The Astrophysical Journal, 972:113 (21pp), 2024 September 1 https://doi.org/10.3847/1538-4357/ad5ae1
© 2024. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

# Theoretical Strong-line Metallicity Diagnostics for the JWST Era

Prerak Garg[1], Desika Narayanan[1,2,3], Ryan L. Sanders[4,5], Romeel Davé[6,7,8], Gergö Popping[9], Alice E. Shapley[10], Daniel P. Stark[11], and Jonathan R. Trump[12]
[1] Department of Astronomy, University of Florida, 211 Bryant Space Sciences Center, Gainesville, FL 32611, USA; prerakgarg@ufl.edu
[2] University of Florida Informatics Institute, 432 Newell Drive, CISE Bldg E251, Gainesville, FL 32611, USA
[3] Cosmic Dawn Center at the Niels Bohr Institute, University of Copenhagen and DTU-Space, Technical University of Denmark, 2200 København N, Denmark
[4] Department of Physics and Astronomy, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA
[5] Department of Physics and Astronomy, University of Kentucky, 505 Rose Street, Lexington, KY 40506, USA
[6] Institute for Astronomy, Royal Observatory, Edinburgh, EH9 3HJ, UK
[7] University of the Western Cape, Bellville, Cape Town 7535, South Africa
[8] South African Astronomical Observatory, Cape Town 7925, South Africa
[9] European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748, Garching, Germany
[10] Department of Physics and Astronomy, University of California, Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA
[11] Steward Observatory, University of Arizona, 933 N Cherry Avenue, Tucson, AZ 85721, USA
[12] Department of Physics, 196 Auditorium Road, Unit 3046, University of Connecticut, Storrs, CT 06269, USA
*Received 2023 October 5; revised 2024 May 30; accepted 2024 June 20; published 2024 August 28*

## Abstract

The ratios of strong rest-frame optical emission lines are the dominant indicators of metallicities in high-redshift galaxies. Since typical strong-line-based metallicity indicators are calibrated on auroral lines at $z = 0$, their applicability for galaxies in the distant Universe is unclear. In this paper, we make use of mock emission-line data from cosmological simulations to investigate the calibration of rest-frame optical emission lines as metallicity indicators at high redshift. Our model, which couples the SIMBA cosmological galaxy formation simulation with CLOUDY photoionization calculations, includes contributions from H II regions, post-asymptotic-giant-branch stars, and diffuse ionized gas (DIG). We find mild redshift evolution in the 12 indicators that we study, which implies that the dominant physical properties that evolve in our simulations do have a discernible impact on the metallicity calibrations at high redshifts. When comparing our calibrations with high-redshift auroral line observations from the James Webb Space Telescope, we find a slight offset between our model results and the observations and find that a higher ionization parameter at high redshifts can be one of the possible explanations. We explore the physics that drives the shapes of strong-line metallicity relationships and propose calibrations for hitherto unexplored low-metallicity regimes. Finally, we study the contribution of DIG to total line fluxes. We find that the contribution of DIG increases with metallicity at $z \sim 0$ for singly ionized oxygen and sulfur lines and can be as high as 70%, making it crucial to include their contribution when modeling nebular emission.

*Unified Astronomy Thesaurus concepts:* Metallicity (1031); Galaxy evolution (594); High-redshift galaxies (734); H II regions (694); Hydrodynamical simulations (767)

## 1. Introduction

Estimating gas-phase metallicity is crucial in understanding the chemical evolution of galaxies across cosmic time. The metal content of galaxies is regulated by enrichment through stellar evolution and the inflow and outflow of gas to the intergalactic medium. Historically, nebular emission lines have been used as one of the primary methods for estimating galaxy metallicities. These lines were first discovered in 1864 by Sir William Huggins, in an early spectrum of the planetary nebula NGC 6543 (the Cat's Eye Nebula). In the spectrum, Huggins saw a set of strong green nebular emission lines at 4959 Å and 5007 Å. Unable to deduce their source, he proposed that they must be originating from an unknown element, which was termed nebulium. Over half a century later, Bowen (1927) found that these lines instead arose from the transition of doubly ionized oxygen (O++), and no new element was needed to explain them. It is now understood that many strong nebular lines primarily arise from the forbidden transition of ionized metal species.

Over the last century, significant effort has been expended to employ these nebular emission lines as diagnostics of the physical conditions in galaxies. One such important tool is the use of rest-frame optical nebular emission-line ratios to derive the ionized gas metallicity. The most accurate method of deriving metallicity in external galaxies is the so-called "direct" method, which uses temperature-sensitive auroral lines like [S II]$\lambda$4072, [O III]$\lambda$4363, [N II]$\lambda$5755, and [S III]$\lambda$6312 to derive the electron temperature ($T_e$; e.g., McLennan & Shrum 1925; Keenan et al. 1996), which, when combined with an ionization correction factor, can then be used to derive an accurate measurement of the ionized gas metallicity (Pagel et al. 1992; Izotov et al. 2006; Osterbrock & Ferland 2006; Pérez-Montero 2017). Recombination-based methods are even better, but they prove too challenging for the majority of extragalactic sources, due to the low luminosities of these emission lines that make it difficult to detect them past the local Universe. Numerous studies have used the "direct" method to derive the ionized gas-phase metallicity in galaxies (Bresolin 2007; Monreal-Ibero et al. 2012; Andrews & Martini 2013; Westmoquette et al. 2013; Maseda et al. 2014; Berg et al. 2015; Croxall et al. 2015, 2016; Mc Leod et al. 2015; Lin et al. 2017).

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Unfortunately, these lines are relatively weak, which makes them extremely hard to observe in all but the nearest galaxies; indeed, auroral line measurements have only recently been taken in about 60 galaxies at redshifts $z \gtrapprox 1$ (Villar-Martín et al. 2004; Yuan & Kewley 2009; Erb et al. 2010; Rigby et al. 2011; Brammer et al. 2012; Christensen et al. 2012a, 2012b; Bayliss et al. 2014; James et al. 2014; Stark et al. 2014; Steidel et al. 2014; Jones et al. 2015; Sanders et al. 2016, 2019, 2024; Kojima et al. 2017; Berg et al. 2018; Patrício et al. 2018; Gburek et al. 2019; Curti et al. 2022; Nakajima et al. 2023; Trump et al. 2023). The fact that almost half of these come from the James Webb Space Telescope (JWST) in the last year shows that it is still nontrivial to get large and representative auroral line measurements at $z \gtrapprox 1$, because current samples are still biased toward a high specific star formation rate (sSFR) relative to the galaxy main sequence.

To overcome this drawback, several studies have calibrated auroral-line-based metallicity measurements against ratios of bright rest-frame optical emission lines, involving [N II]$\lambda$6583, [O III]$\lambda$5007, [O II]$\lambda$3727, 29, [S II]$\lambda$6717, 31, Ne III]$\lambda$3869, H$\alpha$, and H$\beta$ (more commonly called "strong lines"; e.g., McCall et al. 1985; McGaugh 1991; Zaritsky et al. 1994; Kewley & Dopita 2002; Pettini & Pagel 2004; Pilyugin 2005; Maiolino et al. 2008; Dopita et al. 2016; Curti et al. 2017, 2020). These strong-line calibrations are now widely used for deriving metallicity measurements in galaxies where it is hard to observe the auroral lines. However, since these were calibrated using local auroral line observations, their applicability at high redshift remains in question, as the interstellar medium (ISM) ionization conditions might evolve with redshift (Steidel et al. 2016; Strom et al. 2018; Sanders et al. 2019; Shapley et al. 2019; Topping et al. 2020a).

Tackling this issue has now taken center stage, with the launch of JWST. Already in its first year, a number of studies (e.g., Arellano-Córdova et al. 2022; Curti et al. 2022; Katz et al. 2022; Schaerer et al. 2022; Tang et al. 2022; Taylor et al. 2022; Wang et al. 2022; Brinchmann 2023; Heintz et al. 2023; Langeroodi et al. 2023; Matthee et al. 2023; Nakajima et al. 2023; Rhoads et al. 2023; Sanders et al. 2023, 2024; Sun et al. 2023; Tacchella et al. 2023; Trump et al. 2023; Williams et al. 2023) have observed nebular emission lines from high-redshift galaxies using JWST/NIRSpec. The goal of this paper is to assess whether locally calibrated strong-line metallicity estimators apply in the physical conditions appropriate for high-$z$ galaxies. The paper is divided as follows. In Section 2, we describe our methodology. In Section 3, we compare our model calibrations against locally observed auroral-line-based calibrations from the literature. In Section 4, we look at the redshift evolution of strong-line metallicity calibrations, propose a new metallicity calibration, and look at the relative contribution of different ionizing sources. In Section 5, we compare our results in context to other relevant studies and go over some of our model caveats. In Section 6, we summarize our main conclusions. Throughout, we assume a Lambda cold dark matter cosmology, with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 68\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$

## 2. Model Description

### 2.1. Model Overview

To accurately estimate nebular emission in a galaxy, we need to precisely model a variety of different physical processes. For all the photoionization calculations, we make use of CLOUDY version 17.00 (Ferland et al. 2017). We model three main sources of ionizing radiation within the context of galaxy evolution simulations: H II regions, post-asymptotic-giant-branch (post-AGB) stars, and diffuse ionized gas (DIG). The H II regions and post-AGB stars are modeled as constant-density spherical shells with a fixed escape fraction. The escaping ionizing photons then ionize the surrounding gas particles, which give rise to DIG, modeled as gas cells with a plane-parallel geometry. In the following subsections, we describe in detail our modeling strategy for each of these sources, our assumptions, and the model parameters we use for each of them.

### 2.2. SIMBA Simulations

We first simulate galaxies using the SIMBA hydrodynamic cosmological galaxy formation simulation (Davé et al. 2019). SIMBA is an updated version of the MUFASA hydrodynamic simulation (Davé et al. 2016), and uses the GIZMO hydrodynamic solver in the meshless finite-mass mode. SIMBA uses the GRACKLE-3.1 library (Smith et al. 2017) to calculate photoionization and nonequilibrium radiative cooling from H, He, and metals. A self-shielding prescription from Rahmati et al. (2013) is applied to a uniform ionizing background (Haardt & Madau 2012). Star formation occurs in $H_2$ gas, employing a volumetric Schmidt–Kennicutt relation (Kennicutt 1998), which is calculated using the Krumholz et al. (2009) subresolution model for computing the $H_2$ gas fraction, with some modifications (see Davé et al. 2016). The stellar feedback is modeled as a two-phased decoupled kinetic outflow with a 30% hot component (Davé et al. 2016). The mass-loading factors in stellar winds are taken from calibrations from the Feedback In Realistic Environments (FIRE) simulations as measured by Anglés-Alcázar et al. (2017b). A friends-of-friends algorithm is used to seed black holes in galaxies. If a galaxy reaches a stellar mass $(M_*) > 10^{9.5} M_\odot$ and it does not contain a black hole, then the star particle closest to the center is converted to a black hole. This stellar mass threshold was chosen based on studies involving the Evolution and Assembly of GaLaxies and their Environments (EAGLE; Bower et al. 2016; McAlpine et al. 2018), FIRE (Anglés-Alcázar et al. 2017c), and RAMSES-based (Dubois et al. 2015; Habouzit et al. 2017) simulations, which found that black hole growth is suppressed in low-mass galaxies primarily due to stellar feedback. Black hole growth is driven via two accretion modes: Bondi accretion (Bondi & Hoyle 1944), only from the hot halo, and torque-driven cold accretion (Anglés-Alcázar et al. 2017a).

One important feature that differentiates SIMBA from other cosmological simulations is that the model includes an on-the-fly dust production and destruction model. Dust is produced via Type II supernovae and AGB stars, whereas dust can either be destroyed via thermal sputtering and shocks or can be consumed during star formation. Li et al. (2019) demonstrated that this model can successfully match the observed dust-to-gas and dust-to-metals ratios at low- and high-$z$, as well as the cosmic evolution of $\Omega_{\rm dust}$ (Péroux & Howk 2020). Lovell et al. (2021) showed that it can also match the submillimeter galaxy abundances pretty well. Having a sophisticated model of dust production and destruction is important in accurately modeling nebular emission, since some metals are locked up in dust and are therefore not available for ionization. Beyond these key

dust-based scaling relations, overall, SIMBA has been shown to match a number of major observed galaxy scaling relations, including stellar mass functions from $z = 0$–6 and gas and stellar mass–metallicity relations at $z = 0$–2 (Davé et al. 2019). This lends confidence in our ability to use SIMBA for this study.

We run a total of four SIMBA boxes each with a different random seed until $z \sim 0$, with a simulation box size of side length $25\,h^{-1}$ Mpc with $512^3$ particles, giving us a baryon resolution of $2.3 \times 10^6\,M_\odot$. The snapshots from SIMBA are first processed through the CAESAR[13] galaxy catalog generator (Thompson 2014). Once we have identified all the galaxies in a snapshot, we filter out those that do not have at least one star younger than 10 Myr. Enforcing this criterion ensures that our sample galaxies always have some young stars—the major contributor to the ionizing flux in our model. With a star particle mass of $2.3 \times 10^6\,M_\odot$, this means that a galaxy has to have an average star formation rate of $0.23\,M_\odot\,\mathrm{yr}^{-1}$ to have a reasonable probability of having one such star particle. Many small (but still resolved) main-sequence galaxies fall below this. Thus, due to the stochastic nature of the star formation algorithm in SIMBA, this will exclude a lot of small galaxies that actually should have emission lines. We find that the potentially star-forming galaxies that we exclude from our sample due this cutoff are about 5%–10% of those at $M_* > 9.5$, increasing to around 60%–80% at $M_* < 8.0$. This biases our results toward high-mass galaxies and acts as an additional source of uncertainty in our model (see Section 5.2 for more information).

### 2.3. Photoionization Modeling

The nebular emission model that we use here is an updated one of that presented in Garg et al. (2022), with a few notable changes. In particular, while Garg et al. (2022) only considered nebular emission from H II regions around young stars, we have now expanded the model to also include contributions of ionized gas around post-AGB stars and DIG. In the following subsections, we describe in detail our modeling parameters and how we model each of the three sources of nebular emission (H II regions, DIG, and the regions surrounding post-AGB stars).

#### 2.3.1. Stellar Model Parameters

The mass resolution of our parent simulations is of the order of $10^6\,M_\odot$. Because this is larger than the typical mass of observed clusters (Chandar et al. 2014, 2016; Linden et al. 2017; Larson et al. 2020), we employ a subresolution model for breaking these star particles into smaller subunits:

*Cluster Mass Distribution*—if we were to assume a single a H II region around our massive $2.3 \times 10^6\,M_\odot$ clusters, this would result in an overestimation of the ionizing flux per unit H II region. Instead, we consider each star particle in our simulation to represent a collection of unresolved clusters with individual H II regions. We subdivide each star particle into a collection of smaller-mass star particles with the same age and metallicity as the parent star particles, but their masses are derived from a cluster mass distribution of the form:

$$\frac{dN}{dM} \propto M^{\beta}, \tag{1}$$

[13] https://github.com/dnarayanan/caesar

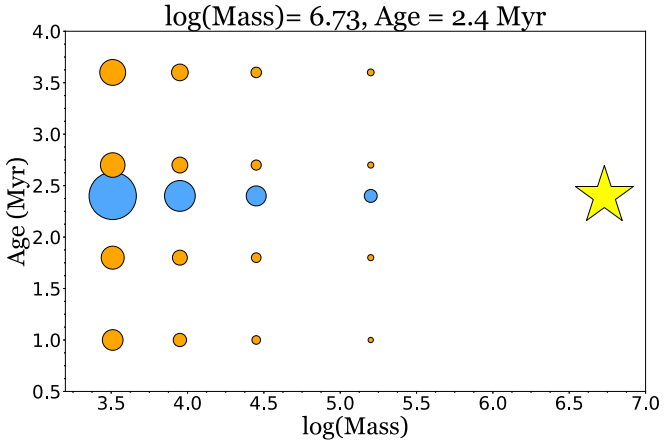


**Figure 1.** An example of our methodology for dividing star particles into a collection of subparticles with varying age and mass based on power-law distributions. The parent star particle (yellow star) has a mass of $10^{6.73}\,M_\odot$ and an age of 2.4 Myr. The particle is first subdivided into four mass bins (blue circles), and each of these subparticles is then further subdivided into four age bins (orange circles). The relative area of the circle represents how many stars there are in each bin. In the end, this one star particle is broken down into 400 star particles, spread across 16 bins of varying ages and masses.

where $M$ is the stellar mass of the star particle, $N$ is the number of star particles, and $\beta$ is the power-law index. While observations have indicated power-law-like mass distributions for observed clusters, there is no consensus yet as to the power-law index, $\beta$. Observations suggest a range from $-2.2 < \beta < -1.7$ (Chandar et al. 2014, 2016; Linden et al. 2017; Larson et al. 2020). We therefore assume $\beta = -2.0$. We subdivide each star particle into four bins, with a mass range from $10^{3.5}$–$10^{5.0}\,M_\odot$. The distribution is done in such a way that the total mass of all the subparticles sums to the mass of the parent star particle, ensuring that the total mass is conserved.

