The mass-metallicity-star formation rate relation at z > 2 with 3D Hubble Space Telescope

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **440,** 2300–2312 (2014) doi:10.1093/mnras/stu443

The mass–metallicity–star formation rate relation at $z \gtrsim 2$ with 3D *Hubble Space Telescope*

F. Cullen,[1]★ M. Cirasuolo,[1,2] R. J. McLure,[1] J. S. Dunlop[1] and R. A. A. Bowler[1]

[1]*SUPA†, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[2]*UK Astronomy Technology Centre, Science and Technology Facilities Council, Royal Observatory, Edinburgh EH9 3HJ, UK*

Accepted 2014 March 5. Received 2014 March 5; in original form 2013 October 3

ABSTRACT

We present new accurate measurements of the mass, metallicity and star-formation rate of a statistically significant sample of 93 galaxies at $z \gtrsim 2$ using near-infrared spectroscopy taken as part of the 3D-*Hubble Space Telescope* survey. We derive a mass–metallicity relation (MZR) for our sample with metallicities based on the oxygen and Hβ nebular emission lines. We find the MZR derived from our data to have the same trend as previous determinations in the range $0 < z < 3$ with metallicity decreasing with stellar mass. However, we find that our MZR is offset from a previous determination at $z \gtrsim 2$ which used metallicities derived from the [N II]/Hα ratio. Incorporating star formation rate information, we find that our galaxies are also offset from the fundamental metallicity relation (FMR) by ∼0.3 dex. Using the Baldwin–Phillips–Terlevich (BPT) diagram we argue that, if the physical conditions of star-forming regions evolve with redshift, metallicity indicators based on [N II] and Hα, calibrated in the local Universe, may not be consistent with the ones based on oxygen lines and Hβ. Our results thus suggest that the evolution of the FMR previously reported at $z \sim 2$–3 may be an artefact of the differential evolution in metallicity indicators and caution against using locally calibrated empirical metallicity relations at high redshift, which do not account for evolution in the physical conditions of star-forming regions.

Key words: galaxies: evolution – galaxies: fundamental parameters – galaxies: high-redshift – galaxies: star formation.

1 INTRODUCTION

Over recent years, evidence has been accumulating for the existence of a tight relationship between mass, gas-phase metallicity (hereafter metallicity) and star formation rate (SFR) in galaxies out to $z \sim 2.5$. This so-called fundamental metallicity relation (FMR) was first proposed by Mannucci et al. (2010) and Lara-Lopez et al. (2010), though the SFR dependence of the mass–metallicity relation (MZR) had been initially noted earlier by Ellison et al. (2008). The existence of an FMR has since been supported by other studies (e.g. Mannucci, Salvaterra & Campisi 2011; Richard et al. 2011; Cresci et al. 2012; Niino 2012; Belli et al. 2013; Henry et al. 2013a,b; Lara-López, López-Sánchez & Hopkins 2013; Stott et al. 2013), though the exact form of the FMR has been shown to be dependent on the methodology adopted (e.g. Yates, Kauffmann & Guo 2012; Andrews & Martini 2013). In their original paper, Mannucci et al. (2010) proposed an FMR which defines a surface in the three-dimensional mass–metallicity–SFR space, and in Mannucci et al. (2011) this original FMR was extended to lower masses. Local Sloan Digital Sky Survey (SDSS) galaxies lie on the FMR surface with a residual dispersion of ∼0.05 dex in metallicity, and their FMR defined locally is seen to extend out to $z \sim 2.5$ (Mannucci et al. 2010). This suggests that the physical processes acting on star-forming galaxies have been consistent over the past ∼10 Gyr of cosmic time. Interestingly, however, Mannucci et al. (2010) observed an evolution of ∼0.6 dex away from the FMR for galaxies at $z > 2.5$, suggesting some change in the processes acting in star-forming galaxies at this epoch.

The original motivation for investigating the relationship between mass, metallicity and SFR was as an attempt to explain some of the features of the more extensively studied MZR. The MZR has been established across a large range in redshift from the local Universe (Tremonti et al. 2004; Kewley & Ellison 2008; Liu et al. 2008; Panter et al. 2008) to $z \sim 1$ (Savaglio et al. 2005; Shapley et al. 2005; Cowie & Barger 2008; Rodrigues et al. 2008; Zahid, Kewley & Bresolin 2011; Roseboom et al. 2012; Zahid et al. 2013), $z \sim 2$ (Erb et al. 2006a; Henry et al. 2013b; Kulas et al. 2013) and out to $z \sim 3$ (Maiolino et al. 2008; Mannucci et al. 2009). A consistent picture has emerged from these studies of an MZR similar in shape but differing in normalization out to high redshifts. At a given redshift, galaxies with lower stellar mass are found to have lower metallicities and at a given stellar mass, galaxies at higher redshift

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

are found to have lower metallicities. In the context of the FMR, the evolution of the MZR with redshift is explained due to the fact that metallicity anticorrelates with SFR, and at higher redshifts galaxies of a given mass have higher SFRs (e.g. Daddi et al. 2007).

Studying the MZR and FMR requires an accurate measure of galaxy metallicity. One outstanding issue with the observational measurements of metallicities at high redshifts is that the only feasible method is to use strong nebular emission lines. These lines are not direct tracers of metallicity and therefore have to be calibrated in some way. One method that has been adopted is to empirically calibrate these emission-line ratios in the local Universe. At low metallicities, the line ratios are calibrated using direct metallicity tracers such as the weak [O III] λ4363 auroral line, whereas at high metallicities, the auroral line becomes too weak and instead local galaxy spectra are fitted with photoionization models of star-forming regions to calibrate the line ratios (see e.g. Nagao, Maiolino & Marconi 2006; Maiolino et al. 2008). Using these local empirical calibrations at high redshift assumes that the calibrations sufficiently account for the change in physical conditions of star-forming galaxies with redshift. However, from data currently available, there is evidence to suggest that conditions in star-forming regions are evolving with redshift (e.g. Brinchmann, Pettini & Charlot 2008; Hainline et al. 2009; Shirazi, Brinchmann & Rahmati 2013; Kewley et al. 2013a,b; Nakajima & Ouchi 2013), and if this is the case, using empirical metallicity calibrations at high redshifts may not be reliable.[1]

The evolution of the physical conditions in star-forming regions with redshift is inferred from studying the relationships between nebular emission lines. One common diagnostic is a plot of the [N II]/Hα versus [O III] λ5007/Hβ line ratios, known as a BPT diagram (Baldwin, Phillips & Terlevich 1981). Star-forming galaxies in the local Universe form a tight sequence in the BPT diagram (e.g. Kauffmann et al. 2003); however, an evolution away from this sequence is observed at high redshifts towards higher [O III] λ5007/Hβ and [N II]/Hα ratios (e.g. Shapley et al. 2005; Erb et al. 2006a; Hainline et al. 2009; Yabe et al. 2012). At the same time, an evolution in the ionization parameter of galaxies, which is a measure of the degree of excitation of H II regions, is observed, in that higher redshift star-forming galaxies are found to have higher ionization parameters (e.g. Lilly, Carollo & Stockton 2003; Hainline et al. 2009; Richard et al. 2011; Nakajima et al. 2013). This evolution towards higher ionization parameters has been used to explain the offset of high-redshift star-forming galaxies from the local Universe relation in the BPT diagram (e.g. Brinchmann et al. 2008; Kewley et al. 2013a,b) and has most recently been suggested as a possible reason for the apparent evolution of galaxies away from the FMR at $z \sim 3$ (Nakajima & Ouchi 2013). Interestingly, using a sample of low-redshift galaxies with elevated ionization parameters, supposed analogues of these high-redshift star-forming galaxies, Ly et al. (2014) observe an offset from the FMR of the order of 0.1–1 dex similar to that observed at $z \sim 3$.

In this paper, we aim to provide additional observational constraints on the MZR and FMR at $z \gtrsim 2$, using a sample of star-forming galaxies with a mean redshift $\langle z \rangle = 2.16$ selected from the 3D-*Hubble Space Telescope* (*HST*) near-infrared (near-IR) spectroscopic data set. 3D-*HST* (Brammer et al. 2012) is a slitless spectroscopic survey with the Wide-Field Camera 3 (WFC3) on the *HST* and provides an ideal data set for obtaining a large sample of $z \gtrsim 2$ galaxies for which metallicities can be measured as the [O II]λ3727, Hβ and [O III] λλ4958,5007 nebular emission fall into the grism spectra in the redshift range $2 < z < 2.3$. We provide a sample of galaxies similar in size to the current largest data set of $z \gtrsim 2$ galaxies with measured metallicities studied by Erb et al. (2006a) ($\langle z \rangle = 2.26$). Since the Erb et al. (2006a) study was based on a sample of ultraviolet (UV)-selected star-forming galaxies observed from the ground, our line-flux-limited sample taken with the *HST* allows us to probe lower SFRs than the Erb et al. (2006a) data. Finally, since Erb et al. (2006a) measure metallicities with the Hα and [N II] nebular lines, we are able to use an independent method with the [O II]λ3727, Hβ and [O III] λλ4958,5007 lines. These are the same nebular emission lines as used for measuring metallicities in the $z \sim 3$ studies (Maiolino et al. 2008; Mannucci et al. 2009), and therefore our data provide useful insight into the observed shift of the FMR between $z \sim 2$ and ∼3.

This paper is organized as follows: in Section 2, we describe the spectroscopic and photometric data, and the measurement of stellar mass, SFRs and metallicities from these data. In Section 3, we describe the results of this paper in the context of the MZR, FMR and photoionization conditions of the galaxies in our sample. We compare our results to those from previous $z \sim 2$ studies and summarize our results in Section 4. Throughout, we assume a cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70$ km s^{-1} Mpc^{-1}.

2 THE DATA

In this paper, we take advantage of two complimentary data sets: a spectroscopic sample drawn from the 3D-*HST* spectroscopic survey (Brammer et al. 2012) and photometric data taken as part of the CANDELS survey (Grogin et al. 2011; Koekemoer et al. 2011). The combination of these data allows us to calculate the required physical properties for the galaxies in our sample.

2.1 3D-*HST* spectroscopic data

The spectroscopic data used in this paper are part of the 3D-*HST* observations described in Brammer et al. (2012). The 3D-*HST* survey provides low-resolution ($R \sim 130$) spatially resolved near-IR grism spectra over the wavelength range 1.1–1.68 μm taken with the WFC3 G141 grism on the *HST*. The survey covers 675 arcmin^2 providing spectroscopic follow-up of three quarters of the deep near-IR CANDELS imaging in five fields (AEGIS, COSMOS, GOODS-S, GOODS-N and UDS).

