MNRAS **000**, 1–10 (2025) Preprint 7 April 2025 Compiled using MNRAS LATEX style file v3.3

METALLICITY GRADIENTS IN MODERN GALAXY SIMULATIONS: TENSION BETWEEN SIMULATIONS AND OBSERVATIONS

Alex M. Garcia[1], Paul Torrey[1]
[1] *Department of Astronomy, University of Virginia, 530 McCormick Road, Charlottesville, VA 22904*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The metallicity of galaxies, and its variation with galactocentric radius, provides key insights into the formation histories of galaxies and the physical processes driving their evolution. In this work, we analyze the radial metallicity gradients of star forming galaxies in the EAGLE, Illustris, IllustrisTNG, and SIMBA cosmological simulations across a broad mass ($10^{8.0} M_\odot \leq M_\star \lesssim 10^{12.0} M_\odot$) and redshift ($0 \leq z \leq 8$) range. We find that all simulations predict strong negative (i.e., radially decreasing) metallicity gradients at early cosmic times, likely due to their similar treatments of relatively smooth stellar feedback allowing for sustained inside-out growth. The strongest redshift evolution occurs in galaxies with stellar masses of $10^{10.0} - 10^{11.0} M_\odot$, while galaxies with stellar masses $< 10^{10} M_\odot$ and $> 10^{11} M_\odot$ exhibit weaker redshift evolution. Our results of negative gradients at high-redshift contrast with the many positive and flat gradients in the $1 < z < 4$ observational literature. At $z > 6$, the negative gradients observed with JWST and ALMA are flatter than those in simulations, albeit with closer agreement than at lower redshift. Overall, we suggest that these smooth stellar feedback galaxy simulations may not sufficiently mix their metal content, and that either stronger stellar feedback or additional subgrid turbulent metal diffusion models may be required to better reproduce observed metallicity gradients.

Key words: keyword1 – keyword2 – keyword3

1 INTRODUCTION

A galaxy's metal content is oftentimes not homogeneously distributed. Galaxies in the local Universe typically have more metals in their nuclear regions than the outskirts (e.g., Searle 1971; Zaritsky et al. 1994; Kewley et al. 2010; Berg et al. 2015; Ho et al. 2015; Belfiore et al. 2017; Grasha et al. 2022; Chen et al. 2023; Khoram & Belfiore 2025). This is most likely a product of galaxies forming inside-out: the stellar populations in the interior form and evolve earlier than those of the outskirts, chemically enriching the inner regions first (Prantzos & Boissier 2000; Pilkington et al. 2012; Pérez et al. 2013; Tissera et al. 2019, etc). This radially decreasing distribution of metals is commonly parameterized by a single linear fit (in logarithmic metallicity) and thus characterized by its gradient – the slope of that regression. Given the simplistic inside-out evolutionary view, galaxies should have "negative" gradients. Typical values for the negative gradients of "normal" observed galaxies in the local Universe are around -0.05 ± 0.05 dex/kpc (e.g., Rupke et al. 2010b; Sánchez et al. 2014; Sánchez-Menguiano et al. 2016; Grasha et al. 2022). Not all galaxies exhibit negative metallicity gradients, however; galaxies – both in simulations and observations – undergoing strong inflow/outflow events, pristine gas inflows, and/or mergers can exhibit flattened or even positive gradients (see, e.g., Rupke et al. 2010a,b; Torrey et al. 2012; Ceverino et al. 2016; Tissera et al. 2022; Venturi et al. 2024). Moreover, observations with limited angular resolution, signal-to-noise, and/or spectral resolution add complexity and seem to artificially flatten gradients (see work by Yuan et al. 2013; Mast et al. 2014; Poetrodjojo et al. 2019; Acharyya et al. 2020; Grasha et al. 2022). Galaxies of different masses also exhibit different gradients. Belfiore et al. (2017), using Sloan Digital Sky Survey (SDSS) IV, suggest that galaxies with lower masses ($M_\star \sim 10^{9.5} M_\odot$) and higher masses ($M_\star \sim 10^{11.0} M_\odot$) have flatter gradients than systems of intermediate mass ($M_\star \sim 10^{10.5} M_\odot$), which have the strongest negative gradients.

The picture becomes less clear at higher redshift. Early studies of galaxies around cosmic noon ($z = 1 - 3$) reported a wide variety of positive, flat, and/or negative gradients (e.g., Cresci et al. 2010; Yuan et al. 2011; Queyrel et al. 2012; Swinbank et al. 2012; Troncoso et al. 2014; Wuyts et al. 2016; Wang et al. 2017; Curti et al. 2020b; Simons et al. 2021; Gillman et al. 2022; Li et al. 2022; Dutta et al. 2024, with gradients typically falling between ±0.1 dex/kpc). Here, the limited spatial and spectral resolution becomes even more of a hindrance without the use of gravitationally lensed systems (as in Jones et al. 2010, 2013; Yuan et al. 2011) or adaptive optics (as in Swinbank et al. 2012). It was therefore extremely difficult, up until recently, to make any robust statements about metallicity gradients during this epoch. As such, no clear consensus on the metallicity gradient evolution of galaxies in the Universe has been established.

The recent launch and successful deployment of JWST serves to provide unprecedented fidelity at higher redshifts to help alleviate the systematic issues with the previous generation of observations. Some work has already been done on characterizing gradients with JWST at $z = 1 - 4$ (Wang et al. 2022; Ju et al. 2025; Morishita et al. 2024; Rodríguez Del Pino et al. 2024) and beyond ($z > 6$; Arribas et al. 2024; Venturi et al. 2024). While the sample sizes remain limited, the tentative scenario is that, indeed, there are a wide assortment of metallicity gradients in the early Universe. However, unlike those of previous $z > 3$ studies, not all observed metallicity gradients are positive. In fact, to date, only one system with a positive gradient has been observed at $z > 6$ (Venturi et al. 2024, though a few have uncertainties that encompass positive gradient values).

The current theoretical understanding of the evolution of gradients from simulations (Gibson et al. 2013; Taylor & Kobayashi 2017; Tissera et al. 2016, 2019, 2022; Hemler et al. 2021; Acharyya et al. 2025; Ibrahim & Kobayashi 2025, etc) and other analytic models (e.g., Mott et al. 2013; Kubryk et al. 2015; Mollá et al. 2019; Sharda et al. 2021b,a) is similarly heterogeneous. Some models predict that gradients should become positive at $z > 3$ (Mott et al. 2013, though this depends on assumptions about star formation efficiencies), some predict that gradients should become more negative with increasing redshift (Gibson et al. 2013; Taylor & Kobayashi 2017; Hemler et al. 2021, though the detailed evolution depends strongly on stellar feedback implementation of the model), and others still predict that there is virtually no evolution of gradients with time (Ma et al. 2017; Tissera et al. 2019; Sharda et al. 2021a; Sun et al. 2024). Moreover, relatively little work has been done analyzing metallicity gradients at very high redshift ($z \gtrsim 3$; although it is not completely unexplored, see, e.g., Taylor & Kobayashi 2017; Ibrahim & Kobayashi 2025). How metallicity gradients evolve through time is thus very much an open question.

In this paper, we begin this effort to provide a more complete and comprehensive view of the spatial variations of the metal content within current galaxy simulation paradigm. This work, which is the first of a series, focuses on several of the current class of large box simulations that contain a large dynamic range of galaxies. The remainder of this paper is organized as follows: in Section 2 we describe each simulation model, our selection criteria, and methodology for deriving metallicity gradients. In Section 3, we present the redshift evolution of the metallicity gradients in each simulation model as well as break down evolution in different stellar mass bins. In Section 4, we compare our results to recent observations.

2 METHODS

2.1 Simulations

This work uses data products from the EAGLE, Illustris, IllustrisTNG, and SIMBA cosmological simulations of galaxy formation and evolution. Each of these simulations models a range of astrophysical processes including gravity, cosmology, star formation, stellar evolution, stellar feedback, chemical enrichment of the ISM, and black hole growth and feedback. We therefore dedicate this section to describing each model, cautioning that this section is not meant to be a complete enumeration of all of the details of each model; rather it is a brief description of each of the physical models with details pertinent to the results presented in this work.

We first note that these models share the commonality of having a "subgrid" prescription for the ISM, owing to their limited resolution. This subgrid prescription, modeled with an effective equation of state, sets the behavior of the cold, dense gas where star formation takes place. As a consequence, all the models in this work have relatively 'smooth' stellar feedback that is persistent in time, yet not terribly destructive. This allows for relatively smooth gas cycling as galaxies assemble (see, e.g., Torrey et al. 2018; Garcia et al. 2024a,b, 2025, though there are differences in the implementations, see Wright et al. 2024 for a careful examination). This is contrasted with high-resolution explicit ISM models that directly model the sites of star formation and feedback, such as the FIRE simulations (Hopkins et al. 2014, 2018, 2023). These explicit ISM models naturally produce much stronger feedback resulting in episodic bursts that eject gas far into the CGM around the galaxy (Muratov et al. 2015, 2017; Anglés-Alcázar et al. 2017; Pandya et al. 2021). We discuss the potential implications of using this smooth feedback paradigm in Section 4.1.1.

