The MOSDEF Survey: Direct-Method Metallicities and ISM Conditions at $z \sim 1.5 - 3.5$

Ryan L. Sanders,$^1$† Alice E. Shapley,$^2$ Naveen A. Reddy,$^{3,4}$ Mariska Kriek,$^5$ Brian Siana,$^3$ Alison L. Coil,$^6$ Bahram Mobasher,$^3$ Irene Shivaei,$^{7,8}$ William R. Freeman,$^3$ Mojegan Azadi,$^9$ Sedona H. Price,$^{10}$ Gene Leung,$^6$ Tara Fetherolf,$^3$ Laura de Groot,$^{11}$ Tom Zick,$^5$ Francesca M. Fornasini,$^9$ and Guillermo Barro$^{12}$

$^1$Department of Physics, University of California, Davis, One Shields Ave, Davis, CA 95616, USA
$^2$Department of Physics & Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA
$^3$Department of Physics & Astronomy, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA
$^4$Alfred P. Sloan Fellow
$^5$Astronomy Department, University of California, Berkeley, CA 94720, USA
$^6$Center for Astrophysics and Space Sciences, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0424, USA
$^7$Department of Astronomy/Steward Observatory, 933 North Cherry Ave, Rm N204, Tucson, AZ, 85721-0065, USA
$^8$Hubble Fellow
$^9$Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
$^{10}$Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, Garching, 85741, Germany
$^{11}$Department of Physics, The College of Wooster, 1189 Beall Avenue, Wooster, OH 44691, USA
$^{12}$Department of Physics, University of the Pacific, 3601 Pacific Ave, Stockton, CA 95211, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We present detections of [O III]λ4363 and direct-method metallicities for star-forming galaxies at $z = 1.7 - 3.6$. We combine new measurements from the MOSFIRE Deep Evolution Field (MOSDEF) survey with literature sources to construct a sample of 18 galaxies with direct-method metallicities at $z > 1$, spanning $7.5 < 12 + \log(O/H) < 8.2$ and $\log(M_*/M_\odot) = 7 - 10$. We find that strong-line calibrations based on local analogs of high-redshift galaxies reliably reproduce the metallicity of the $z > 1$ sample on average. We construct the first mass-metallicity relation at $z > 1$ based purely on direct-method O/H, finding a slope that is consistent with strong-line results. Direct-method O/H evolves by $\lesssim 0.1$ dex at fixed $M_*$ and SFR from $z \sim 0 - 2.2$. We employ photoionization models to constrain the ionization parameter and ionizing spectrum in the high-redshift sample. Stellar models with super-solar O/Fe and binary evolution of massive stars are required to reproduce the observed strong-line ratios. We find that the $z > 1$ sample falls on the $z \sim 0$ relation between ionization parameter and O/H, suggesting no evolution of this relation from $z \sim 0$ to $z \sim 2$. These results suggest that the offset of the strong-line ratios of this sample from local excitation sequences is driven primarily by a harder ionizing spectrum at fixed nebular metallicity compared to what is typical at $z \sim 0$, naturally explained by super-solar O/Fe values at high redshift caused by rapid formation timescales. Given the extreme nature of our $z > 1$ sample, the implications for representative $z \sim 2$ galaxy samples at $\sim 10^{10} M_\odot$ are unclear, but similarities to $z \sim 6$ galaxies suggest that these conclusions can be extended to galaxies in the epoch of reionization.

Key words: galaxies: abundances — galaxies: high-redshift

1 INTRODUCTION

A full understanding of galaxy formation and evolution requires knowledge of the chemical enrichment of galaxies over...
cosmic time. In particular, the gas-phase oxygen abundance (metallicity) of the interstellar medium (ISM) is a sensitive probe of the cycle of baryons into and out of galaxies. The scaling of metallicity with global galaxy properties such as stellar mass ($M_*$) and star-formation rate (SFR) can reveal the role of gas accretion, star formation, feedback, and outflows in shaping galaxy growth (e.g., Finlator & Davé 2008; Peebles & Shankar 2011; Davé et al. 2012; Lilly et al. 2013). The shape and normalization of metallicity scaling relations also provide valuable tests of baryonic physics prescriptions adopted in numerical simulations (e.g., Ma et al. 2016; Davé et al. 2017, 2019; De Rossi et al. 2017; Torrey et al. 2018, 2019). It is thus of critical importance to obtain accurate metallicity measurements of galaxies at low and high redshifts.

A robust determination of oxygen abundance can be obtained based on atomic physics via the “direct” method. In this approach, the electron temperature of the ionized gas is determined from the ratio of a weak auroral emission line (e.g., [O III]λ4363 in the rest-optical or O III]λλ4166,1666 in the rest-ultraviolet) to a strong emission line from the same ion species (e.g., [O III]λ5007). In combination with hydrogen recombination lines and strong oxygen emission features ([O II]λλ3726,3729, [O III]λ5007), the total oxygen abundance can be derived from the electron temperature (e.g., Izotov et al. 2006; Osterbrock & Ferland 2006; Pilyugin et al. 2012; Andrews & Martini 2013; Berg et al. 2015; Croxall et al. 2015, 2016; Pilyugin & Grebel 2016).

Since auroral lines are ∼100× fainter than the strong lines and not typically detected in the spectra of most galaxies in large spectroscopic surveys, empirical calibrations have been constructed relating strong-line ratios to direct-method metallicities. These calibrations have been developed based on photoionization models that reproduce the properties of local H II regions (e.g., Kewley & Dopita 2002; Tremonti et al. 2004; Kobulnicky & Kewley 2004; Dopita et al. 2016). Strong-line metallicity calibrations have led to a detailed understanding of galaxy metallicity scaling relations at $z \sim 0$ (e.g., Tremonti et al. 2004; Kewley & Ellison 2008; Mannucci et al. 2010; Lara-López et al. 2010).

Large samples of galaxies ($N > 1000$) at $z > 1$ with rest-optical strong-line ratio measurements are now available thanks to large spectroscopic surveys including the MOSFIRE Deep Evolution Field (MOSDEF) survey (MOSDEF: Kriek et al. 2015), Keck Baryonic Structure Survey (KBSS; Steidel et al. 2014), FMOS-COSMOS (Kashino et al. 2019), and 3D-HST (Brammer et al. 2012a). Initial work to establish the metallicities of $z > 1$ galaxies across a wide range of M$_*$ and SFR has taken place (e.g., Erb et al. 2006; Cullen et al. 2014; Steidel et al. 2014; Troncoso et al. 2014; Zahid et al. 2014b; Sanders et al. 2015, 2018; Kashino et al. 2017). However, efforts to utilize these large data sets to characterize the metallicity evolution of galaxies as a function of redshift have been hindered by systematic uncertainties associated with the use of $z$-strong-line metallicity calibrations at high redshifts.

The strong-line ratio properties of $z > 1$ star-forming galaxies suggest evolution in the physical conditions of the ionized gas in H II regions (e.g., Kewley et al. 2013a,b; Masters et al. 2014; Steidel et al. 2014; Shapley et al. 2015; Sanders et al. 2016a; Kashino et al. 2017; Strom et al. 2017, 2018). In particular, there is evidence that the ionization parameter, hardness of the ionizing spectrum, ISM density/pressure, and/or N/O ratio evolve with redshift at fixed O/H. Much focus has been given to the position of $z \sim 2$ galaxies in the [O III]λλ5007/Hβ vs. [N II]λλ6584/Hα (BPT; Baldwin et al. 1981) diagram, which is systematically offset from the $z \sim 0$ star-forming sequence (e.g. Shapley et al. 2005; Erb et al. 2006; Hainline et al. 2009; Steidel et al. 2014, 2016; Shapley et al. 2015; Sanders et al. 2016a; Strom et al. 2017, 2018). As a result of these evolving H II region physical conditions, metallicity calibrations produced from $z \sim 0$ empirical datasets or from photoionization models assuming $z \sim 0$ ISM conditions will yield systematically biased metallicities when applied to high-redshift samples. The precise nature of this bias is not currently known given the lack of a consensus on the nature of the ISM conditions at high redshifts.

The most straightforward approach to resolving the problem of measuring metallicities at high redshifts is to obtain determinations of metallicity via the electron temperature method, independent of $z \sim 0$ strong-line calibrations, to characterize the bias of $z \sim 0$ calibrations and understand which low-redshift samples most reliably reproduce high-redshift metallicities. Obtaining electron temperature measurements at $z > 1$ has been observationally difficult due to the extreme faintness of temperature-sensitive auroral emission lines, but efforts over the past decade have produced a sample of $> 10$ auroral-line measurements at $z \sim 1–3$ (Villar-Martín et al. 2004; Yuan & Kewley 2009; Brammer et al. 2012b; Christensen et al. 2012a,b; Stark et al. 2013, 2014; Bayliss et al. 2014; James et al. 2014; Sanders et al. 2016b; Kojima et al. 2017; Berg et al. 2018; Patricio et al. 2018). Many of these $z > 1$ auroral-line detections have taken advantage of strong gravitational lensing to boost line fluxes, or otherwise target particularly bright line emitters in which auroral lines can be detected with current facilities.

In this paper, we present three new detections of [O III]λ4363 and derive temperature-based metallicities for galaxies at $z = 1.7$–$3.6$ from the MOSDEF survey (Kriek et al. 2015), increasing the number of galaxies at $z > 1$ with [O III]λ4363 detections by ∼50%. We combine these measurements with targets from the literature to construct the largest sample of $z > 1$ galaxies with direct-method metallicities to-date ($N = 18$). We utilize this $z > 1$ auroral-line sample to investigate the evolution of strong-line calibrations with redshift, construct metallicity scaling relations based purely on the direct method at $z > 1$ for the first time, and place constraints on the ionization parameters and ionizing spectra present in these galaxies via photoionization modeling.

This paper is organized as follows. In Section 2, we describe the observations, $z > 1$ auroral-line sample, and derived galaxy properties. Using temperature-based metallicities, we characterize the evolution of strong-line metallicity calibrations and metallicity scaling relations, including the mass-metallicity and fundamental metallicity relations in Section 3. In Section 4, we leverage the unique combination of direct-method metallicities and photoionization models to constrain the ionization state of the ionized ISM in star-
forming galaxies at \( z > 1 \). However, our sample of high-redshift auroral-line emitters is not representative of typical \( \sim L^* \) galaxies at \( z \sim 1–3 \). In Section 5, we discuss the implications of our results for representative spectroscopic samples at \( z \sim 2 \), extreme emission-line galaxies at \( z \sim 1–2 \), and typical galaxies in the epoch of reionization at \( z > 6 \). Finally, we summarize our results and conclusions in Section 6.

We adopt the following abbreviations for strong emission-line ratios:

\[
\begin{align*}
O3 &= [O \text{ III}] \lambda 5007/\lambda H\beta \\
O2 &= [O \text{ III}] \lambda 3726,3729/\lambda H\beta \\
Ne3 &= [Ne \text{ III}] \lambda 3869/\lambda H\beta \\
R23 &= ([O \text{ III}] \lambda 4959,5007+[O \text{ II}] \lambda 3726,3729)/\lambda H\beta \\
O32 &= [O \text{ III}] \lambda 5007/[O \text{ II}] \lambda 3726,3729 \\
Ne3O2 &= [Ne \text{ III}] \lambda 3869/[O \text{ II}] \lambda 3726,3729
\end{align*}
\]

Emission-line wavelengths are given in air, and all magnitudes are on the AB scale (Oke & Gunn 1983). We assume a CDM cosmology with \( H_0 = 70 \text{ km s}^{-1}\text{Mpc}^{-1} \), \( \Omega_m = 0.3 \), and \( \Omega_{\Lambda} = 0.7 \).

2 DATA, MEASUREMENTS, & DERIVED QUANTITIES

2.1 The MOSDEF Survey

The MOSDEF Survey was a four-year program that utilized the MOSFIRE spectrograph (McLean et al. 2012) on the 10 m Keck I telescope to obtain near-infrared (rest-frame optical) spectra of \( \sim 1500 \) galaxies at \( 1.37 \leq z \leq 3.80 \) (Kriek et al. 2015). Targets were selected in the five CANDELS fields (AEGIS, COSMOS, GOODS-N, GOODS-S, and UDS; Grogin et al. 2011; Koekemoer et al. 2011) down to limiting magnitudes of \( H = 24.0, 24.5, \) and \( 25.0 \) at \( 1.37 \leq z \leq 1.70, 2.09 \leq z \leq 2.61, \) and \( 2.95 \leq z \leq 3.80 \), respectively, as determined from \( HST/WFC3 \) F160W imaging. Strong rest-optical emission lines fall in near-infrared windows of atmospheric transmission in these three redshift ranges. Target selection was performed using the photometric and spectroscopic catalogs of the 3D-HST survey (Brammer et al. 2012a; Skelton et al. 2014;Momcheva et al. 2016) based on observed \( H \)-band magnitudes and redshifts (spectroscopic when available, otherwise photometric). Completed in 2016, the full survey targeted \( \sim 1500 \) galaxies and yielded \( \sim 1300 \) robust redshifts: \( \sim 300 \) at \( z \sim 1.5, \sim 700 \) at \( z \sim 2.3, \) and \( \sim 300 \) at \( z \sim 3.4 \). Exposure times were typically 1 h per filter for masks targeting galaxies in the lowest redshift bin at \( z \sim 1.5, \) and 2 h per filter for \( z \sim 2.3 \) and \( z \sim 3.4 \) masks. A custom IDL pipeline was used to reduce the data and produce two-dimensional science and error spectra for each slit on a mask that has been flat-fielded, sky-subtracted, cleaned of cosmic rays, rectified, and wavelength- and flux-calibrated. One-dimensional science and error spectra were optimally extracted (Horne 1986) using BNEP (Freeman et al. 2019) and corrected for slit losses based on the \( HST/WFC3 \) F160W image for each object convolved with the average seeing for each filter and mask combination (Reddy et al. 2015). Full details of the survey design, execution, and data reduction can be found in Kriek et al. (2015).

2.2 Measurements & Derived Quantities

We measured emission-line fluxes by fitting Gaussian profiles to the one-dimensional science spectra, and flux uncertainties are estimated using a Monte Carlo method by perturbing the science spectrum according to the error spectrum and remeasuring the flux 1,000 times. The uncertainty in the line flux was taken to be the 68th percentile half-width of the resulting distribution. Redshifts were measured from the highest signal-to-noise (S/N) ratio line, almost always Ha or [O III] \( \lambda 5007 \).

Stellar masses (\( M_\ast \)) were estimated by fitting flexible stellar population synthesis models (Conroy et al. 2009) to rest-frame UV through near-infrared photometry using the spectral energy distribution (SED) fitting code FAST (Kriek et al. 2009). We assumed solar metallicity, constant star-formation histories, the Calzetti et al. (2000) attenuation curve, and a Chabrier (2003) initial mass function (IMF). Redshift is fixed to the spectroscopic redshift determined from strong optical emission lines. Photometry was corrected for emission line contamination based on the measured line fluxes prior to fitting.

The best-fit SED model was used to correct for the effects of stellar absorption on hydrogen Balmer emission line fluxes. Emission-line equivalent widths were measured using the observed line flux relative to the continuum level inferred from the best-fit SED model at the line center (Reddy et al. 2018). We calculated nebular reddening using the available hydrogen Balmer emission lines, assuming a Cardelli et al. (1989) extinction curve and intrinsic ratios of Ha/\( H\beta = 2.86, H\gamma/\lambda H\beta = 0.468, \) and H\( \delta/\lambda H\beta = 0.259 \) (Osterbrock & Ferland 2006). Star-formation rates (SFRs) were calculated from the dust-corrected luminosity of the highest S/N hydrogen Balmer line (always Ha or H\( \beta \)) using the solar metallicity calibration of Hao et al. (2011) converted to a Chabrier (2003) IMF.

For density and [O III] \( \lambda 4363 \) temperature calculations, we adopted the effective collision strengths of Tayal (2007) for O III and Storey et al. (2014) for O III, with transition probabilities for both species taken from the NIST MCHF database (Fischer & Tachiev 2014). For temperature calculations using [O III] \( \lambda \lambda 4361,1666 \), the Storey et al. (2014) collision strengths cannot be used since they include only five levels, and we instead employed the six-level O III collision strengths from Aggarwal & Keenan (1999). Calculations of electron temperature from [O III] \( \lambda 4363 \) using either set of O III collision strengths fall within \( \sim 3\% \) of each other. When both components of the [O III] \( \lambda 3726,3729 \) doublet are detected with S/N \( \gtrsim 3 \), electron density was calculated using a five-level atom approximation (Sanders et al. 2016a) and solved iteratively with the electron temperature. Otherwise, an electron density of \( n_e = 250 \text{ cm}^{-3} \) was assumed, \(^2\) a value typical of star-forming galaxies at \( z \sim 2 \) (e.g., Sanders et al. 2016a; Steidel et al. 2016; Strom et al. 2017). The electron temperature of the O\( ^{2+} \) zone (T3) was

---

\(^1\) Source code and installation instructions available at https://github.com/billfreeman44/bmep

\(^2\) When \( n_e < 3,000 \text{ cm}^{-3} \), changes in density affect inferred O\( ^{2+} \) electron temperatures by at most 1%.
determined using the IRAF routine `nebular.tendens` (Shaw & Dufour 1994). The temperature of the O\(^+\) zone (T\(_2\)) was estimated assuming the T\(_3\)-T\(_2\) relation of Campbell et al. (1986): T\(_2\) = 0.7T\(_3\) + 3,000 K. The equations of Izotov et al. (2006) were used to calculate ionic abundances. The total oxygen abundance was taken to be O/H = O\(^+\)/H + O\(^{2+}\)/H, assuming neutral and O\(^{3+}\) contributions are negligible within the ionized nebula.

Uncertainties on properties calculated from emission line strengths were estimated using a Monte Carlo method in which the measured line fluxes were perturbed according to their uncertainties and all such properties (reddening, equivalent widths, SFR, line ratios, density, temperature, and abundances) were remeasured 1000 times. The uncertainty on each property was taken to be the 16th and 84th percentile of the resulting distribution. In this way, uncertainties on line ratios, SFRs, temperatures, and abundances include errors on line fluxes and nebular reddening. Uncertainties on stellar masses were estimated by perturbing the photometry according to the errors and refitting the SED 500 times.

For objects taken from the literature, all properties excluding stellar masses and equivalent widths were calculated from the observed emission-line fluxes using the same method described above, such that abundances and line ratios can be fairly compared. Any exceptions to this methodology are noted in Appendices A1 and A2. Stellar masses and equivalent widths were taken from literature sources, with all masses converted to a Chabrier (2003) IMF.

