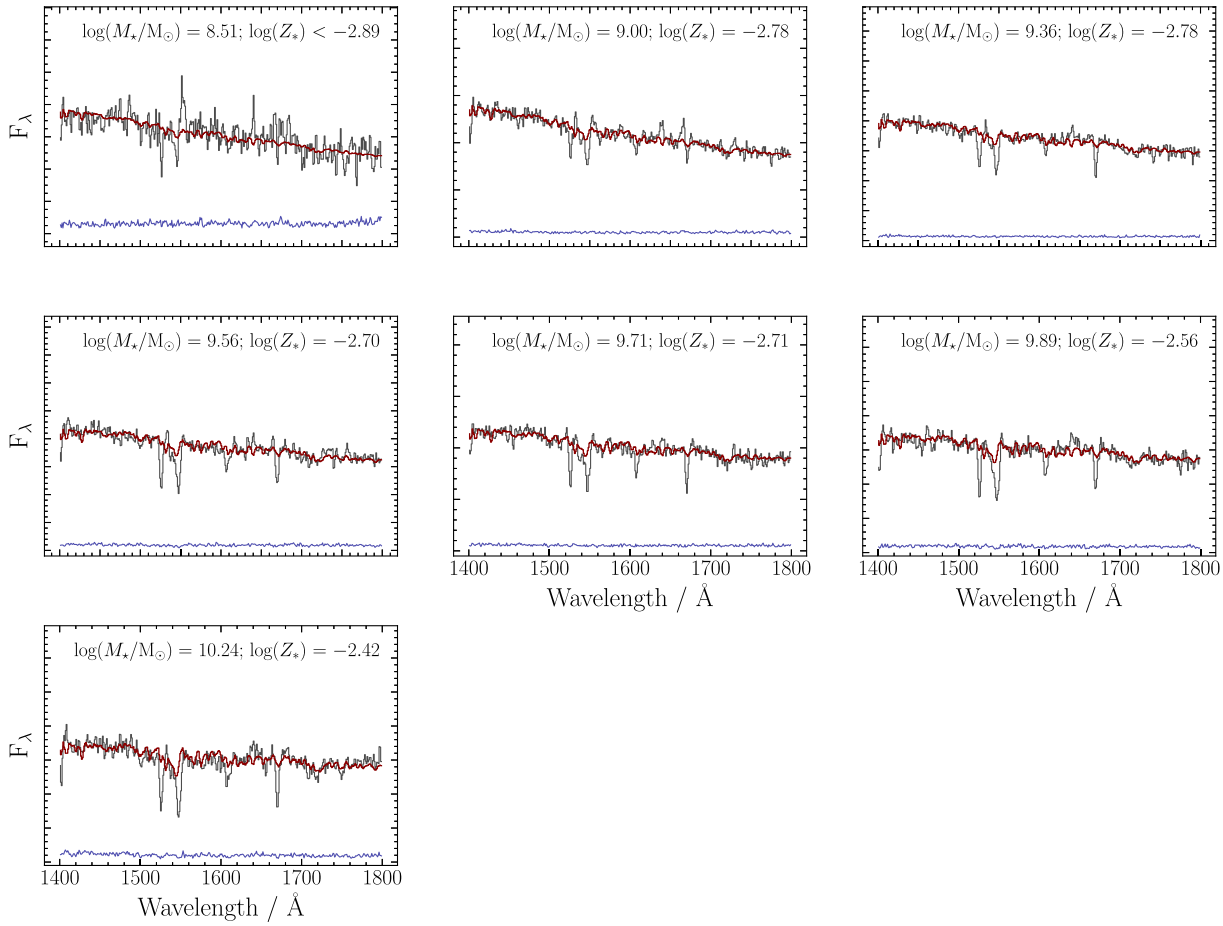

Figure 10. Fits to the VANDELS stellar mass composite spectra at $2.5 < z < 5.0$. Each panel shows one of the seven composites with the average stellar mass and best-fitting metallicity indicated in the legend. The black curves show the observed spectra with the error spectra shown in blue. The best-fitting Starburst99 WM-Basic spectra are overplotted in red. The figure focuses on the wavelength region $1400 < \lambda < 1800$ to illustrate how the increasing metallicity at higher stellar masses causes more pronounced undulations in the continuum. In addition to this, the equivalent widths of metallicity-dependent stellar absorption features such as N IVλ1720 and C IVλ1548 can be seen to decrease at lower stellar masses.

(FiBY) or too small (Simba and FIRE). Given these discrepancies are generally a factor $\lesssim$2, they could be explained by uncertainties in the element yields from CCSNe, SNe Ia, and AGB star winds used in the simulations (e.g. Romano et al. 2010). Indeed, yields in the cosmological simulation Mufasa (Davé, Thompson & Hopkins 2016, the precursor to the Simba simulations) were arbitrarily reduced by a factor of 2 to match published gas-phase oxygen abundance data. The correction became unnecessary in Simba due to a more detailed model for how elements become locked in dust grains (Davé et al. 2019), however, this comparison would suggest that the iron-peak element yields may still not consistent with current data. Another explanation, illustrated in Section 5.2 next, could be related to the strength of galactic outflows in the simulations.

The shape of the relation, on the other hand, appears to be relatively well reproduced by all three simulations. The right-hand panel of Fig. 9 shows the data and simulated relations normalized at $\log(M_*/\mathrm{M}_\odot) = 9.56$ (the stellar mass of the VANDELS-m4 composite). The exact slope of the relation based on the observed data is somewhat uncertain given the upper limit at the lowest stellar mass bin, but the simulations are clearly in relatively good agreement. The Simba data, in particular, match the shape of the relation extremely well.

Determining what governs the overall shape of the stellar mass–metallicity relation in simulations is not straightforward due to the complex interplay of star formation, inflows and outflows that determine the metallicity evolution. However, as we discuss in Section 5.2 next, one of the key physical parameters governing the shape and normalization of the relation is the strength and mass scaling of galactic outflows, parametrized by the mass-loading parameter $\eta = \dot{M}_{\mathrm{outflow}}/\dot{M}_*$. The prescription for how the mass-loading parameter scales with stellar mass (the $\eta{-}M_*$ relation) in Simba is, in fact, based on the results of the high-resolution FIRE zoom simulations, which include a detailed stellar feedback model accounting for the effects of radiation pressure, supernovae, stellar winds, and photoionization and photoelectric heating (Hopkins et al. 2014; Muratov et al. 2015; Anglés-Alcázar et al. 2017). The fact that Simba and FIRE use a similar outflow model could account for the good agreement in the shape of the MZR between the two simulations. The feedback model used in FiBY is based on an earlier prescription for supernova feedback (Dalla Vecchia & Schaye 2012) that results in generally weaker stellar winds (i.e. smaller mass-loading parameters), although crucially the scaling of η with stellar mass is similar to the Simba and FIRE simulations (S. Khochfar, private communication). The similarity of the slope of the $\eta{-}M_*$ relation in each simulation could therefore account