Moreover, none of these simulations have a method by which unresolved turbulence can exchange mass and metals throughout the galaxy. Mass can move along with bulk galactic winds, but the small-scale turbulent eddies that drive diffusion within the ISM (Smagorinsky 1963; Shen et al. 2010; Semenov et al. 2016; Su et al. 2017; Escala et al. 2018; Semenov 2024) are not modeled in any of the simulations in this work. We consider what role this (lack of) unresolved turbulence may play in setting the gradients more carefully in Section 4.1.2.

2.1.1 *EAGLE*

The EAGLE (Crain et al. 2015; Schaye et al. 2015; McAlpine et al. 2016) simulations are built upon a modified version of the smooth particle hydrodynamics (SPH) code GADGET-3 (Springel 2005) code called ANARCHY (Schaye et al. 2015). As mentioned above, the ISM in EAGLE is treated with an effective equation of state (Schaye & Dalla Vecchia 2008). Star formation in EAGLE is thus restricted to gas cells with

$$n_{\rm H} \geq 10^{-1} \left(\frac{Z}{0.002}\right)^{-0.64} \tag{1}$$

and $\log T < \log T_{\rm eos} + 0.5$, where $n_{\rm H}$ is the hydrogen number density of the gas, T is the gas temperature, Z is the metallicity of the gas, and $T_{\rm eos} = 8{\times}10^3$ K (Schaye 2004; Dalla Vecchia & Schaye 2012; Schaye et al. 2015). Stars form from this gas according to a Chabrier (2003) initial mass function (IMF) and evolve according to Wiersma et al. (2009) evolutionary tracks. Mass and metals are sent back into the ISM via asymptotic giant branch (AGB) winds as well as Type Ia and II supernovae. The EAGLE model explicitly tracks the production and evolution of eleven chemical species (H, He, C, N, O, Ne, Mg, Si, Fe, S, and Ca) coming from yield tables in Marigo (2001) for AGB winds, Thielemann et al. (2003) for Type Ia supernovae, and Portinari et al. (1998) for Type II supernovae. The fluid is modeled as discrete particles in SPH codes (like that of EAGLE) and as such newly ejected metals are locked into the cell in which they form.

The full EAGLE suite is comprised of several different runs with varied resolution, box size, and calibrations. In this work, we analyze the flagship $(67.8~{\rm Mpc}/h)^3$ high-resolution box (RefL0100N1504; hereafter synonymous with EAGLE itself), which has 2×1504^3 particles and an initial baryon mass resolution of $1.81 \times 10^6 M_\odot$.

2.1.2 *Illustris*

The original Illustris suite of simulations (Vogelsberger et al. 2013, 2014a,b; Genel et al. 2014; Torrey et al. 2014) was run using the Moving Voronoi Mesh (MVM) code AREPO (Springel 2010). The dense, star forming ISM is treated with the Springel & Hernquist (2003) effective equation of state in Illustris. Stars form stochastically in the dense ($n_{\rm H} > 0.13~{\rm cm}^{-3}$) ISM according to a Chabrier (2003) IMF. The stellar evolutionary tracks are taken from Portinari et al. (1998) and depend on both the mass and metallicity of the stars. The Illustris model explicitly tracks nine chemical species (H, He, C, N, O, Ne, Mg, Si, and Fe). Stars return their mass and metals to the ISM through AGB winds and supernovae. The metal yields for

AGB winds come from Karakas (2010), Type Ia supernovae from Thielemann et al. (2003), and Type II supernovae from Portinari et al. (1998). Finally, we note that the Voronoi mesh structure of AREPO naturally allows for metals to exchange across boundaries as the cells deform and move. However, as mentioned previously, there is no model for allowing metals to do so from unresolved turbulence.

The full Illustris suite is comprised of several $(75\ \mathrm{Mpc}/h)^3$ boxes of varying resolution. Here we analyze the highest resolution run (Illustris-1, hereafter synonymous with Illustris itself) which has 2×1820^3 particles and an initial baryon mass resolution of $1.26 \times 10^6 M_\odot$.

2.1.3 *IllustrisTNG*

IllustrisTNG (hereafter TNG; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018; Pillepich et al. 2019; Nelson et al. 2019a,b) is a suite of cosmological box simulations and is the successor to the original Illustris simulations. The two models are thus very similar in spirit and there are key similarities, as well as differences, between the models that we will enumerate in this section (see Weinberger et al. 2017; Pillepich et al. 2018a, for a detailed comparison of the two models). Star formation in TNG follows from the same Springel & Hernquist (2003) effective equation of state as Illustris. Moreover, the stars formed follow the same Chabrier (2003) IMF. There are changes in the treatment of stellar feedback as well as chemical enrichment, however. The Illustris model set the minimum mass for core-collapse supernovae at $6M_\odot$, whereas TNG raises this to $8M_\odot$. This accounts for a 30% decrease in Type II supernovae in TNG. TNG tracks the same nine chemical species as Illustris, with an additional tenth "other metals" field to account for untracked metals. Many of the metal yield tables are updated from Illustris. TNG adopts yields for Type Ia supernovae from Nomoto et al. (1997), Type II supernovae from Portinari et al. (1998) and Kobayashi et al. (2006), and AGB winds from Karakas (2010), Doherty et al. (2014), and Fishlock et al. (2014). Just as in Illustris, TNG has no built-in model for the unresolved turbulence in the ISM beyond that naturally present in its MVM implementation.

The full suite of TNG simulations is comprised of several different resolution runs as well as varied box sizes. We use the highest resolution run of the $(35\ \mathrm{Mpc}/h)^3$ volume (i.e., TNG50-1), which has $2{\times}2160^3$ resolution elements and an initial baryon mass resolution of $8.5 \times 10^4 M_\odot$. We note that we will use the term TNG synonymously with the TNG50-1 simulation throughout this work.

2.1.4 *SIMBA*

The SIMBA simulations (Davé et al. 2019) are the successor to the MUFASA simulations (Davé et al. 2016) and are run using the meshless finite mass (MFM) code GIZMO (Hopkins 2015). Star formation in SIMBA is set by a molecular column density (and metallicity) of the gas cells with densities $n_{\mathrm{H}} \geq 0.13\ \mathrm{cm}^{-3}$ (adapted from Krumholz, McKee & Tumlinson 2009; Krumholz & Gnedin 2011). Newly formed star particles inherit their metallicity from their natal gas. SIMBA explicitly tracks the same eleven chemical species as EAGLE. Feedback from stars is implemented in the form of AGB winds consistent with a Chabrier (2003) IMF as well as Type Ia and II supernovae. The metal yield tables come from Oppenheimer & Davé (2006) for AGB winds, Iwamoto et al. (1999) for Type Ia supernovae, and Nomoto et al. (2006) for Type II supernovae. Unique to the SIMBA model (among simulations in this work) is that it also includes an explicit tracking of dust production, growth, and destruction (following Dwek 1998). This is important for the purposes of this work since some of the metals produced in stars will be locked in dust and therefore not be accounted for in the gas-phase metallicity of the system. MFM codes, such as GIZMO, also do *not* have a natural scheme for advecting metals from particle to particle.

The full SIMBA suite consists of several different box size runs. In this work, we make use of a $(50\ \mathrm{Mpc}/h)^3$ box with 2×512^3 particles (m50n512; hereafter simply SIMBA) and an initial baryon mass resolution of $1.28 \times 10^7 M_\odot$.

2.2 Selection Criteria

We utilize SUBFIND (Springel et al. 2001) catalogs for each simulation to identify gravitationally bound substructure. We apply a resolution cut of $\sim 10^3$ to both the gas and stellar elements for a galaxy to be considered "well-resolved". In TNG, this corresponds to $M_\star$ (and M_{gas}) $> 10^8 M_\odot$ whereas in EAGLE and Illustris it is $M_\star > 10^9 M_\odot$ and in SIMBA it is $M_\star > 10^{10} M_\odot$. We further require that each galaxy has a SFR > 0 $[M_\odot/\mathrm{yr}]$. We note that our gradient definitions require that the SFR be significant and spatially extended (see discussion below in Section 2.3). This tends to naturally raise the minimum required SFR. Regardless, we do not expect slight changes to this star forming galaxy selection (e.g., a specific SFR cut or star forming main sequence cut as in other works) to significantly change the samples which we draw. Rather, the SFR > 0 $[M_\odot/\mathrm{yr}]$ requirement is a practical cut made such that: (i) the galaxy would likely contain bright emission lines used in observational surveys and (ii) our definitions of the metallicity gradient region are well-posed.

2.3 Metallicity Gradient Definitions

The methodology related to defining metallicity gradients in the simulations derives heavily from a combination of previous theoretical studies by Ma et al. (2017) and Hemler et al. (2021). We utilize these well-tested methods in order to make as fair a comparison to observational studies as possible. We provide a visual summary of the definitions for a TNG galaxy at $z = 5$ in Figure 1.