### 2.3 Sample Selection

#### 2.3.1 MOSDEF [O III]\4363 emitters

We visually inspected the spectra of all MOSDEF targets with robust redshifts and wavelength coverage of 4364.436 \(\AA\) (the vacuum wavelength of [O III]\4363) and searched for detections of the line. We found that several galaxies with formal S/N>2 for [O III]\4363 in fact have spurious detections due to noise associated with sky lines. After culling such false-positives, this search yielded 6 detections of [O III]\4363, the spectra of which are presented in Figure 1. The 6 sources are star-forming galaxies spanning \(z = 1.67 - 3.63\) with [O III]\4363 significance of 2.0-7.7\(\sigma\). While two of the galaxies (GOODS-S-41547 and COSMOS-23895) have low-S/N detections of [O III]\4363 (2.0\(\sigma\) and 2.6\(\sigma\), respectively), these line identifications were made in conjunction with a precise determination of the systemic redshift and line width from multiple strong lines. Additionally, we required identification of both positive and negative traces for [O III]\4363 at the correct \(\gamma\)-pixel locations (Fig. 1). None of these galaxies host an active galactic nucleus (AGN) according to their X-ray and rest-frame infrared properties (Coll et al. 2015; Azadi et al. 2017). All six targets are low-mass (\(log(\mathcal{M}_*/\mathcal{M}_\odot) \lesssim 9.5\)) and highly star-forming (\(log(sSFR/\mathcal{M}_\odot yr^{-1}) \gtrsim 0.75\)).

We cannot determine direct-method metallicities for two of the six targets because of wavelength coverage: AEGIS-32546 at \(z = 2.01\) lacks coverage of H\(\alpha\), H\(\beta\), and [O I]; GOODS-S-35910 at \(z = 2.65\) lacks coverage of [O III]\44959,5007 and [O I]. The other four galaxies have detections of [O III]\4363, [O III]\44959,5007, [O II], H\(\beta\), and at least one other hyrogen Balmer line, comprising the MOSDEF [O III]\4363 emitter sample for which robust determinations of nebular reddening, electron temperature, and nebular oxygen abundance can be obtained. The emission-line fluxes of these four galaxies are given in Table 1, and their properties are presented in Table 2. The [O III]\4363 detection for COSMOS-1908 was previously published in Sanders et al. (2016b). We note that the emission-line fluxes and properties of COSMOS-1908 presented in this work differ slightly due to updates to the data reduction pipeline and SED fitting procedure.

We produced a composite spectrum of the four MOSDEF [O III]\4363 emitters following the median stacking methodology described in Sanders et al. (2018), with the exception that we normalized the spectra by [O III]\45007 flux.
instead of Hα flux to prevent the composite electron temperature measurement from being dominated by the brightest [O III]4950.07 emitter in the sample. The composite spectrum is displayed in Figure 2. The significance of [O III]4363 in the stacked spectrum is 6.2σ. The M∗ and SFR of the composite were taken to be the median M and SFR of the individual galaxies. Emission-line equivalent widths were derived for the composite by dividing the composite line luminosity by the median luminosity density at line center of the set of best-fit SED models of galaxies in the composite. We have confirmed that this methodology reproduces the average equivalent widths of samples of individual MOSDEF galaxies for which the lines of interest are detected. The emission line luminosities and properties derived from the MOSDEF [O III]4363 emitter composite spectrum are given in Tables 1 and 2, respectively.

2.3.2 Literature [O iii]λ4363 and O iii]λλ1661,1666 emitters

We have searched the literature for galaxies at z > 1 with detections of [O III]4363, and found 4 targets for which detections of the requisite lines to determine both dust reddening and total oxygen abundance are additionally available (Brammer et al. 2012b; Christensen et al. 2012b; Stark et al. 2013; James et al. 2014). Based on the work presented in Kojima et al. (2017) and Patrício et al. (2018), we have additionally compiled a sample of 12 star-forming galaxies at z > 1 from the literature with detections of the [O III]λλ1661,1666 doublet (hereafter O III]1663) with the additional detection of lines necessary to determine both the reddening correction and total oxygen abundance (Villar-Martín et al. 2004; Christensen et al. 2012a,b; Bayliss et al. 2014; Stark et al. 2014; Erb et al. 2016; Kojima et al. 2017; Berg et al. 2018). The stellar masses and SFRs of literature targets that are gravitationally lensed have been corrected for magnification, and include uncertainties on the magnification when available. There are 3 literature sources that have both [O III]4363 and O III]1663 detections (J14, C12a, and B18). The metallicities based on either auroral line are consistent with one another in each case. We calc-

![Figure 2](image-url)

**Figure 2.** Composite spectrum of four MOSDEF targets with detections of [O III]4363 and wavelength coverage of [O II]λλ3726,3729, Hβ, and [O III]λλ4959,5007. The gray shaded region denotes the composite error spectrum. The left and middle panels show the strong [O III]4363 emitter composite spectrum are reddening-corrected and normalized to [O III]45007. The significance of [O III]4363 is 6.2σ in the composite spectrum.

### Table 1. Observed emission-line fluxes of MOSDEF [O iii]4363 emitters in units of 10^{-17} erg s^{-1} cm^{-2}.

| Object       | [O II]43726 | [O II]43729 | [Ne II]4969 | Hα  | [O II]4363 | Hβ   | [O III]5007 | Hα  |
|--------------|-------------|-------------|-------------|-----|------------|------|------------|-----|
| AEGIS 1452   | 2.57±0.50   | 4.32±0.66   | 1.92±0.30   | 1.93±0.40 | 3.29±0.64 | 0.82±0.28 | 7.54±0.35 | 36.1±0.29 | 19.7±0.39 |
| GOODS-S 13.3a1.41  | 15.6±0.75   | —           | —           | 5.56±1.79 | 1.40±0.70 | 11.1±1.33 | 57.5±0.53 | 31.7±0.83 | —      |
| COSMOS 1908  | 1.06±0.16   | 1.22±0.14   | 1.74±0.19   | 1.22±0.36 | 2.29±0.28 | 0.53±0.12 | 4.82±0.25 | 33.8±0.65 | —      |
| COSMOS 23895 | 2.17±0.35   | 2.18±0.22   | —           | —         | 1.57±0.30 | 0.26±0.10 | 1.96±0.21 | 13.1±0.28 | —      |
| composite^a  | 0.10±0.01   | 0.17±0.01   | —           | —         | 0.08±0.007 | 0.02±0.003 | 0.17±0.008 | 1.00±0.007 | —      |

^a Line fluxes of the composite spectrum of the four MOSDEF [O III]4363 emitters are reddening-corrected and normalized to [O III]45007.
Table 2. Derived properties of the \( z > 1 \) auroral-line sample.

| Object   | \( z \) | \( \log (\frac{M_{\text{SFR}}}{M_\odot}) \) | \( \log (\frac{\text{SFR}}{\text{Gyr}^{-1}}) \) | \( E(\text{B-V})_{\text{gas}} \) | \( n_e \) | \( T_\text{e} \) | \( 12+\log (\text{O/H}) \) | \( \text{EW}_0([\text{O} \text{iii}]4363) \) | \( \text{EW}_0(\text{H}\beta) \) |
|----------|-------|-----------------|-----------------|-----------------|-------|--------|-----------------|-----------------|-----------------|
| MOSDEF [O iii]4363 |
| AEGIS 11452 | 1.6715 | 9.46±0.08 | 4.09±0.10 | 0.75±0.38 | 0.00±0.00 | 0.00±0.00 | <381 | 16200±300 | 7.7±0.10 | 348±3 | 105±5 |
| GOODS-S 41547 | 2.5451 | 9.3±0.13 | 4.09±0.10 | 1.58±0.14 | 0.27±0.11 | 0.00±0.00 | 278±194 | 16800±400 | 7.84±0.33 | 448±4 | 110±13 |
| COSMOS 1908 | 3.0767 | 8.92±0.01 | 4.09±0.10 | 1.79±0.37 | 0.00±0.23 | 0.00±0.00 | 280±373 | 13700±1800 | 8.92±0.13 | 1845±36 | 343±18 |
| COSMOS 23895 | 3.6372 | 9.43±0.13 | 4.09±0.10 | 1.07±0.08 | 0.00±0.00 | 0.00±0.00 | 531±348 | 15300±2800 | 7.99±0.26 | 491±11 | 98±11 |
| composite | 2.81±0.14 | 9.3±0.01 | 4.09±0.10 | 1.25±0.09 | 0.00±0.00 | 0.00±0.00 | <543 | 15400±2500 | 7.90±0.09 | 770±231 | 129±36 |

| Literature [O iii]4363 |
|---------------------|
| S13 | 1.425 | 8.33±0.14 | 7.1 | 1.48±0.16 | 0.03±0.08 | 0.00±0.00 | 201±12 | 15400±1100 | 7.95±0.07 | >320 | — |
| J14 | 1.433 | 5.1 | 1.38 | 1.44±0.54 | 0.31±0.32 | — | 171±34 | 20900±5000 | 7.46±0.23 | 482±124 | 100±26 |

| Literature [O iii]41663 |
|---------------------|
| S14a | 1.7024 | 7.78±0.11 | 0.21 | 0.52±0.16 | 0.03±0.03 | — | 13300±99 | 8.04±0.04 | 1033±100 | 172±20 |
| B18c | 1.8443 | 8.26±0.12 | 1.24 | 1.87±0.12 | 0.00±0.00 | — | 15400±166 | 7.52±0.08 | 1575±75 | 520±40 |
| C12b | 1.9634 | 10.57±0.08 | 10.2 | 0.5±0.12 | 0.23±0.07 | 82±48 | 12400±400 | 8.14±0.04 | — | — |
| S14c | 2.0596 | 7.44±0.12 | 1.61 | 1.80±0.12 | 0.00±0.00 | — | 22800±770 | 7.30±0.08 | — | — |
| K17 | 2.159 | 9.24±0.15 | 14.1 | 0.9±0.11 | 0.00±0.00 | — | 12300±590 | 8.28±0.30 | — | — |
| E16c | 2.1742 | 9.73±0.1 | 30.1 | 0.75±0.08 | 0.00±0.00 | — | 13800±700 | 7.98±0.08 | 105±34 | 170±10 |
| E16a | 2.1889 | 9.79±0.07 | 67.1 | 1.10±0.07 | 0.37±0.07 | — | 17500±200 | 7.90±0.12 | 698±68 | 80±15 |
| E16b | 2.3054 | 9.47±0.1 | 59.1 | 1.32±0.11 | 0.34±0.04 | — | 19000±2700 | 7.79±0.10 | 759±89 | 102±15 |
| V04 | 3.357 | 7.67±0.10 | 48.1 | 3.01±0.18 | 0.00±0.00 | — | 17600±400 | 7.76±0.04 | — | — |
| C12c | 3.5073 | 9.1±0.16 | 4.8 | 0.39±0.08 | 0.00±0.00 | — | 415±301 | 16200±690 | 7.76±0.03 | — | — |
| B14 | 3.6252 | 9.3±0.15 | 27.1 | 0.93±0.16 | 0.32±0.08 | 560±278 | 13700±1000 | 8.10±0.08 | — | — |
| B14L1 | 2.4036 | 9.8±0.10 | 28.1 | 0.65±0.13 | 0.24±0.05 | 404±97 | 12400±400 | 8.14±0.03 | — | — |

\( a \) 3σ upper limit, \( b \) Median redshift, \( M_\text{e} \), and SFR of the galaxies in the composite. sSFR is calculated from the median \( M_\text{e} \) and SFR. \( c \) Gravitationally-lensed. \( d \) Also has a \( [\text{O} \text{iii}]41663 \) detection reported in James et al. (2014). We adopt the electron temperature and \( \text{O}/\text{H} \) derived from \( [\text{O} \text{iii}]4363 \) (see Appendix A). \( e \) Also has a \( [\text{O} \text{iii}]4363 \) detection reported in Christensen et al. (2012b). We adopt the electron temperature and \( \text{O}/\text{H} \) derived from \( [\text{O} \text{iii}]4463 \) that yields smaller uncertainties. \( f \) Also has a \( [\text{O} \text{iii}]4463 \) detection reported in Brammer et al. (2012b). We adopt the electron temperature and \( \text{O}/\text{H} \) derived from \( [\text{O} \text{iii}]1463 \) that yields smaller uncertainties. \( g \) Lower limit on the total stellar mass and upper limit on sSFR. See Appendix A. \( h \) Composite spectrum of 30 galaxies from the KBSS survey, presented in Steidel et al. (2016). SFR and sSFR are calculated using the dust-corrected Hα luminosity of the composite.
ulate the electron temperature and oxygen abundance using [O iii]λ4363 for J14, [O iii]λ4363 for C12a, and O iii]λ41663 for B18 (see Appendix A). The derived properties of the literature targets are presented in Table 2, and detailed descriptions of the calculations for each literature source are provided in Appendix A.

The literature $z > 1$ auroral sample thus comprises 14 individual galaxies, and the total $z > 1$ auroral-line sample size is 18 when combined with the MOSDEF [O iii]λ4363 emitters. The significance of the auroral-line detections ([O iii]λ4363 or O iii]λ41663) of the total sample spans 2.0 to 103σ, with a median value of 4.8σ. Only two objects have S/N < 2.5 (GOODS-S-41547 and J14), and our results do not significantly change if these objects are excluded. For comparison, we additionally include the KBSS-LM1 composite rest-UV and rest-optical spectrum of 30 galaxies at $2.1 < z < 2.6$ from Steidel et al. (2016), with a 8.6σ O iii]λ41663 detection. The stellar mass of the KBSS-LM1 composite was taken to be the median mass of the individual galaxies. Composites are not included in any fitting function or sample median calculations.

### 3 DIRECT-METHOD METALLICITY CALIBRATIONS & SCALING RELATIONS AT $z > 1$

In this section, we use the largest temperature-based metallicity sample at $z > 1$ to-date to assess the applicability of several strong-line metallicity indicators from $z ~ 0$ to 3.5, construct the mass-metallicity relation based on the direct method for the first time at $z > 1$, and investigate dependence of O/H on SFR.

#### 3.1 Metallicity calibrations at $z > 1$

Using our sample of 18 individual galaxies at $z > 1$ with auroral-line detections and temperature-based metallicities, we investigate whether strong-line metallicity calibrations change with redshift. Recently, Patrício et al. (2018) used a sample of 16 $z > 1$ auroral-line emitters to test the reliability of various sets of strong-line metallicity calibrations at high redshifts by comparing the derived metallicity from each calibration to the temperature-based metallicity. Our approach differs from that of Patrício et al. in that we do not test specific sets of parameterized calibrations, but instead perform a purely empirical comparison of strong-line ratios at fixed direct-method metallicity between various samples to understand which calibrating samples best match the high-redshift galaxies. This empirical approach was employed in earlier works by Jones et al. (2015) at $z ~ 0.8$ and Sanders et al. (2016b) at $z ~ 1$ to 3. We perform our analysis on individual galaxies at $z > 1$ and only include stacks (MOSDEF [O iii]λ4363 emitters and KBSS-LM1) for comparison.

We present strong-line ratios as a function of temperature-based oxygen abundance for the $z > 1$ auroral-line sample in Figure 3. Galaxies are color-coded by redshift, while shapes encode whether [O iii]λ4363 or O iii]λ41663 was used to derive the nebular metallicity. Here, we investigate the behavior of five line ratios: O3, O2, R23, O32, and Ne3O2. We do not explore the strong-line ratios involving [N ii] at this time because only 4 $z > 1$ galaxies have detections of [N ii]λ6584, while over half of the sample (10/18) does not have spectral coverage of [N ii]. It is therefore not possible to draw confident conclusions about nitrogen-based metallicity indicators with the current $z > 1$ auroral-line sample. The five line-ratios presented in Figure 3 have the advantage of being accessible from the ground to $z = 3.80$, and will be observable with JWST/NIRCam and NIRSpec spectroscopy to $z > 9$.

The small sample size ($N = 18$), large typical uncertainties in O/H, and small dynamic range in metallicity ($7.5 < 12 + \log(O/H) < 8.1$), 0.07 - 0.25 $Z_\odot$) of the $z > 1$ galaxies precludes the construction of new calibrations based on the high-redshift sample alone. While we cannot reliably constrain the shape of high-redshift metallicity calibrations, we can use mean sample properties to look for evolution in the normalization of calibrations at $12 + \log(O/H) - 7.9$ ($0.16 Z_\odot$). In each panel of Figure 3, we calculate the median $12 + \log(O/H)$ and line ratio of $z > 1$ galaxies with detections of that line ratio, displayed as black open diamonds. Limits and stacks are not included. We note that there are at most 3 limits in any particular panel. We determine the uncertainty on median values by bootstrap resampling, perturbing the data points according to the uncertainties, and remeasuring the median 500 times. The 1σ confidence bounds are taken from 68th percentile width of the resulting distribution. We compare the $z > 1$ sample and medians to low- and intermediate-redshift samples to investigate potential evolution: individual $z = 0$ H II regions (black line; Pilyugin & Grebel 2016; Sanders et al. 2017), $z ~ 0$ star-forming galaxies from the Sloan Digital Sky Survey (SDSS, gray points; Izotov et al. 2006), and $z ~ 0.8$ star-forming galaxies from DEEP2 (orange points; Jones et al. 2015). The oxygen abundances of each sample have been recalculated using our methodology.

Considering the $z > 1$ sample alone, we find that O3 and R23 are saturated over the range of metallicity spanned by the bulk of the high-redshift galaxies ($7.7 < 12 + \log(O/H) < 8.1$; 0.1 - 0.25 $Z_\odot$), which reside in the turnover regime between the “upper” and “lower” branches of these indices. O3 and R23 thus cannot be utilized as reliable metallicity indicators for high-redshift galaxies falling in this oxygen abundance range, even when used in combination with an additional line-ratio to break the upper/lower branch degeneracy. Earlier works have highlighted similar issues with obtaining metallicities from R23 at $z ~ 2$ (Steidel et al. 2014; Sanders et al. 2018). The $z > 1$ galaxies display a wide dynamic range in O2, but the scatter in line ratio at fixed O/H is largest for O2, suggesting that O2 is a poor metallicity indicator at $z > 1$ as well. The large scatter in O2 at fixed O/H may signify significant galaxy-to-galaxy variation in the nebular attenuation curve at $z > 1$.

In contrast, O32 and NeO2 have a monotonic dependence on direct-method metallicity for all of the samples displayed in Figure 3. The number of $z > 1$ galaxies with [Ne iii] detections is small and thus scatter is difficult to evaluate, but the scatter in O32 at fixed O/H of the $z > 1$ sample is comparable to that of the lower-redshift samples after accounting for measurement uncertainties (typically much larger at $z > 1$). O32 thus appears to be the best strong-line metallicity indicator for $z > 1$ samples with nebular metallicities $\leq 1/3 Z_\odot$, displaying the highest utility over the range $z > 1$. 

