The NIRVANDELS Survey: a robust detection of α-enhancement in star-forming galaxies at $z \simeq 3.4$

F. Cullen [ID],[1]★ A. E. Shapley,[2] R. J. McLure,[1] J. S. Dunlop,[1] R. L. Sanders [ID],[3] M. W. Topping,[2]
N. A. Reddy,[4] R. Amorín [ID],[5,6] R. Begley,[1] M. Bolzonella,[7] A. Calabrò,[8] A. C. Carnall,[1] M. Castellano,[8]
A. Cimatti,[9,10] M. Cirasuolo,[11] G. Cresci,[10] A. Fontana,[8] F. Fontanot [ID],[12] B. Garilli,[13] L. Guaita,[14]
M. Hamadouche,[1] N. P. Hathi [ID],[15] F. Mannucci,[10] D. J. McLeod,[1] L. Pentericci,[8] A. Saxena [ID],[16]
M. Talia [ID][7,9] and G. Zamorani[7]

[1]SUPA†, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
[2]Department of Physics and Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA
[3]Department of Physics and Astronomy, University of California, Davis, One Shields Ave, Davis, CA 95616, USA
[4]Department of Physics and Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
[5]Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile
[6]Departamento de Física y Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile
[7]INAF – Osservatorio Astronomico di Bologna, Via P. Gobetti 93/3, I-40129, Bologna, Italy
[8]INAF − Osservatorio Astronomico di Roma, Via Frascati 33, I-00040 Monte Porzio Catone (RM), Italy
[9]University of Bologna, Department of Physics and Astronomy (DIFA) Via Gobetti 93/2, I-40129 Bologna, Italy
[10]INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125, Firenze, Italy
[11]European Southern Observatory, Karl-Schwarzschild-Str. 2, D-86748 Garching b. München, Germany
[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]Departamento de Ciencias Fisicas, Universidad Andres Bello, Fernandez Concha 700, Las Condes, Santiago, Chile
[15]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[16]Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

ABSTRACT

We present results from the NIRVANDELS survey on the gas-phase metallicity (Z_g, tracing O/H) and stellar metallicity ($Z_\star$, tracing Fe/H) of 33 star-forming galaxies at redshifts $2.95 < z < 3.80$. Based on a combined analysis of deep optical and near-IR spectra, tracing the rest-frame far-ultraviolet (FUV; 1200–2000 Å) and rest-frame optical (3400–5500 Å), respectively, we present the first simultaneous determination of the stellar and gas-phase mass–metallicity relationships (MZRs) at $z \simeq 3.4$. In both cases, we find that metallicity increases with increasing stellar mass ($M_\star$) and that the power-law slope at $M_\star \lesssim 10^{10} \mathrm{M}_\odot$ of both MZRs scales as $Z \propto M_\star^{0.3}$. Comparing the stellar and gas-phase MZRs, we present direct evidence for super-solar O/Fe ratios (i.e. α-enhancement) at $z > 3$, finding (O/Fe) $= 2.54 \pm 0.38 \times$ (O/Fe)$_\odot$, with no clear dependence on $M_\star$.

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

1 INTRODUCTION

The metal content of galaxies is affected by past and current star formation, gas accretion, and galactic winds, and therefore constrains all aspects of the cosmic baryon cycle. Of particular interest is the evolution of galaxy metallicity with cosmic time, and the scaling relations between metallicity and other galaxy properties, most notably stellar mass ($M_\star$). Accurately determined, these relations and their redshift evolution serve as powerful tests of theoretical models of galaxy evolution (Maiolino & Mannucci 2019).

Current constraints on galaxy metallicities primarily come from observations of strong nebular emission lines emitted at rest-frame optical wavelengths ($\simeq$3500–7000 Å) . When measured from integrated spectra of star-forming galaxies, the ratios of these emission lines are sensitive to the gas-phase oxygen abundance (O/H) of galactic H II regions (Kewley, Nicholls & Sutherland 2019). Using this technique, gas-phase metallicities (Z_g) have been measured for sizable samples of galaxies from the local Universe out to $z \simeq 4$ (e.g. Tremonti et al. 2004; Sanders et al. 2020a). Based on these observations, it is now well-established that Z_g is primarily correlated with $M_\star$ at all redshifts (i.e. the mass–metallicity relationship, or MZR) with strong evidence for a secondary dependence on star formation rate (SFR) or molecular gas fraction (i.e. the 'fundamental metallicity relationship', or FMR; Mannucci et al. 2010).

★ E-mail: fc@roe.ac.uk
† Scottish Universities Physics Alliance

An alternative approach to measuring galaxy metallicities at $z > 2$ utilizes observations of the stellar continuum at far-ultraviolet wavelengths (FUV; 1000–2000 Å). In contrast to nebular emission-line measurements, these estimates of the stellar metallicity ($Z_\star$) are sensitive to the photospheric iron abundance (Fe/H) of O- and B-type stars within galaxies (Leitherer et al. 2010). Early efforts to determine stellar metallicities from FUV spectra at high redshift were limited by small samples sizes (e.g. Halliday et al. 2008; Sommariva et al. 2012). However, in Cullen et al. (2019), building upon this early work, we utilized the large number of ultra-deep rest-frame FUV spectra provided by the VANDELS survey (McLure et al. 2018) to publish the first determination of the stellar mass–metallicity relationship at $z > 2$. We found that the stellar mass–metallicity relationship at $z \simeq 3.5$ has a similar shape but lower overall normalization when compared to the local relation, mirroring the redshift evolution observed in the gas-phase relation (see also Calabrò et al. 2021).

Access to deep rest-frame FUV and optical spectra at $z > 2$ (from ground and space-based optical and near-IR spectroscopy, respectively) offers a unique opportunity to move beyond single-element abundances, enabling the study of abundance ratios at early cosmic epochs. Specifically, the combined analysis of FUV+optical spectra allows for the simultaneous determination of Z_g and $Z_\star$, tracing the ratio of oxygen to iron (O/Fe) in young stars and the surrounding ISM (Steidel et al. 2016). The O/Fe ratio is of interest because it is a sensitive probe of the star formation and chemical enrichment history of galaxies. Typical O/Fe ratios at high redshift are expected to be enhanced relative to the solar value due to the fact that the element abundance ratios in relatively young star-forming systems will be dominated by core-collapse supernova (CCSNe) yields (Maiolino & Mannucci 2019). Indeed, this result has already been reported in the literature. Based on a simultaneous rest-frame FUV+optical analysis of 30 star-forming galaxies at $z \simeq 2.4$, Steidel et al. (2016) found O/Fe $\approx 4-5 \times$ O/Fe$_\odot$. Similar levels of O/Fe-enhancement have been reported for individual galaxies at $z \simeq 2.3$ in more recent studies (Topping et al. 2020a, b).

The observed O/Fe-enhancement (also referred to as α-enhancement) has a number of important implications. First, it places robust constraints on the typical star-formation histories at early cosmic epochs, confirming results from previous photometric analyses that the stellar populations at high redshift are typically <1 Gyr old (Reddy et al. 2012). Secondly, a deficit of Fe relative to O will mean that the stellar ionizing spectrum of typical high-redshift galaxies is harder at fixed oxygen abundance compared to galaxies in the local Universe. This relative hardening of the ionizing continuum at fixed O/H offers a natural explanation for the observed offset of $z > 2$ star-forming galaxies relative to $z \simeq 0$ galaxies in the common line ratio diagrams (e.g. the BPT diagram; Topping et al. 2020a; Runco et al. 2021). Non-solar α/Fe ratios will also force us to re-think current stellar population techniques when applied to high-redshift galaxies. Almost universally, current stellar population synthesis models assume solar abundance ratios; accurate analyses of FUV-optical spectra at $z > 6$ with *JWST* will require new models allowing for non-solar abundance ratios.

In this paper, we expand upon a number of previous works at $z \simeq 2.5$ (Steidel et al. 2016; Topping et al. 2020a, b) and present a simultaneous analysis of FUV and optical spectra for a sample of 33 star-forming galaxies drawn from the VANDELS survey at $z \simeq 3.4$. Combining ultra-deep optical VIMOS/VANDELS spectroscopic observations (tracing the rest-frame FUV) with MOSFIRE *H*- and *K*-band near-IR follow-up (tracing the rest-frame optical), our analysis provides the first investigation of Z_g and $Z_\star$ for galaxies at $z > 3$, and we present the first estimates of O/Fe at these redshifts. In addition,

we present both the stellar and gas-phase MZRs for our sample, tracing O/Fe as a function of stellar mass.

The structure of this paper is as follows. In Section 2, we discuss our combined VANDELS+MOSFIRE spectroscopic data set and describe our final $z \simeq 3.4$ galaxy sample (referred to throughout the rest of this paper as the NIRVANDELS sample). In Section 3, we outline the methods used to determine $Z_\star$ and Z_g from the rest-frame FUV and optical spectroscopy, respectively. In Section 4, we present our determination of the stellar and gas-phase MZRs at $z > 3$ along with an estimate of the typical O/Fe ratios of our sample. In Section 5, we discuss some of the implications of our results before summarizing our main conclusions in Section 6. Throughout this paper, metallicities are quoted relative to the solar abundance taken from Asplund et al. (2009), which has a bulk composition by mass of $Z_* = 0.0142$. We assume the following cosmology: $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1}.

2 DATA AND SAMPLE PROPERTIES

2.1 Rest-frame UV VANDELS sample and observations

The star-forming galaxy sample presented in this paper was initially drawn from the VANDELS ESO public spectroscopic survey (Pentericci et al. 2018; McLure et al. 2018). VANDELS is an ultra-deep, optical, spectroscopic survey of the CANDELS CDFS and UDS fields (Grogin et al. 2011; Koekemoer et al. 2011) undertaken using the VIMOS spectrograph on ESO's Very Large Telescope (VLT). The three categories of VANDELS targets were 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 $3.0 \leq z \leq 7.0$, with the main focus being star-forming galaxies at $z > 2.4$ (85 per cent of targets). Observations were obtained using the VIMOS medium-resolution grism covering the wavelength range 4800 Å < λ_{obs} < 10 000 Å at a resolution of $R = 580$ (with 1.0 arcsec slits) and a dispersion of 2.5 Å per pixel.

At the redshifts of interest for our study ($2.95 \leq z \leq 3.80$), the VIMOS spectra cover rest-frame UV emission at $\simeq$ 1000–2000 Å, a wavelength range sensitive to various continuum and emission-line features that trace the properties of young, massive stellar populations in star-forming galaxies (Cullen et al. 2019, 2020; Calabrò et al. 2021). The observations and reduction of the VIMOS spectra are described in detail in the VANDELS data release papers (Pentericci et al. 2018; Garilli et al. 2021). The selection of VANDELS targets for near-IR follow-up with MOSFIRE is described below.

2.2 Rest-frame optical MOSFIRE sample and observations

In order to characterize simultaneously the properties of massive stars and ionized gas in $z \simeq 3.4$ star-forming galaxies, we selected a sample of galaxies from the VANDELS survey for near-infrared spectro-scopic follow-up observations with the Multi-object Spectrometer for Infrared Exploration (MOSFIRE; McLean et al. 2012) on the Keck I telescope. The requirement for strong rest-frame optical features to fall within windows of atmospheric transmission translates into discrete allowed redshift ranges for targets for ground-based near-infrared spectroscopic follow-up, including $2.95 \leq z \leq 3.80$ for $z \sim 3$ targets and $2.09 \leq z \leq 2.61$ for those at $z \sim 2$.

For most of our MOSFIRE mask design (i.e. for masks observed in November 2019), we prioritized VANDELS galaxies with $H_{AB} \leq 25.5$, and robustly measured redshifts (e.g. characterized by redshift flags 3, 4, 9, and 14, as defined in Pentericci et al. 2018) at $2.95 \leq z \leq 3.80$. Slightly higher priority was given to sources with measured

Table 1. Summary of MOSFIRE Observations.

Mask	RA (J2000)	Decl. (J2000)	PA (deg)	K Exptime (s)	H Exptime (s)	N_{targ}	$N_{z \sim 3}$	$N_{z \sim 2}$	Obs. run	Seeing (arcsec)
gs_al1	03:32:43.00	−27:46:25.8	261.7	11520	7080	10	6	2	2018 Oct, 2019 Jan	0.6
gs_van2	03:32:10.99	−27:44:11.7	8.9	7200	7200	23	15	6	2019 Nov	0.5
ud_van7	02:18:00.43	−05:09:59.1	90.1	8640	8640	17	14	2	2019 Nov	0.5

C III] $\lambda\lambda1907$, 1909 rest-UV nebular emission. VANDELS sources with robustly measured redshifts (same redshift flags as above) at $2.09 \leq z \leq 2.61$ were targeted with lower priority, and, finally, VANDELS sources with $H_{AB} > 25.5$ and either $2.95 \leq z \leq 3.80$ or $2.09 \leq z \leq 2.61$ were considered the lowest priority. A slightly different priority scheme was used for the gs_al1 mask, which was designed a year earlier, prior to the completion of the final VANDELS redshift measurements. Accordingly, for this mask, galaxies at $H_{AB} \leq 25.5$ and robustly measured redshifts (again, as defined above) at either $2.95 \leq z \leq 3.80$ or $2.09 \leq z \leq 2.61$ were prioritized, followed by VANDELS targets with redshifts yet to be measured. In attempting to optimize the number of VANDELS sources per MOSFIRE pointing that satisfied the above criteria, we found that the best mask configurations contained 15–20 such targets.

We obtained Keck/MOSFIRE H-, and K-band rest-frame optical spectra for the selected VANDELS targets in the GOODS-S and UDS fields. At $2.95 \leq z \leq 3.8$ H β and [O III] $\lambda\lambda4959$, 5007 fall in the K band, while [O II] $\lambda\lambda3726$, 3729 and [Ne III] $\lambda3870$ fall in the H band. At $2.09 \leq z \leq 2.61$, H α, [N II] $\lambda6584$, and [S II] $\lambda\lambda6717$, 6731 fall in the K band, while H β and [O III] $\lambda\lambda4959$, 5007 fall in the H band. We collected observations of three MOSFIRE slitmasks (gs_al1, ud_van7, and gs_van2) on 2018 October 21, 2019 January 13, 2019 November 4, and 2019 November 13. Conditions were clear during the observations, with seeing ranging from 0.5 to 0.6 arcsec. The slitwidth was 0.7 arcsec, yielding a spectral resolution of $\simeq 3650$ in H and $\simeq 3600$ in K.[1] A total of 50 VANDELS sources were targeted within the three masks. The MOSFIRE observations are summarized in Table 1.

We reduced the raw data to produce two-dimensional science and error spectra using the pipeline described in Kriek et al. (2015). We then optimally extracted one-dimensional science and error spectra from the two-dimensional spectra. Flux calibrations and slit-loss corrections for each filter were applied as described in Kriek et al. (2015) and Reddy et al. (2015). Of the 50 VANDELS sources targeted, we obtained rest-frame optical spectra that yielded robust, science-grade redshifts for 35 sources at $2.95 \leq z \leq 3.8$ and 10 sources at $2.09 \leq z \leq 2.61$. In this paper, we focus exclusively on the $z \geq 2.95$ sources.

