

Evolution of the Stellar Mass–Metallicity Relation. II. Constraints on Galactic Outflows from the Mg Abundances of Quiescent Galaxies

NICHA LEETHOCHAWALIT,^{1,2} EVAN N. KIRBY,¹ RICHARD S. ELLIS,³ SEAN M. MORAN,⁴ AND TOMMASO TREU⁵

¹*Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA*

²*National Astronomical Research Institute of Thailand, MaeRim, Chiang Mai, 50180, Thailand*

³*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

⁴*Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA*

⁵*Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA*

(Accepted September 24, 2019)

ABSTRACT

We present the stellar mass–[Fe/H] and mass–[Mg/H] relation of quiescent galaxies in two galaxy clusters at $z \sim 0.39$ and $z \sim 0.54$. We derive the age, [Fe/H], and [Mg/Fe] for each individual galaxy using a full-spectrum fitting technique. By comparing with the relations for $z \sim 0$ SDSS galaxies, we confirm our previous finding that the mass–[Fe/H] relation evolves with redshift. The mass–[Fe/H] relation at higher redshift has lower normalization and possibly steeper slope. However, based on our sample, the mass–[Mg/H] relation does not evolve over the observed redshift range. We use a simple analytic chemical evolution model to constrain average outflow that these galaxies experience over their lifetime, via the calculation of mass-loading factor. We find that the average mass-loading factor η is a power-law function of galaxy stellar mass, $\eta \propto M_*^{-0.21 \pm 0.09}$. The measured mass-loading factors are consistent with the results of other observational methods for outflow measurements and with the predictions where outflow is caused by star formation feedback in turbulent disks.

Keywords: galaxies: abundances – galaxies: evolution – galaxies: stellar content – galaxies: feedback

1. INTRODUCTION

Over the past five decades, we have made significant progress in measuring gas metallicities in star-forming galaxies. The relation between galaxy luminosity (stellar mass) and interstellar oxygen abundances in extragalactic H II regions was established 50 years ago (e.g. McClure & van den Bergh 1968; Lequeux et al. 1979; Garnett & Shields 1987). Tremonti et al. (2004) used the large statistical sample size of local star-forming galaxies from the Sloan Digital Sky survey (SDSS) and confirmed that the correlation between metallicity and mass is more fundamental than that between metallicity and luminosity. The mass–metallicity relation (MZR) is such that more massive galaxies have higher gas metallicities. In later work, the gas-phase MZR was also found to be present at high redshift and also to evolve with redshift (e.g. Maiolino et al. 2008; Zahid et al. 2013).

Although the gas-phase MZR has been known for almost four decades, its physical drivers are still debated. Early works mostly suggested that galactic winds are the primary agent that drives the relation (e.g., Mathews & Baker 1971; Larson 1974; Garnett 2002; Tremonti

et al. 2004). Lower mass galaxies have shallower potential wells and therefore can retain less of the metals they produced (Dekel & Silk 1986). However, later works argued that the metallicity is regulated by a more complex mechanism that includes the interplay between inflow, outflow, and enrichment rate (e.g. Finlator & Davé 2008; Davé et al. 2011). Spitoni et al. (2010) attempted to use chemical evolution models to explain the observed gas-phase MZR in the SDSS galaxies. They found that a range of models that include either outflow only, both inflow and outflow, or inflow with variable outflow can all explain the observed data equally well. In essence, there is a degeneracy between inflow and outflow rate that cannot be differentiated using only the data from the gas-phase MZR.

However, metals do not reside only in gas. Especially at lower redshift $z < 0.5$, the cold gas fraction is $< 20\%$ in most star-forming galaxies and less than a few percent in quiescent galaxies (e.g. Gobat et al. 2018). The majority of disk and metal mass is in stars (e.g., Werk et al. 2014). When Gallazzi et al. (2005) measured the stellar metallicities of local SDSS galaxies using spec-

tral indices, they found that stellar metallicity also exhibits a tight correlation with stellar mass in galaxies with $M_* > 10^9 M_\odot$. Kirby et al. (2013) measured stellar metallicities of individual stars in the Local Group and found that the correlation extends down to dwarf galaxies with stellar mass as low as $10^3 M_\odot$.

In principle, measuring stellar metallicity in addition to gas-phase metallicity can break the degeneracy between inflow and outflow (e.g., Lu et al. 2015), resulting in better constraints on chemical evolution models. However, recent works that have attempted to incorporate both stellar and gas-phase MZR found that it is difficult to reconcile the two MZR. Lian et al. (2018) found that the stellar metallicities of local galaxies are generally lower than expected based on their gas-phase metallicities. The discrepancies are larger for lower mass galaxies. The only models that can reconcile both MZR have to invoke either a steep initial-mass function (IMF) slope (almost twice the slope of the Salpeter 1955 IMF) or a strong outflow (ejection of all metals produced) at early times.

Determining stellar metallicities in star-forming galaxies is challenging and subject to potentially large biases. First, emission lines from the interstellar medium have to be either subtracted or modeled together with stellar absorption lines. Second, even if the emission lines are modeled perfectly, the measurements of age and metallicity are still subject to bias, especially in the spectra of young stellar populations. The bias generally makes the population appear older and more metal poor than the true values (e.g., Leethochawalit et al. 2018) by as large as ~ 0.5 dex if the priors are not treated carefully (Ge et al. 2018; Cid Fernandes 2018).

Quiescent galaxies provide an alternative and a more convenient way to constrain galaxy chemical evolution models. The populations are older. The contamination from emission lines is less concerning. The spectral analysis is overall less prone to systematic biases. Additionally, the (near) absence of gas in quiescent galaxies simplifies chemical evolution models by eliminating the need to consider metals in the gas.

In this paper, we continue the work of Leethochawalit et al. (2018, hereafter Paper I) in quantifying the stellar MZR of quiescent galaxies as a function of redshift. We expand the sample size, from galaxies in a galaxy cluster at $z = 0.39$ to include an additional sample in a galaxy cluster at $z = 0.54$. We measure their magnesium [Mg/H] abundances in addition to iron [Fe/H] abundances. While Fe is produced by both Type II and Type Ia supernovae, Mg is an alpha element that is mainly produced by Type II supernovae. Mg therefore has a shorter recycling timescale than Fe. By measur-

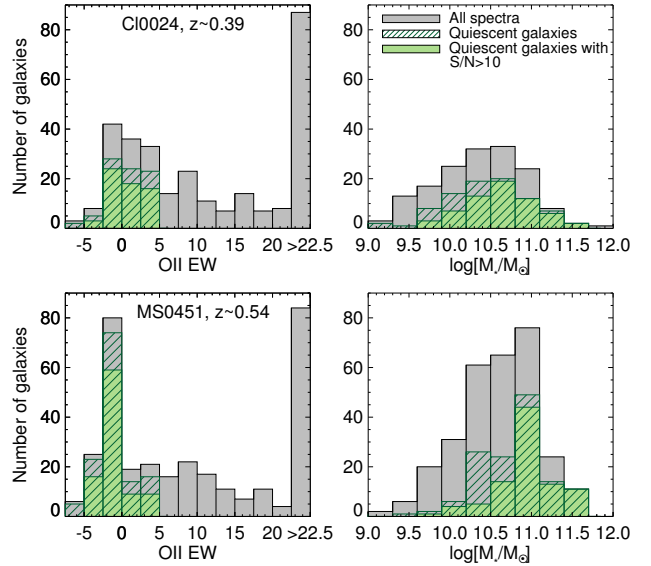


Figure 1. Histograms of the parent and selected sample in this paper. We select quiescent galaxies (defined in Sec. 2) with S/N greater than 10 \AA^{-1} in the rest frame to be in our sample (solid green).

ing the abundance of Mg, we now have an indication of the abundance of metals that is approximately instantaneously recycled, which can be used in simple chemical evolution models.

The highlight of this work is that we propose an archaeological method to constrain galactic outflow, in terms of average mass-loading factors that quiescent galaxies experienced as a function of mass, via the measurements of α element abundances. We build upon the technique introduced in Paper I to trace quiescent galaxies back to their epochs of formation using their ages. We conclude that while the [Fe/H] abundance at a fixed galaxy mass appears to evolve with redshift, the [Mg/H] MZR changes neither with the redshift of formation nor with the redshift of observation.

2. DATA

In Paper I, we presented a stellar MZR based on quiescent galaxies from the galaxy cluster Cl0024+17 at $z = 0.39$. In this paper, we expand the sample by including additional quiescent galaxies from the galaxy cluster MS0451 at $z = 0.54$.

The two galaxy clusters comprised the survey of Moran et al. (2007). The survey provides UV to near-infrared imaging and ground-based optical spectroscopy of member galaxies up to ~ 10 Mpc in diameter centered on both clusters. The available photometric bands are NUV and FUV from the Galaxy Evolution Explorer (GALEX) satellite, *BVRI* bands from the 3.6 m Canada-France-Hawaii telescope (Cl0024) and the

Subaru 8 m telescope (MS0451), F814W ($\sim I$ band) from Hubble Space Telescope, and J and K_s bands from the WIRC camera on the Palomar/Hale 200" telescope. Using the same method as in Paper I, we use this photometry to estimate the stellar masses of the sample with the SDSS KCORRECT software version v4.3 (Blanton & Roweis 2007).

The spectroscopic data for both Cl0024 and MS0451 were obtained with the DEIMOS spectrograph (Faber et al. 2003) on the Keck II Telescope. Most data are part of the original survey. We obtained additional DEIMOS spectroscopy to enhance the S/N of a subset of the galaxy spectra in MS0451. In both clusters, the spectra were obtained with 1" wide slits with a spectral resolution of at least $R = 2000$, spanning rest-frame wavelengths from ~ 3500 to 6000 \AA . The detailed spectroscopic observations of Cl0024 are described by Moran et al. (2005) and Paper I.

We now summarize the previous and new DEIMOS spectroscopy in MS0451. In 2003, Moran et al. (2005) observed 11 slitmasks for 1 hr each with the 600 line mm^{-1} grating centered at 7500 \AA . Based on these initial data, the study identified cluster members and performed a deeper follow-up in 2004–2005 with 10 additional slitmasks with the same grating centered at 6800 \AA . The integration time was 2.5–4 hrs per slitmask. In total, the original survey identified 319 member galaxies in MS0451. Moran et al. (2007) gives further details of these observations. In December 2016 and October 2017, we additionally observed one more slitmask with the same grating centered at 7200 \AA for a final follow-up. In the mask design, we prioritized previously observed quiescent galaxies whose spectra combined over all observing runs could achieve S/N of at least 8 \AA^{-1} within 5 additional hours of integration time. The actual integration time was 4.5 hrs.

All DEIMOS spectra were reduced using the spec2d DEIMOS data reduction pipeline (Newman et al. 2013) adapted by Kirby et al. (2015). Each spectrum was flat-fielded, wavelength-calibrated, sky-subtracted, and telluric-corrected. We did not flux-calibrate the spectra, as we are interested in continuum-normalized spectra.

We selected the final sample in a manner similar to Paper I with the following criteria. First, the member galaxies are quiescent, which we defined as having equivalent widths (EWs) of [O II] $\lambda 3727$ less than 5 \AA and either having rest-frame FUV – V colors larger than 3 or no detection in rest-frame FUV. Second, the signal to noise (S/N) is greater than 10 \AA^{-1} in rest-frame. This resulted in the final sample of 59 galaxies in Cl0024 and 92 galaxies in MS0451. The lowest masses

are $M_* = 10^{9.7}$ and $10^{9.6} M_\odot$, respectively (see Festari figure 1).

Finally, we include a subsample of $z \sim 0$ SDSS quiescent galaxies in our sample to compare with our observed galaxies. We use the same 155 randomly selected SDSS quiescent galaxies from Gallazzi et al.’s (2005) sample in the mass range 10^9 to $10^{11.5} M_\odot$ as in Paper I. To recapitulate, we put limits on broadband color and on the maximum $H\alpha$ EW so that the sample selection criteria are close to what we used for the higher redshift sample. We select roughly equal numbers of SDSS galaxies in each mass bin.

3. MODEL FITTING

We use full spectral fitting to derive ages and stellar metallicities. We develop our own fitting algorithm in this work, although spectral fitting algorithms capable of deriving stellar population properties of stellar population already exist. Recent examples of existing algorithms include FIREFLY (Wilkinson et al. 2017) and `alf` (Conroy et al. 2018). We use our own code because we are interested in the abundances of specific elements. `alf` is the only algorithm that is currently readily capable for this, but the Markov chain Monte Carlo method that it uses to find best-fit parameters is computationally expensive when applied to a large sample size. This is especially pertinent because we wanted to experiment using different spectral masks and continuum normalization techniques. Lastly, we demonstrate in Appendix A that our test results agree reasonably well (within 2σ) with those obtained using `alf`.

As in Leethochawalit et al. (2018), we adopt the single stellar population (SSP) models from the Flexible Stellar Population Synthesis (FSPS, Conroy et al. 2009) version 3.0. We generate the SSP spectra with the Kroupa (2001) IMF, Padova isochrones (Marigo & Girardi 2007), and the MILES spectral library (Sánchez-Blázquez et al. 2006). The models have metallicity and age ranges of $-1.98 < \log Z < 0.2$ and 0.3 Myr to 14 Gyr , respectively. Unlike the model used in Leethochawalit et al. (2018), here we include the enhancements of Mg in addition to [Fe/H]. To do so, we use the theoretical response functions from Conroy et al. (2018), which depend on metallicity and age. They were computed from the Kurucz suite of theoretical model atmospheres and spectra (Kurucz 1993).

