The Astrophysical Journal, 985:24 (18pp), 2025 May 20 https://doi.org/10.3847/1538-4357/adc7b5


# Unveiling the Cosmic Chemistry. II. "Direct" $T_e$-based Metallicity of Galaxies at $3 < z < 10$ with JWST/NIRSpec

Priyanka Chakraborty[1], Arnab Sarkar[2], Randall Smith[1], Gary J. Ferland[3], Michael McDonald[2], William Forman[1], Mark Vogelsberger[2], Paul Torrey[4], Alex M. Garcia[4], Mark Bautz[2], Adam Foster[1], Eric Miller[2], and Catherine Grant[2]

[1] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA; priyanka.chakraborty@cfa.harvard.edu
[2] Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 70 Vassar Street, Cambridge, MA 02139, USA
[3] University of Kentucky, Lexington, KY, USA
[4] University of Virginia, Charlottesville, VA, USA
*Received 2024 December 18; revised 2025 March 20; accepted 2025 March 30; published 2025 May 13*

## Abstract

We report the detection of the [O III] auroral line in 42 galaxies within the redshift range of $3 < z < 10$. These galaxies were selected from publicly available JWST data releases, including the JADES and PRIMAL surveys, and observed using both the low-resolution PRISM/CLEAR configuration and medium-resolution gratings. The measured electron temperatures in the high-ionization regions of these galaxies range from $T_e([\mathrm{O\,III}]) = 12{,}000$ to 24,000 K, consistent with temperatures observed in local metal-poor galaxies and previous JWST studies. In 10 galaxies, we also detect the [O II] auroral line, allowing us to determine electron temperatures in the low-ionization regions, which range between $T_e([\mathrm{O\,II}]) = 10{,}830$ and 20,000 K. The direct $T_e$-based metallicities of our sample span from $12 + \log(\mathrm{O/H}) = 7.2$ to 8.4, indicating these high-redshift galaxies are relatively metal-poor. By combining our sample with 25 galaxies from the literature, we expand the data set to a total of 67 galaxies within $3 < z < 10$, effectively more than doubling the previous sample size for direct $T_e$-based metallicity studies. This larger data set allows us to derive empirical metallicity calibration relations based exclusively on high-redshift galaxies, using six key line ratios: R3, R2, R23, Ne3O2, O32, and O3N2. Notably, we derive a novel metallicity calibration relation for the first time using high-redshift $T_e$-based metallicities: $\hat{R} = 0.18\log \mathrm{R2} + 0.98\log \mathrm{R3}$. This new calibration significantly reduces the scatter in high-redshift galaxies compared to the $\hat{R}$ relation previously calibrated for low-redshift galaxies.

*Unified Astronomy Thesaurus concepts:* High-redshift galaxies (734)

## 1. Introduction

Determining the gas-phase metallicities in galaxies is essential, as metallicity serves as a sensitive tracer of the physical mechanisms that regulate the baryon cycle. It reflects the complex interplay between gas inflows and outflows, star formation, and the subsequent enrichment of the interstellar medium (ISM; e.g., K. Finlator & R. Davé 2008; R. Davé et al. 2011; F. Matteucci 2012; S. J. Lilly et al. 2013; L. J. Kewley et al. 2019).

The mass–metallicity relation (MZR) was first identified by J. Lequeux et al. (1979), with subsequent studies using optical luminosity as a mass proxy confirming a correlation between blue luminosity and metallicity in galaxies (e.g., D. R. Garnett & G. A. Shields 1987; D. Zaritsky et al. 1994). Advances in stellar population synthesis models (G. Bruzual & S. Charlot 2003) enabled more precise stellar mass measurements, leading to the discovery of a robust correlation between stellar mass and gas-phase oxygen abundance in local star-forming galaxies (C. A. Tremonti et al. 2004). Observations indicate that the MZR evolves up to $z \sim 3.5$, with metallicities increasing at lower redshifts for a given stellar mass (R. Maiolino et al. 2008).

The most reliable method for measuring gas-phase metallicity is the "direct $T_e$ method," which determines the electron temperature ($T_e$) using collisionally excited auroral lines like [O III]$\lambda$4363 (M. Peimbert 1967; L. J. Kewley & M. A. Dopita 2002; F. Bresolin et al. 2009). The flux ratio of the weak auroral line [O III]$\lambda$4363 to the stronger nebular line [O III]$\lambda$5007 provides an accurate diagnostic for $T_e$. This method is effective because the electron temperature of the emitting gas is strongly anticorrelated with metallicity, as metal ions play a significant role in radiative cooling. Once $T_e$ is determined, gas-phase metallicity can be directly calculated from the flux ratios of standard emission lines (e.g., D. E. Osterbrock & G. J. Ferland 2006; T. T. Yuan & L. J. Kewley 2009; B. H. Andrews & P. Martini 2013; T. Jones et al. 2015; R. L. Sanders et al. 2016, 2020, 2025; E. Pérez-Montero 2017).

Strong-line metallicity calibrations are established by fitting the relationships between various strong optical nebular emission-line ratios and metallicities determined through the direct $T_e$ method. Extensive studies have been conducted to calibrate metallicity diagnostics using strong-line ratios in samples of H II regions within nearby star-forming galaxies. (M. Pettini & B. E. J. Pagel 2004; T. Nagao et al. 2006; R. A. Marino et al. 2013; J. S. Brown et al. 2016; M. Curti et al. 2017). However, studies aimed at deriving metallicity calibration relations for high-redshift star-forming galaxies ($z > 3$) remain sparse. (I. H. Laseter et al. 2024; R. L. Sanders et al. 2024). Some of the widely used strong-line ratios are

$$\mathrm{R3} = \frac{[\mathrm{O\,III}]\,\lambda 5007}{\mathrm{H}\beta}, \tag{1}$$

$$\mathrm{R2} = \frac{[\mathrm{O\,II}]\,\lambda\lambda 3727, 3729}{\mathrm{H}\beta}, \tag{2}$$

$$\mathrm{R23} = \frac{[\mathrm{O\,II}]\lambda\lambda3727, 3729 + [\mathrm{O\,III}]\lambda\lambda4959, 5007)}{\mathrm{H}\beta}, \quad (3)$$

$$\mathrm{O32} = \frac{[\mathrm{O\,III}]\lambda5007}{[\mathrm{O\,II}]\lambda\lambda3727, 3729}, \quad (4)$$

$$\mathrm{Ne3O2} = \frac{\mathrm{Ne\,III}\lambda3870}{[\mathrm{O\,II}]\lambda\lambda3727, 3729}, \quad (5)$$

$$\mathrm{O3N2} = \frac{\mathrm{O\,III}\lambda5007/H\beta}{[\mathrm{N\,II}]\lambda6585/H\alpha}. \quad (6)$$

The [O III]$\lambda$4363 line is a faint emission line, which restricted the use of the $T_e$ method for measuring metallicity in galaxies at $z > 3$ before the James Webb Space Telescope (JWST) era. Consequently, this method was primarily applied to low-redshift, metal-poor galaxies (e.g., Y. I. Izotov et al. 2006, 2019) or to stacked samples of thousands of high-metallicity galaxies (e.g., M. Curti et al. 2017). Auroral line measurements were scarce for galaxies at $z > 1$ (T. T. Yuan & L. J. Kewley 2009; L. Christensen et al. 2012b, 2012a; T. Kojima et al. 2017; F. Bian et al. 2018; T. Gburek et al. 2019) and even rarer beyond $z > 3$ (R. L. Sanders et al. 2016, 2024).

The advent of JWST has improved this situation by enabling more frequent detections of auroral lines at $z > 3$ (D. Schaerer et al. 2022; K. Nakajima et al. 2023; T. Morishita et al. 2024; R. L. Sanders et al. 2024; M. Stiavelli et al. 2025). Nonetheless, auroral line detections at $z > 3$ with JWST are still limited to around 25 galaxies, posing challenges for establishing robust high-redshift metallicity calibrations. R. L. Sanders et al. (2024) combined a sample of 25 galaxies at $z > 2$ observed with JWST and 21 galaxies within $1.4 < z < 3.7$ observed with ground-based spectroscopy to present the first high-redshift metallicity calibrator within $1.4 < z < 8.7$. Similarly, I. H. Laseter et al. (2024) used a sample from JADES, CEERS, and ERO, also comprising 25 galaxies within $3 < z < 10$, for metallicity calibration.

In this paper, we significantly expand the high-redshift galaxy sample by reporting the detection of novel [O III]$\lambda$4363 auroral emission lines in 42 galaxies within $3 < z < 10$, observed using JWST/NIRSpec medium-resolution gratings from the JADES and PRIMAL surveys. All of these detections are new. We use the dust-corrected [O III]$\lambda$4363 to [O III]$\lambda$5007 line-flux ratio to derive robust gas-phase oxygen abundances using the direct method. This new sample is combined with 25 galaxies in the same redshift range from previous JWST studies with [O III] $\lambda$4363 detections (e.g., R. L. Sanders et al. 2024; K. Nakajima et al. 2023), bringing the total to 67 galaxies within $3 < z < 10$. We use this expanded data set to derive empirical high-redshift metallicity calibrators, enabling robust measurement of gas-phase oxygen abundance from rest-frame optical strong emission-line ratios, as described in Equations (1), (2), (3), (4), (5), and (6). These metallicity calibration relations will facilitate future studies in measuring gas-phase metallicity across a wide redshift range of $3 < z < 10$ and over a metallicity range of $7.2 < 12 + \log(\mathrm{O/H}) < 8.4$. Our sample nearly triples the previously available sample size, providing a stronger foundation for metallicity diagnostics in high-redshift galaxies. The paper is organized as follows. Section 2 details the observed data reduction, analysis, and spectral fitting techniques. In Section 3, we describe the electron temperature and density measurements, followed by the determination of "direct" metallicities. Section 4 presents the high-redshift metallicity calibration relations. Section 5 discusses the implications of these metallicity measurements and introduces a novel metallicity calibration relation. Finally, Section 6 summarizes our findings.

Throughout this paper, we adopt the solar abundance table from M. Asplund et al. (2021), the AB magnitude system (J. B. Oke & J. E. Gunn 1983), and cosmological parameters from Planck Collaboration et al. (2020): the Hubble constant $H_0 = 67.4\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$, matter density parameter $\Omega_M = 0.315$, and dark energy density $\Omega_\Lambda = 0.685$.

## 2. Observations and Data Analysis

The data analyzed in this paper were obtained from multi-object spectroscopy observations using the Micro-Shutter Assembly of NIRSpec on JWST (P. Ferruit et al. 2022). Our analysis focuses on spectra acquired with the NIRSpec low-resolution PRISM/CLEAR configuration, which covers a spectral range of 0.6–5.3 $\mu$m, as well as with medium-resolution gratings ($R \sim 1000$), specifically G140M/F070LP (0.7–1.27 $\mu$m), G235M/F170LP (1.66–3.07 $\mu$m), and G395M/F290LP (2.87–5.10 $\mu$m). We used the following criteria to select galaxies for this study:

1. Each galaxy must have been observed with JWST/NIRSpec using the PRISM/CLEAR configuration and at minimum, medium-resolution gratings that cover the rest-frame wavelength range of 0.35–0.7 $\mu$m.
2. The strong emission lines must be resolved at a $>3\sigma$ significance level, and the [O III]$\lambda$4363 auroral line at a $\geqslant 2\sigma$ significance level.
3. Each galaxy must have a redshift >3, with a detection of the [O III]$\lambda$4363 auroral line that has not been previously reported.

We assembled a sample of galaxies 42 galaxies in the redshift range of $3 < z < 10$ that meet the above criteria, primarily from publicly available data releases, such as 32 galaxies from JADES Data Release 3 (DR3; e.g., B. E. Robertson 2022; A. J. Bunker et al. 2024; E. Curtis-Lake et al. 2023; D. J. Eisenstein et al. 2023; F. D'Eugenio et al. 2025) and 10 galaxies from the JWST-PRIMAL Legacy Survey (K. E. Heintz et al. 2025). Figure 1 (top left) shows the redshift histogram of our sample.

Each target was observed with three micro-shutters activated, utilizing a three-point nodding sequence along the slit to ensure comprehensive coverage and enhanced data quality. For each JADES GOODS-S and GOODS-N target, flux-calibrated 1D and 2D spectra were produced by the JADES team using a custom pipeline developed by the ESA NIRSpec Science Operations Team and Guaranteed Time Observations teams. For detailed descriptions of the data reduction steps, we refer readers to A. J. Bunker et al. (2024), M. Curti et al. (2024), and F. D'Eugenio et al. (2025). In this paper, we use the reduced and flux-calibrated *medium*-tier 1D and 2D spectra of hundreds of targets, publicly released as part of JADES DR3[5] (F. D'Eugenio et al. 2025).

For targets from the JWST-PRIMAL Legacy Survey, we used data from the DAWN JWST Archive (DJA), which includes reduced images, photometric catalogs, and spectroscopic data for publicly available JWST data products.[6,7] The DJA spectroscopic archive (DJA-Spec) includes observations

[5] https://jades-survey.github.io/scientists/data.html
[6] https://dawn-cph.github.io/dja
[7] https://s3.amazonaws.com/msaexp-nirspec/extractions/nirspec_graded_v2.html

