LJMU Research Online

Curti, M, Maiolino, R, Curtis-Lake, E, Chevallard, J, Carniani, S, D'Eugenio, F, Looser, TJ, Scholtz, J, Charlot, S, Cameron, A, Übler, H, Witstok, J, Boyett, K, Laseter, I, Sandles, L, Arribas, S, Bunker, A, Giardino, G, Maseda, MV, Rawle, T, Del Pino, BR, Smit, R, Willott, CJ, Eisenstein, DJ, Hausen, R, Johnson, B, Rieke, M, Robertson, B, Tacchella, S, Williams, CC, Willmer, C, Baker, WM, Bhatawdekar, R, Egami, E, Helton, JM, Ji, Z, Kumari, N, Perna, M, Shivaei, I and Sun, F

 JADES: Insights into the low-mass end of the mass-metallicity-SFR relation at 3 < z < 10 from deep JWST/NIRSpec spectroscopy*

http://researchonline.ljmu.ac.uk/id/eprint/23286/

Article

For more information please contact researchonline@ljmu.ac.uk

Astronomy & Astrophysics

JADES: Insights into the low-mass end of the mass–metallicity–SFR relation at 3 < z < 10 from deep JWST/NIRSpec spectroscopy[⋆]

Mirko Curti[1,2,3], Roberto Maiolino[2,3], Emma Curtis-Lake[4], Jacopo Chevallard[5], Stefano Carniani[6], Francesco D'Eugenio[2,3], Tobias J. Looser[2,3], Jan Scholtz[2,3], Stephane Charlot[7], Alex Cameron[5], Hannah Übler[2,3], Joris Witstok[2,3], Kristian Boyett[8,9], Isaac Laseter[10], Lester Sandles[2,3], Santiago Arribas[11], Andrew Bunker[5], Giovanna Giardino[12], Michael V. Maseda[10], Tim Rawle[13], Bruno Rodríguez Del Pino[11], Renske Smit[14], Chris J. Willott[15], Daniel J. Eisenstein[16], Ryan Hausen[17], Benjamin Johnson[16], Marcia Rieke[18], Brant Robertson[19], Sandro Tacchella[2,3], Christina C. Williams[20], Christopher Willmer[18], William M. Baker[2,3], Rachana Bhatawdekar[13,21], Eiichi Egami[18], Jakob M. Helton[18], Zhiyuan Ji[18], Nimisha Kumari[22], Michele Perna[11], Irene Shivaei[18], and Fengwu Sun[18]

(Affiliations can be found after the references)

ABSTRACT

We analysed the gas-phase metallicity properties of a sample of low-stellar-mass ($\log M_\star/M_\odot \lesssim 9$) galaxies at $3 < z < 10$ observed with JWST/NIRSpec as part of the JADES programme in its deep GOODS-S tier. By combining this sample with more massive galaxies at similar redshifts from other programmes, we study the scaling relations between stellar mass ($M_\star$), oxygen abundance (O/H), and star-formation rate (SFR) for 146 galaxies spanning three orders of magnitude in stellar mass and out to the epoch of early galaxy assembly. We find evidence for a shallower slope at the low-mass end of the mass–metallicity relation (MZR), with $12 + \log(O/H) = (7.72 \pm 0.02) + (0.17 \pm 0.03) \log(M_\star/10^8 \, M_\odot)$, in good agreement with the MZR probed by local analogues of high-redshift systems, such as the 'Green Pea' and 'Blueberry' galaxies. The inferred slope is well matched by models including 'momentum-driven' supernova (SN) winds, suggesting that feedback mechanisms in dwarf galaxies (and at high z) might be different from those in place at higher masses. The evolution in the normalisation is observed to be relatively mild compared to previous determinations of the MZR at $z \sim 3$ ($\sim 0.1 - 0.2$ dex across the explored mass regime). We observe a deviation from the local fundamental metallicity relation (FMR) for our sample at high redshift, especially at $z > 6$, with galaxies significantly less enriched than predicted given their $M_\star$ and SFR (with a median offset in $\log(O/H)$ of ~ 0.5 dex, significant at $\sim 5\sigma$). These observations are consistent with an enhanced stochasticity in the gas accretion and star-formation history of high-redshift systems, prompting us to reconsider the nature of the relationship between $M_\star$, O/H, and SFR in the early Universe.

Key words. ISM: abundances – Galaxy: evolution – Galaxy: fundamental parameters – galaxies: abundances – galaxies: high-redshift – galaxies: ISM

1. Introduction

The *James Webb* Space Telescope (JWST) has already begun to revolutionise our view of the high-redshift Universe, opening a new window onto early galaxy formation. In particular, the high-sensitivity Near Infrared Spectrograph (NIRSpec, Jakobsen et al. 2022; Ferruit et al. 2022; Böker et al. 2023) now allows to spectroscopically characterise the interstellar medium (ISM) properties and constrain the nature of the ionising spectra in primordial galaxies. In this context, the study of the gas-phase metallicity[1] via rest-frame optical spectroscopy, which has been limited for decades to $z \lesssim 3.5$ due to the intrinsic limitations of ground-based observatories, is now possible out to $z \sim 10$. Such observations are providing precious constraints on cosmological simulations and chemical evolution models, enabling scientists to chart the processes that shaped the formation of galaxy structures in the early Universe (Somerville & Davé 2015; Maiolino & Mannucci 2019). For instance, characterising the scaling relation between stellar mass and gas-phase metallicity in galaxies (the so-called mass–metallicity relation (MZR); Lequeux et al. 1979; Tremonti et al. 2004; Lee et al. 2006; Yates et al. 2020; Baker & Maiolino 2023) at high redshift is critical in order to constrain the processes regulating the growth of early galaxies, as this relation is shaped by the interplay between gas accretion, star formation, metal enrichment, and outflows driving the baryon cycle. Furthermore, as a consequence of the long-lasting interplay between these different processes, the metallicity of galaxies has been observed to also correlate with different galactic properties. In particular, an anti-correlation between metallicity and star-formation rate (SFR) at fixed stellar mass has been observed and characterised in detail on both global and local scales in what is usually referred to as the fundamental metallicity relation (FMR; Ellison et al. 2008; Mannucci et al. 2010; Yates et al. 2012; Salim et al. 2015; Telford et al. 2016; Curti et al. 2020b; Baker et al. 2023).

⋆ The high-level data products exploited in this paper for the deep tier of the JADES-GTO Programme in GOODS-South, namely the redshifts, stellar masses, star-formation rates, and metallicities, are reported in Table C.1. Fully reduced spectra and emission line fluxes are available through the MAST database at https://archive.stsci.edu/hlsp/jades, and are described in Bunker et al. (2023a).

[1] As traced by the oxygen abundance, and usually quoted as $12 + \log(O/H)$.

Several studies leveraged the new spectroscopic capabilities of JWST/NIRSpec to start characterising the properties of strong line emitters at $z > 3$, exploiting the data collected in the framework of Early Release Observations (Pontoppidan et al. 2022, ERO), Early Release Science (ERS), and Cycle 1 General Observer (GO) and Guaranteed Time Observations (GTO) programmes, such as GLASS (Proposal ID: 1324; Treu et al. 2022), Cosmic Evolution Early Release Science (CEERS, Proposal ID: 1345; Finkelstein et al. 2023), and JWST Advanced Deep Extragalactic Survey (JADES, Proposal ID: 1210; Eisenstein et al. 2023; Robertson et al. 2023; Curtis-Lake et al. 2023). Recent works have indeed investigated the ionisation properties of galaxies beyond $z = 3$, showing that these sources exhibit emission-line ratios consistent with hard ionising spectra and low metallicities (Mascia et al. 2023; Matthee et al. 2023; Sanders et al. 2023; Cameron et al. 2023b). These kinds of analyses have been pushed to some of the highest-redshift galaxies discovered, with Williams et al. (2023) reporting a $z = 9.5$ galaxy with a very high [O III]/[O II] ratio and relatively low metallicity, while Bunker et al. (2023b) recently reported the detection of multiple emission lines (including several high-ionisation metal lines) in the NIRSpec spectrum of the luminous $z \sim 10.61$ galaxy GN-z11 (Oesch et al. 2016). The wealth of emission lines detected in these high-z galaxy spectra can therefore be exploited to assess the chemical enrichment of their ISM.

In the past, several studies have investigated the evolution in the metallicity scaling relations out to $z \sim 3$, finding signatures of a clear trend of decreasing metallicity with redshift at fixed stellar mass (e.g. Shapley et al. 2005; Erb et al. 2006; Maiolino et al. 2008; Mannucci et al. 2009; Zahid et al. 2011, 2014; Henry et al. 2013a; Yabe et al. 2014; Wuyts et al. 2014; Guo et al. 2016; Sanders et al. 2021; Topping et al. 2021). However, when reported to the framework of the FMR, no evidence of redshift evolution has been observed, suggesting that, on average, galaxies evolved through smooth secular processes over the past ~ 10 Gyr driven by the interplay of gas flows, star formation, and metal enrichment, which is reflected in this latter scaling relation (Mannucci et al. 2010; Belli et al. 2013; Nakajima & Ouchi 2014; Maier et al. 2014; Salim et al. 2015; Hirschauer et al. 2018; Sanders et al. 2021; Hayden-Pawson et al. 2022).

More recent studies attempted to push the investigation of the chemical properties of galaxies up to $z \sim 10$, either from the detection of emission lines in individual sources, or from the analysis of composite spectra, finding initial evidence for a relatively mild evolution of the oxygen abundance with redshift at fixed stellar mass (e.g. Curti et al. 2023; Schaerer et al. 2022; Arellano-Córdova et al. 2022; Taylor et al. 2022; Trump et al. 2023; Rhoads et al. 2023; Langeroodi et al. 2023; Matthee et al. 2023; Heintz et al. 2023; Nakajima et al. 2023; Shapley et al. 2023a). Despite these efforts, the majority of studies focused on relatively high-stellar-mass systems, whereas probing the evolution of the metallicity scaling relations at low stellar masses, a regime strongly sensitive to feedback processes due to the shallower gravitational potential, remained challenging (Wuyts et al. 2012; Henry et al. 2013b), that is until the arrival of JWST (Li et al. 2023).

Moreover, we remind the reader that the vast majority of such metallicity determinations are based on strong emission line ratios calibrated on a local sample of galaxies (Pettini & Pagel 2004; Maiolino et al. 2008; Curti et al. 2017) or on local analogues mimicking the physical conditions of high-z sources (Bian et al. 2018; Nakajima et al. 2022), and the applicability of these methods at high redshift still needs to be carefully assessed, as different ionisation properties in the ISM of high-z galaxies could bias the metallicity measurement (Kewley et al. 2013; Steidel et al. 2014; Strom et al. 2017; Sanders et al. 2018, 2023; Cameron et al. 2023b). Nonetheless, the rate of auroral line detections at high z is now rapidly increasing thanks to the JWST, and novel attempts to recalibrate the classical strong line diagnostics for the high-redshift Universe are being provided (Sanders et al. 2024).

In this paper, we aim to further explore the cosmic evolution of the metallicity scaling relations, leveraging deep spectroscopy with JWST/NIRSpec to probe the low-stellar-mass regime $M_\star \approx 10^{6.5} - 10^{9.5}\,M_\odot$ from $z = 3$ out to the highest redshift in which rest-frame optical nebular lines are accessible to NIRSpec ($z \sim 9.5$). By combining observations from JADES –which enables us to explore the very low-mass regime down to $M_\star \sim 10^{6.5}$– with existing datasets, we assessed the evolution of the MZR slope and normalisation, and its implications on the physical processes in place in early galaxies. Moreover, we tested the FMR framework down to the epoch when the Universe was $\lesssim 500$ Myr old.

The paper is organised as follows. In Sect. 2, we describe observations, data reduction, and spectral fitting procedures. In Sect. 3, we present the analysis we performed to derive the physical quantities of interest; that is, $M_\star$, SFR, and metallicity. In Sect. 4, we describe how we assessed the MZR at $z > 3$, combining our JADES sample with galaxies at higher masses drawn from the CEERS programme, and in Sect. 5 we investigate any evolution in the framework of the fundamental metallicity relation, and discuss possible inferences about the interplay between the gas flows and star formation regulating early galaxy assembly. In Sect. 6, we summarise our conclusions. Throughout this paper, we adopt a Planck Collaboration VI (2020) cosmology, a Chabrier (2003) initial mass function, and a solar metallicity, that is, $12+\log(O/H) = 8.69$ (Allende Prieto et al. 2001; Asplund et al. 2009).

2. Observations and data processing

2.1. Observations

The data presented in this paper were obtained via multi-object spectroscopy observations conducted with the micro-shutter assembly (MSA) of NIRSpec (Jakobsen et al. 2022; Ferruit et al. 2022; Böker et al. 2023) on the JWST. Observations were performed in three visits carried out between 21 and 25 October 2022 (Program ID: 1210; PI: N. Luetzgendorf) in the Great Observatories Origins Deep Survey South (GOODS-S) legacy field (Giavalisco et al. 2004), as part of one of the 'deep' tiers of the JWST-GTO JADES programme (Eisenstein et al. 2023; Bunker et al. 2023a). Each visit consisted of 33 613 s of integration in the PRISM/CLEAR configuration (hereinafter simply 'PRISM'), and 8403 s integration in each of G140M/F070LP, G235M/F170LP, and G395M/F290LP (hereinafter 'medium-resolution gratings'). Across three visits, this totals 28 h of integration in the PRISM, providing continuous spectral coverage from 0.6 to 5.3 μm at $R \sim 30-300$, and ~ 7 h in each of the medium-resolution gratings, providing $R \sim 1000$ across the full spectral range of NIRSpec[2].

Observations within each visit were performed adopting a 3-shutter nodding pattern, with the central pointing of each visit

---

[2] Further 7 h were spent in the high-resolution G395H/F290LP grating, providing $R \sim 2700$ between $\sim 2.8 - 5.1$ μm; however, we are not exploiting such observations in this specific work.

dithered by <1 arcsec to sample different areas of the detector. A total of 253 unique targets were observed with the PRISM within the three pointings; among these, 67 targets were observed in all three MSA configurations, whereas 62 targets featured in two pointings, and the remaining 124 were observed for only one-third of the total exposure time. Each pointing had a bespoke MSA configuration, and target allocation was performed via the eMPT[3] software (Bonaventura et al. 2023) to maximise the number of targets in common between all the three pointings, with special attention for rare objects included in the highest priority classes.

We note that in the medium and high resolution modes, individual spectra are dispersed over a large area on the detector. To minimise the possibility of overlapping emission, in our grating observations we have isolated our highest-priority targets by closing the shutters of low-priority targets on the same row (i.e. targets that could cause overlapping spectra). For this reason, in the grating modes we observe only 198 unique targets.

2.2. Data reduction

Flux-calibrated 2D and 1D spectra have been produced using the pipeline developed by the ESA NIRSpec Science Operations Team (SOT) and the NIRSpec GTO Team. Most of the processing steps in the pipelines adopt the same algorithms included in the official STScI pipeline used to generate the MAST archive products (Ferruit et al. 2022). Initially, we processed the raw data (i.e, level 1a data from the MAST archive) with a ramp-to-slope pipeline that estimates the count rate per pixel by using all unsaturated groups in the ramp, and which is optimised to reject cosmic rays on the basis of the slope of the individual ramps (for more details see Birkmann et al. 2011; Böker et al. 2012; Giardino et al. 2019; Ferruit et al. 2022). All the preprocessed count-rate images were then processed using a data reduction pipeline which includes both ESA NIRSpec SOT codes and bespoke developed NIRSpec GTO algorithms. We briefly outline here the main steps, while for a more detailed description we refer to Bunker et al. (2023a) and to a forthcoming paper of the NIRSpec/GTO collaboration (Carniani et al., in prep.). In brief, the pipeline consists of 11 main steps: (i) identification of non-target galaxies intercepting the open shutters; (ii) pixel-level background subtraction; (iii) extraction of the spectral trace of each target and wavelength and spatial coordinate assignments to each pixel in the 2D maps; (iv) pixel-to-pixel flat-field correction; v) spectrograph optics and disperser correction; (vi) absolute flux calibration; (vii) path-losses correction; (viii) rectification of 2D spectra; (ix) extraction of 1D spectra; (x) combination of 1D spectra generated from each exposure, nod, and pointing; (xi) combination of 2D spectra. Therefore, the data processing workflow returns both a combined 1D and 2D spectrum for each target. We note that the combined 1D spectra are not extracted from the combined 2D maps, but are the result of a weighted average of 1D spectra from all integrations, which allowed us to implement bad pixels and outliers rejection algorithms more efficiently. We adopt an irregular wavelength grid for the 1D and 2D spectra of the PRISM configuration, in order to avoid oversampling of the line spread function at short wavelengths ($\lambda \sim 1\,\mu m$). For the G140M/F070LP grating/filter configuration we extended the calibration of the spectrum up to $1.84\,\mu m$, taking into account the transmission filter throughput beyond the nominal wavelength range of this configuration ($0.70\,\mu m$–$1.27\,\mu m$). Finally, path-loss corrections are applied by

modelling galaxies as point-like sources, taking into account the relative intra-shutter position of each source by leveraging on a full forward-modelling of the telescope and instrument optical paths (Jakobsen et al. 2022; Ferruit et al. 2022).

2.3. Spectral fitting

Both continuum and line emission are modelled simultaneously adopting the penalised pixel fitting algorithm, PPXF (Cappellari 2017, 2023), which models the continuum as a linear superposition of simple stellar-population (SSP) spectra, using non-negative weights and matching the spectral resolution of the observed spectrum. We used the high-resolution ($R = 10\,000$) SSP library combining MIST isochrones (Choi et al. 2016) and the C3K theoretical atmospheres (Conroy et al. 2019). The flux blueward of the Lyman break was manually set to zero. These templates are complemented by a fifth-degree multiplicative Legendre polynomial in order to take into account systematic differences between the SSPs and the data (such as dust extinction, mismatch between the SSP models and high-redshift stellar populations, as well as residual flux calibration issues). The emission lines are modelled as pixel-integrated Gaussians, matching the observed spectral resolution. To reduce the number of degrees of freedom, we divide all emission lines in four kinematic groups, constrained to have the same redshift and intrinsic broadening. These are UV lines (blueward of 3000 Å), the Balmer series of hydrogen emission lines, non-hydrogen optical lines (blueward of 9000 Å), and near-infrared (NIR) lines. The stellar continuum has the same kinematics as the Balmer lines. Furthermore, we tie together doublets that have fixed flux ratios as determined by the atomic physics (such as the [O III] λ 5007/4959 ratio, which is set equal to 2.97), and constrain variable-ratio doublets to their physical ranges. We note that, at the resolution of the PRISM/CLEAR mode, the two components of the [O II] $\lambda\lambda$ 3726, 3729 doublet are completely unresolved and therefore are fit as a single Gaussian component centred at a rest-frame wavelength of 3728.42 Å, whereas for the gratings configurations the two emission lines are fitted separately, and their ratio is constrained to the physical range allowed by the density of the ISM ($\in [0.38-1.46]$); the same applies for the [S II] $\lambda\lambda$ 6718,6732 doublet ($\in [0.44-1.45]$) (Osterbrock & Ferland 2006).