We are interested in 3 parameters: age, [Fe/H], and [Mg/Fe]. We choose to measure [Mg/Fe] for two reasons. First, magnesium is one of the alpha elements that is mainly produced by Type II supernovae. Its recycling time can thus be approximated as instantaneous. Second, magnesium absorption features are dis-

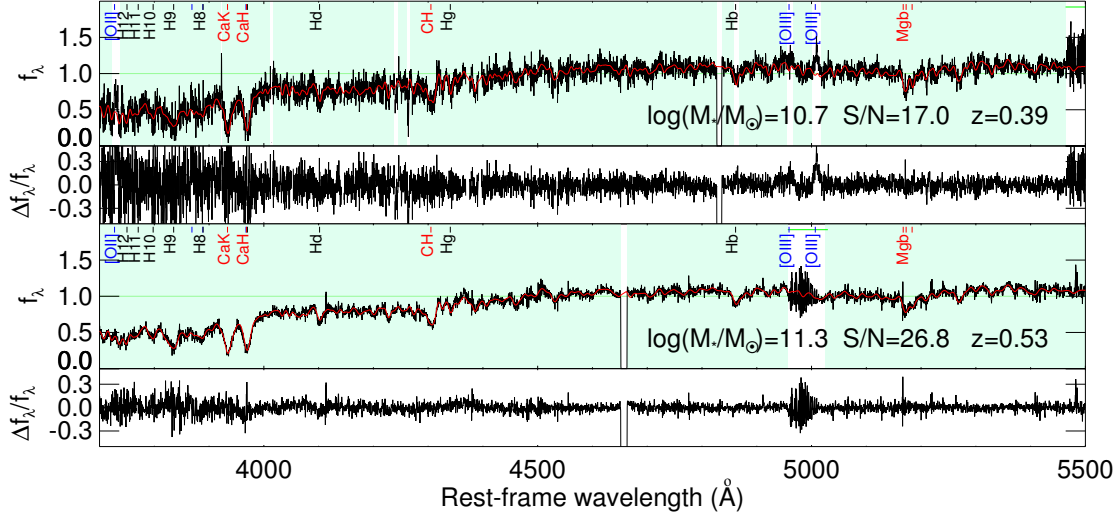


Figure 2. Examples of two observed spectra (black), best fit models (red), and the model residual (bottom panel of each spectrum). The teal background shows the spectral regions used for spectrum modeling while the white background shows the spectral regions that are masked out. The top panels are an example of the spectra at the lower end of the S/N cut from a member of Cl0024 cluster. The bottom panels show a high S/N spectrum from the MS0451 cluster. The best-fit parameters are $\text{Age} = 3.3_{-0.7}^{+1.4}$ and $4.3_{-0.9}^{+1.7}$ Gyr; $[\text{Fe}/\text{H}] = -0.0_{-0.7}^{+0.11}$ and $-0.05_{-0.06}^{+0.08}$ dex; $[\text{Mg}/\text{Fe}] = 0.39_{-0.11}^{+0.13}$ and $0.39_{-0.04}^{+0.15}$ dex, respectively.

tinct. Specifically, the Mg b absorption lines at 5170 Å minimally overlap with absorption features of other elements. This makes the measurement more reliable than other instantaneously recycled elements, especially in low-S/N spectra.

We assume that each galaxy is a single stellar population, i.e., all stars in the galaxy were born at the same time with the same metallicity. Thus, our obtained ages and metallicities are SSP-equivalent values. This means that the measured age is expected to be younger than both light- and mass-weighted ages derived in models with extended star-formation histories (Choi et al. 2014). The metallicities are less affected (Mentz et al. 2016).

In order to measure age, $[\text{Fe}/\text{H}]$, and $[\text{Mg}/\text{Fe}]$, we actually need to fit for 6 parameters, all of which influence the 3 parameters of interest: age, $[\text{Z}/\text{H}]$, velocity dispersion, redshift, $[\text{Mg}/\text{Fe}]$ and $[\text{N}/\text{Fe}]$. Although we are mainly interested in $[\text{Mg}/\text{Fe}]$, we experimented with a few options for the combinations of additional metal enhancements in Appendix A. In summary, we find that by fitting for two additional abundance ratios— $[\text{Mg}/\text{Fe}]$ and $[\text{N}/\text{Fe}]$ —we obtain the values of interest that are most consistent with literature and with the results from fitting with a more elaborate set of metals: $[\text{Mg}/\text{Fe}]$, $[\text{O}/\text{Fe}]$, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$, $[\text{Na}/\text{Fe}]$, $[\text{Si}/\text{Fe}]$, $[\text{Ca}/\text{Fe}]$, and $[\text{Ti}/\text{Fe}]$. This is likely because the flux of the wavelength within the range of $\sim 4000 - 4400$ Å responds simultaneously to the enhancements of Fe, N, and Mg.

We interpret the measured $[\text{Z}/\text{H}]$ as $[\text{Fe}/\text{H}]$ for the following reason. In our case, the measured $[\text{Z}/\text{H}]$ is the metallicity of the base SSP model (without Mg and N enhancement) and therefore not the actual total metallicity of the galaxy. Each base SSP model is the integrated stellar spectrum of stars that lie on the isochrone of a given $[\text{Z}/\text{H}]$ and age. However, the Padova isochrones assume solar-scaled abundances of individual metal elements ($[\text{Fe}/\text{H}] = [\text{Z}/\text{H}]$). In addition, the metallicities in the MILES stellar library were measured in terms of $[\text{Fe}/\text{H}]$. Thus, our measured $[\text{Z}/\text{H}]$ is best interpreted as $[\text{Fe}/\text{H}]$. Besides, based on results in Appendix A, our measured $[\text{Fe}/\text{H}]$ agrees well with $[\text{Fe}/\text{H}]$ measured from more complex models.

The fitting method is the same as that in Paper I. In summary, we first mask out the most prominent telluric band in the 7591–7703 Å wavelength range. In addition, if the spectrum has a positive EW of the $[\text{O II}]\lambda 3727$ emission line, we then also mask out the wavelength range of the $[\text{O II}]\lambda 3727$, $[\text{O III}]\lambda 5007$, 4959 and $\text{H}\beta$ line. We iteratively fit each spectrum with the Levenberg–Marquardt χ^2 -minimization method.

For each spectrum, we fit for at least 100 iterations with the IDL code MPFIT (Markwardt 2012) until the fitting parameters converge. In the first iteration, we fit the continuum-normalized observed spectrum with continuum-normalized model spectra. In the subsequent iterations, we do not alter or continuum-normalize the model spectra. Instead, we apply a synthesized continuum (a B-spline fit to the quotient of the

continuum-normalized spectrum and the best-fit SSP model spectrum from the previous iteration) to the observed continuum-normalized spectrum. We then fit the resulting spectrum with unaltered model spectra. This algorithm avoids having to separately determine the continuum for the observed and model spectra. The method was described in full detail in Paper I. Lastly, to avoid the convergence of parameters on local minima, we fit each spectrum at least 5 times with different initial parameters, and we adopt the results from the models with the least χ^2 .

Finally, we convert the measured enhancement of the [Mg/Fe] from the base FSPS spectra to the actual [Mg/Fe]. We followed the method in Conroy et al. (2018) by adding the abundance pattern of the MILES library stars at a given [Fe/H] to the measured Mg enhancement to obtain the final [Mg/Fe]. Because the [Mg/Fe] pattern of the stellar library in the measured [Fe/H] range only varies within ~ 0.05 dex, this step does not significantly affect the reported value of [Mg/Fe].

As demonstrated in Paper I, the statistical uncertainties obtained from MPFIT underestimate the total uncertainties. Likewise, here we also calculate the systematic uncertainties based on the degeneracies between the parameters of interest and take those as our uncertainties. We first generate a mock spectrum with the measured [Fe/H], [Mg/Fe], and age. We then smooth the spectrum to a fixed velocity dispersion of 250 km/s, convolve it with the observed spectral resolution, and add Gaussian noise to reach the same S/N of the observed spectrum. We thereafter compare the noised spectrum to a 3-dimensional grid of noise-less SSPs (in age, [Fe/H] and [Mg/Fe]), and calculate the uncertainties based on the χ^2 grid.¹

We explore how the uncertainties of the three parameters correlate with each other at a given S/N ratio in Appendix B. In summary, the uncertainties in [Fe/H] and age mostly originate from the highly-correlated degeneracy between age and [Fe/H]. There is also a positive correlation between the uncertainties in [Fe/H] and in [Mg/Fe]. These result in average uncertainties of ~ 0.2 dex at $S/N=10\text{\AA}^{-1}$ and ~ 0.1 dex at $S/N=25\text{\AA}^{-1}$ in age and [Fe/H]; both can be higher at lower [Fe/H] values. The uncertainties of [Mg/Fe] are $\sim 0.2 - 0.3$ at $S/N=10\text{\AA}^{-1}$, and $\sim 0.1 - 0.2$ dex at 25\AA^{-1} .

¹ Each noise-less SSP grid contains $51 \times 51 \times 41$ spectra, covering the range of $[-1.0, 2.5]$ dex in [Fe/H], $[0.1, 12]$ Gyr in age, and $[-0.4, 0.8]$ dex in [Mg/Fe], respectively. The reason for fixing the velocity dispersion at 250 km/s is therefore to keep the size of the SSP grids manageable.

Examples of best-fit models of the spectra at the lower and upper end of the S/N range are shown in Figure 2. We list all measured parameters in Table 1.

4. RESULTS

4.1. The MZR measured from [Fe/H] evolves with redshift

We show the relation between [Fe/H] and galaxy stellar mass in Figure 3. Visually, the galaxies at higher redshift (Cl0024 and MS0451) appear to have lower [Fe/H] abundances than those of local SDSS galaxies, especially in the lower mass range. A small proportion of data points appear to have very low [Fe/H] or [Mg/H] values. These are mostly low S/Ns spectra and younger populations. In the left panel, there are 17 data points whose [Fe/H]’s are smaller than 0.15 dex below the best-fit linear relations. Nevertheless, only eight of them actually have [Fe/H]’s that are inconsistent with the best-fit lines within 2σ of uncertainty. These galaxies are mostly younger than 2 Gyrs, which is the age range that is prone to increased bias (e.g., Ge et al. 2018; Cid Fernandes 2018). As for the right panel, there are 5 galaxies whose 2σ upper limits of [Mg/H] are not consistent with the best-fit line (but all are consistent within 3σ). Interestingly, these galaxies do not overlap with the galaxies with very low [Fe/H].

Next, we use the analysis of covariance to quantify the dependence of MZR on redshift. We first assume that both normalization and slope of the mass-[Fe/H] relation depend on redshift. We fit data from every redshift at once with a single linear multiple regression, where [Fe/H] depends on one quantitative variable (mass) and categorical variables (different redshift samples). We choose to fit with a linear function because we found in Paper I that both a two-degree polynomial and a logarithmic function introduce degrees of freedom that are not justified by the data.

The first model we consider (model 1), allows interactions² between mass and redshift. This means that *normalizations* and *slopes* of the MZRs are allowed to depend on redshift. The model is

$$E([\text{Fe}/\text{H}] | M_{10}, \text{sample}) = \beta_0 + \beta_M M_{10} + \beta_{cl} cl + \beta_{ms} ms + \beta_{cl \cdot M} (cl \cdot M_{10}) + \beta_{ms \cdot M} (ms \cdot M_{10}) \quad (1)$$

cl (or ms) is an indicator or a categorical variable with a value of 1 if the data point belongs to the Cl0024 (or MS0451) sample, and 0 otherwise. M_{10} is the galaxy

² “Interaction” is a term in statistics that describes a situation where the effect of one independent variable depends on the magnitude of the other independent variable. In this case, we allow the evolution of metallicity with redshift to be mass-dependent.

Table 1. Catalog of Measured Age and Metallicities

No.	RA	DEC	$\log(M_*/M_\odot)$	Age (Gyr)	[Fe/H]	[Mg/Fe]	η
Cl0024+17							
1	00 25 51.07	+17 08 42.4	10.8	$2.2^{+0.4}_{-0.2}$	$+0.19^{+0.09}_{-0.13}$	$+0.02^{+0.15}_{-0.13}$	$0.46^{+0.83}_{-0.17}$
2	00 25 54.52	+17 16 26.4	10.4	$3.0^{+1.9}_{-1.1}$	$-0.30^{+0.24}_{-0.16}$	$+0.25^{+0.20}_{-0.22}$	$1.25^{+1.97}_{-0.78}$
3	00 25 57.73	+17 08 01.5	10.4	$3.5^{+2.4}_{-1.0}$	$+0.06^{+0.12}_{-0.09}$	$+0.04^{+0.16}_{-0.18}$	$0.75^{+1.11}_{-0.36}$
4	00 26 04.30	+17 18 46.5	10.3	$2.2^{+0.9}_{-0.7}$	$-0.42^{+0.23}_{-0.14}$	$+0.08^{+0.32}_{-0.30}$	$3.00^{+3.66}_{-2.07}$
5	00 26 04.44	+17 20 00.6	10.5	$2.9^{+1.6}_{-1.4}$	$-0.21^{+0.26}_{-0.14}$	$+0.56^{+0.21}_{-0.22}$	$0.19^{+0.77}_{-0.19}$
...							
MS0451							
1	04 53 18.17	-02 58 57.7	11.1	$3.4^{+2.6}_{-1.0}$	$+0.05^{+0.13}_{-0.09}$	$+0.56^{+0.16}_{-0.15}$	N/A
2	04 53 33.42	-02 56 23.5	10.9	$1.9^{+0.3}_{-0.3}$	$-0.07^{+0.06}_{-0.12}$	$+0.49^{+0.16}_{-0.15}$	$0.08^{+0.56}_{-0.08}$
3	04 53 36.54	-03 04 13.5	11.3	$4.8^{+2.0}_{-1.4}$	$-0.06^{+0.09}_{-0.06}$	$+0.50^{+0.11}_{-0.11}$	$0.06^{+0.40}_{-0.06}$
4	04 53 38.64	-02 54 11.2	10.8	$3.0^{+1.4}_{-0.7}$	$-0.08^{+0.21}_{-0.13}$	$+0.13^{+0.15}_{-0.18}$	$0.92^{+1.36}_{-0.59}$
5	04 53 52.61	-02 55 31.5	10.8	$1.9^{+0.4}_{-0.3}$	$-0.35^{+0.08}_{-0.23}$	$+0.42^{+0.20}_{-0.16}$	$0.84^{+1.35}_{-0.22}$
...							