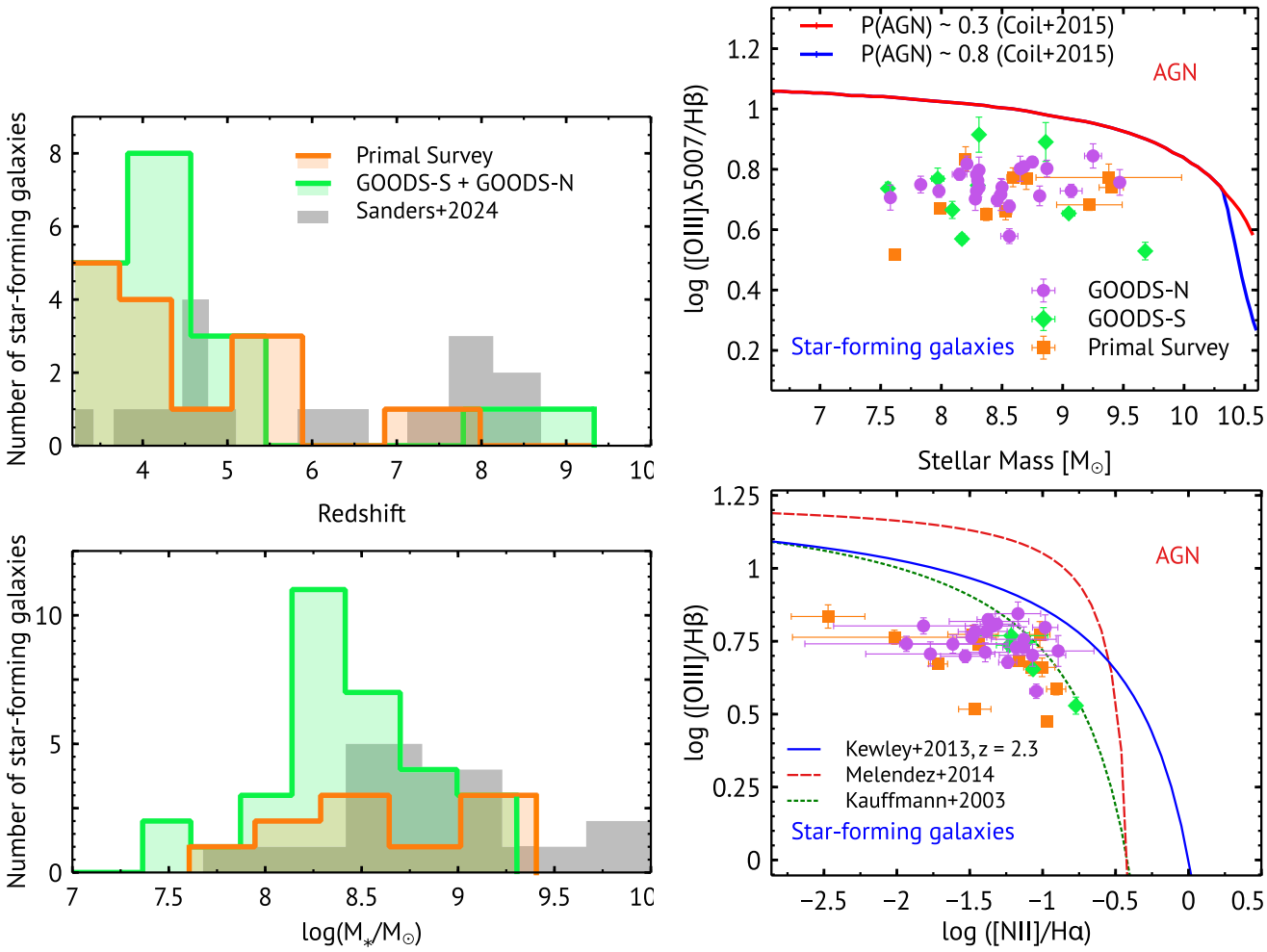


**Figure 1.** Top left: redshift distribution ($3 < z < 10$) of galaxies used in our study that exhibit the auroral [O III]$\lambda$4363 line. The green histogram represents galaxies from GOODS-N and GOODS-S DR3 (F. D'Eugenio et al. 2025), while the orange histogram shows galaxies from the PRIMAL survey (K. E. Heintz et al. 2025). The gray histogram captures JWST-detected galaxies from existing literature within $3 < z < 10$ that also display the [O III]$\lambda$4363 line (R. L. Sanders et al. 2024). Bottom left: the stellar mass distribution of galaxies used in our study is shown, using the same color coding as above. Top right: the MEx diagram for our complete sample, illustrating the relationship between log([O III]$\lambda$5007/H$\beta$) and the stellar mass. This diagnostic, based on X. He et al. (2024) and MEx curves from A. L. Coil et al. (2015), shows probable AGN regions with likelihoods of ∼0.3 (red) and ∼0.8 (blue). We find no evidence of significant AGN contamination in our sample. Bottom right: [O III]/H$\beta$ vs. [N II]/H$\alpha$ BPT diagram for our sample of galaxies. The green dotted curve represents the $z \sim 0$ demarcation line between star-forming galaxies and AGNs from G. Kauffmann et al. (2003). The blue solid and red dashed curves indicate the predicted upper limits for star-forming galaxies at $z \sim 2.3$, as proposed by L. J. Kewley et al. (2013) and the theoretical model from M. Meléndez et al. (2014), respectively.

from major programs, such as CEERS (S. L. Finkelstein et al. 2022), GLASS (T. Treu et al. 2022), JADES (A. J. Bunker et al. 2024), and UNCOVER (R. Bezanson et al. 2024). For further details on the data reduction processes, see K. E. Heintz et al. (2025). The Appendix lists all the high-redshift galaxies studied in this paper.

Additionally, we included 25 JWST-selected star-forming galaxies from the literature, spanning a redshift range of $3 < z < 10$ and a stellar mass range similar to that of our sample. These galaxies display [O III]$\lambda$4363 auroral lines and comprise 16 galaxies from R. L. Sanders et al. (2024) and nine galaxies from I. H. Laseter et al. (2024).

### 2.1. PRISM/CLEAR Spectra

To determine the stellar masses of the 42 newly detected auroral line galaxies, we fit their PRISM spectra using the spectral energy distribution (SED) fitting code `Bagpipes` (A. C. Carnall et al. 2018). `Bagpipes` generates detailed model galaxy spectra, fitting them to both photometric and spectroscopic data (F. Feroz & M. P. Hobson 2008), and provides posterior distributions of galaxy properties for each source in the sample. This versatile code can model galaxies with various star formation histories (SFHs), including delayed-$\tau$, constant, and burst scenarios (e.g., S. Lower et al. 2020; P. Chakraborty et al. 2024).

In this study, we used a constant star formation model (similar to Paper I; A. Sarkar et al. 2025), allowing star formation ages to vary between 0 and 2 Gyr. We adopted stellar population synthesis models based on the 2016 version of the BC03 models (G. Bruzual & S. Charlot 2003). These models assume an initial mass function (IMF) from P. Kroupa (2002) and incorporate nebular line and continuum emissions using `Cloudy` (G. J. Ferland et al. 2017; M. Chatzikos et al. 2023). The SED fitting was conducted over a broad parameter range, with stellar masses $\log(M_*/M_\odot)$ varying between 4 and 13 and stellar metallicities $\log(Z/Z_\odot)$ ranging from 0.005 to 2.5. `Bagpipes` assumes solar abundances from E. Anders & N. Grevesse (1989) and incorporates ISM depletion factors,

along with He and N scaling relations from M. A. Dopita et al. (2000).

The ionization parameter for nebular line and continuum emissions was varied between −4 and −1. We adopted the Calzetti dust attenuation curve (D. Calzetti et al. 2000), with an extinction parameter $A_V$ ranging from 0 to 4. To account for birth cloud dust attenuation, we introduced a multiplicative factor ($1 < \eta < 2$) in the dust model, addressing the increased dust attenuation typically observed around H II regions, which is usually double that of the general ISM during the galaxy's first 10 Myr (A. J. Bunker et al. 2024). To model this effect, we capped the maximum age of the birth cloud at 0.01 Gyr (A. J. Bunker et al. 2024). The resulting stellar masses and other galaxy properties are provided in the Appendix. Figure 1 (bottom left) shows the stellar mass histogram of our sample of galaxies.

### 2.2. Emission-line Flux Measurements

We used medium-resolution grating spectra to measure the line fluxes of several prominent nebular emission lines, when detected, including hydrogen Balmer lines, [O II]$\lambda\lambda$3727,3729, [Ne III]$\lambda$3867, [O III]$\lambda$4363, [O III]$\lambda\lambda$4959,5007, and [N II] $\lambda$6584. The continuum and line emissions were modeled simultaneously using the `MSAEXP` Python modules.[8] `MSAEXP` models the continuum as a linear combination of simple stellar population spectra, matching the spectral resolution of the observed spectrum. For continuum modeling, we used SFHZ templates[9] derived from the Flexible Stellar Population Synthesis (FSPS) models from `EAZY` (G. B. Brammer et al. 2008). Specifically, we adopted the CORR_SFHZ_13 subset of models, which features redshift-dependent SFHs that, at a given redshift, exclude SFHs that begin earlier than the age of the Universe. The maximum allowed attenuation is also constrained by each epoch. Additionally, we included the best-fit template for the JWST-observed extreme emission-line galaxy at $z = 8.5$ (ID4590) from A. C. Carnall et al. (2023), rescaled to match the normalization of the FSPS models, to account for potential emission lines with large equivalent widths. Emission lines were modeled using Gaussian profiles centered at each line's wavelength.

Next, we derived the dust-corrected flux of the key emission lines implementing (e.g., K. Nakajima et al. 2023; A. Sarkar et al. 2025)

$$L_{\rm int}(\lambda) = L_{\rm obs}(\lambda) 10^{0.4k(\lambda)E(B-V)}, \quad (7)$$

where, $L_{\rm int}(\lambda)$ and $L_{\rm obs}(\lambda)$ represent the intrinsic and observed fluxes, respectively, $k_\lambda$ denotes the extinction coefficient at wavelength $\lambda$, and the specific reddening curve $k_\lambda$ was adopted from D. Calzetti et al. (2000). We employed three different approaches to determine the dust-corrected flux:

1. For galaxies at $z < 6.75$ where both H$\alpha$ and H$\beta$ were detected with a signal-to-noise ratio (S/N) $\geqslant 3$, we estimated $E(B-V)$ using the Balmer decrement method. We assumed an intrinsic flux ratio of H$\alpha$/H$\beta = 2.86$ (D. E. Osterbrock & G. J. Ferland 2006) and applied the dust extinction curve from D. Calzetti et al. (2000).
2. For galaxies at $z \geqslant 6.75$, where H$\alpha$ is not observable due to the spectral coverage of NIRSpec, we instead used the H$\gamma$/H$\beta$ ratio, assuming an intrinsic value of H$\gamma$/H$\beta = 0.47$, which corresponds to a temperature of $10^4$ K for Case B recombination (D. E. Osterbrock & G. J. Ferland 2006).
3. If neither H$\gamma$ nor H$\alpha$ was detected, we estimated the nebular dust attenuation using SED fitting on the PRISM spectra, performed with the `Bagpipes` code. This code incorporates a two-component dust attenuation model that accounts for both nebular and stellar emission (see Section 2.1).

To assess the quality of our measurements, we calculated the S/N for each emission line and established a minimum S/N threshold of $\geqslant 3\sigma$ for including an emission line in our metallicity calculations. Specifically, our selection criteria required galaxies to exhibit detectable H$\beta$ and [O III], along with at least one other emission line—[O II]$\lambda\lambda$3727,3729, [N II]$\lambda$6584, or [S II]$\lambda\lambda$6716,31—measured at or above the $3\sigma$ confidence level. This rigorous approach enables robust measurement and validation of emission-line fluxes across our sample, ensuring that only reliable data points are used to derive the gas-phase metallicities of high-redshift galaxies.

### 2.3. Active Galactic Nucleus Contamination: Mass-excitation Diagram and Baldwin–Phillips–Telervich Diagram

In this study, we apply line-flux ratio diagnostics specifically designed to measure gas-phase metallicity in star-forming regions of galaxies. However, active galactic nucleus (AGN)-driven ionization can introduce inaccuracies in these metallicity calibrations by contaminating the emission-line fluxes. To ensure precise metallicity measurements, we carefully examine each galaxy for possible AGN influence. We use three methods to systematically exclude AGN-contaminated sources from our sample, thereby strengthening the reliability of our results.

First, we use the Baldwin–Phillips–Telervich (BPT) diagram to differentiate star-forming galaxies from AGNs, employing the [O III]/H$\beta$ versus [N II]/H$\alpha$ ratios for our target sample, as shown in Figure 1 (bottom right). To identify potential AGNs, we apply three demarcation lines from the literature. We adopt the empirical demarcation line from G. Kauffmann et al. (2003), which distinguishes local star-forming galaxies from AGNs. We then compare our sample with the upper limit for star-forming galaxies at $z \sim 2.3$ proposed by L. J. Kewley et al. (2013) and the theoretical classification scheme from M. Meléndez et al. (2014). As shown in Figure 1, four galaxies in our sample lie above the demarcation line from G. Kauffmann et al. (2003). However, all galaxies remain below the more recent demarcation lines from L. J. Kewley et al. (2013) and M. Meléndez et al. (2014). Therefore, we do not exclude any galaxies from our sample.

Second, we recheck our sample by employing the mass-excitation (MEx) diagnostic diagrams, initially introduced by S. Juneau et al. (2014) and further refined by A. L. Coil et al. (2015), which use the [O III] $\lambda$5007/H$\beta$ (R3) emission-line ratio in conjunction with stellar mass to distinguish AGNs from star-forming galaxies. This diagram provides an alternative to the commonly used BPT diagram (e.g., J. A. Baldwin et al. 1981; L. J. Kewley et al. 2013), which compares the [O III] $\lambda$5007/H$\beta$ and [N II]$\lambda$6584/H$\alpha$ emission-line ratios but can be limited when [N II] or H$\alpha$ lines fall outside the spectral coverage or are blended. As our measurements rely exclusively on emission-line fluxes from medium-resolution grating

[8] https://github.com/gbrammer/msaexp
[9] https://github.com/gbrammer/eazy-photoz/tree/master/templates/sfhz

spectra, we use dust-corrected [O III]$\lambda$5007 and H$\beta$ fluxes for the MEx diagram. Figure 1 (right) displays our JWST samples in the $\log([\mathrm{O\,III}]\lambda5007/\mathrm{H}\beta)$ versus $\log(M_*/M_\odot)$ plane.

In Figure 1 (right), the blue and red curves mark steep gradients at P(AGN) $\sim$ 0.3 and P(AGN) $\sim$ 0.8, respectively, representing the probability that a galaxy hosts an AGN, as established by A. L. Coil et al. (2015) for $z = 2.3$ galaxies and AGNs in the MOSDEF survey (X. He et al. 2024). The positions of our sources on the MEx diagram indicate that our sample is predominantly composed of star-forming galaxies, lying below or near the boundary line. Consequently, no galaxies were excluded due to potential AGN contamination, and we retain all sources in our sample. A similar method was also used by X. He et al. (2024) to distinguish star-forming galaxies from AGNs in the GLASS-JWST sample.

Additionally, we visually inspected each spectrum for signs of broad emission-line regions, following procedures from A. Sarkar et al. (2021).

### 2.4. Samples with an [O III]$\lambda$4363 Auroral Line

In this study, we identified star-forming galaxies displaying an [O III]$\lambda$4363 auroral line with an S/N above 2. We utilized publicly available, reduced, and calibrated 1D grating spectra from the JADES and PRIMAL surveys to assemble a sample of 42 galaxies with significant [O III]$\lambda$4363 emission lines. Figure 2 illustrates a zoomed-in view of the observed [O III] $\lambda$4363 lines with the best-fit continuum and emission-line models. Furthermore, we visually inspected both the 1D and 2D spectra for each galaxy to verify the absence of single-pixel noise artifacts at the location of the [O III]$\lambda$4363 line.

In this study, we broadened our sample by combining 42 newly identified galaxies with auroral line emissions with 25 additional galaxies from existing literature, all exhibiting significant [O III]$\lambda$4363 line fluxes ($\mathrm{S/N} \geqslant 2$). This supplementary group includes 16 galaxies from R. L. Sanders et al. (2024), four galaxies from the SMACS 0723 galaxy cluster field within the ERO program, and five galaxies from the GLASS ERS program (T. Treu et al. 2022).

To estimate the $T_e$-based metallicity for these 25 galaxies, we used the emission-line fluxes reported in the respective studies. Specifically, for galaxies selected from the ERO and GLASS ERS programs, we referred to line fluxes provided in M. Curti et al. (2024 for ERO IDs 4590, 6355, and 10612, K. Nakajima et al. (2023) for ERO IDs 5144 and GLASS IDs 100003, 10021, 150029, and 160133, and T. Jones et al. (2023) for GLASS ID 150008. This results in a robust sample of 67 star-forming galaxies with detected auroral lines, representing a sample size that is at least 2.6 times larger than that used in previous direct $T_e$-based metallicity studies of high-redshift JWST galaxies.

## 3. Direct Metallcity Measurements

We derived electron temperatures and used them to estimate oxygen abundances, with oxygen serving as a proxy for total gas-phase metallicity. This method assumes proportional scaling of all elements and models each galaxy as a single H II region, divided into a high-ionization zone traced by $O^{2+}$ and a low-ionization zone traced by $O^+$. However, this simplified approach does not account for the complexities of temperature distributions and ionization structures within these regions. While a full exploration of these assumptions and their limitations is beyond the scope of this study, we recommend G. Stasińska (2002), R. Maiolino & F. Mannucci (2019), and A. J. Cameron et al. (2023) for further details on these topics.

### 3.1. Electron Temperatures: $T_e$(O II) versus $T_e$(O III)

The thermal structures of H II regions can be characterized by measuring the fluxes of their emission lines. The direct method requires the detection of at least one auroral emission line per ion to determine the temperatures of different ionization zones. Ideally, a complete ionization structure of an H II region would be necessary to measure electron temperatures and densities across all zones; however, this is impractical, so simplified models are commonly applied. Most studies adopt either a two-zone or three-zone model to approximate the H II regions that produce the observed emission lines in galaxies. Here, we use a two-zone model for H II regions: (1) an inner high-ionization zone, characterized by $T_e$([O III]), and (2) an outer low-ionization zone, represented by $T_e$([O II]) (D. R. Garnett 1992).