A modified version of the publicly available pipeline described in Brammer et al. (2012) was used to reduce the 3D-*HST* grism exposures utilizing the AXE software package (Kümmel et al. 2009). Objects are selected from a CANDELS *F*160*W* (H_{160}) mosaic of the given field down to a limiting magnitude of $H_{160} = 26$ (AB mag). The four raw direct images of the 3D-*HST* pointing are combined using the IRAF task MULTIDRIZZLE and positional offset from the CANDELS mosaic are calculated using TWEAKSHIFTS. The four raw grism images are combined using MULTIDRIZZLE and aligned to the reference mosaic using the shifts calculated on the direct image. The standard set of AXE tasks are then used to fully reduce the grism pointings to individual 1D and 2D spectra. A custom method for estimating background subtraction as described in Brammer et al. (2012) is implemented, this takes into account the variation of IR

[1] It is important to note that purely theoretical metallicity calibrations (e.g. Kewley & Dopita 2002) are in principal not subject to the potential bias introduced by using local Universe galaxies to calibrate emission-line ratios, as theoretical models can explore a large range in ionization conditions which may not be sampled in local galaxies. However, the FMR is built upon the local empirical metallicity calibration of Maiolino et al. (2008) and so this calibration will be used throughout this paper.

background across the sky which cannot be fully accounted for in default grism master sky background (Kümmel et al. 2009).

We carefully treat the contamination of the grism spectra. With grism spectroscopy any given spectrum can be contaminated by the spectra of nearby sources. To mitigate this effect, in some grism surveys, exposures are taken at different position angles in the hope that multiple dispersion directions reduce the chances of a spectrum being severely contaminated (e.g. Pirzkal et al. 2013). However, grism exposures in the 3D-*HST* survey are taken at only one position angle, and so a reliable quantitative estimation of spectral contamination is required. We estimate the contamination by using the images in the *F*850*LP*, *F*125*W* and *F*160*W* bands available in the CANDELS fields. We use the AXE task FCUBEPREP to construct model grism images using the photometric and morphological properties derived from the images; the mosaics are extended around the field of view (FOV) of the grism exposure to account for sources outside the FOV whose spectra are dispersed on to the grism image. In this way, a quantitative estimate of the contamination for each spectrum is produced and we subtract the contamination model from the observed spectrum to obtain the final spectrum of each object.

2.1.1 *Redshift determination*

Due to the low resolution of the grism spectra, care must be taken when deriving galaxy redshifts. In many cases, only one emission line is visible in the spectrum and so combining spectral redshift information with photometric redshift estimates is required. We estimate redshifts using a combination of spectral template fitting and photometric redshifts. For spectral fitting, we use galaxy templates from the K20 survey (Mignoli et al. 2005), binned to the resolution of the grism spectra and calculate probability distributions derived from a χ^2 fit. The redshift probability distribution from spectra fitting is then combined with a photometric redshift probability distributions calculated from the broad-band photometry of the galaxies using the EAZY code (Brammer, van Dokkum & Coppi 2008).

2.1.2 *Final spectroscopic sample*

The sample used in this paper is drawn from three of the five fields covered by 3D-*HST*: GOODS-S, COSMOS and UDS. Within a wavelength range of the grism the prominent [O II]λ3727, Hβ and [O III] $\lambda\lambda$4958,5007 emission lines are observable at $2.0 < z < 2.3$. These emission lines can be used for estimating SFRs and metallicities of galaxies as described in Maiolino et al. (2008) and also allow accurate redshift determination. The initial sample is assembled from a visual inspection of 3D-*HST* spectra to identify emission-line galaxies in the correct redshift range. We do not exclude contaminated spectra from our sample unless the contamination is severe or the contamination model clearly erroneous (e.g. leaving residual continuum features after subtraction). This initial sample contains 103 galaxies and extends down to a flux limit in [O III] λ5007 of 5×10^{-18} erg cm^{-2} s^{-1}.

We remove possible active galactic nucleus (AGN) for our sample by first cross-matching with the *Chandra* 4 Ms X-ray catalogue in GOODS-S (Xue et al. 2011) and *Chandra* 1.8 Ms X-ray catalogue in COSMOS (Civano et al. 2012); three galaxies were identified and removed based on the X-ray data. We also followed the method of Henry et al. (2013b) in using the Mass-Extinction (MEx) diagram (Juneau et al. 2011) to separate star-forming galaxies and AGN based on their [O III] λ5007/Hβ ratio and mass (Fig. 1). We follow

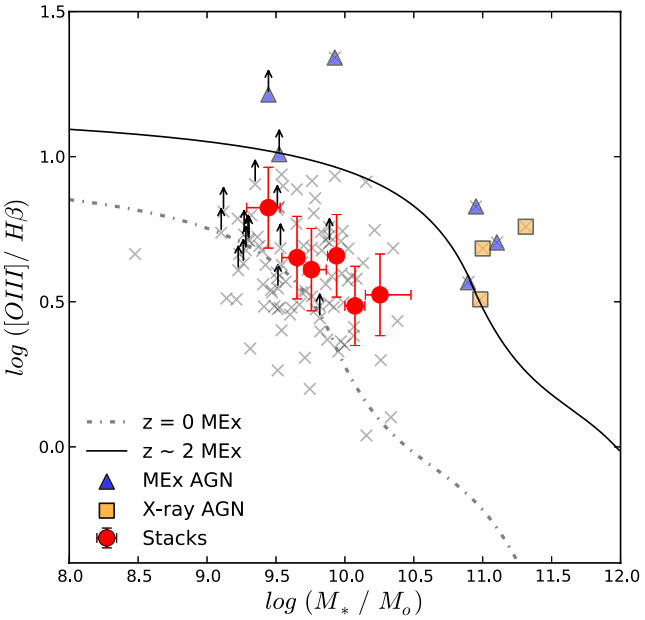

Figure 1. A modified version of the MEx diagram (Juneau et al. 2011) used for selecting AGN in emission-line galaxies. The original form of the separation line given in Juneau et al. (2011) is given by the grey dashed line, and the black solid line shows the modified version of this line used for our $z \gtrsim 2$ sample (see Section 2.1.2 for discussion). The grey crosses show individual galaxies in the sample which are not classified as AGN. The yellow squares show AGN identified from X-ray counterparts and the blue triangles show MEx classified AGN. The black arrows represent individual galaxies with only an upper limit on the [O III] λ5007/Hβ ratio. The red circles with error bars show the position in the MEx diagram of the final stacked galaxies.

Henry et al. (2013b) in shifting the (Juneau et al. 2011) relation by 1 dex in mass to account for the mass–metallicity evolution at high redshift, we also shift by +0.2 dex in [O III] λ5007/Hβ to account for the evolution of the [O III] λ5007/Hβ ratio in star-forming galaxies at $z \sim 2$ due to evolving ionization conditions (see Section 3.3 and Kewley et al. 2013a). Across the range of [N II]/Hα occupied by local star-forming galaxies, +0.2 dex is the average shift in the [O III] λ5007/Hβ ratio between $z \sim 0$ and $\sim$2 based on the Kewley et al. (2013a) models. We note that Hβ is not detected in all individual galaxies, and in this case we can only put a lower limit on the [O III] λ5007/Hβ ratio; however, we do not exclude galaxies with lower limits which fall below the AGN separation line. Using this modified MEx diagram approach, we remove a further six galaxies from the sample, leaving a total of 9/103 galaxies (9 per cent) identified as AGN. This AGN fraction is consistent with other studies at slightly lower redshift ($z \sim 1.5$) (e.g. Stott et al. 2013; Zahid et al. 2013).

Finally, one further galaxy is excluded from the stacking analysis (see Section 2.3.1) as the measured mass of $\log(M_*/M_\odot) = 8.63$ (see Section 2.2.1) is offset by 0.66 dex from the next lowest mass galaxy in the sample and would therefore skew the mass distribution in the lowest mass bin. This leaves a final sample of 93 galaxies on which the following analysis is performed.

2.2 CANDELS photometric data

The photometric data are taken from three separate catalogues for the GOODS-S, UDS and COSMOS fields. All photometry covers the rest-frame UV to mid-IR (*Spitzer*/IRAC 3.6 and 4.5 μm). In both

GOODS-S and UDS, we take the photometry from existing CANDELS catalogues described in Guo et al. (2013) and Galametz et al. (2013), respectively. For COSMOS, we produce a new catalogue by first convolving all the available ground-based and space-based optical and near-IR imaging (see Bowler et al. 2012 for details) to a common point spread function (0.8 arcsec) and then performing aperture photometry using the smoothed $F160W$ image as the detection band. The *Spitzer*/IRAC 3.6 and 4.5 μm photometry was included by de-confusing the IRAC images using the UltraVISTA *Ks*-band imaging (McCracken et al. 2012).

2.2.1 *Stellar masses and SFRs from SED fitting*

Stellar masses and SFRs for individual galaxies are measured using the publicly available code LEPHARE (Ilbert et al. 2006). We run LEPAHRE with solar-metallicity Bruzual & Charlot (2003, hereafter BC03) templates assuming τ model star formation histories with $\tau = 0.3$, 1, 2, 3, 5, 10, 15 and 30 Gyr. The Calzetti et al. (2000) attenuation law is used to account for dust extinction with $E(B-V)$ values ranging from 0 to 0.6. The age of the model is allowed to vary between 0.05 Gyr and the age of the Universe at the spectroscopic redshift of the galaxy. The normalization factor required to scale the best-fitting template to the observed magnitudes in each band gives the stellar mass. In order to compare to Erb et al. (2006a) data, we correct our masses to the total mass of stars formed using the conversions in the BC03 templates; the median increase in mass when applying this correction is +0.18 dex.

2.3 Galaxy metallicities and SFRs from spectra

2.3.1 *Galaxy stacking procedure*

We measure spectral properties (SFR and metallicity) for all the individual galaxies. However, due to the low signal-to-noise (S/N) ratio of some of the spectra, we also stacked the spectra in bins of stellar mass, following the method of Erb et al. (2006a). The mass bins were chosen such that each bin contained a similar number of galaxies. To produce the stacks, each individual spectrum and best-fitting BC03 model were first shifted to the rest frame and de-reddened using the Calzetti et al. (2000) attenuation law. The continuum of each galaxy was then subtracted by interpolating the best-fitting BC03 model SED on to the same wavelength grid and normalizing to the galaxy spectra with the emission lines masked, in this way the continuum could be subtracted whilst accounting for the reddening in the emission lines assuming $E(B-V)_{\rm stars} \sim E(B-V)_{\rm gas}$ (see Section 2.3.3). Examples of continuum fits to galaxies are shown in Fig. 2. This method of continuum subtraction was preferred to fitting a simple low-order polynomial to the galaxy spectra as the SED model can better fit the 4000 Å break crossing the [O II]λ3727 emission line and also accounts for the underlying Hβ absorption in the stellar continuum. Within each mass bin the galaxy spectra were interpolated to a common wavelength grid, and we determine the median flux within each wavelength bin. The final stacked galaxy spectra are shown in Fig. 3. For consistency, we have checked that removing the continuum from each galaxy individually before stacking and removing a stack of SED models from a stack of the observed galaxy spectra does not change our results.

We note that stacking the intrinsic fluxes of each galaxy may be biasing our measurements towards the brightest emission-line galaxies in the stack (Domínguez et al. 2013; Henry et al. 2013b), although taking the median flux in each wavelength bin should mitigate this effect. We have checked that normalizing each individual spectrum by its [O III] λ5007 flux before stacking does not change our results. We find that when the spectra are normalized, the measured metallicities are fully consistent with those measured without normalizing. We use the intrinsic stacked spectra in the remainder of this paper so that we can measure the SFR and metallicity directly from the same spectra.

