Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **532,** 3102–3119 (2024) https://doi.org/10.1093/mnras/stae1705
Advance Access publication 2024 July 11

The NIRVANDELS survey: the stellar and gas-phase mass-metallicity relations of star-forming galaxies at z = 3.5

T. M. Stanton [1]★ F. Cullen [1] R. J. McLure,[1] A. E. Shapley,[2] K. Z. Arellano-Córdova [1] R. Begley [1] R. Amorín [3] L. Barrufet [1] A. Calabrò,[4] A. C. Carnall [1] M. Cirasuolo,[5] J. S. Dunlop,[1] C. T. Donnan [1] M. L. Hamadouche [1] F. Y. Liu,[1] D. J. McLeod [1] L. Pentericci,[4] L. Pozzetti,[6] R. L. Sanders [7] D. Scholte [1] and M. W. Topping [8]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, UK*
[2]*Department of Physics & Astronomy, University of California, 430 Portola Plaza, Los Angeles, CA 90095, USA*
[3]*ARAID Foundation. Centro de Estudios de Física del Cosmos de Aragón (CEFCA), Unidad Asociada al CSIC, Plaza San Juan 1, E–44001 Teruel, Spain*
[4]*INAF – Osservatorio Astronomico di Roma, via Frascati 33, I-00078 Monteporzio Catone, Italy*
[5]*European Southern Observatory, Karl-Schwarzschild-Str 2, D-86748 Garching b. München, Germany*
[6]*INAF – Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, I-40129 Bologna, Italy*
[7]*Department of Physics and Astronomy, University of Kentucky, 505 Rose Street, Lexington, KY 40506, USA*
[8]*Steward Observatory, University of Arizona, 933 N Cherry Avenue, Tucson, AZ 85721, USA*

Accepted 2024 July 9. Received 2024 June 19; in original form 2024 May 3

ABSTRACT

We present determinations of the gas-phase and stellar metallicities of a sample of 65 star-forming galaxies at $z \simeq 3.5$ using rest-frame far-ultraviolet (FUV) spectroscopy from the VANDELS survey in combination with follow-up rest-frame optical spectroscopy from *VLT*/KMOS and Keck/MOSFIRE. We infer gas-phase oxygen abundances (Z_g; tracing O/H) via strong optical nebular lines and stellar iron abundances ($Z_\star$; tracing Fe/H) from full spectral fitting to the FUV continuum. Our sample spans the stellar mass range $8.5 < \log(M_\star/\mathrm{M}_\odot) < 10.5$ and shows clear evidence for both a stellar and gas-phase mass-metallicity relation (MZR). We find that our O and Fe abundance estimates both exhibit a similar mass-dependence, such that $\mathrm{Fe/H} \propto M_\star^{0.30\pm0.11}$ and $\mathrm{O/H} \propto M_\star^{0.32\pm0.09}$. At fixed $M_\star$ we find that, relative to their solar values, O abundances are systematically larger than Fe abundances (i.e. α-enhancement). We estimate an average enhancement of $(\mathrm{O/Fe}) = 2.65 \pm 0.16 \times (\mathrm{O/Fe})_\odot$ which appears to be independent of $M_\star$. We employ analytic chemical evolution models to place a constraint on the strength of galactic-level outflows via the mass-outflow factor (η). We show that outflow efficiencies that scale as $\eta \propto M_\star^{-0.32}$ can simultaneously explain the functional form of of the stellar and gas-phase MZR, as well as the degree of α-enhancement at fixed Fe/H. Our results add further evidence to support a picture in which α-enhanced abundance ratios are ubiquitous in high-redshift star-forming galaxies, as expected for young systems whose interstellar medium is primarily enriched by core-collapse supernovae.

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

1 INTRODUCTION

The secular processes that govern the growth of galaxies are dictated by an overarching cosmic baryon cycle, which encompasses the infall of gas into galaxies, its collapse into and processing by stars, and its release via outflows driven by stellar and active galactic nuclei (AGN) feedback. The metal enrichment of gas and stars in galaxies is intrinsically linked to each of these processes, and therefore the combination of galaxy metallicities with measurements of galaxy properties, such as stellar mass, provides a powerful tool for studying the evolution and growth of galaxies (Maiolino & Mannucci 2019; Péroux & Howk 2020).

The gas-phase abundance of oxygen relative to hydrogen (O/H) is the most common proxy for galaxy metallicity because oxygen dominates the elemental mass distribution and can be readily constrained from rest-frame optical spectra (Kewley, Nicholls & Sutherland 2019). At high redshifts, oxygen abundances are usually inferred from nebular emission line ratios that have been calibrated against direct T_e method abundances (e.g. Maiolino & Mannucci 2019). There are a variety of these calibrations, each of which is sensitive to different metallicity regimes and subject to their own systematic uncertainties (e.g. Curti et al. 2017; Bian, Kewley & Dopita 2018; S24). Even in the local Universe these systematics are large [up to 0.7 dex in log(O/H); e.g. Kewley & Ellison 2008; Arellano-Córdova & Rodríguez 2020], and the situation is compounded at high redshift where significant samples of direct abundances (needed to derive robust empirical calibrations) are only now becoming available with *JWST* observations (e.g. S24).

However, relative abundances derived from a given calibration at a given redshift are presumed to be reliable (Kewley et al. 2019) and, across the literature, gas-phase metallicities (Z_g) have been

★ E-mail: t.stanton@ed.ac.uk

measured from the local Universe out to redshifts $z \sim 10$ (e.g. Curti et al. 2023). At all redshifts, a tight correlation between gas-phase oxygen abundance and galaxy stellar mass ($M_\star$) has been confirmed (i.e. the mass–metallicity relationship, MZR) such that galaxies at higher masses exhibit higher gas-phase metallicities (e.g. Tremonti et al. 2004; Andrews & Martini 2013; Curti et al. 2020). Furthermore, multiple studies have shown that the MZR evolves with redshift, such that at fixed stellar mass galaxies at higher redshifts are less oxygen enriched (e.g. Savaglio et al. 2005; Cullen et al. 2014; Troncoso et al. 2014; Sanders et al. 2021; Li et al. 2023). While the existence of the MZR is robustly established, fully characterizing and understanding the shape and evolution of the MZR with cosmic time remains a key goal in observational astrophysics.

An alternative tracer of the metal content of high-redshift galaxies is the abundance of iron (Fe/H), which can be derived from photospheric line blanketing in the rest-frame far-ultraviolet (FUV) continuum (e.g. Leitherer et al. 2010; Steidel et al. 2016; Cullen et al. 2019). The rest-frame FUV traces young, massive, O- and B-type stars, and FUV-based stellar metallicities reflect the Fe/H of the interstellar medium (ISM) from which they formed, permitting a direct comparison between the Fe/H of these stars and the gas-phase O/H derived the rest-frame optical nebular spectra. Steidel et al. (2016) introduced a novel method for deriving stellar Fe/H from full spectra fitting to rest-frame FUV spectra, and this method has subsequently been developed to robustly determine stellar metallicities up to $z \simeq 3.5$ (e.g. Cullen et al. 2019). A number of independent studies have now confirmed the existence of the stellar MZR at $z \simeq 2 - 3.5$ (e.g. Cullen et al. 2019; Calabrò et al. 2021; Kashino et al. 2022; Chartab et al. 2024) with a shape similar to the local relation, but showing an evolution to lower metallicity (e.g. Gallazzi et al. 2005; Panter et al. 2008; Zahid et al. 2017).

Combining rest-frame FUV and rest-frame optical spectroscopy with the aforementioned methodologies, it is possible to simultaneously determine both O/H and Fe/H in star-forming galaxies, yielding an estimate of the O/Fe abundance ratio (e.g. Steidel et al. 2016; Topping et al. 2020a, b; Cullen et al. 2021). The O/Fe ratio is sensitive to star-formation time-scales due to the fact that oxygen (and α-elements in general) are produced primarily by core-collapse supernovae (CCSNe) over short time-scales, whereas Fe peak elements are released by *both* CCSNe and Type-Ia supernovae (SNIa) probing longer time-scales (e.g. Maoz & Mannucci 2012; Kobayashi, Karakas & Lugaro 2020). Determining O/Fe therefore provides insight on the relative contributions of CCSNe and SNIa, such that galaxies exhibiting super-solar O/Fe ratios (often referred to more generally as α-enhancement) display chemical enrichment patterns dominated by the yields of CCSNe. As such, a number of authors have highlighted a connection between O/Fe ratios and the specific star-formation rate of galaxies (e.g. Matthee & Schaye 2018; Kashino et al. 2022; Chruślińska et al. 2024).

Additionally, chemical evolution models predict that the elemental abundance ratios are sensitive to various macroscopic galaxy processes, such as outflow and star-formation efficiencies (e.g. Weinberg, Andrews & Freudenburg 2017). As a result, the combination of absolute (e.g. Fe/H) and relative (e.g. O/Fe) abundances can be used to explore variations in outflow properties and star-formation histories across the galaxy population.

Moreover, as discussed in Topping et al. (2020a), an Fe deficit relative to O also offers a natural explanation for the evolution in ISM conditions at high redshift, as indicated, for example, by the evolution in classic line ratio diagnostic diagrams (e.g. the [N II] BPT, Steidel et al. 2014; Shapley et al. 2019; the [S II] BPT, Shapley et al. 2019; [Ne III] emission-line diagrams, Jeong et al. 2020; and the [O I] BPT, Clarke et al. 2023). For α-enhanced galaxies, the stellar ionizing spectrum at fixed O abundance is harder, as lower iron abundance leads to less metal-line blanketing in the ionizing continuum at < 912Å. Accurate nebular modelling of galaxies at high redshift therefore requires robust estimates of the typical O/Fe ratio as a function of galaxy properties.

The number of studies exploring O/Fe ratios in high redshift galaxies ($z \simeq 2 - 4$) has continued to grow since the pioneering work of Steidel et al. (2016). To date, all works find ubiquitous evidence for α-enhanced abundance ratios in the range $\simeq 2 - 5 \times (\mathrm{O/Fe})_\odot$ either via estimates of O and Fe abundances for the same sources (e.g. Topping et al. 2020a, b; Cullen et al. 2021; Chartab et al. 2024) or indirect approaches (such as photoionization modeling or comparison of different sources; Sanders et al. 2020; Strom et al. 2022; Kashino et al. 2022).

These direct galaxy estimates of α-enhancement are in good agreement with the signal measured in high-redshift Damped Lyα abosrption line systems (e.g. Cooke et al. 2011; De Cia et al. 2016; Velichko et al. 2024). These results corroborate expectations that young star-forming galaxies at high redshift are α-enhanced and reflect the dominance of CCSNe in star-forming galaxies at these epochs (Maiolino & Mannucci 2019). Interestingly, the estimated level of α-enhancement at these redshifts has been shown to be consistent with stellar archaeological measurements of $10 - 12$ Gyr old stellar populations in the Milky Way (MW; Cullen et al. 2021).

However, despite this progress, the number of individual galaxies with O/Fe estimates is still relatively small (e.g. Topping et al. 2020a, b), and explorations of the variation of α-enhancement with global galaxy properties have been limited. In this work, we compile a sample of 65 star-forming galaxies at $3.0 < z < 3.8$ with simultaneous deep rest-frame optical spectroscopy [from *VLT*/K-band Multi-Object Spectrograph (KMOS) and Keck/MOSFIRE] and FUV spectroscopy (from the VANDELS survey; McLure et al. 2018; Pentericci et al. 2018; Garilli et al. 2021) to determine O and Fe abundances for both individual and composite spectra. Our new analysis follows on from Cullen et al. (2021) who used a sample of 33 galaxies at $z \simeq 3.5$ to determine an average degree of α-enhancement of $(\mathrm{O/Fe}) \simeq 2.5 \times (\mathrm{O/Fe})_\odot$. In this paper, we double the sample size and expand on the analysis in Cullen et al. (2021) by improving constraints on the gas-phase and stellar MZR at $z = 3.5$, as well as the variation of O/Fe with stellar mass. We also demonstrate how our combination of absolute and relative abundance determinations enables us to place novel constraints on the mass scaling of large-scale galactic outflows via comparison with analytic chemical evolution models.

The structure of this paper is as follows. In Section 2, we discuss our sample selection and the primary spectroscopic data sets used in this work.

Section 3 describes the methodologies that we have employed to determine the abundances of O and Fe from the spectra. In Section 4 we present our determinations of the gas-phase and stellar MZRs at $z \simeq 3.5$, along with our inferred measurements of the O/Fe ratios for our sample. We then present constraints on the stellar mass dependence of galactic-scale outflows using analytical chemical evolution models. Our conclusions are summarized in Section 6.

Throughout this work, metallicities are quoted relative to a solar abundance value taken from Asplund et al. (2009), which has a bulk composition by mass of $Z_* = 0.0142$ and relates the widely-used $12 + \log(\mathrm{O/H})$ scale to the Z_g scale via the expression $12 + \log(\mathrm{O/H}) = \log(Z_\mathrm{g}) + 8.69$. We assume the following cosmology: $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70\ \mathrm{kms^{-1}\ Mpc^{-1}}$.

2 DATA AND SAMPLE PROPERTIES

Our analysis is based on Keck/MOSFIRE and *VLT*/KMOS near-infrared (near-IR) spectroscopic follow-up of galaxies selected from the VANDELS survey (McLure et al. 2018). The VANDELS spectroscopy (taken with the *VLT*/VIMOS spectrograph) provides rest-frame FUV spectra for each galaxy while the near-IR spectroscopy traces the rest-frame optical. For $\simeq 50$ per cent of our final galaxy sample we obtained near-IR follow-up with Keck/MOSFIRE and these observations have been previously described in detail in Cullen et al. (2021). The new data presented here consist of 40 h of *VLT*/KMOS observations (ID:108.21Z5.001; PI: Cullen) obtained between October 2021 and February 2023. In the following, we describe each of these data sets in turn, with a particular focus on the selection and reduction of the new KMOS observations used in this paper.

2.1 The VANDELS sample

Star-forming galaxies at $z \simeq 3.5$, observed as part of the VANDELS survey, form the basis of our analysis. VANDELS was an ESO/*VLT* public spectroscopic survey conducted with the VIMOS spectrograph (McLure et al. 2018; Pentericci et al. 2018; Garilli et al. 2021). VANDELS targets were selected from regions centred on the CANDELS CDFS and UDS fields (Grogin et al. 2011; Koekemoer et al. 2011) in three categories: (i) massive passive galaxies at $1.0 \leq z \leq 2.5$, (ii) bright star-forming galaxies at $2.4 \leq z \leq 5.5$ and (iii) fainter star-forming galaxies at $3.0 \leq z \leq 7.0$, with the primary focus being star-forming galaxies at $z > 2.4$ (comprising 85 per cent of all targets). All observations were obtained using the VIMOS medium resolution grism covering the wavelength range $0.48\mu\text{m} < \lambda_{\text{obs}} < 1.0\mu\text{m}$ at a median resolution of R = 580 (with 1 arcsec slits) and a dispersion of 2.5Å per pixel. The median seeing across all observations was 0.7 arcsec.