We show an example fit for a $z = 4.805$ galaxy in Fig. 1. The upper panel shows the fit (in red) to the full PRISM (R100) spectrum in the wavelength region covering the main rest-frame optical emission lines, whereas in the three bottom panels we show a zoom-in of the best-fit to the R1000 spectrum for the same galaxy on the region of [O II] and [Ne III], Hβ and [O III], and Hα, [N II] and [S II] emission lines, respectively.

3. Derived physical quantities

3.1. Gas-phase metallicity

For the purposes of deriving the gas-phase metallicity, we use the observations obtained with both PRISM ($R \sim 100$) and medium resolution gratings ($R \sim 1000$), treating however the two spectral configurations independently. For each spectral configuration, emission line fluxes as measured from PPXF are reddening corrected on the basis of the decrement measured from the available Balmer lines. More specifically, we exploit the Hα/Hβ ratio at redshifts $z < 6.75$, where Hα is spectrally covered by NIRSpec, and Hγ/Hβ otherwise, adopting a Gordon et al. (2003) attenuation law and assuming the theoretical ratios from Case B

---

[3] https://github.com/esdc-esac-esa-int/eMPT_v1
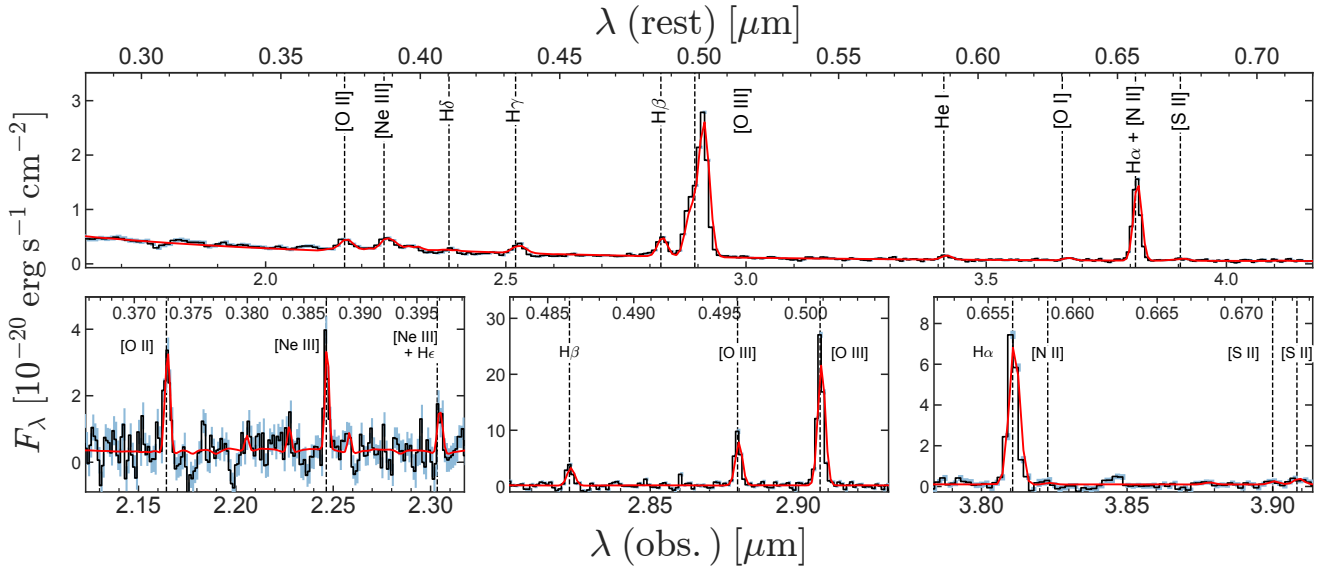
Fig. 1. Example of the spectral fitting procedure for one of the JADES galaxies in the sample, namely JADES-GS+53.16268-27.80237 at $z = 4.805$. *Top panel:* PRISM (R100) spectrum (in black), showing the region of the main rest-frame optical emission lines. The best-fit to the spectrum from PPXF is shown in red. *Bottom panels:* medium-resolution grating (R1000) spectrum (and its best-fit) for the same galaxy. From left to right, the panels show a zoom onto the regions of the [O II] and [Ne III], Hβ and [O III], and Hα [N II] and [S II] emission lines, respectively. In all panels, the error spectrum is marked by the cyan shaded region.

recombination at $T = 1.5 \times 10^4$ K (i.e. Hα/H$\beta = 2.86$; Hγ/H$\beta = 0.47$). In case no Balmer decrement could be measured, we adopt the nebular attenuation inferred from the SED fitting performed on PRISM spectra with the BEAGLE code (Chevallard & Charlot 2016, Chevallard et al., in prep.), which incorporates a two-component dust attenuation model that accounts for the differential attenuation of nebular and stellar emission in a self-consistent way.

The gas-phase metallicity is derived exploiting a revisited version of the calibrations presented in Curti et al. (2020b) (which are anchored to the T_e abundance scale), refined using a sample of local, metal-poor galaxies to better sample the low-metallicity regime. Such calibration set is similar to that presented in Nakajima et al. (2022). In addition, a novel diagnostics labeled R̂ (see Table 1), and based on a linear combination of R2 and R3 which further reduces the scatter of the calibration sample at fixed metallicity and span a wider dynamical range compared to the 'standard' R23 diagnostic, is implemented in place of the latter. For a detailed description of the rationale behind the new diagnostic, and for a more thorough comparison of classical strong-line calibrations against individual auroral line detections within the hereby presented JADES dataset, we refer to Laseter et al. (2024). The coefficients of the adopted calibrations are tabulated in Appendix B. A minimum signal-to-noise (S/N) of 5 is required on Hα and [O III], whereas of 3 on fainter emission lines, in order to consider a given emission line (and its associated diagnostics) in the metallicity derivation.

We visually inspected the spectra to remove sources affected by a poor fitting of the emission lines, as well as two galaxies identified as AGNs based on evidence for broad line emission, as detailed in Maiolino et al. (2023). Moreover, additional 14 AGN candidates identified via significant emission detected in high-ionisation transitions like He IIλ4686 and N IVλ1485 (Scholtz et al. 2023) are also removed from the analysis, as AGN-powered ionisation would compromise the standard metallicity calibrations tuned for star-forming galaxies, as well as

complicate the derivation of stellar masses and star-formation rates. However, it is relevant to note that most of these galaxies still show line ratios consistent with the star-forming population according to the 'BPT' diagram (Baldwin et al. 1981; Kewley et al. 2001; Kauffmann et al. 2003), whose applicability in discriminating among different ionising sources at high redshift has been in fact recently questioned, especially at low metallicity (Übler et al. 2023; Maiolino et al. 2023). Considering the challenges in quantitatively assessing the relative contribution of AGN ionisation with respect to the host galaxy in the observed spectra of these sources, we apply a conservative choice by removing the entire sample of narrow-line AGN candidates. Nonetheless, all the conclusions of the present paper are robust against this choice, and in Appendix A we report the results obtained by including also this subsample of galaxies.

In this work, all available strong-line diagnostics are included for each galaxy in the metallicity calculation, in order to reduce potential biases associated with the use of an individual, specific line ratio. The procedure explores the log(O/H) parameter space with a Markov Chain Monte Carlo (MCMC) algorithm exploiting the EMCEE package (Foreman-Mackey et al. 2013),

Table 1. Definitions of line ratios adopted throughout the paper.

Diagnostics	Line ratio
R2	[O II]$\lambda\lambda$3727,29/Hβ
R3	[O III]λ5007/Hβ
N2	[N II]λ6584/Hα
S2	[S II]$\lambda\lambda$6717,31/Hα
R23	([O II]$\lambda\lambda$3727,29 + [O III]$\lambda\lambda$4959, 5007)/Hβ
R̂	$0.47 \times \log_{10}(R2) + 0.88 \times \log_{10}(R3)$
O32	[O III]λ5007/[O II]$\lambda\lambda$3727,29
Ne3O2	[Ne III]λ3869/[O II]$\lambda\lambda$3727,29
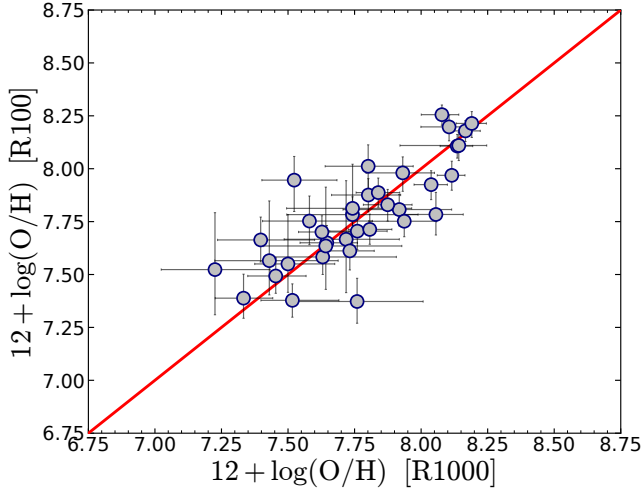
Fig. 2. Comparison between the metallicity derived from medium-resolution gratings and PRISM spectra. We plot only objects for which the requirements described in Sect. 3.1 are satisfied for both configurations. The two distributions scatter across the equality line (in red), with a median offset of 0.01 dex and a standard deviation of 0.15 dex. In 75% of the cases the two measurements are consistent within their 1σ uncertainties.

with the logarithmic likelihood defined as

$$\log(L) \propto \sum_i \frac{(R_{\mathrm{obs},i} - R_{\mathrm{cal},i})^2}{(\sigma_{\mathrm{obs},i}^2 + \sigma_{\mathrm{cal},i}^2)}, \qquad (1)$$

where the sum is performed over the set of available diagnostics used, R_{obs} are the observed line ratios, R_{cal} are the line ratios predicted by each calibration at a given metallicity, σ_{obs} are the uncertainties on the observed line ratios, and σ_{cal} are the dispersions of each calibrated line ratio at fixed metallicity.

For the majority of the sources under study however, we note that the [N II] and [S II] emission lines are either undetected, or falls outside of the NIRSpec wavelength coverage (see also Cameron et al. 2023b); moreover, [N II] $\lambda 6584$ and Hα are severely blended in PRISM spectra at the lowest redshifts probed by our sample. Therefore, we ultimately and effectively involve only diagnostics based on 'alpha' elements (i.e. R3, $\hat{R}$, O32; when available, [Ne III]/[O II] is also included). In some cases, [O III] $\lambda 5007$ and Hβ are the only emission lines significantly (above three-sigma) detected in the spectrum, which means that R3 is the only available diagnostic. In such cases, further information from upper limits on the [N II] $\lambda 6584$ (from R1000 spectra, where available) and on the [O II] λ 3727,29 emission lines can be placed on the N2 and O32 diagnostics, helping to discriminate between the two solutions provided by the double-branched R3 calibration. We here further note however that adopting instead a single, double-branched diagnostics (like R3 or R23) for the whole sample, although increasing the self-consistency of the metallicity derivation, might artificially introduce scatter in the log(O/H) distribution, as high-z galaxies typically tend to occupy the high-excitation region of the calibration, close or above the calibration plateau. This means that, even if the degeneracy between the two branches is somehow broken, the calibration itself carries less information about the metallicity in that region (being almost flat and therefore not very sensitive to metallicity variations). Therefore, forcing either the low- or high-solution introduces a 'gap' in metallicity between populations of galaxies which share very similar line ratios (see

e.g. the discussion in Guo et al. 2016), as the offset between the low- and high-metallicity solution is large at fixed line ratio. This can have a non-negligible impact, for instance, on the inferred slope of the metallicity scaling relations.

The procedure outlined above is applied separately for each galaxy to both PRISM and gratings spectra (where the latter are available). As mentioned above, we note that the N2 diagnostic is not included in any of the metallicity calculations based on PRISM spectra, as the R100 spectral resolution is not enough to resolve [N II] $\lambda 6584$ from Hα; similarly, S2-based diagnostics are not included for PRISM spectra, as they might be affected by [N II] contamination of the Hα line. Regarding the analysis carried out in the present paper, a fiducial metallicity value is then assumed for each source on the basis of the following scheme:

- We take the O/H inferred from medium-resolution (R1000) gratings if at least two diagnostics are available; otherwise
- we take the O/H inferred from the PRISM if at least two diagnostics are available in the PRISM spectrum; otherwise
- we take the O/H inferred from medium-resolution gratings as based on one individual diagnostic (typically R3), exploiting the information from three-sigma upper and lower limits on other line ratios (for instance, N2 or O32) to break the degeneracy of double-valued calibrations; otherwise
- we take the O/H inferred from the PRISM as based on one diagnostic (typically R3) in combination with the information from the upper limits on [O II] $\lambda\lambda 3727,29$.

The hereby adopted scheme leverages the higher resolution grating spectra for galaxies with strong line emission, whereas the much deeper PRISM spectra for fainter targets. We opt for not considering 'hybrid' gratings-to-prism line ratios because of potential issues associated with small wavelength or flux intercalibration uncertainties between the two configurations that are not yet fully understood. However, we compare the metallicity derived from gratings and prism spectra for the 35 galaxies where both measurements are available in the right-hand panel of Fig. 2. The two measurements are in agreement within the 1σ individual uncertainties in ~75% of the cases (and in ~95% within 2σ), with negligible systematic offset (i.e. 0.01 dex median offset towards higher grating metallicities) and a standard deviation of 0.15 dex.

In this work, we focus specifically on galaxies at $z \geq 3$, in order to study the evolution of the metallicity scaling relations beyond the epochs probed by previous surveys from the ground. Therefore, modulo the selections and caveats discussed above, our analysis includes a total of 62 galaxies from JADES with a metallicity measurement. The oxygen abundances measured in the selected JADES sample span between 12+log(O/H) = 7.21–8.54, with an average of 7.77 (i.e. 0.12 $Z_\odot$). In Fig. 3, we report the redshift distribution of our final selected sample of galaxies, compared to the distribution of the sample from the JADES deep GOODS-S tier observations with a spectroscopically confirmed redshift[4], which in turn constitutes ~70 per cent of the total sample introduced in Sect. 2.1; for a more detailed description of the full JADES deep sample, the selection function, and galaxy spectroscopic confirmation, we refer to Bunker et al. (2023a).

In particular, the final sub-sample selected for the current analysis represents the ~57 percent of the JADES deep sample in GOODS-S at $z > 3$ with a spectroscopic redshift, as remaining galaxies either do not have more than one emission line detected in their spectra, show evidence for contamination from AGN, or are lineless sources identified only through prominent spectral breaks (e.g. Curtis-Lake et al. 2023; Looser et al. 2023).

---

[4] i.e. with a redshift confirmed from either R1000 or R100 spectra.
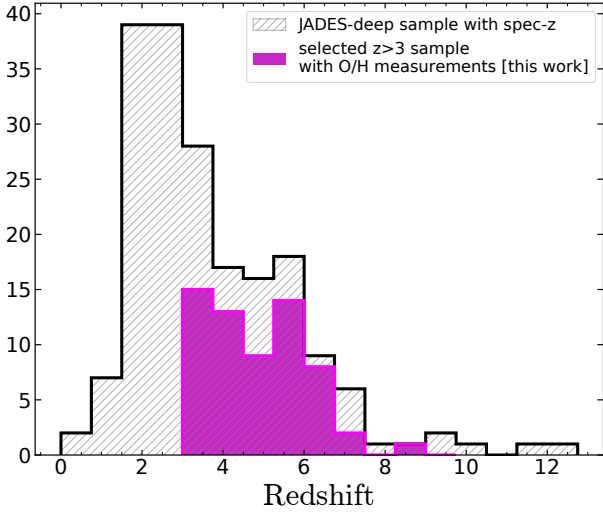
Fig. 3. Redshift distribution of the selected sample of 54 JADES galaxies with metallicity measurements analysed in this paper. Their histogram is compared to the distribution of the full 'JADES-Deep' sample presented in Sect. 2.1 and Bunker et al. (2023a) with a spectroscopic redshift determination.

3.2. Stellar masses and SFRs

To measure the stellar masses for our selected sample of galaxies in JADES, we employ full spectral fitting to PRISM spectra performed with the BEAGLE code (Chevallard & Charlot 2016). The PRISM spectra that have been employed only included slit-loss corrections based on the assumption of a point source, which is not appropriate for all objects in the sample. To address this, where objects have NIRCam photometry from the JADES survey, the BEAGLE fits are run with both PRISM spectroscopy and total, Kron-based photometry (Rieke et al. 2023). In these cases, a low-order calibration polynomial is included in the fit to match the shape and normalisation of the spectrum to the photometry, whereby the total stellar mass and star-formation rate estimates are adjusted to the total fluxes measured for the objects. Where NIRCam photometry is not available we resort to the derived quantities from PRISM spectroscopy alone. We assume a delayed-exponential star-formation history (SFH), a Chabrier (2003) IMF with an upper mass limit of $100\,M_\odot$, and adopt the updated Bruzual & Charlot (2003) stellar population models described in Vidal-García et al. (2017). We assume the total mass currently locked into stars as our fiducial estimate for the stellar mass, which accounts for the fraction of mass returned to the ISM, instead of the integrated SFH. We refer to Chevallard et al. (in prep.) for further information of the BEAGLE fitting procedure.

The star-formation rates of our galaxies are estimated in two different ways, namely from the output of the BEAGLE SED fitting run described above, which provides an estimate of the star-formation rate averaged over the past ten Myr, and from the attenuation-corrected[5] Hα luminosity, following the recipe[6] outlined in Reddy et al. (2022), Shapley et al. (2023b), better suited for low-metallicity galaxies characterised by an increased rate of ionising photons. For galaxies in which Hα is not observed, either because it is shifted out of the spectral coverage of NIR-

Spec (i.e. at $z \gtrsim 7$), or because it falls in one of the detector gaps (in medium resolution gratings), we exploited the dust-corrected Hβ flux rescaled by the theoretical case B recombination factor at $T \sim 1.5 \times 10^4$ K (2.86). For consistency, for any given galaxy we use the Hα(Hβ) flux as measured from either the PRISM or medium resolution gratings spectra according to the configuration exploited for the full metallicity calculation (as described in Sect. 3.1).

