Galaxy and mass assembly (GAMA): the inferred mass–metallicity relation from $z = 0$ to 3.5 via forensic SED fitting

Sabine Bellstedt [1]*, Aaron S. G. Robotham [1,2] Simon P. Driver,[1] Jessica E. Thorne [1]
Luke J. M. Davies [1] Benne W. Holwerda [3] Andrew M. Hopkins,[4] Maritza A. Lara-Lopez,[5]
Ángel R. López-Sánchez[2,4,6] and Steven Phillipps [7]

[1]*ICRAR, The University of Western Australia, 7 Fairway, Crawley, WA 6009, Australia*
[2]*ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia*
[3]*Department of Physics and Astronomy, University of Louisville, Natural Science 102, Louisville, KY 40292, USA*
[4]*Australian Astronomical Optics, Macquarie University, 105 Delhi Rd, North Ryde, NSW 2113, Australia*
[5]*Armagh Observatory and Planetarium, College Hill, Armagh BT61 9DG, UK*
[6]*Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia*
[7]*Astrophysics Group, School of Physics, University of Bristol, Bristol BS8 1TL, UK*

ABSTRACT

We analyse the metallicity histories of ∼4500 galaxies from the GAMA survey at $z < 0.06$ modelled by the SED-fitting code PROSPECT using an evolving metallicity implementation. These metallicity histories, in combination with the associated star formation histories, allow us to analyse the inferred gas-phase mass–metallicity relation. Furthermore, we extract the mass–metallicity relation at a sequence of epochs in cosmic history, to track the evolving mass–metallicity relation with time. Through comparison with observations of gas-phase metallicity over a large range of redshifts, we show that, remarkably, our forensic SED analysis has produced an evolving mass–metallicity relationship that is consistent with observations at all epochs. We additionally analyse the three-dimensional mass–metallicity–SFR space, showing that galaxies occupy a clearly defined plane. This plane is shown to be subtly evolving, displaying an increased tilt with time caused by general enrichment, and also the slowing down of star formation with cosmic time. This evolution is most apparent at lookback times greater than 7 Gyr. The trends in metallicity recovered in this work highlight that the evolving metallicity implementation used within the SED-fitting code PROSPECT produces reasonable metallicity results over the history of a galaxy. This is expected to provide a significant improvement to the accuracy of the SED-fitting outputs.

Key words: galaxies: elliptical and lenticular, cD – galaxies: evolution.

1 INTRODUCTION

An analysis of the chemically enriched nature of galaxies, referred to as their metallicity, has become a commonly used tool to study the evolution of galaxy populations. The relation between galaxy metallicity and luminosity was readily highlighted in the past (McClure & van den Bergh 1968; Rubin, Ford & Whitmore 1984), showing that more luminous galaxies tend to have higher metallicities. It has been determined, however, that the scatter in the mass–metallicity is lower than the scatter in the luminosity–metallicity relation (Berg et al. 2012), an indication that it is stellar mass that is more fundamentally linked than luminosity to a galaxy's metallicity. The gas-phase mass–metallicity relation (MZR), has since been studied in great detail for dwarf galaxies (Lequeux et al. 1979; López-Sánchez 2010; Berg et al. 2012; Calabrò et al. 2017; McQuinn et al. 2020), large statistical samples of galaxies (Tremonti et al. 2004; Lara-López, López-Sánchez & Hopkins 2013c; Curti et al. 2020),

galaxies that fall below the MZR (Peeples, Pogge & Stanek 2009), and integral field data (Sánchez et al. 2013, 2019a), etc. to understand why this relation exists, and what its implications are for galaxy evolution.

One of the earliest explanations for this relation came from Larson (1974), who pointed towards gas loss as a means of suppressing metallicity in lower mass galaxies. The gas fraction of a galaxy does seem to impact its position in the MZR, with galaxies deficient in gas tending to have higher metallicities (Hughes et al. 2013; Lara-Lopez et al. 2013a; Zahid et al. 2014a; Brown et al. 2018). Notably, however, the environment of a galaxy seems to have little or no effect on its position in the MZR, as shown both by Mouhcine, Baldry & Bamford (2007) with an analysis of large-scale environment. Hughes et al. (2013) also investigated the role of the local environment on the MZR, concluding that any environmental trends that could be observed, were likely a second-order effect.

Another key element to interpreting the MZR is understanding how it has evolved with time. Many studies have devoted their attention to the measurement of metallicities in galaxies at high redshifts (including Erb et al. 2006; Mannucci et al. 2010; Henry et al. 2013b;

*★ E-mail: sabine.bellstedt@uwa.edu.au

Yabe et al. 2014; Ly et al. 2016; Huang et al. 2019; Sanders et al. 2020; Weldon, Ly & Cooper 2020) in order to characterize how the shape and normalization of the MZR have evolved with time. These studies have shown that the normalization of the MZR was lower at earlier times, with a similar shape.

In a study of the stellar populations in spiral galaxies, Bell & de Jong (2000) established that the metallicities of galaxies were dependent on both mass and surface density, where the star formation histories (SFHs) of galaxies were driven by their surface densities. This pointed towards a connection between metallicity and star formation rate (SFR). Ellison et al. (2008) showed that, at a given stellar mass, the metallicity of a galaxy was higher for galaxies with a lower SFR. Analysis of the SFRs of galaxies across the MZR revealed that the MZR was actually a projection of the three-dimensional mass–metallicity–SFR relation (for example, Lara-López et al. 2010; Mannucci et al. 2010; Yates, Kauffmann & Guo 2012; Lara-López et al. 2013c; Brown, Martini & Andrews 2016; Brown et al. 2018; Curti et al. 2020). We now understand that the status of chemical enrichment of a galaxy is fundamentally linked to the SFR, and the build-up of stellar mass.

Observational measurements of gas-phase metallicities in galaxies are typically conducted via measurements of nebular emission lines (for example, Savaglio et al. 2005; Erb et al. 2006; Yabe et al. 2014; Huang et al. 2019; Sanders et al. 2020). Depending on which lines are detected for an individual galaxy (typically O III and O II, but also lines such as N II, S II, and S III, in addition to Hα and Hβ), various parameters can be derived which, with the combination of carefully pre-determined calibrations (such as those presented by, for example, Kobulnicky & Kewley 2004; Pettini & Pagel 2004; Bian, Kewley & Dopita 2018) can be converted into a gas-phase metallicity value. The impact of both the strong-line parameters and the metallicity calibrations is an extensive field of research, as there are significant systematic differences between the various implementations, as highlighted by Kewley & Ellison (2008) and López-Sánchez & Esteban (2010). Consequently, when conducting an analysis of the evolution in the MZR or trends with SFR, stellar mass, gas fractions, and environment, the underlying measurement systematics must be carefully considered and accounted for.

Not only can the signatures of a galaxy's gas-phase metallicity be found in the strength of its nebular emission lines, but the optical range of the SED is also sensitive to variations in the galaxy's gas-phase metallicity. In particular, variations in the history of a galaxy's gas-phase metallicity can also influence the SED (as demonstrated by Thorne et al., in preparation). With careful modelling, and sufficiently accurate photometric measurements, it is therefore possible to model the metallicity of a galaxy using SED fitting. Historically, the metallicity evolution implementation within SED fitting codes has been simplified, with the focus of most methods instead being placed on the parametrization of a galaxy's SFH. As shown in recent work that uses the SED fitting code PROSPECT (Robotham et al. 2020) to recover the cosmic SFH (Bellstedt et al. 2020a), there are significant benefits to be gained when carefully modelling the evolving gas-phase metallicity in galaxies, rather than simply assuming this value to be constant over time.

In this paper, we extend the work presented by Bellstedt et al. (2020a) to show that not only can this technique accurately reproduce the SFH of the Universe, but it can also successfully recover the metallicity distributions of galaxy populations. This recovery is important, as it highlights that a more complex approach to metallicity modelling in SED fitting can produce physical results in a broad range of parameter spaces. We describe the GAMA

data in Section 2, and the SED fitting technique in Section 3. Our derived MZR is presented in Section 4, followed by an analysis of the evolving MZR in Section 4.2. We additionally present our derivation of the mass–metallicity–SFR plane evolving through cosmic time in Section 5. We finally discuss these results in Section 6.

For all stellar mass measurements presented in this work, we have utilized a Chabrier (2003) initial mass function (IMF). The cosmology assumed throughout this paper is $H_0 = 67.8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_m = 0.308$, and $\Omega_\Lambda = 0.692$ (consistent with a *Planck* 15 cosmology, Planck Collaboration XIII 2016).

2 DATA

The GAMA survey (Driver et al. 2011; Liske et al. 2015) is a large program that gathered redshifts for $\sim$300 000 galaxies across five different fields spanning 230 deg^2 using the Anglo Australian Telescope. One of the great strengths of the survey is that it achieved a high spectroscopic completeness of 98 per cent above a magnitude limit of $m_r \leq 19.5$[1] (or $m_i \leq 19.0$ in the G23 field).

In this work, as in Bellstedt et al. (2020a), we utilize the spectroscopic and photometric data for 6688 galaxies with $z < 0.06$ and $m_r \leq 19.5$ in the three equatorial fields (G09, G12, and G15) to conduct SED fitting using the code PROSPECT (Robotham et al. 2020). The panchromatic photometry catalogue for the GAMA survey was recently updated to include the KiDS imaging in the optical bands, and also to utilize the PROFOUND source detection software (Robotham et al. 2018). This updated photometry was presented in Bellstedt et al. (2020b). These data include photometry in 19 bands from the far-UV to the far-IR.

As in Bellstedt et al. (2020a), we do not explicitly identify potential AGN in the sample to be removed. We expect that this will have a minimal impact on our results, as the AGN contamination in this sample is expected to be very small (fewer than 30 galaxies, Prescott et al. 2016).

In this work, we use v1 of the GAMAKidsVikingFIR DMU.

3 SED FITTING

We implement the same method as outlined in Bellstedt et al. (2020a), using the GAMA photometry presented in Bellstedt et al. (2020b) and passed into the PROSPECT SED-fitting code (Robotham et al. 2020). For a detailed description of the fitting, we direct the reader to Bellstedt et al. (2020a), however, we provide a brief summary in this section. The stellar templates we use in this analysis are from Bruzual & Charlot (2003), and the SFH is parametrized by the massfunc_snorm_trunc function within PROSPECT. This parametrization takes on the form of a skewed normal distribution, with the peak position (mpeak), peak SFR (mSFR), SFH width (mperiod), and SFH skewness (mskew) set as free parameters. A positive value of the skewness produces a SFH tailing off towards the present day, whereas a negative skewness causes the SFH to tail off towards the start of the Universe. The SFH is anchored to 0 at a lookback time of 13.4 Gyr, deemed in this work to be the age at which galaxies start forming. As outlined by Bellstedt et al. (2020a), this value was selected to correspond with the epoch at which the highest z galaxies are known to exist ($z = 11$, Oesch et al. 2016).

---

[1]With the development of the updated photometric catalogue for the survey (Bellstedt et al. 2020b), the completeness limit has been updated from $m_r \leq$ 19.8 to $m_r \leq 19.5$ in the equatorial fields.

In this PROSPECT analysis, we implement an evolving metallicity where the shape of the stellar mass evolution is linearly mapped on to the shape of the metallicity evolution for each galaxy, given by the `Zfunc_massmap_lin` function. This ensures that chemical enrichment in the galaxies follows the assumed SFR, where increased star formation is associated with an increased rate of metal production in the galaxy. The final metallicity of the galaxy is allowed to be a free parameter, `Zfinal`. We highlight that this value represents the present-day gas-phase metallicity of the object, and correspondingly the metallicity of the youngest stars in the galaxy (as opposed to a time-averaged stellar metallicity). This approach is a significant improvement over the typical approach in SED fitting, which is to assume that the metallicity is constant over a galaxy's history. The impact of the metallicity assumption on the cosmic star formation history (CSFH) was demonstrated in fig. 4 of Bellstedt et al. (2020a). The range in `Zfinal` values is limited by the range of metallicity in the Bruzual & Charlot (2003) stellar templates. The upper limit of these templates is 0.05, and hence our recovered `Zfinal` values cannot extend beyond this value. The resulting CSFH and Stellar Mass Density (SMD) that are derived using this metallicity implementation are shown in appendix B of Bellstedt et al. (2020a).

The fitting outputs used in this work were first presented by Bellstedt et al. (2020a), where they were used to derive the cosmic SFH, and the cosmic metal density evolution. While the main results in that study were derived using a closed-box metallicity implementation in PROSPECT (as given by the `Zfunc_massmap_box` function), appendix B presented the cosmic SFH and cosmic stellar mass density when assuming linear metallicity evolution, as given by the `Zfunc_massmap_lin` function. Because the yields are not assumed to be constant in the linear metallicity implementation (unlike the closed-box implementation), the late-time enrichment of galaxies is slightly reduced when this metallicity evolution is prescribed. While the selected metallicity implementation does not have a large impact on the results, we noted that the number of objects hitting the upper metallicity limit is lower when assuming linear metallicity evolution. We interpret this to indicate that the metallicity outputs derived when using the `Zfunc_massmap_lin` are more physical. This assumption of allowing the metallicity to grow in proportion to the stellar mass growth is similar to what is seen in the chemical enrichment of galaxies in semi-analytic models (SAMs; as seen in, for example, Robotham et al. 2020), providing motivation to use this implementation. See the discussion in Appendix A for a more detailed analysis of the enrichment in SHARK, highlighting the degree to which the proportionality assumption is accurate in this semi-analtyic model. As such, we use the outputs as derived by the `Zfunc_massmap_lin` function in this work.