In our redshift range of interest ($3.0 < z < 3.8$) the VANDELS spectra cover the rest-frame wavelength range 1000 Å $< \lambda <$ 2000 Å. The depth of the VANDELS observations (median 40 h integration per galaxy) enables stellar Fe/H to be derived for both individual and composite spectra of galaxies (e.g. Cullen et al. 2019, 2021; Calabrò et al. 2021) as we describe in Section 3.2. Full descriptions of the VANDELS survey design and target selection can be found in McLure et al. (2018), and observations and data reduction are described Pentericci et al. (2018) and Garilli et al. (2021).

2.2 MOSFIRE observations

For 33 galaxies in our final sample, we obtained rest-frame near-IR spectroscopic follow-up with the Multi-Object Spectrometer for Infrared Exploration (MOSFIRE; McLean et al. 2012) on the Keck I telescope. We observed the VANDELS-selected targets in the redshift range $3.0 < z < 3.8$ with the Keck/MOSFIRE *H* and *K* bands in order to target the [O II] λλ3726, 3729, [Ne III] λ3870, H β, and [O III] λ5007 nebular emission lines necessary to derive O/H estimates in the ionized gas. Observations were conducted using a slit width of 0.7 arcsec, yielding a spectral resolution of 3650 in *H* and 3600 in *K*. The median seeing across all observations was 0.5 arcsec. Despite the narrower MOSFIRE slits compared to VIMOS, the improved seeing for the MOSFIRE observations means that both observations probe comparable intrinsic regions in our target galaxies. Full details of the selection and properties of the Keck/MOSFIRE sample are given in Cullen et al. (2021).

2.3 KMOS observations

The new observations presented here consist of additional near-IR follow-up of VANDELS targets at $3.0 < z < 3.8$ obtained with the KMOS on the *VLT*. Below we describe the sample selection, observations, and data reduction for this sample.

2.3.1 *Sample selection*

The initial targets for *VLT*/KMOS observations were drawn from the VANDELS DR3 catalogue and selected to have: (i) a redshift in the range $3.0 \leq z \leq 3.8$ (so that the [O II] λλ3726, 3729, [Ne III] λ3870, H β, and [O III] λλ4959, 5007 optical nebular emission lines were accessible to the KMOS *HK* grating) and (ii) a redshift quality flag $z_{\text{flag}} \geq 2$ [e.g. a redshift flag 2, 3, 4, 9, or 14, as defined in Pentericci et al. (2018), corresponding to a ≥ 70 per cent probability of the redshift being correct]. From this sample, the highest priority was given to targets according to following criteria: (i) a signal-to-noise ratio (SNR) per resolution element in the VANDELS spectrum of ≥ 3 (to maximize the number of targets for which it would be possible to measure individual Fe/H); (ii) a redshift quality flag of $z_{\text{flag}} = 3$ or 4 (corresponding to a ≥ 95 per cent probability of the redshift being correct; Pentericci et al. 2018); (iii) a redshift such that the [O II] λλ3726, 3729, H β, and [O III] λλ4959, 5007 optical nebular emission lines were predicted to fall on relatively sky-free regions of the KMOS *HK* grating, and (iv) an estimated stellar mass of $\geq 10^9 M_{\odot}$ (to ensure optimal SNR of the optical lines; see Section 2.5.2 for details on the stellar mass estimates). These criteria were imposed to maximize the number of individual galaxies for which Fe/H and O/H could be estimated.

2.3.2 *Observations and data reduction*

KMOS is a near-IR spectrograph consisting of 24 integral field units (IFUs) operating simultaneously within a 7.2 arcmin circular field-of-view (FoV). Each IFU has a 2.8 arcsec × 2.8 arcsec FoV and a uniform spatial sampling of 0.2 arcsec × 0.2 arcsec. The 24 IFUs are fed into three separate detectors with eight IFUs assigned per detector. Across our multiple observing runs (see Table 1) one IFU was continuously out of commission meaning that we only observed 23 sources per pointing. We observed in the KMOS *HK*-band covering the wavelength range 1.48 μm − 2.44 μm with a spectral resolution of R = 1985 at band centre. A total of two pointings were observed, with 20 IFUs per pointing assigned to VANDELS-selected galaxies (described above) and three IFUs per pointing assigned to reference stars (one reference star per detector). These reference stars were used to correct for small detector-dependent shifts in the position of targets between exposures (as described below).

For all objects we adopted an object-sky-object nodding strategy with 300s per exposure (or observing block, OB) and dithered the exposures for sky sampling and subtraction. The resulting maximum on-source integration time per object was 10 h. Across the full observing run, various IFUs periodically malfunctioned and were out of commission on a given night, resulting in a nonuniform observing time across all objects in the sample. We also removed observations taken during poor seeing conditions (> 0.8 arcsec) or objects for which the shifts between OBs could not be accurately calibrated using the reference star. In practice, 94 per cent of the OBs in pointing 1 were usable. However, due to significant disruption caused by out-of-commission pickoff arms, only 40 per cent of OBs in pointing 2 were used. Details of our *VLT*/KMOS observations are given in Table 1.

Table 1. Summary of the *VLT*/KMOS observations.

Pointing	Dates observed	RA (J2000)	Dec (J2000)	N_{gal}	Exposure time (min)	Median full width at half-maximum seeing (arcsec)
p1	Oct−Dec 2021	03:32:54.9	-27:44:42.2	20	566.6	0.56
p2	Dec 2021−Feb 2022/Sept 2022−Feb 2023	03:32:47.2	-27:54:45.7	20	266.6	0.57

We reduced the raw data to produce three-dimensional science and error spectra using the ESOREX-KMOS pipeline (v 4.0.4) described in Davies et al. (2013). For a given IFU, the pipeline extracts the raw image and performs dark and flatfield corrections, illumination corrections, and wavelength calibrations. The pipeline then combines all the reduced images corresponding to the given IFU to form a final datacube. To further improve the quality of the data reduction, we implemented two routines in addition to the standard pipeline routines. First, each OB was combined using a custom algorithm that accounts for small shifts in the positions of targets between each exposure. These small shifts (on average $\lesssim 3$ pixels for the usable OBs) are detector-dependent and were calculated using the position of the bright references stars in each pointing. Second, to derive a more accurate estimate of the sky background, we masked the positions of the nebular emission lines in the datacubes (following Stott et al. 2016) and subtracted the combined median sky cube from all of our final datacubes. The generation of the final datacubes was run in two passes: first without sky subtraction to produce initial datacubes from which accurate KMOS spectroscopic redshifts could be measured, and then with sky subtraction using the measured redshifts. Finally, the data cubes were reconstructed on a spatial scale of 0.1 arcsec × 0.1 arcsec.

For the analysis in this paper, we focus on galaxy-integrated KMOS spectra. To generate the galaxy-integrated one-dimensional spectra, we summed the flux within 1 arcsec diameter apertures centred on the peak of the [O III] λ5007 emission. An aperture size of 1 arcsec was found to be a good compromise between sampling as much of the galaxy as possible while also maximizing the signal-to-noise in the extracted spectra (see also Hayden-Pawson et al. 2022).

It is worth noting that the chosen aperture size should not introduce strong biases in the derived metallicities since galaxies at $z > 3$ are expected to have flat metallicity gradients (e.g. Curti et al. 2020; Venturi et al. 2024). Crucially, a 1 arcsec aperture size matches the VANDELS aperture size, ensuring that we are probing the same regions of our galaxies in the rest-frame optical as in the rest-frame FUV. An example of a slice through one datacube at the position of the [O III] λ5007 line with the associated extraction aperture is shown in Fig. 1. We also show the resulting 1D spectra in the region of the [O II] λλ3726, 3729, [Ne III] λ3870, H β, and [O III] λλ4959, 5007 emission lines.

We evaluated the accuracy of the flux calibration of the final 1D spectra by comparing the integrated flux of the reference stars to photometry from the 3D-HST catalogues (Skelton et al. 2014; Momcheva et al. 2016) in matched apertures. We found good agreement with offsets of < 10 per cent averaged across the six reference stars.

2.4 Final galaxy sample

The final *VLT*/KMOS sample analysed in this paper consists of 32/40 of the targeted galaxies covering the redshift range $3.0 < z < 3.7$. We excluded 7/40 galaxies that lacked a clear ($>3\sigma$) emission line detection required for accurately determining a redshift. An additional galaxy was excluded due to lacking spectral coverage of the [O II] λλ3726, 3729 feature required for our abundance determinations. All galaxy spectra were visually inspected to exclude obvious AGN. We found no galaxies that exhibited extremely broad emission features or high ionization emission features. We also ruled out the presence of significant AGN ionization based on their mid-IR spectral energy distribution (SED) shapes and X-ray properties (McLure et al. 2018). An example galaxy from the sample (KVS-208) is shown in Fig. 1, and the general properties of the KMOS sample are summarized in Table 2.

In addition to the new *VLT*/KMOS sample, we also include the 33 galaxies from the Keck/MOSFIRE sample described in Cullen et al. (2021). As described below, the methodology for deriving key physical parameters (e.g. stellar mass, O/H, Fe/H) is the same between the two samples and therefore, unless explicitly stated, we take estimates of these parameters directly from Cullen et al. (2021; see their table 2).

Combined, our full sample consists of 65 galaxies in the redshift range $2.95 < z < 3.80$ (median $z = 3.5$). Each galaxy has both rest-frame optical and rest-frame FUV spectra, from which we infer O/H and Fe/H, respectively (see Section 3). In the instance of being unable to infer either quantity for a given galaxy, the known spectroscopic redshift permits inclusion into the generation of composite spectra (see Section 2.6). All galaxies are selected from the VANDELS spectroscopic survey. As described in a number of previous studies, the star-forming galaxies in the VANDELS survey are consistent with typical 'main-sequence' galaxies (e.g. McLure et al. 2018; Cullen et al. 2021; Garilli et al. 2021) and we have verified that this is also case for the combined sample presented here. Our final sample can therefore be considered representative of normal star-forming galaxies at $z \simeq 3.5$ within the mass range of our sample (see Section 2.5).

2.5 Measurements and derived quantities

In the following, we describe our measurements of global galaxy properties (e.g. stellar mass, star-formation rate (SFR)) and emission line fluxes for the galaxies in our *VLT*/KMOS sample. The methodology adopted here closely follows the methodology described in Cullen et al. (2021) for the Keck/MOSFIRE observations.

2.5.1 Rest-frame optical emission line fluxes and redshifts

Rest-frame emission line fluxes were measured by fitting Gaussian profiles to the 1D science spectra extracted from the final datacubes. For each individual line in a given galaxy spectrum, we isolated the expected position of the feature using the systemic redshift, within which we applied a Bayesian model fitting algorithm to fit a Gaussian profile characterized by an amplitude, width, and line centroid. All of the fitted line profiles were visually inspected, and the best-fitting fluxes, full width at half-maximums, and centroids were determined by taking the median of the resulting posterior distributions, with uncertainty bounds taken as the 16th and 84th percentiles. The [O II] λλ3726, 3729 emission line doublet was blended at the resolution of our observations, and so for these lines

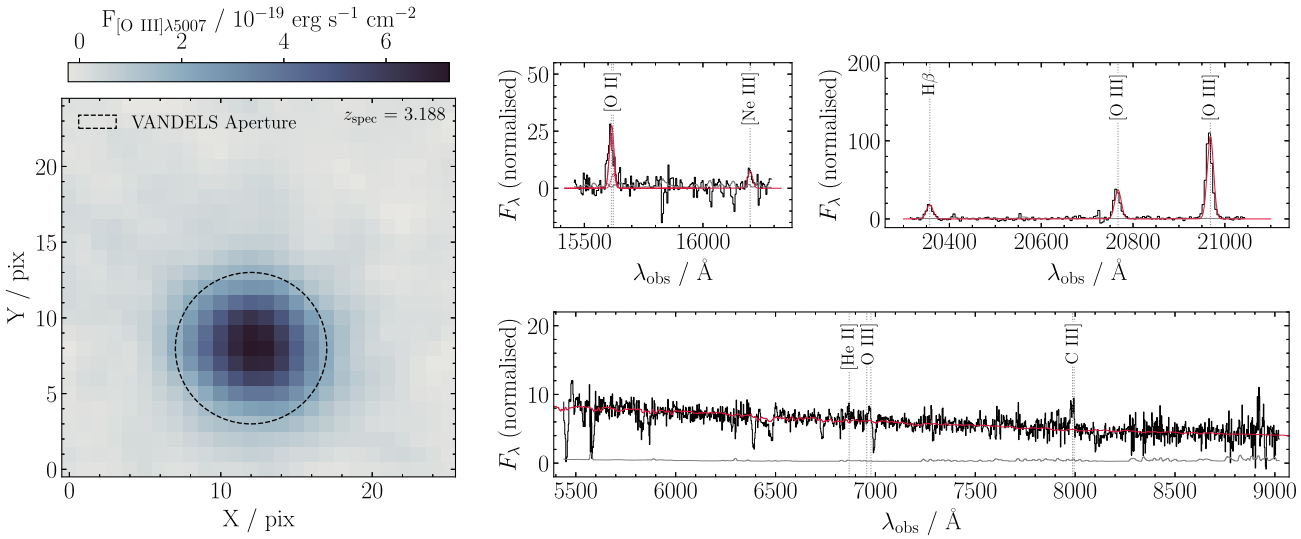

Figure 1. An example galaxy from our VANDELS plus *VLT*/KMOS sample. The left-hand panels shows a 2D slice through the KMOS datacube centred on the wavelength of the [O III] λ5007 emission line for the galaxy KVS-208 ($z = 3.188$; Table 2). The dashed circle represents the 1 arcsecond diameter aperture used to extract the galaxy integrated spectra selected to match to the extraction for the VANDELS spectra. The top two panels in the upper right-hand side show two regions from the extracted KMOS spectrum centred on the [O II] λλ3726, 3729, [Ne III] λ3870, H β, and [O III] λλ4959, 5007 emission lines used to derived gas-phase O/H (see Section 3.1). The black line shows the extracted spectrum and the red line shows the best-fitting model assuming Gaussian line profiles and a flat continuum (see Section 2.5 for details of the emission-line fitting). The lower right-hand panel shows the rest-frame FUV VANDELS spectrum for this object. The VANDELS spectrum is shown in black and the associated error spectrum in grey. The positions of nebular FUV emission lines are marked by vertical dashed lines. The best fitting Starburst99 stellar population model from which we estimate the stellar Fe/H is shown in red (see Section 3.2 for details of the FUV continuum fitting).

two Gaussian profiles of equal widths were fitted simultaneously. An example fit to the galaxy KVS-208 is shown in Fig. 1. Within our 1 arcsec apertures, our data exhibit a median 2σ line flux sensitivity of $\simeq 2 \times 10^{-18}$ erg s^{-1} cm^{-2} Å^{-1}.

For one galaxy in our sample (KVS-248), the observed emission lines exhibited broad wings, and thus were better described by fitting a combination of a broad and narrow Gaussian profiles centred at the same wavelength, but with differing velocity widths. The FUV VANDELS spectrum of this galaxy contained no high ionization emission lines indicating AGN activity, and as such the broad features are attributable to outflows, and the narrow features to the nebular emission of the galaxy. For this galaxy, we used only the narrow nebular feature flux measurements in our analysis.