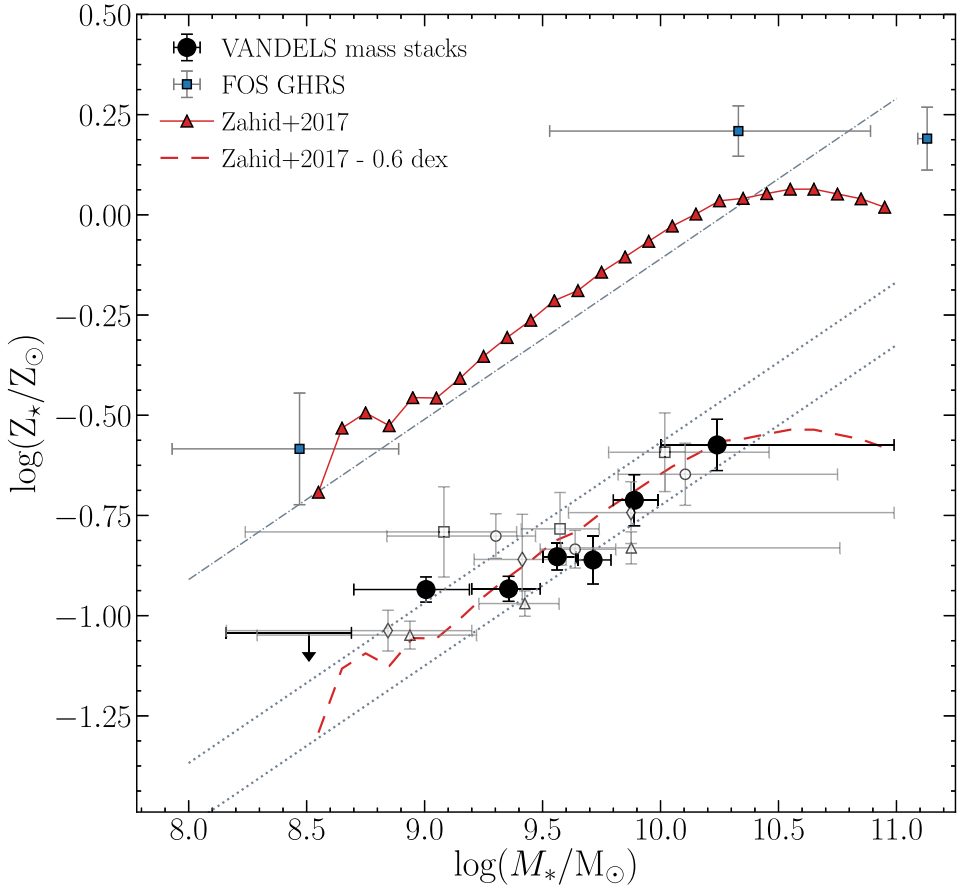

Figure 11. Redshift evolution of the stellar mass versus stellar metallicity relationship. The red triangular data points show the $z = 0$ relation derived from fitting stacked optical continuum spectra of ∼200 000 star-forming galaxies in the Sloan Digital Sky Survey (Zahid et al. 2017). The blue square data points show the FUV-based stellar metallicities for the composite FOS–GHRS spectra of local star-forming regions and starburst galaxies discussed in the text. The VANDELS data at $2.5 < z < 5.0$ are shown as the black/grey data points as in Fig. 9. The dashed red line shows the $z = 0$ relation of Zahid et al. (2017) shifted by −0.6 dex in $\log(Z_*/Z_\odot)$. Simply shifting the local relation down by a factor ≃4 produces remarkably good agreement with the high redshift data. The dotted grey lines show the FIRE prediction at $z = 2.3$ (upper) and $z = 5.0$ (lower) and the grey dot–dashed line is the FIRE prediction at $z = 0$. The FIRE metallicities have been arbitrarily scaled upwards by a factor of 2 (0.3 dex).

for the consistency in their predictions for the shape of the MZR. On the other hand, the differences in the absolute $\log(Z_*/Z_\odot)$ values are likely to be a result of both the different normalizations of the η–M_* relations, as well as differences in the assumed stellar yields.

Finally, we note that the lack of observed redshift evolution between $2.5 < z < 5.0$ is also in reasonable agreement with predictions from simulations. For example, the FIRE simulation predicts ≃0.16 dex of evolution in $\log(Z_*/Z_\odot)$ at fixed $\log(M_*/M_\odot)$ across $2.3 < z < 5.0$ (1.7 Gyr of cosmic time; Ma et al. 2016), a value roughly comparable to the 1–2σ error bars on the derived metallicities for the mass–redshift composites. This is shown by the grey dotted lines in Fig. 11, where we have scaled the FIRE metallicities upwards by a factor of 2 such that the $z = 2.3$ and $z = 5.0$ relations bracket our data. However, given the fact that we are only utilizing ∼65 per cent of the full VANDELS data set, evidence for a redshift evolution will be an area that can be explored in more detail in future work. Fig. 11 also shows the prediction for the redshift evolution of the stellar mass–metallicity relation to $z = 0$ from the FIRE simulation.[6] Again, it can be seen that the predicted evolution of the stellar MZR by ≃0.6 dex is in remarkable agreement with the observations. We also note that an evolution of this magnitude is predicted from the EAGLE simulations (De Rossi et al. 2017), although we do not show this relation in Fig. 11 for clarity. In summary, the current predictions from cosmological simulations seem to be in broad agreement with our data.

5.2 Analytic one-zone chemical evolution models

Although simple analytic models lack the rigour of detailed cosmological simulations, they can still provide useful physical insights into galactic chemical evolution. The primary advantage of using analytic approximations is, perhaps, the ability to quickly determine the sensitivity of chemical evolution to different physical parameters, a procedure that is computationally expensive in the case of detailed simulations. Such analytic models are typically specified by a gas-accretion history, a star formation law, and prescriptions for gas inflows/outflows and nucleosynthetic yields. Numerous analytic models exist (e.g. Lilly et al. 2013; Zhu et al. 2017), however, in this section, we compare our results with the analytic one-zone chemical evolution model presented in Weinberg, Andrews & Freudenburg (2017) (hereafter WAF). The WAF model is particularly suited to our purpose since, as discussed, the FUV stellar metallicities we derive are a proxy for the iron abundance,

[6]Unfortunately, we cannot produce similar predictions from the FIBY and SIMBA simulations since they terminate at redshifts $z = 4$ and $z = 1$, respectively.

and WAF incorporates a realistic delay time distribution (DTD) for Type Ia SNe, allowing the evolution of iron peak elements to be tracked separately to that of the α elements.

Briefly, the key elements of the WAF model are as follows. The SFR ($\dot{M}_*$) is assumed to be proportional to the gas mass ($M_{\rm gas}$) with the star formation efficiency (SFE $= \dot{M}_*/M_{\rm gas}$) assumed to be constant (the so-called 'linear Schmidt law'). For the purpose of this discussion, we will refer to the inverse of the SFE as the gas depletion time-scale ($t_{\rm dep} = M_{\rm gas}/\dot{M}_*$) as this is a quantity that has been extensively studied out to $z \simeq 6$ (e.g. Scoville et al. 2016). Gas outflows are assumed to be a constant multiple of the SFR, with the constant of proportionality commonly referred to as the mass-loading parameter ($\eta = \dot{M}_{\rm outflow}/\dot{M}_*$). Therefore, for a specified SFH, the gas depletion time-scale and mass-loading parameter define the total gas-accretion history. The gas accreted from the IGM is assumed to be pristine ($Z = 0$) and the outflowing gas is assumed to be at the same metallicity as the ambient star-forming ISM.[7] Three SFHs are explicitly solved for in WAF but we only consider the linear−exponential case here.[8] For our purposes, this essentially equates to a linearly rising SFH. Elements produced by CCSNe are instantaneously recycled into the star-forming ISM, whereas enrichment from Type Ia SNe is delayed as specified by the DTD. The form of the DTD is exponentially decreasing in time with a minimum delay time of 0.15 Gyr and an *e*-folding time-scale of 1.5 Gyr.[9] The models assume a Kroupa (2001) IMF and the CCSNe yields of Chieffi & Limongi (2004) and Limongi & Chieffi (2006) that predict that 1.5 $M_\odot$ of oxygen and 0.12 $M_\odot$ of iron are returned to the ISM for every 100 $M_\odot$ of star formation. The net return of iron from Type Ia SNe, assuming the fiducial model parameters, is 0.17 $M_\odot$ for every 100 $M_\odot$ of star formation. These nucleosynthetic yields are assumed to be independent of stellar metallicity. In general, yields will be metallicity dependent, however, for iron, the metallicity dependence is not expected to be strong (e.g. Andrews et al. 2017, fig. 20). A fixed fraction of stellar mass formed into stars is assumed to be recycled from the envelopes of CCSNe progenitors at its original metallicity and is referred to as the recycling parameter ($r = 0.4$). All free parameters excluding the SFR are assumed to be fixed with time.[10] Using these various approximations, it is possible to derive analytic equations for the evolution of chemical abundances. We refer interested readers to the original WAF paper for detailed derivations.