We first center the galaxies by placing the location of the particle with the potential minimum at the origin (using `SubhaloPos` from the SUBFIND catalogs). We then define two characteristic radii: R_{in}, the radius enclosing 5% of the star formation in the galaxy, and R_{out}, the distance enclosing 90% of star formation within 10 kpc. We rotate the galaxy to the face-on orientation by computing the direction of the angular momentum vector of the galaxy via all star forming cells within $R_{\mathrm{in}} < r < R_{\mathrm{out}}$ (i.e., the star-forming disk). We then orient the system so that this vector is pointing in the $+\hat{z}$ direction. Next, we construct 2D mass-weighted metallicity maps of each of the rotated galaxies. We use maps with "pixels" 0.5 kpc × 0.5 kpc.[1] We remove pixels in these maps with gas surface densities of $\Sigma_{\mathrm{gas}} < 10^6 M_\odot\ \mathrm{kpc}^{-2}$ (as these regions of low density gas are unlikely to contain star-forming regions; Ma et al. 2017) and deproject the 2D map into a radial profile (as in the gray background histograms of the bottom panels of Figure 1).

We further reduce the radial profile into a single median relation following directly from Hemler et al. (2021). We generate a median profile in bins of 0.1 kpc by searching a region of $\Delta r = \pm 0.05$ kpc

---

[1] We note that this pixel resolution was chosen to accommodate the SIMBA mass and spatial resolutions. Creating smaller pixels is possible in EAGLE, Illustris, and TNG; however, we confirm that varying the pixel size to smaller values (i.e., 0.25 kpc × 0.25 kpc or 0.1 kpc × 0.1 kpc) does not significantly impact the core results in these models.

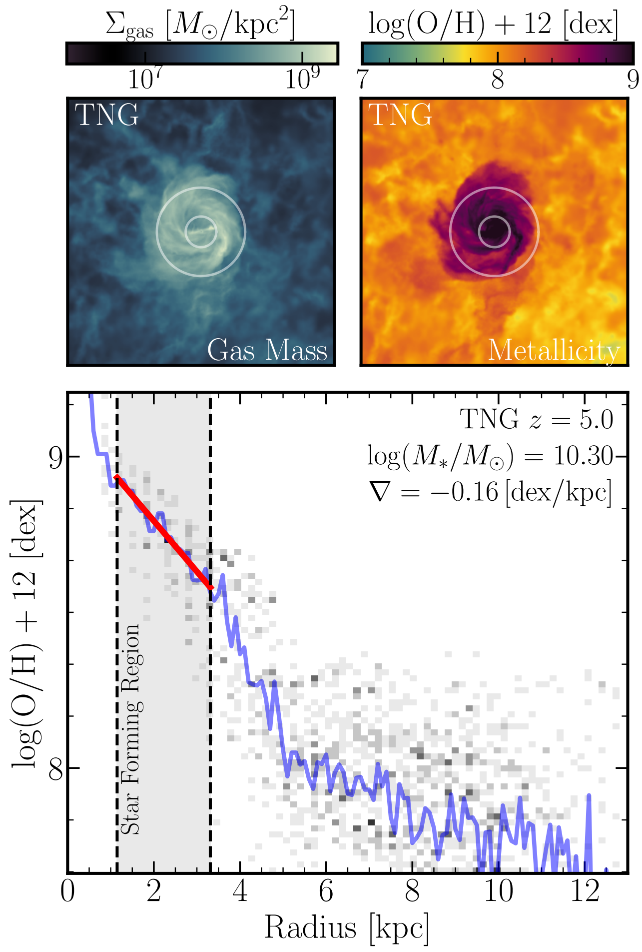

Figure 1. Demonstration of Metallicity Gradient Derivation. Projections of a galaxy at $z = 5$ from TNG in gas mass (top left) and metallicity (top right) as well as the resulting radial profile after creating 0.5 kpc × 0.5 kpc pixel maps of the system (gray 2D histogram in bottom panel). The median 1D metallicity profile is shown in the blue line in the bottom panel. The gray shaded region represents the "star forming region" (see Section 2.3 for definition) and is the region over which we fit the linear regression to obtain the metallicity gradient, ∇ (note that the circles in the top panels also show this star forming region, roughly corresponding to the disk of the galaxy). The straight red line within the star forming region corresponds to the best-fit gradient (in this case $\nabla = -0.16$ dex/kpc).

for 4 populated pixels. If the required number of valid pixels are not found, we increase Δr to ±0.125 kpc, then ±0.25 kpc, then ±0.5 kpc. If this criteria is not satisfied within ±0.5 kpc, that radial bin is removed. The result of this median profile fitting is shown as the blue line in Figure 1.

Finally, we calculate the metallicity gradient using this median metallicity profile. We fit a region of $R'_{\rm in} < r < R_{\rm out}$ (where $R'_{\rm in} = R_{\rm in} + 0.25[R_{\rm out} - R_{\rm in}]$) which we henceforth refer to as the "star forming region" of the galaxy. Not fitting the inner quarter of the star forming disk avoids the central regions of the galaxy that deviate significantly from the outer three-quarters, which can be seen in both bottom panels of Figure 1 and follows directly from previous work (Pilkington et al. 2012; Gibson et al. 2013; Ma et al. 2017; Hemler et al. 2021). Moreover, we stop the fitting at $R_{\rm out}$ as the outer regions also deviate significantly from the star forming disk. Specifically, gradients (both in simulations and observations) seem to flatten out at large galactocentric radii (Bresolin et al. 2009; Sánchez et al. 2014; Belfiore et al. 2016; Sánchez-Menguiano et al. 2018; Grasha et al. 2022; Tapia et al. 2022; Chen et al. 2023; Garcia et al. 2023). Thus, we fit the median profile only in the star forming region with a single linear regression (red line in Figure 1). The slope of this line is what we use as the metallicity gradient of the system.

We also require that the star forming region be at least 1 kpc (in the 1D profile). We note that 1 kpc is potentially large for galaxies at very high redshift (Ormerod et al. 2024); however, even by relaxing this assumption, we find that galaxies below this size tend to fail other selection criteria. This criterion is therefore made for practical reasons. Higher resolution simulations are needed to characterize galaxies with more compact star-forming regions. Finally, we require the median profile to be mostly contiguous by demanding that there be valid data covering at least 90% of the star forming region.

The redshift ranges of each simulation sample are not necessarily consistent owing to the application of the above methodology on varying mass and spatial resolutions. In particular, we report the distribution of gradients in EAGLE at $0 \leq z \leq 7$, Illustris at $0 \leq z \leq 8$, TNG at $0 \leq z \leq 8$, and SIMBA at $0 \leq z \leq 4$.

3 RESULTS

3.1 The Redshift Evolution of Metallicity Gradients

The top panel of Figure 2 shows the distribution of metallicity gradients in each simulation as a function of redshift. We characterize the distributions with their median and their spread with the 16th and 84th percentiles. We note that we require there to be at least 10 galaxies that pass our selection criteria to report the distribution of galaxies. These data are also presented in Table 1 for ease of reference, along with the number of galaxies in each redshift bin that pass the selection criteria. We note that Hemler et al. (2021) quote the peak of a log normal distribution instead of a median. At $z \sim 0$, the distributions are indeed fairly well characterized by a log normal; however, at higher redshift the distributions become significantly less log normally distributed. The net effect of this is that the median is at slightly steeper values than the peak of the log normal distributions. Figure 2 also has a compilation of observed metallicity gradients (gray squares), we make direct comparisons to these gradients in Section 4.1.

Generally speaking, each of the simulations has similar behavior: increasingly negative gradients with increasing redshift. This evolution is roughly linear in redshift space in EAGLE, Illustris, and SIMBA. In TNG, on the other hand, there is a plateau at $z = 4 - 5$ before the gradients start to become flatter towards $z = 8$. We obtain gradient evolutions of -0.015 ± 0.01 dex/kpc/Δz, -0.016 ± 0.001 dex/kpc/Δz, -0.016 ± 0.004 dex/kpc/Δz, and -0.028 ± 0.007 dex/kpc/Δz in EAGLE, Illustris, TNG, and SIMBA, respectively, by fitting a linear regression to the median gradients at each redshift (weighted by the number of galaxies in each redshift bin; see the bottom row of Table 1).

