3D gas-phase elemental abundances across the formation histories of Milky Way-mass galaxies in the FIRE simulations: initial conditions for chemical tagging

Matthew A. Bellardini [1]*, Andrew Wetzel [1], Sarah R. Loebman,[1,2]†
Claude-André Faucher-Giguère [3], Xiangcheng Ma [4] and Robert Feldmann [5]

[1]*Department of Physics and Astronomy, University of California, Davis, One Shields Ave, Davis, CA 95616, USA*
[2]*Department of Physics, University of California, Merced, 5200 Lake Road, Merced, CA 95343, USA*
[3]*Department of Physics and Astronomy and CIERA, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[4]*Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, CA 94720, USA*
[5]*Institute for Computational Science, University of Zurich, Zurich CH-8057, Switzerland*

ABSTRACT

We use FIRE-2 simulations to examine 3D variations of gas-phase elemental abundances of [O/H], [Fe/H], and [N/H] in 11 MW and M31-mass galaxies across their formation histories at $z \leq 1.5$ ($t_{\mathrm{lookback}} \leq 9.4\,\mathrm{Gyr}$), motivated by characterizing the initial conditions of stars for chemical tagging. Gas within 1 kpc of the disc mid-plane is vertically homogeneous to $\lesssim 0.008$ dex at all $z \leq 1.5$. We find negative radial gradients (metallicity decreases with galactocentric radius) at all times, which steepen over time from $\approx -0.01\,\mathrm{dex\,kpc^{-1}}$ at $z = 1$ ($t_{\mathrm{lookback}} = 7.8\,\mathrm{Gyr}$) to $\approx -0.03\,\mathrm{dex\,kpc^{-1}}$ at $z = 0$, and which broadly agree with observations of the MW, M31, and nearby MW/M31-mass galaxies. Azimuthal variations at fixed radius are typically 0.14 dex at $z = 1$, reducing to 0.05 dex at $z = 0$. Thus, over time radial gradients become steeper while azimuthal variations become weaker (more homogeneous). As a result, azimuthal variations were larger than radial variations at $z \gtrsim 0.8$ ($t_{\mathrm{lookback}} \gtrsim 6.9\,\mathrm{Gyr}$). Furthermore, elemental abundances are measurably homogeneous (to $\lesssim 0.05$ dex) across a radial range of $\Delta R \approx 3.5$ kpc at $z \gtrsim 1$ and $\Delta R \approx 1.7$ kpc at $z = 0$. We also measure full distributions of elemental abundances, finding typically negatively skewed normal distributions at $z \gtrsim 1$ that evolve to typically Gaussian distributions by $z = 0$. Our results on gas abundances inform the initial conditions for stars, including the spatial and temporal scales for applying chemical tagging to understand stellar birth in the MW.

Key words: methods: numerical – stars: abundances – ISM: abundances – galaxies: abundances – galaxies: formation – galaxies: ISM.

1 INTRODUCTION

Many current and future observational surveys of stars across the Milky Way (MW) seek to unveil the MW's formation history in exquisite detail. Current surveys, such as the RAdial Velocity Experiment (RAVE; Steinmetz et al. 2006), the Gaia–ESO survey (Gilmore et al. 2012), the Large Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012), GALactic Archaeology with Hermes (GALAH; De Silva et al. 2015), and the Apache Point Galactic Evolution Experiment (APOGEE; Majewski et al. 2017) have measured elemental abundances of hundreds of thousands of stars. Future surveys, such as the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012), the Subaru Prime Focus Spectrograph (PFS; Takada et al. 2014), the Sloan Digital Sky Survey V (SDSS-V; Kollmeier et al. 2017), the 4-m Multi-

Object Spectrograph Telescope (4MOST; de Jong et al. 2019), and the MaunaKea Spectroscopic Explorer (MSE; The MSE Science Team et al. 2019) will increase the samples of spectroscopically observed stars into the millions. A key science driver for these surveys is 'galactic archaeology': to infer the history of the MW using observations of the dynamics and elemental abundances of stars today.

Measurements of stellar dynamics can provide detailed information on the MW's properties and formation history, but the fundamental limitation is that stellar orbits can change over time via mergers, accretion, scattering, and other dynamical perturbations (e.g. Abadi et al. 2003; Brook et al. 2004; Schönrich & Binney 2009; Loebman et al. 2011). However, a star's atmospheric elemental abundances will not change in response to these dynamical processes, providing a key orbital-invariant 'tag'. 'Chemical tagging', introduced in Freeman & Bland-Hawthorn (2002), thus, provides tremendous potential to infer the formation conditions of a star of arbitrary age.

'Strong' chemical tagging represents the most fine-grained scenario, to identify stars born in the same star cluster (e.g. Price-Jones

* E-mail: mbellardini@ucdavis.edu
† Hubble Fellow.

et al. 2020). By contrast, 'weak' chemical tagging seeks to infer the general location and time where/when a stellar population formed, for example, to associate populations of stars to certain birth regions of the galaxy (e.g. Wojno et al. 2016; Anders et al. 2017) or that accreted into the MW from galaxy mergers (e.g. Ostdiek et al. 2020).

Both regimes of chemical tagging rely on sufficiently precise measurements of stellar abundances and on assumptions about the elemental homogeneity (to the measured precision) of the gas from which the stars formed. For example, strong chemical tagging of individual star clusters relies on both the internal homogeneity of the gas cloud from which the stars formed, and on how unique the abundance patterns were in that cloud across space and time. Observational evidence of open star clusters suggests the first criterion is met (De Silva et al. 2007; Ting et al. 2012; Bovy 2016) to measurable precision. Regarding the latter criterion, observations of the MW and external galaxies show radial and azimuthal variations in abundances across the disc (e.g. Sánchez-Menguiano et al. 2016; Mollá et al. 2019b; Wenger et al. 2019; Kreckel et al. 2020), although more work is needed to understand these spatial variations in the context of chemical tagging. Weak chemical tagging is subject to the same assumptions but applied to larger regions of gas across the disc (or in accreting galaxies). For example, if all gas in the disc was measurably homogeneous in all abundances at a given time, chemical tagging would offer no spatially discriminating power. Conversely, the limit of extreme clumpiness, in which each star cluster formed with a measurably unique abundance pattern, would in principle enable detailed chemical tagging, but it significantly would complicate the modelling.

Thus, a key question for chemical tagging is: what are the relevant spatial scales of measurable homogeneity of stars forming at a given time, and how does this evolve across cosmic time? Bland-Hawthorn, Krumholz & Freeman (2010) previously explored this via a toy model, where they show all star clusters $\lesssim 10^4 \, \mathrm{M_\odot}$ and a large fraction of clusters with mass below $\sim 10^5 \, \mathrm{M_\odot}$ are expected to be internally homogeneous. Progress in chemical tagging requires addressing these questions regarding stellar birth before examining the subsequent dynamical evolution of stars after they form.

Many works have examined abundance variations of stars across the MW, generally finding a negative radial gradient in abundances for stars near the plane of the disc, which flattens or turns positive at larger heights (e.g. Cheng et al. 2012; Boeche et al. 2013, 2014; Anders et al. 2014; Hayden et al. 2014; Mikolaitis et al. 2014; Anders et al. 2017). Furthermore, the MW has an observed negative vertical gradient in stars (Carrell, Chen & Zhao 2012; Cheng et al. 2012; Boeche et al. 2014; Hayden et al. 2014), although this slope varies significantly between observations and depends on radius (Hayden et al. 2014). Both the radial and vertical gradients vary with stellar age (Wang et al. 2019b). Additionally, Luck, Kovtyukh & Andrievsky (2006), Lemasle et al. (2008), and Pedicelli et al. (2009) found evidence for azimuthal variations in the abundances of young stars, which may result from patchy star formation (Davies et al. 2009; Luck & Lambert 2011; Genovali et al. 2014). Nieva & Przybilla (2012) also explored the homogeneity of B-type stars in the solar neighbourhood, within 500 pc, of the sun and found scatter on the order of 0.05 dex for [O/H] which they state is comparable to gas-phase abundance scatter out to 1.5 kpc of the sun.

In addition to stellar abundances, many works have characterized trends of gas-phase abundances in the MW. Observations show that the MW has a negative radial gradient in gas-phase abundances, with a slope that varies across the elements (Arellano-Córdova et al. 2020)

and across studies (e.g. Mollá et al. 2019a and references therein). Furthermore, evidence persists for azimuthal variations in this radial gradient, based on H II regions (Balser et al. 2011, 2015; Wenger et al. 2019).

Beyond the MW, observations of nearby MW-mass galaxies also show negative radial gradients in gas-phase abundances (e.g. Pilyugin, Grebel & Kniazev 2014; Sánchez-Menguiano et al. 2016; Belfiore et al. 2017; Poetrodjojo et al. 2018). Furthermore, some observations show azimuthal variations (Sánchez et al. 2015; Vogt et al. 2017; Ho et al. 2017, 2018; Kreckel et al. 2019, 2020; Sánchez-Menguiano et al. 2020), while others show no azimuthal variations within measurement uncertainty ($\lesssim 0.05$ dex) (e.g. Cedrés & Cepa 2002; Zinchenko et al. 2016).

Understanding how these variations change across cosmic time is imperative for chemical-tagging models. Currently no consensus exists, amongst observations (e.g. Curti et al. 2020 and references therein) on the redshift evolution of radial elemental abundance gradients, in part because of angular resolution limitations (Yuan, Kewley & Rich 2013), which some works have addressed via adaptive optics and gravitational lensing (Jones et al. 2010, 2013; Swinbank et al. 2012). Furthermore, different works use different calibrators to measure abundances, which often disagree (Hemler et al. 2020), further complicating our understanding of spatial variations.

Many theoretical works have used simulations to predict the spatial distribution of gas-phase abundances and their evolution. As with observational efforts, there is no consensus for the redshift evolution of abundance gradients in theory (e.g. Mollá et al. 2019a and references therein). Gibson et al. (2013) compared cosmological simulations with MUGS ('conservative' feedback) and MaGICC ('enhanced' feedback) run with the GASOLINE code and found that the strength of feedback in simulations is critical for the evolution of radial gradients of abundances, such that stronger feedback leads to flatter gradients at all times, while galaxies with weaker feedback have gradients that are steep at high redshift and flatten with time. Ma et al. (2017), studying the FIRE-1 suite of cosmological simulations, found that galaxies exhibit a diverse range of radial gradients in abundances, and that these gradients can fluctuate rapidly from steep to shallow (in $\sim 100 \, \mathrm{Myr}$) at high redshift, so measurements of high-redshift gradients may not be indicative of long-term trends. They found that galaxies tend to quickly build up a negative gradient once stellar feedback is no longer sufficient to drive strong outflows of gas. By contrast, analysing star-forming galaxies in the TNG-50 cosmological simulation, Hemler et al. (2020) found that radial gradients in galaxies are steep at high redshift and flatten with time. Several theoretical works also have examined azimuthal variations. Spitoni et al. (2019) developed a 2D model for abundance evolution that follows radial and azimuthal density variations in a MW-like disc and found that azimuthal residuals are strongest at early times and at large radii. Using their S2A model, Spitoni et al. (2019) found azimuthal residuals in [O/H] of ≈ 0.1 dex at $R = 8 \, \mathrm{kpc}$ at $t_{\mathrm{lookback}} = 11 \, \mathrm{Gyr}$ which evolve to ≈ 0.05 dex at present day. Mollá et al. (2019b) explored azimuthal variations in a MW-like disc for five models of 2D abundance evolution and found [O/H] variations that are typically small ($0.05 - 0.1$ dex) and dilute quickly with time. Solar, Tissera & Hernandez-Jimenez (2020) used young star particles as tracers of star-forming gas in the EAGLE cosmological simulation and found an average azimuthal abundance dispersion of ≈ 0.12 dex at $z = 0$ in galaxies with $M_{\mathrm{star}} = 10^9 - 10^{10.8} \, \mathrm{M_\odot}$.

Observations of azimuthal variations of abundances in gas in nearby MW-mass galaxies find scatter that is comparable to observa-

tional measurement uncertainty (~ 0.05 dex) (Zinchenko et al. 2016; Kreckel et al. 2019). This implies that gas in galaxies is well mixed azimuthally at $z = 0$. Consequently, works modelling the abundance evolution of galaxies, which inform chemical tagging, generally assume that gas is well mixed azimuthally in the disc at all times (e.g. Minchev et al. 2018; Mollá et al. 2019a; Frankel et al. 2020), such that the key spatial variation is radial. While galaxies exhibit radial gradients across a range of redshifts (Queyrel et al. 2012; Stott et al. 2014; Wuyts et al. 2016; Carton et al. 2018; Patrício et al. 2019; Curti et al. 2020), radial variations may not always dominate. At early times, in particular, azimuthal variations may be more important. Some abundance-evolution models have begun to explore both azimuthal and radial variations across time (e.g. Acharova et al. 2013; Mollá et al. 2019b; Spitoni et al. 2019). Kawata et al. (2014) and Grand, Kawata & Cropper (2015), using N-body simulations of MW-mass galaxies, and Baba et al. (2016), using baryonic simulations run with the SPH code ASURA-2, found that gas exhibits streaming motion along spiral arms, which could contribute to 2D abundance variations in gas.

More detailed 2D abundance-evolution models, which account for density variations within the disc from spiral arms and bars, result in azimuthal variations in gas-phase abundances. Mollá et al. (2019b) found that arm / inter-arm abundance variations quickly dilute through interactions with spiral structure. Spitoni et al. (2019) also found that azimuthal variations dilute with time, but they found that the strength of azimuthal variations at $z = 0$ approximately agree with the observational results of Kreckel et al. (2019). The simulation analysis of Grand et al. (2015) and Baba et al. (2016) showed that, just like stars (Lynden-Bell & Kalnajs 1972), gas experiences radial migration as a result of spiral structure. Grand et al. (2015) found that this systematic streaming along spiral arms leads to metal-rich gas in the inner galaxy moving to larger radii and metal-poor gas in the outer galaxy moving to the inner galaxy, leading to non-homogeneous abundances at a given radius. However, they found that gas elements quickly exchange abundances after migrating, leading to small azimuthal dispersions in abundance. Sánchez-Menguiano et al. (2020) found that azimuthal variations in abundances are stronger in galaxies with stronger bars and grand-design spirals, which supports non-axisymmetric structure driving azimuthal inhomogeneities.

In this paper, we use FIRE-2 cosmological simulations of MW/M31-mass galaxies to explore the cosmic evolution of 3D abundance patterns of gas, as a first step towards understanding the spatial and temporal scales of applying chemical tagging in a cosmological context. Our analysis of gas represents our first step, to characterize the initial conditions for star-forming regions. In future work, we will examine the resultant trends in stars and their dynamical evolution across time. Here, we seek to quantify the 3D spatial scales over which elemental abundances of gas (and thus the formation of stars) are measurably homogeneous. In Section 2, we describe the simulations used for this analysis. In Section 3, we first explore the radial gradients and compare them against observations of the MW, M31, and nearby MW/M31-mass galaxies. Next, we examine the cosmic evolution of radial, vertical, and azimuthal variations in gas-phase abundances, in particular, to understand which dimension dominates the spatial variations at a given time. We also examine implications of gas (in)homogeneity on current and upcoming observations of the MW. Finally, we examine full distributions of elemental abundances. In Section 4, we summarize the main results of the paper and provide a discussion of their implications.

2 METHODS

2.1 FIRE-2 simulations

We use a suite of MW/M31-mass cosmological zoom-in simulations from the Feedback In Realistic Environments (FIRE) project[1] (Hopkins et al. 2018). We ran these simulations using the FIRE-2 numerical implementations of fluid dynamics, star formation, and stellar feedback. These simulations use the Lagrangian Meshless Finite Mass (MFM) hydrodynamics method in GIZMO (Hopkins 2015). The FIRE-2 model incorporates physically motivated metallicity-dependent radiative heating and cooling processes for gas such as free–free, photoionization and recombination, Compton, photo-electric and dust collisional, cosmic ray, molecular, metal-line, and fine structure processes, accounting for 11 elements (H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe) across a temperature range of $10 - 10^{10}$ K. The simulations also include a spatially uniform, redshift-dependent UV background from Faucher-Giguere et al. (2009). In calculating metallicities throughout this paper, we scale elemental abundances to the solar values in Asplund et al. (2009).

Star particles form out of gas that is self-gravitating, Jeans-unstable, cold ($T < 10^4$ K), dense ($n > 1000 \, cm^{-3}$), and molecular (following Krumholz & Gnedin 2011). Each star particle inherits the mass and elemental abundances of its progenitor gas and represents a single stellar population, assuming a Kroupa (2001) stellar initial mass function. FIRE-2 evolves star particles along standard stellar population models from e.g. STARBURST99 v7.0 (Leitherer et al. 1999), including time-resolved stellar feedback from core-collapse and Ia supernovae, continuous mass-loss, radiation pressure, photoionization, and photo-electric heating. FIRE-2 uses rates of core-collapse and Ia supernovae from STARBURST99 (Leitherer et al. 1999) and Mannucci, Della Valle & Panagia (2006), respectively. The nucleosynthetic yields follow Nomoto et al. (2006) for core-collapse and Iwamoto et al. (1999) for Ia supernovae. Stellar wind yields, sourced primarily from O, B, and AGB stars, are from the combination of models from van den Hoek & Groenewegen (1997), Marigo (2001), Izzard et al. (2004), synthesized in Wiersma et al. (2009).

Critical for this work, these FIRE-2 simulations also explicitly model the sub-grid diffusion/mixing of elemental abundances in gas that occurs via unresolved turbulent eddies (Su et al. 2017; Escala et al. 2018; Hopkins et al. 2018). In effect, this smooths abundance variations between gas elements, assuming that sub-grid mixing is dominated by the largest unresolved eddies. Escala et al. (2018) showed that incorporating this sub-grid model is a necessity to match observed distributions of stellar metallicities. We explore the robustness of our results to variations in the strength of the mixing/diffusion coefficient in Appendix C.

All simulations assume flat ΛCDM cosmologies with parameters broadly consistent with the Planck Collaboration VI (2020): $h = 0.68-0.71$, $\Omega_\Lambda = 0.69-0.734$, $\Omega_m = 0.266-0.31$, $\Omega_b = 0.0455-0.048$, $\sigma_8 = 0.801-0.82$ and $n_s = 0.961-0.97$. For each simulation, we generated cosmological zoom-in initial conditions embedded within cosmological boxes of length $70.4 - 172$ Mpc at $z \approx 99$ using the code MUSIC (Hahn & Abel 2011). We saved 600 snapshots from $z = 99$ to 0, with typical time spacing $\lesssim 25$ Myr.