As discussed in WAF, the main governing parameters in their model are the mass-loading parameter (or outflow efficiency) η and the SFE, parametrized by the familiar gas-depletion time-scale ($t_{\rm dep} = M_{\rm gas}/\dot{M}_*$). The left-hand panel of Fig. 12 shows the predicted time evolution of $\log(Z_*/Z_\odot)$ for different values of η and $t_{\rm dep}$. Here, $\log(Z_*/Z_\odot)$ refers to the iron abundance ([Fe/H]) in the WAF models. It can clearly be seen that the assumed value of the mass-loading parameter has a pronounced effect, strongly influencing the typical $\log(Z_*/Z_\odot)$ value reached after $\simeq$2.5 Gyr of evolution (corresponding to $z \simeq 2.5$). The reason is simply that by increasing the mass-loading parameter (from $\eta = 3-10$ in our example) ISM enrichment is suppressed by the removal of more enriched gas from the galaxy. In contrast, it is difficult to ascribe the low absolute abundances to a combination of relatively weak stellar winds ($\eta = 3$) with long depletion time-scales (i.e. low SFE, which will lead to lower metallicities simply due to the fact that fewer metals are being formed in stars). As can be seen from Fig. 12, even depletion time-scales of 2 Gyr are not sufficient to reach agreement with the data. In reality, time-scales of $\simeq$2 Gyr are more typical of local star-forming galaxies (e.g. Leroy et al. 2013), and it has been found from observations of cold gas in high-redshift galaxies that the depletion time-scale appears to decrease with increasing redshift, with typical values in the range 200–700 Myr across $z = 1-6$ (e.g. Tacconi et al. 2013; Scoville et al. 2016, 2017). Therefore, within the context of the WAF model, the mass of metals produced assuming realistic values of the depletion time-scale is too large to be consistent with the observations unless relatively large mass-loading parameters (i.e. high outflow efficiencies, $\eta \simeq 10$ on average across our sample) are assumed. We note that the importance of strong outflows in shaping both the stellar and gas-phase metallicity is also found in many semi-analytic galaxy evolution models (e.g Hirschmann, De Lucia & Fontanot 2016; Lian et al. 2018a).

Further insights can be gained by investigating the mass dependence of the stellar metallicity. However, since the WAF model is only parametrized in terms of η, to simulate mass dependence a mapping between η and mass is required. Motivated by the outflow scaling relations from the FIRE simulations presented in Muratov et al. (2015),[11] we assume η is related to stellar mass via

$$\eta = \alpha \left(\frac{M_*}{10^{10} {\rm M}_\odot} \right)^{\beta}, \tag{5}$$

where M_* is the galaxy stellar mass, α is the mass-loading parameter at $M_* = 10^{10} {\rm M}_\odot$, and β is the power-law exponent. Muratov et al. (2015) find a redshift-independent relation with $\alpha = 3.55$ and $\beta = -0.351$. Equation (5) can be easily incorporated into the WAF model in order evaluate the values of α and β (and hence the $\eta(M_*)$ relation) most consistent with our data.

The right-hand panel of Fig. 12 shows the difference in metallicity ($\Delta\log(Z_*/Z_\odot)$ as a function of stellar mass for different parametrizations of the stellar mass versus η relationship assuming both the Muratov et al. (2015) normalization ($\alpha = 3.6$) and a slightly higher normalization implied by the WAF models (a value of $\alpha \simeq 6$ is consistent with the highest mass – i.e. highest metallicity – data point in the left-hand panel of Fig. 12). The $\Delta\log(Z_*/Z_\odot)$ values are calculated with respect to $\log(M_*/{\rm M}_\odot) \simeq 9.56$. The data favour an $\eta - M_*$ scaling relation with a similar power-law slope to the Muratov et al. (2015) relation, irrespective of the absolute normalization. A simple χ^2 analysis returns a best-fitting power-law slope of $\beta = (-0.45, -0.40)$ for the $\alpha = (3.6, 6.0)$ cases, respectively. In either case, the mass-loading parameter decreases by roughly a factor of 5 across the stellar mass range $10^8 < M_*/{\rm M}_\odot < 10^{10}$ with a median value of $\eta = 8$ for $\alpha = 3.6$ and $\eta = 15$ for $\alpha = 6.0$. We note there is some degeneracy here between α and the

[7] We note that this is one important simplifying assumption of the model since there is strong evidence that outflows are in fact metal enriched with respect to the average ISM metallicity (e.g. Peeples & Shankar 2011; Chisholm, Tremonti & Leitherer 2018).

[8] In the linear–exponential case, the SFH is modelled as the product of a linear rise and exponential decline: $\dot{M}_*(t) \propto t e^{-t/\tau_{\rm SFH}}$.

[9] Again, it should be noted that this DTD is not necessarily consistent with current observational data (e.g. Maoz, Mannucci & Nelemans 2014) and represents another source of systematic uncertainty when considering the predictions of the WAF model.

[10] The fiducial model parameters are taken from a numerical implementation of the WAF model presented in Andrews et al. (2017), and are adopted here unless explicitly stated. They are calibrated to reproduce the [O/Fe] versus [Fe/H] distribution of local thin disc, thick disc, and halo stars.

[11] We note that these scaling relation have since been updated by Hayward & Hopkins (2017) and Anglés-Alcázar et al. (2017), but we consider this simple parametrization sufficient for our purposes here.

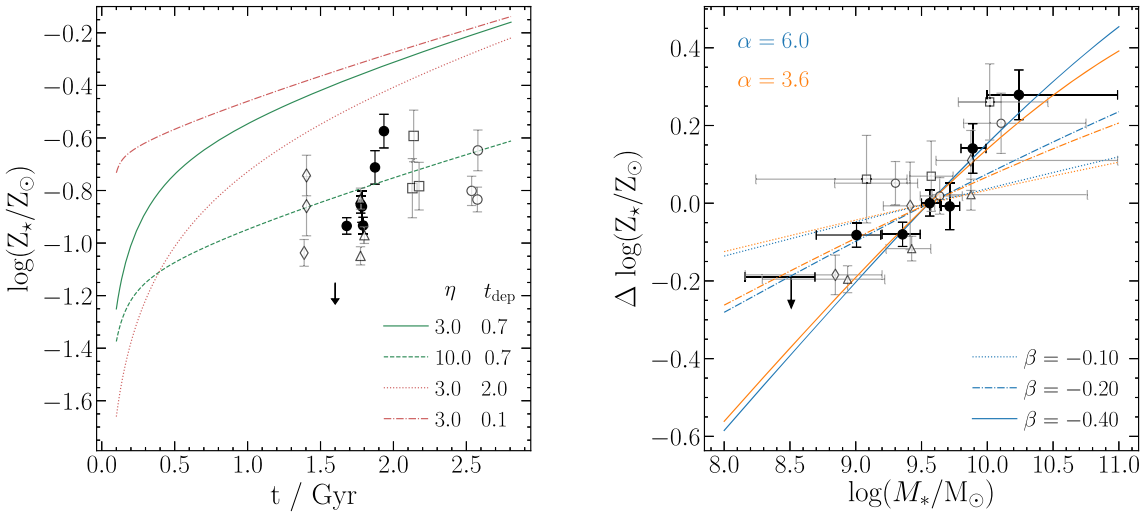

Figure 12. The evolution of $\log(Z_*/Z_\odot)$ as a function of cosmological time (left-hand plot) and $\Delta\log(Z_*/Z_\odot)$ as a function of stellar mass (right-hand plot) compared to predictions from the one-zone analytic chemical evolution models of Weinberg et al. (2017). The left-hand plot shows the time evolution of $\log(Z_*/Z_\odot)$ from $t = 0$ (i.e. the Big Bang) to the age of the Universe at the minimum redshift of our sample ($z = 2.3$; $t = 2.8$ Gyr). The green curves show $\log(Z_*/Z_\odot)$ evolution assuming a constant depletion time-scale ($t_{\rm dep} = 0.7$ Gyr) for two different values of the mass-loading parameter ($\eta = 3$ and 10), while the maroon curves show the evolution assuming a fixed value of $\eta = 3.0$ and two different values of the depletion time-scale ($t_{\rm dep} = 0.1$ and 2 Gyr). The significantly subsolar abundances of our sample (∼10–20 per cent) require mass outflow rates roughly an order of magnitude greater than the star formation rates (averaged across all stellar masses). The right-hand plot shows the difference in $\log(Z_*/Z_\odot)$ as a function of stellar mass assuming different forms of the $M_*-\eta$ relationship. The functional form is taken from the scaling relation presented by Muratov et al. (2015; see equation 5). The orange and blue lines show the $\Delta\log(Z_*/Z_\odot)$-$\log(M_*/{\rm M}_\odot)$ predictions relative to the value at $\log(M_*/{\rm M}_\odot) \simeq 9.5$ for different assumptions of the absolute mass-loading parameter at that stellar mass; $\eta = 3.6$ and $\eta = 6$ (see text for discussion). Regardless of the normalization, the evolution of $\log(Z_*/Z_\odot)$ as a function of stellar mass is consistent with a power-law exponent $\beta < -0.2$ and in good agreement with the Muratov et al. (2015) result ($\beta = -0.36$).

assumed yields, however, the corrections to the WAF would have to be quite large to make $\eta = 3.0$ consistent with the observed data (e.g. *both* the CCSNe and Type Ia SNe yields would have to be decreased by a factor 3).