The $T_e$([O III]) can be derived from the emission-line ratio of [O III]$\lambda$5007 to [O III]$\lambda$4363 (e.g., D. E. Osterbrock & G. J. Ferland 2006; E. Pérez-Montero 2017). We used `PyNeb` (V. Luridiana et al. 2015; C. Morisset et al. 2020) to fit the relationship between this line ratio and electron temperature, assuming a five-level atom model and adopting collisional strengths from P. J. Storey et al. (2014). Since $T_e$([O III]) has only a weak dependence on electron density ($n_e$), we assume $n_e = 300\,\mathrm{cm}^{-3}$ for simplicity in the high-ionization zones. Testing this with $n_e = 1000\,\mathrm{cm}^{-3}$ yielded only a minor temperature reduction of 0.1% (R. L. Sanders et al. 2020). For our sample of 42 galaxies, the derived $T_e$([O III]) values range from approximately 12,000 K to 24,000 K (see the Appendix). Temperature uncertainties were estimated using the affine invariant Markov Chain Monte Carlo (MCMC) sampler in the `emcee` package (D. Foreman-Mackey et al. 2013). Our $T_e$([O III]) measurements are consistent with those of extremely metal-poor local galaxies, as well as with high-redshift galaxies at $z \sim 6$–8 (e.g., K. Z. Arellano-Córdova et al. 2022; D. Schaerer et al. 2022; M. Curti et al. 2024) and galaxies from CEERS at $z \sim 2$–9 (R. L. Sanders et al. 2024).

The $T_e$([O II]) can be estimated from the emission-line ratio of [O II]$\lambda\lambda$3726, 3729 to [O II]$\lambda\lambda$7322, 7332 (D. E. Osterbrock & G. J. Ferland 2006). This calculation is particularly challenging due to (1) the strong dependence of the [O II] line ratio on electron density, and (2) the limited JWST/NIRSpec spectral coverage of the [O II]$\lambda\lambda$7322, 7332 auroral lines for galaxies at $z \geqslant 5.75$. Within our sample of 42 galaxies, we detect [O II]$\lambda\lambda$7322, 7332 auroral lines with an $\mathrm{S/N} \geqslant 2$ in only 10 galaxies. Figure 3 presents the [O II]$\lambda\lambda$7322, 7332 emission lines along with their best-fit continuum + line-emission models.

To estimate $T_e$([O II]) in the 10 galaxies, we first derived $n_e$ from the [S II]$\lambda\lambda$6716, 6731 doublet line ratios (detected with an $\mathrm{S/N} \geqslant 3$) using `PyNeb`, assuming an electron temperature of $10^4$ K. This step is minimally affected by temperature variations, as the resulting $T_e$([O II]) remains largely unchanged even if the electron temperature varies by a factor of 2–3. The measured $n_e$ spans from approximately 50 $\mathrm{cm}^{-3}$ to 1862 $\mathrm{cm}^{-3}$ across these galaxies, with a median of $567 \pm 170\,\mathrm{cm}^{-3}$. We then derived $T_e$([O II]) from the [O II]$\lambda\lambda$3726, 3729 to [O II] $\lambda\lambda$7322, 7332 line ratios, using the measured $n_e$ values in `PyNeb` and adopting $O^+$ collision strengths from R. Kisielius et al. (2009). The resulting $T_e$([O II]) values range from

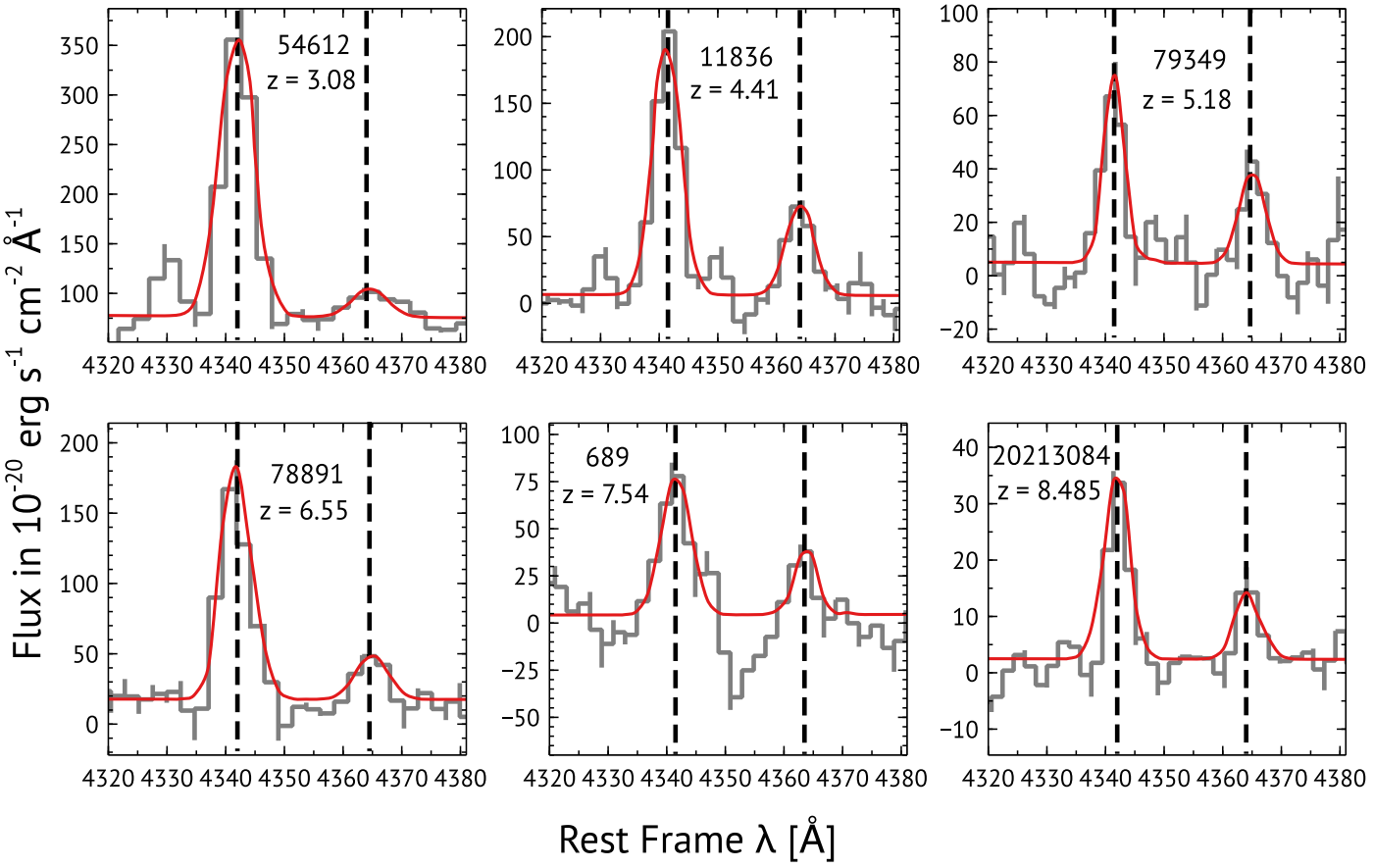


**Figure 2.** The 1D spectra of six galaxies in our sample illustrate the detected [O III]$\lambda$4363 and H$\gamma$ emission lines, representing galaxies across various redshifts. The $x$-axes are converted to rest-frame wavelengths using the best-fit redshifts. The red curve shows the best-fit continuum along with the [O III]$\lambda$4363 and H$\gamma$ emission-line profiles. The black dashed line marks the rest-frame wavelengths of [O III]$\lambda$4363 and H$\gamma$.

10,830 K to 20,000 K across the subsample. Figure 4 displays $T_e$([O II]) as a function of $T_e$([O III]) for these galaxies. We find the best-fit relation of

$$T_e(\mathrm{O\,II}) = (0.58 \pm 0.19) \times T_e(\mathrm{O\,III}) + (4520 \pm 2000)\mathrm{K}. \quad (8)$$

Our best-fit relation aligns well with the widely used $T_e$([O II]) versus $T_e$([O III]) relation for local galaxies from A. Campbell et al. (1986). It also shows consistency with the relation by B. E. J. Pagel et al. (1992) for $T_e$([O III]) $\leqslant$ 23,000 K and with Y. I. Izotov et al. (2006) for low metallicities within $T_e$([O III]) $\leqslant$ 21,000 K.

As shown in Figure 4 and discussed by R. L. Sanders et al. (2024), there is considerable uncertainty in the $T_e$([O II]) versus $T_e$([O III]) relation, even at low redshift ($z = 0$), and this relationship displays a substantial intrinsic scatter.

To assess the robustness of our metallicity measurements, we have adopted three different relations between $T_e$([O II]) and $T_e$([O III]): (1) Equation (8), (2) $T_e$([O II]) = 0.7 $\times$ $T_e$([O III]) + 3000 K (see A. Campbell et al. 1986), (3) $T_e$([O II]) = $T_e$([O III]). These variations did not lead to significant changes in the metallicity estimates, suggesting that the oxygen abundance is primarily governed by the high-ionization region.

### 3.2. Oxygen Abundances

We measured ionic oxygen abundances using `PyNeb`, applying collision strengths for $O^{2+}$ from P. J. Storey et al. (2014) and for $O^{+}$ from R. Kisielius et al. (2009). The electron density was set to the median $n_e$ value of 567 cm$^{-3}$, as determined for 10 galaxies in our sample. This density value aligns with typical $n_e$ values observed in galaxies at redshifts $z = 2$–3 (R. L. Sanders et al. 2016). Since oxygen abundances in the ISM are predominantly influenced by $O^{2+}$ ions, the effect of variations in $n_e$ on abundance measurements is minimal, as $T_e$([O III]) has only a weak dependence on $n_e$ (e.g., Y. I. Izotov et al. 2006; B. H. Andrews & P. Martini 2013; M. Curti et al. 2017; R. L. Sanders et al. 2020). Our choice of density is consistent with studies by R. L. Sanders et al. (2024) and I. H. Laseter et al. (2024), who also fixed $n_e$ values in their $T_e$([O III]) and metallicity calculations. We assumed that within the H II regions, all oxygen exists in either the $O^{2+}$ or $O^{+}$ states, allowing us to simplify the calculation of the total oxygen abundance (O/H):

$$\frac{\mathrm{O}}{\mathrm{H}} = \frac{\mathrm{O}^{2+}}{\mathrm{H}} + \frac{\mathrm{O}^{+}}{\mathrm{H}}. \quad (9)$$

Due to the close ionization potentials of He II (~54.4 eV) and $O^{3+}$ (~55 eV), we did not apply an ionization correction factor (ICF) for $O^{3+}$, as its inclusion would have only a minimal effect on the total oxygen abundance (e.g., Y. I. Izotov et al. 2006; B. H. Andrews & P. Martini 2013; D. A. Berg et al. 2021; M. Curti et al. 2017, 2023). D. A. Berg et al. (2018) found an ICF of 1.055 for the unseen $O^{3+}$ ion. The contribution from $O^{3+}$ is thus negligible (0.2–0.5% in 12+log(O/H)) relative to the uncertainties (~2%–3%) in our total oxygen measurements (D. A. Berg et al. 2021; R. L. Sanders et al. 2024).

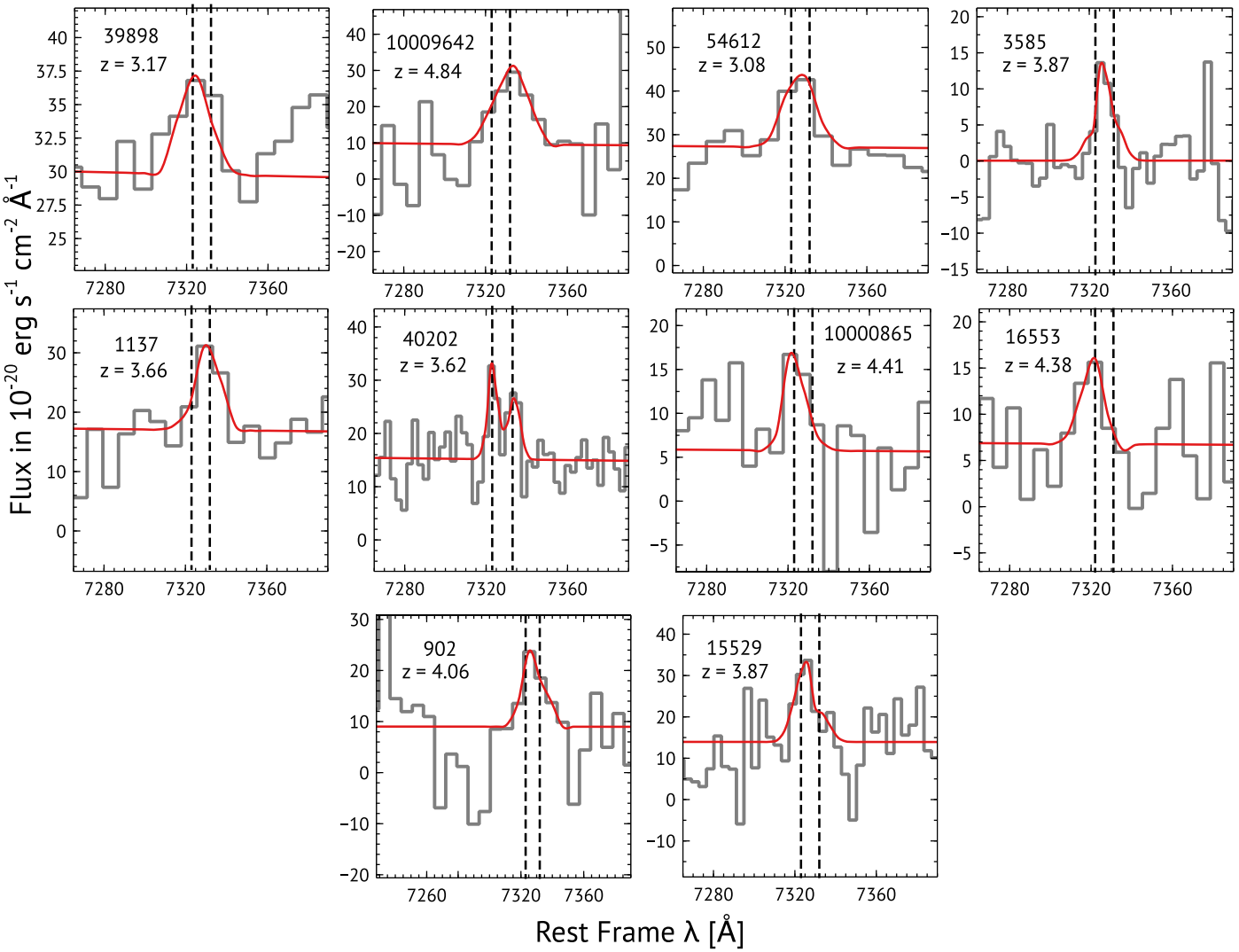


**Figure 3.** The 1D spectra of 10 galaxies in our sample show the detected [O II]$\lambda\lambda$7322, 7332 doublet. The $x$-axes are converted to rest-frame wavelengths using the best-fit redshifts. The gray represents the observed data, while the red curve displays the best-fit continuum along with the emission-line profiles for [O II]$\lambda$7322 and [O II]$\lambda$7332. Vertical black dashed lines indicate the rest-frame wavelengths of these lines.