*Age Distribution*—since clusters within the parent particle that was subdivided may have a distribution of ages, we further subdivide each of the smaller-mass star clusters into multiple individual particles with varying ages, based on a power-law distribution function of the form:

$$\frac{dN}{d\tau} \propto \tau^{\gamma}, \tag{2}$$

where $\tau$ is the age of the star particle, $N$ is the number of star particles, and $\gamma$ is the power-law index. Observations by Chandar et al. (2014, 2016) suggest that $\gamma$ can vary between $-0.55$ and $-0.74$, with only a modest dependence on cluster mass. We therefore fix $\gamma = -0.65$. We divide all the star particles with ages between 2 and 10 Myr into four bins. The age distribution is constructed to ensure that the mean age of all the subparticles is as close to the age of the parent particle as possible. We apply a threshold of 0.1 Myr and if the difference between the average age and the age of the parent star particle is above the threshold, then the width (the difference between the start and end age of the distribution) is decreased at each step until the threshold is met. This is done 100 times at most, and if we fail to meet the threshold even after that then the particle is not broken down.

As an example, in Figure 1, we show how a single star particle of a mass of $10^{6.73}\,M_\odot$ and an age of 2.4 Myr is

subdivided into multiple star particles with varying mass and age, based on the power-law distributions described above. The parent star particle (yellow star) is first broken down into four mass bins, shown by the blue circles, with their relative areas representing how many particles are there in each bin. The particles in each of those mass bins (blue circles) are then further subdivided into four age bins, based on the age distribution function, shown by the orange circles. Thus, in the end, one star particle gets broken down into ∼400 particles spread over 16 age and mass bins.

*Spectral Library and Isochrones*—for all the analysis in this paper, unless otherwise specified, we make use of binary model grids from the Binary Population and Spectral Synthesis (BPASS) library (Eldridge et al. 2017; Stanway & Eldridge 2018), with a Chabrier initial mass function (IMF) and a 100 $M_\odot$ cutoff.

#### 2.3.2. Sources of Ionizing Radiation

*H II Regions*—H II regions around young stars are the primary sources of ionizing radiation in a star-forming galaxy. To model H II regions, we first take all the star particles younger than 10 Myr and subdivide them based on the cluster mass and age distribution functions, as described above. We apply this age cutoff to ensure that only stars that produce an adequate amount of ionizing flux are considered in our calculation (for computational efficiency).

We model H II regions as spherical shells around all the individual subparticles of interest and compute the emergent nebular line emission from these H II regions using CLOUDY v 17.00 (Ferland et al. 2017). The model inputs for the CLOUDY photoionization calculations are the shape of the ionizing radiation field, the ionizing source luminosity, the gas-phase abundance, the hydrogen density, and the inner radius. To compute the ionizing radiation field, the spectral energy distributions (SEDs) for all subdivided star particles are computed using Flexible Stellar Population Synthesis (FSPS;[14] Conroy et al. 2009; Conroy & Gunn 2010). We model each star particle as a simple stellar population (SSP) and set the stellar metallicity to log(Fe/H) rather than the total metal content by mass, so as to mimic the effects of $\alpha$-enhancement (Garg et al. 2022). The age of the star is set based on the age returned from the cosmological simulation.

The ionizing flux for each star particle is computed via

$$Q = M_* \int_{\nu_0}^{\infty} \frac{L_\nu}{h\nu} d_\nu, \tag{3}$$

where $L_\nu$ is in units of ergs $s^{-1}$ $Hz^{-1}$ and is taken directly from the FSPS SED, $M_*$ is the stellar mass, and $Q$ is the rate of ionizing photons in units of inverse seconds. The gas-phase abundances are assumed to be the same as that of the star particle and are taken directly from the simulation. SIMBA keeps track of the abundances of 10 elements (He, C, N, O, Ne, Mg, Si, S, Ca, and Fe) apart from hydrogen, and for all the elements except nitrogen, we make use of the abundances tracked by the simulation. Nitrogen production is fairly complicated to model, since it can be produced via primary or secondary nucleosynthesis and SIMBA does a fairly poor job of matching the observed (N/O) versus (O/H) relation from Pilyugin et al. (2012). To correct for this, we set the nitrogen abundance in our model according to Equation (4). This relation was proposed by Pilyugin et al. (2012), based on the ion abundances calculated using auroral line measurements from 112 H II regions in nearby galaxies. The auroral line fluxes were used to derive the electron temperatures, which were then converted to nitrogen and oxygen abundances, giving us the relation shown below (Equation (4); see Garg et al. 2022 and Hough et al. 2023 for more information):

$$\begin{aligned} \log(\mathrm{N/O}) &= -1.493 \\ &\quad \text{for } 12 + \log(\mathrm{O/H}) < 8.14, \\ &= 1.489 \times [12 + \log(\mathrm{O/H})] - 13.613 \\ &\quad \text{for } 12 + \log(\mathrm{O/H}) > 8.14. \end{aligned} \tag{4}$$

We assume fixed physical conditions for all H II regions in our simulations, irrespective of redshift. This includes an assumed hydrogen density of $n_\mathrm{H} = 30\,\mathrm{cm}^{-3}$ and an escape fraction of 0.4. This is based on studies of resolved galaxies such as Choi et al. (2020) and Della Bruna et al. (2021), which have estimated the escape fraction for H II regions to be anywhere between 25% and 65%. We add dust in our CLOUDY model as graphite and silicate grains with size distributions and abundances appropriate for those along the line of sight to the Trapezium stars in Orion, which are then scaled based on the total metallicity of the star particle.

While the density and escape fraction remain fixed for H II regions in our models, the size and ionization parameter for each H II region are allowed vary between each star particle in the following way. The inner radius is set to be the Strömgren sphere radius ($R_\mathrm{S}$), given by the following equation:

$$R_\mathrm{S} = \left( \frac{3Q}{4\pi\, {n_\mathrm{H}}^2\, \alpha_B} \right)^{1/3}, \tag{5}$$

where $Q$ is the rate of ionizing photons, $n_\mathrm{H}$ is the hydrogen density, and $\alpha_B$ is the case B recombination coefficient at $10^4$ K, which is $\alpha_B = 2.5 \times 10^{-13}\,\mathrm{cm}^3\,\mathrm{s}^{-1}$. We assume the escape fraction to be 0.4. Since we are running CLOUDY on the fly, which is very computationally intensive, we stop the CLOUDY calculation when either the temperature drops below 100 K or if the ionizing fraction drops below 0.1. These stopping conditions were chosen so as to minimize the computation time while having a minimal effect on the final result.

*Post-AGB Stars*—post-AGB stars are stars with initial masses between 0.8 and 8 $M_\odot$ that have left the asymptotic giant branch and are evolving horizontally along the Hertzsprung–Russell diagram. While relatively short-lived, post-AGB stars can become hot enough to ionize their ejected outer envelope, creating planetary nebulae. Byler et al. (2019) showed that post-AGB stars can be an important ionizing source for the large amounts of DIG emission seen in, e.g., early-type galaxies. As discussed in Section 4.2, we find that although post-AGB stars themselves do not contribute significantly to the total line emission budget, they do act as an important ionizing source for the DIG in the galaxy. For modeling post-AGB stars, we follow the same prescription as what was used for H II regions, with a few minor changes. The gas ionized by the post-AGB stars is primarily the ejected outer envelopes of the parent star, which are composed of gas that was enriched by stellar nucleosynthesis over the lifetime of the

[14] https://github.com/cconroy20/fsps

star. Thus, we can no longer assume that the stellar metallicity is equal to the gas metallicity in the case of post-AGB stars.

Stasińska et al. (1998), Stanghellini et al. (2005, 2009), and Henry et al. (2018) have analyzed planetary nebulae in the Milky Way, the Large Magellanic Cloud, and the Small Magellanic Cloud and found that their log(N/O) and log(C/O) was elevated when compared with local observations of H II regions, by about 0.8 and 0.6 dex on average, respectively. Taking a conservative approach, to account for the enrichment, we enhance both the log(C/O) and log(N/O) for post-AGB stars by 0.4 dex in our model. We also assume a higher escape fraction of 0.6 for post-AGB stars, to take into account the fracturing of the gas surrounding the star owing to the stellar evolution.

*DIG*—the DIG component of the ISM can be a major contributor to emission-line fluxes (Sanders et al. 2017; Vale Asari et al. 2019). The ionizing radiation from sources such as H II regions and post-AGB stars can leak out and ionize the surrounding diffuse ISM. To accurately model DIG, we need an estimate of the escape fraction of ionizing photons. Recent studies of resolved galaxies, such as Choi et al. (2020) and Della Bruna et al. (2021), have estimated the escape fraction for H II regions to be anywhere between 25% and 65%. We assume a constant escape fraction of 40% for H II regions and an escape fraction of 60% for post-AGB stars, to take into account the potential fracturing of the ISM over time, due to stellar evolution.

While we would ideally compute the ionization state in DIG by employing bona fide ionizing radiative transfer, this is computationally intractable for the large numbers of galaxies that we aim to simulate in a cosmological context. As a result, we approximate the emission from DIG in the following manner. We first model the shape of the incident ionizing radiation field from nearby ionizing sources (H II regions and post-AGB stars) and then scale the flux from this SED to the total energy deposited in the cell from dust radiative transfer. In detail, we first compute the emitting SED from nearby H II regions and post-AGB stars (within 1 kpc), using the aforementioned CLOUDY calculations. With this, we compute the ionization parameter via

$$U = \frac{Q/6}{4\pi\, n_{\rm H}\, c_{\rm W}{}^2\, c}. \quad (6)$$

Here, $Q$ is the rate of ionizing photons in inverse seconds incident on the cell. Since there are six faces of a cubic cell, we divide this number by six to get the rate of ionizing photons striking each gas cell surface. $c_{\rm W}$ is the cell width, $c$ is the speed of light, and $n_{\rm H}$ is the hydrogen density, which is kept fixed at 10 cm$^3$ for all the gas cells across redshifts. To compute $Q$, we run 3D radiative transfer with the dust radiative transfer code HYPERION[15] (Robitaille 2011; described in more detail shortly). HYPERION only computes the radiative transfer longward of the Lyman Limit, so we use this calculation simply to compute the total energy deposited in every cell (which is dominated by far-UV photons close to the Lyman limit) and scale the shape of the incident SED from the CLOUDY calculations to match this energy. We only run CLOUDY on those cells where log$U$ is greater than $-6.0$, to minimize the computation time. The metallicity of the CLOUDY model is assumed to be the same as that of the gas cell. While this is an approximation to the true incident ionizing flux in a cell of diffuse ISM, fully computing the ionizing radiative transfer in large cosmological volumes for numerous snapshots is computationally intractable.

In detail, we have packaged all of the above processes into the radiative transfer package POWDERDAY (Narayanan et al. 2021),[16] which acts as a high-level wrapper to calculate nebular emission. The general approach that we use is shown via a flowchart in Figure 2. We first run the SIMBA snapshots through the CAESAR galaxy catalog generator. We filter out those galaxies that do not have at least one star particle younger than 10 Myr. This is done to ensure that we are only considering star-forming galaxies for this study. Then, for each galaxy, we take all the star particles less than 10 Myr old and subdivide them based on a cluster mass and age distribution, then assign an SSP using PYTHON-FSPS[17] (Johnson et al. 2021), which is a Python wrapper for FSPS. Each of the subdivided star particles is modeled as an H II region with a spherical shell geometry and a constant density. Once we are done modeling the nebular emission from H II regions, we move on to post-AGB stars. For this, we take all the star particles older than 1 Gyr for $z = 0$ and 1 and older than 100 Myr for $z = 2$ and above, then break them down using a cluster mass and age distribution function. These age limits are used to minimize computation time, all the while ensuring that we capture as much of the old stellar population as possible. These subdivided star particles are then modeled as post-AGB stars, as described above. Once we have finished running the CLOUDY models for all the H II regions and post-AGB stars, we then use HYPERION to do the dust radiative transfer calculation, and this gives us the total energy deposited in each gas cell. We then calculate the shape of the ionizing radiation field for each gas cell by taking a distance-weighted average of the output CLOUDY spectrum of all the H II regions and post-AGB stars that lie within 1 kpc. We filter out gas cells where the log of the ionization parameter is less than $-6.0$ to minimize computation and then run CLOUDY on the remaining gas cells. After adding in the DIG sources, we perform the final dust radiative transfer calculation, and that gives us the line-of-sight SED with nebular emission for each galaxy in our sample. Finally, the line fluxes are extracted by fitting Gaussian profiles. We calculate the metallicity of each galaxy as the [O III]$\lambda$5007-weighted average of the stellar metallicity of all the star particles that emit nebular emission. Throughout this paper, unless otherwise mentioned, we report metallicity in units of $12 + \log({\rm O/H})$. The notation that we adopt for the various emission-line ratios used in this paper is as follows:

***N2***: log([N II]$\lambda$6583/H$\alpha$);
***R3***: log([O III]$\lambda$5007/H$\beta$);
***R2***: log([O II]$\lambda$3727, 29/H$\beta$);
***R23***: log(([O II]$\lambda$3727, 29 + [O III]$\lambda$4959, 5007)/H$\beta$);
***O3N2***: log(([O III]$\lambda$5007/H$\beta$)/ ([N II]$\lambda$6583/H$\alpha$));
***R3N2***: log([O III]$\lambda$5007/[N II]$\lambda$6583);
***N2O2***: log([N II]$\lambda$6583/[O II]$\lambda$3727);
***O3O2***: log([O III]$\lambda$5007/[O II]$\lambda$3727, 29);
***Ne3O2***: log([Ne III]$\lambda$3869/[O II]$\lambda$3727, 29);
***S2***: log([S II]$\lambda$6717, 31/H$\alpha$);
***O3S2***: log(([O III]$\lambda$5007/H$\beta$)/([S II]$\lambda$6717, 31/H$\alpha$)); and

[15] http://www.hyperion-rt.org/

[16] https://powderday.readthedocs.io/en/latest/

[17] http://dfm.io/python-fsps/current/

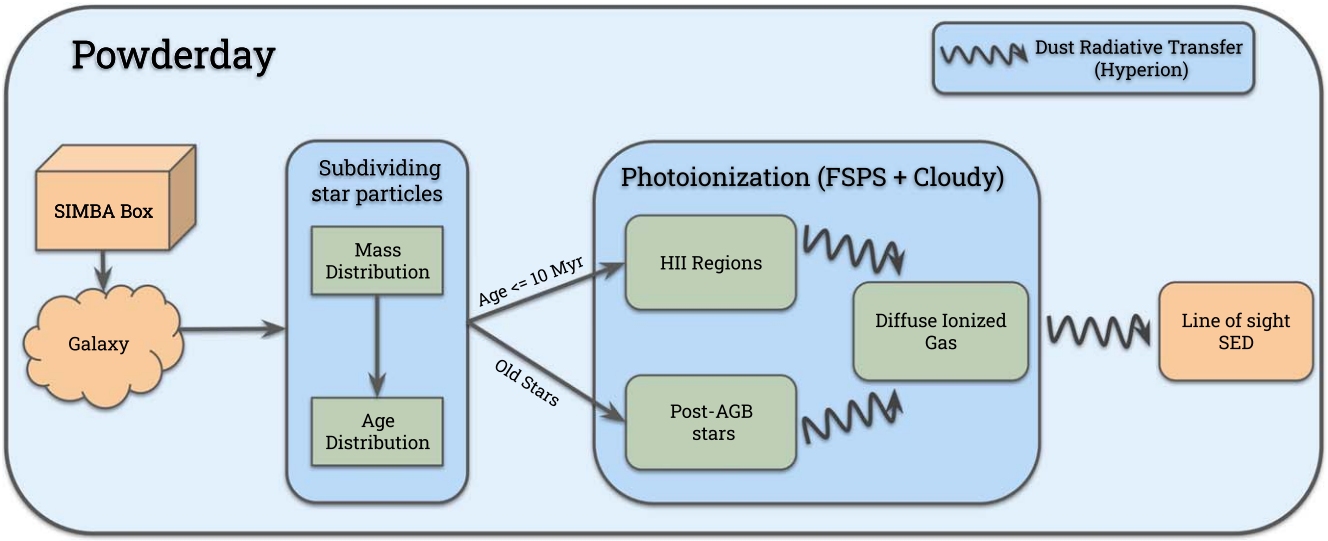


**Figure 2.** Flowchart showing how our nebular emission modeling pipeline integrated within POWDERDAY works. The first step is to identify galaxies within the SIMBA box, using the CAESAR galaxy catalog generator. We then subdivide the star particles into a subresolution mass and age distribution. Star particles younger than 10 Myr are modeled as an H II region and old stars are modeled as post-AGB stars. We then run the first pass of dust radiative transfer using HYPERION to get the energy deposited in each cell and calculate the ionizing radiation flux for each gas particle within 1 kpc of an H II region or post-AGB star. We then apply a filter and only model gas cells with sufficiently high ionizing radiation as DIG. After this, a second pass of dust radiative transfer is performed, giving us the line-of-sight SED with nebular emission for each galaxy.