2.5.2 *Stellar masses and star-formation rates*

To estimate stellar masses and star-formation rates, we utilized the multiwavelength photometry from the VANDELS photometric catalogues (McLure et al. 2018). At the typical redshift of the *VLT*/KMOS sample ($z \simeq 3.5$), the emission line fluxes of the [O II] λλ3726, 3729, H β and [O III] λλ4959, 5007 features will contaminate the broadband photometry in the H- and K- bands (Cullen et al. 2021). As the VANDELS photometry was derived using 2 arcsec diameter apertures, we first scaled our 1 arcsec KMOS line fluxes using the average ratio of 2:1 arcsec flux derived from the reference star data.[1] As discussed in the following, we find that this average correction yields excellent consistency between SED-derived SFRs and SFRs estimated from the dust-corrected H β line flux. For each galaxy, we constructed a model emission-line spectrum based on our line fits, which were then integrated over the appropriate filter profiles. The VANDELS photometry was then adjusted by subtracting these resulting integrated fluxes. In the H-band, the corrections ranged from $-0.23 < \Delta H/\mathrm{mag} < 0.0$ with a median correction of $\Delta H = -0.06$ mag, while corrections in the K-band ranged from $-1.63 < \Delta K/\mathrm{mag} < 0.0$ with a median correction of $\Delta K = -0.21$ mag. These corrections are consistent with the emission line corrections found for the Keck/MOSFIRE sample described in Cullen et al. (2021) who report median corrections of $\Delta H = -0.07$ mag and $\Delta K = -0.35$ mag.

The emission line-corrected photometry was modelled using ⟨0:sc ⟩FAST+ +⟨/0:sc⟩ (Schreiber et al. 2018), a modified SED fitting code based on the original FAST software (Kriek et al. 2009). Following Cullen et al. (2021), we used the Conroy, Gunn & White (2009) flexible stellar population synthesis models, assuming solar metallicities,[2] a Chabrier (2003) initial mass function (IMF), constant star-formation histories and the Calzetti et al. (2000) dust attenuation curve. These SED-fitting parameters were chosen to ensure full consistency between this work and Cullen et al. (2021), as well as with the stellar-to-nebular dust correction calibration derived by Sanders et al. (2021; see Section 2.5.3 below). The final fits yielded an estimate of galaxy stellar mass ($M_\star$), star-formation rate ($\mathrm{SFR_{SED}}$) and a model of the underlying stellar continuum. Using

[1]We note that galaxies are not point sources and will exhibit different scaling factors when using variable aperture sizes. To verify our chosen approach, we measured [O III] line fluxes from our galaxies in 1 and 2 arcsec apertures. Reassuringly, we find an equivalent average scaling factor from this approach.

[2]While we will show in our subsequent analysis (Section 4) that our galaxies exhibit O abundances of $0.2 - 0.7\,\mathrm{Z_\odot}$ and Fe abundances of $0.1 - 0.3\,\mathrm{Z_\odot}$, fitting the broadband photometry with solar metallicity models is required for consistency with the Sanders et al. (2021) $\mathrm{E(B\text{-}V)_{neb}}$ calibration [Section 2.5.3; equation (1)] which we use to dust correct our nebular emission line flues in the absence of a direct Balmer decrement measurements.

Table 2. Redshift, mass, gas-phase, and stellar metallicity measurements for the KMOS sample utilizing the B18 calibration scheme and the S99 stellar population models.

Name	ID	z_{spec}	$\log(M_\star/M_\odot)$	$\log(Z_g/Z_\odot)$	$\log(Z_\star/Z_\odot)$
KVS-006	002584	3.367	9.84	–	$-0.30^{+0.07}_{-0.05}$
KVS-009	237255	3.472	9.60	–	–
KVS-014	001280	3.082	9.68	–	–
KVS-055	105702	3.085	10.26	$-0.06^{+0.05}_{-0.06}$	$-0.67^{+0.06}_{-0.06}$
KVS-067	235637	3.082	8.97	$-0.18^{+0.07}_{-0.08}$	–
KVS-070	027049	3.250	9.04	–	–
KVS-075	012786	3.348	9.36	$-0.31^{+0.08}_{-0.08}$	–
KVS-082	230755	3.185	8.90	$-0.23^{+0.08}_{-0.08}$	–
KVS-085	101998	3.193	9.83	$-0.32^{+0.07}_{-0.09}$	$-0.93^{+0.09}_{-0.07}$
KVS-087	005891	3.081	9.35	–	$-0.64^{+0.06}_{-0.06}$
KVS-093	016368	3.423	9.34	–	–
KVS-100	003496	3.031	9.35	$-0.47^{+0.08}_{-0.08}$	–
KVS-101	027379	3.591	9.54	–	–
KVS-131	014729	3.607	10.03	$-0.46^{+0.08}_{-0.09}$	–
KVS-141	002312	3.470	9.63	$-0.31^{+0.09}_{-0.09}$	–
KVS-150	103010	3.462	9.57	$-0.39^{+0.08}_{-0.09}$	–
KVS-156	019593	3.426	9.58	–	$-0.71^{+0.08}_{-0.08}$
KVS-202	130840	3.182	10.26	$-0.01^{+0.05}_{-0.06}$	–
KVS-204	132586	3.472	9.17	$-0.42^{+0.09}_{-0.10}$	$-0.61^{+0.06}_{-0.06}$
KVS-208	228618	3.188	9.51	$-0.37^{+0.07}_{-0.08}$	$-0.98^{+0.03}_{-0.03}$
KVS-215	207213	3.386	9.20	–	–
KVS-220	232419	3.424	10.44	–	–
KVS-227	017345	3.611	9.10	$-0.43^{+0.07}_{-0.08}$	$-1.06^{+0.04}_{-0.04}$
KVS-248	231194	3.078	9.53	$-0.25^{+0.06}_{-0.06}$	$-0.61^{+0.07}_{-0.06}$
KVS-266	208053	3.089	8.97	–	–
KVS-298	106745	3.465	9.46	$-0.41^{+0.08}_{-0.09}$	–
KVS-312	101861	3.239	9.53	$-0.41^{+0.08}_{-0.08}$	$-0.83^{+0.04}_{-0.03}$
KVS-340	232407	3.411	9.27	$-0.42^{+0.07}_{-0.08}$	–
KVS-361	018915	3.605	9.98	$-0.31^{+0.08}_{-0.09}$	–
KVS-391	131717	3.071	9.68	$-0.23^{+0.06}_{-0.06}$	$-0.94^{+0.04}_{-0.05}$
KVS-414	004151	3.651	9.11	$-0.52^{+0.12}_{-0.13}$	–
KVS-423	234108	3.341	9.55	$-0.50^{+0.07}_{-0.08}$	–

For each galaxy, we also provide the corresponding VANDELS ID.

the best-fitting stellar continuum models we corrected our measured H β flux values for the effect of underlying stellar absorption. This correction resulted in a median increase of 2.8 per cent to our H β flux measurements.

2.5.3 *Dust-correcting nebular emission line fluxes*

Accurate measurements of the gas-phase oxygen abundance require correcting the observed line fluxes for nebular dust attenuation. Conventionally, these corrections are made directly using measurements of the Balmer decrement (H α/H β). However, our observations lacked spectral coverage of the H α feature. As an alternative, we used the calibration between stellar reddening ($E(B\text{-}V)_{stellar}$), SFR_{SED}, redshift and nebular reddening ($E(B\text{-}V)_{neb}$) derived by Sanders et al. (2021),

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

This calibration was derived from observations of galaxies at $z \simeq 2.3$ using measurements of $E(B\text{-}V)_{neb}$ from the Balmer decrement and $E(B\text{-}V)_{stellar}$ from SED fitting (using the same SED fitting assumptions described above). The resulting calibration provides an unbiased estimate of $E(B\text{-}V)_{neb}$ with an intrinsic scatter of 0.23 magnitudes. In 2/32 galaxies, both H γ and H β were detected allowing $E(B\text{-}V)_{neb}$ to be measured from the H γ/H β ratio. The measured $E(B\text{-}V)_{neb}$ are consistent with the $E(B\text{-}V)_{neb}$ derived from equation (1). Based on the estimated $E(B\text{-}V)_{neb}$ values from equation (1) we corrected our observed emission line fluxes assuming a Cardelli, Clayton & Mathis (1989) extinction curve.[3]

The self-consistency of our nebular and stellar dust corrections was verified by deriving star formation rates from the dust-corrected H β line fluxes and comparing to SFR_{SED}. We converted the dust-corrected H β fluxes (scaled to 2 arcsec diameter apertures) into estimates of the intrinsic H α line flux assuming a ratio of H α/H β = 2.86, and derived star formation rates using the Hao et al. (2011) H α-SFR conversion modified for a Chabrier (2003) IMF. The H β-derived SFRs are compared with SFR_{SED} in Fig. 2 and show excellent agreement. The average offset between the two estimates is 0.06 ± 0.09 dex, where the error in the offset is derived from the median absolute deviation ($\sigma = 1.4826 \times$ MAD).

Reassuringly, the SFRs derived from SED fitting and from H β exhibit excellent consistency, highlighting the self-consistency of our stellar and nebular dust correction. Fig. 2 also demonstrates that the *VLT*/KMOS and Keck/MOSFIRE samples exhibit an indistinguishable trend, indicating that our analysis is consistent with that of Cullen et al. (2021).

The final redshift, stellar mass, and star-formation rate distributions of the complete sample are shown in Fig. 3.

2.6 Composite spectra

Of the 32 galaxies in the *VLT*/KMOS sample, we were able to determine individual gas-phase metallicities for 21/32 and stellar metallicities for 11/32, with 8/32 galaxies having determinations of both. (see Table 2 and Section 3 below). In order to fully leverage the entire sample, we therefore employed spectral stacking techniques. In Section 4, we present the results of splitting the sample into two bins split at the median mass of $10^{9.53}\,M_\odot$. To stack the spectra, we converted each spectrum into luminosity density units using its spectroscopic redshift and correct for nebular dust attenuation using its $E(B\text{-}V)_{neb}$ and the Cardelli et al. (1989) extinction law. Each spectrum was normalized using its [O III] λ5007 luminosity to prevent biasing the composites in favour of the brightest objects, and then re-sampled onto a common rest-frame wavelength grid with a spectral sampling of 1Å per pixel. The final stacked spectrum was generated by applying a 3σ clipping and taking the median value at each wavelength, and the associated error spectrum was calculated via bootstrap resampling.

For the Keck/MOSFIRE galaxies, we adopt the sample selection and stack sizes described in Cullen et al. (2021). As described in Cullen et al. (2021), 5/33 galaxies were removed from the stacking sample due to the lack of spectral coverage of the H β line, and two stacks were generated from the remaining 28 galaxies split at the median mass of $10^{9.4}\,M_\odot$. The stacks were generated using an identical methodology to the *VLT*/KMOS composites.

[3]The choice of this extinction law was motivated by evidence of the nebular attenuation of high-redshift star-forming galaxies closely following the Cardelli et al. (1989) extinction curve (e.g. Reddy et al. 2020).

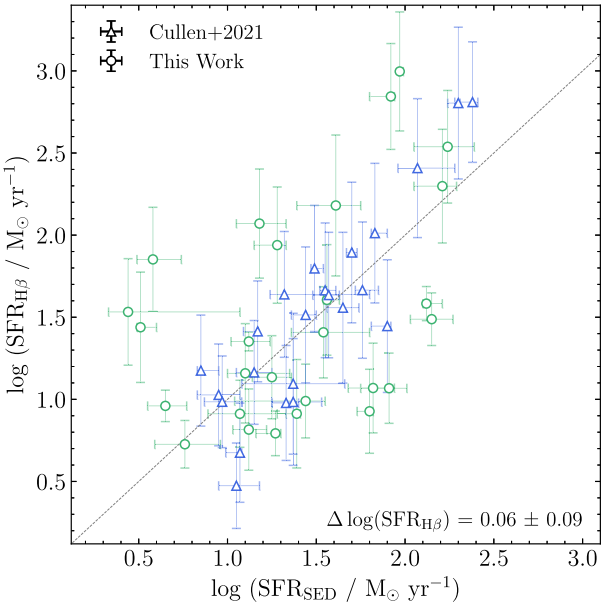

Figure 2. A comparison between the star-formation ratess derived using the emission-line corrected photometry and those derived from dust-corrected H β line fluxes using the Hao et al. (2011) SFR conversion. The dotted black line represents the 1:1 relation. Galaxies from our *VLT*/KMOS sample (green circles) are in good agreement with the trend seen for the Keck/MOSFIRE sample presented in Cullen et al. (2021; blue triangles). Overall, we see excellent agreement between our $\mathrm{SFR}_{\mathrm{SED}}$ and $\mathrm{SFR}_{\mathrm{H}\,\beta}$ estimates. We find an average offset between our SED derived SFRs and our H β derived SFRs of 0.06 ± 0.09 dex from the 1:1 line indicating that our stellar and nebular dust correction are fully self-consistent.

As the emission features in the composite spectra were not necessarily well described by Gaussian profiles, we measured the integrated line fluxes via direct integration after subtracting any local continuum. To infer uncertainties on line fluxes, we perturbed the composite spectrum by its associated error spectrum 500 times and re-measured the line fluxes. For each line, the resulting flux and uncertainty were taken as the median and 68th percentile width of the resulting distribution of measurements. To correct for stellar absorption in the H β line, we multiplied the measured H β line fluxes by the median correction factor for the individual galaxies ($\simeq \times 1.028$). The properties of the composites are summarized in Table 3.

For the generation of the VANDELS stacks we followed a similar procedure except that in this case we normalized each individual spectrum by the median flux within the wavelength window $1420 \leq \lambda_{\mathrm{rest}} \leq 1480$. We also do not explicitly correct for dust attenuation when constructing the VANDELS stacks as the attenuation is accounted for in the stellar metallicity fitting methodology described in Section 3.2. Each individual spectrum was resampled onto a common rest-frame wavelength grid between 1200 and 2000 Å with a spectral sampling of 1 Å. The final stacked spectrum was generated by applying a 3σ clipping and taking the median value at each wavelength, and the associated error spectrum was calculated via bootstrap resampling.

3 DETERMINING GAS-PHASE AND STELLAR METALLICITIES

The primary focus of this study is the determination of the gas-phase metallicities (Z_{g}, tracing O/H) and stellar metallicities ($Z_\star$, tracing Fe/H) as a function of galaxy stellar mass for our sample. For the remainder of the paper we will primarily use the notation Z_{g} and $Z_\star$ when referring to the respective metallicity estimates. In the following section, we describe in detail how each of these properties was measured from the spectroscopic data.