In addition to the five free parameters specifying the star formation and metallicity histories, we include four free parameters to describe the dust contribution to the SED. The dust is assumed to exist in two forms; either in birth clouds formed around young stars, or distributed as a screen in the interstellar medium. For each of these components, we include two free parameters, describing the dust opacity (`tau_birth`, `tau_screen`), and the dust radiation field intensity (`alpha_birth`, `alpha_screen`). Hence, we model the SED in our work using a total of nine free parameters. The fitting ranges and priors are presented in table 2 of Bellstedt et al. (2020a).

Of the 6688 galaxies analysed in the $z < 0.06$ sample presented in Bellstedt et al. (2020a), in this analysis we focus only on a subset of these objects that have a reasonable constraint on the metallicity parameter from SED fitting. In order to determine this
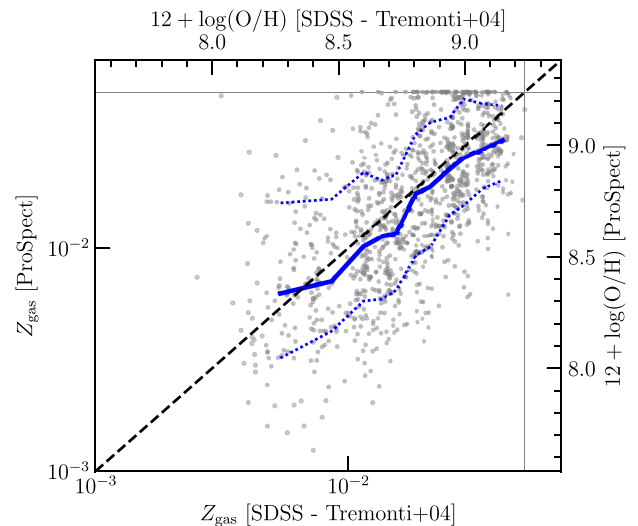
Figure 1. Comparison of the PROSPECT-derived metallicity values with the corresponding measured metallicites from Tremonti et al. (2004) for a matched sample. The black-dashed line shows the one-to-one, and the thin grey vertical and horizontal lines show the upper metallicity limit of the Bruzual & Charlot (2003) templates. The solid blue line shows the running median, with the 1σ scatter indicated by the blue-dotted lines. The scatter in our metallicity recovery is $\sim 0.25\,$dex, with an offset of $-0.06\,$dex.

subset, we remove from analysis any objects for which the 1σ uncertainty from the MCMC sampling is greater than 0.5 dex for the `Zfinal` parameter. After removing these objects, we are left with a sample of 4531 galaxies for which we have constrained metallicity estimates.

3.1 Comparison to SDSS metallicities

A subset of the objects analysed in this work have spectroscopically derived metallicities from SDSS[2] (Brinchmann et al. 2004; Tremonti et al. 2004). We compare our derived metallicity values against these observational measurements as an assessment of the accuracy of our SED-fitting approach to modelling this quantity. The subset consists of 2220 objects, and a comparison of the metallicities is presented in Fig. 1. Here, we show that the SED-derived values follow the spectroscopically derived values generally well, with a mean offset of $-0.06\,$dex and a scatter of $\sim 0.25\,$dex. Interestingly, those values in our sample that are hitting the upper metallicity limit as governed by the upper limit of the Bruzual & Charlot (2003) templates cover a range of values as determined by Tremonti et al. (2004). Although the Tremonti et al. (2004) measurements are not restricted to the same upper limit as the Bruzual & Charlot (2003) templates, the highest metallicities are similar to this limit.

The SED-derived uncertainties are significantly larger than the spectroscopically derived uncertainties, which is indicative of the lower constraint that broad-band photometry is able to provide. The broad agreement between the observationally measured values from SDSS and the inferred values from our SED fitting highlight that the values we recover are not simply 'nuisance' parameters, and are instead physically meaningful (albeit with significant uncertainty).

[2]https://www.mpa.mpa-garching.mpg.de/SDSS/DR7/oh.html
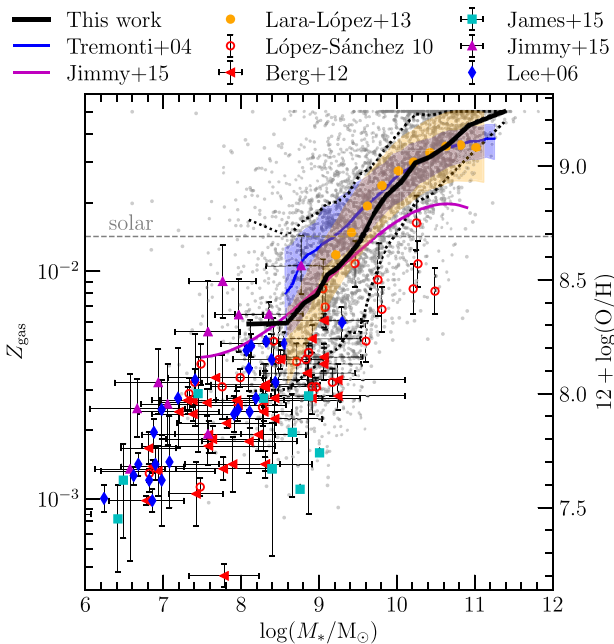
Figure 2. The resulting mass–metallicity relationship when assuming a metallicity that evolves with the star formation history. The blue line indicates the median MZR recovered by Tremonti et al. (2004), the orange points indicate the binned metallicity measurements by Lara-López et al. (2013b), and the magenta line shows the fit to the MZR by Jimmy et al. (2015). We additionally include observations that extend to lower stellar masses from Lee et al. (2006), López-Sánchez (2010), Berg et al. (2012), James et al. (2015), and Jimmy et al. (2015).

4 THE MASS–METALLICITY RELATION

While there was no fitting prior set on the resulting gas-phase metallicity values in our implementation, we recover in our analysis a mass–metallicity distribution that is consistent with trends recovered by other observations. This is shown in Fig. 2, where for each galaxy in our sample, we plot the resulting stellar mass against the fitted `Zfinal` value. The overall trend of our MZR is shown via the solid black line showing the moving median, and the black-dashed lines that indicate the 1σ range in metallicity at any given stellar mass. In calculating these values, we demand that each bin includes at least 300 galaxies. The scatter in the relation at stellar masses below 10^{10} M$_\odot$ is significant, but this scatter reduces at larger stellar masses. We see a bending of this relation at $M_* \sim 10^{10}$ M$_\odot$. Below $M_* \sim 10^9$ M$_\odot$ our sample becomes increasingly incomplete, so the MZR at these masses is prone to bias.

A clear artefact in this image is the upper limit in the range of metallicity values at 0.05. This is particuarly stark at $M_* > 10^{10.5}$ M$_\odot$, where the upper region of our 1σ range is at this limit. This limit is the highest metallicity template present in the Bruzual & Charlot (2003) stellar population templates, and hence our application of PROSPECT is not sensitive to gas-phase metallicity values larger than $Z_{\rm gas} = 0.05$.

To compare against observational trends, we include measurements of metallicity made for GAMA galaxies by Lara-López et al. (2013b), and the MZR fit with 1σ scatter produced using SDSS galaxies by Tremonti et al. (2004). We additionally include the fit to the MZR presented by Jimmy et al. (2015), in which more massive galaxies were determined using the ALFALFA/SDSS sample. Note that this MZR is significantly offset to both the relation recovered by

PROSPECT, and also to the observational measurements of Tremonti et al. (2004) and Lara-López et al. (2013b). We summarize the parameters and calibrators used by each of these studies to determine metallicities in Table 1.

While the aforementioned MZR measurements are made using massive galaxies, we also include in Fig. 2 a comparison to metallicity measurements made for galaxies in the dwarf regime by Lee et al. (2006), López-Sánchez (2010),[3] Berg et al. (2012), James et al. (2015), and Jimmy et al. (2015), in each case correcting the stellar masses to a Chabrier IMF where necessary using the conversion factors presented in table 1 of Driver et al. (2013). We find that, while these observations overlap with the distribution of points derived by PROSPECT, these metallicity measurements (all derived using T_e methods) all have on average slightly lower values than those we derive in the low-mass range. Similarly, these values are also systematically lower than the observations by Tremonti et al. (2004), Lara-López et al. (2013b), and Jimmy et al. (2015) for the overlapping mass range.

At the highest stellar masses ($M_* > 10^{10.5}$ M$_\odot$), our median MZR is greater than that measured by other observations. Observations by Tremonti et al. (2004) and Lara-López et al. (2013b) are necessarily restricted to galaxies with star formation, as the metallicity measurement is made on the emission lines that are produced by star formation. Early-type galaxies such as ellipticals and lenticulars are very metal-rich and dominate the massive end of the MZR, and as such the high-mass MZR may be biased high in our analysis. We determine that a cut in specific SFR does not cause the median metallicity value at high-mass to reduce, however, and therefore the presence of early-type galaxies in our analysis is unlikely to account for the larger metallicities that we derive at large stellar masses.

We show how galaxies with different visual morphologies contribute to the MZR in Fig. 3. Here, elliptical galaxies are shown in red, `S0-Sa` galaxies in orange, `Sab-Scd` galaxies in green, and `Sd-Irr` galaxies in blue. The histogram above the main panel of the plot shows how the morphologies are distributed with stellar mass, whereas the histogram to the right of the main panel shows how they are distributed with gas-phase metallicity. Fig. 3 highlights that early-type galaxies dominate the MZR at high mass and high metallicity, whereas the low-metallicity portion of the plot is almost entirely occupied by late-type galaxies. Additionally, Fig. 3 shows that the dispersion of the MZR in the low-metallicity regime (dominated by `Sd-Irr` galaxies) is much higher than the high-mass regime.
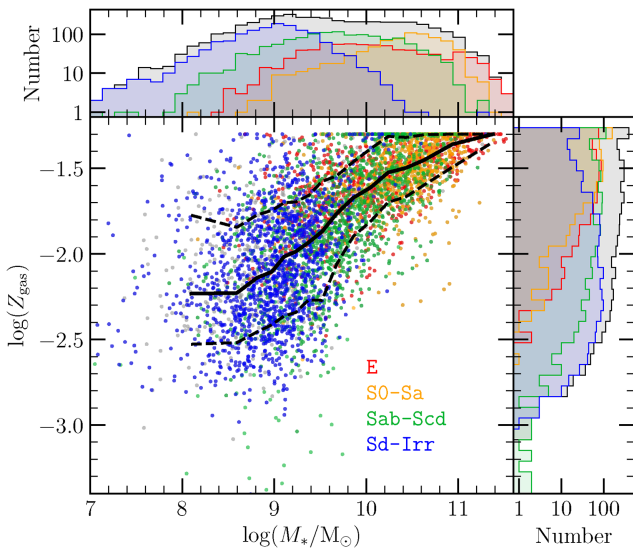
4.1 Comparison with simulations

The MZR has historically been very difficult for simulations and SAMs to reproduce, due to the intricate nature of chemical evolution in galaxies.

In recent years, there has been an increased reporting of studies that are producing MZR trends more like observations. In Fig. 4, we briefly compare our derived MZR with those prodocued by leading simulations/SAMs. We include the MZR derived by the cosmological, hydrodynamic simulations Illustris (Torrey et al. 2014), IllustrisTNG (Torrey et al. 2019), MUFASA (Davé et al. 2017), SIMBA (Davé et al. 2019), and EAGLE (Zenocratti et al. 2020), and the SAMs GAEA (De Lucia et al. 2020) and SHARK (Lagos et al. 2018).

---

[3]We exclude from Fig. 2 the galaxies from López-Sánchez (2010) that are undergoing mergers.

Table 1. Metallicity indicators and calibrations applied in each of the observational studies presented in Fig. 2. T_e refers to the electron temperature, which is derived using auroral lines, and is regarded as a 'direct' method of measuring metallicity. PP04 refers to Pettini & Pagel (2004), and CL01 refers to Charlot & Longhetti (2001).