We examine 11 MW/M31-mass galaxies from two suites of simulations. We select only galaxies with a stellar mass within a factor of ≈ 2 of the MW, $\approx 5 \times 10^{10} \, M_\odot$ (Bland-Hawthorn & Gerhard 2016). Five of our galaxies are from the *Latte* suite of

---

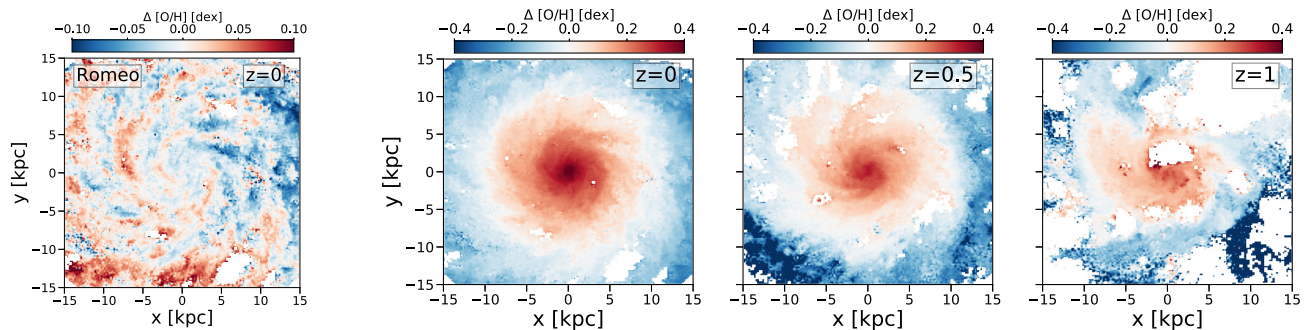[1]FIRE project website: http://fire.northwestern.edu

Figure 1. Face-on image of all gas within ± 1 kpc of the galactic mid-plane of Romeo, one of FIRE-2 simulations that we analyse. We colour-code gas by Δ[O/H], its deviation from the azimuthally averaged [O/H]. Δ[Fe/H] (not shown) looks nearly identical to Δ[O/H], to within $\lesssim 0.02$ dex at all times. The left-hand panel emphasizes azimuthal variations by showing the deviation from the azimuthally averaged [O/H] *at each radius*, that is, subtracting off the radial gradient, thus highlighting the enhanced enrichment correlated with spiral arms. The three right-hand panels show the deviation from the mean [O/H] of all gas at $R \leq 15$ kpc at each redshift. White regions have highly diffuse gas in which we do not report a measured abundance. The radial gradient in [O/H] dominates over azimuthal variations at late times, but at early times the azimuthal variations are the most significant.

isolated individual MW/M31-mass haloes, introduced in Wetzel et al. (2016). (We exclude m12w, because it has an unusually compact gas disc at $z = 0$, with $R_{90}^{gas} = 7.4$ kpc). *Latte* galaxies have halo masses $M_{200m} = 1 - 2 \times 10^{12}$ M$_\odot$, for which M_{200m} refers to the total mass within the radius containing 200 times the mean matter density of the Universe. These simulations have dark matter (DM) particle masses of 3.5×10^4 M$_\odot$ and initial baryonic particle masses of 7070 M$_\odot$ (though because of stellar mass-loss, star particles at $z = 0$ typically have masses of ~ 5000 M$_\odot$). Star and DM particles have fixed gravitational force softening lengths of 4 and 40 pc (Plummer equivalent), comoving at $z > 9$ and physical thereafter. Gas elements have adaptive gas smoothing and gravitational force softening lengths that reach a minimum of 1 pc. We also include six galaxies from the 'ELVIS on FIRE' suite (Garrison-Kimmel et al. 2014, 2019). These simulate LG-like MW+M31 pairs. ELVIS hosts have halo masses $M_{200m} = 1 - 3 \times 10^{12}$ M$_\odot$, with ≈ 2 times better mass resolution than the *Latte* suite.

In general, we find few systematic differences in any of our results for the isolated galaxies versus the LG-like pairs, the only notable difference being the relative strength of azimuthal scatter to radial gradient strength at large radius and high z, so we combine these suites in all of our results.

3 RESULTS

Fig. 1 shows face-on images of the gas disc of one of our simulations, Romeo, at several redshifts. We colour-code gas by its variation in [O/H], to visualize key trends that we explore in this work. We do not show results for [Fe/H], because they are qualitatively consistent with [O/H]. The left-hand panel shows the deviation of the local [O/H] from the mean value *at each radius* for radial bins of width 200 pc at $z = 0$, that is, we subtract off the overall radial gradient. This highlights the variations along the azimuthal direction at each radius, showing enhancement in [O/H] along spiral structure (Orr et al., in preparation will present a detailed analysis of metallicity enhancements along spiral arms).

The three right-hand panels show the difference between the local [O/H] and the mean [O/H] across all gas at $R \leq 15$ kpc and within ± 1 kpc of the galactic mid-plane at each redshift. This highlights the importance of both radial and azimuthal abundance variations in gas. At late times, the gas disc shows a clear negative radial gradient that is much stronger than the azimuthal variations. However, at earlier

times, the gas disc is azimuthally more asymmetric, including cavities from local star-forming and feedback regions. A radial gradient is less pronounced. As we will show, at $z \gtrsim 0.8$ ($t_{lookback} \gtrsim 6.9$ Gyr), the azimuthal variations in abundance at a given radius are typically larger than the radial change across the disc.

Because the absolute normalization of any elemental abundance in our simulations is uncertain, given uncertainties in underlying nucleosynthetic rate and yield models, throughout this paper, we focus on *relative* abundance variations, including spatial variations, evolution, and the shapes of abundance distributions rather than absolute normalizations of abundances.

As Fig. 2 shows, gas [O/H] in our simulations is super-solar at all radii (out to 15 kpc). Over comparable radial ranges, our hosts have an average gas-phase [O/H] ≈ 0.56 dex larger than what is observed in the MW (Fernández-Martín et al. 2017; Wenger et al. 2019; Arellano-Córdova et al. 2020) and M31 (Z12; S12). Most likely the primary cause of this high normalization is our modelling of core-collapse supernovae in FIRE-2. We assume all core-collapse supernovae to have identical (IMF-averaged) yields, but in reality, different mass progenitors produce different yields, as Muley et al. (2020) explored in the context of FIRE-2.

Furthermore, in FIRE-2, we assume nucleosynthetic yields for core-collapse supernovae from Nomoto et al. (2006), but more recent compilations (Nomoto, Kobayashi & Tominaga 2013; Pignatari et al. 2016; Sukhbold et al. 2016; Limongi & Chieffi 2018) suggest ~ 2 times lower O yields (IMF-averaged), as we will show in detail with our next-generation FIRE model (Hopkins, in preparation). In that work, we also will show that the overall stellar-wind mass-loss rates are likely $\sim 2 \times$ smaller than in FIRE-2, which further contributes to lower [O/H]. Finally, as we will explore in Gandhi et al. (in preparation), our assumed supernovae Ia rates in FIRE-2 (Mannucci et al. 2006) are ~ 2 times lower than more recent Ia rate constraints (e.g. Maoz & Graur 2017). Thus, these updates will account for up to $\sim 0.3 - 0.4$ dex lower normalization in our predicted [O/H], and up to $0.4 - 0.5$ dex lower normalization in our [O/Fe].

3.1 Radial profiles at $z = 0$

First, we examine the radial profiles of [O/H], [Fe/H], and [O/Fe] in gas for all 11 galaxies at $z = 0$. We time-average each galaxy's profile across ~ 50 Myr by stacking three snapshots to reduce short-time fluctuations. We present all results in physical radii; in Appendix A,
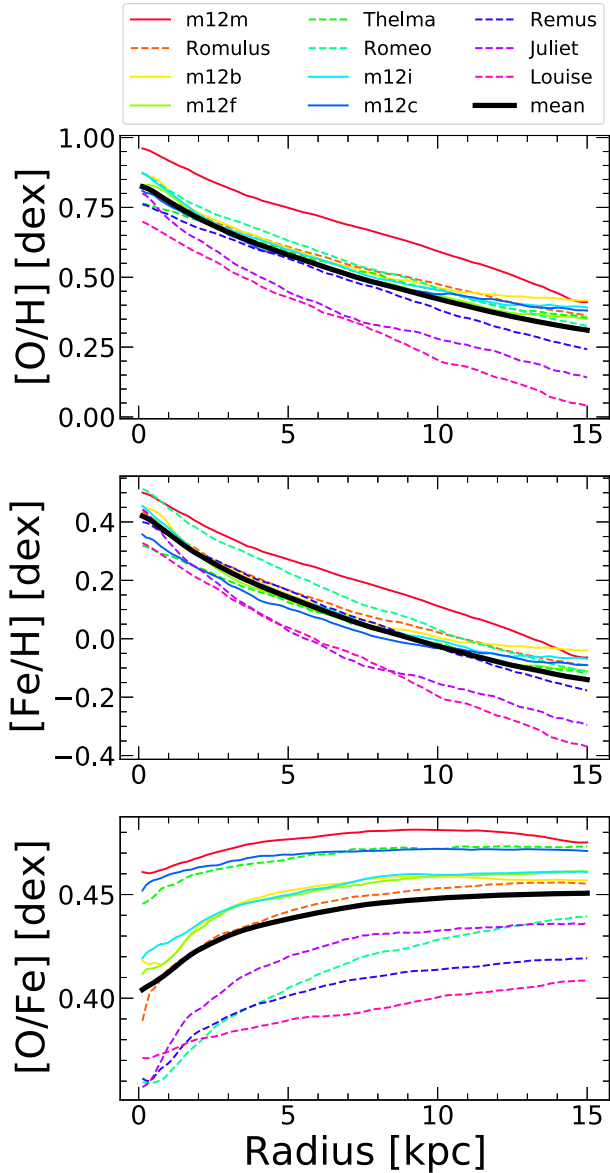
Figure 2. Radial profiles of gas-phase elemental abundances within the discs of our 11 galaxies at $z = 0$, listed by decreasing stellar mass. Each simulation includes all gas, averaging across three snapshots (≈ 50 Myr) within a disc height of ± 1 kpc. The black line shows the mean across all hosts. The normalization of the radial abundances scales with the stellar mass of the galaxy (see Table: 1). For [O/H] and [Fe/H], the profiles exhibit negative radial gradients, with a mean change across $0 - 15$ kpc of ~ 0.51 dex for [O/H] and ~ 0.56 dex for [Fe/H]. The [O/Fe] profile is approximately flat at radii $\gtrsim 4$ kpc, indicating enrichment dominated by core-collapse supernovae, but it exhibits a small positive gradient in the inner galaxies, from the increasing importance of Fe from Ia supernovae.

we examine these trends scaling to various galactic scale radii, finding that the host-to-host scatter in our suite is minimized when examining gradients in physical units. These profiles contain all gas within a vertical height $Z \pm 1$ kpc from the disc and we use radial bins of width 0.25 kpc. We calculate the mass-weighted mean of the gas-phase abundance in each bin. We show profiles out to $R = 15$ kpc; our gas discs generally extend beyond this radius, but because our primary motivation is chemical tagging, we examine only regions with significant star formation (see Table 1).

Fig. 2 (top two panels) shows that [O/H] and [Fe/H] decrease monotonically with radius. The mean gradient is ≈ -0.03 dex kpc^{-1} for both [O/H] and [Fe/H], and the mean change in abundance from $0 - 15$ kpc is ≈ 0.51 dex for [O/H] and ≈ 0.56 dex for [Fe/H]. These negative gradients in gas reflect the decreasing ratio of stars (the sources of enrichment) to gas towards the outer disc, and show that these gas discs are not radially well mixed at $z = 0$.

Across our 11 galaxies, the host-to-host standard deviation is ≈ 0.09 dex for [O/H] and ≈ 0.07 dex for [Fe/H]. The legend of Fig. 2 lists the host galaxies in decreasing order of stellar mass, highlighting that the abundance at a given radius correlates strongly with the galaxy's mass. Table 1 shows that stellar mass drops by a factor of ~ 4 from m12m to Louise; given the slope of the gas-phase mass–metallicity relation from Ma et al. (2016), ≈ 0.4 dex, the scatter in [O/H] normalization for our mass range should be ≈ 0.24 dex, almost exactly that in Fig. 2. In other words, the scatter across our suite primarily reflects the mass–metallicity relation (see Lequeux et al. 1979; Tremonti et al. 2004; Mannucci et al. 2010; Andrews & Martini 2013).

In Fig. 2, dashed lines show the LG-like hosts, and while they show typically lower abundance at a fixed radius than the isolated hosts, this is because they have somewhat lower stellar mass on average. We find no systematic differences between LG-like and isolated hosts beyond this, despite the fact that the LG-like hosts form their in-situ stars systematically earlier than the isolated hosts (Santistevan et al. 2020). Thus, this difference in formation history does not imprint itself on gas-phase abundances at $z \leq 1.5$. As a result, we will combine these samples in all subsequent analyses.

Fig. 2 (bottom panel) shows profiles for [O/Fe], which are nearly flat at all radii. The mean change in [O/Fe] from $0 - 15$ kpc is ≈ -0.046 dex. [O/Fe] shows the strongest (positive) gradient in the inner ≈ 4 kpc, highlighting the increasing importance of enrichment from (more delayed) Ia supernovae towards the Galactic Center, which underwent the longest period of enrichment. However, the outer disc, beyond ≈ 4 kpc, reflects relatively similar enrichment from core-collapse and Ia supernovae at each radius. We find a host-to-host standard deviation of ≈ 0.027 dex for [O/Fe]. We measure this for [Mg/Fe] (not shown here) as another tracer of core-collapse versus Ia supernovae enrichment. The mean change in [Mg/Fe] from $0 - 15$ kpc is $\approx -.111$ dex and the host-to-host standard deviation is ≈ 0.02 dex. These differences are likely attributable to our stellar wind model having a metallicity dependent yield for O and not for Mg. This leads to more O production at small radii, where the metallicity is higher, thus a flatter profile.

Fig. 3 shows the radial gradients of [O/H] in our simulations at $z = 0$ and includes observations of the MW, M31, and nearby MW/M31-mass galaxies. We fit the gradients in our simulations using a least-squares fit of the [O/H] abundance across $4 - 12$ kpc. As Fig. 2 shows, including the inner region of our discs, where the bulge dominates ($R \lesssim 4$ kpc), gives a profile not well approximated by a single linear fit (the bulge is steeper), so we exclude it in fitting this profile, to measure the 'disc' component. The range $4 - 12$ kpc covers the inner and outer disc and generally exhibits a single power-law profile. The solid blue line shows the median (-0.028 dex kpc^{-1}) across our 11 galaxies, while the shaded regions show the 68th percentile and the full distribution. The latter ranges from -0.042 to -0.024 dex kpc^{-1}. [Fe/H] gradients show similar results, with the full distribution spanning -0.044 to -0.024 dex kpc^{-1}.

Fig. 3 shows [O/H] gradients observed in nearby MW/M31-mass galaxies as box-and-whisker plots, with the box showing the 68th percentile and the whiskers showing the full observed range. We apply a cut on the stellar masses of these observed samples to

Table 1. Properties of the stellar and gas discs of our simulated MW/M31-mass galaxies $z = 0$. The first column lists the name of the galaxy: 'm12' indicates isolated galaxies from the *Latte* suite, while the other galaxies are Local Group (LG) analogues from the ELVIS on FIRE suite. R_{25} and R_{50} is the radius where the cumulative mass of the disc is 25% and 50%, respectively, within a height ± 3 kpc out of the mid-plane, relative to the total stellar/gas mass of the disc within 20 kpc. We fit R_{90} and Z_{90} as the radius and height, where the cumulative mass of the stellar/gas disc are 90% of the total mass of stars/gas within a sphere of 20 kpc. M_{90} is the total stellar/gas mass contained within both R_{90} and Z_{90}. The gas fraction, f_{90}^{gas}, is the ratio of gas mass to total baryonic mass within R_{90}^{star} and Z_{90}^{star}.

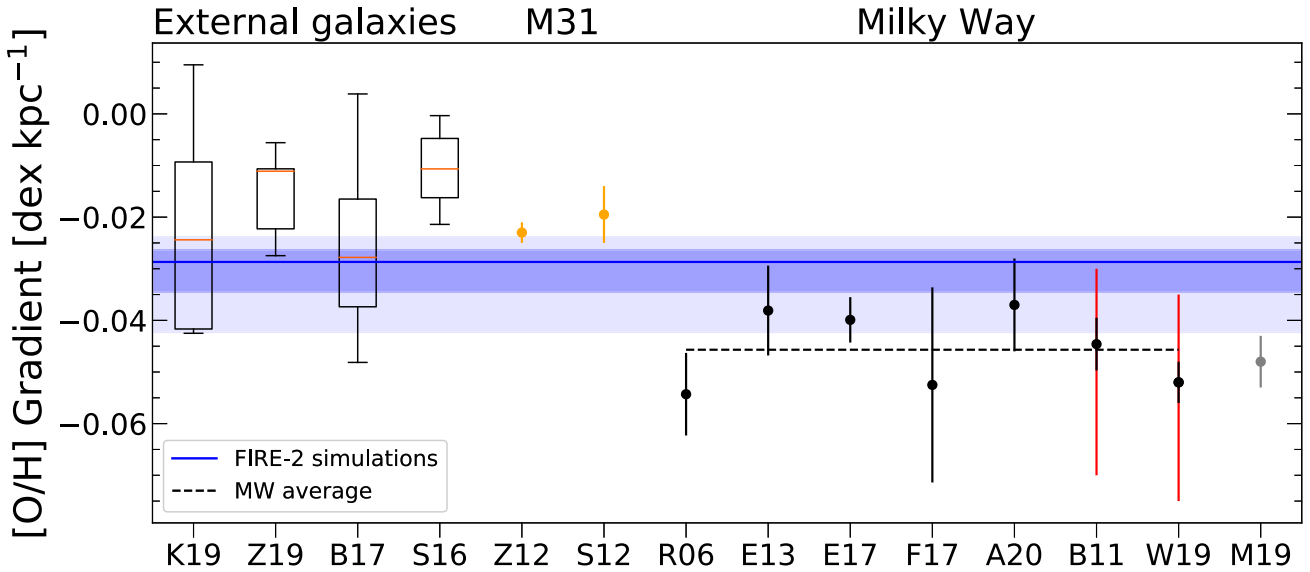| Simulation	M_{90}^{star} $(10^{10}\,\mathrm{M_\odot})$	R_{25}^{star} (kpc)	R_{50}^{star} (kpc)	R_{90}^{star} (kpc)	Z_{90}^{star} (kpc)	M_{90}^{gas} $(10^{10}\,\mathrm{M_\odot})$	R_{25}^{gas} (kpc)	R_{50}^{gas} (kpc)	R_{90}^{gas} (kpc)	Z_{90}^{gas} (kpc)	f_{90}^{gas} $(< R_{90}^{\mathrm{star}}, < Z_{90}^{\mathrm{star}})$
m12m	10.0	1.9	4.3	11.6	2.3	2.1	6.6	10.3	15.0	1.2	0.13
Romulus	8.0	1.2	3.2	12.9	2.4	2.7	9.0	13.1	18.3	2.3	0.16
m12b	7.3	1.0	2.2	9.0	1.8	1.7	6.4	9.6	15.0	1.5	0.11
m12f	6.9	1.2	2.9	11.8	2.1	2.3	8.7	12.6	17.8	2.4	0.14
Thelma	6.3	6.3	3.4	11.2	3.2	2.6	7.5	12.1	17.6	3.1	0.16
Romeo	5.9	1.6	3.6	12.4	1.9	1.8	8.0	12.2	18.1	1.5	0.16
m12i	5.3	1.1	2.6	9.8	2.3	1.7	7.1	10.2	16.7	1.7	0.15
m12c	5.1	1.3	2.9	9.1	2.0	1.5	5.4	8.4	14.6	2.4	0.15
Remus	4.0	1.2	2.9	11.0	2.2	1.5	7.4	11.8	18.0	1.8	0.22
Juliet	3.3	0.8	1.8	8.1	2.2	1.5	6.9	11.3	18.6	3.1	0.14
Louise	2.3	1.2	2.8	11.2	2.2	1.4	8.0	12.6	18.5	2.0	0.34

Figure 3. Radial gradients in gas-phase [O/H] across our 11 galaxies and observed in the MW, M31, and in nearby MW-mass galaxies. The blue horizontal line shows the median across our 11 galaxies, with the dark shaded region showing the 68th percentile and the light-shaded region the full distribution. We also show observations of radial gradients in external galaxies, from Kreckel et al. (2019, K19), Zinchenko et al. (2019, Z19), Belfiore et al. (2017, B17), and Sánchez-Menguiano et al. (2016, S16), via box-and-whisker, where the box displays the 68th percentile, the whiskers display the full distribution, and the orange horizontal line is the median. Orange circles show observed abundance gradients for M31 derived from H II regions by Zurita & Bresolin (2012, Z12) and Sanders et al. (2012, S12). Black circles show observed abundance gradients for the MW derived from H II regions from Rudolph et al. (2006, R06), Balser et al. (2011, B11), Esteban et al. (2013, E13), Esteban et al. (2017, E17), Fernández-Martín et al. (2017, F17), Wenger et al. (2019, W19), and Arellano-Córdova et al. (2020, A20). Mollá et al. (2019a, M19), shown in grey, is the gradient derived from a compilation of the data from R06, B11, E13, E17, and F17. We show uncertainties for all points. For W19 and B11, the red shows the variation in gradient observed by looking along different azimuths. The dashed line shows the best-fitting MW gradient (-0.046 dex kpc^{-1}), based on the gradients the observations presented here (excluding M19). The median [O/H] gradient across our galaxies is -0.028 dex kpc^{-1} and the standard deviation is 0.005 dex kpc^{-1}, in agreement with K19 and B17 to within 1σ, and with Z19 to within 2σ, but not in agreement with S16. Our simulations also agree with observations of M31 and some of the MW observations.

be comparable to our simulations. The Kreckel et al. (2019, K19) sample includes five galaxies from the PHANGS-MUSE survey with $10.2 \leq \log_{10} M_{\mathrm{star}}/\mathrm{M_\odot} \leq 10.6$, the Zinchenko et al. (2019, Z19) sample includes seven galaxies from CALIFA DR3 with $10.2 \leq \log_{10} M_{\mathrm{star}}/\mathrm{M_\odot} \leq 10.8$, the Belfiore et al. (2017, B17) sample includes 13 galaxies from the MaNGA survey with $10.2 \leq \log_{10} M_{\mathrm{star}}/\mathrm{M_\odot} \leq 11$, and the Sánchez-Menguiano et al. (2016, S16) sample includes 20 galaxies from the CALIFA survey with $10.2 \leq \log_{10} M_{\mathrm{star}}/\mathrm{M_\odot}$

≤ 11. In addition to the mass cut, we select galaxies that have gradients measured across a radial range comparable to the range in our analysis (the measured ranges all fall within $2 - 14$ kpc except for K19 which falls within $1 - 11$ kpc). While all of these observed samples show almost exclusively negative gradients in [O/H], their abundance gradients are typically flatter than in our simulations. Our simulations are consistent at the 1σ level with K19 and B17, and at the 2σ level with Z19. However, our sample does not overlap with S16.