On average, the Hα-based SFRs are ~0.14 dex lower than those derived by BEAGLE, with a dispersion in the deviation between the two measurements of ~0.3 dex. Such discrepancy is possibly driven by (i) the different underlying stellar population synthesis models (i.e. BPASS for Reddy et al. 2022, revised Bruzual & Charlot 2003 for BEAGLE), (ii) different prescriptions for the star-formation history, (iii) residual contamination of the Hα flux from [N II] in PRISM spectra. We adopt the SED-based measurements as our fiducial value of the total SFR of a galaxy for the analysis presented in the current paper, more consistent with the stellar mass derivation. Nonetheless, we verified that assuming either of the two SFR values does not alter any of the conclusions presented in this work, particularly when discussing the evolution in the fundamental metallicity relation (Sect. 5), as shown in Appendix A.

3.3. High-redshift galaxy samples from the literature

The specific selection function of the deep tier of the JADES spectroscopic campaign provides preferential coverage of the low-end of the stellar mass distribution (i.e. $M_\star/M_\odot \approx 10^{6.5}$–$10^9$), a regime poorly probed before, especially at such early cosmic epochs. In order to extend the parameter space over which the metallicity scaling relations are analysed, here we complement our JADES sample with a sample of 80, $z = 4$–10 galaxies from the CEERS programme as analysed and presented by Nakajima et al. (2023), which are preferentially distributed across the $\log(M_\star/M_\odot) \sim 8.5$–$10$ regime.

To preserve a good level of self-consistency in the analysis, we have recomputed the metallicity for these galaxies starting from the emission line ratios reported in Table B1 of Nakajima et al. (2023), applying the procedure discussed in Sect. 3.1. Star-formation rates reported in Nakajima et al. (2023) are based on dust-corrected Hβ luminosity and the calibration reported by Kennicutt & Evans (2012). We scale down these values by the median offset (i.e. 0.23 dex) between the Hα-based SFR (this time derived assuming the Kennicutt & Evans 2012 formula) and the BEAGLE-SFR as measured for the JADES galaxies, in order to minimise systematic offsets between the SFRs compiled for CEERS galaxies and the fiducial values assumed for JADES objects.

Stellar masses in the CEERS sample are based on fits to NIRCAM photometry performed with the PROSPECTOR code (Johnson et al. 2021). A full assessment of the systematics involved in the stellar mass determination as based on different spectral fitting codes goes beyond the scope of this paper. Nonetheless, we note that the statistical uncertainties on $M_\star$ quoted by Nakajima et al. (2023) are much larger than those associated with BEAGLE measurements for JADES galaxies. To reduce the impact of such differential measurement uncertainties, we do not apply inverse variance weighting based on $M_\star$ uncertainties in any of our fitting analyses in the paper.

In addition, we include in the analysis four targets whose metallicity have been derived with the 'direct' T_e-method, namely three galaxies observed in the framework of the Early Release Observations programme as described in Curti et al. (2023),

---

[5] As described in Sect. 3.1, we exploited the measured Balmer decrements with a Gordon et al. (2003) law where available, whereas the output Av from BEAGLE otherwise.
[6] Assuming a conversion factor of $10^{-41.67}$ $(M_\odot\,\mathrm{yr}^{-1})/(\mathrm{erg\,s}^{-1})$.
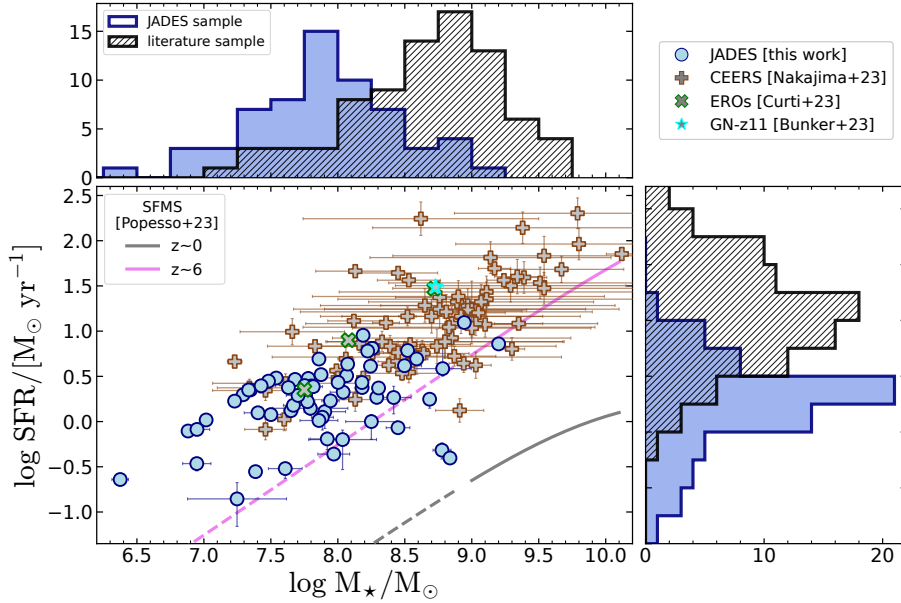
Fig. 4. Distribution in the stellar-mass–SFR plane for our combined JWST sample. This includes both JADES galaxies presented in this work and the literature sample compiled from Nakajima et al. (2023; CEERS), Curti et al. (2023; EROs), and Bunker et al. (2023b; GN-z11), respectively. The top and right-hand inset panels show the histograms of the distribution in both parameters. The SFR for JADES galaxies is derived from BEAGLE fitting to both PRISM spectra and NIRCAM photometry (where available, and to PRISM spectra only otherwise). For CEERS galaxies, the SFRs are compiled from Nakajima et al. (2023) as inferred from the Hβ luminosity and the Kennicutt & Evans (2012) calibration, but have been here scaled down by 0.23 dex to account for the mean offset between the Hα(Hβ)-based and BEAGLE-based SFRs measured for the JADES sample (we refer to Sect. 3.2 for more details). For both ERO galaxies and GN-z11 the $M_\star$ and SFR have been derived via BEAGLE fitting, consistently with what done for JADES sources. The parametrisation of the main sequence of star-formation (SFMS) at $z \sim 0$ and $z \sim 6$ from Popesso et al. (2023; and their extrapolation at low stellar mass, dashed line) is also shown for reference.

and the galaxy GN-z11 observed in the framework of JADES and presented in Bunker et al. (2023b), Tacchella et al. (2023a), though the nature of its ionising source is debated and the presence of an AGN has been proposed (Maiolino et al. 2024). This brings the total number of galaxies analysed in this paper to 146.

For the three ERO obejcts, $M_\star$ and SFR are derived via BEAGLE fitting to NIRCAM photometry as described in Curti et al. (2023), whereas the metallicities have been recomputed following the same data reduction implemented for the JADES galaxies analysed in this paper, as also described in Laseter et al. (2024). For GN-z11, fiducial values for $M_\star$ and SFR are taken from Bunker et al. (2023b) and are based on BEAGLE fitting to PRISM spectrum assuming a Chabrier (2003) IMF with an upper mass cut-off of 100 $M_\odot$, whereas the oxygen abundance is measured exploiting the detection of [O III] λ4363 in the PRISM spectrum and its ratio over [Ne III] λ3869 and Hγ, as detailed in Cameron et al. (2023a), which quote a fiducial value of 12+log(O/H) = 7.82. For comparison, we also obtain an indirect estimate of the [O III] λ5007 flux (which is not covered in the spectra of GN-z11) based on the observed [Ne III]/[O II] (from medium resolution gratings) and the Witstok et al. (2021) [Ne III]/[O II] versus [O III]/[O II] calibration, which we then use to derive the R̂ diagnostic and use it conjunction with Ne3O2 and the methodology described in Sect. 3.1: this approach provides $12+\log(O/H) = 7.67^{+0.16}_{-0.10}$. At the same time, using the same indirect estimate of the [O III] λ5007 flux to apply the T_e method (and adopting Pilyugin et al. 2009 to derive the temperature of the low-ionisation zone) delivers a much higher 12+log(O/H) = 8.42. Given the large systematic uncertainties affecting both methods, we here assume the value reported by Cameron et al. (2023a) (7.82) as our fiducial value, with a conservative uncertainty of 0.35 dex.

Figure 4 shows the distribution of both the JADES sample introduced in this work and the compiled sample from the literature discussed above, in the stellar mass versus SFR diagram; the marginalised histograms of the distribution of the two quantities are also shown in the upper and right-hand inset panels. We note the two samples are quite complementary, and especially including galaxies from Nakajima et al. (2023) allows us to extend the mass regime probed up to $M_\star \sim 10^{9.5} M_\odot$ ($\langle\log(M_\star/M_\odot)\rangle_{\rm CEERS} = 8.72$; $\langle\log(M_\star/M_\odot)\rangle_{\rm JADES} = 7.83$). Galaxies selected from CEERS are also distributed towards higher SFRs ($\langle\log({\rm SFR})\rangle_{\rm CEERS} = 1.3$; $\langle\log({\rm SFR})\rangle_{\rm JADES} = 0.45$), though the two galaxy samples probe similar specific star-formation rates (sSFR = SFR/$M_\star$). For reference, we plot the parametrisation of the 'Star Forming Main Sequence' (SFMS, Noeske et al. 2007; Speagle et al. 2014; Renzini & Peng 2015) at $z \sim 0$ and $z \sim 6$ from Popesso et al. (2023), noting that the steepness of such parametrisation causes its low-mass extrapolation at $z \sim 6$ (dashed line) to systematically underestimate the observed star-formation rate for the low-mass JWST galaxy sample (by ~0.5 dex on average at $M_\star = 10^8 M_\odot$), given our assumed fiducial $M_\star$ and SFR measurements.

4. Evolution of the mass–metallicity relation beyond z = 3

4.1. The observed MZR at 3 < z < 10

In Fig. 5, we show our combined JWST galaxy sample on the mass-metallicity plane. Each point is marked with a different symbol according to its parent sample (as reported in the legend), and it is colour-coded on the basis of the following redshift binning scheme: yellow at $z \in [3-6]$, and blue at $z \in [6-10]$. The
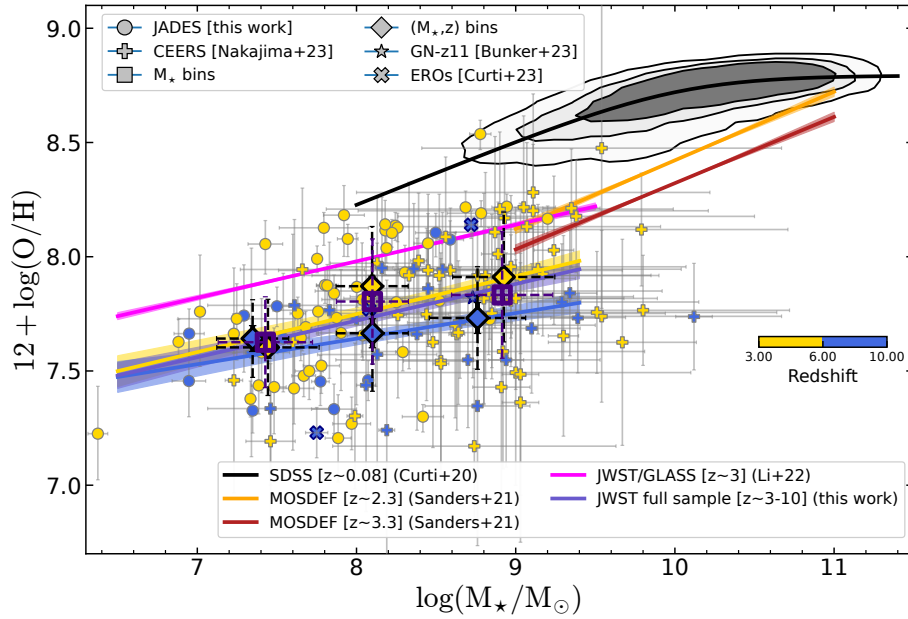
Fig. 5. Mass-metallicity relation (MZR) for our full JWST sample. Filled circles represent individual galaxies presented in this work from JADES, whereas filled crosses are galaxies observed in the framework of the CEERS programme and compiled from Nakajima et al. (2023), with metallicity derived as detailed in Sect. 3.1. The star symbol reports the JADES/NIRSpec observations of GN-z11 (Bunker et al. 2023b, at $z = 10.603$), whereas 'X' symbols mark galaxies from the EROs as compiled from Curti et al. (2023) and Laseter et al. (2024). Large squared and diamond symbols mark the median values computed in bins of $M_\star$ (full sample, in purple), and $(M_\star z)$, as described in Table 2. An orthogonal linear regression fit to the median values in bins of $M_\star$ for the different redshift sub-samples is shown by the purple (full sample), yellow ($z = 3-6$) and blue ($z = 6-10$) lines, respectively. We include a comparison with previous determinations of the MZR at lower redshifts from Curti et al. (2020b; SDSS at $z \sim 0.07$), and Sanders et al. (2021; MOSDEF at $z \sim 2-3$), as well as the best-fit of the low-mass end of the MZR at $z \sim 3$ provided by Li et al. (2023) and based on JWST/NIRISS slitless spectroscopy. The MZR curves at $z \sim 2-3$ have been scaled down by ~0.1 dex to account for the systematics differences between the metallicity calibrations used in this work and the Bian et al. (2018) calibrations adopted in the original papers.

median redshift of the whole sample is $\langle z \rangle = 5.10$, whereas the median redshift of the $z \in 3-6$ sub-sample is $\langle z \rangle_{3-6} = 4.76$ and that of the $z \in 6-10$ sub-sample is $\langle z \rangle_{6-10} = 6.73$. In particular, filled circle points in the plot report galaxies from the JADES survey, while filled crosses symbols mark galaxies from CEERS (Nakajima et al. 2023), whose metallicity has been remeasured as described in Sect. 3.1. For both the total sample and each sub-sample in redshift we show the median (large diamond markers), error on the median (solid errorbar), and standard deviation (dashed errorbar) of the metallicity within three different bins in stellar mass. The size of the stellar mass bins is not uniform to maintain a reasonable (i.e. at least 10) number of galaxies in each bin. The average properties of the binned samples are reported in Table 2.

Our galaxies present a median offset of ~−0.5 dex (−0.48 dex and −0.64 dex at $z = 3-6$ and $z = 6-10$, respectively) compared to the low-mass end (and its extrapolation) of the MZR in the local Universe (based on individual SDSS galaxies, Curti et al. 2020b). We also compare our results with previous realisations of the MZR at $z \sim 2-3$. In particular, we first compare our observations with the mass-metallicity relation derived from stacked spectra of galaxies from the MOSDEF Survey as reported by Sanders et al. (2021). To minimise the impact of systematics introduced by the use of different metallicity calibrations (the calibrations from Bian et al. 2018 are adopted in Sanders et al. 2021), we have recomputed the metallicity for the MOSDEF stacked spectra with the same methodology outlined in Sect. 3.1, finding an average offset of ~0.088 dex towards lower metallicities. Therefore, in Fig. 5 the mass-metallicity relations at $z \sim 2.2$ and $z \sim 3.3$ from Sanders et al. (2021) are lowered by

0.088 dex compared to their original parametrisation. Overall, the evolution in normalisation probed by our full galaxy sample at $z = 3-10$ appears relatively mild if compared to the extrapolation at low stellar mass of the $z \sim 3.3$ MZR from Sanders et al. (2021), with a mean offset for the full sample of 0.045 dex (0.13 dex if we assume the fiducial MZR from Sanders et al. 2021 based on the Bian et al. 2018 calibrations). However, such deviation is mass-dependent and is more prominent at higher $M_\star$, being almost zero for $\log(M_\star/M_\odot) \lesssim 8$, while ~0.15 dex at $\log(M_\star/M_\odot) \sim 9$. This behaviour suggests an evolution in the slope of the mass-metallicity relation in the mass and redshift regimes probed by this work.

In the attempt to characterise the evolution in the slope of the MZR at these redshifts, we perform an orthogonal linear regression fit to both the full JWST sample of individual galaxies, and within the two redshift bins separately, in the functional form

$$12+\log(O/H) = \beta_z \log(M_\star/10^8\, M_\odot) + Z_{m8}, \qquad (2)$$

where β_z is the slope at a given redshift interval, and Z_{m8} is the normalisation at $\log(M_\star/M_\odot) = 8$. The results are displayed in Fig. 5, as shown by the solid purple (full sample), yellow ($z = 3-6$), and blue ($z = 6-10$) lines, whereas the shaded regions represent the 1-σ confidence interval of each fit, derived from bootstrapping each sample (with replacement) and repeating the fitting procedure 300 times. The best-fit parameters are reported in Table 3. We note that the best-fit MZR (based on individual data points) agrees well with the median values computed in bins of $M_\star$ for each redshift sub-sample.

We find indications of a flattening of the slope of the low-mass end of the MZR with redshift, with the best-fit slope for

Table 2. Median values, error on the median (and standard deviation) in stellar mass, SFR, and metallicity for our sample of galaxies in each of the ($M_\star$–redshift) bins considered in this work.

[log $M_\star$]$_{\rm bin}$	$N_{\rm gal}$	$\langle z \rangle$	log $M_\star/M_\odot$	log SFR [$M_\odot$ yr^{-1}]	12+log(O/H)
Redshift $z \in [3,10]$					
[6 ; 7.75]	29	5.80	7.43 ± 0.06 (0.32)	0.22 ± 0.11 (0.46)	7.63 ± 0.04 (0.20)
[7.75 ; 8.5]	52	5.18	8.10 ± 0.03 (0.23)	0.48 ± 0.06 (0.39)	7.80 ± 0.04 (0.27)
[8.5 ; 10]	64	5.00	8.92 ± 0.04 (0.32)	1.16 ± 0.07 (0.53)	7.83 ± 0.03 (0.28)
Redshift $z \in [3,6]$					
[6 ; 7.75]	20	5.25	7.44 ± 0.08 (0.36)	0.15 ± 0.12 (0.51)	7.60 ± 0.05 (0.36)
[7.75 ; 8.5]	36	4.41	8.10 ± 0.04 (0.24)	0.41 ± 0.06 (0.35)	7.87 ± 0.04 (0.24)
[8.5 ; 10]	52	4.78	8.93 ± 0.04 (0.33)	1.16 ± 0.0 (0.56)	7.91 ± 0.04 (0.33)
Redshift $z \in [6,10]$					
[6 ; 7.75]	10	6.71	7.35 ± 0.08 (0.24)	0.34 ± 0.09 (0.30)	7.64 ± 0.07 (0.24)
[7.75 ; 8.5]	15	6.54	8.10 ± 0.06 (0.22)	0.63 ± 0.10 (0.40)	7.67 ± 0.06 (0.22)
[8.5 ; 10]	11	6.93	8.76 ± 0.09 (0.30)	1.28 ± 0.10 (0.32)	7.73 ± 0.08 (0.30)

Notes. The number of galaxies in each bin and their median redshift is also reported.