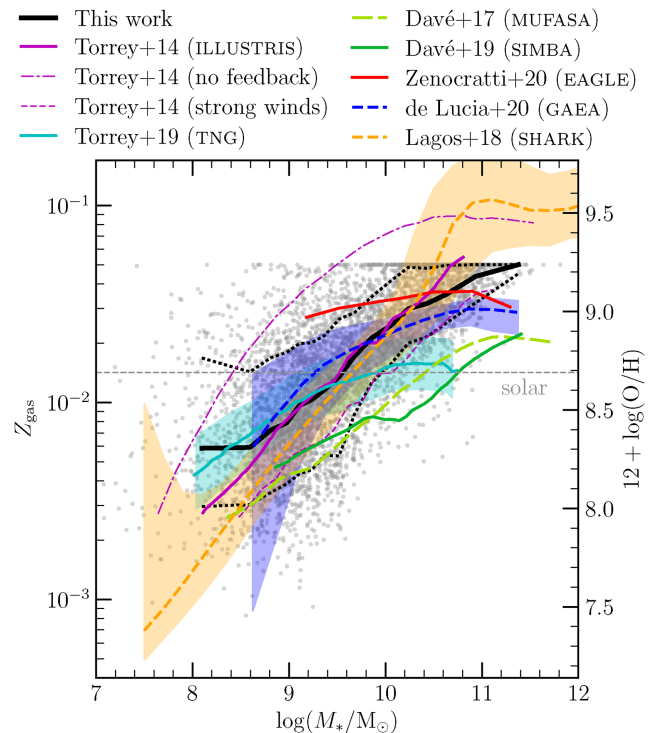| Study	Parameters/Emission lines	Calibration	Comments
Tremonti et al. (2004)	[O II], H β, [O III], H α, [N II], [S II]		Simultaneous fit using CL01
Lee et al. (2006)	[O III]λ4363		T_e derivation
López-Sánchez (2010)	[O III](λ4959 + λ5507)/λ4363, [N II](λ6548 + λ6583)/λ5755, [OII](λ3727+λ3729)/(λ7319 + λ7330)		T_e derivation
Lara-López et al. (2013b)	O3N2	PP04	Strong line derivation
Berg et al. (2012)	[O III]λ4363		T_e derivation
Jimmy et al. (2015)	N2	Denicoló, Terlevich & Terlevich (2002)	Strong line derivation
James et al. (2015)	[O II]$\lambda\lambda$3727, 3729		T_e derivation

Figure 3. The mass–metallicity relation, as divided by the visually classified morphological types.

Figure 4. Comparison of our MZR with simulations. Here, we include the cosmological hydrodynamic simulations Illustris (Torrey et al. 2014, both the default model, and also variations of the feedback model including no feedback, and string winds), IllustrisTNG (Torrey et al. 2019), MUFASA (Davé et al. 2017), SIMBA (Davé et al. 2019), EAGLE (Zenocratti et al. 2020), and the semi-analytic models GAEA (De Lucia et al. 2020) and SHARK (Lagos et al. 2018).

Torrey et al. (2014) demonstrate with Illustris that the adopted feedback model has a dramatic influence on the resulting MZR, with no feedback resulting in much higher metallicities (shown in Fig. 4 by the magenta-dashed line that extends beyond the upper limit of our probed metallicity range), whilst strong feedback reduces the normalization of the MZR. While the bending of the MZR was recovered by Illustris when feedback was removed, the bending of the MZR at high stellar masses was not present when feedback was included. For the default feedback model, however, the agreement between the Illustris MZR and the MZR recovered in this work is consistent at stellar masses $M_* < 10^{10.5} M_\odot$.

Unlike Illustris, where little bending was observed, the MZR for IllustrisTNG (Torrey et al. 2019, shown in Fig. 4 in cyan) does recover a saturation metallicity. Both the stellar mass at which this bending occurs, as well as the metallicity value, is significantly lower in IllustrisTNG than we recover using PROSPECT.

In addition to a systematically lower normalization, the shape of the MZR recovered by SIMBA (Davé et al. 2019, shown in green) is different to that of other simulations. It displays a dip at around $10^{10} M_\odot$, without an indication of a saturation in metallicity at the highest stellar masses. Interestingly a predecessor of SIMBA, MUFASA (Davé et al. 2017, dashed light green), displays an MZR shape that is more consistent with the other trends presented in Fig. 4.

This is despite the fact that Davé et al. (2019) describe the MUFASA MZR as being too steep.

We also include in Fig. 4 the MZR from the EAGLE simulations by Zenocratti et al. (2020, solid red line). On average, EAGLE galaxies seem to have a higher metallicity than other simulations and observations, with supersolar values at all stellar masses. Furthermore, EAGLE does not recover the characteristic MZR shape with lower metallicities in low-mass galaxies, displaying instead a relatively constant metallicity with varying stellar mass.

The GAEA MZR (De Lucia et al. 2020, dashed blue line) recovers the typical lower metallicities for low-mass galaxies, and a saturation of metallicities for high-mass galaxies, although the difference in metallicity between low-mass and high-mass systems is less extreme

than what we derive. The scatter from De Lucia et al. (2020) was reported to be much larger than observations at lower stellar masses, although we note that their scatter is only slightly larger than the 1σ range that we derive. The MZR from the SAM SHARK (Lagos et al. 2018, the orange-dashed line) is consistent in the stellar mass range $10^9 < M_*/M_\odot < 10^{10.5}$, however, at the high-mass end the predicted MZR is significantly higher than what we infer.

Simulations frequently compare against observations of the MZR as a way of assessing how closely the physical implementation of the simulation can reproduce reality. One set of observations that are frequently used for comparison is the SDSS data set by Tremonti et al. (2004), shown in Figs 1 and 2. Due to the biases introduced by different methods of metallicity calibration (as emphasized in the work by Kewley & Ellison 2008), the absolute normalization of the Tremonti et al. (2004) data set is debated in the literature. When compared against simulations in the works by Torrey et al. (2019), Davé et al. (2017, 2019), the Tremonti et al. (2004) vaues were scaled downwards by 0.26 dex, to be consistent with the calibration of Pettini & Pagel (2004). As a result, the agreement between the observations and simulations in those works appears closer than the agreement presented in our Fig. 4. The complexity of comparing observational metallicity measurements with simulations is discussed in detail in Lagos et al. (2012).

In all of the MZRs recovered by simulations, the lower mass limit in each comparison is determined by the resolution limit of the underlying dark matter in each simulation. As a result, the simulated behaviour at low stellar masses cannot be determined. As such, there is no way of comparing whether the low-mass turnover in the MZR we recover is also replicated by either hydrodynamic simulations or SAMs.

The important aspect to consider when comparing different simulations, is that current state-of-the-art models (either cosmological or semi-analytic) of galaxy evolution still predictly widely varying mass–metallicity relations. In fact, the variation between these MZR predictions is significantly greater than the discrepancies we see between our inferred MZR and that of observations. This highlights that our approach to metallicity evolution in galaxies is competitive, if not perfect.

4.2 Evolution in the MZR

Due to the forensic nature of our analysis, we not only extract the $z = 0$ mass–metallicity relation, but also the evolution of this relation over cosmic time, by analysing the star formation and metallicity histories derived using PROSPECT for individual galaxies. As such, at any arbitrary value of lookback time, we can determine the forensically inferred gas-phase metallicity, SFR, and stellar mass at that epoch. This enables us to contruct the MZR of the sample not only at the epoch of observation, but also at other epochs in the Universe's history. The following sections will present an analysis of the evolving MZR that we derive.

The manner in which we use our metallicity and SFHs to trace back the MZR is shown in Fig. 5. For four example galaxies (7551, 425755, 3895257, and 7839), we present the SFH and metallicity history (ZH) as modelled by PROSPECT, in a two-panel subplot, with the corresponding model fit to the SED shown above. The MCMC sampled distribution is shown as 1000 grey lines, and the median SFH and ZH are shown as the solid coloured lines. For each of these histories, we indicate the value at 1 Gyr intervals using coloured points. In the top middle panel, we show the mass–metallicity plot at $z = 0$, where a track has been added for each example galaxy. These example galaxies vary from low-mass galaxies that form their

stars more recently, to massive galaxies that formed their stars early in the Universe. This is evident in the mass–metallicity plot, where each galaxy has a track in a different part of the parameter space. It is interesting to note that these tracks bear much resemblance to the various tracks for galaxies of different morphologies presented in fig. 5 by Calura et al. (2009), who produced theoretical chemical enrichment models of galaxies with different morphologies by applying assumptions about chemical enrichment via star formation, inflows, and outflows.

4.2.1 Comparison with observations

The changing mass–metallicity relation at 1 Gyr intervals is shown in each subpanel of Fig. 6. For each time interval, the median and 1σ values are shown in the solid and black-dashed lines, respectively. In generating the median and 1σ ranges, we demand that each bin includes at least 300 galaxies. For comparison, in each subpanel we show the $z \sim 0$ MZR in cyan. Wherever possible, we have included observations of the mass–metallicity relation at the relevant epochs. These studies include Savaglio et al. (2005), Erb et al. (2006), Maiolino et al. (2008), Mannucci et al. (2009), Henry et al. (2013a,b), Lara-López et al. (2013b), Yabe et al. (2014), Zahid et al. (2014b), Ly et al. (2016), Huang et al. (2019), Cameron et al. (2019), Weldon et al. (2020), and Gillman et al. (2021), converting from the published $12 + \log(\text{O/H})$ values to absolute metallicity via

$$Z_{\text{gas}} = Z_\odot \times 10^{[12+\log(\text{O/H})]-[12+\log(\text{O/H})]_\odot}, \qquad (1)$$

where $Z_\odot = 0.0142^4$ and $[12 + \log(\text{O/H})]_\odot = 8.69$ (Asplund et al. 2009). We additionally correct all stellar mass measurements to be consistent with a Chabrier (2003) IMF, employing correction factors as derived by Driver et al. (2013).[5]

A number of different methods were used to make the observed metallicity measurements presented in Fig. 6. The R_{23} parameter (Pagel et al. 1979) and the N2 parameter were used, as well as the O3N2 parameter and the O_{32} parameter. Even if the same parameters are applied to measure metallicity, they can be differently calibrated. The definitions of these parameters are

$$R_{23} \equiv \frac{[\text{O\,II}]\lambda 3727 + [\text{O\,III}]\lambda 4959, \lambda 5007}{\text{H}\beta}, \qquad (2)$$

$$O_{32} \equiv \frac{[\text{O\,III}]\lambda 4959 + [\text{O\,III}]\lambda 5007}{[\text{O\,II}]\lambda 3727}, \qquad (3)$$

$$\text{N2} \equiv \log\left(\frac{[\text{N\,II}]\lambda 6584}{\text{H}\alpha}\right), \quad \text{and} \qquad (4)$$

$$\text{O3N2} \equiv \log\left(\frac{[\text{O\,III}]\lambda 5007/\text{H}\beta}{[\text{N\,II}]\lambda 6584/\text{H}\alpha}\right). \qquad (5)$$

Frequently applied calibrations include the N2 and O3N2 calibrations by Pettini & Pagel (2004), which were later updated by Tremonti et al. (2004), and the calibration of the R_{23} parameter by Kobulnicky & Kewley (2004). We summarize the parameters and calibrators used by each of these studies in Table 2. Note that in the high-mass regime, strong line derivations are generally employed, whereas in the low-mas regime, electron temperature-based derivations are employed. Some scatter between the observations themselves is to be expected due to these different indicators alone, as highlighted by Kewley & Ellison (2008), López-Sánchez & Esteban (2010), and López-Sánchez et al. (2012).

---

[4]Note that this value differs from the commonly assumed value of 0.02.