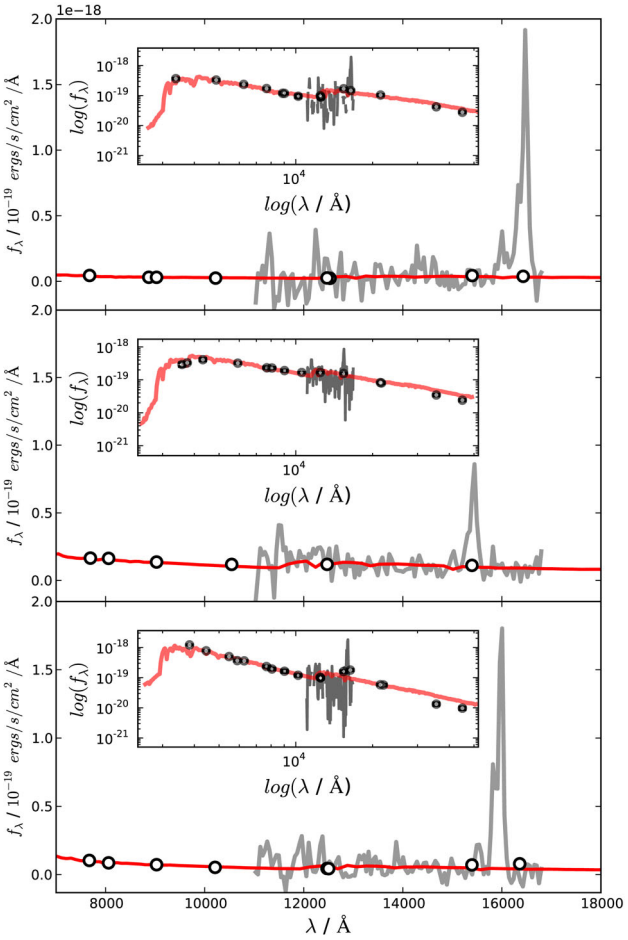

Figure 2. The continuum fits to three galaxies in our sample. Each panel shows a grism spectra in the rest frame with the continuum fit from the SED plotted in red and the photometry points plotted as black empty circles. Inset in each panel shows the full best-fitting SED of the galaxy with the photometry and grism spectra overplotted.

2.3.2 *Line fluxes*

We measure line fluxes from the continuum-subtracted grism spectra in the following way. For the [O III] λλ4958,5007 doublet and Hβ lines, we fit a triple Gaussian allowing the height of each Gaussian to vary. Since the [O III] λ5007 line always has the highest S/N, we fit the centroid of the [O III] λ5007 line and fix the centroids of [O III] λ4958 and Hβ given the best-fitting [O III] λ5007 centroid and redshift. We allow the fit to vary the width of the [O III] λ5007 line; the width of the [O III] λ4958 and Hβ lines are fixed to the best-fitting width of the [O III] λ5007 line. The [O III] λλ4958,5007 lines are blended due to the low resolution of the grism spectra and to measure the flux in these lines, we used two methods. First, we fit two Gaussians and use the best-fitting parameters for each line directly

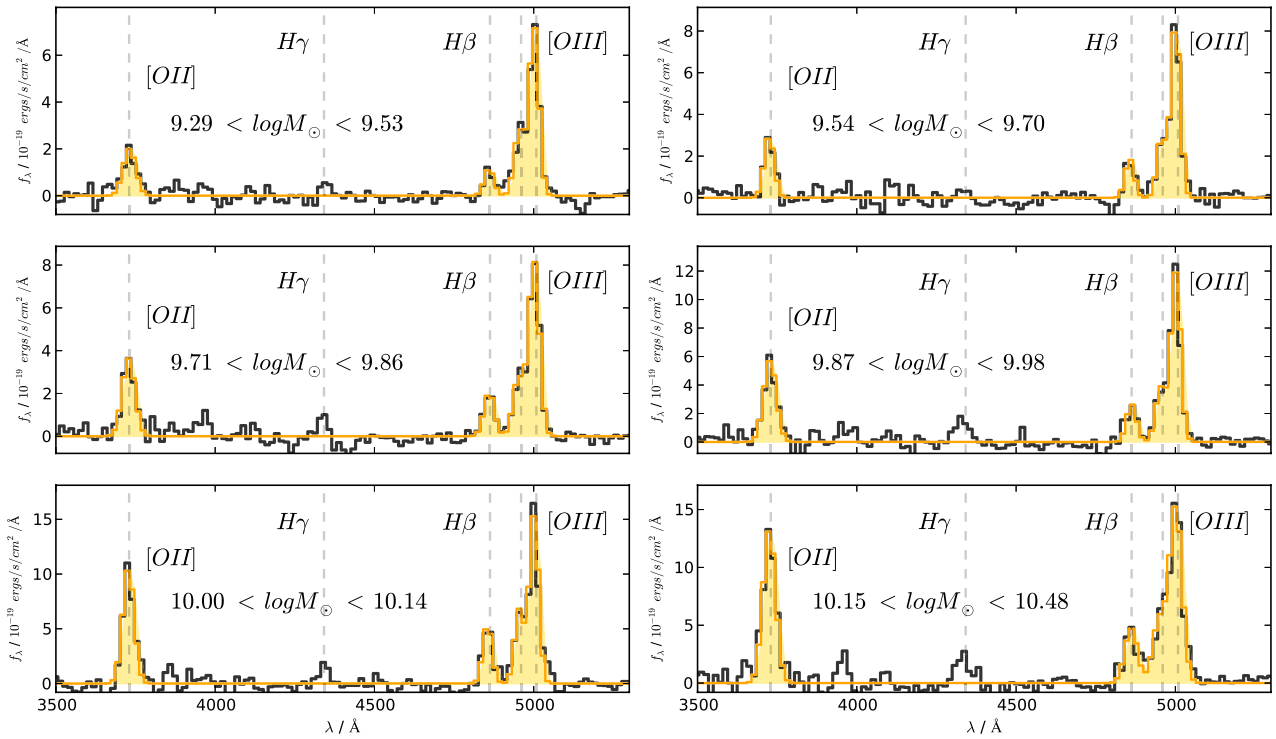

Figure 3. The rest-frame stacked spectra of galaxies in our sample split into six bins of increasing stellar mass. Each panel is labelled with the range of stellar mass in each bin and the [O II]λ3727, Hβ and blended [O III] λλ4958,5007 emission lines are marked by dashed lines. The stacked spectra are shown in black with the fit to the emission lines overlaid in orange.

Table 1. Measured data for the galaxy stacks in our sample.

ID	N_{gal}[a]	$\langle z \rangle$[b]	$f_{[O\,III]5008}$[c]	$f_{H\beta}$[c]	$f_{[O\,II]}$[c]
1	15	2.22	2.9 ± 0.3	0.5 ± 0.1	1.0 ± 0.1
2	15	2.11	3.1 ± 0.3	0.7 ± 0.1	1.2 ± 0.2
3	16	2.15	3.5 ± 0.4	0.9 ± 0.1	1.8 ± 0.2
4	16	2.14	4.6 ± 0.5	1.1 ± 0.1	2.9 ± 0.3
5	16	2.17	6.1 ± 0.6	2.0 ± 0.3	4.8 ± 0.5
6	15	2.20	7.7 ± 0.8	2.3 ± 0.3	6.7 ± 0.7

[a]Number of individual galaxies in the stack.
[b]Average redshift of the stack.
[c]Dust-corrected flux in units of 10^{-17} erg s^{-1} cm^{-2}.

from the fit; secondly, we sum the two Gaussians from the triple Gaussian and take the [O III] λ5007 flux to be 3/4 of the total (Storey & Zeippen 2000), this follows previous methods for measuring the [O III] λλ4958,5007 line fluxes from grism spectra (Trump et al. 2013). We have checked that both methods are fully consistent and do not affect the results of this paper. For the [O II]λ3727 line, we fitted a single Gaussian allowing the centroid, line width and height to vary. Examples of the line fits to the stacked spectra are shown in Fig. 3. Dust-corrected line fluxes for the stacked spectra are given in Table 1.

2.3.3 *Star formation rates*

We measure SFRs from the Hβ line flux, first converting to luminosity and then using the common conversion factor Hα/Hβ = 2.86 (Kennicutt 1998). We then scale down by a factor 1.7 to convert to a Chabrier (2003) initial mass function (IMF). Use of the Hα/Hβ conversion factor relies on the assumption that the galaxies in the stacks have $E(B-V) = 0$ after the de-reddening of individual spectra; therefore, a good estimate of the extinction suffered by emission lines in the galaxy is required. There has been much debate in the literature as to whether the extinction returned from SED fitting, which corresponds to the extinction in the stellar continuum, is the same as the extinction suffered by emission lines, since emission lines emanate from dusty star-forming regions and therefore could potentially suffer larger extinction than stars.

The extra extinction of nebular emission lines is parametrized by a factor f where $E(B-V)_{stars} = f^* E(B-V)_{gas}$. Calzetti et al. (2000) find that $f = 0.44 \pm 0.03$ for local starburst galaxies, and similar corrections have been reported for galaxies out to $z \sim 2$ (e.g. Förster Schreiber et al. 2009; Cresci et al. 2012; Yabe et al. 2012). However, at intermediate redshifts ($z \sim 1.5$) there is some evidence for an evolution towards larger f values (e.g. Kashino et al. 2013; Price et al. 2013), and at $z \sim 2$, some studies find an apparent evolution to $f \sim 1$ (e.g. Erb et al. 2006b; Hainline et al. 2009).

The method of de-reddening spectra described in Section 2.3.1 assumes $E(B-V)_{stars} \sim E(B-V)_{gas}$, so to get a sense as to whether we are making sensible dust corrections, we can compare our best-fitting SFR from SED fitting to that derived from the de-reddened line flux. The comparison is shown in Fig. 4. The figure shows that the two independent measures of SFR are in good agreement for the galaxy stacks, for the Hβ and SED methods, respectively, the median and median absolute deviation (MAD) of the SFRs of the whole galaxy sample are 17.2 ± 8.8 and 13.0 ± 12.3 M$_\odot$ yr^{-1}. Given the good agreement between the SFRs, we adopt the $E(B-V)$ values returned by SED fitting for correcting the emission lines; therefore, following previous studies at $z \sim 2$ in assuming that $E(B-V)_{stars} \sim E(B-V)_{gas}$.