---

**References**

1. Patrício, J. et al. (2018). A&A, 599, A7.
2. Jones, A. B. et al. (2015). ApJ, 804, 12.
3. Steidel, C. C. et al. (2014). A&A, 568, A114.
4. Sanders, D. B. et al. (2017). ApJ, 843, 142.
5. Izotov, I. Y. et al. (2006). ApJ, 642, 279.
6. Pilyugin, L. I. & Grebel, E. K. (2016). MNRAS, 455, 2047.
7. Sanders, D. B. et al. (2018). ApJ, 864, 132.
of metallicities in our sample. Given the tight relation between O32 and Ne3O2 at sub-solar metallicities (Levesque & Richardson 2014; Strom et al. 2017), Ne3O2 should perform equally well and has the additional advantage of not requiring a reddening-correction.

The $z > 1$ auroral-line sample displays significantly higher excitation than local H II regions at fixed O/H, with higher O3, O32, and Ne3O2 and lower O2. The $z > 1$ sample also has significantly higher R23 values than are typical of the local H II regions. While in reasonable agreement at 12+log(O/H) ~ 8.3, the H II regions diverge from the $z \sim 0$ and $z \sim 1$ galaxy samples at 12+log(O/H) < 8.0, flattening in O32 and Ne3O2 and dropping off steeply in O3 and R23. This divergent behavior may indicate that the local H II region sample is significantly incomplete in the metal-poor regime, lacking high-excitation star-forming regions that must be present in star-forming galaxies with high O32. We caution against utilizing empirical calibrations based on individual local H II regions at oxygen abundances below 0.2 Z⊙ (12+log(O/H)=8.0) for either low- or high-redshift galaxies.

Bian et al. (2018) recently constructed metallicity calibrations for use at high redshifts by stacking spectra of $z \sim 0$ galaxies from SDSS that fall in the same region of the BPT diagram as $z \sim 2$ galaxies, and measuring temperature-based metallicities from these stacks. Their approach differs from past studies that have selected analogs based on global galaxy properties (e.g., sSFR; Brown et al. 2016; Cowie et al. 2016) in that Bian et al. have selected high-redshift analogs based directly on strong-line ratios that should be more closely linked to the local gas conditions in H II regions. Stacks of a local reference sample that traces the average $z \sim 0$ BPT sequence were also constructed. Functional fits to the Bian et al. (2018) high-redshift analogs and $z \sim 0$ reference sample are shown in Figure 3 (cyan and magenta lines, respectively). We find that the median of the $z > 1$ sample (black diamonds) falls above the $z \sim 0$ reference sample and below the high-redshift analogs in O3, O2, O32, and Ne3O2.
at fixed O/H, being ~1σ consistent with both samples. However, the z > 0 reference sample does not have R23 as high as the z > 1 sample even at the maximum value, while the high-redshift analogs closely match the z > 1 median R23 at fixed oxygen abundance.

Interestingly, the Izotov et al. (2006) z ~ 0 SDSS sample does not follow the Bian et al. (2018) local reference sample, but instead lies in between the reference and high-redshift analog lines. This trend is due to the requirement of a [O III] λ4363 detection by Izotov et al. (2006), which preferentially selects high-sSFR and high-excitation galaxies with brighter [O III] λ4363. Such galaxies likely have ISM conditions that are shifted away from the local average and towards what is typical at high redshift. Thus, the Izotov et al. (2006) sample is not representative of z ~ 0 galaxies, but instead matches the z > 1 sample suitably well within the current uncertainties. Jones et al. (2015) have previously shown that their [O III] λ4363 sample at z ~ 0.8 matches the Izotov et al. (2006) sample in these line-ratio vs. O/H spaces, and we find that the z ~ 0.8 galaxies also agree with the z > 1 sample on average.

In summary, the median O3, O2, R23, O32, and Ne3O2 of 18 individual z > 1 galaxies agree with the values for the Izotov et al. (2006) z ~ 0, Jones et al. (2015) z ~ 0.8, and Bian et al. (2018) high-redshift analog samples at fixed temperature-based O/H. In contrast, the z > 1 sample displays systematically higher excitation at fixed O/H compared to the SDSS local reference sample and individual z ~ 0 H II regions. Calibrations based on z = 0 H II regions are not reliable at z > 1 for 12+log(O/H) ≤ 8.0 (2.0 Z⊙). Calibrations based on typical z ~ 0 galaxies tend to underestimate O/H, although this bias is typically small (~0.1 dex). The Bian et al. (2018) high-redshift analog O32 and Ne3O2 calibrations display the highest utility for application at high redshifts, reproducing the metallicities of our z > 1 auroral-line sample to within ~0.1 dex on average while maintaining a monotonic sensitivity to O/H.

We stress that these conclusions only apply over the range of metallicities probed by our z > 1 auroral-line sample (7.5 ≤ 12+log(O/H) ≤ 8.1: 0.07–0.25 Z⊙), and only for galaxies with similar properties (i.e., low M*, and high sSFR). The z > 1 sample presented in Figure 3 is not representative of typical galaxies at z ~ 1–3 in large rest-optical spectroscopic surveys (e.g., MOSDEF, KBSS, 3D-HST, FMOS-COSMOS), which have log(M*/M⊙) ~ 10.0 and lie on the mean M–sSFR relation (Kriek et al. 2015; Steidel et al. 2014; Brammer et al. 2012a; Kashino et al. 2019). Based on commonly applied z ~ 0 strong-line methods, the majority of z ~ 2 galaxies at log(M*/M⊙) > 9.5 are expected to have higher O/H than our z > 1 auroral-line sample (Steidel et al. 2014; Sanders et al. 2015, 2018), even if there is a factor of two systematic uncertainty. We discuss implications for the typical z ~ 1–3 galaxy population in Section 5.1. Additionally, our conclusions regarding calibration evolution are based on the mean properties of the z > 1 sample, and a detailed analysis of the scatter awaits a larger sample. We thus caution that the stated ~0.1 dex precision of the Bian et al. (2018) high-redshift analog calibrations is applicable for population averages only, and that the metallicity of an individual galaxy may not be well-determined using this method.

### 3.2 Direct-method metallicity scaling relations at z > 1

The z > 1 auroral-line sample spans a wide dynamic range in stellar mass, with log(M*/M⊙) = 7.5 – 10.0. We use this sample to explore the dependence of gas-phase oxygen abundance on stellar mass based entirely on direct-method metallicities for the first time at z > 1, and constrain the evolution of the mass-metallicity relation (MZR) from z ~ 0 to z ~ 2. Two galaxies in the z > 1 auroral-line sample (C12b and C12c) are identified as possible ongoing mergers due to morphology and close companions (Christensen et al. 2012a; Patrício et al. 2016). Mergers in the local universe have been shown to deviate from the mean z ~ 0 MZR, having systematically lower metallicities and higher SFRs at fixed M* (Ellison et al. 2008; Scudder et al. 2012). In contrast, recent theoretical and observational work suggests that the SFR and O/H of close galaxy pairs may not be systematically offset from isolated galaxies at z ~ 2 (Fensch et al. 2017; Wilson et al. 2019). However, it is not clear if the two potential mergers in our sample fall into the merger stage and pair separation on which such conclusions are valid. To avoid introducing systematic biases from these merging systems, we exclude them from the sample used to construct metallicity scaling relations with global galaxy properties (SFR and M*). We show these two excluded mergers as open triangles in the figures in this section and discuss their behavior in comparison to the rest of the z > 1 sample where appropriate. We also exclude J14, which lacks a reliable M* estimate, and V04, which does not have a total M* determination available (see Appendix A).

In the sections below, we investigate the shape and evolution of the MZR and its dependence on SFR. In Section 3.2.1, we present the MZR at z ~ 2.2 based purely on direct-method metallicities. We begin with the correlation between direct-method O/H and M* and the mass-metallicity relation (MZR) from Ellison et al. (2008; Scudder et al. 2012). In contrast, recent theoretical and observational work suggests that the SFR and O/H of close galaxy pairs may not be systematically offset from isolated galaxies at z ~ 2 (Fensch et al. 2017; Wilson et al. 2019). However, it is not clear if the two potential mergers in our sample fall into the merger stage and pair separation on which such conclusions are valid. To avoid introducing systematic biases from these merging systems, we exclude them from the sample used to construct metallicity scaling relations with global galaxy properties (SFR and M*). We show these two excluded mergers as open triangles in the figures in this section and discuss their behavior in comparison to the rest of the z > 1 sample where appropriate. We also exclude J14, which lacks a reliable M* estimate, and V04, which does not have a total M* determination available (see Appendix A).

### 3.2.1 The direct-method z ~ 2.2 mass-metallicity relation

With the remaining high-redshift auroral-line sample of 14 galaxies, we construct the direct-method MZR at z > 1 for the first time, shown in the left panel of Figure 4. The median redshift of the z > 1 MZR sample is zmold = 2.17. We calculate mean values in two bins of stellar mass divided at log(M*/M⊙) = 8.75, with 5 galaxies in the low-mass bin and 9 galaxies in the high-mass bin. The binned means are shown in Figure 4 as black open diamonds, with log(M*/M⊙) = [7.90 ± 0.16, 9.42 ± 0.10] and 12+log(O/H) = [7.66 ± 0.14, 7.95 ± 0.08] for the low- and high-mass bins, respectively. We observe a positive correlation between direct-method O/H and M*, for both the means and individual galaxies, though with significant scatter for the latter. Using an orthogonal distance regression, we fit a linear relation to the 14 individual z > 1 galaxies (excluding
Figure 4. **LEFT:** The stellar mass—gas-phase metallicity relation (MZR) at $z \sim 2.2$ using electron temperature measurements. Points with errorbars are as in Figure 3. Hollow triangles denote the two galaxies identified as probable mergers (C12b, C12c), which are excluded when fitting the MZR. The solid blue line shows the best-fit linear relation to individual $z > 1$ galaxies (equation 7), with the 1σ confidence interval given by the blue-shaded region. The dashed and dotted blue lines denote the best-fit relations to the subset of $z > 1$ galaxies with only [O iii]$\lambda 4363$ or [O iii]$\lambda 14663$ detections, respectively, and are not significantly different than the best fit to the total sample. The mean values in two bins of stellar mass divided at log($M_*/M_\odot$) = 8.75 are displayed as black diamonds. The orange line shows the direct-method MZR at $z \sim 0.8$ from Ly et al. (2016). The $z \sim 0$ M$_*$-binned stacked of Andrews & Martini (2013) is presented as gray squares, where the filled filled squares have been corrected for DIG emission and flux-weighting effects using the models of Sanders et al. (2017). The best-fit $z \sim 0$ MZR given by the gray line. **RIGHT:** The MZR based on direct-method metallicities, where O/H has been corrected for biases in SFR of each sample relative to the mean M$_*$-SFR relation at each redshift. Corrected O/H values should represent typical galaxies on the M$_*$-SFR relation at each redshift. The blue, dashed, and dotted lines in the right panel show the best-fit $z \sim 2.2$ MZR to the SFR-corrected total, [O iii]$\lambda 4363$, and O iii]$\lambda 14663$ samples. The uncorrected best-fit MZR from the left panel is displayed as a black dot-dashed line. The solid black line displays the best-fit $z \sim 2.2$ MZR after correcting for the low-redshift bias of the low-mass bin, and is the most robust representation of the MZR at $z \sim 2.2$ presented here.

The open triangles and stacks, obtaining:

$$12 + \log(O/H) = (0.201 \pm 0.058) \times \log(M_*/M_\odot) + (6.07 \pm 0.52)$$

(7)

with a strong covariance between slope and intercept of $\rho = 0.9975$. The best-fit line is shown in the left panel of Figure 4 as a solid blue line. We also fit lines to the subset of galaxies with [O iii]$\lambda 4363$ detections and the subset with O iii]$\lambda 14663$ detections, and maintain consistent results (dashed and dotted blue lines). The $z > 1$ mergers do not appear to be strong outliers compared to the full $z > 1$ sample.

We compare to two $z < 1$ samples with auroral-line detections to investigate metallicity evolution at fixed M$_*$. The M$_*$-binned $z \sim 0$ SDSS stacks of Andrews & Martini (2013) are shown as gray squares in the left panel of Figure 4. The open squares show the values derived by Andrews & Martini (2013), while the filled squares denote the derived metallicities after correcting for contamination from diffuse ionized gas according to Sanders et al. (2017), shifted $\sim 0.05$ dex lower in O/H. The orange line shows the best-fit MZR to a sample of $z \sim 0.8$ galaxies with [O iii]$\lambda 4363$ detections from Ly et al. (2016). The $z \sim 0.8$ MZR lies approximately midway between the $z \sim 0$ and $z \sim 2.2$ MZRs. Compared to $z \sim 0$, we find that O/H at fixed M$_*$ is 0.3 dex lower at $z \sim 0.8$ and 0.65 dex lower at $z \sim 2.2$ at log($M_*/M_\odot$) $\sim 9.5$, where the majority of the $z > 1$ auroral-line sample lies. This amount of metallicity evolution is much larger than is seen in studies using strong-line methods on representative samples of galaxies at the same M$_*$, which find 0.3 – 0.4 dex evolution between $z \sim 0$ and $z \sim 2.3$ (Erb et al. 2006; Steidel et al. 2014; Sanders et al. 2015, 2018).

A major concern when comparing the gas-phase metallicities of different samples is how representative the SFRs are over the range of M$_*$. In the local universe, there is a well-established relationship between M$_*$, SFR, and O/H such that galaxies with higher SFR have lower O/H at fixed M$_*$, known as the “fundamental metallicity relation” (FMR; e.g., Mannucci et al. 2010; Lara-López et al. 2010; Yates et al. 2012; Andrews & Martini 2013; Salim et al. 2014; Cresci et al. 2018). A M$_*$-SFR-O/H relation has also been shown to exist at $z \sim 1 − 2.5$ using strong-line metallicities with a similar strength of SFR dependence as at $z \sim 0$ (Zahid et al. 2014b; Sanders et al. 2018). Thus, if a sample is biased in SFR at fixed M$_*$, then its metallicity will not be representative of typical galaxies at the same M$_*$. Samples selected to have detections of [O iii]$\lambda 4363$ are often biased towards higher SFR at fixed M$_*$ due to the faintness of the feature, and thus the MZR evolution displayed in the left panel of Figure 4 may be artifically large.

In Figure 5, we place the $z > 1$ auroral-line sample on the M$_*$-SFR diagram. We compare to the best-fit M$_*$-SFR relation from Sanders et al. (2018) for a representative sample of $z \sim 2.3$ star-forming galaxies with log($M_*/M_\odot$) = $9.0 - 11.0$ from MOSDEF, with SFRs derived from dust-corrected H$\alpha$ luminosities (dashed blue line):

$$\log(\text{SFR}/M_\odot \text{yr}^{-1}) = 0.67 \times \log(M_*/M_\odot) - 5.33$$

(8)
Given the redshift range of the $z > 1$ auroral-line sample (1.4 < $z$ < 3.7) and the evolution of the $M_\ast$-SFR relation with redshift, we fix the slope of the $M_\ast$-SFR relation and shift the normalization to estimate mean $M_\ast$-SFR relations at $z \sim 1.5$ (green dashed line) and $z \sim 3.4$ (red dashed line) with offsets calculated according to the best-fit $M_\ast$-$z$ relation of Speagle et al. (2014) ($-0.21$ dex for $z \sim 1.5$ and +0.15 dex for $z \sim 3.4$). We find that the $z > 1$ galaxies have significantly higher SFR at fixed $M_\ast$ than is typical for galaxies at the same redshifts, with only one galaxy in the $z > 1$ MOSDEF sample falling below its corresponding $M_\ast$-SFR relation.

We plot the median $M_\ast$ and SFR values of the Andrews & Martini (2013) stacks in Figure 5, and find that their SDSS sample is significantly biased in SFR at log($M_\ast/M_\odot$) < 8.75. We fit a linear relation to the Andrews & Martini (2013) stacks above log($M_\ast/M_\odot$) = 9.0 to determine the mean $z \sim 0$ $M_\ast$-SFR relation (gray dashed line), recovering a slope that is nearly identical to that of the $z \sim 2.3$ relation from Sanders et al. (2018). The $z \sim 0.8$ MOSDEF sample has been shown to be biased in SFR $\sim 0.3$ dex higher than representative samples at the same redshift (Ly et al. 2016).

If the dependence of O/H on SFR at fixed $M_\ast$ is known, then SFR biases can be corrected for to obtain representative MZR$s$ for each sample. The strength of the SFR dependence of the FMR can be quantified by the slope of the anticorrelation between residuals around the $M_\ast$-SFR relation ($\Delta$log(SFR)) and residuals around the MZR ($\Delta$log(O/H))

(e.g., Salim et al. 2015; Davé et al. 2017; Kashino et al. 2017; Sanders et al. 2018). Using the Andrews & Martini (2013) $z \sim 0$ SDSS stacks binned in $M_\ast$+SFR and corrected for diffuse ionized gas (Sanders et al. 2017), we fit $\Delta$log(O/H) vs. $\Delta$log(SFR) using direct-method metallicities (gray points and black dashed line in Figure 6) and find that

$$\Delta$$log(O/H) $\sim$ $-0.29\Delta$log($M_\ast/M_\odot$ yr$^{-1}$) 

This dependence of O/H on SFR is approximately twice as strong as is seen when using strong-line metallicity calibrations (Salim et al. 2014; Sanders et al. 2018), but the FMR is known to have a stronger SFR dependence when direct-method metallicities are employed (Andrews & Martini 2013; Sanders et al. 2017). This best-fit relation can be used to correct the metallicity of samples in the left panel of Figure 4 for SFR biases, assuming the SFR dependence of direct-method O/H at fixed $M_\ast$ does not change with redshift.

We correct the oxygen abundances of the $z \sim 0$ and $z > 1$ MZR samples for SFR biases according to equation 9 and $\Delta$log(SFR) relative to the $M_\ast$-SFR relation at each redshift (Figure 5). For the $z \sim 0.8$ MZR of Ly et al. (2016), we shift their best-fit MZR adopting an average $\Delta$log(SFR)=0.3 dex. The resulting SFR-corrected MZR at $z \sim 0$−2.2 is presented in the right panel of Figure 4. Correcting for the SFR biases present in each sample shifts the relations in the the left panel of Figure 4 such that they are now representative of galaxy populations falling on the mean $M_\ast$-SFR relation at each redshift. We fit the SFR-corrected $z \sim 2.2$ MZR and obtain

$$12 +\text{log}(O/H) = \left(0.206 \pm 0.065\right)\times\text{log}(M_\ast/M_\odot) + \left(6.18 \pm 0.58\right)$$
with a covariance of $\rho = 0.998$ between the slope and intercept. The SFR-corrected $z \sim 2.2$ MZR fit is $-0.15$ dex higher in metallicity than the uncorrected fit, but has a similar slope (equation 7). The mean SFR-corrected metallicity of the low- and high-mass bin is $12 + \log(O/H)_{\text{med}} = 7.81 \pm 0.13$ and $8.11 \pm 0.07$, respectively.