The $O^{2+}/H$ ratio was calculated from the observed [O III] $\lambda$5008/H$\beta$ ratio using $T_e$([O III]), while $O^{+}/H$ was derived from the dust-corrected [O III]$\lambda\lambda$3726, 29/H$\beta$ ratio, applying $T_e$([O II]) when directly measurable or estimated via Equation (8) otherwise. Uncertainties were assessed using the affine invariant MCMC sampler implemented in the `emcee` package (D. Foreman-Mackey et al. 2013), performing 10,000 realizations based on normal distributions of the measured line fluxes for [O III]$\lambda\lambda$5007, 4959, 4363, H$\beta$, and [O III] $\lambda\lambda$3726, 29 with their associated 1$\sigma$ uncertainties. The metallicities for our sample of galaxies range from $12 + \log(O/H) = 7.2$ to 8.4, suggesting that these high-redshift galaxies are relatively metal-poor, consistent with previous findings from JWST observations. The metallicities and their 1$\sigma$ uncertainties are listed in Tables 3 and 4.

## 4. Metallicity Diagnostic for Strong Lines at High Redshift

### 4.1. Gas-phase Metallicity

We utilize a sample of 67 galaxies within the redshift range of $3 < z < 10$ (42 new galaxies and 25 from the literature) to determine $T_e$-based metallicities. This compiled data set significantly expands the current sample of high-redshift galaxies, incorporating 2.6 times the number used in previous JWST studies for calibrating $T_e$-based metallicities and empirical diagnostics (e.g., R. L. Sanders et al. 2024) within $3 < z < 10$. In Figure 5, we display the relationships between key strong-line metallicity indicators (outlined in Section 1) and the gas-phase oxygen abundance for the entire data set based on the direct $T_e$ method. To estimate the average strong-line ratios, we binned the data set into four bins: 7.2–7.5, 7.5–7.8, 7.8–8.1, and $>$ 8.1, as shown with large filled red stars in Figure 5.

To establish our new calibrations, we adopt a polynomial fitting approach, characterized by the following functional form:

$$\log(R) = \sum_{n=0}^{N} c_n\, x^n, \tag{10}$$

where $R$ represents the line ratio under consideration, and $x$ is defined as $x = 12 + \log(O/H) - 8.0$ (C. Allende Prieto et al. 2001). The $c_n$'s are the fit coefficients. Figure 5 shows the best-

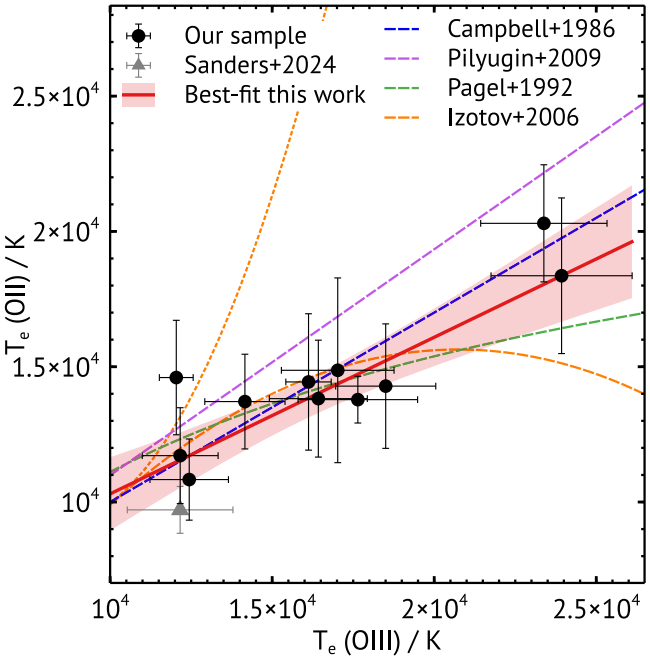


**Figure 4.** The relation between the electron temperatures, $T_e$([O II]) and $T_e$([O III]), is shown for the 10 galaxies in our sample with detected [O II] $\lambda\lambda$7322, 7332 auroral lines. Black filled circles represent our sample, while the filled gray triangle indicates the one galaxy from R. L. Sanders et al. (2024) at $z = 3.302$. The red curve shows the best-fit linear regression line with $1\sigma$ uncertainties. Blue, violet, and green dashed lines display $T_e$([O II]) vs. $T_e$([O III]) relations for local galaxies from A. Campbell et al. (1986), L. S. Pilyugin et al. (2009), and B. E. J. Pagel et al. (1992). The orange dashed and dotted curves, adapted from Y. I. Izotov et al. (2006), represent metallicities $12 + \log(\mathrm{O/H}) = 7.2$ and 8.2, respectively.

fit polynomials with $1\sigma$ uncertainties alongside the observed line ratios. The best-fit coefficients are listed in Table 1.

The R23 and R3 parameters are widely utilized diagnostics for deriving metallicity (e.g., B. E. J. Pagel et al. 1979; M. G. Edmunds & B. E. J. Pagel 1984; S. Charlot & M. Longhetti 2001; L. S. Pilyugin 2001; K. Nakajima et al. 2023). R23 ratio is long-known to have two metallicity solutions, when metallicity approaches $12+\log(\mathrm{O/H}) \geqslant 8.0$ (e.g, B. E. J. Pagel et al. 1979; L. J. Kewley & M. A. Dopita 2002). Studies focusing on the local universe ($z \sim 0$) and local analogs (e.g., M. Curti et al. 2017; F. Bian et al. 2018; M. Curti et al. 2020; K. Nakajima et al. 2022), as well as prior high-redshift ($z > 3$) investigations, albeit with limited sample sizes (I. H. Laseter et al. 2024; R. L. Sanders et al. 2024), show that R3 parameters can also yield two distinct metallicity solutions: one corresponding to a low-metallicity regime and the other to a high-metallicity regime, with a turnover of around $12 + \log(\mathrm{O/H}) \sim 8.0$. Our analysis of galaxies within the redshift range of $3 < z < 10$ also exhibits a similar pattern, consistent with earlier studies, where R23 and R3 show a decline at lower metallicities, particularly at $12 + \log(\mathrm{O/H}) < 7.7$ (K. Nakajima et al. 2022).

R2 is typically not utilized as a standalone metric for measuring metallicity since it is primarily sensitive to the ionization parameter in addition to metallicity (e.g., S. Charlot et al. 2002; L. J. Kewley et al. 2004, 2019). R2 is used as a supplementary tool to resolve degeneracies associated with the R23 and R3 calibrations. For R2, the turnover manifests at $12 + \log(\mathrm{O/H}) \sim 8.5$ in the low-redshift universe (M. Curti et al. 2020; K. Nakajima et al. 2022). Given that our sample is constrained to the metallicity range of $12 + \log(\mathrm{O/H}) < 8.4$, we detect an increasing trend in R2 with metallicity.

O32 is an indirect metallicity indicator that can show considerable scatter, affecting its precision, as seen in Figure 5. However, it is useful for distinguishing between metallicity solutions from diagnostics like R23 (e.g., L. J. Kewley & M. A. Dopita 2002; T. Nagao et al. 2006; R. Maiolino & F. Mannucci 2019; K. Nakajima et al. 2023). We find a negative trend between the O32 ratio and metallicity, showing no signs of reaching a plateau, with a Spearman coefficient of $\rho_s = -0.47$ and a $p$-value of 0.001, indicating a weak correlation.

Ne3O2 index can be a good metallicity indicator in the H II regions due to its monotonic trend with metallicity and minimal sensitivity to reddening (F. Shi et al. 2007). For Ne3O2, the Spearman correlation yields a coefficient of $\rho_s = -0.63$ with a $p$-value of $7.4 \times 10^{-7}$, indicating a statistically significant negative correlation, which can be used to break the degeneracy between two metallicity solutions from R23 and R3 diagnostics. We present, for the first time, diagnostic tools for the O3N2 line ratio tailored to high-redshift galaxies. Figure 5 displays the O3N2 line ratio for our sample of galaxies. Our best-fit curve shows an initial increase in line ratio with metallicity, up to $12 + \log(\mathrm{O/H}) \sim 7.9$, followed by a decline, resulting in doubly degenerate metallicity values.

### 4.2. Redshift Evolution of Strong-line Ratio Metallicity Indicators

In this section, we explore the potential redshift evolution of strong-line ratio metallicity indicators within $3 < z < 10$. We only use the R23 line ratio for this purpose, as they demonstrate the least scatter in the data (see Figure 5) To facilitate this analysis, we divided our entire sample of galaxies into two subsamples: one consisting of 36 galaxies with redshifts between $z = 3$ and $z = 5$, and the other comprising 23 galaxies with redshifts between $z = 5$ and $z = 10$. Figure 6 illustrates the redshift evolution for the ranges $z = 3$–5 and $z = 5$–10. The black dashed line represents the best-fit polynomial for the subsample within $3 < z < 5$, while the black dotted line represents the best-fit polynomial for the subsample with $5 < z < 10$. Figure 6 also shows the modeled $z > 4$ relation for IllustrisTNG100 galaxies from M. Hirschmann et al. (2023), shown as the thick red line.

For R23, the subsample with $3 < z < 5$ shows metallicity differences of less than 0.02 dex compared to the calibration in the $5 < z < 10$ range, indicating a negligible variation. This aligns with both simulations and previous studies. M. Hirschmann et al. (2023) used the TNG100 galaxy population to study metallicity calibrations up to redshift 8 and observed minimal variation between $z = 4$ and $z = 8$ in R23. Similarly, P. Garg et al. (2024) found no significant variation in R23 calibration between $z = 3$ and $z = 5$ using mock emission-line data from the SIMBA simulations. R. L. Sanders et al. (2024) examined the residuals around the best-fit R23 calibrations as a function of redshift and did not find significant trends, implying limited redshift evolution up to $z \sim 8$. While they highlighted the potential benefits of additional $T_e$ measurements for refining the calibration across this range, our expanded sample to approximately three times larger within $3 < z < 10$, allows us to divide the data into two redshift bins. This larger sample size provides further evidence supporting the lack of significant redshift evolution in R23.

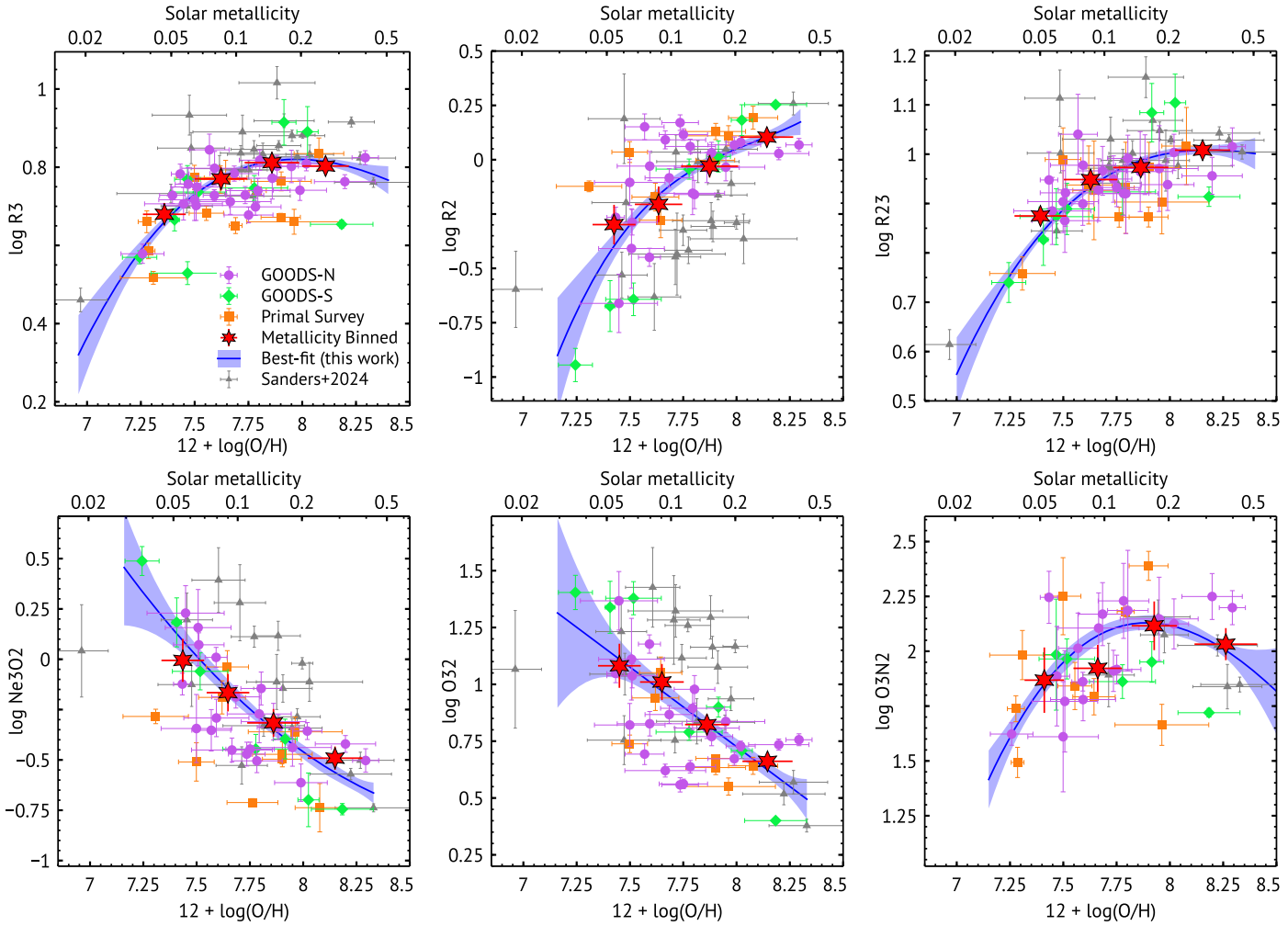


**Figure 5.** The relationship between $T_e$-based metallicity and strong-line ratios (R3, R2, R23, O32, Ne3O2, and O3N2) is shown for a sample of 67 individual galaxies within the redshift range of $3 < z < 10$. Purple and green data points represent galaxies from the GOODS-N and GOODS-S fields, respectively, based on the JADES DR3 (F. D'Eugenio et al. 2025). Orange data points correspond to galaxies from the PRIMAL survey (K. E. Heintz et al. 2025), while gray data points indicate JWST-detected galaxies compiled from the literature (R. L. Sanders et al. 2024). Large red data points denote the metallicity-averaged line ratios. Blue curves represent our best-fit polynomials with $1\sigma$ uncertainties for each diagnostic.

**Table 1**
Best-fit Values of the Metallicity Calibrators Shown in Figure 5

| Diagnostic | $N_{gal}$ | $c_0$ | $c_1$ | $c_2$ | $c_3$ | rms | $\sigma$ |
|---|---|---|---|---|---|---|---|
| R3 | 65 | 0.819 | −0.022 | −0.334 | 0.143 | 0.07 | 0.01 |
| R2 | 59 | 0.049 | 0.41 | −0.20 | 0.80 | 0.17 | 0.02 |
| R23 | 59 | 1.0 | 0.074 | −0.245 | 0.128 | 0.067 | 0.004 |
| O32 | 55 | 0.726 | −0.70 | ⋯ | ⋯ | 0.211 | 0.023 |
| Ne3O2 | 54 | −0.46 | −0.75 | 0.4 | ⋯ | 0.194 | 0.020 |
| O3N2 | 38 | 2.12 | −0.22 | −0.94 | 0.33 | 0.21 | 0.04 |
| $\hat{R}$ | 59 | 0.796 | 0.082 | −0.216 | 0.472 | 0.06 | 0.018 |

## 5. Discussion

### 5.1. MZR: *Direct* $T_e$ *Based*

Characterizing the scaling relation between stellar mass and gas-phase metallicity in galaxies, known as the MZR (e.g., J. Lequeux et al. 1979; C. A. Tremonti et al. 2004; H. Lee et al. 2006; A. M. Garcia et al. 2025; A. Sarkar et al. 2025), is essential for understanding the processes that regulate the growth of early galaxies. This relation reflects the interplay between gas accretion, star formation, metal enrichment, and outflows that shape the baryon cycle. Here, we present, for the first time, the MZR for galaxies in the redshift range of $3 < z < 10$, with metallicities derived from the direct $T_e$ method and stellar masses obtained from continuum fitting, as detailed in Sections 2.1 and 3. We rescaled our stellar mass measurements to a common G. Chabrier (2003) IMF. Figure 7 shows the MZR for our complete sample of 67 galaxies, including 25 from the literature.