3.1 Determination of gas-phase metallicity

We estimated Z_{g} using ratios of the strong nebular emission lines measured from the rest-frame optical spectra, employing a variety of empirical strong-line calibrations. As discussed above, our *H*- and *K*-band spectroscopic observations cover the [O II] λλ3726, 3729,

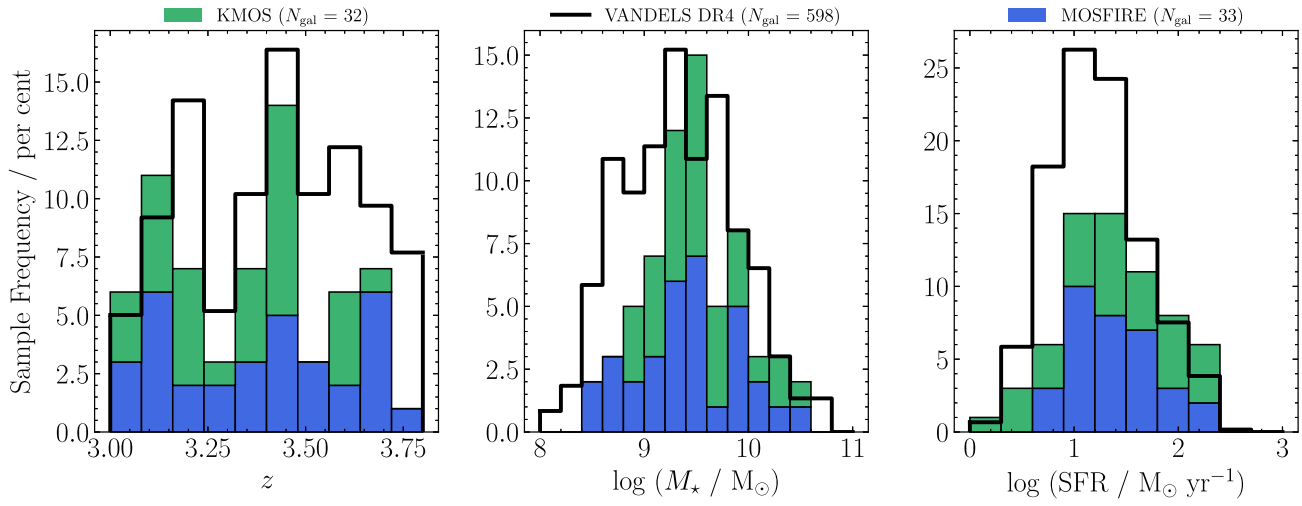

Figure 3. Plots showing, from left to right, the redshift, stellar mass, and star-formation rate distributions for the *VLT*/KMOS (green) and Keck/MOSFIRE (blue) samples. The combined sample has an average redshift of $z \simeq 3.5$, an average stellar mass of $\log(M_\star/\mathrm{M}_\odot) = 9.5 \pm 0.5$ and an average SFR of $\log(\mathrm{SFR}/\mathrm{M}_\odot \mathrm{yr}^{-1}) = 1.4 \pm 0.5$. In all three panels, it can be seen that the distributions for both samples are comparable. Moreover, these distributions are consistent with the entire VANDELS DR4 $3.0 < z < 3.8$ star-forming sample shown in black, though the median SFR of our sample is marginally higher than the median SFR of the VANDELS sample ($\log(\mathrm{SFR}/\mathrm{M}_\odot \mathrm{yr}^{-1}) = 1.2 \pm 0.4$). The comparison highlights the fact that our sample is consistent with typical VANDELS star-forming galaxies and therefore representative of normal, main-sequence, star-forming galaxies at $z \simeq 3.5$ (McLure et al. 2018).

Table 3. Properties of the low-$M_\star$ and high-$M_\star$ composite spectra from the *VLT*/KMOS sample, utilizing the B18 calibrations and S99 SPS models.

Stack	Mass Range	Median $\log(M_\star/M_\odot)$	$\log([O\,III]/H\,\beta)$	$\log([O\,III]/[O\,II])$	$\log(([O\,III]+[O\,II])/H\,\beta)$	$\log(Z_g/Z_\odot)$	$\log(Z_\star/Z_\odot)$
Low-$M_\star$	$8.9-9.53$	9.23	0.78 ± 0.15	0.54 ± 0.09	1.02 ± 0.15	-0.46 ± 0.05	-0.95 ± 0.05
High-$M_\star$	$9.53-10.44$	9.68	0.57 ± 0.09	0.36 ± 0.07	0.85 ± 0.08	-0.32 ± 0.04	-0.77 ± 0.08

We have excluded [Ne III]/[O II] due to neither composite yielding a significant detection of [Ne III]λ3870. The same information for the low- and high-$M_\star$ composite spectra from the Keck/MOSFIRE sample are given in table 1 of Cullen et al. (2021).

[Ne III] λ3870, H β and [O III] λλ4959, 5007 emission lines, which permit the use of the following line ratio diagnostics:

(i) O32 = [O III] λλ4959, 5007/[O II] λλ3726, 3729
(ii) O3 = [O III] λ5007/H β
(iii) O2 = [O II] λλ3726, 3729/H β
(iv) R23 = ([O III] λλ4959, 5007 + [O II] λλ3726, 3729)/H β
(v) Ne3O2 = [Ne III] λ3870/[O II] λλ3726, 3729

To estimate Z_g from these ratios, we follow the methodology described in Cullen et al. (2021) and use the empirical calibrations of Bian et al. (2018; hereafter B18) as our primary calibration set. These calibrations are based on local analogues of high-redshift galaxies. We have additionally implemented the calibrations of Curti et al. (2020; hereafter C20) to which we added the new Ne3O2 line ratio calibration from Curti et al. (2023), derived from galaxies from SDSS. Finally, we also implemented the calibration scheme of Sanders et al. (2024; hereafter S24), which are derived from early *JWST* measurements of auroral lines to determine direct O/H measurements of galaxies from $2 \leq z \leq 9$. As Cullen et al. (2021) only report B18-based oxygen abundances for the Keck/MOSFIRE sample, we have calculated O/H for the Keck/MOSFIRE sample using these additional calibration schemes. The oxygen abundances corresponding to these alternative calibration schemes are reported in Table A1.

For a given calibration scheme, we calculated the following χ^2 equation:

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

where the sum over i represents the above set line ratios, $x = 12 + \log(O/H)$,[4] $R_{obs,i}$ is the logarithm of the ith line ratio, $R_{cal,i}$ predicted value of $R_{obs,i}$ at x, and $\sigma_{cal,i}$ and $\sigma_{obs,i}$ are the uncertainties on the calibration and measured line ratio respectively. We converted equation (2) into a likelihood function ($\propto e^{-\chi^2/2}$) which was then maximized over x using the Bayesian nested sampling package DYNESTY (Speagle 2020) assuming a uniform prior across the range $6.7 \leq 12 + \log(O/H) \leq 9.0$ ($\simeq 0.01 - 2Z_\odot$). The final metallicity was taken to be the median of the resulting metallicity posterior distribution, with the 1σ uncertainty derived from the 68th percentile width of the posterior probability distribution. The resulting best-fitting $12 + \log(O/H)$ was converted to a gas-phase metallicity using $12 + \log(O/H) = \log(Z_g/Z_\odot) + 8.69$.

We note that our assumed prior extends beyond the formal range of the B18, C20, and S24 calibrations, which are $7.8 \leq 12 + \log(O/H) \leq 8.4$, $7.6 \leq 12 + \log(O/H) \leq 8.9$, and $7.4 \leq 12 + \log(O/H) \leq 8.3$, respectively. These extrapolations were necessary because for some objects, the observed line ratios fell outside of the range of the calibrations. For the B18 and C20 calibrations, the number of such cases were low. For the B18 calibration, 8/41 individual galaxies fell into the extrapolated metallicity range; however, all of the composites estimates fell within the original range. For C20 calibration, all of the individual galaxies and composite fell within the defined metallicity range. The S24 calibrations yielded the largest number of best-fit metallicities beyond the formal range (17/41), however all of the composites were consistent with the original range within $\simeq 2\sigma$.

Because our main results are based primarily on the composite spectra, these extrapolations do not affect our main conclusions.

For a given object, not all emission lines were significantly detected ($>3\sigma$). We determined that the minimum requirement for a robust metallicity estimate was significant detections of the [O III] λ5007 and [O II] λλ3726, 3729 features, since the O_{32} ratio is needed to break the degeneracy of the double-valued O_3 calibration. Ne_3O_2 is also a monotonic ratio and thus could be used alone to infer metallicity (Shapley et al. 2017), but due to the weaker line strength of [Ne III] λ3870 we found that there were no instances in which this was the only available line ratio.

For our *VLT*/KMOS sample, 11/32 galaxies were fit using all line ratios, 9/32 galaxies were fit using all excluding Ne3O2, and 1/32 galaxies were fit using only O_{32}; 11/32 galaxies do not have individual metallicity determinations due to the lack of a significant estimate of the O_{32} ratio. For both the low- and high-$M_\star$ composite spectra, we found that all lines except for [Ne III] were significantly detected. Consequently, the composite spectra did not have accurate estimates of Ne3O2, and were fit using all other available ratios. The resulting gas-phase metallicity measurements for the individual and composite spectra can be found in Tables 2 and 3, respectively. Incorporating the estimates from the Cullen et al. (2021) Keck/MOSFIRE sample adds an additional 21 individual galaxies and 2 composite spectra with Z_g estimates. In total therefore our sample contains 42 galaxies with individual Z_g estimates and four Z_g estimates from composite spectra binned by stellar mass.

3.2 Determination of stellar metallicity

We estimated $Z_\star$ from the VANDELS rest-frame FUV spectra following the methodology described in Cullen et al. (2019, 2021). Within the wavelength range $\lambda = 1221 - 2000$ Å, SPS models were fit to the full spectrum, masking out regions defined by Steidel et al. (2016) as containing nebular emission and ISM emission/absorption features.

As our fiducial model for the FUV spectral fitting we use the Starburst99 (S99; Leitherer et al. 1999, 2010). We assumed constant star-formation over time-scales of 100 Myr,[5] and generated models

[4] $x = 12 + \log(O/H) - 8.0$ for the S24 calibration.

[5] We note that this choice of star-formation time-scale is somewhat arbitrary as the FUV SEDs reach an equilibrium after a few $\times 10^7$ yr (see the discussion in Topping et al. 2020b).

adopting the weaker-wind Geneva tracks with single-star evolution and no stellar rotation. Five models were generated with metallicities of $Z_\star = (0.001, 0.002, 0.008, 0.014, 0.040)$, which were then convolved to match the dispersion of the VANDELS spectra (2.5 Å) and resampled onto a grid of 1 Å per pixel. In order to properly populate the $Z_\star$ parameter space, we constructed models of intermediate metallicities by interpolating between models of known metallicity. To constrain the overall continuum shape of the observed spectra, we additionally fit for dust attenuation using the FUV attenuation curve parametrization of Salim, Boquien & Lee (2018; following Noll et al. 2009), which utilizes three parameters A_V, δ and B that characterize the normalization, slope and UV bump strength of the dust attenuation curve respectively. Full details of the model fitting and examples are given in Cullen et al. (2019, 2021).

In addition to our fiducial model we also constructed S99 models including the effect of stellar rotation in order to assess the level of systematic uncertainty arising from SPS model choices. The iron abundances corresponding to the models including stellar rotation can be found in Table A1. We chose not to employ the Binary Population and Stellar Synthesis (BPASS) models (v2.2.1 Stanway & Eldridge 2018), though Cullen et al. (2019) explored the effect of substituting the S99 models for BPASS, finding that iron abundances were reduced by ~ 0.1 dex. Although the newer subset of BPASS models (Byrne et al. 2022) include α enhancement, it is not accounted for in the spectra of O/B type stars (Chartab et al. 2024). For our purposes of fitting the FUV, these newer models are essentially unchanged from BPASS v2.2.1.

We used the PYTHON package DYNESTY to fit the spectral models and dust attenuation profile to the FUV spectra. We used a log-likelihood function of the form,

$$\ln(L) = K - \frac{1}{2}\sum_i \left[\frac{(f_i - f(\theta)_i)^2}{\sigma_i^2}\right] = K - \frac{1}{2}\chi^2, \quad (3)$$

where K is a constant, f is the observed flux, $f(\theta)$ is the model flux given a set of parameters θ, and σ is the error on the observed flux. We sum the likelihood over all of the pixels in the spectra corresponding to the stellar continuum, using 'mask 1' from Steidel et al. (2016) to exclude wavelength pixels contaminated by ISM absorption and/or nebular emission lines. For each parameter, the final value was taken as the 50th percentile of the posterior probability distribution, and the 1σ uncertainties were taken as the 68 percentile width. Based on the results of Topping et al. (2020b), we do not report $Z_\star$ estimates for individual spectra with a SNR of < 5.6 per resolution element which, as well as being poorly constrained, are also likely to return biased $Z_\star$ estimates.

From the *VLT*/KMOS sample, 11/32 galaxies yielded robust individual estimates of $Z_\star$, as well as both of the low- and high-$M_\star$ composite spectra. The results of the individual and composite fitting for the *VLT*/KMOS galaxies can be found in Tables 2 and 3, respectively. For consistency, we also re-fit the individual and composite galaxy spectra from the Cullen et al. (2021) Keck/MOSFIRE sample, yielding a further six individual $Z_\star$ estimates. In total therefore our sample contains 17 galaxies within individual $Z_\star$ estimates and four $Z_\star$ estimates from composite spectra binned by stellar mass.

4 THE STELLAR AND GAS-PHASE MZRS

The aim of this work is to constrain the mass-metallicity scaling relations for O/H (tracing the gas phase, Z_g) and Fe/H (tracing massive stars, $Z_\star$) of our sample, and to investigate the typical O/Fe ratio at $z \simeq 3.5$. In this section, we summarize our findings regarding both gas-phase and stellar MZRs (shown in Fig. 4) and draw comparisons with relations from the literature (Fig. 5). Subsequently, for galaxies with simultaneous determinations of both O/H and Fe/H, we evaluate O/Fe in order to ascertain the degree of α-enhancement present in our sample, as well as evidence for any dependence on stellar mass (Fig. 6).

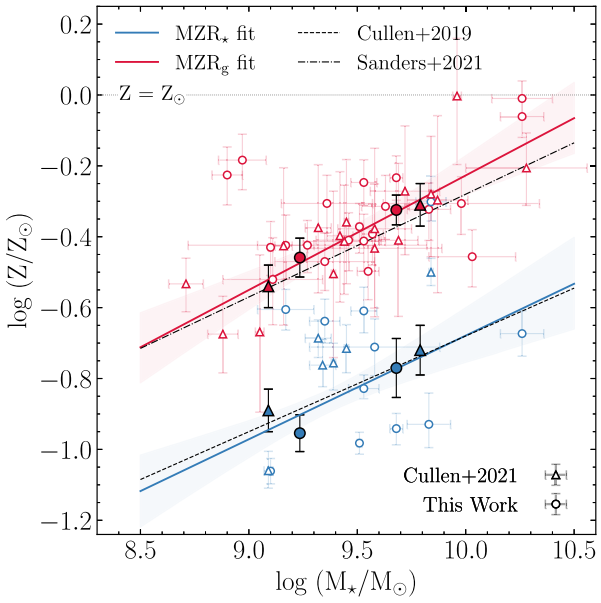

Figure 4. The gas-phase metallicities (red, tracing O/H) and stellar metallicities (blue, tracing Fe/H) for the *VLT*/KMOS (circles) and Keck/MOSFIRE galaxies (triangles; Cullen et al. 2021) as a function of stellar mass. The individual galaxy determinations are shown by unfilled data points whilst the composite determinations are shown by the filled and outlined data points. The best fitting gas-phase and stellar MZRs (MZR$_g$ and MZR$_\star$) are shown by red and blue solid lines, respectively. 1σ uncertainties on each relationship are shown by shaded regions in their respective colours. For comparison, we plot literature relations at similar redshifts: the black dashed lines shows the MZR$_g$ from Sanders et al. (2021) and the black dot–dashed line shows the MZR$_\star$ from Cullen et al. (2019). The solar metallicity value (i.e. $\log(Z_\star/Z_\odot) = 0.0$) is marked by the horizontal grey dashed line.