NOTE—Full table is available online in machine-readable form. Note that the [Fe/H]’s listed here are different from those in Paper I, which were calculated by masking out the Mg b lines and assuming the solar abundance ratios. See Appendix A for the comparison between the two results. The mass-loading factors, η , are calculated based on Equation 7 using the fiducial yield and return fraction. Where η is entered as N/A, the [Mg/H] is larger than the adopted yield - see Section 5.3 for a more detailed discussion.

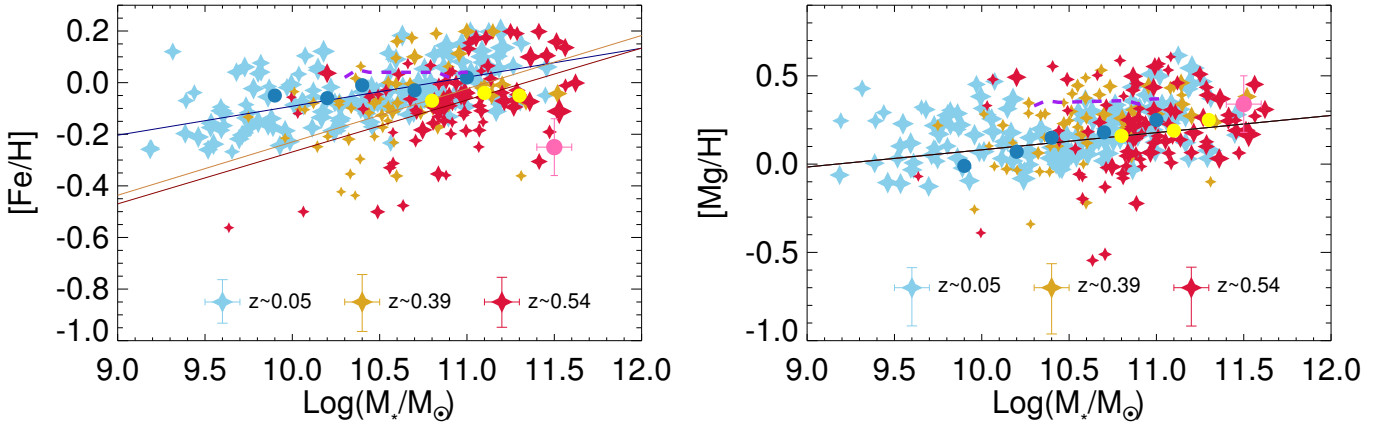


Figure 3. Stellar mass–metallicity relations measured from [Fe/H] (left) and [Mg/H] (right). Measurements in this paper are in diamonds. The labels indicate different redshift samples and their median uncertainties. The size of the diamonds corresponds to the S/N of the observed spectra, ranging from 10 \AA^{-1} (smallest) to 45 \AA^{-1} (largest). Best-fit linear models from Table 2 are plotted in solid lines (model 1 for [Fe/H] and model 3 for [Mg/H]). Measurements based on stacked spectra of similar redshifts by Choi et al. (2014) are in dark blue and yellow circles (for $z = 0.1\text{--}0.2$ and $z = 0.4\text{--}0.55$ redshift bins, respectively). The pink circle represents a measurement of a $z \sim 2.1$ quiescent galaxy from Kriek et al. (2016). Magenta dashed lines are predictions based on the $z \sim 0$ quiescent galaxies with $10^{10.2} \lesssim M_*/M_\odot \lesssim 10^{11}$ in the IllustrisTNG simulation (Naiman et al. 2018). We found a significant evolution with redshift in the [Fe/H]–mass relation, but not in the [Mg/H]–mass relation.

stellar mass, defined as $M_{10} = \log[M_*/10^{10}M_\odot]$ so that the normalization is at $10^{10}M_\odot$. To digest this equation, we can view that β_0 is the normalization and β_M is the slope of the MZR for the $z \sim 0$ SDSS population; we call this the baseline MZR. The rest of the parameters are the evolution terms. β_{cl} is the change in the normalization, while $\beta_{cl.M}$ is the change in the slope from those of local galaxies to those of $z \sim 0.4$ Cl0024 galaxies. β_{ms} and $\beta_{ms.M}$ can be interpreted similarly.

We use the Monte Carlo (MC) technique with 1000 iterations to find the best-fit β_i parameters. In each iteration, we draw all samples randomly according to the measured probability distribution of [Fe/H]. We use the IDL code `regress` to find the best-fit parameters from each iteration. The final best-fit parameters (the mean and standard deviation) are listed as model 1 in Table 2.

Based on model 1, the MZRs of the higher redshift samples are likely steeper than that of local galaxies. The $\beta_{cl.M}$ and $\beta_{ms.M}$ terms suggest that the mean change in [Fe/H] for a 1 dex increase in stellar mass is ~ 0.9 dex more for higher redshift galaxies than for local galaxies. However, their significance values are only $2-3\sigma$.

We further formally test the necessity of using the interaction model (model 1) over a simpler model (model 2) without the interactive terms, i.e., the model where the slopes are fixed to the same value regardless of redshift:

$$E([\text{Fe}/\text{H}] | M_{10}, \text{sample}) = \beta_0 + \beta_M M_{10} + \beta_{cl} cl + \beta_{ms} ms \quad (2)$$

The best-fit parameters are also listed in Table 2. The F-test for the R^2 change between the two models suggests that we can reject the null hypothesis that the simpler model 2 is sufficient to describe the data with a p -value of 0.01 (also $> 2\sigma$). Despite these borderline values of $2-3\sigma$, for [Fe/H], we conclude that the interaction model (model 1) performs better than the simpler model 2 and we will work only with model 1 for further interpretations.

We found that the normalizations for the MZRs of the two higher redshift samples are statistically significantly lower ($> 3\sigma$) than the normalization of the local SDSS galaxies. With the more detailed spectral fitting model and larger sample size in this work, we confirm the finding in Paper I that there is an evolution in the MZR when the metal indicator is [Fe/H]. The best estimate of normalizations at $10^{10}M_\odot$ for the $z \sim 0.39$ and the $z \sim 0.54$ samples are ~ 0.14 and 0.18 ± 0.05 dex lower than the local samples. These values translate to the evolution of 0.04 ± 0.01 dex per observed Gyr, consistent with what we reported in Paper I.

In conclusion, we found that model 1 is the most appropriate to describe [Fe/H] evolution. We observe an evolution in [Fe/H] in both slope (likely at $2 - 3\sigma$) and normalization ($> 3\sigma$). This means that the amount of the evolution with observed redshift depends on mass. Over the same redshift range, lower mass quiescent galaxies evolve more strongly in [Fe/H] at fixed mass than higher mass quiescent galaxies. Further interpretation of [Fe/H] in Section 5 will be mainly based on the results from this model.

4.2. The MZR measured from [Mg/H] does not evolve with observed redshift

Now, we look at the evolution of the relation between [Mg/H] and M_* . First, we repeat the fitting with model 1 but substitute [Fe/H] with [Mg/H]. We found that the mass-[Mg/H] relations of each redshift sample have the same slope within the $\sim 1\sigma$ uncertainties. The best-fit MZRs at different redshifts are roughly parallel. This suggests that the simpler model 2, in which slopes are fixed to a common value, is more appropriate to describe the mass-[Mg/H] relation than the more complex model 1.

We continue to fit with model 2 and also found no significant evolution in the normalization of the MZR with observed redshift. As shown in Table 2, the values of β_{cl} and β_{ms} in model 2 are both consistent with zero. We found no significant evolution in [Mg/H] with observed redshift.

The results from the first two models suggest that we should proceed to fit a simple linear equation without redshift-dependent parameters (model 3). The parameters are also listed in Table 2. This model is most appropriate for the mass-[Mg/H] relation.

In summary, we did not detect an evolution of the stellar mass-[Mg/H] relation with redshift, neither in terms of slope nor normalization. This is in contrast to the $> 3\sigma$ detected in the evolution of the stellar mass-[Fe/H] relation with redshift. We can explicitly separate the best-fit parameters for both Fe and Mg abundances into the predicted MZRs at each redshift. The MZRs based on the measurements in this work are the following:

$$\begin{aligned} [\text{Fe}/\text{H}](M_{10}, z \sim 0) &= (-0.09 \pm 0.01) + (0.11 \pm 0.02)M_{10} \\ [\text{Fe}/\text{H}](M_{10}, z \sim 0.39) &= (-0.23 \pm 0.05) + (0.20 \pm 0.06)M_{10} \\ [\text{Fe}/\text{H}](M_{10}, z \sim 0.54) &= (-0.27 \pm 0.05) + (0.20 \pm 0.04)M_{10} \\ [\text{Mg}/\text{H}](M_{10}) &= (+0.08 \pm 0.02) + (0.10 \pm 0.02)M_{10} \end{aligned} \quad (3)$$

where $M_{10} = \log[M_*/10^{10}M_\odot]$.

4.3. Intrinsic Scatter

Table 2. Model parameters for MZR

model	parameter/description	[Fe/H] vs mass			[Mg/H] vs mass			
		value	t	sig.	value	t	sig.	
1	β_0	constant	-0.09 ± 0.01	7.9	<0.0005	0.09 ± 0.02	4.5	< 0.0005
	β_M	mass	0.11 ± 0.02	7.2	<0.0005	0.10 ± 0.03	3.8	< 0.0005
	β_{cl}	cl0024	-0.14 ± 0.05	2.9	0.002	-0.06 ± 0.07	0.8	0.2
	β_{ms}	ms0451	-0.18 ± 0.05	3.9	<0.0005	-0.01 ± 0.07	0.2	0.4
	β_{cl-M}	cl0024*mass	0.09 ± 0.06	1.6	0.05	0.12 ± 0.09	1.4	0.08
	β_{ms-M}	ms0451*mass	0.09 ± 0.04	2.0	0.02	-0.03 ± 0.07	0.4	0.4
2	β_0	constant	-0.10 ± 0.01	9.3	< 0.0005	0.09 ± 0.02	4.2	< 0.0005
	β_M	mass	0.14 ± 0.02	9.3	< 0.0005	0.11 ± 0.02	4.5	< 0.0005
	β_{cl}	cl0024	-0.09 ± 0.02	3.9	< 0.0005	0.02 ± 0.03	0.5	0.6
	β_{ms}	ms0451	-0.11 ± 0.02	6.4	< 0.0005	-0.04 ± 0.03	1.4	0.2
3	β_0	constant				0.08 ± 0.02	4.2	< 0.0005
	β_M	mass				0.10 ± 0.02	4.4	< 0.0005
Is model 1 necessary over model 2?			likely		0.01	no		0.3
Is model 2 necessary over model 3?						no		0.2

NOTE—The significance values (sig.) are the p-values from one-tailed t tests. The values will be twice for the two-tailed test. We use model 1 for [Fe/H] and model 3 for [Mg/H] relations with mass for further discussion.

We measured the intrinsic scatter (σ_Z) in the [Fe/H] and [Mg/H] relations with mass (Figure 3). We used a method similar to that in Paper I. In short, we convolved a normalized Gaussian with size σ_Z to the existing probability of each data point. The measured intrinsic scatter is the σ_Z that maximizes the sum of the resulting log-likelihood evaluated at the predicted [Fe/H] (or [Mg/H]) according to the relations in Equation 3. The uncertainties of the intrinsic scatter are derived via the jackknife resampling technique.

The measured intrinsic scatter in [Fe/H] is consistent with what we found in Paper I: 0.06 ± 0.01 dex, compared to 0.07 ± 0.01 dex in Paper I. The nominal intrinsic scatter in [Mg/H] is smaller, but it has a large uncertainty: 0.05 ± 0.17 dex, which is consistent with both zero and with the intrinsic scatter of [Fe/H]. As a test, if we remove the two data points whose [Mg/H] is below -0.4 dex, the measured σ_Z in [Mg/H] becomes 0.03 ± 0.05 dex, which is still consistent with both zero and the intrinsic scatter of [Fe/H]. If we expect that the true intrinsic scatter is non-zero, our result, by being consistent with zero, indicates that we may have overestimated the uncertainties in [Mg/H]. In addition, if the true intrinsic scatter in [Mg/H] is small (e.g., less than 0.08 dex), we may have a small number of “outliers” in the data for which we underestimated the uncertainties.

In sum, our results indicate that the intrinsic scatter in [Mg/H] is likely smaller than the intrinsic scatter in [Fe/H]. However, we refrain from any further interpretation due to the large uncertainties in our measurements.

5. DISCUSSION

In this section, we will interpret the measured MZR and their evolution with redshift, or lack thereof, using galactic chemical evolution models. Specifically, we use the measured MZR to constrain average galactic outflows in terms of the mass-loading factor.

5.1. Comparison to galaxy simulations and semi-analytic models

The relatively shallow slopes of the MZR estimated from our samples are consistent with other observations that specifically sample quiescent galaxies. Our slopes for both [Fe/H] and [Mg/H] are consistent within 2σ with the slopes derived from simple linear fits to the measurements by Choi et al. (2014), which were based on stacked spectra of field quiescent galaxies at similar redshifts. They are also consistent with the slopes for the quiescent galaxies reported by Gallazzi et al. (2014), in which metallicities (Z) were measured from a combination of H, Fe, and Mg absorption lines.

Gallazzi et al. (2014) found that the slopes of the MZR of quiescent galaxies are shallower than those of

star-forming galaxies.³ Do theoretical models reproduce the shallower slope?