Note that the calibrator used for determining the abundances varies from survey to survey. Using different calibrators can give drastically different abundance measurements (Hemler et al. 2020), which could contribute to discrepancies between the different surveys, and to differences with our simulations. Note that the difference between our simulations and these observations are comparable to the differences between surveys themselves.

Fig. 3 also shows observed abundance gradients in M31 and the MW from H II regions. The orange points show observed gradients in M31 from Zurita & Bresolin (2012, Z12) and Sanders et al. (2012, S12). These gradients are slightly shallower than in our simulations, though they agree within 2σ. This may be a consequence of M31 gradient measurements spanning $\approx 4 - 25$ kpc: from our analysis, including the outer regions of a gas disc flattens the inferred gradient.

The black points show measured gradients of the MW. Mollá et al. (2019a), shown in grey, is the best-fitting measurement of the MW abundance gradient based on the combined data of Rudolph et al. (2006), Balser et al. (2011), Esteban et al. (2013, 2017), and Fernández-Martín et al. (2017). We show uncertainties for all samples. The red error bars for Balser et al. (2011) and Wenger et al. (2019) show the impact of measuring the radial gradient along different galactic azimuths. Balser et al. (2011) finds gradients ranging from -0.03 to -0.07 dex kpc^{-1} and Wenger et al. (2019) find gradients ranging from -0.035 to -0.075 dex kpc^{-1} which highlights that measurements of the MW radial gradient are strongly sensitive to azimuthal variations. The different samples include different radial ranges, so they are not exactly comparable to each other or our analysis. Most measurements of [O/H] gradients in the MW overlap with our simulations, though our simulations generally have shallower gradients.

While not included in Fig. 3, Hernandez et al. (2021) recently measured the radial [O/H] gradient in neutral and ionized gas in M83. The gradients were measured out to ≈ 5.5 kpc. They found the gradients in neutral gas to be substantially steeper than the gradients in ionized gas. As most observations target ionized gas around H II regions, one might expect that our measured gradients shown in Fig. 3 are flatter than expected. However, Hernandez et al. (2021) measured gradients primarily in the bulge, which we exclude in this analysis. Their gradient for neutral gas external to the bulge is ≈ -0.02 dex kpc^{-1} and for ionized gas is ≈ -0.03 dex kpc^{-1}, in good agreement with our values.

As a whole, the radial gradients in our simulations are somewhat steeper than in external galaxies but somewhat shallower than in the MW. The MW may be an outlier: as Boardman et al. (2020) note, its gradient is typically steeper than those observed in MW analogues. These differences are likely the result of a combination of different factors, such as: measuring over different radial ranges or using different calibrators. For example, B17 also measure the gradient in their MaNGA observations using a different calibrator for [O/H] (O3N2, not shown here, as opposed to R23, as Fig. 3 shows), which results in a median gradient that is ≈ 0.008 dex kpc^{-1} shallower. Thus, given that S16 used the O3N2 calibrator applied to their CALIFA observations, this may explain the discrepancy between S16 and B17. We defer a more detailed comparison via synthetic observations of our simulations, tailored to each observation, to future work. Rather, Fig. 3 provides a broad comparison, highlighting that the radial gradients of gas-phase [O/H] within our simulations lie within the scatter across the MW, M31, and nearby MW-mass galaxies.

Also, while we do not show it, we compared the [N/H] gradients of our sample at $z = 0$ to observations of M31 and the MW. Our mean [N/H] gradient is -0.039 dex kpc^{-1}, with a standard deviation of 0.007, and the full distribution across hosts spans -0.028 to -0.057 dex kpc^{-1}. This agrees well with values measured for the MW in Esteban & García-Rojas (2018) (comparing three different ionization correction factors, they found [N/H] gradients of -0.047 to -0.050 dex kpc^{-1} with uncertainties of ≈ 0.008) and in Arellano-Córdova et al. (2020) (-0.049 ± 0.007). Our [N/H] gradients also agree with the value measured in M31 by S12) (-0.0303 ± 0.0049 dex kpc^{-1}). However, [N/H] gradients measured in Rudolph et al. (2006) (-0.071 ± 0.010 dex kpc^{-1} in the optical and -0.085 ± 0.010 dex kpc^{-1} in the far-infrared) and Fernández-Martín et al. (2017) (-0.080 ± 0.019 dex kpc^{-1}) are substantially steeper than in our hosts. As with the [O/H] gradients, this may be because the gradients are not measured over the same radial range.

3.2 Evolution of radial gradients

We next explore the evolution of gas-phase radial gradients of [O/H], [Fe/H], and [N/H] at $z \leq 1.5$, over the last ~ 10 Gyr, during the primary epoch of disc assembly, to understand the initial conditions for star formation and chemical tagging of stars. In summary, we find that at earlier times, the gas disc was more radially homogeneous (flatter gradients), so chemical tagging offers less discriminating power for radial birth location at earlier times.

Similar to Fig. 2, Fig. 4 shows radial profiles of [O/H], [Fe/H], and [O/Fe] in gas at different redshifts. The solid line shows the mean across our 11 galaxies, while the shaded region shows the 1σ scatter. At all radii, [O/H] and [Fe/H] increase with time, as the gas mass declines while more stars enrich the interstellar medium (ISM). This evolution agrees with the observed gas-phase galaxy mass–metallicity relation (Tremonti et al. 2004). Ma et al. (2016) explored this evolution across a wide galaxy mass range in the FIRE-1 simulations: they found that as galaxies grow more massive, the mass-loading factor of their winds decreases, and metals are more easily held in/near the galaxy as opposed to being driven into the halo (see also Muratov et al. 2015, 2017; Anglés-Alcázar et al. 2017).

Fig. 4 also shows that both [O/H] and [Fe/H] have negative radial gradients at all times. Ma et al. (2017) also found primarily negative gradients in the FIRE-1 suite, because the high star formation efficiency in the inner discs of galaxies with well ordered rotation leads to sustained negative radial gradients. At $z = 0$, our average change in [O/H] from $0 - 15$ kpc is ≈ 0.51 dex, while this is ≈ 0.56 dex for [Fe/H]. At $z = 1.5$ ($t_{lookback} = 9.4$ Gyr), the average change in abundance from $0 - 15$ kpc is ≈ 0.24 dex for [O/H] and 0.28 dex for [Fe/H]. Furthermore, as expected given the scatter in formation history, we find larger host-to-host scatter (averaged over all radii) at earlier times: 0.09 dex for [O/H] and 0.07 dex for [Fe/H] at $z = 0$, but at $z = 1.5$, this was 0.2 dex for both elements.

Fig. 4 also shows that the abundance profiles were flatter (more homogeneous) at earlier times, because the abundance at smaller radii evolves more rapidly than at larger radii. Increased accretion/merger rates, coupled with higher star formation rates and stronger gas turbulence, drove more efficient radial mixing at earlier times (Ma et al. 2017). In FIRE simulations, early galaxies experience bursty, stellar feedback-driven outflows that radially mix the ISM, in addition to local turbulence. The profiles steepen with time, because as the gas disc settles down and becomes more rotationally supported, it is capable of sustaining stronger radial gradients given less radial mixing (Ma et al. 2017). At $z \lesssim 1$ ($t_{lookback} \lesssim 7.8$ Gyr), the radial profiles generally show an up-turn in the innermost bulge-dominated region. Beyond ≈ 4 kpc, the profiles are well fit by a linear relation at all redshifts.
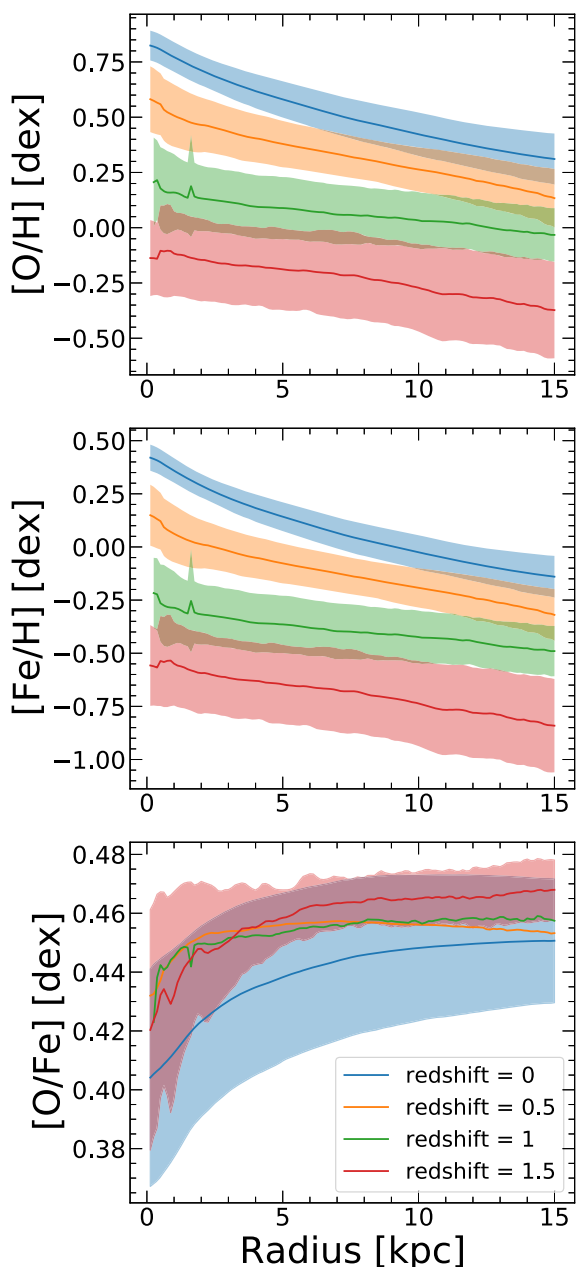
Figure 4. Radial profiles of gas-phase elemental abundances in our simulations at different redshifts. The solid lines show the mean and the shaded regions show the 1σ scatter across our 11 galaxies. The top and middle panels show steepening radial profiles of [O/H] and [Fe/H] with time, as the gas disc becomes more rotation dominated, with less radial turbulence, thus sustaining stronger radial gradients. The bottom panel shows that the innermost regions of the gas disc have lower [O/Fe] than the outer disc, this indicates that the inner disc is more evolved, that is, has had more Ia supernovae that produce more Fe than α-elements (like O), than the outer disc.

Fig. 4 (bottom panel) shows that [O/Fe] tends to decline over time at all radii, because at early times, core-collapse supernovae dominate the enrichment, which preferentially produce α-elements like O. However, the mean trends in Fig. 4 (bottom panel) are not necessarily true for individual hosts. The dynamic range of [O/Fe] is small, so minor upticks in star formation rates lead to enhanced [O/Fe] as the rate of core collapse supernovae temporarily increases. Individual galaxies show overall increases in [O/Fe] following periods of

increased star formation, and different radii of individual hosts show relative increases or decreases in [O/Fe], likely correlated with radial variations in star formation rates.

On average, at later times, the (more delayed) Ia supernovae preferentially enrich the galaxy in Fe, driving down [O/Fe]. However, the typical change in gas-phase [O/Fe] at fixed radius from $z = 1.5$ ($t_{lookback} = 9.4\,\text{Gyr}$) to 0 is only $\approx 0.02 - 0.03$ dex. The [O/Fe] radial gradients are positive at all times, because the outer disc is always younger than the inner disc/bulge, though the gradients are weak at larger radii. The [O/Fe] radial profiles steepen at small radii at later times, at least at $z \lesssim 1$ ($t_{lookback} \lesssim 7.8\,\text{Gyr}$). Unlike the profiles of individual elements, the host-to-host standard deviation of [O/Fe] increases at later times, from 0.015 dex at $z = 1.5$ to 0.037 dex at $z = 0$. Overall, [O/Fe] does not provide strong discrimination power for chemical tagging at any radii or time that we examine.

While not shown here, we also measure the evolution of [Mg/Fe], which is more significant than [O/Fe]. In the outer disc, ($R = 12\,\text{kpc}$) [Mg/Fe] decreases from ≈ 0.3 dex at $z = 1.5$ to ≈ 0.22 dex at $z = 0$. In the inner disc ($R = 4\,\text{kpc}$) the evolution is larger, from ≈ 0.29 dex to 0.18 dex over the same redshifts. The stronger evolution seen in [Mg/Fe] likely results from stellar winds in our simulations producing relatively little Mg. The stellar-wind model in FIRE-2 has a metallicity-dependent yield for O. At lower redshifts, as the gas disc becomes more enriched, O production also increases, leading to less evolution in [O/Fe].

Fig. 5 shows the evolution of the mean radial gradients of [O/H], [Fe/H], and [N/H] at redshifts 0, 0.2, 0.5, 0.75, 1, 1.1, 1.2, 1.25, 1.3, 1.4, and 1.5. The shaded regions show the 1σ scatter. We also checked that the evolution of each galaxy qualitatively agrees with this mean evolution. We measure radial gradients between $4 < R < 12\,\text{kpc}$ at $z < 1$ ($t_{lookback} < 7.8\,\text{Gyr}$), consistent with Fig. 3, and between $0 < R < 8\,\text{kpc}$ at $z \geq 1$ ($t_{lookback} \geq 9.4\,\text{Gyr}$). (For a few hosts with bulge-like upturns in their profiles at $\lesssim 2\,\text{kpc}$, we measure their gradients between 2 and 8 kpc, where the gradient is nearly linear). We use a redshift-dependent radial range, because we are exploring the gas from which stars are forming. We select gas approximately contained in R_{90}^{star}, which is $\approx 11\,\text{kpc}$ at $z = 0$ and $\lesssim 8\,\text{kpc}$ at $z \gtrsim 0.5$. At $z < 1$, we measure the gradient at $4 - 12\,\text{kpc}$, because the inner disc is dominated by the bulge region and has a steeper profile, as Fig. 4 shows. Thus, at late times, we measure at $4 - 12\,\text{kpc}$, which encompasses both the inner and outer disc and exhibits a nearly linear profile. We tested our analysis measuring the gradient over varying radial ranges and found that, while the normalization varies somewhat, the shape of the profile and the evolution are consistent, as Fig. 4.

As Fig. 5 shows, the strength of the radial gradient generally decreases over time. The minimum magnitude of the gradient is $\approx 0.01\,\text{dex}\,\text{kpc}^{-1}$ and occurs at $z \approx 1.5$ ($t_{lookback} \approx 9.4\,\text{Gyr}$). We find just two galaxies that achieve a flat gradient at this time. At $z \lesssim 1$ ($t_{lookback} \lesssim 7.8\,\text{Gyr}$), the radial gradients gradually steepen to $-0.03\,\text{dex}\,\text{kpc}^{-1}$ at $z = 0$. The gradients prior to $z \approx 1$ are approximately constant with redshift. While there are fluctuations in the gradient at high z, these are transient features in the simulations driven primarily through minor mergers, flybys of satellite galaxies, and starbursts. The host-to-host scatter is smallest at $z = 0$ and is largest at $z = 0.75$ ($t_{lookback} = 6.6\,\text{Gyr}$), in part because of one galaxy (m12f) that experiences a major merger at this time.

Fig. 5 also compares the evolution of the radial gradients in [O/H] and [Fe/H] with [N/H]. Consistent with most results in this paper, we find little-to-no difference between [O/H] and [Fe/H], despite their differing origins, from primarily core-collapse and primarily Ia supernovae, respectively. However, we find systematically stronger
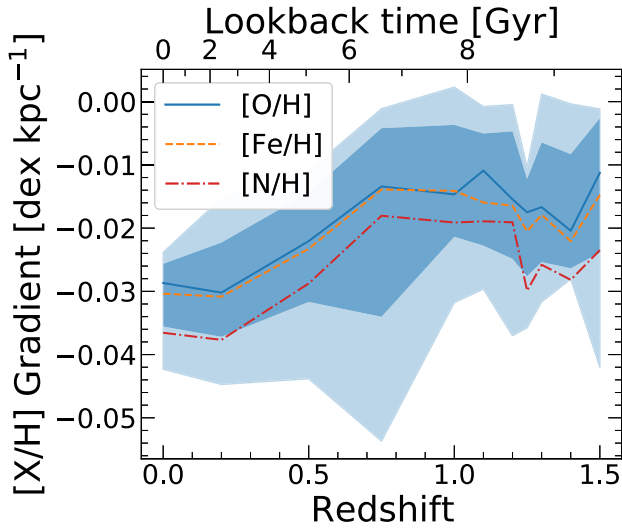
Figure 5. Evolution of radial gradients in gas elemental abundances. The lines show the means for [O/H], [Fe/H], and [N/H]. The dark shaded region shows the 1σ scatter and the light shaded region shows the total scatter in [O/H] across our 11 galaxies. At each redshift, we calculate the gradient via a linear fit across a redshift-dependent radial range: $4 < R < 12$ kpc for $z < 1$ ($t_{lookback} < 7.8$ Gyr) and $0 < R < 8$ kpc for $z \geq 1$. The gradient is flattest (≈ -0.015 dex kpc^{-1}) at $z = 1.5$ ($t_{lookback} = 9.4$ Gyr), likely because high merger rates drive large-scale radial turbulence, erasing the radial gradients of the discs. This leads to the azimuthally averaged metallicity at a given radius showing little radial dependence although stochastic enrichment/accretion events may lead to locally over or underenriched regions. Shallow gradients persist until $z \approx 1$. At lower redshifts, the gas disc becomes more rotationally supported and is capable of sustaining stronger radial gradients, becoming steepest (≈ -0.03 dex kpc^{-1}) at $z = 0$. The gradients of [O/H] and [Fe/H] are almost identical, despite being sourced primarily by core-collapse and Ia supernovae, respectively. However, the [N/H] gradient, sourced primarily through stellar winds from massive stars, which have a metallicity dependent mass-loss rate, is steeper at all times.

radial gradients in [N/H] at all times. Unlike O and Fe, which are sourced primarily through supernovae, N is sourced primarily by stellar winds in the FIRE-2 simulations, and the wind mass-loss rate from massive stars (in the first 3.5 Myr) depends roughly linearly on progenitor metallicity. (The N yield from core-collapse supernovae also increases linearly with progenitor metallicity in the FIRE-2 model, but this effect is subdominant, because most N comes from stellar winds.) The progenitor metallicity dependence of N (often called secondary production of N) results in enhanced N production in regions that are already more metal rich, and thus it drives a steeper gradient for N (by ≈ 0.015 dex kpc^{-1}) than O or Fe at all times.