***N2S2***: $\log(([\mathrm{N\,II}]\lambda6583/[\mathrm{S\,II}]\lambda6717,\ 31) + 0.264([\mathrm{N\,II}]\lambda6583/\mathrm{H}\alpha))$.

## 3. Local Strong-line Metallicity Calibrations

Before we can use our model to extend the strong-line metallicity calibrations to high-redshift galaxies, we first ensure that our model can accurately reproduce the locally observed auroral-line-based metallicity calibrations from the literature. In Figure 3, we compare the theoretical strong-line metallicity calibrations derived using the SIMBA $z = 0$ snapshot against different local observational and theoretical calibrations from the literature. We consider all 12 modeled line ratios and perform a fit of a third-degree polynomial of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$—where $x$ is the line ratio and $y$ is the metallicity $(12 + \log(\mathrm{O/H}))$—to the data, using an ordinary least squares regression, represented by the thick red line. The red shaded area represents the region of the 90% confidence interval, while the other solid color lines are the observational auroral-line-based strong-line calibrations. The dashed lines are theoretical calibrations that were used to extend the auroral line calibrations in the regime where no observational data were present. As can be seen, our model provides a reasonable match to the observed calibrations for the majority of the line ratios. We now discuss the individual line ratios in turn.

*N2*—we compare our calibration fit for N2 against the calibrations from Denicoló et al. (2002), Kewley & Dopita (2002), Pettini & Pagel (2004), Tremonti et al. (2004), Nagao et al. (2006), Maiolino et al. (2008), Marino et al. (2013), Brown et al. (2016), and Curti et al. (2020). As can be seen in Figure 3, our model performs well in matching the calibrations from the literature across the entire metallicity range. The nature of the curve can be understood by the fact that as the metallicity decreases, the H II region temperature rises, due to a decrease in metal cooling. This causes more of the nitrogen to occupy higher ionization states. This, coupled with the decreasing abundance of nitrogen with decreasing metallicity, leads to [N II]$\lambda$6583 being a monotonically decreasing function of metallicity.

*R3, R2, and R23*—these line ratios are all double-valued functions of metallicity. This is because, at high metallicities, the temperatures become too low to collisionally excite the valence electrons of oxygen ions into higher electron energy levels. As the metallicity decreases, the increase in ionization parameter and changes in the ionization spectral shape/hardness lead to more O++ emission, and as a result, the line ratios increase. We can see that [O II]$\lambda$ 3727, 29/H$\beta$ reaches a peak around a metallicity of $12 + \log(\mathrm{O/H}) \sim 8.5$, whereas [O III]$\lambda$5007/H$\beta$ reaches a peak at a lower metallicity of $12 + \log(\mathrm{O/H}) \sim 8.0$, due to it being a doubly ionized species. As we move to lower metallicities, the decrease in the amount of oxygen available to excite becomes the primary limiting factor and we see that the line ratios start to decrease, giving rise to the double-valued shape. In Figure 3, we compare our calibration for R2, R3, and R23 against the calibrations from Kewley & Dopita (2002), Tremonti et al. (2004), Nagao et al. (2006), Maiolino et al. (2008), Jones et al. (2015), and Curti et al. (2020). In general, for R3 and R23, our calibrations underestimate the line ratio at lower metallicities $(12 + \log(\mathrm{O/H}) < 8)$, but the overall nature of the curve is similar to the literature calibrations. As for R2, our calibrations work well at low metallicities, but are unable to reach high enough values at high metallicities $(12 + \log(\mathrm{O/H}) > 8)$. As discussed in Section 4.2 (Figure 5), we find that DIG is a substantial contributor to the total line emission budget of [O II]$\lambda$3727, 29 at high metallicities $(12 + \log(\mathrm{O/H}) > 8)$, accounting for as much as 70% of the total line emission budget. As such, the mismatch at high metallicities might be explained by the uncertainties in our DIG modeling.

*O3N2 and R3N2*—we compare our calibrations for O3N2 and R3N2 against the observed auroral line and theoretical calibrations from Kewley & Dopita (2002), Pettini & Pagel (2004), Tremonti et al. (2004), Nagao et al. (2006), Maiolino et al. (2008), Marino et al. (2013), Brown et al. (2016), and

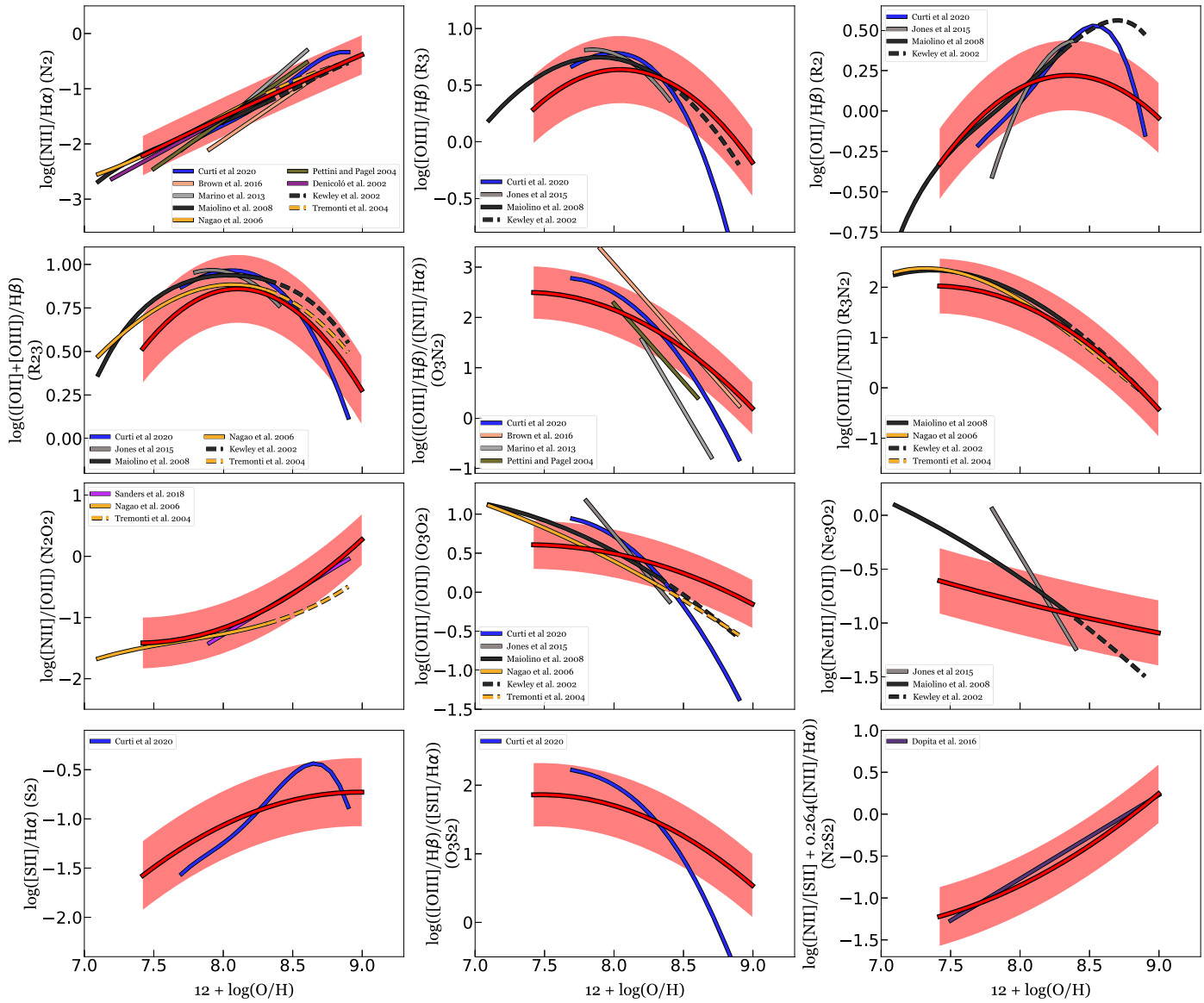


**Figure 3.** A comparison between the theoretical strong-line metallicity calibrations for the 12 line ratios using our model on the SIMBA z = 0 galaxies against different local observational and theoretical calibrations from the literature (Denicoló et al. 2002; Kewley & Dopita 2002; Pettini & Pagel 2004; Tremonti et al. 2004; Nagao et al. 2006; Maiolino et al. 2008; Marino et al. 2013; Jones et al. 2015; Brown et al. 2016; Dopita et al. 2016; Sanders et al. 2018; Curti et al. 2020). In all the panels, the thick red line shows the calibration from SIMBA and the red shaded region shows the 90% confidence interval. We fit polynomials of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$, where $x$ is the line ratio and $y$ is the metallicity ($12 + \log(\mathrm{O/H})$). The fits for all the 12 line ratios at z = 0 are listed in Table 1. The metallicity calibrations from the literature that were calibrated using auroral line measurements are shown by the solid colored lines, whereas the dashed lines represent the theoretical strong-line calibrations that were used to extend them in the regime where no observational data were available.

Curti et al. (2020). We find that for R3N2, our calibrations are an excellent match to the observations, whereas for O3N2, we do a reasonable job at low metallicities, but at high metallicities ($12 + \log(\mathrm{O/H}) > 8.5$), we tend to somewhat overestimate the line ratio. That said, there is significant uncertainty in the observational calibrations themselves. Both of these calibrations are decreasing functions of metallicity, with decreasing slopes at lower metallicities. This is because as we move to metallicities lower than ∼8.0 in units of $12 + \log(\mathrm{O/H})$, the [O III]$\lambda$5007 starts to decrease, which causes the curve to flatten out and even to eventually move to a downward trend at even lower metallicities.

*N2O2*—we compare our calibration for N2O2 against the calibrations from Tremonti et al. (2004), Nagao et al. (2006), and Sanders et al. (2018). Note that the two observed calibrations do not agree among themselves, because Sanders et al. (2018) corrected for DIG contribution, whereas Nagao et al. (2006) did not. As a result of this, the calibration from Nagao et al. (2006) flattens out at low metallicity, whereas there is no such flattening observed in the calibration from Sanders et al. (2018). On the other hand, at high metallicities, the calibration from Sanders et al. (2018) has a much higher slope than the calibration from Nagao et al. (2006). We find that on the high-metallicity end, we find a close match to the calibrations from Sanders et al. (2018). We would like to note that it is somewhat surprising that our models that include DIG provide such a good match to Sanders et al. (2018), who, as mentioned above, corrected the Sloan Digital Sky Survey calibration by removing the contribution from DIG. On the low-metallicity end ($12 + \log(\mathrm{O/H}) < 8$), we find that our

model closely matches the calibrations presented in Nagao et al. (2006). This resemblance is observed in the flattening of the curve, as the [O II] emission line exhibits a turning point and begins to decrease at lower metallicities.

*O3O2 and Ne3O2*—we compare our calibrations for O3O2 and Ne3O2 against the calibrations from Kewley & Dopita (2002), Tremonti et al. (2004), Nagao et al. (2006), Maiolino et al. (2008), Jones et al. (2015), and Curti et al. (2020). As stated earlier, we tend to underestimate R3 at low metallicities ($12 + \log(\mathrm{O/H}) < 8$), while at high metallicities ($12 + \log(\mathrm{O/H}) > 8.5$), we underestimate R2. The net result of this is that we find that our calibration for O3O2 is flatter. Similarly, we find that our calibration for Ne3O2 also has a lower slope than the observed calibrations. That said, there is a large variation in slope across the different observed calibrations for both the line ratios as well, though none are as flat as the simulations. This mismatch with observations may potentially be attributed to the uncertainties in neon abundance yields as implemented in SIMBA, and further exacerbated by uncertainties in the metal deposition rates onto dust grains. Further, this flatness can be be explained by the fact that although we find that an anticorrelation relation between log$U$ and O/H exists in our models, it is a lot flatter than what is reported in some other studies (e.g., Perez-Montero 2014).

*S2 and O3S2*—we find that our calibration for S2 matches the observed calibration from Curti et al. (2020) fairly well at low metallicities ($12 + \log(\mathrm{O/H}) < 8.2$). The overall nature of the curve is much flatter than the observed calibration, which means that at high metallicities ($12 + \log(\mathrm{O/H}) > 8.2$), our calibration does not reach values as high as those observed. Similarly, for O3S2, we find that our calibration has the same general nature, with the line ratio increases with decreasing metallicity, albeit with a flatter slope, which can be attributed to the underestimation of [S II]$\lambda$6717, 31 at high metallicities. The mismatch at high metallicities might be explained by the uncertainties in modeling the DIG emission (see Section 5.2 for more information). Similar to [O II]$\lambda$3727, 29, we find that DIG can account for as much as 70% of the total ionizing flux of [S II]$\lambda$6717, 31 at high metallicities (Figure 5).

*N2S2*—in the last panel of Figure 3, we compare our calibration for N2S2 against the calibration from Dopita et al. (2016). We can see that our model performs reasonably well at matching the observed calibration, although we do see slight deviations at the extreme ends.

In summary, we find that our model does a fairly reasonable job of reproducing the general natures of the curves for most of the line ratios. We can match the observed auroral line calibrations fairly well at low metallicities. However, for line ratios like R2, S2, and O3S2, which involve singly ionized oxygen and sulfur lines, our model underperforms at high metallicities ($12 + \log(\mathrm{O/H}) > 8$), and the fact that DIG dominates the line emission budget for these lines at high metallicities, coupled with the fact that our current model is unable to accurately capture that emission, might explain the mismatch.

## 4. Strong-line Metallicity Calibrations across Redshifts

In the previous section, we discussed how our metallicity calibrations fare against redshift $z \sim 0$ constraints anchored in auroral-line-based metallicity indicators. We now proceed in applying our model to high-redshift galaxies in order to study and understand the redshift evolution of strong-line metallicity indicators. We remind the reader that the elements of our model that vary with redshift are:

1. Stellar and gas metallicities;
2. Spectra of ionizing photons in H II regions and post-AGB stars;
3. The incident radiation field for calculating the DIG emission;
4. The number of young and post-AGB stars; and
5. The amount of gas available for ionization.

In Figure 4, we plot the 12 line ratios we consider at integer redshifts from $z = 0$–5 (a more detailed version of this figure, with the 90% confidence intervals plotted for each line ratio across redshifts, is available in Appendix B: Figure 7). In short, we find mild redshift evolution for all of the line ratios that we model, aside from the fact that the majority of high-$z$ galaxies lie in a lower-metallicity regime of each line-ratio metallicity diagnostic. This suggests that the changes in stellar and gas metallicities, incident radiation fields, and the number of star and gas particles with redshift do have a discernible effect on the strong-line calibrations, causing them to vary from low to high redshift in our model. Recent JWST observations of auroral line galaxies spanning a redshift range from ∼2–9 (Curti et al. 2022; Trump et al. 2023; Nakajima et al. 2023; Sanders et al. 2024) also indicate that the metallicity indicators might evolve with redshift. We discuss this in detail, along with the uncertainties and caveats of our model, in Section 5.