4.1 The gas-phase MZR

In Fig 4, we show the gas-phase MZR (hereafter MZR$_g$) for our full sample assuming our fiducial strong line calibration scheme (B18). It can be seen that for both individual galaxies and composite spectra, we find a clear trend between Z_g and $M_\star$, such that gas-phase O/H increases with increasing stellar mass. Fitting a linear relation in log-log space to the composite data yields a relationship of the form,

$$\log(Z_g/Z_\odot) = (0.32 \pm 0.09)m_{10} - (0.23 \pm 0.05), \quad (4)$$

where $m_{10} = \log(M_\star/10^{10}M_\odot)$. This relation is in excellent agreement with the $z \simeq 3.3$ MZR$_g$ of Sanders et al. (2021) and Cullen et al. (2021), which infer a slope and normalization of 0.29 ± 0.02 and -0.28 ± 0.03, respectively (also using the calibration of B18).

As can be seen from Fig. 5, the slope of the $z = 3.5$ relation is consistent with the low mass slope (below $M_\star \simeq 10^{10}\,M_\odot$) of the $z \sim 0$ relations of Sanders et al. (2021) and Andrews & Martini (2013; the latter based on a direct T_e-based determination). This comparison implies an approximately constant offset of $\simeq 0.3$ dex in $\log(Z_g)$ (i.e. a factor 2 in O/H) between $z = 0$ and $z = 3.5$ in the stellar mass range $M_\star = 10^{8.5} - 10^{10.5}\,M_\odot$. At $z = 3.5$, we find that the oxygen abundance across this stellar mass range varies from

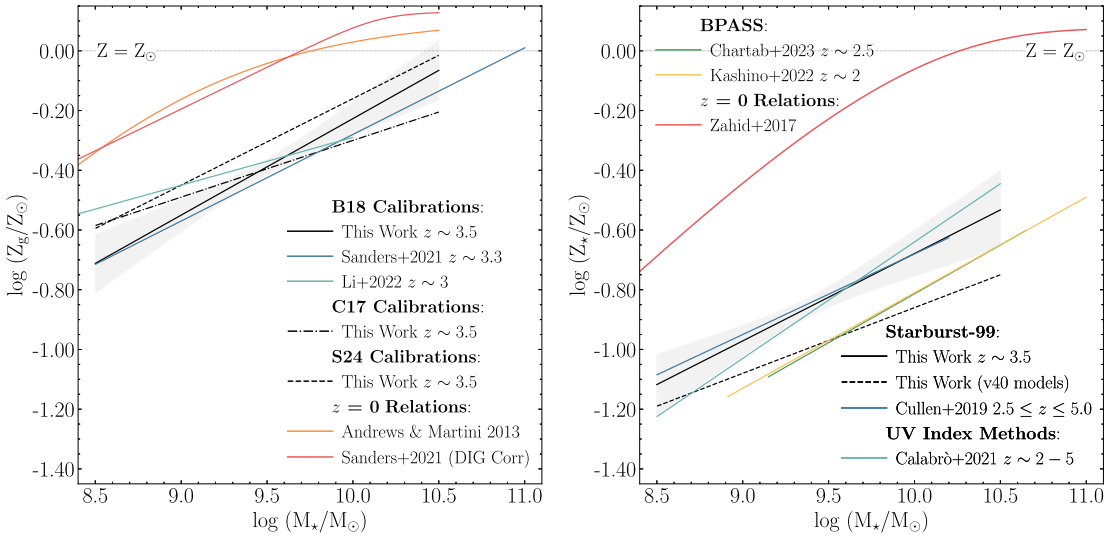

Figure 5. A comparison between the gas-phase (left-hand panel) and stellar (right-hand panel) MZRps of this work and other literature relations at $z = 0$ and $z \simeq 2 - 3$. Our fiducial relations (from Fig. 4) are plotted as solid black lines in each panel. In the left-hand panel we show the MZR_g determined from our sample using the C20 and S24 calibrations as the black dot–dashed and black dashed lines, respectively. Various other literature relations are shown at $z = 3.5$ and $z = 0$, as indicated in the legend (see text for discussion). Despite systematic differences, the gas-phase O abundances at $z = 3.5$ typically fall within the range $\simeq 20 - 70$ per cent solar across the stellar-mass range of our sample, a factor $\simeq 2\times$ lower than local galaxies of the same stellar mass. In the right-hand panel, we show our fiducial $MZR_\star$ along with the relation derived using S99 models including stellar rotation (black dashed line). We show various other literature relations at $z \simeq 2 - 3$ and $z = 0$, as indicated in the legend. Again, despite some differences, the high-redshift stellar Fe abundances typically fall within the range $\simeq 5 - 30$ per cent solar, in this case a factor $\simeq 4 - 5\times$ lower than local galaxies of the same stellar mass.

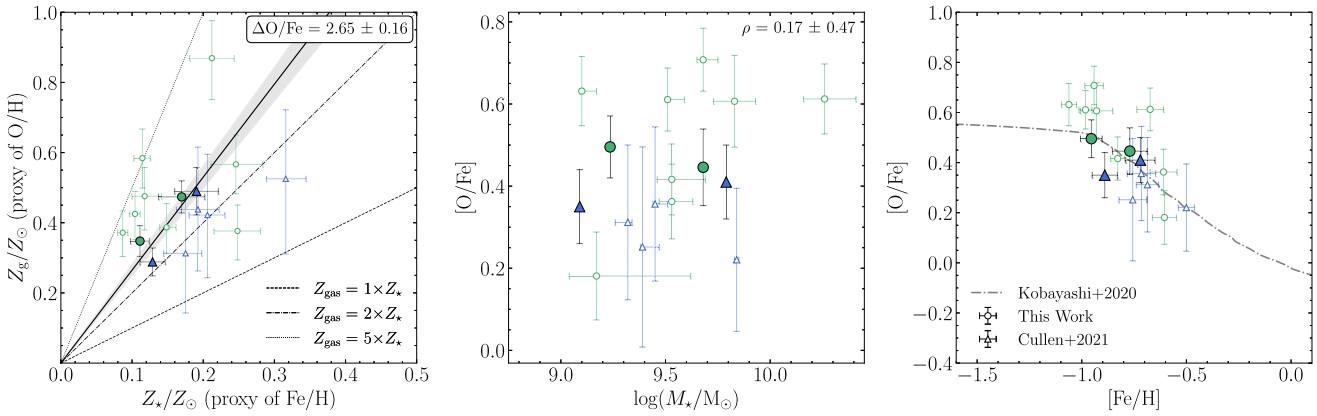

Figure 6. The O/Fe enhancement in our full $z \simeq 3.5$ star-forming galaxy sample. In all panels, galaxies from our *VLT*/KMOS sample are shown as green circles, and galaxies from our Keck/MOSFIRE sample are shown as blue triangles. The filled data points represent values derived from composite spectra, and the unfilled data points show individual galaxy determinations. The left-hand panel shows the $Z_g/Z_\odot$ (tracing O) versus $Z_\star/Z_\odot$ (tracing Fe) relation for our sample. The three black lines denote constant ratios from $1\times$ to $5\times$ the solar O/Fe value (see legend). Our sample lies consistently above the 1:1 line, consistent with an average enhancement of $\langle O/Fe\rangle = 2.65 \pm 0.16(O/Fe)_\odot$, as shown by the solid black line (with the grey shaded region highlighting the 68 per cent confidence interval). The centre panel shows the [O/Fe] versus $M_\star$ relation for our sample. We find that [O/Fe] is not strongly stellar mass-dependent in the range $10^9 < M_\star/M_\odot < 10^{10}$. In the right-hand panels we show the position of our sample in the [O/Fe] versus [Fe/H] diagram (see text for definitions) compared to the model evolutionary sequence for the MW taken from Kobayashi et al. (2020; dot–dashed line). The galaxies in our sample overlap with the knee of the Kobayashi et al. (2020) relation, indicating similarities between our high-redshift systems and the expected chemical evolution of MW-like galaxies.

$0.2 - 0.7\,Z_\odot$, with an average value of $\simeq 0.4\,Z_\odot$ at the median stellar mass of our sample ($M_\star \simeq 10^{9.5}\,M_\odot$). At $z = 0$, the typical O/H at the same stellar mass is $\simeq 0.8\,Z_\odot$. Interestingly, these relations suggest that star-forming galaxies within the first $\simeq 15$ per cent of the Universe's history already contain roughly half the amount of oxygen present in the ISM of local $z = 0$ star-forming galaxies with the same stellar mass.

In Fig. 5 we also show how the form of the MZR_g for our $z = 3.5$ sample changes depending on the choice of strong-line calibration. For the S24 calibration, we measure a slope of 0.29 ± 0.13 consistent with our fiducial model, but offset to higher metallicities with a normalization of -0.16 ± 0.07. For the C20 calibrations, we measure a similar normalization to our fiducial model (-0.30 ± 0.03) but a shallower slope of 0.19 ± 0.07. Clearly, subtle strong-line systematics are present, which we discuss further below. Nevertheless, the range of metallicities is fairly consistent, with O/H $\simeq 40 - 50$ per cent solar of the solar value at $M_\star \simeq 10^{9.5}\,M_\odot$.

We note that below $M_\star \simeq 10^9\,M_\odot$, we currently lack data, and the extrapolation of MZR_g given by equation (4) is naturally highly uncertain in this stellar mass regime. Indeed, recent *JWST* studies, using similar strong-line approaches, find evidence for a flattening of the MZR_g slope at $M_\star \lesssim 10^9\,M_\odot$ (e.g. He et al. 2023; Li et al. 2023). For the purposes of this work we restrict ourselves to the simple linear form of the MZR_g (in log-log space), noting that our results are most robust in the stellar mass range $M_\star \simeq 10^9 - 10^{10}\,M_\odot$.

4.2 The stellar MZR

In Fig 4, we also show the stellar MZR (hereafter $MZR_\star$) for our full sample assuming our fiducial S99 model. Despite fewer individual $Z_\star$ constraints, both the individual galaxies and the composites again show clear evidence for a stellar MZR. Fitting to the composite data yields a relationship of the form,

$$\log(Z_\star/Z_\odot) = (0.30 \pm 0.11)m_{10} - (0.68 \pm 0.07), \quad (5)$$

where $m_{10} = \log(M_\star/10^{10}M_\odot)$. This relation is in excellent agreement with the results presented Cullen et al. (2019), who measured a slope and normalization of 0.27 ± 0.06 and -0.68 ± 0.04, respectively, based on fitting composite spectra generated from a much larger VANDELS sample ($\simeq 700$ individual galaxies) using S99 models (see also Cullen et al. 2021). Across the stellar mass range of our sample, we find that stellar Fe abundances vary from $0.08 - 0.3\,Z_\odot$, with a mean value of $\simeq 0.15\,Z_\odot$ at $M_\star \simeq 10^{9.5}\,M_\odot$.

From Fig. 5 it can again be seen that for $M_\star \lesssim 10^{10}\,M_\odot$ the $MZR_\star$ slope at $z = 3.5$ is consistent with the local Universe determination Zahid et al. (2017). Other local estimates (not shown) yield similar results (e.g. Gallazzi et al. 2005). As with the gas-phase relation, we see an offset in the normalization of the $MZR_\star$, such that the galaxy distribution evolves towards higher metallicity at fixed mass at later times. However, the evolution of stellar Fe abundances is more pronounced than gas-phase O abundances, increasing by $\simeq 0.6$ dex (i.e. a factor of 4) between $z = 3.5$ and $z = 0$ (cf. a factor 2 for O/H). The reason for this is naturally explained by the fact that galaxies at $z = 3.5$ are typically α-enhanced (i.e. Fe-deficient), as we discuss in the following.

We compare our results to other estimates of $MZR_\star$ at $z > 2$ in Fig. 5. We find good agreement between the slope of our relation at $M_\star \lesssim 10^{10}\,M_\odot$ and the relations of Kashino et al. (2022) at $z \simeq 2.2$ (slope $= 0.32 \pm 0.03$) and Chartab et al. (2024) at $z \simeq 2.5$ (slope $= 0.32 \pm 0.02$).[6] Both Kashino et al. (2022) and Chartab et al. (2024) use a methodology similar to ours (i.e. full spectra fitting to rest-frame FUV spectra) but use the BPASS SPS models and analyse two independent data sets: zCOSMOS (Lilly et al. 2007) and LATIS (Newman et al. 2020) respectively. We also find good agreement with the $2 < z < 5$ relation of Calabrò et al. (2021), who estimate stellar metallicities using a subset rest-frame UV absorption features for a sample of VANDELS galaxies. The consistency in the resulting $MZR_\star$ shape regardless of the stellar model choice and methodology is therefore reassuring. The normalization of both $z \simeq 2.5$ relations is lower by $\simeq 0.12$ dex compared to our fiducial relation, but this can be reconciled by the fact that fitting with BPASS models is known to return a lower $Z_\star$ by roughly 0.1 dex (e.g. Cullen et al. 2019). We note that fitting with the S99 models including stellar rotation also yields lower estimates of $Z_\star$ and reduces the difference with the BPASS estimates (Fig. 5). Nevertheless, across a variety of modelling assumptions and data sets, a consistent pictures emerges of galaxies at $z \simeq 2 - 3$, with $M_\star \lesssim 10^9 - 10^{10}\,M_\odot$ having stellar Fe abundances of $\simeq 5$ to 20 per cent of the solar value. Crucially, the estimates *all* fall below the MZR_g at the same redshift.

[6]We note that Chartab et al. (2024) do not provide a linear fit to their data, so this value represents our own fit to their published data.

4.3 Enhanced O/Fe abundance ratios

A comparison the MZR_g and $MZR_\star$ relations in Fig. 4 shows that the gas-phase relation (tracing O/H) is offset by $\simeq +0.4$ dex compared to the stellar relation (tracing Fe/H) across the full stellar mass range. This offset represents direct evidence of enhanced O/Fe ratios (relative to the solar value) in star-forming galaxies at $z = 3.5$. Throughout the discussion below (and the remainder of the paper), we adopt the following definitions:

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

(i.e. the Fe abundance relative to the solar value) and,

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

(i.e. the O/Fe ratio relative to the solar value). In the context of this work, $\log(Z_\star/Z_\odot)$ is an estimate of [Fe/H] and $\log(Z_g/Z_\star)$ is an estimate of [O/Fe]. A value of $[\mathrm{O/Fe}] > 0$ represents an enhancement of O/Fe relative to the solar value (i.e. α-enhancement).