Several semi-analytical (SA) and hydrodynamical simulation models have predicted stellar MZR and their evolution with redshift (e.g. Lu et al. 2014; Guo et al. 2016; Ma et al. 2016; Taylor & Kobayashi 2016). However, most results from simulations do not report metallicities based on galaxy star formation properties. The predictions are for the total populations, which are mixtures of star-forming and quiescent galaxies, depending on the passive fraction at each redshift.

It is therefore important to be cautious of interpreting the differences between observations and simulations. In Paper I, we compared the slope of [Fe/H] measured in quiescent galaxies to the slopes predicted in SA models in Lu et al. (2014) at face value. We argued that the observational results agree the most with the model in which the outflow mass-loading factor from Type II SNe is independent of mass. However, the galaxies simulated by Lu et al. have gas fractions on average ranging from 20% in the highest mass bin to 50% in the lowest mass bin at $z \sim 0$. In other words, portions of the simulated galaxies are still forming stars, which is not the case for our observed samples.

One study that specifically predicts the MZR of quiescent galaxies is by Naiman et al. (2018), using the IllustrisTNG suite of simulations. The median of their simulated stellar abundances at $z \sim 0$ is plotted as magenta lines in Figure 3. We calculated the simulated [Mg/H] based on their median [Fe/H] and [Mg/Fe]. The slope in the simulated [Fe/H] is flatter than the observations. The normalization is consistent with our observations at $\sim 10^{11} M_{\odot}$ but is larger at lower masses. The relatively flat slope in the simulations was already explained by Naiman et al. as a result of the sampling criteria. Specifically, they excluded smaller mass galaxies in their sample. If those were included, they reported that the predicted mass–[Fe/H] relation could have been steeper. For [Mg/H], the simulated slope is ~ 0.04 dex per $\log(\text{mass})$, which is roughly half of what we observed. Because we calculated [Mg/H] from [Mg/Fe] and [Fe/H], it is likely that the cause of the difference in [Mg/H] slopes between the observations and simulations is the same as for [Fe/H].

5.2. Analytic chemical evolution model

As an alternative to more complex simulations, we can schematically understand the stellar metallicities of quiescent galaxies using an analytic galactic chemical evolution model.

In this paper, we present a simple model that connects the metal mass in a quiescent galaxy to the average outflow it experienced. The model is based on the work of Lu et al. (2015). In this model, the assumptions are

1. Star formation occurs in the interstellar medium (ISM) that is perfectly mixed.
2. Metals are instantaneously recycled.
3. Outflows and inflows are permitted but the inflow do not contribute significantly to the total metal budget.⁴

We start by tracking the change in the total metal mass in the ISM ($dM_{Z,g}$) as follows:

$$dM_{Z,g} = ydM_* - Z_g dM_* - \frac{\eta}{1-R} Z_g dM_* \quad (4)$$

where M_* is the mass of long-lived stars. The first term on the right side of Equation 4 represents the metal mass that is newly produced and returned to the ISM by a generation of forming stars. y is the chemical yield, defined as the mass of metals returned to the ISM per mass turned into low-mass stars and remnants (dM_*). The second term is the metal mass in the ISM that is locked into stars. The last term is the metal mass that is lost to outflows, which is equal to the mass outflow rate times gas-phase metallicity Z_g . The outflow rate is parameterized in terms of the mass-loading factor (η), defined as the ratio of the mass outflow rate to the star formation rate (SFR). In this equation, the SFR is written in terms of dM_* and the return mass fraction R , which is the fraction of the mass of a stellar generation that returns—with its original composition—to the ISM from short-lived stars and stellar winds.

With the assumption that the ISM is well mixed, the change in the total metal mass locked in long-lived stars during the time interval in which a mass dM_* of long-lived stars formed is $dM_{Z,*} = Z_g dM_*$. We can substitute this to the last two terms of Equation 4 to get

$$dM_{Z,g} = ydM_* - \left(1 + \frac{\eta}{1-R}\right) dM_{Z,*} \quad (5)$$

We then integrate this equation, divide both sides by M_* , and rearrange the terms to get a description for the stellar metallicity of a galaxy:

$$Z_* = \frac{y - r_g Z_g}{1 + \frac{\langle \eta \rangle}{1-R}} \quad (6)$$

⁴ We also derive an estimate for stellar metallicity when the inflow is enriched and when the outflow has different metallicity than the ISM, as well as evaluate these effects on our results in Appendix C.

³ This is however not true for Local Group dwarf galaxies (Kirby et al. 2013).

r_g is the cold gas fraction M_g/M_* . We adopt the definition of gas- and stellar-phase metallicity as $Z_g = M_{Z,g}/M_g$ and $Z_* = M_{Z,*}/M_*$. Therefore, in this equation, Z_g is the current gas-phase metallicity and Z_* is already an average metallicity, specifically the mass-weighted metallicity of the galaxy. Note that during the integration, we consider y and R as constant. $\langle\eta\rangle$ can be viewed as the average mass-loading factor weighted by $dM_{Z,*}$, which means that it is most sensitive to feedback at the galaxy's peak of star formation.

Now, we consider stellar metallicities of quiescent galaxies based on Equation 6. We can see that at the limit where there is no gas left ($r_g = 0$), the second term in the numerator is zero and the stellar metallicity is a function of the mass-loading factor. This situation is appropriate to describe most quiescent galaxies.

We estimate that the product of gas fraction and gas-phase metallicity ($r_g Z_g$) in quiescent galaxies is at most 10% of the yield. First, quiescent galaxies have low gas fractions. In the local universe, early-type galaxies typically have molecular gas fractions less than a percent (Boselli et al. 2014). Galaxies at higher redshifts have higher molecular gas fractions. However, even at $z \sim 1.8$, an average early-type galaxy has a gas fraction that is less than 10% (Gobat et al. 2018). Second, quiescent galaxies do not have particularly high gas-phase metallicities. Griffith et al. (2019) measured gas-phase metallicities in 3 quiescent galaxies and found that their Z_g is roughly consistent with their stellar metallicities. This suggests that we can assume quiescent galaxies to have gas fractions r_g less than 10% and gas-phase metallicities less than the yield, y . Therefore, the $r_g Z_g$ term for quiescent galaxies in Equation 6 is at least 1 dex smaller than the yield term.

Considering the discussion above, we estimate the stellar metallicities of quiescent galaxies to be

$$Z_{*,\text{quiescent}} \approx \frac{y}{1 + \frac{\langle\eta\rangle}{1-R}} \quad (7)$$

The yield and return fraction are fundamentally properties of stars. Though they can depend on IMF, they are often treated as constant and not a function of galaxy mass (e.g. Pagel 1997). The only other variable in Equation 7 is $\langle\eta\rangle$ (hereafter referred to simply as η). This means that a non-zero slope of the MZR of quiescent galaxies implies that the mass-loading factor is a function of galaxy mass.

5.3. Constraints on the mass-loading factor

We can now convert measured stellar metallicities of quiescent galaxies into mass-loading factors via Equation 7. Because our model assumes instantaneous re-

cycling, we apply our model to the observed Mg abundances. Mg is an α element produced in core-collapse supernovae and is appropriate for the assumption of instantaneous recycling. In addition, comparing to other α elements, it tracks oxygen the closest (Conroy et al. 2014).

We estimate the mass ratio $M_{Mg,*}/M_*$ from the measured [Mg/H] using the solar abundance of Mg from Asplund et al. (2009) and the mass number of Mg (24) as the average Mg to H atomic weight in stars. The yield is set to 3 times the solar Mg abundance (Nomoto et al. 2006). Following Lu et al. (2015), we set the return fraction to $R = 0.46$.

We found that mass-loading factor is a power-law function of galaxy mass. The result is plotted in Figure 4, where we convert each measurement of [Mg/H] to a mass-loading factor. We apply the MC technique to the probability distribution of [Mg/H] directly to obtain the best linear fit to the relation between $\log \eta$ and \log mass:

$$\log \eta = (-0.21 \pm 0.09)M_{10} - (0.26 \pm 0.07) \quad (8)$$

The mass-loading factor scales with galaxy stellar mass, $\eta \propto M_*^{-0.21 \pm 0.09}$. A galaxy that is 10 times less massive would have approximately 1.6 times larger outflow rate per SFR.

We note that in fitting the linear relation between $\log \eta$ and \log mass, we follow the convention of using fixed values for the yield and the return mass fraction. We exclude 18 galaxies whose best estimated [Mg/H]'s are larger than the adopted yield. Although the yield is found to be quite independent of metallicity, it can vary with different IMFs and the upper mass cutoffs of the IMF (Vincenzo et al. 2016). If we instead measure η with an extremely high yield (6 times the solar abundance of Mg), we do not have to exclude those 18 galaxies. The resulting power-law index would be lower ($\eta \propto M_*^{-0.13 \pm 0.05}$), but is still consistent with the current estimate. The normalization would be ~ 0.5 dex higher.

We also experiment on using a different value of return mass fraction. We reduce the return mass fraction to $R = 0.23$ to be roughly consistent with the return mass fraction when the Salpeter (1955) IMF is adopted (Vincenzo et al. 2016). The power-law index does not change (0.21 ± 0.09) but the normalization drops to -0.11 ± 0.07 dex. In short, choices of the yield and the return mass fraction do not seem to change the resulting power-law index significantly, but they can change the normalization by a factor of a few.

We further explore the massive galaxies in Figure 4 with very low mass-loading factors. There are a total

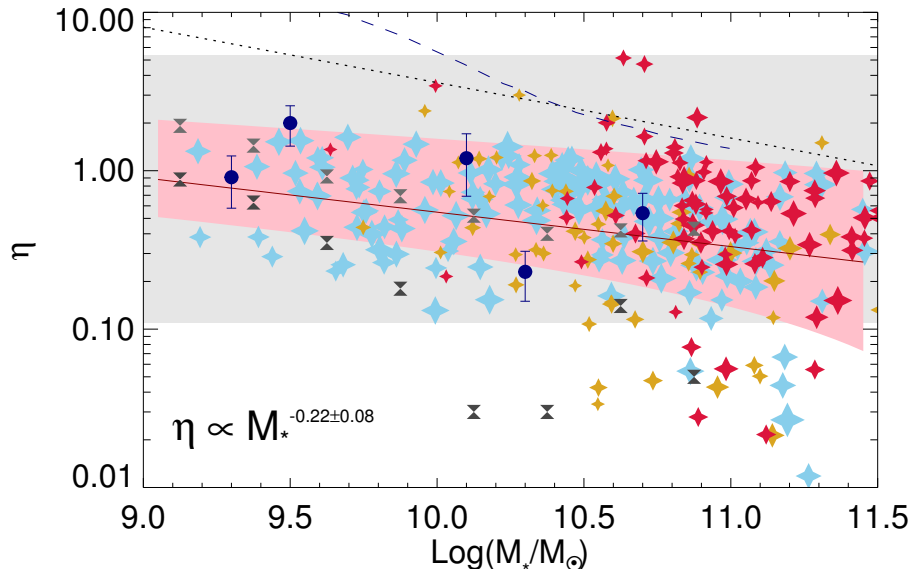


Figure 4. The relation between mass-loading factor (η) and galaxy stellar mass. Star-shaped symbols show η measured from our quiescent galaxy sample, color coded by redshift as in Figure 3. The median for the uncertainties in the estimated η is 0.1 dex. The line is the linear fit to individual measurements of η . Pink shading represents the best-fit and the scatter in the mass-[Mg/H] relation in Figure 3. The gray shaded line and navy dashed line are the estimations based on gas-phase metallicity from Spitoni et al. (2010) and Lu et al. (2015) (upper limits), respectively. The gray hourglasses are the estimations from gas-phase metallicity gradients of nearby star-forming galaxies. The light and dark gray indicate the results using different metallicity calibrations (Belfiore et al. 2019). Navy circles are the mass-loading factors measured from UV absorption lines (Chisholm et al. 2017). Results from the hydrodynamical simulation of Muratov et al. (2015) are shown as the dotted line.

of 13 galaxies whose measured mass-loading factors lie more than 0.8 dex below the best-fit line. They almost appear to be closed-box systems since, using the adopted yield values, we estimate their mass-loading factors to be less than 5%. These galaxies are generally massive, $M_* > 10^{10.5} M_\odot$. Their mean [Fe/H] (0.03 ± 0.11 dex) and age (3.8 ± 2.1 Gyr) are not significantly different from those of other galaxies also with $M_* > 10^{10.5} M_\odot$ (whose mean [Fe/H] and age are -0.04 ± 0.12 dex and 5.1 ± 2.6 Gyr respectively). However, they are more [Mg/Fe] enhanced (at the mean [Mg/Fe] of 0.42 ± 0.11 dex for those galaxies as compared to the mean of 0.23 ± 0.14 dex for other galaxies). This suggests that they had shorter formation timescales. Moreover, they appear to be near the center of the clusters; most are at less than half the virial radius in projection. It may be interesting to subsequently study in detail the star-formation histories of these galaxies.

5.4. Comparing the measured mass-loading factors with other measurements and with theoretical predictions

We proposed an alternative way to constrain the mass-loading factor using the metallicities of quiescent galaxies. Our results, using the fiducial yield and the fidu-

cial return mass fraction, agree reasonably well with other works that use chemical evolution models for star-forming galaxies. In Figure 4, we plot as a blue dashed line the constraints from Lu et al. (2015), whose model is very similar to ours (and actually was the model that inspired our work). Their results are upper limits because of the uncertainties in gas fractions, different gas-phase metallicity calibrations, and a conservative choice of yield (5 times solar). Our estimates of η are consistent with the upper limit. We also plot an earlier estimate from Spitoni et al. (2010) (in gray shading). Their estimate is based on a chemical evolution model of the gas-phase MZR that allows both infall and outflow.