3.3 Vertical profiles across time

We next examine the vertical profiles (in absolute height) of elemental abundances, for all gas near the solar circle, within a cylindrical radius of $7 < R < 9$ kpc. We normalize the vertical profiles by subtracting the mid-plane abundance at each redshift. Fig. 6 shows the vertical profiles for [O/H]. The solid line shows the mean and the shaded regions show the 1σ scatter; we only show the scatter at $z = 1.5$ ($t_{lookback} = 9.4$ Gyr) and $z = 0$ for clarity.

Fig. 6 shows that any systematic trends in abundance with height to 1 kpc is $\lesssim 0.01$ dex kpc^{-1} absolute on average at all times, and the 1σ scatter is $\lesssim 0.01$ dex kpc^{-1} at $z = 0$ and $\lesssim 0.02$ dex kpc^{-1} at $z = 1.5$. Thus, the gas disc is well mixed vertically. In most of our
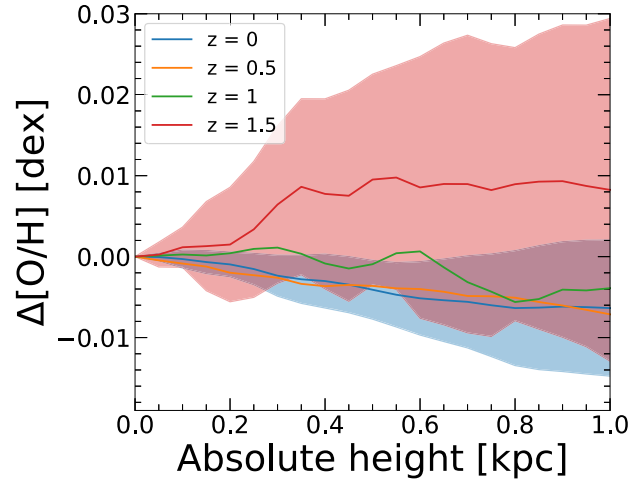

Figure 6. The vertical change in [O/H] relative to the mid-plane at various redshifts. We measure the vertical profile at $R = 7 - 9$ kpc. The solid lines show the mean while the shaded region shows the 1σ standard deviation across 11 galaxies. [O/H] shows minimal variation with height: the mean deviation at 1 kpc above the mid-plane is ≈ 0.008 dex at $z = 1.5$ ($t_{lookback} = 9.4$ Gyr) and ≈ -0.006 dex at $z = 0$. The majority of stars form within a few hundred parsecs of the galactic mid-plane at low z and within ~ 1.5 kpc of the mid-plane at the highest redshift we examine. Close to the plane of the disc ($|Z| \leq 200$ pc), the variation in [O/H] is even smaller at all times ($|\Delta[O/H]| \lesssim 0.001$ dex at $z = 1.5$ and $|\Delta[O/H]| \lesssim 0.002$ dex at $z = 0$). At $z < 1$ ($t_{lookback} < 7.8$ Gyr), we find a weak but systematically negative gradient. We conclude that vertical gradients are not measurably significant for chemical tagging at any redshift.

galaxies, the deviations in abundance increase with distance from the mid-plane, that is, they show a systematic gradient with height. Over time, these vertical gradients become increasingly (if weakly) more negative, which supports the idea of 'upside-down' disc formation (e.g. Bird et al. 2013, 2021; Ma et al. 2017), such that stars formed in a more vertically extended disc at higher redshifts, leading to more enrichment at larger heights, at later times stars formed in a thinner disc and gas farther out of the mid-plane became relatively less enriched. At $z \gtrsim 1$, the absolute strength of this vertical gradient is in fact comparable to the radial gradient (Fig. 5), while at $z = 0$, the vertical gradient is ≈ 3 times weaker than the radial gradient. This is because the time-scale for vertical mixing is short, given gas turbulence, and that the vertical scale-height of the gas is itself set by the maximum Jeans length at that time. Furthermore, with implications for chemical tagging, the majority of star formation in our simulations is limited to $\lesssim 500$ pc of the mid-plane at $z < 0.5$, and $\lesssim 1.5$ kpc up to $z < 1.5$, and Fig. 6 shows that vertical variations in abundance are minimal on those scales. The vertical trends in [Fe/H] (not shown here) are consistent with [O/H] within ≈ 0.01 dex.

3.4 Azimuthal variations across time

We next investigate azimuthal variations of elemental abundances in gas, including its evolution. We, thus, test a common assumption in galactic evolution, that gas is well mixed azimuthally within a given annulus (e.g. Frankel et al. 2018, 2020).

Fig. 7 shows the standard deviation in [O/H] and [O/Fe] along angular bins of varying length at fixed radius. Specifically, we compute the standard deviation within a given angular bin, and Fig. 7 shows the mean standard deviation across all bins of a given angular size for all 11 simulations. We stack three snapshots ($\Delta t \approx 50$ Myr)
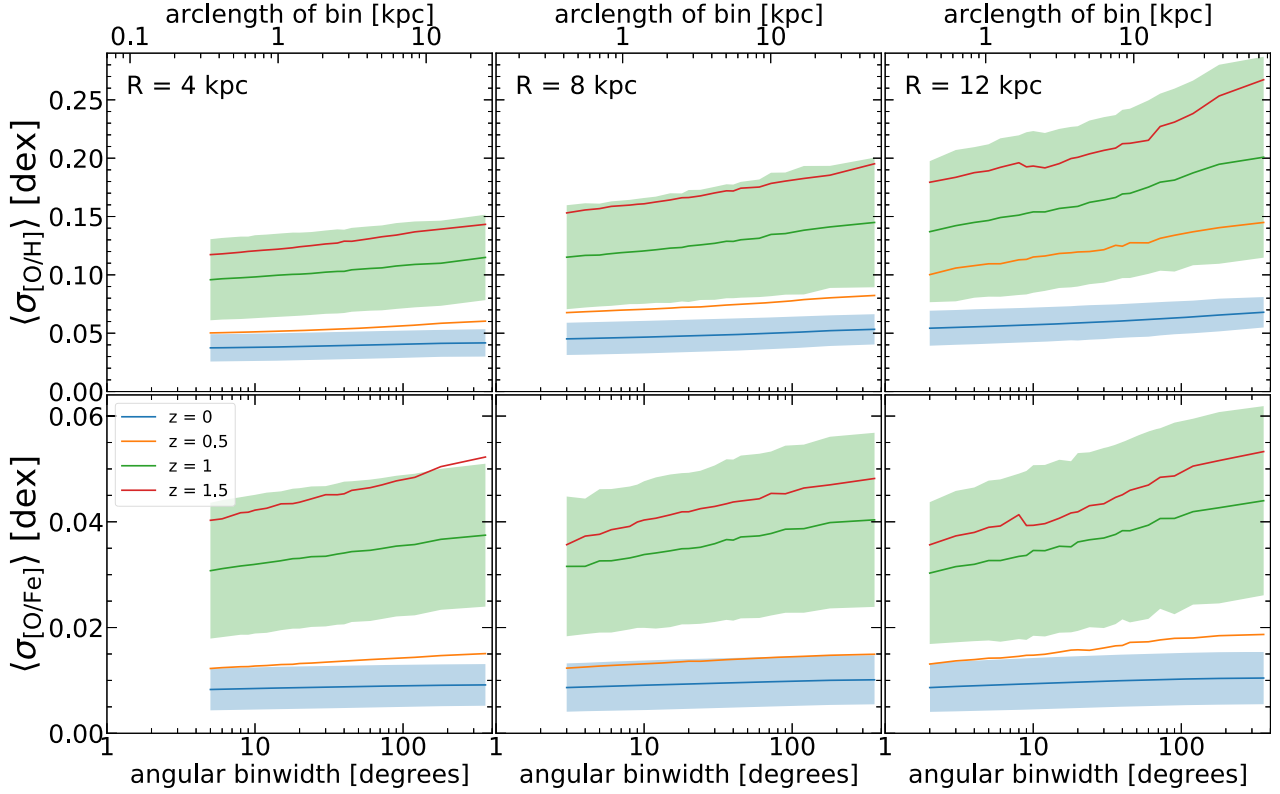
Figure 7. Azimuthal scatter in elemental abundances for [O/H] and [O/Fe] in gas, as a function of angular scale, at different redshifts and different radii. The solid lines show the mean and the shaded regions show the 1σ scatter across our suite of 11 galaxies: we show scatter only at $z = 1$ ($t_{lookback} = 7.8$ Gyr) and 0. The scatter increases as a function of angular bin size at all redshifts and at all radii. At $z = 0$, near the solar circle ($R = 8$ kpc), the average azimuthal scatter across the disc is ≈ 0.053 dex for [O/H] and [Fe/H] (not shown) and ≈ 0.009 dex for [O/Fe]. For all angular bin sizes, the average scatter increases with redshift: at earlier cosmic times, the gas discs were less well mixed within a given annulus. At $z = 0$, the scatter across the disc is ≈ 0.2 dex for [O/H] and [Fe/H] (not shown) and ≈ 0.05 dex for [O/Fe]. The scatter also increases with angular bin size at all redshifts, although the increase is minimal at late times. At low z, this means that azimuthal variations are dominated by local (and not global) fluctuations in the disc. Finally, the azimuthal scatter increases with radius for individual abundances: gas is azimuthally better mixed in the inner disc, likely a result of shorter orbital times leading to faster mixing.

for each redshift. We use an annulus of gas ±0.3 kpc out of the plane of the disc because as shown in Fig. 6, gas-phase abundances are effectively homogeneous within this height. We also measure within a radius ±0.15 kpc of the selected cylindrical radius, to ensure that the angular length dominates over the radial length for our smallest angular bins, that is, to ensure that the radial gradient does not induce significant scatter. We show the 1σ scatter for $z = 0$ and 1 ($t_{lookback} = 7.8$ Gyr). We exclude m12c at $z = 1.5$ for angular scales $\Delta\phi \leq 8°$, because its angular bins contain too few gas particles.

At $z = 0$ and $R = 8$ kpc (near the solar circle), the typical azimuthal scatter across the gas disc is $\lesssim 0.053$ dex for [O/H], $\lesssim 0.055$ dex for [Fe/H] (not shown), and $\lesssim 0.01$ dex for [O/Fe]. This value for [O/H] agrees well with MW observations (Wenger et al. 2019) and observations of external galaxies (Sakhibov et al. 2018; K19; Kreckel et al. 2020), though we emphasize that we are not measuring azimuthal scatter in the same way: those observations typically measure differences in abundances between arm and inter-arm regions or measure abundance variations between H II regions within an aperture of a given size.

Our azimuthal scatter decreases with smaller angular bin size, with a minimum of ≈ 0.045 dex for [O/H] (≈ 0.046 dex for [Fe/H]) and ≈ 0.009 dex for [O/Fe] at the smallest angular scales. Interestingly, this minimal scatter remains well above 0 dex as $\Delta\phi$ goes to 0. We emphasize that our analysis does not zoom-in on giant molecular

clouds (GMCs) or individual star-forming regions, but rather we examine all of the ISM centered on (effectively) random positions. Thus, our results on small scales do not immediately inform the homogeneity of individual GMCs, especially given their short lifetimes ($\lesssim 7$ Myr) in our simulations (Benincasa et al. 2020), and we will examine GMC homogeneity in future work. Appendix B also examines how small-scale variations depend on our choice of diffusion coefficient for sub-grid turbulent mixing in gas.

The 1σ host-to-host scatter is approximately independent of bin size and is $\lesssim 0.014$ dex for [O/H], $\lesssim 0.015$ dex for [Fe/H], and $\lesssim 0.005$ dex for [O/Fe]. Thus, at $z = 0$, gas within all of our galaxies is well mixed, that is, the azimuthal scatter is comparable to typical measurement uncertainties (~ 0.05 dex) for elemental abundances.

Fig. 7 shows that, at all radii and at all angular bin sizes, the azimuthal scatter was more significant at earlier times, that is, gas was less azimuthally mixed than it is now. This is likely because higher accretion and star formation rates combined with burstier star formation leads to more pronounced local pockets of enrichment in gas. At $R = 8$ kpc and at $z = 1.5$ ($t_{lookback} = 9.4$ Gyr), the azimuthal scatter across the disc is $\lesssim 0.2$ dex for [O/H] and [Fe/H] (not shown) and $\lesssim 0.05$ dex for [O/Fe]. The scatter does not drop below ≈ 0.15 dex for either [O/H] or [Fe/H] at the smallest azimuthal scales (0.035 dex for [O/Fe]). The 1σ host-to-host scatter is $\lesssim 0.05$ dex for [O/H] and [Fe/H] and $\lesssim 0.016$ dex for [O/Fe].
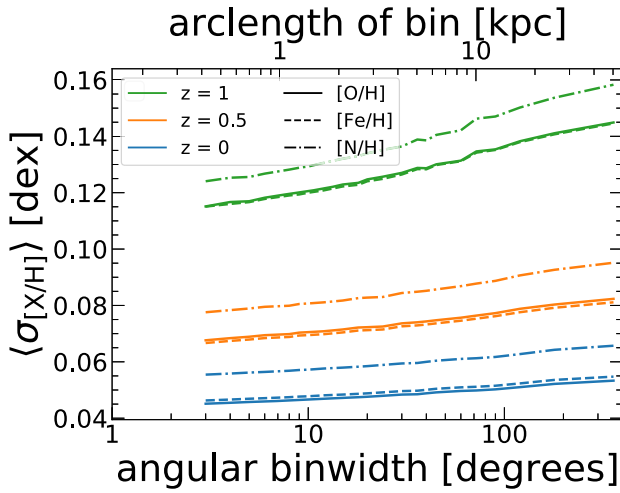
Figure 8. Azimuthal scatter, as in Fig. 7, at $R = 8\,\mathrm{kpc}$ for [O/H] (solid), [Fe/H] (dashed), and [N/H] (dash dot). As with the radial gradients, the azimuthal variations of [O/H] and [Fe/H] are almost identical, despite being sourced primarily by core-collapse and Ia supernovae, respectively, while the variation in [N/H], which is sourced primarily through metallicity-dependent stellar winds from massive stars, is higher at all redshifts.

Additionally, Fig. 7 shows that the difference in scatter between large and small angular scales varies with time. This difference is more significant at earlier times: at $z \gtrsim 1$ ($t_{\mathrm{lookback}} \gtrsim 7.8\,\mathrm{Gyr}$), this change is $\approx 0.042\,\mathrm{dex}$ at $R = 8\,\mathrm{kpc}$. Thus, at early times, galaxy-scale fluctuations are more important in driving azimuthal scatter (as visible in Fig. 1). However, at $z \sim 0$, the azimuthal scatter across small versus large angular scales differs by only $\approx 0.008\,\mathrm{dex}$, so small-scale fluctuations drive most of the azimuthal scatter (also visible in Fig. 1). These results at low redshift are useful from an observational perspective, because they mean that one can generalize smaller scale observations of gas-phase abundances to overall azimuthal trends at fixed radius.

Fig. 7 also shows that the azimuthal scatter depends on radius. The azimuthal scatter increases with increasing radius for a given angular bin size, and in fact, this is true for both fixed angular and physical bin size. At $z = 0$, the azimuthal scatter across the entire disc at $R = 4\,\mathrm{kpc}$ is $\lesssim 0.042\,\mathrm{dex}$ for [O/H] ($\lesssim 0.046\,\mathrm{dex}$ for [Fe/H], not shown), and this increases to $\lesssim 0.062\,\mathrm{dex}$ for [O/H] and [Fe/H] at $R = 12\,\mathrm{kpc}$. At $z = 1.5$ ($t_{\mathrm{lookback}} = 9.4\,\mathrm{Gyr}$), the azimuthal scatter ranges from $\lesssim 0.015\,\mathrm{dex}$ to $0.25\,\mathrm{dex}$ for [O/H] and [Fe/H]. We do not find radial dependence in [O/Fe], which has a maximal scatter $\lesssim 0.01\,\mathrm{dex}$ ($\lesssim 0.052\,\mathrm{dex}$) at all radii at $z = 0$ ($z = 1.5$). Most likely, the radial dependence in the azimuthal scatter of [O/H] and [Fe/H] results from the increase of the orbital time-scale, and hence the time-scale for mixing, with radius. Furthermore, cosmic accretion of underenriched gas also likely contributes to this increase in azimuthal scatter with radius, especially at earlier times, when the increase with radius is stronger.

Fig. 8 shows the azimuthal scatter of [O/H], [Fe/H], and [N/H] at $R = 8\,\mathrm{kpc}$ at three redshifts. As we discussed above, metallicity-dependent stellar winds from massive stars, rather than supernovae, primarily source N, so this compares the azimuthal mixing of elements sourced by these different processes. The azimuthal scatter of [N/H] is larger than that of [O/H] and [Fe/H] for all bins at each redshift. On the scale of the entire annulus, the scatter in [N/H] is approximately 0.015 dex larger at $z = 1$ ($t_{\mathrm{lookback}} = 7.8\,\mathrm{Gyr}$) and approximately 0.01 dex larger at $z = 0$. This discrepancy is slightly
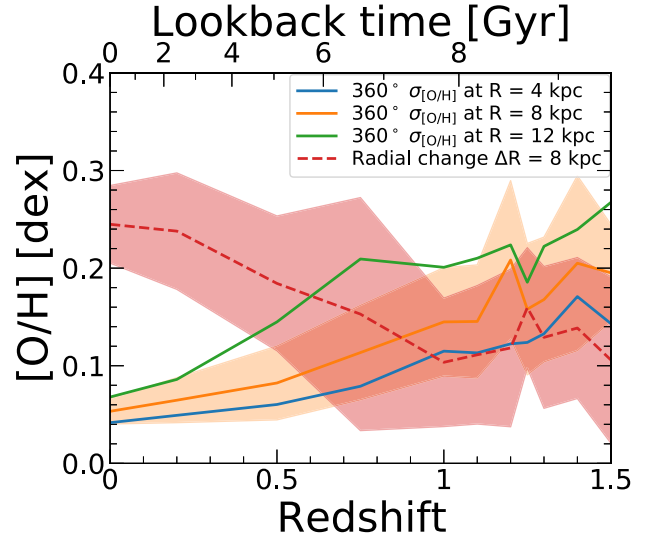

Figure 9. The evolution of variations in [O/H] in gas, both radially and azimuthally. (We do not show [Fe/H], its trends are consistent with [O/H] to $\lesssim$ 0.01 dex.) The solid lines show the mean scatter, across our 11 galaxies, for the full (360°) annulus of gas at $R = 4\,\mathrm{kpc}$ (blue), $R = 8\,\mathrm{kpc}$ (yellow), and $R = 12\,\mathrm{kpc}$ (green). The red dashed line shows the mean radial change in [O/H] across a radial distance of 8 kpc. The shaded regions show the 1σ scatter. While the radial gradient dominates the spatial variations at late cosmic times, azimuthal variations were more significant than the radial gradient at earlier cosmic times ($z \gtrsim 0.9$, or $t_{\mathrm{lookback}} \gtrsim 7.4\,\mathrm{Gyr}$, at $R = 8\,\mathrm{kpc}$). Larger radii transition to radially dominated abundance variations at later times (see also Fig. 10). Thus, elemental evolution models should not assume azimuthal homogeneity of abundances at early times; instead, azimuthal variations are the primary source of spatial information for chemical tagging of stars forming at early times.

smaller for smaller angular scales at $z = 1$ (approximately 0.01 dex difference), but the difference in azimuthal scatter is independent of scale at $z = 0$. As previously stated, our stellar-wind rate depends linearly on progenitor metallicity, which likely drives these differences for N as compared with O and Fe. One might expect Fe, being sourced primarily by Ia supernovae, to have less scatter at small scales than O, because Ia are caused by preferentially older stars than core-collapse supernovae, which occur closer to stellar birth location. A possible cause of our similarity comes from our assumed Ia delay-time distribution (Mannucci et al. 2006), which has a significant component from prompt Ia, at ages $\lesssim 100\,\mathrm{Myr}$. A Ia delay-time distribution with a more significant contribution from older stellar ages (Maoz & Graur 2017) may lead to less small-scale scatter (e.g. Gandhi et al., in preparation).