Table 3. Best-fit parameters of the mass–metallicity relation from Eq. (2) for the full sample, and the two redshift bins at $z = 3-6$ and $z = 6-10$, for which we report both the slope β_z and the normalisation at $M_\star = 10^8 \, M_\odot$, Z_{m8}.

Sample	β_z	Z_{m8}
$z \in [3,10]$	0.17 ± 0.03	7.72 ± 0.02
$z \in [3,6]$	0.18 ± 0.03	7.75 ± 0.03
$z \in [6,10]$	0.11 ± 0.05	7.65 ± 0.04

the full sample equal to $\beta_{3-10} = 0.17 \pm 0.03$. If we focus more specifically on the two redshift bins described above, we find $\beta_{3-6} = 0.18 \pm 0.03$, and $\beta_{6-10} = 0.11 \pm 0.05$. The low-mass end slope of the MZR as probed by our sample is flatter (at 3.6σ significance) than both observed for the $z \sim 2-3$ MZR at higher stellar masses ($\beta_{2-3} = 0.29$), and also in the local Universe (0.28, as measured in Curti et al. 2020b). Such apparent invariance between $z \sim 0$ and $z \sim 3$ suggests that the same physical processes, and in particular how the metal removal efficiency of stellar winds scales with $M_\star$, governs the MZR slope over the last ~ 12 Gyr of cosmic time (Sanders et al. 2021). We note however that most previous assessments of the MZR at high-z are based on near-infrared spectroscopic surveys from the ground, and were therefore limited in the stellar mass range they could probe, struggling to access the regime below $\log(M_\star) < 10^9 \, M_\odot$ and thus to provide a sample matched in stellar mass to the JWST sample discussed in the present work. Further insights on the low-mass end slope of the MZR at $z \sim 2-3$ come from recent JWST/NIRISS observations within the GLASS-ERS programme, have been presented in Li et al. (2023), and report similar evidence for a shallower slope ($\beta_{2-3} = 0.17 \pm 0.03$, as shown in magenta in Fig. 5) compared to the simple extrapolation of the $z \sim 3.3$ relation inferred by Sanders et al. (2021). In this work, we push the investigation of the MZR in the dwarf regime to much higher redshift, finding a slope at z_{3-6} which is consistent with that reported by Li et al. (2023). Compared to that parametrisation of the MZR, we observe a more prominent evolution in the normalisation of ~ 0.25 dex, after once again correcting for the systematics between the Bian et al. (2018) calibrations and the set of calibrations adopted in this work.

It is also interesting to observe that the slope inferred at $z = 6-10$ (in blue in Fig. 5) is even flatter (and consistent with zero within $\sim 2\sigma$), suggesting that the mass-metallicity relation could be in an initial build-up phase. However, the high-redshift bins are only sparsely populated and are probably subject to strong selection biases, and it is therefore difficult to draw any strong conclusion at this stage relative to any possible evolution in the MZR slope between $z = 3-6$ and $z > 6$.

Finally, we note that the distribution of our data in the MZR plane is characterised by a large amount of scatter, at any given mass and redshift. Although the sample considered in the present work is not as complete and unbiased as required to perform a robust assessment of the scatter of the scaling relation, we can tentatively measure the amount of intrinsic scatter in our galaxy sample as $\sigma_{\rm MZR} = \sqrt{\sigma_{\rm obs}^2 - \sigma_{\rm meas}^2}$, where $\sigma_{\rm obs}$ is the standard deviation of the full JWST sample around the best-fit relation and $\sigma_{\rm meas}$ is the average measurement uncertainty associated with the metallicity determination[7]. We find $\sigma_{\rm MZR} \sim 0.073$ dex for our full sample, consistent with the scatter measured in the local MZR at higher masses ($\log(M_\star/M_\odot) > 9.5$) ($\sigma_{{\rm MZR},z=0} = 0.075$; Curti et al. 2020b). We note that the scatter at high masses in local galaxies is likely dominated by the flattening of the MZR at high masses ($\log(M_\star/M_\odot) \geq 10$), interpreted in classical gas-regulator models (Finlator & Davé 2008; Lilly et al. 2013) as a consequence of the metallicity approaching the stellar yield, whereas no clear 'turnover' nor high-mass flattening is probed at higher redshifts.

4.2. Local analogues of high-redshift galaxies

The best-fit relations (and median values in bins of $M_\star$ and redshift) reported in this work show remarkable agreement with the MZR probed by samples of local star-forming galaxies with extreme line emission properties, such as the 'Green Pea' and 'Blueberry' galaxies compiled from Yang et al. (2017a,b), whose metallicity has been derived with the 'direct' T_e method; these objects are marked by green and cyan symbols, respectively, in the left panel of Fig. 6. In particular, the MZR slope inferred from the combined sample of 'Green Pea' and 'Blueberry' galaxies is 0.18 ± 0.02, fully consistent with the slope

---

[7] The dispersion of the adopted strong-line calibrations are included directly at the metallicity derivation step.
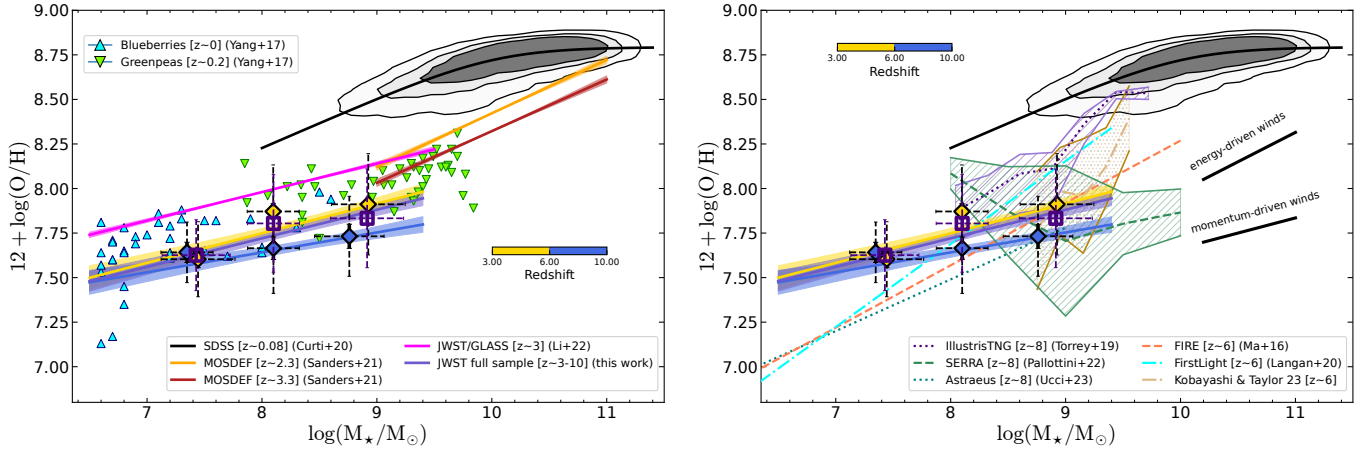
Fig. 6. Comparison with local analogues and cosmological simulations. *Left-hand panel:* the best-fit MZR presented in this work (for both the full JWST sample in purple, and in the $z = 3−6$ and $z = 6−10$ redshift binsin yellow and blue, respectively) is compared to the T_e-based mass-metallicity relation from local, metal-poor 'Blueberry' and 'Green Pea' galaxies (Yang et al. 2017a,b), which share similar excitation and emission line properties to our high-redshift galaxy sample (Cameron et al. 2023b). Their distribution agrees well in slope and normalisation (with ~0.11 dex offset in log(O/H)) with the average relation inferred in this work for our low-mass JWST sample. *Right-hand panel:* our best-fit MZRs are compared with the predictions from different suites of cosmological simulations at $z ∼ 6−8$, namely FIRE (Ma et al. 2016), IllustrisTNG (Torrey et al. 2019), FirstLight (Langan et al. 2020), Astraeus (Ucci et al. 2023), SERRA (Pallottini et al. 2022), and with the chemodynamical simulations from Kobayashi & Taylor (2023). In addition, the typical MZR slopes predicted by (semi-) analytical chemical evolution models (e.g. Davé et al. 2012) implementing feedback via either 'energy-driven' or 'momentum-driven' winds are also shown (the normalisation is assumed arbitrary to aid visualisation). The latter agrees well with the slope observed at low stellar mass for galaxies at $z ≥ 3$.

measured from our sample, whereas JWST galaxies have, on average, 0.12 dex lower metallicity. These low-redshift sources are well matched in stellar mass with our JWST sample $(\log(M_\star/M_\odot) ≈ 6.5–9.5)$, are very compact, characterised by very high equivalent width of the emission lines, low metallicities, and high ionisation parameters, and therefore have been long identified as potential analogues of very high-redshift sources. The good agreement between the emission line and metallicity properties in our $z ∼ 3−10$ JWST sample and those observed in this sample of local, extreme line emitters corroborates such interpretation (see also Cameron et al. 2023b).

4.3. Comparison with models and simulations

Here, we compare the observed MZR in our $z = 3−10$ galaxy sample with the predictions of different cosmological simulations, as depicted in the right-hand panel of Fig. 6. In particular, we report the mass-metallicity relationship at $z ∼ 6$ (close to the median redshift of our full sample; i.e 5.86) from FIRE (Ma et al. 2016), IllustrisTNG (TNG100) (Torrey et al. 2019), FirstLight (Langan et al. 2020), and from the chemo-dynamical simulations by Kobayashi & Taylor (2023), as well as predictions at higher redshift $(z ∼ 8)$ from Astraeus (Ucci et al. 2023), and Serra (Pallottini et al. 2022). For IllustrisTNG simulations we followed the approach described in Torrey et al. (2019) and already followed in Curti et al. (2023), considering central galaxies and assuming that oxygen comprises 35 per cent of the SFR-weighted metal mass fraction within twice the stellar half-mass radius. Within FIRE simulations the gas-phase metallicity is defined as the mass-weighted metallicity of all gas particles that belong to the ISM, assuming solar abundance ratios (Asplund et al. 2009). We note that the FIRE results at $z = 6$ are renormalised to match the normalisation of the MZR in the local Universe. Results from the FirstLight simulations at $z = 6$ are compiled from Langan et al. (2020, and Langan, priv. comm.), and based on the assumption that the unresolved nebular region around each star particle shares the same mass ratio of met-

als produced in Type II SN explosions as in the star particle. The galaxy metallicity is defined as the mass-weighted average nebular metallicity including all star particles younger than 100 Myr. The predictions from chemo-dynamical simulations by Kobayashi & Taylor (2023) are instead based on the SFR-weighted gas-phase metallicity. Finally, the Astraeus simulations (Ucci et al. 2023) accounts for the oxygen mass in the halo without any weighting, whereas in SERRA gas and stellar metallicity are coupled and the metallicity is tracked as the sum of all heavy elements (Pallottini et al. 2022).

Overall, most simulations appear to reasonably match the normalisation of the observed MZR at $M_\star ≈ 10^{8−9}$, though under-predicting the abundances observed at $\log(M_\star/M_\odot) < 8$, with the theoretical MZRs characterised by steeper slopes than inferred from our sample. Only the extrapolation of IllustrisTNG predictions to lower $M_\star$ seems to match reasonably well (in normalisation) the average distribution of galaxies in the mass-metallicity plane below $M_\star = 10^8 M_\odot$, but diverges more dramatically at higher masses. Interestingly, the lack of a clear correlation (and the large scatter) at $z ∼ 8$ predicted by SERRA seems in line with our observations in the highest redshift bin $(z > 6)$.

We also compare our inferred MZR slopes with those predicted by (semi-)analytical chemical evolution models under the assumption of different modes of stellar feedback. In particular, in the framework of 'equilibrium' models (e.g. Davé et al. 2012; Lilly et al. 2013), an 'energy-driven winds' scenario predict a slope for the MZR of $\beta_{en} ∼ 0.33$, whereas for 'momentum-driven winds' the predicted slope is shallower, i.e. $\beta_{mom} ∼ 0.17$ (Guo et al. 2016); these two cases are shown in the right-hand panel of Fig. 6 with thick black lines (here the normalisation is assumed arbitrary, and the two curves are artificially shifted along the x-axis to aid the visualisation and comparison of the various trends). We observe the former to be in good agreement with the MZR slope probed by galaxies at intermediate and high masses $(\log(M_\star/M_\odot) ≥ 9)$ and up to $z ∼ 3$, whereas the

latter is more consistent with the observed MZR slope as probed by our JWST sample at the low-mass end and at higher redshift. Although preliminary, these results are consistent with a scenario in which the dominant feedback mechanism in galaxies changes in different mass regimes, producing a shallower MZR slope at low $M_\star$. At the same time, observing a similar MZR slope at $M_\star \sim 10^7 - 10^9$ between $z \sim 3$ (from Li et al. 2023) and $z \sim 6$ (from this work) suggests that the scaling of the mass-loading factor of outflows with stellar mass is invariant in this mass regime during this epoch.

4.4. Can the evolution in the MZR normalisation be explained by evolving the gas fraction?

Another critical parameter which contributes to setting the scaling relation between $M_\star$ and metallicity is the gas fraction. The average gas fraction in galaxies is observed to increase with redshift at fixed $M_\star$, as inferred from large existing datasets up to $z \sim 2-3$ (e.g. Saintonge et al. 2016; Scoville et al. 2017; Tacconi et al. 2018, 2020). If the evolution proceeds at a similar rate even at earlier epochs, the impact of metal dilution is expected to become increasingly more important with redshift, overcoming the impact of metal loss due to outflows at fixed stellar mass (Sanders et al. 2021). However, this seems apparently hard to reconcile with the relatively mild evolution in the MZR normalisation observed in our JWST sample. In this section, we briefly test the impact of evolving the gas fraction on the predicted MZR evolution at low $M_\star$, under a set of simple assumptions.

In the framework of analytical chemical evolution models that allow the gas fraction to vary (see e.g. Peeples & Shankar 2011; Lilly et al. 2013), and under the assumption of (almost) pristine gas inflows[8], the metallicity of the ISM is the result of the balance between the nucleosynthetic stellar yield, the gas dilution, and the amount of metals lost due to outflows. Here we adopt the formalism of the Peeples & Shankar (2011) model (their Eqs. (9), (10), and (11)) to predict the evolution in the metallicity at fixed $M_\star$ in the low-mass regime (specifically, at $M_\star = 10^8 M_\odot$) as the consequence of the sole evolution in the gas fraction with redshift, keeping other parameters such as oxygen yields, return mass fraction, and mass-loading factor of outflows, fixed.

As for the oxygen yield y_O and return fraction R[9], we adopt $y_O = 0.032$ and $R = 0.463$ as fiducial values, following (Vincenzo et al. 2016) and assuming a Chabrier (2003) IMF with upper-mass cutoff of $100 M_\odot$, stellar yields from Nomoto et al. (2013), and metallicities typical of our JWST galaxy sample. We assume then that the redshift evolution of the gas fraction μ_{gas} (here defined as $\mu_{gas} = M_{gas}/M_\star$) follows with the same scaling up to $z \sim 6$ as inferred up to $z \sim 3$ (i.e. $\log(\mu_{gas}) \propto 2.49 \log(1+z)$) as based on the linear parametrisation of Eq. (6) and Table 3 in Tacconi et al. 2018[10]. Furthermore, in modelling the observed median MZR for the full JWST sample, we assume that galaxies lie on the SFMS (i.e. $\delta(MS) = 0$) and on the $M_\star$-R_e relation from van der Wel et al. (2014) evaluated at the median redshift of the sample at $\log(M_\star/M_\odot) = 8$ ($z \sim 5.2$), the only dependence of the gas fraction being therefore on stellar mass and redshift.

We performed our calculations for two different scenarios. In the first case, we extrapolate the MZR at $z \sim 3.3$ from Sanders et al. (2021; with a slope of 0.30) to $\log(M_\star/M_\odot) = 8$ to fix the 'observed' metallicity, and solve Eq. (10) from Peeples & Shankar (2011) for the mass-loading factor of outflows ζ_w, finding $\log(\zeta_w) = 1.46$; this value is higher than what predicted by Sanders et al. (2021) if extrapolating their Eq. (16) to lower $M_\star$, mainly reflecting our different choice of the fiducial y_O. Then, assuming ζ_w invariant with redshift between $z = 3$ and $z = 5.2$ (the median redshift of the full JWST sample), we derive the expected metallicity at $M_\star = 10^8 M_\odot$ by evolving the gas fraction μ_{gas} according to the scaling relations of Tacconi et al. (2018). We predict an evolution in metallicity of 0.17 dex, larger than the (almost negligible) evolution of 0.02 dex probed by the extrapolation of the Sanders et al. (2021) MZR at $z \sim 3.3$.

In the second case, we instead solve for ζ_w in order to match slope (0.17) and normalisation ($12+\log(O/H) = 7.98$ at $M_\star = 10^8 M_\odot$) of the MZR at $z = 3$ in the dwarf regime from Li et al. (2023), finding $\log(\zeta_w) = 1.16$. Fixing again ζ_w with redshift, we predict a decrease in $\log(O/H)$ as driven purely by an evolution in the gas fraction of 0.22 dex, which is in good agreement with that observed in such mass regime between the $z = 3$ MZR from Li et al. (2023) and our median MZR at $z = 5.2$ (i.e. 0.26 dex).