Recently, Belfiore et al. (2019) applied chemical evolution models to gas-phase metallicity gradients measured from nearby star-forming galaxies in the MANGA survey (Belfiore et al. 2017; Bundy et al. 2015). Their favored model assumes the inside-out growth formalism and that the star formation efficiency is inversely proportional to the orbital timescale. We plot their mass-loading factors as gray hourglass data points in Figure 4. The light-gray hourglasses are obtained by adopting the Pettini & Pagel (2004) O3N2 metallicity calibration, while the dark-gray hourglasses are obtained by adopting the Maiolino et al. (2008)

R23 calibration. Our measured mass-loading factors are more consistent with the results using the former metallicity calibration. Their slopes (-0.58 ± 0.20 and -0.29 ± 0.07 for the two calibrations respectively) are nominally steeper than our estimation but are consistent with ours within uncertainties.

Our estimates for the mass-loading factors are also consistent with mass loading factors measured in individual galaxies currently driving outflows. Another way to constrain mass-loading factor is to directly observe outflows, via emission lines from the outflow gas or absorption features by the outflow material seen in galaxy or background quasar spectra. From these features, outflow velocities can be determined quite directly. However translating them to mass outflow rate requires a number of assumptions on the ionization structure and the geometry of the outflow. The results from applying these techniques to local star-forming galaxies are still limited in terms of sample size. Nonetheless, they suggest a range in mass-loading factor from 0.1 to several tens (e.g., Bolatto et al. 2013; Bouché et al. 2012). In figure 4, we plot in navy circles the results based on UV absorption lines from Chisholm et al. (2017), in which stellar mass measurements were readily available. Those results agree with ours not only in terms of slope but also normalization. In other words, our archaeologically derived mass-loading factors, which are rather of time-average quantities, are consistent with the instantaneous mass-loading factors that are measured in star-forming galaxies.

Given the simplicity of our chemical evolution model, the estimated power-law index of the relation between η and M_* agrees surprisingly well with both analytic and hydrodynamical simulations. Hayward & Hopkins (2017) presented an analytic model for galactic outflows driven by stellar feedback. In the model, the SFR is self-regulated. It is set by an equilibrium between the momentum injection rate from stellar feedback and the dissipation rate by turbulence and occasional outflows. The outflow occurs when the momentum accumulated within a coherence time is large enough given the gas density of the ISM patch. Their model predicts that the mass-loading factor scales with gas fraction and stellar mass. For the stellar mass component, their predicted scaling relation is $\eta \propto M_*^{-0.23}$, which is consistent with our observation.

However, the observed exponent of the scaling relation is smaller than predicted in an earlier semi-analytical model by Lagos et al. (2013). In this model, the basic assumptions are similar to those of Hayward & Hopkins: the ISM disk is supported by turbulent pressure, and stellar feedback drives gas outflow. However, the

timescale of the Lagos et al. simulation is set by the lifetime of giant molecular clouds, which can be larger than the crossing time adopted by Hayward & Hopkins, and their SFR is set by an empirical relation. They also found that the mass-loading factor scales with gas density and galaxy mass with an exponent of $\sim -1.1 \pm 0.5$ (based on estimation from their Figure 14), which is larger than what we observed.

Hydrodynamical simulations predict scaling relations of the mass-loading factor that are consistent in slope but with somewhat larger normalization than the observational result. The dotted line in figure 4 shows the scaling relation presented by Muratov et al. (2015) based on the Feedback in Realistic Environments (FIRE) simulations. The slope is consistent with our observation but it is $\sim 3 - 7$ times larger in normalization than our best-fit line.

It is normally difficult for simulations to define mass-loading factors that match observations. Muratov et al. (2015) calculated the mass-loading factors from outflow flux (all gas with outward radial velocity) at $0.25R_{vir}$. This does not capture the effect of recycling from galactic fountains, whereas our measurement does. Thus, their results are likely to be larger than our measurements by construction. Based on the IllustrisTNG results from Pillepich et al. (2018), in which their definition of the mass-loading factor was explicitly remarked as not applicable to observations and was calculated from wind energy at injection, their “mass-loading factor” scales with halo mass M_{200} with the power law index of ~ -0.8 to -0.6 . If we convert halo mass to stellar mass using the stellar-to-halo mass relation $M_* \propto M_{200}^{2.27}$, found in the same suite of simulations (Niemiec et al. 2018) for galaxies with $M_* < 10^{11.2} M_\odot$, the power-law index with stellar mass is approximately -0.35 to -0.26 , which is still roughly consistent with what we measured.

5.5. Remarks on the MZR evolution with redshift

In Paper I, we argued that the evolution of metallicity with formation redshift in quiescent galaxies is more fundamental than the evolution with observed redshift. Galaxies of the same mass that formed at the same redshift should have similar metallicities regardless of when we observe them, assuming that most quiescent galaxies evolve passively.⁵ As in Paper I, we trace each galaxy back to its formation time by adding the age of the uni-

⁵ If the sample size were large enough, we could also use this technique to trace whether quiescent galaxies evolve passively in metal abundance. This could be done by comparing MZRs of galaxies that formed at the same formation redshift but are observed at different redshifts.

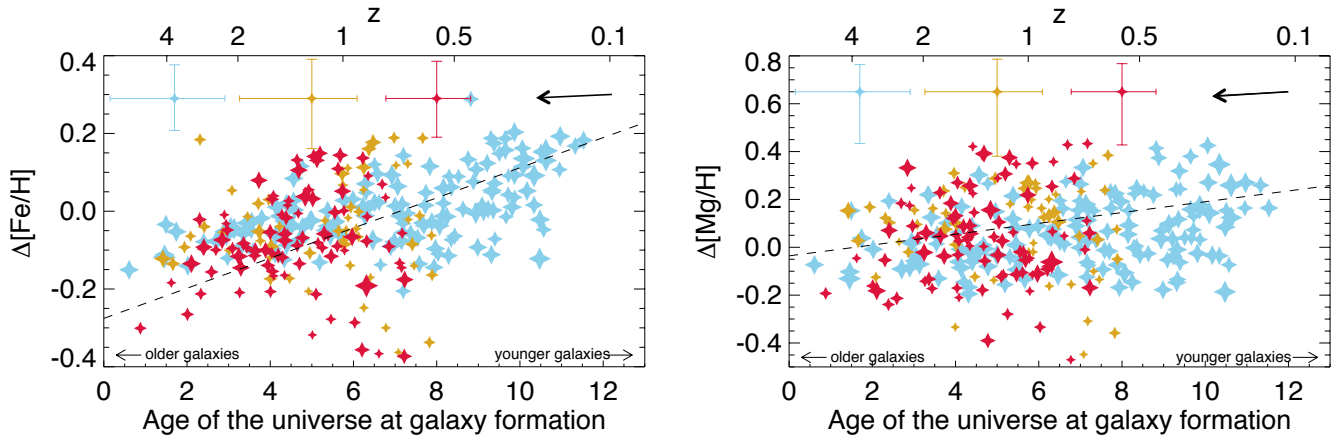


Figure 5. The dependence of $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ on formation redshift when the dependence on mass is removed (see Section 5.5). The median uncertainties are shown in the top of each panel. The color scheme follows previous plots. The slope on the left is statistically significant, but the slope on the right is not. The arrows at the top right corners show the slopes in the age-metallicity planes that are induced by the age-metallicity degeneracy.

verse when it is observed to the age of the galaxy. We plot the deviation in metallicities from the $z \sim 0$ MZR (Equation 3) as a function of formation time in Figure 5. The plots essentially show the dependence of metallicities on formation redshift (age), while the dependence on mass is removed.

The deviation in the $[\text{Fe}/\text{H}]$ abundances is consistent with what we found in Paper I. The $[\text{Fe}/\text{H}]$ evolution with formation time is 0.04 ± 0.01 dex per formation Gyr. However, we do not find a significant dependency of $[\text{Mg}/\text{H}]$ on formation time. As shown in the right of Figure 5, the best-fit slope for the evolution is 0.02 ± 0.01 dex per formation Gyr, a significance less than 2σ at face value.

We further estimate the impact of the age-metallicity degeneracy on the evolution of $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ with formation time. In a similar manner as that in Paper I, we create a set of mock observations by assuming that each galaxy of the observed mass and age has the intrinsic $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ values that follow the $z \sim 0$ MZR, i.e., hypothetically there is no evolution with redshift. We create SSP spectra with these ages and metallicities, add Gaussian noise with the same flux uncertainty array as the observed spectra, and fix the velocity dispersion to 250 km/s (to keep the computation time reasonable). We then calculate a χ^2 grid for each spectrum by comparing to the noiseless grid used to calculate systematic uncertainties in Section 3. The “observed” age, $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$ of each mock galaxy is then selected according to the probability from the χ^2 grid. These mock galaxies do not have any intrinsic deviation from the $z \sim 0$ MZR. Therefore, if there is any evolution with formation time, it is caused by the age-metallicity degeneracy.

Based on these mock observations, we find that the age-metallicity degeneracy cannot explain the dependence of $\Delta[\text{Fe}/\text{H}]$ on formation time and even lowers the significance for the dependence of $\Delta[\text{Mg}/\text{H}]$ on formation time. Similar to what we found in Paper I, the age-metallicity degeneracy induces a small positive slope of $[\text{Fe}/\text{H}]$ at 0.004 ± 0.002 dex per formation Gyr. For $[\text{Mg}/\text{H}]$, it induces 0.010 ± 0.004 dex per formation Gyr. These slopes are illustrated at the top-right corner of each panel in Figure 5. This confirms that the evolution of $[\text{Fe}/\text{H}]$ with formation time is significant at $> 3\sigma$ but the evolution of $[\text{Mg}/\text{H}]$ is less than 2σ significant. An evolution of $[\text{Mg}/\text{H}]$ with formation time would imply that fundamental properties of galaxies such as yield or mass-loading factors depend on time. Since we did not detect any significant evolution of $[\text{Mg}/\text{H}]$ with both observed and formation redshift, these fundamental properties are not required to depend on redshift.

The absence of the evolution of $[\text{Mg}/\text{H}]$ with redshift together with the existence of the evolution of $[\text{Fe}/\text{H}]$ with redshift suggest that the latter is caused by the delayed time of the Type Ia supernovae. The main difference between the two elements is that Mg is approximately instantaneously recycled while the Fe can have a delay time up to several Gyrs (e.g., Maoz et al. 2012). Thus, the evolution of $[\text{Fe}/\text{H}]$ with redshift in quiescent galaxies that we found is not surprising and is perhaps expected due to selection effects. We restricted our sample to be passive. The galaxies at higher redshift are required to finish forming star early, while the galaxies at lower redshift could have quenched at a later time. Quiescent galaxies at higher redshift had less time to become enriched by Fe from Type Ia supernovae.

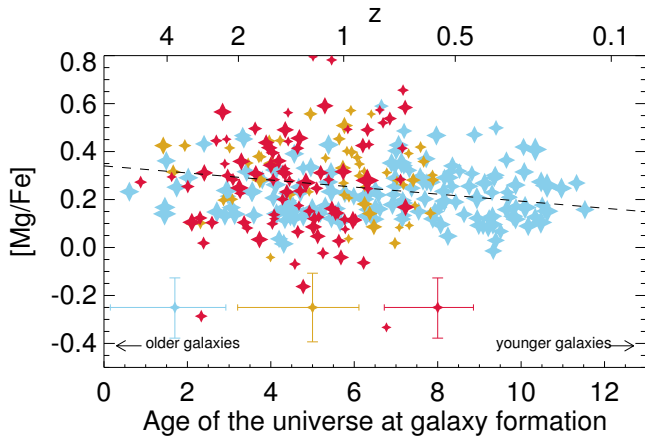


Figure 6. The dependence of $[\text{Mg}/\text{Fe}]$ on formation redshift. The median uncertainties are shown in the bottom. The color scheme follows previous plots (indicating different observed redshifts). A slight negative slope suggests that quiescent galaxies that form earlier on average had shorter star formation duration.

We plot $[\text{Mg}/\text{Fe}]$ as a function of formation time in Figure 6. This is to further investigate the conjecture that the evolution of $[\text{Fe}/\text{H}]$ with redshift is due to shorter star formation histories of quiescent galaxies at an earlier time. Since the ratio between alpha elements and Fe is an indication of star formation duration (Thomas et al. 2005), we expect that, on average, galaxies that form earlier have higher $[\text{Mg}/\text{Fe}]$. The slope in Figure 6 is -0.015 ± 0.006 dex per formation Gyr, suggesting that galaxies that form earlier indeed have higher $[\text{Mg}/\text{Fe}]$ than galaxies that form later. The slope is also as expected, roughly equal to the difference in the slope in the $\Delta [\text{Mg}/\text{H}]$ with formation time and the slope in the $\Delta [\text{Fe}/\text{H}]$ with formation time in Figure 5.

The evolution of $[\text{Fe}/\text{H}]$ with redshift we found in this work does not necessarily contradict the non-detection of evolution of MZR with redshift in Gallazzi et al. (2014); Estrada-Carpenter et al. (2019); Saracco et al. (2019). This is mainly due to large measurement uncertainties. We first note that most of these works measure total metallicity $[\text{Z}/\text{H}]$ without α enhancement. However, most of the spectra either do not cover the Mg b lines, or the absorption indices were chosen to be less sensitive to α enhancement. Therefore these $[\text{Z}/\text{H}]$ should be close to our $[\text{Fe}/\text{H}]$ measurements. Regardless, Gallazzi et al. (2014), Estrada-Carpenter et al. (2019) and Saracco et al. (2019) found that the $[\text{Z}/\text{H}]$ -mass relations in quiescent galaxies (both in the field and in clusters) at $z \sim 0.7 - 1.6$ are not different from that of local galaxies. These studies are quite limited in terms of sample size, especially for the lower mass galaxies, and either small wavelength coverage or coarse spectral

resolution. These aspects result in large measurement uncertainties that can disguise the expected evolution.

