



HAL
open science

The Stellar Mass versus Stellar Metallicity Relation of Star-forming Galaxies at $1.6 \leq z \leq 3.0$ and Implications for the Evolution of the α -enhancement

Daichi Kashino, Simon J. Lilly, Alvio Renzini, Emanuele Daddi, Giovanni Zamorani, John D. Silverman, Olivier Ilbert, Ying-Jie Peng, Vincenzo Mainieri, Sandro Bardelli, et al.

► To cite this version:

Daichi Kashino, Simon J. Lilly, Alvio Renzini, Emanuele Daddi, Giovanni Zamorani, et al.. The Stellar Mass versus Stellar Metallicity Relation of Star-forming Galaxies at $1.6 \leq z \leq 3.0$ and Implications for the Evolution of the α -enhancement. The Astrophysical Journal, 2022, 925, 10.3847/1538-4357/ac399e . insu-03667483

HAL Id: insu-03667483

<https://insu.hal.science/insu-03667483>

Submitted on 13 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



The Stellar Mass versus Stellar Metallicity Relation of Star-forming Galaxies at $1.6 \leq z \leq 3.0$ and Implications for the Evolution of the α -enhancement

Daichi Kashino^{1,2}, Simon J. Lilly³, Alvio Renzini⁴, Emanuele Daddi⁵, Giovanni Zamorani⁶, John D. Silverman^{7,8}, Olivier Ilbert⁹, Ying-jie Peng^{10,11}, Vincenzo Mainieri¹², Sandro Bardelli⁶, Elena Zucca⁶, Jeyhan S. Kartaltepe¹³, and David B. Sanders¹⁴

¹ Institute for Advanced Research, Nagoya University, Nagoya 464-8601, Japan

² Division of Particle and Astrophysical Science, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan; kashino.daichi@b.mbox.nagoya-u.ac.jp

³ Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland

⁴ INAF-Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padova, Italy

⁵ CEA, Irfu, DAp, AIM, Université Paris-Saclay, Université de Paris, CNRS, F-91191 Gif-sur-Yvette, France

⁶ INAF—Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, I-40129, Bologna, Italy

⁷ Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo, Kashiwa, 277-8583, Japan (Kavli IPMU, WPI)

⁸ Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

⁹ Aix Marseille Université, CNRS, LAM, Laboratoire d'Astrophysique de Marseille, Marseille, France

¹⁰ Kavli Institute for Astronomy and Astrophysics, Peking University, 5 Yiheyuan Road, Beijing 100871, People's Republic of China

¹¹ Department of Astronomy, School of Physics, Peking University, 5 Yiheyuan Road, Beijing 100871, People's Republic of China

¹² European Southern Observatory, Karl-Schwarzschild-Strasse 2, Garching bei München, Germany

¹³ School of Physics and Astronomy, Rochester Institute of Technology, 84 Lomb Memorial Drive, Rochester, NY 14623, USA

¹⁴ Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

Received 2021 September 13; revised 2021 October 29; accepted 2021 November 9; published 2022 January 26

Abstract

We measure the relationship between stellar mass and stellar metallicity for 1336 star-forming galaxies at $1.6 \leq z \leq 3.0$ using rest-frame far-ultraviolet spectra from the zCOSMOS-deep survey. High signal-to-noise ratio composite spectra containing stellar absorption features are fit with stellar population synthesis model spectra of a range of stellar metallicity. We find stellar metallicities, which mostly reflect instantaneous iron abundances, scaling as $[\text{Fe}/\text{H}] = -(0.81 \pm 0.01) + (0.32 + 0.03)\log(M_*/10^{10}M_\odot)$ across the stellar mass range of $10^9 \lesssim M_*/M_\odot \lesssim 10^{11}$. The instantaneous oxygen-to-iron ratio (α -enhancement) inferred using the gas-phase mass–metallicity relation is on average found to be $[\text{O}/\text{Fe}] \approx 0.47$, being higher than the local $[\text{O}/\text{Fe}] \approx 0$. The observed changes in $[\text{O}/\text{Fe}]$ and $[\text{Fe}/\text{H}]$ are reproduced in simple gas-regulator models with steady star formation histories. Our models show that the $[\text{O}/\text{Fe}]$ is determined almost entirely by the instantaneous specific star formation rate alone while being independent of the mass and the characteristic of the gas regulation. We also find that the locations of $\sim 10^{10}M_\odot$ galaxies at $z \sim 2$ in the $[\text{O}/\text{Fe}]$ –metallicity planes are in remarkable agreement with the sequence of low-metallicity thick-disk stars in our own Galaxy. This manifests a beautiful concordance between the results of Galactic archeology and observations of high-redshift Milky Way progenitors. There remains, however, a question of how and when the old metal-rich, low α/Fe stars seen in the bulge had formed by $z \sim 2$ because such a stellar population is not seen in our data and is difficult to explain in the context of our models.

Key words: High-redshift galaxies – Metallicity – Abundance ratios – Galaxy chemical evolution – Galaxy evolution – Chemical enrichment – Chemical abundances

1. Introduction

The metallicity of galaxies can be measured for either stars or the interstellar medium (ISM) through analysis of galaxy spectra, either the stellar absorption lines in the integrated light of the stellar population(s) or the nebular emission lines from gaseous H II regions (see Maiolino & Mannucci 2019 for a recent review). The gas-phase and stellar metallicities reflect different aspects of the evolutionary history of the galaxies.

The gas-phase metallicity refers usually to the abundance of oxygen relative to hydrogen, which is often measured using emission lines in the rest-frame optical wave band that are produced by the ionized gas in star-forming regions. The so-called “direct method” is based on the detection of faint auroral

lines (e.g., $[\text{O III}] \lambda 4363$) and determines the electron temperature and metallicity with high accuracy (e.g., Andrews & Martini 2013; Ly et al. 2014; Kashino & Inoue 2019; Kojima et al. 2020; Sanders et al. 2020). The so-called “strong-line” methods, which use empirical relations between the metallicity and the ratios of strong optical emission lines (e.g., $([\text{O II}] + [\text{O III}])/\text{H}\beta$), have been widely applied to estimate metallicities from spectra with low or moderate signal-to-noise ratio (S/N; e.g., Pettini & Pagel 2004; Nagao et al. 2006; Maiolino et al. 2008; Curti et al. 2017). The gas-phase metallicities reflect the “instantaneous” oxygen abundance in star-forming regions at the time of observation.

The stellar metallicity can be measured through absorption lines caused by metal ions, such as iron and magnesium, in the photospheres of stars. Measurement is generally carried out by comparing observed spectra with synthetic spectra from stellar population synthesis models. Some standardized indices, which represent the absorption depths for particular, relatively strong absorption features, or a combination of absorption features, have

been conventionally used for estimating the average stellar metallicity (e.g., the 1978 Å index; Rix et al. 2004; see also Halliday et al. 2008; Onodera et al. 2015). More recently, full spectral fitting that uses all the information contained in the spectra has been employed (Onodera et al. 2015; Steidel et al. 2016; Zahid et al. 2017; Leethochawalit et al. 2018; Kriek et al. 2019; Harikane et al. 2020; Topping et al. 2020a, 2020b). In any case, these measurements require a high-S/N detection of the continuum emission and are thus generally more expensive than the gas-phase metallicity measurements using the strong-line methods.

An important point is that, in contrast to the gas-phase metallicity, the stellar metallicity is measured as the luminosity-weighted average value across all the different stellar populations that contribute to the integrated light of the galaxy at a particular wavelength. The inferred metallicity may therefore also depend on which portion of the spectrum is used. For example, the metallicity derived from the rest-frame optical light reflects the light from older (i.e., possibly lower-metallicity) populations, whereas that from the far-ultraviolet (FUV) spectrum is more weighted toward younger (i.e., possibly higher-metallicity) populations and should thus be closer to the abundance in the gas phase.

Another key aspect is that oxygen and iron, usually traced by gas-phase and stellar metallicities, respectively, form through different channels: oxygen or the α -elements are supplied mainly through core-collapse supernovae (CCSNe), while the Type Ia supernovae (SNe Ia) are the main supplier of the iron-peak elements. Therefore, the past star formation history (SFH) of the galaxies is imprinted in the abundance pattern between these species, often called α -enhancement, due to the time delay of SNe Ia (from 40 Myr to several Gyr) since the formation of their progenitor stars.

The overall relationship between galaxy stellar mass (M_*) and metallicity, often called the mass–metallicity relation (MZR), has long been thought to be a fundamental measurement to constrain models of galaxy evolution (e.g., Lequeux et al. 1979). In the local universe, a tight correlation between these two quantities has been robustly established both for the gas-phase metallicity (Tremonti et al. 2004; Andrews & Martini 2013; Curti et al. 2020) and for the metallicity of the stellar component (Gallazzi et al. 2005; Zahid et al. 2017) using the Sloan Digital Sky Survey (SDSS; York et al. 2000). At high redshifts, the gas-phase MZR has been measured back to $z \sim 4$ by many authors, mostly using strong-line methods (e.g., Erb et al. 2006; Zahid et al. 2011; Yabe et al. 2012; Zahid et al. 2014a, 2014b; Sanders et al. 2015; Kashino et al. 2017) with only a few cases where the direct method has been used (Ly et al. 2016; Sanders et al. 2020). The evolution of the gas MZR is established with the metallicity monotonically decreasing at fixed (observed) M_* with redshift.

In contrast, the measurement of the stellar mass–stellar metallicity (Z_*) relation (hereafter stellar MZR) beyond the local universe is to date very limited (Cullen et al. 2019; Calabrò et al. 2021). A notable work was recently carried out by Cullen et al. (2019), who presented an M_* – Z_* correlation over $10^{8.5} \lesssim M_*/M_\odot \lesssim 10^{10.5}$ using a large statistical sample of star-forming galaxies at $z = 2.5$ – 5.0 (see also Cullen et al. 2020). We are, however, still a long way from being able to constrain the evolution of the stellar MZR through cosmic time. Given the limited number of the existing measurements, independent measurements based on a different data set are highly desired.

In this work, we measure the stellar MZR for a large sample of star-forming galaxies at $1.6 \leq z \leq 3.0$ by utilizing the rest-frame

FUV spectra obtained with the Visible Multi-Object Spectrograph (VIMOS) mounted on the Very Large Telescope (VLT) UT3 in the zCOSMOS-deep survey (Lilly et al. 2007; S. J. Lilly et al., in preparation). We then explore the evolution of the oxygen-to-iron abundance pattern that is inferred from the comparison with the gas-phase metallicity measurements.

The paper is organized as follows. Section 2 presents an overview of the observations and describes the sample selection. Section 3 describes our spectral analysis for estimating the stellar metallicities. The results are presented in Section 4. Section 5 presents further attempts for interpreting the observations by using gas-regulated chemical evolution models to track the iron and oxygen chemical enrichment. We then compare our results and models with data of the Galactic stars to explore the link with the Galactic archeology in Section 6. Section 7 provides a summary of the paper.

We adopt the solar metallicity values of $12 + \log(\text{O}/\text{H})_\odot = 8.69$ and $Z_\odot = 0.0142$ (Asplund et al. 2009). Here Z denotes the overall metal mass fraction. We use Z_{Fe} and Z_{O} when specifying the element, either iron or oxygen. The notation $[A/B]$ denotes the logarithmic elemental abundance ratio normalized to solar, i.e., $[A/B] = \log(N_A/N_B) - \log(N_A/N_B)_\odot$. The solar oxygen and iron mass fractions are $Z_{\text{O},\odot} = 0.00561$ and $Z_{\text{Fe},\odot} = 0.00126$, respectively. Magnitudes are quoted on the AB system. The Chabrier (2003) initial mass function (IMF) is used throughout. This paper uses a standard flat cosmology ($h = 0.7$, $\Omega_{\text{M}} = 0.3$, $\Omega_{\Lambda} = 0.7$).

2. Data and Galaxy Sample

2.1. Observations

The zCOSMOS-deep redshift survey has observed around 10^4 galaxies in the central $\sim 0.8 \text{ deg}^2$ of the COSMOS field (Scoville et al. 2007). Here we provide a brief description and refer the reader to Lilly et al. (2007) and Kashino et al. (2021) for more details.

The observations were carried out using VLT/VIMOS (Le Fèvre et al. 2003) with the low-resolution blue grism with $1''0$ slits, yielding a spectral resolution of $R \sim 200$ and a spectral coverage of ≈ 3600 – 6700 Å . The selection of the targets was performed based on a then-current version of the COSMOS photometric catalog. All of the objects were color-selected through a *BzK* (Daddi et al. 2004) or *ugr* (Steidel et al. 2004) method with a blue magnitude cut $B_{\text{AB}} < 25.25$. These selection criteria isolate star-forming galaxies in a range $1.4 \lesssim z \lesssim 3.0$ (Lilly et al. 2007). Redshifts were visually inspected in 2D and 1D reduced spectra by identifying multiple prominent spectral features in the rest-frame FUV window or Ly α emission line and break.

Figure 1 shows six representative examples of individual galaxy spectra spanning a range of redshifts and the confidence class. Note that the additional flux calibration described in Section 3.1 has been applied to these spectra. The wavelength regions without data are those impacted by strong sky lines or zeroth-order contamination. Those of class = 3 or 4 present robust redshift identification based on several strong absorption lines and/or a strong Ly α emission line. The identification of such spectral features becomes less secure for class = 2 spectra.

2.2. Sample Selection

We constructed the sample used in this paper from the full catalog of the zCOSMOS-deep survey (S. J. Lilly et al., in preparation). The sample is limited to those having a clear

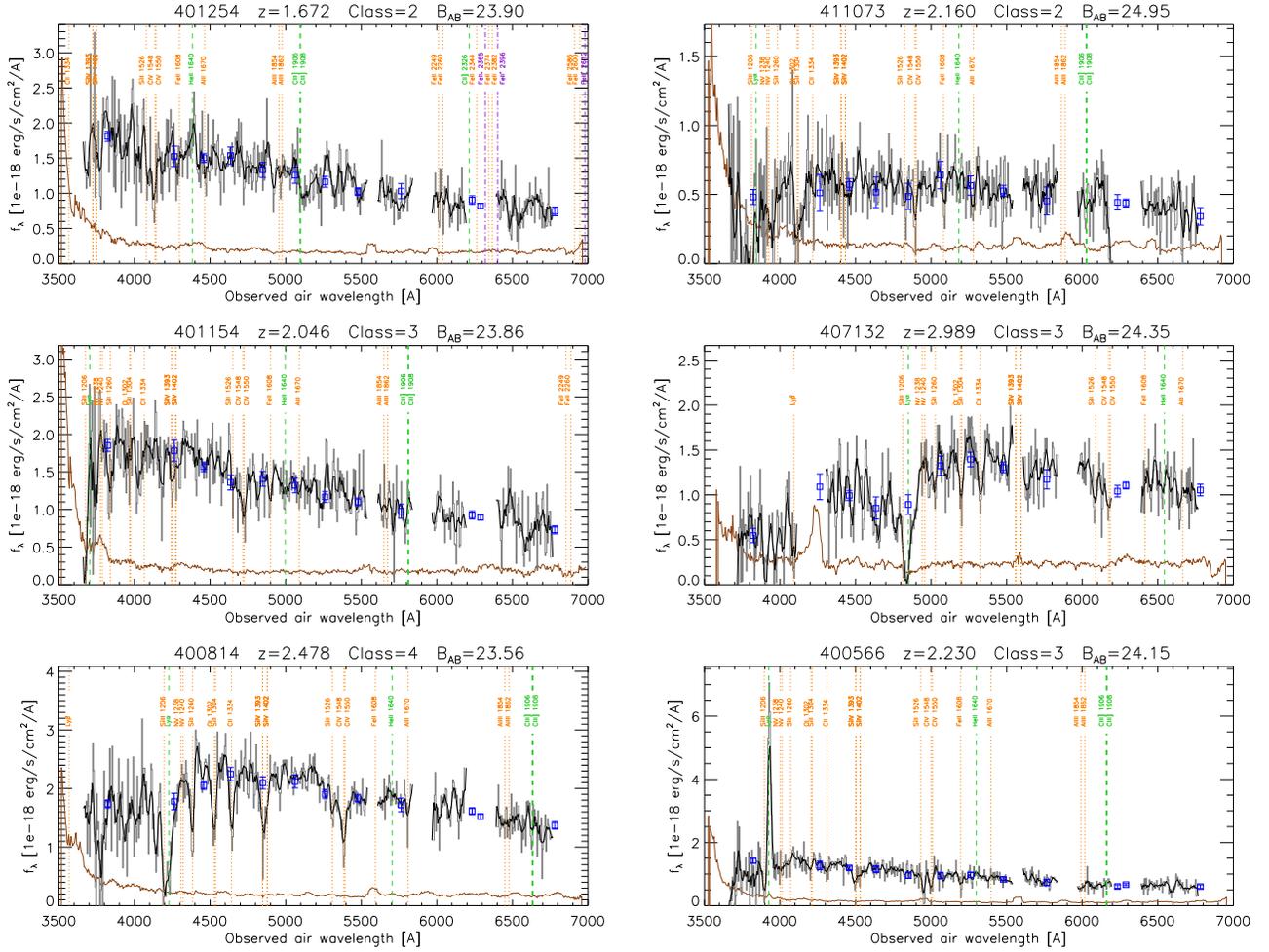


Figure 1. Examples of representative individual galaxy spectra, after the application of the flux calibration that is described in Section 3.1. Typical spectra are shown for each of the redshift confidence classes (2, 3, and 4). The bottom right panel presents an example of a strong Ly α emitter. The gray lines indicate the observed spectra. The black lines present spectra smoothed with an arbitrary Gaussian kernel of $\sigma = 5$ pixel. The brown lines indicate the 1σ error spectra. The blue squares indicate the photometric fluxes within the 12 bands that are used for flux calibration. Positions of some strong spectral features are marked by horizontal lines.

photometric counterpart in the COSMOS2015 photometric catalog (Laigle et al. 2016). Galaxies detected in X-rays (Civano et al. 2016) are also excluded to remove possible active galactic nuclei (AGNs) from the sample.

The redshift range for the current analysis is limited to $1.6 \leq z \leq 3.0$ so that the VIMOS spectrum covers the range of $\lambda_{\text{rest}} \approx 1400\text{--}1700$ Å for all the galaxies. We adopt all objects with a very secure zCOSMOS-deep redshift (Confidence Class = 3 or 4¹⁵) within the redshift range. For those with Class = 2, we use only those that are consistent to within $|z_{\text{phot}} - z_{\text{spec}}| / (1 + z_{\text{spec}}) \leq 0.1$ of the photometric redshift in the COSMOS2015 catalog.¹⁶

The redshifts in these two categories are both estimated to be $\geq 99\%$ reliable (Lilly et al., in preparation). We do not use any of the objects with less secure redshifts, nor any of those with broad emission lines (i.e., Class +10). Finally, we visually inspected individual sources and excluded 34 sources (2.5% of

the remaining sample) for which the FUV continuum is barely detected (which are likely to be Ly α emitters with very faint continua), or that show obvious AGN signatures (e.g., strong NV $\lambda 1240$ emission lines), or that suffered from severe spectral contamination from companion sources that fall in the same slit.

The final sample consists of 1336 galaxies. Figure 2 shows the redshift distribution of the sample with the median redshift $\langle z \rangle_{\text{med}} = 2.22$. We note that the redshifts were all determined with strong absorption lines owing to mostly carbon and/or silicon ions in the ISM, but not with any kind of stellar iron lines. Here we ignore any possibility that our sample with secure redshifts may be biased toward those with strong iron features, although the strengths of the ISM absorption lines could be correlated at some level with the overall gas-phase metallicity and thus the ISM absorption strengths (Faisst et al. 2016).

2.3. Stellar Mass Estimation

We derived stellar masses (M_*) for individual galaxies through spectral energy distribution (SED) fitting based on the photometry from the COSMOS2015 catalog, together with the precise spectroscopic redshifts. It should be noted that the

¹⁵ The definition of the quality flags follows Lilly et al. (2007, see Section 4.3). The evaluation of the reliability will be detailed in S. J. Lilly et al. (in preparation).

¹⁶ This threshold corresponds to $\approx 3 \times$ the standard deviation of $(z_{\text{phot}} - z_{\text{spec}}) / (1 + z_{\text{spec}})$ and thus is meant to include all sources whose photometric redshift is broadly consistent with the spectroscopic redshift while excluding catastrophic failures.