The bottom panel of Figure 2 shows the scatter of each distribution (taken as the difference between the 84th and 16th percentiles of the distributions). EAGLE, Illustris, and SIMBA share the most similar trends in the scatter about their distributions. In EAGLE, the scatter increases roughly linearly with increasing redshift. The scatter at $z = 0$ is 0.08 dex/kpc and increases to > 0.3 dex/kpc at $z = 7$. So, too, does the scatter in Illustris increase roughly linearly with increasing redshift starting quite small at ~ 0.03 dex/kpc to ~ 0.25 dex/kpc

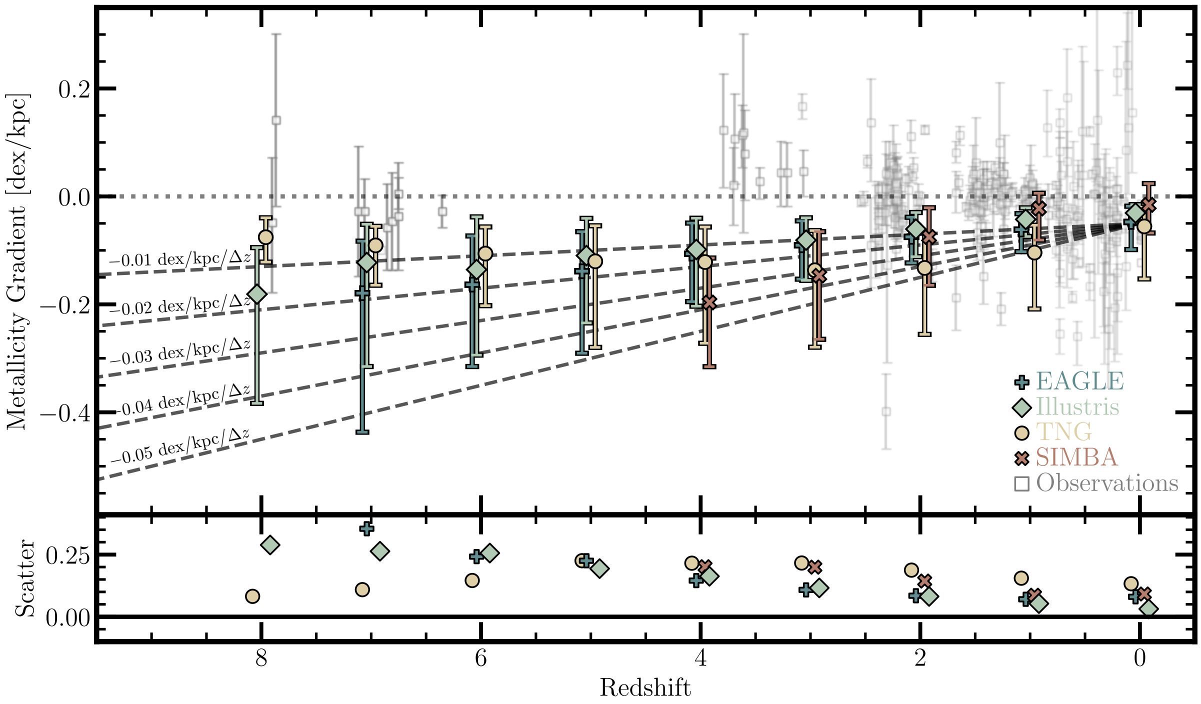

Figure 2. The Redshift Evolution of Metallicity Gradients in EAGLE, Illustris, TNG, and SIMBA. *Top:* The median, 16th and 84th percentile of the gradient distributions in EAGLE (plus), Illustris (diamond), TNG (circle), and SIMBA (x). We also include evolutionary lines of −0.01, −0.02, −0.03, −0.04, and −0.05 dex/kpc/Δz (with intercepts of −0.05 dex/kpc) as a point of reference. The observed metallicity gradients (unfilled squares) are a collection of Rupke et al. (2010b); Queyrel et al. (2012); Swinbank et al. (2012); Jones et al. (2013, 2015); Troncoso et al. (2014); Leethochawalit et al. (2016); Wang et al. (2017, 2019, 2022); Carton et al. (2018); Förster Schreiber et al. (2018); Curti et al. (2020b); Grasha et al. (2022); Li et al. (2022); Arribas et al. (2024); Ju et al. (2025); Vallini et al. (2024); Venturi et al. (2024). We emphasize that the errorbars on the observations are uncertainty on individual gradient measurements, whereas the errorbars on the simulation points are the spread of the distributions. Additionally, we note that the simulation data points are slightly offset from their respective redshifts for aesthetic purposes; however, all simulation data points are at integer redshifts. *Bottom:* The scatter about the median for each redshift in each simulation, defined as magnitude of difference between 84th and 16th percentiles. We note that the median, 16th percentile, and 84th percentile for each simulation can also be found in Table 1.

	EAGLE		Illustris		TNG		SIMBA	
	N	∇ [dex/kpc]	N	∇ [dex/kpc]	N	∇ [dex/kpc]	N	∇ [dex/kpc]
$z = 0$	5851	$-0.047^{-0.018}_{-0.098}$	19101	$-0.031^{-0.015}_{-0.047}$	3507	$-0.056^{-0.020}_{-0.153}$	389	$-0.016^{+0.024}_{-0.068}$
$z = 1$	7673	$-0.061^{-0.032}_{-0.102}$	16105	$-0.043^{-0.021}_{-0.074}$	4675	$-0.104^{-0.053}_{-0.209}$	342	$-0.022^{+0.006}_{-0.081}$
$z = 2$	5762	$-0.075^{-0.039}_{-0.123}$	10364	$-0.061^{-0.030}_{-0.112}$	4661	$-0.133^{-0.068}_{-0.256}$	167	$-0.075^{-0.021}_{-0.165}$
$z = 3$	3087	$-0.091^{-0.045}_{-0.153}$	5557	$-0.081^{-0.040}_{-0.156}$	3651	$-0.137^{-0.063}_{-0.279}$	77	$-0.147^{-0.065}_{-0.265}$
$z = 4$	1432	$-0.106^{-0.049}_{-0.195}$	2591	$-0.098^{-0.040}_{-0.203}$	2317	$-0.122^{-0.057}_{-0.272}$	29	$-0.197^{-0.114}_{-0.316}$
$z = 5$	461	$-0.139^{-0.065}_{-0.291}$	982	$-0.110^{-0.041}_{-0.234}$	1243	$-0.120^{-0.054}_{-0.280}$		
$z = 6$	150	$-0.164^{-0.073}_{-0.315}$	325	$-0.136^{-0.038}_{-0.294}$	583	$-0.106^{-0.057}_{-0.202}$		
$z = 7$	34	$-0.179^{-0.083}_{-0.437}$	82	$-0.123^{-0.052}_{-0.315}$	230	$-0.091^{-0.055}_{-0.165}$		
$z = 8$			14	$-0.181^{-0.095}_{-0.384}$	73	$-0.076^{-0.040}_{-0.122}$		
dex/kpc/Δz	-0.015 ± 0.001		-0.016 ± 0.001		-0.016 ± 0.004		-0.028 ± 0.007	

Table 1. Average Gradients in the Samples. The number of galaxies (N) and median metallicity gradient (∇) for every redshift where there are more than 10 galaxies that pass our selection criteria (outlined in Section 2.2). The subscripts and superscripts on the median gradients are the 16th and 84th percentiles of the distributions. The bottom row shows the average evolution of the gradients in each simulation (regression via medians weighted by the number of galaxies in each bin). The quoted uncertainties are the square root of the variance of the slope, taken from the covariance matrix. These data are also shown graphically in Figure 2.

at $z = 8$. Finally, SIMBA follows the pattern with a roughly linearly increasing scatter with increasing redshift. The 16th percentile of the distribution in particular changes more significantly than the 84th percentile, suggesting there are more steep negative gradients at higher redshift than lower in EAGLE, Illustris, and SIMBA. TNG breaks the pattern by increasing linearly with increasing redshift to around $z \sim 4$ (around 0.25 dex/kpc), plateauing to $z = 5$, and then decreasing back to $z = 8$ (to a small 0.08 dex/kpc). This behavior is qualitatively very similar to that of the overall gradients in TNG, which become steeper back to $z = 4$, plateau to $z = 5$, and then become flatter back to $z = 8$. It is likely that this behavior, both in the scatter and the medians, is a result of the mass distribution of the sample changing with increasing redshift, as we will discuss in more detail in the next section.

It should be remarked that, for the most part, the qualitative behavior of the simulation models is really quite similar, despite the quantitative differences. The level of qualitative agreement is, perhaps, expected based on previous results (see, e.g., Gibson et al. 2013). Each of the models of this work employs an ISM treatment that gives rise to smooth stellar feedback (see Section 2.1) which allows for (i) the enhancement of nuclear metallicity through inside-out growth and (ii) the persistence of the radially decreasing metallicity profiles through cosmic time.

3.2 Gradient Evolution By Mass

The direct comparison of gradients in different simulations above makes two implicit assumptions: (i) that the evolution of gradients is the same in galaxies across subpopulations and (ii) that the samples in each simulation contain the same subpopulations. These are not necessarily the case, however. As noted in Section 2.2, each simulation sample has a slightly different minimum stellar mass owing to the varied mass resolutions. To that end, this section first addresses how the gradients of galaxies in different mass bins evolve and then investigates the implications this has on the interpretation of the results of Section 3.1.