3.5 Azimuthal versus radial variations across time

We now compare the relative importance of radial gradients versus azimuthal scatter in gas-phase abundances. Fig. 9 shows the evolution of the radial variations in [O/H], from multiplying the radial gradient (as calculated in Section 3.2) of each simulation at each redshift by 8 kpc, to show the change from the inner to the outer disc. This measures the change from 4 to 12 kpc at $z < 1$ ($t_{\mathrm{lookback}} < 7.8\,\mathrm{Gyr}$ and from 0 to 8 kpc at $z \geq 1$, where the radial profile is approximately linear (see Fig. 4). Fig. 9 also shows evolution of the azimuthal scatter around each disc (360°) at 3 radii.

Fig. 9 shows that at early times, $z \gtrsim 1$ ($t_{\mathrm{lookback}} \gtrsim 7.8\,\mathrm{Gyr}$), the 360° angular scatter in [O/H] at all radii is larger than the radial
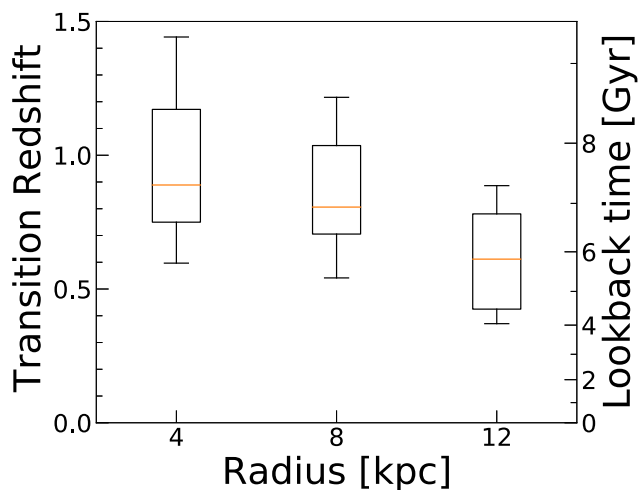
Figure 10. Following Fig. 9, the redshift below which radial variations in [O/H] dominate over azimuthal variations at 3 radii in our 11 simulations (the intersection of the dashed and solid lines). The horizontal line shows the median, the box shows the 68th percentile, and the whiskers show the full distribution. This transition redshift is the last time the azimuthal variation is stronger than the radial variation. This transition occurs earlier at smaller radii, where azimuthal variations are smaller (Fig. 7), given the shorter time-scale for mixing at smaller radii. Before these transition redshifts, any model of chemical tagging should account for azimuthal variations as the primary source of spatial variation.

variation, with the angular scatter at large radii being approximately two times higher. However, the radial variation dominates at all radii at $z \lesssim 0.6$ ($t_{lookback} \lesssim 5.8\,\text{Gyr}$). At $z = 0$, the radial variations in [O/H] is approximately a factor of 4 higher than the azimuthal variation. Thus, at $z \sim 0$, one can approximate the gas disc variations primarily via the radial gradient, but at earlier times, the azimuthal variations dominate.

Fig. 10 shows the redshift (and lookback time) when the radial variation begins to dominate over the azimuthal scatter in [O/H] (the intersection of the solid and dashed lines in Fig. 9), for 3 radii. We measure this transition redshift separately for each simulation: the horizontal lines show the median, and the boxes and whiskers show the 68th percentile and the full distribution. The radial variation starts to dominate earlier at smaller radii, because the azimuthal scatter increases with radius at all redshifts (Fig. 7). The median transition redshift, after which the radial variation dominates, is $z \approx 0.9$ ($t_{lookback} \approx 7.4\,\text{Gyr}$) at $R = 4\,\text{kpc}$, $z \approx 0.8$ ($t_{lookback} \approx 6.9\,\text{Gyr}$) at $R = 8\,\text{kpc}$, and $z \approx 0.6$ ($t_{lookback} \approx 5.8\,\text{Gyr}$) at $R = 12\,\text{kpc}$.

While we, in general, find no systematic differences between our LG-like hosts and our isolated hosts, the transition redshift of the LG-like simulations is systematically higher than that of the isolated simulations at large radii. At $R = 12\,\text{kpc}$, the the median transition of the LG suite occurs at $z \approx 0.8$ ($t_{lookback} \approx 6.9\,\text{Gyr}$) compared to $z \approx 0.5$ ($t_{lookback} \approx 5.1\,\text{Gyr}$) for the isolated hosts. The overall distribution of transition redshifts for the isolated hosts falls within the distribution of transition redshifts for the LG suite, with the 68th percentiles almost entirely overlapping. The slight tendency for LG-like hosts to transition to radial domination earlier is unsurprising, given the results in Santistevan et al. (2020), who found that the main progenitors of our LG-like galaxies formed earlier than those of the isolated galaxies. This earlier formation may be responsible for the earlier transition redshift as the discs undergo more orbital times, which smooths azimuthal variation, and have a longer time to build up strong radial gradients.
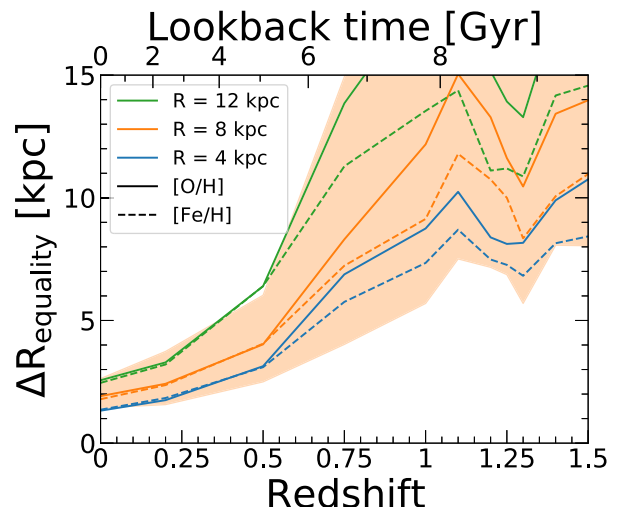

Figure 11. $\Delta R_{equality}$ is the ratio of the azimuthal scatter to the radial gradient: it indicates the characteristic radial scale below which azimuthal variations dominate over radial variations, for [O/H] (solid) and [Fe/H] (dashed). We compute the azimuthal scatter using the full (360°) annulus of gas at $R = 4$, 8, and 12 kpc. Lines show the median, and shaded regions show the 68th percentile across our 11 galaxies. At early times, $\Delta R_{equality}$ is comparable to the size of the gas disc, when azimuthal scatter dominated over radial variation, but after $z \approx 0.5$ ($t_{lookback} = 5.1\,\text{Gyr}$) the disc is well mixed azimuthally. At $z = 0$, the median $\Delta R_{equality}$ near the solar circle is $\approx 1.8\,\text{kpc}$ for [O/H] and $\approx 1.7\,\text{kpc}$ for [Fe/H], while at $z = 1.5$ ($t_{lookback} = 9.4\,\text{Gyr}$), it is $\approx 14\,\text{kpc}$ for [O/H] and $\approx 11\,\text{kpc}$ for [Fe/H]. For radial scales less than $\Delta R_{equality}$, the primary source of inhomogeneity of elemental abundances in gas is azimuthal variations, which complicate/limit the use of elemental abundances to infer a star's birth location to a radial precision smaller than $\Delta R_{equality}$.

Fig. 11 shows another metric for comparing the importance of azimuthal versus radial variations across time. For each simulation, we compute $\Delta R_{equality}$, the ratio of the 360° azimuthal scatter (at a given radius) to the radial gradient. This represents the characteristic radial scale over which the radial and azimuthal abundance variations are equal. The lines show the median $\Delta R_{equality}$ for [O/H] and [Fe/H], measuring the azimuthal variation at 3 radii, and the shaded region shows the 68th percentile of [O/H] at $R = 8\,\text{kpc}$. We apply boxcar smoothing to the data to minimize the significant stochasticity in these values at early times, when the radial gradient fluctuated over short time-scales (see Section 3.2).

The median $\Delta R_{equality}$ for [O/H] is $\lesssim 14\,\text{kpc}$ at $z = 1.5$ ($t_{lookback} = 9.4\,\text{Gyr}$) and $\lesssim 1.8\,\text{kpc}$ at $z = 0$ at the solar circle. This corresponds to the radial range over which azimuthal scatter dominates the variations in abundance, rather than the radial gradient. Thus, for the purposes of chemical tagging, this represents a limit for the radial precision that chemical tagging (of a single element) can place on the formation location of a star without also modelling azimuthal location. At early times, $\Delta R_{equality}$ is comparable to or greater than the size of the disc, meaning that the azimuthal coordinate determines the abundance of newly forming stars more than the radial position.

$\Delta R_{equality}$ is largest at $z = 1.5$ ($t_{lookback} = 9.4\,\text{Gyr}$) and then decreases with time, given the decreasing azimuthal scatter and increasing strength of the radial gradient with time. Also, over time the scatter across our 11 galaxies decreases. The high scatter at high redshifts is a result of the large scatter in both radial gradients and azimuthal variations at these times, given scatter in formation history.
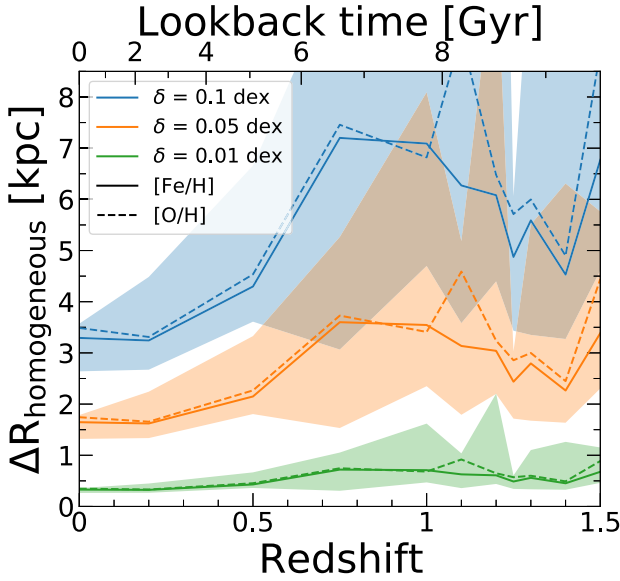
Figure 12. Assuming a given measurement uncertainty in an elemental abundance, $\Delta R_{homogeneous}$ is the ratio of this measurement uncertainty to the radial gradient in that abundance; it is the characteristic radial scale below which the gas disc is effectively radially homogeneous in a measurable sense. We show $\Delta R_{homogeneous}$ versus redshift, for three measurement uncertainties, motivated by observational surveys of stellar abundances. The lines show the median across our 11 galaxies for [Fe/H] (solid) and [O/H] (dashed), and the shaded regions show the 68th percentiles for [Fe/H]. At $z = 1.5$ ($t_{lookback} = 9.4$ Gyr), the gas disc is measurably homogeneous (for $\delta = 0.05$ dex) across significant radial scales, $\Delta R \lesssim 3$ kpc, though this decreases to $\lesssim 1.6$ kpc at $z = 0$. For high-precision abundance measurements ($\delta = 0.01$ dex), the gas disc is measurably homogeneous at $\Delta R \lesssim 0.6$ kpc at $z = 1.5$ and $\Delta R \lesssim 0.3$ kpc at $z = 0$. This highlights the limitations from just measurement uncertainties on chemical tagging to infer a star's birth radius; it does not include the additional complications from azimuthal variations (Fig. 11), which can be more important.

Fig. 11 also shows the dependence of $\Delta R_{equality}$ on radius. This ratio slightly increases as a function of radius, because the azimuthal scatter increases with radius at all times (Fig. 7). This means that modelling chemical tagging of stellar birth radius is more challenging at larger radii.

In summary, any models for chemical tagging should incorporate azimuthal scatter in abundance especially at $z \gtrsim 0.6$ ($t_{lookback} \gtrsim 5.8$), because the azimuthal scatter in gas dominates at these early times.

3.6 Radial scale of measurable homogeneity

We next explore observational implications of our measured radial gradients, by comparing them against typical measurement uncertainties in elemental abundances, to understand the radial scales over which gas is effectively homogeneous in a measurable sense. Thus, in this sub-section, we ignore azimuthal variations and focus just on radial gradients. While we examine gas-phase abundances, the chemical tagging of stars (that form out of this gas) ultimately motivates our work, so we examine measurement uncertainties typical of MW stellar surveys.

Motivated by observational surveys of stellar abundances, we select three observational measurement uncertainties of $\delta_m = 0.01$, 0.05, and 0.1 dex. Fig. 12 shows $\Delta R_{homogeneous}$, the ratio of δ_m to the radial gradient in abundance, versus redshift. Unlike Fig. 11, which shows radial scales of homogeneity at different radii, which

depends on the azimuthal scatter at each radius, Fig. 12 depends only on the radial gradient measured across the whole disc. Thus, the radial scale in Fig. 12 represents the average radial scale of observed homogeneity based on an azimuthally averaged radial gradient. The solid line shows the median and the shaded region shows the 68th percentile across our 11 hosts. $\Delta R_{homogeneous}$ represents the radial scale of measurable homogeneity: below this radial length scale, the change in abundance from the radial gradient is less than this measurement uncertainty. The larger $\Delta R_{homogeneous}$ is, the less precisely measurements can pinpoint a star's birth radius.

For a fiducial measurement uncertainty of $\delta = 0.05$ dex at $z = 1.5$ ($t_{lookback} = 9.4$ Gyr) $\Delta R_{homogeneous} \approx 3.1$ kpc, which drops to ≈ 1.6 kpc at $z = 0$ for [Fe/H]. [O/H] is consistent to within ≈ 1 kpc at all redshifts, except for $z = 1.5$. At early times, $\Delta R_{homogeneous}$ is larger and has large scatter. The largest scatter, at $z = 1.5$, comes from the radial gradients being flattest: some galaxies have gradients approaching 0 dex kpc^{-1}, as Fig. 5 shows. Fig. 12 shows essentially an inverse gradient, so the short time fluctuations at early times (see Section 5) also lead to rapid and significant variations in $\Delta R_{homogeneous}$. After $z = 0.75$ ($t_{lookback} = 6.6$ Gyr), $\Delta R_{homogeneous}$ decreases over time. This means that chemical tagging with measured abundances can identify the birth radius of more recently formed stars more precisely.

Comparing Fig. 11 with Fig. 12 shows that, in terms of limitations on chemical tagging for a star's birth radius, at $z \gtrsim 0.5$ ($t_{lookback} \gtrsim 5.1$ Gyr), azimuthal variations dominate over observational uncertainties in the inner disc, for a fiducial uncertainty of $\delta = 0.05$ dex. In the outer disc ($R \gtrsim 8$ kpc), azimuthal variations are larger than observational uncertainties for all redshifts. For higher precision measurements, $\delta = 0.01$ dex, azimuthal variations dominate at all times at all radii. This implies that, if a primary motivation for chemical tagging is inferring the birth location of a star, there is not much benefit in pushing to higher precision, because azimuthal variations dominate. In fact, Fig. 9 can indicate the maximum precision in elemental abundance that one should aim to measure stars of a given age for this purpose, given our predicted azimuthal scatter, unless a given chemical tagging approach includes modelling azimuthal variations. We will explore possible models in future work.

3.7 Distributions of elemental abundances

Finally, we explore the full distributions of elemental abundances in gas that our simulations predict. Again, we emphasize that our FIRE-2 simulations explicitly model the sub-grid diffusion/mixing of elemental abundances in gas via unresolved turbulent eddies, which is necessary to match observed distributions of abundances in the local Universe (Su et al. 2017; Escala et al. 2018; Hopkins et al. 2018).

We measure [O/H] and [Fe/H] distribution at $R = 4$ kpc, 8 kpc, and 12 kpc at $z = 1$ ($t_{lookback} = 7.8$ Gyr), $z = 0.5$ ($t_{lookback} = 5.1$ Gyr), and $z = 0$ for all galaxies. We fit these with a skew normal distribution, using the LevMarLSQ fitter in ASTROPY (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018)

$$\frac{dF}{dx} = A \times \exp\left(-0.5\left(\frac{x-\mu}{\sigma}\right)^2\right) \times \frac{1 + erf\left(\alpha \times \frac{x-\mu}{\sqrt{2}\sigma}\right)}{2}, \quad (1)$$

where μ is the mean, σ is the standard deviation, and α is the skewness. Fig. 13 shows representative example distributions of [Fe/H], for good and bad fits to this distribution, for a single simulation, and we list the percent of galaxies and radii that fall into each category.
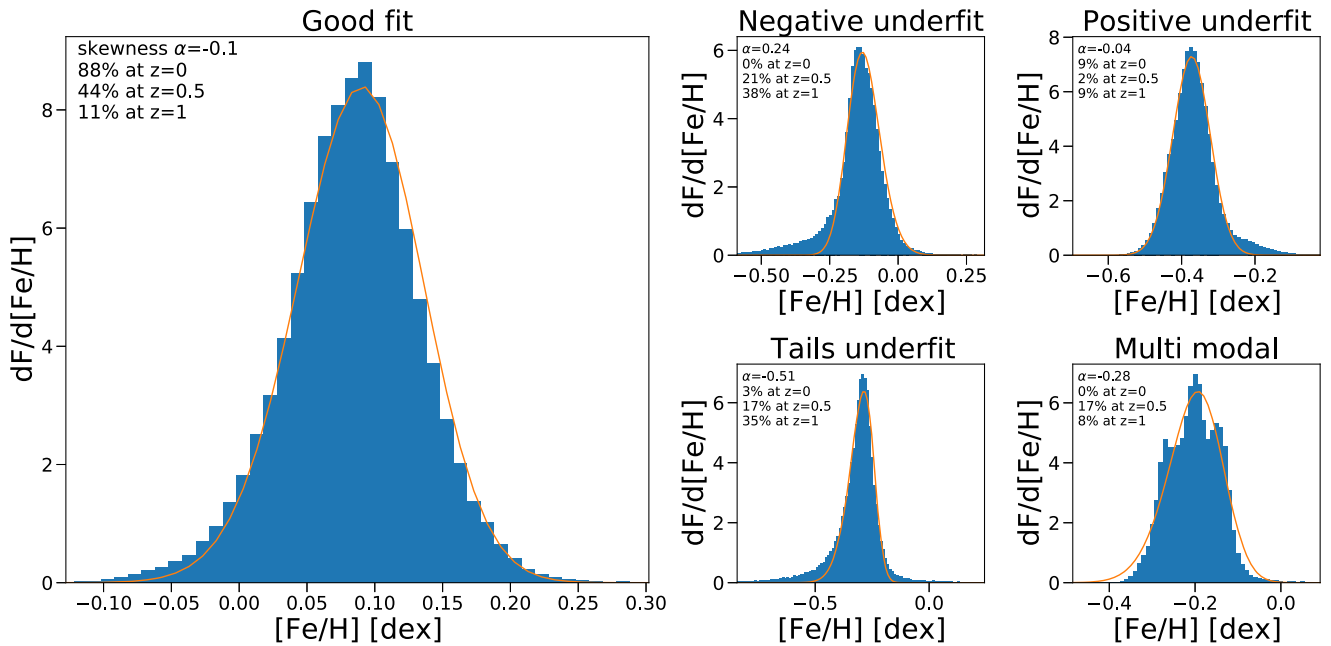
Figure 13. Example distributions of [Fe/H] from our 11 galaxies. Each panel shows the elemental distribution for a single simulation at a single radius. For each distribution, we stack three consecutive snapshots ($\Delta t \approx 50$ Myr) and measure all gas within ± 0.2 kpc of 3 radii ($R = 4$, 8, and 12 kpc) and within a height ± 1 kpc of the disc. The solid line shows a skew normal fit to the distribution. The left-hand panel shows a well fit skew normal distribution. The right-hand panel shows typical failure modes of fitting a skew normal distribution: underfitting negative skewness, underfitting positive skewness, underfitting the tails of the distribution, and multimodal distributions. Each panel shows the fitted skewness for the example distribution along with the percentage of fits that fall into the category at each redshift across all radii and for both [O/H] and [Fe/H]. In general, the simulations are well fit by a skew normal at $z = 0$, but it provides a worse fit at higher redshift, when the (negative) tails are preferentially underfit. In general, the fit to the distribution of [O/H] and [Fe/H] for the same R and redshift fall into the same category.