In summary, whilst acknowledging the caveats associated with simple analytic models, we find that the WAF model supports a scenario in which the significantly subsolar iron abundances of the galaxies in our sample, and the dependence of iron abundance on stellar mass, are a consequence of strong outflows that scale with $\log(M_*/{\rm M}_\odot)$ to the power $\simeq -0.4$. In this scenario, the mass of gas outflowing from the galaxy, at $M_* = 10^9\,{\rm M}_\odot$, is predicted to be roughly an order of-magnitude greater than the gas mass being incorporated into new stars. Finally, we note that the inferences drawn from the WAF model are broadly similar to the those drawn from hydrodynamical simulations. Although these methods are not fully independent, since simulations still rely some analytic subgrid recipes, the agreement is encouraging and highlights the utility of simple analytic prescriptions for chemical evolution.

5.3 Enhanced O/Fe ratios at high redshift

Important insights can be gained by comparing the stellar metallicities of our sample to published gas-phase metallicities at similar redshifts. The comparison is of interest because, while the stellar metallicities derived here are sensitive to elements dominating photospheric absorption in massive stars (namely iron), the gas-phase metallicities are determined from the nebular emission lines emitted from the surrounding H II regions and are therefore sensitive to the dominant nebular coolants (namely oxygen). Under the assumption that the chemical abundances of massive stars should be similar to the surrounding H II region gas − out of which they presumably formed − then comparing stellar and gas-phase metallicities should be a good proxy for the O/Fe ratio in these galaxies (Steidel et al. 2016).

Non-solar O/Fe ratios should be expected for galaxies with constant or rising SFHs and relatively young ages (<500 Myr−1 Gyr). This results from the fact that the release of Fe into the ISM is determined by both the rate of Type Ia SNe and CCSNe, while the release of O (and other α elements) is determined only by the rate of CCSNe. Crucially, the Type Ia SNe rate is sensitive to the SFR of a galaxy over a large range of epoch before the time of observations (from ≃50 Myr to ≃10 Gyr; see e.g. Maoz & Mannucci 2012; Maiolino & Mannucci 2019), while CCSNe enrich the ISM almost instantaneously (within a few Myr of a given star formation episode). Therefore, for constant or rising SFHs, the abundance of the ISM will always be dominated by CCSNe products at young ages. The yields of Nomoto et al. (2006) predict, for a Salpeter IMF (Salpeter 1955), that CCSNe yields produce $({\rm O/Fe}) \simeq 4-6 \times ({\rm O/Fe})_\odot$ for initial stellar metallicities $\simeq 0.1-2 \times {\rm Z}_\odot$. The yields of the Chieffi & Limongi (2004) and Limongi & Chieffi (2006) fiducial model predict that CCSNe produce 1.5 $\rm M_\odot$ of O and 0.12 $\rm M_\odot$ of Fe for every 100 $\rm M_\odot$ of star formation assuming a Kroupa IMF (Kroupa 2001), equivalent to $({\rm O/Fe}) \simeq 3 \times ({\rm O/Fe})_\odot$. Given the growing observational and theoretical evidence for constant or rising SFHs at the redshift of our sample (e.g. Papovich et al. 2011; Finlator et al. 2011; Reddy et al. 2012; see also Fig. 5), we might therefore expect to see $({\rm O/H}) \simeq 3-6 \times ({\rm Fe/H})$.

Indeed, this result has already been reported for star-forming galaxies at $\langle z \rangle = 2.3$ from the KBSS-MOSFIRE sample. Strom et al. (2018) present an analysis of the optical spectra of 148 galaxies allowing them to constrain [Fe/H] and [O/H] for individual galaxies, finding an average [O/Fe] = 0.42 [i.e. $({\rm O/H}) \simeq 2.6 \times ({\rm Fe/H})$], somewhat lower than the initial [O/Fe] ≃ 0.7 reported for a stack of 30 star-forming galaxies at the same redshift derived from a com-

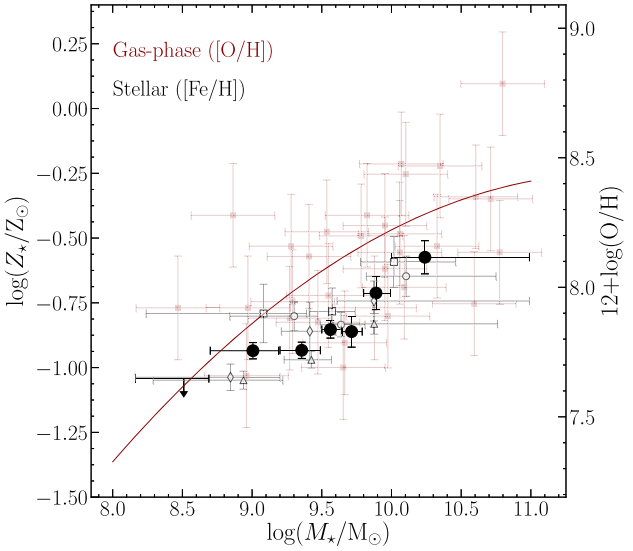

Figure 13. A comparison of the stellar and gas-phase mass–metallicity relationships at $z \sim 3.5$. The gas-phase data, tracing [O/H], are taken from the AMAZE/LSD survey of star-forming galaxies at $3 < z < 5$ with $\langle z \rangle = 3.4$ presented in Troncoso et al. (2014). The left-hand y-axis labels give metallicity in units of $\log(Z_*/Z_\odot)$, while the right-hand labels give the equivalent 12+log(O/H) units commonly used for reporting gas-phase oxygen abundances. The red square data points with error bars show the individual galaxy data and the solid red curve is their best-fitting relationship. The black/grey data points are the stellar metallicities of our VANDELS sample plotted as in Fig. 9. The comparison provides some evidence for alpha enhancement in star-forming galaxies at $z \sim 3.5$ of the order $(\mathrm{O/Fe}) \gtrsim 1.8 \times (\mathrm{O/Fe})_\odot$.

bined analysis of FUV and optical spectra presented in Steidel et al. (2016). Nevertheless, both values are comparable to the expected range of [O/Fe] from the CCSNe yield models described above.

Although we are unable to directly derive [O/H] for the galaxies in our sample, it is still instructive to compare the stellar metallicities reported here to [O/H] values at similar redshifts from the literature. To this end we show, in Fig. 13, the gas-phase mass–metallicity relationship for a sample of 34 galaxies at $3 < z < 5$ from the AMAZE/LSD survey (Maiolino et al. 2008; Mannucci et al. 2009; Troncoso et al. 2014) compared to the stellar metallicities derived for our VANDELS sample. The AMAZE/LSD sample spans the stellar mass range $8.5 < \log(M_*/\mathrm{M}_\odot) < 11.0$ and is composed of Lyman-break selected galaxies with a bias to the most highly star-forming galaxies at a given mass (Maiolino et al. 2008). The metallicities were derived from a combination of the [O II], H β, and [O III] nebular emission lines using the semi-empirical metallicity calibration of Maiolino et al. (2008; M08). Fig. 13 shows that the stellar metallicities (Fe/H) of star-forming galaxies at $z \simeq 3.5$ are systematically lower than the gas-phase metallicities (O/H) by $\simeq$0.25 dex at all stellar masses compared to the mean relation found by Troncoso et al. (2014). This value is comparable to the prediction of the WAF chemical evolution model described above, which, assuming the best-fitting $\eta - M_*$ relation, predicts [O/Fe] $\simeq$ 0.2 dex at $z = 3.5$ (i.e. after $\simeq$1.8 Gyr of evolution) across all stellar masses.[12]

[12]This value is lower than that predicted directly from the yields because it incorporates the assumed SFH and Type Ia SNe DTD.