Figure 3 shows the evolution of metallicity gradients of all four simulations broken down into four different stellar mass bins: $10^{8.0}M_\odot \leq M_\star < 10^{9.0}M_\odot$ (top left), $10^{9.0}M_\odot \leq M_\star < 10^{10.0}M_\odot$ (top right), $10^{10.0}M_\odot \leq M_\star < 10^{11.0}M_\odot$ (bottom left), and $10^{11.0}M_\odot \leq M_\star < 10^{12.0}M_\odot$ (bottom right). We first note that each simulation has qualitatively similar behavior. Generally, we find that galaxies in the lowest mass bins ($M_\star < 10^9 M_\odot$) tend to have virtually no redshift evolution ($\sim$ 0.00 dex/kpc/Δz). Galaxies in the highest mass bins ($10^{11.0}M_\odot \leq M_\star < 10^{12.0}M_\odot$; bottom right) also show very little redshift evolution. The caveat here, of course, is that there are virtually no galaxies in any of the simulations here at stellar masses of greater than $10^{11.0}M_\odot$ at $z > 3$. The extent to which this trend holds at $z > 3$ is therefore not clear. The intermediate mass bins ($10^{9.0}M_\odot \leq M_\star < 10^{10.0}M_\odot$; top right, and $10^{10.0}M_\odot \leq M_\star < 10^{11.0}M_\odot$; bottom left) have much stronger evolution. While there are quantitative differences (which we will discuss more below), the generality of this result across the different simulation models is quite remarkable.

Looking more quantitatively, we can fit the evolution of the gradients with a linear regression in each simulation and in each mass bin in the same way as in Section 3.1 (shown in Table 2). The quantitative trends confirm those of the qualitative trends: the $10^{8.0}M_\odot \leq M_\star < 10^{9.0}M_\odot$ bin has the weakest redshift evolution[2], $10^{10.0}M_\odot \leq M_\star < 10^{11.0}M_\odot$ bin has the strongest, with the other two mass ranges having moderate evolution. Moreover, this trend is qualitatively similar as that of Belfiore et al. (2017) at $z \sim 0$ in SDSS, who find that gradients are the most negative in galaxies with stellar masses of $10^{10.0}M_\odot < M_\star < 10^{10.5}M_\odot$ (see, e.g., their Figure B1). Although, it should be noted that in the simulations at $z = 0$ galaxies with stellar masses ranging from $10^{9.0}M_\odot < M_\star < 10^{10.0}M_\odot$ tend to have the most negative gradients.

4 DISCUSSION

4.1 Comparison with Observations

As a point of comparison, we also include the observational measurements of metallicity gradients in Figure 2 (unfilled gray squares; data from Rupke et al. 2010b; Swinbank et al. 2012; Jones et al. 2013, 2015; Troncoso et al. 2014; Leethochawalit et al. 2016; Wang et al. 2017, 2019, 2022; Carton et al. 2018; Curti et al. 2020b; Grasha et al. 2022; Arribas et al. 2024; Ju et al. 2025; Vallini et al. 2024; Venturi et al. 2024). Broadly speaking, the observed metallicity gradients fall into three different categories: (i) a wide diversity of gradients at $z \lesssim 2.5$ (Rupke et al. 2010b; Swinbank et al. 2012; Jones et al. 2013, 2015; Leethochawalit et al. 2016; Wang et al. 2017, 2019; Carton et al. 2018; Curti et al. 2020b; Grasha et al. 2022; Ju et al. 2025), (ii) positive gradients at $z \approx 3-4$ (Troncoso et al. 2014; Wang et al. 2022) and (iii) mostly negative, albeit relatively flat, gradients at $z > 6$ (Arribas et al. 2024; Vallini et al. 2024; Venturi et al. 2024). Notably, there are presently no measured metallicity gradients in the range $4 \lesssim z \lesssim 6$. In the following subsections, we make detailed comparisons between the simulations and observations at $0 < z < 4$ and then at $z > 6$.

We first caution that the metallicity values measured in these simulations are mass-weighted, whereas those of observations are light-weighted. To help accommodate for this difference, we required all valid pixels in our 2D maps construction to have significant gas content that is likely to host star forming gas (see Section 2.3 and Ma et al. 2017). The pixels we obtain are therefore all would likely contain the emission lines required to obtain metallicities; however, they are not weighted by this factor in our analysis.

4.1.1 Comparison with Observed Metallicity Gradients at $0 < z < 4$

In the local Universe ($z \sim 0$), the observed metallicity gradients are typically negative with values of ~ -0.05 dex/kpc (Rupke et al. 2010b; Belfiore et al. 2017; Grasha et al. 2022). The results from the various simulations presented here are well in line with the observations. Indeed, the median gradients in EAGLE and TNG at $z = 0$ (-0.047 dex/kpc and -0.056 dex/kpc, respectively) are remarkably good matches. However, it has also been observed that the strength of the negative gradients does seem to depend on the mass of the galaxies (Belfiore et al. 2017). In particular, galaxies with stellar masses of $10^{10.0}M_\odot < M_\star < 10^{10.5}M_\odot$ have the strongest negative gradients. The mass dependence is seen in the simulations, too, albeit with a slight shift where the most negative gradients tend to be in slightly lower mass bins (i.e., $10^{9.0}M_\odot < M_\star < 10^{10.0}M_\odot$).

At increasing redshift, observed galaxies tend to have a wider diversity of negative, flat, and positive metallicity gradients (Swinbank et al. 2012; Jones et al. 2013, 2015; Leethochawalit et al. 2016; Wang et al. 2017, 2019; Förster Schreiber et al. 2018; Curti et al. 2020b; Li et al. 2022; Dutta et al. 2024; Ju et al. 2025). In the range $0 < z \lesssim 1$ gradients are still predominantly negative (Swinbank et al. 2012; Carton et al. 2018). The metallicity gradients from the four simulation models are broadly similar to the observations in this redshift range, although it should be noted that the simulations do not produce as many positive gradients. At even higher redshifts ($1 \lesssim z \lesssim 2.5$), the observational trend persists with even more gradients being either flat or positive (Queyrel et al. 2012; Swinbank et al. 2012; Jones et al. 2015; Leethochawalit et al. 2016; Wang et al. 2017; Curti et al. 2020b; Li et al. 2022; Ju et al. 2025, although strong negative gradients have been observed, e.g., Jones et al. 2013; Wang et al. 2017;

---

[2] It should be noted though that there is a slight 'dip' in the evolution of the gradients at $z \sim 2-3$ in the $10^{8.0}M_\odot \leq M_\star < 10^{9.0}M_\odot$ bin. This implies that galaxies of this mass should have their steepest negative gradients at $z \sim 2-3$.

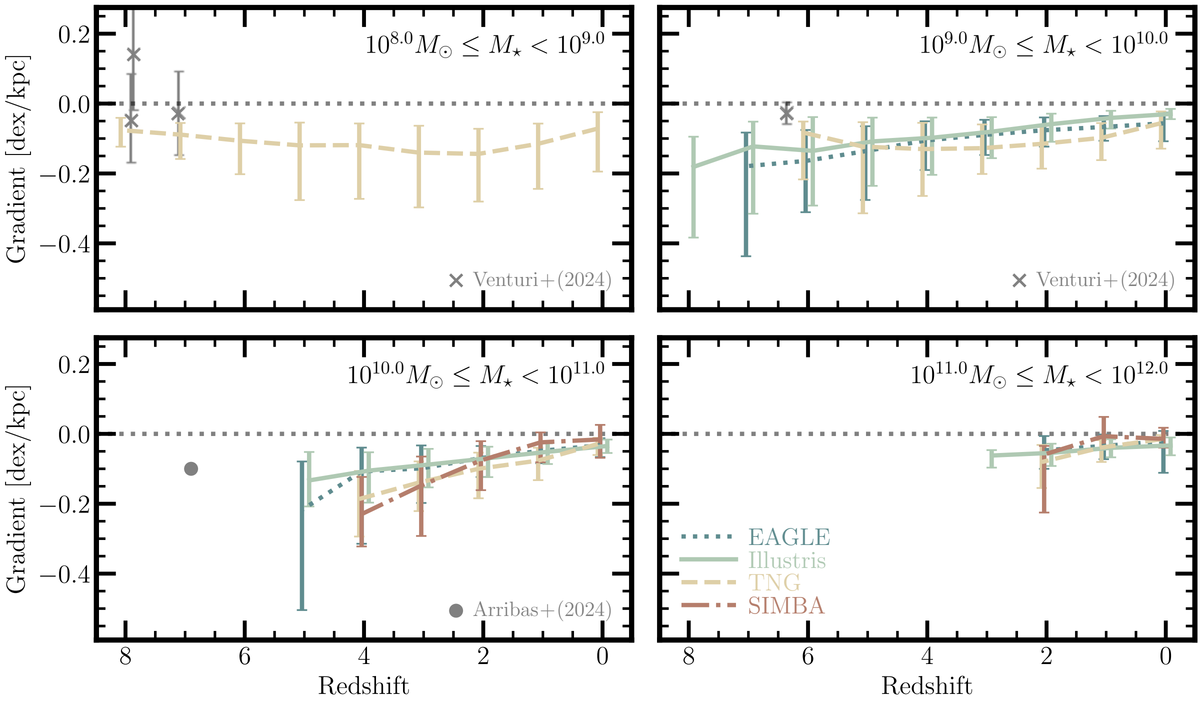

Figure 3. Metallicity Gradient Evolution by Stellar Mass. The gradient evolution of star forming galaxies in a combination of EAGLE (dotted), Illustris (solid), TNG (dashed), and SIMBA (dot dash) broken into four stellar mass bins of width 1.0 dex ranging from $10^{8.0}$ $M_\odot$ to $10^{12.0}$ $M_\odot$ as labeled in each panel. The errorbars represent the width of the distributions (16th and 84th percentiles). The gray data points are observed gradients at high-redshift ($z > 6$) from Arribas et al. (2024; circles) and Venturi et al. (2024; Xs). The observed data points fall in their respective mass bins.