We test whether the SFR dependence described by equation 9 was appropriate to apply to the $z > 1$ sample by plotting the high-redshift galaxies in Figure 6, where $\delta \log(\text{SFR})$ is determined relative to the $M_*$-SFR relation matched in redshift for each galaxy (Fig. 5) and $\delta \log(O/H)$ is determined using the uncorrected metallicities (i.e., those displayed in Table 2) relative to the SFR-corrected MZR of equation 10. We find a weak anticorrelation between $\delta \log(O/H)$ and $\delta \log(\text{SFR})$, with a significance of $\approx 2\sigma$. We fit a linear relation to the $z > 1$ sample, finding a slope of $-0.40 \pm 0.16$, consistent with the $z \sim 0$ slope. This relation is tentative evidence for the existence of a $M_*$-SFR-O/H relation at $z \sim 2.2$ based purely on direct-method metallicities, although the current sample has only one galaxy falling below the mean $M_*$-SFR relation (i.e., with $\delta \log(\text{SFR}) < 0$). We conclude that equation 9 and our method of correcting for SFR biases are applicable at high redshifts.

The scatter of the $z > 1$ sample around the best-fit MZR is smaller after correcting for SFR. Indeed, the most significant outliers in the left panel of Figure 4 were also outliers in SFR compared to the rest of the sample, and are no longer outliers in the right panel. The $z > 1$ mergers (open triangles) are not strong outliers in the left panel, but fall well below the best-fit $z \sim 2.2$ MZR in the right panel, suggesting that they do not have the same relation between $M_*$, SFR, and O/H as the rest of the sample.

The best-fit $z \sim 2.2$ MZR in equation 10 has a shallower slope than is typically seen in strong-line samples. This flattening is due to the redshifts of the galaxies in each mass bin. The median redshifts of the galaxies in the high-mass bin is $2.31$, while the median redshift in the low-mass bin is $1.83$. It is thus expected that the high-mass bin will have evolved farther from the $z \sim 0$ MZR than the low-mass bin, artificially flattening the observed MZR of our $z > 1$ auroral-line sample. Consequently, the true $z \sim 2.2$ MZR is steeper than equation 10.

In the high-mass bin, we find that O/H at $\log(M_*/M_\odot) \approx 9.5$ decreases by 0.2 dex from $z \sim 0$ to $z \sim 0.8$ and by $0.5 \pm 0.1$ dex from $z \sim 0$ to $z \sim 2.3$. At $\log(M_*/M_\odot) \approx 7.9$, O/H is lower by 0.2 dex at $z \sim 0.8$ and $0.35 \pm 0.15$ dex at $z \sim 1.8$ compared to $z \sim 0$. These values suggest a roughly linear evolution in $\log(O/H)$ at fixed $M_*$ with redshift of $\delta \log(O/H)/\delta z \approx 0.25$ that does not display a strong dependence on $M_*$ below $\sim 10^{9.5} M_\odot$ out to $z \sim 2$. Accordingly, we find that the low-mass bin would have 0.12 dex lower O/H when “evolved” from $z \sim 1.83$ to $z \sim 2.31$. We alter the slope and intercept of equation 10 such that the line has the same O/H at $\log(M_*/M_\odot) = 9.42$ but yields a value 0.12 dex lower at $\log(M_*/M_\odot) = 7.90$, obtaining:

$$12 + \log(O/H) = 0.285 \times \log(M_*/M_\odot) + 5.44$$

where the uncertainties in equation 10 can be applied to the slope and intercept here.

After correcting for SFR and sample redshift biases, we find that O/H at $\log(M_*/M_\odot) \approx 9.5$ decreases by 0.2 dex from $z \sim 0$ to $z \sim 0.8$ and by $0.5 \pm 0.1$ dex from $z \sim 0$ to $z \sim 2.2$. This MZR evolution is consistent with what is found using strong-line methods at $z \sim 0.5$–1 (Savaglio et al. 2005; Maiolino et al. 2008; Zahid et al. 2014a; Troncoso et al. 2014) and slightly larger than strong-line studies at $z \sim 2$ that find $\sim 0.25$ – $0.4$ dex evolution in O/H at fixed $M_*$ (e.g., Erb et al. 2006; Maier et al. 2014; Steidel et al. 2014; Sanders et al. 2015, 2018). This tension with strong-line results at $z \sim 2.3$ is not significant given the current uncertainties.

Our final $z \sim 2.2$ slope of 0.285$\pm$0.065 is consistent within the uncertainties with previous results at $z \sim 2$ based on a number of strong-line metallicity indicators (e.g., Erb et al. 2006; Henry et al. 2013; Maier et al. 2014; Sanders et al. 2015, 2018). Sanders et al. (2018) found a $z \sim 2.3$ MZR slope of 0.26, 0.30, 0.34, and 0.50 when using O32, O3N2, N2, and N2O2 strong-line calibrations, respectively. Steidel et al. (2014) find a shallow $z \sim 2.3$ MZR slope of 0.20 based on the N2 and O3N2 indicators, in contrast to other strong-line works. Our results are consistent at the 1.3$\sigma$ level with the Steidel et al. (2014) slope, so we cannot rule out a shallow high-redshift MZR. A larger $z > 1$ auroral-line sample is needed to tightly constrain the high-redshift MZR slope via the direct-method.

### 3.2.2 The direct-method fundamental metallicity relation from $z \sim 0$–$z \sim 2.2$

The $z \sim 0$ FMR has been claimed to be redshift independent, such that galaxies at $z \sim 0$–2.5 all follow the same $M_*$-SFR-O/H relation, with high-redshift galaxies having lower metallicities in proportion to their higher SFRs at fixed $M_*$ (e.g., Mannucci et al. 2010; Cresci et al. 2018). In Sanders et al. (2018), we showed that $z \sim 2.3$ galaxies were offset $\sim 0.1$ dex below the $z \sim 0$ FMR using a direct comparison of strong-line metallicities at fixed $M_*$ and SFR. Whether the FMR is in fact redshift invariant remains an open question given the uncertainties associated with strong-line metallicities at high redshifts. It is thus of interest to test the universality of the FMR using direct-method metallicities. We cannot perform a direct comparison following Sanders et al. (2018) because there are not enough local analogs with [O iii]4363 detections matched in $M_*$ and SFR to our $z > 1$ sample. In lieu of a direct comparison, we rely on parameterized forms of the FMR.

Mannucci et al. (2010) constructed a planar parameterization of the $z \sim 0$ $M_*$-SFR-O/H relation that is described by O/H as a function of $\mu_0 = \log(M_*) - \alpha \log(\text{SFR})$. Using strong-line metallicities, Mannucci et al. found that $\alpha = 0.32$ minimized the scatter in O/H at fixed $\mu_0$. We plot $12 + \log(O/H)$ vs. $\mu_{0.32}$ for the $z > 1$ and low-redshift MZR samples in the top panel of Figure 7. There are offsets between the $z \sim 0$, $z \sim 0.8$, and $z \sim 2.2$ samples, such that galaxies at higher redshift fall lower in O/H at fixed $\mu_{0.32}$. However, the direct-method FMR has been found to have stronger SFR dependence than when using strong-line metallicities. Andrews & Martini (2013) find a value of $\alpha = 0.66$ minimizes the FMR scatter using direct-method metallicity, while Sanders et al. (2017) find a best-fit value of $\alpha = 0.63$ after correcting for diffuse ionized gas contamination. We adopt $\alpha = 0.63$, and present the resulting FMR projection in the bottom panel of Figure 7. The $z \sim 2.2$ sample is consistent with both the $z \sim 0$ and $z \sim 0.8$ samples to within $\sim 0.1$ dex at fixed $\mu_{0.63}$. This comparison is the first
test of the universality of the FMR at high redshifts using temperature-based metallicities. This result suggests that, on average, O/H does not vary by more than ~0.1 dex at fixed $M_*$ and SFR from $z = 0$ to $z \sim 2.2$, in agreement with results using strong-line metallicities (Sanders et al. 2018; Cresci et al. 2018). The mean O/H of the high-mass $z > 1$ bin at $\mu_{0.63} \approx 8.5$ is shifted ~0.1 dex below $z \sim 0$ galaxies at the same $\mu_{0.63}$, hinting at a slight evolution of the $M_*$-SFR-O/H relation, as found by Sanders et al. (2018) using strong-line metallicities. A larger $z > 1$ auroral-line sample is needed to resolve such a small shift with statistical significance.

### 3.2.3 Systematic effects on O/H, SFR, and $M_*$

A concern when comparing metallicity scaling relations across multiple galaxy samples is whether the metallicities, $M_*$, and SFRs of each sample can be fairly compared. The metallicities of all samples in this work have been determined via the direct-method based on either [O III]4363 or O III]14663 and can be directly compared in an unbiased manner, even when the samples span a large range in redshift. The SFRs of all samples have been determined using reddening-corrected Balmer emission lines and calibrations within ±10% of one another in normalization, such that the comparison of SFRs between our low- and high-redshift samples is robust. Stellar mass determinations require a number of assumptions regarding the stellar population templates, attenuation curve, IMF, and star-formation history. We have shifted all stellar masses and SFRs to a Chabrier (2003) IMF.

Even when utilizing the same IMF, differences in SED-fitting assumptions can result in systematic uncertainties of ~0.3 dex in $M_*$ (e.g., Reddy et al. 2012). Details of the SED fitting were not included in all of the references from which we obtained $M_*$ for the $z > 1$ aurora-line sample. However, we have confirmed that over half of the $z > 1$ sample was fit with Bruzual & Charlot (2003) stellar population synthesis models, a constant or declining star-formation history, a Calzetti et al. (2000) attenuation curve, and metallicities ranging from 0.2–1.0 $Z_\odot$. Based on changing the SED-fitting assumptions for the MOSDEF $z > 2$ star-forming galaxy sample, stellar masses under the range of assumptions described above typically vary by ~0.1–0.2 dex compared to $M_*$ derived under our fiducial assumptions (Sec. 2.2). Furthermore, this range of assumptions is similar to those used in $M_*$ determination of the $z \sim 0$ and $z \sim 0.8$ samples (Brinchmann et al. 2004; Andrews & Martini 2013; Ly et al. 2016). Given our best-fit $z \sim 2$ MZR slope of $\alpha = 0.30$, a systematic offset of 0.2 dex in $M_*$ between the samples at each redshift would introduce a systematic bias of ~0.05 dex in the inferred O/H evolution. We conclude that the metallicity evolution results presented herein are not significantly affected by systematic biases associated with the determination of O/H, SFR, or $M_*$ for each sample.

### 4 PHOTOIONIZATION MODELING: STELLAR METALLICITY & IONIZATION PARAMETER

Photoionization models of H II regions utilizing state-of-the-art stellar models are powerful tools for extracting information about the ionization state and chemical abundances of ionized gas from observed nebular emission-line strengths (e.g., Kewley & Dopita 2002; Steidel et al. 2014; Dopita et al. 2016; Strom et al. 2018; Kashino & Inoue 2019). However, degeneracies between the ionization state and nebular metallicity are significant when only the strong optical emission-line ratios are available, especially when just a subset of the strong optical nebular emission lines are detected, as is the case for nearly all high-redshift galaxies.

Earlier observational and theoretical work has suggested that the Fe/H of massive stars may be deficient relative...
to the O/H of the ionized gas on a solar abundance scale in \( z > 1 \) galaxies (i.e., \( \text{O}/\text{Fe}>\text{O}/\text{Fe}^\odot \); Steidel et al. 2016; Matteucci & Scaye 2018; Strom et al. 2018). This O/Fe enhancement is thought to be caused by the rapid formation timescales (\( \lesssim 1 \) Gyr; e.g., Reddy et al. 2012, 2018) of high-redshift galaxies, in which chemical enrichment is dominated by Type II SNe that yield super-solar O/Fe ratios (Nomoto et al. 2006; Kobayashi et al. 2006). This enhancement in O/Fe relative to a solar abundance pattern is brought about because the majority of Fe is produced in Type Ia SNe, most of which occur on timescales \( \sim 1 \) Gyr after a star-formation event, while O is produced promptly in core-collapse Type II SNe on timescales of \( \lesssim 10 \) Myr, the lifetimes of massive stars.

Using the unique direct-method nebular metallicity constraints of our \( z > 1 \) auroral-line sample, we fix the input nebular metallicity (i.e., \( O/\text{H} \)) in grids of photoionization models, allowing us to unambiguously constrain the stellar metallicity (i.e., \( \text{Fe}/\text{H} \)) and the ionization parameter, \( U \), for this set of high-redshift star-forming galaxies.

### 4.1 Description of the photoionization models and fitting procedure

We utilize the code Cloudy (v17.01; Ferland et al. 2017) to produce a grid of photoionization models for use in constraining the ionization parameter (\( U \)) and stellar ionizing spectral shape for the \( z > 1 \) auroral-line sample. We construct simple H II region models assuming a plane-parallel geometry ionized by a single star cluster, where the calculation is stopped at the edge of the fully-ionized H zone. The input parameters of the models are the ionization parameter \( U \), electron density \( n_e \), nebular metallicity \( Z_{\text{neb}}^{-}\), and stellar metallicity \( Z_* \). In practice, we fix \( n_e \) to a constant value and vary \( U \), \( Z_{\text{neb}}^{-} \), and \( Z_* \) for the calculation of the model grids, as described below. Using this set of model grids, we fix \( Z_{\text{neb}}^{-} \) to a value matched to the direct-method metallicity for each object and unambiguously fit for \( U \) and \( Z_* \).

#### 4.1.1 Photoionization model grids

The intensity of the radiation field is set by the dimensionless ionization parameter, \( U = n_e \text{H} / n_H \), where \( n_e \text{H} \) is the volume density of hydrogen-ionizing photons and \( n_H \) is the hydrogen gas volume density, which can be approximated by the electron density \( n_e \). We vary \( \log(U) \) from \(-1.0 \) to \(-4.0 \) in 0.1 dex steps. We assume a constant density of \( n_e = 250 \text{ cm}^{-3} \), a typical value for \( z \sim 2 \) star-forming galaxies (e.g., Sanders et al. 2016a; Steidel et al. 2016; Strom et al. 2017).

The gas-phase elemental abundances are assumed to follow a solar abundance pattern, such that the abundances of each element is set by the oxygen abundance, \( O/\text{H} \). To avoid confusion with the nebular metallicity measured via the direct-method from observations (\( Z_{\text{neb}} \)), we refer to the model input nebular metallicity as \( Z_{\text{neb}}^{-} \). While the abundances of elements with secondary production channels (C, N) depend on \( Z_{\text{neb}} \), we do not utilize any lines of these elements in the fitting procedure described below, and thus simply leave them fixed at a solar scale for convenience. This assumption has a negligible effect on the line ratios of O, Ne, and H across the range of parameters explored here. We adopt a range of values for the nebular metallicity: \( Z_{\text{neb}}^{-} = [0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.25, 1.5] \times Z_\odot \) (7.4\( \leq 12+\log(O/\text{H}) \leq 8.9 \)).

Following Steidel et al. (2016) and Strom et al. (2018), the stellar metallicity, \( Z_* \), is allowed to vary freely from \( Z_{\text{neb}}^{-} \). Only the hydrogen-ionizing photons (\( h > 13.6 \text{ eV} \); \( \lambda < 912 \AA \)) produced by the stellar ionizing source affect the emission-line production of the ionized nebula, as these photons are responsible for liberating and heating of the free electron population. The spectrum of massive stars at \( \lambda < 912 \AA \) is primarily governed by the opacity of the photosphere from line-blanketing of iron-peak elements. This opacity, driven by Fe/H, strongly affects the evolution of massive stars. Larger opacity due to higher Fe/H results in stronger stellar winds, driving larger mass loss rates and decreasing rotation. Accordingly, stars with lower Fe/H produce a larger number of ionizing photons, have a harder spectral shape due to a hotter effective temperature, and can potentially continue producing copious amounts of ionizing photons for a longer time period due to effects such as quasi-homogeneous evolution (Eldridge et al. 2011; Eldridge & Stanway 2012). We assume that \( Z_* \), which sets the shape of the ionizing stellar spectrum, effectively traces Fe/H. Thus, decoupling \( Z_* \) and \( Z_{\text{neb}}^{-} \) allows us to investigate non-solar O/Fe ratios and vary the hardness of the ionizing spectrum at fixed O/H.

We adopt the “Binary Population and Spectral Synthesis” v2.2.1 models (BPASS; Stanway & Eldridge 2018), which include the effects of binarity on the evolution of massive stars. We use the binary BPASS models with a high-mass IMF slope of \(-2.35 \) and a high-mass cutoff of 100 \( \text{M}_\odot \) as our fiducial set of ionizing spectra. We vary the BPASS stellar metallicities over the range \( Z_* = 10^{-5} - 0.020 \) (0.0007-1.4 \( Z_\odot \) based on \( Z_\odot = 0.014 \) from Asplund et al. 2009). For comparison, we also utilize single-star models from Starburst99 (SB99; Leitherer et al. 2014) with the same IMF slope and cutoff, with a metallicity range of 0.01 – 0.040 (0.0007-2.8 \( Z_\odot \)). The input stellar populations are assumed to have continuous star formation, with an age of 100 Myr. The ionizing spectra of both BPASS and SB99 models reach equilibrium after 10 Myr of continuous star formation, so the nebular emission-line predictions are appropriate for systems with stellar populations older than 10 Myr.