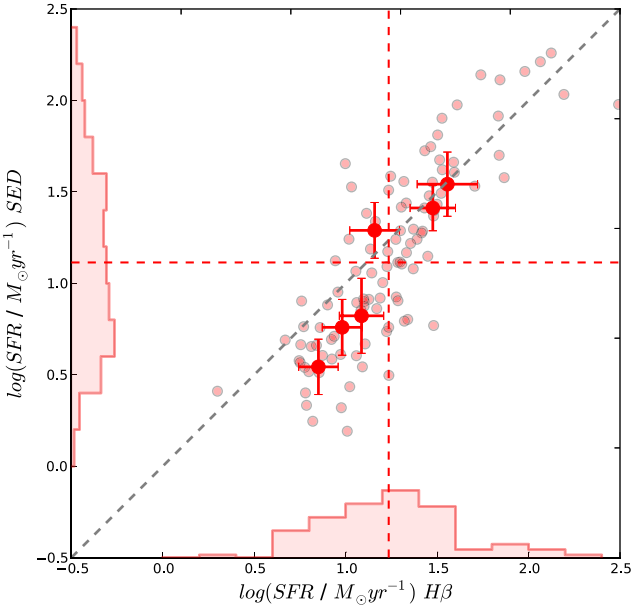

Figure 4. A comparison of SFR measured from the Hβ line with SFR as derived from SED fitting. The galaxy stacks are shown as the large red points with error bars. The SED SFRs for the stacks are taken as the median of the SED-derived SFRs of individual galaxies within the stack with measured Hβ. The small red points represent the individual galaxies. On both axes, histograms of the values for the individual galaxies are plotted and the dotted red lines show the median values for each distribution. Respectively, the median and MAD of the SFRs of the whole galaxy sample are 17.2 ± 8.8 and 13.0 ± 12.3 $M_\odot$ yr^{-1}.

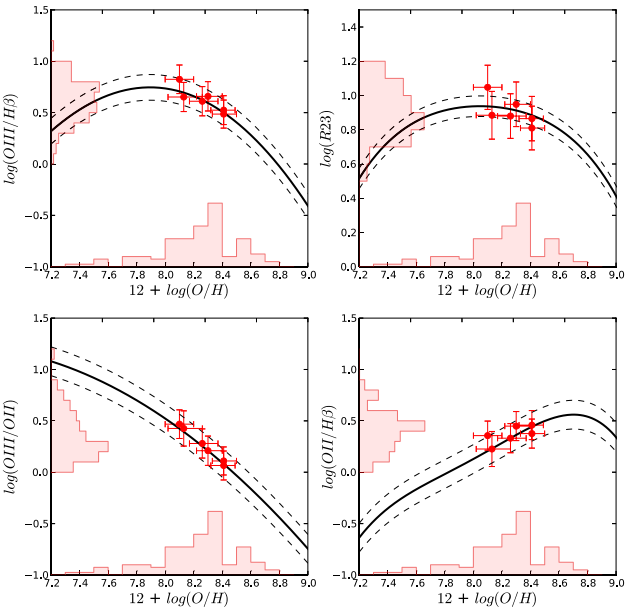

Figure 5. Relationship between line ratio and derived metallicity values for each of the four metallicity calibrations used to determine galaxy metallicities in this paper. The calibration curves are taken from Maiolino et al. (2008). Shown as red solid circles are the metallicities derived using a combination of all the four line ratios as described in Section 2.3.4. The red histograms represent the distributions of individual galaxies in the sample with detections in the appropriate emission lines.

2.3.4 *Gas-phase metallicity*

The metallicity of a galaxy is a measure of the abundance of metals relative to hydrogen in the interstellar medium and is most commonly quoted in terms of the oxygen abundance ratio 12+log(O/H). Measuring the metallicity at high redshifts relies on using ratios of various strong, optical emission lines emanating from H II regions. These line ratios are a combination of strong hydrogen recombination lines such as Hα and Hβ, and collisionally excited forbidden transitions in metals such as [N II], [O II]λ3727 and [O III] $\lambda\lambda$4958,5007.

Unfortunately, the ratios of the strengths of these lines do not depend solely on the element abundances but have other dependences (e.g. gas density, hardness of ionizing radiation, etc.), and therefore, must be calibrated against direct metallicity tracers. At low metallicities, in the local Universe, the electron temperature T_e method (Pettini & Pagel 2004) can be used to calibrate the nebular lines. At higher metallicities, the electron temperature methods cannot be used and we rely on photoionization modelling of H II regions (Kewley & Dopita 2002). Various attempts have been made to calibrate these strong nebular line ratios across a wide range of metallicity values (e.g. Nagao et al. 2006; Kewley & Ellison 2008; Maiolino et al. 2008). These authors give various recipes for calculating the metallicity of a galaxy from line ratios of optical emission lines. It is important to note that inferred values of metallicities depend crucially on the calibration adopted, and therefore when making comparisons, all results must be converted to a consistent calibration (see Kewley & Ellison 2008 for a detailed discussion). In this paper, we use the local empirical calibrations described in Maiolino et al. (2008) as this calibration was used in the original investigation of the FMR (Mannucci et al. 2010). The Maiolino et al. (2008) calibrations are derived from a sample for low-metallicity galaxies for which metallicities can be measured directly using the T_e method, and a sample of SDSS spectra at higher metallicity for which metallicities are derived from photoionization modelling.

We measure metallicities using the [O II]λ3727, [O III] λ4958, [O III] λ5007 and Hβ emission lines via a method similar to that used for the $z \sim 3$ galaxies in the AMAZE and LSD surveys (Maiolino et al. 2008; Mannucci et al. 2009). However, in contrast to these studies, we do not fit to metallicity and extinction simultaneously across all line ratios. However, both Maiolino et al. (2008) and Mannucci et al. (2009) note that their fit, whilst constraining the metallicity, does not simultaneously provide good constraints on the dust extinction, implying the derived metallicity is weakly dependent on the value of dust extinction adopted. We measure all line ratios available which have metallicity calibrations given in Maiolino et al. (2008; fig. 5 of their paper). Specifically, the line ratios are R_{23} (=[O II]λ3727 + [O III] λ4958 + [O III] λ5007/Hβ), [O III] λ5007/Hβ, [O III] λ5007/[O II]λ3727 and [O II]λ3727/Hβ. The line ratios and calibration curves are shown in Fig. 5.

To measure the metallicity of a galaxy stack, we first calculate the observed line ratio in each of the four diagrams. At each value of metallicity, within a large range ($7 < \log(\mathrm{O/H}) + 12 < 9.5$), we calculate, for each calibration curve, a χ^2 statistic from the difference between the observed line ratio and calibration line ratio at that metallicity. Thus, for all four calibration curves, we construct a χ^2 versus $\log(\mathrm{O/H}) + 12$. We combine the χ^2 values from each of the four diagrams to construct an overall χ^2 versus $\log(\mathrm{O/H}) + 12$. We take the metallicity at the minimum of the combined χ^2 as the best-fitting metallicity. We take the 1σ confidence intervals to be within $\Delta\chi^2 = 1$ of the best-fitting value. The metallicities of

Table 2. Derived data for the galaxy stacks in our sample.

ID	$\log(M/\mathrm{M}_\odot)^a$	$12+\log(\mathrm{O/H})^b$	$\log(\mathrm{SFR}_{\mathrm{H}\beta}/\mathrm{M}_\odot\ \mathrm{yr}^{-1})^c$	$\log(\mathrm{SFR}_{\mathrm{SED}}/\mathrm{M}_\odot\ \mathrm{yr}^{-1})^d$	$E(B-V)^e$
1	$9.44^{+0.09}_{-0.16}$	$8.10^{+0.10}_{-0.11}$	0.86 ± 0.16	0.57 ± 0.15	0.04 ± 0.04
2	$9.65^{+0.05}_{-0.11}$	$8.13^{+0.13}_{-0.12}$	0.97 ± 0.17	0.78 ± 0.22	0.09 ± 0.04
3	$9.75^{+0.11}_{-0.05}$	$8.26^{+0.11}_{-0.09}$	1.09 ± 0.12	0.81 ± 0.36	0.08 ± 0.07
4	$9.94^{+0.04}_{-0.07}$	$8.30^{+0.09}_{-0.09}$	1.17 ± 0.14	1.28 ± 0.31	0.19 ± 0.09
5	$10.07^{+0.07}_{-0.07}$	$8.41^{+0.09}_{-0.09}$	1.47 ± 0.16	1.37 ± 0.28	0.18 ± 0.07
6	$10.25^{+0.22}_{-0.11}$	$8.41^{+0.09}_{-0.08}$	1.56 ± 0.17	1.62 ± 0.43	0.23 ± 0.13

[a]Median mass of the stack, error bars represent the range of stellar masses in each bin.
[b]Metallicities of the stacks derived from the Maiolino et al. (2008) calibrations.
[c]The SFR of the stacks measured from the Hβ flux corrected for dust extinction.
[d]Mean and standard deviation of SED SFR of the individual galaxies in the stack.
[e]Mean and standard deviation of $E(B-V)$ of the individual galaxies in the stack.

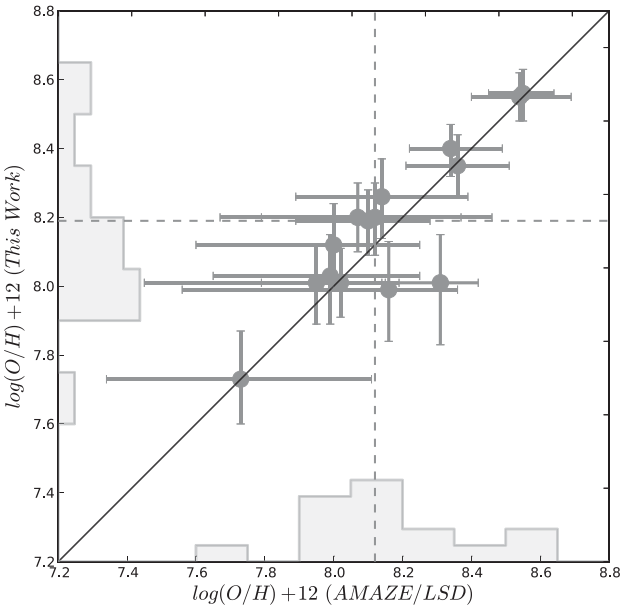

Figure 6. A comparison of our measurement of metallicity for the $z \sim 3$ galaxies in Maiolino et al. (2008) and Mannucci et al. (2009) compared to the metallicities quoted in those papers. The mean and median offset of the measured metallicities are 0.013 and 0.004 dex, respectively. On each axis, the histograms represent the distribution of the individual points and the dashed lines represent the median of those distributions.

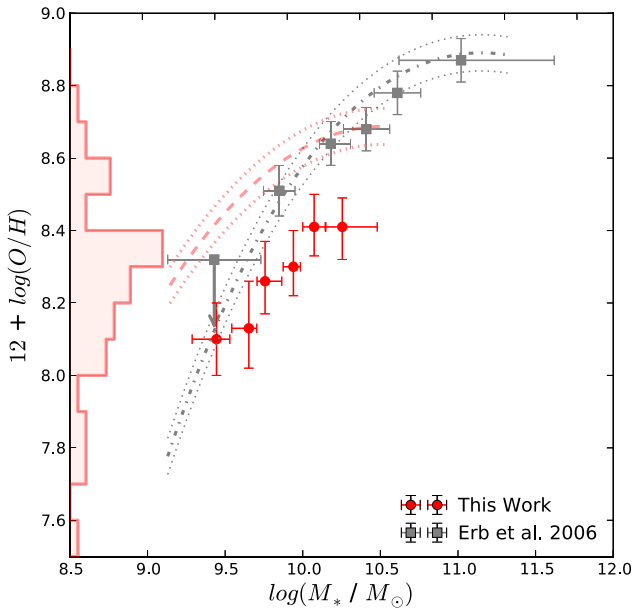

Figure 7. The MZR for the $z \gtrsim 2$ galaxies in our sample. The red circles represent the galaxy stacks presented in this paper with metallicities derived from the [O II]λ3727, [O III] $\lambda\lambda$4958,5007 and Hβ nebular emission lines. The red histogram shows the distribution of metallicities for the individual galaxies in the stacks with a measured metallicity. The grey squares represent the $z \gtrsim 2$ MZR from Erb et al. (2006a) with metallicities derived from [N II] and Hα. The red dashed line (our sample) and the grey dot–dashed line (Erb et al. 2006a) show the predicted positions of the two data sets in the MZ plane derived from the FMR (see Section 3.2).

the stacked spectra along with all other derived data are given in Table 2.