		Stellar Mass Bin [log $M_\odot$]			
Simulation		8.0 – 9.0	9.0 – 10.0	10.0 – 11.0	11.0 – 12.0
EAGLE	[dex/kpc/Δz]	–	-0.011 ± 0.001	-0.019 ± 0.001	-0.013 ± 0.000
Illustris	[dex/kpc/Δz]	–	-0.016 ± 0.001	-0.017 ± 0.000	-0.009 ± 0.002
TNG	[dex/kpc/Δz]	-0.007 ± 0.005	-0.024 ± 0.004	-0.037 ± 0.004	-0.027 ± 0.005
SIMBA	[dex/kpc/Δz]	–	–	-0.030 ± 0.011	-0.022 ± 0.021

Table 2. Average Redshift Evolution of Metallicity Gradients By Mass. The best-fit linear regression parameters (weighted by number of galaxies) to the evolution of gradients for every simulation analyzed in this work. The quoted uncertainties are the square root of the variance of the slope, taken from the covariance matrix. We find that galaxies with stellar masses of $10^{10}M_\odot \leq M_\star < 10^{11}$ tend to have the strongest dex/kpc/Δz gradient evolution.

Förster Schreiber et al. 2018). In contrast to this, the simulated gradients ubiquitously become more and more negative from $z = 1 - 3$. This tension was highlighted in Hemler et al. (2021) between TNG and observations, but it appears to be a more general prediction of the smooth stellar feedback models.

Interestingly, all of the galaxies measured in the literature thus far at $z = 3 - 4$ show positive gradients across a wide range of stellar masses ($10^{8.6}M_\odot \lesssim M_\star \lesssim 10^{11.0}M_\odot$, Cresci et al. 2010; Troncoso et al. 2014; Wang et al. 2022). These positive gradients stand in stark contrast to the models in this work. We caution that we are not claiming there are *no* positive gradients in these simulations at these redshifts. In fact, at $z = 3$, 77 (2.41%) gradients are positive in EAGLE, 31 (0.56%) in Illustris, 18 (0.49%) in TNG, and 2 (1.60%) in SIMBA. At $z = 4$, 47 (3.01%) gradients are positive in EAGLE, 23 (0.88%) in Illustris, 10 (0.43%) in TNG, and 2 (3.64%) in SIMBA. There is, however, a clear qualitative difference in the sample of observed metallicity gradients at $z = 3 - 4$ from metallicity gradients of these simulations.

Although the tension at $1 < z < 4$ seems quite significant, it should be noted that there are some observational systematics to be taken into account in these gradients. Several studies have examined the impact of systematically degrading the quality of high spatial resolution, high signal-to-noise data (Yuan et al. 2013; Mast et al. 2014; Poetrodjojo et al. 2019; Acharyya et al. 2020; Grasha et al. 2022; Metha et al. 2024). Outside of lensed systems, there are typically only a few resolution elements per galaxy, even with the best telescopes. Generally, metallicity gradients are found to systematically flatten with lowered angular resolution and cannot be meaningfully constrained with sufficiently low signal-to-noise ratio. Lower spatial resolution observations can "smear" measurements of a single Hii region with either nearby Hii regions or diffuse ionized gas (DIG). Coarser spatial resolution that combines multiple Hii regions will over-weight the spectra towards the regions of stronger emission, causing an overestimation of the metallicity in the outskirts of galaxies (Yuan et al. 2013). DIG-dominated regions tend to have different physical conditions and emission line ratios than Hii regions (Blanc et al. 2009; Zhang et al. 2017). Any contamination of DIG in a spectrum systematically alters the derived metallicity (Poetrodjojo et al.

2019). Fortunately, ongoing and future studies in the new generation of optical and infrared telescopes (i.e., JWST and future ELTs) should help alleviate some of these issues with increased spatial resolution at these high redshifts. However, the gradients from Wang et al. (2022) and Ju et al. (2025) *do* come from JWST and are virtually all flat or positive. The extent to which the large population of gradients can be simply explained away by poor angular resolution is therefore unclear.

Beyond the impact of observational systematics, each of the simulations employs a subgrid equation of state for the dense, star forming ISM (see Section 2.1 and references therein for more details). An equation of state is not the unique method by which the dense ISM is modeled, however. Simulations such as FIRE (Hopkins et al. 2014, 2018, 2023) and SMUGGLE (Marinacci et al. 2019) are higher resolution and can directly model giant molecular cloud scales ($M_{\rm baryon} \lesssim 10^4 M_\odot$) of the ISM. These treatments, known collectively as "explicit" ISM models, have feedback regulated star formation, which can lead to episodic blow outs of gas in a short period of time (see, e.g., Muratov et al. 2015, 2017; Anglés-Alcázar et al. 2017; Pandya et al. 2021). These episodic bursts of feedback work to systematically flatten (or even temporarily invert) metallicity gradients by rapidly redistributing material from the inner regions to the outskirts (Ma et al. 2017; Muratov et al. 2017; Bellardini et al. 2021; Sun et al. 2024). It is therefore possible that the dearth of positive gradients in this redshift range in EAGLE, Illustris, TNG, and SIMBA is indicative that subgrid ISM models do not sufficiently model galaxy feedback at $z \sim 1-4$.

4.1.2 *Comparison with Observed Metallicity Gradients at $z > 6$*

Most of the gradients measured at $z > 6$ are at $z \sim 7$ with 7 galaxies to date (from Vallini et al. (2024): COS-2987 at $z = 6.8$, COS-3018 at $z = 6.8$, UVISTA-Z-001 at $z = 7.0$, UVISTA-Z-007 at $z = 6.7$, UVISTA-Z-019 at $z = 6.7$; from Venturi et al. (2024): BDF-3299 at $z = 7.1$; from Arribas et al. (2024): SPT0311-58 E at $z = 6.9$). Meanwhile, only one gradient has been measured at $z \sim 6$ (COSMOS24108 at $z = 6.3$ from Venturi et al. 2024) and two at $z \sim 8$ (A2744-YD4 at $z = 7.8$ and A2744-YD1 at $z = 7.8$ both from Venturi et al. 2024). All of the high redshift galaxies display relatively shallow negative gradients (with the lone exception of A2744-YD4 at $z = 7.8$ from Venturi et al. 2024, which has a strong positive gradient). Broadly speaking, this is consistent with simulations in the redshift range, although in detail they appear slightly flatter than simulations.

We will make a more direct comparison to observed gradients at $z \sim 7$ as that is the most populated redshift in the recent observations. The median observed gradient at $z = 7$ is approximately -0.04 dex/kpc. This is significantly flatter than the medians of -0.179 dex/kpc, -0.123 dex/kpc, and -0.091 dex/kpc in EAGLE, Illustris, and TNG at $z = 7$ (respectively). Naïvely taking the full distributions from the simulations, the observational median of ~ -0.03 dex/kpc is outside the 84th percentiles of all of the EAGLE, Illustris, and TNG distributions. The tension is not as strong as at $3 < z < 4$, but still persists. Recall, though, that the evolution of gradients is not constant with time (see Section 3.2). Lower mass galaxies tend to have flatter metallicity gradients than high mass galaxies at $z = 7$, in particular. Figure 3 also includes these $z > 6$ gradients in their respective mass bins. Broadly speaking, these galaxies span a wide range of masses: SPT0311-58 E (from Arribas et al. (2024) has a stellar mass of $\sim 10^{10.5} M_\odot$, BDF-3299 (from Venturi et al. 2024) has a stellar mass of $10^{8.21} M_\odot$, and galaxies from Vallini et al. (2024) have masses ranging from $10^{8.0} M_\odot < M_\star < 5 \times 10^{10} M_\odot$.[3] Considering the stellar masses of these systems, there is perhaps hints of a tension between the observed gradients and those of the simulations analyzed in this work, in particular with the Arribas et al. (2024) SPT0311-58 E galaxy. It should be noted that spatially resolved maps of galaxies at these high redshifts are highly expensive observationally. The earliest observed targets (such as the ones we compare to here) are likely the brightest objects. It is possible that they are atypical in some regard, perhaps, a recent merger or large gas accretion event (both of which could systematically flatten the gradient, see, e.g., Rupke et al. 2010b; Torrey et al. 2012; Ceverino et al. 2016). Larger samples of galaxies at these extreme redshifts are therefore critical for getting a statistical understanding of the extent to which the tension between observations and smooth feedback models is significant.