As the left-hand panel of Fig. 13 shows, a skew normal distribution reasonably fits these distributions in most cases at $z \sim 0$. However, there are several common failure modes. We categorize them as: failing to capture the positive or negative tails of the distribution, failing to capture the width of the distribution, or the distribution being multimodal. The right-hand panel of Fig. 13 shows examples of each of these failures, along with the percentage of fits ([Fe/H] and [O/H] combined) that we identify to fall into each category at each redshift, stacking all galaxies and radii at that redshift. In general, the fit to [O/H] and [Fe/H] at a given redshift and radius falls into the same category. At $z = 0$, the vast majority (≈ 88 per cent) of the distributions are well fit. The most common failure is a positive underfit, given pockets of high metal enhancement from feedback. However, at $z = 1$ ($t_{\text{lookback}} = 7.8$ Gyr), the failures are more common, and only ≈ 11 per cent are well fit. Most common is having a negative underfit or both tails underfit, likely driven by more rapid accretion of low-metallicity gas at earlier times.

While not perfect, especially at earlier times, a skew normal fit does provide a simple characterization of the full distribution. Thus, Fig. 14 shows the fit parameters for [Fe/H] ([O/H], not shown, is consistent with this) at different radii and redshifts. The box-and-whisker plots show the median, 68th percentile, and the full distribution. The top row shows the fitted standard deviation, while the bottom row shows the fitted skewness, and left- to right-hands show increasing redshift.

At all radii, σ decreases over time. At $R = 8$ kpc, near the solar circle, the median standard deviation decreases from ≈ 0.12 dex at $z = 1$ ($t_{\text{lookback}} = 7.8$ Gyr) to ≈ 0.07 at $z = 0$ for [Fe/H] (≈ 0.11 to ≈ 0.06 for [O/H]). Thus, consistent with the results for azimuthal scatter, gas at a given radius becomes more homogeneous over time.

Also consistent with our results for azimuthal scatter, Fig. 14 shows that at all redshifts, σ increases with radius, that is, metals are less well mixed at larger radii. At $z = 0$, the median is $\sigma \approx 0.05$ at $R = 4$ kpc and increases to ≈ 0.07 at $R = 12$ kpc for [Fe/H] (≈ 0.04 to ≈ 0.06 for [O/H]).

Fig. 14 (bottom row) shows that the distributions are preferentially negatively skewed at earlier times, but they trend towards Gaussian over time. At $R = 8$ kpc, the median skewness is $\alpha \approx -0.54$ at $z = 1$ ($t_{\text{lookback}} = 7.8$ Gyr) and $\alpha \approx -0.21$ at $z = 0$ for [Fe/H] ($\alpha \approx -0.52$ to ≈ -0.23 for [O/H]). At earlier times, skewness decreases with radius, from ≈ -0.19 at $R = 4$ kpc to ≈ -0.58 at $z = 1$ (≈ -0.15 to ≈ -0.5 for [O/H]). At earlier times, higher rates of cosmic accretion of pristine gas can skew the distributions negatively, especially at large radii, where enrichment also is more stochastic given lower star formation rates and orbital/mixing times are longer. At later times, as the gas accretion and star formation rates decrease, the distributions tend towards Gaussian, as metals become better mixed within each annulus. At $z \sim 0$, all radii show abundance distributions consistent with no skewness at the 1σ level.

4 SUMMARY AND DISCUSSION

4.1 Summary

We use a suite of FIRE-2 cosmological zoom-in simulations of 11 MW/M31-mass galaxies to explore the 3D spatial variations and evolution of elemental abundances [O/H], [Fe/H], and [N/H] in gas at $z \leq 1.5$ ($t_{\text{lookback}} \leq 9.4$ Gyr), to understand the birth conditions of stars to inform the efficacy of chemical tagging of stars. While
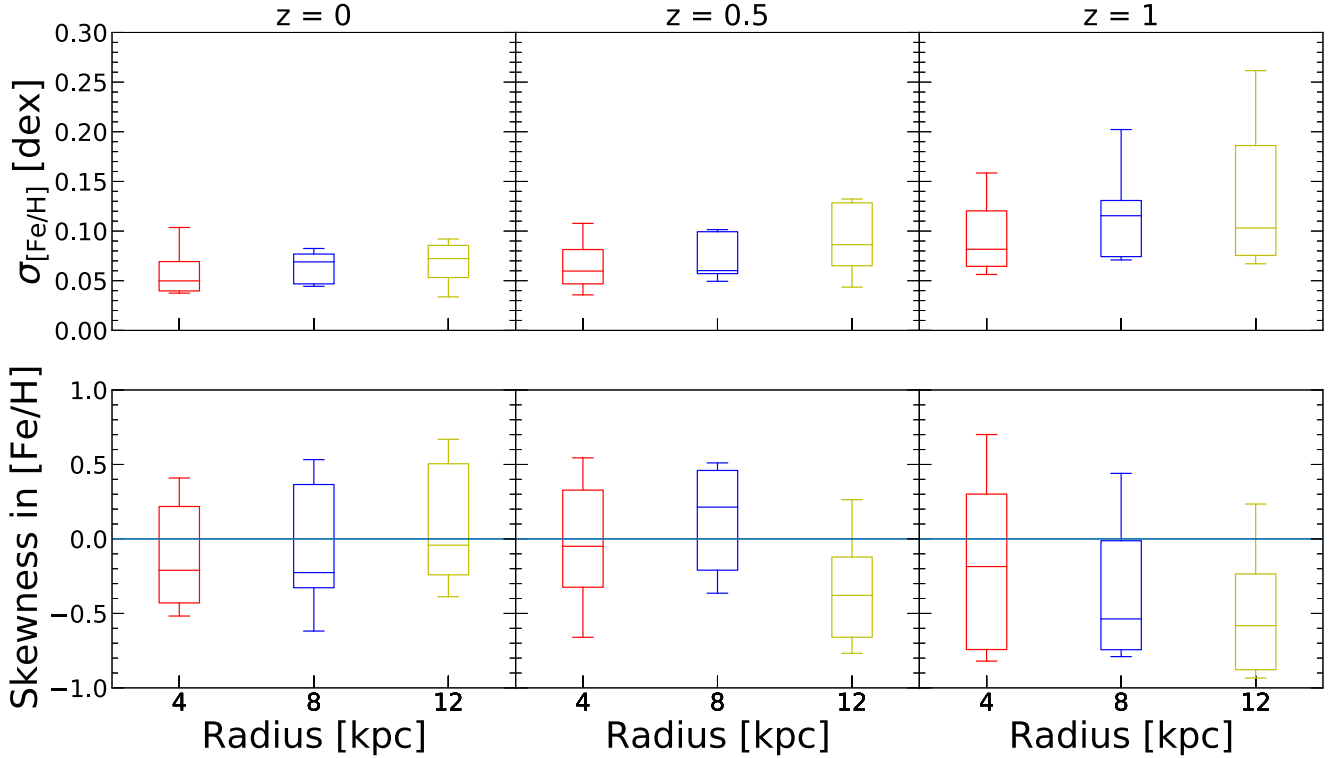
Figure 14. Top panel: standard deviation of the fitted elemental distribution of gas [Fe/H] at 3 radii, for increasing redshift (left- to right-hand panel). For each distribution, we stack three consecutive snapshots ($\Delta t \approx 50$ Myr) and measure all gas within ± 0.2 kpc of each radius and within a height ± 1 kpc of the disc. Each box shows the 68th percentile of the distribution and the whiskers show the full distribution of standard deviations. The standard deviation increases with radius and decreases with time. Bottom panel: skewness of the fitted elemental distributions of gas [Fe/H]. At earlier times, the gas disc had stronger negative skewness, but the disc relaxes to near zero skewness at $z = 0$. The skewness shows a slight radial dependence at both $z = 1$ ($t_{lookback} = 7.8$ Gyr) and $z = 0$. At $z = 1$, the distributions at larger radii were more negatively skewed, whereas at $z = 0$, the distributions at smaller radii are more negatively skewed.

many stars form prior to $z = 1.5$, the last ~ 10 Gyr mark the primary epoch of disc assembly, which is where we are primarily interested in chemically tagging stars. Our main results are as follows:

(i) *Vertical gradients*: are negligible. Abundances in gas are well mixed vertically at all times. At $R = 8$ kpc, the mean deviation in [O/H] at 1 kpc from the galactic mid-plane is < 0.01 dex at all times. The inner ~ 200 pc of the discs, where the majority of star formation for $z < 0.5$ occurs, is approximately uniform in abundance ($|\Delta[O/H]| \lesssim 0.002$ dex) at all times. The inner ~ 1.5 kpc of the discs, where the majority of star formation for $z > 0.5$ occurs has minimal vertical variation in abundance $|\Delta[O/H]| \lesssim 0.01$ dex at all times. Thus, there is minimal vertical information for chemical tagging.

(ii) *Radial gradients*: are negative at all times and for all abundances, with a maximum steepness of ≈ -0.03 dex kpc^{-1} at $z = 0$ and a minimum of ≈ -0.01 dex kpc^{-1} at $z \gtrsim 1$ ($t_{lookback} \gtrsim 7.8$ Gyr). Radial gradients become steeper over time, because the discs become more rotationally supported and are better able to sustain a gradient against radial mixing, as noted in analysis of FIRE-1 simulations in Ma et al. (2017). [N/H] has a steeper gradient at all times, because their production is dominated by stellar winds, whose mass-loss rates increase with metallicity in our simulations, enhancing the discrepancies between metal-rich and metal-poor regions. [O/Fe] shows little variation with redshift and is approximately flat across the disc indicating it provides limited discriminating power for chemically tagging birth radii. Our [O/H] gradients broadly agree with most observations of nearby MW-mass galaxies, including M31,

at the 1 or 2σ level, though our gradients are somewhat steeper on average. By contrast, our gradients are somewhat shallower than most observations of the MW, though they agree at the 1 or 2σ level with 9 of 13 MW observations.

(iii) *Azimuthal scatter*: systematically decreases over time for all abundances, from ≈ 0.2 dex at $z = 1.5$ ($t_{lookback} = 9.4$ Gyr) to ≈ 0.05 dex at $z = 0$ for [O/H] and [Fe/H] around the entire disc at $R = 8$ kpc. This evolution is a result of higher gas accretion and also star formation rates at earlier times, which lead to stronger variations in abundances on small scales, especially at larger radii, where orbital/mixing time-scales are longer. The azimuthal scatter in [N/H] is larger (by ≈ 0.01 dex at $R = 8$ kpc) at all times than in [Fe/H] or [O/H], for the same reasons as above. Azimuthal variations reduce somewhat with smaller azimuthal aperture. However, even in angular bins as small as ≈ 350 pc, they remain ≈ 0.04 dex at $z = 0$ and ≈ 0.1 dex at $z = 1$ ($t_{lookback} = 7.8$ Gyr) for [O/H] and [Fe/H]. We emphasize that our azimuthal bins do not centre on GMCs or star-forming regions, so our results probe the homogeneity of effectively random patches of gas. We find good agreement between our azimuthal scatter in [O/H] in gas at $z = 0$ (≈ 0.05 dex) and observations of nearby galaxies (Sakhibov et al. 2018; K19; Kreckel et al. 2020).

(iv) *Azimuthal versus radial scatter*: At early times, the azimuthal scatter was larger than the radial variation for all abundances. We quantify the redshifts when the radial variation (across $\Delta R = 8$ kpc) first dominates over the azimuthal scatter, finding a median of $z \approx 0.9$ ($t_{lookback} \approx 7.4$ Gyr) at $R = 4$ kpc and $z \approx 0.6$ ($t_{lookback} \approx 5.8$ Gyr) at $R = 12$ kpc. Before this time, stars born at the same radius could

have the same difference in metallicity as stars born $\Delta R \gtrsim 8\,\mathrm{kpc}$ apart. We also quantify across time the radial range over which the radial and azimuthal variations are comparable, $\Delta R_{\mathrm{equality}}$. At $z \sim 0$, $\Delta R_{\mathrm{equality}}$ is small at $\approx 1.8\,\mathrm{kpc}$, but at $z \gtrsim 1$ ($t_{\mathrm{lookback}} \gtrsim 7.8\,\mathrm{Gyr}$), $\Delta R_{\mathrm{equality}}$ is larger than the size of the disc. These results indicate that azimuthal variations in abundances provide the *dominant* information content for chemical tagging for stars formed $\gtrsim 6\,\mathrm{Gyr}$ ago, so future approaches to chemical tagging of stars should start to incorporate/model these significant azimuthal variations.

(v) *Measurable homogeneity*: We quantified the radial scales across which our gas discs are effectively homogeneous in a measurable sense, given representative measurement uncertainties. For an uncertainty in elemental abundance of 0.05 dex, our gas discs are measurably homogeneous across $\Delta R \approx 1.7\,\mathrm{kpc}$ at $z = 0$ and $\Delta R \approx 3.5\,\mathrm{kpc}$ at $z \gtrsim 0.75$ ($t_{\mathrm{lookback}} \gtrsim 6.6\,\mathrm{Gyr}$). Moreover, azimuthal variations at $R \gtrsim 8\,\mathrm{kpc}$ are larger than 0.05 dex at all times. Thus, for any measurement uncertainty at or below this, using chemical tagging to measure birth radius is limited not by measurement uncertainty but instead by azimuthal variations. These results inform the needed precision for observations, given targeted precision for chemical tagging of stars across age/time. For example, if one only cares about modelling birth radius, there is little-to-no benefit in measuring a stellar abundance to better than $\approx 0.05\,\mathrm{dex}$.

(vi) *Elemental abundance distributions*: We measured the full distributions of elemental abundances in radial annuli and fit them with skew normal distributions. The skew normal distributions fit these distributions reasonably well, but there are failure modes that become more common at higher redshift, most notably underfitting the negative tails of the distribution and simultaneously underfitting the positive and negative tails. We find typically negatively skewed normal distributions at $z \gtrsim 1$ ($t_{\mathrm{lookback}} \gtrsim 7.8\,\mathrm{Gyr}$), with stronger negative skewness at larger radii. The distributions evolve towards approximately Gaussian distributions at all radii by $z = 0$.

4.2 Discussion

The primary goal of this paper is to understand the homogeneity of gas as a proxy for the birth conditions of stars across space and time, as a first step to understanding the efficacy of chemical tagging in a cosmological context. There are caveats to our analysis, though. Namely, our analysis is performed looking at individual elements, with the exception of [O/Fe]. Examining multi-element abundance distributions may well offer more discerning power. Our analysis of [O/Fe] suggests that this may be limited. Furthermore, uncertainty in our fiducial diffusion coefficient leads to uncertainty in our small-scale azimuthal abundance scatter, as seen in Appendix. C.

Additional complications arise when comparing our simulations to observational data. We present results in the context of constraining chemical tagging in the MW, but our simulations are not exact MW-analogues. Also, when comparing the redshift evolution of our results to observations of external galaxies, we track the evolutionary history of individual galaxies across time, as opposed to measuring properties of different galaxies at fixed mass across time. For all of our comparisons to observations, we explore all gas whereas observers typically measure abundances in H II regions specifically. However, Hernandez et al. (2021) compared observations of ionized and neutral gas-phase abundance gradients in M83, finding gradients for neutral gas to be $-0.17\,\mathrm{dex\,kpc^{-1}}$ and gradients for ionized gas to be $-0.03\,\mathrm{dex\,kpc^{-1}}$. This might imply that our measured gradients are much flatter than one would expect, given observations. Hernandez et al. (2021) did the same analysis excluding the nuclear region of M83 and found the neutral gas to be in much better agreement

with ionized gas (a gradient of $-0.02\,\mathrm{dex\,kpc^{-1}}$). Additionally, we compare to observations which have measurements in broadly similar physical regions to those we analyse in the simulations, but they are not exactly the same.

We compared against observations of radial gradients in the MW, M31, and similar-mass galaxies at $z = 0$, finding broad agreement. We also connect our evolutionary trends with high-redshift observations of gas-phase abundances. In particular, we find that our MW/M31-mass galaxies all have negative radial gradients at $z \sim 0$ but had nearly flat radial gradients at $z \gtrsim 1$ (where the average stellar mass of the hosts is $M_{90} \approx 1.74 \times 10^{10}\,\mathrm{M_\odot}$). This trend agrees well with many observations of comparable mass galaxies at these higher redshifts (e.g. Queyrel et al. 2012; Stott et al. 2014; Wuyts et al. 2016; Patrício et al. 2019; Curti et al. 2020). However, some observational works have found strong negative radial gradients at these masses and redshifts (Wuyts et al. 2016; Carton et al. 2018; Wang et al. 2019a). Furthermore, while less common than negative radial gradients, some observations find some positive radial gradients at these redshifts as well (Queyrel et al. 2012; Wuyts et al. 2016; Carton et al. 2018; Wang et al. 2019a), which we do not find in any of our galaxies. In general, we find that the steepening of radial gradients with time in our simulations is consistent with observational results and follows the quantitative trends in other theoretical analyses, both in simulations (e.g Ma et al. 2017) and in metallicity-evolution modelling (Sharda et al. 2021).