For example, based on the mean age of 2.0 ± 0.8 Gyr of 7 quiescent galaxies at $z \sim 1.22$ in Saracco et al.’s (2019) sample and the amount of evolution with formation time found in this work, we expect them to have an average 0.15 ± 0.06 lower metallicity than the local galaxies. This number is smaller than their mean uncertainty in $[\text{Z}/\text{H}]$ measurements (0.22 dex), standard deviation within the sample (0.18 dex), and uncertainty in the normalization of the best-fit linear relation with stellar mass (0.15 dex).

5.6. Limitations of our approach

The formalism for the stellar metallicities of quiescent galaxies adopted here is simple and subject to certain limitations. Aside from those listed in the explicit assumptions regarding the metallicities of inflow and outflow, which are discussed in Appendix C, there are additional limitations involved in simplifying Equation 5 to Equation 7.

First is the rapid truncation of star formation by non-star forming activities when the galaxy is still gas rich, i.e., r_g may not be as negligible as assumed. This includes active galactic nuclei (AGN) feedback and ram-pressure stripping. It is still unclear whether AGNs are able to eject significant mass fractions out of the galaxies. There is evidence that AGNs can drive large outflow in individual cases (e.g. Maiolino et al. 2012; Nyland et al. 2013), or in more extreme objects such as ULIRGs and QSOs (e.g. Ciccone et al. 2014; Rupke et al. 2017). Nevertheless, many statistical studies found that AGN-driven winds are usually not strong enough to escape the galaxies (e.g. Sarzi et al. 2016; Concas et al. 2017; Roberts-Borsani & Saintonge 2019). We defer the discussion on ram-pressure stripping to the next section on galaxy environment.

Moreover, the formalism adopted here does not consider mergers. The most concerning case is a dry merger, as it increases the stellar mass but does not increase the stellar metallicity. There is evidence that the most massive galaxies have experienced multiple minor mergers in their outskirts, which is responsible for their size growth (e.g. van Dokkum et al. 2010; Greene et al. 2012, 2013). The effect of this is beyond our study. Nonetheless, we can consider a simpler case of a major merger of two quiescent galaxies with equal masses and metallicities that lie on the MZR. In this case, the descendant galaxy will double in stellar mass (an increase of 0.3 dex), while the metallicity remains the same. However, we estimate that the effect in this case is quite small. Because the MZR of quiescent galaxies has a gentle slope ($\lesssim 0.2$ dex per log mass in our finding), the descendant galaxy of a

major merger will have a metallicity that is just $\lesssim 0.06$ dex below the MZR relation.

5.7. Effect of galaxy environment

As already discussed in Paper I, we do not expect our finding to be affected by the environment. In short, a cluster environment mainly affects the quenched fraction of galaxies in the sense that cluster galaxies tend to be more quenched than field galaxies. However, in terms of chemical composition, the effect is either small (< 0.06 dex, e.g., Cooper et al. 2008; Jørgensen et al. 2018), or not significant (e.g., Harrison et al. 2011; Kacprzak et al. 2015; Fitzpatrick & Graves 2015).

These findings may be relevant to the role of ram-pressure stripping in galaxy clusters. Based on the SDSS group/cluster catalogues and a cosmological simulation, Wetzel et al. (2013) suggest that most satellite galaxies can retain their molecular gas in the disks after falling into clusters. These galaxies can continue their star formation histories in a similar manner (within 10%) to those of central galaxies. The authors suggest that this is consistent with the observations of molecular gas in the satellites of the Virgo cluster (Kenney & Young 1986; Young et al. 2011). Moreover, based on a different cosmological simulation, Bahé & McCarthy (2015) find that the role of ram-pressure stripping is most relevant to the removal of the atomic gas in the outer halo just before the star formation stops.

6. CONCLUSION

We measured ages and metallicities ($[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$) of individual quiescent galaxies in two galaxy clusters at $z \sim 0.39$ and $z \sim 0.54$ using full-spectrum fitting. The sample consists of 62 galaxies at $z = 0.39$ and 92 galaxies at $z = 0.54$ in the galaxy clusters C10024 and MS0451. We also used a subsample of 155 SDSS quiescent galaxies as reference $z \sim 0$ galaxies. We used the SPS models and response functions from Conroy et al. (2009) and Conroy et al. (2018) to fit the observed spectra. By expanding the sample size from Leethochawalit et al. (2018) and measuring Mg abundances in addition to Fe, we were able archaeologically to derive the dependence of mass-loading factor on mass based on the mass–metallicity relation of quiescent galaxies. The summary of the findings this paper are the following.

- We confirmed the finding in Paper I that the stellar mass– $[\text{Fe}/\text{H}]$ relation evolves with redshift. At a fixed stellar mass, quiescent galaxies at higher redshift have lower $[\text{Fe}/\text{H}]$ than quiescent galaxies at lower redshift. The slope of the relation is also likely steeper at higher redshift.

- We found no evolution in the mass– $[\text{Mg}/\text{H}]$ relation with redshift. This suggests that the evolution observed in the $[\text{Fe}/\text{H}]$ abundance is due to the delay time of Type Ia supernovae.
- We constrained the mass-loading factor using an analytic chemical evolution model for quiescent galaxies, assuming that Mg is an indicator of instantaneously recycled elements. We found that the mass-loading factor is a power-law function of galaxy stellar mass, $\eta \propto M_*^{-0.21 \pm 0.09}$, over the observed mass range $\sim 10^{9.5}$ to $10^{11.5} M_\odot$.
- Our constraint on the mass-loading factor is consistent with an analytic prediction in which outflow is caused by star-formation feedback in a turbulent disk (Hayward & Hopkins 2017). It is also consistent with the results from direct measurements of outflow using UV absorption lines.

There are still many open questions that can be explored beyond this paper. These include confirming the stellar mass–metallicity relationship (in both $[\text{Fe}/\text{H}]$ and alpha elements) with a large sample of field galaxies, especially down to low masses ($\sim 10^{9.5} M_\odot$). A large sample of abundance measurements in field quiescent galaxies at different redshifts would allow us not only to even more firmly establish the evolution of MZRs, but also, as suggested in Section 5.5, to determine the degree of passive evolution of quiescent galaxies in terms of their chemical composition. This can be done by comparing the MZRs of galaxies that formed at the same redshift but are observed at different redshifts. Beyond this, precise and consistent measurements of stellar metallicities in star-forming galaxies (especially as a function of gas fraction), will help bridge and strengthen the findings in our work with those by Gallazzi et al. (2005) and Lu et al. (2015).

We thank the referee for comments and suggestions that significantly improved the paper. The authors acknowledge Charlie Conroy, Philip Hopkins, and Gwen Rudie for useful feedback and the use of their models and fitting codes.

Facilities: Keck:II (DEIMOS), Sloan, GALEX, HST (WFPC2), Subaru (Suprime-Cam), CFHT (CFH12k), Hale (WIRC)

Software: KCORRECT (Blanton & Roweis 2007), FSPS, (Conroy et al. 2009), alf (Conroy et al. 2018), MPFIT (Markwardt 2012)

REFERENCES

- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *Annual Review of Astronomy and Astrophysics*, 47, 481
- Bahé, Y. M., & McCarthy, I. G. 2015, *MNRAS*, 447, 969
- Belfiore, F., Maiolino, R., Maraston, C., et al. 2017, *MNRAS*, 466, 2570
- Belfiore, F., Westfall, K. B., Schaefer, A., et al. 2019, arXiv e-prints, arXiv:1901.00866
- Blanton, M. R., & Roweis, S. 2007, *AJ*, 133, 734
- Bolatto, A. D., Warren, S. R., Leroy, A. K., et al. 2013, *Nature*, 499, 450
- Boselli, A., Cortese, L., Boquien, M., et al. 2014, *A&A*, 564, A66
- Bouché, N., Hohensee, W., Vargas, R., et al. 2012, *MNRAS*, 426, 801
- Bundy, K., Bershad, M. A., Law, D. R., et al. 2015, *ApJ*, 798, 7
- Chisholm, J., Tremonti, C., & Leitherer, C. 2018, *MNRAS*, 481, 1690
- Chisholm, J., Tremonti, C. A., Leitherer, C., & Chen, Y. 2017, *MNRAS*, 469, 4831
- Choi, J., Conroy, C., Moustakas, J., et al. 2014, *ApJ*, 792, 95
- Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102
- Churchill, C. W., Kacprzak, G. G., Steidel, C. C., et al. 2012, *ApJ*, 760, 68
- Cicone, C., Maiolino, R., Sturm, E., et al. 2014, *A&A*, 562, A21
- Cid Fernandes, R. 2018, *MNRAS*, 480, 4480
- Concas, A., Popesso, P., Brusa, M., et al. 2017, *A&A*, 606, A36
- Conroy, C., Graves, G. J., & van Dokkum, P. G. 2014, *ApJ*, 780, 33
- Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, 699, 486
- Conroy, C., Villaume, A., van Dokkum, P. G., & Lind, K. 2018, *ApJ*, 854, 139
- Cooper, M. C., Tremonti, C. A., Newman, J. A., & Zabludoff, A. I. 2008, *MNRAS*, 390, 245
- Davé, R., Finlator, K., & Oppenheimer, B. D. 2011, *MNRAS*, 416, 1354
- Dekel, A., & Silk, J. 1986, *ApJ*, 303, 39
- Estrada-Carpenter, V., Papovich, C., Momcheva, I., et al. 2019, *ApJ*, 870, 133
- Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4841, *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes*, ed. M. Iye & A. F. M. Moorwood, 1657–1669
- Finlator, K., & Davé, R. 2008, *MNRAS*, 385, 2181
- Fitzpatrick, P. J., & Graves, G. J. 2015, *MNRAS*, 447, 1383
- Gallazzi, A., Bell, E. F., Zibetti, S., Brinchmann, J., & Kelson, D. D. 2014, *ApJ*, 788, 72
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, *MNRAS*, 362, 41
- Garnett, D. R. 2002, *ApJ*, 581, 1019
- Garnett, D. R., & Shields, G. A. 1987, *ApJ*, 317, 82
- Ge, J., Yan, R., Cappellari, M., et al. 2018, *MNRAS*, 478, 2633
- Gobat, R., Daddi, E., Magdis, G., et al. 2018, *Nature Astronomy*, 2, 239
- Greene, J. E., Murphy, J. D., Comerford, J. M., Gebhardt, K., & Adams, J. J. 2012, *ApJ*, 750, 32
- Greene, J. E., Murphy, J. D., Graves, G. J., et al. 2013, *ApJ*, 776, 64
- Griffith, E., Martini, P., & Conroy, C. 2019, *MNRAS*, 484, 562
- Guo, Q., Gonzalez-Perez, V., Guo, Q., et al. 2016, *MNRAS*, 461, 3457
- Hafen, Z., Faucher-Giguère, C.-A., Anglés-Alcázar, D., et al. 2017, *MNRAS*, 469, 2292
- Harrison, C. D., Colless, M., Kuntschner, H., et al. 2011, *MNRAS*, 413, 1036
- Hayward, C. C., & Hopkins, P. F. 2017, *MNRAS*, 465, 1682
- Jørgensen, I., Chiboucas, K., Webb, K., & Woodrum, C. 2018, *AJ*, 156, 224
- Kacprzak, G. G., Yuan, T., Nanayakkara, T., et al. 2015, *ApJ*, 802, L26
- Kenney, J. D., & Young, J. S. 1986, *ApJL*, 301, L13
- Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, *ApJ*, 779, 102
- Kirby, E. N., Simon, J. D., & Cohen, J. G. 2015, *ApJ*, 810, 56
- Kochanek, C. S., Eisenstein, D. J., Cool, R. J., et al. 2012, *ApJS*, 200, 8
- Kriek, M., Conroy, C., van Dokkum, P. G., et al. 2016, *Nature*, 540, 248
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kurucz, R. 1993, *SYNTHÉ Spectrum Synthesis Programs and Line Data*. Kurucz CD-ROM No. 18. Cambridge, 18
- Lagos, C. d. P., Lacey, C. G., & Baugh, C. M. 2013, *MNRAS*, 436, 1787
- Larson, R. B. 1974, *MNRAS*, 169, 229
- Leethochawalit, N., Kirby, E. N., Moran, S. M., Ellis, R. S., & Treu, T. 2018, *ApJ*, 856, 15 (Paper I)
- Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, *A&A*, 500, 145
- Lian, J., Thomas, D., Maraston, C., et al. 2018, *MNRAS*, 474, 1143
- Lu, Y., Blanc, G. A., & Benson, A. 2015, *ApJ*, 808, 129