An important point to note is that the utility of the line ratios changes, depending on what metallicity range or redshift one is interested in. This, in turn, affects the applicability of the line ratios across redshifts. For example, the line ratio N2S2 (Figure 4, last panel) is a monotonically increasing function of metallicity at high metallicity/low redshifts, but becomes a double-valued function of metallicity at low metallicities/high redshifts. Thus, a line ratio that works well at low redshifts might not be an appropriate choice for high-redshift galaxies. In the following subsections, we will look at the 12 line ratios in detail and explore the general nature of the curves. We will mainly focus on exploring the low- and high-metallicity regimes and discussing where different metallicity indicators work well and the regimes where they must be avoided.

### 4.1. Discussion of Current Standard Calibrators

*N2*—we find that logN2 varies by about 0.2–0.4 dex from redshift 0 to 5 and that the variation decreases with metallicity. However, we would like to point out that this variation is smaller than the 90% confidence intervals at $z = 0$ that span a width of about 0.7 dex (see Figure 3). N2 is a reasonably robust metallicity indicator across all redshifts. N2 is a single-valued function with a steep slope, which makes it an excellent indicator over the whole metallicity range, except at low metallicities ($12 + \log(\mathrm{O/H}) < 7.0$), where the slope starts to flatten out. Nonetheless, at low metallicities (common at high-$z$), it is very challenging to detect [N II]$\lambda$6583 and it can even be weaker than the [O III]$\lambda$4363 auroral line in many cases at these redshifts. Thus, in practice, not many very-high-$z$ galaxies ($z > 4$) have [N II]$\lambda$6583 detections in JWST spectroscopy, which limits its utility (see Sanders et al. 2023 and Shapley et al. 2023b for more information). Further, it is important to recall a significant caveat concerning the line ratios involving [N II]$\lambda$6583 in our model. SIMBA is unable to reproduce the observed log(N/O) versus log(O/H) relation in nearby galaxies

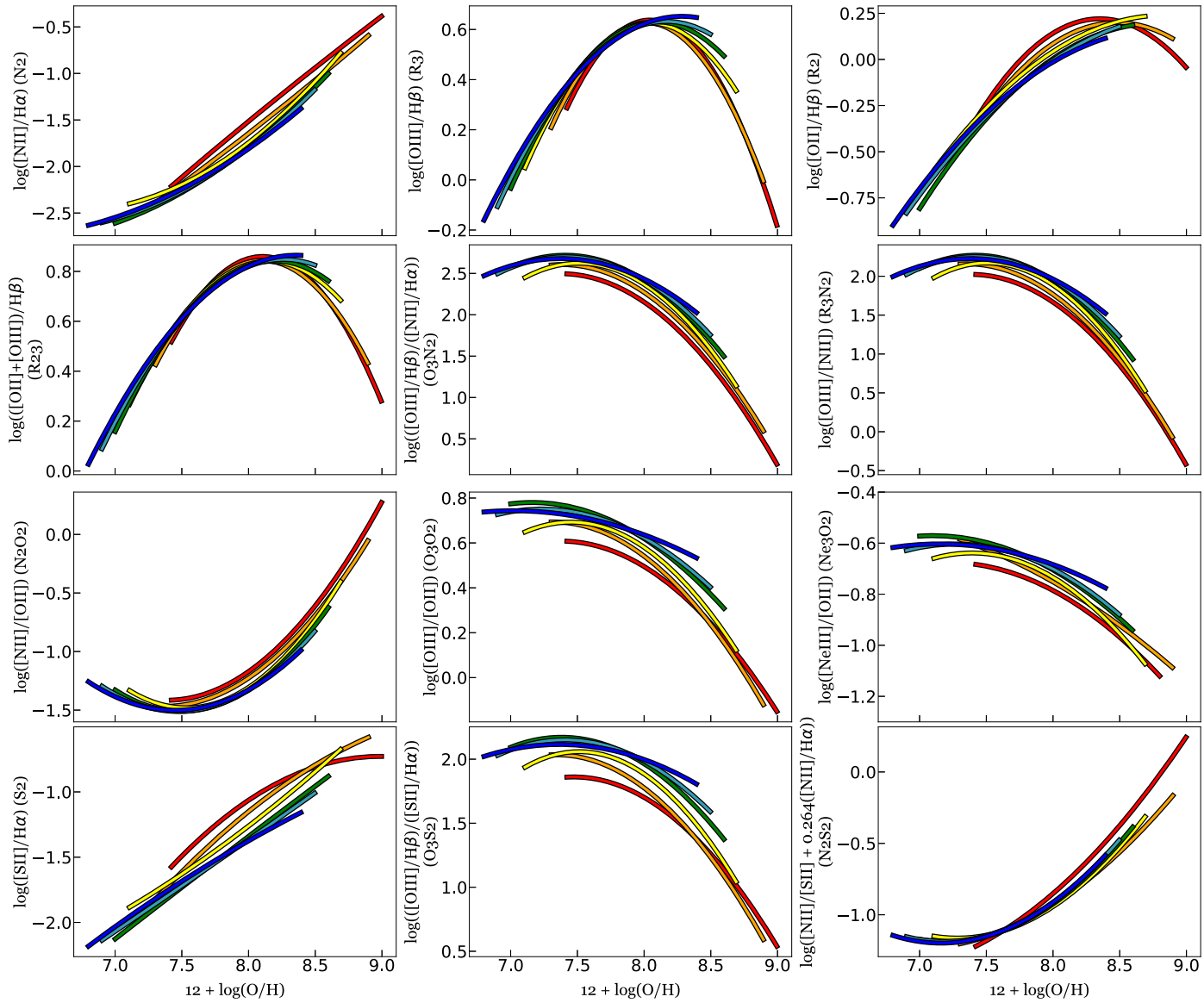


**Figure 4.** The redshift evolution of the strong-line metallicity calibrations using our nebular emission model. For the 12 line ratios, the calibrations going from z = 0 to z = 5 are shown by the different-colored lines, going from red to blue with increasing redshift. We fit polynomials of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$, where $x$ is the line ratio and $y$ is the metallicity (12 + log(O/H)). We find that the N2, O3N2, R3N2, N2O2, O3O2, Ne3O2, S2, and O3S2 line ratios show mild evolution of about 0.2–0.4 dex with redshift. In contrast, indicators like R2, R3, R32, and N2S2 show little redshift evolution. We provide another version of this figure with 90% confidence intervals plotted on top as shaded regions in Appendix B (Figure 7).

(Pilyugin et al. 2012; Garg et al. 2022). Therefore, in our model, we fix the nitrogen abundances according to Equation (4) to match the observed abundances. If it turns out that the log(N/O) versus log(O/H) relation evolves with redshift, then that will introduce uncertainty into our calibrations involving [N II]$\lambda$6583 at high redshifts (see Sections 2.3.2 and 5.2 for more information).

*R3, R2, and R23*—R3, R2, and R23 all involve oxygen emission lines and behave similarly; we find almost no evolution in any of these metallicity indicators across redshifts. [O III]$\lambda$5007 is one of the brightest lines in the spectrum, making it easy to observe even in faraway galaxies, which makes these ratios some of the most widely accessible across redshifts. Unfortunately, we find these line ratios, especially R3 and R23, to be quite poor metallicity calibrations, so they should be used only if a better alternative is not available, especially at low redshifts. This is because at metallicities above 7.0 in units of 12 + log(O/H), these are double values functions and at the same value of line ratio one can have two solutions for metallicity. Therefore, in the absence of any other constraint on metallicity, like stellar mass, there is no way to know which value is correct. However, as we go to lower metallicities or higher redshifts, these line ratios eventually do become single-valued and can be used as metallicity indicators reasonably well.

*O3N2, R3N2, and N2O2*—we find that O3N2, R3N2, and N2O2 all show mild variations of about 0.2–0.3 dex with redshift. O3N2 and R3N2 increase at the same metallicity with redshift, whereas N2O2 decreases with redshift at the same metallicity. We find that they are all reasonable metallicity indicators at high metallicities/low redshifts, with an average 90% confidence interval spread of about 0.7 dex at redshift 0, which is greater than the variation seen across redshifts. As we move to the low-metallicity regime, their utility becomes fairly

limited. Below $12 + \log(\mathrm{O/H}) \approx 8.0$, the curves start to flatten out and eventually start to turn over and become double-valued, which limits the scenarios in which these metallicity indicators can be utilized.

*O3O2 and Ne3O2*—in general, we find that O3O2 and Ne3O2 show an increase of about 0.2 dex with increasing redshift at the same metallicities. Both the indicators are a monotonic function of metallicity in the low-redshift/high-metallicity regime, while they become double-valued at metallicities below 7.5 in units of $12 + \log(\mathrm{O/H})$ as the curve starts to flatten out. Both these line ratios have a fairly wide 90% confidence interval, which, coupled with a low slope, hampers the usefulness as a metallicity indicator at all but extremely high metallicities.

*S2*—we find that S2 decreases by about 0.4 dex between $z = 0$ and $z = 5$ at the same metallicities. In general, we find S2 to be an excellent metallicity indicator across the whole metallicity range, except at very high metallicities ($12 + \log(\mathrm{O/H}) > 8.5$), where the curve starts to flatten out. That said, due to the significant contribution of the DIG to the total line flux at high metallicities, coupled with the uncertainties involved with our modeling of the DIG, one should be careful when using S2 at very high metallicities (see Section 4.2 for more information). S2 is a monotonically increasing function of metallicity and has a fairly high and almost constant slope, with 0.8 dex and a 90% confidence interval.

*O3S2 and N2S2*—we find that O3S2 increases by about 0.3–0.4 dex from $z = 0$ to $z = 5$ at the same metallicity, whereas we find almost no evolution in N2S2 except at very high metallicities. Both the O3S2 and N2S2 metallicity indicators work well at high metallicities ($12 + \log(\mathrm{O/H}) > 8.0$), but at metallicities lower than 8.0, they are double-valued functions of metallicity, and unless there is a way to break the degeneracy, this limits their application as metallicity indicators in the low-metallicity/high-redshift regime.

In summary, we find that indicators like N2, O3N2, R3N2, N2O2, O3O2, Ne3O2, S2, and O3S2 show mild evolution of about 0.2–0.4 dex with redshift, whereas indicators like R2, R3, R32, and N2S2 show almost no redshift evolution. We find S2 to be an excellent metallicity indicator across a metallicity range of 6.8–8.5 in units of $12 + \log(\mathrm{O/H})$, and then it starts to flatten out, whereas N2 works fairly well at all metallicities, except at the extremely low end. Line ratios such as R2, R3, and R23 work fairly well at low metallicities/high redshifts, but become a double-valued function at high metallicities, which hampers their utility at low redshifts. On the other hand, R3N2, O3N2, N2O2, O3S2, and N2S2 all work fairly well at low redshifts, but are double-valued at low metallicities, limiting their applicability at high redshifts. In contrast to observations, relatively shallow slopes and fairly wide confidence intervals hamper the usefulness of O3O2 and Ne3O2 across the whole metallicity range. New multidimensional metallicity indicators similar to N2S2, for example, can help break this degeneracy. Leveraging artificial intelligence models or techniques similar to that of Mirkwood Gilda et al. (2021), a machine-learning-based SED fitting tool, one can probe the higher-dimensional metallicity indicator space and find new indicators that are more robust and monotonic across a wider metallicity and redshift range. Implementing these methods is outside the scope of this paper and we defer this to future work.

**Table 1**
Theoretical Strong-line Calibrations Using Our Model at $z = 0$ for Different Line Ratios

| Line Ratio | Fit | Lower CI | Upper CI | Metallicity Range |
|---|---|---|---|---|
| N2 | $a_0$: −13.9<br>$a_1$: 1.918<br>$a_2$: −0.046 | $a_0$: −13.417<br>$a_1$: 1.887<br>$a_2$: −0.044 | $a_0$: −14.382<br>$a_1$: 1.949<br>$a_2$: −0.048 | 7.4–9.0 |
| R3 | $a_0$: −57.889<br>$a_1$: 14.552<br>$a_2$: −0.905 | $a_0$: −57.482<br>$a_1$: 14.525<br>$a_2$: −0.903 | $a_0$: −58.296<br>$a_1$: 14.578<br>$a_2$: −0.906 | 7.4–9.0 |
| R2 | $a_0$: −43.87<br>$a_1$: 10.558<br>$a_2$: −0.632 | $a_0$: −43.574<br>$a_1$: 10.539<br>$a_2$: −0.631 | $a_0$: −44.166<br>$a_1$: 10.577<br>$a_2$: −0.633 | 7.4–9.0 |
| R23 | $a_0$: −47.251<br>$a_1$: 11.872<br>$a_2$: −0.732 | $a_0$: −46.982<br>$a_1$: 11.854<br>$a_2$: −0.731 | $a_0$: −47.521<br>$a_1$: 11.89<br>$a_2$: −0.733 | 7.4–9.0 |
| O3N2 | $a_0$: −45.251<br>$a_1$: 12.947<br>$a_2$: −0.878 | $a_0$: −44.543<br>$a_1$: 12.902<br>$a_2$: −0.875 | $a_0$: −45.959<br>$a_1$: 12.993<br>$a_2$: −0.88 | 7.4–9.0 |
| R3N2 | $a_0$: −49.279<br>$a_1$: 13.895<br>$a_2$: −0.941 | $a_0$: −48.537<br>$a_1$: 13.847<br>$a_2$: −0.938 | $a_0$: −50.021<br>$a_1$: 13.944<br>$a_2$: −0.944 | 7.4–9.0 |
| N2O2 | $a_0$: 34.201<br>$a_1$: −9.642<br>$a_2$: 0.652 | $a_0$: 34.765<br>$a_1$: −9.678<br>$a_2$: 0.655 | $a_0$: 33.638<br>$a_1$: −9.605<br>$a_2$: 0.65 | 7.4–9.0 |
| O3O2 | $a_0$: −15.091<br>$a_1$: 4.258<br>$a_2$: −0.289 | $a_0$: −14.67<br>$a_1$: 4.231<br>$a_2$: −0.287 | $a_0$: −15.511<br>$a_1$: 4.285<br>$a_2$: −0.29 | 7.4–9.0 |
| Ne3O2 | $a_0$: 3.986<br>$a_1$: −0.877<br>$a_2$: 0.035 | $a_0$: 4.412<br>$a_1$: −0.907<br>$a_2$: 0.037 | $a_0$: 3.559<br>$a_1$: −0.847<br>$a_2$: 0.033 | 7.4–9.0 |
| S2 | $a_0$: −28.699<br>$a_1$: 6.227<br>$a_2$: −0.347 | $a_0$: −28.226<br>$a_1$: 6.196<br>$a_2$: −0.345 | $a_0$: −29.171<br>$a_1$: 6.257<br>$a_2$: −0.348 | 7.4–9.0 |
| O3S2 | $a_0$: −29.935<br>$a_1$: 8.511<br>$a_2$: −0.57 | $a_0$: −29.305<br>$a_1$: 8.47<br>$a_2$: −0.567 | $a_0$: −30.565<br>$a_1$: 8.552<br>$a_2$: −0.572 | 7.4–9.0 |
| N2S2 | $a_0$: 11.406<br>$a_1$: −3.871<br>$a_2$: 0.292 | $a_0$: 11.882<br>$a_1$: −3.901<br>$a_2$: 0.294 | $a_0$: 10.931<br>$a_1$: −3.84<br>$a_2$: 0.29 | 7.4–9.0 |

**Note.** We fit polynomials of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$, where $x$ is the line ratio and $y$ is the metallicity ($12 + \log(\mathrm{O/H})$). For fits to higher-redshift galaxies, see Appendix A.

We recommend using multiple metallicity indicators whenever possible, to mitigate the uncertainty.

In Tables 1 and 2 in Appendix A, we list the polynomial fits along with the fits for the lower and upper 90% confidence intervals to all the metallicity indicators discussed above, at $z = 0$ and $z = 1$–5, respectively.