[5]To convert from a Salpeter (1955) IMF we multiply by a factor of 0.65, and for a Baldry & Glazebrook (2003) IMF we use a factor of 1.2.
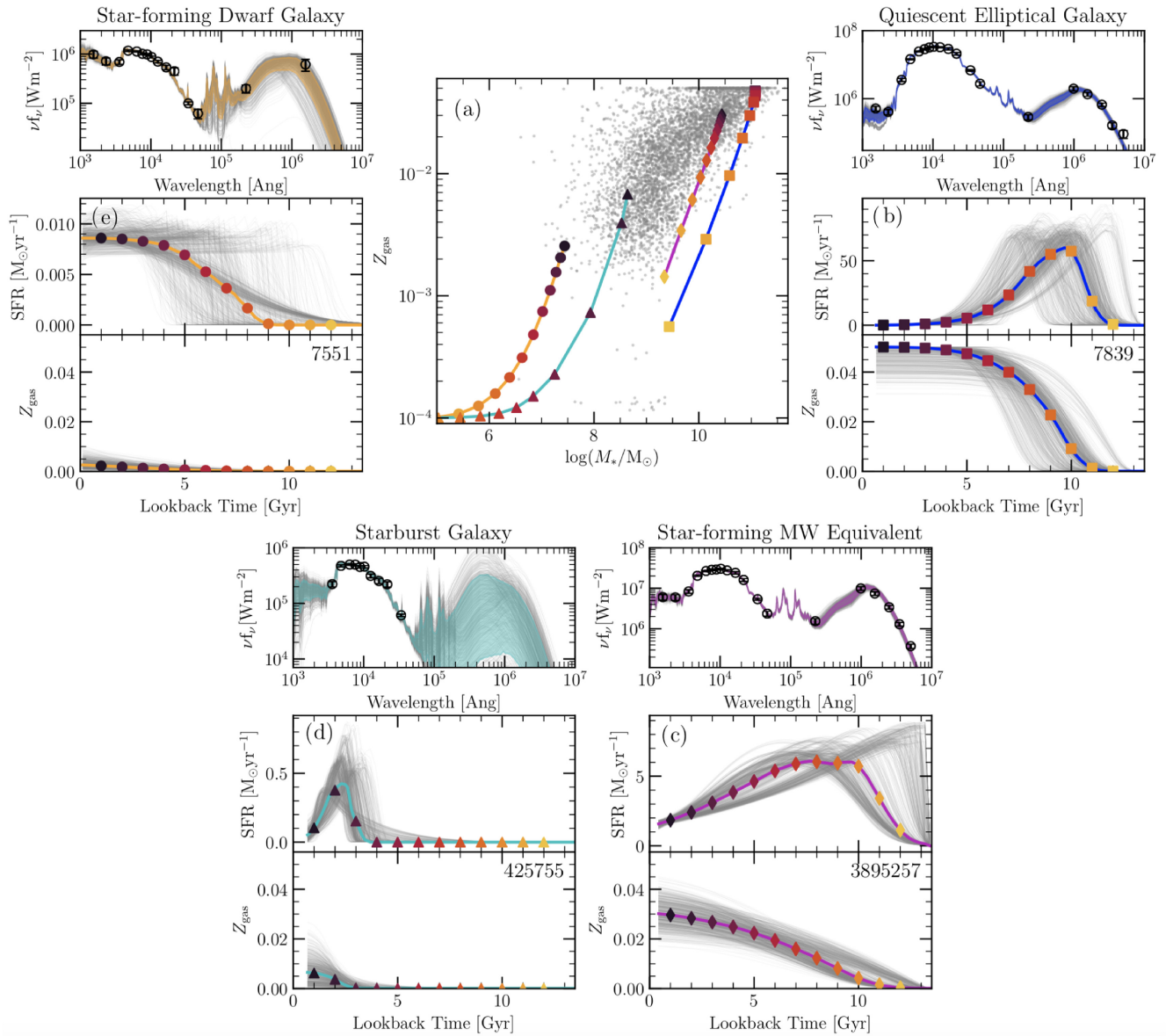
Figure 5. Tracks of four individual galaxies in the mass–metallicity plot (a) over cosmic time. For each example galaxy, we show the modelled star formation histories and metallicity histories, as well as the model fit to the SED. These galaxies are (b) 7839, (c) 3895257, (d) 425755, and (e) 7551. The SFR and metallicity values at 1 Gyr intervals are shown in coloured points, with the corresponding position on the mass–metallicity plot shown in panel (a). Symbols are used to differentiate between different example galaxies. The distribution of mass–metallicity points for our full sample is shown as the grey points in panel (a).

The agreement between the trends recovered by observations and our forensically determined MZR at each epoch is remarkable, at all stellar mass ranges. Note that in classical SED fitting approaches, where metallicity is assumed to be constant with time, the inferred metallicity distribution would be recovered to be constant across all subpanels of Fig. 6, with only stellar masses evolving with time. This would be in clear tension with observations, highlighting the improvement that has been gained through the implementation of an evolving metallicity parametrization in our modelling. We note that the metallicities derived for the most massive galaxies in our sample ($M_* > 10^{10.5} M_\odot$) tend to be slightly higher than observed metallicities at these stellar masses, and hence the flattening of our derived MZR is slightly weaker than observed relations. Observations between 9 and 11 Gyr (Erb et al. 2006; Yabe et al. 2014;

Zahid et al. 2014b; Gillman et al. 2021) recover metallicities lower than our forensic values in the highest stellar mass bins. This looks likely to be the result of metallicity saturation from the N2 indicator, which is known to occur for metallicities above $12 + \log(O/H) \geq 8.8$ (corresponding to $Z_{\rm gas} \geq 0.018$), as discussed by Lara-López et al. (2013b). Similarly, in the 7 Gyr bin, the observations of both Ly et al. (2016) and Weldon et al. (2020) flatten at higher stellar masses, whereas other metallicity measurements focussing on a larger mass range do recover significantly higher metallicities in the overlapping mass range. This also suggests that saturation is occurring when using $T_{\rm e}$-based metallicity measurements.

The potential presence of saturation in observational metallicity measurements makes a comparison between our values and observations for massive galaxies at 9–11 Gyr potentially biased.
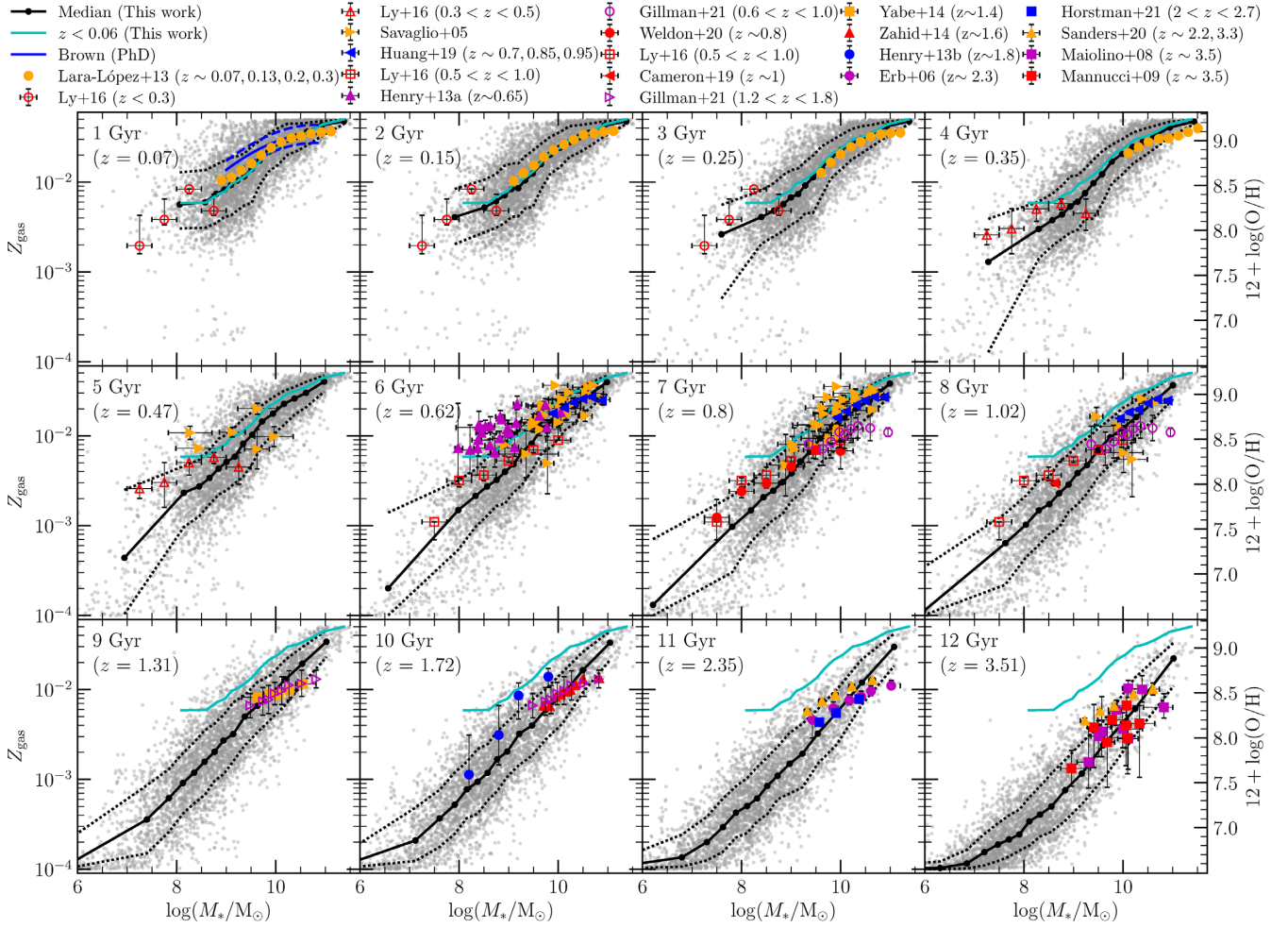
Figure 6. The resulting mass–metallicity relation at 1Gyr intervals resulting from the PROSPECT fits. At each interval, the running median is indicated with a solid black line and the black points, and the 1σ range in the scatter is shown in the black-dotted lines. Where possible, we have included observational measurements at the relevant epochs as a comparison. These studies include those of Savaglio et al. (2005), Erb et al. (2006), Maiolino et al. (2008), Mannucci et al. (2009), Henry et al. (2013a,b), Lara-López et al. (2013b), Yabe et al. (2014), Zahid et al. (2014b), Ly et al. (2016), Huang et al. (2019), Cameron et al. (2019), Horstman et al. (2021), Weldon et al. (2020), Sanders et al. (2020), and Gillman et al. (2021). For measurements of the MZR that span a large redshift range (including those of Ly et al. 2016), we plot the same values over multiple subpanels, and show the data using the open symbols.

Table 2. Metallicity indicators and calibrations applied in each of the observational studies presented in Fig. 6. PP04 refers to Pettini & Pagel (2004), KK04 refers to Kobulnicky & Kewley (2004), and M08 refers to Maiolino et al. (2008). T_e refers to the electron temperature, which is derived using auroral lines, and is regarded as a 'direct' method of measuring metallicity.

Study	Parameters/emission lines	Calibration	Comments
Savaglio et al. (2005)	R_{23}, O_{32}	KK04	Strong line derivation
Erb et al. (2006)	N2	PP04	Strong line derivation
Maiolino et al. (2008)	[O III]λ5007/Hβ, [O III]λ5007/[O II]λ3237	Nagao, Maiolino & Marconi (2006)	T_e at low-Z, and photoionization at high-Z
Mannucci et al. (2009)	R_{23}, [O III]λ5007/Hβ	M08	
Yabe et al. (2014)	N2	PP04	Strong line derivation
Lara-López et al. (2013b)	O3N2	PP04	Strong line derivation
Henry et al. (2013a)	R_{23}	KK04	Strong line derivation
Henry et al. (2013b)	R_{23}	M08	Strong line derivation
Zahid et al. (2014b)	N2	PP04	Strong line derivation
Ly et al. (2016)	[O II]λ3726, λ3729/Hβ, [O III]λ4959, λ5007/Hβ		T_e derivation
Cameron et al. (2019)	O3N2	PP04	Strong line derivation
Huang et al. (2019)	R_{23}, O_{32}	KK04	Strong line derivation
Horstman et al. (2021)	N2, O3N2	PP04	Strong line derivation
Weldon et al. (2020)	[O II]λ3726, λ3729/Hβ, [O III]λ4959, λ5007/Hβ		T_e derivation
Sanders et al. (2020)	[O III]λ5007/Hβ, O_{32}, [Ne III]λ3869/[O II]λ3727	Bian et al. (2018)	Strong line derivation
Gillman et al. (2021)	N2	PP04	Strong line derivation
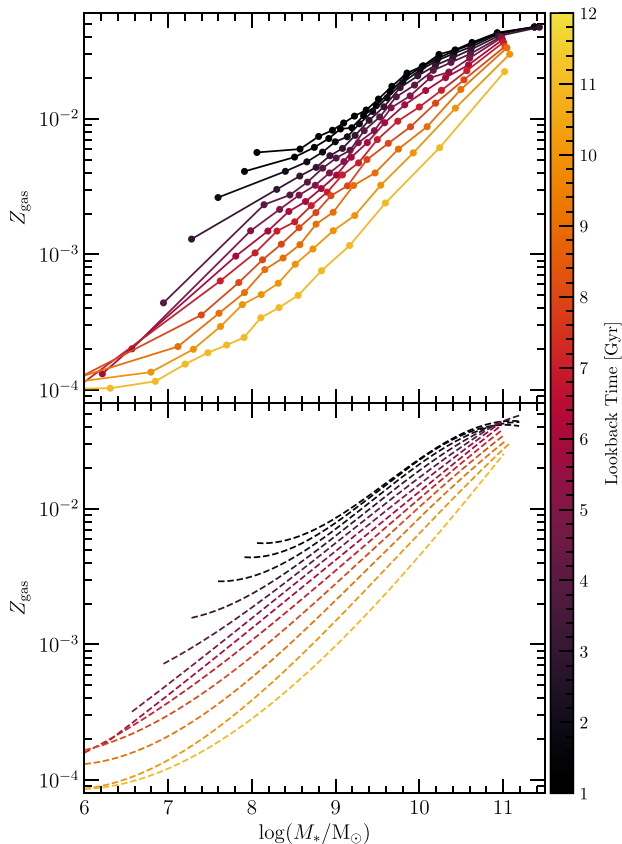
Figure 7. The top panel shows the median MZR measured at each epoch (as shown in each subpanel of Fig. 6), coloured by lookback time. The bottom panel shows the functional form of the evoving MZR, as given by equation (6).

4.2.2 Evolving shape and normalization

To visually demonstrate how the form of the MZR evolves with time, we present the median MZR at each epoch in Fig. 7, coloured by lookback time. As expected, the MZR reduces in normalization with increasing lookback time, with the bending at higher masses becoming less pronounced at earlier times. This changing shape (albeit subtle) seems to suggest a 'saturation' of metallicities occurring in galaxies at recent times.