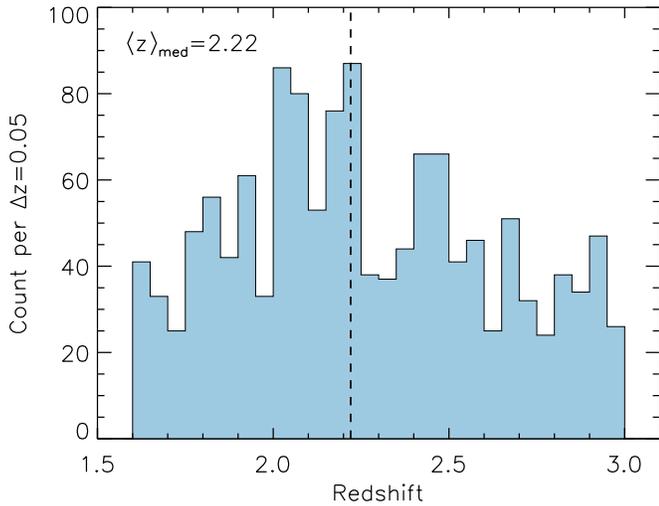


Figure 2. Spectroscopic redshift distribution for the entire sample of 1336 galaxies. Counts per $\Delta z = 0.05$ are shown. The vertical dashed line indicates the median redshift of $\langle z \rangle = 2.22$.

stellar mass here denotes the mass of living stars at the time of observation, rather than the integral of the star formation rate (SFR). The SED fitting outlined in this section is done to the broadband photometry and is meant to derive stellar masses. We will independently perform another fitting to the composite FUV spectra using high-resolution model spectra to estimate the stellar metallicities (see Section 3.5).

Our SED fitting procedure uses the photometric fluxes measured within 32 broad-, intermediate-, and narrowband filters from GALEX near-UV to Spitzer/IRAC ch4, as listed in Table 3 of Laigle et al. (2016). For CFHT, Subaru, and UltraVISTA photometry, we used the fluxes measured in a $3''$ -diameter aperture and applied offsets provided in the catalog to convert them to the total fluxes. All the photometric bands whose rest-frame central wavelengths are within $1960 \text{ \AA} \leq \lambda_{\text{cen}}/(1+z) \leq 2440 \text{ \AA}$ were excluded in order to ensure that the SED fitting is not affected by the possible 2175 Å bump feature in the SED of the galaxies (see Kashino et al. 2021).

The stellar masses and SFRs are estimated using a Python code CIGALE (Burgarella et al. 2005; Noll et al. 2009; Boquien et al. 2019) for the SED fitting and adopted the stellar population synthesis models of Bruzual & Charlot (2003) with a Chabrier (2003) IMF. We considered delayed SFHs with an additional recent burst in order to model the long-term star formation that has formed the bulk of the stellar mass and the latest episode of star formation (e.g., Ciesla et al. 2016, 2017; Pearson et al. 2017),

$$\text{SFR}(t) = \text{SFR}_{\text{delayed}}(t) + \text{SFR}_{\text{burst}}(t), \quad (1)$$

where $\text{SFR}_{\text{delayed}}(t) \propto te^{-t/\tau_0}$ and $\text{SFR}_{\text{burst}}(t) \propto e^{-(t-t_1)/\tau_1}$ if $t > t_1$ and $\text{SFR}_{\text{burst}}(t) = 0$ otherwise. The parameter t denotes the elapsed time since the onset of star formation, t_1 the galaxy age when the late episode of star formation onsets, and τ_0 and τ_1 the e -folding times of the main stellar population and the late starburst population.

In CIGALE, we also accounted for the effect on the photometry of nebular emission lines assuming a common ionization parameter $\log U = -2.8$ and dust emission based on the templates from Dale et al. (2014). The dust attenuation is

Table 1
Input Parameters of the SED Fitting with CIGALE

Parameter	Values
Delayed+Burst SFH	
Age (main population) (Myr)	1000–4500 in steps of 250
e -folding time of the delayed SFH, τ_0 (Myr)	1000, 2000, 3000
Age (starburst population) (Myr)	50, 100, 150, 200
e -folding time of the late starburst, τ_1 (Myr)	10,000
Mass fraction of the late burst population	0.001, 0.003, 0.010, 0.020, 0.040, 0.100, 0.200, 0.400
Stellar Population: Bruzual & Charlot (2003)	
Initial mass function	Chabrier (2003)
Metallicity, Z	0.008
Separation age between young and old populations (Myr)	10
Dust Attenuation: Charlot & Fall (2000)	
A_V in the diffuse ISM	0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.5, 3.0, 3.5
δ_{ISM} (power-law slope for the ISM)	−0.7
δ_{BC} (power-law slope for birth clouds)	−1.3
$A_V^{\text{ISM}}/(A_V^{\text{BC}} + A_V^{\text{ISM}})$	0.44

accounted for using the prescription of Charlot & Fall (2000). The full list of the input parameters is presented in Table 1.

It is known that SED fitting often leads to unrealistically young ages (a few $\times 100$ Myr), and thus spuriously low stellar masses, for high- z star-forming galaxies when there is no limitation on the age. This is because the SED of such galaxies is dominated by the youngest stellar populations (e.g., Maraston et al. 2010). Moreover, it is not plausible that $\sim 10^{10} M_{\odot}$ galaxies at $z \sim 2$ – 3 have started to form stars just before we happen to observe them, given the fact that we can find star-forming galaxies continuously at higher redshifts. It has been shown that limiting the minimum age to a Gyr or more better recovers the stellar masses of high- z star-forming galaxies (Maraston et al. 2010; Pforr et al. 2012). To avoid artificially inferring unrealistically young ages, we thus limit the age of the main stellar population to be ≥ 1 Gyr. This corresponds to a formation redshift $z \gtrsim 5$ for galaxies at $z \sim 3$ and should therefore not be far from reality.

The estimated stellar masses span the range $9.23 \leq \log M_*/M_{\odot} \leq 10.85$ (the 2.5th–97.5th percentiles) with a median value of $\langle \log(M_*/M_{\odot}) \rangle_{\text{med}} = 10.0$. The reduced χ^2 values range mostly between 0.4 and 3.3 with the median value of 0.95. Figure 3 shows the distribution of the inferred M_* and instantaneous SFR. Note that there is a (spurious) sharp upper boundary of the SFR distribution in Figure 3. This is produced by the limited range of SFR probed by the adopted SFHs. This should be of no consequence, as we do not use the SFR values of the individual galaxies in the following analyses.

For reference, the main-sequence M_* –SFR relations at similar redshifts are taken from Whitaker et al. (2014, $z = 2.0$ – 2.5). We also plot the local main sequence adapted from Renzini & Peng (2015). Overall, our sample is in good agreement with the main sequence at the earlier epoch, indicating that it should be a representative sample at these

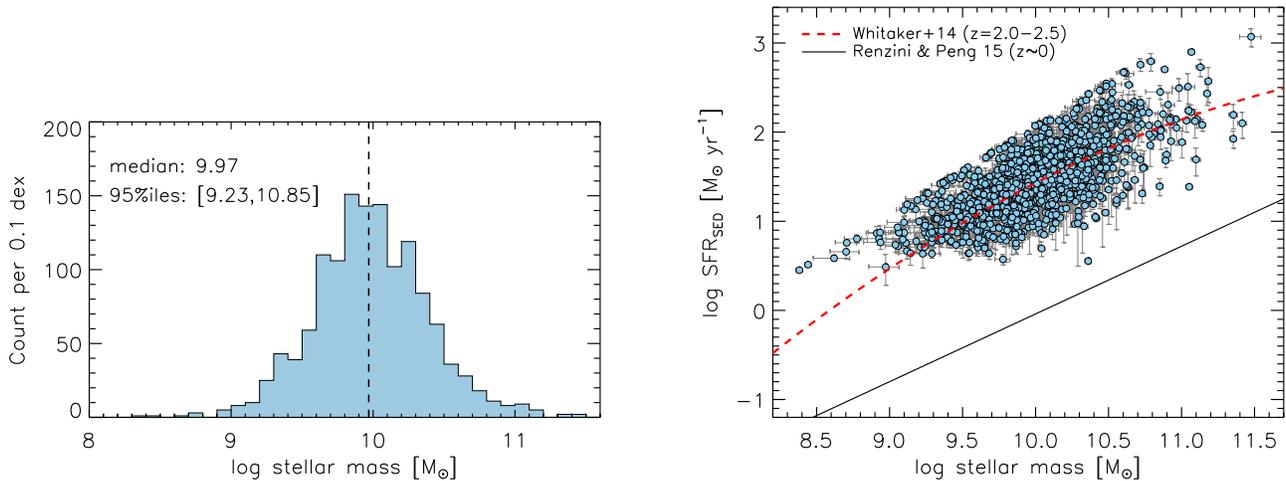


Figure 3. Results of the SED fitting. Left panel: distribution of estimated stellar masses. Right panel: stellar mass vs. instantaneous SFR for the entire sample of 1336 galaxies. For reference, the main-sequence relations are shown, taken from Whitaker et al. (2014, $z = 2.0\text{--}2.5$; red dashed line) and Renzini & Peng (2015, $z \approx 0.1$; solid line).

redshifts. We note that the magnitude limit, $B_{AB} < 25.25$, for the parent sample emerges as a cut in SFR at $\text{SFR} \sim 5 M_{\odot} \text{yr}^{-1}$ in Figure 3 (right panel). At $\log M_{*}/M_{\odot} \lesssim 9.6$, we are missing galaxies with lower SFRs owing to their low observed flux, and thus the results may be biased at the low-mass end.

3. Stellar Metallicity Measurement

In this section, we describe the method to measure the stellar metallicities. We start with the spectrophotometric calibration of each single spectrum and then improve the accuracy of the spectroscopic redshifts of the sources by fitting a common spectral template to each single spectrum.

3.1. Flux Calibration of the VIMOS Spectra

We adopt a spectrophotometric correction to every single spectrum of our sample following the method described in Kashino et al. (2021). The spectra that are produced through the standard zCOSMOS-deep reduction pipeline were flux-calibrated based on standard-star observations. The nominal flux calibration, however, cannot perfectly correct for the effects of finite slit width, imperfect slit centering, and the effects of atmospheric dispersion.

We correct each spectrum with a smooth function of wavelength (see Equation (8) of Kashino et al. 2021) that is constructed using the differences between the actual photometric fluxes in four broadband and eight intermediate-band filters (shown in Figure 4) and the fluxes computed from the pipeline-processed spectrum in these same 12 filter bands. In doing so, we excluded any photometric bands that sampled the rest-frame wavelength of the $\text{Ly}\alpha$ emission line. This is because the flux of the strong $\text{Ly}\alpha$ line may be differently affected than the continuum flux owing to the possible extended shape of the emission and/or the overestimate of the sky level at these particular wavelengths.

Examples of the flux-calibrated spectra are shown in Figure 1, demonstrating that the spectra are in agreement with the photometric fluxes. The left panel of Figure 4 shows how this spectrophotometric calibration works in the observed frame by comparing the spectral broad- and intermediate-band fluxes (F_i^{spec} , where i denotes a filter) recomputed in the flux-corrected spectra with the photometric fluxes (F_i^{phot}) from the

COSMOS2015 catalog. The median values of the corrected F_i^{spec} -to- F_i^{phot} ratios in each band are all within ± 0.015 dex, and the scatter is $\sim 0.02\text{--}0.05$ dex, depending on the filters. This overall scatter (seen in the rest frame) is quite smaller than before correction ($\sim 0.1\text{--}0.15$ dex). Note that there is still scatter for a given filter because the correction used a smooth function of wavelength, so the effects of photometric noise in the different filters are still seen.

The right panel of Figure 4 shows the same data but now shifted to the rest-frame wavelengths. There is no systematic trend in the corrected flux ratios across the entire wavelength range for the stellar metallicity measurement (1274–2053 Å). The running medians with a window size of 100 Å are all within 0.005 dex over this wavelength range of interest. On the other hand, a significant systematic undercorrection is seen around 1216 Å most likely because of sky-subtraction issues associated with strong $\text{Ly}\alpha$ lines mentioned above.

This flux calibration is important to obtain composite spectra that correctly reflect the average shape of the galaxies’ SEDs. We note, however, that the precision of this calibration is unlikely to be critical for our conclusions, because the stellar metallicities are measured based on the detailed shape of the spectra that results from the blending of numerous narrow stellar absorption lines, whereas the overall shape of the smooth continuum is fit with an arbitrary multiplicative λ -dependent function (see Section 3.5). In the remainder of the paper, the term “observed VIMOS spectra” will always refer to these accurately spectrophotometrically recalibrated spectra.

3.2. Fine Adjustment of Spectroscopic Redshifts

Precise determination of the spectroscopic redshifts is important to reduce the loss of the potential spectral resolution when spectra are stacked. The spectroscopic redshifts in the zCOSMOS-deep catalog have been determined by visually inspecting prominent emission and absorption features in each spectrum, but no systematic spectral fitting has been performed. As a consequence, the spectral redshifts may be more uncertain than the best estimates that can potentially be achieved from the existing spectra; thus, some improvements are possible.

We therefore made small adjustments to the spectroscopic redshifts by fitting a common template spectrum to each of the

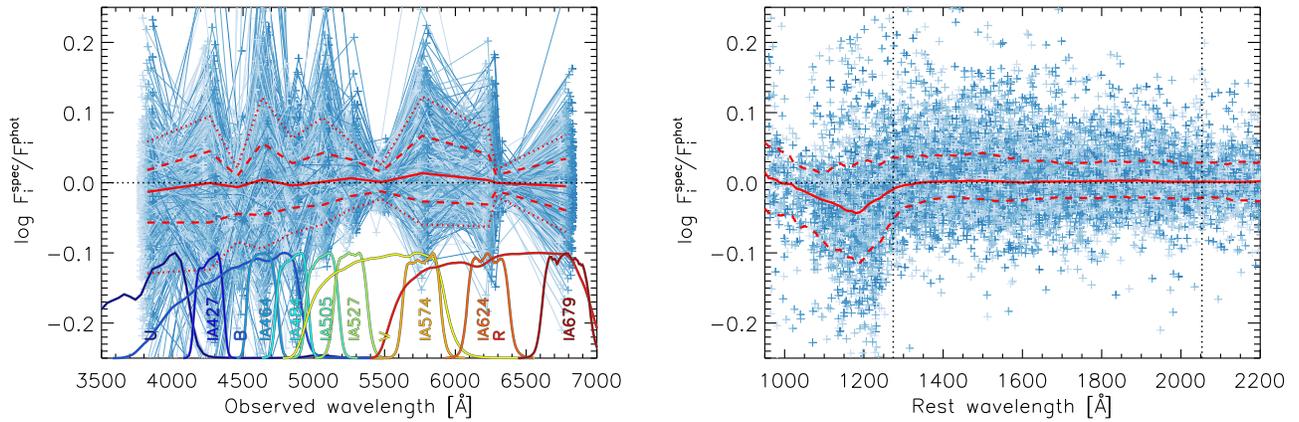


Figure 4. Left panel: flux ratios, $F_i^{\text{spec}}/F_i^{\text{phot}}$, after correction for the sample of 1336 galaxies. Symbols for each single spectrum are connected by a line. Different spectra are colored differently for display purposes. The red solid, dashed, and dotted lines indicate, respectively, the median values, the 16th–84th percentiles, and the 5th–95th percentiles. The horizontal dotted line indicates $F_i^{\text{spec}}/F_i^{\text{phot}} = 1$. The transmission curves of the relevant photometric bands are shown at the bottom. Right panel: the same flux ratios, but as a function of rest-frame wavelength. The red solid and dashed lines indicate running medians and 16th–84th percentiles with a window size of 100 Å. The vertical dotted lines denote the lower and upper limits of the wavelength range used for the stellar metallicity measurement.

individual spectra. We constructed this template by stacking the observed spectra of the entire sample of 1336 galaxies used in this work by using the original spectroscopic redshifts from the catalog. In this template fitting, we applied an arbitrary normalization and a wavelength-dependent multiplicative factor that is intended to mimic the Calzetti et al. (2000) dust attenuation.

The differences between the revised and original spectroscopic redshifts, $(z_{\text{orig}} - z_{\text{new}})/(1 + z_{\text{new}})$, have a Gaussian-like distribution with a standard deviation of 8.4×10^{-4} (252 km s^{-1}) and median of 6.4×10^{-6} . We found that, by using the revised spectroscopic redshifts, the resulting stacked spectra are noticeably improved, showing sharper spectral features in both emission and absorption than those seen in the stacks based on the original redshifts. However, the adjustment of the spectroscopic redshifts has little effect on the metallicity measurements, and our conclusions do not change if the stacking is done using the original redshift values.

3.3. Stacking Procedure

To infer the stellar metallicity as a function of stellar mass, we rely on stacked spectra from subsamples of galaxies defined by their stellar mass. The observed spectra are co-added as follows. We first transform all the individual spectra to the rest-frame wavelength based on their adjusted redshift (see Section 3.2) and rebin them to a common wavelength grid with a spacing of 1 Å. This is slightly oversampled relative to the pixel resolution of the original spectra ($5.35 \text{ Å pixel}^{-1}$ in the observed frame) and equivalent to the wavelength grid spacing of the model spectra used for spectral fitting (see Section 3.4).

Each spectrum is then normalized by dividing by a fitted continuum of the form $\lambda^\beta \times 10^{-0.4A_\lambda}$, where A_λ is the dust attenuation of the Charlot & Fall (2000) prescription. Here the overall dust attenuation is a free parameter for continuum fitting, independent of the result from the SED fitting in Section 2.3. By doing so, the effects of the variable overall shapes of the spectra are mitigated. We then take the mean value of the individual continuum-divided spectra at each wavelength grid while ignoring any spectral regions that are missing and/or contaminated, for example, by zeroth-order contamination or strong sky lines. Finally, to recover the global shape of the average SED, the stacked spectrum is multiplied

by the mean of the fitted continua. We note that normalizing the individual spectra to their continuum level at $\lambda_{\text{rest}} \sim 1550 \text{ Å}$ produces very similar results. Stacking without any normalization, however, yields more noisy composite spectra and more residuals with respect to the fitted models.

Figure 5 shows the composite spectrum (middle panel) of the entire sample.¹⁷ The number of spectra used at each wavelength is indicated in the top panel. Note that, given the redshift range of our sample, the rest-frame wavelength range of 1400–1700 Å is covered by nearly all the input spectra, while shorter and longer wavelengths are less well sampled. Some prominent spectral features are clearly identified as marked by vertical lines.

3.4. Stellar Population Synthesis Models for Metallicity Estimation

To derive stellar metallicities for the galaxies, we compare our observed composite spectra with model spectra, following the approaches described in Steidel et al. (2016) and Cullen et al. (2019).

We utilize the latest public data release of the population synthesis code “Binary Population and Spectral Synthesis” (BPASSv2.2.1; Eldridge et al. 2017; Stanway & Eldridge 2018). The package provides sets of single stellar population synthesis model spectra at a pixel resolution 1 Å pixel^{-1} as a function of stellar ages for different IMFs and for discrete (overall) stellar metallicities ($Z_* = 10^{-5}, 10^{-4}, 0.001, 0.002, 0.003, 0.004, 0.006, 0.008, 0.010, 0.014, 0.020, 0.030,$ and 0.040). It is important to note that although the BPASS models adopt a fixed abundance ratio based on the solar abundances, the model fit to the rest-frame FUV spectrum is mostly sensitive to the iron abundance of the young stellar component of the galaxies. This is because stellar photospheric absorption features in the wavelength range of interest are primarily due to transitions of highly ionized iron (Dean & Bruhweiler 1985; Brandt et al. 1998). We can therefore translate the inferred metallicity Z_* into $[\text{Fe}/\text{H}]$.

¹⁷ The composite spectrum and those in mass bins are available in electronic form from https://github.com/kashinod/zCOSMOS-deep/tree/main/Kashino_et_al_2021b.