### 4.2. Relative Contributions of Different Ionizing Sources

In this section, we discuss how the relative contributions of the different ionizing sources in a galaxy—namely, the H II regions around young stars, post-AGB stars, and DIG—vary with metallicities for different emission lines. In Figure 5, we

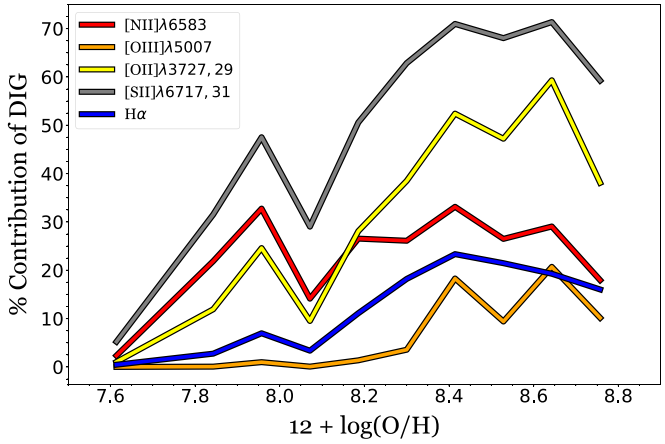


**Figure 5.** The relative contribution of DIG to the total line flux with respect to metallicity (12 + log(O/H) for SIMBA $z = 0$ galaxies. The DIG contributions for [N II]$\lambda$6583, [O III]$\lambda$5007, [O II]$\lambda$3727, 29, [S II]$\lambda$6717, 31, and H$\alpha$ are shown in red, orange, yellow, gray, and blue, respectively. We find that DIG is an important source of nebular emission and can account for as much as 70% of the total line flux in extreme cases.

plot the percentage of the total line flux that is emitted by DIG for [N II]$\lambda$6583, [O III]$\lambda$5007, [O II]$\lambda$3727, 29, [S II]$\lambda$6717, 31, and H$\alpha$. To simplify interpretation, in Figure 5, we only plot the percentage contribution of DIG against metallicity at $z = 0$: in all cases, we find that the post-AGB stars contribute exceedingly little to the total line emission budget and therefore ignore their contribution. However, their main contribution comes in the fact that they act as an important source of ionizing radiation for producing DIG. Vale Asari et al. (2019) and Sanders et al. (2017) have argued that DIG can be a substantial contributor and can account for as much as 50% of the line emission budget in local galaxies. Similarly, we find that DIG gas plays a significant role in generating nebular emission and that the contribution of DIG is positively correlated with metallicity for all five lines. For [N II]$\lambda$6583 and [O III]$\lambda$5007, the H II regions around young stars are the main contributor to the line flux at all metallicities, but DIG can still account for as much as 20%–30% of the total line flux in some cases. This is true even for H$\alpha$, which is independent of the complex abundance pattern and ionization state issues that affect the forbidden metal lines. In the cases of [O II]$\lambda$3727, 29 and [S II]$\lambda$6717, 31, we find that the contribution of DIG can account for as much as 70% of the total flux for high metallicities at $z = 0$. We find that the DIG contribution increases as the redshift increases, by a factor of 2 by $z = 5$, although these numbers remain very uncertain at high redshift for drawing any conclusion, given our lack of ionizing radiative transfer modeling in our simulations. Thus, DIG is an important source of nebular emission and must not be omitted when modeling nebular emission in galaxy-integrated spectra.

We observe a positive correlation between the percentage of DIG contribution to the nebular flux and the metallicity. This may be due to a number of possible reasons. First, this may just be due to galaxies with higher metallicity having larger stellar masses. This would mean that high-metallicity galaxies can have more ionizing photons escaping from H II regions and post-AGB stars that can then ionize the surrounding gas particles, therefore increasing the relative contribution of DIG. On the other hand, there can be a variety of other factors—for example, local observations from Oey et al. (2007) suggest that the fractional DIG contribution is related to the star formation rate surface density ($\Sigma_S$FR), which may be in turn depend on the sSFR. Thus, a varying sSFR with metallicity, with high-metallicity galaxies tending to have a lower sSFR or an evolving volume-filling fraction in the ISM, can also impact the relative contribution of DIG.

## 5. Discussion

### 5.1. Comparison to Other Theoretical Models and Relevant Observations

In this section, we review some of the recent observations and theoretical studies that look at the redshift evolution of nebular line ratios. Recently, major JWST programs—such as the Cosmic Evolution Early Release Survey, GLASS-ERS, and the Advanced Deep Extragalactic Survey—have observed nebular lines in a number of galaxies spanning a redshift range from $z \sim 5$–9 (Curti et al. 2022; Trump et al. 2023; Cameron et al. 2023; Nakajima et al. 2023; Sanders et al. 2024, 2023). Some of these studies (Curti et al. 2022; Trump et al. 2023; Nakajima et al. 2023; Sanders et al. 2024; Shapley et al. 2023a, 2023b; Laseter et al. 2024) were sufficiently sensitive to be able to detect the [O III]$\lambda$4363 auroral line in approximately 40 galaxies and obtain an estimate of the gas-phase metallicity using the direct $T_e$ measurement for the first time at such high redshifts.

In Figure 6, we plot the observations from Curti et al. (2022), Trump et al. (2023), Nakajima et al. (2023), Sanders et al. (2024), and Laseter et al. (2024) against our model calibrations at $z = 5$ (the red line). The observations span a wide range from $z \sim 5$–9, but our models do not go beyond $z = 5$. Thus, to make a fair comparison, we only focus on the observed galaxies that lie within $z = 4$–5 (shown as the square points). As can be seen, for the R2 line ratio, the base model at $z = 5$ shows a good match with the observed galaxies. On the other hand, we have shown that O3O2 and Ne3O2 do evolve with redshift, showing an increase of about 0.2 dex between $z = 0$ and $z = 5$. We find that our calibrations at $z = 5$ for O3O2 and Ne3O2 do not reach values high enough to match the observed line ratios. Our model underestimates the line ratios by about 0.5–0.7 dex. As for R3 and R32, we find that most of the observed galaxies lie either outside or at the edges of the 90% confidence interval. These preliminary observations suggest that there may be some evolution of the nebular metallicity indicators that is not captured by our models.

The origin of this offset may be due to factors like a varying ionization parameter, increasing hydrogen density with redshift, or increasing hardness of the ionizing spectrum at fixed nebular oxygen abundance with redshift, which is supported by studies like Strom et al. (2018), Sanders et al. (2019), Topping et al. (2020a, 2020b), and Runco et al. (2021) as being the primary driver behind the evolution of nebular lines with redshift. We are limited in our modeling by the things we can control. Therefore, we choose to manually vary the ionization parameter and see how it affects the line ratios. We find that increasing the ionization parameter can be one of the possible explanations for the observed redshift evolution of the line ratios. This is supported by the fact that we systematically underproduce higher ionization lines like [O III]$\lambda$5007 and overproduce lower ionization lines like [O II]$\lambda$3727, 29, which implies that high-redshift galaxies are evolving toward having a higher ionization parameter as compared to their local counterparts.

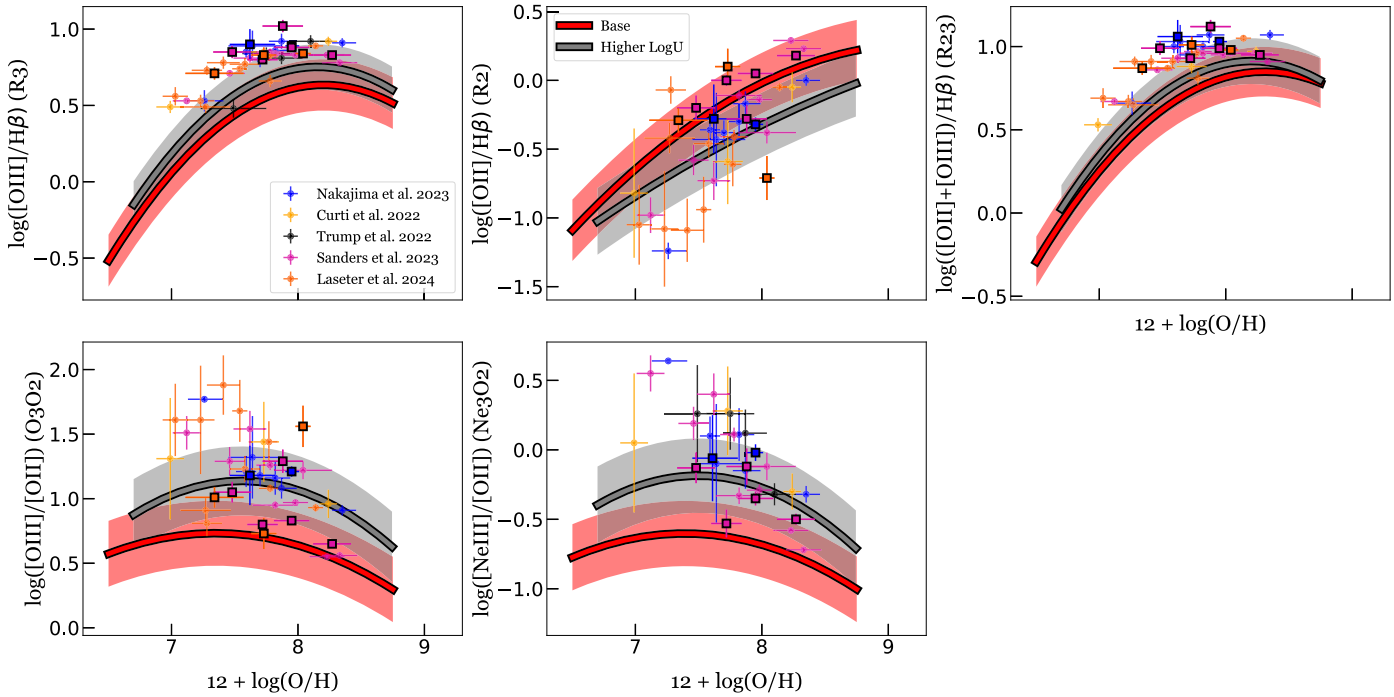


**Figure 6.** A comparison of our calibrations using SIMBA galaxies with the base model at $z = 5$ and an increased ionization parameter model at $z = 5$ against recent JWST observations of high-$z$ galaxies. We fit polynomials of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$, where $x$ is the line ratio and $y$ is the metallicity ($12 + \log(\mathrm{O/H})$). As the observed galaxies span a redshift range from $z \sim 5$–9, to make a fair comparison with our $z = 5$ model, we focus only on the observed galaxies that lie within $z = 4$–5, shown here by the square points. We find that the increased-ionization-parameter model provides a much better match to the observations in most cases, as compared to our base model, implying that the metallicity calibrations might evolve with redshift toward a higher ionization parameter. The observations from Curti et al. (2022), Trump et al. (2023), Nakajima et al. (2023), Sanders et al. (2024), and Laseter et al. (2024) are shown by the yellow, black, blue, pink, and orange points, respectively. In all the panels, the red line shows the calibration using our base model and the gray line shows our calibrations for an increased ionization parameter model. The shaded red and gray regions denote the 90% confidence intervals for the two models.

The ionization parameter encapsulates the ionization state of the nebula, with a higher ionization parameter corresponding to more elements being in a higher-ionization state. In our model, we define the ionization parameter as follows:

$$U = \left( \frac{Q}{4\pi \, {R_s}^2 \, n_{\mathrm{H}} \, c} \right), \tag{7}$$

where $Q$ is the rate of ionizing photons in inverse seconds, $n_{\mathrm{H}}$ is the hydrogen density in inverse cubic centimeters, and $R_{\mathrm{S}}$ is the Strömgren sphere radius given by Equation (5). On substituting for the defined Strömgren sphere radius, we get

$$U \propto Q^{1/3} {n_{\mathrm{H}}}^{1/3}. \tag{8}$$

To test whether having a higher ionization parameter improves the agreement between our calibrations and the observations, we rerun our model with an increased hydrogen density of $100\,\mathrm{cm}^{-3}$ and we set the inner radius of the cloud to be $10^{17}$ cm, which changes the cloud geometry from a thick shell to spherical. As we move to a spherical geometry, the cloud gets thicker and the inner radius moves closer to the star particle. This leads to an increase in the higher-ionization lines like [O III]$\lambda$5007, which are mainly produced in the inner regions of the cloud, whereas the opposite is true for singly ionized lines like [O II]$\lambda$3727, 29. In effect, moving to a spherical geometry and increasing the density lead to an increase in the ionization parameter by about 1.5 dex on average. For more information on how the line ratios are affected by cloud geometry, see Garg et al. (2022). The resulting calibrations with a higher ionization parameter are shown as the gray lines in Figure 6. We tabulate the fits to the metallicity calibrations with a higher ionization parameter in Table 2 (see Appendix A).

As can be seen, having a higher ionization parameter leads to a much better match with the observations for all line ratios, except R2. This suggests that metallicity calibrations might evolve with redshift toward a higher ionization parameter. This is also supported by a number of studies (e.g., Brinchmann et al. 2004; Liu et al. 2008; Hainline et al. 2009; Erb et al. 2010; Wuyts et al. 2012; Nakajima et al. 2013; Shirazi et al. 2014; Hayashi et al. 2015; Onodera et al. 2016; Kashino et al. 2017; Reddy et al. 2023a, 2023b) that concluded that galaxies at $z \sim 2$ have a higher ionization parameter compared to their local counterparts. Further, studies such as Curti et al. (2022), Trump et al. (2023), and Cameron et al. (2023) have found that nebular line ratios might evolve with redshift, since none of these metallicity indicators can simultaneously match the observations. Trump et al. (2023), using photoionization modeling, concluded that the line ratios might be explained with an ISM with high ionization, low metallicity, and very high pressure, suggesting extreme conditions in these high-$z$ galaxies. On the other hand, Nakajima et al. (2023) found no strong redshift evolution for R2, R3, and R23, whereas Sanders et al. (2024) used a combined sample of 41 galaxies at redshifts ranging from 1.4 to 8.7 and found a clear evolution in high-redshift metallicity indicators when compared to the $z = 0$ sample. However, similar to Sanders et al. (2024, 2023), we find no clear evolution among metallicity indicators from $z = 2$ onward.

Thus, the answer to the question of whether the nebular line ratios evolve with redshift is still unclear, but preliminary observations do point to an evolution toward an increasing

ionization parameter with redshift. Ongoing observations from JWST and future observations from upcoming space- and ground-based telescopes should provide a much more definite answer. The origin of this putative rise in a higher ionization parameter is unclear. A nonexhaustive list includes increased hydrogen density in H II regions (Reddy et al. 2023a, 2023b), evolving cloud geometries, or an increased ionizing photon production rate (due to stellar processes such as binary star evolution or stellar IMF variation).

On the theoretical end, Hirschmann et al. (2023) recently investigated the redshift evolution of strong-line metallicity indicators in $z \sim 0$–8 galaxies from the IllustrisTNG simulation. Their nebular emission model takes into account contributions from young stars, post-AGB stars, active galactic nuclei (AGNs), and, for the first time, fast-radiating shocks. That said, there are some significant differences between their methodology and the one employed in this study. For example, Hirschmann et al. (2023) do not include dust radiative transfer and the contribution from DIG, whereas we do not include the contribution from radiating shocks and AGNs in our model. One major difference is that Hirschmann et al. (2023) model nebular emission from H II regions and post-AGB stars by assigning each galaxy a single model that is selected based on the galaxy average properties, like the metallicity, ionization parameter, and the C/O ratio, using a precomputed set of lookup tables. In contrast to this, we compute the nebular line emission on a particle-by-particle basis, where each particle is assigned as an H II region, a post-AGB star, or a DIG source, which is modeled based on the properties of that particle. In the end, Hirschmann et al. (2023) conclude that strong-line metallicity indicators evolve with redshift, and the evolution can be explained by high-redshift galaxies having a higher ionization parameter, which can be a byproduct of them having a larger sSFR and higher assumed H II region densities.

### 5.2. Caveats and Uncertainties

*N/O Ratio*—Garg et al. (2022) found that the SIMBA simulation is unable to reproduce the observed log(N/O) versus log(O/H) relation from Pilyugin et al. (2012), which may be attributed to the complexity involved in modeling nitrogen nucleosynthesis. The complexity arises because nitrogen can be produced via two pathways. In stars, nitrogen is produced through the CNO cycle, where the reaction $^{14}$N(P, $\gamma$)$^{15}$O is the slowest in the chain, causing nitrogen to build up over time. If carbon synthesized within the star is used as a catalyst for this reaction, then the nitrogen produced is called primary nitrogen. Alternatively, if pre-existing carbon is used, then it is called secondary nitrogen (Meynet & Maeder 2002).

In our models, we sidestepped the inability of SIMBA to match the local N/O versus O/H observed relationship by manually setting the nitrogen abundance as per Equation (4). That said, if it turns out that the log(N/O) versus log(O/H) relation evolves with redshift, then our model calibrations will not apply at high redshift. The possible evolution (or lack thereof) of N/O ratios with redshift is currently under debate. Some studies, such as Steidel et al. (2016), find no evolution with redshift, whereas others find an increase in log(N/O) of anywhere between 0.1 and 0.4 dex at redshifts of about 1.5 (Masters et al. 2014; Strom et al. 2018; Bian et al. 2020).