- Lu, Y., Wechsler, R. H., Somerville, R. S., et al. 2014, *ApJ*, 795, 123
- Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, *MNRAS*, 456, 2140
- Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, 488, 463
- Maiolino, R., Gallerani, S., Neri, R., et al. 2012, *MNRAS*, 425, L66
- Maoz, D., Mannucci, F., & Brandt, T. D. 2012, *MNRAS*, 426, 3282
- Marigo, P., & Girardi, L. 2007, *A&A*, 469, 239
- Markwardt, C. 2012, MPFIT: Robust non-linear least squares curve fitting, *Astrophysics Source Code Library*, , ascl:1208.019
- Mathews, W. G., & Baker, J. C. 1971, *ApJ*, 170, 241
- McClure, R. D., & van den Bergh, S. 1968, *AJ*, 73, 1008
- Mentz, J. J., La Barbera, F., Peletier, R. F., et al. 2016, *MNRAS*, 463, 2819
- Moran, S. M., Ellis, R. S., Treu, T., et al. 2005, *ApJ*, 634, 977
- . 2007, *ApJ*, 671, 1503
- Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al. 2015, *MNRAS*, 454, 2691
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, 477, 1206
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *The Astrophysical Journal Supplement Series*, 208, 5
- Niemiec, A., Jullo, E., Giocoli, C., Limousin, M., & Jauzac, M. 2018, arXiv e-prints, arXiv:1811.04996
- Nomoto, K., Tominaga, N., Umeda, H., Kobayashi, C., & Maeda, K. 2006, *NuPhA*, 777, 424
- Nyland, K., Alatalo, K., Wrobel, J. M., et al. 2013, *ApJ*, 779, 173
- Pagel, B. E. J. 1997, *Nucleosynthesis and Chemical Evolution of Galaxies*
- Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, 348, L59
- Pillepich, A., Springel, V., Nelson, D., et al. 2018, *MNRAS*, 473, 4077
- Roberts-Borsani, G. W., & Saintonge, A. 2019, *MNRAS*, 482, 4111
- Rupke, D. S. N., Gültekin, K., & Veilleux, S. 2017, *ApJ*, 850, 40
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, *MNRAS*, 371, 703
- Saracco, P., La Barbera, F., Gargiulo, A., et al. 2019, *MNRAS*, 484, 2281
- Sarzi, M., Kaviraj, S., Nedelchev, B., et al. 2016, *MNRAS*, 456, L25
- Spitoni, E., Calura, F., Matteucci, F., & Recchi, S. 2010, *A&A*, 514, A73
- Taylor, P., & Kobayashi, C. 2016, *MNRAS*, 463, 2465
- Thomas, D., Maraston, C., Bender, R., & Mendes de Oliveira, C. 2005, *ApJ*, 621, 673
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, 613, 898
- van Dokkum, P. G., Whitaker, K. E., Brammer, G., et al. 2010, *ApJ*, 709, 1018
- Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, *MNRAS*, 455, 4183
- Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, *ApJ*, 792, 8
- Wetzel, A. R., Tinker, J. L., Conroy, C., & van den Bosch, F. C. 2013, *MNRAS*, 432, 336
- Wilkinson, D. M., Maraston, C., Goddard, D., Thomas, D., & Parikh, T. 2017, *MNRAS*, 472, 4297
- Young, L. M., Bureau, M., Davis, T. A., et al. 2011, *MNRAS*, 414, 940
- Zahid, H. J., Geller, M. J., Kewley, L. J., et al. 2013, *ApJ*, 771, L19

APPENDIX

A. CHOOSING THE BEST COMBINATION OF RESPONSE FUNCTIONS

Although the FSPS code is able to measure many elemental abundances simultaneously, it requires more computation time to measure more elements. Every additional element requires one additional spectral response function. To achieve reliable measurements of $[\text{Mg}/\text{Fe}]$ within a reasonable computation time, we tested several combinations of response functions to be included in the fitting models. We performed these tests on two sets of data. First, we used our fitting code to fit the high-S/N, stacked spectra of [Choi et al. \(2014\)](#). We then compared our results to the [Choi et al.](#)'s measurements. Second, we used the full-spectrum fitting code `alf` ([Conroy et al. 2018](#)) to re-measure abundances for a subset of our sample. We compared the `alf` measurements to the results described according to the method described in Section 3.

A.1. Comparison with Choi14

In this section, we test our measurements of age and metal abundance on 10 high-S/N stacked spectra from [Choi et al. \(2014\)](#). The spectra were stacked from individual spectra in the AGN and Galaxy Evolution Survey (AGES, [Kochanek et al. 2012](#)) based on their redshifts, $z = 0.3$ to 0.7 , and masses, $M_* = 10^{10.2}$ to $10^{11.3} M_\odot$. The spectral resolution is 6 \AA with the wavelength range of 4000 to 5500 \AA .

We consider 5 spectral models with different sets of response functions in addition to the four baseline parameters: age, $[\text{Z}/\text{H}]$, velocity dispersion, and redshift. The combinations are

1. Mg only,
2. alpha elements (Mg, Si, Ca, Ti, and O) all fixed to the same value,
3. Mg and O,
4. Mg and N,
5. “full” combination having the same list of 9 elements as [Choi et al. \(2014\)](#): Mg, N, Fe, O, C, N, Si, Ca and Ti.

We compare the results from each model with the measurements from [Choi et al. \(2014\)](#) (see Figure 7). We find that three of five the models agree within 0.1 dex with the measurements from [Choi et al.](#) These models are those that include the response functions of (3) Mg and O, (4) Mg and N, and (5) the “full” combination.

In terms of the age measurements, all models perform equally well. The agreement holds when the age is older than ~ 3.5 Gyr. For younger populations, the age measurements obtained here are ~ 0.05 dex older than the ages reported by [Choi et al.](#) This trend of the differences in the age measurements is the same as what we found in Paper I when we fit for only ages and $[\text{Z}/\text{H}]$. We attributed the difference to the lack of age-sensitive higher-order Balmer lines ($<4000 \text{ \AA}$) in the spectra and possibly the differences in the models used (see Appendix of Paper I).

However, we find agreement in $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$ with [Choi et al.](#) only in the three combinations of response functions mentioned above. When the enhancements of all alpha elements are fixed to the same value or when only the $[\text{Mg}/\text{Fe}]$ response function is included, the $[\text{Fe}/\text{H}]$ is generally slightly over-predicted while the $[\text{Mg}/\text{H}]$ is under-predicted (red triangles or orange squares in Figure 7). The former case (all alpha elements fixed to the same value) is not surprising because alpha elements generally do not have the same values, and they do not always track each other. [Conroy et al. \(2014\)](#) found that Mg tracks O closely, but Mg does not track heavier α elements such as Si, Ca, and Ti. The latter case (Mg response function only) is more interesting. We cannot measure $[\text{Mg}/\text{Fe}]$ or $[\text{Fe}/\text{H}]$ well when we only include the response function of Mg. This is probably because the Mg enhancement not only strengthens the absorption lines in the Mg b $\lambda 5170$ region but also weakens the absorption lines at $\sim 4000\text{--}4400 \text{ \AA}$. The wavelength range also responds to the enhancement of Fe, O, and N (See Figures 7 and 8 of [Conroy et al. 2018](#)). For this reason, the agreements improve to within 0.1 dex when we include the response functions of either O or N.

In summary, we found that our fitting code yields results that agree reasonably well (within 0.1 dex) with the literature when we use the following models: Mg and N, Mg and O, or the “full” combination. However, the “full” combination model, which includes 13 parameters, is costly in terms of the computational time. Therefore, it is difficult to apply to a large sample. Thus, we restrict further consideration to the two simpler models that include the response functions of either Mg and N or Mg and O.

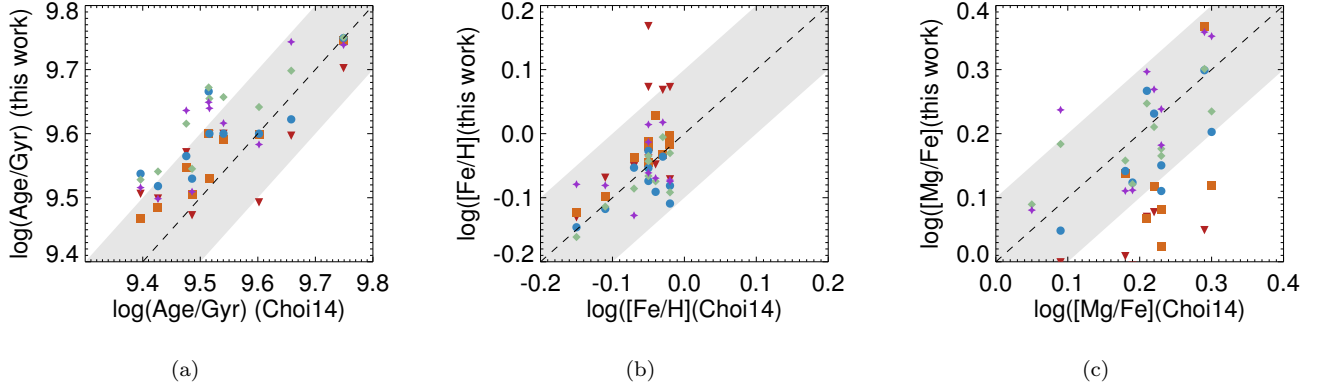


Figure 7. Comparison with the measurements of Choi et al. (2014) using different combinations of metal elements in the response functions. The combinations are (1) [Mg/Fe] only (orange squares), (2) alpha elements [Mg,Si,Ca,Ti,O/Fe] all fixed to the same value (red downward triangles), (3) [Mg/Fe] and [O/Fe] (blue circles), (4) [Mg/Fe] and [N/Fe] (green diamonds), and (5) Mg,N,Fe,O,C,N,Si,Ca and Ti/Fe (purple stars). The latter three combinations show the best agreement with Choi et al. (2014).

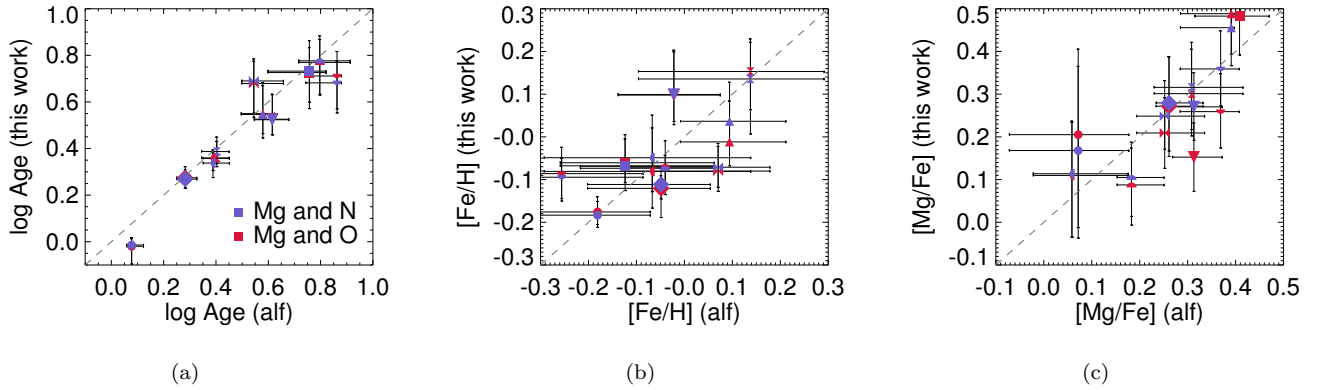


Figure 8. Comparison of parameters of 10 galaxies measured with `alf` (Conroy et al. 2018), which fits for the abundances of nine individual elements, and measure with our code using the combination of response functions of (1) [Mg/Fe] and [N/Fe] and (2) [Mg/Fe] and [O/Fe]. We chose combination (1) in our final fitting code to apply to the full sample.

A.2. Comparison to the `alf` full-spectrum fitting code

In order to choose the better of the remaining two models, we compare the results from our fitting code to the results from the publicly available “absorption line fitting code” (`alf`, Conroy et al. 2018). `alf` is a fitting algorithm that uses the Markov Chain Monte Carlo method of parameter estimation. It constructs empirical SSP spectra from the MIST isochrones (Choi et al. 2016), the MILES and Extended IRTF spectra libraries, and theoretical response functions of individual elements based on the Kurucz suite of routines. Our code uses a different set of empirical SSP spectra (Conroy et al. 2009) but uses the same response functions. All of the `alf` results presented here were modeled with the Kroupa IMF and the “simple” mode of fitting, which fits for 13 parameters including stellar age, $[Z/H]$, and abundances of Fe, C, N, O, Na, Mg, Si, Ca, and Ti.

Because the `alf` computational time for a single spectrum is long ($\gtrsim 100$ cpu hours), we selected the 10 highest S/N spectra from our sample in MS0451 ($z \sim 0.54$). Figure 8 shows the comparison between the two methods. Our measurements are remarkably consistent with the results from `alf`, given that our code is simpler. However, We found that in some of the cases, the model that includes the response functions of Mg and N yields better agreement, especially in terms of [Mg/Fe], than the model that includes the response functions of Mg and O. Based on the results in this section, we chose the model that includes the response functions of Mg and N to apply to our data.

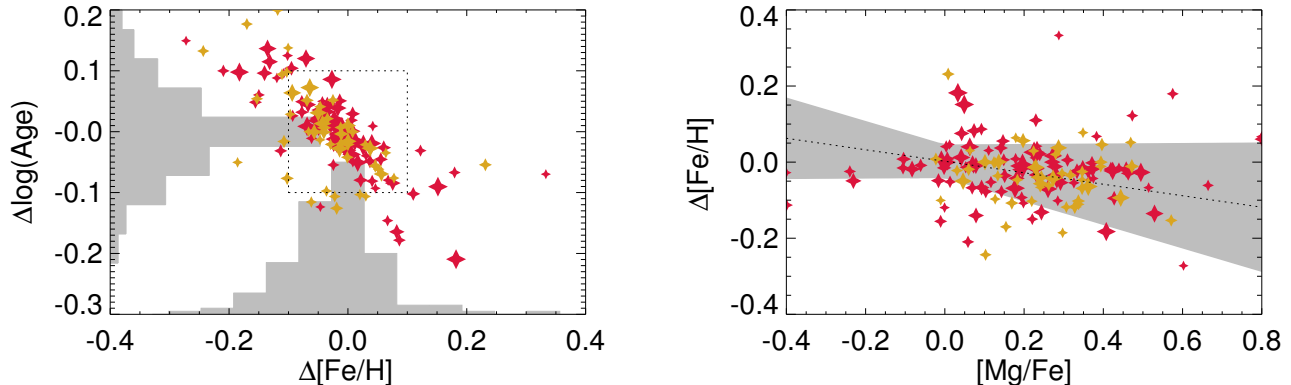


Figure 9. Comparison of the age and $[\text{Fe}/\text{H}]$ measurements using the models in Paper I (same models but without α enhancement) and the models used here. The left panel shows the differences in the measured age and $[\text{Fe}/\text{H}]$ with underlying histograms in gray color. The color (yellow and red) represents the $z \sim 0.39$ and the $z \sim 0.55$ sample, respectively. The dotted box delineates uncertainties of 0.1 dex. The right panel shows the difference in $[\text{Fe}/\text{H}]$ with the best linear fit and its uncertainty (dotted line and gray shading). The uncertainties in $[\text{Fe}/\text{H}]$ and age lie along the age-metallicity degeneracy direction with a slight dependence on $[\text{Mg}/\text{Fe}]$.