2.3 Measurements and derived quantities

2.3.1 Rest-frame optical emission-line fluxes and redshifts

Measurements of rest-frame optical emission-line fluxes were obtained by fitting Gaussian profiles to the flux-calibrated one-

dimensional MOSFIRE spectra. The 1σ uncertainty on each line flux was estimated by perturbing the spectra 500 times according to the error spectra, remeasuring the line flux, and taking the standard deviation of the resulting distribution. The absolute flux calibration of the MOSFIRE spectra is accurate to within 18 per cent and the relative calibration between H-, and K-band filters is accurate to within 13 per cent (Kriek et al. 2015). These line flux measurements therefore provide robust estimates of absolute line luminosity (e.g. for estimating SFRs) and cross-filter line ratios (e.g. [O III]/[O II]). The same error simulations were used to determine line centroids and errors, and the associated redshift and redshift uncertainty for each emission line. The individual line redshifts were combined using a weighted average to determine the final spectroscopic redshift and its associated uncertainty.

2.3.2 Stellar masses and star formation rates

We estimated stellar masses using multiwavelength photometry from the VANDELS photometric catalogues (McLure et al. 2018), taking into consideration the fact that some photometric filters could be contaminated by rest-frame optical emission-line flux. At high redshift ($z \gtrsim 2$), the large equivalent widths of the rest-frame optical emission lines (e.g. typical rest-frame [O III] $\lambda5007$ equivalent-width values of ≥ 300 Å at $10^{10} M_{\odot}$; Reddy et al. 2018) can contaminate broad-band photometry and result in an overestimation of galaxy masses. In order to derive more accurate stellar masses, we first corrected the H-, and K-band photometry for the emission-line fluxes measured from the MOSFIRE spectra. To perform this correction, we constructed a model emission-line-only spectrum for each galaxy based on the line fits described above. The flux contributed by emission lines to the H-, and K-band photometry was then determined by integrating the model over the appropriate filter profiles, and this flux was subtracted from the original photometry. Corrections to the H- and K-band photometry ranged from $-0.3 < \Delta H/$mag < 0.0 (median $\Delta H = -0.07$ mag) and $-1.8 < \Delta K/$mag < 0.0 (median $\Delta K = -0.35$ mag), respectively.

The emission-line corrected photometry was modelled using FAST++[2] (Schreiber et al. 2018), an SED-fitting code closely based on the original FAST software (Kriek et al. 2009). We adopted Conroy, Gunn & White (2009) flexible stellar population synthesis models and assumed solar metallicity,[3] a Chabrier (2003) initial mass

---

[1]While the MOSFIRE slitwidth (0.7 arcsec) is narrower than the VIMOS slitwidth (1.0 arcsec), the typical seeing in our MOSFIRE observations (0.5–0.6 arcsec) was also better than the typical seeing for the VANDELS VIMOS observations (0.7 arcsec). In fact, the VIMOS and MOSFIRE spectroscopic observations presented here probe comparable intrinsic regions of our target galaxies.

[2]https://github.com/cschreib/fastpp

[3]Although our analysis later in the paper (Section 4) places the α-element abundance of our sample at $\simeq 0.3-0.5$ $Z_{\odot}$ and the Fe/H abundance at $\simeq 0.1-0.2$ $Z_{\odot}$, adopting solar metallicity models is crucial for dust-correcting the nebular emission lines using the relation derived by Sanders et al. (2020a) (a different metallicity would bias the best-fitting continuum dust attenuation, see Section 2.3.3). This metallicity assumption will inevitably introduce some bias into the stellar mass determination (see e.g. Theios et al. 2019); however, it ensures our results and those of Sanders et al. (2020a) are directly comparable.

function (IMF), constant star formation histories, and the Calzetti et al. (2000) dust attenuation curve. The redshift was fixed to the measured spectroscopic redshift. These SED-fitting parameters were chosen to facilitate a direct comparison between our results and the recent study of gas-phase metallicities at $z \simeq 3.3$ by Sanders et al. (2020a). The resulting fits yielded an estimate of the galaxy stellar mass ($M_\star$), star formation rate (SFR$_{SED}$), stellar continuum dust attenuation $E(B - V)_\star$, and a model of the stellar continuum. Using the stellar continuum model, we corrected the H β line fluxes for underlying stellar absorption, resulting in a median increase of $\simeq 3$ per cent to the original flux values.

2.3.3 Dust-correcting nebular emission-line fluxes

In order to determine accurate optical line ratios for estimating gas-phase metallicities, the observed line fluxes need to be corrected for nebular extinction. Ideally, the nebular dust correction is determined directly using the Balmer decrement. However, this requires the detection of both the H α and H β emission lines and is therefore not possible for our $z \simeq 3.4$ sample. Instead, we used the best-fitting value of the stellar continuum attenuation to dust-correct the emission-line fluxes for each galaxy. We employed the calibration between stellar attenuation, SFR$_{SED}$, redshift and nebular extinction described in Sanders et al. (2020a), given by

$$E(B - V)_{neb} = E(B - V)_{stellar} - 0.604$$
$$+ 0.538 \times [\log(\text{SFR}_{SED}) - 0.20 \times (z - 2.3)]. \quad (1)$$

This calibration is based on observations of galaxies at $z \simeq 2.3$ with both $E(B - V)_{neb}$ (measured via the Balmer decrement) and $E(B - V)_\star$, and yields an unbiased estimate of $E(B - V)_{neb}$ with an intrinsic scatter of 0.23 mag and is directly applicable to our sample since we assume the same SED-fitting parameters used in the derivation of the calibration. Based on the estimate of $E(B - V)_{neb}$ given by equation (1), the observed emission line fluxes were corrected for reddening assuming the Cardelli, Clayton & Mathis (1989) extinction law.[4]

To assess the reliability of the nebular dust correction we derived star-formation rates from the dust-corrected H β line fluxes, assuming an intrinsic ratio of H α/H $\beta = 2.86$ and applying the Hao et al. (2011) H α-SFR conversion modified for a Chabrier (2003) IMF.[5] In Fig. 1, we show the resulting comparison between the H β-derived SFRs and the original photometrically derived SFR estimate for galaxies with a $>2\sigma$ detection of the H β line. The agreement is generally excellent, with a median offset of 0.06 dex in log(SFR) and a scatter of $\sigma = 0.3$ dex, where σ is derived from the median absolute deviation (MAD) ($\sigma = 1.4826 \times$ MAD).

2.3.4 The effect of emission lines on SED-derived stellar masses and star formation rates at $z \simeq 3.4$

In many instances, studies that rely on stellar mass estimates at $2.95 \leq z \leq 3.8$ do not have access to spectroscopic line flux measurements and as a result lack accurate corrections to the H- and K-band photometry. It is therefore interesting to assess
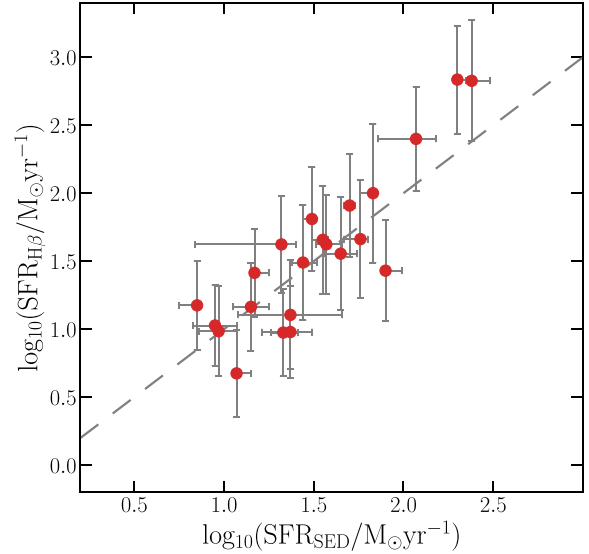
Figure 1. A comparison between star formation rate derived from the photometry and the dust-corrected H β line flux for the 22 galaxies in our sample with a $>2\sigma$ detection of the H β line. The grey dashed line shows the one-to-one relationship.

the effect of contamination from nebular emission lines on the derived stellar masses and SFRs at these redshifts. In Fig. 2, we compare the values of $M_\star$ and SFR(SED) estimated from corrected and uncorrected photometry. If $M_\star$ and SFR are derived using uncorrected H- and K-band photometry, we find that, on average, $\log(M_\star)$ is overestimated by 0.26 ± 0.16 dex and $\log(\text{SFR})$ is underestimated by 0.11 ± 0.10 dex. These offsets are consistent with previous results in the literature. In particular, the overestimation of $M_\star$ is a well-known effect of the contamination of broad-band photometry by high-equivalent width rest-frame optical emission lines at high redshift (e.g. Schaerer, de Barros & Sklias 2013; Amorín et al. 2015; Onodera et al. 2016). As can be seen from Fig. 2, the effect is strongest at low $M_\star$ where the equivalent widths are largest (e.g. Reddy et al. 2018). Indeed, this bias due to optical line contamination is expected to be ubiquitous at $z > 2$ due to the known evolution towards larger emission-line equivalent widths as a result of increasing specific star-formation rates and lower metallicity (Mármol-Queraltó et al. 2016; Reddy et al. 2018).

However, as can be seen from Fig. 2, unbiased estimates of $M_\star$ and SFR can be obtained by simply excluding the photometric bands that are known to include strong optical emission lines (in the case of $z \simeq 3$ galaxies, the H and K bands). With the contaminated bands excluded, the offsets in $\log(M_\star)$ and $\log(\text{SFR})$ are consistent with zero (-0.06 ± 0.13 and 0.05 ± 0.15 dex, respectively). In our particular case, the fact that we can recover these properties reliably in the absence of photometric anchors at rest-frame optical wavelengths (between 3000 and 6000 Å) is due to the data at longer wavelengths provided by the *Spitzer* IRAC 3.6 and 4.5 μm imaging (covering rest-frame wavelengths between 8000 and 10000 Å). In the absence of accurate line flux corrections, and with photometric data redwards of the K band, our results suggest that − for the SED fitting assumptions described above − simply excluding contaminated photometric bands from the SED-fitting process yields unbiased $M_\star$ and SFR estimates that are consistent with the emission line-corrected values to within $\simeq 10$–15 per cent.

---

[4]As shown by Reddy et al. (2020), the nebular attenuation law in high-redshift star-forming galaxies closely follows the Cardelli et al. (1989) Milky Way extinction curve.
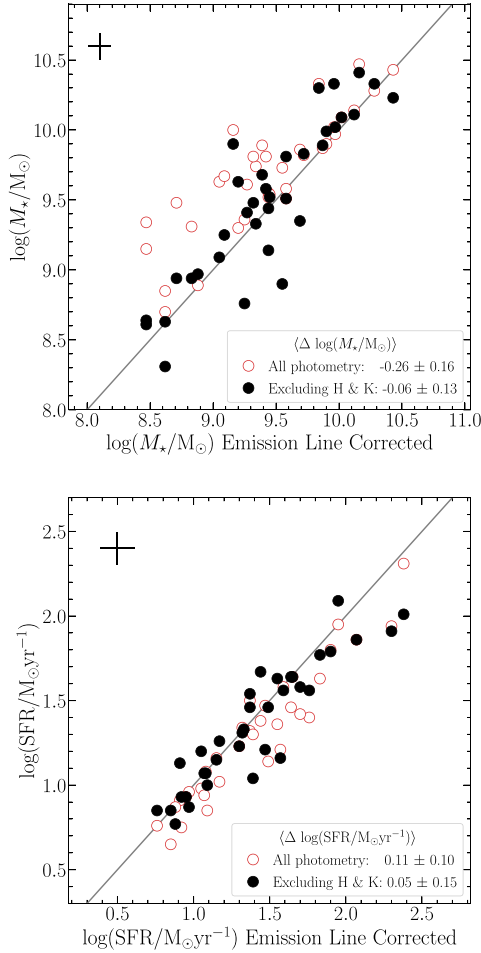
[5]To convert SFRs from the Kroupa (2001) IMF assumed in Hao et al. (2011) to a Chabrier (2003) IMF we divide by 1.06.

Figure 2. The top panel shows SED-derived stellar masses after correcting the *H*- and *K*-band photometry for emission-line contamination (see Section 2.3.2) plotted against stellar masses derived from the original, non-corrected, photometry in which the *H* and *K* bands are included (open red circles) and excluded (filled black circles). The bottom panel shows the same for SED-derived star formation rates. In each panel, the grey solid line is the one-to-one relation and median error bars are shown in the top left-hand corner. The mean, and standard deviation, of the offset from the one-to-one relation is indicated in the legend. The true, emission-line-corrected values of $M_\star$ and SFR can be reasonably recovered by simply excluding the contaminated *H* and *K* bands are excluded from the fit.

2.4 The $z \sim 3.4$ NIRVANDELS sample

Our final sample was drawn from the 35 galaxies with spectroscopic redshifts in the range $3.0 \leq z \leq 3.8$ (Table 1). We identified one galaxy in which the presence of active galactic nucleus (AGN) ionization was indicated by strong emission from the high-ionization species N V $\lambda\lambda$1238, 1242, C IV $\lambda\lambda$1548, 1550, and He II λ1640 in the rest-frame UV spectrum. This galaxy was removed from our sample. No further AGN was identified based on rest-frame UV and optical spectral properties. Furthermore, for the remaining sample, we ruled out the presence of significant AGN ionization based on their mid-IR SED shapes and X-ray properties (see McLure et al. 2018). We also removed one galaxy in which the only optical emission feature detected was [O II] $\lambda\lambda$3726, 3729. There were two reasons for this decision. First, it is not possible to determine a gas-phase metallicity from the [O II] $\lambda\lambda$3726, 3729 doublet alone, and therefore this object could not be included in our individual galaxy analysis.

Secondly, in our rest-frame optical stacking analysis (described in Section 2.4.1) we required the detection of the [O III] λ5007 line in order to normalize the spectra, which was not available for this object.

Our final NIRVANDELS sample therefore consisted of 33 galaxies with secure spectroscopic redshifts in the range $2.95 \leq z \leq 3.8$. Each object in the sample has MOSFIRE spectra in the *H* and *K* bands covering a number of rest-frame optical emission lines sufficient for deriving gas-phase metallicity (Z_g), and a VIMOS/VANDELS rest-frame FUV spectrum from which the stellar metallicity ($Z_\star$) can be determined.

In Fig. 3, we show the normalized redshift and mass distributions of the NIRVANDELS galaxies compared to the full VANDELS sample of $N = 791$ star-forming galaxies in the redshift range $3.0 \leq z \leq 3.8$. The stellar masses of the full sample have been derived using the same SED fitting procedure described in Section 2.3.2. As the photometry for the full sample cannot be corrected for emission-line contamination, we have excluded the *H*- and *K*-band photometry from these fits for the reasons described in Section 2.3.4 (all objects benefit from longer-wavelength *Spitzer* IRAC data). The properties of the sample with respect to the full $z \simeq 3.4$ star-forming galaxy population are discussed further in Section 2.4.2.