To check for consistency, we take the line ratios from the Maiolino et al. (2008) and Mannucci et al. (2009) papers, de-redden via the best-fitting extinctions quoted and measure the metallicities in the same way as described above. We compare these measured metallicities to the metallicities quoted in those papers. Fig. 6 shows the results of this test; our method agrees very well across the all 15 data points with median and average difference of 0.004 and 0.013 dex, respectively.

3 RESULTS: MZR AND FMR

In this section, we present the results of our study of the MZR and FMR, exploring the position of the 3D-*HST* $z \gtrsim 2$ emission-line galaxies in the MZ plane and on the FMR surface. Throughout, we compare to the previous $z \gtrsim 2$ sample of Erb et al. (2006a) as it represents the largest single sample of $z \gtrsim 2$ galaxy metallicities to date and directly follows our method of stacking low-S/N galaxies into bins of stellar mass. We also compare in Section 3.2 our results with the complimentary study of Henry et al. (2013b) at lower redshift ($z \sim 1.8$) with metallicities measured from stacked grism spectra using the same nebular emission lines.

3.1 Mass–metallicity relation

Fig. 7 shows the MZR for our data compared to the MZR from Erb et al. (2006a). Erb et al. (2006a) measured metallicities from the [N II]/Hα ratio using the calibration of Pettini & Pagel (2004), so we convert their data to the Maiolino et al. (2008) calibration to keep the metallicity scales consistent. It is important that strong line metallicity diagnostics are compared using the same calibration

since it has been noted that mismatches between different calibration scales can cause systematic offsets in derived metallicities (e.g. Kewley & Ellison 2008). Erb et al. (2006a) used a Chabrier IMF to derive stellar masses so that no mass conversion is necessary for consistency with our analysis. As a cross-check, we have run the photometric data from their paper through LEPHARE and confirmed that we derive similar stellar masses.

Our data support the existence of the MZR at $z \sim 2$. It can be seen from Fig. 7 that we find, in agreement with many other studies of the MZR, a decrease in metallicity with decreasing stellar mass. Across the range of stellar mass $9.3 < \log(M/\mathrm{M}_\odot) < 10.5$, we observe a decrease of ∼0.3 dex in metallicity. Our data extend the Erb et al. (2006a) study as we probe masses below $\log(M/\mathrm{M}_\odot) \sim 10.0$, where their data could only place an upper limit on the metallicity. However, at a given value of stellar mass we measure lower metallicities than Erb et al. (2006a). This discrepancy is of the order of ∼0.2–0.3 dex in the highest mass bins. This offset is large considering the similar redshifts of the two samples; in Section 3.2, we explore a possible explanation for this offset given by the FMR.

3.2 Fundamental metallicity relation

To investigate the FMR, the SFR of the galaxy stacks were measured from the Hβ line flux as described in Section 2.3.3. We use the original FMR equation given in Mannucci et al. (2010) as opposed to its extension in Mannucci et al. (2011) as we find that Mannucci et al. (2010) FMR was derived using a sample that best matches the mass–SFR parameter space of our sample and the Erb et al. (2006a) sample. The Mannucci et al. (2011) FMR extension was derived by adding low-mass ($\log(M/\mathrm{M}_\odot) < 9.2$), low-SFR galaxies which are not representative of the galaxies in our data set. We find that using Mannucci et al. (2011) FMR parametrization produces discontinuities in the MZR at the high masses and high SFRs of our sample. However, we note that adopting the Mannucci et al. (2011) FMR parametrization does not significantly change the results discussed below.

According to the FMR, measuring lower metallicity in a given stellar mass bin should indicate that the average SFR in that bin is higher. Therefore, given that we measure lower metallicities than Erb et al. (2006a) across our stellar mass bins (Fig. 7), we should also observe elevated average SFRs in those bins. However, as illustrated in Fig. 8, the average SFR within our stellar mass bins is lower than the Erb et al. (2006a) data, converging to similar values towards the higher mass bins in our sample ($\sim 10^{10}\,\mathrm{M}_\odot$). This implies at least one of these data sets is in contradiction to the FMR. The problem is not alleviated by taking our SFRs from SED fitting as these are on average slightly lower than the Hβ derived SFRs as discussed in Section 2.3.3.

The FMR offset of various samples is illustrated in Fig. 9. This figure shows the difference between the metallicity observed and the metallicity predicted from the FMR given the Hβ derived SFR and the median values of stellar mass for each galaxy stack. It can be seen that the galaxies presented in this paper lie offset from the FMR by an average of ∼0.3 dex (dashed red line), whilst the Erb et al. (2006a) data are consistent with the FMR. This is also illustrated in Fig. 7, where the predicted positions of the two samples, based on the FMR, are shown in the MZ plane.

Also shown in Fig. 9 is the ∼0.5 dex offset of the AMAZE/LSD galaxies (Maiolino et al. 2008; Mannucci et al. 2009) at $z \sim 3$. The offset was calculated by taking the median values of mass, metallicity and SFR of the individual galaxies quoted in those papers. We convert the stellar mass from Maiolino et al. (2008) to be consistent with the Chabrier IMF used in this analysis. This offset of the $z \sim 3$ galaxies was also noted in Mannucci et al. (2010). As discussed in Section 2.3.4, we use the same calibration and the same set of emission lines to measure metallicities as the AMAZE/LSD surveys. Furthermore, our method of metallicity measurement is shown to return consistent results (see Fig. 6).

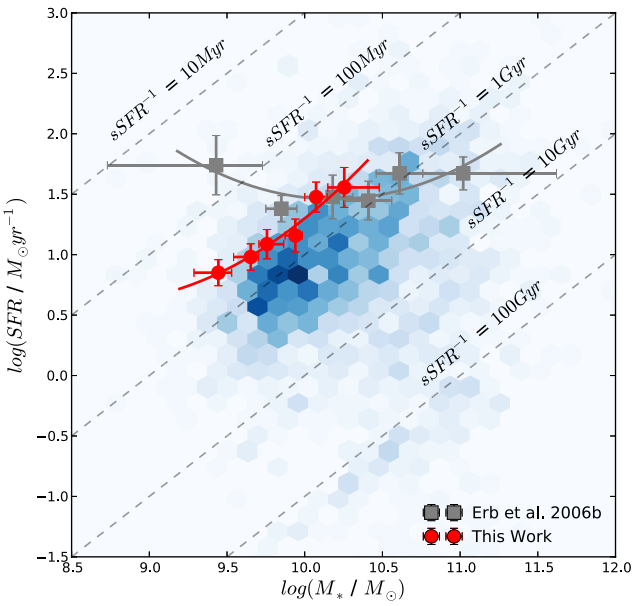

Figure 8. The position of the galaxies in our sample in the M_*–SFR plane. The red circles represent SFRs for the stacked galaxies presented in this paper with the SFR measured from the Hβ flux of the stacked spectra, converted to Hα via the common conversion factor Hα/Hβ = 2.86 (Kennicutt 1998). The grey squares represent SFRs for the stacked galaxies from Erb et al. (2006a; grey squares) with the SFR measured directly from Hα. The points are overlaid on a 2D histogram (in blue) of ∼3000 galaxies at $2 < z < 2.5$ from Whitaker et al. (2012) to provide a visual reference for the region of the M_*–SFR plane we are sampling. Polynomial fits to the Erb et al. (2006a) data (grey line) and the data presented here (red line) are indicated on the plot; these lines represent two slices along the FMR surface.

The fact that the $z \gtrsim 2$ galaxies presented here are offset from the FMR, as are the $z \sim 3$ galaxies from AMAZE/LSD, whilst the $z \gtrsim 2$ galaxies of Erb et al. (2006a) are in agreement, suggests that the choice of metallicity indicator may be affecting the measured metallicities. At first sight, it appears that this should not be the case, since all line ratios are calibrated to the same metallicity scale. However, these calibrations were made using local Universe star-forming galaxies which may not be representative of typical star-forming galaxies at high redshift. In Section 3.3, we investigate whether a change in the ionization conditions of H II regions at high redshifts affects the consistency of metallicity measurements using different sets of line ratios.

We also compare our results with data from Henry et al. (2013b) in Fig. 9 (blue triangles). The Henry et al. (2013b) sample consists of 83 grism spectra in the redshift range $1.3 < z < 2.3$, the spectra are stacked in four bins of stellar mass and metallicities are measured from the oxygen and Hβ lines using the Maiolino et al. (2008) calibrations. In this respect, our data and methods are very similar. The main difference between our data sets is that their sample spans a much wider redshift range resulting in a lower median redshift ($z = 1.76$), and the Henry et al. (2013b) sample probes to lower stellar masses ($\log(M/\mathrm{M}_\odot) < 10^{9.2}$). In Fig. 9, we only include the mass bins which overlap with our sample. Despite using the same

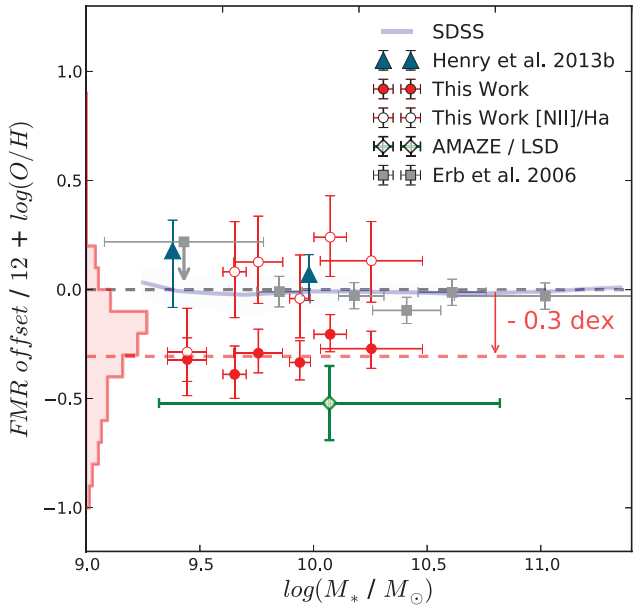

Figure 9. The difference in the observed values of metallicity from those predicted by the FMR as a function of stellar mass for various samples. The blue line is the original SDSS sample of Mannucci et al. (2010) with the blue shaded region representing the 1σ dispersion. The filled red circles represent the galaxy stacks presented in this paper. The green diamond shows a combination of $z \sim 3$ galaxies from the AMAZE/LSD surveys (Maiolino et al. 2008; Mannucci et al. 2009) and the grey squares show the Erb et al. (2006a) galaxies. The navy blue triangles show data from Henry et al. (2013b) for the mass bins in their data which overlap with our sample. Our results, within the error bars, are not consistent with the FMR derived in the local Universe persisting out to $z \gtrsim 2$, and yield an average offset of $\sim$0.3 dex. The open red circles represent the FMR offset predicted from the [N II]/Hα ratio of our sample using theoretical predictions for the evolution of ionization conditions in star-forming regions with redshift (see Section 3.3 for discussion)