One of the most important aspects of our analysis is quantifying azimuthal variations in gas abundances and comparing their strength relative to radial gradients across cosmic time. With the advent of integral field spectroscopy, observations have begun characterizing 2D abundance distributions in nearby galaxies. These works (Sánchez et al. 2015; Vogt et al. 2017; Ho et al. 2017, 2018) all find non-trivial azimuthal variations in nearby galaxies, for example, K19 found variations of $\approx 0.05\,\mathrm{dex}$ at fixed radius, which agrees well with our results. However, some observations (e.g. Zinchenko et al. 2016) found no evidence for large-scale azimuthal variation in nearby galaxies. One of our key results/predictions is the evolution of azimuthal variations, which we predict were stronger at higher redshifts. Observations of gravitationally lensed systems now allow sub-kpc measurements at high redshift (Jones et al. 2013, 2015), making it possible to test this predicted evolution in more detail.

Kreckel et al. (2020) examined azimuthal variations in gas-phase [O/H] across eight nearby galaxies using PHANGS-MUSE optical integral field spectroscopy. While our technique for measuring azimuthal variations are not exactly comparable to their methods, we find similar results. In our analysis, we focus on scatter in all gas by measuring a mean scatter in angular bins of varying size, so we in effect measure the azimuthal inhomogeneities of random patches of gas at a given radius. In contrast to this, Kreckel et al. (2020) measure abundances specifically in H II regions and determine scatter by first subtracting off the radial gradient and then centering apertures of various sizes on individual H II regions and measuring the scatter between the H II regions contained within the aperture. They find a slight scale dependence associated with the scatter, which we also see at $z = 0$, with the scatter on scales larger than $\approx 3\,\mathrm{kpc}$ being $\approx 0.05\,\mathrm{dex}$. The small-scale scatter in Kreckel et al. (2020) ($\approx 0.02\,\mathrm{dex}$) is slightly smaller than the $z = 0$ scatter that we observe, but this could be attributed to the discrepancy in our methods. H II regions are likely better mixed in abundances than random patches of gas, so our analysis may be artificially inflating the typical azimuthal scatter of the gas from which stars are forming. However, centering on H II regions is beyond the scope of our analysis, and in future

M. A. Bellardini et al.

work, we will examine azimuthal variations in newly formed stars, which may be closer to the values in H II regions.

Krumholz & Ting ([2018](#)) derived the expected correlation function of metal distribution in galaxies across space and time using a stochastic diffusion model. While we did not explore the correlation function of metals, we did examine homogeneity as a function of azimuthal scale, which we can compare broadly with their work. They found that gas-phase abundances produced primarily through core-collapse supernovae, in MW-like conditions near the solar circle, are correlated on scales of $\approx 0.5 - 1$ kpc giving an expected scatter of $0.04 - 0.1$ dex. From fully cosmological simulations, our results for azimuthal scatter on scales of ≈ 1 kpc near the solar cylinder agree well with their predicted range.

All of our results agree with Ma et al. ([2017](#)), who analysed radial gradients of abundances in the FIRE-1 simulations across a much wider galaxy mass range. In comparing with other theoretical/simulation works, our gradients in [O/H] at $z = 0$ fall between the gradients Hemler et al. ([2020](#)) measured in the TNG50 simulations (≈ -0.02 dex kpc^{-1}) and the gradients Gibson et al. ([2013](#)) measured in the MaGICC and MUGS simulations (≈ -0.04 dex kpc^{-1}). In particular, Hemler et al. ([2020](#)) found a gradual flattening of the gradients with time, which could come from an 'inside-out' growth of galaxies wherein star formation, hence elemental enrichment, proceeds from the inner galaxy to the outer galaxy (e.g. Prantzos & Boissier [2000](#); Bird et al. [2013](#)). The flat(ter) radial gradients at earlier times in our galaxies result from higher turbulence and outflows that frequently eject much of the ISM at those times, perturbations such as mergers and rapid gas infall result in the velocity dispersion of gas particles dominating over their rotational velocity leading to galaxy-scale radial mixing (Ma et al. [2017](#)). As the disc settles over time, it becomes more rotationally supported, so stronger radial gradients can develop/persist. Our results qualitatively agree with those of the EAGLE simulations (Tissera et al. [2019](#)), though as with TNG50, Tissera et al. ([2019](#)) found [O/H] gradients that are slightly shallower than ours, ≈ -0.011 dex kpc^{-1} at $z = 0$.

The evolution of our gas-phase abundance gradients disagrees with Agertz et al. ([2021](#)), who analysed the VINTERGATAN simulation of the m12i initial conditions, performed using the adaptive mesh refinement (AMR) code RAMSES. They found that the gas-phase profile of [Fe/H] becomes shallower over time (their fig. 7), compared with our steepening with time. One possible explanation is the difference in hydrodynamic solvers: we use the mesh-free finite-mass (MFM) quasi-Lagrangian method in GIZMO, coupled with explicit modelling of sub-grid mixing, while the AMR simulation of VINTERGATAN induces significantly more mixing in gas (complete mixing on the scale of an individual cell), which may contribute to the flattening of their gradient over time. However, as shown in Appendix. C, the qualitative steepening of the gradient we observe with time is independent of the strength of our diffusion coefficient.

Galactic evolution models often simplify the abundance distributions of gas in galaxies to a 1D model (e.g. Minchev et al. [2018](#); Mollá et al. [2019a](#); Frankel et al. [2020](#)), with azimuthal scatter assumed from measurements at $z = 0$. While this is a useful first step in understanding the abundance evolution of galaxies and testing chemical tagging, our results mean that this simplifying assumption overestimates the radial information content in elemental abundances, including how well chemical tagging can constrain the birth radius of a star. On the one hand, the non-trivial azimuthal scatter that we find, especially at earlier times, complicates modelling the abundances of stars at a given radius. On the other hand, this likely makes individual GMCs more elementally distinct at fixed radius, providing greater discriminating power, as we will explore

in future work. However, we do not explore the homogeneity of individual GMCs in this work. Recent work has started to pursue 2D abundance evolution models (e.g. Mollá et al. [2019b](#)) which may address this question, works such as Spitoni et al. ([2019](#)) find azimuthal abundance variations on the order of 0.1 dex, twice what we find at $z = 0$ in our simulation suite.

Related to our analysis of a transition epoch, after which radial variations dominate over azimuthal scatter as the disc settles, Yu et al. ([2021](#)) examine the transition epoch from 'bursty' to 'steady' star formation and disc settling in the same simulations. We checked that their measurement of this bursty/steady transition agrees moderately well with the transition epoch that we present here, at least at smaller radius (4 kpc). We find weaker agreement for our transition times at larger radii (8 and 12 kpc). Furthermore, our transition times are consistently earlier (~ 3 Gyr at $R = 4$ kpc and ~ 1 Gyr at $R = 12$ kpc) than those in Yu et al. ([2021](#)), with the transition times on average being more similar at larger radii. Thus, we find a broad correlation between the transition from bursty to smooth star formation and the transition from azimuthal to radial abundance variations, but with significant scatter and some time delay.

Our simulations show the importance of considering azimuthal variations in addition to radial variations when studying gas-phase elemental abundance distributions. This is particularly important in the context of chemical tagging; in order to accurately identify the birth locations of stars using elemental abundances, the initial conditions of stars need to be well defined. As we showed in Section. 3.5, azimuthal variations in abundance are greater than or comparable to radial variations at earlier times, so chemical tagging models that only account for radial variations will fail to accurately capture the scatter in abundances at a given radius. This could lead to incorrectly assigning stars as co-natal, or vice versa. We also fit the elemental distributions of our galaxies at different radii at different times, finding that they shift from negative to zero skewness over time. While skew-normal fits are not a perfect fit for the elemental distributions of our galaxies at all redshifts, they more accurately represent the distributions than a Gaussian. Thus, using our measured distributions would be a useful step in building more accurate abundance evolution models including for chemical tagging.

Next generation telescopes are crucial for testing the predictions of gas-phase abundance homogeneity presented in this work, particularly the predictions for azimuthal scatter at high redshifts. Current measurements of azimuthal scatter in abundances have been restricted to nearby galaxies. However, with the advent of JWST NIRSPec IFU and next-generation adaptive-optic IFUs on telescopes like IRIS and TMT, spatially resolved measurements of metallicities in distant galaxies are feasible, providing tests of our predictions for azimuthal scatter and the transition redshifts when it becomes sub-dominant.

This work is the first step of testing the limits of chemical tagging in the FIRE simulations. In the future, we will examine the degree to which these results for gas are mirrored in newly formed stars across time. Combining those results with measurements of the dynamical evolution of stars in our simulations, we more directly will test the efficacy of chemical tagging of stars in the FIRE simulations.

ACKNOWLEDGEMENTS

We thank Francesco Belfiore for providing abundance gradient measurements for galaxies from the MaNGA survey. We also thank Tucker Jones, Ryan Sanders, Joss Bland-Hawthorn, Trey Wenger, and Dana Balser for useful discussions that improved this

Downloaded from https://academic.oup.com/mnras/article/505/3/4586/6293871 by Galter Health Sciences Library user on 22 July 2021

MNRAS **505**, 4586–4607 (2021)

DATA AVAILABILITY

Full simulation snapshots at $z = 0$ are available for m12i, m12f, and m12m at ananke.hub.yt. A public version of the GIZMO code is available at http://www.tapir.caltech.edu/phopkins/Site/GIZMO.html. Additional data including simulation snapshots, initial conditions, and derived data products are available at https://fire.northwestern.edu/data/. Some of the PYTHON code used to analyse these data includes the publicly available packages https://bitbucket.org/awetzel/gizmo_analysis (Wetzel & Garrison-Kimmel 2020) and https://bitbucket.org/awetzel/utilities.

REFERENCES

Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003, ApJ, 597, 21
Acharova I., Gibson B. K., Mishurov Y. N., Kovtyukh V., 2013, A&A, 557, A107
Agertz O. et al., 2021, MNRAS, 503, 5826
Anders F. et al., 2014, A&A, 564, A115
Anders F. et al., 2017, A&A, 600, A70
Andrews B. H., Martini P., 2013, ApJ, 765, 140
Anglés-Alcázar D., Faucher-Giguère C.-A., Kereš D., Hopkins P. F., Quataert E., Murray N., 2017, MNRAS, 470, 4698
Arellano-Córdova K. Z., Esteban C., García-Rojas J., Méndez-Delgado J. E., 2020, MNRAS, 496, 1051
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Astropy Collaboration et al., 2013, A&A, 558, A33
Baba J., Morokuma-Matsui K., Miyamoto Y., Egusa F., Kuno N., 2016, MNRAS, 460, 2472
Balser D. S., Rood R. T., Bania T. M., Anderson L. D., 2011, ApJ, 738, 27
Balser D. S., Wenger T. V., Anderson L. D., Bania T. M., 2015, ApJ, 806, 199
Belfiore F. et al., 2017, MNRAS, 469, 151
Benincasa S. M. et al., 2020, MNRAS, 497, 3993
Bird J. C., Kazantzidis S., Weinberg D. H., Guedes J., Callegari S., Mayer L., Madau P., 2013, ApJ, 773, 43
Bird J. C., Loebman S. R., Weinberg D. H., Brooks A. M., Quinn T. R., Christensen C. R., 2021, MNRAS, 503, 1815
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Bland-Hawthorn J., Krumholz M. R., Freeman K., 2010, ApJ, 713, 166
Boardman N. et al., 2020, MNRAS, 491, 3672
Boeche C. et al., 2013, A&A, 559, A59
Boeche C. et al., 2014, A&A, 568, A71
Bovy J., 2016, ApJ, 817, 49
Brook C. B., Kawata D., Gibson B. K., Freeman K. C., 2004, ApJ, 612, 894
Carrell K., Chen Y., Zhao G., 2012, AJ, 144, 185
Carton D. et al., 2018, MNRAS, 478, 4293
Cedrés B., Cepa J., 2002, A&A, 391, 809
Cheng J. Y. et al., 2012, ApJ, 746, 149
Colbrook M. J., Ma X., Hopkins P. F., Squire J., 2017, MNRAS, 467, 2421
Cui X.-Q. et al., 2012, Res. Astron. Astrophys., 12, 1197
Curti M. et al., 2020, MNRAS, 492, 821
Dalton G. et al., 2012, Proc. SPIE Conf. Ser. Vol. 8446, Ground-Based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P
Davies B., Origlia L., Kudritzki R.-P., Figer D. F., Rich R. M., Najarro F., Negueruela I., Clark J. S., 2009, ApJ, 696, 2014
de Jong R. S. et al., 2019, The Messenger, 175, 3
De Silva G. M., Freeman K. C., Bland-Hawthorn J., Asplund M., Bessell M. S., 2007, AJ, 133, 694
De Silva G. M. et al., 2015, MNRAS, 449, 2604
Escala I. et al., 2018, MNRAS, 474, 2194
Esteban C., García-Rojas J., 2018, MNRAS, 478, 2315
Esteban C., Carigi L., Copetti M. V. F., García-Rojas J., Mesa-Delgado A., Castañeda H. O., Péquignot D., 2013, MNRAS, 433, 382
Esteban C., Fang X., García-Rojas J., Toribio San Cipriano L., 2017, MNRAS, 471, 987
Faucher-Giguere C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416
Fernández-Martín A., Pérez-Montero E., Vílchez J. M., Mampaso A., 2017, A&A, 597, A84
Frankel N., Rix H.-W., Ting Y.-S., Ness M., Hogg D. W., 2018, ApJ, 865, 96
Frankel N., Sanders J., Ting Y.-S., Rix H.-W., 2020, ApJ, 896, 15
Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487
Garrison-Kimmel S., Boylan-Kolchin M., Bullock J. S., Lee K., 2014, MNRAS, 438, 2578
Garrison-Kimmel S. et al., 2019, MNRAS, 487, 1380
Genovali K. et al., 2014, A&A, 566, A37
Gibson B. K., Pilkington K., Brook C. B., Stinson G. S., Bailin J., 2013, A&A, 554, 47
Gilmore G. et al., 2012, The Messenger, 147, 25
Grand R. J. J., Kawata D., Cropper M., 2015, MNRAS, 447, 4018
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Hayden M. R. et al., 2014, AJ, 147, 116
Hemler Z. S. et al., 2020, MNRAS, preprint (arXiv:2007.10993)
Hernandez S. et al., 2021, ApJ, 908, 226
Ho I.-T. et al., 2017, ApJ, 846, 39
Ho I.-T. et al., 2018, A&A, 615, A64
Hopkins P. F., 2015, MNRAS, 450, 53
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F.-K., 1999, ApJS, 125, 439
Izzard R. G., Tout C. A., Karakas A. I., Pols O. R., 2004, MNRAS, 350, 407
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
Kawata D., Hunt J. A. S., Grand R. J. J., Pasetto S., Cropper M., 2014, MNRAS, 443, 2757
Keller B. W., Wadsley J. W., Wang L., Kruijssen J. M. D., 2019, MNRAS, 482, 2244
Kollmeier J. A. et al., 2017, preprint (arXiv:1711.03234)
Kreckel K. et al., 2019, ApJ, 887, 80
Kreckel K. et al., 2020, MNRAS, 499, 193
Kroupa P., 2001, MNRAS, 322, 231
Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36
Krumholz M. R., Ting Y.-S., 2018, MNRAS, 475, 2236
Leitherer C. et al., 1999, ApJS, 123, 3
Lemasle B., François P., Piersimoni A., Pedicelli S., Bono G., Laney C. D., Primas F., Romaniello M., 2008, A&A, 490, 613
Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, A&A, 500, 145

Limongi M., Chieffi A., 2018, ApJS, 237, 13

Loebman S. R., Roškar R., Debattista V. P., Ivezić Ž., Quinn T. R., Wadsley J., 2011, ApJ, 737, 8

Luck R. E., Lambert D. L., 2011, AJ, 142, 136

Luck R. E., Kovtyukh V. V., Andrievsky S. M., 2006, AJ, 132, 902

Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1

Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140

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

Majewski S. R. et al., 2017, AJ, 154, 94

Mannucci F., Della Valle M., Panagia N., 2006, MNRAS, 370, 773

Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115

Maoz D., Graur O., 2017, ApJ, 848, 25

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

Mikolaitis Š. et al., 2014, A&A, 572, A33

Minchev I. et al., 2018, MNRAS, 481, 1645

Mollá M. et al., 2019a, MNRAS, 482, 3071

Mollá M. et al., 2019b, MNRAS, 490, 665

Muley D. A., Wheeler C. R., Hopkins P. F., Wetzel A., Emerick A., Keres D., 2020, preprint (arXiv:2008.04901)

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

Nieva M. F., Przybilla N., 2012, A&A, 539, A143

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

Nomoto K., Kobayashi C., Tominaga N., 2013, ARA&A, 51, 457

Ostdiek B. et al., 2020, A&A, 636, A75

Patrício V. et al., 2019, MNRAS, 489, 224

Pedicelli S. et al., 2009, A&A, 504, 81

Pignatari M. et al., 2016, ApJS, 225, 24

Pilyugin L., Grebel E., Kniazev A., 2014, AJ, 147, 131

Planck Collaboration et al., 2020, A&A, 641, A6

Poetrodjojo H. et al., 2018, MNRAS, 479, 5235

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

Price-Jones N. et al., 2020, MNRAS, 496, 5101

Price-Whelan A. M. et al., 2018, AJ, 156, 123

Queyrel J. et al., 2012, A&A, 539, 93

Rennehan D., Babul A., Hopkins P. F., Davé R., Moa B., 2019, MNRAS, 483, 3810

Rudolph A. L., Fich M., Bell G. R., Norsen T., Simpson J. P., Haas M. R., Erickson E. F., 2006, ApJS, 162, 346

Sakhibov F., Zinchenko I. A., Pilyugin L. S., Grebel E. K., Just A., Vílchez J. M., 2018, MNRAS, 474, 1657

Sánchez S. F. et al., 2015, A&A, 573, 105

Sánchez-Menguiano L. et al., 2016, A&A, 587, A70 (S16)

Sánchez-Menguiano L., Sánchez S. F., Pérez I., Ruiz-Lara T., Galbany L., Anderson J. P., Kuncarayakti H., 2020, MNRAS, 492, 4149

Sanders N. E., Caldwell N., McDowell J., Harding P., 2012, ApJ, 758, 133

Santistevan I. B., Wetzel A., El-Badry K., Bland-Hawthorn J., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., Benincasa S., 2020, MNRAS, 497, 747

Schönrich R., Binney J., 2009, MNRAS, 396, 203

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

Solar M., Tissera P. B., Hernandez-Jimenez J. A., 2020, MNRAS, 491, 4894

Spitoni E., Cescutti G., Minchev I., Matteucci F., Silva Aguirre V., Martig M., Bono G., Chiappini C., 2019, A&A, 628, A38

Steinmetz M. et al., 2006, AJ, 132, 1645

Stott J. P. et al., 2014, MNRAS, 443, 2695

Sukhbold T., Ertl T., Woosley S. E., Brown J. M., Janka H. T., 2016, ApJ, 821, 38

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

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

Takada M. et al., 2014, PASJ, 66, R1

The MSE Science Team et al., 2019, preprint (arXiv:1904.04907)

Ting Y.-S., De Silva G. M., Freeman K. C., Parker S. J., 2012, MNRAS, 427, 882

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

Tremonti C. A. et al., 2004, ApJ, 613, 898

van den Hoek L. B., Groenewegen M. A. T., 1997, A&AS, 123, 305

Vogt F. P. A., Pérez E., Dopita M. A., Verdes-Montenegro L., Borthakur S., 2017, A&A, 601, 61

Wang X. et al., 2019a, ApJ, 900, 18

Wang C. et al., 2019b, MNRAS, 482, 2189

Wenger T. V., Balser D. S., Anderson L. D., Bania T. M., 2019, ApJ, 887, 114

Wetzel A., Garrison-Kimmel S., 2020, Astrophysics Source Code Library, record ascl:2002.015