A.3. Comparison between the fitting method in Paper I and in this paper

Here we compare the $[\text{Fe}/\text{H}]$ and age measurements derived from the method in Paper I to those derived in this paper. In Paper I, we masked out the Mg b absorption region in the spectra and measured $[\text{Fe}/\text{H}]$ and age with and without α enhancement. In this paper, we fit the full spectra with models that treat N and Mg abundances as free parameters. We use both methods to fit the higher-redshift spectra in this paper.

Figure 9 shows the comparisons of the results from both methods. The x - and y - axes represent the results from this paper subtracted from the results using the method in Paper I. The left panel shows that the differences in $[\text{Fe}/\text{H}]$ and age lies along the age-metallicity degeneracy, with means of -0.01 ± 0.01 dex and 0.00 ± 0.01 dex, respectively. The mode of the distributions are zero. I.e., there are no significant systematic differences in the $[\text{Fe}/\text{H}]$ nor age measured by either method. Most of the discrepancies are within 0.1 dex in either parameter, as shown by the dotted box. However, the distribution of the differences in measured $[\text{Fe}/\text{H}]$ ($\Delta[\text{Fe}/\text{H}]$) is slightly asymmetric, with a heavier distribution toward negative values. This means that we are more likely to have underestimated $[\text{Fe}/\text{H}]$ in this paper compared to Paper I.

The right panel shows $\Delta[\text{Fe}/\text{H}]$ vs. $[\text{Mg}/\text{Fe}]$ along with the best-fit line. Although the best-fit slope is statistically consistent with zero, its best-fit value suggests that when we $[\text{Mg}/\text{Fe}]$ is high, we possibly underestimate $[\text{Fe}/\text{H}]$ with the method in Paper I. The reason for this is also likely due to the response function of Fe and Mg. The spectrum in the 4000–4400 Å range responds to Mg and Fe in the opposite direction. Therefore, the models without elemental enhancements are likely to underestimate the $[\text{Fe}/\text{H}]$ when fit to a Mg-enhanced spectrum.

B. DEGENERACIES BETWEEN AGE, $[\text{Fe}/\text{H}]$ AND $[\text{Mg}/\text{Fe}]$

In this section, we explore the systematic uncertainties based on the degeneracies between the three parameters of interest: age, $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$. To do so, we select a set of reference spectra from the same noiseless SSP grid generated in Section 3. Then, we add Gaussian noise to reach S/N of 10 \AA^{-1} (lower S/N limit of our data) and 25 \AA^{-1} . We then calculate their χ^2 s by comparing them to the noiseless SSP grid. The results are shown in Figure 10. Each figure shows contours of 1σ uncertainty based on the 2-D posterior probability distribution functions of the degeneracy between the two parameters on the x-y axis. The third parameter is fixed at $[\text{Mg}/\text{Fe}] = 0.1$ dex, Age = 3 Gyr, and $[\text{Fe}/\text{H}] = -0.15$ dex (from left to right respectively).

Based on Figure 10, the correlation between uncertainties of the measured parameters is strongest for $[\text{Fe}/\text{H}]$ and age (left column). This is known as the age-metallicity degeneracy. When the $[\text{Fe}/\text{H}]$ is underestimated, the age is overestimated (and vice versa). In the middle column, we observe a slight positive correlation between the uncertainties of $[\text{Mg}/\text{Fe}]$ and of $[\text{Fe}/\text{H}]$. If $[\text{Mg}/\text{Fe}]$ is overestimated, $[\text{Fe}/\text{H}]$ will likely also be overestimated. This is due to the overlapping response functions of both elements in the 4000–4400 Å range. This is consistent with what we found

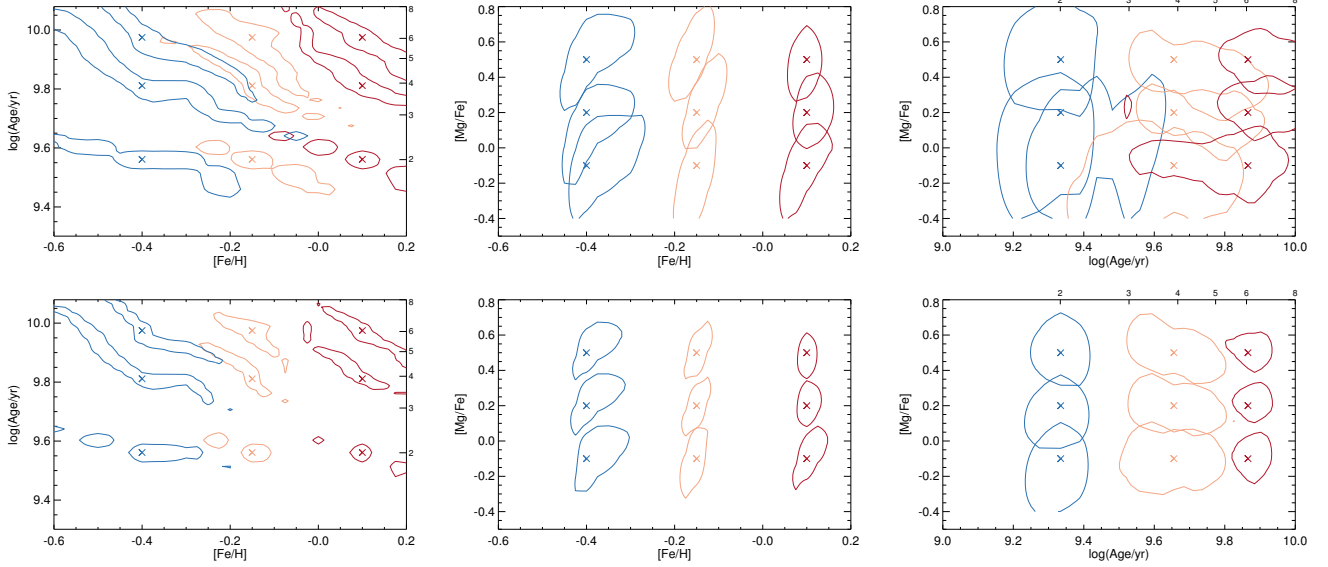


Figure 10. Uncertainties in the measured ages, $[\text{Fe}/\text{H}]$'s, and $[\text{Mg}/\text{Fe}]$'s according to degeneracies between the three parameters. Each contour shows the 1σ range in uncertainties of the x-y parameters, while the third parameter is fixed at the correct value (at $[\text{Mg}/\text{Fe}] = 0.1$ dex, Age = 3 Gyr, and $[\text{Fe}/\text{H}] = -0.15$ dex from left to right). The true parameters of the x-y axis are shown as cross marks. The S/N of each spectrum is 10 \AA^{-1} (upper panels) and 25 \AA^{-1} (lower panels).

in Appendix A.3. Lastly, as shown in the left column, the uncertainties in $[\text{Mg}/\text{Fe}]$ and age are the most radially symmetrical. This suggests that $[\text{Mg}/\text{Fe}]$ and age have the least degeneracy.

The dominant source of the systematic uncertainty in $[\text{Fe}/\text{H}]$ is the degeneracy between $[\text{Fe}/\text{H}]$ and age (left column). This degeneracy contributes ~ 0.2 dex to the total uncertainty of $[\text{Fe}/\text{H}]$ at $\text{S/N} = 10 \text{ \AA}^{-1}$ (this reduces to ~ 0.1 dex at $\text{S/N} = 25 \text{ \AA}^{-1}$). The uncertainty in $[\text{Fe}/\text{H}]$ also depends on the $[\text{Fe}/\text{H}]$ value. It is larger at lower $[\text{Fe}/\text{H}]$'s and can be as large as 0.3 dex at low S/N. The degeneracy between $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{Fe}]$ (middle column) is a smaller source of $[\text{Fe}/\text{H}]$ uncertainty, contributing less than half of the uncertainty that is generated by the degeneracy between $[\text{Fe}/\text{H}]$ and age (left column). As for the uncertainty in age, the dominant source is also the degeneracy between $[\text{Fe}/\text{H}]$ and age. However, at low $[\text{Mg}/\text{Fe}]$ values (less than solar), the systematic uncertainty in age increases significantly. Nevertheless, we need not be too concerned because most of the quiescent galaxies are $[\text{Mg}/\text{Fe}]$ enhanced. Finally, the uncertainty in $[\text{Mg}/\text{Fe}]$ comes equally from the degeneracy of $[\text{Mg}/\text{Fe}]$ with age and with $[\text{Fe}/\text{H}]$. Both contribute ~ 0.2 - 0.3 dex at low S/N and ~ 0.1 dex at high S/N.

C. THE IMPACT OF ENRICHED OUTFLOW AND ENRICHED INFLOW ON MEASURED MASS-LOADING FACTORS

In this section, we separately investigate the effect of the assumed pristine inflow and the assumed gas-phase metallicity of the outflow on our measured mass-loading factors. We first explore the case of an enriched inflow. If the infalling gas has a metallicity of Z_{inf} , we can modify Equation 4 to

$$dM_{Z,g} = ydM_* - Z_g dM_* - \frac{\eta}{1-R} Z_g dM_* + Z_{\text{inf}} dM_{\text{inf}} \quad (\text{C1})$$

where dM_{inf} is the mass inflow rate. We can substitute $Z_g dM_*$ with $dM_{Z,*}$, integrate the equation, and divide all terms by dM_* to get

$$Z_g r_g = y - \left(1 + \frac{\eta}{1-R}\right) Z_* + \langle Z_{\text{inf}} \rangle \frac{M_{\text{inf}}}{M_*} \quad (\text{C2})$$

Assuming the galaxy started with no gas, we can estimate the total infall mass for quiescent galaxies as the current stellar mass plus the total outflow mass: $M_{\text{inf}} = \left(1 + \frac{\eta}{1-R}\right) M_*$. The equation for quiescent galaxies is then

$$Z_* \approx \frac{y}{1 + \frac{\langle \eta \rangle}{1-R}} + \langle Z_{\text{inf}} \rangle \quad \text{or} \quad \langle \eta \rangle \approx \left(\frac{y}{Z_* - \langle Z_{\text{inf}} \rangle} - 1 \right) (1-R) \quad (\text{C3})$$

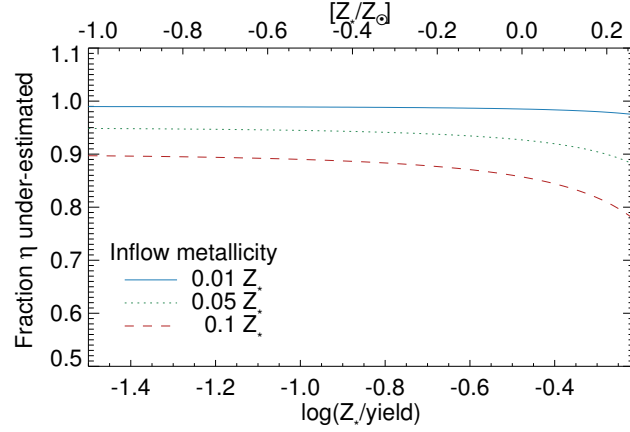


Figure 11. The underestimation of the mass-loading factor (the ratio between η derived assuming pristine inflowing gas to η derived with enriched inflow) as a function of stellar metallicity. The bottom x -axis is the metallicity in units of the yield. The top x -axis shows the metallicity when the yield is assumed to be $3Z_{\odot}$. Each line corresponds to a different assumed metallicity of the inflow.

The equation above suggests that we would underestimate the mass-loading factor by assuming that the inflowing gas is pristine. The amount of the underestimation is shown in Figure 11. The y -axis is the ratio between the η derived assuming pristine inflow to the η derived from the model with enriched inflow. The amount of the underestimation depends on both inflow and stellar metallicity. For simplicity in quantifying the amount of the underestimation, we parameterize the inflow metallicity as a fixed fraction of the final stellar metallicity: $Z_{\text{inf}} = \alpha Z_*$. (This is similar to the parameterization used by [Finlator & Davé 2008](#) but with the stellar metallicity instead of the gas-phase metallicity.)

The underestimation of the mass-loading factor when gas inflow is assumed to be pristine is estimated to be at most 20% on a linear scale. Studies of infalling H I absorbing complexes generally have found that H I filaments and the circum-galactic medium are metal-poor, $\log Z/Z_{\odot} < -1$ (e.g., [Churchill et al. 2012](#); [Hafen et al. 2017](#)). Thus, based on Figure 11, even in a very enriched inflow case at the high-mass end ($Z_{\text{inf}} = 0.1Z_* \sim -1$ dex), the estimated η would be at most underestimated by 20%, or less than 0.1 dex.

Now, we investigate the assumption of the perfectly mixed ISM, i.e., the outflow metallicity is the same as the ISM metallicity. There is observational evidence that the outflows might be more metal-rich than the ISM. For example, [Chisholm et al. \(2018\)](#) measured the outflow metallicities of five nearby star-forming galaxies based on rest-frame ultraviolet absorption lines. The authors found that on average, the outflow metallicities are 2.6 times larger than the ISM with no significant dependence on galaxy mass.

If the outflow is more enriched than the ISM, we will overestimate the mass-loading factors by assuming that the outflow metallicity is the same as the ISM metallicity. Let's assume that the outflow metallicity is βZ_g instead of Z_g , we can modify Equation 4 to

$$dM_{Z,g} = ydM_* - Z_g dM_* - \frac{\eta}{1-R} \beta Z_g dM_* \quad (\text{C4})$$

This means that our estimated η in this paper are actually $\beta\eta$. Therefore, if β is 2.6 then we would overestimate the mass-loading factors by the same amount, which is 0.4 dex in log scale. If β is really independent of mass, as found in [Chisholm et al. \(2018\)](#), this should have no effect on the derived power-law index of the relation between η and M_* .