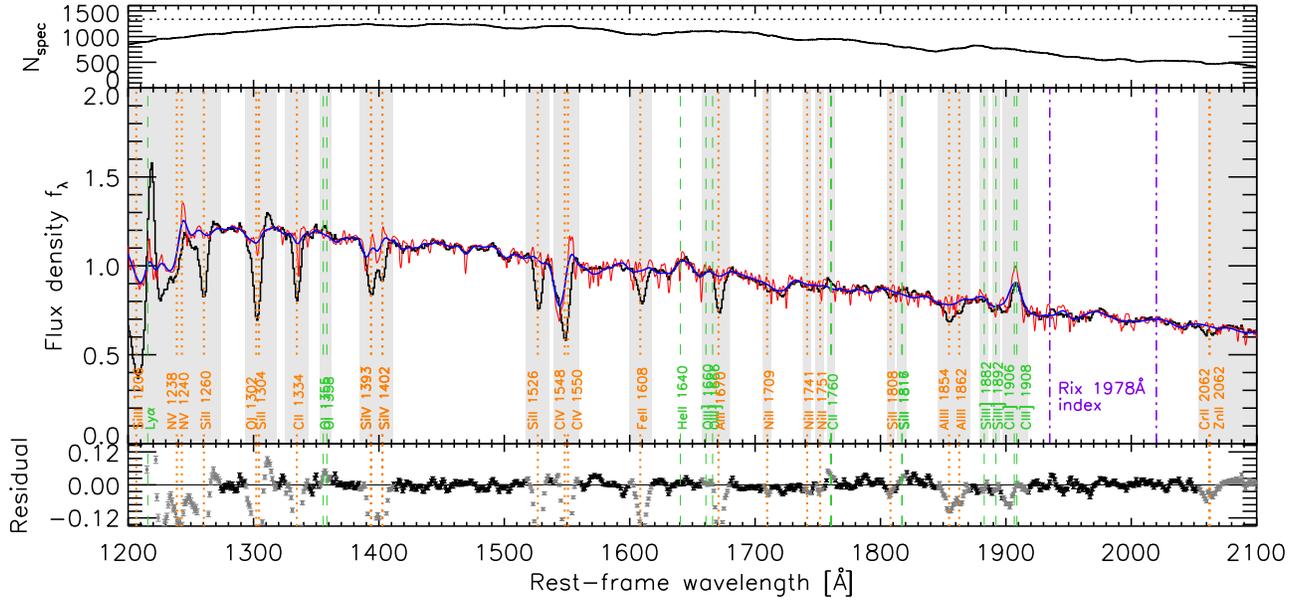


Figure 5. Composite VIMOS spectrum of the entire sample of 1336 galaxies at $1.6 \leq z \leq 3.0$. Top panel: the number of spectra that have been stacked at each wavelength grid. The horizontal dotted line indicates the number of galaxies in the stack. Middle panel: the stacked spectrum (black line), in which some prominent absorption and emission features are identified as marked by color-coded labels (interstellar absorption features—orange; nebular emission lines—green). The blue line indicates the best-fit BPASS model obtained from the MCMC analysis, and the red line indicates the best-fit model before smoothing. The gray regions indicate the wavelength ranges that are masked out in the fitting (see Table 2). The purple dotted–dashed lines mark the wavelength region that is used for the Rix et al. (2004) 1978 Å index. Bottom panel: residuals from the best fit. The vertical error bars correspond to the associated 1σ errors. The data points that are not included in the fit are shown in gray.

We adopted the BPASSv2.2.1 model that used a Chabrier (2003) IMF with a high-mass cutoff of $100 M_{\odot}$ and included binary star evolution, and we considered a continuous SFH with a duration of 100 Myr to construct a set of rest-frame UV template spectra. Note that the FUV spectrum is almost independent of the past SFH at $\gtrsim 100$ Myr previously, as it is dominated by massive young stars. Indeed, we checked that the results hardly change for different duration times between 10 and 1000 Myr. Figure 6 shows examples of the model spectra at different metallicities, demonstrating that the detailed shapes of the galaxy spectra across the rest FUV window are sensitive to the stellar metallicity. A feature of the BPASS binary models is that they model the broad He $\lambda 1640$ emission line that originates in the winds from very massive stars (see Section 3.5 for a relevant description).

We added nebular continuum and line emission to the templates, although it has only a minor contribution of $\sim 5\%$ in the total continuum flux. The nebular continuum was computed using CLOUDY v17.00 (Ferland et al. 2017) adopting the BPASS spectrum itself as the incident spectrum. We assumed the electron density to be $n_e = 300 \text{ cm}^{-3}$ and the ionization parameter to be $\log U = -2.8$. These values are consistent with the recent estimates in $z \sim 2-3$ star-forming galaxies (Sanders et al. 2016; Kashino et al. 2017; Strom et al. 2017, 2018) and local star-forming galaxies with specific SFR (sSFR) as high as our sample ($\sim 5 \text{ Gyr}^{-1}$; Kashino & Inoue 2019). We change the gas-phase metallicity to calculate nebular emission according to the stellar metallicity, with an offset of $\log(Z_{\text{gas}}/Z_{*}) = 0.42$ as reported for $z \sim 2-3$ galaxies by Strom et al. (2018). In other words, we considered an enhanced gas-phase metallicity, $12 + \log(\text{O}/\text{H}) = 8.69 + \log(Z_{*}/Z_{\odot}) + 0.42$, relative to each Z_{*} of the BPASS template. We will indeed measure consistent $\log(Z_{\text{gas}}/Z_{*})$ or $[\text{O}/\text{Fe}]$ values in Section 4.3 and discuss the evolution of $[\text{O}/\text{Fe}]$ in Section 5.

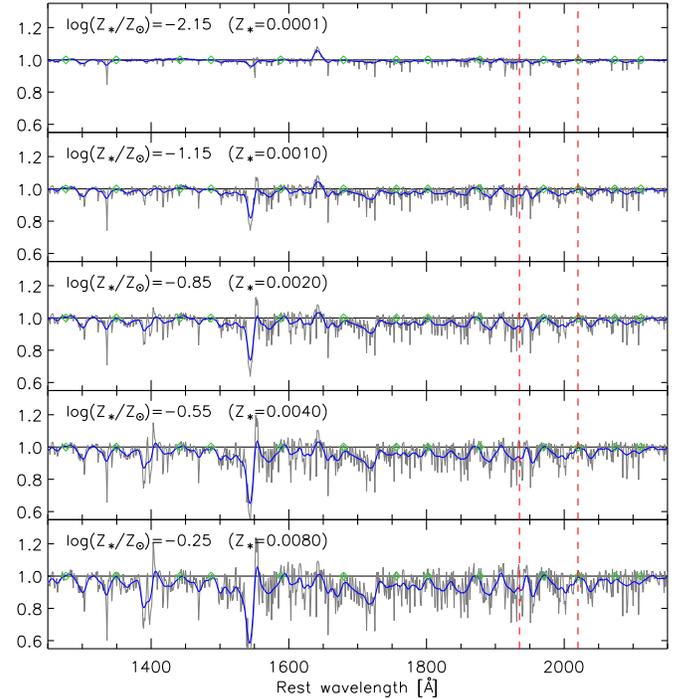


Figure 6. The BPASSv2.2.1 stellar population models. All models assume a constant SFR over 100 Myr. Each panel shows the model spectrum in the rest FUV window at $\log(Z_{*}/Z_{\odot}) = -2.15$ to -0.25 from the top to the bottom. The gray lines show the full resolution model spectra, which are here normalized by the pseudocontinuum constructed from a spline fit to the selected points (green diamonds; see Table 3 of Rix et al. 2004). The blue lines indicate the models smoothed to the VIMOS resolution ($\sigma = 3.3 \text{ \AA}$, or 2500 km s^{-1} in FWHM; see Section 4). It is clearly seen, even at the VIMOS resolution, that the absorption features are stronger with increasing stellar metallicity. The vertical dashed lines indicate the range of the traditional 1978 Å index (Rix et al. 2004) for reference.

Table 2
Rest-frame Wavelength Ranges Excluded from Fitting^a

λ_{\min} (Å)	λ_{\max} (Å)	Interstellar ^b Spectral Features
1294	1318	O I λ 1302.17, Si II λ 1304.37, Si II λ 1309.28
1326	1343	C II λ 1334.53
1353	1361	O I λ 1355.60, O I λ 1358.51
1385	1411	Si IV λ 1393.76, 1402.77
1518	1535	Si II λ 1526.71
1540	1559	C IV λ 1548.19, 1550.77
1600	1617	Fe II λ 1608.45
1658	1679	O III] λ 1660.81, 1666.15, Al II λ 1670.79
1707	1712	Ni II λ 1709.60
1739	1744	Ni II λ 1741.55
1749	1754	Ni II λ 1751.91
1758	1763	C II λ 1760.5 (blended)
1806	1811	Si II λ 1808.01
1814	1820	Si II λ 1816.93, 1817.45
1846	1871	Al III λ 1854.72, 1862.79
1880	1885	Si III] λ 1882.71
1890	1895	Si III] λ 1892.03
1898	1917	C III] λ 1906.68, 1908.73

Notes.

- ^a The wavelength range used for the fit is limited to $\lambda_{\text{rest}} = 1274\text{--}2053 \text{ \AA}$.
^b Wavelengths are all given in vacuum.

3.5. Fitting Synthesis Model Spectra

Our goal is to estimate average stellar metallicities for binned subsamples of galaxies by using their composite spectra. To do so, we limited the wavelength range between $1274 \text{ \AA} \leq \lambda_{\text{rest}} \leq 2053 \text{ \AA}$ and excluded some narrow wavelength regions where the stellar continuum is impacted by interstellar absorption lines or nebular emission lines. The wavelength regions excluded from the fitting are summarized in Table 2. We note that the region of the He λ 1640 line is not excluded because the BPASS models include the broad He λ 1640 stellar emission line that originates from very hot stars. In fact, as shown in Section 4, the observed He λ 1640 feature in the composite spectra could be attributed almost entirely to the contribution from stars that is predicted in the BPASS models.

An added advantage of the full spectral fitting over using the traditional indices (e.g., the 1978 Å index) is that the former does not need to identify the pseudosmooth continuum. Contrarily, the latter requires it for measuring equivalent widths (EWs) of particular stellar absorption lines. The measurements of the EWs of faint features are quite sensitive to the assumed continuum level, while the determination of the pseudocontinuum has to be based on very limited wavelength regions free from narrow features, thus being easily affected by noise. The full spectral fitting is free from these difficulties.

We fit the constructed BPASS models to our composite spectra based on the maximum likelihood estimation. We assume that the associated errors are Gaussian and independent, so the logarithm of the likelihood \mathcal{L} is given by

$$\ln \mathcal{L} = -\frac{1}{2} \sum_i \left[\frac{(f_i^{\text{obs}} - f_i^{\text{model}})^2}{\sigma_i^2} + \ln(2\pi\sigma_i^2) \right], \quad (2)$$

where f_i^{obs} is the observed composite spectrum at the i th wavelength grid, f_i^{model} is the model spectrum for a given set of

parameters, and σ_i is the error on the observed flux. The summation is over all wavelength grids that are included in the fit.

For our models, we adopted eight free parameters: the logarithm of the overall stellar metallicity ($\log Z_*$), the Gaussian smoothing kernel σ_{smooth} applied to the model spectra, and six parameters for a fifth-order polynomial function that is multiplied by the template spectrum:

$$f_i^{\text{model}}(Z_*) = f_i^{\text{BPASS}}(Z_*) \times \sum_{k=0}^5 p_k (\lambda[\text{\AA}] - 1650)^k, \quad (3)$$

where $f_i^{\text{BPASS}}(Z_*)$ is the template spectrum at given Z_* in arbitrary units, in which the nebular emission is accounted for. The polynomial term is ideally intended to reflect the change of the overall shape of the spectra due to dust attenuation.¹⁸

To sample the posterior probability distribution of the model parameters, we employed a Markov Chain Monte Carlo (MCMC) technique using the `emcee` package for Python (Foreman-Mackey et al. 2013). In our fitting, we adopted a uniform prior probability function for each parameter. For stellar metallicity, a flat prior is adopted in log space (i.e., $\log(Z_*)$) between $-5 \leq \log(Z_*) \leq \log(0.040)$ (which corresponds to $-3.15 \leq \log(Z_*/Z_\odot) \leq 0.44$) according to the possible choices in the BPASS models. Since the models are only provided at 13 discrete metallicity values (see Section 3.4), we interpolated the flux values in $\log Z_*\text{--}\log(\text{flux})$ space between the models. This enables us to generate a model at any metallicity value within the available range.

In the fitting, we need to match the spectral resolution of the model templates to that of the observed composite spectra. However, the spectral resolution of the composite spectra is not precisely known. It may vary from galaxy to galaxy within a subsample because of different velocity dispersions of the galaxies and different slit illumination profiles. We thus leave the smoothing scale σ_{smooth} , in units of Å,¹⁹ as a free parameter rather than using a predetermined value. In the fitting, the model spectra are smoothed by a Gaussian kernel with σ_{smooth} , and the smoothing scale that gives the best fit to the data is constrained along with other parameters.

3.6. Accuracy of the Metallicity Measurements

Before applying our procedure to the data, we evaluate the potential systematic uncertainties in the metallicity measurements that may come from the noise and the limited spectral resolution of the spectra.

As we will show below, the width of the smoothing kernel, σ_{smooth} , is estimated from the fits to be $\sim 3.3 \text{ \AA}$. This is large enough to wash out individual narrow stellar absorption features in the model spectra. Still, the model spectra with different metallicities will retain their different characteristic shapes that result from blending of the narrow features after smoothing. Thus, the metallicity measurement is indeed

¹⁸ We tried an alternative, more physically motivated dust model assuming an attenuation curve and obtained constraints on Z_* and σ_{smooth} that are equivalent to the fiducial results. However, the degeneracy between the parameters, e.g., overall scaling, overall dust attenuation, and the slope of an attenuation curve, causes poorer convergence in the MCMC chains, and therefore we adopted the fifth-order polynomial for our fiducial analysis.

¹⁹ Note that the wavelength sampling is 1 \AA pixel^{-1} .

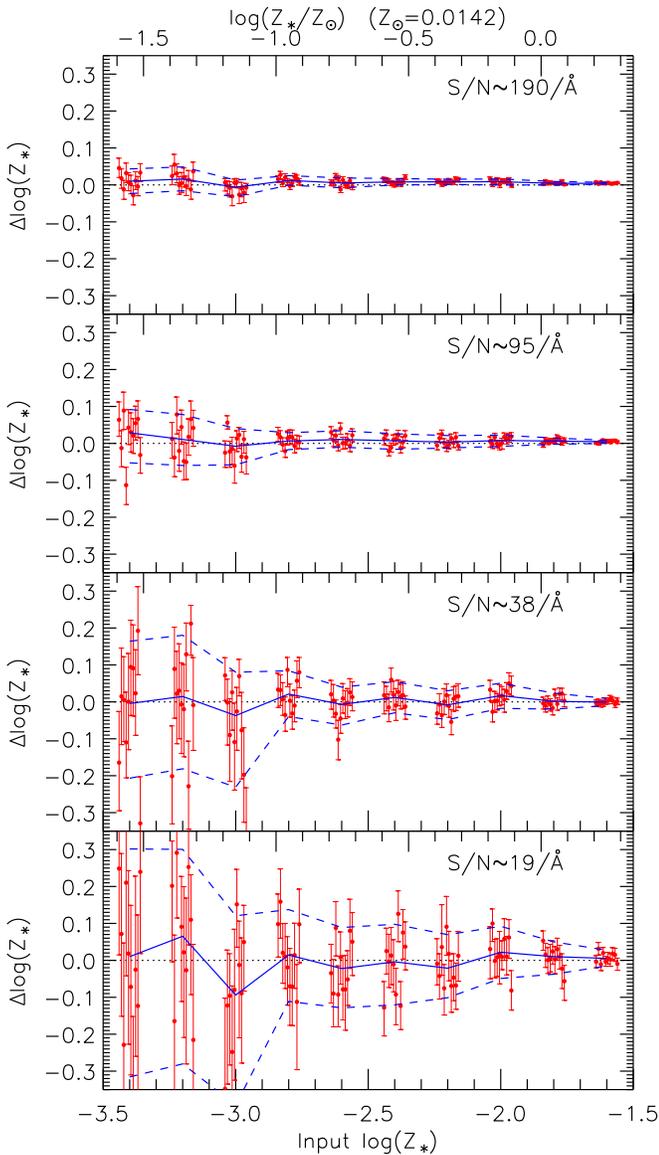


Figure 7. The results of testing how accurately the input metallicities can be recovered with realistic noise and the spectral resolution of the actual data. The difference between the derived metallicity and the input metallicity is shown as a function of the input value. The four panels show the results for different levels of random noise, from top to bottom, the same level as the stack of the entire sample ($S/N \approx 190 \text{ \AA}^{-1}$), and twice, five times, and 10 times higher than the top one ($S/N \approx 95$, 38 , and 19 \AA^{-1} , respectively). Red symbols indicate individual results of the MCMC analysis at different input metallicities, each for 1 of 10 realizations of the random noise. The results are shown with small offsets along the x -axis for display purposes. Blue solid and dashed lines indicate, respectively, the median and the 16th–84th percentiles of the posterior probability distribution compiling all 10 trials.

possible. The possible systematic uncertainties, however, should be evaluated.

For this purpose, we constructed a set of artificial spectra using the BPASS models at different metallicities. We smoothed them with $\sigma_{\text{smooth}} = 3.3 \text{ \AA}$ and modified the overall shape using the representative polynomial function as in Equation (3). We then added Gaussian noise with the moderate wavelength dependence that replicates the 1σ error spectra. We considered noise with four different levels: the best had the noise seen in the composite spectrum of the entire sample, i.e., the S/N at $\lambda_{\text{rest}} \sim 1450 \text{ \AA}$, $S/N_{1450} \approx 190$ per 1 \AA . The other

three are degraded to have twice ($S/N_{1450} \approx 95$), five times ($S/N_{1450} \approx 38$), or 10 times ($S/N_{1450} \approx 19$) higher noise than the first one. The stacked spectra of the subsamples selected by stellar mass have noise levels between these values, depending on the number of galaxies in the bin and their average brightness.

We then attempted to measure the stellar metallicity in the same way as from the data in order to evaluate how accurately the input metallicities can be recovered. We repeated each setup 10 times with different random noise realizations. Here only a single fiducial duration of star formation (100 Myr) is considered.

Figure 7 shows the results of this exercise, separately, for four different noise levels. The difference between the inferred and input metallicities is shown as a function of the input value. The error bars indicate the 16th–84th percentiles of the inferred posterior for individual measurements. The results from the 10 trials with different random noises are shown separately with slight offsets in the x -axis. We also compile the posterior probability distributions of $\log Z_*$ from all 10 trials and derive the median and the 16th–84th percentiles that are shown by blue solid and dashed lines.

Generally, the uncertainties reduce with increasing metallicity because the stellar absorption features in the spectrum become more prominent and sensitive to the metallicity value at higher metallicity. It can be seen that, with all these noise levels, the systematic biases in the metallicity measurements are negligible with no trend across the metallicity range of interest.

4. Results

Figure 8 shows the posterior distributions of the parameters for the composite stack of the entire sample of 1336 galaxies.²⁰ The stellar metallicity, Z_* , and smoothing scale, σ_{smooth} , are constrained with a single preferred solution ($\log(Z_*/Z_\odot) = -0.812 \pm 0.008$ and $\sigma_{\text{smooth}} = 3.30 \pm 0.07 \text{ \AA}$). Here the uncertainties denote the statistical 16th–84th confidence intervals in the posterior probability distribution functions inferred from MCMC fitting and do not therefore include any systematic uncertainties ascribed to the methodology, e.g., the choice of the stellar population synthesis model used in the fits. The constraints on the coefficients p_k in Equation (3) also show single peaks in the posterior distributions. This also holds when the sample is divided into stellar mass bins.

Figure 5 shows the best-fit model for the stack of the entire sample. The detailed shape of the FUV continuum is well reproduced after smoothing the original BPASS model spectrum with $\sigma_{\text{smooth}} = 3.3 \text{ \AA}$. The recovered smoothing width σ_{smooth} is marginally less than the nominal resolution of the VIMOS spectrograph with the LR blue grism and $1''.0$ arcsec slits of $\approx 3.4 \text{ \AA}$, possibly due to nonuniform illumination of the slit.²¹

For comparison to a standardized metallicity indicator, we also measured the “1978 Å index” of Rix et al. (2004), i.e., the EW across 1935–2020 Å, to be $2.62 \pm 0.08 \text{ \AA}$. The calibration of Rix et al. (2004, see Equation (8)) converts this into

²⁰ This figure was created using the Python module `corner.py` (Foreman-Mackey 2016).

²¹ The resolution $R = 200$ in FWHM at the center of the spectral window ($\sim 5150 \text{ \AA}$) corresponds to $\sigma \approx 3.4 \text{ \AA}$ in the rest frame for the median redshift of 2.2.