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23

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

Wojno J. et al., 2016, MNRAS, 461, 4246

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

Yu S. et al., 2021, MNRAS, 505, 889

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

Zinchenko I. A., Pilyugin L. S., Grebel E. K., Sánchez S. F., Vílchez J. M., 2016, MNRAS, 462, 2715

Zinchenko I., Just A., Pilyugin L., Lara-Lopez M., 2019, A&A, 623, A7

Zurita A., Bresolin F., 2012, MNRAS, 427, 1463

APPENDIX A: SCALED RADIAL PROFILES

Fig. A1 compares the host-to-host scatter in radial gradients of [O/H] in gas in our simulated galaxies when scaling these gradients to various galaxy scale radii at $z = 0$. We scale each galaxy's profile
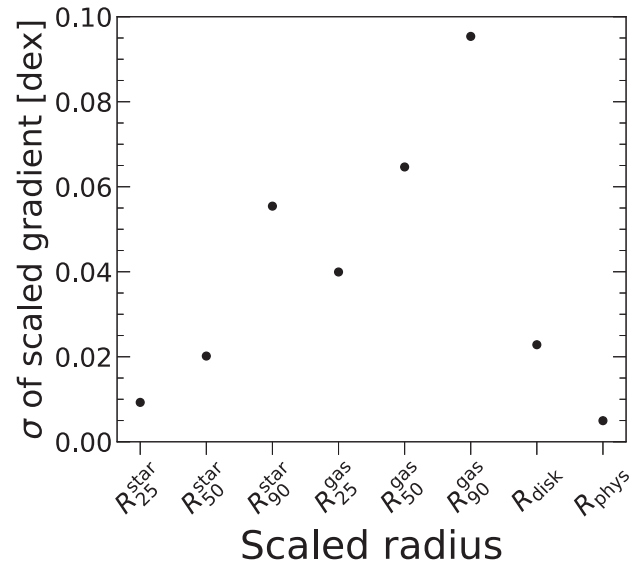

Figure A1. The 1σ host-to-host scatter of the scaled radial gradients for different scale radii at $z = 0$. We define R_{25}, R_{50}, and R_{90} for the gas and stars in Table 1. $R_{\rm disc}$ is the exponential scale length of the stellar disc determined via a two-component fit to the surface density, and $R_{\rm phys}$ is the physical radial coordinates of the disc, i.e. unscaled coordinates. For each scale radius, we measure the gradient of all galaxies across an equal radial range that corresponds to $4 - 12\,{\rm kpc}$ physical for the galaxy with the median scale length. The 1σ host-to-host scatter is smallest when measuring the gradients in physical space, which motivates our choice for our analysis in this paper.

using: R_{25}, R_{50}, and R_{90} for the gas and the stars, along with the exponential scale length, R_{disc}, from a two-component (sérsic plus exponential) fit to the surface density. Table 1 lists the values for each galaxy. We also compare these scaled gradients to the gradients in physical radii (as measured in Section 3.2). We bin each profile equally in scaled radius, defining the bin width such that the galaxy with the median scale length has a binwidth of 0.25 kpc physical. We measure the radial gradient of each galaxy across an equal radial range for each scale radius. We define this radial range such that we measure the galaxy with the median scale length across a physical range $4 - 12$ kpc. This range corresponds to: $\approx 3.3 - 9.8 R_{25}^{\text{star}}$, $\approx 1.4 - 4.1 R_{50}^{\text{star}}$, $\approx 0.4 - 1.1 R_{90}^{\text{star}}$, $\approx 0.5 - 1.6 R_{25}^{\text{gas}}$, $\approx 0.3 - 1.0 R_{50}^{\text{gas}}$, $\approx 0.2 - 0.7 R_{90}^{\text{gas}}$, and $\approx 1.0 - 3.1 R_{\text{disc}}$.

Measuring the gas abundance radial gradient (from $4 - 12$ kpc) in physical space minimizes the host-to-host scatter, to $\sigma \approx 0.005$ dex. R_{25}^{star} has the next smallest 1σ scatter with $\sigma \approx 0.009$ dex. The gradients are the least self-similar when scaled by R_{90}^{gas}, for which the 1σ scatter is ≈ 0.1. The self-similarity of the radial profiles in physical space motivates our choice in this paper, because there is no compelling reason to scale the profiles of our galaxies. We emphasize, though, that this may be a result of the small mass range of our suite (halo masses are $M_{\text{200m}} = 1 - 3 \times 10^{12} \, \text{M}_\odot$, stellar masses are in Table 1) and may not be generalizable to galaxies across a wider mass range.

APPENDIX B: ALL GAS VERSUS STAR-FORMING GAS

In this paper, we examine elemental abundances in all gas, as initial conditions for chemical tagging of stars. We choose to measure all gas, in part, because star-forming gas represents only a small fraction of all gas elements at a given snapshot, leading to significant Poisson noise. In principal, we could attempt to identify photoionized (H II) regions near young star particles to compare with gas-phase measurements via nebular emission lines, but doing this correctly requires generating synthetic observations via ray-tracing, which is beyond the scope of our analysis. In future work (Bellardini et al., in preparation), we will compare in detail the spatial variations in abundance of star particles that form out of this gas to the gas itself. Here, we explore the impact of measuring only star-forming gas instead of all gas.

Fig. B1 compares measuring [O/H] in star forming versus all gas at $z = 1$ ($t_{\text{lookback}} = 7.8$ Gyr) and $z = 0$. For each galaxy, we select gas elements at $4 < R < 12$ kpc and $|Z| < 1$ kpc, and we stack this measurement across 10 snapshots (≈ 200 Myr), because at any single snapshot there are few star-forming gas elements. For reference, for these same simulations at $z = 0$, Benincasa et al. (2020) find typical GMC lifetimes, and hence lifetimes of star-forming regions, of $5 - 7$ Myr. We first measure the difference in the average abundance
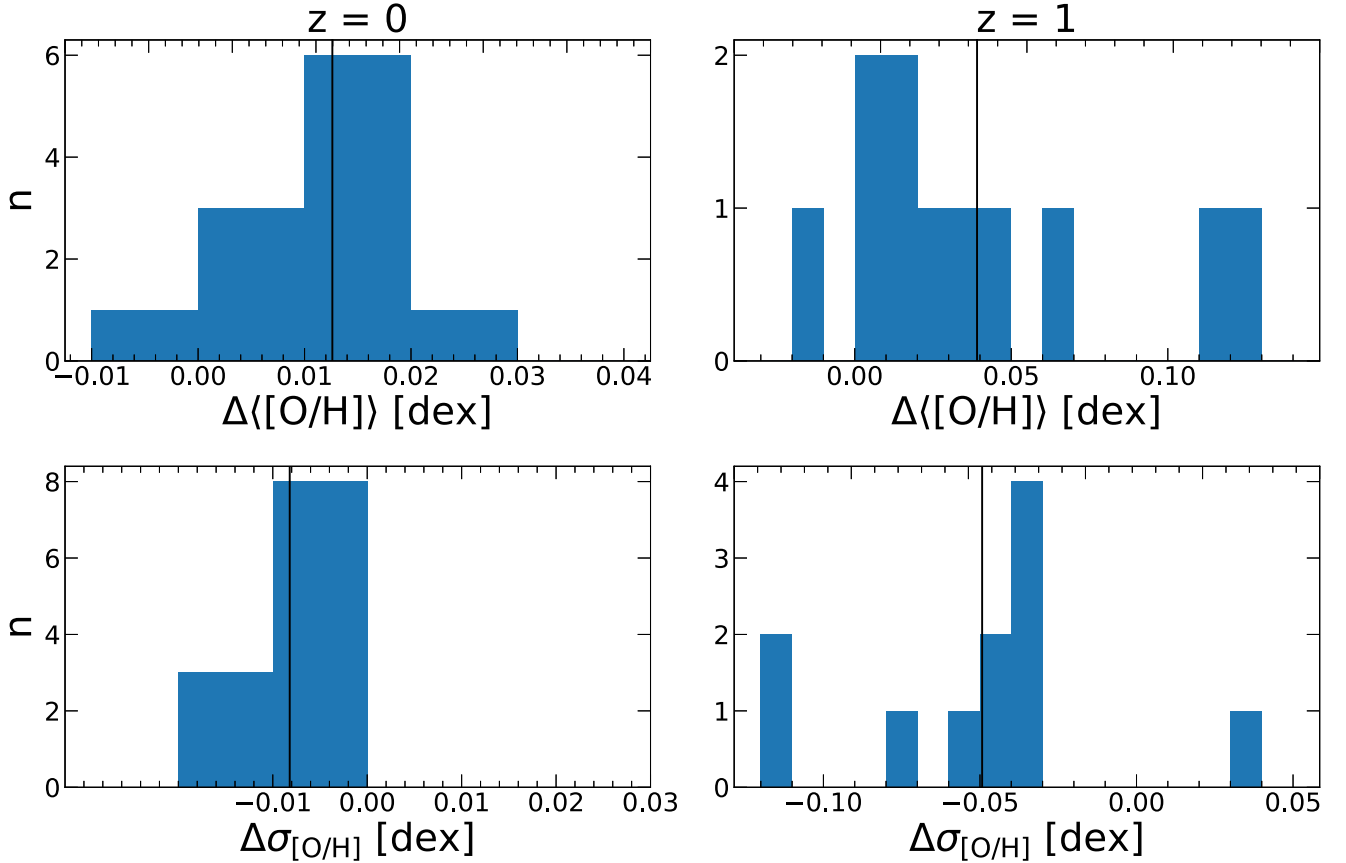

Figure B1. A comparison of [O/H] measured in all gas versus only star-forming gas, at $4 < R < 12$ kpc and $|Z| < 1$ kpc. For each host, we measure the mean and standard deviation of its [O/H] (stacking 10 consecutive snapshot across ≈ 200 Myr to boost the number of star-forming gas elements), and we compute the galaxy-wide difference between star-forming gas and all gas. Top panels show histograms of the difference in the mean [O/H], while bottom panels show histograms of the difference in the standard deviation of [O/H]. Left-hand panels show $z = 0$ and right-hand panels show $z = 1$ ($t_{\text{lookback}} = 7.8$ Gyr). The solid vertical lines show the mean of each difference. Star-forming gas is on average more metal rich than all gas by ≈ 0.04 dex at $z = 1$ and ≈ 0.01 dex at $z = 0$. Furthermore, star-forming gas is slightly better mixed (with less scatter), with $\sigma_{\text{[O/H]}} \approx 0.05$ dex smaller at $z = 1$ and ≈ 0.008 dex smaller at $z = 0$.

between star forming and all gas for each galaxy. Fig. B1 (top row) shows a histogram of this offset in the mean [O/H]. A positive value means star-forming gas has a higher [O/H] than all gas for that galaxy. The black vertical line shows the mean value of the histogram. On average, star-forming gas has modestly higher [O/H] than all gas by ≈ 0.04 dex at $z = 1$ and ≈ 0.01 dex at $z = 0$. The difference in [O/H] is typically $\lesssim 0.02$ dex for $z = 0$ and always less than 0.03 dex. The discrepancy is larger at higher redshift, the difference is typically $\lesssim 0.04$ dex and always less than 0.13 dex. This is likely because cosmic accretion and star formation rates are higher at earlier times, leading to less efficient small-scale mixing of gas. Of course, a simple offset in the [O/H] normalization does not alone mean that spatial variations are different.

Fig. B1 (bottom row) shows the difference in the standard deviation of star forming versus all gas. Again, the black line shows the mean value. On average, [O/H] for star-forming gas has slightly smaller standard deviation than for all gas. This difference is larger at $z = 1$ than at $z = 0$. However, the difference is typically small, $\lesssim 0.05$ dex. This suggests that the azimuthal variations of star-forming gas may be smaller than that of all gas, especially if the scatter is driven primarily by radial variations in abundance. Thus, chemically tagging the birth radii of stars may be complicated by azimuthal variations for redshifts higher than we show in Section 3.5, which we will explore further in Bellardini et al., in prep.

We also explore the differences in the radial gradients for star-forming versus all gas (not shown). At $z = 0$, the radial gradients of star forming and all gas are consistent to within ± 0.005 dex kpc^{-1}, less than the host-to-host scatter in Fig. 5. At $z = 1$, for the galaxies with sufficient star-forming gas to measure a reliable radial gradient, they agree with the gradients for all gas to within ± 0.002 dex kpc^{-1}. We also compare compare the radial profiles of [O/H] for newly formed stars (in age bins of 200 Myr) with that of all gas at the same snapshots. For $z \lesssim 0.5$, these profiles overlap to within uncertainty. For $z \gtrsim 0.5$, the stellar and gas abundance profiles start to diverge, such that the young stars tend to have higher [O/H] and flatter gradients than all gas, which we will explore further in Bellardini et al., in prep.

In summary, while analysing all gas is a reasonable, if not perfect, proxy for star-forming gas in our simulations. Given the short

lifetimes of star-forming gas clouds (Benincasa et al. 2020) and the strict conditions for particles to be star forming (Hopkins et al. 2018), only a small fraction of gas elements are star forming at a given snapshot ($\lesssim 1$ per cent of gas particles at the redshifts we observe), so analysing all gas greatly reduces the statistical uncertainty. In the future, we will study the abundances of star particles that form from this gas, to compare with these results in detail.

APPENDIX C: IMPACT OF DIFFUSION COEFFICIENT

Our FIRE-2 simulations model the sub-grid diffusion/mixing of metals in gas via unresolved turbulent eddies (Su et al. 2017; Escala et al. 2018; Hopkins et al. 2018)

$$\frac{\partial Z_i}{\partial t} + \nabla \cdot (D \nabla Z_i) = 0, \qquad (C1)$$

where Z_i is the mass fraction of a metal in gas element i, and D is the diffusion coefficient. While there is some uncertainty in the exact value to choose for this coefficient, our fiducial value is physically motivated based on tests of the metal diffusion implementation in FIRE-2 on idealized, converged turbulent box simulations by Colbrook et al. (2017) and other more extensive studies by Rennehan et al. (2019). Here, we compare our key results using our fiducial diffusion coefficient D in m12i against a re-simulations of m12i with all identical physics/parameters, except one has a diffusion coefficient that is 10 times higher (that is, faster mixing) and the other includes no subgrid mixing.

Fig. C1 compares the vertical, radial, and azimuthal variations for m12i. The left-hand panel shows the vertical gradient in gas [O/H], similar to Fig. 6. At $z = 0$, we find no differences within 200 pc and at most ~ 0.015 dex difference at 1 kpc. The differences are stronger at $z = 1$ for 10 times higher diffusion and stronger at $z = 0$ for the simulation with no diffusion, though again, not at a significant level to change our interpretations, especially within 200 pc.

Fig. C1 (centre) shows the radial profile in gas [O/H], normalized to the abundance at $R = 4$ kpc (given a strong upturn at smaller R). The radial gradients, measured over our fiducial radial ranges, vary by $\lesssim 0.005$ dex kpc^{-1} between the 10 times diffusion simulation
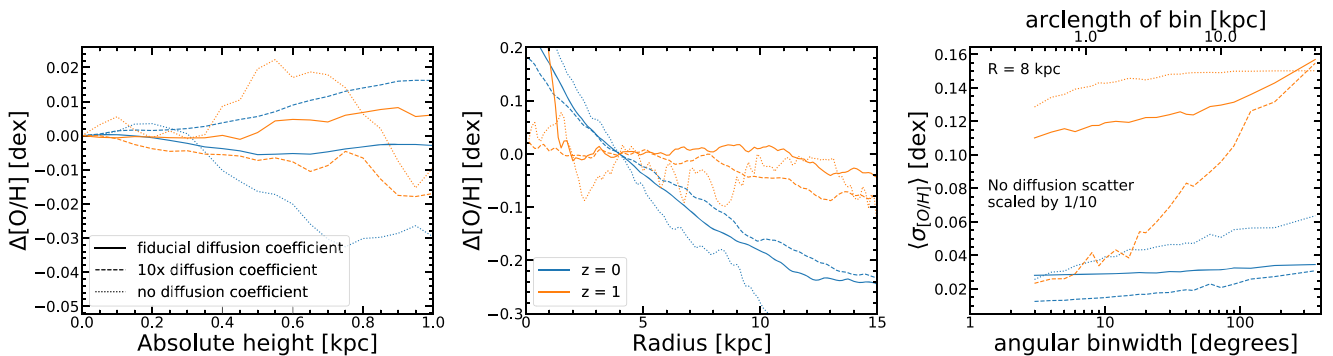

Figure C1. Vertical (left-hand panel), radial (middle), and azimuthal variations (right-hand panel) in [O/H] between our fiducial simulation of m12i, a version with no subgrid metal diffusion, and a re-simulation increasing the diffusion coefficient by 10 times. The vertical profiles show no clear systematic variations at a level important for our analysis. We normalize the radial profiles to 4 kpc (the approximate edge of the bulge) for clarity in comparison. The radial gradients (measured from $4 - 12$ kpc for $z = 0$, $2 - 8$ kpc for $z = 1$) vary by no more than ≈ 0.005 dex kpc^{-1} between our fiducial simulation and the simulation with 10 times higher metal diffusion, while the simulation with no metal diffusion has a ≈ 0.13 dex kpc^{-1} steep gradient at $z = 0$. In the right-hand panel, we scaled down the azimuthal scatter in the simulation with no metal diffusion by a factor of 10, for comparison. Thus, neglecting metal diffusion/mixing leads to 10 times higher azimuthal scatter, and moreover, this scatter does not depend much on azimuthal scale. The enhanced metal diffusion re-simulation shows smaller azimuthal scatter at small azimuthal scales, given the enhanced mixing rate on these small scales. However, that simulation shows similar scatter at large azimuthal scales, indicating that disc-wide azimuthal scatter is not sensitive to the detailed choice of diffusion coefficient.

and the fiducial simulation. The gradients vary by less than 0.014 dex between the fiducial simulation and the one with no metal diffusion at $z = 0$. The simulation with no subgrid diffusion has a steeper gradient at $z = 0$, potentially because the metals are less efficient at spreading from a given radius, in the absence of subgrid diffusion, once the disc has become rotationally dominated and radial turbulence is no longer efficient at moving the gas particles. We, thus, conclude that the radial gradients are reasonably robust to choices of the strength of the diffusion coefficient, however, in the unphysical case of no subgrid diffusion, the gradient can be (unphysically) steeper.

Fig. C1 (right-hand panel) compares the azimuthal variations versus angular bin width, at $R = 8$ kpc. The simulation with no subgrid diffusion has 10 times higher azimuthal scatter, so we scale down its values in Fig. C1 by 10 times for visual comparison. Using no subgrid diffusion leads to scatter that is largely independent of azimuthal scale at high z, but that increases with azimuthal scale at low z. Without subgrid diffusion, a small number of gas elements can absorb most of the metals, while a significant number of (neighbouring) elements can remain nearly un-enriched. This is a patently unphysical scenario and it yields azimuthal scatter that disagrees with observations by an order of magnitude. At both redshifts, using a higher diffusion coefficient leads to smaller azimuthal variations at small scales, because diffusion smooths variations between nearby gas elements on scales approaching the resolution (Escala et al. 2018). However, the azimuthal variations are nearly unchanged on large azimuthal scales. Therefore, our results on small azimuthal scales are likely sensitive to the exact choice of diffusion coefficient, but the large-scale azimuthal variations are robust. An important caveat to this comparison is that it is only one simulated galaxy, and individual simulations with the same initial conditions and physics can show non-trivial stochastic variations from random number generators, floating-point roundoff, and chaotic behaviour (e.g. Keller et al. 2019). Indeed, we find minor fluctuations between these simulations, for example, in the exact timing of mergers, which can affect all panels in Fig. C1. We consider it likely that the differences in azimuthal variations on small scales are robust, but any other differences in Fig. C1 are potentially stochastic.

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