2.4.1 Composite spectra

Our study is focused on determining stellar metallicities ($Z_\star$) and gas-phase metallicities (Z_g), and, wherever possible, we estimated these quantities on an individual galaxy basis. However, in order to include galaxies for which such measurements were not possible, we also made use of stacked spectra. In order for an object to be included in the stacking sample, we required coverage of all rest-frame optical emission lines in the MOSFIRE spectra ([O II] $\lambda\lambda$3726, 3729, [Ne III] λ3870, H β, and [O III] $\lambda\lambda$4959, 5007). In total, five objects were removed from the stacking sample due to fact that the H β line was not covered by the detector.[6] The final stacking sample therefore contained only 28/33 galaxies, and is fully representative of the complete sample.

In this paper, we focus on the correlation between metallicity and stellar mass, and we therefore constructed stacks in two bins of $M_\star$: high-$M_\star$ and low-$M_\star$, split at the median $M_\star$ of the stacking sample ($M_\star = 10^{9.4} M_\odot$). We employed the method described in Cullen et al. (2019) to produce FUV stacks from the VANDELS data and refer readers to that paper for further details. For the MOSFIRE spectra, we first converted each spectrum into luminosity density units using its spectroscopic redshift and corrected for nebular dust attenuation using $E(B - V)_{neb}$, assuming the Cardelli et al. (1989) Milky-Way extinction curve. To avoid biasing the stacks in favour of the brightest objects, we normalized each spectrum using its measured [O III] λ5007 luminosity.[7] The normalized spectra were then re-sampled on to a common rest-frame wavelength grid, and

---

[6]We note that these five galaxies were detected in [O III] λ5007 and [O II] $\lambda\lambda$3726, 3729, allowing Z_g to be determined on an individual basis using the [O III] λ5007/[O II] $\lambda\lambda$3726, 3729 ratio (see Section 3). However, including these objects in the stacks would clearly bias the Hβ-dependent line ratios and Hβ-derived average SFRs.

[7]We note that this choice has the potential to introduce subtle biases due to the exact details of the sample, and the metallicity dependence of the chosen line, etc. However, we have confirmed that this choice does not have a strong effect on the measured line ratios and derived gas-phase metallicities; combining the galaxies without this normalization does not change the results of this paper.
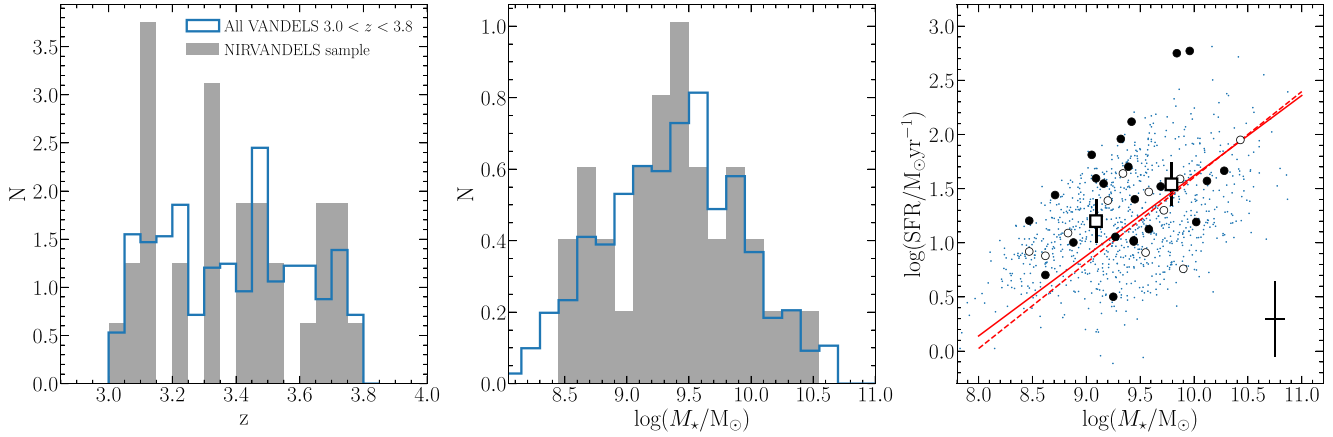
Figure 3. The normalized redshift distribution (left), normalized stellar mass distribution (centre), and SFR$-M_\star$ relation (right) for the star-forming galaxies in our NIRVANDELS sample. In the left-hand and centre panels, the open blue histogram represents all galaxies at $3.0 < z < 3.8$ in the VANDELS parent sample ($N = 980$) and the filled grey histogram shows the subset of galaxies with MOSFIRE follow-up analysed here ($N = 33$). In the right-hand panel, the black filled circular points are galaxies with a H β detection and therefore an SFR estimate based on the H β flux; the open circular points represent galaxies without a H β detection and therefore a SED-derived SFR estimate. The median uncertainty on $M_\star$ and SFR is displayed in the lower right corner. The square data points show the high-$M_\star$ and low-$M_\star$ stacks (see Section 2.4.1). The small background data points shown the full VANDELS sample. The solid and dashed red lines show two parametrizations of the SFR–$M_\star$relation at $z = 3.3$ derived by Sanders et al. (2020a) and Speagle et al. (2014), respectively.

the final stacked spectrum was produced by taking the median value at each wavelength. The error at each wavelength was calculated via a bootstrap re-sampling of the individual values. Finally, the normalized stacks were converted back into absolute luminosity density units by multiplying by the median [O III] $\lambda5007$ luminosity of the contributing galaxies.

Due to the fact that the optical emission-line profiles in the stacked spectra were not necessarily well-described by a single Gaussian profile, line luminosities for the stacks were determined via direct integration after subtracting off the local stellar continuum. The stacked H β line luminosities were corrected for stellar absorption using the median correction of the galaxies contributing to the stack. These correction factors were relatively small, at the $\simeq 3$ per cent level.

To estimate uncertainties on the various properties derived from the stacked spectra (line ratios/luminosities, $Z_\star$, Z_g), we employed a re-sampling methodology. The $M_\star$ values of galaxies in the stacking sample were perturbed according to their 1σ uncertainties and the high-$M_\star$ and low-$M_\star$ bins re-populated with replacement. The $E(B - V)_{neb}$ values for each galaxy were also perturbed according to their 1σ uncertainties. Finally, the 1D spectra were perturbed using their error spectra. The stacks were then re-constructed using the methodology described above and the various quantities of interest were remeasured. This process was repeated 500 times. Uncertainties on all quantities measured from stacked spectra were then calculated using the 16th and 84th percentiles of the resulting distributions.

2.4.2 A representative sample?

If the results of this study are to be applied in general to star-forming galaxies at $z \simeq 3.4$, it is necessary to demonstrate that our sample is not a highly-biased subset of the general galaxy population at this epoch. In particular, when considering the mass–metallicity relation, it is important to assess whether the galaxy sample is biased in terms of SFR at fixed $M_\star$. Numerous studies in the literature, both at $z = 0$ and higher redshifts, have found strong evidence for an

anticorrelation between SFR and metallicity at fixed $M_\star$ (i.e. the FMR; Mannucci et al. 2010; Sanders et al. 2018). Therefore, in order to avoid SFR biases, the galaxies in our sample should be representative of the 'star-forming main sequence' (SFR–$M_\star$relation) at $z \simeq 3.4$.

In the right-hand panel of Fig. 3, we show the location of the NIRVANDELS galaxies on the SFR–$M_\star$plane compared to two literature determinations of the SFR–$M_\star$relation at $z \simeq 3.4$ (Speagle et al. 2014; Sanders et al. 2020a). We also show the underlying distribution of all VANDELS galaxies at $2.95 \leq z \leq 3.80$ to illustrate the typical range of SFRs observed at fixed $M_\star$. Above $M_\star \simeq 10^{9.4} M_\odot$(the median stellar mass of our sample), the scatter of the individual NIRVANDELS galaxies about the main sequence is consistent with that of the underlying population and there is no evidence that our sample is strongly biased in terms of SFR at these stellar masses. Moreover, the high-$M_\star$ stack is consistent with the Speagle et al. (2014) main-sequence relation within 1σ. Therefore, for galaxies with $M_\star \gtrsim 10^{9.4} M_\odot$ (our high-$M_\star$ sample), we conclude that the results presented in this paper are applicable to the general $z \simeq 3.4$ star-forming population.

At $M_\star < 10^{9.4} M_\odot$, the individual NIRVANDELS galaxies predominantly fall above the literature relations. The low-$M_\star$ stack also sits above the main sequence by $\simeq 0.3$ dex. Although the stack is formally consistent with the Speagle et al. (2014) relation within 2σ, we clearly cannot rule out the possibility that our sample is biased towards high SFR at the lowest stellar masses. This in turn could lead to a bias in the MZR. Based on the results of Sanders et al. (2018) the Z_g bias at fixed $M_\star$ due to an offset from the main-sequence can be estimated as $\Delta Z_g \simeq -0.15 \times \Delta \log(\mathrm{SFR})$. An offset of 0.3 dex therefore represents a bias of $\simeq -0.05$ dex in Z_g.[8] We chose not to correct for this potential bias in this paper because (i) the magnitude of the offset is similar to the typical uncertainties on our individual Z_g measurements and (ii) the main sequence offset in itself is not highly

---

[8]We note that the offset is slightly larger ($\simeq -0.08$ dex) assuming another recent determination of the FMR presented in Curti et al. (2020).

significant ($<2\sigma$). Nevertheless, we caution that at $M_\star \lesssim 10^{9.4}$ M$_\odot$ our sample may include a small bias towards objects with elevated SFRs and lower metallicities.

3 DETERMINATION OF GALAXY METALLICITIES

The primary focus of this study is the determination of gas-phase metallicities (Z_g) and stellar metallicities ($Z_\star$) as a function of galaxy stellar mass at $z \simeq 3.4$. Below we describe in detail how each of these parameters was measured from the spectroscopic data.

Before going into the details of these methods, it is worth again emphasizing that Z_g and $Z_\star$, as measured in this paper, are not sensitive to the same element abundances (see also Section 1). Gas-phase metallicities, determined from rest-frame optical nebular emission lines, are sensitive to the elements that act to cool the $T \simeq 10^4$ K ionized gas, namely oxygen (O/H). On the other hand, stellar metallicities derived from rest-frame FUV spectra are sensitive to the elements that dominate the FUV opacity of massive stars, namely iron (Fe/H) (Leitherer et al. 2010). We will discuss the implications of this difference in detail in Section 4.3.

3.1 Determination of the gas metallicity (Z_g)

Estimates of the gas-phase metallicity (Z_g, or O/H) were derived using the ratios of rest-frame optical emission lines measured in the MOSFIRE spectra. In the redshift range investigated in this paper ($2.95 \le z \le 3.80$), the [O II] $\lambda\lambda3726$, 3729 and [Ne III] $\lambda3870$ lines are available in the H band, while the H β and [O III] $\lambda\lambda4959$, 5007 lines are available in the K band. We selected three independent metallicity diagnostics from these lines: [O III] $\lambda5007$/H β, [O III] $\lambda\lambda4959$, 5007/[O II] $\lambda\lambda3726$, 3729 (O_{32}), and [Ne III] $\lambda3870$/[O II] $\lambda\lambda3726$, 3729. To estimate Z_g from these ratios, we followed the method described in Sanders et al. (2020a) using the high-redshift analogue calibrations of Bian, Kewley & Dopita (2018). These calibrations are known to be consistent with the (currently small) sample of $z \simeq 2$ galaxies for which direct-method O/H estimates are available (Sanders et al. 2020b). Using these calibrations, we calculated the best-fitting value of Z_g via a χ^2 minimization using

$$\chi^2(x) = \sum_i \frac{(R_{\mathrm{obs},i} - R_{\mathrm{cal},i}(x))^2}{(\sigma^2_{\mathrm{obs},i} + \sigma^2_{\mathrm{cal},i})}, \qquad (2)$$

where $x = 12 + \log(\mathrm{O/H}) = \log(Z_g/Z_\odot) + 8.69$, the sum over i represents the line ratios used, $R_{\mathrm{obs},i}$, is the logarithm of the ith observed line ratio, $R_{\mathrm{cal},i}(x)$ is the predicted value of R_i at x from the Bian et al. (2018) calibrations, $\sigma_{\mathrm{obs},i}$ is the uncertainty on $R_{\mathrm{obs},i}$, and $\sigma_{\mathrm{cal},i}$ is the calibration uncertainty. The best-fitting Z_g solution is found by minimizing the value of χ^2 in equation (2). Uncertainties were estimated by perturbing the observed line ratios by their 1σ error values and recalculating Z_g 500 times. The 1σ uncertainty on Z_g was derived from the 68th percentile width of the resulting distribution.

In practice, not all lines were detected for a given object. In this case, the minimum requirement for a metallicity estimate was the detection of the [O III] $\lambda5007$ and [O II] $\lambda\lambda3726$, 3729 lines. It is possible to estimate Z_g using only the O_{32} ratio because the $O_{32} - Z_g$ calibration is monotonic. The [O III] $\lambda5007$/H β calibration, on the other hand, is double valued and therefore cannot be used without another ratio to break the degeneracy. The

Table 2. Properties of the individual galaxies in the NIRVANDELS sample, including the derived stellar and gas-phase metallicities.

Name	z_{neb}	$\log(M_\star/\mathrm{M}_\odot)$	$\log(Z_g/Z_\odot)$	$\log(Z_\star/Z_\odot)$
NIRV-25444	3.7019	9.87	$-0.29^{+0.26}_{-0.22}$	–
NIRV-25568	3.2109	10.12	–	–
NIRV-25732	3.1447	8.83	$-0.49^{+0.12}_{-0.13}$	–
NIRV-28864	3.5172	9.55	–	–
NIRV-29419	3.7048	10.28	$-0.24^{+0.15}_{-0.19}$	–
NIRV-30119	3.7713	8.88	$-0.68^{+0.09}_{-0.09}$	–
NIRV-30602	2.9832	9.72	$-0.28^{+0.22}_{-0.21}$	–
NIRV-30845	3.1452	9.69	$-0.43^{+0.21}_{-0.24}$	–
NIRV-31538	3.3131	10.02	–	–
NIRV-31982	3.4702	9.44	$-0.45^{+0.13}_{-0.10}$	–
NIRV-33568	3.3085	8.47	$-0.77^{+0.06}_{-0.08}$	–
NIRV-33613	3.7076	9.58	$-0.36^{+0.24}_{-0.22}$	–
NIRV-33644	3.2049	9.32	$-0.44^{+0.15}_{-0.12}$	$-0.93^{+0.22}_{-0.12}$
NIRV-34030	3.0750	9.90	–	–
NIRV-34438	3.1383	9.09	–	$-0.93^{+0.36}_{-0.20}$
NIRV-34449	3.4714	9.58	$-0.45^{+0.23}_{-0.26}$	–
NIRV-34500	3.6008	9.42	$-0.44^{+0.17}_{-0.19}$	–
NIRV-34591	3.3229	9.25	–	–
NIRV-34777	3.4042	9.39	$-0.59^{+0.18}_{-0.16}$	$-0.73^{+0.26}_{-0.31}$
NIRV-34889	3.6712	9.96	$-0.01^{+0.16}_{-0.19}$	–
NIRV-35212	3.4024	9.84	$-0.38^{+0.17}_{-0.19}$	$-0.54^{+0.10}_{-0.10}$
NIRV-35557	3.3470	9.27	$-0.47^{+0.09}_{-0.07}$	–
NIRV-35865	3.4017	10.43	–	–
NIRV-35915	3.1159	8.62	–	–
NIRV-36951	3.0668	8.71	$-0.56^{+0.06}_{-0.06}$	–
NIRV-37402	3.1360	8.47	–	–
NIRV-38315	3.4986	9.05	$-0.72^{+0.09}_{-0.09}$	–
NIRV-38451	3.5091	9.16	$-0.55^{+0.08}_{-0.09}$	–
NIRV-41105	3.0039	8.62	–	–
NIRV-42161	3.1113	9.34	–	$-0.77^{+0.21}_{-0.18}$
NIRV-44246	3.6770	9.44	$-0.62^{+0.09}_{-0.08}$	–
NIRV-45896	3.6747	9.20	–	–
NIRV-46857	3.3490	9.45	$-0.44^{+0.16}_{-0.16}$	$-0.72^{+0.18}_{-0.18}$

[Ne III] $\lambda3870$/[O II] $\lambda\lambda3726$, 3729 ratio is also monotonic and could, in principle, be used to infer metallicity (e.g. Shapley et al. 2017) but in practice there were no instances in which this was the only line ratio available.[9]

For each galaxy, we utilized the maximum number of line ratios possible, which resulted in 6/33 metallicities determined using all three line ratios, 11/33 using O_{32} and [O III] $\lambda5007$/H β, and 4/33 using O_{32} alone. Using more line ratios improves the constraints on Z_g but does not bias the solution on average (Sanders et al. 2020a). The remaining 12/33 galaxies do not have individual Z_g determinations; in all cases, this was due to a non-detection of the [O II] $\lambda\lambda3726$, 3729 line. For both the low-$M_\star$ and high-$M_\star$ composite spectra, all lines were detected. The gas-phase metallicities estimates resulting from this procedure are listed in Tables 2 and 3. Examples of the nebular emission-line spectra for individual objects are shown Fig. 4 and a comparison between the two composite spectra are shown in Fig. 5.