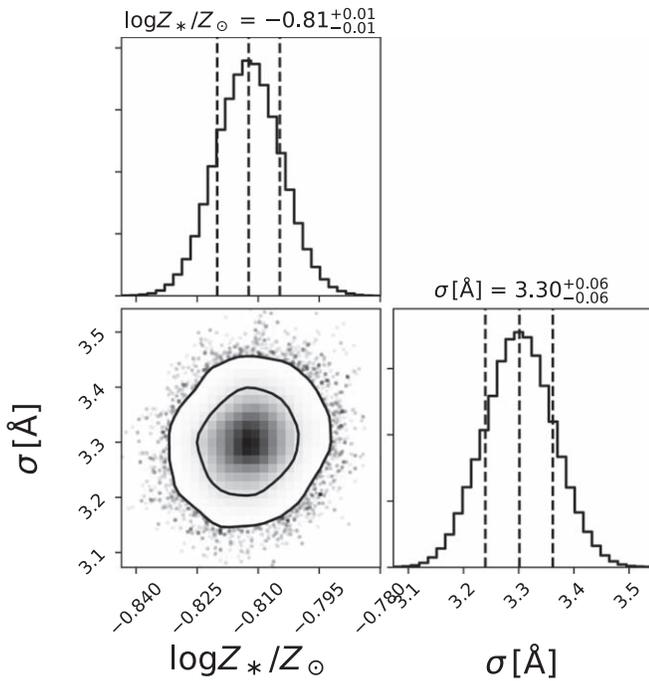


Figure 8. Posterior probability distribution functions for $\log(Z_*/Z_\odot)$ and smoothing scale σ_{smooth} . The contours on the 2D plots correspond to enclosing 68% and 95% of the posterior probability. The vertical lines in the 1D plots indicate the 16th, 50th (median), and 84th percentiles. Here we omit the posteriors of the coefficients p_k in Equation (3) that were simultaneously fitted.

$\log(Z_*/Z_\odot) = -0.93 \pm 0.05$, which is consistent within ~ 0.1 dex with the fiducial result.²² Note that the statistical uncertainty here is about six times larger than that of the fiducial result because the spectral regions used for the 1978 Å index are limited.

In the following, we present the measurement of the stellar MZR based on the stacks in stellar mass bins. We here recall that the metallicity estimated from the photospheric absorption in the FUV spectra reflects the iron abundance (e.g., Steidel et al. 2016) in short-lived, recently formed stars and thus can be regarded as being almost equivalent to the gas-phase value.

4.1. Stellar Mass versus Stellar Metallicity

To investigate the relation between stellar mass and stellar metallicity, we divided the sample of 1336 star-forming galaxies into bins of stellar mass in two different binning schemes: first we equally split the sample into six mass bins, and then we divided the sample into bins with a constant width of $\Delta \log M_* = 0.2$ dex. After stacking, we performed the MCMC analysis to fit the BPASS template to the stacked spectra as described in Section 3.5.

Figure 9 shows the composite spectra and the corresponding best-fit models in the six equally populated subsamples. It is noticeable that the overall slope of the spectra becomes shallower for higher M_* owing to the increasing average dust attenuation. The overall shape and detailed features are both well reproduced across the entire wavelength range of interest by the model spectra.

The results in different bins are summarized in Table 3. The stellar metallicities are measured to range over $-1.3 \lesssim$

$\log(Z_*/Z_\odot) \lesssim -0.4$, increasing with M_* . Note that the BPASS models cover this observed range of metallicity with a sufficiently large margin at either end, and thus our MCMC analysis should not be affected by the artificial limit of the explored metallicity range. We obtained consistent results from the fixed-width binning scheme.

Figure 10 shows the relationship between stellar mass and stellar metallicity. The x -axis values of the points correspond to the median stellar masses within the bin, and the horizontal error bars indicate the minimum and maximum M_* values in each bin. The y -axis values and the error bars indicate the median and the 16th–84th percentiles of the posterior distribution of $\log(Z_*/Z_\odot)$. There is a tight correlation between these quantities, though the $\log(Z_*/Z_\odot)$ values in the lowest mass bins are relatively insecure. A linear fit to the results in log–log space, obtained from binning the sample into 0.2 dex width bins, yields

$$\log(Z_*/Z_\odot) = -(0.81 \pm 0.01) + (0.32 \pm 0.03) \log[M_*/(10^{10} M_\odot)], \quad (4)$$

where the errors in the parentheses denote the nominal 1σ errors. This is shown as the blue line in Figure 10.

This result is generally in good agreement with that of Cullen et al. (2019), who used stacks of in total 681 star-forming galaxies at $z = 2.5\text{--}5$ ($\langle z \rangle \approx 3.5$). There is a difference at the low-mass end, where the Cullen et al. (2019) data show a tentative sign of flattening at $8.8 \lesssim \log M_*/M_\odot \lesssim 9.5$, which, however, is not seen in our data (it should be noted that their lowest mass point is only an upper limit). Overall, our metallicities are slightly lower at given M_* , which appears to be counter to the expected redshift evolution since the Cullen et al. (2019) sample is generally at somewhat higher redshifts. This apparent discrepancy may be accounted for by a systematic effect from the methodology that we now explain.

The method in Cullen et al. (2019) is very similar to ours and was based on stacks of rest-frame FUV spectra and model templates with various metallicities. The authors, however, used the Starburst99 high-resolution WM-basic stellar population models for their fiducial results. They compared their fiducial estimates to the ones derived using the BPASSv2.1 models and found that the use of the BPASS models leads to systematically lower metallicities by ~ 0.1 dex. Accounting for this systematic offset will then bring the two results into better agreement. Note that the lower-metallicity limit of the Starburst99 model is $\log(Z_*/Z_\odot) = -1.15$, which is not sufficiently low for unbiased fitting for our sample, and thus we adopted the BPASS model in this work.

Our results also appear to be lower by ~ 0.4 dex than the average measurement of $\log(Z_*/Z_\odot) = -0.425 \pm 0.159$ ²³ at $M_* \sim 10^{10} M_\odot$ obtained by Halliday et al. (2008) using the Rix et al. (2004) 1978 Å index. This offset suggests that there possibly remain substantial systematic uncertainties in the stellar metallicity measurements for different samples and methodologies.

It should also be noted that the stellar metallicities may be underestimated because the integrated stellar emission would be biased toward stars formed in the less obscured and thus lower-metallicity environments. Such systematic biases in the stellar metallicity measurements will be explored in future papers.

²² Here we adopted $Z_\odot = 0.0142$, while Rix et al. (2004) assumed $Z_\odot = 0.020$.

²³ This value is converted to $Z_\odot = 0.0142$.

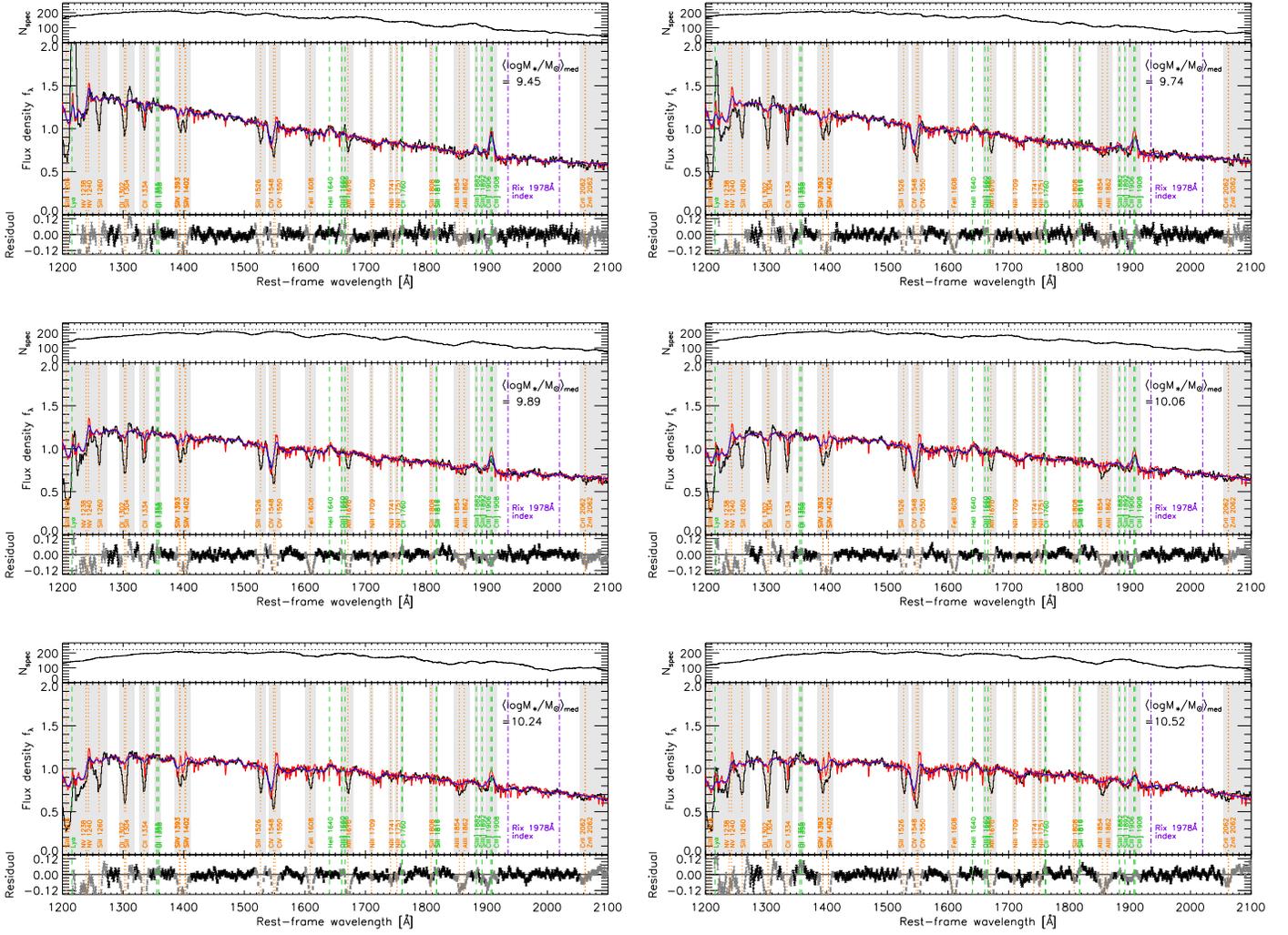


Figure 9. Composite VIMOS spectra and the best-fit BPASS models are shown, separately, for the six subsamples equally separated by stellar mass. The median masses are indicated in each panel. Each panel is in the same format as Figure 5.

4.2. Comparison with the Local Stellar MZR

We now compare our result with other work at low redshifts to see how the stellar MZR evolves through cosmic time. Zahid et al. (2017) established the stellar MZR in the range $10^{8.5} \lesssim M_*/M_\odot \lesssim 10^{11}$ using $\sim 2 \times 10^5$ star-forming galaxies at $z < 0.25$ (median $\langle z \rangle = 0.08$) from SDSS. The authors employed a full spectral fitting approach using the observed composite spectra in the rest-frame optical. Their result is in agreement with the measurements obtained from spectroscopy of individual stars in nearby galaxies compiled by Kudritzki et al. (2016). Furthermore, comparing to the stellar metallicities of dwarf galaxies, measured by Kirby et al. (2013), revealed that the M_* - Z_* correlation holds smoothly down to $M_* \sim 10^4 M_\odot$.

We note that the metallicities from Kirby et al. (2013) are intended to purely reflect the iron abundance (Fe/H). We assume that other local metallicity measurements also reflect the iron abundance. Particularly, this approximation can be justified for the Zahid et al. (2017) measurements that are obtained from comparison between the observed rest optical spectra and the models assuming solar α/Fe . This is because the inferred total metallicity (Z_*) is known to be related to [Fe/H] via $[\text{Fe}/\text{H}] = \log(Z_*/Z_\odot) - A[\alpha/\text{Fe}]$ with $A \approx 0.8$ – 0.9

(Trager et al. 2000). Indeed, we will find $[\alpha/\text{Fe}] \approx [\text{O}/\text{Fe}] \sim 0$ for those galaxies.

Figure 11 compares these low-redshift measurements with our results at high redshift. Our high-redshift stellar MZR is clearly offset below the local relation (at our sampled masses) with $\Delta \log Z \approx 0.8$ dex at a given mass. We should note that the local MZR is not yet corrected for the underestimation that arises from the measurements based on the rest-frame optical integrated light of the galaxies. We will take account of this correction, which is about $+0.1$ dex (see Appendix), in the subsequent sections, where we will compare high- and low-redshift MZRs more precisely.

We employed the empirical parameterization introduced by Curti et al. (2020, Equation (2)) to express the local stellar MZR:

$$\log\left(\frac{Z}{Z_\odot}\right) = Z_0 - \frac{\gamma}{\beta} \log\left(1 + \left(\frac{M_*}{M_0}\right)^{-\beta}\right), \quad (5)$$

where Z_0 is the asymptotic metallicity at the massive end, M_0 is a characteristic mass where the relation begins to flatten, γ is the power-law slope of the relation at $M_* \ll M_0$, and β determines the width of the transition region between the two extremes. Fitting the measurements from Zahid et al. (2017)

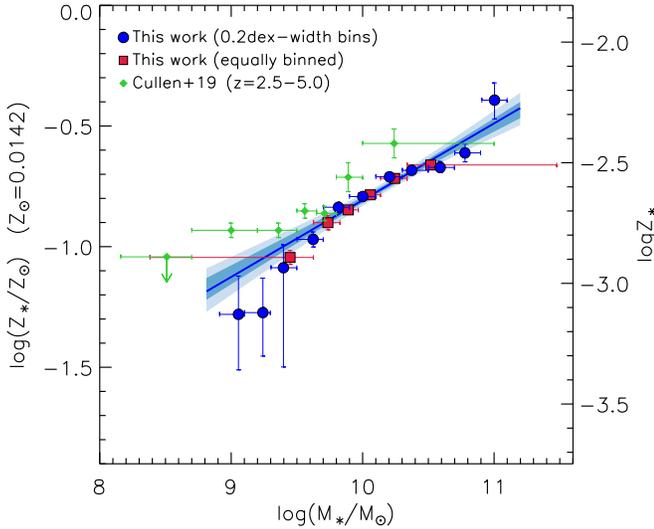


Figure 10. Stellar metallicity as a function of stellar mass, i.e., the stellar MZR, for our sample of star-forming galaxies at $1.6 \leq z \leq 3.0$ (median $\langle z \rangle = 2.22$). Here the stellar metallicity mostly reflects the iron abundance ($[\text{Fe}/\text{H}]$). The red squares and blue circles show the results based on two different binning schemes (see text). The blue solid line indicates the best linear fit to our data (Equation (4)) with the 1σ (dark-blue region) and 2σ (light-blue region) confidence limits. For comparison, the relation for a galaxy sample at $2.5 \leq z \leq 5.0$ ($\langle z \rangle = 3.5$; green diamonds) is taken from Cullen et al. (2019).

Table 3
Stack Statistics and Metallicity Estimates^a

$\log(M_*/M_\odot)$		N	S/N_{1450}^b	$\log(Z_*/Z_\odot)^c$ ($[\text{Fe}/\text{H}]$)
Median	Min–Max			
9.97	8.38–11.48	1336	189	$-0.812^{+0.008}_{-0.008}$
Binned Equally into Six Subsamples in M_*				
9.45	8.38–9.62	222	71	$-1.045^{+0.029}_{-0.029}$
9.74	9.63–9.83	222	72	$-0.901^{+0.029}_{-0.030}$
9.89	9.83–9.97	223	81	$-0.848^{+0.017}_{-0.018}$
10.06	9.97–10.14	223	86	$-0.785^{+0.017}_{-0.018}$
10.24	10.14–10.34	223	82	$-0.717^{+0.017}_{-0.017}$
10.52	10.34–11.48	223	73	$-0.662^{+0.016}_{-0.012}$
Binned into the 0.2 dex Width Intervals in M_*				
9.06	8.91–9.10	13	13	$-1.281^{+0.159}_{-0.230}$
9.24	9.10–9.30	35	28	$-1.273^{+0.143}_{-0.180}$
9.40	9.30–9.50	82	46	$-1.087^{+0.095}_{-0.412}$
9.62	9.50–9.70	169	67	$-0.970^{+0.031}_{-0.031}$
9.82	9.70–9.90	257	83	$-0.837^{+0.017}_{-0.017}$
10.00	9.90–10.10	287	98	$-0.792^{+0.016}_{-0.016}$
10.20	10.10–10.30	221	86	$-0.711^{+0.017}_{-0.017}$
10.37	10.30–10.50	147	62	$-0.684^{+0.011}_{-0.017}$
10.59	10.50–10.70	64	42	$-0.672^{+0.025}_{-0.021}$
10.78	10.70–10.90	29	27	$-0.611^{+0.037}_{-0.037}$
11.00	10.91–11.10	17	20	$-0.393^{+0.072}_{-0.078}$

Notes.

^a The top row is for the stack of the entire sample. The middle set of rows is for six equally populated bins of M_* , and the bottom set of rows is for stacks in 0.2 dex fixed-width M_* bins.

^b The S/N per unit pixel (1 \AA) of the stacked spectra around 1450 \AA , represented by the median S/N within $\lambda_{\text{rest}} = 1430\text{--}1470 \text{ \AA}$.

^c Metallicity estimates, represented by the median value of the posterior probability distribution function obtained from the MCMC analysis. The associated errors correspond to the 16th–84th percentiles. Because the fit is sensitive to the iron abundance, this can be translated into $[\text{Fe}/\text{H}]$.

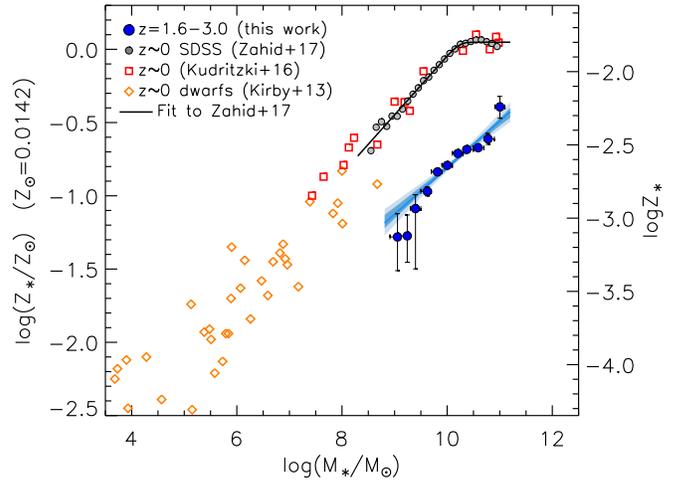


Figure 11. The stellar MZR for our sample (blue circles) at $\langle z \rangle = 2.22$ in comparison with those for $z \sim 0$ galaxies: SDSS galaxies (gray circles; Zahid et al. 2017), individual stars in nearby galaxies (red open squares; Kudritzki et al. 2016), and dwarf galaxies (orange open diamonds; Kirby et al. 2013). These stellar MZRs refer to iron abundances. The blue solid line indicates the best linear fit to our data (Equation (4)) with the 1σ (dark-blue region) and 2σ (light-blue region) confidence limits. The black solid line indicates the best-fit relation of Equation (5) to the local sample. The stellar metallicities of $z \sim 2.2$ galaxies are lower by ~ 0.8 dex than the local galaxies at a given stellar mass. Note that the local stellar metallicities are not corrected for the reduction caused by the measurements based on the rest-frame optical spectra (see Section 4.3).

yields $(Z_0, M_0, \gamma, \beta)_{z \sim 0} = (0.049, 10.24, 0.40, 6.36)$. In Figure 11, this best-fit relation clearly represents the local measurements.

Comparing with the local relation, our result at $z = 1.6\text{--}3.0$ is found to have a slightly shallower slope, 0.32 ± 0.03 (see above Equation (4)). As noted above, the high-redshift relation is offset to lower metallicities and, perhaps for this reason, does not show evidence of a saturation at the high-mass end.

4.3. The $[\text{O}/\text{Fe}]$ –Metallicity Relations in Low- and High-redshift Galaxies

As already noted, the gas-phase and stellar “metallicities” can be more or less translated, at both high and low redshifts, into oxygen and iron abundances, respectively. Comparison of the two may therefore give some insight into the dependence and evolution of the relative abundance ratio, O/Fe , often used as a proxy of the α -enhancement.

The gas-phase metallicity reflects the instantaneous oxygen abundance at the time of observation. Measuring the gas-phase iron abundance is, however, quite challenging because of the faintness of iron emission lines, as well as the uncertainties in dust depletion factors; we therefore need to rely on the stellar metallicities. The stellar metallicities of our high- z galaxies estimated from the rest-frame FUV spectra, which are dominated by short-lived, recently formed stars, offer a more “instantaneous” measurement, similar to the gas-phase O/H . As we discuss in Appendix, we find that the FUV-weighted $[\text{Fe}/\text{H}]$ is consistent within ~ 0.02 dex with the gas-phase values. We can thus adopt the ratio of the gas-phase oxygen abundance to the stellar iron abundance of $z \sim 2.2$ galaxies, $[\text{O}/\text{Fe}]$, as a proxy of the instantaneous α -enhancement of the gas in the galaxies for the remainder of the paper.