This evolution in the metallicity can be described as a function of stellar mass and lookback time ($t_{\rm lb}$) by

$$\log(Z_{\rm gas})(M_*, t_{\rm lb}) = \sum_{i=0}^{3} f_i(t_{\rm lb}) m^i, \quad (6)$$

where

$$m = \log(M_*/{\rm M_\odot}) - 10, \quad (7)$$

$$f_i(t_{\rm lb}) = \sum_{j=0}^{5} a_{i,j} t_{\rm lb}^j, \quad (8)$$

and the $a_{i,j}$ coefficients are provided in Table 3. These fits are shown in the lower panel of Fig. 7. The scatter (in dex) can be described by the same functional form. The corresponding coefficients are also shown in Table 3.

The evolution of the MZR shown in Fig. 7 shows that the chemical enrichment of a $10^8\,{\rm M_\odot}$ galaxy is significantly more different at high-z versus low-z, than that of a $10^{10}\,{\rm M_\odot}$ galaxy, which displays

less variation in metallicity with cosmic time. This behaviour is attributed to the 'galaxy downsizing' phenomenon that is recovered in this analysis (as highlighted by Bellstedt et al. 2020a). Massive galaxies tend to assemble earlier, and hence their late-time chemical enrichment is minimal, whereas low-mass galaxies continue to form stellar mass at the present day. Consequently, this stellar mass range experiences more prolonged chemical enrichment.

4.3 Trends with SFR

For each cosmic epoch shown in Fig. 6, we now assess how the galaxy SFRs change across the mass–metallicity relation. In each subpanel in Fig. 8, each galaxy is now coloured by log(SFR), with galaxies lighter in colour representing objects that are more rapidly forming stars. Fig. 8 shows clearly that the position of a galaxy within the MZR is highly dependent on the SFR. In particular, we highlight that this trend is strongest at earlier times, and becomes less distinct in the last 5 Gyr, where galaxies start to saturate in gas-phase metallicity. This plot shows that in the early Universe, metallicity enrichment and SFR are directly linked to stellar mass, such that at fixed stellar mass, the SFR is relatively constant for different metallicity values. A vertical gradient in colour (a dependence of SFR on metallicity), only starts becoming apparent at a lookback time of 10 Gyr – the same epoch in which the bending of the MZR starts to become apparent.

It is conceivable that the recovered behaviour is simply a consequence of the parametrization of SFHs that we implement, and that therefore the trends in the early Universe are highly simplified. Additionally, the actual constraint on the SFH and ZH at such large lookback times is also very small when applying SED fitting, so there may not actually be any real signal here – only what our models are telling us. Note, however, that if this behaviour really is simply the consequence of the adopted SFH, then the fact that the evolving MZR is so well recovered provides confidence that the adopted SFH and ZH parametrization is indeed appropriate.

We highlight that Fig. 8 also conveys the clear mass-dependence of the CSFH. In particular, at all redshifts, the highest mass systems have the greatest SFRs. The value of this maximum SFR changes with redshift, in accordance with the overall decline in the CSFH over the past ~10 Gyr (as shown by Madau & Dickinson 2014; Bellstedt et al. 2020a). This is evident when observing the changing SFR within galaxies at a fixed stellar mass over time: At a lookback time of 11 Gyr, galaxies with $M_* = 10^{10}\,{\rm M_\odot}$ have SFRs $\sim 10\,{\rm M_\odot\,yr^{-1}}$, whereas by 1 Gyr lookback time the star-forming galaxies in this mass range have SFRs $\sim 1\,{\rm M_\odot\,yr^{-1}}$.

We do not apply a selection in Fig. 8 based on sSFR, which means the plot features both star-forming and quenched galaxies. The presence of quenched/quenching systems is first apparent at a lookback time of 10 Gyr, where the first high-mass galaxies – $\log(M_*/{\rm M_\odot}) > 10$ – are becoming dark purple in colour, corresponding to $\log({\rm SFR/M_\odot\,yr^{-1}}) < -1$. By 7–8 Gyr lookback time, these objects become much more prevalent. The presence of these system causes a noticeable visual dilution of the SFR trends, which is maximized by 1 Gyr, when a significant fraction of the high-mass galaxies is quenched. As such, there is only a minimal trend with SFR in the 1 Gyr panel of this plot.

To assess how the trends in Fig. 8 change with sSFR, we present the same parameter space coloured by sSFR in Fig. 9. In addition to the running median in each subpanel of the total population, we include the running median for the star-forming population (as defined by $\log({\rm sSFR/yr^{-1}}) > -11.31$, Davies et al. 2019). We highlight here that the consequence of including passive systems in our analysis are essentially negligible.

Table 3. Coefficients to describe the time evolution of the MZR median and scatter, as used in equation (6).

| | Median evolution coefficients | | | | | |
	$a_{i,0}$	$a_{i,1}$	$a_{i,2}$	$a_{i,3}$	$a_{i,4}$	$a_{i,5}$
$a_{0,j}$	-1.67	6.32×10^{-2}	-3.08×10^{-2}	3.62×10^{-3}	-1.57×10^{-4}	0
$a_{1,j}$	4.25×10^{-1}	-3.69×10^{-3}	3.86×10^{-3}	-4.16×10^{-4}	2.38×10^{-5}	0
$a_{2,j}$	1.42×10^{-2}	-1.38×10^{-1}	5.24×10^{-2}	-6.88×10^{-3}	3.37×10^{-4}	-3.78×10^{-6}
$a_{3,j}$	-2.67×10^{-2}	-7.20×10^{-2}	3.58×10^{-2}	-6.00×10^{-3}	4.20×10^{-4}	-1.05×10^{-5}
	Scatter evolution coefficients					
	$a_{i,0}$	$a_{i,1}$	$a_{i,2}$	$a_{i,3}$	$a_{i,4}$	$a_{i,5}$
$a_{0,j}$	-3.65×10^{-1}	-5.85×10^{-3}	3.56×10^{-4}	1.04×10^{-4}	1.61×10^{-7}	0
$a_{1,j}$	-3.34×10^{-1}	4.30×10^{-2}	-8.47×10^{-3}	9.06×10^{-4}	-2.92×10^{-5}	0
$a_{2,j}$	-1.66×10^{-1}	4.26×10^{-2}	-3.29×10^{-2}	8.93×10^{-3}	-8.92×10^{-4}	2.99×10^{-5}
$a_{3,j}$	-1.05×10^{-2}	5.16×10^{-3}	-1.62×10^{-2}	4.92×10^{-3}	-5.05×10^{-4}	1.72×10^{-5}
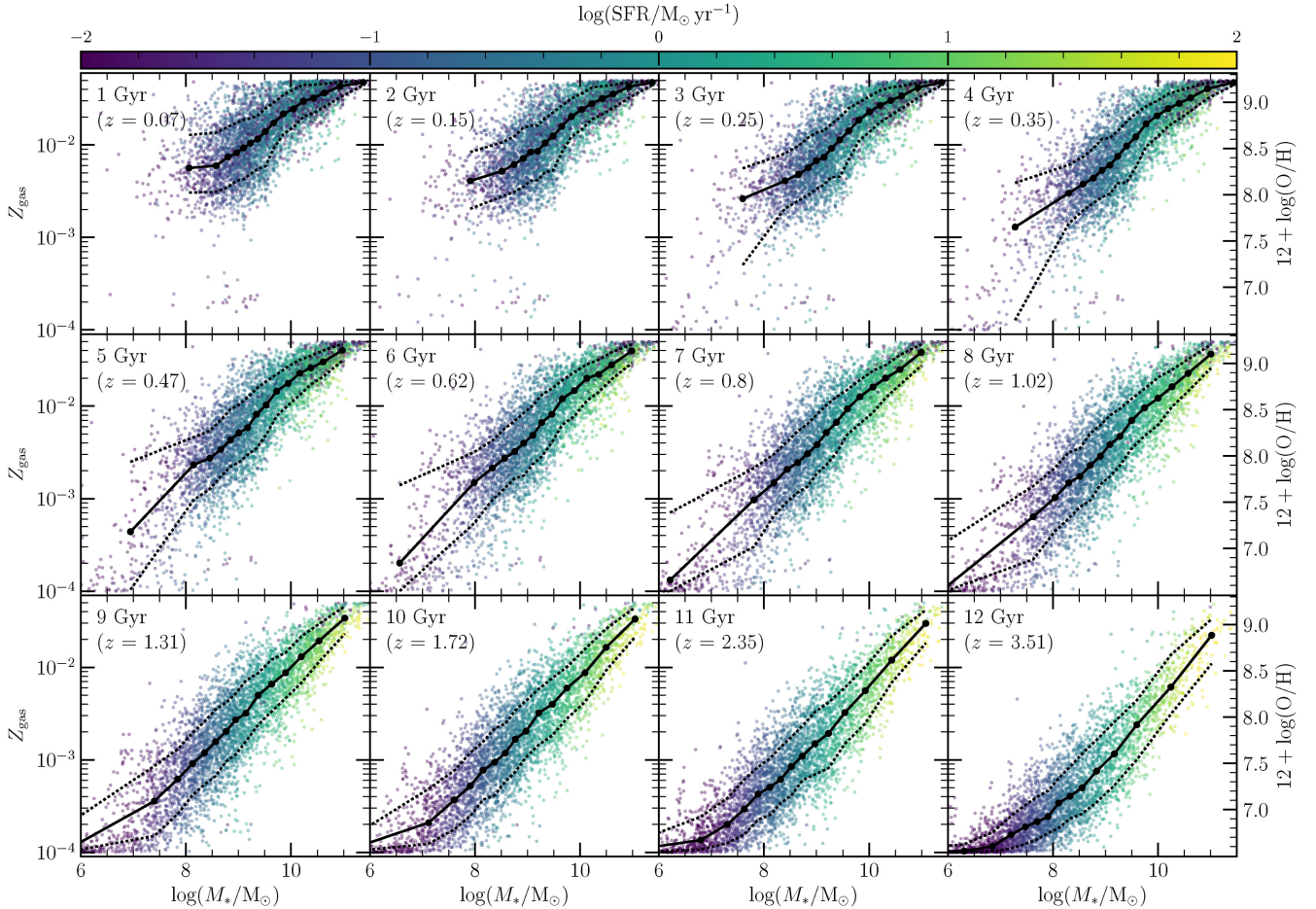
Figure 8. The mass–metallicity relation in 1 Gyr intervals for the same galaxy population, coloured by the SFR at that epoch. As in Fig. 6, the median and 1σ MZR are shown in the solid and black-dashed lines in each subpanel.

An important caveat to note, which will impact the trends shown in Fig. 8, is that we neglect the impact of potential historical mergers in this analysis. If a galaxy has undergone a major merger in the past, then instead of a single progenitor galaxy, there would be two. As a result, for this object we are likely to have overestimated the mass (and, by extension, the SFR) of the galaxy at high lookback times. It is therefore possible that the high-mass high-SFR objects in the panels at high lookback times could be an artefact of this assumption, and that in reality such galaxies do not exist.

5 THE MASS–METALLICITY–SFR PLANE

It is now understood that the MZR is a projection of the three-dimensional mass–metallicity–SFR plane (Magrini et al. 2012; Lara-López et al. 2013c; Peeples & Somerville 2013). Using the outputs of SED fitting, we can present our data in this plane, with a broad range of stellar masses, metallicities, and SFRs.