An additional consideration is the metallicity diagnostics at these high redshift. Vallini et al. (2024) use Atacama Large Millimeter/submillimeter Array (ALMA) observations using (rest) infrared lines ([OIII] 88μm and [CII] 158μm). This is in contrast to the observations from JWST that use (rest) optical lines: the Curti et al. (2017, 2020a) calibrations in the case of Arribas et al. (2024) and the Laseter et al. (2024) $\hat{R}$ in Venturi et al. (2024). There may therefore be some systematics not taken into account with the use of the rest infrared lines over the optical lines, as Vallini et al. (2024) point out. Moreover, it is yet to be seen as to whether the low-redshift calibrated optical line relations are valid at higher redshifts (though work has been done to calibrate these relations at high redshift, Garg et al. 2024; Curti et al. 2023; Hirschmann et al. 2023; Sanders et al. 2023; Trump et al. 2023; Übler et al. 2023; Chakraborty et al. 2024, Laseter et al. 2024, Backhaus et al. 2025).

Whereas producing the $3 < z < 4$ positive gradients from observations would require substantial changes to the galactic winds and ISM of the models, at $z > 6$ only a subtle change (if, indeed, one is needed) would be required to flatten simulated gradients to their observed counterparts. This correction could come in the form of turbulent metal diffusion between gas elements within the ISM. EAGLE and SIMBA – owing to their SPH and MFM implementations, respectively – do not have any metals advected from one element to the next. Metals in these models are locked in the gas particle in which they form. This has been shown to produce metal distributions that are more inhomogeneous than observed (see, e.g., Aguirre et al. 2005). For Illustris and TNG, the MVM implementation of AREPO naturally allows metals to advect as the cells deform and reshape. Beyond the cell deformation, however, there is no implementation for the transport of metals from small, unresolved turbulent eddies that drive diffusion in the ISM (as in, e.g., Smagorinsky 1963; Shen et al. 2010; Semenov et al. 2016; Su et al. 2017; Escala et al. 2018; Semenov 2024). The addition of these turbulent eddies in these models would have the effect of redistributing metals through the ISM, potentially flattening the metallicity gradient in the process. Bellardini et al. (2021) show this concretely using the FIRE model and a range of metal diffusion coefficients. Bellardini et al. (2021) find that the lack of a diffusion coefficient significantly steepens a gradient while higher coefficients allow for flattening. It is thus possible that the subtle differences between the metallicity gradients at these high redshifts indicate that these subgrid models require a metal diffusion model to redistribute the metals of the ISM.

In summary, it is as of yet uncertain whether the tension between

---

[3] The stellar masses of the galaxies in Vallini et al. (2024) are not reported exactly for each galaxy. We therefore do not report their gradients in any panel.

the simulations and observations at $z > 6$ is meaningful. Current and upcoming observational campaigns with JWST and ALMA should help clarify the picture by providing additional galaxies for more robust comparisons.

5 CONCLUSIONS

In this work, we analyze the gas-phase metallicity gradients of star forming galaxies across a wide mass ($10^{8.0}M_\odot < M_\star \lesssim 10^{12.0}M_\odot$) and redshift range ($0 \leq z \leq 8$) in EAGLE, Illustris, IllustrisTNG, and SIMBA. We construct face-on metallicity maps and reduce them into a metallicity radial profile that we fit with a single linear regression in the "star forming region" of the galaxies (see Figure 1).

Our conclusions are as follows:

- We find that the evolution of metallicity gradients in EAGLE, Illustris, TNG, and SIMBA are all very similar, with more negative gradients further back in time (Section 3.1 and Figure 2). We speculate that this is likely owing to the relatively smooth implementation of stellar feedback that arises naturally from the subgrid ISM prescription common amongst the models analyzed in this work.
- In more detail, we find that different stellar mass bins have different metallicity gradient evolution (Section 3.2 and Figure 3). We find that galaxy stellar masses of $10^{8.0}M_\odot - 10^{9.0}M_\odot$ have virtually no redshift evolution out to $z = 8$ (although gradients are slightly more negative around cosmic noon). On the other hand, galaxies with stellar masses of $10^{10.0}M_\odot - 10^{11.0}M_\odot$ have the strongest redshift evolution. Meanwhile, galaxies of lower intermediate ($10^{9.0}M_\odot - 10^{10.0}M_\odot$) and very high masses ($10^{11.0}M_\odot - 10^{12.0}M_\odot$) have a moderate amount of redshift evolution.
- Finally, we compare the simulation results to those of observations (Section 4.1). We find that our results are in contrast with observations at low-to-intermediate redshifts ($1 < z < 4$) which exhibit a larger fraction of positive and flat gradients (see Section 4.1.1). Comparing with higher redshift ($z > 6$) JWST and ALMA observations, however, the tension is more subtle: observed gradients at $z > 6$ are only slightly flatter than the simulations (see Section 4.1.2).. These comparisons suggest that metals may be under-mixed in many widely-used ISM subgrid models.

The spatial distribution of metals within the ISM of galaxies is critically sensitive to the underlying physics, particularly the feedback. The upcoming prospects for high-redshift observations of metallicity gradients with, e.g., JWST provide an exciting opportunity to understand the processes driving galactic evolution as well as constrain future simulation models of galaxies. A particular opportunity presents itself: no galaxy metallicity gradients have been observed in the redshift range of $4 < z < 6$ to date. This redshift range is where the smooth feedback simulated galaxies ubiquitously have strong negative metallicity gradients, regardless of stellar mass. Filling in this observational gap would therefore provide key insights into the level of metal mixing required in the ISM of simulated galaxies.

6 ACKNOWLEDGMENTS

AMG acknowledges support from a Virginia Space Grant Consortium Graduate STEM Research Fellowship.

REFERENCES

Acharyya A., Krumholz M. R., Federrath C., Kewley L. J., Goldbaum N. J., Sharp R., 2020, MNRAS, 495, 3819
Acharyya A., et al., 2025, ApJ, 979, 129
Aguirre A., Schaye J., Hernquist L., Kay S., Springel V., Theuns T., 2005, ApJ, 620, L13
Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698
Arribas S., et al., 2024, A&A, 688, A146
Backhaus B. E., et al., 2025, arXiv e-prints, p. arXiv:2502.03519
Belfiore F., et al., 2016, MNRAS, 461, 3111
Belfiore F., et al., 2017, MNRAS, 469, 151
Bellardini M. A., Wetzel A., Loebman S. R., Faucher-Giguère C.-A., Ma X., Feldmann R., 2021, MNRAS, 505, 4586
Berg D. A., Skillman E. D., Croxall K. V., Pogge R. W., Moustakas J., Johnson-Groh M., 2015, ApJ, 806, 16
Blanc G. A., Heiderman A., Gebhardt K., Evans II N. J., Adams J., 2009, ApJ, 704, 842
Bresolin F., Ryan-Weber E., Kennicutt R. C., Goddard Q., 2009, ApJ, 695, 580
Carton D., et al., 2018, MNRAS, 478, 4293
Ceverino D., Sánchez Almeida J., Muñoz Tuñón C., Dekel A., Elmegreen B. G., Elmegreen D. M., Primack J., 2016, MNRAS, 457, 2605
Chabrier G., 2003, PASP, 115, 763
Chakraborty P., et al., 2024, arXiv e-prints, p. arXiv:2412.15435
Chen Q.-H., Grasha K., Battisti A. J., Kewley L. J., Madore B. F., Seibert M., Rich J. A., Beaton R. L., 2023, MNRAS, 519, 4801
Crain R. A., et al., 2015, MNRAS, 450, 1937
Cresci G., Mannucci F., Maiolino R., Marconi A., Gnerucci A., Magrini L., 2010, Nature, 467, 811
Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Esposito S., 2017, MNRAS, 465, 1384
Curti M., Mannucci F., Cresci G., Maiolino R., 2020a, MNRAS, 491, 944
Curti M., et al., 2020b, MNRAS, 492, 821
Curti M., et al., 2023, MNRAS, 518, 425
Dalla Vecchia C., Schaye J., 2012, MNRAS, 426, 140
Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Doherty C. L., Gil-Pons P., Lau H. H. B., Lattanzio J. C., Siess L., 2014, MNRAS, 437, 195
Dutta R., et al., 2024, A&A, 691, A236
Dwek E., 1998, ApJ, 501, 643
Escala I., et al., 2018, MNRAS, 474, 2194
Fishlock C. K., Karakas A. I., Lugaro M., Yong D., 2014, ApJ, 797, 44
Förster Schreiber N. M., et al., 2018, ApJS, 238, 21
Garcia A. M., et al., 2023, MNRAS, 519, 4716
Garcia A. M., et al., 2024a, MNRAS, 529, 3342
Garcia A. M., et al., 2024b, MNRAS, 531, 1398
Garcia A. M., et al., 2025, MNRAS, 536, 119
Garg P., Narayanan D., Sanders R. L., Davé R., Popping G., Shapley A. E., Stark D. P., Trump J. R., 2024, ApJ, 972, 113
Genel S., et al., 2014, MNRAS, 445, 175
Gibson B. K., Pilkington K., Brook C. B., Stinson G. S., Bailin J., 2013, A&A, 554, A47
Gillman S., et al., 2022, MNRAS, 512, 3480
Grasha K., et al., 2022, ApJ, 929, 118
Hemler Z. S., et al., 2021, MNRAS, 506, 3024
Hirschmann M., Charlot S., Somerville R. S., 2023, MNRAS, 526, 3504
Ho I. T., Kudritzki R.-P., Kewley L. J., Zahid H. J., Dopita M. A., Bresolin F., Rupke D. S. N., 2015, MNRAS, 448, 2030
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581
Hopkins P. F., et al., 2018, MNRAS, 480, 800
Hopkins P. F., et al., 2023, MNRAS, 519, 3154
Ibrahim D., Kobayashi C., 2025, arXiv e-prints, p. arXiv:2501.11209

Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F.-K., 1999, ApJS, 125, 439

Jones T., Ellis R., Jullo E., Richard J., 2010, ApJ, 725, L176

Jones T., Ellis R. S., Richard J., Jullo E., 2013, ApJ, 765, 48

Jones T., et al., 2015, AJ, 149, 107

Ju M., et al., 2025, ApJ, 978, L39

Karakas A. I., 2010, MNRAS, 403, 1413

Kewley L. J., Rupke D., Zahid H. J., Geller M. J., Barton E. J., 2010, ApJ, 721, L48

Khoram A. H., Belfiore F., 2025, A&A, 693, A150

Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, ApJ, 653, 1145

Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36

Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216

Kubryk M., Prantzos N., Athanassoula E., 2015, A&A, 580, A127

Laseter I. H., et al., 2024, A&A, 681, A70

Leethochawalit N., Jones T. A., Ellis R. S., Stark D. P., Richard J., Zitrin A., Auger M., 2016, ApJ, 820, 84

Li Z., et al., 2022, ApJ, 929, L8

Ma X., Hopkins P. F., Feldmann R., Torrey P., Faucher-Giguère C.-A., Kereš D., 2017, MNRAS, 466, 4780

Marigo P., 2001, A&A, 370, 194

Marinacci F., et al., 2018, MNRAS, 480, 5113

Marinacci F., Sales L. V., Vogelsberger M., Torrey P., Springel V., 2019, MNRAS, 489, 4233

Mast D., et al., 2014, A&A, 561, A129

McAlpine S., et al., 2016, Astronomy and Computing, 15, 72

Metha B., Trenti M., Battisti A., Chu T., 2024, MNRAS, 529, 104

Mollá M., Díaz Á. I., Cavichia O., Gibson B. K., Maciel W. J., Costa R. D. D., Ascasibar Y., Few C. G., 2019, MNRAS, 482, 3071

Morishita T., et al., 2024, ApJ, 971, 43

Mott A., Spitoni E., Matteucci F., 2013, MNRAS, 435, 2918

Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, MNRAS, 454, 2691

Muratov A. L., et al., 2017, MNRAS, 468, 4170

Naiman J. P., et al., 2018, MNRAS, 477, 1206

Nelson D., et al., 2018, MNRAS, 475, 624

Nelson D., et al., 2019a, Computational Astrophysics and Cosmology, 6, 2

Nelson D., et al., 2019b, MNRAS, 490, 3234

Nomoto K., Iwamoto K., Nakasato N., Thielemann F. K., Brachwitz F., Tsujimoto T., Kubo Y., Kishimoto N., 1997, Nuclear Phys. A, 621, 467

Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, Nuclear Phys. A, 777, 424

Oppenheimer B. D., Davé R., 2006, MNRAS, 373, 1265

Ormerod K., et al., 2024, MNRAS, 527, 6110

Pandya V., et al., 2021, MNRAS, 508, 2979

Pilkington K., et al., 2012, MNRAS, 425, 969

Pillepich A., et al., 2018a, MNRAS, 473, 4077

Pillepich A., et al., 2018b, MNRAS, 475, 648

Pillepich A., et al., 2019, MNRAS, 490, 3196

Poetrodjojo H., D'Agostino J. J., Groves B., Kewley L., Ho I. T., Rich J., Madore B. F., Seibert M., 2019, MNRAS, 487, 79

Portinari L., Chiosi C., Bressan A., 1998, A&A, 334, 505

Prantzos N., Boissier S., 2000, MNRAS, 313, 338

Pérez E., et al., 2013, ApJL, 764, L1

Queyrel J., et al., 2012, A&A, 539, A93

Rodríguez Del Pino B., et al., 2024, A&A, 684, A187

Rupke D. S. N., Kewley L. J., Barnes J. E., 2010a, ApJ, 710, L156

Rupke D. S. N., Kewley L. J., Chien L. H., 2010b, ApJ, 723, 1255

Sánchez-Menguiano L., et al., 2016, A&A, 587, A70

Sánchez-Menguiano L., et al., 2018, A&A, 609, A119

Sánchez S. F., et al., 2014, A&A, 563, A49

Sanders R. L., Shapley A. E., Topping M. W., Reddy N. A., Brammer G. B., 2023, ApJ, 955, 54

Schaye J., 2004, ApJ, 609, 667

Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

Schaye J., et al., 2015, MNRAS, 446, 521

Searle L., 1971, ApJ, 168, 327

Semenov V. A., 2024, arXiv e-prints, p. arXiv:2410.23339

Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2016, ApJ, 826, 200

Sharda P., Krumholz M. R., Wisnioski E., Forbes J. C., Federrath C., Acharyya A., 2021a, MNRAS, 502, 5935

Sharda P., Krumholz M. R., Wisnioski E., Acharyya A., Federrath C., Forbes J. C., 2021b, MNRAS, 504, 53

Shen S., Wadsley J., Stinson G., 2010, MNRAS, 407, 1581

Simons R. C., et al., 2021, ApJ, 923, 203

Smagorinsky J., 1963, Monthly Weather Review, 91, 99

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., White M., Hernquist L., 2001, ApJ, 549, 681

Springel V., et al., 2018, MNRAS, 475, 676

Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, MNRAS, 471, 144

Sun X., et al., 2024, arXiv e-prints, p. arXiv:2409.09290

Swinbank A. M., Sobral D., Smail I., Geach J. E., Best P. N., McCarthy I. G., Crain R. A., Theuns T., 2012, MNRAS, 426, 935

Tapia B., Tissera P. B., Sillero E., Casanueva C., Pedrosa S., Bignone L., Dominguez Tenreiro R., Padilla N., 2022, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 63, 256

Taylor P., Kobayashi C., 2017, MNRAS, 471, 3856

Thielemann F. K., et al., 2003, Nuclear Phys. A, 718, 139

Tissera P. B., Pedrosa S. E., Sillero E., Vilchez J. M., 2016, MNRAS, 456, 2982

Tissera P. B., Rosas-Guevara Y., Bower R. G., Crain R. A., del P Lagos C., Schaller M., Schaye J., Theuns T., 2019, MNRAS, 482, 2208

Tissera P. B., Rosas-Guevara Y., Sillero E., Pedrosa S. E., Theuns T., Bignone L., 2022, MNRAS, 511, 1667

Torrey P., Cox T. J., Kewley L., Hernquist L., 2012, ApJ, 746, 108

Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, MNRAS, 438, 1985

Torrey P., et al., 2018, MNRAS, 477, L16

Troncoso P., et al., 2014, A&A, 563, A58

Trump J. R., et al., 2023, ApJ, 945, 35

Übler H., et al., 2023, A&A, 677, A145

Vallini L., et al., 2024, MNRAS, 527, 10

Venturi G., et al., 2024, A&A, 691, A19

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, MNRAS, 436, 3031

Vogelsberger M., et al., 2014a, MNRAS, 444, 1518

Vogelsberger M., et al., 2014b, Nature, 509, 177

Wang X., et al., 2017, ApJ, 837, 89

Wang X., et al., 2019, ApJ, 882, 94

Wang X., et al., 2022, ApJ, 938, L16

Weinberger R., et al., 2017, MNRAS, 465, 3291

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574

Wright R. J., Somerville R. S., Lagos C. d. P., Schaller M., Davé R., Anglés-Alcázar D., Genel S., 2024, MNRAS, 532, 3417

Wuyts E., et al., 2016, ApJ, 827, 74

Yuan T. T., Kewley L. J., Swinbank A. M., Richard J., Livermore R. C., 2011, ApJ, 732, L14

Yuan T. T., Kewley L. J., Rich J., 2013, ApJ, 767, 106

Zaritsky D., Kennicutt Jr. R. C., Huchra J. P., 1994, ApJ, 420, 87

Zhang K., et al., 2017, MNRAS, 466, 3217

This paper has been typeset from a TEX/LATEX file prepared by the author.