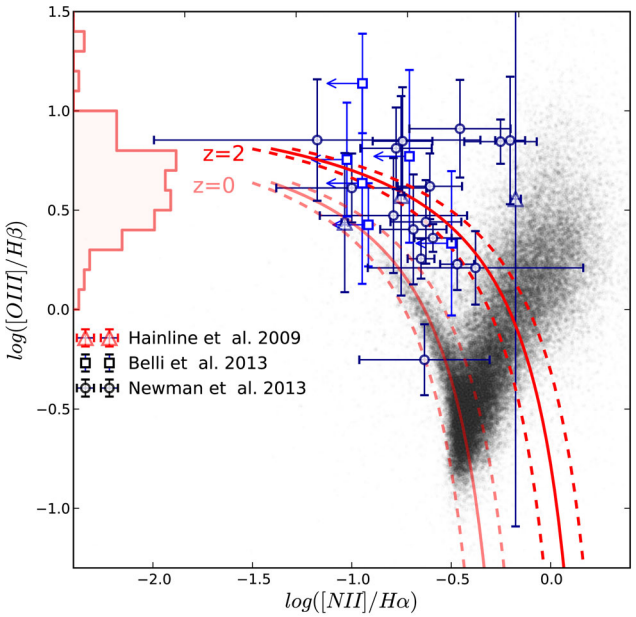

Figure 10. The BPT diagram of log([N II]/Hα) versus log([O III]/Hβ) used for diagnozing H II regions via emission-line ratios. The grey points show a sample of local SDSS galaxies from Brinchmann et al. (2013). The blue data points with error bars show the current data for $z \sim 2$ galaxies with measurements in all four emission lines, the open triangles show three galaxies from Hainline et al. (2009), the open squares show a sample of gravitationally lensed galaxies from Belli et al. (2013) and the open circles from Newman et al. (2013). A histogram of the distribution of the [O III] λ5007/Hβ ratio of the individual galaxies in our sample with measured Hβ is shown on the y-axis. The two red lines show the star-forming sequence of galaxies in the BPT diagram at $z = 0$ and 2, respectively, taken from the theoretical calibration of Kewley et al. (2013b).

line diagnostics, the data from Henry et al. (2013b) lie on the FMR relation, in disagreement with our results. However, their results are obtained by stacking over a wide redshift range, with a median redshift of $z = 1.76$, which is significantly lower than ours ($z = 2.16$); again an evolution in ionization conditions may explain the inconsistency as we discuss in Section 3.3.

3.3 Photoionization conditions

The metallicity calibrations outlined in Maiolino et al. (2008) are calibrated against local star-forming galaxies. The line-flux ratios used in these calibrations are known to have other dependences (e.g. ionization parameter and gas density), and therefore to use them at high redshift, we have to assume that the conditions in star-forming galaxies at high redshift do not differ significantly from those at low redshift. However, it is known that many properties of galaxies change at high redshift; for example at $z \sim 2$, galaxies are on average more compact and have higher SFR (e.g. Buitrago et al. 2008; Whitaker et al. 2012; McLure et al. 2013). These changes may affect the physical conditions within these galaxies. Below, we discuss the evolution of physical conditions in H II regions via the BPT diagram.

3.3.1 BPT diagram

The most commonly used diagnostic of photoionization conditions in H II regions is the BPT diagram (Baldwin et al. 1981) which has traditionally been used to separate star-forming galaxies from AGN in local galaxies. In recent years, the BPT diagram has been used to suggest that the ionization conditions of H II regions at $z > 1$ are different from those in the local Universe (e.g. Shapley et al. 2005; Erb et al. 2006b; Liu et al. 2008; Hainline et al. 2009; Nakajima et al. 2013). The BPT diagram is shown in Fig. 10. Plotted in black are a sample of galaxies from the SDSS DR7 MPA–JHU data release (Brinchmann et al. 2013) in the redshift range $0.04 < z < 0.5$. Fig. 10 also shows a handful of galaxies at $z \sim 2$ which have [O III] λ5007, Hβ, Hα and [N II] measurements and can therefore be placed in the BPT diagram (Hainline et al. 2009; Belli et al. 2013; Newman et al. 2013). It is evident that $z \sim 2$ galaxies are offset from the local BPT relation for star-forming galaxies, particularly striking are the large values of the [O III] λ5007/Hβ ratio seen at these redshifts, clearly offset from the locus of the SDSS galaxies. The offset is also seen at lower redshifts, for example in the DEEP2 survey of $1 < z < 1.5$ star-forming galaxies of Shapley et al. (2005) and the Trump et al. (2013) study of $z \sim 1.5$ galaxies using MOSFIRE and 3D-*HST*. For our sample, since we can only measure the [O III] λ5007/Hβ ratio, we show a histogram of the values for our full galaxy sample on the y-axis of Fig. 10. Our results are consistent with previously measured $z \sim 2$ ratios of individual objects; the median [O III] λ5007/Hβ ratio of our sample is 0.59. Of course, elevated values of the [O III] λ5007/Hβ ratio are not necessarily an indication of change in ionization conditions in H II regions as the ratio is also strongly metallicity dependent.

However, Brinchmann et al. (2008) and Kewley et al. (2013a,b) have investigated the evolution of the star-forming sequence with

redshift and claim that a larger ionization parameter, higher electron density and harder ionizing radiation field may be the cause of this evolution. Trump et al. (2013) on the other hand claim to detect nuclear activity in 2/3 of a sample of $z \sim 1.5$ galaxies, arguing that the majority of high-redshift galaxies show evidence for nuclear activity, such that high-redshift galaxies are offset from the star-forming line in the BPT diagram in the same way as AGN are offset in local galaxies. However, it is not clear that AGN diagnostics for local Universe galaxies can be directly applied to galaxies at high redshift given a change in ionization conditions (Kewley et al. 2013a). Most recently, Kewley et al. (2013b) used cosmological hydrodynamical simulations to investigate how the positions of star-forming galaxies in the BPT diagram evolve given different sets of assumptions for the ionization conditions within galaxies, concluding that the position changes depending on the hardness of ionizing radiation field, ionization parameter and electron densities. Kewley et al. (2013b) provide a redshift-dependent equation giving the position for the star-forming sequence of galaxies in the BPT diagram, the position of this main sequence at $z = 0$ and 2 is shown in Fig. 10. These star-forming sequences can be seen to be consistent with both the SDSS and current $z \sim 2$ data.

We note that the $z \sim 2$ star-forming sequence taken from Kewley et al. (2013b) is derived using the upper limit of the hardness of the ionizing radiation field and therefore represents the most extreme star-forming conditions realized in their models at this redshift (the $z = 0$ sequence being the lower limit at any redshift). As noted in Kewley et al. (2013b), until more statistically significant samples of galaxies become available at high redshift, it is only possible to assume an upper and lower limit of the star-forming sequence at any given redshift.

3.3.2 *O32 versus R_{23} diagram*

We can investigate the ionization state of our $z \gtrsim 2$ galaxies directly using the O32 versus R_{23} diagram (Lilly et al. 2003). The [O III] λλ4958,5007/[O II]λ3727 ratio (O32) is ionization parameter sensitive (Kewley & Dopita 2002; Brinchmann et al. 2008), and we can use this to test the typical ionization conditions in our galaxy sample. However, O32 is also sensitive to metallicity (Nagao et al. 2006), so a second more metallicity sensitive ratio must be used to break the degeneracy. We follow the method employed in various studies in the literature by plotting O32 versus R_{23} (e.g. Lilly et al. 2003; Hainline et al. 2009; Nakajima et al. 2013). The comparison is shown in Fig. 11, where we again plot the SDSS sample plus a sample at intermediate redshift ($0.47 < z < 0.92$) from Lilly et al. (2003). Our data seem to support the conclusions of Hainline et al. (2009) and Nakajima et al. (2013) that galaxies at $z > 2$ are systematically offset towards larger values of O32 at a given value of R_{23}. As discussed by those authors, this is evidence for higher ionization parameters at fixed metallicity. Therefore, as our data are consistent with an elevated ionization parameter relative to local galaxies, indicative of a harder ionization radiation field at $z \sim 2$, the galaxies in our sample should lie offset from the $z = 0$ star-forming galaxy sequence, and on, or towards, the $z = 2$ star-forming sequence as shown in Fig. 10.

3.3.3 *Implications for metallicity measurements*

An offset in the position of star-forming galaxies in the BPT diagram has implications for the metallicities derived from either the oxygen and Hβ or [N II] and Hα emission-line ratios. Since the metallicity calibrations of Maiolino et al. (2008) were made using local star-forming galaxies, metallicities derived from either oxygen and Hβ lines, or [N II]/Hα, will only be consistent for galaxies which lie on the $z = 0$ star-forming sequence. Fig. 12 illustrates this by comparing the [N II]/ Hα ratios inferred in three different ways for the individual galaxies in our sample with detections in all emission lines. First, we use the measured metallicities from oxygen and Hβ lines and work backwards from the Maiolino et al. (2008) calibrations to infer a [N II]/Hα ratio via

$$\log([\mathrm{N\,II}]/\mathrm{H}\alpha) = -0.7732 + 1.2357x - 0.2811x^2 - 0.7201x^3 - 0.3330x^4, \quad (1)$$

where $x = 12+\log(\mathrm{O/H}) - 8.69$.

Secondly and thirdly, we take the measured values of the [O III] λ5007/Hβ ratio and use the BPT star-forming sequence of Kewley et al. (2013b) at $z \sim 2$ and 0, respectively, to infer a [N II]/Hα ratio via

$$\log([\mathrm{N\,II}]/\mathrm{H}\alpha) = 0.1833z - 0.08 + \frac{0.61}{\log([\mathrm{O\,III}]/\mathrm{H}\beta) - 1.1 - 0.03z}. \quad (2)$$

Though we note, as described in Section 3.3, that at $z \sim 2$ this represents an upper limit on the [N II]/Hα ratio assuming extreme star formation conditions.