To compare the slope of the our MZR relation with the previous studies, we fit the measured average MZR with the relation, as described in several previous studies (e.g., R. L. Sanders et al. 2021; K. Nakajima et al. 2023;

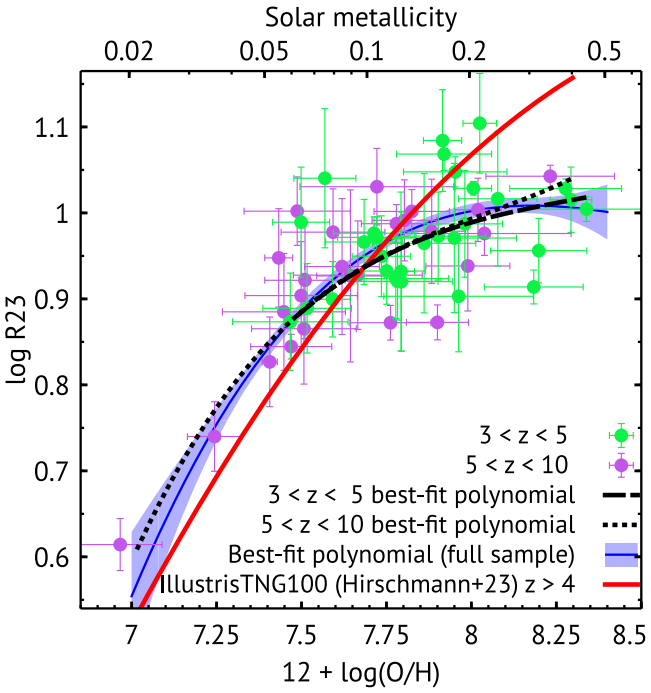


**Figure 6.** The R23 line ratio vs. metallicity is shown for two subsamples: galaxies within $3 < z < 5$ (filled green circles) and galaxies within $5 < z < 10$ (filled purple circles). The black dashed curve represents the best-fit metallicity calibration for the $3 < z < 5$ subsample, while the black dotted curve shows the best-fit calibration for the $5 < z < 10$ subsample. The blue curve represents the overall metallicity calibration for the full sample, covering $3 < z < 10$, with $1\sigma$ uncertainty. The red solid curve displays the R23 vs. metallicity relationship derived from $z > 4$ IllustrisTNG100 galaxies, adapted from M. Hirschmann et al. (2023).

M. Curti et al. 2024c; A. Sarkar et al. 2025),

$$12 + \log(\mathrm{O/H}) = \gamma \times \log\left(\frac{M_\star}{10^{10} M_\odot}\right) + Z_{10}, \qquad (11)$$

where the slope $\gamma$ and offset $Z_{10}$ (the gas-phase metallicity at a stellar mass of $10^{10} M_\odot$) are derived by fitting to the observed MZR. To estimate the average MZR, we divided our sample into three stellar mass bins: $\log(M_*/M_\odot) = 7.5$–$8.25$, $8.25$–$9$, and $> 9$, ensuring that each bin contains at least 15 galaxies. We determined a best-fit slope of $\gamma = 0.211 \pm 0.120$ and a metallicity intercept of $Z_{10} = 7.986 \pm 0.205$. The average points and the best-fit regression line, along with their $1\sigma$ uncertainties, are shown in Figure 7.

This study presents the first MZR at high redshift derived using the $T_e$-based method with JWST. Prior studies such as JADES + Primal (A. Sarkar et al. 2025), CEERS (K. Nakajima et al. 2023), and JADES (M. Curti et al. 2024) have employed calibration-based techniques to estimate metallicities. Table 2 provides a comparison of our best-fit slope ($\gamma$) and metallicity intercept ($Z_{10}$) with those from previous studies. While the slope and normalization for our sample are consistent within $1\sigma$ uncertainty with other high-redshift studies that used local calibrations, we find that at a stellar mass of $M_* = 10^8\, M_\odot$, our best-fit curve lies approximately 0.2 dex lower than those derived using calibration-based methods. The methodological distinction between $T_e$-based and calibration-based approaches may account for the systematic offset observed in metallicity measurements between our work and these studies.

We also compared our results with previous MZR studies conducted at lower redshifts using the $T_e$-based method. The metallicity at a given stellar mass in our sample is noticeably lower than the metallicity curves at $z \sim 0$ from B. H. Andrews & P. Martini (2013) and $z \sim 0.08$ from M. Curti et al. (2020). This difference varies with stellar mass, showing an offset of about 0.6 dex at $10^8\, M_\odot$ and 0.7 dex at $10^9\, M_\odot$, consistent with previous high-redshift galaxy studies (e.g., K. Nakajima et al. 2023; M. Curti et al. 2024; A. Sarkar et al. 2025). Additionally, we find that our direct $T_e$-based metallicity is approximately 0.3 dex lower than the $z \sim 3.3$ MZR curve reported by R. L. Sanders et al. (2021) for galaxies with stellar masses of $10^9\, M_\odot$.

Recent JWST studies reveal considerable intrinsic scatter in the properties of high-redshift galaxies. However, performing a statistically robust assessment of this scatter within scaling relations remains challenging due to the limited size of our sample. We determine the intrinsic scatter following the methodology outlined in A. Sarkar et al. (2025):

$$\sigma_{\mathrm{scatter}} = \sqrt{\sigma^2_{\mathrm{obs}} - \sigma^2_{\mathrm{measured}}}, \qquad (12)$$

where $\sigma_{\mathrm{obs}}$ represents the observed scatter in the full sample around the best-fit MZR line, and $\sigma_{\mathrm{measured}}$ indicates the average uncertainty in the metallicity measurement. We estimate $\sigma_{\mathrm{scatter}}$ for our entire galaxy sample to be approximately 0.09 dex. This value aligns with the intrinsic scatter of 0.08 dex reported for low-stellar-mass galaxies ($\log M_\star/M_\odot \lesssim 9$) at redshifts $3 < z < 10$ (M. Curti et al. 2024). However, it is lower than the intrinsic scatter of 0.16 dex observed in JADES + Primal survey galaxies (A. Sarkar et al. 2025) and the 0.16–0.18 dex range found in dwarf galaxies at redshifts $z = 2$–$3$ with stellar masses between $10^8$ and $10^9\, M_\odot$ (M. Li et al. 2023).

### 5.2. Metallicity Calibrations at z ∼ 0

Figure 8 presents a comparison between our high-redshift strong-line calibrators and those derived from local metallicity calibrators at $z \sim 0$ (e.g., M. Curti et al. 2020; R. L. Sanders et al. 2021; K. Nakajima et al. 2022). The R3 and R23 diagnostics exhibit consistently larger line ratios at a given metallicity compared to local calibrations. Both R3 and R23 show a turnover around $12 + \log(\mathrm{O/H}) \sim 8.0$, which is consistent with the turnover observed in the local R3 and R23 diagnostics. For R3, the high-redshift calibrations either underpredict or overpredict metallicity by $\sim$0.2–0.4 dex compared to the $z \sim 0$ calibrators, depending on whether they lie on the left or right side of the turnover. In a similar manner, the high-redshift calibrations for R23 exhibit metallicity differences (either positive or negative) of around 0.1–0.3 dex when compared to the $z \sim 0$ calibrators.

The R3 ratio at high redshift largely aligns with the $z \sim 0$ calibrations within the metallicity range of $7.7 < 12 + \log(\mathrm{O/H}) < 8.4$, exhibiting marginally higher line ratios at a given metallicity compared to $z \sim 0$. Conversely, for $12 + \log(\mathrm{O/H}) < 7.7$, the R2 ratio at high redshift is consistently lower than that of the local calibrations at the same metallicity. At a fixed O/H, high-redshift calibrations show elevated Ne3O2 ratio compared to local Ne3O2 diagnostics. The O32 ratio, an ionization-sensitive diagnostic commonly employed as a degeneracy breaker in the low-$z$ universe, consistently exhibits higher values at the same

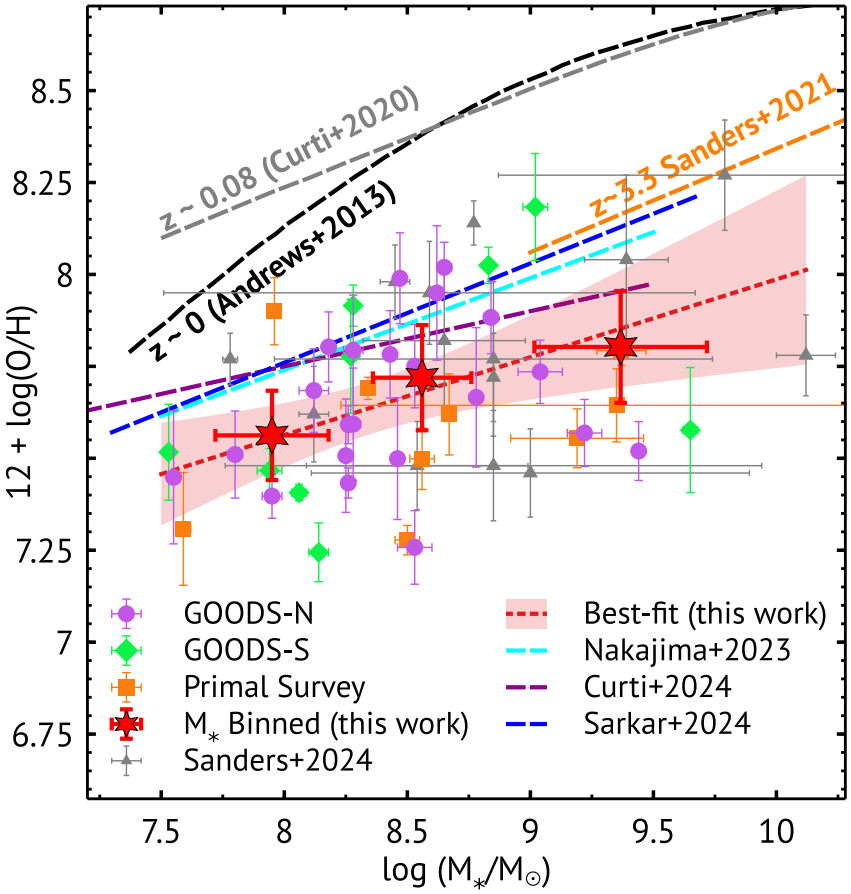


**Figure 7.** The MZR is shown, where metallicity is derived using the direct $T_e$-based method. Filled green diamonds, purple circles, and orange squares represent data from the GOODS-S, GOODS-N, and Primal surveys, respectively. Gray triangles denote a sample adapted from literature. Red stars show the stellar mass-averaged metallicities. The red dashed curve illustrates our best-fit MZR with a $1\sigma$ uncertainty, with best-fit parameters $\gamma = 0.211 \pm 0.120$ and $Z_{10} = 7.986 \pm 0.205$. For comparison, MZR curves from other calibration-based high-redshift JWST studies by M. Curti et al. (2024), K. Nakajima et al. (2023), and A. Sarkar et al. (2025) are included, as well as MZR curves from low-redshift studies by M. Curti et al. (2020), B. H. Andrews & P. Martini (2013), and R. L. Sanders et al. (2021).

**Table 2**
Comparing MZR with Different Studies Following Equation (11)

| Study | $z$ Range | $\gamma$ | $Z_{10}$ |
|---|---|---|---|
| This work (direct) | 3–10 | $0.21 \pm 0.03$ | $7.99 \pm 0.21$ |
| A. Sarkar et al. (2025) | 4–10 | $0.27 \pm 0.02$ | $8.28 \pm 0.08$ |
| K. Nakajima et al. (2023) | 4–10 | $0.25 \pm 0.03$ | $8.24 \pm 0.05$ |
| M. Curti et al. (2024) | 3–10 | $0.17 \pm 0.03$ | $8.06 \pm 0.18$ |
| R. L. Sanders et al. (2021) | 0 | $0.28 \pm 0.01$ | $8.77 \pm 0.01$ |
| | 2.3 | $0.30 \pm 0.02$ | $8.51 \pm 0.02$ |
| | 3.3 | $0.29 \pm 0.02$ | $8.41 \pm 0.03$ |
| M. Li et al. (2023) | 2 | $0.16 \pm 0.02$ | $8.50 \pm 0.13$ |
| | 3 | $0.16 \pm 0.01$ | $8.40 \pm 0.06$ |
| K. E. Heintz et al. (2023) | 7–10 | 0.33 | 7.95 |
| X. He et al. (2024) | 1.9 | $0.23 \pm 0.03$ | $8.54 \pm 0.12$ |
| | 2.88 | $0.26 \pm 0.04$ | $8.57 \pm 0.15$ |

**Note.** $Z_8$ is converted to $Z_{10}$ using $Z_{10} = 2\gamma + Z_8$.

metallicity in high-$z$ galaxies compared to local O32 calibrations. However, the substantial scatter observed in the high-$z$ O32 ratios—spanning up to an order of magnitude—results in a weak correlation between O32 and metallicity (see Section 4.1), raising concerns about the reliability of O32 as a diagnostic tool for resolving degeneracies in the high-$z$ universe. This scatter has been previously attributed to the significant diversity in ISM conditions observed within a sample of 25 high-redshift galaxies, and our sample, which is approximately $> 2.5$ times larger, further reinforces this interpretation. The calibrations derived based on $T_e$ metallicities in H II regions in the local universe have been found to be subject to systematic discrepancies (L. J. Kewley et al. 2019). The underlying factors contributing to the observed discrepancies are complex and challenging to unravel. These discrepancies stem from sample biases, the assumptions made during the estimation of metallicity, and several other factors discussed in Section 5.5, resulting in variations of up to 1 dex in metallicity estimates (see Figure 8 in L. J. Kewley et al. 2019). The sample biases introduced by these differences may be attributed to assumptions that H II regions will always display the auroral lines if observed for an adequate duration and that H II regions showing auroral lines are representative of all H II regions and galaxies.