Within the described framework, an increase in the gas reservoir of low-mass galaxies between redshifts ~3 and ~5 is almost sufficient to entirely explain the (relatively mild) observed evolution of the MZR normalisation, without any adjustment to the contribution of the metal-loading factor of outflows with redshift at fixed $M_\star$ (whereas the scaling of ζ_w with $M_\star$ might be varying, as reflected in the apparent flattening of the MZR slope). Indeed, solving directly for ζ_w at $z = 5.2$ by imposing the measured slope and normalisation of the MZR at $\log(M_\star/M_\odot) = 8$ as derived in this work delivers $\log(\zeta_w) = 1.22$, only slightly higher than the observational constraints inferred from the MZR at $z \sim 3$ from Li et al. (2023) ($\log(\zeta_w)=1.16$, as derived above). On the other hand, an extrapolation of the Sanders et al. (2021) MZR in the low-mass regime (i.e. fixing the MZR slope) would be consistent with almost no evolution in the normalisation, requiring either a much shallower evolution in the μ_{gas} versus redshift scaling relation at $z > 3$, an increase in the oxygen yields, and/or a non-negligible metallicity of the infalling gas compared to that of the ISM in order to reconcile the model-predicted MZR with the observed median MZR at $\log(M_\star/M_\odot) = 8$.

There are nonetheless a number of strong assumptions in place underlying this simple modelling. First, in analytical chemical evolution models the predicted normalisation of the MZR is highly sensitive to the assumed oxygen yield and the shape of the upper mass cutoff of the IMF (Vincenzo et al. 2016). For instance, assuming different values for the oxygen yield in the range 0.015–0.04 (which are typical values spanned in the literature for different IMFs) propagates into a shift of $^{+0.04}_{-0.14}$ dex in the predicted $\log(O/H)$ at $M_\star = 10^8 M_\odot$. Furthermore, we have assumed that the gas mass budget is dominated by hydrogen in the molecular phase and took $M_{gas} = M_{H_2}$ (Tacconi et al. 2018). Although different from what is observed in relatively higher mass star-forming galaxies in the local Universe (Saintonge et al. 2017; Catinella et al. 2018), there is evidence for the HI mass to be the sub-dominant component already at $z \sim 2.5$ (contributing no more than ~20% to the total gas mass), as based on dynamical mass arguments (Wuyts et al. 2016; Price et al. 2016). In addition, we have ignored the fraction of metals expelled from outflows that might be reaccreted onto galaxies via galactic fountains from the enriched

---

[8] Or that, at least, the metallicity of the infalling gas is much lower than that of the outflowing gas.

[9] Defined as the total mass fraction returned into the ISM by a stellar generation.
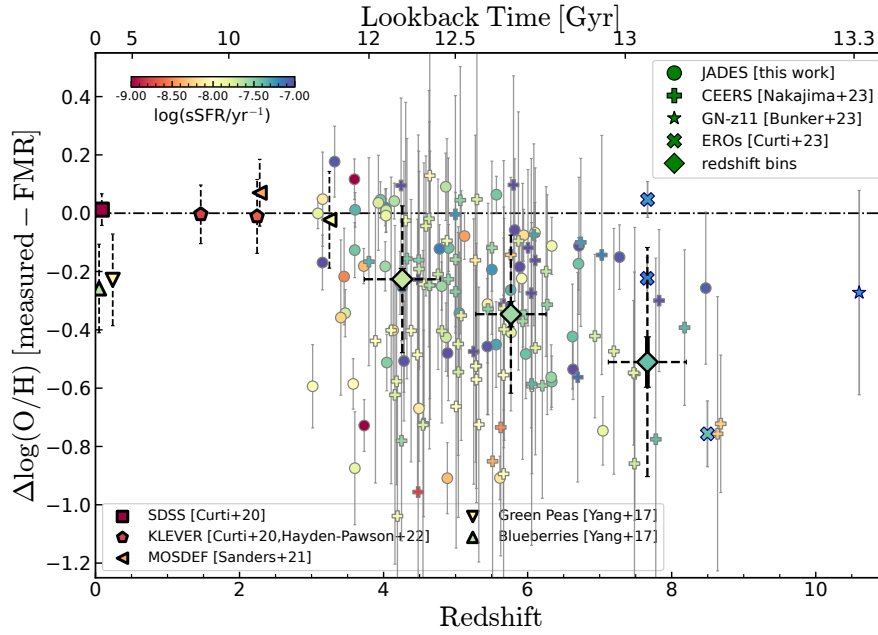
[10] We note that assuming the quadratic form of the scaling relation between μ_{gas} and redshift instead would produce a turnover when extrapolating at $z > 3$, that is the gas fraction would be decreasing with redshift.

Fig. 7. Deviations of the JWST sample from the predictions of the local FMR plotted as a function of redshift. Marker symbols are as in Fig. 5, with each galaxy colour-coded by its sSFR. The weighted average of the deviation in three redshift bins (colour-coded by the mean sSFR of the binned sample) for both the JWST sample and for galaxy samples at lower redshift (SDSS at $z \sim 0$, KLEVER at $z \sim 1.5-2.5$, MOSDEF at $z \sim 3$) are marked by larger symbols. Dashed error bars report the dispersion in $\Delta\log(O/H)$ of each binned sample, whereas thick, solid bars report the standard error on the mean (which is often smaller than the marker's size). Overall, $z > 3$ galaxies observed by JWST are scattered across the local FMR predictions, but show evidence of being preferentially offset towards lower metallicity values than expected with increasing redshift, with an average offset of -0.4 dex at $z > 6$.

CGM (Fraternali & Binney 2008; Anglés-Alcázar et al. 2017; Péroux et al. 2020), although accreted metals are generally modelled to be relevant for setting the MZR shape only at much lower redshift than explored here (Davé et al. 2011). Finally, systematics in the adopted metallicity calibrations can also shift the normalisation of the observed MZR by non-negligible (~ 0.1–0.15 dex) amounts. We defer a more detailed modelling of the MZR evolution in the framework of analytical chemical evolution models, as well as a proper assessment of the involved systematics, to a forthcoming paper of this collaboration, which will exploit the improved statistics provided by the full JADES dataset.

5. Possible evidence of evolution in the fundamental metallicity relation

It is known that the observed distribution of galaxies on the mass-metallicity plane reflects selection effects associated with the average star-formation rate probed by the samples under study (Yates et al. 2012; Salim et al. 2015; Telford et al. 2016; Cresci et al. 2019). In this sense, a least biased picture of the chemical evolution stage of these early galaxies might be given by considering the full relationship between $M_\star$, O/H, and SFR, in what is usually referred to as the fundamental metallicity relation (FMR; Mannucci et al. 2010), which takes into account the impact of the current SFR on the expected level of chemical enrichment of the ISM, as driven by the interplay between gas accretion, star formation, and outflows.

In Fig. 7 we report the deviation of the measured metallicity in our sample, as a function of redshift, from the predictions of the locally defined FMR as parametrised in Curti et al. (2020b). Individual points in Fig. 7 are colour-coded by their specific star-

formation-rate (sSFR=SFR/$M_\star$). The error bars on the y-axis includes not only the uncertainty on the individual metallicity measurements, but also an additional term (co-added in quadrature) associated with the FMR predictions as derived from the standard deviation of the distribution of log(O/H) obtained by varying one hundred times the input $M_\star$ and SFR of each galaxy within their uncertainties. The weighted median values in three bins of redshift ($z \in [3-5]$; $z \in [5-7]$; $z \in [7-10]$) for our combined JWST sample are shown by larger diamond markers to highlight possible trends, with solid and dashed errorbars representing the error on the median (as derived from bootstrapping by varying each point 300 times within its uncertainty) and the dispersion of the sample in each bin, respectively. The average offset from the FMR as measured by galaxy samples at lower redshifts ($z < 3$) as compiled from the literature (SDSS at $z \sim 0$, Curti et al. 2020b; 'Blueberry' and 'Green Pea' galaxies at $z \sim 0.2$, Yang et al. 2017a,b; KLEVER at $z \sim 1.5-2.5$, Curti et al. 2020a; Hayden-Pawson et al. 2022; MOSDEF at $z \sim 2-3$, Sanders et al. 2021) are also shown with different symbols as outlined in the legend, and colour-coded by the average sSFR value of the relative sample.

Individual galaxies are mostly offset below the FMR at any redshift $z > 3$, with 83 over 146 sources deviating from the FMR predictions at more than 1σ (48 percent of galaxies at $\langle z \rangle \sim 4$, 54 percent at $\langle z \rangle \sim 6$, and 87 percent at $\langle z \rangle \sim 8$), and 25 at more than 3σ (19 percent of galaxies at $\langle z \rangle \sim 4$, 14 percent at $\langle z \rangle \sim 6$, and 19 percent at $\langle z \rangle \sim 8$). We perform a two-sample Kolmogorov-Smirnov (KS) test on the two distributions, namely the JWST galaxy sample and the full SDSS sample used to parametrise the FMR from Curti et al. (2020b; which is well approximated by a normal distribution with *mean* = 0 and σ = 0.054 dex), which are found to be significantly different (p-value $\ll 0.01$), both for the global sample and within all individual redshift bins.
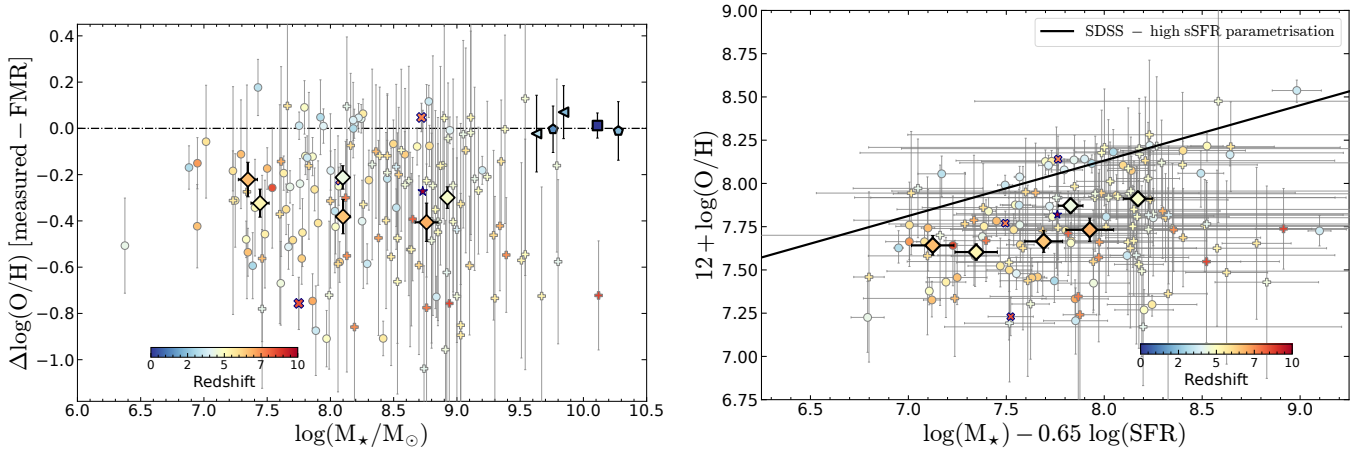
Fig. 8. FMR offset as a function of stellar mass and the projection of minumum scatter. *Left-hand panel:* deviations from the FMR plotted as a function of stellar mass. Symbols are as in Figs. 5 and 7, with individual points (and $M_\star$–redshift bins) colour-coded according to redshift. At fixed stellar mass, JWST galaxies are preferentially offset towards lower O/H with incresing redshift, and a similar trend is also seen with increasing $M_\star$ in the highest redshift bin. *Right-hand panel:* two-dimensional projection of the $M_\star$-SFR-O/H space in the O/H vs $\mu = \log(M_\star)$-$\alpha\log(\mathrm{SFR})$ plane, assuming the parametrisation for high sSFR ($\log(\mathrm{sSFR/yr^{-1}}) > -9.5$) galaxies inferred from the analysis of the SDSS sample ($\alpha = 0.65$, Curti et al. 2020b). Most of the individual galaxies, as well as the median values in the $M_\star$-redshift bins, are offset below the relation, further suggesting that the local FMR parametrisation does not hold at the very high-z and low-$M_\star$ probed in this work.

Looking more closely at the trends in the redshift bins, the median $\Delta\log(\mathrm{O/H})$ at $\langle z \rangle \sim 4$ is $= -0.23 \pm 0.04$ dex (and $1 - \sigma$ dispersion of 0.29 dex), the median deviation at $\langle z \rangle \sim 4$ is $= -0.34 \pm 0.05$ (with a dispersion of 0.25 dex), whereas the median deviation at $\langle z \rangle \sim 8$ is $= -0.50 \pm 0.10$ (with a dispersion of 0.38 dex). A Z-test confirms that these are all significantly offset at more than 5σ from the FMR predictions. We note how-ever that we are here neglecting any contribution from additional systematic uncertainties on the measured quantities ($M_\star$, SFR, O/H), as well as further uncertainties associated with the extrap-olation of the FMR parametrisation outside the parameter space of its calibration sample. However, including a further 0.25 dex systematic uncertainty on the individual metallicity predictions evaluated at the average $M_\star$ and SFR of the JWST sample (fol-lowing Fig. 11 in Curti et al. 2020b), the significance of the devi-ation of the median values from the predictions of the local FMR only slightly decreases to $\sim 4\sigma$ in each redshift bin.

Furthermore, we identify a tentative trend in which galax-ies are seen to sit preferentially below the FMR predictions with increasing redshift, with in particular sources at $z > 6$ that are significantly less enriched than their $M_\star$ and SFR would predict were they local galaxies. A Spearman correlation test between $\Delta\log(\mathrm{O/H})$ and redshift z on the full JWST sample provides evi-dence for a weak, though significant negative correlation, with $\rho = -0.18$, p-value $= 0.03$. Moreover, high redshift galaxies are also characterised, on average, by higher specific star-formation rates (sSFR) than lower redshift systems, forming stars at a higher pace while rapidly building-up their mass. A Spearman correlation test between $\Delta\log(\mathrm{O/H})$ and sSFR indeed reveals a moderate ($\rho = 0.29$), significant correlation (p-value < 0.01).

Finally, in the left-hand panel of Fig. 8 we plot the same metallicity deviation from the FMR as a function of stellar mass, reporting both individual galaxies and median values in the same $M_\star$-z bins described in Table 2, here colour-coded by their (median) redshift. Although no clear trend can be identified from individual points, with galaxies showing different levels of offset from the FMR regardless of their stellar mass, from the median values we find weak evidence for higher redshift galax-ies to be preferentially offset from the FMR at fixed stellar mass,

especially in the two highest $M_\star$ bins, where the offset between median points at different redshift is significant at 2.3σ and 1.2σ, respectively. No clear trend is instead identified with $M_\star$, at fixed redshift.

Although a more statistically representative sample of early galaxies assembled by JWST is required to either robustly con-firm or revisit such trends, these observations already suggest a few possible interpretations. On the one hand, they might reveal prominent accretion of pristine gas at high-redshift, happening on timescales shorter than the gas depletion, star-formation, ISM enrichment and mixing timescales compared to lower redshift galaxies, increasing the gas reservoir and diluting the galaxy metal content at fixed $M_\star$ and SFR (Dekel et al. 2009; Lilly et al. 2013; Somerville & Davé 2015; Davé et al. 2017). This could also be reflected into an increased level of stochasticity in the star-formation history during such early phases of galaxy forma-tion, especially at low stellar masses. If the assembly of early galaxies is primarily driven by stochastic gas accretion from the cosmic web, as often modelled with a scatter term in the baryonic accretion rate parameter (e.g. Forbes et al. 2014), the relatively longer timescales over which the star-formation and enrichment from supernovae can balance the variations in the gas reservoir would make a system more likely to be observed out of the equilibrium which is at the baseline of the scaling rela-tions observed at lower redshift (Davé et al. 2011; Zahid et al. 2012). At the same time, the efficiency of metal removals could also be enhanced, especially in galaxies with high specific star-formation rate, such that the metal loading factor of outflows is higher, at fixed SFR, than that of galaxies following the observed FMR at lower redshifts (Baker et al. 2023).

A possible way to differentiate between these different sce-narios is to study the morphology of galaxies with the posi-tion relative to the FMR. For instance, Tacchella et al. (2023b) analysed the three galaxies at $z \sim 8$ from the EROs, find-ing that the galaxy with the lowest gas-phase metallicity is very compact and consistent with rapid gas accretion. The two other objects with relatively higher gas-phase metallicity show more complex multi-component morphologies on kpc scales, indicating that their recent increase in star-formation rate is

driven by mergers or internal gravitational instabilities. More recently, Langeroodi & Hjorth (2023) investigated the correlation between the deviation from the FMR at high redshift and the 'compactness' of galaxies, finding more compact sources to be preferentially offset, and interpreting this result as further evidence for enhanced stochasticity in the accretion and star-formation episodes, happening on timescales much shorter than those associated with the ISM enrichment and therefore diluting the metal content of such compact galaxies more than the average scaling relation would predict.

As seen from a different perspective, the apparent inconsistency between the measured oxygen abundance and the FMR predictions reveals the inadequacy of the local formalism when extrapolated to such early epochs and low stellar masses. Within the framework originally described in Curti et al. (2020b) in fact, the SFR dependence is only embedded in the 'turnover mass' term (i.e. the value in $M_\star$ at which the slope flattens), which is described to scale linearly with the star-formation rate. Therefore, no direct dependence of the MZR slope on the SFR is included in such FMR parametrisation, nor it is, by definition, in any of the two-dimensional projections explored in the literature in the form $\mu_\alpha = \log(M_\star) - \alpha\log(SFR)$ versus $\log(O/H)$ (e.g. Mannucci et al. 2010; Andrews & Martini 2013), which implicitly assumes no SFR-dependence of the MZR slope. In fact, a rotation of the $M_\star$ and SFR axis in the 3D space removes the apparent secondary dependence of the MZR on the star-formation rate only if the slopes of the different mass-metallicity relations are invariant for samples with different average star-formation rates. Then, if galaxies follow the same scaling relations between $M_\star$, SFR, and O/H over cosmic time (as in the original FMR framework), such a projection is capable to cancel out any apparent evolution of the MZR with redshift. On the contrary, if such an assumption does not hold anymore, and the low-mass end slope of the MZR is SFR-dependent, then a single 2D projection is not capable of capturing the full interplay between gas flows, star-formation an chemical enrichment which manifests in the existence of a relationship between the observational quantities $M_\star$, SFR, and O/H.