*Subresolution Effects*—current state-of-the-art large-scale cosmological galaxy formation simulations (including SIMBA) are unable to achieve a high enough resolution to directly model individual H II regions or DIG clouds. Due to this, we have to resort to subgrid modeling, which introduces uncertainties into our model. We enforce three major assumptions that could substantively impact our model results:

1. *Mass and Age Distribution*—a single particle in our SIMBA box is of the order of $2.3 \times 10^6\,M_\odot$, which is too massive to be modeled as a single H II region. Further, SIMBA has a stochastic implementation of stellar ages for young star particles. To counter these drawbacks, we divide each star particle into smaller-mass particles with varying ages based on power-law distributions. We have kept the slope of this distribution fixed across redshifts and different physical conditions, which introduces a level of uncertainty into our modeling. Apart from this, since we model galaxies that have at least one star younger than 10 Myr, this introduces a bias into our results. To quantify the additional uncertainty due to excluding low-mass galaxies, we took a subsample of galaxies at $z = 0$ and relaxed the constraint to include all galaxies that have star particles younger than 50 Myr instead of 10 Myr. We found that doing so added an uncertainty of about 0.05–0.3 dex, which is well within the 90% confidence interval for all metallicity indicators.
2. *Geometry*—another major subresolution model is in our assumption of spherical H II regions: we are unable to model the true geometry and escape fractions of H II regions surrounding massive stars. Similarly, the detailed geometry of the DIG is unresolved, and we are instead forced to assume a plane-parallel geometry.
3. *Density*—we assume a fixed density for H II regions and DIG gas that does not vary with redshift. Increasing the hydrogen density will lead to an increase in the average ionization parameter, thus the flux of singly ionized species like [N II]$\lambda$6583, [S II]$\lambda$6717, 31 will decrease, whereas the flux of doubly ionized species like [O III] $\lambda$5007 will increase.

*Other Sources of Ionizing Radiation*—though we have expanded our base model from Garg et al. (2022) to include nebular emission not only from H II regions, but also from post-AGB stars and DIG, we are still missing some other sources of ionizing radiation in galaxies, namely AGNs and fast-radiating shocks. The ionized gas present within the narrowline regions of AGNs can be a substantial source of nebular emission. AGNs produce a much harder radiation field when compared to star-forming regions, which leads to an increase in higher-order ionized line ratios like R3. Including the effect of AGNs is important to accurately model emission lines across redshifts. Some theoretical studies (e.g., Feltre et al. 2016; Hirschmann et al. 2017, 2023) have modeled nebular emission from AGNs using CLOUDY lookup tables, and we defer an implementation of a similar approach to future work.

*DIG Modeling*—the main uncertainty in our DIG model comes from our approximation of the shape of the input radiation field impinging on the gas cell. As described in detail in Section 2.3.2, we compute the shape of the radiation field by taking a distance-weighted average of the CLOUDY output spectrum of the ionizing radiation sources (the H II regions around massive stars and post-AGB stars) within 1 kpc. This implicitly assumes the optically thin transfer of those photons to the incident DIG cell. We additionally are required to make assumptions regarding the escape fractions from H II regions

and post-AGB stars, which, as described in Section 2.3.2, are kept fixed at 0.4 and 0.6, respectively. This, of course, does not capture the wide variety of environments that these ionizing sources can be found in.

Beyond this, our mass resolution in cosmological simulations prevents us from modeling directly the geometry of DIG gas cells. The SIMBA snapshots used in this study have a resolution of about $10^6\,M_\odot$, which means that we can model only a collection of gas clouds at a time, which fails to capture any potential substructure. In reality, individual gas clouds span a wide range of densities and geometries.

Taken together, these uncertainties in our DIG modeling may explain why our calibrations for SIMBA $z = 0$ galaxies involving [O III]$\lambda$5007, [O II]$\lambda$3727, 29, and [S II]$\lambda$6717, 31, like R2, R3, R23, O3O2, and S2, match the auroral line observations fairly well at low metallicities, but are not as good a match at high metallicities (see Figure 3 and Section 3 for more details). This, coupled with the fact that our models are a reasonably good match to N2, which does not have a substantial DIG contribution at high metallicities, strongly suggests that the mismatch might be explained by the inability of our model to accurately capture the DIG emission.

## 6. Summary

In summary, in this study, we have investigated the redshift evolution of 12 strong-line metallicity indicators. We have modeled nebular emission on a particle-by-particle basis on galaxies from SIMBA, using the CLOUDY photoionization code. We have expanded our model to not only include the contribution from H II regions, but also post-AGB stars and DIG. The main results of this study are as follows:

1. We find that our model reasonably matches the observed auroral line calibrations and reproduces the general nature of the curves for most of the line ratios (Section 3, Figure 3, and Table 1).
2. We run our models on the SIMBA cosmological simulation at integer redshifts from $z = 0$–5, and find mild redshift evolution for all the 12 line ratios (Section 4 and Figure 4). This suggests that our results are affected by properties that vary in our model with redshift, including the spectra of ionizing photons in H II regions and post-AGB stars, the incident radiation field for calculating the DIG emission, the stellar IMF, and the amount of gas available for ionization, which are likely to have a discernible impact on the metallicity indicators.
3. We use these results to assess strong-line calibrations at high redshift (Section 4, Figure 4, and Appendix A). We find that the ratios R3, R2, and R32 show no evolution with redshift, whereas other line ratios, like N2, O3N2, R3N2, N2O2, O3O2, Ne3O2, S2, O3S2, and N2S2, show a an evolution ranging from 0.2–0.5 dex between $z = 0$ and $z = 5$.
4. We find that at low metallicities/high redshifts, the line ratios R3, R2, S2, and R32 work fairly well, depending on the metallicity range one is interested in, with R2 and S2 being mostly monotonic across the metallicity range. At low redshifts/high metallicities, the line ratios N2, O3N2, R3N2, N2O2, O3S2, and N2S2 work well, with N2 being the best among them. On the other hand, we find that O3O2 and Ne3O2 have limited applicability, owing to their flat slopes and fairly wide confidence intervals.
5. We find that for local galaxies, the DIG contribution increases with metallicity for singly ionized oxygen and sulfur lines. It can account for as much as 70% of the total line emission in extreme cases (Figure 5). This demonstrates the importance of including the contribution from DIG when modeling nebular emission.
6. On comparing our metallicity calibrations against recent JWST auroral line measurements, we see a slight mismatch and find that having a higher ionization parameter at high redshifts might be one of the possible explanations (Figure 6). The increased ionization parameter could be caused by increasing hydrogen density, evolving cloud geometry, or an evolving (more top-heavy) IMF with redshift. Current observations are limited to a handful of galaxies and more JWST observations in the coming years will enable us to conclusively answer whether the metallicity indicators evolve with redshift or not.

## Acknowledgments

The authors would like to thank Allison Strom for fruitful discussions that guided the analysis of this paper. Some of the SIMBA simulation boxes used in this work were run on the Hipergator computing cluster at the University of Florida. The flagship SIMBA simulation was run using the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility. The equipment was funded by BEIS capital funding, via STFC capital grants ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University, and STFC operations grant ST/R000832/1. R.D. acknowledges support from the Wolfson Research Merit Award program of the U.K. Royal Society. P.G. and D.N. were funded by JWST-AR-01883.001.

*Software:* caesar (Thompson 2014), cloudy (v17.00; Ferland et al. 2017), cloudyfsps (Byler et al. 2017), fsps (Conroy et al. 2009; Conroy & Gunn 2010), powderday (Narayanan et al. 2021), python-fsps (Johnson et al. 2021), simba (Davé et al. 2019).

## Appendix A
## Theoretical Strong-line Calibrations Using Our Model at High Redshifts

As discussed in Section 4, we list the polynomial fits along with the fits for the lower and upper 90% confidence intervals to all the metallicity indicators at higher redshifts ($z$ = 1–5) in Table 2.

**Table 2**
Theoretical Strong-line Calibrations for $z = 1$–5 Using the Base Model and for $z = 5$ Using the Model with a Higher Ionization Parameter for Different Line Ratios

| Line Ratio | Redshift | Fit | Lower CI | Upper CI | Metallicity Range |
|---|---|---|---|---|---|
| N2 | 1 | $a_0$: −6.662<br>$a_1$: 0.138<br>$a_2$: 0.061 | $a_0$: −6.233<br>$a_1$: 0.105<br>$a_2$: 0.063 | $a_0$: −7.091<br>$a_1$: 0.17<br>$a_2$: 0.059 | 7.3–8.9 |
| | 2 | $a_0$: 16.247<br>$a_1$: −5.606<br>$a_2$: 0.42 | $a_0$: 16.596<br>$a_1$: −5.632<br>$a_2$: 0.421 | $a_0$: 15.899<br>$a_1$: −5.581<br>$a_2$: 0.418 | 7.1–8.7 |
| | 3 | $a_0$: 9.368<br>$a_1$: −3.923<br>$a_2$: 0.316 | $a_0$: 9.682<br>$a_1$: −3.942<br>$a_2$: 0.317 | $a_0$: 9.055<br>$a_1$: −3.904<br>$a_2$: 0.315 | 7.0–8.6 |
| | 4 | $a_0$: 10.87<br>$a_1$: −4.264<br>$a_2$: 0.335 | $a_0$: 11.181<br>$a_1$: −4.281<br>$a_2$: 0.336 | $a_0$: 10.559<br>$a_1$: −4.248<br>$a_2$: 0.334 | 6.9–8.5 |
| | 5 | $a_0$: 6.902<br>$a_1$: −3.176<br>$a_2$: 0.261 | $a_0$: 7.212<br>$a_1$: −3.195<br>$a_2$: 0.262 | $a_0$: 6.593<br>$a_1$: −3.156<br>$a_2$: 0.26 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: 17.348<br>$a_1$: −5.817<br>$a_2$: 0.423 | $a_0$: 17.933<br>$a_1$: −5.9<br>$a_2$: 0.429 | $a_0$: 16.762<br>$a_1$: −5.733<br>$a_2$: 0.418 | 6.8–8.4 |
| R3 | 1 | $a_0$: −52.137<br>$a_1$: 13.162<br>$a_2$: −0.821 | $a_0$: −51.754<br>$a_1$: 13.133<br>$a_2$: −0.819 | $a_0$: −52.521<br>$a_1$: 13.192<br>$a_2$: −0.823 | 7.3–8.9 |
| | 2 | $a_0$: −41.669<br>$a_1$: 10.511<br>$a_2$: −0.653 | $a_0$: −41.425<br>$a_1$: 10.494<br>$a_2$: −0.652 | $a_0$: −41.913<br>$a_1$: 10.529<br>$a_2$: −0.654 | 7.1–8.7 |
| | 3 | $a_0$: −35.338<br>$a_1$: 8.88<br>$a_2$: −0.548 | $a_0$: −35.114<br>$a_1$: 8.867<br>$a_2$: −0.547 | $a_0$: −35.562<br>$a_1$: 8.894<br>$a_2$: −0.549 | 7.0–8.6 |
| | 4 | $a_0$: −30.177<br>$a_1$: 7.546<br>$a_2$: −0.462 | $a_0$: −29.971<br>$a_1$: 7.535<br>$a_2$: −0.461 | $a_0$: −30.384<br>$a_1$: 7.557<br>$a_2$: −0.463 | 6.9–8.5 |
| | 5 | $a_0$: −24.688<br>$a_1$: 6.118<br>$a_2$: −0.369 | $a_0$: −24.473<br>$a_1$: 6.104<br>$a_2$: −0.368 | $a_0$: −24.902<br>$a_1$: 6.132<br>$a_2$: −0.37 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −26.244<br>$a_1$: 6.585<br>$a_2$: −0.401 | $a_0$: −25.914<br>$a_1$: 6.538<br>$a_2$: −0.399 | $a_0$: −26.573<br>$a_1$: 6.632<br>$a_2$: −0.404 | 6.8–8.4 |
| R2 | 1 | $a_0$: −35.496<br>$a_1$: 8.42<br>$a_2$: −0.497 | $a_0$: −35.222<br>$a_1$: 8.399<br>$a_2$: −0.495 | $a_0$: −35.771<br>$a_1$: 8.441<br>$a_2$: −0.498 | 7.3–8.9 |
| | 2 | $a_0$: −21.727<br>$a_1$: 4.973<br>$a_2$: −0.281 | $a_0$: −21.466<br>$a_1$: 4.954<br>$a_2$: −0.28 | $a_0$: −21.989<br>$a_1$: 4.992<br>$a_2$: −0.283 | 7.1–8.7 |
| | 3 | $a_0$: −24.442<br>$a_1$: 5.618<br>$a_2$: −0.32 | $a_0$: −24.188<br>$a_1$: 5.603<br>$a_2$: −0.319 | $a_0$: −24.696<br>$a_1$: 5.634<br>$a_2$: −0.321 | 7.0–8.6 |
| | 4 | $a_0$: −19.516<br>$a_1$: 4.391<br>$a_2$: −0.244 | $a_0$: −19.254<br>$a_1$: 4.377<br>$a_2$: −0.243 | $a_0$: −19.777<br>$a_1$: 4.405<br>$a_2$: −0.245 | 6.9–8.5 |
| | 5 | $a_0$: −19.95<br>$a_1$: 4.556<br>$a_2$: −0.258 | $a_0$: −19.667<br>$a_1$: 4.538<br>$a_2$: −0.257 | $a_0$: −20.233<br>$a_1$: 4.574<br>$a_2$: −0.259 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −11.204<br>$a_1$: 2.311<br>$a_2$: −0.119 | $a_0$: −10.668<br>$a_1$: 2.235<br>$a_2$: −0.114 | $a_0$: −11.74<br>$a_1$: 2.388<br>$a_2$: −0.124 | 6.8–8.4 |
| R23 | 1 | $a_0$: −42.197<br>$a_1$: 10.625<br>$a_2$: −0.656 | $a_0$: −41.955<br>$a_1$: 10.606<br>$a_2$: −0.655 | $a_0$: −42.439<br>$a_1$: 10.643<br>$a_2$: −0.657 | 7.3–8.9 |