In contrast, the iron abundance based on the rest-frame optical spectra could be lower than the instantaneous value because of the substantial contribution to the integrated light

from older stars. This is the case of the stellar MZR at $z \sim 0$ from Zahid et al. (2017). Therefore, the comparison between the two $[\text{Fe}/\text{H}]$ measurements is not straightforward. As shown in Appendix, the offset with respect to the instantaneous $[\text{Fe}/\text{H}]$ is probably around ≈ -0.1 dex for low-redshift galaxies. In the subsequent sections, we thus shift the local stellar MZR by $\Delta \log(Z_*) = +0.1$ dex so that it better reflects the instantaneous iron metallicity. This correction, however, does not have any significant impact on our conclusions.

We take the gas-phase $[\text{O}/\text{H}]$ MZRs at $z \sim 0$ and $z \sim 2.2$, respectively, from Curti et al. (2020, Equation (2)) and Sanders et al. (2020). The former uses the latest accurate calibration between the optical strong-line ratios and the metallicity determined from the direct method. The latter is purely based on the direct method O/H measure in individual 18 galaxies at $z = 1.7\text{--}3.6$ (median $\langle z \rangle = 2.17$). An important caveat is that the sample of Sanders et al. (2020) is not representative of our zCOSMOS-deep sample, but rather a compilation from the literature of different surveys. The following analysis is thus based on an assumption that the result of Sanders et al. (2020) and ours both independently represent the same, typical star-forming galaxy population at these epochs.

Sanders et al. (2020) corrected their O/H measurements for the residuals around the epoch’s main sequence, $\Delta_{\text{MS}} \log(\text{SFR})$, of their sample galaxies to obtain more representative metallicities at given redshift. The corrections achieve ~ 0.2 dex on average. These corrections are, however, questionable at some level owing to systematic uncertainties in determining the shape of the main sequence and the $\Delta_{\text{MS}} \log(\text{SFR})$ dependence of the metallicity. In particular, the authors adopted a locally derived relation, $\Delta(\text{O}/\text{H}) = -0.29 \times \Delta_{\text{MS}} \log(\text{SFR})$. However, this would be a function of redshift, and indeed a weaker dependence, or none at all, has been claimed for $z \sim 2\text{--}3$ galaxies in other papers (Onodera et al. 2016; Sanders et al. 2021). We therefore adopt here their direct measurements of O/H (Equation (7) of Sanders et al. 2020) with a moderate constant correction of $\Delta \log(\text{O}/\text{H}) = +0.1$. Although this choice is more or less arbitrary, it is indeed within the statistical error of the direct measurement.

A similar bias might also be expected in our own sample. However, we do not find a significant offset from the main sequence at this epoch given by Whitaker et al. (2014) in the M_* –SFR diagram of Figure 3. We therefore assume that the observed galaxies and their stellar metallicities are reasonably representative of galaxies of the given M_* at these epochs.

In the following analysis, we ignore the possible depletion of oxygen onto dust grains in the ISM, which would cause some underestimate of the gas-phase O/H. The depletion of O is estimated to be $\sim -0.09 \pm 0.05$ dex in $z \sim 2$ star-forming galaxies (Steidel et al. 2016). This is within the observed uncertainties.

In Figure 12, we compare the empirical fits to the four MZRs (stellar and gas phase; $z \sim 0$ and $z \sim 2.2$). All the metallicities are here normalized to the solar values. It can be seen that the evolution in $[\text{Fe}/\text{H}]$ between $z \sim 0$ and $z \sim 2.2$ is larger than that seen in $[\text{O}/\text{H}]$. At $M_* \sim 10^{10} M_\odot$, the $[\text{Fe}/\text{H}]$ metallicities at $z \sim 2.2$ are about 0.9 dex lower than locally (including the $+0.1$ dex shift in the local stellar MZR; see above), whereas the $[\text{O}/\text{H}]$ metallicities are only about 0.4 dex lower than that at $z \sim 0$. In other words, the offset between the instantaneous $[\text{O}/\text{H}]$ and $[\text{Fe}/\text{H}]$ metallicities evidently increases with redshift; put another way, at $M_* = 10^{10} M_\odot$, $[\text{O}/\text{Fe}] \approx 0$ is

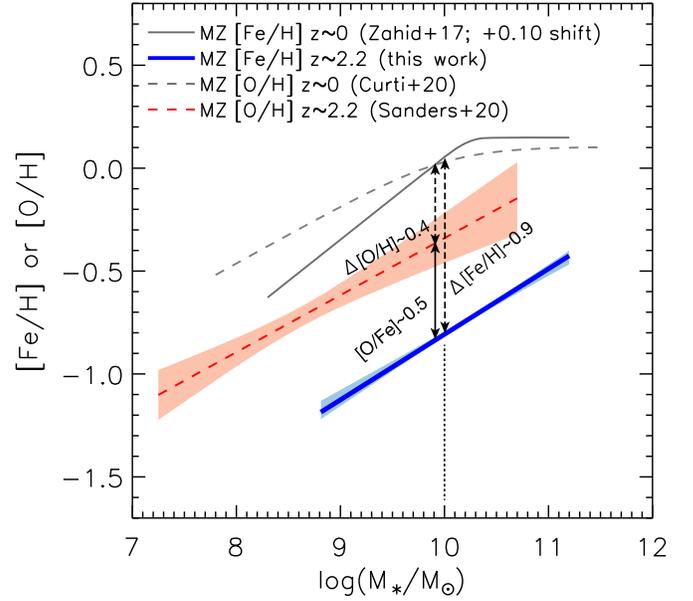


Figure 12. Comparison between the stellar ($[\text{Fe}/\text{H}]$) and gas-phase ($[\text{O}/\text{H}]$) MZRs at low and high redshifts: stellar MZR at $z \sim 0$ (gray solid line; Zahid et al. 2017 fitted with Equation (5)), stellar MZR at $z \sim 2.2$ (thick solid blue line; our data, Equation (5)), gas MZR at $z \sim 0$ (gray dashed line; Curti et al. 2020), gas MZR at $z \sim 2.2$ (red dashed line; Sanders et al. 2020). Here the local stellar MZR is shifted by $\log Z_* = +0.1$ dex so that it reflects better the instantaneous iron abundance in the gas phase. No correction is applied for the $z \sim 2.2$ stellar MZR, as it is based on the rest-frame FUV spectra. The light-blue and red shaded regions indicate, respectively, the 68% confidence limits of the fits.

indicated locally, but $[\text{O}/\text{Fe}] = 0.47 \pm 0.12$ at $z \sim 2.2$. Given the similar slopes of the Fe and O MZRs at $z \sim 2.2$, the average $[\text{O}/\text{Fe}]$ is consistent with being independent of M_* . The implied value of $[\text{O}/\text{Fe}]$ at $z \sim 2.2$ is close to the average $[\text{O}/\text{Fe}] \approx 0.4$ reported for star-forming galaxies at $z \sim 2.3$ (Strom et al. 2018; Cullen et al. 2021). It also approaches the predicted maximum value (~ 0.6) that is predicted for metal enrichment from CCSNe (Nomoto et al. 2006; Kobayashi et al. 2020) alone, suggesting that the production of iron by SNe Ia had not progressed far at these redshifts.

We can in principle eliminate M_* from the gas and stellar MZRs (taken from Figure 12) to yield the relations between $[\text{O}/\text{Fe}]$ and either $[\text{O}/\text{H}]$ or $[\text{Fe}/\text{H}]$. This is done in Figure 13, which shows the $[\text{O}/\text{Fe}]$ as a function of $[\text{Fe}/\text{H}]$ (left panel) and $[\text{O}/\text{H}]$ (right panel). The mass ranges of the $z \sim 0$ and $z \sim 2.2$ relations are limited to $\log(M_*/M_\odot) = 8.5\text{--}11.0$ and $8.8\text{--}10.7$, respectively.

Of course, this procedure of eliminating M_* from the MZR fits may not produce the same $[\text{O}/\text{Fe}]$ versus metallicity relations as would be obtained by considering individual galaxies, because it does not consider the scatter in the observed quantities at a given M_* . In other words, this procedure is tantamount to assuming that galactic mass is the primary driver of the variations in the $[\text{Fe}/\text{H}]$ and $[\text{O}/\text{H}]$ within the observed sample. In the current study, we have no alternative to this procedure, because we relied on stacked spectra (in stellar mass bins) for our metallicity measurements.

The results show that these relations have negative slopes in both panels and at both redshifts. Interestingly, the low-redshift $[\text{O}/\text{Fe}]$ – $[\text{Fe}/\text{H}]$ (and $[\text{O}/\text{Fe}]$ – $[\text{O}/\text{H}]$) relations in Figure 13 appear to broadly line up with the overall evolutionary vectors from $z \sim 2.2$ to $z \sim 0$. The high-redshift $[\text{O}/\text{Fe}]$ – $[\text{Fe}/\text{H}]$ (and

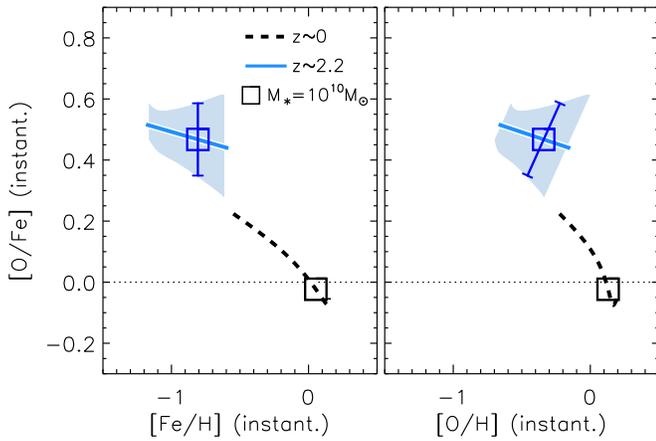


Figure 13. The implied $[O/Fe]$ as a function of the iron abundance $[Fe/H]$ (left panel) and the oxygen abundance $[O/H]$ (right panel) obtained by eliminating M_* from the M_* - $[Fe/H]$ and M_* - $[O/H]$ relations shown in Figure 12. The blue solid and black dashed lines indicate the inferred relations at $z \sim 2.2$ and $z \sim 0$, respectively. The mass ranges are limited to $\log(M_*/M_\odot) = 8.5$ – 11.0 ($z \sim 0$) and 8.8 – 10.7 ($z \sim 2.2$), respectively. For the $z \sim 2.2$ relations, the shaded regions indicate the 68% confidence limit of each relation. The squares indicate the values at $M_* = 10^{10} M_\odot$ at each redshift. The error bar corresponds to the 1σ error. Note that the error in $[O/Fe]$ is correlated with that in the x -axis values; thus, the error bar is nearly vertical in the left panel while being tilted in the right panel because it is dominated by the error in $[O/H]$, while the error in $[Fe/H]$ is much smaller than the symbol size.

$[O/Fe]$ - $[O/H]$) relation is also consistent with having the same slope, but this slope is quite uncertain.

Uncertainties in the slopes of the input MZR can propagate to have a large effect on the resultant slopes of the inferred $[O/Fe]$ versus metallicity relations. We constructed confidence intervals on these slopes by considering the range of possible linear relations that are obtained using random combinations of fits (within the uncertainties) to the input M_* - $[O/H]$ and M_* - $[Fe/H]$ relations. These are shown in Figure 13. Flat or even positive correlations between $[O/Fe]$ and overall metallicities are evidently allowed by our available high-redshift data and analysis methods. We do not try to evaluate the range of possible slopes at low redshift because, while the statistical errors in the fits are very small, it is difficult for us to assess any systematic uncertainties in those taken from the independent studies.

Given the large uncertainties in the slopes of these relations, we focus instead on the $[O/Fe]$ and $[Fe/H]$ (or $[O/H]$) values at a single fiducial mass of $M_* = 10^{10} M_\odot$ (close to the middle of our high- z mass range) as being representative of each redshift. These representative values for $M_* = 10^{10} M_\odot$ at high and low redshifts are shown by the two blue and black squares in each of the panels of Figure 13.

5. Modeling of Iron and Oxygen Abundances

In this section, we explore the observed evolution in both $[O/Fe]$ and $[Fe/H]$ (and $[O/H]$) in the context of “flow-through” gas-regulated models of galaxies. The goal is to demonstrate that the simple chemical evolution model explains well all the observed changes in these quantities from $z \sim 2$ to $z \sim 0$ while assuming that the galaxies have followed the evolving main sequence through cosmic time. In particular, we will derive the evolutionary tracks in the $[O/Fe]$ - $[Fe/H]$ (and $[O/H]$) planes to show that all galaxies must follow limited

evolutionary paths in these diagrams and the locations of the galaxies are determined almost entirely by sSFR alone.

All the metallicities, $[Fe/H]$ and $[O/H]$ (and thus $[O/Fe]$), refer to the instantaneous (gas-phase) values if not specified throughout the section.

5.1. Model Framework

In what follows, we adopt the gas-regulator model of Lilly et al. (2013) in which the SFR is instantaneously regulated by the mass of gas (M_{gas}) present in some reservoir, via the star formation efficiency ($\text{SFE} = \text{SFR}/M_{\text{gas}}$), and with a wind-driven mass loss that scales with the SFR via a “mass-loading” factor η . Mass conservation then straightforwardly gives (see Equation (9) of Lilly et al. 2013)

$$\Phi = (1 - r + \eta) \cdot \text{SFR} + \dot{M}_{\text{gas}}, \quad (6)$$

where Φ is the mass inflow rate and r is the recycling factor (or called the return fraction), i.e., the fraction of mass that is formed into stars and then at later times returned to the ISM.

Obviously, once the two parameters, SFE and mass loading η , are specified (possibly as a function of mass and/or redshift), the gas accretion history (and gas content history $M_{\text{gas}}(t)$) of a given system follows *completely* from its SFH, since \dot{M}_{gas} will also be given by the change in $\text{SFR}(t)$. Note that it is the changing gas reservoir that distinguishes this “gas-regulator” model from the “bathtub” models of Bouché et al. (2010) and Davé et al. (2012), in which \dot{M}_{gas} is set to be zero.

This means that, within the context of the gas-regulator model, the instantaneous metallicity of the gas reservoir will also be completely determined by the SFH once the (possibly mass- and/or epoch-dependent) SFE and η are specified, along with an assumption of the metallicity of the inflowing gas; we will here assume for simplicity that this is zero.

In the following, we consider the instantaneous gas-phase metallicity as the ratio of metal mass in the gas phase and gas mass at given time. We do not consider metal depletion onto dust grains, and thus the gas-phase metal mass represents all metals except those locked in surviving stars and those metals gone in the wind.

We can assume that the oxygen is produced only by CCSNe that occur “promptly” (with zero time delay) after the birth of the progenitor stars. The change of oxygen mass M_{O} in the gas phase is therefore expressed as

$$\frac{dM_{\text{O}}}{dt} = y_{\text{O}}^{\text{CC}} \cdot \text{SFR} - Z_{\text{O}}(1 - r + \eta) \cdot \text{SFR} + \Phi Z_{\text{O}}^{\text{inf}}, \quad (7)$$

where y_{O}^{CC} is the IMF-weighted oxygen yield defined as the oxygen mass synthesized and then returned into the ISM per unit mass formed,²⁴ $Z_{\text{O}} (= M_{\text{O}}/M_{\text{gas}})$ is the oxygen abundance, and $Z_{\text{O}}^{\text{inf}}$ is the oxygen abundance of the infalling gas.

We use Equation (7) to numerically track the chemical evolution. However, it is also useful to formalize the metallicity too. The change in Z_{O} is then obtained by eliminating Φ using

²⁴ This definition of the yield is different from that in some other literature where the yield is denoted in units of mass that is locked up into long-lived stars and remnants. The difference between these definitions is thus a factor of $(1 - r)$.

Equation (6) as

$$\frac{dZ_{\text{O}}}{dt} = y_{\text{O}}^{\text{CC}} - (Z_{\text{O}} - Z_{\text{O}}^{\text{inf}})(1 - r + \eta)\text{SFE} - (Z_{\text{O}} - Z_{\text{O}}^{\text{inf}})\frac{\dot{M}_{\text{gas}}}{M_{\text{gas}}}. \quad (8)$$

The metallicity in an equilibrium condition is thus derived by setting dZ_{O}/dt to zero, i.e.,

$$Z_{\text{O}}^{\text{eq}} = Z_{\text{O}}^{\text{inf}} + \frac{y_{\text{O}}^{\text{CC}}}{1 - r + \eta + \text{SFE}^{-1}((1 - r)\text{sSFR} + \dot{M}_{\text{gas}}/M_{\text{gas}})}. \quad (9)$$

Lilly et al. (2013) showed that the timescale for driving Z_{O} toward Z_{O}^{eq} is shorter than the timescale on which the equilibrium conditions are varying. Assuming equilibrium is therefore a good approximation.

For iron, we must consider the substantial amount of iron that is produced by SNe Ia, which occur with some considerable delay after the birth of their stellar progenitors. In this analysis, we adopt the expression for the distribution of the SN Ia delay time, $f_{\text{Ia}}(t)$, that was formalized by Greggio (2005) for a single stellar population (see also Greggio et al. 2008; Greggio 2010). We consider contributions from the single-degenerate and double-degenerate channels. Note that $f_{\text{Ia}}(t)$ is normalized so that the time integration equals 1. Figure 14 shows the adopted delay time distribution (DTD), compared with a simple exponential parameterization used in Andrews et al. (2017) and Weinberg et al. (2017). The adopted one is more sensitive to SFR at earlier times in the past.

Using f_{Ia} , the change of the gas-phase iron mass M_{Fe} is written as

$$\frac{dM_{\text{Fe}}}{dt} = y_{\text{Fe}}^{\text{CC}} \cdot \text{SFR} - Z_{\text{Fe}}(1 - r + \eta) \cdot \text{SFR} + \Phi Z_{\text{Fe}}^{\text{inf}} + y_{\text{Fe}}^{\text{Ia}} \int_0^t \text{SFR}(t') f_{\text{Ia}}(t - t') dt', \quad (10)$$

where $y_{\text{Fe}}^{\text{CC}}$ is the CCSN iron yield and $Z_{\text{Fe}}^{\text{inf}}$ is the iron abundance of the infalling gas. The last term denotes the contribution from SNe Ia, where $y_{\text{Fe}}^{\text{Ia}}$ is the time-integrated SN Ia yield of iron for unit mass formed.²⁵

The steady-state iron metallicity is then written as

$$Z_{\text{Fe}}^{\text{eq}} = Z_{\text{Fe}}^{\text{inf}} + \frac{y_{\text{Fe}}^{\text{CC}} + (y_{\text{Fe}}^{\text{Ia}}/\text{SFR}) \int_0^t \text{SFR}(t') f_{\text{Ia}}(t - t') dt'}{1 - r + \eta + \text{SFE}^{-1}((1 - r)\text{sSFR} + \dot{M}_{\text{gas}}/M_{\text{gas}})}. \quad (11)$$

Assuming $Z_{\text{O}}^{\text{inf}} = Z_{\text{Fe}}^{\text{inf}} = 0$, Equations (9) and (11) yield

$$\frac{Z_{\text{O}}^{\text{eq}}}{Z_{\text{Fe}}^{\text{eq}}} = \frac{y_{\text{O}}^{\text{CC}}}{y_{\text{Fe}}^{\text{CC}} + (y_{\text{Fe}}^{\text{Ia}}/\text{SFR}) \int_0^t \text{SFR}(t') f_{\text{Ia}}(t - t') dt'} \quad (12)$$

²⁵ The SN Ia yield $y_{\text{Fe}}^{\text{Ia}}$ is usually expressed as the product $K_{\text{Fe}}^{\text{Ia}} R_0$, where R_0 ($\sim 10^{-3} M_{\odot}^{-1}$; Greggio et al. 2008; Maoz & Mannucci 2012) is the time-integrated number of SNe Ia per unit stellar mass formed and $K_{\text{Fe}}^{\text{Ia}}$ is the average mass of iron from an individual SN Ia.