[9]This is due to the fact that [Ne III] $\lambda3870$ is the faintest line considered in this paper. All galaxies with a [Ne III] $\lambda3870$ detection were also detected in [O III] $\lambda5007$ and [O II] $\lambda\lambda3726$, 3729.

Table 3. Properties of the low-$M_\star$ and high-$M_\star$ composites.

Stack	log($M_\star$/M$_\odot$) range	Median log($M_\star$/M$_\odot$)	[O III]/Hβ	[O III]/[O II]	[Ne III]/[O II]	log(Z_g/Z$_\odot$)	log($Z_\star$/Z$_\odot$)
low-$M_\star$	8.50−9.40	9.09	0.88 ± 0.06	0.55 ± 0.10	−0.64 ± 0.15	−0.54 ± 0.06	−0.89 ± 0.06
high-$M_\star$	9.40−10.40	9.81	0.58 ± 0.06	0.22 ± 0.09	−0.76 ± 0.19	−0.31 ± 0.06	−0.72 ± 0.07

3.2 Determination of the stellar metallicity ($Z_\star$)

Stellar metallicities ($Z_\star$ or Fe/H) were estimated from the rest-frame FUV spectra using a full spectral fitting technique within the wavelength range 1221−2000 Å as described in Cullen et al. (2019, 2020). The method involves fitting stellar population synthesis models to all portions of the FUV spectra dominated by stellar continuum emission (avoiding regions contaminated by ISM absorption lines and/or nebular emission lines). The metallicity of the model ($Z_\star$) is constrained, while marginalizing over three nuisance parameters related to dust attenuation. Below we briefly describe this technique but refer readers to Cullen et al. (2019) for full details.

For the stellar population models, we adopted the Starburst99 (SB99) high-resolution WM-Basic models described in Leitherer et al. (2010). To construct the models we assumed constant star formation over timescales of 100 Myr and adopted the weaker-wind Geneva tracks with stellar rotation and single-star evolution at the following metallicities: $Z_\star$ = (0.001, 0.002, 0.008, 0.014, 0.040). The models were fitted using the nested sampling algorithm implemented in the python package `dynesty` (Speagle 2020). The free parameters in the fit were $Z_\star$ and three nuisance parameters that define the shape of the FUV attenuation curve using the parametrization described in Salim, Boquien & Lee (2018). As the models are provided at five fixed $Z_\star$ values, we linearly interpolated the logarithmic flux values between the models in order to generate a model for any $Z_\star$ value within the prescribed range. The models were then convolved to the resolution of the VANDELS spectra and appropriately re-sampled.

We used a log-likelihood function of the form,

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

$$= K - \frac{1}{2}\chi^2, \tag{3}$$

where K is a constant, 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 likelihood was computed using only those wavelength pixels free from ISM absorption or nebular emission-line contamination. For this purpose, we adopted the 'Mask 1' windows defined in Steidel et al. (2016). The `dynesty` algorithm provides estimates of the posterior probability distributions for each of the free parameters in the fit. For a given fit, the best-fitting $Z_\star$ value was calculated from the 50th percentile of the resulting $Z_\star$ posterior distribution and the uncertainty derived from the 68th percentile width.

Deriving stellar metallicities for individual galaxies is generally more difficult than deriving gas-phase metallicities due to the requirement for a high S/N detection of the continuum. In this paper, we only report $Z_\star$ values for the six individual galaxies for which the average S/N per resolution element of the VANDELS spectrum is ≥5 in the relevant wavelength range. At lower S/N, we find that the typical 1σ uncertainties on $Z_\star$ are > 50 per cent. Moreover, below a certain threshold S/N estimates of $Z_\star$ can become biased. For example, Topping et al. (2020b) find that unbiased estimates of $Z_\star$ require S/N ≥5.6 per resolution element (five of the six galaxies

with individual $Z_\star$ estimates satisfy this slightly higher threshold). Stellar metallicity estimates resulting from this procedure are listed in Tables 2 and 3, and examples of fits to the individual and stacked FUV spectra are shown in Figs 4 and 6. In Fig. 6, we also show a zoom-in of continuum-normalized versions of the stacks in the region of the '1719' index, recently advocated by Calabrò et al. (2021) as a useful $Z_\star$-sensitive index in the FUV. In agreement with the results of Calabrò et al. (2021), we find that the higher metallicity, high-$M_\star$, stack shows increased absorption in the '1719' index region.

4 RESULTS

4.1 The gas-phase mass–metallicity relation at $z \simeq 3.4$

4.1.1 Trends with emission-line ratios

Since the gas-phase metallicity is estimated directly from emission-line ratios, we first examine the empirical trends between the observed dust-corrected line ratios, stellar mass and redshift. In Fig. 7, we show the three emission-line ratios used to estimate Z_g ([O III]/Hβ, O$_{32}$ and [Ne III]/[O II]) as a function of $M_\star$ for our sample. Also shown is an independent data set of star-forming galaxies at $z \simeq 3.3$ from Sanders et al. (2020a), and a sample of local star-forming galaxies drawn from the Sloan Digital Sky Survey (SDSS). The SDSS galaxies are taken from the Andrews & Martini (2013) sample as described in Sanders et al. (2020a). To facilitate a direct comparison with our $z \simeq 3.4$ galaxies, the SDSS line ratios have been corrected, where possible, for diffuse ionized gas (DIG) emission.[10] DIG is not associated with H II regions and therefore biases nebular emission line flux measurements in the integrated spectra of galaxies at $z \simeq 0$ (Oey et al. 2007; Sanders et al. 2017; Vale Asari et al. 2019). A DIG correction is not required for the $z \simeq 3.4$ sample because the DIG contribution becomes negligible at the typical star-formation surface densities of high-redshift galaxies (Sanders et al. 2017; Shapley et al. 2019).

Fig. 7 clearly demonstrates that each of the three line ratios increases with decreasing stellar mass. The negative correlation between line ratio and $M_\star$ applies to both the galaxies at $z \simeq 3.4$ and the local SDSS sample. Based on the assumption that the galaxies lie on the upper-branch of the [O III]/Hβ–Z_g relation,[11] these trends indicate that Z_g increases with increasing $M_\star$. Crucially, this correlation holds irrespective of exactly which emission-line calibration is adopted (e.g. Maiolino et al. 2008; Bian et al. 2018; Kewley et al. 2019; Curti et al. 2020). For our $z \simeq 3.4$ sample, the trend is clearly evident in the [O III]/Hβ–$M_\star$ and O$_{32}$–$M_\star$ diagrams for both the individual galaxies and stacks, and is fully consistent with

---

[10]For the [O III]/Hβ and [O III]/[O II] ratios, the SDSS data are DIG corrected; however, this is not possible for [Ne III]/[O II] due to the lack of an accurate DIG-correction for the [Ne III] line (Sanders et al. 2020a).

[11]We can assert this with confidence based on the observed [O III]/[O II] ratios, which rule out galaxies lying on the lower branch of the [O III]/Hβ–Z_g relation assuming the Bian et al. (2018) calibration is applicable to our sample (see also Curti et al. 2020).
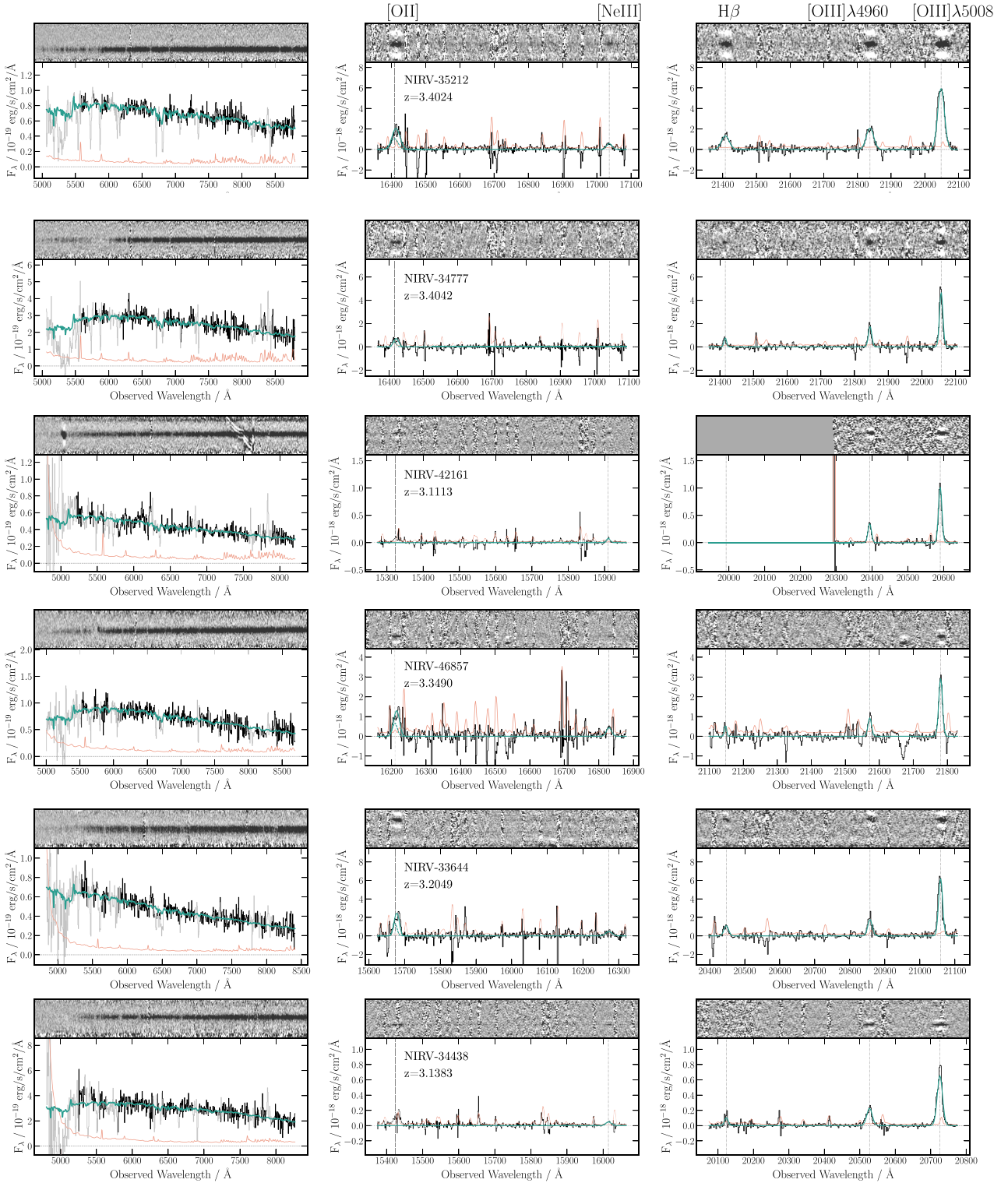
Figure 4. From left to right: VANDELS rest-frame FUV spectrum, MOSFIRE *H*-band spectrum, and MOSFIRE *K*-band spectrum for the $N = 6/33$ individual galaxies in our sample for which were able to estimate $Z_\star$. These galaxies have a median SNR per resolution element $>=5$ in the VANDELS rest-frame FUV spectrum. In each panel, both 1D and 2D spectra are shown in the lower and upper portion of the panel, respectively. For the 1D spectra, the observed spectrum is shown in black, with the best-fitting model overplotted in blue and the error spectrum shown in orange. For the VANDELS rest-frame FUV 1D spectra, regions of the spectrum dominated by ISM absorption lines/nebular emission lines are masked out (lighter shading) as these wavelength regions are not included in the stellar continuum fits used to determine $Z_\star$. In the MOSFIRE panels, dotted vertical lines indicate the positions of the nebular emission lines used to determine Z_g ([O II], [Ne III], H β, [O III]); the line labels are given at the top of the figure.
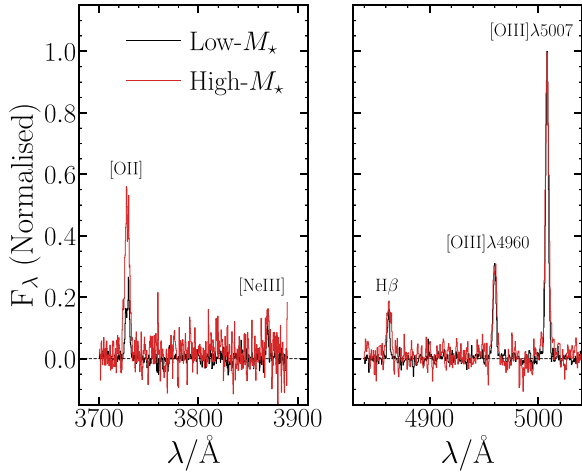
Figure 5. Portions of the MOSFIRE composite spectra in the *H* band (left) and *K* band (right), covering the [O II], [Ne III], H β, and [O III] nebular emission lines. The low-$M_\star$ composite is shown in black and the high-$M_\star$ stack in red. Both composites have been normalized by the peak [O III] $\lambda 5007$ flux of the low-$M_\star$ composite spectrum.

the Sanders et al. (2020a) data. The consistency with the Sanders et al. (2020a) line ratios, which are based on a much larger sample of $N = 245$ galaxies drawn from the star-forming main sequence, provides further evidence that our sample is not a highly biased subset of the star-forming galaxy population at $z \simeq 3.4$. As an aside, we also note that the position of our sample in the [O III]/Hβ–$M_\star$ plane overlaps with the star-forming galaxy region of the Mass-Excitation (MEx) diagnostic diagram proposed by Juneau et al. (2014), providing further evidence that our sample is not strongly contaminated by AGN excitation.