However, there are a number of caveats that should be acknowledged with respect to this comparison. First, we note that the galaxies in the Troncoso et al. (2014) sample have, on average, higher SFRs at fixed stellar mass than the galaxies in our sample (the median sSFR of the AMAZE/LSD galaxies is 5.4 Gyr^{-1} compared to 4.4 Gyr^{-1} for our sample). Assuming a stellar mass–SFR–metallicity relation (e.g. Mannucci et al. 2010) is in place at high redshifts (as claimed recently by Sanders et al. 2018 at $z = 2.3$) then, due to their higher SFRs, the AMAZE/LSD gas-phase metallicities will be biased low with respect to the true gas-phase metallicities of our sample. In this case, the offset of [O/Fe] = 0.25 dex would be a lower limit. Secondly, the comparison is affected by zero-point offsets in both the stellar and gas-phase metallicity values. For example, in Appendix C we show how using an alternative stellar population model (in this case BPASSv2.2; Eldridge et al. 2017) can result in a systematic shift of $\simeq -0.1$ dex in $\log(Z_*/Z_\odot)$. Additionally, the zero-point offset in gas-phase metallicity calibrations, which has been well studied in the local Universe (e.g. Kewley & Ellison 2008), implies that the calibration adopted by Troncoso et al. (2014; M08) returns metallicities intermediate between other calibrations with offsets of up to $\simeq \pm 0.2$ dex across the stellar mass range shown in Fig. 13 (e.g see fig. 3 of Sánchez et al. 2019).

Clearly, this comparison is currently limited to qualitative statements, and is hampered by the different selection biases affecting the VANDELS and AMAZE/LSD samples. A more definitive answer will have to wait until rest-frame optical observations of VANDELS galaxies are available, from which [O/H] can be directly estimated. Nevertheless, adopting currently published values suggests that the O/Fe ratio is enhanced in high-redshift galaxies relative to the solar values, with $(\mathrm{O/Fe}) \gtrsim 1.8 \times (\mathrm{O/Fe})_\odot$. One major practical implication of this result is that stellar population models that assume solar chemical abundance ratios are likely not valid for modelling the full spectra energy distribution of galaxies at high redshift (e.g. Steidel et al. 2016).[13]

6 CONCLUSIONS

In this paper, we have presented the results of a study of the stellar mass–metallicity relationship for a sample of 681 star-forming galaxies at $2.3 < z < 5.0$ spanning the stellar masses range $8.2 < \log(M_*/\mathrm{M}_\odot) < 11.0$, based on deep rest-frame FUV spectroscopic data from the VANDELS survey (McLure et al. 2018b; Pentericci et al. 2018). Stellar metallicities (to first order a proxy for the iron abundance) have been derived by fitting high-resolution theoretical SB99 WM-Basic stellar population synthesis models (Leitherer et al. 2010) to high S/N composite spectra in bins of stellar mass and redshift. Our method is tested using both synthetic spectra generated from simulation data and FUV spectra of galaxies in the local Universe. Finally, we have investigated the normalization and shape of the stellar mass–metallicity relationship at $2.3 < z < 5.0$, and compared our results to predictions from state-of-the-art cosmological simulations and simple one-zone analytic models for chemical evolution. Our main results can be summarized as follows:

(i) We find a strong correlation between stellar metallicity and stellar mass at $2.3 < z < 5.0$ ($\langle z \rangle = 3.5$) with $\log(Z_*/Z_\odot)$

[13]Although the SB99 models used here assume a solar abundance pattern, the fact that the FUV spectral features are primarily determined by Fe abundance means our recovered $\log(Z_*/Z_\odot)$ values should not be strongly dependent on the detailed abundance pattern. Nevertheless, in future it would clearly be desirable to test this assumption.

monotonically increasing from <-1.04 for galaxies with $\langle M_* \rangle = 3.2 \times 10^8 \mathrm{M}_\odot$ to -0.57 at $\langle M_* \rangle = 1.7 \times 10^{10} \mathrm{M}_\odot$. Across the full mass range, the metallicities (iron abundances) we derive are significantly subsolar with $Z_* \lesssim 0.25 \mathrm{Z}_\odot$.

(ii) We do not observe a strong relationship between stellar metallicity and redshift within our current sample. Within the error bars, the metallicities derived for composite spectra binned in redshift and stellar mass are consistent with scattering around the stellar-mass-only relationship.

(iii) The stellar mass–metallicity relation we derive at $\langle z \rangle = 3.5$ is offset to lower metallicities by $\simeq$0.6 dex compated to the $z = 0$ relation for star-forming galaxies derived from optical spectroscopic data (Zahid et al. 2017). There are some systematic uncertainties related to this comparison (e.g. probing different stellar populations and/or abundance types), but similar results are obtained by comparing to the FUV-based metallicities at $z = 0$ from FOS–GHRS data.

(iv) The normalization of the $\langle z \rangle = 3.5$ stellar mass–metallicity relation is not well reproduced by simulations. Absolute metallicity values tend to fall either systematically above (e.g FIBY) or below (e.g. SIMBA, FIRE) our data. However, given the relatively large uncertainties in stellar yields, it is hard to draw strong conclusions from this observation. On the other hand, the shape of the relation is generally well recovered. We postulate that this is likely due to the fact that the scaling between outflow efficiency (parametrized by the mass-loading parameter $\eta = \dot{M}_{\mathrm{outflow}}/\dot{M}_*$) and stellar mass is similar across all of the simulations.

(v) We investigate the parameters affecting the shape and normalization of the stellar mass–metallicity relation using detailed one-zone analytic chemical evolution models (the WAF models; Andrews et al. 2017; Weinberg et al. 2017). We find that these models can reproduce the observed shape the relation by assuming a power-law scaling relation between η and M_* that is similar to that derived from fully cosmological simulations (e.g. Muratov et al. 2015) and predicted from analytic models of momentum-driven winds (Hayward & Hopkins 2017). A comparison of the models and simulations to our data suggest that $\eta \propto M_*^{\beta}$ with $\beta \simeq -0.4$.

(vi) Furthermore, simulations that include state-of-the-art prescriptions for stellar feedback (FIRE and SIMBA), as well as the WAF model, suggest that an average mass-loading parameter of $\langle \eta \rangle \simeq 10$ is required for consistency with our data, ranging from $\eta \simeq 6$ at $M_* = 10^{10} \mathrm{M}_\odot$, to $\eta \simeq 40$ at $M_* = 10^8 \mathrm{M}_\odot$. Although the absolute normalization of the $\eta - M_*$ relation is subject to yield uncertainties, the reasonable agreement with predictions from cosmological simulations is encouraging.

(vii) Finally, by comparing the stellar metallicities derived here (which trace Fe/H) to published gas-phase metallicities at similar redshifts (Troncoso et al. 2014), we find that the gas-phase abundances (which trace O/H) are likely enhanced by at least a factor of $\simeq$2. This comparison provides further evidence that star-forming galaxies at these epochs are α-enhanced systems, although it will require direct [O/H] estimates for the galaxies in our sample to provide a more robust measurement.

ACKNOWLEDGEMENTS

FC, RJM, JSD, SK, AC, and DJM acknowledge the support of the UK Science and Technology Facilities Council. This work is based on data products from observations made with ESO Telescopes at La Silla Paranal Observatory under ESO programme ID 194.A-2003(E-Q). AC acknowledges the grants PRIN MIUR 2015, ASI n.I/023/12/0, and ASI n.2018-23-HH.0. GC has been supported by the INAF PRIN-SKA 2017 program 1.05.01.88.04. This research has used ASTROPY, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013), NumPy and SciPy (Oliphant 2007), Matplotlib (Hunter 2007), IPython (Pérez & Granger 2007), and NASA's Astrophysics Data System Bibliographic Services.