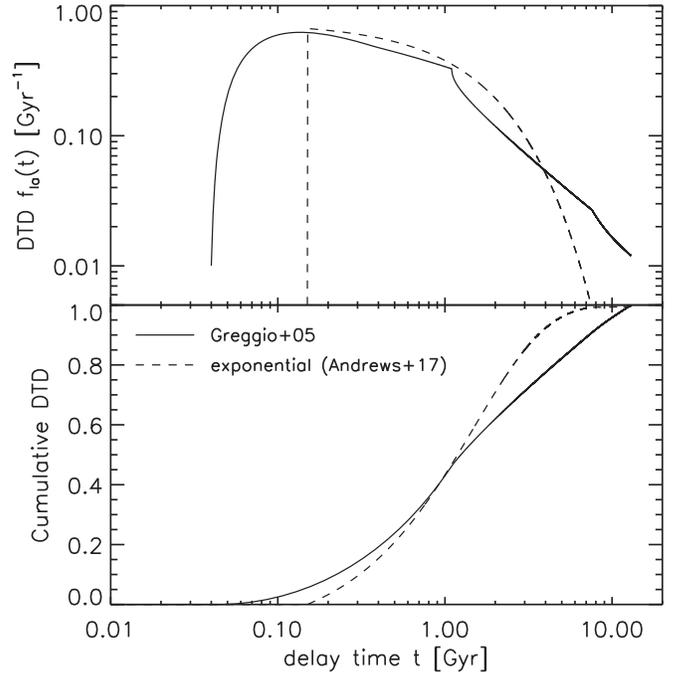


Figure 14. The adopted SN Ia DTD adapted from Greggio (2005) is compared to the exponential DTD adopted in Andrews et al. (2017). Both are normalized so that the total number of events equals 1. The bottom panel shows the cumulative DTDs.

$$\approx \frac{y_{\text{O}}^{\text{CC}} \text{SFR}}{y_{\text{Fe}}^{\text{Ia}} \int_0^t \text{SFR}(t') f_{\text{Ia}}(t - t') dt'}, \quad (13)$$

where the approximation holds when the SNe Ia dominate the iron production. This indicates that the $[\text{O}/\text{Fe}]$ is approximately proportional to the number ratio of the CCSNe and SNe Ia at any time. In other words, as the denominator is some kind of average SFR in the past, the $[\text{O}/\text{Fe}]$ would be tightly correlated with the sSFR. We will see this in Section 5.3.

The parameters in these equations are not very well constrained from observations. We therefore basically follow the “fiducial” choice of Andrews et al. (2017) and Weinberg et al. (2017): $r = 0.4$ and $y_{\text{O}}^{\text{CC}} = 0.017$ (see also Vincenzo et al. 2016 for the IMF-weighted yield and return mass fraction). As discussed below (Section 5.2.2), we adjusted the values of the CCSN and SN Ia iron yields, taking $y_{\text{Fe}}^{\text{CC}} = 0.00081$ (instead of 0.0012) and $y_{\text{Fe}}^{\text{Ia}} = 0.0022$ (instead of 0.0017), together with the parameters determining the SFE and η . As noted above, for simplicity we set both $Z_{\text{O}}^{\text{inf}}$ and $Z_{\text{Fe}}^{\text{inf}}$ to zero.

5.2. Calculating Chemical Evolutionary Tracks

5.2.1. Choice of Star Formation Histories

Again for simplicity, we construct a representative set of SFHs by integrating the evolving main sequence of star-forming galaxies across cosmic time. We adopted the local M_{*} -SFR relation ($z \approx 0.08$) derived by Renzini & Peng (2015) and the redshift evolution at fixed M_{*} as follows:

$$\log \text{SFR} (M_{\odot} \text{ yr}^{-1}) = 0.76 \times \log M_{*}(M_{\odot}) - 7.55 + 3.0 \times \log(1 + z) \quad (14)$$

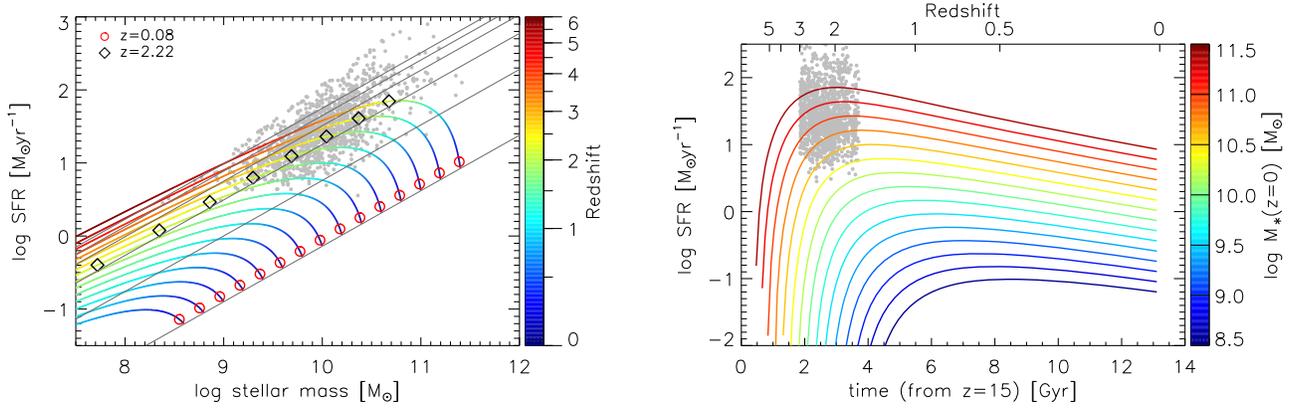


Figure 15. Left panel: the evolving main sequence of star-forming galaxies shown by gray lines at $z = 1, 2, 3, 4$ and 5 from bottom to top. The curves show the evolutionary tracks of galaxies at given stellar masses in steps of 0.2 dex at $z = 0$, color-coded by redshift. The black diamonds and red circles mark the time steps of $z = 2.22$ and $z = 0.08$, respectively, for all shown tracks. Gray dots indicate our sample galaxies at $1.6 \leq z_{\text{spec}} \leq 3.0$. Right panel: the corresponding SFHs. Each line is color-coded by stellar mass at $z = 0$. The gray dots indicate SFR versus z_{spec} for our sample galaxies.

for $z \leq 2.4$ and

$$\begin{aligned} \log \text{SFR} (M_{\odot} \text{ yr}^{-1}) &= \log \text{SFR}(z = 2.4) \\ &+ 1.2 \times \log \left(\frac{1+z}{1+2.4} \right) \end{aligned} \quad (15)$$

for $z > 2.4$ (see, e.g., Lehnert et al. 2015; Tasca et al. 2015).

The inferred evolutionary tracks in the M_{*} –SFR plane and SFHs are shown in Figure 15. In the former, we mark the values of our representative galaxies at $z = 0.08$ and $z = 2.22$. These simulated galaxies at $z = 2.22$ are in broad agreement with our sample galaxies, as shown by gray dots.

5.2.2. Parameters of the Regulator Systems and Iron Yields

In order to derive the time-dependent gas content of each model galaxy from their individual SFHs and thereby compute the corresponding chemical evolution, we need to define the two parameters of the gas-regulator system, the mass-loading factor η and SFE. We assume that these values scale with the instantaneous M_{*} of the system, but that they are not redshift dependent, i.e.,

$$\eta = \max \{ \eta_{10} \times (M_{*}/(10^{10} M_{\odot}))^a, \eta_{\min} \}, \quad (16)$$

$$\text{SFE} = \text{SFE}_{10} \times (M_{*}/(10^{10} M_{\odot}))^b. \quad (17)$$

Here we consider the minimum value for η at high masses (assuming a negative a) for better representation of the saturation feature of the local MZR at the high-mass end. Although this is an arbitrary treatment, the bending in the average η toward high masses is seen owing to effects of AGNs in simulations (Nelson et al. 2019).

Now Equations (7) and (10) can be used to compute the evolution with time of $[\text{O}/\text{H}]$ and $[\text{Fe}/\text{H}]$ for any arbitrary SFR (t), i.e., for each of the representative SFH identified above. However, using the fiducial iron yields given by Andrews et al. (2017) and Weinberg et al. (2017) gives an $[\text{Fe}/\text{H}]$ MZR that is slightly higher at $z \sim 2.2$ but lower locally than those observed at each redshift shown in Figure 12. We also notice that the value of $y_{\text{Fe}}^{\text{CC}} = 0.0012$ adopted in the above papers gives $[\text{O}/\text{Fe}] (=0.5$ with $y_{\text{O}}^{\text{CC}} = 0.017)$ from pure CCSNe that is smaller than the values seen in the literature (~ 0.6 ; e.g., Nomoto et al. 2006). We therefore allowed the CCSN and SN Ia iron yields to vary in order to better reproduce the observed MZRs.

We used an MCMC algorithm to determine values of the six parameters to be $(\eta_{10}, a, \eta_{\min}, \text{SFE}_{10}(\text{Gyr}^{-1}), b, y_{\text{Fe}}^{\text{CC}}, y_{\text{Fe}}^{\text{Ia}}) = (2.58, -0.267, 2.34, 0.310, 0.043, 0.00083, 0.00215)$ that, using our representative SFHs, well reproduce the M_{*} – $[\text{O}/\text{H}]$ and M_{*} – $[\text{Fe}/\text{H}]$ relations at $z \sim 0$ and $z \sim 2.2$ shown in Figure 11.

We note that these values are not far from those suggested from observations and/or simulations. The mass-loading factor and its scaling are similar to what is found in Muratov et al. (2015, see their Equation (8)). The typical (total gas) depletion timescales ($t_{\text{dep}} = 1/\text{SFE}$) of several Gyr have been observed (Bigiel et al. 2008) and reproduced in simulations (Semenov et al. 2017). From Equations (14) and (17), we obtain $M_{\text{gas}} \propto M_{*}^{0.72}$, which is in broad agreement with the scaling relation between the molecular gas mass and stellar mass ($\sim M_{*}^{0.59}$; Tacconi et al. 2020). The corrections of the iron yields are also small ($\sim 30\%$) with respect to the fiducial values.

5.2.3. Calculated Chemical Evolutionary Tracks

In Figure 16, we show the chemical evolution tracks for the input SFHs in the iron and oxygen MZR diagrams. The individual lines correspond to each of the SFHs that were shown in Figure 15. As there, the positions of the galaxies at $z = 2.22$ and at $z = 0.08$ are marked (again, black diamonds and red circles, respectively). At $z = 2.22$ we limit those to having M_{*} higher than $\approx 10^{8.8} M_{\odot}$, which is the lower mass limit of our high- z sample.

This figure illustrates how the models well reproduce the observed MZRs in both oxygen and iron, including their slopes, at both $z \sim 0$ and $z \sim 2.2$. This was, of course, to be expected since both the $\text{SFE}(M_{*})$ and $\eta(M_{*})$ functions of the regulator model and the iron yields have been adjusted so as to match these overall MZRs at both $z \sim 2$ and $z \sim 0$, although we stress that the adopted parameters are, in our view, completely reasonable.

Rather, our interest in constructing these models is in examining and understanding the *expected* relations between the overall metallicity and the α -enhancement at these two redshifts, especially in comparison with the evolutionary offset between them. We address this in the following sections of the paper.

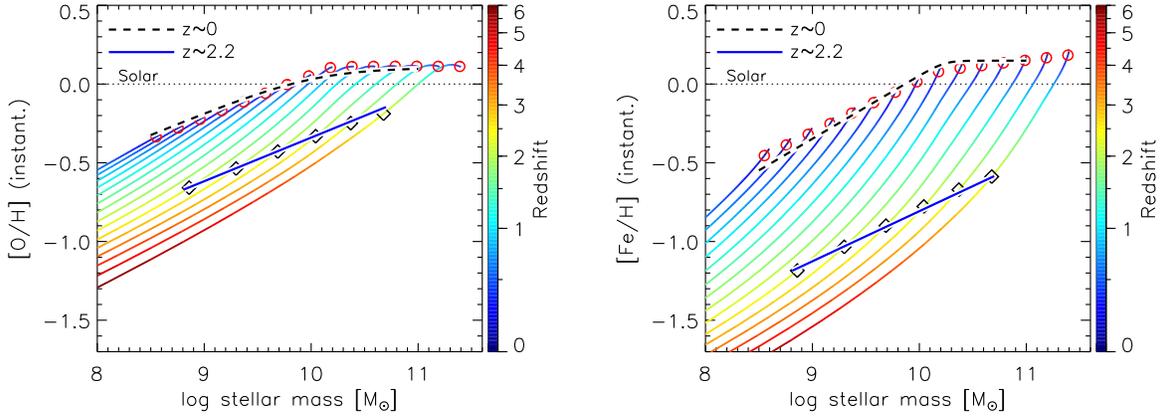


Figure 16. Left panel: the calculated evolutionary tracks in the M_* –[O/H] diagram in comparison with the observations; Curti et al. (2020) for $z \sim 0$ (black dashed line) and Sanders et al. (2020) for $z \sim 2.2$ (blue solid line; shifted by +0.1 dex as mentioned in Section 4.3). The color-coding is according to the redshift. Black diamonds mark those with $M_* \gtrsim 10^{8.8} M_\odot$ at $z = 2.22$, and red circles mark those at $z = 0.08$. Right panel: same as the left panel, but in the M_* –[Fe/H] diagram. The observed relations come from Zahid et al. (2017) at $z \sim 0$ (shifted by -0.1 dex as mentioned in Section 4.3) and our result (Equation (4)).

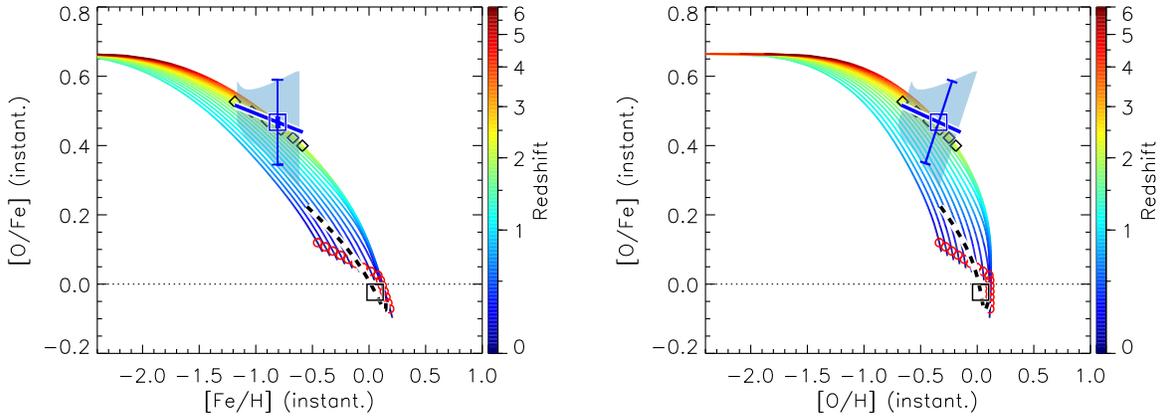


Figure 17. The calculated evolutionary tracks in the [Fe/H]–[O/Fe] (left) and [O/H]–[O/Fe] (right) diagrams in comparison with the observations (same as Figure 13). Visualization of the evolutionary tracks is the same as in Figure 16. The horizontal dotted line marks [O/Fe] = 0 (i.e., the solar value) in both panels.

5.3. The Expected [O/Fe]–Metallicity Relation

In Figure 17, we show the α -enhancement versus overall metallicity of these same models, i.e., [O/Fe] as a function of either [Fe/H] (left panel) or [O/H] (right panel). As in Figure 16, the model galaxies with $M_* \gtrsim 10^9 M_\odot$ are marked at $z = 2.2$ and at $z = 0$ in order to compare with the observations (as in Figure 13). Both the evolutionary tracks in this diagram and the variations within the population at fixed epoch (locus of black and red points) may be read from this diagram.

It can be seen that the modeled evolutionary tracks are in good agreement with the overall change in both metallicity and [O/Fe] between $z \sim 2.2$ and the present epoch. This is not surprising because the model parameters were tuned (within reasonable ranges) to fit the oxygen and iron MZR, and thus implicitly the [O/Fe], at both redshifts.

The models present variations across mass that show a negative slope between these quantities. The observed slopes are similar to what are observed at these redshifts. However, as mentioned in Section 4.3, large uncertainties in the slopes do not enable us to make any robust statement. We thus focus on the representative [O/Fe] and [Fe/H] (or [O/H]) values at our single fiducial mass of $M_* = 10^{10} M_\odot$ at each redshift (squares in Figure 17).

It is noticeable that, at least for the range of representative SFHs considered, the model tracks all follow a relatively narrow path in the [O/Fe]–[Fe/H] plane. The tightness of this

path is enhanced because, at a given epoch, the loci of (model) galaxies of different masses are close to being parallel to the individual evolutionary tracks.

We now turn to investigating whether the observed offset between $z \sim 0$ and $z \sim 2.2$ (i.e., the “evolution vector”) in this diagram can be explained by the change with redshift in some other parameter, and whether that parameter, if present, might also be responsible for producing the variation within the population at a single epoch. While the [O/Fe] (or generally α /Fe) is often loosely considered as being indicative of the “age” of a stellar system, because of the time delay in producing much of the iron via SNe Ia, it should really reflect, especially in a continuous “flow-through” scenario, the sSFR. The sSFR reflects the ratio of the current SFR to some average SFR in the past. This ratio will thus determine the relative number of CCSNe and SNe Ia at any point in time, which will then determine the “instantaneous” α -enhancement of the gas (see Equation (13)).

To explore this, we show in Figure 18 the same evolutionary tracks (as in Figure 17) but now color-coded by the instantaneous sSFR (top panels) and M_* (bottom panels) of the model galaxy. Now the different dependences of the evolutionary tracks on these quantities are quite obvious: it is clear that the [O/Fe] is very tightly correlated with the sSFR across the whole range of these evolutionary tracks and is largely unaffected by either the metallicity or the mass of the

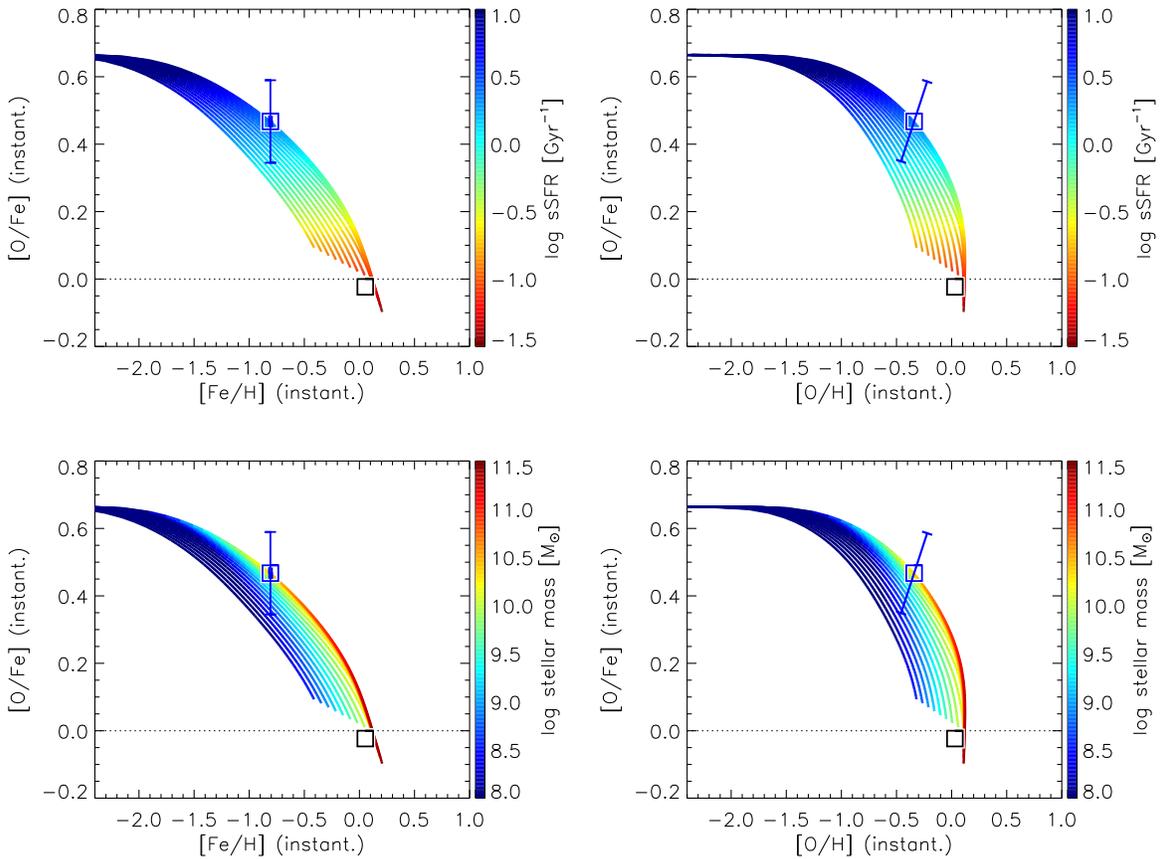


Figure 18. Same as Figure 17, but the evolutionary tracks are here color-coded by the instantaneous sSFR (top panels) and stellar mass (bottom panels). We omit all the symbols except the representative measurements at two redshifts.

system. This is indicated by the striking horizontal banding of the sSFR-coded colors in the top panels. In contrast, the inclined M_* -coded color banding in the bottom panels shows that the metallicity (whether [O/H] or [Fe/H]) depends on both the stellar mass and [O/Fe]. In other words, using the tight relation between [O/Fe] and sSFR, the metallicity depends on the *mass* and *sSFR* together.