It can be seen from Fig. 12 that assuming the Maiolino et al. (2008) calibrations leads to systematically lower predicted values of the [N II]/Hα ratio compared to those derived using the Kewley et al. (2013b) star-forming sequence at $z \sim 2$. By contrast, the Maiolino et al. (2008) calibrations are in good agreement with [N II]/Hα ratios derived using Kewley et al. (2013b) star-forming sequence at $z = 0$. This implies the Maiolino et al. (2008) calibrations may

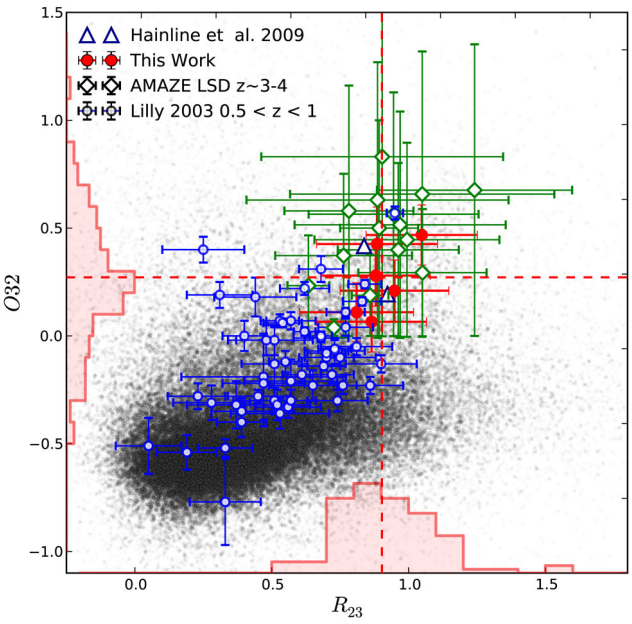

Figure 11. The O32 versus R_{23} diagnostic diagram used to differentiate change in ionization conditions from a change in metallicity of galaxies (Lilly et al. 2003; Hainline et al. 2009; Nakajima et al. 2013). The stacked galaxies in our sample are represented by filled red circles and the distribution of the individual galaxies are shown as histograms on each axis. The open blue circles represent a sample of $0.5 < z < 1$ galaxies taken from Lilly et al. (2003), the open green diamonds represent the AMAZE/LSD sample of $z \sim 3$–4 galaxies from Maiolino et al. (2008) and Mannucci et al. (2009).

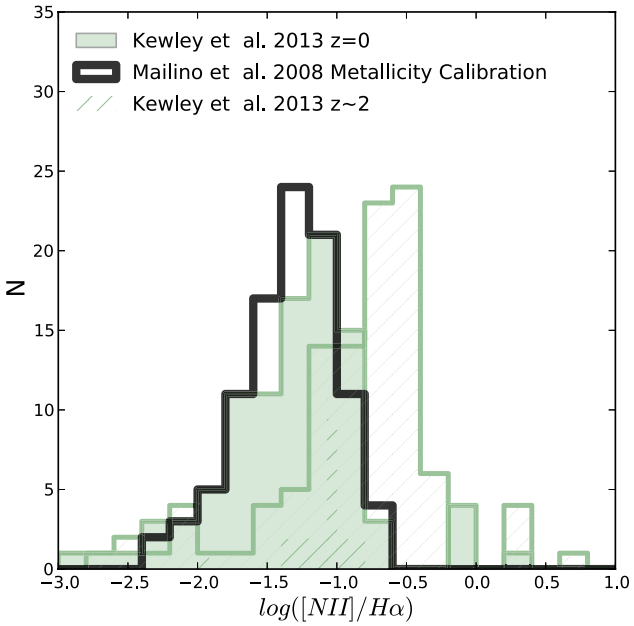

Figure 12. Histograms showing the inferred [N II]/Hα ratios for the individual galaxies in our sample with detections in all lines. The open black histogram is the [N II]/Hα ratio derived using the measured metallicity with the Maiolino et al. (2008) metallicity calibrations (equation 1). The filled green histogram is the ratio derived from the measured [O III] λ5007/Hβ ratio using the BPT diagram star-forming galaxy main-sequence equation at $z = 0$ from Kewley et al. (2013b; equation 2). The median and MAD of the ratios derived from Maiolino et al. (2008) and Kewley et al. (2013b) ($z = 0$), respectively, are -1.34 ± 0.2 and -1.26 ± 0.3 dex. The hatched green histogram is derived in the same way as the filled green histogram except using the galaxie's redshift in equation 2 ($z \sim 2$). The median and MAD of this distribution is -0.74 ± 0.3 dex.

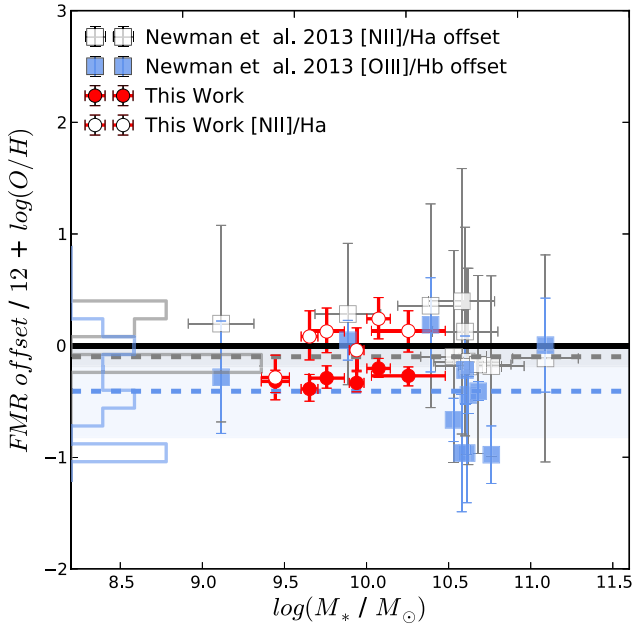

Figure 13. The FMR offset for a sample of $z \sim 2$ star-forming galaxies taken from Newman et al. (2013). The open grey squares show the metallicities derived from the [N II]/Hα ratio and the filled blue squares show the metallicities derived from the [O III] λ5007/Hβ ratio. On the y-axis, the grey and blue histograms show the distribution of the individual points with the horizontal dashed lines representing the median values of the distributions. The median and MAD for the [N II]/Hα ratio and [O III] λ5007/Hβ ratio, respectively, are -0.10 ± 0.09 and -0.44 ± 0.22 dex; the MAD is shown by the shaded region around the median line. For reference, the galaxies presented in this paper are plotted as filled red and open red circles as in Fig. 9.

not be applicable to high-redshift galaxies since, given a metallicity measured from the oxygen and Hβ lines, they predict [N II]/Hα ratios consistent with the local star-forming galaxy sequence, whereas $z \sim 2$ galaxies lie offset from this sequence towards higher [N II]/Hα ratios.

The effect of assuming the Kewley et al. (2013b) star-forming sequence on the FMR is shown in Fig. 9 where the open red circles represent the metallicities derived from the [N II]/Hα ratio assuming the Kewley et al. (2013b) conversion at the median redshift of the stacks ($z \sim 2$). Metallicities are then derived from the [N II]/Hα ratio using the Maiolino et al. (2008) calibrations. These theoretical [N II]/Hα ratios return systematically higher metallicities than those derived from oxygen and Hβ lines. This suggests that metallicities at high redshift based on the oxygen and Hβ lines are not directly comparable to metallicities measured from the [N II]/Hα ratio using local empirical calibrations. It can also be seen from Fig. 9 that these theoretical [N II]/Hα metallicities are in better agreement with the FMR. Thus, the consistency of the Erb et al. (2006a) data with the FMR, and the inconsistency of our data with it, may be explained by this evolution of galaxies in the BPT diagram with redshift.

The evolution in ionization conditions of galaxies with redshift may also explain the apparent discrepancy between our data and the Henry et al. (2013b) results. Since their sample spans a much larger range in redshifts ($1.3 < z < 2.3$), and therefore a larger range of ionization conditions at a given metallicity, it is perhaps not surprising that in a given mass bin our metallicities are offset. However, a detailed analysis of the effect of stacking across a large range in redshift is beyond the scope of this work, and we focus here on the discrepancy between [N II]/Hα and oxygen and Hβ metallicities using local empirical calibrations at high redshift.

3.3.4 *Newman et al. (2013) data*

To further investigate the discrepancy between line indicators, we took a sample of 11 $z \sim 2$ galaxies from Newman et al. (2013) which have flux measurements in the [N II], Hα, [O III] λ5007 and Hβ lines, and have measured masses and SFRs. The masses were derived using a Chabrier (2003) IMF, so no conversion was necessary for comparison with our results. Metallicities were derived from the [N II]/Hα and [O III] λ5007/Hβ line ratios individually and the offset from the predicted FMR value computed as described in Section 3.2.

Note that the full set of oxygen and Hβ calibrations used on our data could not be used on the Newman et al. (2013) data, so oxygen and Hβ metallicities were derived from the [O III] λ5007/Hβ ratio alone. When used independently, the [O III] λ5007/Hβ metallicity calibration is double valued (see Maiolino et al. 2008 and Fig. 5), so to derive a metallicity from the [O III] λ5007/Hβ ratio solely, it is necessary to choose between the upper and lower solutions. For each galaxy in the Newman et al. (2013) sample, the upper solution was taken, as it was in agreement with the high metallicities implied from the [N II]/Hα solution. Nevertheless, even when taking the upper solution for [O III] λ5007/Hβ, the [N II]/Hα ratio returns systematically higher values of metallicity in better agreement with the FMR. This is illustrated in Fig. 13. The median and MAD for the [N II]/Hα ratio and [O III] λ5007/Hβ ratio, respectively, are

-0.10 ± 0.09 and -0.44 ± 0.22 dex. It is also interesting to note that the Newman et al. (2013) data probe a higher mass regime, similar to the Erb et al. (2006a) sample, indicating that this effect is seen across a wide range in mass $\sim 9.5 < \log(M/\mathrm{M}_{\odot}) < 11.5$.

To summarize, these arguments suggest that if spectra are obtained in the *K* band for our galaxy sample such that the [N II]/Hα ratio can be measured, this ratio will return a higher metallicities by $\sim$0.3–0.5 dex over metallicities derived from the oxygen and Hβ lines using the Maiolino et al. (2008) metallicity calibrations. Thus, from the [N II]/Hα ratio, one will observe better consistency with the FMR as in Erb et al. (2006a), whilst from the oxygen and Hβ lines an offset will be observed as in this work and at $z \sim 3$ (Maiolino et al. 2008; Mannucci et al. 2009). A detailed comparison of the two line ratios as metallicity indicators is beyond the scope of this work, though resolving this issue is clearly of key importance if we want to conclude anything more definitive regarding the FMR at high redshift.

4 SUMMARY AND CONCLUSIONS

We have selected a sample of 93 galaxies in the redshift range $2.0 < z < 2.3$ from an independent reduction of the 3D-*HST* spectroscopic grism survey. In this redshift range, the [O II]λ3727, [O III] $\lambda\lambda$4958,5007 and Hβ emission lines fall within the wavelength range of the grism spectra. Our aim is to use those emission lines to measure the metallicities of galaxies and to measure their masses from the ancillary CANDELS photometry available in the 3D-*HST* survey fields. We stack the galaxies in bins of stellar mass and construct an MZR for this sample, which we can directly compare with the previous $z \gtrsim 2$ study of Erb et al. (2006a). We then measure the SFR from the Hβ line and use this to investigate the FMR at these redshifts. Below is a summary of our results.