REFERENCES

Abel T., Anninos P., Zhang Y., Norman M. L., 1997, New Astron., 2, 181
Amorín R. et al., 2017, Nat. Astron., 1, 0052
Andrews B. H., Weinberg D. H., Schönrich R., Johnson J. A., 2017, ApJ, 835, 224
Anglés-Alcázar D., Faucher Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Astropy Collaboration, 2013, A&A, 558, A33
Barrera-Ballesteros J. K., Sánchez S. F., Heckman T., Blanc G. A., The MaNGA Team, 2017, ApJ, 844, 80
Brandt J. C. et al., 1998, AJ, 116, 941
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buchner J. et al., 2014, A&A, 564, A125
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Carnall A. C., Leja J., Johnson B. D., McLure R. J., Dunlop J. S., Conroy C., 2018a, ApJ, 873, 44
Carnall A. C., McLure R. J., Dunlop J. S., Davé R., 2018b, MNRAS, 480, 4379
Chieffi A., Limongi M., 2004, ApJ, 608, 405
Chisholm J., Tremonti C., Leitherer C., 2018, MNRAS, 481, 1690
Cid Fernandes R. et al., 2014, A&A, 561, A130
Conroy C., 2013, ARA&A, 51, 393
Conroy C., Villaume A., van Dokkum P. G., Lind K., 2018, ApJ, 854, 139
Cullen F., Cirasuolo M., McLure R. J., Dunlop J. S., Bowler R. A. A., 2014, MNRAS, 440, 2300
Cullen F., Cirasuolo M., Kewley L. J., McLure R. J., Dunlop J. S., Bowler R. A. A., 2016, MNRAS, 460, 3002
Cullen F., McLure R. J., Khochfar S., Dunlop J. S., Dalla Vecchia C., 2017, MNRAS, 470, 3006
Cullen F. et al., 2018, MNRAS, 476, 3218
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
De Rossi M. E., Bower R. G., Font A. S., Schaye J., Theuns T., 2017, MNRAS, 472, 3354
Dean C. A., Bruhweiler F. C., 1985, ApJS, 57, 133
Dessauges-Zavadsky M., D'Odorico S., Schaerer D., Modigliani A., Tapken C., Vernet J., 2010, A&A, 510, A26
Eldridge J. J., Stanway E. R., 2012, MNRAS, 419, 479
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, Publ. Astron. Soc. Aust., 34, e058
Faber S. M., Friel E. D., Burstein D., Gaskell C. M., 1985, ApJS, 57, 711
Faisst A. L. et al., 2016, ApJ, 822, 29
Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761
Ferland G. J. et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Finlator K., Oppenheimer B. D., Davé R., 2011, MNRAS, 410, 1703
Fletcher T. J., Tang M., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2018, preprint (arXiv:1806.01741)
Galametz A. et al., 2013, ApJS, 206, 10
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
Galli D., Palla F., 1998, A&A, 335, 403
Garilli B., Paioro L., Scodeggio M., Franzetti P., Fumana M., Guzzo L., 2012, PASP, 124, 1232

González Delgado R. M. et al., 2014, ApJ, 791, L16
Grazian A. et al., 2017, A&A, 602, A18
Grogin N. A. et al., 2011, ApJS, 197, 35
Guo Y. et al., 2013, ApJS, 207, 24
Halliday C. et al., 2008, A&A, 479, 417
Hayward C. C., Hopkins P. F., 2017, MNRAS, 465, 1682
Hirschmann M., De Lucia G., Fontanot F., 2016, MNRAS, 461, 1760
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., 2017, preprint (arXiv:1712.01294)
Hopkins P. F., Kereš D., Oñorbe J., Faucher Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F.-K., 1999, ApJS, 125, 439
Johnson J. L., Dalla Vecchia C., Khochfar S., 2013, MNRAS, 428, 1857
Kashino D. et al., 2017, ApJ, 835, 88
Kennicutt Jr. R. C., 1998, ApJ, 498, 541
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, ApJ, 779, 102
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Kriek M., Conroy C., 2013, ApJ, 775, L16
Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221
Kroupa P., 2001, MNRAS, 322, 231
Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36
Kudritzki R.-P., Puls J., 2000, ARA&A, 38, 613
Leitherer C., Ortiz Otálvaro P. A., Bresolin F., Kudritzki R.-P., Lo Faro B., Pauldrach A. W. A., Pettini M., Rix S. A., 2010, ApJS, 189, 309
Leitherer C., Tremonti C. A., Heckman T. M., Calzetti D., 2011, AJ, 141, 37
Leitherer C., Ekström S., Meynet G., Schaerer D., Agienko K. B., Levesque E. M., 2014, ApJS, 212, 14
Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2019, ApJ, 876, 3
Leroy A. K. et al., 2013, AJ, 146, 19
Lian J., Thomas D., Maraston C., Goddard D., Comparat J., Gonzalez-Perez V., Ventura P., 2018a, MNRAS, 474, 1143
Lian J. et al., 2018b, MNRAS, 476, 3883
Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, ApJ, 772, 119
Limongi M., Chieffi A., 2006, ApJ, 647, 483
Ma X., Hopkins P. F., Faucher Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140
Maier C., Lilly S. J., Ziegler B. L., Contini T., Pérez Montero E., Peng Y., Balestra I., 2014, ApJ, 792, 3
Maiolino R., Mannucci F., 2019, A&AR, 27, 3
Maiolino R. et al., 2008, A&A, 488, 463
Mannucci F. et al., 2009, MNRAS, 398, 1915
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Maoz D., Mannucci F., 2012, Publ. Astron. Soc. Aust., 29, 447
Maoz D., Mannucci F., Nelemans G., 2014, ARA&A, 52, 107
McGaugh S. S., Schombert J. M., 2014, AJ, 148, 77
McLure R. J. et al., 2018a, MNRAS, 476, 3991
McLure R. J. et al., 2018b, MNRAS, 479, 25
Mollá M., García-Vargas M. L., Bressan A., 2009, MNRAS, 398, 451
Muratov A. L., Kereš D., Faucher Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691
Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nucl. Phys. A, 777, 424
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Oliphant T. E., 2007, Comput. Sci. Eng., 9, 10
Onodera M. et al., 2016, ApJ, 822, 42
Oppenheimer B. D., Davé R., 2008, MNRAS, 387, 577
Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544
Panter B., Jimenez R., Heavens A. F., Charlot S., 2008, MNRAS, 391, 1117
Papovich C., Finkelstein S. L., Ferguson H. C., Lotz J. M., Giavalisco M., 2011, MNRAS, 412, 1123
Peeples M. S., Shankar F., 2011, MNRAS, 417, 2962
Pentericci L. et al., 2018, A&A, 616, A174
Pérez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Pettini M., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., 2000, ApJ, 528, 96
Puls J., Vink J. S., Najarro F., 2008, A&AR, 16, 209
Quider A. M., Pettini M., Shapley A. E., Steidel C. C., 2009, MNRAS, 398, 1263
Reddy N. A., Pettini M., Steidel C. C., Shapley A. E., Erb D. K., Law D. R., 2012, ApJ, 754, 25
Reddy N. A. et al., 2018, ApJ, 853, 56
Rix S. A., Pettini M., Leitherer C., Bresolin F., Kudritzki R.-P., Steidel C. C., 2004, ApJ, 615, 98
Romano D., Karakas A. I., Tosi M., Matteucci F., 2010, A&A, 522, A32
Salim S., Lee J. C., Davé R., Dickinson M., 2015, ApJ, 808, 25
Salim S., Boquien M., Lee J. C., 2018, ApJ, 859, 11
Salpeter E. E., 1955, ApJ, 121, 161
Sánchez S. F. et al., 2019, MNRAS, 484, 3042
Sanders R. L. et al., 2015, ApJ, 799, 138
Sanders R. L. et al., 2016, ApJ, 816, 23
Sanders R. L. et al., 2018, ApJ, 858, 99
Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210
Schaye J. et al., 2010, MNRAS, 402, 1536
Schmidt M., 1959, ApJ, 129, 243
Schreiber C. et al., 2018, A&A, 611, A22
Scodeggio M. et al., 2005, PASP, 117, 1284
Scott N. et al., 2017, MNRAS, 472, 2833
Scoville N. et al., 2016, ApJ, 820, 83
Scoville N. et al., 2017, ApJ, 837, 150
Shapley A. E. et al., 2015, ApJ, 801, 88
Skilling J., 2006, Bayesian Anal., 1, 833
Smith B. D. et al., 2017, MNRAS, 466, 2217
Sommariva V., Mannucci F., Cresci G., Maiolino R., Marconi A., Nagao T., Baroni A., Grazian A., 2012, A&A, 539, A136
Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289
Steidel C. C. et al., 2014, ApJ, 795, 165
Steidel C. C., Strom A. L., Pettini M., Rudie G. C., Reddy N. A., Trainor R. F., 2016, ApJ, 826, 159
Steidel C. C., Bogosavljević M., Shapley A. E., Reddy N. A., Rudie G. C., Pettini M., Trainor R. F., Strom A. L., 2018, ApJ, 869, 123
Strom A. L., Steidel C. C., Rudie G. C., Trainor R. F., Pettini M., Reddy N. A., 2017, ApJ, 836, 164
Strom A. L., Steidel C. C., Rudie G. C., Trainor R. F., Pettini M., 2018, ApJ, 868, 117
Tacconi L. J. et al., 2013, ApJ, 768, 74
Thomas D., Maraston C., Bender R., 2003, MNRAS, 339, 897
Toribio San Cipriano L., Domínguez-Guzmán G., Esteban C., García-Rojas J., Mesa-Delgado A., Bresolin F., Rodríguez M., Simón-Díaz S., 2017, MNRAS, 467, 3759
Troncoso P. et al., 2014, A&A, 563, A58
Trussler J., Maiolino R., Maraston C., Peng Y., Thomas D., Goddard D., Lian J., 2018, preprint (arXiv:1811.09283)
Vidal-García A., Charlot S., Bruzual G., Hubeny I., 2017, MNRAS, 470, 3532
Walcher J., Groves B., Budavári T., Dale D., 2011, Ap&SS, 331, 1
Weinberg D. H., Andrews B. H., Freudenburg J., 2017, ApJ, 837, 183
Wiersma R. P. C., Schaye J., Smith B. D., 2009, MNRAS, 393, 99
Wuyts E. et al., 2014, ApJ, 789, L40
Yoon S.-C., Langer N., 2005, A&A, 443, 643
Yoshida N., Omukai K., Hernquist L., Abel T., 2006, ApJ, 652, 6
Zahid H. J., Kudritzki R.-P., Conroy C., Andrews B., Ho I.-T., 2017, ApJ, 847, 18
Zetterlund E., Levesque E. M., Leitherer C., Danforth C. W., 2015, ApJ, 805, 151
Zhu G. B., Barrera-Ballesteros J. K., Heckman T. M., Zakamska N. L., Sánchez S. F., Yan R., Brinkmann J., 2017, MNRAS, 468, 4494