#### 4.1.2 \( U \) and \( Z_* \) fitting procedure

Utilizing the model grids described above, we produce constraints on the ionization state of the \( z > 1 \) auroral-line sample by fixing \( Z_{\text{neb}} \) to a value appropriate for each galaxy. Thus, a critical step to utilizing these photoionization models to constrain \( U \) and \( Z_* \) is matching the measured direct-method \( Z_{\text{neb}} \) to \( Z_{\text{neb}}^{-} \) of a particular set of models. The simplest assumption is \( Z_{\text{neb}}^{-} = Z_{\text{neb}} \), that is, the direct-method O/H represents the “true” gas-phase metallicity in an unbiased manner. In the left column of Figure 8, we compare the measured O3 and O2 line ratios of \( z > 1 \) galaxies to model grids directly matched in metallicity. We find that both BPASS and SB99 grids significantly underpredict O3 and O2 of the majority of \( z > 1 \) galaxies when \( Z_{\text{neb}}^{-} = Z_{\text{neb}} \), even at the lowest \( Z_* \) (hardest ionizing spectrum).
Figure 8. Comparison of photoionization model grids to observed line-ratios (O3 vs. O2) at z > 1, in three bins of metallicity. Grids are shown from the BPASS (blue) and SB99 (red) models at fixed nebular metallicity, with varying stellar metallicity and ionization parameter. Lines of constant stellar metallicity are solid, while lines of constant ionization parameter are dotted. In the left column, the input nebular metallicity to the models is assumed to be the same as the direct-method metallicity of the observed galaxies ($\Delta \log(O/H)|_{T_e} = 0.0$), such that the input nebular metallicity for the top-, middle-, and bottom-left panels is $12+\log(O/H) = 7.7$, 7.9, and 8.1, respectively. In the right column, the input nebular metallicity to the models is instead assumed to be larger than the observed direct-method metallicity by 0.24 dex ($\Delta \log(O/H)|_{T_e} = 0.24$). The mean difference between electron-temperature and recombination line metallicities of $z \sim 0$ H II regions. The input nebular metallicity of the models in the top-, middle-, and bottom-right panels is thus $12+\log(O/H) = 7.94$, 8.14, and 8.34, respectively. The SB99 models fail to produce large enough O3 and O2 values to match observations in all panels. The BPASS models fail to match observations when $\Delta \log(O/H)|_{T_e} = 0.0$. The ability of the BPASS models to reproduce the range of O3 and O2 values observed at fixed O/H significantly improved when assuming $\Delta \log(O/H)|_{T_e} = 0.24$. We adopt this as a fiducial assumption when matching models to observations.

There is evidence that the temperature-based metallicities are systematically biased and thus do not represent the “true” gas-phase metallicity. There is a long-standing disagreement between temperature-based metallicities from collisionally-excited auroral lines (direct method) and metallicities determined using O recombination lines (RL method), known as the abundance discrepancy factor (ADF) problem (e.g., Peimbert 1967; Esteban et al. 2014). Direct-method metallicities are systematically lower than RL-method metallicities measured for the same targets. The value of the ADF appears to be roughly constant as a function of metallicity, with the direct method yielding $\pm 0.24$ dex lower metallicities than the RL method on average (e.g., Esteban et al. 2014; Blanc et al. 2015). The ADF has been proposed to be caused by fluctuations in the temperature field of the ionized nebula, such that hotter regions dominate the auroral-line emission and bias temperature measurements high and, consequently, direct-method O/H measurements low. There is an ongoing debate whether the direct-method or RL metallicities represent the true gas-phase abundance scale. Attempts have been made to solve this problem by comparing the stellar oxygen abundance of young A and B supergiants to the RL and direct method metallicities of the H II regions they occupy, with conflicting results. The stellar O/H agrees with the RL method in some H II regions and the direct method in others (Bresolin et al. 2016; Torbjo San Cipriano et al. 2016, 2017). It is not clear which method yields the true $Z_{\text{neb}}$.

Regardless of the remaining uncertainties regarding direct vs. RL methods, we have found that the models cannot reproduce the observations when $Z_{\text{neb}}^{\text{ini}} = Z_{\text{neb}}$, where the latter is determined using the direct method (Fig. 8, left column). In the right column of Figure 8, we show the resulting model grids when we apply the typical ADF of 0.24 dex between the RL and direct methods to our measured $z > 1$ direct-method O/H values ($\log(Z_{\text{ini}}^{\text{neb}}) = \log(Z_{\text{neb}}) + 0.24$). The BPASS grids (blue) now span the range of measured O3 and O2 of the majority of the $z > 1$ sample. Steidel et al. (2016) similarly found that applying a 0.24 dex ADF to the direct-method O/H of their stack of $z \sim 2.3$ star-forming galaxies was necessary for model grids to match measured line ratios. We proceed under the assumption that $\log(Z_{\text{ini}}^{\text{neb}}) = \log(Z_{\text{neb}}) + 0.24$, and consider the comparison in Figure 8 circumstantial evidence for the existence of temperature fluctuations in H II regions at $z > 1$. We note that the SB99 grids fail to overlap the $z > 1$ galaxies even under this assumption because they do not produce hard enough ionizing spectra even at the lowest available metallicity, $Z_e \sim 0.07 Z_{\odot}$. Therefore, we conclude that SB99 ionizing spectra cannot reproduce the emission-line properties of our $z > 1$ sample and only show results using the BPASS binary models for the remainder of the analysis.

We fit for the ionization parameter and stellar metallicity of each galaxy using the following procedure. We first interpolate the photoionization model grids in nebular metallicity to construct a grid of strong-line ratios as a function of $U$ and $Z_e$ at fixed $Z_{\text{neb}}^{\text{ini}}$ according to $\log(Z_{\text{ini}}^{\text{neb}}) = \log(Z_{\text{neb}}) + 0.24$, where $Z_{\text{neb}}$ is the measured direct-method O/H given in Table 2. We fit for $U$ and $Z_e$ by performing a $\chi^2$ minimization using the measured reddening-corrected strong-line ratios simultaneously. For the fitting procedure, we do not.
include ratios involving lines of N because of uncertainty in the N/O vs. O/H relation at high redshifts (Masters et al. 2014; Steidel et al. 2014; Sanders et al. 2015, 2016a; Shapley et al. 2015; Strom et al. 2018), and also exclude line ratios involving S because it has an ionization potential lower than that of H and may be emitted in partially-ionized regions that our photoionization models do not include. We use line ratios of collisionally-excited metal lines relative to Hβ only. Thus, the line ratios used in the fitting process include [O iii]λ45007/Hβ (O3), [O ii]λ3726/Hβ (O2), and [Ne iii]λ3869/Hβ (Ne3). Given the tight relation between O3 and Ne3 at sub-solar metallicities (Levesque & Richard-son 2014), a galaxy must have at minimum detections of O3 and O2 to fit for U and Z, though Ne3 is also used when available. Examples of this fitting process are shown in the left panels of Figures 9 to 12 for the four MOSDEF [O iii]λλ4363 emitters.

We estimate confidence intervals on the best-fit U and Z, by perturbing the line ratios and direct-method O/H according to the uncertainties, and repeating the fitting procedure above on each realization, 5000 times in total. The uncertainties on U and Z, thus include uncertainties in the direct-method O/H that sets Z_{n eb}. The resulting distributions of U, Z, and Z_{n eb}/Z_{n eb} are displayed in the right panels of Figures 9 to 12. The red ‘X’ and lines denote the best fit in each panel.

Of the 18 z > 1 auroral-line emitters in our sample, 4 have limits on the [O ii] flux and one lacks [O ii] coverage and thus cannot be fit for U and Z. For the other 13 objects and two composites, the ionization parameter is always constrained, typically within ±0.2 dex, and has a weak co-variance with Z. The stellar metallicity is well-constrained (i.e., 1σ confidence intervals within the explored parameter space) for 6 z > 1 galaxies (see, e.g., Figs. 9 and 11), as well as the Steidel et al. (2016) composite. Six of the z > 1 galaxies and the MOSDEF stack have one-sided constraints on Z, (e.g., Figs. 10 and 12), which we display as 3σ upper limits. No targets produced lower limits on Z, although U could still be constrained with larger uncertainties. The best-fit U and Z, values assuming the BPASS grids are given in Table 3.

Our methodology of fixing the nebular metallicity of the models to fit for Z, is complementary to the method of Steidel et al. (2016), who fixed Z, according to the stacked FUV continuum spectrum and fit for nebular metallicity. For the Steidel et al. (2016) composite, we obtain a best-fit log(U) = −2.85, consistent with their best-fit value of log(U) = −2.8 when employing BPASS binary models with a 100 M⊙ IMF cutoff. We find a best-fit stellar metallicity of Z/Z⊙ = 0.01, somewhat lower than the best-fit value of Z/Z⊙ = 0.07 from Steidel et al. (2016). We note that Steidel et al. only considered BPASS models with Z/Z⊙ ≥ 0.07 (Z ≥ 0.001), such that the best-fit stellar metallicity was at the lower bound of the parameter space. In this work, we allow the stellar metallicity of the BPASS models to be as low as Z/Z⊙ = 0.0007 (Z = 10^{-5}). It is possible that the FUV
Continuum of the KBSS-LM1 stack could favor an even lower $Z_*$ if the range of permitted metallicities was increased. We conclude that both our method and that presented in Steidel et al. (2016) reliably constrain $U$, but further investigation is required to understand whether both methods infer similar $Z_*$. In the future, we will analyze individual $z > 1$ galaxies with both auroral-line and FUV spectral continuum measurements to provide a more rigorous validation of these techniques and test of the stellar models.

### 4.2 Stellar metallicity and O/Fe at $z > 1$

We now utilize these best-fit values of the ionization parameter and stellar metallicity to explore the ionization state
of \( z > 1 \) star-forming regions and the redshift evolution of these properties. In Figure 13, we compare the best-fit \( Z_\star \) (i.e., \( \text{Fe}/\text{H} \)) with the nebular metallicity \( Z_{\text{n neb}}^{\text{in}} \) (i.e., \( \text{O}/\text{H} \)) of each \( z > 1 \) galaxy. The solid black line shows a one-to-one relation where \( Z_\star = Z_{\text{n neb}}^{\text{in}} \) on a solar scale or, equivalently, \( \text{O}/\text{Fe} = \text{O}/\text{Fe}_\odot \). Overall, we find a strong preference for super-solar \( \text{O}/\text{Fe} \) (\( Z_\star < Z_{\text{n neb}}^{\text{in}} \)). Only two galaxies have constraints or limits that are consistent with \( Z_\star = Z_{\text{n neb}}^{\text{in}} \), and the one that is not an upper limit has a large uncertainty towards lower \( Z_\star \). The remaining 11 individual galaxies and two composites favor \( Z_\star < Z_{\text{n neb}}^{\text{in}} \).

Six galaxies and the Steidel et al. (2016) stack have inferred \( Z_\star \) less than 1/5 \( Z_{\odot} \), where \( 5\times\text{O}/\text{Fe}_\odot \) is the theoretical limit for \( \alpha \)-enhancement from pure Type II supernova enrichment (dotted line in Fig. 13; Nomoto et al. 2006). These galaxies are thus in tension with the standard picture of chemical enrichment through star formation. It is possible that uncertainties on elemental yields in very metal-poor stars (\( Z_* < 0.03 \; Z_{\odot} \)) could relieve this tension if metal-poor supernovae have lower Fe yields relative to O. However, some objects fall > 1 dex below the Type II SNe limit. The ionizing spectra of metal-poor massive stars is highly uncertain, and few observational constraints currently exist (Schenzlea et al. 2017, 2019; Stanway & Eldridge 2019). If metal-poor massive stars produce harder ionizing spectra than the current stellar models predict, then the true stellar metallicities would in fact be higher.

Alternatively, a top-heavy IMF at high redshifts may relieve tension with the theoretical Type II SNe limit. The limit of \( 5\times\text{O}/\text{Fe}_\odot \) is an IMF-integrated value assuming a

\[ \text{Figure 12.} \] Same as Figure 9 for GOODS-S-41547. The best-fit solution is at the lower bound of the explored parameter space for \( Z_\star \), thus only a 3σ lower limit on \( Z_\star \) is obtained.

| Table 3. | Constraints on the ionization parameter and stellar metallicity for the \( z > 1 \) auroral-line sample. |
|-----------|-----------------------------------------------|
| Object    | \log(U) | \log \left( \frac{Z_\star}{Z_{\odot}} \right) | \log \left( \frac{Z_{\text{n neb}}^{\text{in}}}{Z_{\odot}} \right) |
| AEGIS 11452 | -2.34\text{^+0.20}_{-0.14} | -0.73\text{^+0.51}_{-2.41} | -0.73\text{^+0.21}_{-0.15} |
| GOODS-S 41547 | -2.76\text{^+0.16}_{-0.06} | <0.14^a | <0.09\text{^+0.05}_{-0.02} |
| COSMOS 1908 | -1.84\text{^+0.04}_{-0.18} | <1.55^a | <0.57\text{^+0.25}_{-0.17} |
| COSMOS 23895 | -2.67\text{^+0.10}_{-0.06} | <1.55^a | <0.43\text{^+0.25}_{-0.17} |
| S13 | -2.30\text{^+0.05}_{-0.04} | <1.95^a | <0.49\text{^+0.07}_{-0.20} |
| J14 | -2.28\text{^+0.08}_{-0.14} | -0.85\text{^+0.12}_{-1.29} | -0.32\text{^+0.05}_{-0.21} |
| C12a | -2.35\text{^+0.22}_{-0.05} | <1.39^a | <0.54\text{^+0.05}_{-0.10} |
| C12c | -2.30\text{^+0.05}_{-0.05} | <2.02\text{^+0.01}_{-0.06} | <0.30\text{^+0.05}_{-0.21} |
| E16c | -2.05\text{^+0.05}_{-0.05} | <0.85\text{^+0.17}_{-0.14} | <0.46\text{^+0.05}_{-0.21} |
| E16d | -2.61\text{^+0.11}_{-0.05} | <1.39^a | <0.54\text{^+0.05}_{-0.10} |
| E16b | -2.29\text{^+0.04}_{-0.04} | <2.09\text{^+0.14}_{-1.04} | <0.60\text{^+0.05}_{-0.21} |
| C12e | -2.27\text{^+0.01}_{-0.01} | <2.42^a | <0.54\text{^+0.05}_{-0.21} |
| B14 | -2.23\text{^+0.01}_{-0.01} | <2.42^a | <0.54\text{^+0.05}_{-0.21} |

\[ ^a \text{3σ upper limit.} \text{^b} \text{The 1σ limits on Z}_\star \text{spanned the entire parameter space, thus no constraint or limit on Z}_\star \text{ was obtained.} \]
The MOSDEF Survey: Direct-Method Metallicities at z $\sim$ 1.5 – 3.5

...canonical high-mass IMF slope of $\approx -2.3$. SNe of very massive ($\gtrsim 25 M_\odot$) stars yield $> 10\times O/Fe$ (Nomoto et al. 2006; Kobayashi et al. 2006), such that the IMF-integrated O/Fe can be larger than $5\times O/Fe$ for top-heavy IMFs (i.e., with high-mass slope shallower than $-2.3$). A top-heavy IMF would also imply a harder ionizing spectrum at fixed metallicity due to the increased relative abundance of hotter, more massive stars. Thus, for the case of a top-heavy IMF, $> 5\times O/Fe$ is allowed and a higher $Z_*$ (i.e., $Fe/H$) would best fit the observed line ratios compared to the values of $Z_*$ inferred under our fiducial model (assuming a Chabrier 2003 IMF), lowering the inferred O/Fe. Recent observational and theoretical work has suggested that the IMF grows more top-heavy in low-metallicity and high-SFR environments common among high-redshift galaxies (Jerabkova et al. 2018; Schneider et al. 2018). The potential impact of IMF variations on the ionizing spectrum in high-redshift star-forming regions should be further investigated in future work.

Despite the large uncertainty on the ionizing spectra of metal-poor stellar populations at high redshift, our results still strongly disfavor a scenario in which $Z_*=Z_{\text{neb}}^\text{in}$, which would occur at stellar metallicities where better observational constraints are available ($\sim 1/3 - 1/2 Z_\odot$). A third possibility is that the sources below the Type II SNe limit have additional ionization mechanisms beyond star formation that harden the galaxy-averaged ionizing spectrum, such as shocks or weak AGN. The galaxies in the $z > 1$ sample do not exhibit any obvious signatures of AGN, but a low-luminosity AGN in a highly star-forming galaxy may only slightly shift the observed line ratios. However, it is unlikely that the presence of shocks or AGN are responsible for the position of the Steidel et al. (2016) stack of 30 galaxies unless such a phenomenon is ubiquitous among typical star-forming galaxies at $z \sim 2$. Ultimately, we require more confidence in stellar models of metal-poor ($\lesssim 0.2 Z_\odot$) massive stars to understand the behavior of galaxies falling below the Type II SNe limit.

We show the O/Fe ratio as a function of stellar population age for the MOSDEF [O III]$\lambda 4363$ emitters and stack in Figure 14. The stellar population age is derived from the best-fit SED models, and O/Fe is on a solar scale such that [O/Fe]$=\log\left(\frac{O/Fe}{O/Fe}_\odot\right)$. We find that all four galaxies have young stellar populations ($< 300$ Myr), with three younger than $100$ Myr. Two of the MOSDEF galaxies are constrained to have super-solar O/Fe, while the other two do not have strong constraints. The composite spectrum yields a $3\sigma$ lower-limit of $O/Fe=3\times O/Fe_\odot$. The ages of the MOSDEF [O III]$\lambda 4363$ emitters are consistent with the presence of super-solar O/Fe from an enrichment history dominated by Type II SNe, with solar O/Fe expected to be reached only at ages $> 1$ Gyr. The [O III]$\lambda 4363$ emitters have ages younger than $90\%$ of MOSDEF star-forming galaxies matched in $M_*$ and redshift. While the precise stellar population ages are significantly uncertain due to systematics associated with the assumed star-formation history, attenuation law, and metallicity for SED fitting, relative ages should be robust under our assumption of constant star formation. Furthermore, the large sSFRs and emission-line equivalent widths of the $z > 1$ sample (Table 2) suggest very young ages, in agreement with results from the SED fitting. The four MOSDEF [O III]$\lambda 4363$ emitters appear to be very young star-forming...
galaxies in which the delayed Fe enrichment from Type Ia SNe has not yet occurred for the bulk of past star formation.

4.3 Ionization parameter at $z > 1$

The $z > 1$ star-forming galaxy population has considerably higher ionization parameters than are typical of $z \sim 0$ samples (e.g., SDSS), as evidenced by significantly higher O32 and O3 at fixed M$_*$ (e.g., Nakajima & Ouchi 2014; Sanders et al. 2016a, 2018; Dickey et al. 2016; Holden et al. 2016; Kashino et al. 2017; Strom et al. 2017; Kashino et al. 2019). sSFR has been shown to strongly correlate with O32 and O3 (e.g., Nakajima & Ouchi 2014; Sanders et al. 2016a; Kewley et al. 2015; Dickey et al. 2016; Kashino & Inoue 2019), and is consequently closely tied to $U$. However, $U$ is anticorrelated with O/H according to the $M_*$-SFR-O/H relation present at low and high redshifts: galaxies with higher SFR at fixed $M_*$ (i.e., higher sSFR) have lower O/H (e.g., Manucci et al. 2010; Lara-López et al. 2010; Andrews & Martini 2013; Sanders et al. 2018). Furthermore, $U$ is tightly anticorrelated with O/H in the local universe (Dopita et al. 2006b,a; Pérez-Montero 2014; Sánchez et al. 2015; Kashino & Inoue 2019). It is therefore possible that the high values of $U$ inferred at high redshifts are simply a byproduct of the lower metallicities and higher sSFRs of high-redshift galaxies compared to the local universe. It has been proposed that the redshift evolution of strong-line ratio sequences is primarily driven by enhanced $U$ at high redshifts compared to typical $z \sim 0$ values at fixed O/H (Kewley et al. 2015, 2016; Kashino et al. 2017, 2019; Kaasinen et al. 2018). In other words, the $U$-O/H relation evolves with redshift such that $U$ increases at fixed O/H with increasing redshift. We test this scenario by comparing $U$, sSFR, and direct-method O/H of our $z > 1$ sample to the corresponding measurements of local star-forming galaxies.