(i) We measure the metallicities of our galaxies from the [O II]λ3727, Hβ and [O III] $\lambda\lambda$4958,5007 emission lines via the calibrations of Maiolino et al. (2008). We find an MZR in our galaxysample consistent with the MZR reported elsewhere in the literature (e.g. Erb et al. 2006a; Mannucci et al. 2009; Yuan, Kewley & Richard 2013); in that we observe a decrease in metallicity with a decrease in mass of our galaxy stacks (Fig. 7). However, our MZR is offset to lower metallicities at a given stellar mass from the $z \gtrsim 2$ MZR of Erb et al. (2006a).

(ii) We investigate this metallicity offset using the FMR proposed by Mannucci et al. (2010) which incorporates mass, metallicity and SFR in an attempt to explain the scatter and redshift evolution of the MZR. However, we find that our data are apparently inconsistent with the FMR. We measure metallicities lower by $\sim$0.3 dex from those predicted from the FMR given our measured masses and SFRs (Fig. 9). The previous $z \gtrsim 2$ data of Erb et al. (2006a) are consistent with the FMR as discussed in Mannucci et al. (2010); therefore, there is a discrepancy between the current $z \gtrsim 2$ data. One difference between the Erb et al. (2006a) study and our own is the metallicity indicator used. Erb et al. (2006a) use the [N II]/Hα ratio whereas our metallicities are based on the oxygen and Hβ lines. Interestingly, our method for metallicity measurement follows previous $z \sim 3$ MZR studies (Maiolino et al. 2008; Mannucci et al. 2009), which find a similar offset from the FMR using the same set of emission lines.

(iii) We investigate the ionization conditions of our galaxies to attempt to explain the FMR offset. We construct a O32 versus R_{23} diagram following the method of Lilly et al. (2003), Hainline et al. (2009) and Nakajima et al. (2013; Fig. 11). We find, consistent with previous $z \gtrsim 2$ data, evidence for an enhancement of the O32 ratio at a fixed value of R_{23} indicative of an enhanced ionization parameter in these galaxies at fixed metallicity.

(iv) We note that from small samples of $z \sim 2$ star-forming galaxies with [O II]λ3727, Hβ, [O III] $\lambda\lambda$4958,5007, [N II] and Hα measured (e.g. Hainline et al. 2009; Belli et al. 2013; Newman et al. 2013), there is evidence that higher ionization parameters cause an offset from the star-forming galaxy sequence in the BPT diagram of local galaxies (Fig. 10). This offset has recently been quantified by Kewley et al. (2013b) who provide an equation for the evolution of the star-forming sequence with redshift (Fig. 10 and equation 2). Given the evidence for an increased ionization parameter in our galaxies, and [O III] λ5007/Hβ ratios consistent with other $z \gtrsim 2$ samples, we conclude that our galaxies would most likely also lie offset from the $z = 0$ star-forming galaxy sequence.

(v) Working backwards from our metallicity measurements, we can infer the expected [N II]/Hα ratio of our galaxies from the Maiolino et al. (2008) calibrations (equation 1). We can also infer the expected [N II]/Hα ratio from the Kewley et al. (2013b) BPT main-sequence calibration at $z \sim 2$ and 0 (equation 2). The inferred ratios from the Maiolino et al. (2008) calibrations are systematically lower than the inferred ratios from Kewley et al. (2013b) at $z \sim 2$ but consistent with those at $z = 0$ (Fig. 12). This implies the [O III] λ5007/Hβ and [N II]/Hα metallicity calibrations are not comparable for galaxies significantly offset from the star-forming main sequence of the BPT diagram.

(vi) Taking the inferred [N II]/Hα ratios from the Kewley et al. (2013b) calibration at $z \sim 2$, we find that our galaxies would fall into better agreement with the FMR (Fig. 9). For further investigation, we also take a sample of 11 $z \sim 2$ galaxies with measured [N II], Hα, [O III] λ5007 and Hβ fluxes from Newman et al. (2013) and confirm that the metallicities derived using the Maiolino et al. (2008) calibrations are systematically higher than when using the [N II]/Hα ratio (Fig. 13), following the trend in our data inferred using Kewley et al. (2013b). This further supports the conclusion that metallicities derived from the oxygen and Hβ lines are not equivalent to those derived using [N II]/Hα at high redshift.

(vii) We conclude that the discrepancy between our data and the FMR, as well as the $z \sim 3$ data of Maiolino et al. (2008) and Mannucci et al. (2009), may be the result of using locally calibrated empirical metallicity relations. If this is the case, the evolution of the FMR at high redshift cannot be quantified using these empirical calibrations which do not fully account for the change in ionization conditions in star-forming galaxies with redshift.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous referee for their useful comments that have helped in improving the paper. FC and MC acknowledge the support of the Science and Technology Facilities Council (STFC) via the award of an STFC Studentship and an STFC Advanced Fellowship, respectively. RJM acknowledges the support of the European Research Council via the award of a Consolidator Grant (PI McLure). JSD acknowledges the support of the European Research Council via the award of an Advanced Grant and the support of the Royal Society via a Wolfson Research Merit Award. This work is based on observations taken by the 3D-*HST* Treasury Programme (GO 12177 and 12328) with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. This work is based (in part) on observations made with the *Spitzer* Space Telescope, which is operated by the Jet Propulsion Laboratory, California

Institute of Technology under a contract with NASA. This research made use of Astropy, a community-developed core PYTHON package for astronomy (Astropy Collaboration et al. 2013), NumPy and SciPy (Oliphant 2007), Matplotlib (Hunter 2007), IPython (Pérez & Granger 2007) and NASA's Astrophysics Data System Bibliographic Services.

REFERENCES

Andrews B. H., Martini P., 2013, ApJ, 765, 140
Astropy Collaboration et al., 2013, A&A, 558, A33
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Belli S., Jones T., Ellis R. S., Richard J., 2013, ApJ, 772, 141
Bowler R. A. A. et al., 2012, MNRAS, 426, 2772
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brammer G. B. et al., 2012, ApJS, 200, 13
Brinchmann J., Pettini M., Charlot S., 2008, MNRAS, 385, 769
Brinchmann J., Charlot S., Kauffmann G., Heckman T., White S. D. M., Tremonti C., 2013, MNRAS, 432, 2112
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buitrago F., Trujillo I., Conselice C. J., Bouwens R. J., Dickinson M., Yan H., 2008, ApJ, 687, L61
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Chabrier G., 2003, PASP, 115, 763
Civano F. et al., 2012, ApJS, 201, 30
Cowie L. L., Barger A. J., 2008, ApJ, 686, 72
Cresci G., Mannucci F., Sommariva V., Maiolino R., Marconi A., Brusa M., 2012, MNRAS, 421, 262
Daddi E. et al., 2007, ApJ, 670, 156
Domínguez A. et al., 2013, ApJ, 763, 145
Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, ApJ, 672, L107
Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006a, ApJ, 644, 813
Erb D. K., Steidel C. C., Shapley A. E., Pettini M., Reddy N. A., Adelberger K. L., 2006b, ApJ, 647, 128
Förster Schreiber N. M. et al., 2009, ApJ, 706, 1364
Galametz A. et al., 2013, ApJS, 206, 10
Grogin N. A. et al., 2011, ApJS, 197, 35
Guo Y. et al., 2013, ApJS, 207, 24
Hainline K. N., Shapley A. E., Kornei K. A., Pettini M., Buckley-Geer E., Allam S. S., Tucker D. L., 2009, ApJ, 701, 52
Henry A., Martin C. L., Finlator K., Dressler A., 2013a, ApJ, 769, 148
Henry A. et al., 2013b, ApJ, 776, L27
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ilbert O. et al., 2006, A&A, 457, 841
Juneau S., Dickinson M., Alexander D. M., Salim S., 2011, ApJ, 736, 104
Kashino D. et al., 2013, ApJ, 777, L8
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Kennicutt R. C. J., 1998, A&AR, 36, 189
Kewley L. J., Dopita M. A., 2002, ApJS, 142, 35
Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183
Kewley L. J., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013a, ApJ, 774, L10
Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013b, ApJ, 774, 100
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Kulas K. R. et al., 2013, ApJ, 774, 130
Kümmel M., Walsh J. R., Pirzkal N., Kuntschner H., Pasquali A., 2009, PASP, 121, 59
Lara-Lopez M. A. et al., 2010, A&A, 521, L53
Lara-López M. A., López-Sánchez Á. R., Hopkins A. M., 2013, ApJ, 764, 178
Lilly S. J., Carollo C. M., Stockton A. N., 2003, ApJ, 597, 730
Liu X., Shapley A. E., Coil A. L., Brinchmann J., Ma C.-P., 2008, ApJ, 678, 758
Ly C., Malkan M. A., Nagao T., Kashikawa N., Shimasaku K., Hayashi M., 2014, ApJ, 780, 122
Maiolino R. et al., 2008, A&A, 488, 463
Mannucci F. et al., 2009, MNRAS, 398, 1915
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Mannucci F., Salvaterra R., Campisi M. A., 2011, MNRAS, 414, 1263
McCracken H. J. et al., 2012, A&A, 544, A156
McLure R. J. et al., 2013, MNRAS, 428, 1088
Mignoli M. et al., 2005, A&A, 437, 883
Nagao T., Maiolino R., Marconi A., 2006, A&A, 459, 85
Nakajima K., Ouchi M., 2013, preprint (arXiv:1309.0207)
Nakajima K., Ouchi M., Shimasaku K., Hashimoto T., Ono Y., Lee J. C., 2013, ApJ, 769, 3
Newman S. F. et al., 2013, ApJ, 781, 21
Niino Y., 2012, ApJ, 761, 126
Oliphant T. E., 2007, Comput. Sci. Eng., 9, 10
Panter B., Jimenez R., Heavens A. F., Charlot S., 2008, MNRAS, 391, 1117
Pérez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Pettini M., Pagel B. E. J., 2004, MNRAS, 348, L59
Pirzkal N. et al., 2013, ApJ, 772, 48
Price S. H. et al., 2013, preprint (arXiv:1310.4177)
Richard J., Jones T., Ellis R., Stark D. P., Livermore R., Swinbank M., 2011, MNRAS, 413, 643
Rodrigues M. et al., 2008, A&A, 492, 371
Roseboom I. G. et al., 2012, MNRAS, 426, 1782
Savaglio S. et al., 2005, ApJ, 635, 260
Shapley A. E., Coil A. L., Ma C.-P., Bundy K., 2005, ApJ, 635, 1006
Shirazi M., Brinchmann J., Rahmati A., 2013, preprint (arXiv:1307.4758)
Storey P. J., Zeippen C. J., 2000, MNRAS, 312, 813
Stott J. P. et al., 2013, MNRAS, 436, 1130
Tremonti C. A. et al., 2004, ApJ, 613, 898
Trump J. R. et al., 2013, ApJ, 763, L6
Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, ApJ, 754, L29
Xue Y. Q. et al., 2011, ApJS, 195, 10
Yabe K. et al., 2012, PASJ, 64, 60
Yates R. M., Kauffmann G., Guo Q., 2012, MNRAS, 422, 215
Yuan T. T., Kewley L. J., Richard J., 2013, ApJ, 763, 9
Zahid H. J., Kewley L. J., Bresolin F., 2011, ApJ, 730, 137
Zahid H. J. et al., 2013, preprint (arXiv:1310.4950)

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