To fit the plane in three dimensions, we employ the R package HYPERFIT (Robotham & Obreschkow 2015). When fitting a plane to our derived values of stellar mass, SFR, and metallicity, we have converted our absolute metallicity values to oxygen abundances, for
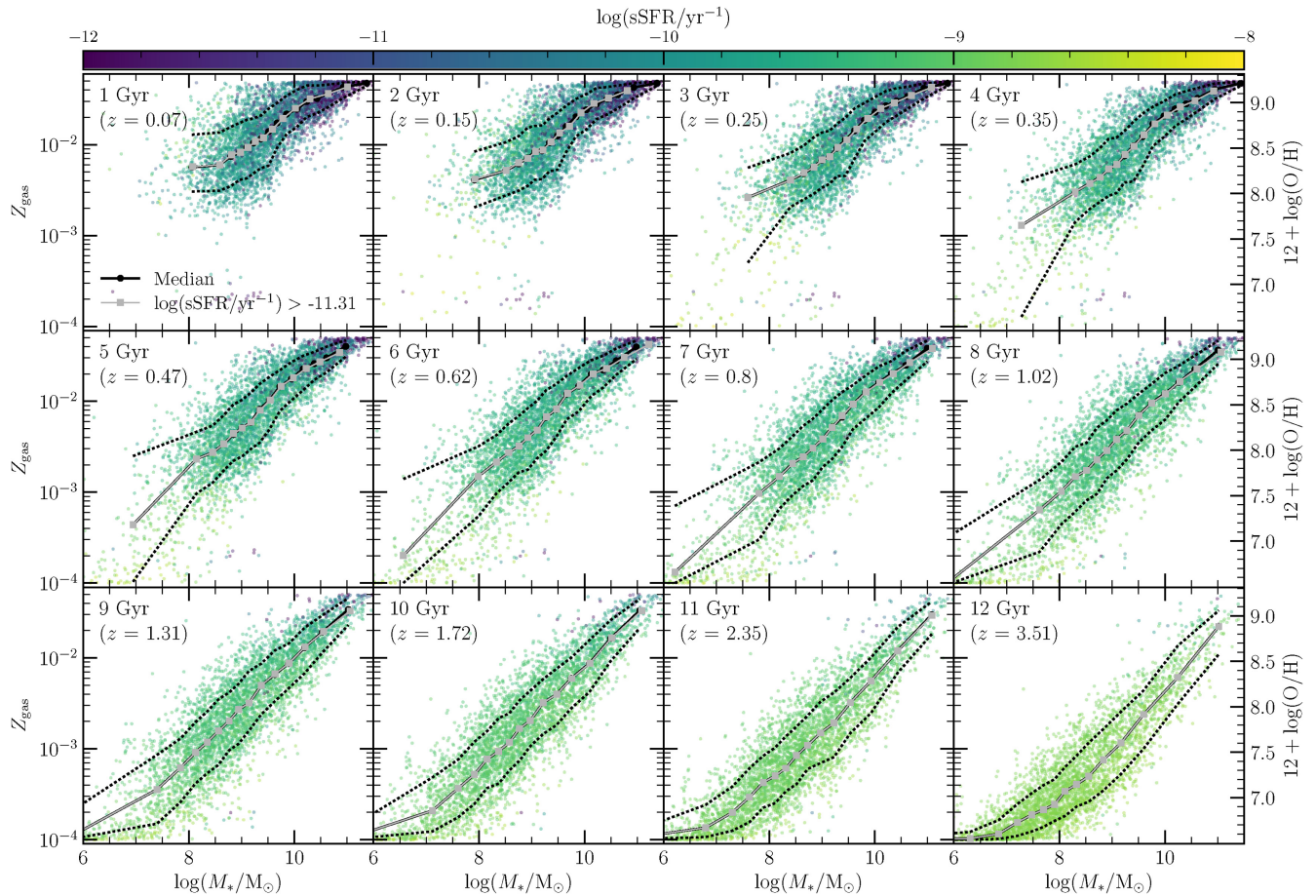
Figure 9. The mass–metallicity relation in 1 Gyr intervals for the same galaxy population, coloured by the sSFR at that epoch. As in Fig. 6, the median and 1σ MZR are shown in the solid and black-dashed lines in each subpanel. The running median for the star-forming population (as defined by $\log(\mathrm{sSFR/yr^{-1}}) > -11.31$) is shown in grey.

the sake of easy comparison against observations. We additionally restrict our fit to galaxies with specific SFR values $\log(\mathrm{sSFR/yr^{-1}}) > -11.5$, so that we do not include quenched and quenching galaxies in our fit.

Our fit to the $z \sim 0$ mass–metallicity–SFR plane can be described with the functional form:

$$[12 + \log(\mathrm{O/H})] = \alpha \left(\frac{\mathrm{SFR}}{\mathrm{M_\odot\, yr^{-1}}} \right) + \beta \log \left(\frac{\mathrm{M_*}}{\mathrm{M_\odot}} \right) + \gamma, \qquad (9)$$

where

$\alpha = -1.426$

$\beta = 1.628$

$\gamma = -7.592.$

In Fig. 10, we plot the plane as fitted by Lara-López et al. (2013c) in orange. This plane intersects with ours at high masses and metallicities, however, it diverges in the lower mass and lower metallicity regimes. This is a consequence of the Lara-López et al. (2013c) sample, which focused on galaxies with stellar masses of $\sim 10^9 \mathrm{M_\odot}$, and as such a discrepancy of the plane at low stellar masses is unsurprising. We additionally present the three-dimensional fitted surface by Mannucci et al. (2010) in cyan, which only covers the high-mass, high-metallicity regime of the three-dimensional space. Similarly, we show the three-dimensional surface fitted by Curti et al. (2020) is shown in blue. We find that the surfaces from Mannucci

et al. (2010) and Curti et al. (2020) have a signficantly stronger tilt than ours, and also saturate at a much lower metallicity that the maximum metallicities we derive. The structure in the surfaces fitted by both Mannucci et al. (2010) and Curti et al. (2020) do not visually provide a better fit to the mass–metallicity–SFR relation we derive than a simple plane does. In fact, the plane derived by Lara-López et al. (2013c) is in much better agreement with the distribution of our derived data than either of the two other surfaces.

We highlight here that, because the Zfinal parameter is a free parameter in our implementation, the planar structure of our mass–metallicity–SFR plane was not prescribed by our evolving metallicity model.

5.1 Evolving plane

Our forensic analysis also allows us to measure the evolution of the mass–metallicity–SFR plane. The evolution of the planes with cosmic time is shown in Fig. 11. Each subpanel shows the mass–metallicity–SFR plane in a specific lookback time interval. The projection of the points on to each of the three related two-dimensional spaces (i.e. mass–metallicity, mass–SFR, and SFR–metallicity) is indicated on the edges of the plot with the blue/green points. A consequence of the functional form of the SFH is that the early build-up of galaxies in this sample is very similar due to the truncation of each SFH at early times. As a result, the main sequence

(the two-dimensional projection at the bottom of each subpanel) can be seen to be extremely narrow at high lookback times. At this epoch, the resulting stellar masses are limited by the SFR, and hence they are very tightly correlated.

The plane can be seen to evolve with cosmic time. In particular, with time the plane tilts towards high-mass, low-SFR systems with increasing time. The low-metallicity portion of the plane tilts towards higher SFRs, however, this is unconstrained at recent times due to the lower mass limit of our observations. This evolution is most notable in lookback times beyond 7 Gyr, in the 9 and 11 Gyr panels. We highlight that, as mentioned above, this epoch is highly influenced by our choice of SFH parametrization, and hence the measurement of this evolution must be assessed with this caveat.

The evolution of the plane with lookback time can be expressed as

$$[12 + \log(\text{O/H})] = \alpha(t_{\text{lb}}) \log \left(\frac{\text{SFR}}{\text{M}_\odot \text{yr}^{-1}} \right)$$
$$+ \beta(t_{\text{lb}}) \log \left(\frac{M_*}{\text{M}_\odot} \right) + \gamma(t_{\text{lb}}), \tag{10}$$

where

$\alpha(t_{\text{lb}}) = -0.9202 + 0.2916t_{\text{lb}} - 0.05082t_{\text{lb}}^2$,

$\beta(t_{\text{lb}}) = 1.198 - 0.2658t_{\text{lb}} + 0.04884t_{\text{lb}}^2$, and

$\gamma(t_{\text{lb}}) = -2.785 + 2.268t_{\text{lb}} - 0.4267t_{\text{lb}}^2$.

Here, t_{lb} is the lookback time in Gyr. The scatter in the plane is seen to evolve as

$\sigma(t_{\text{lb}}) = 0.427 - 0.02979t_{\text{lb}} + 0.0026182t_{\text{lb}}^2$.

This fit to the evolving plane parameters is shown in Fig. B1.

In an analysis of stellar masses, SFRs and metallicities spanning a wide redshift range, Mannucci et al. (2010) concluded that there was no evolution in the mass–metallicity–SFR plane from $z = 0$ to $z = 2.5$. This is in constrast with the subtle evolution that we derive in our analysis. Note that the study by Mannucci et al. (2010) focused on galaxies with $M_* > 10^{9.5} \text{M}_\odot$, and hence the stellar mass range may well have been too small to detect the evolution that we infer. Lara-López et al. (2010) also concluded in their analysis that there was no evidence for an evolution in the plane.

6 DISCUSSION

There are numerous caveats associated with the analysis presented in this work, both in the manner in which our values are compared against the observed metallicity values, and also inherent to the modelling method we have applied. We discuss these issues in this section.

A major challenge associated with any comparison of modelled metallicities to spectroscopically derived values is that there are significant biases that accompany spectroscopic measurements. This is the result of relying on different emission lines between different data sets, different strong line parameters, and finally different calibration methods to transform the strong line measurements into abundances. This challenge was highlighted by Sánchez et al. (2019b), who circumvented some of these systematics by electing to use 11 different calibrators in their analysis, rather than only selecting one. The resulting MZRs that they derive not only vary in normalization by up to 0.5 dex, but the resulting slopes also vary dramatically. A similar depiction of this challenge was presented earlier by Kewley & Ellison (2008), López-Sánchez & Esteban (2010), and López-Sánchez et al. (2012).
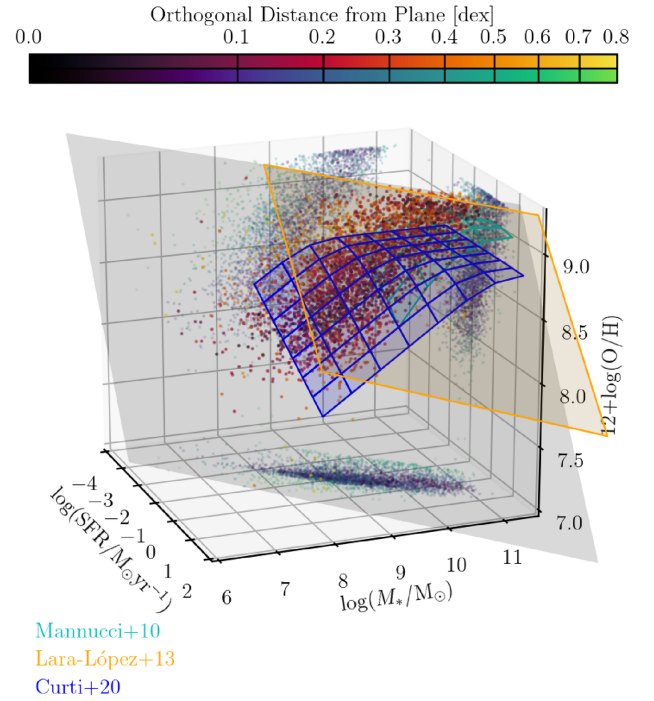
Figure 10. The mass–metallicity–SFR plane as determined by the fitted parameters. The values in three-dimensional space, as well as their projections on to the three two-dimensional spaces, are all plotted. Colours indicate the orthogonal distance of the point to the place. three-dimensional values are shown in red/orange hues, while their corresponding projections are shown in blue/green hues, to allow them to be distinguished. The fit made to the plane by Lara-López et al. (2013c) is shown in orange. The tilt of the plane in this fit is significantly larger than the tilt we measure. Note that the disagreement is greatest at low metallicities, where observations are sparse. The cyan surface is the measurement by Mannucci et al. (2010), and the blue surface is the fit to the three-dimensional space by Curti et al. (2020). Here, the surface has a significantly larger tilt, and the saturation metallicity is lower than the maximum metallicities we recover. See online version for three-dimensional video.

Not only is measuring the absolute oxygen abundance challenging, but there is extra uncertainty associated with the conversion of oxygen abundance to total metallicity, which is the quantity we model in our analysis. This was highlighted by Gallazzi et al. (2005) who, using stellar population models to infer the total metallicities of SDSS spectra, compared the total metallicities against oxygen abundances determined for the same sample as Tremonti et al. (2004). While there was a clear correlation between these parameters, the scatter was also significant. Consequently, scatter between the observed and our modelled metallicities could also arise from the scaling between oxygen abundances and total metallicities.