For the [Ne III]/[O II] ratio, the situation is less clear due to the difficulty in detecting the faint [Ne III] line. Nevertheless, for the small number of individual detections, a trend is apparent that is consistent with the slope and normalization of the Sanders et al. (2020a) data. Moreover, the observed [Ne III]/[O II] ratio of the $M_\star$ stacks are also consistent with the Sanders et al. (2020a) $z \simeq 3.4$ relation within the uncertainties. We note that the reason the individual detections lie predominantly above the stacked value is due to selection effects (i.e. the stacks also contain objects that are not individually detected in [Ne III] and/or [O II]). This selection effect is also present in the other two line ratio diagrams, but does not affect the general trends described above.

It can also be seen from Fig. 7 that all of the $z \simeq 3.4$ line ratios are elevated with respect to SDSS galaxies at fixed $M_\star$. Again, this empirical trend strongly suggests an evolution towards lower Z_g at higher redshift, at all stellar masses. However, when interpreting the redshift evolution of emission lines, the known evolution towards more extreme H II region conditions at high redshift must also be considered (e.g. Steidel et al. 2014; Shapley et al. 2015). Most recent results, as well as the results presented in this paper, suggest that the primary cause of this evolution is the harder ionizing spectra emitted by oxygen-enhanced, low-metallicity (i.e. iron-poor), stars at high redshift (e.g. Steidel et al. 2016; Strom et al. 2017; Topping et al. 2020b, see Section 4.3 for further discussion). As a consequence, not all of the observed redshift evolution in line ratios can be attributed purely to changes in Z_g (e.g. Cullen et al. 2016). Nevertheless, these empirical line-ratio diagrams provide useful qualitative indications of the likely evolution of Z_g with $M_\star$ and redshift, which we discuss quantitatively below.

4.1.2 *The* $M_\star$–Z_g *relation*

The NIRVANDELS $z \simeq 3.4$ gas-phase mass–metallicity relationship (MZR$_g$) is shown in Fig. 8. It can be seen that the NIRVANDELS galaxies follow a clear MZR$_g$, evident for both the individual galaxies and the stacks. The difference in $\log(Z_g)$ between the high-$M_\star$ and low-$M_\star$ stacks is 0.25 ± 0.08, representing an increase in Z_g of 1.78 ± 0.35 across $\simeq 1$ dex in stellar mass ($\simeq 10^9$–10^{10} M$_\odot$). It is also clear from Fig. 8 that the NIRVANDELS sample is fully consistent with the Sanders et al. (2020a) MZR$_g$ determination at $z \simeq 3.3$. This is unsurprising given the similar $M_\star$-dependence of the emission-line ratios illustrated in Fig. 7. Given this excellent agreement, we do not attempt to refit our new data here. The functional form of the Sanders et al. (2020a) MZR$_g$ (converted into units of $\log(Z_g/Z_\odot)^{12}$) is

$$\log(Z_g/Z_\odot) = (0.29 \pm 0.02)m_{10} - (0.28 \pm 0.03), \qquad (4)$$

where $m_{10} = \log(M_\star/10^{10}$ M$_\odot)$.

It can also be seen from Fig. 8 that the slope of the $z \simeq 3.4$ MZR$_g$ is also fully consistent with the low-mass slope at $z \simeq 0$ (0.28 ± 0.01; Sanders et al. 2020a). As a result, there is a constant offset of $\simeq -0.36$ dex in $\log(Z_g)$ (a factor of 0.44 in Z_g) between $z = 0$–3.4 in the relevant mass range ($\simeq 10^{8.5} - 10^{10.5}$ M$_\odot$). We note that the size of this offset does critically depend on the Z_g calibration used; to this end, the $z \simeq 0$ relation of Sanders et al. (2020a) was derived using a local calibration that should be non-biased with respect to Z_g measurements at $z > 1$ derived from the Bian et al. (2018) calibration, effectively accounting for the harder ionizing radiation field at high redshifts discussed above in Section 4.1.1, and later in Section 4.3 (see also Sanders et al. (2020a) for full details). Finally we note that, although not shown in Fig. 8, a number of other previous studies at $z > 3$ have derived $M_\star$–Z_g relations with a lower overall normalization due to the use of a different metallicity calibration (e.g. Maiolino et al. 2008; Mannucci et al. 2009; Troncoso et al. 2014; Onodera et al. 2016, see Sanders et al. 2020a for a full discussion); crucially, however, the slope of these other literature relations are generally consistent with the results presented here, indicating that the scaling of metallicity with stellar mass is independent of the chosen calibration.

### 4.2 The stellar mass–metallicity relation at $z \simeq 3.4$

The NIRVANDELS $z \simeq 3.4$ stellar mass–metallicity relationship (MZR$_\star$) is also shown in Fig. 8. Again, the individual galaxies and stacks appear to follow a clear MZR$_\star$. The evolution in $\log(Z_\star)$ between the low-$M_\star$ and high-$M_\star$ stack is 0.18 ± 0.10 dex (a factor of 1.51 ± 0.35). In contrast to MZR$_g$, the trend for individual objects is less clear due to the fact that the error bars on the individual measurements are large, and the number of objects much reduced. Nevertheless, these individual measurements are formally consistent with the trend observed in the stacks.

Also shown in Fig. 8 is a determination of MZR$_\star$ derived from the much larger sample of $N = 681$ VANDELS galaxies presented in (Cullen et al. 2019). The $M_\star$ values in Cullen et al. (2019) were derived using *H* - and *K* - band photometry that had not been corrected for optical emission line contamination, and used a set of SED-fitting assumptions different to the ones that are assumed in this paper. Therefore, the relation shown in Fig. 8 is a re-derivation of the Cullen et al. (2019) MZR$_\star$ with $M_\star$ values derived excluding *H*- and *K*-band photometry and using the same

---

[12] The conversion between $\log(Z_g/Z_\odot)$ and the widely used $12 + \log(O/H)$ is simply $12 + \log(O/H) = \log(Z_g/Z_\odot) + 8.69$.
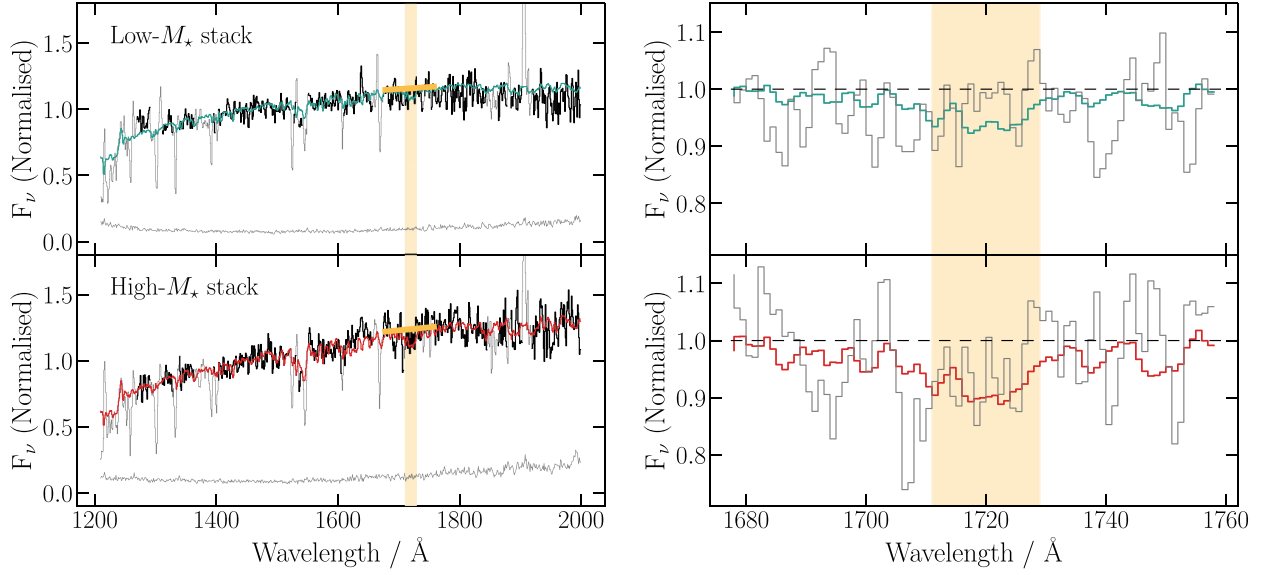
Figure 6. Stacked FUV composite spectra for the low-$M_\star$ (top panels) and high-$M_\star$ (bottom panels) sample. The left-hand panels show the full FUV spectrum in the range 1200–2000 Å, normalized by the median flux in the range 1420–1480 Å. Bootstrapped error spectra are shown in grey. The best-fitting Starburst99 models are superposed on top of the composites (in turquoise and red for the low-$M_\star$ the high-$M_\star$ composites, respectively). Regions of the composite spectra in a lighter shade highlight the position of strong interstellar absorption features or nebular emission lines that are not included in the model fitting. The yellow shaded region shows the position of the $Z_\star$-sensitive '1719' index (e.g. Calabrò et al. 2021). The right-hand panels show continuum-normalized versions of both spectra in the region of the '1719' index. The continuum normalization was performed using the surrounding Rix et al. (2004) windows and the level of the continuum is shown is shown by the thick yellow line in both of the left-hand panels. It can be seen that the high-$M_\star$ composite shows stronger absorption across the '1719' index, indicating higher $Z_\star$ (Calabrò et al. 2021).
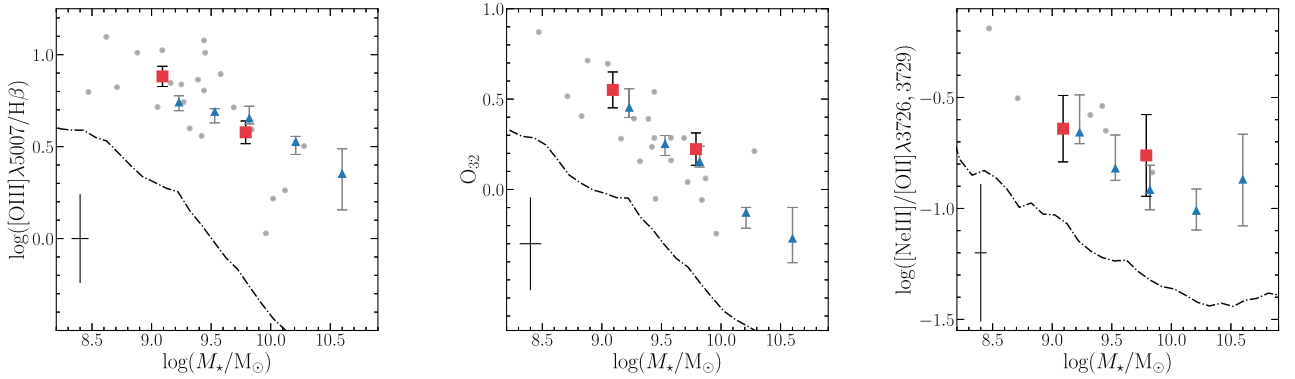
Figure 7. Dust-corrected emission lines ratios versus $M_\star$ for galaxies at $z \simeq 3.4$ and $z \simeq 0$. Each panel shows one of the three emission-line ratios used to determine Z_g in this paper (see Section 3). From left to right, these line ratios are [O III]/Hβ, O_{32}, and [Ne III]/[O II]. The red squares show the low-$M_\star$ and high-$M_\star$ NIRVANDELS stacks (see also Fig. 5) and the grey circles show the individual NIRVANDELS galaxies with the necessary line detections. The median error bar on the individual detections is shown in the bottom left-hand corner. The blue triangular data points show stacked measurements from the $z \simeq 3.3$ MOSDEF sample presented in Sanders et al. (2020a). The dot–dashed lines represent the running medians for $z \simeq 0$ galaxies from the SDSS sample of Andrews & Martini (2013). These $z \simeq 0$ line ratios have been corrected, where possible, for diffuse ionized gas (DIG) contamination that can bias line flux measurements in local galaxies [see the text and Sanders et al. (2020a) for further details].

methodology described in Section 2.3.2. This relation is in good agreement with our NIRVANDELS data, and is essentially consistent with the original relation shifted to lower $M_\star$. The functional form of this MZR$_\star$ is given by

$$\log(Z_\star/Z_\odot) = (0.27 \pm 0.06)m_{10} - (0.68 \pm 0.04), \quad (5)$$

where $m_{10} = \log(M_\star/10^{10}\,\mathrm{M_\odot})$.

There are two striking aspects to the comparison between MZR$_\star$ and MZR$_g$ in Fig. 8. First, MZR$_\star$ is offset to lower values at all stellar masses by $\simeq -0.4$ dex. This is a result of the fact that Z_g

traces O/H, while $Z_\star$ traces Fe/H and will be discussed in detail in Section 4.3. Secondly, the slopes of both relations are consistent: $\mathrm{d}(\log Z)/\mathrm{d}(\log M) \simeq 0.3$. Moreover, we note that the low-mass slope of the MZR$_\star$ measured in local star-forming galaxies is $\simeq 0.38$ (Zahid et al. 2017).[13] This result, combined with the results discussed in Section 4.1.2, imply that the power-law slope of the $M_\star-Z$ relation

---

[13] This value was calculated by fitting a relation of the form given by equation (4) to the Zahid et al. (2017) data below $10^{10}\,\mathrm{M_\odot}$ (i.e. data below the high mass turnover).
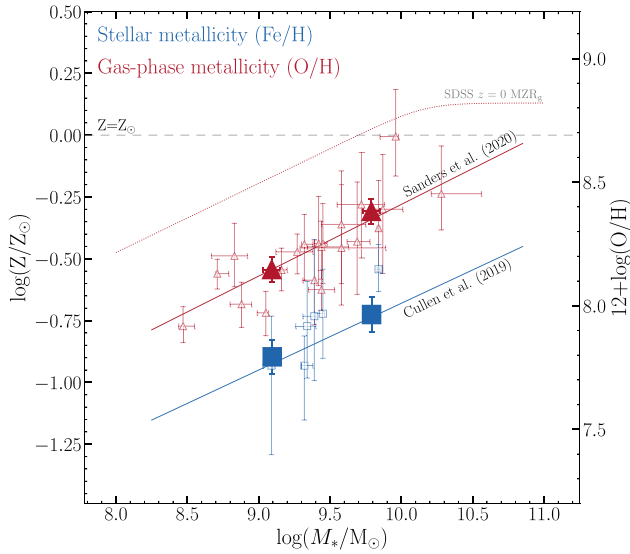
Figure 8. The mass–metallicity relation for stars ($Z_\star$, blue squares) and nebular gas (Z_g, red triangles) within the NIRVANDELS galaxies at $z \simeq 3.4$. The small open data points show $Z_\star$ and Z_g for individual galaxies, and the large filled data points show the low-$M_\star$ and high-$M_\star$ stacks. The blue and red solid lines show, respectively, previous determinations of MZR$_\star$ and MZR$_g$ at $z \simeq 3.4$ from Cullen et al. (2019) and Sanders et al. (2020a). The Cullen et al. (2019) MZR$_\star$ has been re-calculated to account for shifts in $M_\star$ as described in the text (Section 4.1.2). The grey dotted line shows the DIG-corrected MZR$_g$ at $z \simeq 0$ from Sanders et al. (2020a), and the horizontal dashed grey line indicates the value of solar metallicity on the y-axis. On the right-hand side of the y-axis, we show the value of log(O/H) + 12 for the Z_g data (the conversion is given by $12 + \log(O/H) = \log(Z_\star/Z_\odot) + 8.69$).

is very similar for both massive stars and H II region gas, and is also not strongly redshift dependent. This in turn would suggest that the physical processes responsible for determining the power-law slope of the $M_\star - Z$ relations below the turnover do not evolve strongly, out to at least $z \simeq 4$ (a lookback time of $\simeq 12$ Gyr).