We find that the low- and high-mass composites from the *VLT*/KMOS sample have highly consistent [O/Fe] ratios of 0.49 ± 0.07 and 0.45 ± 0.09 respectively. For the Keck/MOSFIRE stacks of Cullen et al. (2021), we measure [O/Fe] values of 0.35 ± 0.09 for the low-mass composite and 0.41 ± 0.09 for high-mass composite. Each of these values is statistically significant at the $\gtrsim 4\sigma$ level and also consistent with the inverse variance-weighted mean value 0.5 ± 0.03 of the 12/65 individual galaxies with robust O/H and Fe/H abundances in our sample. In the left-hand panel of Fig. 6 we show a linear fit to our composites in the $Z_\star$-Z_g plane, from which we derive a typical value of $\mathrm{O/Fe} = 2.65 \pm 0.16 \times (\mathrm{O/Fe})_\odot$ (i.e. $[\mathrm{O/Fe}] = 0.42 \pm 0.03$). Our new constraint is in excellent agreement with the value of $\mathrm{O/Fe} = 2.57 \pm 0.38 \times (\mathrm{O/Fe})_\odot$ derived from the Keck/MOSFIRE sample in Cullen et al. (2021), and reduces the uncertainty on the degree of enhancement by a factor of ~ 2. Overall, we find that our sample is ubiquitously α-enhanced.

The consistent slopes of the MZR_g and $MZR_\star$ suggest no stellar-mass dependence, which we show explicitly in the middle panel of Fig 6. In the stellar mass range $M_\star = 10^9 - 10^{10}\,M_\odot$ we see no evidence for a deviation from the full-sample average. At face value, this result suggests that the same physical processes are responsible for setting the shape of the low mass slope of the stellar and gas-phase MZRs (namely large scale galactic outflows; e.g. Cullen et al. 2019, Sanders et al. 2021), and that the typical star-formation histories are similar for galaxies with $M_\star = 10^9\,M_\odot$ and $M_\star = 10^{10}\,M_\odot$. We discuss the physical interpretation of our results in more detail in Section 5. It is worth noting, however, that this conclusion is calibration-dependent. Applying the strong-line calibration of C20 yields a shallower MZR_g (Fig. 5) such that lower-mass galaxies are more α-enhanced. In Fig. A1 and Appendix A we explore the extent to which the choice of strong-line calibration affects the derived value of O/Fe. Crucially, whilst the degree of α-enhancement varies across the tested calibration schemes and SPS models, all cases yield [O/Fe] ratios enhanced by a factor of $\gtrsim 2.5$.

In the right-hand panel of Fig. 6 we show the location of our $z = 3.5$ star-forming galaxies in the [O/Fe] versus [Fe/H] plane compared to the one-zone Galactic Chemical Evolution model of Kobayashi et al. (2020). The Kobayashi et al. (2020) model is based on an assumed MW-like star formation history (SFH) and is shown

to match a variety of stellar archaeological data in the MW. For our fiducial strong-line calibration and SPS model assumptions, our data provide a good match to the predicted evolution of MW-like galaxies. The $z = 3.5$ galaxies fall close to the 'knee' in the relation at [Fe/H] $\simeq -1.0$ (i.e. the point at which Fe enrichment from SN Ia initiates the transition to lower O/Fe). The consistency between our high-redshift systems and the trend of MW stars represented by the Kobayashi et al. (2020) is encouraging.

Our results are generally in good agreement with other works in the literature that have estimated O/Fe at $z > 2$ (e.g. Steidel et al. 2016; Cullen et al. 2019; Topping et al. 2020a; Kashino et al. 2022; Strom et al. 2022; Chartab et al. 2024). Of the studies shown in Fig. 5, Kashino et al. (2022) and Strom et al. (2022) estimate O/Fe $= 2.12 \pm 0.54 \times (\mathrm{O/Fe})_{\odot}$ and O/Fe $\simeq 2.2 \times (\mathrm{O/Fe})_{\odot}$ respectively; Chartab et al. (2024) measure an average enhancement of O/Fe $\simeq 3.7 \pm 0.4 \times (\mathrm{O/Fe})_{\odot}$. Steidel et al. (2016) and Topping et al. (2020a) report larger values of up to $5 - 7 \times (\mathrm{O/Fe})_{\odot}$ but crucially all studies agree that star-forming galaxies at $z > 2$ are commonly α-enhanced (generally, studies utilizing the BPASS SPS models suggest lower [Fe/H] and hence larger [O/Fe]; Cullen et al. 2021). Combined, the results of all these studies suggest a picture in which the ISM of star-forming galaxies at $z > 2$ is dominated by CCSNe enrichment.

5 CONSTRAINING THE MASS-DEPENDENCE OF GALAXY OUTFLOWS

Previous studies have demonstrated how the power-law slope of the MZR (both stellar and gas-phase) below $\log(M_{\star}) \simeq 10.5$ can be used to constrain the mass-dependence of galaxy-scale outflows (e.g. Cullen et al. 2019; Sanders et al. 2021; Chartab et al. 2024). The strength of galaxy-scale outflows is also predicted to affect the degree of α-enhancement in the ISM (e.g. Weinberg et al. 2017, hereafter W17). In this section, we employ a chemical evolution model to simultaneously interpret the scaling of O and Fe abundances with stellar mass as well as the typical O/Fe ratios in our sample.

5.1 An analytical chemical evolution model

We employ the analytical one-zone chemical evolution model of W17 which incorporates a realistic delay time distribution (DTD) for Type-Ia SNe, allowing the evolution of iron-peak elements to be tracked separately to α-elements. The models assume a constant (i.e. time-independent) star-formation efficiency (SFE; SFE $= \dot{M}_{\star}/M_{\mathrm{gas}}$) in accordance with the 'linear Schmidt law' and mass outflow rates are assumed to scale with the star-formation rate, with the scaling factor described by the mass-loading factor ($\eta = \dot{M}_{\mathrm{out}}/\dot{M}_{\star}$). Therefore, the accretion and loss of gas (i.e. the baryon cycle) as a function of time is defined by the SFH, the star formation efficiency (which we characterize by the gas depletion time-scale $t_{\mathrm{dep}} = 1/\mathrm{SFE}$), and the mass loading factor η. We adopt a linearly rising SFH to best match the predicted SFH of star-forming galaxies (e.g. Reddy et al. 2012; Ciesla, Elbaz & Fensch 2017), and use the t_{dep}-$M_{\star}$ scaling relation of Tacconi et al. (2018) to set the star formation efficiency.

Unless otherwise stated, we adopt the fiducial model parameters from W17. The Type-Ia DTD is exponentially declining with a minimum delay time of 0.15 Gyr and an *e*-folding time-scale of 1.5 Gyr. We assume a Kroupa (2001) IMF, and the CCSNe yields of Chieffi & Limongi (2004) and Limongi & Chieffi (2006) in which $1.5\mathrm{M}_{\odot}$ of O and $0.12\mathrm{M}_{\odot}$ of Fe are returned to the ISM for every $100\mathrm{M}_{\odot}$ of star formation. The corresponding return of Fe from Type-Ia SNe is 0.17 $\mathrm{M}_{\odot}$. The model assumes that gas accretion from the IGM is pristine, that outflowing gas is of equal metallicity to the metallicity of the star-forming ISM, and that enriched material from CCSNe are instantaneously returned and mixed into the star-forming ISM. The resulting time evolution of the abundances of O and Fe is given by equations 55 − 58 in W17.

Our aim is to explore whether, using this model, we can find a simple scaling between η and $M_{\star}$ that can simultaneously explain both the stellar and gas-phase MZRs and the O/Fe ratios of our sample. Both theory and observations support a scenario in which lower-mass galaxies exhibit higher η values compared to higher-mass galaxies (i.e. gas is more efficiently ejected in lower mass galaxies; Hayward & Hopkins 2017; Llerena et al. 2023). Ejection is also more likely to occur at velocities greater than their respective escape velocities, because momentum sources such as supernovae are more effective in lower mass galaxies (Chisholm et al. 2017). We employ the simple scaling $\eta = \alpha(M_{\star}/10^{10}\mathrm{M}_{\odot})^{\beta}$ suggested by the FIRE simulations (Muratov et al. 2015). This simple parametrization has been shown to effectively reproduce the power-law slope of the MZR (e.g. Cullen et al. 2019).

We also found that it was necessary to increase the CCSNe oxygen yields to match the data. The default W17 yields result in a maximum of [O/Fe] $\simeq 0.4$ which falls below a number of our estimates (as well as other [O/Fe] estimates at high-redshift; e.g. Steidel et al. 2016; Chartab et al. 2024). Evidence from stellar data suggests that [O/Fe] can reach values up to $\simeq 0.6$ (e.g. APOGEE DR17 abundances; Abdurro'uf et al. 2022). Alternative stellar yield estimates are also consistent with a [O/Fe] $\simeq 0.6$ plateau (e.g. Nomoto et al. 2006; Kobayashi et al. 2006, 2020). Rather than fixing our yields using an alternative prescription, we include a free parameter in our model which scales the W17 CCSNe O yield, denoted $f_{\mathrm{O_{CC}}}$. We note that the need to scale the oxygen yields is in part a consequence of the underlying simplifying assumptions of the model, namely the assumption that inflowing gas is pristine and that outflows are not more enriched than the ISM. For example, including enriched inflows from previous cycles of star formation would increase the degree of α-enhancement, as the composition of the reaccreted gas in the inflows would reflect an earlier ISM enriched primarily by CCSNe. These enriched inflows might mitigate the need to enhance the CCSNe O yields. However, it is unlikely that the effect of these assumptions will change the mass scaling of the outflows, which is driven primarily by the mass scaling of the individual abundances.

5.2 Constraints on the η-$M_{\star}$ relation

For a given composite galaxy and set of model parameters (α, β, and $f_{\mathrm{O_{CC}}}$), we use our model to predict both the O and Fe abundances at the average redshift (i.e. time) and $M_{\star}$. We sampled the entire parameter space and derive the best-fitting values of the parameters with DYNESTY, employing a Gaussian likelihood (equation 3). We assumed a flat prior on each parameter: $0 < \alpha < 12$; $-1.4 < \beta < 0.4$ and $0.5 < \mathrm{f_{O_{CC}}} < 4$. The best-fitting parameters are shown in the right-hand panel of Fig. 7.

The consistency between the best-fitting model and data is highlighted in the left-hand and centre panels of Fig. 7. In the left-hand panel, we have converted the model into $M_{\star} - \mathrm{O/H}$ and $M_{\star} - \mathrm{Fe/H}$ relations at the median redshift of the sample ($z = 3.5$). It can be seen that both the slope and normalization of both relations are reproduced well. Similarly, our model predicts minimal differences from the measured [O/Fe] across all of our individual galaxies and composites. We show this comparison in the centre panel as a function of stellar mass, again highlighting the excellent agreement. Overall, we find that this relatively simple chemical evolution model, which

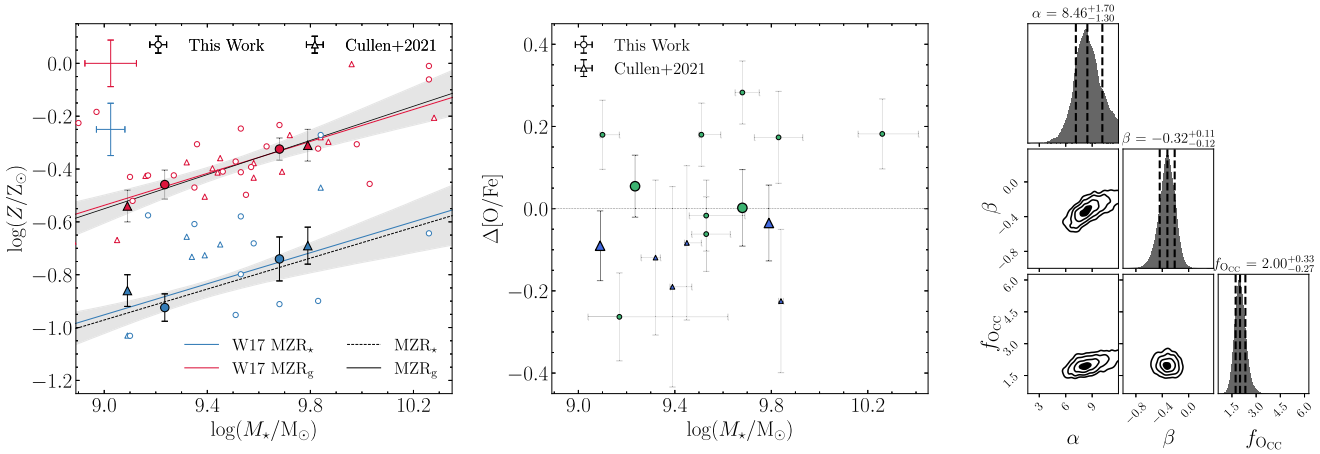

Figure 7. The results of chemical evolution model fitting to the O and Fe abundances in our sample, showing our constraints on the stellar-mass scaling of galaxy-scale outflows. Our principal result is that our data are well described by a model (following the framework of W17) in which the mass-loading parameter η ($= \dot{M}_{\rm out}/\dot{M}_\star$) scales with stellar mass as $\eta \propto M_\star^\beta$ where $\beta = -0.32 \pm 0.12$. In the left-hand and centre panels, we show a comparison between the best-fitting model and our data. In each panel the *VLT*/KMOS galaxies are shown as circular data points, and Keck/MOSFIRE galaxies as triangular data points. Composite galaxies are shown as large filled points with error bars, while individual galaxies are small open points. The median uncertainty for all individual galaxies is indicated in each panel. In the left-hand panel, we show our data and the resulting gas-phase and stellar MZRs at $z \simeq 3.5$ for our best-fitting model as the red and blue solid lines, respectively. The resulting relations are in excellent agreement with the data. In the centre panel, we show the difference between our measured [O/Fe] and the predicted [O/Fe] from our model as a function of galaxy stellar mass. Our best fitting model accurately predicts the [O/Fe]-enhancement of our sample and supports the stellar mass independence of α-enhancement. The right-hand panel shows a corner plot giving the posterior probability distributions for the three parameters in our chemical evolution model fitting.

fundamentally assumes (i) linearly rising star formation histories, (ii) the Tacconi et al. (2018) $t_{\rm dep} - M_\star$ relation, and (iii) a power-law anticorrelation between η and $M_\star$, does a remarkably good job of reproducing the O- and Fe-abundance scaling relations and the flat [O/Fe] ratios of our sample.

In Fig. 8 we show the predicted time evolution of O and Fe abundances as a function of stellar mass for the best-fitting model, where some of the fundamental physical trends are apparent. First, galaxies with lower $M_\star$ have lower [Fe/H]; in the context of our modelling, this is due to more efficient removal of enriched content at lower $M_\star$ (i.e. larger η), and higher gas fractions diluting the ISM at lower $M_\star$ (Tacconi et al. 2018). Secondly, [O/Fe] is approximately constant at all [Fe/H], which is a consequence of the shallow $t_{\rm dep}$-$M_\star$ relation and implies similar star formation time-scales for galaxies in the stellar mass range $10^9 - 10^{10}\,{\rm M}_\odot$. We find a best-fitting oxygen yield enhancement of $f_{\rm OCC} = 2.00^{+0.33}_{-0.27}$, which results in an [O/Fe] plateau of [O/Fe] $= 0.73 \pm 0.06$. This value is somewhat higher than the [O/Fe] $\simeq 0.6$ plateau of the Kobayashi et al. (2020) models, but is broadly consistent within the uncertainties and probably acceptable given the significant uncertainties in supernova yields (e.g. Romano et al. 2010).