In Curti et al. (2020b), such a possibility was briefly discussed, and two different regimes were identified and explored, providing two different parametrisations of the FMR based on the sSFR of galaxies. In this scenario, the slope of the MZR is still invariant with SFR, but the relative strength of the SFR-dependence is significantly different in the two regimes, being much more prominent for high sSFR galaxies, which translates into two different μ_α projections, with $\alpha = 0.22$ at low sSFR, and $\alpha = 0.65$ at high sSFR[11]. However we stress that, being based on SDSS galaxies, these parametrisations are still statistically dominated by galaxies at higher $M_\star$ compared to the regime explored in the present work. For reference, we plot our combined JWST sample presented in this work in the O/H versus μ_α projection, for $\alpha = 0.65$, in the right-hand panel of Fig. 8: such relation, though calibrated on the highest sSFR galaxies in the SDSS, is not yet capable to account for the full redshift evolution in the MZR as seen at $z = 3-10$, and most of the objects (as well as their median binned values) clearly fall below the best-fit line (in black).

Adopting different realisations of the FMR can impact the level of offset of high-redshift galaxies (and its significance), but does not cancel out nor invert the observed trend. In this sense, a comparison between the observed metallicities for the JWST sample and the FMR parametrisation from Andrews & Martini (2013) is presented and discussed in Appendix A.

---

[11] The best-fit α value for the global sample is $\alpha = 0.55$.

6. Summary and conclusions

We analysed the metallicity properties of a sample of low-mass ($\log(M_\star/M_\odot) \lesssim 8.5$) galaxies at $3 < z < 10$, exploiting deep NIRSpec spectroscopic observations from the JADES programme. Detection of multiple emission lines in both PRISM and medium-resolution grating spectra allowed us to derive the gas-phase metallicity in these sources, which we complemented with a sample of sources at $\log(M_\star/M_\odot) \sim 8.5-10$ from other programmes and compiled from the literature (CEERS, Nakajima et al. 2023; ERO, Curti et al. 2023; and GN-z11, Bunker et al. 2023b) to provide insights into the cosmic evolution of the metallicity scaling relations in the low-mass regime up to the epoch of early galaxy assembly. Our main findings can be summarised as follows:

– We find evidence for a mild evolution in the normalisation of the MZR at $z > 3$, as previously suggested by similar studies (Langeroodi et al. 2023; Heintz et al. 2023; Matthee et al. 2023; Nakajima et al. 2023; Shapley et al. 2023a). Our sample of galaxies spans a wide range in 12+log(O/H) (corresponding to $Z \sim 0.03-0.6\,Z_\odot$), with a median metallicity of 12+log(O/H) = 7.77 ($Z = 0.12\,Z_\odot$). Compared to a simple extrapolation in the low-$M_\star$ regime of the MZR $z \sim 3.3$ from (Sanders et al. 2021), the JWST sample shows an average offset in log(O/H) of only 0.05 dex towards lower metallicity (negligible at $\log(M_\star/M_\odot) \lesssim 8$, more prominent at $\log(M_\star/M_\odot) \sim 9$).

– An orthogonal distance regression fit performed on the full JWST sample suggests a flattening of the slope at the low-mass end of the MZR at $z = 3-10$ (Fig. 5). This is in agreement with previous findings at $z \sim 3$ in the same stellar-mass regime as reported by Li et al. (2023), compared to which we nevertheless measure a more prominent evolution in the MZR normalisation of $\sim$0.25 dex. The flattening of the MZR slope at low masses could indicate a change in the dominant mechanisms regulating supernovae-driven outflows compared to higher mass galaxies. The slope measured for our combined sample of galaxies at $z = 3-10$, $\beta_{3-10} = 0.17 \pm 0.03$, is in good agreement with predictions from chemical evolution models implementing a 'momentum-driven' feedback mode for SNe winds (right panel of Fig. 6).

– Theoretical simulations of the high-redshift Universe generally predict steeper MZR slopes than observed, broadly matching the normalisation of the relation at $\log(M_\star/M_\odot) \sim 8-9$ but struggling to reproduce the metallicity at the lowest masses probed by the present work (right panel of Fig. 6). In the framework of analytical chemical evolution models (e.g. Peeples & Shankar 2011), evolving the gas fraction with redshift according to the same scaling relations inferred up to $z \sim 3$ (Tacconi et al. 2018) would almost entirely account (fixing all other parameters) for the evolution in the MZR normalisation at $\log(M_\star/M_\odot) \sim 8$ as observed between $z \sim 3$ and $z \sim 5.2$, once the flattening of the MZR slope in this mass and redshift regime is considered.

– Notably, our galaxy sample is distributed over a similar region (and with a fully consistent inferred slope) of the mass–metallicity plane as low-redshift, low-$M_\star$, metal-poor systems with extreme ionisation conditions, such as the 'Blueberry' and 'Green Pea' galaxies (Yang et al. 2017a,b) (left panel of Fig. 6). These objects have been identified as local analogues of high-redshift galaxies, and we observe that they share similar metallicity properties with those pertaining to the high-z sources analysed in this paper (see also Cameron et al. 2023b).

– We find evidence for a deviation of our JWST sample from the framework of the fundamental metallicity relation (FMR) as described for local galaxies. In particular, galaxies appear significantly metal deficient compared to local galaxies matched in $M_\star$ and SFR, and this trend is observed more prominently as increased redshift (Fig. 7). This suggests that in these objects metals are either more efficiently removed by SNe-driven outflows, or that the gas is strongly diluted by stochastic accretion of (nearly) pristine gas, happening on timescales shorter than those associated with the enrichment from subsequent star formation.

– This behaviour also highlights potential inconsistencies in the assumed FMR parametrisation at very high z, with the strength of the metallicity–SFR dependence at fixed $M_\star$ being no longer redshift invariant. In fact, if the low-mass end slope of the MZR depends on the average SFR of galaxies, a simple two-dimensional projection of the FMR in the log(O/H) versus μ_α plane is not capable of capturing the metallicity variations as caused by the interplay of dilution, enrichment, and metal-loaded outflows driven by star formation (right panel of Fig. 8).

The observations presented in this paper represent a step towards understanding the chemical enrichment processes in place during early galaxy formation. Deep JWST observations have already proven to be capable of extending the mass and luminosity regimes probed by previous facilities by orders of magnitudes, extending our investigations up to the highest redshift galaxies ever discovered. Improving the statistical robustness of these samples in the near future will be crucial for forthcoming studies in order to further reduce the large systematic uncertainties and to provide a more comprehensive picture of galaxy formation across different cosmic epochs, a goal fully within the reach of JWST capabilities.

References

Allende Prieto, C., Lambert, D. L., & Asplund, M. 2001, ApJ, 556, L63
Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140
Anglés-Alcázar, D., Faucher-Giguère, C.-A., Kereš, D., et al. 2017, MNRAS, 470, 4698
Arellano-Córdova, K. Z., Berg, D. A., Chisholm, J., et al. 2022, ApJ, 940, L23
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Baker, W. M., & Maiolino, R. 2023, MNRAS, 521, 4173
Baker, W. M., Maiolino, R., Belfiore, F., et al. 2023, MNRAS, 519, 1149
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Belli, S., Jones, T., Ellis, R. S., & Richard, J. 2013, ApJ, 772, 141
Bian, F., Kewley, L. J., & Dopita, M. A. 2018, ApJ, 859, 175
Birkmann, S. M., Böker, T., Ferruit, P., et al. 2011, SPIE Conf. Ser., 8150, 81500B
Böker, T., Birkmann, S., de Marchi, G., et al. 2012, SPIE Conf. Ser., 8442, 84423F
Böker, T., Beck, T. L., Birkmann, S. M., et al. 2023, PASP, 135, 038001
Bonaventura, N., Jakobsen, P., Ferruit, P., Arribas, S., & Giardino, G. 2023, A&A, 672, A40
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bunker, A. J., Cameron, A. J., Curtis-Lake, E., et al. 2023a, A&A, submitted [arXiv:2306.02467]
Bunker, A. J., Saxena, A., Cameron, A. J., et al. 2023b, A&A, 677, A88
Cameron, A. J., Katz, H., Rey, M. P., & Saxena, A. 2023a, MNRAS, 523, 3516
Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023b, A&A, 677, A115
Cappellari, M. 2017, MNRAS, 466, 798
Cappellari, M. 2023, MNRAS, 526, 3273
Catinella, B., Saintonge, A., Janowiecki, S., et al. 2018, MNRAS, 476, 875
Chabrier, G. 2003, PASP, 115, 763
Chevallard, J., & Charlot, S. 2016, MNRAS, 462, 1415
Choi, J., Dotter, A., Conroy, C., et al. 2016, ApJ, 823, 102
Conroy, C., Naidu, R. P., Zaritsky, D., et al. 2019, ApJ, 887, 237
Cresci, G., Mannucci, F., & Curti, M. 2019, A&A, 627, A42
Curti, M., Cresci, G., Mannucci, F., et al. 2017, MNRAS, 465, 1384
Curti, M., Maiolino, R., Cirasuolo, M., et al. 2020a, MNRAS, 492, 821
Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020b, MNRAS, 491, 944
Curti, M., D'Eugenio, F., Carniani, S., et al. 2023, MNRAS, 518, 425
Curtis-Lake, E., Carniani, S., Cameron, A., et al. 2023, Nat. Astron., 7, 622
Davé, R., Finlator, K., & Oppenheimer, B. D. 2011, MNRAS, 416, 1354
Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98
Davé, R., Rafieferantsoa, M. H., Thompson, R. J., & Hopkins, P. F. 2017, MNRAS, 467, 115
Dekel, A., Birnboim, Y., Engel, G., et al. 2009, Nature, 457, 451
Eisenstein, D. J., Willott, C., Alberts, S., et al. 2023, ApJS, submitted, [arXiv:2306.02465]
Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, ApJ, 672, L107
Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813
Ferruit, P., Jakobsen, P., Giardino, G., et al. 2022, A&A, 661, A81
Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, ApJ, 946, L13
Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181
Forbes, J. C., Krumholz, M. R., Burkert, A., & Dekel, A. 2014, MNRAS, 443, 168
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Fraternali, F., & Binney, J. J. 2008, MNRAS, 386, 935
Giardino, G., Birkmann, S., Robberto, M., et al. 2019, PASP, 131
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93
Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279
Guo, Y., Koo, D. C., Lu, Y., et al. 2016, ApJ, 822, 103
Hayden-Pawson, C., Curti, M., Maiolino, R., et al. 2022, MNRAS, 512, 2867
Heintz, K. E., Brammer, G. B., Giménez-Arteaga, C., et al. 2023, Nat. Astron., 7, 1517
Henry, A., Martin, C. L., Finlator, K., & Dressler, A. 2013a, ApJ, 769, 148
Henry, A., Scarlata, C., Domínguez, A., et al. 2013b, ApJ, 776, L27
Hirschauer, A. S., Salzer, J. J., Janowiecki, S., & Wegner, G. A. 2018, AJ, 155, 82
Jakobsen, P., Ferruit, P., Alves de Oliveira, C., et al. 2022, A&A, 661, A80
Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, ApJS, 254, 22
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 54
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., Maier, C., Yabe, K., et al. 2013, ApJ, 774, L10
Kobayashi, C., & Taylor, P. 2023, arXiv e-prints [arXiv:2302.07255]
Langan, I., Ceverino, D., & Finlator, K. 2020, MNRAS, 494, 1988

Langeroodi, D., & Hjorth, J. 2023, arXiv e-prints [arXiv:2307.06336]
Langeroodi, D., Hjorth, J., Chen, W., et al. 2023, ApJ, 957, 39
Laseter, I. H., Maseda, M. V., Curti, M., et al. 2024, A&A, 681, A70
Lee, H., Skillman, E. D., Cannon, J. M., et al. 2006, ApJ, 647, 970
Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155
Li, M., Cai, Z., Bian, F., et al. 2023, ApJ, 955, L18
Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119
Looser, T. J., D'Eugenio, F., Maiolino, R., et al. 2023, Nature, submitted [arXiv:2302.14155]
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140
Maier, C., Lilly, S. J., Ziegler, B. L., et al. 2014, ApJ, 792, 3
Maiolino, R., & Mannucci, F. 2019, A&A Rev., 27, 3
Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463
Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2023, A&A, submitted, [arXiv:2308.01230]
Maiolino, R., Scholtz, J., Witstok, J., et al. 2024, Nature, 627, 59
Mannucci, F., Cresci, G., Maiolino, R., et al. 2009, MNRAS, 398, 1915
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Mascia, S., Pentericci, L., Calabrò, A., et al. 2023, A&A, 672, A155
Matthee, J., Mackenzie, R., Simcoe, R. A., et al. 2023, ApJ, 950, 67
Nakajima, K., & Ouchi, M. 2014, MNRAS, 442, 900
Nakajima, K., Ouchi, M., Xu, Y., et al. 2022, ApJS, 262, 3
Nakajima, K., Ouchi, M., Isobe, Y., et al. 2023, ApJS, 269, 33
Noeske, K. G., Faber, S. M., Weiner, B. J., et al. 2007, ApJ, 660, L47
Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457
Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, ApJ, 819, 129
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei, 2nd edn. (University Science Books)
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2022, MNRAS, 513, 5621
Peeples, M. S., & Shankar, F. 2011, MNRAS, 417, 2962
Péroux, C., Nelson, D., van de Voort, F., et al. 2020, MNRAS, 499, 2462
Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59
Pilyugin, L. S., Mattsson, L., Vílchez, J. M., & Cedrés, B. 2009, MNRAS, 398, 485
Planck Collaboration VI. 2020, A&A, 641, A6
Pontoppidan, K. M., Barrientes, J., Blome, C., et al. 2022, ApJ, 936, L14
Popesso, P., Concas, A., Cresci, G., et al. 2023, MNRAS, 519, 1526
Price, S. H., Kriek, M., Shapley, A. E., et al. 2016, ApJ, 819, 80
Reddy, N. A., Topping, M. W., Shapley, A. E., et al. 2022, ApJ, 926, 31
Renzini, A., & Peng, Y.-J. 2015, ApJ, 801, L29
Rhoads, J. E., Wold, I. G. B., Harish, S., et al. 2023, ApJ, 942, L14
Rieke, M. J., Robertson, B., Tacchella, S., et al. 2023, ApJS, 269, 16
Robertson, B. E., Tacchella, S., Johnson, B. D., et al. 2023, Nat. Astron., 7, 611
Saintonge, A., Catinella, B., Cortese, L., et al. 2016, MNRAS, 462, 1749
Saintonge, A., Catinella, B., Tacconi, L. J., et al. 2017, ApJS, 233, 22
Salim, S., Lee, J. C., Dav'e, R., & Dickinson, M. 2015, ApJ, 808, 25
Sanders, R. L., Shapley, A. E., & Kriek, M., et al. 2018, ApJ, 858, 99
Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, ApJ, 914, 19
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2023, ApJ, 955, 54
Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, ApJ, 962, 24
Schaerer, D., Marques-Chaves, R., Barrufet, L., et al. 2022, A&A, 665, L4
Scholtz, J., Maiolino, R., D'Eugenio, F., et al. 2023, A&A, submitted, [arXiv:2311.18731]
Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, ApJ, 837, 150
Shapley, A. E., Coil, A. L., Ma, C.-P., & Bundy, K. 2005, ApJ, 635, 1006
Shapley, A. E., Reddy, N. A., Sanders, R. L., Topping, M. W., & Brammer, G. B. 2023a, ApJ, 950, L1
Shapley, A. E., Sanders, R. L., Reddy, N. A., Topping, M. W., & Brammer, G. B. 2023b, ApJ, 954, 157
Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165
Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164
Tacchella, S., Eisenstein, D. J., Hainline, K., et al. 2023a, ApJ, 952, 74
Tacchella, S., Johnson, B. D., Robertson, B. E., et al. 2023b, MNRAS, 522, 6236
Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179
Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, ARA&A, 58, 157
Taylor, A. J., Barger, A. J., & Cowie, L. L. 2022, ApJ, 939, L3
Telford, O. G., Dalcanton, J. J., Skillman, E. D., & Conroy, C. 2016, ApJ, 827, 35
Topping, M. W., Shapley, A. E., Sanders, R. L., et al. 2021, MNRAS, 506, 1237
Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, MNRAS, 484, 5587
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Treu, T., Roberts-Borsani, G., Bradac, M., et al. 2022, ApJ, 935, 110
Trump, J. R., Arrabal Haro, P., Simons, R. C., et al. 2023, ApJ, 945, 35
Übler, H., Maiolino, R., Curtis-Lake, E., et al. 2023, A&A, 677, A145
Ucci, G., Dayal, P., Hutter, A., et al. 2023, MNRAS, 518, 3557
van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, ApJ, 788, 28
Vidal-García, A., Charlot, S., Bruzual, G., & Hubeny, I. 2017, MNRAS, 470, 3532
Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, MNRAS, 455, 4183
Williams, H., Kelly, P. L., Chen, W., et al. 2023, Science, 380, 416
Witstok, J., Smit, R., Maiolino, R., et al. 2021, MNRAS, 508, 1686
Wuyts, E., Rigby, J. R., Sharon, K., & Gladders, M. D. 2012, ApJ, 755, 73
Wuyts, E., Kurk, J., Förster Schreiber, N. M., et al. 2014, ApJ, 789, L40
Wuyts, E., Förster Schreiber, N. M., Wisnioski, E., et al. 2016, ApJ, 831, 149
Yabe, K., Ohta, K., Iwamuro, F., et al. 2014, MNRAS, 437, 3647
Yang, H., Malhotra, S., Gronke, M., et al. 2017a, ApJ, 844, 171
Yang, H., Malhotra, S., Rhoads, J. E., & Wang, J. 2017b, ApJ, 847, 38
Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215
Yates, R. M., Schady, P., Chen, T. W., Schweyer, T., & Wiseman, P. 2020, A&A, 634, A107
Zahid, H. J., Kewley, L. J., & Bresolin, F. 2011, ApJ, 730, 137
Zahid, H. J., Bresolin, F., Kewley, L. J., Coil, A. L., & Davé, R. 2012, ApJ, 750, 120
Zahid, H. J., Dima, G. I., Kudritzki, R.-P., et al. 2014, ApJ, 791, 130

1 European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany
   e-mail: mirko.curti@eso.org
2 Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
3 Cavendish Laboratory – Astrophysics Group, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK
4 Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK
5 Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
6 Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
7 Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
8 School of Physics, University of Melbourne, Parkville 3010, VIC, Australia
9 ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Mount Stromlo Rd, Australian Capital Territory 2611 AU, Australia
10 Department of Astronomy, University of Wisconsin-Madison, 475 N. Charter St., Madison, WI 53706, USA
11 Centro de Astrobiología (CAB), CSIC-INTA, Cra. de Ajalvir Km. 4, 28850 Torrejón de Ardoz, Madrid, Spain
12 ATG Europe for the European Space Agency, ESTEC, Noordwijk, The Netherlands
13 European Space Agency (ESA), European Space Astronomy Centre (ESAC), Camino Bajo del Castillo s/n, 28692 Villanueva de la Cañada, Madrid, Spain
14 Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK
15 NRC Herzberg, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada
16 Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA
17 Department of Physics and Astronomy, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA
18 Steward Observatory University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA
19 Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA
20 NSF's National Optical-Infrared Astronomy Research Laboratory, 950 North Cherry Avenue, Tucson, AZ 85719, USA
21 European Space Agency, ESA/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
22 AURA for European Space Agency, Space Telescope Science Institute, 3700 San Martin Drive., Baltimore, MD 21210, USA

Appendix A: Metallicity scaling relations under different assumptions as to sample selection and derivation of galaxy properties

In the main body of the paper, we discuss the mass–metallicity relation (MZR) and the fundamental metallicity relation (FMR) for a combined sample of galaxies observed with the *JWST* from different observational programmes. Here, we investigate whether any of the main results presented in the paper change if we adopt different choices from our set of fiducial assumptions, particularly regarding sample selection, SFR measurement, and FMR parametrisation.