A key simplification in the SED fitting approach that we have taken is that mergers that may have occurred in a galaxy's past are ignored. It would be reasonable to assume that a massive galaxy that only has a single progenitor that has epochs of extremely high star formation would likely have different metallicities to a massive galaxy that formed as the result of a series of major mergers, each progenitor of which may have had smaller SFRs. In our approach, we would not differentiate such systems, as we simply model the history of all the stars currently present in a galaxy. In an analysis focusing on the metallicities of merging galaxies, Horstman et al. (2021) found that galaxies in mergers had suppressed metallicities versus isolated galaxies at the same stellar mass at $2 < z < 2.7$, highlighting that mergers will definitely have an impact on the
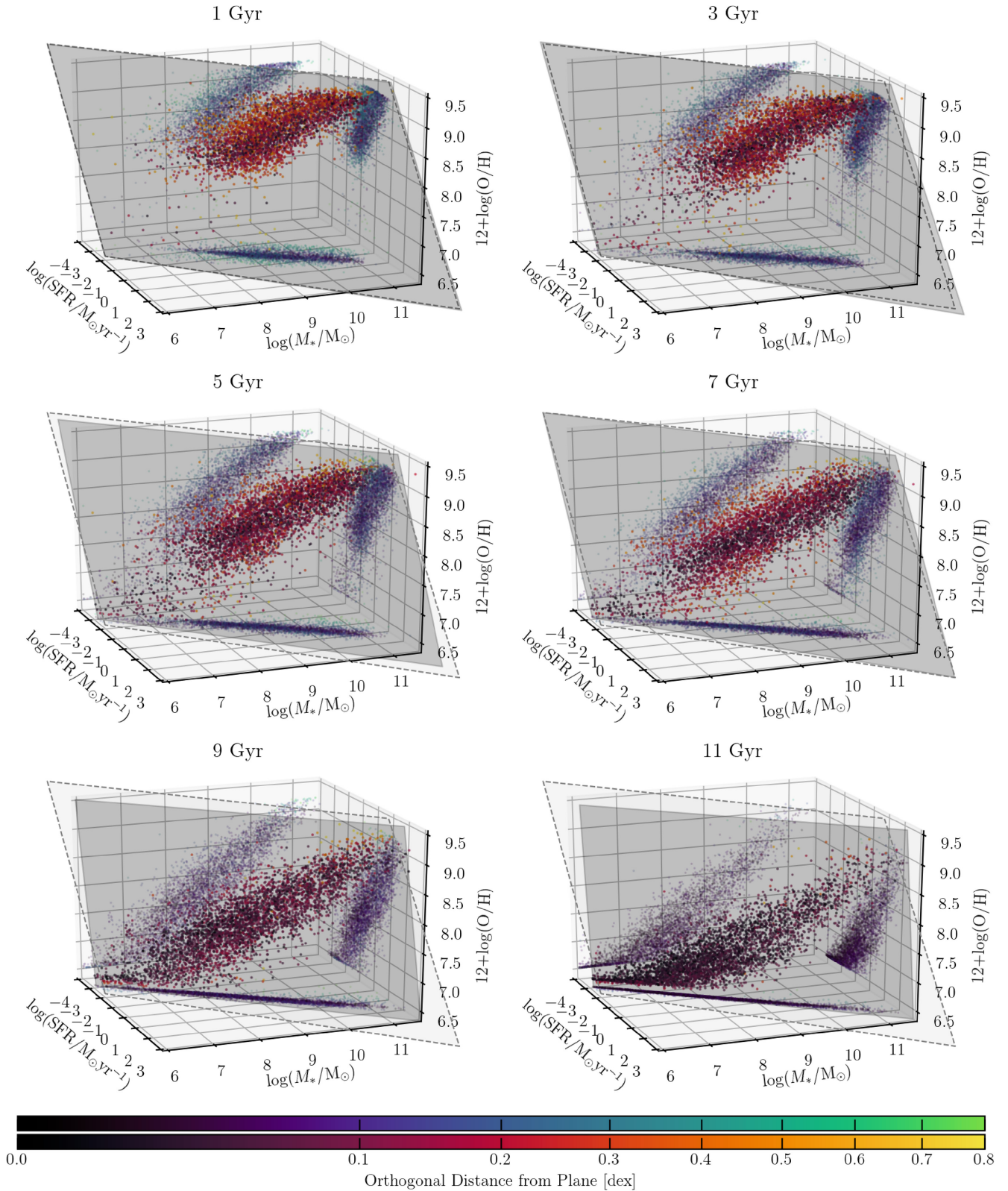
Figure 11. The mass–metallicity–SFR plane at regular lookback time intervals between 1 and 11 Gyr. The values themselves are plotted in red/orange hues, while their corresponding projections on to the three two-dimensional spaces is plotted in blue/green hues, for clarity. As in Fig. 10, points are coloured according to their orthogonal distance from the fitted plane. The plane at each epoch is shown in solid grey, and for comparison the $z \sim 0$ plane is indicated with the dashed line for reference. The tilting of the mass–metallicity–SFR plane with cosmic time is evident. Additionally, it is clear that the scatter around the plane increases with cosmic time. See online version for three-dimensional video.

metallicity of an object. Additionally, the associated stellar masses and SFRs are likely to be biased to be high in our analysis. The extent of this bias would depend on the merger rates throughout cosmic time. A quantification of this potential bias is beyond the scope of this work, however, this is a likely explanation for the very high-mass objects presented in our work at earlier lookback times, that are higher mass than the typical galaxy observations at that epoch.

Furthermore, our SED-fitting approach does not explicitly account for any interactions between a galaxy and its surroundings, including phenomena such as feedback (either stellar or AGN) or any gas inflows. These are mechanisms that could have the effect of either enriching or diluting the metal content of gas in galaxies (as demonstrated by, for example, López-Sánchez et al. 2015). Some allowance for the occurrence of gas inflows is provided by the linear metallicity evolution implementation, unlike the closed-box model (which was utilized in the main body of Bellstedt et al. 2020a, see Robotham et al. 2020 for further details), however, this is not explicit. Work by Rupke, Kewley & Chien (2010) demonstrated that the metallicity gradients of galaxies undergoing interactions are significantly flatter than non-interacting galaxies caused by interaction-induced gas inflow. Note that in our approach we cannot model the spatially resolved nature of metallicity within galaxies. Chisholm, Tremonti & Leitherer (2018) looked at outflows as a mechanism for producing the shape of the MZR. Based on an extensive modelling of UV spectral observations for 7 local galaxies, Chisholm et al. (2018) determined that the metal outflow rate in galaxies linearly correlates with their stellar mass. Outflows were also invoked in the theoretical analysis of Dayal, Ferrara & Dunlop (2013) to describe the fundamental metallicity relation. Contrastingly, Calura et al. (2009) determined that winds were not required to explain the shape of the MZR, and that it could instead be explained by the lower star formation efficiency of lower mass systems. The analysis presented in this work suggests that an explicit treatment of outflows is not required to reproduce the general shape of the MZR.

An additional phenomenon that is unconsidered in our approach is that of galaxy starvation, in which the gas inflow to a galaxy is disrupted, leading to a gradual truncation of star formation within a galaxy. This process was discussed in detail by Peng, Maiolino & Cochrane (2015), who in particular noted that in a starvation scenario, the final star formation epoch is likely to be more chemically enriched than any previous star formation, due to the lack of dilution caused by gas inflows. This effect is predicted to be responsible for the discrepancy between *stellar* metallicities of passive and star-forming galaxies at fixed stellar mass (as presented by, for example, Peng et al. 2015; Trussler et al. 2020). Such a late-time acceleration of chemical enrichment of galaxies would not be considered in our approach.

While the present-day gas-phase metallicities of our galaxies are free parameters within our analysis, the metallicity at the beginning of each galaxy's history is fixed to the lower limit of the Bruzual & Charlot (2003) stellar population templates. It is conceivable that a galaxy forming later in the Universe is formed out of non-pristine gas, and therefore we might be underestimating its initial metallicity. Contrastingly, this lower metallicity limit of 10^{-4} is likely too high for galaxies that form very early in the Universe, where the formation gas is likely to be pristine. It is unknown to what extent these simplifications impact the metallicity evolution of galaxies that we infer, however, we point the reader to the discussion in Appendix A, where we highlight a potential consequence of the initial metallicity value on the assumption of linearity in the metallicity evolution prescription.

7 CONCLUSIONS

SED-fitting techniques are widespread in use, and are frequently employed to measure the stellar masses of galaxies. Historically, SED-fitting techniques have employed a simple parametrization of the SFH, and have typically modelled the corresponding ZH with a single value held constant over the history of the galaxy. This simplification has a significant impact on not only the star formation histories of galaxies (as discussed in Bellstedt et al. 2020a), but as a consequence the derived stellar masses can also be affected. Using the SED fitting code PROSPECT (Robotham et al. 2020), we have used a simple prescription to model an evolving metallicity for individual galaxies (used in Bellstedt et al. 2020a, to extract a cosmic SFH consistent with observational measurements) to derive the MZR for a sample of ∼4500 $z < 0.06$ galaxies form the GAMA survey.

Despite the caveats discussed in the previous section, the agreement that we observe between our inferred MZR and observations across a wide range of stellar masses, metallicities, and cosmic time is remarkable. We show the evolution of the median MZR with cosmic time, and show this evolution in a functional form in equation (6).

Furthermore, we use consistently derived metallicities, stellar masses, and SFRs to make a planar fit to the three-dimensional mass–metallicity–SFR space. We show that this space can be well described by a plane over cosmic time, and show that as galaxies increase in stellar mass and metallicity, and reduce the rate at which they are forming stars, the mass–metallicity–SFR plane tilts. This evolution mostly occurs at lookback times greater than 7 Gyr. We present the evolution of this plane in functional form in equation (10).

Combined with the accurate cosmic SFH derived using this implementation of PROSPECT (as presented in Bellstedt et al. 2020a), the analysis presented in this work reaffirms that galaxy stellar populations can be modelled accurately using SED fitting of broad-band photometry alone, if careful consideration is given to the evolution of the gas phase metallicity in addition to the evolution of the SFH.

used ASTROPY,[6] a community-developed core PYTHON package for astronomy (Astropy Collaboration 2013, 2018), PANDAS (McKinney 2010), and NUMPY (Harris et al. 2020).

8 DATA AVAILABILITY

The PROSPECT catalogue will be collated into a DMU, and can be accessed via a GAMA collaboration request.[7]

REFERENCES

Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481

Astropy Collaboration, 2013, A&A, 558, A33

Astropy Collaboration, 2018, AJ, 156, 123

Baldry I. K., Glazebrook K., 2003, ApJ, 593, 258

Bell E. F., de Jong R. S., 2000, MNRAS, 312, 497

Bellstedt S. et al., 2020a, MNRAS, 498, 5581

Bellstedt S. et al., 2020b, MNRAS, 496, 3235

Berg D. A. et al., 2012, ApJ, 754, 98

Bian F., Kewley L. J., Dopita M. A., 2018, ApJ, 859, 175

Brinchmann J., Charlot S., Heckman T. M., Kauffmann G., Tremonti C., White S. D. M., 2004, preprint (arXiv:astro–ph/0406220)

Brown J. S., Martini P., Andrews B. H., 2016, MNRAS, 458, 1529

Brown T., Cortese L., Catinella B., Kilborn V., 2018, MNRAS, 473, 1868

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Calabrò A. et al., 2017, A&A, 601, A95

Calura F., Pipino A., Chiappini C., Matteucci F., Maiolino R., 2009, A&A, 504, 373

Cameron A. J., Jones T., Yuan T., Trenti M., Bernard S., Henry A., Hoag A., Vulcani B., 2019, ApJ, 882, 116

Chabrier G., 2003, PASP, 115, 763

Charlot S., Longhetti M., 2001, MNRAS, 323, 887

Chisholm J., Tremonti C., Leitherer C., 2018, MNRAS, 481, 1690

Curti M., Mannucci F., Cresci G., Maiolino R., 2020, MNRAS, 491, 944

Davé R., Rafieferantsoa M. H., Thompson R. J., Hopkins P. F., 2017, MNRAS, 467, 115

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Davies L. J. M. et al., 2019, MNRAS, 483, 1881

Dayal P., Ferrara A., Dunlop J. S., 2013, MNRAS, 430, 2891

De Lucia G., Xie L., Fontanot F., Hirschmann M., 2020, MNRAS, 498, 3215

Denicoló G., Terlevich R., Terlevich E., 2002, MNRAS, 330, 69

Driver S. P. et al., 2011, MNRAS, 413, 971

Driver S. P., Robotham A. S. G., Bland-Hawthorn J., Brown M., Hopkins A., Liske J., Phillipps S., Wilkins S., 2013, MNRAS, 430, 2622

Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, ApJ, 672, L107

Erb D. K., Shapley A. E., Pettini M., Steidel C. C., Reddy N. A., Adelberger K. L., 2006, ApJ, 644, 813

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41

Gillman S. et al., 2021, MNRAS, 500, 4229

Harris C. R. et al., 2020, Nature, 585, 357

Henry A., Martin C. L., Finlator K., Dressler A., 2013a, ApJ, 769, 148

Henry A. et al., 2013b, ApJ, 776, L27

Horstman K. et al., 2021, MNRAS, 501, 137

Huang C. et al., 2019, ApJ, 886, 31

Hughes T. M., Cortese L., Boselli A., Gavazzi G., Davies J. I., 2013, A&A, 550, A115

Hunter J. D., 2007, Comput. Sci. Eng., 9, 90

James B. L., Koposov S., Stark D. P., Belokurov V., Pettini M., Olszewski E. W., 2015, MNRAS, 448, 2687

Jimmy, Tran K.-V., Saintonge A., Accurso G., Brough S., Oliva-Altamirano P., 2015, ApJ, 812, 98

Kewley L. J., Ellison S. L., 2008, ApJ, 681, 1183

Kobulnicky H. A., Kewley L. J., 2004, ApJ, 617, 240

Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, MNRAS, 426, 2142

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573

Lara-López M. A. et al., 2010, A&A, 521, L53

Lara-Lopez M. A. et al., 2013a, MNRAS, 433, L35

Lara-López M. A. et al., 2013b, MNRAS, 434, 451

Lara-López M. A., López-Sánchez Á. R., Hopkins A. M., 2013c, ApJ, 764, 178

Larson R. B., 1974, MNRAS, 169, 229

Lee H., Skillman E. D., Cannon J. M., Jackson D. C., Gehrz R. D., Polomski E. F., Woodward C. E., 2006, ApJ, 647, 970

Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, A&A, 500, 145

Liske J. et al., 2015, MNRAS, 452, 2087

López-Sánchez Á. R., 2010, A&A, 521, A63

López-Sánchez Á. R., Esteban C., 2010, A&A, 517, A85

López-Sánchez Á. R., Dopita M. A., Kewley L. J., Zahid H. J., Nicholls D. C., Scharwächter J., 2012, MNRAS, 426, 2630

López-Sánchez Á. R., Westmeier T., Esteban C., Koribalski B. S., 2015, MNRAS, 450, 3381