**Table 2**
(Continued)

| Line Ratio | Redshift | Fit | Lower CI | Upper CI | Metallicity Range |
|---|---|---|---|---|---|
| | 2 | $a_0$: −34.364<br>$a_1$: 8.641<br>$a_2$: −0.53 | $a_0$: −34.18<br>$a_1$: 8.628<br>$a_2$: −0.529 | $a_0$: −34.548<br>$a_1$: 8.655<br>$a_2$: −0.531 | 7.1–8.7 |
| | 3 | $a_0$: −31.514<br>$a_1$: 7.9<br>$a_2$: −0.482 | $a_0$: −31.324<br>$a_1$: 7.889<br>$a_2$: −0.481 | $a_0$: −31.704<br>$a_1$: 7.911<br>$a_2$: −0.483 | 7.0–8.6 |
| | 4 | $a_0$: −27.149<br>$a_1$: 6.777<br>$a_2$: −0.41 | $a_0$: −26.973<br>$a_1$: 6.768<br>$a_2$: −0.41 | $a_0$: −27.325<br>$a_1$: 6.787<br>$a_2$: −0.411 | 6.9–8.5 |
| | 5 | $a_0$: −23.269<br>$a_1$: 5.776<br>$a_2$: −0.346 | $a_0$: −23.077<br>$a_1$: 5.764<br>$a_2$: −0.345 | $a_0$: −23.46<br>$a_1$: 5.788<br>$a_2$: −0.346 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −24.758<br>$a_1$: 6.233<br>$a_2$: −0.378 | $a_0$: −24.446<br>$a_1$: 6.189<br>$a_2$: −0.376 | $a_0$: −25.07<br>$a_1$: 6.278<br>$a_2$: −0.381 | 6.8–8.4 |
| O3N2 | 1 | $a_0$: −45.45<br>$a_1$: 13.017<br>$a_2$: −0.881 | $a_0$: −44.783<br>$a_1$: 12.966<br>$a_2$: −0.878 | $a_0$: −46.116<br>$a_1$: 13.068<br>$a_2$: −0.885 | 7.3–8.9 |
| | 2 | $a_0$: −57.392<br>$a_1$: 15.982<br>$a_2$: −1.064 | $a_0$: −56.912<br>$a_1$: 15.947<br>$a_2$: −1.062 | $a_0$: −57.871<br>$a_1$: 16.016<br>$a_2$: −1.066 | 7.1–8.7 |
| | 3 | $a_0$: −44.706<br>$a_1$: 12.803<br>$a_2$: −0.864 | $a_0$: −44.303<br>$a_1$: 12.779<br>$a_2$: −0.863 | $a_0$: −45.11<br>$a_1$: 12.827<br>$a_2$: −0.866 | 7.0–8.6 |
| | 4 | $a_0$: −41.047<br>$a_1$: 11.811<br>$a_2$: −0.797 | $a_0$: −40.658<br>$a_1$: 11.789<br>$a_2$: −0.796 | $a_0$: −41.436<br>$a_1$: 11.832<br>$a_2$: −0.798 | 6.9–8.5 |
| | 5 | $a_0$: −31.59<br>$a_1$: 9.294<br>$a_2$: −0.63 | $a_0$: −31.208<br>$a_1$: 9.27<br>$a_2$: −0.629 | $a_0$: −31.972<br>$a_1$: 9.318<br>$a_2$: −0.632 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −43.591<br>$a_1$: 12.402<br>$a_2$: −0.825 | $a_0$: −42.887<br>$a_1$: 12.302<br>$a_2$: −0.818 | $a_0$: −44.295<br>$a_1$: 12.502<br>$a_2$: −0.831 | 6.8–8.4 |
| R3N2 | 1 | $a_0$: −48.618<br>$a_1$: 13.809<br>$a_2$: −0.939 | $a_0$: −47.876<br>$a_1$: 13.752<br>$a_2$: −0.935 | $a_0$: −49.359<br>$a_1$: 13.866<br>$a_2$: −0.942 | 7.3–8.9 |
| | 2 | $a_0$: −64.199<br>$a_1$: 17.678<br>$a_2$: −1.177 | $a_0$: −63.674<br>$a_1$: 17.64<br>$a_2$: −1.175 | $a_0$: −64.724<br>$a_1$: 17.716<br>$a_2$: −1.18 | 7.1–8.7 |
| | 3 | $a_0$: −50.615<br>$a_1$: 14.258<br>$a_2$: −0.961 | $a_0$: −50.181<br>$a_1$: 14.232<br>$a_2$: −0.959 | $a_0$: −51.049<br>$a_1$: 14.284<br>$a_2$: −0.963 | 7.0–8.6 |
| | 4 | $a_0$: −45.803<br>$a_1$: 12.962<br>$a_2$: −0.874 | $a_0$: −45.39<br>$a_1$: 12.94<br>$a_2$: −0.873 | $a_0$: −46.215<br>$a_1$: 12.984<br>$a_2$: −0.875 | 6.9–8.5 |
| | 5 | $a_0$: −35.253<br>$a_1$: 10.157<br>$a_2$: −0.688 | $a_0$: −34.853<br>$a_1$: 10.132<br>$a_2$: −0.686 | $a_0$: −35.653<br>$a_1$: 10.182<br>$a_2$: −0.69 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −46.413<br>$a_1$: 13.038<br>$a_2$: −0.868 | $a_0$: −45.683<br>$a_1$: 12.934<br>$a_2$: −0.861 | $a_0$: −47.143<br>$a_1$: 13.143<br>$a_2$: −0.874 | 6.8–8.4 |
| N2O2 | 1 | $a_0$: 31.541<br>$a_1$: −8.959<br>$a_2$: 0.608 | $a_0$: 32.024<br>$a_1$: −8.996<br>$a_2$: 0.61 | $a_0$: 31.057<br>$a_1$: −8.922<br>$a_2$: 0.606 | 7.3–8.9 |
| | 2 | $a_0$: 44.395<br>$a_1$: −12.176<br>$a_2$: 0.808 | $a_0$: 44.698<br>$a_1$: −12.198<br>$a_2$: 0.809 | $a_0$: 44.091<br>$a_1$: −12.154<br>$a_2$: 0.807 | 7.1–8.7 |

**Table 2**
(Continued)

| Line Ratio | Redshift | Fit | Lower CI | Upper CI | Metallicity Range |
|---|---|---|---|---|---|
| | 3 | $a_0$: 39.719<br>$a_1$: −10.996<br>$a_2$: 0.733 | $a_0$: 39.964<br>$a_1$: −11.011<br>$a_2$: 0.734 | $a_0$: 39.474<br>$a_1$: −10.982<br>$a_2$: 0.732 | 7.0–8.6 |
| | 4 | $a_0$: 35.141<br>$a_1$: −9.807<br>$a_2$: 0.656 | $a_0$: 35.37<br>$a_1$: −9.819<br>$a_2$: 0.657 | $a_0$: 34.912<br>$a_1$: −9.794<br>$a_2$: 0.655 | 6.9–8.5 |
| | 5 | $a_0$: 30.515<br>$a_1$: −8.595<br>$a_2$: 0.577 | $a_0$: 30.767<br>$a_1$: −8.611<br>$a_2$: 0.578 | $a_0$: 30.264<br>$a_1$: −8.579<br>$a_2$: 0.576 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: 31.373<br>$a_1$: −8.764<br>$a_2$: 0.585 | $a_0$: 31.833<br>$a_1$: −8.83<br>$a_2$: 0.589 | $a_0$: 30.913<br>$a_1$: −8.699<br>$a_2$: 0.581 | 6.8–8.4 |
| O3O2 | 1 | $a_0$: −16.862<br>$a_1$: 4.797<br>$a_2$: −0.328 | $a_0$: −16.364<br>$a_1$: 4.759<br>$a_2$: −0.325 | $a_0$: −17.36<br>$a_1$: 4.835<br>$a_2$: −0.33 | 7.3–8.9 |
| | 2 | $a_0$: −19.924<br>$a_1$: 5.534<br>$a_2$: −0.371 | $a_0$: −19.548<br>$a_1$: 5.507<br>$a_2$: −0.37 | $a_0$: −20.3<br>$a_1$: 5.561<br>$a_2$: −0.373 | 7.1–8.7 |
| | 3 | $a_0$: −10.896<br>$a_1$: 3.262<br>$a_2$: −0.228 | $a_0$: −10.573<br>$a_1$: 3.242<br>$a_2$: −0.227 | $a_0$: −11.219<br>$a_1$: 3.281<br>$a_2$: −0.229 | 7.0–8.6 |
| | 4 | $a_0$: −10.661<br>$a_1$: 3.155<br>$a_2$: −0.218 | $a_0$: −10.345<br>$a_1$: 3.138<br>$a_2$: −0.217 | $a_0$: −10.978<br>$a_1$: 3.172<br>$a_2$: −0.219 | 6.9–8.5 |
| | 5 | $a_0$: −4.737<br>$a_1$: 1.562<br>$a_2$: −0.111 | $a_0$: −4.414<br>$a_1$: 1.541<br>$a_2$: −0.11 | $a_0$: −5.061<br>$a_1$: 1.582<br>$a_2$: −0.113 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −15.04<br>$a_1$: 4.274<br>$a_2$: −0.283 | $a_0$: −14.453<br>$a_1$: 4.19<br>$a_2$: −0.277 | $a_0$: −15.627<br>$a_1$: 4.358<br>$a_2$: −0.288 | 6.8–8.4 |
| Ne3O2 | 1 | $a_0$: −3.828<br>$a_1$: 1.067<br>$a_2$: −0.085 | $a_0$: −3.489<br>$a_1$: 1.04<br>$a_2$: −0.084 | $a_0$: −4.166<br>$a_1$: 1.095<br>$a_2$: −0.087 | 7.3–8.9 |
| | 2 | $a_0$: −14.735<br>$a_1$: 3.814<br>$a_2$: −0.258 | $a_0$: −14.416<br>$a_1$: 3.79<br>$a_2$: −0.256 | $a_0$: −15.054<br>$a_1$: 3.837<br>$a_2$: −0.259 | 7.1–8.7 |
| | 3 | $a_0$: −8.895<br>$a_1$: 2.345<br>$a_2$: −0.165 | $a_0$: −8.611<br>$a_1$: 2.328<br>$a_2$: −0.164 | $a_0$: −9.179<br>$a_1$: 2.362<br>$a_2$: −0.166 | 7.0–8.6 |
| | 4 | $a_0$: −10.372<br>$a_1$: 2.686<br>$a_2$: −0.185 | $a_0$: −10.084<br>$a_1$: 2.671<br>$a_2$: −0.184 | $a_0$: −10.659<br>$a_1$: 2.702<br>$a_2$: −0.186 | 6.9–8.5 |
| | 5 | $a_0$: −6.314<br>$a_1$: 1.597<br>$a_2$: −0.112 | $a_0$: −6.011<br>$a_1$: 1.578<br>$a_2$: −0.11 | $a_0$: −6.618<br>$a_1$: 1.616<br>$a_2$: −0.113 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −14.295<br>$a_1$: 3.768<br>$a_2$: −0.252 | $a_0$: −13.706<br>$a_1$: 3.684<br>$a_2$: −0.247 | $a_0$: −14.884<br>$a_1$: 3.852<br>$a_2$: −0.257 | 6.8–8.4 |
| S2 | 1 | $a_0$: −20.957<br>$a_1$: 4.132<br>$a_2$: −0.207 | $a_0$: −20.434<br>$a_1$: 4.093<br>$a_2$: −0.205 | $a_0$: −21.48<br>$a_1$: 4.172<br>$a_2$: −0.209 | 7.3–8.9 |
| | 2 | $a_0$: −1.276<br>$a_1$: −0.78<br>$a_2$: 0.098 | $a_0$: −0.853<br>$a_1$: −0.811<br>$a_2$: 0.1 | $a_0$: −1.699<br>$a_1$: −0.75<br>$a_2$: 0.096 | 7.1–8.7 |
| | 3 | $a_0$: −8.202<br>$a_1$: 0.94<br>$a_2$: −0.01 | $a_0$: −7.847<br>$a_1$: 0.918<br>$a_2$: −0.009 | $a_0$: −8.557<br>$a_1$: 0.961<br>$a_2$: −0.012 | 7.0–8.6 |

**Table 2**
(Continued)

| Line Ratio | Redshift | Fit | Lower CI | Upper CI | Metallicity Range |
|---|---|---|---|---|---|
| | 4 | $a_0$: −6.348<br>$a_1$: 0.532<br>$a_2$: 0.011 | $a_0$: −6.012<br>$a_1$: 0.514<br>$a_2$: 0.013 | $a_0$: −6.685<br>$a_1$: 0.55<br>$a_2$: 0.01 | 6.9–8.5 |
| | 5 | $a_0$: −10.901<br>$a_1$: 1.8<br>$a_2$: −0.076 | $a_0$: −10.565<br>$a_1$: 1.779<br>$a_2$: −0.075 | $a_0$: −11.236<br>$a_1$: 1.821<br>$a_2$: −0.078 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −4.682<br>$a_1$: 0.242<br>$a_2$: 0.016 | $a_0$: −4.134<br>$a_1$: 0.164<br>$a_2$: 0.021 | $a_0$: −5.229<br>$a_1$: 0.32<br>$a_2$: 0.012 | 6.8–8.4 |
| O3S2 | 1 | $a_0$: −30.333<br>$a_1$: 8.815<br>$a_2$: −0.6 | $a_0$: −29.541<br>$a_1$: 8.754<br>$a_2$: −0.596 | $a_0$: −31.124<br>$a_1$: 8.875<br>$a_2$: −0.604 | 7.3–8.9 |
| | 2 | $a_0$: −39.426<br>$a_1$: 11.041<br>$a_2$: −0.735 | $a_0$: −38.874<br>$a_1$: 11.001<br>$a_2$: −0.732 | $a_0$: −39.978<br>$a_1$: 11.081<br>$a_2$: −0.737 | 7.1–8.7 |
| | 3 | $a_0$: −27.136<br>$a_1$: 7.94<br>$a_2$: −0.538 | $a_0$: −26.695<br>$a_1$: 7.914<br>$a_2$: −0.536 | $a_0$: −27.578<br>$a_1$: 7.967<br>$a_2$: −0.54 | 7.0–8.6 |
| | 4 | $a_0$: −23.829<br>$a_1$: 7.014<br>$a_2$: −0.473 | $a_0$: −23.419<br>$a_1$: 6.992<br>$a_2$: −0.472 | $a_0$: −24.239<br>$a_1$: 7.036<br>$a_2$: −0.475 | 6.9–8.5 |
| | 5 | $a_0$: −13.787<br>$a_1$: 4.318<br>$a_2$: −0.293 | $a_0$: −13.392<br>$a_1$: 4.293<br>$a_2$: −0.292 | $a_0$: −14.182<br>$a_1$: 4.343<br>$a_2$: −0.295 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: −21.562<br>$a_1$: 6.343<br>$a_2$: −0.418 | $a_0$: −20.948<br>$a_1$: 6.256<br>$a_2$: −0.412 | $a_0$: −22.176<br>$a_1$: 6.431<br>$a_2$: −0.423 | 6.8–8.4 |
| N2S2 | 1 | $a_0$: 12.618<br>$a_1$: −3.979<br>$a_2$: 0.286 | $a_0$: 13.126<br>$a_1$: −4.018<br>$a_2$: 0.288 | $a_0$: 12.11<br>$a_1$: −3.941<br>$a_2$: 0.283 | 7.3-8.9 |
| | 2 | $a_0$: 21.596<br>$a_1$: −6.25<br>$a_2$: 0.429 | $a_0$: 21.867<br>$a_1$: −6.269<br>$a_2$: 0.43 | $a_0$: 21.325<br>$a_1$: −6.23<br>$a_2$: 0.428 | 7.1–8.7 |
| | 3 | $a_0$: 20.043<br>$a_1$: −5.898<br>$a_2$: 0.41 | $a_0$: 20.269<br>$a_1$: −5.912<br>$a_2$: 0.411 | $a_0$: 19.818<br>$a_1$: −5.885<br>$a_2$: 0.409 | 7.0–8.6 |
| | 4 | $a_0$: 20.088<br>$a_1$: −5.922<br>$a_2$: 0.412 | $a_0$: 20.293<br>$a_1$: −5.933<br>$a_2$: 0.413 | $a_0$: 19.883<br>$a_1$: −5.911<br>$a_2$: 0.411 | 6.9–8.5 |
| | 5 | $a_0$: 19.625<br>$a_1$: −5.814<br>$a_2$: 0.406 | $a_0$: 19.842<br>$a_1$: −5.828<br>$a_2$: 0.407 | $a_0$: 19.409<br>$a_1$: −5.801<br>$a_2$: 0.405 | 6.8–8.4 |
| | 5 (high log$U$) | $a_0$: 26.609<br>$a_1$: −7.594<br>$a_2$: 0.519 | $a_0$: 27.039<br>$a_1$: −7.656<br>$a_2$: 0.522 | $a_0$: 26.179<br>$a_1$: −7.533<br>$a_2$: 0.515 | 6.8—8.4 |

**Note.** We fit polynomials of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$, where $x$ is the line ratio and $y$ is the metallicity ($12 + \log(\mathrm{O/H})$).

## Appendix B
## Redshift Evolution of the Strong-line Metallicity Calibrations

Figure 7 is an extension of Figure 4, where we plot the redshift evolution of the 12 line ratios, along with shaded regions representing the 90% confidence interval.