Figure 15 compares $U$, sSFR, and temperature-based O/H for the $z > 1$ auroral-line sample and $z \sim 0$ M$_*$-binned composites of Andrews & Martini (2013). We determine $U$ for the Andrews & Martini (2013) stacks using the same models and fitting procedure described in Section 4.1, except that we fix Z$_{aur}$ = Z$_{med}^{21}$ as expected for $z \sim 0$ galaxies with more extended star-formation histories than high-redshift galaxies. We note that allowing Z$_{aur}$ to vary freely does not significantly change the results. Median values of the $z > 1$ sample of individual galaxies are also displayed in each panel as a open black diamond.

As expected, we find that $U$ is anticorrelated with O/H for $z \sim 0$ galaxies, as shown in the top panel of Figure 15. We find our mean $U$-O/H relation for $z \sim 0$ SDSS stacks closely matches that found by Pérez-Montero (2014) for $z \sim 0$ H ii regions (shaded light gray region). The $z > 1$ auroral-line sample scatters above and below the local relation defined by the $z \sim 0$ stacks. We do not have a suitable sample size or dynamic range in O/H to look for an anticorrelation among the high-redshift galaxies alone. The $z > 1$ galaxies display a significant scatter in $U$ at fixed O/H, but the intrinsic scatter is similar to the range spanned by local H ii regions after accounting for measurement uncertainties, ~ 0.25 dex in $U$ at fixed O/H. The median values of the $z > 1$ sample are $12+\log(O/H)^{med} = 7.96^{+0.04}_{-0.12}$ and $(\log(U)^{med} = -2.33^{+0.06}_{-0.19}$. The median of the $z > 1$ sample is in remarkable agreement with the median of the $z \sim 0$ sample.
with the \( z \sim 0 \) \( U - O/H \) relation of both local galaxies and individual H II regions.

We show \( U \) vs. sSFR in the middle panel of Figure 15. We find a correlation between \( U \) and sSFR among the \( z \sim 0 \) stacks. Once again, the \( z > 1 \) auroral-line sample scatters both above and below the \( z \sim 0 \) relation. Interestingly, the lowest-sSFR galaxy in the \( z > 1 \) sample (C12b; Christensen et al. 2012a,b), with \( sSFR \sim 1.5 \) dex below the sample average, falls directly on the \( z \sim 0 \) relation and has a particularly tightly-constrained \( U \) and \( O/H \). The lowest-mass Andrews & Martini (2013) bins have similar sSFR to the range spanned by the \( z > 1 \) sample (Fig. 5). The median \( z > 1 \) values fall on the \( U \)-sSFR relation defined by the \( z \sim 0 \) composites.

In the bottom panel of Figure 15, we present \( O/H \) vs. sSFR, finding the anticorrelation between \( O/H \) and sSFR at \( z \sim 0 \) as expected from the FMR. The \( z > 1 \) galaxies follow the same relation between \( O/H \) and sSFR displayed by the \( z \sim 0 \) galaxies. This result is expected based on the agreement between \( z \sim 0, z \sim 0.8, \) and \( z \sim 2.2 \) samples in the FMR projection shown in the bottom panel of Figure 7. The \( z > 1 \) median is very closely-matched to \( z \sim 0 \) stacks with similar sSFR.

In summary, Figure 15 shows that, on average, the \( z > 1 \) galaxies fall on the local relationships between \( U, O/H, \) and sSFR. The high ionization parameters of the \( z > 1 \) sample are therefore as expected in accordance with their low metallicities and high sSFRs, and are not in excess of what is expected at fixed \( O/H \) compared to \( z \sim 0 \) galaxies. This result suggests that variations in \( U \) beyond local relations is not a significant driver of the evolving strong-line ratios of galaxies. Combining these results with those presented in Section 4.2, we conclude that the primary cause of strong-line ratio evolution at fixed \( O/H \) (Fig. 3) for our \( z > 1 \) auroral-line sample is a harder ionizing spectrum at fixed nebular metallicity compared to what is typical in \( z \sim 0 \) galaxy populations. The harder spectrum at fixed \( O/H \) in the \( z > 1 \) auroral-line sample is naturally explained by the youth of their stellar populations, leading to chemically-immature galaxies that have super-solar \( O/Fe \) due to the dominance Type II SNe enrichment that promptly produce O and a lack of time-delayed Type Ia SNe that are the main source of Fe. The Fe-poor massive stars produce harder ionizing spectra than their counterparts at solar \( O/Fe \).

5 DISCUSSION

In this work, we have used a sample of 18 individual \( z > 1 \) galaxies with direct-method metallicities to test strong-line metallicity indicators, construct the mass-metallicity relationship, investigate the redshift invariance of the \( M_* \)-SFR-\( O/H \) relation, and demonstrate that the high ionization state of this sample is driven by an ionizing spectrum that is harder than that of \( z \sim 0 \) galaxies at fixed \( O/H \). In this section, we discuss the implications of these results for typical galaxy populations at \( z \sim 1 - 3 \), extreme emission-line galaxy samples at \( z \sim 1 - 3 \), and galaxies in the epoch of reionization at \( z > 6 \).

5.1 Implications for typical \( z \sim 1 - 3 \) galaxies

Obtaining accurate metallicity estimates from strong-line ratios for large existing spectroscopic survey data sets at \( z \sim 1 - 3 \) (e.g., MOSDEF, Kriek et al. 2015; KBSS, Steidel et al. 2014; 3D-HST, Brammer et al. 2012a; FMOS-COSMOS, Kashino et al. 2019) requires knowledge of the ionization state of the ISM in typical high-redshift galaxies. We have shown that our \( z > 1 \) auroral-line sample has significantly higher SFR at fixed \( M_* \) than galaxies falling on the mean \( M_* \)-SFR relation at the same redshifts (Fig. 5). We now put the \( z > 1 \) auroral-line sample in the context of a representative sample of star-forming galaxies at \( z \sim 2.3 \), and discuss which conclusions, if any, can be extended to the typical galaxy population at \( z > 1 \).

In Figure 16, we present SFR, \( O_32 \), and \( EW_0([O III]_{5007}) \) as a function of stellar mass. We compare the \( z > 1 \) auroral-line sample (red points) to a sample of \( z \sim 2.3 \) star-forming galaxies from MOSDEF representative of the typical population at \( \log(M_*/M_\odot) = 9 - 11 \) from Sanders et al. (2018) (open blue circles). Black circles show the mean properties of this sample obtained via spectral stacking in four \( M_* \) bins, where we have derived the mean \( EW_0([O III]_{5007}) \) in each bin according to the method described in Section 2.3.1. We find that all but two galaxies in the auroral-line sample (median redshift of 2.2) fall above the mean \( z \sim 2.3 \) \( M_* \)-SFR relation, with a mean offset of \( \Delta \log(SFR) \) \sim 0.6 dex. The \( O_32 \) values of the typical \( z \sim 2.3 \) sample are much lower than those of the auroral-line sample, but the typical sample has a significantly higher average stellar mass. When comparing at \( \log(M_*/M_\odot) = 9.0 - 9.5 \) where the two samples overlap, the auroral-line sample has \( O_32 \) values \sim 0.5 dex higher on average. The equivalent widths of the auroral-line sample span \( EW_0([O III]_{5007}) = 300 - 2000 \) \AA, with a median value of \sim 500 \AA. This value is much larger than the mean value of the \( z \sim 2.3 \) comparison sample \sim 100 \AA. It is important to take into account the fact that \( EW_0([O III]_{5007}) \) has a strong dependence on stellar mass (Reddy et al. 2018). However, even when comparing at fixed \( M_* \), the auroral-line sample displays \( EW_0([O III]_{5007}) \) that is larger by a factor of 3.

We show \( O_32 \) and \( EW_0(H\beta) \) vs. \( EW_0([O III]_{5007}) \) in Figure 17, and once again find that the auroral-line sample is nearly disjoint from the typical \( z \sim 2.3 \) sample, having higher emission-line equivalent widths suggestive of younger ages and higher \( O_32 \) values implying lower \( O/H \) and higher \( U \). We select the subset of the \( z \sim 2.3 \) comparison sample with \( EW_0([O III]_{5007}) > 300 \) \AA (matched to the auroral-line sample), displayed as filled blue circles in Figures 16 and 17. The EW-matched \( z \sim 2.3 \) sample reproduces the properties of the auroral-line sample in the range of stellar mass overlap, selecting a subset lying above the \( M_* \)-SFR relation at \( \log(M_*/M_\odot) = 9.0 - 9.5 \) with similarly high \( O_32 \) values. However, this EW-matched subset makes up only 11\% (22/203) of the typical \( \log(M_*/M_\odot) \geq 9.0 \) star-forming population at \( z \sim 2.3 \) from MOSDEF.

Based on the high sSFRs, \( O_32 \) values, and emission-line equivalent widths, the \( z > 1 \) auroral-line sample is made up of galaxies that are younger and more metal-poor than the bulk of the typical \( z \sim 1 - 3 \) galaxy population at \( \log(M_*/M_\odot) > 9.0 \), the mass range that most high-redshift
log(O$_{32}$)

Figure 16. SFR (top), O$_{32}$ (middle), and EW$_0$([O III]λ5007) (bottom) vs. M$_*$ for high-redshift galaxy samples. The z > 1 auroral-line sample is shown as red points, with shapes as in Fig. 3. Open blue circles display a representative sample of z ~ 2.3 star-forming galaxies from MOSDEF (Sanders et al. 2018), while filled blue circles denote the subset of that sample with EW$_0$([O III]λ5007) > 300 Å. The z ~ 2 extreme emission-line galaxy sample of Tang et al. (2018) is presented as cyan circles. In each panel, the cyan and blue error bar denotes the mean uncertainty of the Tang et al. (2018) and MOSDEF samples, respectively. Composite spectra are shown in black for the stack of MOSDEF [O III]λ4363 emitters (star), Steidel et al. (2016, ; triangle), and typical z ~ 2.3 MOSDEF galaxies (circles; Sanders et al. 2018). The blue dashed line in the top panel shows the best-fit M$_*$-SFR relation of equation 8. Gray points and dashed line show mean z ~ 0 relations (Andrews & Martini 2013). The mean z ~ 2.3 EW$_0$([O III]λ5007) vs. M$_*$ relation from Reddy et al. (2018) is shown as the solid black line in the bottom panel.

Figure 17. O$_{32}$ (top) and EW$_0$([Hβ]) (bottom) as a function of EW$_0$([O III]λ5007). Points are as in Figure 16. Both panels display remarkably tight correlations over 2.5 orders of magnitude in EW$_0$([O III]λ5007).

spectroscopic surveys probe. Accordingly, it is not clear that our conclusions regarding metallicity calibration evolution and O/Fe enhancement apply to the bulk of the z ~ 2 star-forming population. Based on the high-redshift analog O$_{32}$ strong-line calibration of Bian et al. (2018), ~90% of the z ~ 2.3 comparison sample has oxygen abundances higher than 12 + log(O/H) = 8.14, the highest direct-method O/H of the auroral-line sample. Therefore, the strong-line indicators considered in Figure 3 have not been tested over the metallicity range of the bulk of the log(M$_*/$M$_\odot$) > 9.0 galaxy population at z ~ 2. However, given that the auroral-line sample represents the most extreme subset of the MOSDEF z ~ 2.3 sample, it is expected that the strong-line ratio vs. O/H relations of typical z ~ 2 galaxies will show at most the same level of evolution as the auroral-line sample (~0.1 dex in O/H at fixed strong-line ratio). The Bian et al. (2018) high-redshift analog calibrations are likely good to within the same amount for the typical z ~ 2 population when calculating sample average abundances.

The very young ages of the auroral-line sample implied by large emission-line equivalent widths and sSFRs are consistent with the super-solar O/Fe values found from the photoionization models. However, typical z ~ 2 galaxies at log(M$_*/$M$_\odot$) > 9.0 appear to have older stellar populations.
Star-forming galaxies with stellar populations 300 Myr to ~1 Gyr old may still display some level of O/Fe enhancement relative to solar values, while the abundance ratios of galaxies older than ~1 Gyr will lie close to O/Fe⊙. The age of the universe at z = 2.3 is 2.8 Gyr, and can thus accommodate stellar populations older than 1 Gyr. Reddy et al. (2018) find that galaxies with EW0([O iii]λ5007) < 100 Å (the typical value at log(M*/M⊙) ~ 10.0 and z ~ 2.3) have ages >1 Gyr according to SED fitting. It is therefore possible that roughly half of the z ~ 2.3 MOSDEF sample has O/Fe⊙ while the other half has enhanced O/Fe. However, ages inferred from SED fitting are sensitive to assumptions regarding the stellar population models, stellar metallicity, attenuation curve, and star-formation history, such that there are large systematic uncertainties on the ages.

To confirm or deny the presence of O/Fe enhancement in the typical z ~ 2 population requires measurements of either auroral-lines or high-S/N rest-FUV continuum, from which Fe/H can be inferred, for individual galaxies spanning the M∗-SFR relation. Detecting [O ii]λ3727 is currently out of reach for galaxies on the mean M∗-SFR relation where fluxes are expected to be ~5 ~ 200 times weaker than for the MOSDEF [O ii]λ3727 emitters in this work. With current facilities, obtaining high-S/N spectra of the rest-FUV is the most viable option over a wide range of M∗, SFR, and metallicity, especially for massive metal-rich galaxies.

If the z ~ 2 star-forming population is “mixed” as implied by the ages in Reddy et al. (2018), with chemically-immature low-mass galaxies having super-solar O/Fe and mature high-mass galaxies having solar O/Fe, it carries implications for measurements of metallicity scaling relations. Assuming a single metallicity calibration in one of the two extremes (i.e., the high-redshift analog vs. z ~ 0 reference calibrations of Bian et al. 2018) would lead to systematic biases in the slope of the MZR, such that the measured MZR would be artificially steepened. If such a scenario holds, a different metallicity calibration must be applied in the low- and high-mass regimes, or else age-dependent calibrations of Bian et al. (2018) would lead to systematic biases in the slope of the MZR, such that the measured MZR would be artificially steepened. If such a scenario holds, a different metallicity calibration must be applied in the low- and high-mass regimes, or else age-dependent calibrations must be constructed. There is likely an age gradient above and below the M∗-SFR relation as well, which will bias measurements of the M∗-SFR-O/H relation for a mixed population. The strong-line metallicities of the highest-SFR local galaxies, analogous to z ~ 2 systems, may also be biased. If these are extreme, the evolution of gas-phase metallicity with global galaxy properties, we must understand what role redshift O/Fe enhancement begins affecting the star-forming population, the relation of O/Fe with M∗ and SFR, and at what redshift the full star-forming population is dominated by Type II SNe enrichment.

Steidel et al. (2016) find super-solar O/Fe using a composite spectrum of 30 star-forming galaxies at z ~ 2.3 falling on the M∗-SFR relation, and we find consistent results when applying our methodology. This finding may signify that many z ~ 2 galaxies on the M∗-SFR relation are O/Fe enhanced, but the sample-averaged measurement cannot address how O/Fe varies across the sample, a pressing question given that stellar population age (and thus O/Fe) varies with M∗ and sSFR. Additionally, there are concerns that the Steidel et al. (2016) KBSS-LM1 stack may be biased such that it is not representative of typical z ~ 2 samples. In Figure 15, the KBSS-LM1 stack (black triangle) has a peculiar low ionization parameter for its direct-method O/H and sSFR, falling significantly below the z ~ 0 relations and lower than any galaxies in our z > 1 sample. Expectations are that high-redshift galaxies will have at least as high U as local galaxies at fixed O/H and sSFR. The offset of the KBSS-LM1 stack is much more significant than individual galaxies scattering low in these relations since it is an average of 30 galaxies. It is possible that the stacking procedure has systematically biased the rest-optical lines towards lower-excitation ratios. Individual determinations of O/Fe (via auroral lines or rest-UV continuum) at higher M∗ and lower sSFR than our z ~ 1 auroral-line sample are sorely needed to validate stacking methods and explore O/Fe as a function of galaxy properties.

Steidel et al. (2016) and Strom et al. (2018) find that harder ionizing spectra at fixed O/H produced by super-solar O/Fe stellar populations are responsible for driving the well-known offset of z ~ 2 galaxies from the local star-forming sequence in the [N ii] BPT diagram. Strom et al. (2018) further find that elevated N/O at fixed O/H is required to explain the most highly-offset z ~ 2 objects. Our results are consistent with this scenario. However, with the current sample, we cannot evaluate whether this is the case for the more massive galaxies for which the [N ii] BPT offset is observed, if O/Fe enhancement alone can fully account for the observed offset, and what the role of N/O variation is. Our current z > 1 auroral-line sample lacks [N ii] coverage and detections in most galaxies. Future high-redshift auroral-line samples spanning a wider dynamic range in M∗ and metallicity will directly address the nature of the [N ii] BPT diagram offset.

5.2 Implications for z ~ 1 ~ 3 extreme emission-line galaxies

Extreme emission-line galaxies (EELGs) are a population of galaxies with spectra dominated by high equivalent width emission lines, including the “green peas” in the local universe (Cardamone et al. 2009). The large equivalent widths of our z > 1 auroral-line sample imply similarities to EELGs. We show the EELG sample of Tang et al. (2018) at 1.3 < z < 2.4 in Figures 15 and 16 (cyan points). This sample was selected to have EW0([O iii]λ5007) > 300 Å. We find that the z > 1 auroral-line sample spans a similar range of SFR, O32, and EW0([O iii]λ5007) at fixed M∗ as the EELGs, which are high-excitation galaxies lying above the M∗-SFR relation at log(M∗/M⊙) ~ 7 ~ 9.5. The high-EW z ~ 2.3 MOSDEF subsample overlaps with the most massive EELGs.