Our capacity to evaluate metallicity diagnostics for values above $12 + \log(\mathrm{O/H}) > 8.4$ is presently constrained by the absence of data points beyond this range. The apparent absence of a turnover in R2 may be attributed to the lack of data points at $12 + \log(\mathrm{O/H}) > 8.4$. To better understand the behavior of high-redshift metallicity calibrations at higher metallicities

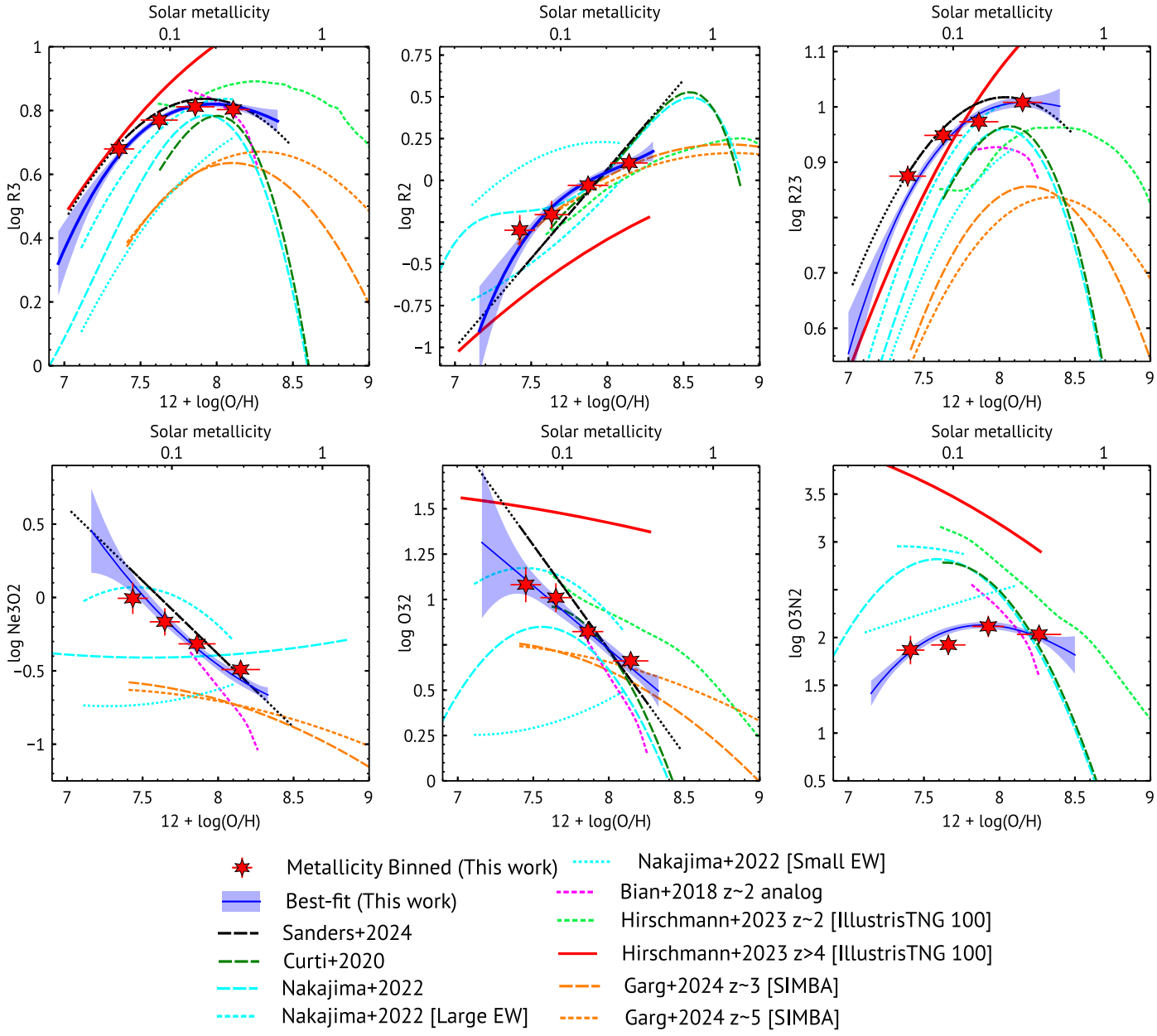


**Figure 8.** Comparison of our strong-line metallicity calibration for high-redshift galaxies in the range of $3 < z < 10$ (blue curve with shaded regions showing the $1\sigma$ uncertainty) with previous observational calibrations and simulations, as color coded in the legend. The comparison includes strong-line calibrations for $z \sim 0$ from T. Jones et al. (2015), M. Curti et al. (2020), K. Nakajima et al. (2022), along with local analogs that occupy similar regions to $z \sim 2$ star-forming galaxies on the BPT diagram (F. Bian et al. 2018). High-redshift calibrations for galaxies in the range of $2 < z < 9$ are also shown (R. L. Sanders et al. 2024), alongside strong-line metallicity calibrations derived from simulations, including the SIMBA cosmological galaxy formation model at $z \sim 3$ and $z \sim 5$ (P. Garg et al. 2024), and the TNG100 galaxy population simulation at $z \sim 2$ and $z > 4$ (M. Hirschmann et al. 2023).

$(12 + \log(\mathrm{O/H}) > 8.4)$, a more comprehensive analysis, incorporating a larger high-$z$ [O III]$\lambda$4363 sample and stacked spectra from hundreds of galaxies, is needed.

### 5.3. High-redshift Metallicity Calibrations

Before the JWST era, several efforts aimed to expand the sample of high-$z$ galaxies with robust oxygen abundance measurements using the direct method. This was often achieved by stacking spectra of local analogs of high-$z$ star-forming galaxies to develop calibrations consistent with the high-$z$ regime (T. Jones et al. 2015; F. Bian et al. 2018; E. Pérez-Montero et al. 2021) or by obtaining direct metallicity measurements from the [O III]$\lambda$4363 auroral line (R. L. Sanders et al. 2020). High-redshift metallicity calibrations were initially explored using a sample of 46 star-forming galaxies covering a redshift range of $z = 1.4$ to $z = 8.7$, including 25 JWST-detected galaxies at $z > 2$ (R. L. Sanders et al. 2024). This study provided $T_e$-based calibrations for strong emission lines such as R23, R2, R3, O32, and Ne3O2. A subsequent study by I. H. Laseter et al. (2024) generally found agreement with these calibrations for most strong emission lines, though they identified some deviations for R2 and O32, which were attributed to significant scatter in the data. To enhance the robustness of these calibrations, I. H. Laseter et al. (2024) emphasized the importance of obtaining larger samples of [O III]$\lambda$4363 detections for the high-redshift universe.

We revisited the $T_e$-based metallicities using a sample of 67 [O III]$\lambda$4363 lines at $z > 3$, incorporating 42 new detections from our analysis. The solid blue lines in Figure 8 represent our calibration, which is compared to the calibrations of R. L. Sanders et al. (2024), shown with black lines. For R23 and R3, the expanded sample indicates a good agreement with the calibrations of R. L. Sanders et al. (2024), with the metallicity offset being $\sim<0.05$ dex. The Ne3O2 calibrator also shows reasonable agreement, with metallicity offsets of $\sim<0.07$ dex. For R2, while there is generally reasonable alignment, there is a maximum metallicity offset of around $\sim$0.15 dex. Whereas R. L. Sanders et al. (2024) suggest a linear relationship for the metallicity calibrator, our findings indicate a potential nonlinear trend, as shown in Figures 5 and 8. For O32, we note a slight difference in the slopes of the metallicity calibrators between our calculations and those presented by R. L. Sanders et al. (2024), with the maximum metallicity offset reaching approximately $\sim$0.4 dex.

### 5.4. New Metallicity Calibrators

The large intrinsic scatter in individual line-flux ratios and the $T_e$-based metallicity for high-redshift galaxies makes the calibration a challenging task. I. H. Laseter et al. (2024) proposed a novel metallicity calibration relation, $\hat{R}$ based on observations from 465 low-metallicity objects compiled from several previous observations. $\hat{R}$ is defined as

$$\hat{R} = \cos(\phi)\,\log(\mathrm{R2}) + \sin(\phi)\,\log(\mathrm{R3}), \tag{13}$$

which is equivalent to a rotation of the R2-R3 plane around the O/H axis. The process to find $\phi$ involves fitting a fourth-order polynomial to the resulting $\hat{R}$ ratio versus the metallicity, in the form of $\hat{R} = \sum_{n=0}^{N} c_n x^n$, where $x = 12 + \log(\mathrm{O/H}) - 8.0$, and the angle $\phi$ minimizes the scatter in metallicity from the best-fit relation. I. H. Laseter et al. (2024) found a best fit $\phi$ of 61.82 deg, which gives $\hat{R} = 0.47\,\log(\mathrm{R2}) + 0.88\,\log(\mathrm{R3})$.

In this paper, we tested this new metallicity calibration for our full sample of galaxies within $3 < z < 10$. Figure 9 illustrates the estimated $\hat{R}$ of our full sample along with the best-fit curve predicted by I. H. Laseter et al. (2024). We next follow the same prescription as I. H. Laseter et al. (2024) to constrain the $\phi$ value for our high-redshift sample. We found $\phi = 79.8 \pm 2.4$ deg, which gives

$$\hat{R} = 0.18\,\log(\mathrm{R2}) + 0.98\,\log(\mathrm{R3}). \tag{14}$$

Figure 9 illustrates this new calibration relation for our full sample with the best-fit curve. The best-fit coefficients for our new $\hat{R}$ calibration {0.796, 0.082, −0.216, 0.472} are listed in Table 2. To quantify the scattering in both $\hat{R}$ relations, we measure the rms scatter ($\sigma_{\mathrm{rms}}$) of derived $\hat{R}$ from their respective best-fit curve. For I. H. Laseter et al. (2024), we find $\sigma_{\mathrm{rms}} = 0.14$ dex and for our new $\hat{R}$ we find $\sigma_{\mathrm{rms}} = 0.06$ dex.

Our new $\hat{R}$ calibration significantly reduces the scatter in the observed line ratio versus metallicity, as clearly seen in Figure 9, and therefore can be used to estimate metallicity in high-redshift galaxies. However, we caution that our new $\hat{R}$ calibration is derived from a limited sample of 59 high-redshift galaxies. Users should apply these new calibrations within the range of $7.2 \leqslant 12 + \log(\mathrm{O/H}) \leqslant 8.2$ to ensure reliable metallicity estimates. Though our new $\hat{R}$ calibration hints toward the challenges of using low-redshift metallicity calibrators in deriving metallicity in high-redshift galaxies, more high-quality data is needed to put stronger constraints on the high-redshift metallicity calibrators.

### 5.5. Caveats on $T_e$-based Metallicity and Calibrations

The $T_e$-based method offers a relatively direct approach to measuring gas-phase metallicity compared to estimates derived from local calibration relations. However, it is not without limitations and can suffer from significant systematic uncertainties due to various factors (L. J. Kewley et al. 2019). One major challenge lies in the simplified multizone models used to describe the temperature structure within ionized nebulae. These models often lack critical information about the temperature of the $O^+$ zone, as the required auroral emission-line doublet at [O II]$\lambda\lambda$7322,7332 is frequently too faint to detect or falls outside the wavelength range of NIRSpec. As a result, the $O^+/H$ ratio, which can represent a substantial fraction ($\sim$22%–30%) of the total oxygen abundance, must be inferred indirectly. This is typically done using a proxy temperature, such as $T_e$(N II), or through empirical relations linking $T_e$(O II) to $T_e$(O III). Additionally, the [O III]$\lambda$4363 auroral line, which is critical for $T_e$-based metallicity estimates, is predominantly produced in hot, metal-poor gas (e.g., M. Peimbert 1967; L. J. Kewley & S. L. Ellison 2008). In metal-rich regions, the gas temperature is lower, and collisional excitation of the [O III] lines is less efficient. This leads to significant temperature gradients and fluctuations within H II regions, which systematically bias $T_e$ measurements to higher values and result in underestimates of metallicity (e.g., G. Stasińska 2005; R. M. Yates et al. 2020).

Accurate determinations of electron temperature ($T_e$) require the detection of multiple faint emission lines, such as [O II], [S II], and [N II]. These lines must be measured with an intensity accuracy better than 5% relative to Balmer recombination lines (G. F. Hägele et al. 2006). More deep observations of high-redshift galaxies with JWST/NIRSpec are needed to address these challenges and improve the reliability of $T_e$-based metallicity measurements.

Another significant source of systematic uncertainty in metallicity estimates stems from the assumptions made in the calculation of the ionization correction factor, particularly for unobserved ionization stages such as O IV. Standard ICF calculations typically assume a uniform electron temperature and a simplified H II region composition, often considering only a single or a small subset of ionic species. However, electron temperatures can vary significantly—by as much as 2000–3000 K—across different H II regions (e.g., G. F. Hägele et al. 2006; L. J. Kewley et al. 2019). H II regions are complex, multizone environments containing a wide variety of atomic and ionic species. The ionization states of elements are strongly influenced by the stellar radiation field, which itself depends on detailed stellar atmosphere models. In the absence of accurate data, the assumed ICF can deviate substantially from the true distribution of unobserved ionization stages, particularly in low-excitation H II regions. This deviation can result in significant underestimation of metallicity, with errors reaching up to 0.2 dex when using the $T_e$-based method (G. F. Hägele et al. 2008). Such uncertainties highlight the need for improved models and future JWST observations to refine ICF calculations and enhance the reliability of metallicity estimates.

The metallicity calibration diagnostics for certain line ratios, such as R2, O32, and Ne3O2, are known to be sensitive to

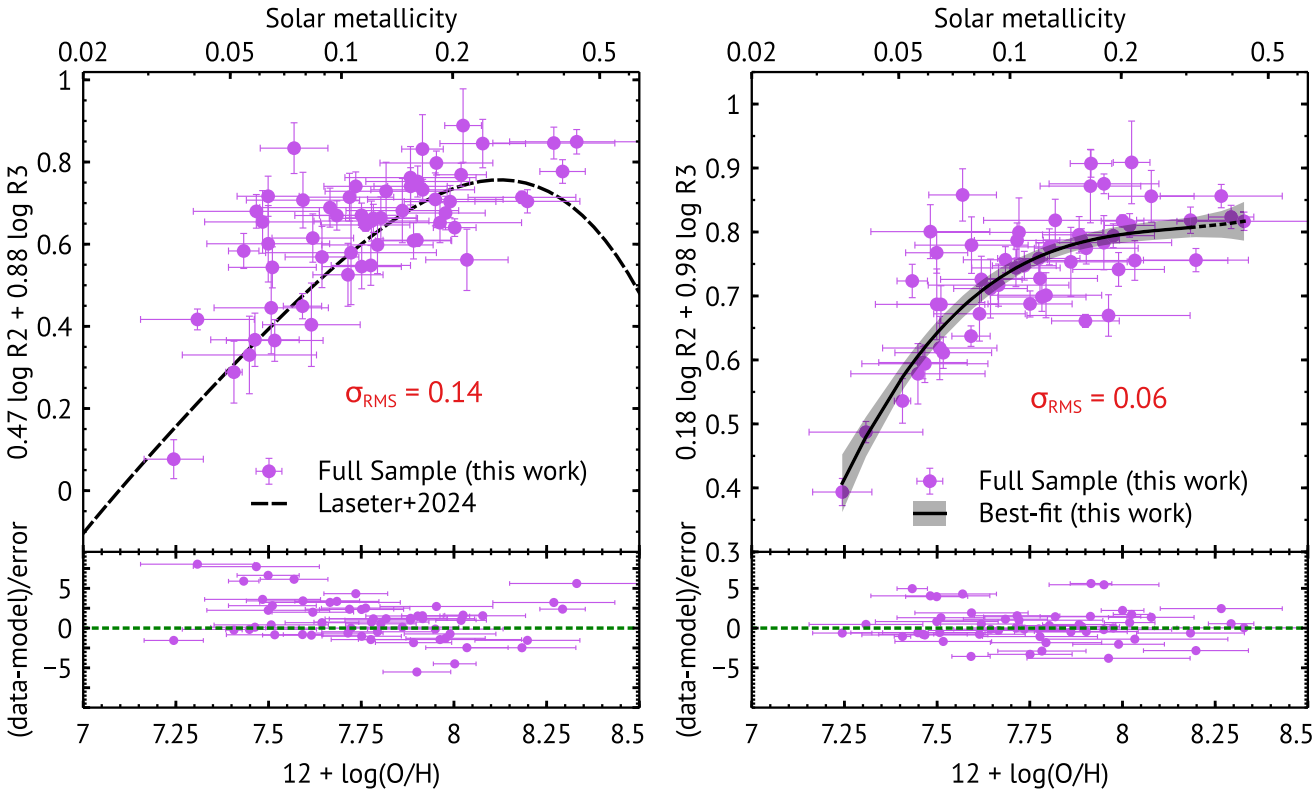


**Figure 9.** The novel metallicity calibration relation, $\hat{R}$, is displayed. In both panels, filled purple circles represent our entire sample. Left: $\hat{R}$ derived using the best-fit relation, $\hat{R} = 0.47$ log R2 + 0.88 log R3, from I. H. Laseter et al. (2024). Their best-fit polynomial curve is indicated by the dashed line. Right: $\hat{R}$ derived exclusively from our high-redshift sample, using the relation $\hat{R} = 0.18$ log R2 + 0.98 log R3. The best-fit polynomial curve for this calibration is shown as a black solid line and dashed line with $1\sigma$ uncertainties. The solid line represents the best fit within the range where the statistical significance is higher, supported by a larger number of data points. Bottom panels in both figures illustrate the $\Delta\chi$ values. Our new calibration substantially reduces the intrinsic scatter in the data, offering a more reliable tool for measuring metallicity at higher redshifts.