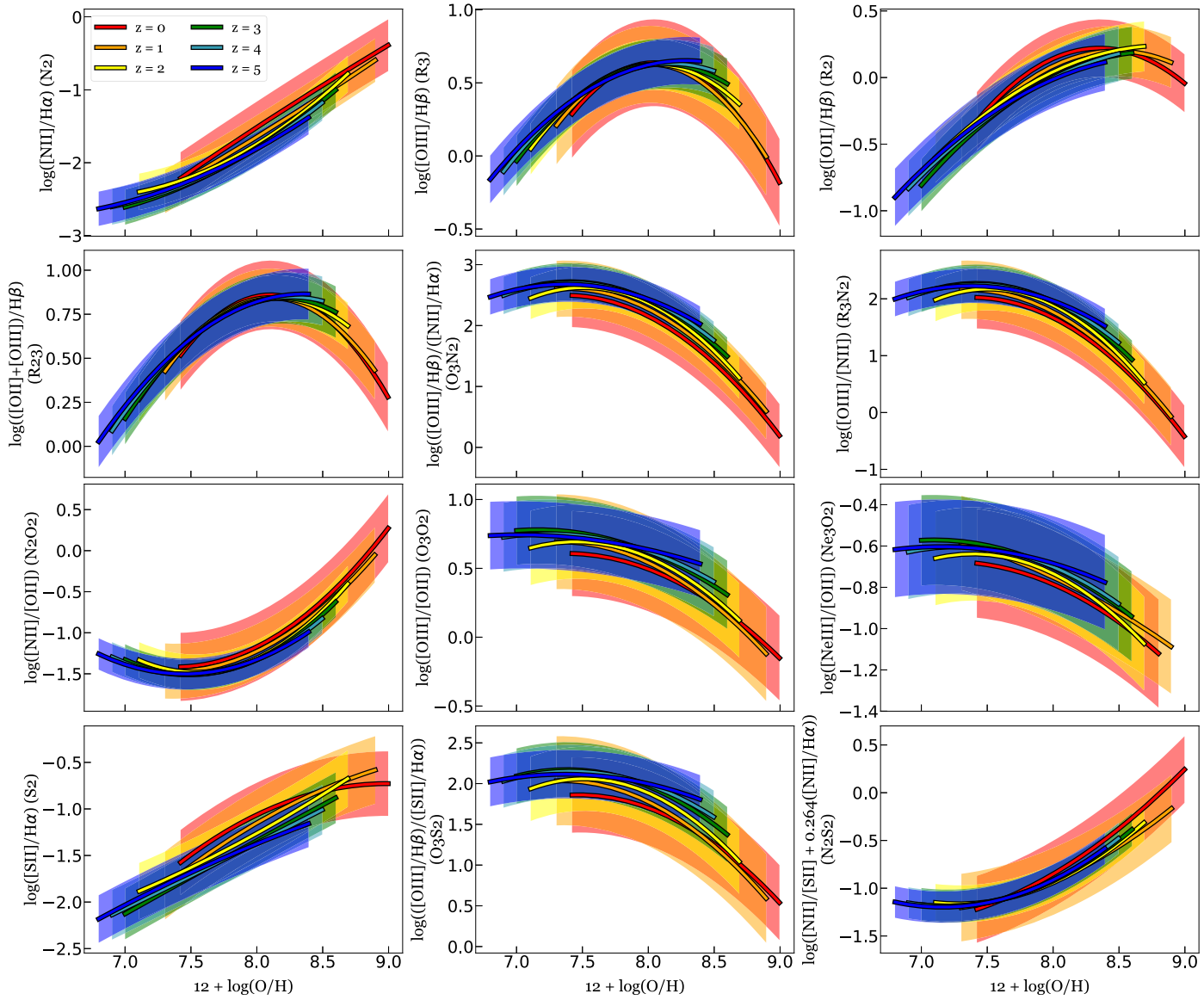


**Figure 7.** The redshift evolution of the strong-line metallicity calibrations using our nebular emission model. For the 12 line ratios, the calibrations going from $z = 0$ to $z = 5$ are shown by the different-colored lines, going from red to blue with increasing redshift. We fit polynomials of the form $y = a_2^* x^2 + a_1^* x^1 + a_0$, where $x$ is the line ratio and $y$ is the metallicity ($12 + \log(O/H)$). The shaded regions denote the 90% confidence intervals in each case. We find that the N2, O3N2, R3N2, N2O2, O3O2, Ne3O2, S2, and O3S2 line ratios show a mild evolution of about 0.2–0.4 dex with redshift. In contrast, indicators like R2, R3, R32, and N2S2 show little redshift evolution. Note: this figure is identical to Figure 4, except the 90% confidence intervals are also shown as shaded regions.

## ORCID iDs

Prerak Garg https://orcid.org/0000-0002-5923-2151
Desika Narayanan https://orcid.org/0000-0002-7064-4309
Ryan L. Sanders https://orcid.org/0000-0003-4792-9119
Romeel Davé https://orcid.org/0000-0003-2842-9434
Gergö Popping https://orcid.org/0000-0003-1151-4659
Alice E. Shapley https://orcid.org/0000-0003-3509-4855
Daniel P. Stark https://orcid.org/0000-0001-6106-5172
Jonathan R. Trump https://orcid.org/0000-0002-1410-0470

## References

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Anglés-Alcázar, D., Davé, R., Faucher-Giguère, C.-A., Özel, F., & Hopkins, P. F. 2017a, MNRAS, 464, 2840
Anglés-Alcázar, D., Faucher-Giguère, C.-A., Kereš, D., et al. 2017b, MNRAS, 470, 4698
Anglés-Alcázar, D., Faucher-Giguère, C.-A., Quataert, E., et al. 2017c, MNRAS, 472, L109
Arellano-Córdova, K. Z., Berg, D. A., Chisholm, J., et al. 2022, ApJL, 940, L23

Bayliss, M. B., Rigby, J. R., Sharon, K., et al. 2014, ApJ, 790, 144
Berg, D. A., Erb, D. K., Auger, M. W., Pettini, M., & Brammer, G. B. 2018, ApJ, 859, 164
Berg, D. A., Skillman, E. D., Croxall, K. V., et al. 2015, ApJ, 806, 16
Bian, F., Kewley, L. J., Groves, B., & Dopita, M. A. 2020, MNRAS, 493, 580
Bondi, H., & Hoyle, F. 1944, MNRAS, 104, 273
Bowen, I. S. 1927, Natur, 120, 473
Bower, R. G., Schaye, J., Frenk, C. S., et al. 2016, MNRAS, 465, 32
Brammer, G. B., Sánchez-Janssen, R., Labbé, I., et al. 2012, ApJ, 758, L17
Bresolin, F. 2007, ApJ, 656, 186
Brinchmann, J. 2023, MNRAS, 525, 2087
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Brown, J. S., Martini, P., & Andrews, B. H. 2016, MNRAS, 458, 1529
Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44
Byler, N., Dalcanton, J. J., Conroy, C., et al. 2019, AJ, 158, 2
Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023, A&A, 677, A115
Chandar, R., Whitmore, B. C., Calzetti, D., & O'Connell, R. 2014, ApJ, 787, 17
Chandar, R., Whitmore, B. C., Dinino, D., et al. 2016, ApJ, 824, 71
Choi, Y., Dalcanton, J. J., Williams, B. F., et al. 2020, ApJ, 902, 54
Christensen, L., Laursen, P., Richard, J., et al. 2012b, MNRAS, 427, 1973
Christensen, L., Richard, J., Hjorth, J., et al. 2012a, MNRAS, 427, 1953
Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833
Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486
Croxall, K. V., Pogge, R. W., Berg, D. A., Skillman, E. D., & Moustakas, J. 2015, ApJ, 808, 42
Croxall, K. V., Pogge, R. W., Berg, D. A., Skillman, E. D., & Moustakas, J. 2016, ApJ, 830, 4
Curti, M., Cresci, G., Mannucci, F., et al. 2017, MNRAS, 465, 1384
Curti, M., D'Eugenio, F., Carniani, S., et al. 2022, MNRAS, 518, 425
Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, MNRAS, 491, 944
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265
Della Bruna, L., Adamo, A., Lee, J. C., et al. 2021, A&A, 650, A103
Denicoló, G., Terlevich, R., & Terlevich, E. 2002, MNRAS, 330, 69
Dopita, M. A., Kewley, L. J., Sutherland, R. S., & Nicholls, D. C. 2016, Ap&SS, 361, 61
Dubois, Y., Volonteri, M., Silk, J., et al. 2015, MNRAS, 452, 1502
Eldridge, J., Stanway, E., Xiao, L., et al. 2017, PASA, 34, e058
Erb, D. K., Pettini, M., Shapley, A. E., et al. 2010, ApJ, 719, 1168
Feltre, A., Charlot, S., & Gutkin, J. 2016, MNRAS, 456, 3354
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Garg, P., Narayanan, D., Byler, N., et al. 2022, ApJ, 926, 80
Gburek, T., Siana, B., Alavi, A., et al. 2019, ApJ, 887, 168
Gilda, S., Lower, S., & Narayanan, D. 2021, ApJ, 916, 43
Haardt, F., & Madau, P. 2012, ApJ, 746, 125
Habouzit, M., Volonteri, M., & Dubois, Y. 2017, MNRAS, 468, 3935
Hainline, K. N., Shapley, A. E., Kornei, K. A., et al. 2009, ApJ, 701, 52
Hayashi, M., Ly, C., Shimasaku, K., et al. 2015, PASJ, 67, 80
Heintz, K. E., Giménez-Arteaga, C., Fujimoto, S., et al. 2023, ApJL, 944, L30
Henry, R. B. C., Stephenson, B. G., Miller Bertolami, M. M., Kwitter, K. B., & Balick, B. 2018, MNRAS, 473, 241
Hirschmann, M., Charlot, S., Feltre, A., et al. 2017, MNRAS, 472, 2468
Hirschmann, M., Charlot, S., & Somerville, R. S. 2023, MNRAS, 526, 3504
Hough, R. T., Rennehan, D., Kobayashi, C., et al. 2023, MNRAS, 525, 1061
Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955
James, B. L., Pettini, M., Christensen, L., et al. 2014, MNRAS, 440, 1794
Johnson, B., Foreman-Mackey, D., Sick, J., et al. 2021, dfm/python-fsps: python-fsps v0.4.1rc1, Zenodo, doi:10.5281/zenodo.4737461
Jones, T., Martin, C., & Cooper, M. C. 2015, ApJ, 813, 126
Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, ApJ, 835, 88
Katz, H., Saxena, A., Cameron, A. J., et al. 2022, MNRAS, 518, 592
Keenan, F. P., Aller, L. H., Bell, K. L., et al. 1996, MNRAS, 281, 1073
Kennicutt, R. C., Jr. 1998, ApJ, 498, 541
Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35
Kojima, T., Ouchi, M., Nakajima, K., et al. 2017, PASJ, 69, 44
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216
Langeroodi, D., Hjorth, J., Chen, W., et al. 2023, ApJL, 957, L39
Larson, K., Díaz-Santos, T., Armus, L., et al. 2020, ApJ, 888, 92
Laseter, I. H., Maseda, M. V., Curti, M., et al. 2024, A&A, 681, A70
Li, Q., Narayanan, D., & Davé, R. 2019, MNRAS, 490, 1425
Lin, Z., Hu, N., Kong, X., et al. 2017, ApJ, 842, 97
Linden, S., Evans, A., Rich, J., et al. 2017, ApJ, 843, 91
Liu, X., Shapley, A. E., Coil, A. L., Brinchmann, J., & Ma, C.-P. 2008, ApJ, 678, 758
Lovell, C. C., Geach, J. E., Davé, R., Narayanan, D., & Li, Q. 2021, MNRAS, 502, 772
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
Maseda, M. V., van der Wel, A., Rix, H.-W., et al. 2014, ApJ, 791, 17
Masters, D., McCarthy, P., Siana, B., et al. 2014, ApJ, 785, 153
Matthee, J., Mackenzie, R., Simcoe, R. A., et al. 2023, ApJ, 950, 67
Mc Leod, A. F., Weilbacher, P. M., Ginsburg, A., et al. 2015, MNRAS, 455, 4057
McAlpine, S., Bower, R. G., Rosario, D. J., et al. 2018, MNRAS, 481, 3118
McCall, M. L., Rybski, P., & Shields, G. 1985, ApJS, 57, 1
McGaugh, S. S. 1991, ApJ, 380, 140
McLennan, J. C., & Shrum, G. M. 1925, RSPSA, 108, 501
Meynet, G., & Maeder, A. 2002, A&A, 381, L25
Monreal-Ibero, A., Walsh, J. R., & Vílchez, J. M. 2012, A&A, 544, A60
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., Shimasaku, K., et al. 2013, ApJ, 769, 3
Narayanan, D., Turk, M. J., Robitaille, T., et al. 2021, ApJS, 252, 12
Oey, M. S., Meurer, G. R., Yelda, S., et al. 2007, ApJ, 661, 801
Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei (Mill Valley, CA: Univ. Science Books)
Pagel, B. E. J., Simonson, E. A., Terlevich, R. J., & Edmunds, M. G. 1992, MNRAS, 255, 325
Patrício, V., Christensen, L., Rhodin, H., Cañameras, R., & Lara-López, M. A. 2018, MNRAS, 481, 3520
Perez-Montero, E. 2014, MNRAS, 441, 2663
Pérez-Montero, E. 2017, PASP, 129, 043001
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pilyugin, L., Vílchez, J., Mattsson, L., & Thuan, T. 2012, MNRAS, 421, 1624
Pilyugin, L. S. 2005, A&A, 436, L1
Péroux, C., & Howk, J. C. 2020, ARA&A, 58, 363
Rahmati, A., Pawlik, A. H., Raičevic, M., & Schaye, J. 2013, MNRAS, 430, 2427
Reddy, N. A., Sanders, R. L., Shapley, A. E., et al. 2023b, ApJ, 951, 56
Reddy, N. A., Topping, M. W., Sanders, R. L., Shapley, A. E., & Brammer, G. 2023a, ApJ, 952, 167
Rhoads, J. E., Wold, I. G. B., Harish, S., et al. 2023, ApJ, 942, L14
Rigby, J. R., Wuyts, E., Gladders, M., Sharon, K., & Becker, G. 2011, ApJ, 732, 59
Robitaille, T. P. 2011, A&A, 536, A79
Runco, J. N., Shapley, A. E., Sanders, R. L., et al. 2021, MNRAS, 502, 2600
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, ApJ, 825, L23
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2018, ApJ, 858, 99
Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2019, MNRAS, 491, 1427
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, ApJ, 962, 24
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2023, ApJ, 955, 54
Sanders, R. L., Shapley, A. E., Zhang, K., & Yan, R. 2017, ApJ, 850, 136
Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, A&A, 665, L4
Shapley, A. E., Reddy, N. A., Sanders, R. L., Topping, M. W., & Brammer, G. B. 2023a, ApJL, 950, L1
Shapley, A. E., Sanders, R. L., Reddy, N. A., Topping, M. W., & Brammer, G. B. 2023b, ApJ, 954, 157
Shapley, A. E., Sanders, R. L., Shao, P., et al. 2019, ApJ, 881, L35
Shirazi, M., Brinchmann, J., & Rahmati, A. 2014, ApJ, 787, 120
Smith, B. D., Bryan, G. L., Glover, S. C., et al. 2017, MNRAS, 466, 2217
Stanghellini, L., Lee, T.-H., Shaw, R. A., Balick, B., & Villaver, E. 2009, ApJ, 702, 733
Stanghellini, L., Shaw, R. A., & Gilmore, D. 2005, ApJ, 622, 294
Stanway, E., & Eldridge, J. 2018, MNRAS, 479, 75
Stark, D. P., Richard, J., Siana, B., et al. 2014, MNRAS, 445, 3200
Stasińska, G., Richer, M. G., & McCall, M. L. 1998, A&A, 336, 667
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, ApJ, 826, 159
Strom, A. L., Steidel, C. C., Rudie, G. C., Trainor, R. F., & Pettini, M. 2018, ApJ, 868, 117
Sun, F., Egami, E., Pirzkal, N., et al. 2023, ApJ, 953, 53
Tacchella, S., Johnson, B. D., Robertson, B. E., et al. 2023, MNRAS, 522, 6236
Tang, M., Stark, D. P., Chen, Z., et al. 2022, MNRAS, 526, 1657
Taylor, A. J., Barger, A. J., & Cowie, L. L. 2022, ApJL, 939, L3

Thompson, R. 2014, pyGadgetReader: GADGET Snapshot Reader for Python, ascl:1411.001
Topping, M. W., Shapley, A. E., Reddy, N. A., et al. 2020a, MNRAS, 499, 1652
Topping, M. W., Shapley, A. E., Reddy, N. A., et al. 2020b, MNRAS, 495, 4430
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Trump, J. R., Haro, P. A., Simons, R. C., et al. 2023, ApJ, 945, 35
Vale Asari, N., Couto, G., Cid Fernandes, R., et al. 2019, MNRAS, 489, 4721
Villar-Martín, M., Cerviño, M., & González Delgado, R. M. 2004, MNRAS, 355, 1132
Wang, X., Jones, T., Vulcani, B., et al. 2022, ApJL, 938, L16
Westmoquette, M. S., James, B., Monreal-Ibero, A., & Walsh, J. R. 2013, A&A, 550, A88
Williams, H., Kelly, P. L., Chen, W., et al. 2023, Sci, 380, 416
Wuyts, E., Rigby, J. R., Gladders, M. D., et al. 2012, ApJ, 745, 86
Yuan, T.-T., & Kewley, L. J. 2009, ApJ, 699, L161
Zaritsky, D., Kennicutt, R. C., Jr, & Huchra, J. P. 1994, ApJ, 420, 87