In Figure 16, the auroral-line sample again aligns with the EELG sample, with similar O32 and EW0(Hβ) at fixed EW0([O iii]λ5007). The EELGs appear to be a continuation of the relation between O32, EW0(Hβ), and EW0([O iii]λ5007) defined by the z ~ 2.3 MOSDEF sample, forming remarkably tight sequences over 2.5 orders of magnitude in EW0([O iii]λ5007). We fit these relations, obtaining:

\[ \log(O32) = 0.77x - 1.50 \]  
\[ \log(EW_0(H\beta)) = 0.12x^2 + 0.12x + 0.77 \]  

where \( x = \log(EW_0([O\text{ iii}]\lambda5007)) \), with 0.18 and 0.11 dex scatter in O32 and EW0(Hβ), respectively. These relations can be used to obtain rough information about the O/H...
and ionization parameter of galaxies with large equivalent widths where \(E_{W}(\text{O III}+\text{H} \beta)\) can be inferred from broadband photometry.

Given the common properties of the \(z \approx 2\) EELGs and \(z > 1\) auroral-line sample, we conclude that the EELGs are also young metal-poor galaxies with super-solar O/Fe, driving extremely hard ionizing spectra in their H II regions. By modeling the photometry and rest-optical spectra simultaneously, Tang et al. (2018) find ages younger than 300 Myr for the majority of their sample. Tang et al. (2018) suggest EELGs have hard ionizing spectra and high ionization parameters based on their large O32 values. The O32 high-redshift analog calibration of Bian et al. (2018) yields metallicities spanning \(\log(O/H)=7.7-8.4\) for the EELG sample, with a median of \(12+\log(O/H)_{\text{med}}=8.07\), similar to the O/H distribution of the auroral-line sample. At these metallicities, the \(z = 0\) \(U\)-O/H relation predicts high ionization parameters of \(\log(U)=-2.0\) to \(-2.5\) (Pérez-Montero 2014). When modeling the nebular spectra of EELGs, we suggest using BPASS models including binaries, and assuming \(Z_{\odot}=1.5Z_{\text{med}}\) (the pure Type II SNe limit) and the \(z = 0\) \(U\)-O/H relation.

5.3 Implications for typical \(z > 6\) galaxies and reionization

Star-forming galaxies at \(z > 6\) are thought to be the sources of ionizing photons that powered reionization (e.g., Bouwens et al. 2012, 2015; Robertson et al. 2015; Stark 2016). Through SED fitting and broadband photometric excess, these galaxies have been found to have very large \(E_{W}(\text{O III}+\text{H} \beta)\), with typical values of \(\approx 670\) Å extending up to \(\approx 1500\) Å, and high sSFRs \((\sim 10\ Gyr^{-1})\) (Labbé et al. 2013; Smit et al. 2014, 2015; Roberts-Borsani et al. 2016). These properties are similar to those of the \(z > 1\) auroral-line sample and the \(z \approx 2\) EELGs. Indeed, the Tang et al. (2018) sample was selected to be analogs of epoch of reionization galaxies. Given the similar properties of these two samples, we suggest that typical star-forming galaxies at \(z > 6\) have super-solar O/Fe.

The age of the universe at \(z = 6\) is 900 Myr. Assuming the first galaxies began forming \(\approx 200\ Myr\) after the big bang, the oldest possible stellar populations at \(z = 6\) would have had an age of \(\approx 700\ Myr\). This maximum age would decrease at higher redshifts when reionization was in progress \((\approx 400\ and\ 250\ Myr\ at\ z = 8\ and\ 10,\ respectively)\). A spectroscopically-confirmed \(z \sim 10\) galaxy has been found to have an age of 340 Myr, close to the age limit (Hoag et al. 2018). Labbé et al. (2013) find a younger age of 100 Myr for a stacked SED of \(z \approx 8\) galaxies. Due to the maximum possible age of galaxies at \(z > 6\), every galaxy in the epoch of reionization should host an O/Fe-enhanced stellar population. The properties of our \(z > 1\) auroral-line sample provide observational evidence for this expectation. Just as for \(z \approx 1-3\) EELGs, we suggest employing BPASS binary stellar models with \(Z_{\odot}=1.5Z_{\text{med}}\) and the local relation between \(U\) and \(O/H\) when modeling the nebular spectra of \(z > 6\) galaxies. Based on the results in Figure 3, we find that the most useful strong-line ratios for determining gas-phase oxygen abundance from rest-optical spectroscopy of \(z > 6\) galaxies are O32 and Ne3O2, which will be accessible out to \(z > 9\) and 12, respectively, with \(\text{JWST/NIRSpec\ and\ NIRCam\ spectroscopy}\). While it is possible that [N ii]-based indicators may also be useful, they can only be observed out to \(z \approx 6.6\) with NIRSpec and NIRCam.

Super-solar O/Fe values carry important implications for reionization. Massive stars with lower Fe abundances produce harder spectra and a larger number of ionizing photons per unit stellar mass. The latter makes it easier for star-forming galaxies to reionize the universe, requiring lower escape fractions. Low-mass, faint galaxies are thought to provide the bulk of the ionizing photons due to their number (e.g., Bouwens et al. 2012, 2015; Finkelstein et al. 2012; Stark 2016). Such galaxies should form latest and thus have the youngest stellar populations with maximal O/Fe\(=5xO/Fe_{\odot}\) based on current Type II SNe yields (Nomoto et al. 2006). Assuming \(z > 6\) galaxies follow the relations in Figure 17, we use equations 12 and 13 to estimate O32 values of \(z > 6\) galaxies. Labbé et al. (2013) use stacked broadband photometry to estimate \(E_{W}(\text{O III}+\text{H} \beta)\)=760 Å of \(L^*\) galaxies at \(z \sim 8\). This equivalent width corresponds to \(E_{W}(\text{O III}]=45000]/=420\ Â, \(E_{W}(H \beta]=80\ Â, \log(O32]=0.52, \text{and}\ 12+\log(O/H]=8.15\ (0.3\ Z_{\odot})\), assuming \(Z_{\odot}=0.06\ Z_{\odot}\). Roberts-Borsani et al. (2016) find \(E_{W}(\text{O III}]+H \beta]=1500\ Â\) for four bright galaxies at \(z \sim 7-9\), implying \(E_{W}(\text{O III}]=45000]/=1000\ Â, \(E_{W}(H \beta]=160\ Â, \log(O32]=0.8, \text{and}\ 12+\log(O/H]=8.0\ (0.2\ Z_{\odot})\). For the Roberts-Borsani et al. (2016) sample, maximal O/Fe enhancement implies \(Z_{\odot}=0.04\ Z_{\odot}\). These rough estimates suggest that metal-poor massive stars with \(\lesssim 5\%\ Z_{\odot}\) power the ionizing spectra of the star-forming galaxies responsible for reionizing the universe.

6 SUMMARY & CONCLUSIONS

We have compiled a sample of 18 individual star-forming galaxies at \(z = 1.4-3.7\) with detections of auroral [O III],[4363] or O III],[4663], for which temperature-based nebular oxygen abundances can be derived independent of strong-line ratio calibrations. Four of these are galaxies with [O III],[4363] detections from the MOSDEF survey, three of which have not been previously published. We utilized this sample to investigate the chemical abundances and ionization states of star-forming regions at \(z > 1\). For the first time, we have constructed metallicity scaling relations, including the mass-metallicity relation and fundamental metallicity relation, at \(z > 1\) based purely on direct-method metallicities. The independently-constrained nebular metallicities enable us to unambiguously constrain the stellar metallicities and ionization parameters of these galaxies using photoionization models, providing unprecedented insight into the gas physical conditions in \(z > 1\) H II regions. We summarize our main results and conclusions below.

(i) We investigated whether strong-line metallicity calibrations change with redshift by directly comparing the strong-line ratios O3, O2, R23, O32, and Ne3O2 at fixed direct-method metallicity for samples at \(z \approx 0-2.2\) (Fig. 3). Over \(12+\log(O/H]=7.7-8.1\), we find that O3 and R23 are saturated and are not useful metallicity indicators. The median values of the \(z > 1\) auroral-line sample closely match the \(z \approx 0\) high-redshift analogs of Bian et al. (2018) to
within \( \sim 0.1 \) dex in O/H at fixed line ratio. We suggest the use of the Bian et al. (2018) high-redshift analog calibrations at \( z > 1 \), with O32 and Ne3O2 providing the most utility as these ratios are monotonic with metallicity over \( 7.7 < z < 12 + \log(O/H) < 8.1 \). With JWST spectroscopic measurements of O32 and Ne3O2 can be made to \( z \sim 9 \) and \( z \sim 12 \), respectively. Local H II region calibrations should not be utilized at \( 12 + \log(O/H) \leq 8.0 \) for this set of strong-line ratios. Our auroral-line sample did not have sufficient [N II] coverage and detections to investigate N/O and nitrogen-based metallicity indicators.

(ii) We construct the stellar mass—gas-phase metallicity relation using temperature-based metallicities for the first time at \( z > 1 \) (Fig. 4). We find a correlation between \( M_\* \) and direct-method O/H (eq. 11). After correcting for SFR biases in the \( z > 1 \) sample, we find that metallicity decreases by \( 0.5 \pm 0.1 \) dex from \( z \sim 0 \) to \( z \sim 2.2 \) at \( \log(M_\*/M_\odot) = 9.5 \), slightly larger than found with strong-line methods. We find tentative evidence for a \( M_\*-\text{SFR}-O/H \) relation at \( z \sim 2.2 \) based on direct-method metallicities (Fig. 6). Galaxies at \( z > 1 \) have the same oxygen abundance of galaxies at \( z \sim 0 \) at fixed \( M_\* \) and SFR to within 0.1 dex, suggesting that the FMR is nearly redshift invariant (Fig. 7), although a slight shift towards lower O/H with increasing redshift at fixed \( M_\* \) and SFR is present.

(iii) Using photoionization models, we uniquely leverage the direct-method nebular metallicity measurements to obtain constraints on the ionization parameter, \( U \), and stellar metallicity of massive stars, \( Z_\* \), in star-forming regions in the \( z > 1 \) auroral-line galaxies (Sec. 4). For the majority of the \( z > 1 \) auroral-line sample, we find that \( Z_\* \), which traces Fe/H, is lower than the nebular metallicity, following O/H, on a solar abundance scale. This result implies super-solar O/Fe values, in contrast to typical \( [\text{O} II] \) values, with \( Z_\* = Z_{\text{neb}} \). The super-solar O/Fe values can be explained by a physically-motivated picture in which the young stellar populations of these galaxies have been enriched by Type II SNe that occur promptly following a star-formation event and are the dominant production channel for O, but have not yet been fully enriched by Type Ia SNe that are delayed \( \sim 0.3 - 1 \) Gyr and are the dominant producer of Fe-poor elements. The four MOSDEF \( [\text{O} II] \lambda4363 \) emitters have ages \( \lesssim 300 \) Myr, and are younger than 90% of MOSDEF galaxies matched in redshift and \( M_\* \).

(iv) On average, the \( z > 1 \) auroral-line sample has the same \( U \) as local galaxies and H II regions with the same O/H, and the same \( U \) and O/H as \( z \sim 0 \) galaxies matched in sSFR (Fig. 15). The high ionization parameters of the \( z > 1 \) sample are driven by their high sSFRs and low oxygen abundances, with \( U \) values as expected from local relations between \( U \), sSFR, and O/H. We conclude that evolution in \( U \) at fixed O/H is not a significant driver of the evolution of strong-line ratios with redshift. Harder ionizing spectra at fixed O/H compared to the local universe are a cause of strong-line ratio evolution for our \( z > 1 \) auroral-line sample due to super-solar O/Fe. Due to a lack of [N II] coverage and detections, we are unable to evaluate the role of N/O with the current sample and cannot draw conclusions about the positions of \( z \sim 2 \) galaxies in the [N II] BPT diagram.

(v) The \( z > 1 \) auroral-line sample has significantly higher sSFR, O32, and emission-line equivalent widths than typical \( z \sim 2.3 \) galaxies (Figs. 16 and 17), indicating they are younger and more metal-poor than the bulk of the \( z \sim 2 \) star-forming population. It is therefore unclear if our results carry over to typical galaxies at these redshifts, in particular because stellar population age is expected to vary with \( M_\* \) and sSFR. It is possible that \( z > 2 \) galaxies are a mixed population, where low-mass and starburst galaxies are young and have super-solar O/Fe, while more massive galaxies have lower levels of O/Fe enhancement, potentially solar O/Fe at \( \log(M_\*/M_\odot) \geq 10.5 \). As a result of such variation, metallicity measurements would be biased as a function of \( M_\* \) and SFR when assuming a single calibration for the full sample, affecting the measured shape of the MZR and \( M_\*-\text{sSFR}-O/H \) relation. O/Fe constraints for galaxies falling on the \( M_\*-\text{sSFR} \) relation and more massive, metal-rich galaxies are needed to obtain robust constraints on metallicity scaling relations in the high-redshift universe.

(vi) The properties of the \( z > 1 \) auroral-line sample are closely matched to those of extreme emission-line galaxies at \( z > 2 \) (Tang et al. 2018). We conclude that EELGs have super-solar O/Fe, but similar \( U \) to local galaxies with the same O/H. We find that typical \( z > 2.3 \) MOSDEF galaxies and EELGs form a tight sequence in O32 and EW\(_{\beta}(\text{H}\beta)\) as a function of EW\(_{\beta}(\text{[O III]\lambda5007})\) across 2.5 orders of magnitude (eqs. 12 and 13).

(vii) The emission-line equivalent widths and sSFRs of \( z > 6 \) galaxies also match the properties of the \( z > 1 \) auroral-line sample. We extend our results to typical star-forming galaxies in the epoch of reionization, implying more efficient production of ionizing photons from Fe-poor massive stars. Because of a maximum stellar population age imposed by the age of the universe, all galaxies at \( z > 6 \) are expected to have super-solar O/Fe. We use the O32 and EW\(_{\beta}(\text{H}\beta)\) vs. EW\(_{\beta}(\text{[O III]\lambda5007})\) relations to estimate the nebular and stellar metallicities of UV-luminous \( z > 6 \) galaxies, finding typical values of \( Z_{\text{neb}} \sim 0.25 Z_\odot \) and \( Z_\* \sim 0.05 Z_\odot \).

Ultimately, a larger sample of \( z > 1 \) auroral-line emitters spanning a wider dynamic range in oxygen abundance is needed to make a full assessment of metallicity indicators across the range of metallicities spanned by large spectroscopic survey datasets. Such a sample would allow for tests of calibration shapes, not just normalization as performed in this work. Our results suggest that finding galaxies with \( [\text{O} II] \lambda4363 \) that can be detected with current facilities can be achieved by identifying galaxies with EW\(_{\beta}(\text{[O III]\lambda5007})\geq 300 \) \( \AA \) and high SFR (\( \gtrsim 10 M_\odot \) yr\(^{-1} \)), but this selection will only yield metal-poor, low-mass galaxies. Obtaining auroral \( [\text{O} II] \lambda7320,7330 \) measurements for massive, moderately metal-rich galaxies with high SFRs presents a potential avenue to extend the metallicity baseline, improving calibration tests and constraints on the mass-metallicity relation.

The ionizing spectrum of massive stars is so strongly tied to the production of emission lines that it is imperative to understand the properties of massive stars at low and high redshifts in order to extract robust star-formation rates, metallicities, and ionization state information from spectra. To gain a more complete picture of massive stars at high redshifts, we must constrain O/Fe of individual galaxies spanning the \( M_\* \) and SFR range of the star-forming population. Reaching this goal will require auroral-line measurements and high-S/N FUV continuum spectroscopy of a large...
number of high-redshift galaxies over a wide dynamic range of properties. The increased sensitivity of JWST and 30 m-class telescopes will be needed to achieve this goal. Until that time, we must continue extrapolating from small samples with detailed observations to infer the chemical enrichment of the large spectroscopic datasets currently available at $z \sim 1$–3.

ACKNOWLEDGEMENTS

Based on data obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, and NASA, and was made possible by the generous financial support of the W.M. Keck Foundation. We acknowledge support from NSF AAG grants AST-1312780, 1312547, 1312764, and 1313171, archival grant AR-13907 provided by NASA through the Space Telescope Science Institute, and grant NNX16AF54G from the NASA ADAP program. We also acknowledge a NASA contract supporting the AÄJW-FIRST Extragalactic Potential Observations (EXPO) Science Investigation TeamÅÎ (15-WFIRST15-0004), administered by GSFC. We additionally acknowledge the 3-D-HST collaboration for providing spectroscopic and photometric catalogs used in the MOSDEF survey. We wish to extend special thanks to those of Hawaiian ancestry on whose sacred mountain we are privileged to be guests. Without their generous hospitality, the work presented herein would not have been possible.