4.3 Enhanced O/Fe ratios at high redshift

As derived in this paper, $Z_\star$ traces Fe/H in massive stars while Z_g traces O/H in ionized gas (see Section 3). Crucially, the two quantities are linked via the fact that ionized gas exists in the vicinity of massive stars, and is therefore likely to have very similar chemical abundance properties. Indeed, due to the short lifetimes of O- and B-type stars, the ionized gas is probably the remnant of their original stellar birth clouds. In this sense, O/H in ionized gas should also be representative of O/H in massive stars. Therefore, to a reasonable approximation, combining optically-derived estimates of Z_g with FUV-based estimates of $Z_\star$ traces the O/Fe abundance ratio of the massive stellar populations in star-forming galaxies (Steidel et al. 2016; Topping et al. 2020b). More generally, because O is an α-process element, the observed O/Fe-enhancement traces the enhancement of α elements, and is also referred to as α-enhancement.

If the NIRVANDELS galaxies have solar-like abundance ratios, we would expect that $Z_g = Z_\star$. However, it is clear from Fig. 8 that $Z_g > Z_\star$ across the full range of $M_\star$. That is, the O abundances of the NIRVANDELS galaxies are consistently larger than the Fe abundances, implying enhanced O/Fe ratios relative to the solar value. The values of $\log(Z_g/Z_\odot) - \log(Z_\star/Z_\odot)$ for the low-$M_\star$ and high-$M_\star$ stacks are 0.35 ± 0.09 and 0.46 ± 0.09, respectively. These
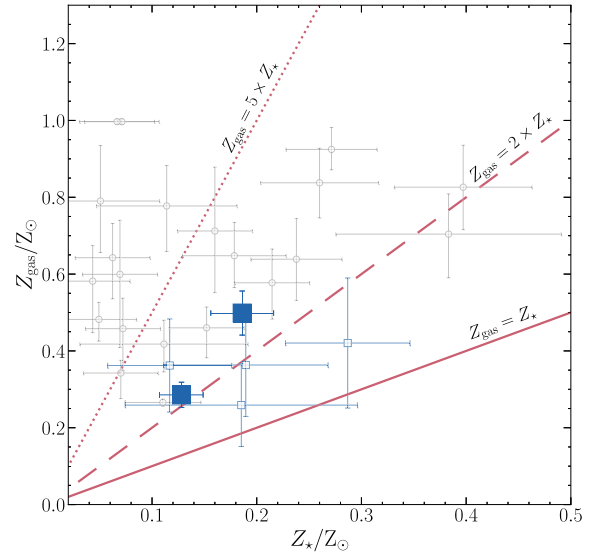
Figure 9. Z_g versus $Z_\star$ for star-forming galaxies at $z = 2-3$. Small square data points show the individual galaxies in our sample with a measurement of both quantities. Large filled square data points represent the low-$M_\star$ and high-$M_\star$ stacks. The grey circular data points show the data for star-forming galaxies at $z \simeq 2.3$ presented in Topping et al. (2020b). The various lines show constant $Z_g/Z_\star$ ratios as indicated by the labels, with the solid line representing the one-to-one relation.

values are both highly significant in themselves ($\approx 4-5\sigma$), and are consistent with each other within 1.2σ, with an average value of 0.41 ± 0.06. Assuming the average value of the two stacks, we find that the offset translates into an enhancement of O/Fe relative to the solar value of (O/Fe) $\simeq 2.57 \pm 0.38 \times$ (O/Fe)$_\odot$. Within the uncertainties, our results indicate that this offset is constant as a function of $M_\star$ between 10^9 and 10^{10} M$_\odot$.

This result is clarified further in Fig. 9 where we plot, on a linear scale, $Z_\star$ against Z_g (in units of solar metallicity). It is clear from Fig. 9 that both stacks sit well above the one-to-one relation that would indicate a solar O/Fe ratio (red solid line). Moreover, each of the four individual galaxies with a measurement of both $Z_\star$ and Z_g also sit above the one-to-one relation. Given the large uncertainties the individual results are not highly significant ($<2\sigma$); nevertheless, based on the more robust stacked measurements, Figs 8 and 9 present strong evidence for enhanced O/Fe ratios in typical star-forming galaxies at $z > 3$ with $M_\star \lesssim 10^{10}$ M$_\odot$.

We note that the evidence presented here for O/Fe-enhancement is consistent with other findings in the literature at slightly lower redshift ($z \simeq 2.4$, see below), although our results represent the first direct demonstration that (i) O/Fe-enhancement exists at $z > 3$ and (ii) is not strongly dependent on stellar mass below 10^{10} M$_\odot$.

4.3.1 Comparison with the literature

The most directly comparable works in the literature are the analyses of Steidel et al. (2016) and Topping et al. (2020a, b) both of whom performed a similar joint FUV + optical spectral analysis to the one presented in this paper. Steidel et al. (2016) analysed a stack of 30 star-forming galaxies at $z \simeq 2.4$ with $M_\star = 10^{9.8}$ M$_\odot$, finding (O/Fe) $\simeq 4-5 \times$ (O/Fe)$_\odot$, while Topping et al. (2020a, Topping et al. 2020b) analysed 30 individual galaxies at $z \simeq 2.3$ finding O/Fe-enhancements ranging from $1.8 \times$ (O/Fe)$_\odot$ to $>5 \times$ (O/Fe)$_\odot$. A direct comparison to the Topping et al. (2020b) results is shown

in Fig. 9 where it can be seen that the basic result (i.e. $Z_g > Z_\star$) is fully consistent with the result presented here. However, the median enhancement in the Topping et al. (2020b) sample ($\approx 4.5 \times$ (O/Fe)$_\odot$) is clearly larger than the value we derive ($\approx 2.5 \times$ (O/Fe)$_\odot$). We discuss the main reason for this difference in more detail in the following below.

In addition to the direct method presented here, other indirect methods based on reconciling photoionization models with observed optical emission-line ratios have been used to investigate O/Fe-enhancement at high redshift. In these methods, $Z_\star$ is constrained mapping Fe/H to the shape of the ionizing continuum spectra used to generate the predicted emission line ratios. Using such an indirect approach, Strom et al. (2018) found (O/Fe) $\simeq 2.6 \times$ (O/Fe)$_\odot$ for a sample of 150 main-sequence star-forming galaxies at $z \simeq 2.3$. Similarly, Sanders et al. (2020b) also found evidence for significant O/Fe-enhancement in a small sample of extreme emission line galaxies with direct temperature-based Z_g estimates at $z = 1 - 4$.

We note that although the precise value of the enhancement varies between studies – due to a combination of selection effects, and systematics related to measuring both Z_g and $Z_\star$ – the basic result remains consistent. That is, star-forming galaxies at high redshift are enhanced in O/Fe, with (O/Fe) $\gtrsim 2 \times$ (O/Fe)$_\odot$. Moreover, as discussed below, there is no known systematic that can easily explain away this result.

4.3.2 A discussion of systematic effects

The observed (O/Fe)-enhancement is subject to systematic uncertainties affecting the measurement of both Z_g and $Z_\star$. Sytematics related to the measurement of Z_g are well-known and have been discussed extensively in the literature (e.g. Kewley & Ellison 2008). The Bian et al. (2018) calibration used in this paper is arguably the most reliable calibration for galaxies at high redshift, having been built using local analogues of high-redshift galaxies; furthermore, it is known to be consistent with the (currently small) sample of $z \simeq 2$ galaxies for which direct-method O/H estimates are available (Sanders et al. 2020b). The Bian et al. (2018) calibration returns systematically larger estimates of Z_g compared to the commonly-adopted calibration of Maiolino et al. (2008), which has typically been the calibration of choice when determining Z_g at $z > 3$ (Mannucci et al. 2009; Troncoso et al. 2014; Onodera et al. 2016). However, even assuming the Maiolino et al. (2008) calibration, the Z_g values reported here only decrease by a factor $\simeq 1.4$, resulting in (O/Fe) $\simeq 1.8 \times$ (O/Fe)$_\odot$. We find similar offset using the updated version of these calibrations described in Curti et al. (2017). In addition, we find that another commonly-used calibration based on the [O III], [O II] and H β lines – the calibration of Kobulnicky & Kewley (2004) – returns values consistent with Bian et al. (2018) for our sample. In summary, all standard high-redshift Z_g calibrations produce estimates of Z_g that imply (O/Fe) > (O/Fe)$_\odot$.

Another crucial systematic related to Z_g is the choice of fundamental abundance scale. The Bian et al. (2018) calibration (and indeed the vast majority of empirical Z_g calibrations) is tied to temperature-based metallicities measured from auroral lines (often referred to as 'direct' method metallicities). However, there is a well-known discrepancy between Z_g determined via the direct method and Z_g measured using an alternative method based on oxygen recombination lines (the RL method). Estimates of Z_g using the RL method are typically $\simeq 0.24$ dex higher than Z_g estimated using the direct method, with this so-called abundance discrepancy factor

(ADF) being roughly constant at all metallicities (Peimbert 1967; Esteban et al. 2014). It is worth nothing that the results of Steidel et al. (2016) discussed above are tied to the RL abundance scale, and this in part explains their larger (O/Fe) estimate. Reasons for favouring the RL abundance scale are discussed in detail in Steidel et al. (2016) and Strom et al. (2017). Again, however, accounting for the ADF would only serve to increase Z_g (by $\simeq$ factor of 2) and therefore strengthen the observed (O/Fe)-enhancement in our sample.

With respect to $Z_\star$, the major source of systematic uncertainty is the choice of SPS model used to fit to the observed stellar continuum. Along with the SB99 models used in this paper, the Binary Population and Spectra Synthesis (BPASS) models (Eldridge et al. 2017; Stanway & Eldridge 2018) are also commonly used in the analysis of high-redshift galaxies. The primary way in which the BPASS models differ from SB99 is via the inclusion of massive binary star evolution, which can have a strong effect on the predicted UV spectrum, particularly at low metallicities. In Cullen et al. (2019), we found that the BPASSv2.1 models return systematically lower $Z_\star$ values compared to SB99 by $\simeq 0.1$ dex (see also Chisholm et al. 2019). Fitting our current sample with the latest BPASSv2.2 models (including models at stellar metallicities down to 0.1 per cent solar) yields $\log(Z_\star) = -3.41 \pm 0.15$ dex and $\log(Z_\star) = -2.61 \pm 0.10$ dex for the low-$M_\star$ and high-$M_\star$ stacks respectively. This represents a much larger offset with respect to the SB99 value for the low-$M_\star$ stack ($\Delta \log(Z_\star) \simeq -0.6$ dex) but is consistent with the $Z_\star$ values reported in Topping et al. (2020b) ($\lesssim 5$ per cent solar). We conclude that the systematically lower metallicities derived using the BPASS models are the primary cause of the difference between our data and those of Topping et al. (2020b) in Fig. 9. Crucially, however, use of the BPASS models would only decrease the measured $Z_\star$ and hence increase the implied (O/Fe)-enhancement of our sample.

Finally, another concern is the depletion of O on to dust grains in the ISM, which would cause Z_g to be an underestimate of the true O/H in stars. Steidel et al. (2016) estimate that O-depletion causes O/H determined from nebular emission lines to be biased by $\simeq -0.09$ dex ($\simeq 20$ per cent) in high-redshift galaxies, with an uncertainty of ± 0.05 dex. Overall, the effect of O-depletion would be an underestimation of the true gas-phase O/H, resulting in an underestimation of O/Fe. Again, incorporating the effect of dust depletion would only strengthen the (O/Fe)-enhancement reported here.

In summary, while the exact value of (O/Fe)-enhancement is clearly subject to number of systematic uncertainties, with current estimates in the range (O/Fe) $\approx 2-10 \times$ (O/Fe)$_\odot$, we find no systematic effects that are large enough to explain away the basic result. Accordingly, we can state with some confidence that the abundance ratios in typical high-redshift star-forming galaxies are α-enhanced.

5 DISCUSSION

In this section, we focus on the implications of the main result of this work: O/Fe-enhancement (or α-enhancement) in high-redshift star-forming galaxies. As discussed in previous works (e.g. Steidel et al. 2016; Cullen et al. 2019), O/Fe-enhancement is actually expected for galaxies with constant/rising star-formation histories and relatively young ages (<500 Myr $- 1$ Gyr). This is a result of the fact that while O is produced predominantly via core-collapse supernovae (CCSNe), Fe is produced via a mixture of *both* CCSNe and Type-Ia SNe. Therefore, as a result of the delayed onset of Type-Ia SNe after the start of star formation, young galaxies will initially be underabundant in Fe relative to O, only reaching the solar ratio

on timescales $\gtrsim 1\,\mathrm{Gyr}$ (e.g. Maoz & Mannucci 2012; Maiolino & Mannucci 2019). Given the substantial evidence to suggest that star-forming galaxies at $z > 2$ are characterised by constant/rising SFHs and ages $\lesssim 1\,\mathrm{Gyr}$ (Reddy et al. 2012; Topping et al. 2020b), we therefore expect enhanced O/Fe ratios at this cosmic epoch. Below we discuss how the observed (O/Fe)-enhancement might relate to local observations, and its consequences with respect to excitation conditions in high-redshift galaxies.

5.1 Abundance patterns compared to Milky Way stars

It is interesting to consider the link between high-redshift stellar populations and local stellar populations that we know (based on their age) formed at high redshift. For example, if (O/Fe)-enhancement is ubiquitous among galaxies at $z > 2$, then we would expect to observe enhanced abundance ratios in local stars that formed at these cosmic epochs (i.e. in stars with ages $\gtrsim 10\,\mathrm{Gyr}$). In the subsequent discussion, we use the following definitions for consistency with local conventions:

$$[\mathrm{Fe/H}] = \log(\mathrm{Fe/H}) - \log(\mathrm{Fe/H})_{\odot}, \qquad (6)$$

and

$$[\mathrm{O/Fe}] = [\mathrm{O/H}] - [\mathrm{Fe/H}]. \qquad (7)$$

In this work, $\log(Z_{\star}/Z_{\odot})$ is a proxy for [Fe/H] and $\log(Z_g/Z_{\star})$ is a proxy for [O/Fe].