variations in the ionization parameter (e.g., L. J. Kewley & S. L. Ellison 2008; I. H. Laseter et al. 2024). To test this for our sample of galaxies, we derived the ionization parameter, defined as the ratio of hydrogen-ionizing photon flux to the hydrogen density in the H II region, using observed O32 line ratios. We adopted the linear relation between log(O32) and the ionization parameter from C. Papovich et al. (2022):

$$\log q = (0.86 \pm 0.07)\,\mathrm{logO32} + (7.53 \pm 0.02), \qquad (15)$$

where $q$ represents the ionization parameter in units of centimeters per second. Figure 10 illustrates the derived ionization parameters as a color gradient in Ne3O2 versus metallicity and $\hat{R}$ versus metallicity plots. For the Ne3O2 ratio, the ionization parameter varies by $\sim$1 dex for a given metallicity in the range of 7.5–7.8, indicating that the Ne3O2 diagnostic may be affected by the log $q$–log(O/H) relation. However, the limited statistical power of our sample restricts a more detailed investigation of this dependency, such as splitting the entire sample into two or more $q$ bins. A more robust data set with higher statistical power is necessary for such an analysis. Future JWST surveys are expected to provide larger high-redshift galaxy samples with high-quality spectra, enabling a more comprehensive study. In contrast, the $\hat{R}$ diagnostic shows weak dependence on the ionization parameter, highlighting its robustness as a metallicity calibration diagnostic for high-redshift galaxies.

## 6. Summary

We report the discovery of [O III] auroral lines ([O III]$\lambda$4363) in 42 galaxies observed with JWST/NIRSpec within the redshift range of $3 < z < 10$. Combined with previous observations of 25 galaxies featuring [O III] auroral lines within the same redshift range, this combined data set of 67 galaxies significantly expands the sample of high-redshift galaxies available for study, nearly tripling the number used in prior JWST studies for calibrating $T_e$-based metallicities and empirical diagnostics for $3 < z < 10$. Our main findings are summarized below.

1. The 42 galaxies in this study were selected from a sample within the redshift range of $3 < z < 10$, primarily derived from publicly available data releases, including JADES/GOODS-S, JADES/GOODS-N, and JWST-PRIMAL. These galaxies were observed using the NIRSpec low-resolution PRISM/CLEAR configuration as well as medium-resolution gratings. Stellar masses were determined through spectral fitting of the PRISM spectra using `Bagpipes`, with the sample spanning a mass range of $10^{7.5}$–$10^{10}\,M_{\odot}$.
2. To measure emission-line fluxes, we utilized the medium-resolution grating spectra for each galaxy, focusing on key nebular emission lines when detected, including the hydrogen Balmer lines, [O II]$\lambda\lambda$3727,3729, [Ne III] $\lambda$3867, [O III]$\lambda$4363, [O III]$\lambda\lambda$5007,4959, and [N II] $\lambda$6584. The line fluxes were corrected for dust extinction

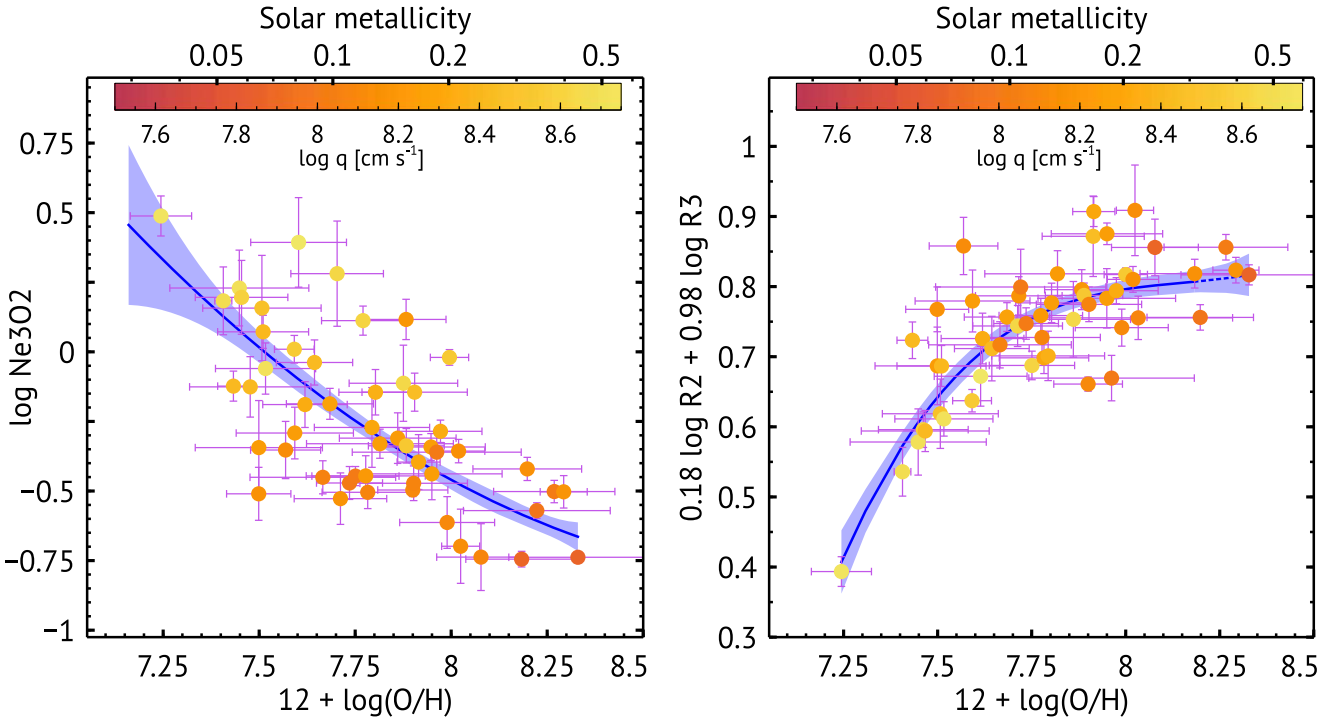


**Figure 10.** Left: the Ne3O2 line ratio vs. metallicity is shown, with derived ionization parameters for individual galaxies based on the O32 line ratio. Ionization parameters are represented as a color gradient, with the color bar displayed at the top. The best-fit polynomial for the entire sample (as listed in Table 1) is shown as a blue solid line, along with the $1\sigma$ uncertainty. The ionization parameter varies by $\sim$1 dex for a given metallicity of around 7.6, indicating that Ne3O2 diagnostics may be affected by the log $q$–log(O/H) relation. Right: similar to the left panel but for the $\hat{R}$ calibration. The $\hat{R}$ diagnostic is less impacted by the log $q$–log(O/H) relation, demonstrating the robustness of this calibration relation.

using the curve from D. Calzetti et al. (2000). To distinguish star-forming galaxies from AGN-dominated ones, we applied the MEx diagram.

3. The electron temperatures in the high-ionization regions, $T_e$([O III]), were derived from the flux ratio of [O III] $\lambda$5007 to [O III]$\lambda$4363, assuming an electron density of 300 cm$^{-3}$. For our sample of 42 galaxies, $T_e$([O III]) ranges from 12,000 to 24,000 K, consistent with extremely metal-poor local galaxies and previous JWST surveys. Among these, 10 galaxies also show [O II] $\lambda\lambda$7322,32 auroral lines with an S/N $\geqslant$ 2, allowing us to use the [O II]$\lambda\lambda$3727,3729 to [O II]$\lambda\lambda$7322,32 line ratios to determine the electron temperature in the low-ionization region, $T_e$([O II]). Electron densities for the low-ionization region were calculated using the [S II] $\lambda\lambda$6716,31 doublet ratios. For these 10 galaxies, $T_e$([O II]) ranges between 10,830 and 20,000 K. We established a best-fit relation between $T_e$([O II]) and $T_e$([O III]): $T_e([\mathrm{O\,II}]) = (0.58 \pm 0.19) \times T_e([\mathrm{O\,III}]) + (4520 \pm 2000)$ K, which aligns with observations of local galaxies. Oxygen abundances were calculated for our sample using `PyNeb`, based on the measured $T_e$([O III]) and either measured or inferred $T_e$([O II]). The resulting metallicities range from 12 + log(O/H) = 7.2 to 8.4, indicating that these high-redshift galaxies are relatively metal-poor, in agreement with prior JWST findings.
4. We present, for the first time, the MZR for galaxies in the redshift range of $3 < z < 10$, with metallicities entirely derived using the direct $T_e$ method. To estimate the average MZR, we divided our sample into three stellar mass bins: log($M_*/M_\odot$) = 7–8, 8–9, and 9–10, ensuring that each bin contained at least 15 galaxies. Our analysis yielded a best-fit slope of $\gamma = 0.211 \pm 0.120$ and a metallicity intercept of $Z_{10} = 7.986 \pm 0.205$. The slope and normalization of our MZR are consistent within $1\sigma$ uncertainty with previous high-redshift studies where metallicities were derived using local calibrations, including JADES + Primal (A. Sarkar et al. 2025), CEERS (K. Nakajima et al. 2023), and JADES (M. Curti et al. 2024). However, our best-fit curve is approximately 0.2 dex lower at $M_* = 10^8\,M_\odot$ compared to these studies. Additionally, when compared with MZR studies at lower redshifts, we find that metallicities for a given stellar mass in our sample are noticeably lower than those reported at $z \sim 0$ by B. H. Andrews & P. Martini (2013) and at $z \sim 0.08$ by M. Curti et al. (2020). We estimate the scatter, $\sigma_{\mathrm{scatter}}$, in our MZR to be approximately 0.09 dex, which aligns well with the intrinsic scatter of 0.08 dex observed in low-stellar-mass (log $M_\star/M_\odot \sim <9$) galaxies at redshifts $3 < z < 10$ (M. Curti et al. 2024).
5. We utilized our full sample of 67 galaxies (42 from this study and 25 from previous literature) within the redshift range of $3 < z < 10$ to develop new empirical metallicity calibrators tailored to high-redshift galaxies. A polynomial fitting approach, $\log(R) = \sum c_n x^n$, was used to derive calibrators based on six line ratios: R3, R2, R23, Ne3O2, O32, and O3N2, spanning metallicities of 12 + log(O/H) = 7.2–8.4. The best-fit polynomials for each line ratio are presented in Table 1.
6. We tested the novel metallicity calibration relation proposed by I. H. Laseter et al. (2024) for local galaxies, defined as $\hat{R} = \cos(\Phi)\ \log(\mathrm{R2}) + \sin(\Phi)\ \log(\mathrm{R3})$ with

$\Phi = 61.^{\circ}82$, on our high-redshift sample. For this relation, we found an rms scatter of $\sigma_{\rm rms} = 0.14$ when comparing the best-fit Rcap values from I. H. Laseter et al. (2024) with our observational data. Following the methodology of I. H. Laseter et al. (2024), we optimized the $\Phi$ parameter for our high-redshift sample, finding an optimal $\Phi = 79.8 \pm 2.^{\circ}4$, which yields $\hat{R} = 0.18$ log(R2) + 0.98 log(R3). For our optimized Rcap, the rms scatter reduced significantly to $\sigma_{\rm rms} = 0.06$, indicating a tighter calibration compared to the original Rcap for local galaxies and confirming a large intrinsic scatter among high-redshift galaxies.

## Acknowledgments

We extend our heartfelt gratitude to Dr. Lisa Kewley for her invaluable insights and stimulating discussions. P.C. acknowledges support from Chandra grants 16619325 under contract GO3-24124X, 16619354 under contract GO4-25094X, and 16619349 under contract GO3-24024X.

This work is based on observations conducted with the NASA/ESA/CSA James Webb Space Telescope (JWST). The data were obtained from the Mikulski Archive for Space Telescopes (MAST) hosted by the Space Telescope Science Institute (STScI). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST operations. We acknowledge the JADES repository at MAST (M. Rieke et al. 2023). The JADES data products used in this study were retrieved from the DR3 archive, available at https://jades-survey.github.io/scientists/data.html. Additionally, the Primal Survey data products were obtained from the DAWN JWST Archive (DJA). The DJA is an initiative of the Cosmic Dawn Center, funded by the Danish National Research Foundation under grant DNRF140.

## Appendix

We used publicly available JWST/NIRSpec data from the JADES and JWST-PRIMAL surveys, as listed in Tables 3 and 4.