In Figure A.1 we show the MZR (left-hand panel) and deviations from the FMR (right-hand panel) for our sample of galaxies which now includes 14 objects from JADES identi- fied as AGN candidates ((Scholtz et al. 2023), highlighted in cyan). For these galaxies, we here assume a mild contribu- tion of AGN ionisation to the emission line spectrum, so that the standard metallicity calibrations adopted for 'normal' star- forming galaxies can be applied. Stellar masses and SFRs are also derived assuming negligible AGN contribution to the con- tinuum level; for what concern the SFR however, any con- tribution to Hα emission from the narrow-line region would imply the inferred SFR have to be considered as upper lim- its. No significant change to the observed evolution in the metallicity scaling relations is found when including these objects in the sample, with a Spearman correlation $\rho = -0.18$, p-value=0.017.

In Figure A.2 instead we report, similarly to Figure 7, the deviations from the predictions of the local FMR, but this time assuming different estimates for the SFR of our entire *JWST* sample. In particular, we adopt the calibration for metal poor galaxies described in Reddy et al. (2022), Shapley et al. (2023b) in the left-hand panel, whereas the calibration from Kennicutt & Evans (2012), and suited for solar metallicities, is assumed in the right-hand panel. As already discussed in Section 3.2 of the main text, once estimated on JADES galax- ies the former provides on average lower (s)SFRs by 0.14 dex, whereas the latter delivers higher (s)SFR by 0.23 dex, compared to our fiducial values based on BEAGLE fitting to both PRISM spectra and photometry. This means that the predicted metal- licities from the FMR (i.e. Δlog(O/H)) are, on average, higher (lower) than our fiducial case when the low-metallicity (solar metallicity) calibration to convert the Hα flux to SFR is adopted. Nonetheless, the median trends and the observed evolution in the FMR at high-z are robust against the choice of the SFR calibra- tion (Spearman $\rho = -0.17$, p-value= 0.04 in both cases.)

Furthermore, in Figure A.3 we compare the metallicity for our *JWST* sample with the predictions based on the FMR described as in Andrews & Martini (2013), which is parametri- sation in the form O/H versus μ_α, with μ_α=0.66. Compared to Curti et al. (2020b), the Andrews & Martini (2013) FMR is

based on 'direct' (Te-based) metallicities derived from stacked spectra in bins of M$_\star$ and SFR, rather than on individual galaxies with strong-line-based (though T_e-calibrated) metallicities. On the one hand, this allows to extend the parametrisation towards lower masses, while on the other, sample selection and weight- ing effects can possibly bias low the metallicity of stacked spec- tra dominated by few bright [O III] λ4363 emitters. As a conse- quence, the inferred SFR-dependence of O/H at (fixed) low M$_\star$ is stronger for Andrews & Martini (2013; with a slope of the lin- ear O/H vs. μ_α relation of 0.43), which manifests as a lower pre- dicted O/H at low μ_α, and therefore a smaller offset, for high-z galaxies (see also Nakajima et al. 2023).

Nonetheless, z > 3 galaxies still appear to lie preferentially below the FMR predictions, as shown in the left-hand panel of Figure A.3, with median values in redshift bins offset by - 0.07 dex at z~ 4 (though now marginally consistent at 2σ with the FMR given the error on the median), -0.21 dex at z~ 6 (6σ significance), and -0.39 dex at z~ 8 (5σ significance).

In the middle panel of Figure A.3 we show instead the median values in M$_\star$ and redshift bins for the *JWST* sam- ple on the O/H vs μ_α projection. Median points at z~ 3 − 6 are still offset below the FMR by 0.135 dex, 0.073 dex, and 0.178 dex at log(M$_\star$)=7.4, 8.1, and 8.9, respectively (at 2.1σ, 1.5σ, and 3.6σ significance, respectively), whereas higher red- shift bins (z~ 6 − 10) show larger offsets (0.22 dex and 0.25 dex, at 3σ significance) at log(M$_\star$)>7.5, then only exception being the high-z bin at log(M$_\star$)=7.35 which is in agreement with the Andrews & Martini (2013) FMR.

Finally, it is important to remark once again that a dif- ferent choice of the SFR calibration can also affect the pre- dicted FMR-metallicity: this is further shown in the right-hand panel of Fig. A.3, where the same O/H vs μ_α projection for the Andrews & Martini (2013) FMR is shown when adopting the Hα-based SFR with the solar metallicity calibration from Kennicutt & Evans (2012). We note that adopting the Hα-based SFR and the calibration from Kennicutt & Evans (2012) for all our galaxies, which significantly reduces the number of ionis- ing photons per unit SFR compared to both our fiducial and the Hα-based, low-metallicity calibration scenario, provides a bet- ter agreement with the FMR predictions of Andrews & Martini (2013). This is in fact driven by the increase in the inferred SFR, which reduces μ_α and consequently moves the points towards the left along the x-axis of the diagram, and it observed in par- ticular for median-binned points at $3 < z < 6$ (which are now only mildy offset from the FMR predictions, and consistent with it within $\sim 2 - 3\sigma$), whereas higher redshift bins (z> 6) at log(M$_\star$)> 7.5 keep showing an offset below the FMR at $> 3\sigma$ significance. The choice of the SFR-calibration should therefore be evaluated as carefully as that of the metallicity calibrations, as it might introduce biases and produce tensions among different assessments of the evolution of these important scaling relations (Nakajima et al. 2023).
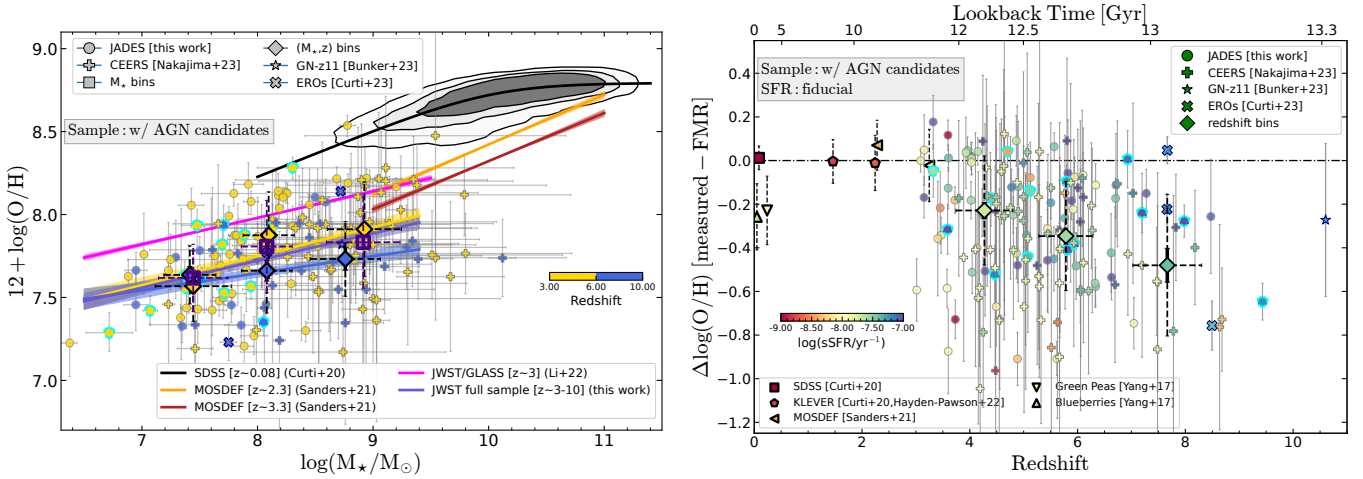
Fig. A.1. Similar to Figure 5 (left-hand panel) and Figure 7 (right-hand panel), but including the sample of 14 (narrow-line) AGN candidates in *JADES-deep* Scholtz et al. 2023, here highlighted with cyan contours. No clear change in the median trends of the metallicity scaling relations is found, and the main results reported in the paper are unaffected.
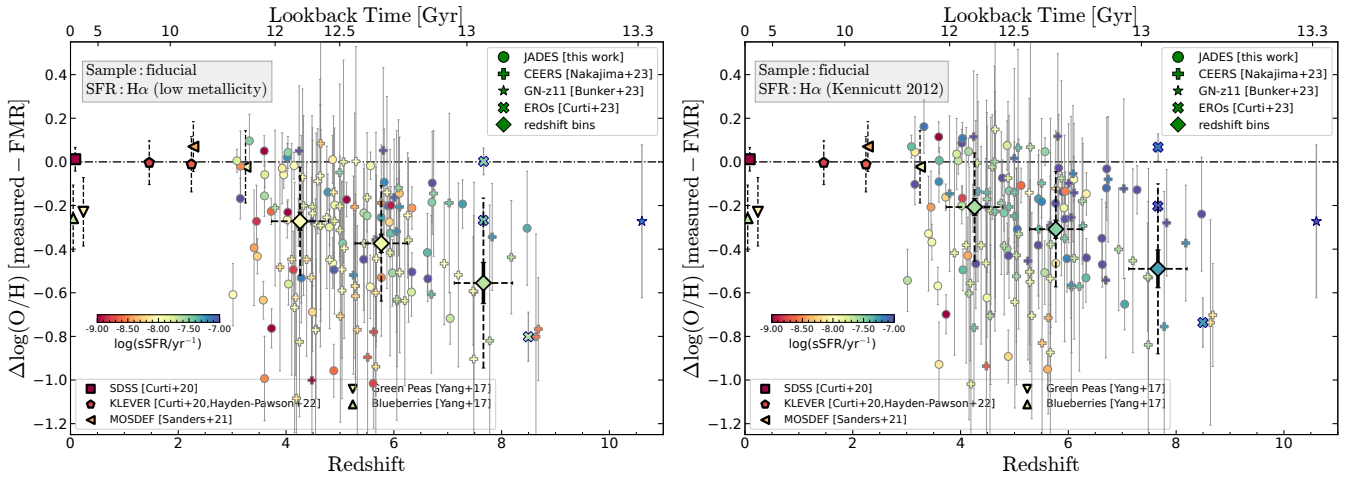
Fig. A.2. Same as Figure 7, but with SFR inferred from the (attenuation corrected) Hα flux (or Hβ, at $z \gtrsim 7$). The calibration for low metallicity galaxies described in Reddy et al. 2022; Shapley et al. 2023b is adopted for the entire *JWST* sample in the left-hand panel, while in the right-hand panel the SFR is derived following the calibration from Kennicutt & Evans 2012 for solar metallicity.
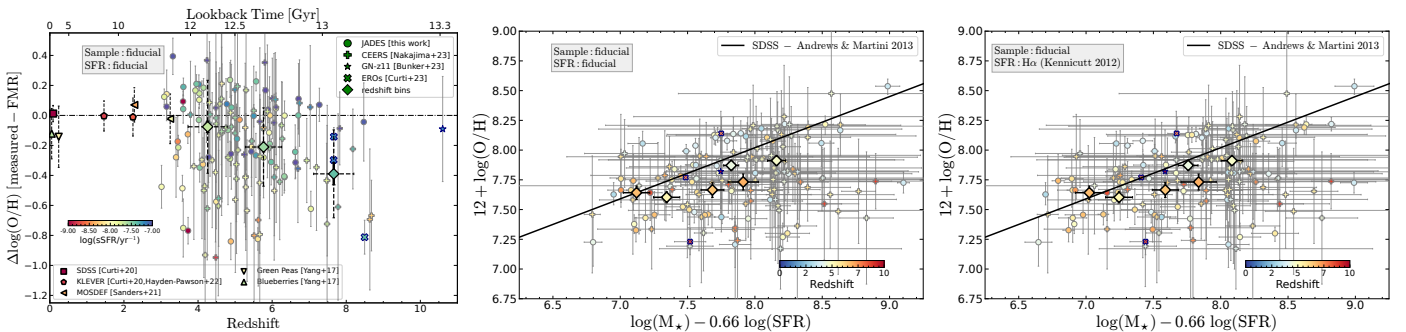
Fig. A.3. Comparison between observed metallicity for the *JWST* sample and the predictions of the FMR, here adopting the formalism from Andrews & Martini 2013. The three panels show, from left to right, the offset from the FMR as a function of redshift, and the measured versus FMR-predicted metallicities based on two different SFR estimates, namely the fiducial value adopted throughout the paper (and based on BEAGLE fitting to both R100 spectra and photometry), and the Hα-based SFR based on the calibration from Kennicutt & Evans 2012, respectively. In the latter case, the SFRs are higher on average by ~ 0.23 dex than our fiducial values, and by ~ 0.37 dex compared to the SFR-calibration for metal poor galaxies, decreasing μ_α and improving the agreement with the predictions of the Andrews & Martini 2013 parametrisation of the FMR, especially at $z \sim 4$; nonetheless, higher redshift ($z > 6$) galaxies still sit preferentially below the FMR, appearing metal deficient compared to local galaxies at fixed M$_\star$ and SFR.

Appendix B: Adopted metallicity calibrations

As discussed in Section 3.1, in this work we have adopted the set of strong-line metallicity calibrations presented in Curti et al. (2017) and Curti et al. (2020b), whose polynomial functions have been however slightly revised to better probe the low-metallicity regime. In addition, we have exploited the novel diagnostics $\hat{R}$ introduced by Laseter et al. (2024). In table B.1 we report the best-fit coefficients for each of the polynomial calibrations adopted in the paper. Each calibrations is parametrised in the form

$$\log(R) = \sum_{i=0}^{n} c_i x^i \,, \tag{B.1}$$

where $x = 12 + \log(O/H) - 8.69$, c_i are the coefficients, and R is the diagnostic line ratio (see Table 1).

Table B.1. Best-fit polynomial coefficients for the strong-line metallicity calibrations adopted in this paper.

Diagnostics	c_0	c_1	c_2	c_3	c_4	c_5
R2	0.4326	-1.0751	-5.1141	-5.5321	-2.3009	-0.2850
R3	-0.2768	-3.1422	-2.73	-0.6003	–	–
$\hat{R}$	-0.0478	-3.0707	-3.4164	-1.0034	-0.0379	–
R23	0.5145	-1.4633	-1.3891	-0.2847	–	–
O32	-0.6915	-2.6856	-1.1642	–	–	–
Ne3O2	-1.632	-2.0606	-0.46088	–	–	–
N2	-0.4725	1.0428	-5.2680	-10.1788	-6.699	-1.4815
S2	-0.4420	-0.3194	-6.6187	-10.5222	-6.5198	-1.4172

Appendix C: Summary of properties for JADES galaxies

We here report the derived redshift, stellar mass, star-formation rate, and metallicity for the full JADES galaxy sample analysed in this work. In addition, we indicate whether the metallicity has been derived from PRISM or grating spectra, which diagnostics have been used, if NIRCAM photometry has been included in the $M_\star$ and SFR derivation with BEAGLE, and whether a galaxy has been identified as a type-2 AGN candidate (Scholtz et al. 2023).

Table C.1. Summary of derived properties for JADES galaxies analysed in this paper.