Following the argument described above, the relationship between [O/Fe] and [Fe/H] (or more generally any [α/Fe]−[Fe/H] relation) can be used to constrain star formation time-scales in galaxies due to the different production time-scales for the α elements and Fe. For example, directly after the onset of star formation, at low [Fe/H], the value of [O/Fe] is set solely by CCSNe yields, as well as the IMF of the first generation of stars (Tolstoy, Hill & Tosi 2009). Current galactic chemical evolution models − constrained by observations of stars in the Milky Way − predict an initial plateau at [O/Fe] $\simeq$ 0.6 (Fig. 10; e.g. Kobayashi, Karakas & Lugaro 2020) for stars in our Galaxy. Other estimates of CCSNe yields give similar values, typically within the range [O/Fe] $\simeq 0.6-0.75$ depending on the assumed stellar metallicity and IMF (e.g. Limongi & Chieffi 2003; Chieffi & Limongi 2004; Nomoto et al. 2006). Assuming no strong redshift evolution in either the IMF or CCSNe yields as a function of stellar metallicity, this plateau value should be relatively stable at all cosmic epochs.

At later times, Fe enrichment of the ISM becomes enhanced by Type Ia SNe, causing a knee in the relation and subsequent evolution towards [O/Fe] <0.6. In the Milky Way, this knee occurs at [Fe/H] $\simeq -1$. The sequence in [Fe/H] − [O/Fe] can therefore be thought of in terms of age, with older stars (i.e. those that formed earlier) located at high [O/Fe]. In Fig. 10, we show one recent model of the [Fe/H]−[O/Fe] relation for the Milky Way from Kobayashi et al. (2020), highlighting these trends.

The model of Kobayashi et al. (2020), and other so-called Galactic Chemical Evolution Models, are constrained by observations of Milky Way stellar abundance ratios. For example, studies of FKG stars in the solar neighborhood have revealed a tight correlation between [α/Fe] and age, with stars on the knee/plateau having formed 10–12 Gyr ago (i.e. $z = 2 - 4$; Haywood et al. 2013, 2019). These older, high-[α/Fe] populations, are predominantly located in the thick disk and halo (Bensby, Feltzing & Lundström 2003; Haywood et al. 2013; Zhao et al. 2016; Kobayashi et al. 2020). Evidence also exists for a distinct high-α population in the bulge at large-scale heights (Bensby et al. 2013; Lian et al. 2020), and is thought to indicate an
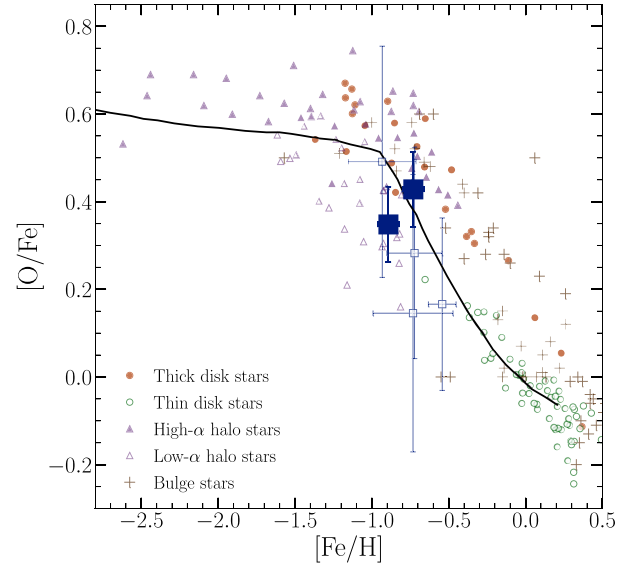
Figure 10. A comparison between the [Fe/H] − [O/Fe] relation for stars in the Milky Way and star-forming galaxies at $z \simeq 3.4$. The open square points show the individual galaxies in our sample for which it was possible to measure both $Z_{\star}$ and Z_g. The filled square points represent low-$M_{\star}$ and high-$M_{\star}$ stacks (with the low-$M_{\star}$ stack being the most metal-poor i.e. lowest [Fe/H]). All other data points represent individual Milky Way stars drawn from the halo, thick disk, thin disk and bulge as indicated in the legend. These data have been compiled from Bensby et al. (2013) (micro-lensed dwarfs and K-giants), and Zhao et al. (2016) and Amarsi et al. (2019) (F- and G-type stars). The black solid line shows a model for the average [Fe/H] − [O/Fe] relation for stars in the Milky Way, representing the evolutionary history of the galaxy (Kobayashi et al. 2020).

intimate link between the bulge and disc populations (Haywood et al. 2018). To illustrate these various populations, we show in Fig. 10 the position of a selection of F- and G-type stars drawn from the thin disc, thick disc and halo of the Milky Way using data from Zhao et al. (2016) and Amarsi, Nissen & Skúladóttir (2019), as well as a selection of micro-lensed dwarf and K-giant stars drawn from the bulge taken from Bensby et al. (2013). As can be seen from Fig. 10, stars in the thick disk and halo are found to have enhanced [O/Fe] while the stars with solar-like abundances are drawn pre-dominantly from the thin disc. A similar sequence is also evident for the bulge population, with stars at higher [O/Fe] having larger scale heights (Amarsi et al. 2019).

It can also be seen from Fig. 10 that our observed [Fe/H] − [O/Fe] values are remarkably consistent with the sequence defined locally, corresponding roughly to the location of the knee in the Milky Way sequence dominated by old stars ($\simeq$10–12 Gyr) drawn primarily from the thick disk, halo and high-α bulge populations of the Milky Way (e.g. Haywood et al. 2018, 2019). This similarity suggests that these local 10–12 Gyr old stellar populations likely formed from similar material (in terms of enrichment properties) to the massive O/B-type stellar populations we observe in the rest-frame FUV spectra of galaxies at $z = 3.5$ (lookback time 11.7 Gyr). The basic interpretation of this result would be that our $z = 3.5$ observations probe the formation of the stellar populations that dominate the largest scale heights (thick disc, subsets of the halo and bulge) in spiral galaxies at $z = 0$. We note that the potential systematic effects discussed in Section 4.3.2 do not affect this observation. In fact, converting our Z_g measurements to a RL abundance scale would shift [O/Fe] by $\simeq$
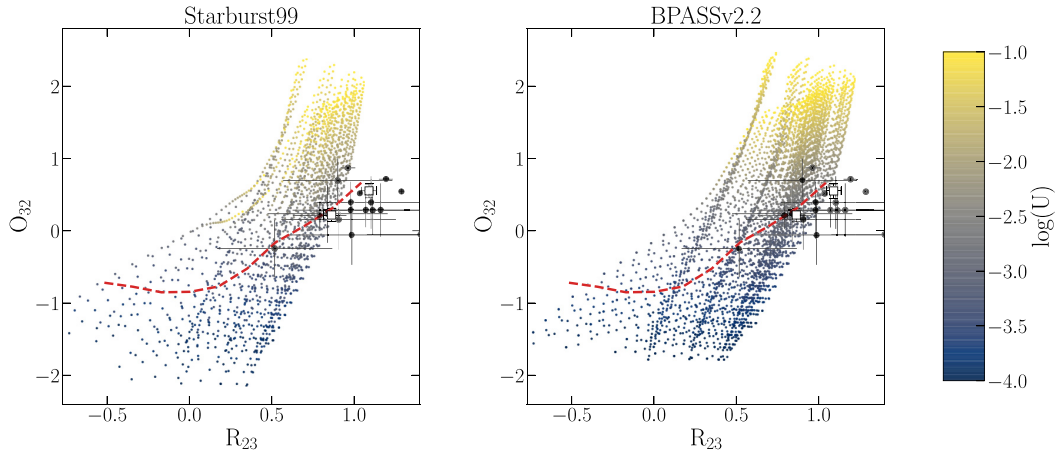
Figure 11. Comparing theoretical model predictions to observations in the R_{23}–O_{32} emission-line ratio diagram. In each panel, the coloured points show the photoionization model grids generated via the method described in the Appendix. The left-hand panels show grids generated assuming Starburst99 models for the input stellar ionizing continuum, while the right-hand panels show grids generated assuming the BPASSv2.2 models. The colour of the photoionization model points corresponds to the logarithm of the ionization parameter ($\log(U)$) as shown in the colour bar. In each panel the black circular data points show individual galaxies in our sample with detections of the required lines, and the open black squares show the low-$M_{\star}$ and high-$M_{\star}$ stacks. The red dashed line shows the running median for a sample of 1050 local H II regions (see the text for description).

+0.24 dex, somewhat improving the overlap between our data and the Milky Way halo/bulge population.

These results can also be compared to the element abundance ratios of damped Lyman-α systems (DLAs) at $z > 2$, which also exhibit α-enhanced abundance ratios at low metallicities ([Fe/H] $\lesssim 1.5$) (e.g. Cooke, Pettini & Jorgenson 2015; De Cia et al. 2016). Interestingly, however, Cooke et al. (2015) report a gentle decline towards solar [α/Fe] values by [Fe/H] $\simeq -1.0$ for DLAs, in contrast to the significant O/Fe-enhancement seen in Fig. 10 at similar metallicity. Based on their observations, Cooke et al. (2015) argue that the typical star-formation and chemical enrichment histories of DLA's are not representative of the galaxy population that formed the Milky Way stellar halo. Following the same argument, our results indicate that stellar populations drawn from the main sequence of star-formation at $z \simeq 3.4$ are representative of at least a fraction of MW halo stars.

Improved observations of [Fe/H]−[O/Fe] at high redshift have the potential to provide further constraints on the typical star-formation and chemical evolution histories of early galaxies. For example, the combined FUV+optical observations presented here can be performed at $z = 2-4$, and offer the opportunity to observe the evolution of [Fe/H]−[O/Fe] across ≈ 2 Gyr of cosmic time (and a wide range in stellar mass), placing unique constraints on the early evolutionary pathways of galaxies and providing unique clues as to the connection between various local and high-redshift stellar populations.

5.2 Excitation conditions at high redshift

The observed (O/Fe)-enhancement in $z \simeq 3$ NIRVANDELS galaxies also has important implications for our understating of the excitation conditions in high-redshift H II regions. Below, we briefly examine our results in relation to the much-discussed offset between high-redshift and local star-forming galaxies in classical line ratio diagnostic diagrams. In addition, we explore whether (O/Fe)-enhancement can resolve the previously reported difficulty in reproducing $z > 3$ emission-line ratios using standard photoionization modelling analyses.

5.2.1 The evolution of optical emission line ratios with redshift

It has been known for some time that high-redshift star-forming galaxies are offset from local star-forming galaxies in various nebular emission-line diagnostics diagrams (most notably the N2 BPT diagram; e.g. Steidel et al. 2014; Shapley et al. 2015). A number of physical explanations have been invoked to explain this effect, including an evolution in ionization parameter and ISM pressure (Cullen et al. 2016; Kashino et al. 2017), N/O abundance ratios (Masters et al. 2014; Shapley et al. 2015; Sanders et al. 2016) and harder ionizing spectra (Steidel et al. 2016; Strom et al. 2017; Sanders et al. 2020a; Topping et al. 2020b). In Fig. 11, we show the position of our $z \simeq 3.4$ sample in the $R_{23}-O_{32}$ nebular emission-line ratio diagram compared to a sample of 1050 local H II regions, where $R_{23} = \log(([O\,III]\,\lambda\lambda 4959, 5007 + [O\,II]\,\lambda\lambda 3726, 3729)/H\,\beta)$. Following Sanders et al. (2017) and Jeong et al. (2020), our H II sample is drawn from the catalogue of Pilyugin & Grebel (2016) and supplemented with additional data from Croxall et al. (2016) and Toribio San Cipriano et al. (2016). We choose to use local H II regions (as opposed to SDSS spectra) to avoid the problem of DIG contamination in the integrated spectra of local star-forming galaxies (e.g. Sanders et al. 2017).

It can be seen from Fig. 11 that, in agreement with much of the previous literature, we find that the $z \simeq 3.4$ line ratios are, in general, systematically offset from the median values at $z = 0$. The results presented in this paper, as well as previous observations of (O/Fe)-enhancement at $z > 2$, would suggest that one reason for this offset is due to the fact that, at fixed oxygen abundance, star-forming galaxies at $z > 3$ have harder ionizing continuum spectra (i.e. lower [Fe/H]) compared to star-forming galaxies in the local Universe. Indeed, Topping et al. (2020a, b) have shown explicitly how the increase in the excitation conditions within H II regions due to (O/Fe)-enhancement is directly related to the observed N2 BPT offset in $z \simeq 2.4$ galaxies. Our results do not rule out contributions from other physical effects, but highlight the importance of a harder ionizing continuum at high redshift (at fixed O/H).

5.2.2 Photoionization models

Another interesting question is whether – under the assumption of α-enhanced chemical abundances – current stellar-population synthesis models, combined with photoionization modelling of H II regions, can accurately reproduce the observed $z > 3$ nebular emission-line ratios. Previous literature results have highlighted the fact that photoionization models generally struggle to reproduce the observed ([O III] + [O II])/Hβ ratios in galaxies at $z > 3$ (e.g. Onodera et al. 2016). However, these previous comparison have generally been based on the assumption of assumed solar abundance ratios, coupling the metallicity of the ionizing stars to the metallicity of the gas [i.e. $Z_\star = Z_g$, or (O/Fe) = (O/Fe)$_\odot$].

In Fig. 11, we also show the results from a photoionization modelling analysis of the $R_{23} - $ [O III]/[O II] diagram, in which we have relaxed this assumption and decoupled the stellar and gas-phase metallicities. Full details of the photoionization modelling analysis are given in the Appendix. It can be seen from Fig. 11 that, even when accounting for a harder ionizing continuum at fixed O/H, the photoionization models still fail to reproduce the full range of observed emission-line ratios. Crucially, the offset between models and observations is evident (although smaller) even when assuming the harder ionizing continuum spectra predicted by the BPASSv2.2 models.[14] These results indicate that, even when accounting for a harder ionizing continuum at fixed O/H (i.e. α-enhancement), the shape of the ionizing spectrum predicted by current SPS models is still not sufficient to account for the full range typical emission-line ratios observed at $z > 3$.

The fact that the BPASSv2.2 models lie closer to the observed line ratios clearly indicates that increasing the strength of the ionizing continuum at fixed $Z_\star$ helps to alleviate this problem and that an insufficiently hard ionizing continuum is likely to be contributing to the discrepancy. Indeed, this conclusion has also frequently been reached with respect to modelling the He II emission line in metal-poor star-forming galaxies across a large range of redshifts (e.g. Plat et al. 2019; Saxena et al. 2020). Additionally, the offset may be related to the fact that most photoionization models treat galaxies as single H II regions, when in reality the observed spectra are a line luminosity-weighted averages of many H II regions, each with distinct physical properties. Our analysis adds to the growing body of literature that suggests that the full range of emission-line ratios observed at high-redshift are still not fully captured by simple photoionization models (even when including the effect of α-enhanced abundance ratios). Solving this problem will be crucial for accurately interpreting of the rest-frame optical spectra of galaxies at $z > 4$ with *JWST*.