In Fig. 9 we show our best-fitting relation between η and $M_\star$ in comparison to predictions from various cosmological simulations and the recent constraint of Chartab et al. (2024). The Chartab et al. (2024) constraint was derived by applying the W17 modelling framework to their Fe/H $- M_\star$ relation derived from the LATIS survey. In the mass regime to which we are sensitive ($M_\star \lesssim 10^{10.3}\,{\rm M}_\odot$), the shape of the relations are in excellent agreement. Chartab et al. (2024) find $\eta \propto M_\star^{-0.35\pm0.01}$ compared to our constraint of $\eta \propto M_\star^{-0.32\pm0.12}$. Unfortunately, our data do not probe the mass regime in which Chartab et al. (2024) find evidence for an upturn in η which they attribute to AGN-driven winds (at $M_\star \gtrsim 10^{10.5}\,{\rm M}_\odot$). We find systematically larger values of η at fixed $M_\star$ by $0.1 - 0.2$ dex but note that the absolute normalization of the relation is strongly

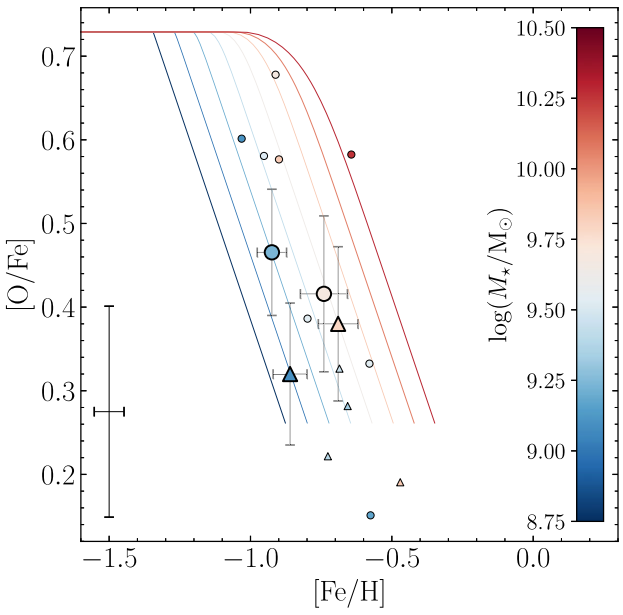

Figure 8. The resulting tracks in the [O/Fe] versus [Fe/H] plane derived from our chemical evolution model fitting to the O and Fe abundances in our sample. *VLT*/KMOS galaxies are shown as circular data points and Keck/MOSFIRE galaxies as triangular points. Composite galaxies are shown as large filled points with error bars while individual galaxies are smaller open points, with their average errors given by the error bar in the bottom left of the plot. Galaxies are colour-coded according to their stellar mass, shown by the inset colourbar on the right. The coloured lines show the mass-dependent chemical evolution tracks for our best-fitting model parameters. These tracks are also temporal sequences, meaning that it is possible to trace the chemical abundances of each element as a function of time from the onset of star formation. In general, all data points fall close to the track corresponding to their stellar mass.

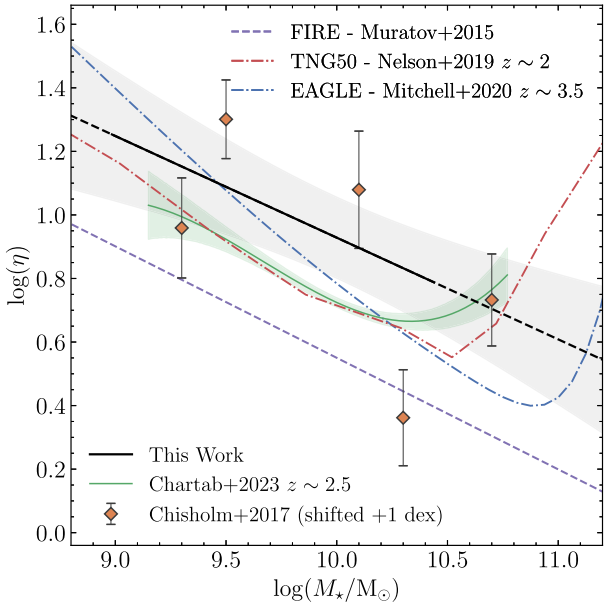

Figure 9. The mass loading factor (η) versus stellar mass ($M_\star$) relation derived from fitting to the O and Fe abundances in our sample within the chemical evolution modelling framework of Weinberg et al. (2017, hereafter W17; black line). The dashed black line shows the extrapolation of our fitted relation outside the mass regime of our sample. The green curve shows the best-fit relation from Chartab et al. (2024), derived by fitting to their stellar MZR relation at $z \simeq 2.5$ (also using the W17 analytic models). We also show the $\eta - M_\star$ relations derived from the FIRE (Muratov et al. 2015; dashed purple), EAGLE (Muratov et al. 2015; dot–dashed blue) and TNG50 (Nelson et al. 2019; dot-dashed blue) cosmological simulations. The orange diamond data points show the mass loading factors and stellar masses of four local star-forming galaxies taken from Chisholm et al. (2017). The Chisholm et al. (2017) estimates for η were derived using ionized FUV absorption lines, and we find that the absolute values need to be scaled by a factor of $\simeq 10$ to match our relation (see text for discussion). In general, both the observational data and the simulation predictions show a consistent $\eta - M_\star$ scaling, with $-0.4 \lesssim \beta \lesssim -0.3$ (where $\eta \propto M_\star^\beta$). Differences in the overall normalization are relatively minor considering the various systematic uncertainties involved (e.g. yields, abundance estimates, the radius at which η is defined in simulation analyses).

dependent on the measured O and Fe abundances and subject to larger systematic uncertainties. The mass dependence of η derived here is also consistent with other similar models in the literature; for example, Cullen et al. (2019) and Sanders et al. (2021), which find $\beta \simeq -0.4$ and $\beta = -0.35 \pm 0.01$, respectively.

In Fig. 9 we also show data from Chisholm et al. (2017) who estimated η for a sample of local star-forming galaxies via ionized FUV ISM absorption lines. Chisholm et al. (2017) find a mass scaling of $\eta \propto M_\star^{-0.4}$ but with a normalization lower by a factor of $\simeq 10$. In Fig. 9 we have scaled their η estimates by this factor to compare with our relation and highlight the consistent mass dependence. The discrepancy in normalization is not surprising, since the Chisholm et al. (2017) relation traces only the ionized gas phase. Multiple studies have found that the ionized phase accounts for only a small fraction of the total outflowing gas, which is probably dominated by the molecular phase (e.g. Fluetsch et al. 2021; Roberts-Borsani 2020; Concas et al. 2022; Llerena et al. 2023; Weldon et al. 2024). Scaling η derived from the ionized gas by a factor $\sim 10 - 100$ to estimate the total mass loading factor is not out of the question.

We compare with the predictions of the TNG50 (Nelson et al. 2019), EAGLE (Mitchell et al. 2020) and FIRE (Muratov et al. 2015) cosmological simulations in Fig. 9. Generally, the agreement is encouraging. In particular, the mass scaling derived from the simulations is in good agreement with our derivation for $M_\star \lesssim 10^{10.5}\,\mathrm{M}_\odot$. Above this mass threshold, simulations that include AGN feedback (TNG-50, EAGLE) predict an upturn on η; however, our data are not sensitive to this mass regime. Differences in normalization are evident (e.g. compared to the FIRE predictions) but we note that the normalization derived from simulations is dependent on the radius at which η is defined. Coupled with the yield and abundance uncertainties that affect our derived relation, we do not consider these discrepancies significant. Crucially, the form of the $\eta - M_\star$ relationship derived from a relatively simple chemical evolution model fit to O and Fe abundances of galaxies at high redshift is fully consistent with the predictions of more detailed physical simulations.

6 CONCLUSIONS

In this work be have performed an analysis of 65 star-forming galaxies at $z = 3.5$ with ground-based near-IR spectra tracing rest-frame optical emission lines and ground-based optical spectra tracing the rest-frame FUV continuum. For both individual galaxies and composite spectra, we have derived estimates of the oxygen abundances from the [O III] λλ4959, 5007, [O II] λλ3726, 3729 and H β nebular emission lines and iron abundances from full spectral fitting to the rest-frame FUV continuum. Combining our abundance determinations with stellar masses inferred from photometric SED fitting, we have derived scaling relations for both oxygen and iron abundances with stellar mass. For the galaxies and composite spectra with both oxygen and iron abundance estimates, we have determined the O/Fe ratio and investigated its dependence on stellar mass. Finally, we have utilized one-zone chemical evolution models to interpret our results in terms of star-formation time-scales and the mass-dependence of large-scale outflows. The main results of our study can be summarized as follows:

(i) Combining dust-corrected strong emission line ratios from our rest-frame optical spectra with the strong-line calibration scheme of B18, we derive gas-phase metallicities (Z_g, a proxy of O/H) for our individual galaxies and composites. Comparing the Z_g measurements of our sample with their stellar masses, we observe a clear gas-phase mass-metallicity ($\mathrm{MZR_g}$) relationship (Fig. 4). Our relationship exhibits a slope of $\mathrm{d}(\log Z_\mathrm{g})/\mathrm{d}(\log M_\star) = 0.32 \pm 0.09$, with Z_g increasing from $\simeq 0.2\mathrm{Z}_\odot$ at $M_\star = 10^{8.5}\,\mathrm{M}_\odot$ to $\simeq 0.7\,\mathrm{Z}_\odot$ at $M_\star = 10^{10.5}\,\mathrm{M}_\odot$.

(ii) From fitting to the rest-frame FUV continuum spectra we derive stellar metallicities ($Z_\star$, a proxy for Fe/H) and observe a clear stellar mass-metallicity ($\mathrm{MZR}_\star$) relationship in our data (Fig. 4). The slope of the $\mathrm{MZR}_\star$ is consistent with the slope of the $\mathrm{MZR_g}$, but exhibits a constant offset to lower $Z_\star$ across the fitting range of stellar mass. Across the same range of mass ($M_\star = 10^{8.5-10.5}\,\mathrm{M}_\odot$), we observe that $Z_\star$ increases from $\simeq 0.08\,\mathrm{Z}_\odot$ to $\simeq 0.5\mathrm{Z}_\odot$.

(iii) For the galaxies and composites for which we have determinations of both $Z_\star$ and $Z_\mathrm{g}$we calculate the O/Fe ratio. We find that our sample displays super-solar O/Fe ratios (i.e. α-enhancement; Fig. 6) and find no evidence for a dependence on stellar mass in the range $M_\star = 10^9 - 10^{10}\,\mathrm{M}_\odot$. Fitting to our data yields an average enhancement of $(\mathrm{O/Fe}) = 2.65 \pm 0.16 \times (\mathrm{O/Fe})_\odot$. Our new estimates add further support to a picture in which star-forming galaxies at $z > 2$ are ubiquitously α-enhanced.

(iv) Using the one-zone analytical chemical evolution models of W17 we fit for the relative O and Fe abundances of our sample in the stellar mass range $M_\star = 10^9 - 10^{10}\,\mathrm{M}_\odot$. We find that a model in

which the gas depletion scales with stellar mass following Tacconi et al. (2018), and the mass-loading of stellar winds scales as $\eta \propto M_{\star}^{-0.32\pm0.12}$ provides an excellent description of our data. In this modelling framework, it is an increasing efficiency of stellar winds at lower stellar mass that is driving the gas-phase and stellar MZRs. The enhanced O/Fe ratios are a product of delayed Fe enrichment from Typa-Ia supernovae. Interestingly, this $\eta - M_{\star}$ scaling is in good agreement with the predictions of hydrodynamical simulations (e.g. Muratov et al. 2015; Nelson et al. 2019; see Fig. 9).

The sample sizes with both $Z_{\star}$ and $Z_{\rm g}$ determinations remain small, and while we have addressed some of the systematics involved in our determinations, complete high-redshift strong-line calibration schemes and stellar population models remain active areas of research and development. Future surveys, such as AURORA (PI: Shapley, Sanders; GO1914) and EXCELS (PI: Carnall, Cullen; GO3543) will supply additional data and aid the assessment of these systematic effects, with which we will be able to further assess the degree of α-enhancement in high-redshift star-forming galaxies.

ACKNOWLEDGEMENTS

TMS, FC, KZA-C, and DS acknowledge support from a UKRI Frontier Research Guarantee Grant (PI :Cullen; grant reference: EP/X021025/1). RJM, RB, LB, JSD, CTD, MLH, FYL, and DJM acknowledge the support of the Science and Technology Facilities Council. JSD also thanks the Royal Society for their support through a Royal Society Research Professorship. ACC thanks the Leverhulme Trust for their support via the Leverhulme Early Career Fellowship scheme.

This research utilized ASTROPY, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2022), NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), DYNESTY (Speagle 2020) and NASA's Astrophysics Data System Bibliographic Services.

DATA AVAILABILITY

The VANDELS spectroscopic data utilized in this work is publicly available, and can be accessed through the ESO Science Portal. All other data will be shared by the corresponding author upon reasonable request.