**Table 3**
GOODS-S and GOODS-N JWST/NIRSpec Sample of Galaxies Analyzed in this Paper[a]

| NIRSpec ID | R.A. | Decl. | $z$ | $\log(M_*/M_\odot)$ | 12+log(O/H) (direct) | $T_e$[b] (K) |
|---|---|---|---|---|---|---|
| 54612 | 53.14468 | −27.771185 | 3.08 | 9.05 ± 0.05 | 8.18 ± 0.15 | 12,441 ± 1211 |
| 39898 | 53.1324005 | −27.8091354 | 3.17 | 9.68 ± 0.03 | 7.47 ± 0.17 | 23,929 ± 2177 |
| 26113 | 189.0996212 | 62.2635629 | 3.23 | 8.56 ± 0.07 | 7.26 ± 0.1 | 21,483 ± 2005 |
| 4550 | 189.1924782 | 62.2388249 | 3.24 | 8.15 ± 0.06 | 7.68 ± 0.11 | 18,735 ± 2103 |
| 56785 | 189.3147109 | 62.2023856 | 3.44 | 9.25 ± 0.07 | 7.57 ± 0.09 | 23,800 ± 2082 |
| 1137 | 189.105766 | 62.283372 | 3.66 | 9.07 ± 0.09 | 7.74 ± 0.09 | 18,503 ± 1545 |
| 3683 | 189.0936083 | 62.2918914 | 3.66 | 8.31 ± 0.04 | 7.79 ± 0.15 | 15,962 ± 1725 |
| 10009453 | 53.1787077 | −27.7989084 | 3.71 | 7.56 ± 0.02 | 7.52 ± 0.13 | 19,818 ± 2591 |
| 1048 | 189.0583155 | 62.2725583 | 3.87 | 7.98 ± 0.04 | 7.40 ± 0.06 | 21,767 ± 1100 |
| 15529 | 189.215044 | 62.2770075 | 3.87 | 8.75 ± 0.03 | 8.29 ± 0.06 | 12,041 ± 521 |
| 902 | 189.1932763 | 62.2537271 | 4.06 | 8.50 ± 0.02 | 7.99 ± 0.12 | 14,159 ± 1238 |
| 1165 | 189.111861 | 62.2877299 | 4.38 | 8.30 ± 0.02 | 8.20 ± 0.14 | 12,162 ± 1170 |
| 16553 | 189.1436028 | 62.2805455 | 4.38 | 8.46 ± 0.02 | 7.78 ± 0.12 | 16,424 ± 1512 |
| 11836 | 189.2205875 | 62.2636751 | 4.41 | 8.29 ± 0.02 | 7.59 ± 0.05 | 18,470 ± 975 |
| 10000865 | 189.2629857 | 62.2501062 | 4.41 | 8.56 ± 0.02 | 7.75 ± 0.11 | 17,023 ± 1738 |
| 206035 | 53.1581721 | −27.7864763 | 4.77 | 8.30 ± 0.02 | 7.78 ± 0.16 | 16,672 ± 2248 |
| 61321 | 53.154274 | −27.7524204 | 4.84 | 8.31 ± 0.02 | 7.92 ± 0.06 | 17,116 ± 912 |
| 10009642 | 53.1926798 | −27.7842212 | 4.84 | 8.86 ± 0.02 | 8.02 ± 0.05 | 16,122 ± 700 |
| 920 | 189.0981446 | 62.2555524 | 4.88 | 8.65 ± 0.03 | 7.95 ± 0.18 | 14,910 ± 2109 |
| 70920 | 189.2509624 | 62.1603815 | 5.04 | 8.29 ± 0.02 | 7.43 ± 0.04 | 23,973 ± 1119 |
| 607 | 189.1169473 | 62.2220788 | 5.18 | 8.31 ± 0.03 | 7.59 ± 0.15 | 21,021 ± 2046 |
| 721 | 189.1153168 | 62.2340987 | 5.18 | 7.58 ± 0.015 | 7.45 ± 0.18 | 20,728 ± 2267 |
| 59412 | 189.1563218 | 62.2100022 | 5.18 | 8.81 ± 0.02 | 7.67 ± 0.19 | 18,812 ± 3021 |
| 79349 | 189.2096823 | 62.207252 | 5.18 | 7.83 ± 0.03 | 7.51 ± 0.12 | 21,021 ± 2491 |
| 131737 | 53.1990421 | −27.7725801 | 5.89 | 7.97 ± 0.05 | 7.47 ± 0.06 | 21,050 ± 1102 |
| 78891 | 189.2258239 | 62.2042147 | 6.55 | 8.68 ± 0.03 | 8.02 ± 0.07 | 14,569 ± 821 |
| 1967 | 189.1650306 | 62.3001933 | 6.56 | 8.49 ± 0.02 | 7.50 ± 0.17 | 21,347 ± 2140 |
| 38428 | 189.1792747 | 62.2758955 | 6.71 | 9.47 ± 0.02 | 7.47 ± 0.06 | 20522 ± 1100 |
| 38432 | 189.18617 | 62.2708636 | 6.72 | 8.87 ± 0.03 | 7.88 ± 0.1 | 16,107 ± 1415 |
| 38420 | 189.17514 | 62.2822634 | 6.73 | 8.28 ± 0.02 | 7.51 ± 0.15 | 19,854 ± 2839 |
| 18536 | 189.1553143 | 62.2864715 | 6.81 | 8.21 ± 0.02 | 7.80 ± 0.1 | 16,892 ± 1495 |
| 20213084 | 53.1589064 | −27.765076 | 8.49 | 8.09 ± 0.03 | 7.41 ± 0.02 | 20,761 ± 512 |
| 265801[c] | 53.1124351 | −27.7746258 | 9.43 | 8.17 ± 0.04 | 7.36 ± 0.08 | 21,176 ± 1916 |

**Notes.**
[a] F. D'Eugenio et al. (2024).
[b] Temperature measured using the [O III]λ4363 emission line.
[c] Metallicity of JADES-GS-z9-0 at $z = 9.4327$ is also reported by M. Curti et al. (2024b), consistent with our new measurements.

**Table 4**
PRIMAL JWST/NIRSpec Sample of Galaxies Analyzed in this Paper[a]

| NIRSpec ID | R.A. | Decl. | $z$ | $\log(M_*/M_\odot)$ | 12+log(O/H) (Direct) | $T_e$[b] (K) |
|---|---|---|---|---|---|---|
| 4385 | 215.1795990 | 53.0620667 | 3.42 | 8.20 ± 0.04 | 8.08 ± 0.12 | 14,596 ± 1339 |
| 13491 | 53.1501095 | −27.8197035 | 3.47 | 8.37 ± 0.02 | 7.69 ± 0.03 | 14,815 ± 1166 |
| 3585 | 215.0232093 | 53.0079722 | 3.87 | 8.59 ± 0.05 | 7.50 ± 0.08 | 23,383 ± 1949 |
| 9489 | 53.1687170 | −27.8151743 | 4.02 | 7.99 ± 0.02 | 7.90 ± 0.09 | 14,346 ± 1062 |
| 40066 | 3.5997164 | −30.4318948 | 4.02 | 9.40 ± 0.10 | 7.79 ± 0.04 | 14,536 ± 316 |
| 1173 | 215.1542076 | 52.9558470 | 5.00 | 9.38 ± 1.12 | 7.64 ± 0.10 | 18,642 ± 1870 |
| 110000 | 3.5706428 | −30.4146380 | 5.76 | 9.22 ± 0.27 | 7.55 ± 0.08 | 17,397 ± 1243 |
| 9842 | 53.1540870 | −27.7660620 | 5.80 | 7.62 ± 0.03 | 7.31 ± 0.15 | 22,186 ± 2881 |
| 2782 | 214.8234525 | 52.8302813 | 5.24 | 8.53 ± 0.05 | 7.28 ± 0.04 | 22,777 ± 900 |
| 689 | 214.9990525 | 52.9419767 | 7.55 | 8.70 ± 0.01 | 7.62 ± 0.11 | 19,339 ± 1700 |

**Notes.**
[a] K. E. Heintz et al. (2024).
[b] Temperature measured using [O III]$\lambda$4363 emission line.

## ORCID iDs

Priyanka Chakraborty https://orcid.org/0000-0002-4469-2518
Arnab Sarkar https://orcid.org/0000-0002-5222-1337
Randall Smith https://orcid.org/0000-0003-4284-4167
Gary J. Ferland https://orcid.org/0000-0003-4503-6333
Michael McDonald https://orcid.org/0000-0001-5226-8349
William Forman https://orcid.org/0000-0002-9478-1682
Mark Vogelsberger https://orcid.org/0000-0001-8593-7692
Paul Torrey https://orcid.org/0000-0002-5653-0786
Alex M. Garcia https://orcid.org/0000-0002-8111-9884
Mark Bautz https://orcid.org/0000-0002-1379-4482
Adam Foster https://orcid.org/0000-0003-3462-8886
Eric Miller https://orcid.org/0000-0002-3031-2326
Catherine Grant https://orcid.org/0000-0002-4737-1373

## References

Allende Prieto, C., Lambert, D. L., & Asplund, M. 2001, ApJL, 556, L63
Anders, E., & Grevesse, N. 1989, GeCoA, 53, 197
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Arellano-Córdova, K. Z., Berg, D. A., Chisholm, J., et al. 2022, ApJL, 940, L23
Asplund, M., Amarsi, A. M., & Grevesse, N. 2021, A&A, 653, A141
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Berg, D. A., Chisholm, J., Erb, D. K., et al. 2021, ApJ, 922, 170
Berg, D. A., Erb, D. K., Auger, M. W., Pettini, M., & Brammer, G. B. 2018, ApJ, 859, 164
Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2024, ApJ, 974, 92
Bian, F., Kewley, L. J., & Dopita, M. A. 2018, ApJ, 859, 175
Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
Bresolin, F., Gieren, W., Kudritzki, R.-P., et al. 2009, ApJ, 700, 309
Brown, J. S., Martini, P., & Andrews, B. H. 2016, MNRAS, 458, 1529
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bunker, A. J., Cameron, A. J., Curtis-Lake, E., et al. 2024, A&A, 690, A288
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cameron, A. J., Katz, H., & Rey, M. P. 2023, MNRAS, 522, L89
Campbell, A., Terlevich, R., & Melnick, J. 1986, MNRAS, 223, 811
Carnall, A. C., McLure, R. J., Dunlop, J. S., & Davé, R. 2018, MNRAS, 480, 4379
Carnall, A. C., McLure, R. J., Dunlop, J. S., et al. 2023, Natur, 619, 716
Chabrier, G. 2003, PASP, 115, 763
Chakraborty, P., Sarkar, A., Wolk, S., et al. 2024, arXiv:2406.05306
Charlot, S., Kauffmann, G., Longhetti, M., et al. 2002, MNRAS, 330, 876
Charlot, S., & Longhetti, M. 2001, MNRAS, 323, 887
Chatzikos, M., Bianchi, S., Camilloni, F., et al. 2023, RMxAA, 59, 327
Christensen, L., Laursen, P., Richard, J., et al. 2012a, MNRAS, 427, 1973
Christensen, L., Richard, J., Hjorth, J., et al. 2012b, MNRAS, 427, 1953
Coil, A. L., Aird, J., Reddy, N., et al. 2015, ApJ, 801, 35
Curti, M., Cresci, G., Mannucci, F., et al. 2017, MNRAS, 465, 1384
Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, MNRAS, 518, 425
Curti, M., Maiolino, R., Curtis-Lake, E., et al. 2024, A&A, 684, A75
Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944
Curti, M., Witstok, J., Jakobsen, P., et al. 2024b, arXiv:2407.02575
Curtis-Lake, E., Carniani, S., Cameron, A., et al. 2023, NatAs, 7, 622
Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 415, 11
D'Eugenio, F., Cameron, A. J., Scholtz, J., et al. 2025, ApJS, 277, 4
Dopita, M. A., Kewley, L. J., Heisler, C. A., & Sutherland, R. S. 2000, ApJ, 542, 224
Edmunds, M. G., & Pagel, B. E. J. 1984, MNRAS, 211, 507
Eisenstein, D. J., Willott, C., Alberts, S., et al. 2023, arXiv:2306.02465
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Feroz, F., & Hobson, M. P. 2008, MNRAS, 384, 449
Ferruit, P., Jakobsen, P., Giardino, G., et al. 2022, A&A, 661, A81
Finkelstein, S. L., Bagley, M. B., Arrabal Haro, P., et al. 2022, ApJL, 940, L55
Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Garcia, A. M., Torrey, P., Ellison, S. L., et al. 2025, MNRAS, 536, 119
Garg, P., Narayanan, D., Sanders, R. L., et al. 2024, ApJ, 972, 113
Garnett, D. R. 1992, AJ, 103, 1330
Garnett, D. R., & Shields, G. A. 1987, ApJ, 317, 82
Gburek, T., Siana, B., Alavi, A., et al. 2019, ApJ, 887, 168
Hägele, G. F., Díaz, ÁI., Terlevich, E., et al. 2008, MNRAS, 383, 209
Hägele, G. F., Pérez-Montero, E., Díaz, Á. I., Terlevich, E., & Terlevich, R. 2006, MNRAS, 372, 293
He, X., Wang, X., Jones, T., et al. 2024, ApJL, 960, L13
Heintz, K. E., Brammer, G. B., Giménez-Arteaga, C., et al. 2023, NatAs, 7, 1517
Heintz, K. E., Brammer, G. B., Watson, D., et al. 2025, A&A, 693, A60
Hirschmann, M., Charlot, S., & Somerville, R. S. 2023, MNRAS, 526, 3504
Izotov, Y. I., Guseva, N. G., Fricke, K. J., & Henkel, C. 2019, A&A, 623, A40
Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955
Jones, T., Martin, C., & Cooper, M. C. 2015, ApJ, 813, 126
Jones, T., Sanders, R., Chen, Y., et al. 2023, ApJL, 951, L17
Juneau, S., Bournaud, F., Charlot, S., et al. 2014, ApJ, 788, 88
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Geller, M. J., & Jansen, R. A. 2004, AJ, 127, 2002
Kewley, L. J., Maier, C., Yabe, K., et al. 2013, ApJL, 774, L10
Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, ARA&A, 57, 511
Kisielius, R., Storey, P. J., Ferland, G. J., & Keenan, F. P. 2009, MNRAS, 397, 903
Kojima, T., Ouchi, M., Nakajima, K., et al. 2017, PASJ, 69, 44
Kroupa, P. 2002, Sci, 295, 82
Laseter, I. H., Maseda, M. V., Curti, M., et al. 2024, A&A, 681, A70
Lee, H., Skillman, E. D., Cannon, J. M., et al. 2006, ApJ, 647, 970
Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155
Li, M., Cai, Z., Bian, F., et al. 2023, ApJL, 955, L18

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Lower, S., Narayanan, D., Leja, J., et al. 2020, ApJ, 904, 33
Luridiana, V., Morisset, C., & Shaw, R. A. 2015, A&A, 573, A42
Maiolino, R., & Mannucci, F. 2019, A&ARv, 27, 3
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Marino, R. A., Rosales-Ortega, F. F., Sánchez, S. F., et al. 2013, A&A, 559, A114
Matteucci, F. 2012, Chemical Evolution of Galaxies (Berlin: Springer)
Meléndez, M., Heckman, T. M., Martínez-Paredes, M., Kraemer, S. B., & Mendoza, C. 2014, MNRAS, 443, 1358
Morishita, T., Stiavelli, M., Grillo, C., et al. 2024, ApJ, 971, 43
Morisset, C., Luridiana, V., García-Rojas, J., et al. 2020, Atoms, 8, 66
Nagao, T., Maiolino, R., & Marconi, A. 2006, A&A, 459, 85
Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, ApJS, 269, 33
Nakajima, K., Ouchi, M., Xu, Y., et al. 2022, ApJS, 262, 3
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Sausalito, CA: Univ. Science Books)
Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, MNRAS, 189, 95
Pagel, B. E. J., Simonson, E. A., Terlevich, R. J., & Edmunds, M. G. 1992, MNRAS, 255, 325
Papovich, C., Simons, R. C., Estrada-Carpenter, V., et al. 2022, ApJ, 937, 22
Peimbert, M. 1967, ApJ, 150, 825
Pérez-Montero, E. 2017, PASP, 129, 043001
Pérez-Montero, E., Amorín, R., Sánchez Almeida, J., et al. 2021, MNRAS, 504, 1237
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pilyugin, L. S. 2001, A&A, 369, 594
Pilyugin, L. S., Mattsson, L., Vílchez, J. M., & Cedrés, B. 2009, MNRAS, 398, 485
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Rieke, M., Robertson, B., Tacchella, S., et al. 2023, Data from the JWST Advanced Deep Extragalactic Survey (JADES), STScI/MAST, doi:10.17909/8TDJ-8N28
Robertson, B. E. 2022, ARA&A, 60, 121
Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, ApJ, 914, 19
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, ApJL, 825, L23
Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2020, MNRAS, 491, 1427
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, ApJ, 962, 24
Sarkar, A., Chakraborty, P., Vogelsberger, M., et al. 2025, ApJ, 978, 136
Sarkar, A., Ferland, G. J., Chatzikos, M., et al. 2021, ApJ, 907, 12
Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, A&A, 665, L4
Shi, F., Zhao, G., & Liang, Y. C. 2007, A&A, 475, 409
Stasińska, G. 2002, RMxAA, 12, 62
Stasińska, G. 2005, A&A, 434, 507
Stiavelli, M., Morishita, T., Chiaberge, M., et al. 2025, ApJ, 981, 136
Storey, P. J., Sochi, T., & Badnell, N. R. 2014, MNRAS, 441, 3028
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Roberts-Borsani, G., Bradac, M., et al. 2022, ApJ, 935, 110
Yates, R. M., Schady, P., Chen, T. W., Schweyer, T., & Wiseman, P. 2020, A&A, 634, A107
Yuan, T. T., & Kewley, L. J. 2009, ApJL, 699, L161
Zaritsky, D., Kennicutt Robert, C. J., & Huchra, J. P. 1994, ApJ, 420, 87