6 SUMMARY AND CONCLUSIONS

In this paper, we have presented the results of an analysis of the gas-phase and stellar metallicities of $N = 33$ typical star-forming galaxies at $3.0 \leq z \leq 3.8$ drawn from the VANDELS survey. Using gas-phase metallicity measurements derived from rest-frame optical emission-line ratios and stellar metallicity measurements derived from rest-frame FUV continuum fitting, we have investigated the scaling of both parameters with galaxy stellar mass. In addition, we have directly compared the gas-phase and stellar metallicities of our

sample to investigate evidence for (O/Fe)-enhancement in typical star-forming galaxies at $z \simeq 3.4$. Finally, we have discussed how our observations may be related to chemical abundance ratios of stars in the Milky Way. The main results of this study can be summarized as follows:

(i) The three rest-frame optical emission-line ratios covered by our MOSFIRE H- and K-band observations ([O III]/Hβ, [O III]/[O II], and [Ne III]/[O II]) all show a negative correlation with $M_\star$. The slope of the $z \simeq 3.4$ line ratio correlations are consistent with those in the local Universe, but with a higher normalization at fixed $M_\star$. These qualitative trends are to be expected if gas-phase metallicities increase with $M_\star$ and decrease with redshift.

(ii) Converting the dust-corrected line ratios into gas-phase metallicities (Z_g, or O/H), we observe a clear gas-phase mass–metallicity relationship (MZR$_g$) using the calibration of Bian et al. (2018). The relation has a gradient of $d(\log Z_g)/d(\log M_\star) \simeq 0.3$, with the average value of Z_g increasing from $\simeq 0.2\,Z_\odot$ to $\simeq 0.7\,Z_\odot$ across the stellar mass range $M_\star \simeq 10^{8.5}$–$10^{10.5}\,M_\odot$.

(iii) We derived stellar metallicities ($Z_\star$, or Fe/H) for our sample based on full FUV continuum fitting in the wavelength range 1250–2000 Å (rest frame). As with Z_g, we observe an increase in $Z_\star$ with increasing $M_\star$. The slope of the MZR$_\star$ relation is consistent with the MZR$_g$ relation, but $Z_\star$ is offset to lower values at fixed stellar mass. We find $Z_\star$ increases from $\simeq 0.08\,Z_\odot$ to $\simeq 0.25\,Z_\odot$ across the stellar-mass range $M_\star \simeq 10^{8.5} - 10^{10}\,M_\odot$.

(iv) We find an approximately constant offset between Z_g and $Z_\star$ at fixed $M_\star$, implying enhanced O/Fe relative to the solar value (i.e. α-enhancement), with (O/Fe) = $2.54 \pm 0.38 \times$ (O/Fe)$_\odot$. This result is consistent with the α-enhancement observed in galaxy samples at slightly lower redshift ($z \simeq 2.4$; e.g. Steidel et al. 2016; Topping et al. 2020b) and is robust against known systematics. Within the uncertainties of our observations, there is no strong evidence that (O/Fe)-enhancement is a function of $M_\star$ at $M_\star < 10^{10}\,M_\odot$.

(v) We compare our measurements to the average [Fe/H]−[O/Fe] sequence of stars in the Milky Way. Under the assumption that $\log(Z_\star/Z_\odot)$ is a proxy for [Fe/H] and $\log(Z_g/Z_\star)$ is a proxy for [O/Fe], we find the position of our $z \simeq 3.4$ sample in the [Fe/H]−[O/Fe] plane is consistent with the knee of the Milky Way sequence. In the Milky Way, the knee is predominantly populated by old stars consistent with a formation redshift of $z \sim 2 - 4$ (i.e. 10–12 Gyr old) found at large scale heights above the plane of the disc. Improved constraints on $Z_\star$ and Z_g at high redshift could shed light on the link between high-redshift stellar populations and the structural components in local galaxies using chemical abundance patterns.

(vi) Finally, we demonstrate that our results are consistent with previous results in the literature suggesting that the observed evolution in rest-frame optical emission-line ratios with redshift is caused, at least in part, by an increase in the hardness of the stellar ionizing continuum at fixed oxygen abundance (e.g. Steidel et al. 2016; Topping et al. 2020b). We also show, via a photoionization modelling analysis accounting for (O/Fe)-enhancement, that the full range of observed line ratios at $z > 3$ remain difficult to reproduce with standard models. Resolving this tension between the models and observations will be crucial for accurately interpreting upcoming spectroscopic data of high-redshift galaxies from *JWST*.

---

[14]The effects of binary star evolution in the BPASS models results in a significant boost in the ionizing flux at fixed $Z_\star$ with respect to the Starburst99 models, particularly at $Z_\star < 0.3\,Z_\odot$ (see e.g. Stanway, Eldridge & Becker 2016).

DATA AVAILABILITY

The VANDELS spectra used in this work a publicly available and can be accessed via the ESO science portal (http://archive.eso.org/scienceportal/home). All other data contributing to this article will be shared on reasonable request to the corresponding author.

REFERENCES

Amarsi A. M., Nissen P. E., Skúladóttir Á., 2019, A&A, 630, A104
Amorín R. et al., 2015, A&A, 578, A105
Andrews B. H., Martini P., 2013, ApJ, 765, 140
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Astropy Collaboration et al., 2018, AJ, 156, 123
Bensby T., Feltzing S., Lundström I., 2003, A&A, 410, 527
Bensby T. et al., 2013, A&A, 549, A147
Bian F., Kewley L. J., Dopita M. A., 2018, ApJ, 859, 175
Calabrò A. et al., 2021, A&A, 646, A39
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carnall A. , 2021, SpectRes: Simple spectral resampling (ascl:2104.019)
Chabrier G., 2003, PASP, 115, 763
Chieffi A., Limongi M., 2004, ApJ, 608, 405
Chisholm J., Rigby J. R., Bayliss M., Berg D. A., Dahle H., Gladders M., Sharon K., 2019, ApJ, 882, 182
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Cooke R. J., Pettini M., Jorgenson R. A., 2015, ApJ, 800, 12
Croxall K. V., Pogge R. W., Berg D. A., Skillman E. D., Moustakas J., 2016, ApJ, 830, 4
Cullen F., Cirasuolo M., Kewley L. J., McLure R. J., Dunlop J. S., Bowler R. A. A., 2016, MNRAS, 460, 3002
Cullen F. et al., 2019, MNRAS, 487, 2038
Cullen F. et al., 2020, MNRAS, 495, 1501
Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Esposito S., 2017, MNRAS, 465, 1384
Curti M., Mannucci F., Cresci G., Maiolino R., 2020, MNRAS, 491, 944
De Cia A., Ledoux C., Mattsson L., Petitjean P., Srianand R., Gavignaud I., Jenkins E. B., 2016, A&A, 596, A97
Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, PASA, 34, e058
Esteban C., García-Rojas J., Carigi L., Peimbert M., Bresolin F., López-Sánchez A. R., Mesa-Delgado A., 2014, MNRAS, 443, 624
Ferland G. J. et al., 2017, RMxAA, 53, 385
Garilli B. et al., 2021, A&A, 647, A150
Grogin N. A. et al., 2011, ApJS, 197, 35
Halliday C. et al., 2008, A&A, 479, 417
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Haywood M., Di Matteo P., Lehnert M. D., Katz D., Gómez A., 2013, A&A, 560, A109
Haywood M., Di Matteo P., Lehnert M., Snaith O., Fragkoudi F., Khoperskov S., 2018, A&A, 618, A78
Haywood M., Snaith O., Lehnert M. D., Di Matteo P., Khoperskov S., 2019, A&A, 625, A105
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jeong M.-S. et al., 2020, ApJ, 902, L16
Juneau S. et al., 2014, ApJ, 788, 88
Kashino D. et al., 2017, ApJ, 835, 88
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Kewley L. J., Nicholls D. C., Sutherland R. S., 2019, ARA&A, 57, 511
Kobayashi C., Karakas A. I., Lugaro M., 2020, ApJ, 900, 179
Kobulnicky H. A., Kewley L. J., 2004, ApJ, 617, 240
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Kriek M., van Dokkum P. G., Labbé I., Franx M., Illingworth G. D., Marchesini D., Quadri R. F., 2009, ApJ, 700, 221
Kriek M. et al., 2015, ApJS, 218, 15
Kroupa P., 2001, MNRAS, 322, 231
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., Ekström S., Meynet G., Schaerer D., Agienko K. B., Levesque E. M., 2014, ApJS, 212, 14
Lian J. et al., 2020, MNRAS, 497, 3557
Limongi M., Chieffi A., 2003, ApJ, 592, 404
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, PASA, 29, 447
Mármol-Queraltó E., McLure R. J., Cullen F., Dunlop J. S., Fontana A., McLeod D. J., 2016, MNRAS, 460, 3587
Masters D. et al., 2014, ApJ, 785, 153
McLean I. S. et al., 2012, Proc. SPIE, 8446, 8446
McLure R. J. et al., 2018, MNRAS, 479, 25
Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nucl. Phys. A, 777, 424
Oey M. S. et al., 2007, ApJ, 661, 801
Oliphant T. E., 2007, Comput. Sci. Eng., 9, 10
Onodera M. et al., 2016, ApJ, 822, 42
Peimbert M., 1967, ApJ, 150, 825
Pentericci L. et al., 2018, A&A, 616, A174
Pérez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Pilyugin L. S., Grebel E. K., 2016, MNRAS, 457, 3678
Plat A., Charlot S., Bruzual G., Feltre A., Vidal-García A., Morisset C., Chevallard J., Todt H., 2019, MNRAS, 490, 978
Reddy N. A., Pettini M., Steidel C. C., Shapley A. E., Erb D. K., Law D. R., 2012, ApJ, 754, 25
Reddy N. A. et al., 2015, ApJ, 806, 259
Reddy N. A. et al., 2018, ApJ, 869, 92
Reddy N. A. et al., 2020, ApJ, 902, 123
Rix S. A., Pettini M., Leitherer C., Bresolin F., Kudritzki R.-P., Steidel C. C., 2004, ApJ, 615, 98
Runco J. N. et al., 2021, MNRAS, 502, 2600
Salim S., Boquien M., Lee J. C., 2018, ApJ, 859, 11
Sanders R. L. et al., 2016, ApJ, 816, 23
Sanders R. L., Shapley A. E., Zhang K., Yan R., 2017, ApJ, 850, 136
Sanders R. L. et al., 2018, ApJ, 858, 99
Sanders R. L. et al., 2020a, preprint (arXiv:2009.07292)
Sanders R. L. et al., 2020b, MNRAS, 491, 1427
Saxena A. et al., 2020, A&A, 636, A47
Schaerer D., de Barros S., Sklias P., 2013, A&A, 549, A4
Schreiber C. et al., 2018, A&A, 618, A85
Shapley A. E. et al., 2015, ApJ, 801, 88
Shapley A. E. et al., 2017, ApJ, 846, L30
Shapley A. E. et al., 2019, ApJ, 881, L35
Sommariva V., Mannucci F., Cresci G., Maiolino R., Marconi A., Nagao T., Baroni A., Grazian A., 2012, A&A, 539, A136
Speagle J. S., 2020, MNRAS, 493, 3132
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Stanway E. R., Eldridge J. J., 2018, MNRAS, 479, 75
Stanway E. R., Eldridge J. J., Becker G. D., 2016, MNRAS, 456, 485
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

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

Theios R. L., Steidel C. C., Strom A. L., Rudie G. C., Trainor R. F., Reddy N. A., 2019, ApJ, 871, 128

Tolstoy E., Hill V., Tosi M., 2009, ARA&A, 47, 371

Topping M. W., Shapley A. E., Reddy N. A., Sanders R. L., Coil A. L., Kriek M., Mobasher B., Siana B., 2020a, MNRAS, 495, 4430

Topping M. W., Shapley A. E., Reddy N. A., Sanders R. L., Coil A. L., Kriek M., Mobasher B., Siana B., 2020b, MNRAS, 499, 1652

Toribio San Cipriano L., García-Rojas J., Esteban C., Bresolin F., Peimbert M., 2016, MNRAS, 458, 1866

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

Troncoso P. et al., 2014, A&A, 563, A58

Vale Asari N., Couto G. S., Cid Fernandes R., Stasińska G., de Amorim A. L., Ruschel-Dutra D., Werle A., Florido T. Z., 2019, MNRAS, 489, 4721

Zahid H. J., Kudritzki R.-P., Conroy C., Andrews B., Ho I. T., 2017, ApJ, 847, 18

Zhao G. et al., 2016, ApJ, 833, 225

APPENDIX A: PHOTOIONIZATION MODELS

The photoionization model grids discussed in Section 5, and shown in Fig. 11, were generated using CLOUDY v17.01 (Ferland et al. 2017). To generate these grids, we assumed a simple plane-parallel geometry and adopted a constant H II region electron density of $n_e = 350$ cm^{-3} [based on the median [O II] doublet ratio for our sample and using the conversion given in Sanders et al. (2016)]. The three variable input parameters were Z_g, $Z_\star$, and the ionization parameter, U. The ionization parameter sets the intensity of the ionization radiation field

and is defined as

$$U = n_{LyC}/n_H, \tag{A1}$$

where n_{LyC} is the volume density of hydrogen-ionizing photons and n_H is the volume density of hydrogen gas ($n_H \approx n_e$). We varied $\log(U)$ between -1.0 to -4.0 in steps of 0.1 dex. For the gas-phase metallicity, we adopted the following set of values $Z_g/Z_\odot = 0.05$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.25, 1.5 ($7.4 \leq 12 + \log(O/H) \leq 8.9$).

In order to account for non-solar O/Fe ratios, we decouple the stellar metallicity (which sets the shape of the ionizing continuum spectrum, and is primarily dependent on Fe/H) from the gas-phase metallicity (tracing O/H). To generate the input ionizing continuum at a given $Z_\star$ we adopted two fiducial models, one based on Starburst99 (Leitherer et al. 2014) and the other BPASSv2.2 (Eldridge et al. 2017; Stanway & Eldridge 2018). In both cases, we assumed continuous star formation with an age of 100 Myr and a Chabrier (2003) IMF. For the Starburst99 models we assumed the same weak-wind Geneva tracks with stellar rotation used to fit the FUV continuum spectra (see Section 3). The available $Z_\star$ values for these models are $Z_\star/Z_\odot = 0.07$, 0.14, 0.56, 0.99, 2.8. For the BPASSv2.2 models, we assumed the binary star evolution pathway with an upper-mass IMF cut-off of 300M$_\odot$. The values of $Z_\star$ available for the BPASSv2.2 models are $Z_\star/Z_\odot = 0.0007$, 0.007, 0.07, 0.14, 0.21, 0.28, 0.42, 0.56, 0.70, 0.99, 1.4, 2.1. For each value of $Z_\star$, photoionization grid points were generated across the full range Z_g and $\log(U)$ values. The number of individual grid points for the Starburst99 and BPASSv2.2 grids were therefore 2015 and 4836, respectively. The resulting photoionization model grids in the $R_{23}-O_{32}$ plane are shown in Fig. 11.

This paper has been typeset from a T$_E$X/L^{A}T$_E$X file prepared by the author.