REFERENCES

Aggarwal K. M., Keenan F. P., 1999, ApJS, 123, 311
Andrews B. H., Martini P., 2013, ApJ, 765, 140
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Atek H., et al., 2011, ApJ, 743, 121
Azadi M., et al., 2017, ApJ, 855, 27
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Bayliss M. B., Rigby J. R., Shawr K., Wuyts E., Florian M., Gladders M. D., Johnson T., Oguri M., 2014, ApJ, 790, 144
Berg D. A., Skillman E. D., Croxall K. V., Pogge R. W., Moustakas J., Johnson-Groh M., 2015, ApJ, 806, 16
Berg D. A., Erb D. K., Auger M. W., Pettini M., Brammer G. B., 2018, ApJ, 859, 164
Bian F., Kewley L. J., Dopita M. A., 2018, ApJ, 859, 175
Blanc G. A., Kewley L., Vogt F. P. A., Dopita M. A., 2015, ApJ, 798, 99
Bouwens R. J., et al., 2012, ApJ, 752, L5
Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, ApJ, 811, 140
Brammer G. B., et al., 2012a, ApJS, 200, 13
Brammer G. B., et al., 2012b, ApJ, 758, L17
Bressolin F., Kudritzki R.-P., Urbaneja M. A., Gieren W., Ho I.-T., Pietrzyński G., 2016, ApJ, 830, 64
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1153
Brown J. S., Martini P., Andrews B. H., 2016, MNRAS, 458, 1529
Brualdi G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Campbell A., Terlevich R., Melnick J., 1986, MNRAS, 223, 811
Cardamone C., et al., 2009, MNRAS, 399, 1191
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Chabrier G., 2003, PASP, 115, 763
Christensen L., et al., 2012a, MNRAS, 427, 1953
Christensen L., et al., 2012b, MNRAS, 427, 1973
Coil A. L., et al., 2015, ApJ, 801, 35
Conroy C., Gunn E. J., White M., 2009, ApJ, 699, 486
Cowie L. L., Barger A. J., Songaila A., 2016, ApJ, 817, 57
Cresci G., Mannucci F., Curti M., 2018, arXiv e-prints,
Croxall K. V., Pettini M., Steidel C. C., Moustakas J., 2015, ApJ, 808, 42
Croxall K. V., Pettini M., Steidel C. C., Moustakas J., 2016, ApJ, 830, 4
Cullen F., Cirasuolo M., McLoore R. J., Dunlop J. S., Bowler R. A. A., 2014, MNRAS, 440, 2300
Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Espostio S., 2017, MNRAS, 465, 1384
Davé R., Finlator K., Oppenheimer B. D., 2012, MNRAS, 421, 98
Davé R., Rafieferantssoa M. H., Thompson R. J., Hopkins P. F., 2017, MNRAS, 467, 115
Davé R., Anglés-Alcazar D., Narayanan D., Li Q., Rafieferantssoa M. H., Appleby S., 2019, MNRAS, 486, 2827
De Rossi M. E., Bower R. G., Font A. S., Schaye J., Theuns T., 2017, preprint, (arXiv:1704.00006)
Dickey C. M., et al., 2016, ApJ, 828, L11
Dopita M. A., et al., 2006a, ApJS, 167, 177
Dopita M. A., et al., 2006b, ApJ, 647, 244
Dopita M. A., Kewley L. J., Sutherland R. S., Nichols D. C., 2016, Ap&SS, 361, 61
Eldridge J. J., Stanway E. R., 2012, MNRAS, 419, 479
Eldridge J. J., Langer N., Tout C. A., 2011, MNRAS, 414, 3501
Ellison S. L., Patton D. R., Simard L., McConnell A. W., 2008, ApJ, 672, L107
Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813
Erb D. K., Pettini M., Shapley A. E., Steidel C. C., Law D. R., Reddy N. A., 2010, ApJ, 719, 1168
Erb D. K., Pettini M., Steidel C. C., Strom A. L., Rudie G. C., Trainor R. F., Shapley A. E., Reddy N. A., 2016, ApJ, 830, 52
Esteban C., García-Rojas J., Carigi L., Peimbert M., Bressolin F., López-Sánchez A. R., Mesa-Delgado A., 2014, MNRAS, 443, 624
Fensch J., et al., 2017, MNRAS, 465, 1934
Ferland G. J., et al., 2017, RMxAA, 53, 385
Finlator K., Davé R., 2008, MNRAS, 385, 2181
Fischer C. F., Tachiev G., 2014, MCHF/MCDHF Collection, Version 2, Ref No. 10 & 20, Available online at http://physics.nist.gov/mchf
Fosbury R. A. A., 2014, MNRAS, 440, 2303
Geer E., Allam S. S., Tucker D. L., 2009, ApJ, 701, 52
Gomez F. M., et al., 2011, ApJS, 193, 35
Hainline K. N., Shapley A. E., Kornei K. A., Pettini M., Buckley-Geer E., Alam S. M., Tucker D. L., 2009, ApJ, 701, 52
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Henry A., et al., 2013, ApJ, 775, L27
Hoag A., et al., 2018, ApJ, 854, 39
Holden B. P., et al., 2016, ApJ, 820, 73
Horne K., 1986, PASP, 98, 699
Izotov Y. I., Stasińska G., Meynet G., Guseva N. G., Thuan T. X., 2017, Ap&SS, 361, 48
Izotov Y. I., Stasińska G., Meynet G., Guseva N. G., Thuan T. X., 2017, Ap&SS, 361, 48
James B. L., et al., 2014, MNRAS, 440, 1794
Jerabkova T., Hasani Zonoizi A., Kroupa P., Becki G., Yan Z., Vazdekis A., Zhang Y., 2018, A&A, 620, A39
APPENDIX A: LITERATURE $z > 1$

AURORAL-LINE SAMPLE

In this appendix, we describe the details of the oxygen abundance and galaxy property calculations on a case-by-case basis for each of the [O\textsc{iii}]4363 and [O\textsc{ii}]1663 emitters taken from the literature. We also explain why we do not include the claimed [O\textsc{iii}]4363 detection from Yuan & Kewley (2009) in our analysis.

A1 Literature [O\textsc{iii}]4363 emitters

Brammer et al. (2012b) report a marginal 1.7σ detection of [O\textsc{iii}]4363 for SL25J02176-0515 (B18), a star-forming galaxy at $z = 1.847$ magnified by a factor of 19. We estimate zero nebular reddening according to the observed Balmer ratio of Hβ/Hδ = 0.18 ± 0.03, assuming an intrinsic value of Hδ/Hβ = 0.259. The [O\textsc{iii}]143726,3729 doublet is not included in the HST/WFC3 G141 grism wavelength coverage. Following Brammer et al. (2012b), we instead assume a uniform distribution of [2.5, 9.0] for the [O\textsc{iii}]145007/[O\textsc{iii}]143726,3729 ratio (Erb et al. 2010; Atek et al. 2011) and the mean value of [O\textsc{iii}]145007/[O\textsc{ii}]143726,3729 = 5.75 for calculating the metallicity. This range of [O\textsc{iii}]14/[O\textsc{ii}] values is typical of galaxies at the same stellar mass as B18 (see Fig. 17).

The oxygen abundance under this set of assumptions is 12+log(O/H) = 7.56^{+0.35}_{-0.35}. We note that O* makes up at most 30% of O under the range of assumed [O\textsc{iii}]14/[O\textsc{ii}] values, and the total oxygen abundance only varies by ±0.06 dex. Properties of B18 are also reported in Berg et al. (2018) (see A2).

Christensen et al. (2012a,b) report a 3.4σ detection of [O\textsc{iii}]4363 for a lensed galaxy at $z = 1.833$, Abell 1689 arc ID 31.1 (C12a), magnified by a factor of 26.6. Nebular reddening is estimated from Hγ and Hδ, with Hδ/Hβ = 0.40±0.07 corresponding to E(B-V)$_{\text{gas}}$ = 0.31^{+0.06}_{-0.09}, assuming an intrinsic ratio of Hδ/Hβ = 0.468. The direct-method oxygen abundance is 12+log(O/H) = 7.46^{+0.23}_{-0.21}. We adopt the rest-frame equivalent width reported in Stark et al. (2014, ID 876,330 therein) of EW$_{(O\textsc{iii})}$1663,3729+Hβ = 740±190 Å, and determine EW$_{(O\textsc{iii})}$143007 and EW$_{(H\beta)}$ based on the measured O3 ratio.

Stark et al. (2013) report a detection of [O\textsc{ii}]1663 for CSWA 141 (S13) at $z = 1.425$ with a significance of 7.3σ. S13 is magnified by a factor of 5.5. Nebular reddening is estimated from Hβ, Hγ, and Hδ, and is found to be small (E(B-V)$_{\text{gas}}$ = 0.03). The total oxygen abundance is 12+log(O/H) = 7.96^{+0.07}_{-0.07}. The stellar mass of S13 is log(M$_{*}$/M$_{\odot}$) = 8.33^{+0.10}_{-0.14} (private communication, D. Stark & R. Maiolino).

A marginal 1.7σ detection of [O\textsc{iii}]4363 is reported by James et al. (2014) for CSWA 20 (J14), a star-forming galaxy at $z = 1.433$ magnified by a factor of 11.5. Hα, Hβ, and Hγ are used to determine nebular reddening. The total oxygen abundance of J14 based on [O\textsc{ii}]1663 is 12+log(O/H) = 8.13^{+0.35}_{-0.21}. Patricio et al. (2018) report a stellar mass of log(M$_{*}$/M$_{\odot}$) = 10.3 ± 0.3 for J14, estimated by fitting the galaxy SED obtained from SDSS broadband photometry, with z-band being the reddest filter at $\lambda_{\text{rest}} = 3700$ Å. James et al. (2014) measure a line width of 34.9 ± 0.7 km s$^{-1}$ for the narrow systemic component of strong nebular emission features in the spectrum of J14, which also display a broad blueshifted component. The high stellar mass of Patricio et al. (2018) is inconsistent with such narrow line widths for standard disk and spheroid geometries, requiring unphysically large virial coefficients to reconcile the two in dynamical modeling. Additionally considering systematic uncertainties in lens modeling and the lack of broadband photometry blueward of the Balmer break, we conclude that the reported stellar mass of J14 from Patricio et al. (2018) is not reliable and likely overestimates the true stellar mass. We do not report M$_{*}$ or sSFR for J14 in Table 2.

A2 Literature O\textsc{iii}\lambda1661,1666 emitters

Christensen et al. (2012a,b) report O\textsc{iii}\lambda1663 detections for three strongly lensed galaxies at $z > 1$. Abell 1689 arc ID 31.1 (C12a) has detections of both O\textsc{iii}\lambda1663 and O\textsc{iii}\lambda1663. C12a has a total oxygen abundance of 12+log(O/H) = 7.34^{+0.44}_{-0.70} based on O\textsc{iii}\lambda1663. Despite a remarkable 21.2σ detection of O\textsc{iii}\lambda1663, the uncertainty on the metallicity is very large due to the large uncertainty of the dust correction and the extreme sensitivity of O\textsc{iii}\lambda1663/O\textsc{ii}\lambda5007 to reddening. The oxygen abundance derived using O\textsc{iii}\lambda1663 is consistent with that based on [O\textsc{iii}]4363 (see A1), although the uncertainty is significantly lower when using [O\textsc{iii}]4363. We adopt the total oxygen abundance based on [O\textsc{iii}]4363 for this target.

SMACS J0304 (C12b) at $z = 1.9634$ has a 7.5σ detection of O\textsc{iii}\lambda1663 (Christensen et al. 2012a,b). Nebular reddening is estimated using Hα, Hβ, Hγ, and Hδ. The direct-method oxygen abundance is 12+log(O/H) = 8.14^{+0.04}_{-0.04}. Christensen et al. (2012a) note C12b as a merging system based on its complex morphology, with 5 distinct components at close projected separation ($\leq 20$ kpc, uncorrected for magnification) in HST imaging. A 12.7σ O\textsc{iii}\lambda1663 detection is reported for M2031 (c12c; Christensen et al. 2012a,b). Hβ and He suggest little nebular reddening assuming an intrinsic ratio of Hδ/Hβ = 0.159 (Osterbrock & Ferland 2006). The total oxygen abundance is 12+log(O/H) = 7.77^{+0.30}_{-0.30}. C12c is an interacting system, with a companion located at the same redshift ($\Delta v < 50$ km s$^{-1}$) and 11.2 kpc projected separation in the source plane (Patricio et al. 2016).

James et al. (2014) report a 2.4σ detection of O\textsc{iii}\lambda1663 for the gravitationally-lensed galaxy CSWA 20 (J14), which also has a marginal detection of [O\textsc{iii}]4363 (see A1). The direct-method oxygen abundance derived from O\textsc{iii}\lambda1663 is 12+log(O/H) = 7.86^{+0.15}_{-0.14}. The metallicity based on O\textsc{iii}\lambda4363 is $\sim 1σ$ consistent with this value. Given that both O\textsc{iii}\lambda4363 and O\textsc{iii}\lambda1663 are marginally detected (1.7σ and 2.4σ, respectively), the derived O/H uncertainties are comparable, and the large sensitivity to reddening correction for O\textsc{iii}\lambda1663, we utilize the metallicity based on O\textsc{iii}\lambda4363 for J14. Our results do not change when instead utilizing the O\textsc{iii}\lambda1663-based metallicity.

Bayliss et al. (2014) present a 6.9σ O\textsc{iii}\lambda1663 detection for a lensed star-forming galaxy at $z = 3.6252$ (B14). Hβ and Hγ are both detected, but Hγ is contaminated by skylines. The measured Hγ/Hβ = 0.52 ± 0.02 is 2.5σ inconsistent with the maximum value of 0.468 in the case of no nebular reddening (Osterbrock & Ferland 2006), again suggesting Hγ is unreliable. Following Bayliss et al. (2014), we adopt A_V = 1.0 ± 0.2 from their SED fitting and assume that...
the nebular and stellar reddening is equal. This assumption yields 12+log(O/H) = 8.10±0.09.

Stark et al. (2014) present 3.5σ detections of O [III]41663 for two lensed galaxies: Abell 860.359 at z = 1.7024 (S14a) and MACS 0451 ID 1.1b at z = 2.0596 (S14b). Nebular reddening is estimated from Ha and Hβ for each source. [O [III]43726,3729 is not detected for either galaxy, but 2σ upper limits are reported. We adopt the 2σ [O [II] upper limit values when calculating the metallicity, and a uniform distribution between zero and 3σ for [O [II]] when creating realizations for uncertainty estimation. Under this assumption, both objects are still dominated by doubly-ionized oxygen, with ∼ 20% of O in O II. Varying the assumed [O [II]] values between zero and 3σ changes the total oxygen abundances by < 0.07 dex. The derived direct-method oxygen abundances are 12+log(O/H) = 8.05±0.11 for S14a and 12+log(O/H) = 7.31±0.14 for S14b. The EW(O [III]λ4959) and EW(Hβ) from 860.359 is inferred from the reported EW(O [III]4349,4363+Hβ) = 1550 ± 150 A measured on the O3 ratio.

Three unlensed galaxies from Erb et al. (2016) with detections of O [III]41663 are included in the auroral-line sample of Patrício et al. (2018). These galaxies are BXT4 (E16a), BX418 (E16b), and BX660 (E16c) at z = 2.1889, 2.3054, and 2.1742, respectively, with 12σ, 3.5σ, and 5.2σ detections of O [III]41663. For all three galaxies, we take rest-optical line fluxes and stellar masses from Erb et al. (2016), and determine nebular reddening using Ha and Hβ. We use the O [III]41663 flux from Erb et al. (2016) for E16a. For E16b and E16c, we infer the O [III]41663 fluxes from the reddened intensities relative to Hβ presented in Patrício et al. (2018) in combination with the Hβ fluxes from Erb et al. (2016). The total oxygen abundances are 12+log(O/H) = 7.91±0.11 for E16a, 12+log(O/H) = 7.74±0.13 for E16b, and 12+log(O/H) = 7.98±0.08 for E16c.

Kojima et al. (2017) report a 6.5σ detection of O [III]41663 for COSMOS 12805 (K17), an unlensed galaxy at z = 2.159. A 2σ upper limit is reported for O [III]43726,3729. We adopt the same strategy as for the two galaxies from Stark et al. (2014), finding a total oxygen abundance of 12+log(O/H) = 8.26±0.03. Assuming the 2σ [O [II]] upper limit for the abundance calculation gives 34% of total O in the singly-ionized state. The significant uncertainty of the [O [II]] flux and, consequently, the O abundance is due to the large lower error on O/H.

Berg et al. (2018) present a very high S/N rest-UV spectrum of SL2S J02176-0513 (B18) with S/N>50 in each component of the O III]41661,41666 doublet. Rest-optical HST/grism spectroscopy has also revealed a marginal [O [III]4363 detection (Brammer et al. 2012b, see A1). We adopt the updated magnification factor and stellar mass of B18 from Berg et al. (2018). The rest-UV spectrum reveals an extremely high level of ionization based on strong nebular C iv and Ne ii, and models presented in Berg et al. (2018) suggest <2% of O is in O+ with a non-negligible fraction in O++. We employ their oxygen ionization correction factor of O/H=1.055xO++/H and find the total oxygen abundance to be 12+log(O/H) = 7.3±0.05 using O [III]41663. We adopt this value for B18, which is consistent with but better constrained than the abundance based on the marginal [O [III]4363 detection from Brammer et al. (2012b).

The O III]41663 doublet is detected at 14.3σ significance for the highly-magnified Lynx arc at z = 3.357 (V04; Bosbury et al. 2003; Villar-Mártn et al. 2004). Reddening cannot be estimated from Balmer lines since only Hβ is available. Following Patrício et al. (2018), we assume zero nebular reddening based on SED modeling of the rest-optical and rest-UV spectrum (Villar-Mártn et al. 2004). In the absence of a detection of [O II], we follow the same procedure as for the two galaxies from Stark et al. (2014). We note that the Lynx arc is highly ionized, with O+ making up at most 4% of O (3σ upper limit), such that the [O II] non-detection has a negligible impact on the nebular metallicity determination. The total oxygen abundance is 12+log(O/H) = 7.76±0.04. While a stellar mass is given in Villar-Mártn et al. (2004), it is important to note that they only observe one bright knot of the highly-magnified arc with a particularly low mass-to-light ratio (possibly a single lensed H II region) and that the stellar mass is that of an instantaneous burst simple stellar population able to power the Hβ emission. The stellar mass from Villar-Mártn et al. (2004) thus only represents the mass of a star cluster ionizing the H II region, not the global stellar mass of the galaxy, and is not suitable for studying metallicity scaling relations with global galaxy properties.

A3 The case of the Yuan & Kewley (2009) [O III]4363 detection

A detection of [O III]4363 for Abell 1689 Lens 22.3 at z = 1.706 was presented in Yuan & Kewley (2009), with a magnification of 15.5. The reported [O III]4363 flux is 0.27±0.1 on a normalized scale where the Hβ flux is 1.0. Yuan & Kewley (2009) measure a redshift from [O III]4363 of z = 1.696 that is ∆z = 0.009 lower than the redshift measured from strong rest-optical lines, z = 1.705. This redshift offset corresponds to a velocity offset of ~ 1000 km s⁻¹ (∆Δz = 15 A in the rest-frame), and calls into question the legitimacy of the [O III]4363 detection given that [O III]4363 emission should be at rest with respect to other rest-optical nebular emission lines. Based on the redshift from the strong lines (z = 1.705), the rest-frame centroid of the line in question is 4348 Å, close to the wavelength of Hγ (4342 Å). Given that the spectrum in Yuan & Kewley (2009) has a low spectral resolution of R ~ 500 (9 Å resolution in the rest-frame at λ = 4350 Å), we argue that Hγ was misidentified as [O III]4363. Indeed, Hγ is almost always significantly stronger than [O III]4363 (the case for every other source in our z > 1 auroral-line sample), and should be easily detected if [O III]4363 is also detected unless Hγ is contaminated by a sky line. Yuan & Kewley (2009) report a detector response-corrected ratio of Hγ/Hβ = 5.03. This Balmer ratio implies a nebular reddening of E(B-V)gas = 0.57. Based on this E(B-V)gas value, the expected Hγ/Hβ ratio is 0.34, consistent with the reported [O III]4363/Hβ = 0.27 ± 0.1 from Yuan & Kewley (2009). We conclude that the claimed [O III]4363 detection in Yuan & Kewley (2009) is in fact a misidentification of Hγ, and exclude this source from our literature z > 1 auroral-line sample. Ghurek et al. (2019, in prep.) have recently obtained a deep R ~ 3000 spectrum of Lens 22.3 and confirmed that the line reported in Yuan & Kewley (2009) is Hγ and no significant [O III]4363 is present.

This paper has been typeset from a TeX file prepared by the author.