REFERENCES

Abdurro'uf et al., 2022, ApJS, 259, 35
Andrews B. H., Martini P., 2013, ApJ, 765, 140
Arellano-Córdova K. Z., Rodríguez M., 2020, MNRAS, 497, 672
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Astropy Collaboration, 2022, ApJ, 935, 167
Bian F., Kewley L. J., Dopita M. A., 2018, ApJ, 859, 175 (B18)
Byrne C. M., Stanway E. R., Eldridge J. J., McSwiney L., Townsend O. T., 2022, MNRAS, 512, 5329
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
Chabrier G., 2003, PASP, 115, 763
Chartab N., Newman A. B., Rudie G. C., Blanc G. A., Kelson D. D., 2024, ApJ,960,73
Chieffi A., Limongi M., 2004, ApJ, 608, 405
Chisholm J., Tremonti C. A., Leitherer C., Chen Y., 2017, MNRAS, 469, 4831
Chruślińska M., Pakmor R., Matthee J., Matsuno T., 2024, A&A, 686, A186
Ciesla L., Elbaz D., Fensch J., 2017, A&A, 608, A41
Clarke L. et al., 2023, ApJ, 957, 81
Concas A. et al., 2022, MNRAS, 513, 2535
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Cooke R., Pettini M., Steidel C. C., Rudie G. C., Nissen P. E., 2011, MNRAS, 417, 1534
Cullen F., Cirasuolo M., McLure R. J., Dunlop J. S., Bowler R. A. A., 2014, MNRAS, 440, 2300
Cullen F. et al., 2019, MNRAS, 487, 2038
Cullen F. et al., 2021, MNRAS, 505, 903
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 (C20)
Curti M. et al., 2023, MNRAS, 518, 425
Davies R. I. et al., 2013, A&A, 558, A56
De Cia A., Ledoux C., Mattsson L., Petitjean P., Srianand R., Gavignaud I., Jenkins E. B., 2016, A&A, 596, A97
Fluetsch A. et al., 2021, MNRAS, 505, 5753
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Garilli B. et al., 2021, A&A, 647, A150
Grogin N. A. et al., 2011, ApJS, 197, 35
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Harris C. R. et al., 2020, Nature, 585, 357
Hayden-Pawson C. et al., 2022, MNRAS, 512, 2867
Hayward C. C., Hopkins P. F., 2017, MNRAS, 465, 1682
He X. et al., 2023, ApJL, 960, L13
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jeong M.-S. et al., 2020, ApJ, 902, L16
Kashino D. et al., 2022, ApJ, 925, 82
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., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, ApJ, 653, 1145
Kobayashi C., Karakas A. I., Lugaro M., 2020, ApJ, 900, 179
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
Kroupa P., 2001, MNRAS, 322, 231
Leitherer C. et al., 1999, ApJS, 123, 3
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
Li M. et al., 2023, ApJ, 955, L18
Lilly S. J. et al., 2007, ApJS, 172, 70
Limongi M., Chieffi A., 2006, ApJ, 647, 483
Llerena M. et al., 2023, A&A, 676, A53
McLean I. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham. p. 84460J
McLure R. J. et al., 2018, MNRAS, 479, 25
Maiolino R., Mannucci F., 2019, A&AR, 27, 3
Maoz D., Mannucci F., 2012, PASA, 29, 447
Matthee J., Schaye J., 2018, MNRAS, 479, L34
Mitchell P. D., Schaye J., Bower R. G., Crain R. A., 2020, MNRAS, 494, 3971
Momcheva I. G. et al., 2016, ApJS, 225, 27
Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691
Nelson D. et al., 2019, MNRAS, 490, 3234
Newman A. B. et al., 2020, ApJ, 891, 147
Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793

Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nucl. Phys. A, 777, 424
Panter B., Jimenez R., Heavens A. F., Charlot S., 2008, MNRAS, 391, 1117
Pentericci L. et al., 2018, A&A, 616, A174
Péroux C., Howk J. C., 2020, ARA&A, 58, 363
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., 2020, ApJ, 902, 123
Roberts-Borsani G. W., 2020, MNRAS, 494, 4266
Romano D., Karakas A. I., Tosi M., Matteucci F., 2010, A&A, 522, A32
Salim S., Boquien M., Lee J. C., 2018, ApJ, 859, 11
Sanders R. L. et al., 2020, MNRAS, 491, 1427
Sanders R. L. et al., 2021, ApJ, 914, 19
Sanders R. L., Shapley A. E., Topping M. W., Reddy N. A., Brammer G. B., 2024, ApJ, 962, 24 (S24)
Savaglio S. et al., 2005, ApJ, 635, 260
Schreiber C. et al., 2018, A&A, 618, A85
Shapley A. E. et al., 2017, ApJ, 846, L30
Shapley A. E. et al., 2019, ApJ, 881, L35
Skelton R. E. et al., 2014, ApJS, 214, 24
Speagle J. S., 2020, MNRAS, 493, 3132
Stanway E. R., Eldridge J. J., 2018, MNRAS, 479, 75
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
Stott J. P. et al., 2016, MNRAS, 457, 1888
Strom A. L., Rudie G. C., Steidel C. C., Trainor R. F., 2022, ApJ, 925, 116
Tacconi L. J. et al., 2018, ApJ, 853, 179
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
Tremonti C. A. et al., 2004, ApJ, 613, 898
Troncoso P. et al., 2014, A&A, 563, A58
Velichko A., De Cia A., Konstantopoulou C., Ledoux C., Krogager J.-K., Ramburuth-Hurt T., 2024, A&A, 685, A103
Venturi G. et al., 2024, preprint (arXiv:2403.03977)
Virtanen P. et al., 2020, Nature Methods, 17, 261
Weinberg D. H., Andrews B. H., Freudenburg J., 2017, ApJ, 837, 183
Weldon A. et al., 2024, MNRAS, 531, 4560
Zahid H. J., Kudritzki R.-P., Conroy C., Andrews B., Ho I. T., 2017, ApJ, 847, 18

APPENDIX A: SYSTEMATIC UNCERTAINTIES IN DETERMINING GALAXY ABUNDANCES

The primary results of this work utilized the B18 calibration scheme and S99 non-rotating models (v00). Here, we show how different choices can affect the primary result of this work. In Fig. A1 we show versions of our main results for the following combinations of strong-line calibration and stellar population model: (i) the C20 calibrations [derived from local SDSS star-forming galaxies, to which we have added the Curti et al. (2023) Ne3O2 ratio calibration] with the Starbust99-v00 models; (ii) the S24 calibration (defined using auroral line measurements from early *JWST* measurements of galaxies at $2 \leq z \leq 9$) with the Starbust99-v00 models and, (iii) our fiducial B18 calibration and Starburst99-v40 (i.e. rotating) models. The abundances corresponding to these systematic choices for the individual and composite samples can be found in Table A1.

We find that slope of the stellar MZR is the same for both the rotating and non-rotating models, but that the stellar Fe abundances of the rotating models are $\simeq 0.1$ dex lower. For the O abundances derived from strong line calibrations, we find that all three calibrations yield a consistent median value of $\mathrm{dlog}(Z_g/Z_\odot) \simeq 0.4$ but that the slope and scatter of the $\mathrm{MZR_g}$ are clearly affected. We find that the C20 calibrations yield the smallest scatter and shallowest slope with $\mathrm{d}(\log Z_g)/dM_\star = 0.16 \pm 0.07$. Ultimately, all combinations predict ubiquitous α-enhancement with average values int he range $\langle \mathrm{O/Fe} \rangle \simeq 2 - 3 \times (\mathrm{O/Fe})_\odot$. However, the shallow slope given by the C20 calibrations implies a decrease in O/Fe at higher stellar masses, which would have important consequences for interpreting the outflow strengths and star-formation time-scales.

In the near future *JWST* programs such as AURORA (PI: Shapley, Sanders; GO1914) and EXCELS (PI: Carnall, Cullen; GO3543) will help to resolve this issue by combining direct T_e-based O/H and stellar Fe/H estimates for samples of galaxies at $z > 2$.

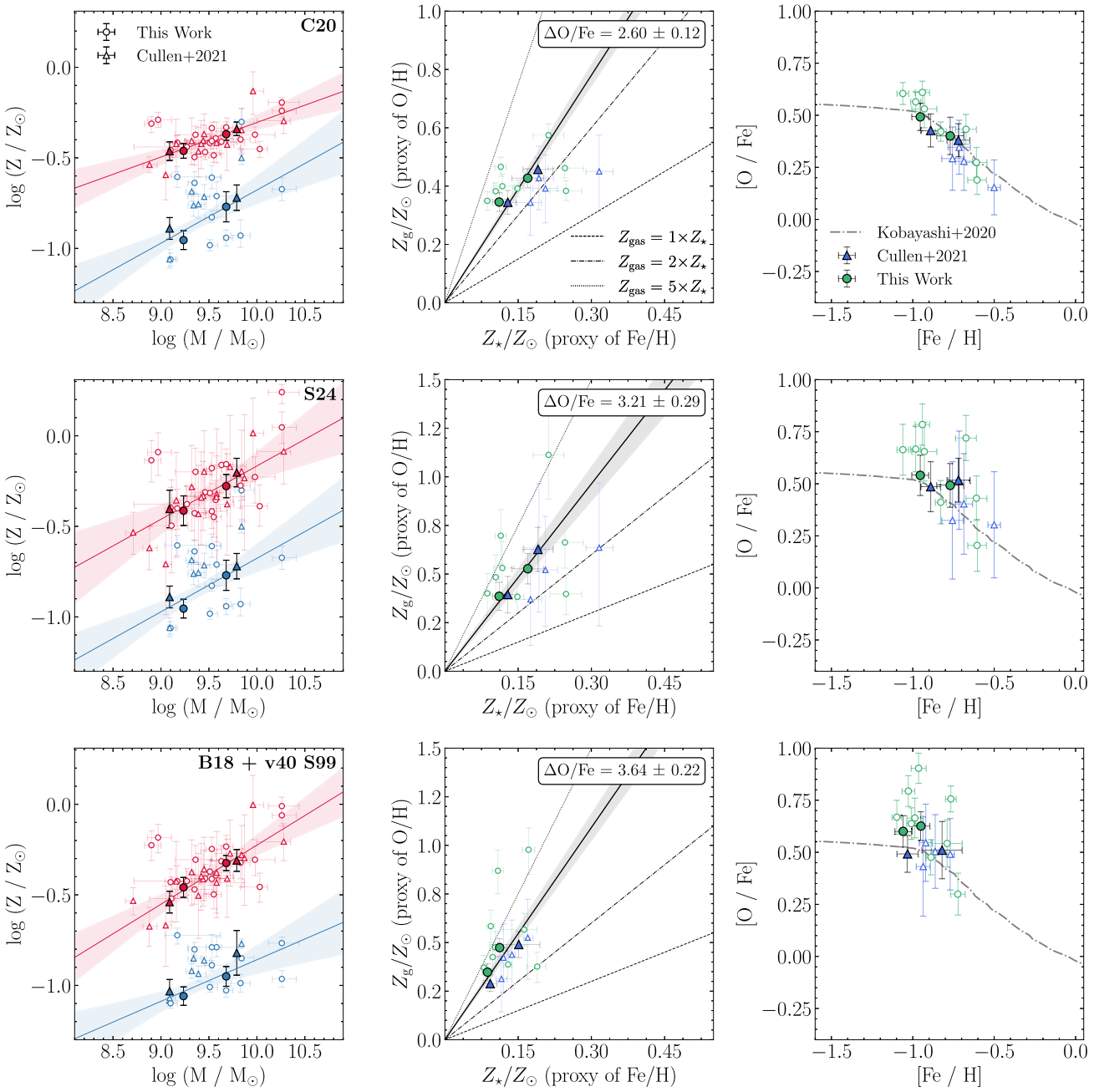

Figure A1. Comparison of the effects of different strong-line calibration scheme and stellar population model choices on the on the key results of this work. The first row corresponds to the C20 calibration scheme, and the second corresponds to the S24 calibration scheme. The third row uses the B18 calibrations, but substitutes the no-rotation S99 models for the v40-rotation S99 models. The left column emulates Fig. 4, and the central and right columns emulate the left-hand and right-hand panels of Fig. 6, respectively. The C20 calibration finds a shallower MZ_gR and hence a lower degree of alpha enhancement, whilst the S24 calibrations yield a much steeper MZ_gR and hence a high degree of enhancement. Switching to rotational stellar population models reduces the normalization of the stellar MZR, increasing the degree of alpha enhancement. All three configurations maintain the overlap of our high-z systems with the knee of the Kobayashi et al. (2020) relation.

Table A1. Gas-phase and stellar metallicity measurements for the composite and individual KMOS sample, utilizing the S24 and C20 calibration schemes and the S99 v40 rotational models.

Name	S24 $\log(Z_g/Z_\odot)$	C20 $\log(Z_g/Z_\odot)$	v40 $\log(Z_\star/Z_\odot)$
KVS-006	–	–	$-0.85^{+0.01}_{-0.01}$
KVS-009	–	–	–
KVS-014	–	–	–
KVS-055	$0.05^{+0.09}_{-0.09}$	$-0.24^{+0.03}_{-0.03}$	$-0.96^{+0.05}_{-0.05}$
KVS-067	$-0.09^{+0.11}_{-0.13}$	$-0.29^{+0.04}_{-0.04}$	–
KVS-070	–	–	–
KVS-075	$-0.20^{+0.13}_{-0.14}$	$-0.37^{+0.05}_{-0.05}$	–
KVS-082	$-0.14^{+0.11}_{-0.12}$	$-0.31^{+0.04}_{-0.05}$	–
KVS-085	$-0.27^{+0.11}_{-0.12}$	$-0.40^{+0.05}_{-0.05}$	$-0.99^{+0.05}_{-0.05}$
KVS-087	–	–	$-0.80^{+0.04}_{-0.03}$
KVS-093	–	–	–
KVS-100	$-0.43^{+0.13}_{-0.15}$	$-0.50^{+0.05}_{-0.04}$	–
KVS-101	–	–	–
KVS-131	$-0.39^{+0.12}_{-0.11}$	$-0.45^{+0.04}_{-0.05}$	–
KVS-141	$-0.16^{+0.12}_{-0.15}$	$-0.41^{+0.06}_{-0.07}$	–
KVS-150	$-0.35^{+0.11}_{-0.12}$	$-0.39^{+0.04}_{-0.04}$	–
KVS-156	–	–	$-0.79^{+0.06}_{-0.05}$
KVS-202	$0.24^{+0.04}_{-0.06}$	$-0.19^{+0.03}_{-0.03}$	$-0.77^{+0.03}_{-0.03}$
KVS-204	$-0.40^{+0.11}_{-0.12}$	$-0.42^{+0.04}_{-0.04}$	$-0.72^{+0.04}_{-0.05}$
KVS-208	$-0.32^{+0.10}_{-0.11}$	$-0.42^{+0.03}_{-0.03}$	$-1.01^{+0.03}_{-0.03}$
KVS-215	–	–	–
KVS-220	–	–	–
KVS-227	$-0.40^{+0.11}_{-0.12}$	$-0.46^{+0.04}_{-0.04}$	$-1.10^{+0.03}_{-0.03}$
KVS-248	$-0.18^{+0.09}_{-0.09}$	$-0.34^{+0.03}_{-0.03}$	$-0.79^{+0.06}_{-0.05}$
KVS-266	–	–	–
KVS-298	$-0.31^{+0.12}_{-0.14}$	$-0.47^{+0.04}_{-0.04}$	–
KVS-312	$-0.42^{+0.10}_{-0.09}$	$-0.41^{+0.03}_{-0.03}$	$-0.89^{+0.03}_{-0.03}$
KVS-340	$-0.38^{+0.10}_{-0.12}$	$-0.44^{+0.04}_{-0.04}$	–
KVS-361	$-0.23^{+0.11}_{-0.14}$	$-0.37^{+0.04}_{-0.04}$	–
KVS-391	$-0.16^{+0.08}_{-0.09}$	$-0.33^{+0.03}_{-0.03}$	$-1.03^{+0.04}_{-0.04}$
KVS-414	$-0.50^{+0.25}_{-0.26}$	$-1.22^{+0.71}_{-0.52}$	–
KVS-423	$-0.45^{+0.11}_{-0.11}$	$-0.49^{+0.04}_{-0.04}$	–
	KMOS composites		
Low-$M_\star$	$-0.41^{+0.08}_{-0.08}$	$-0.46^{+0.04}_{-0.04}$	$-1.06^{+0.05}_{-0.05}$
High-$M_\star$	$-0.28^{+0.06}_{-0.06}$	$-0.37^{+0.03}_{-0.03}$	$-0.95^{+0.06}_{-0.05}$

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