APPENDIX A: FOS–GHRS LOCAL SAMPLE

The local FUV spectra are taken from a complication of 28 local starbursts and star-forming galaxies presented in Leitherer et al. (2011; L11). L11 present 46 rest-frame UV spectra observed with the FOS and the GHRS of the *HST*. The spectral resolution of the individual spectra spans the range 0.5–3 Å depending on the instrument, aperture size, and physical extent of the object being observed. For consistency with the VANDELS data, we smooth all spectra to a common 3 Å resolution. We also require that the spectra have wavelength coverage in the interval $1410 \leq \lambda \leq 1450$ Å to enable an analysis of normalized composite spectra, and finally that the corresponding galaxy has a measurement of absolute *K*-band magnitude that we can use to estimate the stellar mass. This selection leaves 26 of 46 of the original spectra from 18 of 28 of the original local starbursts and star-forming galaxies presented in L11. A list of the individual spectra used in this work is presented in Table A1.

Table A1. Sample of local starbursts and star-forming galaxies taken from Leitherer et al. (2011). We use the M/L relations from McGaugh & Schombert (2014) to convert from M_K to $\log(M_*/\mathrm{M}_\odot)$.

Galaxy	M_K	$\log(M_*/\mathrm{M}_\odot)$
NGC2363-a	−17.1	7.93
NGC2363-b	−17.1	7.93
NGC1705	−18.0	8.29
SBS0335-052	−18.0	8.29
NGC1569	−18.9	8.65
NGC4214	−19.4	8.85
NCG5253-UV1	−19.5	8.89
NCG5253-UV2	−19.5	8.89
He2-10a	−21.1	9.53
He2-10b	−21.1	9.53
NGC4670	−21.4	9.65
NGC1741	−21.8	9.81
IC3639	−23.0	10.29
NGC7714	−23.2	10.37
NGC5457-Searle5	−23.8	10.61
NGC7552	−24.2	10.77
NGC4038-405	−24.5	10.89
NGC4038-442	−24.5	10.89
NGC3690	−25.0	11.09
NGC5135	−25.0	11.09
NGC7130	−25.0	11.09
NGC1068-pos1	−25.1	11.13
NGC1068-pos3	−25.1	11.13
NGC1068-pos5	−25.1	11.13
NGC1068-pos8a	−25.1	11.13

APPENDIX B: EFFECT OF REDSHIFT UNCERTAINTIES

One potential source of bias in our analysis relates to the effect of redshift uncertainties when stacking. The redshifts for the individual VANDELS spectra are constrained primarily by the strong interstellar absorption features and/or Ly α emission line. These features originate from outflowing gas and are therefore often redshifted or blueshifted by up to hundreds of kms^{-1} with respect to the systemic redshift of the galaxy. For example, from a sample of 89 galaxies at $z \sim 2-3$, Steidel et al. (2010) find an average velocity offset for redshifts derived from interstellar absorption lines of $\langle \Delta v_{\mathrm{IS}} \rangle = -164 \pm 16$ kms^{-1}, and of $\langle \Delta v_{\mathrm{Ly}\,\alpha} \rangle = +445 \pm 27\,\mathrm{km\,s}^{-1}$ for redshifts derived from Ly α emission. Each galaxy in our sample will therefore have some unknown redshift offset that could potentially result in a systematic bias when deriving metallicities. In particular, the weak continuum features that are used to constrain the metallicity via continuum fitting could be 'washed out' to make the continuum look increasingly featureless, and therefore bias measurements towards lower metallicity solutions.

We tested the effect of redshift uncertainties using simulated galaxy spectra. For each of the five default SB99 WM-Basic templates ($Z_* = 0.001, 0.002, 0.008, 0.014$, and 0.040), we produced 100 versions (i.e. roughly the number of galaxies in our mass stacks) each with a random velocity offset distributed uniformly within the range $-150 < \Delta v < 500\,\mathrm{km\,s}^{-1}$. The 100 spectra were then median combined to mimic stacking spectra with random, unknown, redshift uncertainties. These spectra were convolved to the resolution of the VANDELS data and Gaussian noise was added assuming an S/N per pixel of 15. The metallicity was then derived using the method applied to the observed data as described in Section 3. For each template the process was repeated 100 times and the distribution of recovered metallicities is shown in Fig. B1. It can be seen that although there is a systematic offset to lower metallicities for all but the lowest metallicity template (for which it is impossible to measure a lower metallicity), the effect is relatively small ($\leq$10 per cent). We therefore conclude that the low metallicities derived for our VANDELS sample are robust against the effect of redshift uncertainties.