Figure 19 shows the same data in a different way, demonstrating that, at least in the gas-regulated models and representative SFHs considered in this paper, there is remarkably little scatter between [O/Fe] and the instantaneous sSFR of the model galaxies. The [O/Fe] scales as $\approx \text{sSFR}^{0.34}$ at $\log(\text{sSFR} [\text{Gyr}^{-1}]) < 0.5$.

These results support the idea that the tight relation between [O/Fe] and sSFR is quite fundamental, being almost independent of the epoch, the shape of the SFHs, and thus, in our modeling, the present-day mass and thus the values of SFE and η .

Lastly, we compare these results to the concept of the so-called “fundamental metallicity relation” (FMR; e.g., Mannucci et al. 2010; Lara-López et al. 2010; see also Andrews & Martini 2013; Lara-López et al. 2013). The FMR established that the SFR appears to be a second parameter in the gas-phase [O/H] MZR at low redshift, in the sense that higher-SFR galaxies at a given mass have lower oxygen metallicities. Further, it was then shown that the observed evolutionary change in [O/H] to high redshift (at a given mass) was the same as that obtained by simply extrapolating the trend with SFR established at $z \sim 0$ to the much higher SFR seen at high redshift.

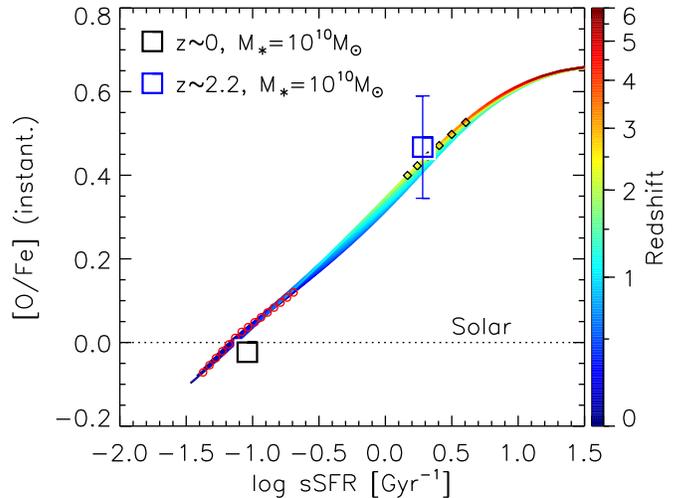


Figure 19. [O/Fe] as a function of sSFR for all the model tracks. The squares indicate the representative measurements at $M_* = 10^{10} M_\odot$ at the two redshifts. Visualization of the evolutionary tracks is the same as in Figure 16.

An important insight into the FMR came from the introduction of the gas-regulated model of galaxies (Lilly et al. 2013). Assuming a quasi-equilibrium state, the (gas-phase) metallicity of the galaxy is set “instantaneously” by two considerations. The first is the specific rate at which the system is being fed by gas ($dM_{\text{gas}}/dt/M_{\text{gas}}$), which itself can be inferred from the sSFR. The second is the values of the two regulator parameters: the SFE and mass loading η . The latter parameters determine the gas content of the system that is

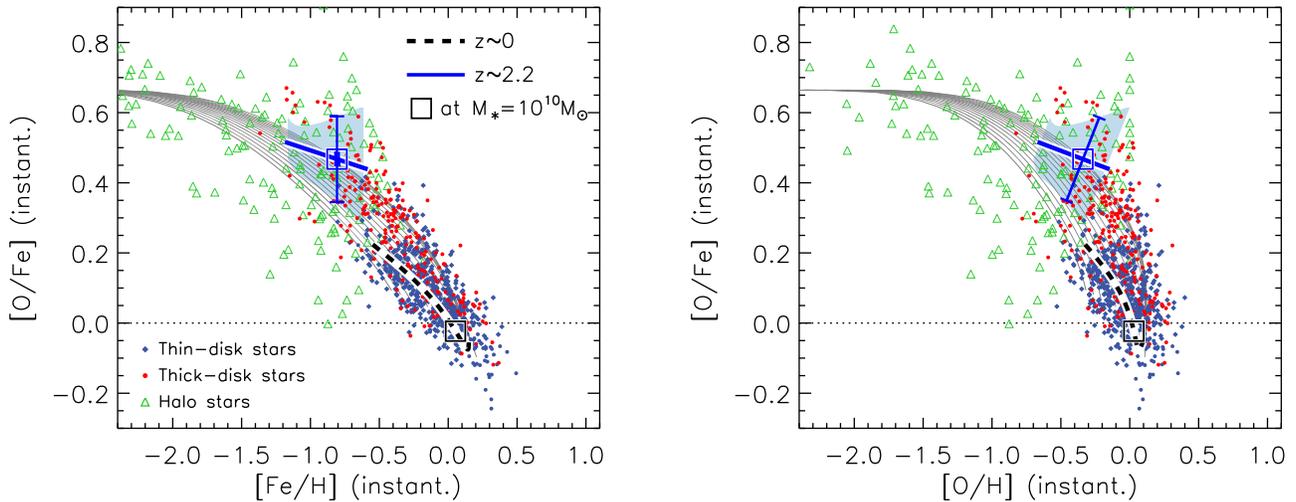


Figure 20. The observed data for $z \sim 2$ and $z \sim 0$ galaxies (same as in Figure 17) are compared to the data for the individual Galactic stars (Ramírez et al. 2013; Amarsi et al. 2019) in the [O/Fe]–metallicity diagrams. The stars are separated into thin-disk stars (dark-blue diamonds), thick-disk stars (red circles), and halo stars (green triangles). The gray curves show the model evolutionary tracks that are shown in Figure 17.

necessary to achieve the required sSFR (see Lilly et al. 2013 for details and discussion). The evolving sSFR of main-sequence galaxies can be viewed in this framework as the primary driver of the observed redshift evolution of the gas metallicity, but the positive MZR at a given epoch is the consequence of the change in sSFR along the main sequence, together with the mass dependences of the SFE and mass loading η .

It is important to clarify that the dependence of [O/H] on the mass-dependent SFE and η (see Lilly et al. 2013) contrasts with the fact that [O/Fe] is determined almost entirely by the sSFR alone, acting as a proxy of the number ratio of CCSNe and SNe Ia at a given time. The variations in SFE and η (with mass) cause the variations in O/H at fixed sSFR (thus at fixed O/Fe) and thus the range in the evolutionary tracks in the [O/Fe]–metallicity diagrams (Figure 17). Note that, in the context of the gas-regulator model, if SFE and η are constant, the gas-phase metallicity is also determined by sSFR alone. The [O/Fe]–metallicity tracks in Figure 17 will thus be completely independent of SFHs, having no scatter, while, at a single epoch, the values of [O/Fe] and metallicities vary along this single path if the sSFR changes with mass along the main sequence, or a population has a scatter in SFR at fixed mass as expected in the real universe.

6. Comparison with Galactic Stars

An important and revealing comparison may be made between our high-redshift galaxies and those of individual Galactic stars. Are we seeing at high redshift the formation of stars of the same type as seen today in the Galaxy? This enables a direct confrontation between “Galactic archeology” and observations of galaxies at high redshifts.

6.1. Thick-disk Stars and High- z Galaxies

Figure 20 shows data for solar neighbor FGK stars in the Milky Way (MW) adapted from Ramírez et al. (2013) and Amarsi et al. (2019). These stars are separated into the thin-disk, thick-disk, and halo populations. It is known that stars in the so-called “thick disk” are typically older, have lower metallicity, and are α -enhanced (i.e., have higher [O/Fe]) compared to stars in the “thin disk.” It is indeed clear that thin-

and thick-disk stars occupy on average different regions in the [O/Fe]–metallicity plane. Halo stars present large scatters in both metallicity and α -enhancement but mostly have lower Z and higher α than the thin-disk stars.

Galactic stars, especially the thick-disk stars, clearly lie along the locus extending between the two $M_* \sim 10^{10} M_\odot$ squares in Figure 20. In particular, it can be seen that our best-estimate locations of $M_* = 10^{10} M_\odot$ galaxies at $z \sim 2$ in the two [O/Fe]–metallicity diagrams lie at the upper tail of the sequence of thick-disk stars in our own Galaxy.

Figure 20 also shows that the simple models constructed in Section 5 successfully reproduce the locus of these Galactic stars. This gives added confidence that they are reasonable.

Martig et al. (2016) found the median ages of the thick-disk population of stars to be ≈ 9 –5 Gyr, decreasing from the inner to the outer disk. Kilic et al. (2017) also estimated the average age of the thick-disk stars to be ≈ 9 Gyr. It is thus quite plausible that those Galactic thick-disk stars located around our high- z measurement (i.e., the relatively metal-poor α -enhanced stars) formed around $z \sim 2$ –3, i.e., some 10 Gyr ago and some 3.5 Gyr after the big bang. We have already discussed how, in terms of the simple flow-through scenario of chemical evolution discussed in Section 5, it is natural to get these high α -enhancements ~ 10 Gyr ago. Note that, in our modeling, present-day galaxies with the same mass as the Galaxy ($6.4 \times 10^{10} M_\odot$; McMillan 2011) had a mass around this fiducial value of $M_* = 10^{10} M_\odot$ at $z \sim 2$.

The evident agreement of [O/Fe] and [Fe/H] (or of [O/Fe] and [O/H]) between the old stars in the Galactic thick disk and the high-redshift galaxies with masses comparable to those expected for MW progenitors provides, in our opinion, a beautifully direct link between the results of Galactic archeology and observations of galaxies in the high- z universe that has long been assumed but rarely established directly (see Cullen et al. 2021 for a similar conclusion).

6.2. The α -dichotomy and Old, Low- α Stars in the MW Bulge

In the subsection above, we found a general agreement between the chemical abundances seen in high- z ($z \sim 2$) galaxies and the MW thick-disk population, plus the estimated ages of the latter, and it supports the idea that, when observing

$z \sim 2$ galaxies, we are witnessing the formation of stars that will constitute the thick-disk population in MW-like galaxies. However, as we will see below, some caveats and questions are posed when a detailed comparison is made with bulge stars in the Galaxy.

Recently, Queiroz et al. (2020) have presented a detailed study of chemical abundances from APOGEE data (Abolfathi et al. 2018) combined with stellar distances from Gaia (Gaia Collaboration et al. 2018). They constructed detailed distributions in $[\alpha/\text{Fe}]$ versus $[\text{Fe}/\text{H}]$ plots at several radial distances from the center of the Galaxy and at three vertical distances from the Galactic plane (see their Figure 6). Note that they measure $[\alpha/\text{Fe}]$ using different multiple α -elements, and the comparison to $[\text{O}/\text{Fe}]$ probably needs some offset (order 0.1 dex; McWilliam et al. 2008; Jönsson et al. 2020). They showed that there are basically two disjoint populations in the $[\alpha/\text{Fe}]-[\text{Fe}/\text{H}]$ plot producing, at each spatial location, a bimodal distribution of stars (the so-called α -dichotomy). Broadly speaking, these two distinct populations comprise a metal-rich (solar or more), low- α (close to solar) component and a lower-metallicity, high- α component. Globally, $\sim 50\%$ of bulge stars belong to the metal-rich component (Zoccali et al. 2017).

In the bulge region ($R < 2$ kpc), the supersolar metal-rich component with $[\alpha/\text{Fe}]$ spread around zero dominates. Moving radially outward in the plane of the disk, the metal-rich component shifts slightly to lower (i.e., subsolar) metallicities without significant change in $[\alpha/\text{Fe}]$, while the metal-poor, high- α component progressively vanishes. Instead, moving vertically away from the plane, it is the metal-rich component that tends to vanish, leaving the metal-poor component dominant, without much change in its abundances. This variation in the spatial manifestation of the bimodality may therefore be related to the “thin” and “thick” disk components, at least at intermediate radii. However, the dominant metal-rich component in the central bulge region challenges this simple picture and poses questions for the beautiful concordance between $z \sim 2$ galaxies and the thick-disk stars shown in Section 6.1.

It is clear that we find no evidence in our zCOSMOS-deep sample at $z \sim 2.2$ for star formation with abundances comparable to the metal-rich low- α component. Some studies, however, have indicated that this metal-rich component is dominated by stars of age ~ 10 Gyr or older, as demonstrated by multiband Hubble Space Telescope photometry extending from the near-UV to the near-IR (Brown et al. 2009; Renzini et al. 2018).²⁶ Based on these indications, this component seems to have formed at $z \sim 2$. The obvious question is, why are such abundances not seen in our $z \sim 2$ zCOSMOS-deep galaxies? If we suppose that MW progenitor(s) should indeed have been found in our sample of main-sequence star-forming galaxies at $z \sim 2.2$, then there are some possible explanations for this observational discrepancy.

The first possibility is that the central metal-rich component in the MW had not in fact formed by $z \sim 2$, i.e., that the ages of these stars have been overestimated. This possibility would be consistent with our observational results and also with our models, since these suggest that galaxies with SFHs that follow the evolution of the main sequence cannot have $[\alpha/\text{Fe}] \sim 0$ at

$z \sim 2$. The main problem with this first possibility is the observational evidence for old ages for these metal-rich low- α stars. One way out could come from assuming that the central regions of the MW are unusual and not representative of the general galaxy population seen in the high-redshift universe.

A second possibility is that metal-rich stars are indeed forming at $z \sim 2$, as indicated by their ages, but that these metal-rich stars will not have contributed to the rest-frame FUV spectra on account of being heavily obscured by dust, which may be common in a supersolar-metallicity environment.

A second possibility is that metal-rich stars are indeed forming at $z \sim 2$, as indicated by their ages, but that these metal-rich stars have contributed to the rest-frame FUV spectra on account of being heavily obscured by dust, which may be common in a supersolar-metallicity environment. The global dust obscuration in the FUV estimated from SED fitting is on average $A_{\text{FUV}} \approx 1.4$ at $M_* = 10^{10} M_\odot$, increasing with M_* from ~ 0 ($\sim 10^9 M_\odot$) to ~ 3 ($\sim 10^{11} M_\odot$). This means that on average 70%–80% of the light from SFR in the galaxies at $M_* = 10^{10} M_\odot$ has been absorbed. If there is a large variation in dust obscuration across the galaxy, increasing toward regions that are more metal-rich and solar- α , the light from this SFR may have been preferentially, and possibly completely, removed from the observed FUV spectra.

We can, however, test the hypothesis of substantial heavily obscured star formation making almost no contribution to the FUV light by examining the far-infrared emission of our sample using the public super-deblended catalog (Jin et al. 2018). Only 54 sources out of the entire sample of 1336 galaxies have FIR luminosity detected at $> 3\sigma$. Even for these FIR-selected galaxies the nominal FIR-based SFRs (median $\sim 140 M_\odot \text{ yr}^{-1}$; estimated in the same way as in Kashino et al. 2021) are only 50% higher than the SED-based SFR estimates ($\sim 100 M_\odot \text{ yr}^{-1}$). There is therefore no compelling evidence with the available data of substantial “hidden” SFR across the entire sample.

The third possibility is that the formation of the metal-rich component was completed well before $z \sim 2$, and that such stars are present in the zCOSMOS-deep galaxies but already old enough so as not to contribute to the ultraviolet light. This may be partly supported by observations that have revealed metal-rich (\gtrsim solar) stars in quiescent galaxies at $z \sim 2$ (Estrada-Carpenter et al. 2019; Morishita et al. 2019). However, such a scenario seems implausible to us for several reasons. Pushing star formation to earlier epochs will clearly tend to *increase* the α -enhancement, not decrease it, exacerbating the problems identified in the previous paragraph. Not least, the stars would have had to form in a still shorter time at even higher sSFR. Indeed, Onodera et al. (2015) found an enhanced average $[\alpha/\text{Fe}]$ (≈ 0.3), along with supersolar $[\text{Z}/\text{H}]$ (≈ 0.24), for quenched galaxies at $z \sim 1.6$.

Lastly, we should also consider a bias in the selection of our sample, namely, that galaxies containing metal-rich star formation may be not represented in our blue-selected ($B_{\text{AB}} \lesssim 25$) zCOSMOS-deep sample, which excludes objects with low FUV continuum. It is naturally expected that galaxies either dominated by metal-rich star formation or having experienced starbursts are dust rich, and thus their rest-frame UV emission may well disappear below the selection limit. A highly complete study, including UV faint sources, is desired to gain a more robust conclusion.

²⁶ In this respect we note that Bensby et al. (2017) derive a broad age distribution for the supersolar low- α stars in the bulge, but still a fraction of them are given ages older than 8–10 Gyr.

We return to emphasize the basic difficulties (required in both the second and third possibilities discussed above) of achieving low (solar) α -enhancement at $z \sim 2$, at least using our fiducial SN Ia DTD that provides the satisfactory match presented in Section 5. We showed in Section 5.3 that the SFHs that are obtained by integrating the main sequence simply do not yield gas abundances with low α -enhancements at these redshifts. This is ultimately because of the high sSFR of the main sequence at $z \sim 2$. Except in extremely contrived scenarios, low α -enhancement stars must be formed at low sSFR. If a significant mass of stars is to be formed, then this star formation must be maintained over an extended period of time, i.e., over timescales of order sSFR^{-1} . This is naturally achieved at late epochs in the universe, but it is hard to see how this can happen at much earlier times. Quite independent of our own observational results, this conundrum represents a basic puzzle about the origin of the low α -enhancement metal-rich central population identified in the central kiloparsec of our Galaxy by Queiroz et al. (2020) if these stars are truly over 10 Gyr old rather than $\lesssim 4$ Gyr old (i.e., formed after $z \sim 0.5$) as indicated by our model (Figure 17).

It is clear that our models, which employ representative smooth SFHs, produce one continuous sequence of stars in the α -metallicity plane, rather than the two distinct components as observed (Queiroz et al. 2020). This suggests that the construction of the MW was a more complex phenomenon. Recently, an SFH with an extended gap (~ 6 Gyr) between the early formation of the metal-poor high- α and the metal-rich low- α components has been proposed to reproduce the α -dichotomy of the MW stars (Lian et al. 2020; Queiroz et al. 2021). As star formation of the metal-poor component quenched, the ISM was enriched only in Fe by ongoing production of SNe Ia on a ~ 1 Gyr timescale, and the metal-rich, low- α populations were produced from the low- α gas when star formation resumed after the gap. In this scenario, the MW would have been in the gap at $z \sim 2-3$. But this is manifestly not the case for the majority of our sample galaxies, as they actually have high sSFRs. This would suggest that the α -dichotomy may be a characteristic of the MW rather than a common feature, though some cases of “rejuvenation” after a hiatus in star formation have been proposed (Mancini et al. 2019). We should also mention another work by Clarke et al. (2019), who proposed clumpy star formation led by disc fragmentation as the origin of high- α populations, and claimed that the α -dichotomy is common in massive galaxies as high- z galaxies commonly show clumpy morphologies.

Lastly, there is another possibility to be mentioned, namely, that the SN Ia DTD may shift to shorter delays with increasing metallicity, in particular in the supersolar regime. For example, in the double-degenerate scenario for SN Ia progenitors, the event results from the merging of two white dwarfs (WD) as they spiral in owing to gravitational wave radiation. Thus, the DTD is controlled by the distribution of the binary WD separations as they emerge from their last common envelope event. As the delay time scales as the fourth power of such separation (e.g., Greggio 2005), a $\sim 30\%$ reduction in the distribution of WD separations would result in a shift by a factor of ~ 5 in the DTD, e.g., moving the median delay time from, say, ~ 1 Gyr to ~ 200 Myr. The amount of orbital shrinkage during common envelope events is notoriously hard to predict; nevertheless, we note that stellar sizes are larger at high metallicity, and hence stars fill their Roche lobe at an

earlier evolutionary phase compared to the case at lower metallicity. An effect of metallicity on the WD orbital separations and on the resulting DTD is therefore to be expected, though hard to quantitatively predict, and the few attempts in this direction remain inconclusive (Meng et al. 2011; Meng & Yang 2012; Kistler et al. 2013). The old ages, high metallicity, and low $[\alpha/\text{Fe}]$ of bulge stars could then be more easily reconciled if the DTD were to move to substantially shorter delays in the supersolar regime.

It is clear to us that integrated light spectroscopy of high-redshift galaxies, such as presented in this paper, can highlight the problem, but it is unlikely to enable us to understand the complexities revealed by individual star abundances across the body of the MW. It is possible that high-resolution imaging and integrated field spectroscopy with the forthcoming JWST instruments will finally give clues to the solution of this puzzle.