Ly C., Malkan M. A., Rigby J. R., Nagao T., 2016, ApJ, 828, 67

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Magrini L. et al., 2012, MNRAS, 427, 1075

Maiolino R. et al., 2008, A&A, 488, 463

Mannucci F. et al., 2009, MNRAS, 398, 1915

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

McClure R. D., van den Bergh S., 1968, AJ, 73, 1008

McKinney W., 2010, in van der Walt S., Millman J., eds, Proc. 9th Python in Science Conference. Austin, Texas, p. 51

McQuinn K. B. W. et al., 2020, ApJ, 891, 181

Mouhcine M., Baldry I. K., Bamford S. P., 2007, MNRAS, 382, 801

Nagao T., Maiolino R., Marconi A., 2006, A&A, 459, 85

Oesch P. A. et al., 2016, ApJ, 819, 129

Pagel B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, MNRAS, 189, 95

Peeples M. S., Somerville R. S., 2013, MNRAS, 428, 1766

Peeples M. S., Pogge R. W., Stanek K. Z., 2009, ApJ, 695, 259

Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192

Pettini M., Pagel B. E. J., 2004, MNRAS, 348, L59

Planck Collaboration XIII, 2016, A&A, 594, A13

Prescott M. et al., 2016, MNRAS, 457, 730

R Core Team, 2017, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/

Robotham A. S. G., Obreschkow D., 2015, Publ. Astron. Soc. Aust., 32, e033

Robotham A. S. G., Davies L. J. M., Driver S. P., Koushan S., Taranu D. S., Casura S., Liske J., 2018, MNRAS, 476, 3137

Robotham A. S. G., Bellstedt S., Lagos C. d. P., Thorne J. E., Davies L. J., Driver S. P., Bravo M., 2020, MNRAS, 495, 905

Rubin V. C., Ford W. K. J., Whitmore B. C., 1984, ApJ, 281, L21

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

Salpeter E. E., 1955, ApJ, 121, 161

Sánchez S. F. et al., 2013, A&A, 554, A58

Sánchez S. F. et al., 2019a, MNRAS, 482, 1557

Sánchez S. F. et al., 2019b, MNRAS, 484, 3042

Sanders R. L. et al., 2020, preprint (arXiv:2009.07292)

Savaglio S. et al., 2005, ApJ, 635, 260

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

Torrey P. et al., 2019, MNRAS, 484, 5587

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

Trussler J., Maiolino R., Maraston C., Peng Y., Thomas D., Goddard D., Lian J., 2020, MNRAS, 491, 5406

Weldon A., Ly C., Cooper M., 2020, MNRAS, 491, 2254

[6] http://www.astropy.org

[7] http://www.gama-survey.org/collaborate/

Yabe K. et al., 2014, MNRAS, 437, 3647

Yates R. M., Kauffmann G., Guo Q., 2012, MNRAS, 422, 215

Zahid H. J., Dima G. I., Kudritzki R.-P., Kewley L. J., Geller M. J., Hwang H. S., Silverman J. D., Kashino D., 2014a, ApJ, 791, 130

Zahid H. J. et al., 2014b, ApJ, 792, 75

Zenocratti L. J., De Rossi M. E., Lara-López M. A., Theuns T., 2020, MNRAS, 496, L33

## APPENDIX A: VALIDITY OF PROPORTIONALLY EVOLVING METALLICITIES

In order to assess the validity of our 'linearly evolving' metallicity evolution parametrization, we turn to outputs from the SAM SHARK (Lagos et al. 2018). For a sample of ∼6000 galaxies sampled from one subvolume of the $z = 0.07$ snapshot (Bravo et al., in preparation), we extract the mass build-up, and gas-phase metallicity histories for each galaxy. In order to test how well these galaxies are, on average, approximated by our model, we need to identify the extent to which the mass build-up and metallicity evolution are proportional to each other. To do this we scale the mass profiles to the metallicity profiles until the differences between them are minimised, and then we quantify the difference between these profiles in log space. If a galaxy has purely proportional evolution, then this value would be 0 at all epochs. Some examples of how this behaves for galaxies in SHARK are shown in Fig. A1. Note that the metallicity build-up does in general follow the stellar mass build-up fairly closely, with the greatest discrepancies typically displayed in the first few billion years of the galaxy's lifetime. Note that this discrepancy is apparent in the scaled Mass/Z ratio shown in the bottom panels of the figure.

Although the lower limit of gas-phase metallicity prescribed within SHARK is 10^{-7}, the value tends to be ∼10^{-4} as soon as star formation has started. This is close to the initial value of 10^{-4} that we are forced to implement due to the limit of the Bruzual & Charlot (2003) stellar population templates.
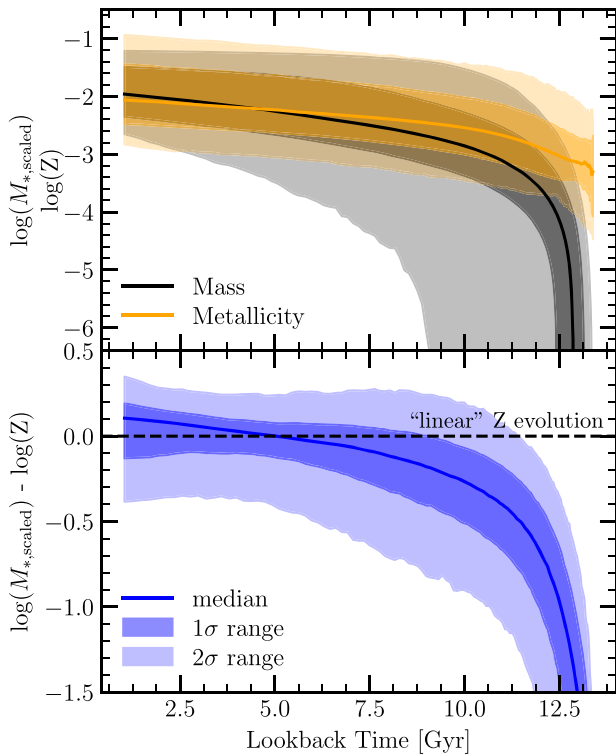
The result of this analysis over the full sample of galaxies is shown in Fig. A2. The relative mass and metallicity build-up are shown in the top panel, and their ratio is shown in in the bottom panel, with the 1σ and 2σ regions shown in shading. This plot shows that the proportional approximation is consistent with the evolution in SHARK in the most recent 8 Gyr, with deviations from this approximation occuring in the early Universe. We highlight that there are complexities in the nature of the modelling itself within SHARK that make this comparison difficult. Instantaneous recycling of metals will likely cause the metallicity to build-up more quickly than in reality, as an example. From the perspective of SED fitting, the age–metallicity degeneracy is greatest at this epoch where the discrepancy between the linear model and SHARK is at its maximum. Furthermore, because the stellar populations formed early on in the history of a galaxy contribute to such a small fraction of the galaxy's observed flux, the constraining power of the SED fitting process is very small for this epoch. These factors make distinguishing between metallicity evolution models in the early Universe very difficult. Further work would be required with a larger range of semi-analytic and cosmological simulations to properly conclude the extent to which 'linearity' is physical, however, such an analysis is well beyond the scope of this work.

We reiterate that the implementation of the linear metallicity evolution is a significant improvement over models in which the metallicity is assumed to be constant over the duration of a galaxy's history. As discussed in Section 3, the linear metallicity evolution model is currently favourable over the closed-box model (presented in detail in Bellstedt et al. 2020a), as the resulting metallicity
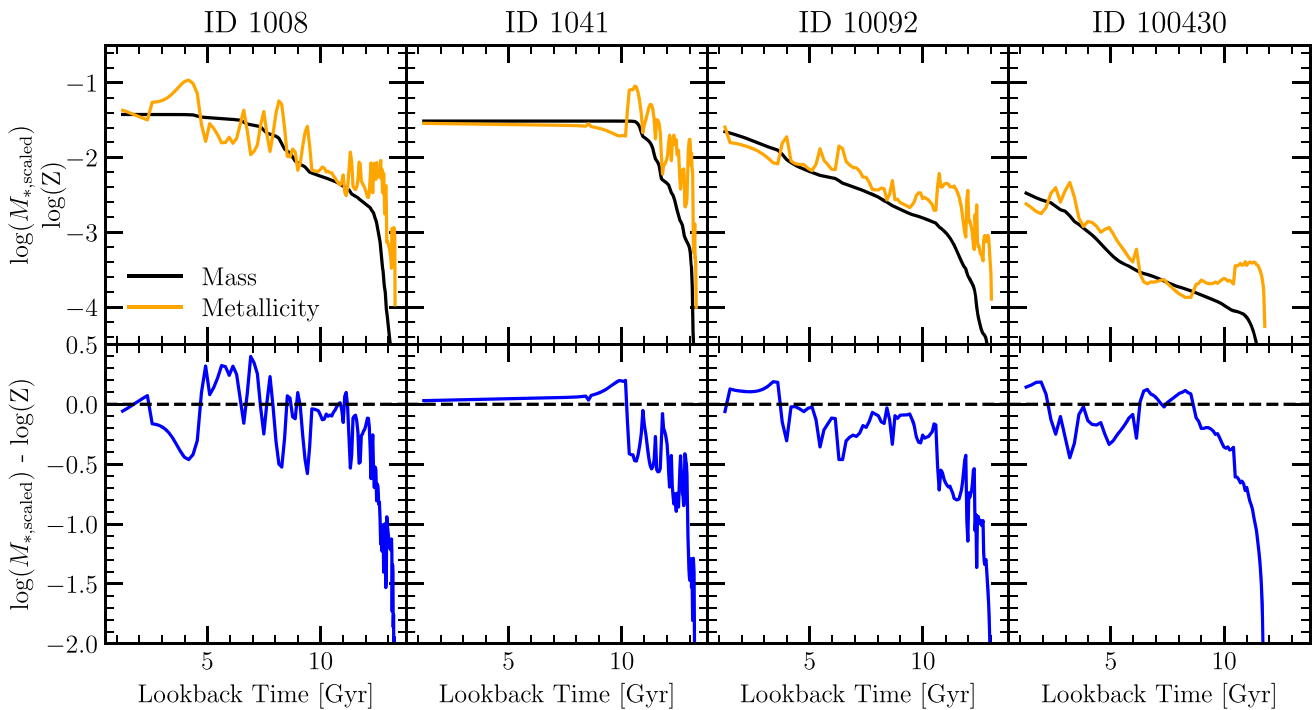
Figure A1. The relative stellar mass and metallicity build-up of four example SHARK galaxies (top panel), and the resulting ratio between these (bottom). The black-dashed line indicates proportional metallicity evolution.
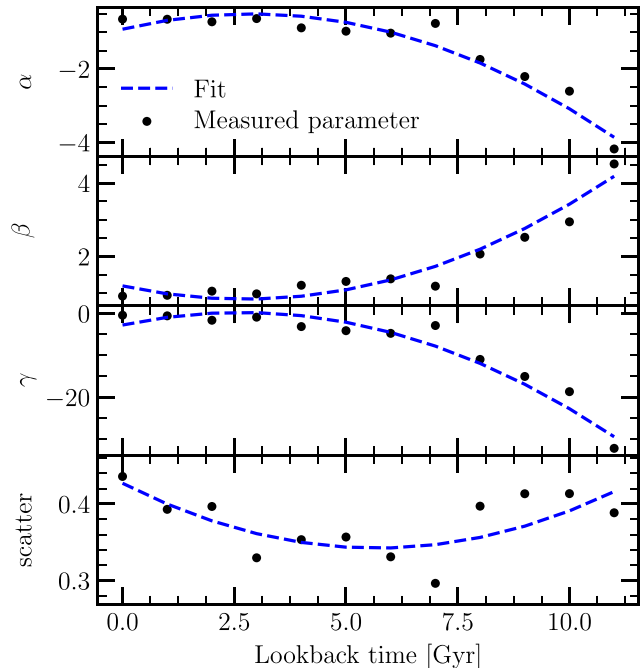
evolution, and therefore this is straightforward to adapt in the future if desired.

APPENDIX B: FITTING THE EVOLVING PLANE

The fit to the evolving plane parameters, as presented by equation (10) is presented in Fig. B1. Note that the evolution in the parameters is only significant beyond a lookback time of 7 Gyr.

Figure A2. Top: Overall relative build-up of mass (black) and metallicity (orange) for a sample of ∼6000 galaxies from the semi-analytic model SHARK. Bottom: Distribution of the corresponding stellar mass build-up to gas-phase metallicity evolution ratios (blue, with shaded regions indicating the 1σ and 2σ ranges). The black-dashed line indicates the assumption of our proportional metallicity evolution (given by the Zfunc_massmap_lin function). Note that SHARK is consistent with linear evolution to lookback times of ∼8 Gyr.

distributions are more physical. This implies that the most recent epoch of evolution (to which SED fitting is most sensitive), is better modelled by linear metallicity evolution than the closed box. PROSPECT can be implemented with any model of metallicity

Figure B1. Measured parameters describing the fit to the plane in each epoch between 1 and 11 Gyr lookback times. The fit, showing the time evolution of the plane and presented in equation (10) is shown in blue.

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