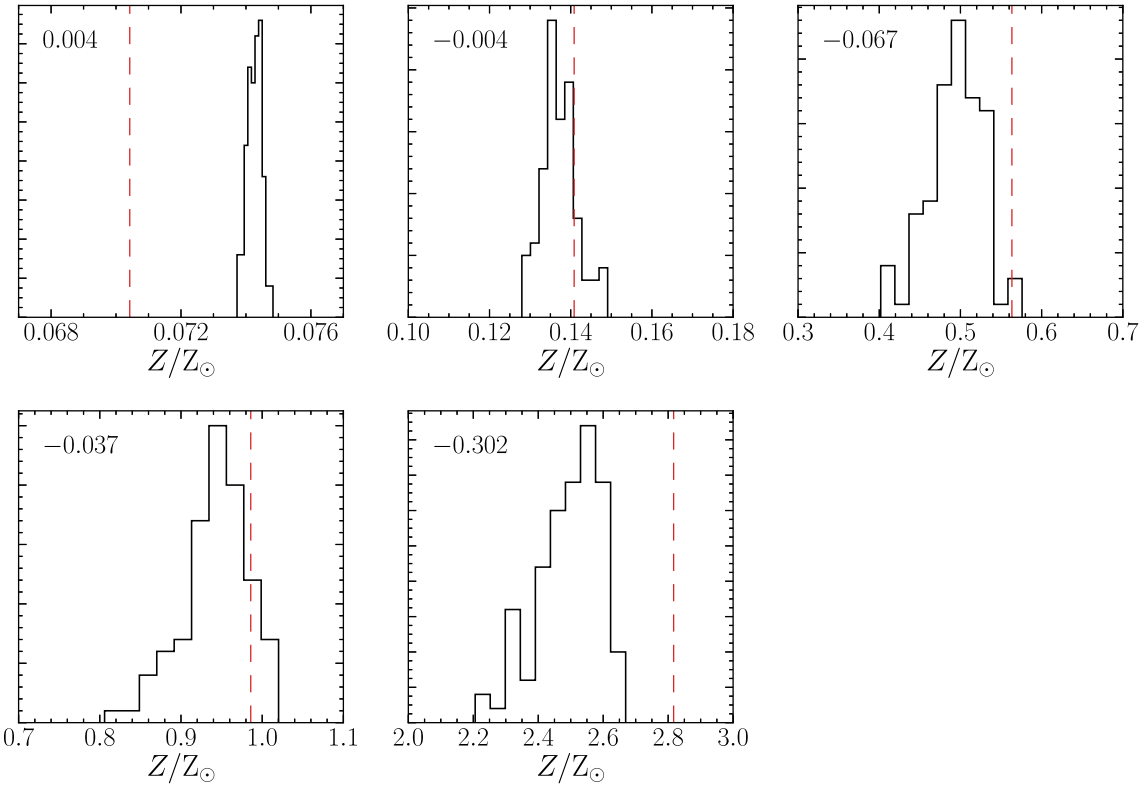

Figure B1. The distribution of recovered metallicity for each of the five default Starburst99 WM-Basic template spectra ($Z_* = 0.001$, 0.002, 0.008, 0.014, and 0.040) after mimicking the effect of redshift uncertainties when creating composite spectra (see text for details). Each panel shows one of the five default templates with the true metallicity indicated by the dashed vertical red line. The distribution of recovered metallicities is shown in black. The offset between the median recovered metallicity and the true metallicity is quoted in the top left-hand corner of each panel.

APPENDIX C: BPASSV2.1 STELLAR POPULATION SYNTHESIS MODELS

In our main analysis, we focused on using the SB99 WM-Basic stellar population synthesis models to fit the VANDELS spectra for the reasons outlined in Section 3. However, we have also performed the same analysis using the BPASSv2.1 stellar population synthesis models described in Eldridge et al. (2017). It is well known that using different models will cause systematic offsets in derived parameters (e.g. Conroy 2013; Cid Fernandes et al. 2014) and it is useful to have an estimate of how significant this offset might be in the case of FUV-derived metallicities.

The unique feature of the BPASSv2.1 models is the inclusion of massive binary star evolution, which can have a strong effect on the predicted UV spectrum, particularly at low metallicities. For example, at $Z_*/Z_\odot < 0.35$, a phenomenon known as 'quasi-homogeneous evolution' can occur (QHE; e.g. Eldridge & Stanway 2012), which results in stars living longer on the main sequence and becoming hotter than single stars at similar mass and metallicity. QHE can have a profound effect on the estimated ages and ionizing photon output of stellar populations. Stellar metallicities should be less affected, particularly since the O-star models in BPASSv2.1 are also generated with WM-Basic code, albeit with a different parameter set to SB99. Nevertheless, it worth testing this assumption.

We considered a 100 Myr constant star formation BPASSv2.1 model that included binary star evolution and had an IMF cut-off of 100 $M_\odot$, with IMF index of −1.3 between 0.1–0.5 $M_\odot$ and −2.35 above 0.5 $M_\odot$. The model is referred to as BPASSv2.1-

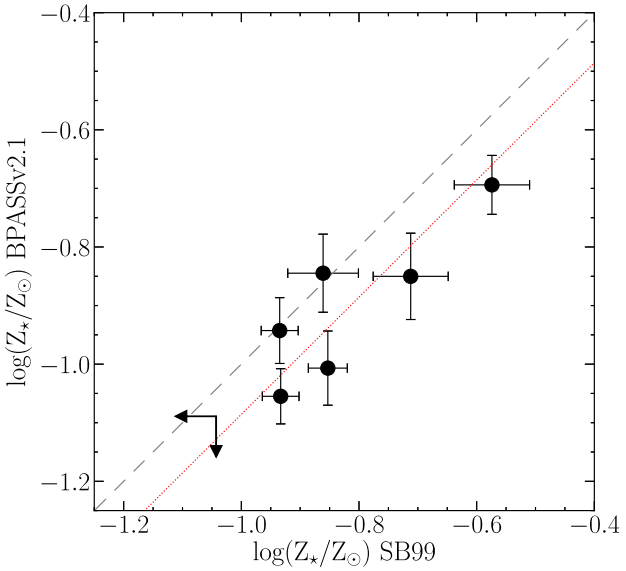

Figure C1. A comparison between stellar metallicities derived from the Starburst99 WM-Basic and BPASSv2.1 models for the VANDELS mass-stacks. The grey dashed line shows the 1:1 relation and the red dotted line shows the best-fitting constant offset from the 1:1 line. Both estimates are clearly correlated with the BPASSv2.1 metallicities offset to slightly lower metallicity values by ∼0.09 dex. For both Starburst99 and BPASSv2.1, the metallicity derived for the lowest mass bin is an upper limit (indicated by the arrows in the figure). The figure illustrates that adopting the BPASSv2.1 metallicity values as opposed to the Starburst99 values would not affect our main results.

100bin. We considered the following set of metallicity values $Z_* = (0.001, 0.002, 0.003, 0.004, 0.006, 0.008, 0.010, 0.014, 0.020, 0.030, 0.040)$, interpolating between the models in the same way as for SB99. Each of the seven mass stacks were fitted with the BPASSv2.1-100bin model in the same way as was done for the SB99 models (see Section 3 for details). Fig. C1 compares the metallicities derived from the BPASSv2.1-100bin model and the metallicities derived from SB99. It can be seen that the metallicity estimates are clearly correlated, with the BPASSv2.1 values offset to slightly lower metallicity by ∼0.09 dex (roughly a factor of 1.2). However, it is clear from this figure that adopting the BPASSv2.1 model estimates would not affect our main results.

[1]*SUPAScottish Universities Physics Alliance, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[2]*University of the Western Cape, Bellville, Cape Town 7535, South Africa*
[3]*South African Astronomical Observatories, Observatory, Cape Town 7925, South Africa*
[4]*Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile*
[5]*Departamento de Física y Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile*
[6]*INAF - Osservatorio Astronomico di Bologna, via P. Gobetti 93/3, I-40129 Bologna, Italy*
[7]*INAF−Osservatorio Astronomico di Roma, Via Frascati 33, I-00040 Monte Porzio Catone (RM), Italy*
[8]*Department of Physics and Astronomy (DIFA), University of Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy*
[9]*INAF - Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*
[10]*European Southern Observatory, Karl-Schwarzschild-Str 2, D-86748 Garching b. München, Germany*
[11]*The Cosmic Dawn Center, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen Ø, Denmark*
[12]*INAF-Astronomical Observatory of Trieste, via G.B. Tiepolo 11, I-34143 Trieste, Italy*
[13]*INAF-IASF Milano, via Bassini 15, I-20133 Milano, Italy*
[14]*Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile*
[15]*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[16]*European Southern Observatory, Alonso de Cordova, Santiago, Chile*
[17]*Department of Physics and Astronomy, University of California, 430 Portola Plaza, Los Angeles, CA 90095, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.