NIRSpec ID	JADES ID JADES-GS	Redshift	$\log M_\star$ [$M_\odot$]	SFR [$M_\odot\,\mathrm{yr}^{-1}$]	SFR [Hα] [$M_\odot\,\mathrm{yr}^{-1}$]	$12 + \log(\mathrm{O/H})$	C ($\star$)	Metallicity diagnostics	NC (†)	AGN (‡‡)
2923	+53.15407−27.82094	3.015	$7.39^{+0.05}_{-0.05}$	$0.28^{+0.02}_{-0.02}$	0.23 ± 0.02	$7.44^{+0.16}_{-0.13}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	1	
21150	+53.16995−27.76840	3.088	$8.18^{+0.01}_{-0.01}$	$2.38^{+0.09}_{-0.08}$	2.57 ± 0.04	$8.12^{+0.05}_{-0.05}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
18322	+53.13171−27.77412	3.151	$6.88^{+0.02}_{-0.02}$	$0.79^{+0.02}_{-0.02}$	0.80 ± 0.02	$7.63^{+0.10}_{-0.09}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10040	+53.12304−27.79564	3.154	$7.92^{+0.13}_{-0.13}$	$0.64^{+0.17}_{-0.11}$	0.26 ± 0.08	$8.18^{+0.13}_{-0.18}$	M	R3, R2, $\hat{\mathrm{R}}$, O32	1	
19431	+53.16292−27.77141	3.320	$7.43^{+0.11}_{-0.11}$	$2.49^{+0.11}_{-0.11}$	0.86 ± 0.05	$8.06^{+0.10}_{-0.13}$	M	R3, R2, $\hat{\mathrm{R}}$, O32	1	
10013597	+53.12325−27.76939	3.323	$7.48^{+0.13}_{-0.13}$	$0.77^{+0.20}_{-0.12}$	0.28 ± 0.01	$7.94^{+0.10}_{-0.12}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	0	‡‡
3322	+53.12574−27.81921	3.406	$8.04^{+0.25}_{-0.25}$	$0.63^{+0.48}_{-0.15}$	0.39 ± 0.09	$7.81^{+0.12}_{-0.25}$	P	R3	1	
7629	+53.16983−27.80344	3.452	$8.45^{+0.09}_{-0.09}$	$0.86^{+0.06}_{-0.07}$	0.41 ± 0.02	$8.06^{+0.13}_{-0.20}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	1	
3184	+53.15010−27.81971	3.468	$8.52^{+0.11}_{-0.11}$	$6.11^{+0.66}_{-0.61}$	1.85 ± 0.03	$7.81^{+0.07}_{-0.08}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
7809	+53.12972−27.80274	3.578	$8.29^{+0.05}_{-0.05}$	$1.85^{+0.31}_{-0.23}$	0.98 ± 0.09	$7.58^{+0.09}_{-0.08}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10035295	+53.11434−27.81549	3.589	$7.39^{+0.07}_{-0.07}$	$2.34^{+0.06}_{-0.05}$	1.94 ± 0.03	$7.56^{+0.12}_{-0.10}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	‡‡
6460	+53.12808−27.80739	3.595	$8.78^{+0.07}_{-0.07}$	$0.48^{+0.11}_{-0.09}$	0.20 ± 0.05	$8.54^{+0.06}_{-0.06}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	0	
10009506	+53.15301−27.80607	3.599	$7.82^{+0.07}_{-0.07}$	$2.43^{+0.14}_{-0.13}$	1.65 ± 0.02	$7.87^{+0.09}_{-0.09}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
4282	+53.12051−27.81523	3.601	$7.88^{+0.08}_{-0.08}$	$1.12^{+0.36}_{-0.28}$	0.23 ± 0.06	$7.21^{+0.19}_{-0.19}$	P	R3	1	
19519	+53.16609−27.77126	3.604	$7.75^{+0.02}_{-0.02}$	$2.50^{+0.14}_{-0.16}$	0.99 ± 0.01	$7.99^{+0.06}_{-0.06}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
18970	+53.15650−27.77227	3.725	$9.20^{+0.00}_{-0.00}$	$7.22^{+0.15}_{-0.14}$	3.95 ± 0.08	$8.17^{+0.06}_{-0.07}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
18028	+53.12869−27.77482	3.731	$8.84^{+0.04}_{-0.04}$	$0.40^{+0.03}_{-0.02}$	0.25 ± 0.05	$7.73^{+0.09}_{-0.09}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10016186	+53.11695−27.80283	3.928	$8.18^{+0.01}_{-0.01}$	$2.70^{+0.15}_{-0.21}$	1.14 ± 0.02	$8.14^{+0.11}_{-0.22}$	M	R3, R2, $\hat{\mathrm{R}}$, O32	0	
10015193	+53.10889−27.79840	3.957	$8.22^{+0.04}_{-0.04}$	$6.03^{+0.31}_{-0.35}$	1.50 ± 0.02	$8.10^{+0.06}_{-0.06}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10013578	+53.11875−27.77811	4.022	$8.00^{+0.05}_{-0.05}$	$2.71^{+0.50}_{-0.23}$	0.91 ± 0.06	$7.87^{+0.08}_{-0.09}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
4270	+53.16871−27.81516	4.023	$8.19^{+0.02}_{-0.02}$	$9.00^{+0.25}_{-0.26}$	9.12 ± 0.11	$8.04^{+0.06}_{-0.06}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10015344	+53.11438−27.77158	4.032	$8.94^{+0.02}_{-0.02}$	$12.45^{+1.31}_{-1.47}$	0.65 ± 0.01	$8.22^{+0.06}_{-0.06}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10013545	+53.12413−27.79914	4.035	$7.95^{+0.06}_{-0.06}$	$1.70^{+0.18}_{-0.17}$	2.63 ± 0.08	$8.08^{+0.06}_{-0.08}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
7507	+53.15548−27.80388	4.044	$7.67^{+0.07}_{-0.07}$	$1.51^{+0.10}_{-0.08}$	0.80 ± 0.04	$7.48^{+0.10}_{-0.09}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	1	
17777	+53.16743−27.77548	4.134	$8.25^{+0.19}_{-0.19}$	$1.00^{+0.45}_{-0.18}$	0.29 ± 0.02	$7.80^{+0.17}_{-0.16}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	1	
7762	+53.11333−27.80299	4.149	$8.24^{+0.02}_{-0.02}$	$4.10^{+0.29}_{-0.28}$	1.87 ± 0.03	$8.14^{+0.06}_{-0.06}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
7892	+53.15003−27.80250	4.229	$7.76^{+0.03}_{-0.03}$	$1.96^{+0.05}_{-0.05}$	1.17 ± 0.02	$7.76^{+0.11}_{-0.10}$	M	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
6519	+53.15832−27.80724	4.260	$7.68^{+0.04}_{-0.04}$	$2.92^{+0.08}_{-0.06}$	1.71 ± 0.03	$7.69^{+0.07}_{-0.07}$	P	R3, R2, $\hat{\mathrm{R}}$, O32, Ne3O2	0	
10001916	+53.13228−27.79811	4.283	$6.38^{+0.06}_{-0.06}$	$0.23^{+0.02}_{-0.02}$	0.16 ± 0.02	$7.23^{+0.21}_{-0.20}$	M	R3	1	
8073	+53.15119−27.80195	4.392	$7.70^{+0.05}_{-0.05}$	$1.68^{+0.24}_{-0.17}$	0.50 ± 0.04	$7.82^{+0.13}_{-0.13}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	1	‡‡
10000626	+53.14700−27.81303	4.464	$6.72^{+0.20}_{-0.20}$	$0.34^{+0.02}_{-0.02}$	0.29 ± 0.02	$7.29^{+0.11}_{-0.12}$	M	R3	1	‡‡
7304	+53.16083−27.80455	4.490	$7.61^{+0.13}_{-0.13}$	$0.30^{+0.05}_{-0.05}$	0.24 ± 0.02	$7.42^{+0.20}_{-0.16}$	P	R3	1	
17072	+53.17022−27.77739	4.702	$8.30^{+0.04}_{-0.04}$	$0.75^{+0.08}_{-0.06}$	0.53 ± 0.06	$8.28^{+0.06}_{-0.07}$	P	R3, R2, $\hat{\mathrm{R}}$, O32	1	‡‡

Table C.1. (cont.)

NIRSpec ID	JADES ID JADES-GS	Redshift	log $M_\star$ [$M_\odot$]	SFR [$M_\odot$ yr^{-1}]	SFR [Hα] [$M_\odot$ yr^{-1}]	$12 + \log(O/H)$	C (★)	Metallicity diagnostics	NC (†)	AGN (‡)
18090	+53.16718-27.77462	4.775	$7.86^{+0.03}_{-0.03}$	$4.91^{+0.10}_{-0.10}$	3.93 ± 0.29	$7.84^{+0.08}_{-0.08}$	M	R3, R2, R̂, O32, Ne3O2	0	
7938	+53.16268-27.80237	4.806	$8.07^{+0.03}_{-0.03}$	$3.22^{+0.07}_{-0.07}$	1.69 ± 0.03	$7.81^{+0.08}_{-0.08}$	M	R3, R2, R̂, O32, Ne3O2	0	
5457	+53.11667-27.81093	4.863	$7.79^{+0.28}_{-0.28}$	$1.40^{+0.26}_{-0.20}$	0.48 ± 0.02	$8.13^{+0.07}_{-0.08}$	P	R3, R2, R̂, O32, Ne3O2	1	
4009	+53.15705-27.81629	4.866	$8.04^{+0.10}_{-0.10}$	$2.11^{+0.41}_{-0.37}$	1.27 ± 0.07	$7.66^{+0.12}_{-0.11}$	P	R3, R2, R̂, O32, Ne3O2	1	
17260	+53.12689-27.77689	4.885	$7.97^{+0.12}_{-0.12}$	$0.44^{+0.05}_{-0.04}$	0.23 ± 0.03	$7.27^{+0.11}_{-0.11}$	P	R3	1	
10005217	+53.17351-27.77187	4.888	$7.33^{+0.01}_{-0.01}$	$2.24^{+0.04}_{-0.05}$	1.84 ± 0.01	$7.38^{+0.08}_{-0.08}$	P	R3, R2, R̂, O32, Ne3O2	0	
8113	+53.12300-27.80176	4.903	$7.80^{+0.05}_{-0.05}$	$2.48^{+0.25}_{-0.16}$	0.87 ± 0.04	$7.88^{+0.08}_{-0.08}$	P	R3, R2, R̂, O32, Ne3O2	0	
5759	+53.14946-27.80979	5.052	$7.40^{+0.10}_{-0.10}$	$1.25^{+0.08}_{-0.09}$	0.83 ± 0.04	$7.58^{+0.24}_{-0.18}$	M	R3, R2, R̂, O32	1	
10015338	+53.11535-27.77289	5.076	$7.94^{+0.04}_{-0.04}$	$3.64^{+0.15}_{-0.19}$	1.89 ± 0.05	$7.57^{+0.08}_{-0.08}$	P	R3, R2, R̂, O32, Ne3O2	0	‡‡
9452	+53.11583-27.79755	5.122	$8.13^{+0.17}_{-0.17}$	$4.17^{+1.17}_{-0.97}$	1.19 ± 0.11	$7.92^{+0.21}_{-0.22}$	P	R3	0	‡‡
4902	+53.11852-27.81297	5.123	$8.69^{+0.03}_{-0.03}$	$1.77^{+0.41}_{-0.24}$	0.50 ± 0.02	$8.22^{+0.09}_{-0.07}$	P	R3, R2, R̂, O32	1	
9743	+53.12300-27.79661	5.440	$7.48^{+0.08}_{-0.08}$	$2.80^{+0.28}_{-0.23}$	3.21 ± 0.32	$7.43^{+0.26}_{-0.20}$	M	R3	0	
9343	+53.12874-27.79787	5.443	$7.25^{+0.37}_{-0.37}$	$0.14^{+0.10}_{-0.06}$	0.39 ± 0.09	$7.73^{+0.15}_{-0.18}$	P	R2, Ne3O2	0	
10016374	+53.11572-27.77496	5.503	$7.63^{+0.05}_{-0.05}$	$2.38^{+0.09}_{-0.07}$	1.17 ± 0.03	$7.75^{+0.07}_{-0.07}$	P	R3, R2, R̂, O32, Ne3O2	0	
6246	+53.12972-27.80818	5.562	$7.78^{+0.11}_{-0.11}$	$3.00^{+0.55}_{-0.53}$	0.59 ± 0.03	$7.52^{+0.16}_{-0.12}$	M	R3, R2, R̂, O32	1	
16745	+53.13002-27.77839	5.567	$8.26^{+0.03}_{-0.03}$	$6.45^{+0.26}_{-0.27}$	2.19 ± 0.04	$8.13^{+0.06}_{-0.06}$	P	R3, R2, R̂, O32, Ne3O2	0	
6384	+53.13059-27.80771	5.615	$8.42^{+0.13}_{-0.13}$	$1.84^{+0.77}_{-0.53}$	0.44 ± 0.07	$7.30^{+0.05}_{-0.07}$	P	R3	1	
4404	+53.11537-27.81477	5.764	$7.87^{+0.04}_{-0.04}$	$3.30^{+0.19}_{-0.13}$	3.72 ± 0.12	$7.73^{+0.09}_{-0.09}$	M	R3, R2, R̂, O32, Ne3O2	0	
3968	+53.14505-27.81643	5.769	$7.90^{+0.16}_{-0.16}$	$1.29^{+0.55}_{-0.24}$	0.26 ± 0.02	$7.67^{+0.28}_{-0.25}$	P	R3, R2, R̂, O32, Ne3O2	1	
22251	+53.15407-27.76607	5.798	$7.89^{+0.03}_{-0.03}$	$4.68^{+0.12}_{-0.13}$	4.75 ± 0.11	$7.88^{+0.08}_{-0.07}$	M	R3, R2, R̂, O32, Ne3O2	0	‡‡
10056849	+53.11351-27.77284	5.814	$7.07^{+0.06}_{-0.06}$	$1.12^{+0.05}_{-0.04}$	1.98 ± 0.17	$7.42^{+0.09}_{-0.09}$	P	R3, R2, R̂, O32, Ne3O2	0	‡‡
10005113	+53.16730-27.80287	5.821	$7.02^{+0.03}_{-0.03}$	$1.04^{+0.05}_{-0.05}$	0.65 ± 0.12	$7.76^{+0.25}_{-0.24}$	M	R3	1	
19606	+53.17655-27.77111	5.889	$7.23^{+0.03}_{-0.03}$	$1.68^{+0.04}_{-0.04}$	1.33 ± 0.04	$7.66^{+0.09}_{-0.08}$	P	R3, R2, R̂, O32, Ne3O2	1	
10013620	+53.12259-27.76057	5.918	$8.30^{+0.05}_{-0.05}$	$2.35^{+0.20}_{-0.17}$	1.18 ± 0.06	$7.93^{+0.12}_{-0.14}$	M	R3, R2, R̂, O32, Ne3O2	0	
9422	+53.12175-27.79763	5.936	$7.71^{+0.00}_{-0.00}$	$5.35^{+0.04}_{-0.03}$	20.38 ± 0.78	$7.53^{+0.07}_{-0.07}$	M	R3, R2, R̂, O32, Ne3O2	0	‡‡
6002	+53.11041-27.80892	5.937	$7.65^{+0.05}_{-0.05}$	$1.32^{+0.09}_{-0.08}$	1.79 ± 0.08	$7.65^{+0.10}_{-0.10}$	M	R3, R2, R̂, O32, Ne3O2	1	
10013618	+53.11911-27.76080	5.944	$8.78^{+0.21}_{-0.21}$	$3.85^{+0.75}_{-2.19}$	0.74 ± 0.04	$8.19^{+0.05}_{-0.06}$	M	R3, R2, R̂, O32, Ne3O2	0	
19342	+53.16062-27.77161	5.974	$7.70^{+0.09}_{-0.09}$	$1.95^{+0.19}_{-0.17}$	1.39 ± 0.06	$7.50^{+0.18}_{-0.12}$	M	R3	1	
17566	+53.15613-27.77584	6.102	$8.50^{+0.13}_{-0.13}$	$4.15^{+0.59}_{-0.44}$	1.48 ± 0.07	$8.11^{+0.09}_{-0.11}$	M	R3, R2, R̂, O32	0	
18976	+53.16660-27.77240	6.327	$7.77^{+0.06}_{-0.06}$	$1.66^{+0.13}_{-0.12}$	1.06 ± 0.06	$7.45^{+0.11}_{-0.11}$	M	R3, R2, R̂, O32	1	
18846	+53.13492-27.77271	6.335	$8.07^{+0.03}_{-0.03}$	$4.32^{+0.09}_{-0.08}$	11.39 ± 0.40	$7.46^{+0.09}_{-0.09}$	P	R3, R2, R̂, O32, Ne3O2	0	
18179	+53.17582-27.77446	6.335	$8.59^{+0.08}_{-0.08}$	$4.93^{+0.68}_{-0.64}$	1.32 ± 0.08	$8.07^{+0.08}_{-0.10}$	P	R3, R2, R̂, O32	1	

Table C.1. (cont.)

NIRSpec ID	JADES ID JADES-GS	Redshift	log M★ [M☉]	SFR [M☉ yr⁻¹]	SFR [Hα] [M☉ yr⁻¹]	12 + log(O/H)	C (★)	Metallicity diagnostics	NC (†)	AGN (‡)
10005447	+53.16288-27.76928	6.623	$6.95^{+0.11}_{-0.11}$	$0.34^{+0.06}_{-0.06}$	0.38±0.04	$7.46^{+0.20}_{-0.16}$	P	R3	1	
16625	+53.16904-27.77884	6.631	$7.35^{+0.03}_{-0.03}$	$2.24^{+0.09}_{-0.09}$	2.23±0.07	$7.33^{+0.11}_{-0.10}$	M	R3	0	
3334	+53.15138-27.81917	6.706	$7.50^{+0.08}_{-0.08}$	$1.19^{+0.12}_{-0.10}$	1.03±0.08	$7.78^{+0.08}_{-0.09}$	P	R3, R2, R̂, O32, Ne3O2	1	
4297	+53.15579-27.81520	6.713	$7.29^{+0.03}_{-0.03}$	$1.98^{+0.07}_{-0.05}$	2.42±0.16	$7.74^{+0.22}_{-0.25}$	M	R3	0	
10013609	+53.11730-27.76408	6.929	$7.71^{+0.06}_{-0.06}$	$3.85^{+0.19}_{-0.18}$	5.39±0.31	$7.94^{+0.12}_{-0.14}$	M	R3, R2, R̂, O32, Ne3O2	0	‡
20961	+53.13423-27.76891	7.045	$7.86^{+0.14}_{-0.14}$	$1.03^{+0.12}_{-0.09}$	1.50±0.33	$7.33^{+0.11}_{-0.11}$	M	R3, R2, R̂, R, O32	1	
10013905	+53.11833-27.76901	7.197	$7.37^{+0.05}_{-0.05}$	$2.20^{+0.14}_{-0.12}$	4.98±0.67	$7.63^{+0.09}_{-0.09}$	P	R3, R2, R̂, O32, Ne3O2	0	‡
10013682	+53.16746-27.77201	7.275	$6.95^{+0.09}_{-0.09}$	$0.82^{+0.08}_{-0.07}$	0.47±0.06	$7.66^{+0.11}_{-0.10}$	P	R3, R2, R̂, O32, Ne3O2	1	
21842	+53.15682-27.76716	7.981	$7.51^{+0.04}_{-0.04}$	$3.21^{+0.25}_{-0.25}$	1.72±0.08	$7.61^{+0.07}_{-0.07}$	P	R3, R2, R̂, O32, Ne3O2	1	‡
8013	+53.16446-27.80218	8.473	$7.54^{+0.08}_{-0.08}$	$3.03^{+0.32}_{-0.26}$	1.61±0.37	$7.64^{+0.29}_{-0.23}$	M	R3	1	
10058975	+53.11243-27.77461	9.433	$8.05^{+0.03}_{-0.03}$	$6.62^{+0.20}_{-0.20}$	13.79±1.36	$7.35^{+0.08}_{-0.08}$	M	R3, R2, R̂, O32, Ne3O2	0	‡

Notes: ★ Spectral configuration (P=prism; M=medium resolution gratings) considered for deriving the metallicity. See Section 3.1 for more details. † NIRCAM photometry: for galaxies with flag=1 the photometry has been included in the M★ and SFR derivation with BEAGLE; objects with flag=0 have M★ and SFR measured only from fitting to the PRISM spectra. ‡ Galaxies identified as candidate (narrow-lines) AGN based on the criteria described in Scholtz et al. (2023). These objects are removed from the fiducial analysis presented in the main body of the paper due to the unknown AGN contribution and the potential biases introduced in the M★, SFR, and metallicity derivation, nonetheless their main properties (estimated self-consistently as for the other sources) are included here (and shown in Section A).