7. Summary

We have measured the stellar metallicities of galaxies for 1336 star-forming galaxies at $1.6 \leq z \leq 3.0$ using high-S/N stacked low-resolution ($R \sim 200$) rest-frame FUV spectra from the zCOSMOS-deep survey. The metallicities were estimated using fits of high-resolution model spectra constructed from stellar population synthesis models across a range of the stellar metallicity. These metallicity estimates enabled us to construct the relationship between stellar mass and stellar metallicity, i.e., the stellar MZR at $z \sim 2.2$.

The measured stellar metallicities, which mostly reflect the iron abundance, range between $-1.3 \lesssim \log Z_*/Z_\odot \lesssim -0.4$ across the stellar mass range $10^9 \lesssim M_*/M_\odot \lesssim 10^{11}$. Because they are based on the spectra of short-lived massive stars, we argue that this iron abundance should be representative of the gas phase in these galaxies. Our measurements are consistent with the one previous work on similarly high- z galaxies (Cullen et al. 2019).

A clear positive correlation between stellar mass and stellar metallicity is established, in which the metallicity scales as $\log(Z_*/Z_\odot) = -0.81 + 0.32 \log(M_*/10^{10}M_\odot)$ (Equation (4)). Comparing with iron metallicity data at $z \sim 0$ from the literature, we find that the $z \sim 2.2$ stellar MZR is offset by ~ 0.8 dex below the local relation.

Adding published [O/H] measurements at both high and low redshifts, we implied [O/Fe] ratios (i.e., α -enhancement) to be $\approx 0.47 \pm 0.12$ at $z \sim 2$ for $M_* \sim 10^{10}M_\odot$ galaxies. This is considerably enhanced against the local value of [O/Fe] ~ 0 and, in fact, approaches the [O/Fe] ~ 0.6 limit imposed by the yields of CCSNe. This indicates that SNe Ia have not yet contributed very much to the iron supply at these epochs.

These results are then compared with the expectations of “flow-through” gas-regulator models, especially in the context of evolutionary tracks in the [O/Fe]-metallicity plane. In constructing these, it is assumed that the SFHs follow those implied by the evolving main sequence of star-forming galaxies. Adjusting the regulator parameters and iron yields within very reasonable ranges, it is found that the models can reproduce the evolution of [O/Fe] and [Fe/H] (or [O/H]) from $z \sim 2$ to $z \sim 0$. The models predict that galaxies at $M_* \sim 10^9-10^{11}$ locally should lie on a relatively narrow locus in the [O/Fe]-[Fe/H] (or [O/H]) plane if they have been continuously forming stars and following the main sequence in the past.

An important insight obtained from this modeling is that the instantaneous (gas-phase) α -enhancement is determined almost entirely by the instantaneous sSFR of the galaxy. This is because the sSFR is a good proxy of the instantaneous number ratio of CCSNe to SNe Ia, which is what effectively determines the gas-phase α -enhancement in flow-through models. The variations in $[\text{O}/\text{Fe}]$ among a galaxy population at a single epoch arise as a result of the mass dependence of the main-sequence sSFR, while the variations in the evolutionary tracks in the $[\text{O}/\text{Fe}]$ –metallicity planes arise as a result of the dependence of the regulator parameters (SFE and mass loading η) on the stellar mass.

We found that $z \sim 2$ galaxies at a representative mass of $10^{10} M_{\odot}$ have similar α -enhancement and metallicity to the low-metallicity thick-disk stars in our own Galaxy. These Galactic stars were probably formed around 10 Gyr ago, when the MW presumably had a stellar mass of around this same value, $10^{10} M_{\odot}$. This observation therefore provides an unusually direct concordance between the results of Galactic archaeology and observations of presumed MW progenitors at high redshift.

There remains, however, an open question about the formation of the population of old metal-rich stars seen in the MW bulge with low (roughly solar) α -enhancement. Our rest-frame FUV data at $z \sim 2$ show no evidence of such high metallicities and low- α enhancements. Three possible explanations of this observational discrepancy have been considered: (i) These central Galactic stars may not be as old as so far estimated with a variety of methods, and formed well after $z \sim 2$. Besides being in contrast with current age estimates, this option would predict the existence of supersolar star-forming galaxies at lower redshifts, which have not been observed so far. (ii) They may indeed be forming at $z \sim 2$ but doing so in highly obscured environments as one may expect in a supersolar-metallicity regime. (iii) They may have already formed well before $z \sim 2$. We discussed how both the second and third possibilities are problematic in terms of achieving low α -enhancements by $z \sim 2$ and, as a possible solution, the possibility that the SNIa delay time distribution shifts to substantially shorter delays in the supersolar-metallicity regime.

Spatially resolved imaging and spectroscopy, which will be enabled by JWST, may be able to constrain the formation scenarios of the possible metal-rich bulges and the detailed process of chemical evolution through cosmic history. Furthermore, the forthcoming multiobject spectrographs, i.e., Subaru/PFS and VLT/MOONS, will enable less biased studies with highly complete samples, including UV faint sources that were excluded in this study.

This research is based on observations undertaken at the European Southern Observatory (ESO) Very Large Telescope (VLT) under the Large Program 175.A-0839 and has been supported by the Swiss National Science Foundation (SNF) and JSPS KAKENHI grant No. JP21K13956. This work made use of v2.2.1 of the Binary Population and Spectral Synthesis (BPASS) models as described in Eldridge et al. (2017) and Stanway & Eldridge (2018). Y.P. acknowledges National Science Foundation of China (NSFC) grant Nos. 12125301, 11773001, and 11991052.

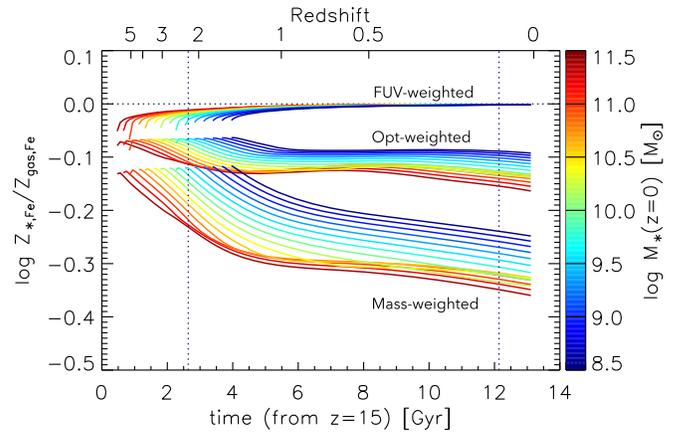


Figure 21. The stellar-to-gas iron metallicity ratio as a function of cosmic time for representative SFHs adopted in Section 5. The stellar $Z_{*,\text{Fe}}$ values are inferred weighting by stellar mass, FUV luminosity, or optical luminosity, as labeled. The evolutionary tracks for the individual SFHs are color-coded by the present-day stellar mass. Two redshifts, $z = 2.2$ of our high- z sample and $z = 0.08$ of the local sample, are marked by vertical dotted lines. The horizontal dotted line marks $Z_{*,\text{Fe}} = Z_{\text{gas,Fe}}$.

Appendix A

Offset between Gas-phase and Stellar Fe/H

It is straightforward to calculate mass-weighted stellar metallicities, $Z_{\text{Fe},*}^M$, once SFR(t) and (gas-phase) $Z_{\text{Fe}}(t)$ are specified:

$$Z_{\text{Fe},*}^M(t) = \frac{\int_0^t Z_{\text{Fe}}(t') \text{SFR}(t') dt'}{\int_0^t \text{SFR}(t') dt'}. \quad (\text{A1})$$

To calculate luminosity-weighted stellar metallicities, we additionally need a library of the spectra for a single stellar population as a function of age. We adopted the same BPASSv2.2.1 template spectra as in the main analysis and calculated the stellar metallicity weighted by the average luminosity around 1500 Å (FUV) and 5500 Å (optical):

$$Z_{\text{Fe},*}^L(t) = \frac{\int_0^t Z_{\text{Fe}}(t') L_{\lambda}^{\text{SSP}}(t - t') \text{SFR}(t') dt'}{\int_0^t L_{\lambda}^{\text{SSP}}(t - t') \text{SFR}(t') dt'}, \quad (\text{A2})$$

where $L_{\lambda}^{\text{SSP}}(t)$ ($\lambda = 1500$ or 5500 Å) is the luminosity density at a particular wavelength of a single stellar population of age t . Note that, in this equation, we ignore the effects of possible differential dust attenuation between the younger and longer-lived stellar components.

We calculated Equations (A1)–(A2) for the same evolutionary tracks obtained in Section 5. Figure 21 shows the ratios of the stellar $[\text{Fe}/\text{H}]$, weighted by either mass, L_{1500} (FUV), or L_{5500} (optical), to the instantaneous gas-phase $[\text{Fe}/\text{H}]$ as a function of cosmic time. The different evolutionary tracks correspond to the different SFHs and are color-coded by the present-day stellar mass. The two redshifts $z = 0.08$ and 2.2 are marked.

As expected, the L_{1500} -weighted $[\text{Fe}/\text{H}]$ is similar to the gas-phase value within ~ 0.02 dex at $z = 2.2$, and almost equivalent at $z = 0$ when the metallicity change is slow. In contrast, the mass-weighted and L_{5500} -weighted $[\text{Fe}/\text{H}]$ values show substantial offsets and some scatters for different SFHs. At $z = 0.08$, the offsets in the L_{5500} -weighted values are in a range

of $-(0.08-1.5)$ dex, being larger (in negative) for larger masses, i.e., lower sSFR. We adopted -0.1 dex, the value for $M_*(z \sim 0) = 10^{10} M_\odot$, as a representative value for correcting the local stellar $[\text{Fe}/\text{H}]$ from Zahid et al. (2017), so that they reflect better instantaneous values for comparison with our FUV-based $[\text{Fe}/\text{H}]$ and the model values in Section 4.3 and later.

ORCID iDs

Daichi Kashino  <https://orcid.org/0000-0001-9044-1747>
 Simon J. Lilly  <https://orcid.org/0000-0002-6423-3597>
 Alvio Renzini  <https://orcid.org/0000-0002-7093-7355>
 Emanuele Daddi  <https://orcid.org/0000-0002-3331-9590>
 Giovanni Zamorani  <https://orcid.org/0000-0002-2318-301X>
 John D. Silverman  <https://orcid.org/0000-0002-0000-6977>
 Olivier Ilbert  <https://orcid.org/0000-0002-7303-4397>
 Vincenzo Mainieri  <https://orcid.org/0000-0002-1047-9583>
 Sandro Bardelli  <https://orcid.org/0000-0002-8900-0298>
 Elena Zucca  <https://orcid.org/0000-0002-5845-8132>
 Jeyhan S. Kartaltepe  <https://orcid.org/0000-0001-9187-3605>
 David B. Sanders  <https://orcid.org/0000-0002-1233-9998>

References

- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, 235, 42
 Amarsi, A. M., Nissen, P. E., & Skúladóttir, Á. 2019, *A&A*, 630, A104
 Andrews, B. H., & Martini, P. 2013, *ApJ*, 765, 140
 Andrews, B. H., Weinberg, D. H., Schönrich, R., & Johnson, J. A. 2017, *ApJ*, 835, 224
 Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
 Brandt, J. C., Feltzing, S., Gould, A., et al. 2017, *A&A*, 605, A89
 Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, 136, 2846
 Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, *A&A*, 622, A103
 Bouché, N., Dekel, A., Genzel, R., et al. 2010, *ApJ*, 718, 1001
 Brandt, J. C., Heap, S. R., Beaver, E. A., et al. 1998, *AJ*, 116, 941
 Brown, T. M., Sahu, K., Zoccali, M., et al. 2009, *AJ*, 137, 3172
 Bruzual, G., & Charlot, S. 2003, *MNRAS*, 344, 1000
 Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, *MNRAS*, 360, 1413
 Calabrò, A., Castellano, M., Pentericci, L., et al. 2021, *A&A*, 646, A39
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, 533, 682
 Chabrier, G. 2003, *PASP*, 115, 763
 Charlot, S., & Fall, S. M. 2000, *ApJ*, 539, 718
 Ciesla, L., Boselli, A., Elbaz, D., et al. 2016, *A&A*, 585, A43
 Ciesla, L., Elbaz, D., & Fensch, J. 2017, *A&A*, 608, A41
 Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, 819, 62
 Clarke, A. J., Debattista, V. P., Nidever, D. L., et al. 2019, *MNRAS*, 484, 3476
 Cullen, F., McLure, R. J., Dunlop, J. S., et al. 2019, *MNRAS*, 487, 2038
 Cullen, F., McLure, R. J., Dunlop, J. S., et al. 2020, *MNRAS*, 495, 1501
 Cullen, F., Shapley, A. E., McLure, R. J., et al. 2021, *MNRAS*, 505, 903
 Curti, M., Cresci, G., Mannucci, F., et al. 2017, *MNRAS*, 465, 1384
 Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, *MNRAS*, 491, 944
 Daddi, E., Cimatti, A., Renzini, A., et al. 2004, *ApJ*, 617, 746
 Dale, D. A., Helou, G., Magdis, G. E., et al. 2014, *ApJ*, 784, 83
 Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, *MNRAS*, 421, 98
 Dean, C. A., & Bruhweiler, F. C. 1985, *ApJS*, 57, 133
 Eldridge, J. J., Stanway, E. R., Xiao, L., et al. 2017, *PASA*, 34, e058
 Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, *ApJ*, 644, 813
 Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2019, *ApJ*, 870, 133
 Faissst, A. L., Capak, P. L., Davidzon, I., et al. 2016, *ApJ*, 822, 29
 Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, *RMxAA*, 53, 385
 Foreman-Mackey, D. 2016, *JOSS*, 1, 24
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, 616, A1
 Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
 Greggio, L. 2005, *A&A*, 441, 1055
 Greggio, L. 2010, *MNRAS*, 406, 22
 Greggio, L., Renzini, A., & Daddi, E. 2008, *MNRAS*, 388, 829
 Halliday, C., Daddi, E., Cimatti, A., et al. 2008, *A&A*, 479, 417
 Harikane, Y., Laporte, N., Ellis, R. S., & Matsuoka, Y. 2020, *ApJ*, 902, 117
 Jin, S., Daddi, E., Liu, D., et al. 2018, *ApJ*, 864, 56
 Jönsson, H., Holtzman, J. A., Allende Prieto, C., et al. 2020, *AJ*, 160, 120
 Kashino, D., & Inoue, A. K. 2019, *MNRAS*, 486, 1053
 Kashino, D., Lilly, S. J., Silverman, J. D., et al. 2021, *ApJ*, 909, 213
 Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, *ApJ*, 835, 88
 Kilic, M., Munn, J. A., Harris, H. C., et al. 2017, *ApJ*, 837, 162
 Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, *ApJ*, 779, 102
 Kistler, M. D., Stanek, K. Z., Kochanek, C. S., Prieto, J. L., & Thompson, T. A. 2013, *ApJ*, 770, 88
 Kobayashi, C., Karakas, A. I., & Lugaro, M. 2020, *ApJ*, 900, 179
 Kojima, T., Ouchi, M., Rauch, M., et al. 2020, *ApJ*, 898, 142
 Kriek, M., Price, S. H., Conroy, C., et al. 2019, *ApJL*, 880, L31
 Kudritzki, R. P., Castro, N., Urbaneja, M. A., et al. 2016, *ApJ*, 829, 70
 Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
 Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, *A&A*, 521, L53
 Lara-López, M. A., Hopkins, A. M., López-Sánchez, A. R., et al. 2013, *MNRAS*, 434, 451
 Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, *Proc. SPIE*, 4841, 1670
 Leethochawalit, N., Kirby, E. N., Moran, S. M., Ellis, R. S., & Treu, T. 2018, *ApJ*, 856, 15
 Lehnert, M. D., van Driel, W., Le Tiran, L., Di Matteo, P., & Haywood, M. 2015, *A&A*, 577, A112
 Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, *A&A*, 80, 155
 Lian, J., Thomas, D., Maraston, C., et al. 2020, *MNRAS*, 497, 2371
 Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, *ApJ*, 772, 119
 Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
 Ly, C., Malkan, M. A., Nagao, T., et al. 2014, *ApJ*, 780, 122
 Ly, C., Malkan, M. A., Rigby, J. R., & Nagao, T. 2016, *ApJ*, 828, 67
 Maiolino, R., & Mannucci, F. 2019, *A&ARv*, 27, 3
 Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, 488, 463
 Mancini, C., Daddi, E., Juneau, S., et al. 2019, *MNRAS*, 489, 1265
 Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, 408, 2115
 Maoz, D., & Mannucci, F. 2012, *PASA*, 29, 447
 Maraston, C., Pforr, J., Renzini, A., et al. 2010, *MNRAS*, 407, 830
 Martig, M., Minchev, I., Ness, M., Fouesneau, M., & Rix, H.-W. 2016, *ApJ*, 831, 139
 McLure, R. J., Pentericci, L., Cimatti, A., et al. 2018, *MNRAS*, 479, 25
 McMillan, P. J. 2011, *MNRAS*, 414, 2446
 McWilliam, A., Matteucci, F., Ballero, S., et al. 2008, *AJ*, 136, 367
 Meng, X., & Yang, W. 2012, *A&A*, 543, A137
 Meng, X. C., Li, Z. M., & Yang, W. M. 2011, *PASJ*, 63, 31
 Morishita, T., Abramson, L. E., Treu, T., et al. 2019, *ApJ*, 877, 141
 Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al. 2015, *MNRAS*, 454, 2691
 Nagao, T., Maiolino, R., & Marconi, A. 2006, *A&A*, 459, 85
 Nelson, D., Pillepich, A., Springel, V., et al. 2019, *MNRAS*, 490, 3234
 Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, *A&A*, 507, 1793
 Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C., & Maeda, K. 2006, *NuPhA*, 777, 424
 Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, *ApJ*, 822, 42
 Onodera, M., Carollo, C. M., Renzini, A., et al. 2015, *ApJ*, 808, 161
 Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, *A&A*, 603, A102
 Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, 348, L59
 Pforr, J., Maraston, C., & Tonini, C. 2012, *MNRAS*, 422, 3285
 Queiroz, A. B. A., Anders, F., Chiappini, C., et al. 2020, *A&A*, 638, A76
 Queiroz, A. B. A., Chiappini, C., Perez-Villegas, A., et al. 2021, *A&A*, 656, A156
 Ramírez, I., Allende Prieto, C., & Lambert, D. L. 2013, *ApJ*, 764, 78
 Renzini, A., Gennaro, M., Zoccali, M., et al. 2018, *ApJ*, 863, 16
 Renzini, A., & Peng, Y.-j. 2015, *ApJL*, 801, L29
 Rix, S. A., Pettini, M., Leitherer, C., et al. 2004, *ApJ*, 615, 98
 Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, *ApJ*, 914, 19
 Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, *ApJ*, 799, 138
 Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, *ApJ*, 816, 23
 Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2020, *MNRAS*, 491, 1427
 Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, *ApJS*, 172, 38
 Semenov, V. A., Kravtsov, A. V., & Gnedin, N. Y. 2017, *ApJ*, 845, 133
 Stanway, E. R., & Eldridge, J. J. 2018, *MNRAS*, 479, 75
 Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, *ApJ*, 604, 534
 Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, *ApJ*, 826, 159

- Strom, A. L., Steidel, C. C., Rudie, G. C., Trainor, R. F., & Pettini, M. 2018, *ApJ*, **868**, 117
- Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, *ApJ*, **836**, 164
- Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, *ARA&A*, **58**, 157
- Tasca, L. A. M., Le Fèvre, O., Hathi, N. P., et al. 2015, *A&A*, **581**, A54
- Topping, M. W., Shapley, A. E., Reddy, N. A., et al. 2020a, *MNRAS*, **495**, 4430
- Topping, M. W., Shapley, A. E., Reddy, N. A., et al. 2020b, *MNRAS*, **499**, 1652
- Trager, S. C., Faber, S. M., Worthey, G., & González, J. J. 2000, *AJ*, **119**, 1645
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, **613**, 898
- Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, *MNRAS*, **455**, 4183
- Weinberg, D. H., Andrews, B. H., & Freudenburg, J. 2017, *ApJ*, **837**, 183
- Whitaker, K. E., Franx, M., Leja, J., et al. 2014, *ApJ*, **795**, 104
- Yabe, K., Ohta, K., Iwamuro, F., et al. 2012, *PASJ*, **64**, 60
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zahid, H. J., Dima, G. I., Kudritzki, R.-P., et al. 2014a, *ApJ*, **791**, 130
- Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014b, *ApJ*, **792**, 75
- Zahid, H. J., Kewley, L. J., & Bresolin, F. 2011, *ApJ*, **730**, 137
- Zahid, H. J., Kudritzki, R.-P., Conroy, C., Andrews, B., & Ho, I.-T. 2017, *ApJ*, **847**, 18
- Zoccali, M., Vasquez, S., Gonzalez, O. A., et al. 2017, *A&A*, **599**, A12