

JADES: the mass-metallicity relation at $z = 1 - 10$. New calibrations, extremely metal-poor galaxies, and chemical diversity

Yuki Isobe,^{1,2,3*} Mirko Curti,⁴ Roberto Maiolino,^{1,2,5} Qiao Duan,^{1,2} William McClymont,^{1,2} Dávid Puskás,^{1,2} Francesco D'Eugenio,^{1,2} Pierluigi Rinaldi,⁶ James A. A. Trussler,⁷ Jan Scholtz,^{1,2} Tobias J. Looser,⁷ Erica Nelson,⁸ Xihan Ji,^{1,2} Danial Langeroodi,^{1,2,9} Sandro Tacchella,^{1,2} Gareth C. Jones,^{1,2} Ignas Juodžbalis,^{1,2} Robert G. Pescalau,^{1,2} Tiger Yu-Yang Hsiao,^{10,11} Hannah Übler,¹² William M. Baker,⁹ Andrew J. Bunker,¹³ Stefano Carniani,¹⁴ Stéphane Charlot,¹⁵ Emma Curtis-Lake,¹⁶ Sophia Geris,^{1,2} Maria Koller,^{1,2} Jianwei Lyu,¹⁷ Brant Robertson,¹⁸ Christina C. Williams,¹⁹ and Zihao Wu,⁷

¹Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK

²Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge, CB3 0HE, UK

³Waseda Research Institute for Science and Engineering, Faculty of Science and Engineering, Waseda University, 3-4-1, Okubo, Shinjuku, Tokyo 169-8555, Japan

⁴INAF, Osservatorio di Astrofisica e Scienza dello Spazio, Via P. Gobetti 93/3, I-40129 Bologna, Italy

⁵Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

⁶Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA

⁷Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge MA 02138 USA

⁸Department for Astrophysical and Planetary Science, University of Colorado, Boulder, CO 80309, USA

⁹DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155A, 2200 Copenhagen, Denmark

¹⁰Department of Astronomy, The University of Texas at Austin, 2515 Speedway, Austin, Texas 78712, USA

¹¹Cosmic Frontier Center, The University of Texas at Austin, Austin, TX 78712, USA

¹²Max-Planck-Institut für extraterrestrische Physik (MPE), Gießenbachstraße 1, 85748 Garching, Germany

¹³Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

¹⁴Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

¹⁵Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France

¹⁶Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK

¹⁷Steward Observatory, University of Arizona, 933 N. Cherry Avenue, Tucson, AZ 85721, USA

¹⁸Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz CA 96054, USA

¹⁹NSF National Optical-Infrared Astronomy Research Laboratory, 950 North Cherry Avenue, Tucson, AZ 85719, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We present gas-phase metallicities of star-forming galaxies at $z = 1-10$ with deep *JWST*/NIRSpec spectra from the JADES full data release, Dark Horse, and OASIS programmes. We stack ~ 1500 medium-resolution spectra, yielding detections of the $[\text{O III}]\lambda 4363$ auroral line down to $12 + \log(\text{O}/\text{H}) = 7.0$ to derive stack-based strong-line calibrations over the metallicity range $12 + \log(\text{O}/\text{H}) = 7.0-8.7$. At a fixed metallicity, our stacks exhibit $[\text{O III}]\lambda 5007/\text{H}\beta$ and $[\text{O III}]\lambda 5007/[\text{O II}]\lambda 3726, 3729$ values generally lower than calibrations based on high- z individual auroral-line emitters, suggesting an observational bias towards higher excitation introduced when requiring auroral line detections in individual spectra. Based on our new calibrations, we obtain canonical mass-metallicity relations (MZR) at $z=1-10$, identifying a decrease in metallicities from $z \sim 0$ to $z=4-10$, without significant change in slope. Moreover, we identify 50 promising candidates of extremely metal-poor galaxies (EMPGs) with $12 + \log(\text{O}/\text{H}) = 6.7-7.3$ (1–4% solar metallicity) at $z = 1.2-9.1$. The MZR of EMPGs are characterised by a large scatter, with those having lower metallicities generally exhibiting lower sSFRs, opposite of what expected from the local Fundamental Metallicity Relation. These results support a stochastic star-formation history involving gas consumption/ejection and metal-poor inflow, strongly affecting metallicities of low-mass galaxies. Furthermore, we identify two Little Red Dots in our EMPG candidates, both exhibiting broad $\text{H}\alpha$ and prominent $\text{Ly}\alpha$, offering insights into the early black-hole growth in extremely metal-poor environments.

Key words: galaxies: high-redshift – ISM: abundances – galaxies: star formation

1 INTRODUCTION

Understanding the early stages of star formation hinges on the study of chemically pristine galaxies. The first generation of stars (Popu-

* E-mail: yi264@cam.ac.uk (YI)

lation III; PopIII stars) forms from primordial gas composed only of hydrogen (H) and helium (He). Subsequent stellar populations contain elements heavier than He (so-called metals), enriched by earlier generations of stars. Metal-poor gas is less efficient at cooling, which can lead to stellar initial mass functions (IMFs) that are more top-heavy than those of local galaxies (e.g., [Hirano et al. 2015](#); [Stacy et al. 2016](#); [Chon et al. 2024](#)). Observations of extremely metal-poor galaxies¹ (EMPGs; e.g., [Kojima et al. 2020](#); [Isobe et al. 2022](#)) may hint at this initial mode of star formation.

The unprecedented sensitivity of Near-Infrared Spectrograph (NIRSpec; [Jakobsen et al. 2022](#)) on the *James Webb Space Telescope* (*JWST*; [Gardner et al. 2023](#)) allows us to investigate gas-phase metallicities, quantified by $12 + \log(\text{O}/\text{H})$, at high redshifts. The most metal-poor galaxies observed at $z \sim 0$ exhibit $12 + \log(\text{O}/\text{H}) = 6.8\text{--}7.0$ (i.e., $\sim 0.01\text{--}0.02 Z_{\odot}$; [Izotov et al. 2018](#); [Kojima et al. 2020](#); [Scholte et al. 2026](#)), while *JWST* has recently identified EMPGs at $z > 3$ whose inferred metallicities are even lower (e.g., [Asada et al. 2026](#); [Cai et al. 2025](#); [Chemerynska et al. 2024](#); [Cullen et al. 2025](#); [Hsiao et al. 2025](#); [Mowla et al. 2024](#); [Nakajima et al. 2025](#); [Vanzella et al. 2023, 2024, 2025](#); [Willott et al. 2025](#); [Maiolino et al. 2025](#)), possibly even $12 + \log(\text{O}/\text{H}) < 6$ ([Morishita et al. 2025a](#)). The NIRSpec integral-field unit reveals satellite components whose inferred metallicities are even lower than their main galaxies at $z = 8.5\text{--}10.2$ ([Pascalau et al. 2026](#); [Koller et al. 2026](#)). Remarkably, [Übler et al. \(2026\)](#) and [Maiolino et al. \(2026\)](#) identify a $z = 10.6$ system with $\text{H}\gamma$ and $\text{He II } \lambda 1640$ emission without any metal lines, suggesting a pristine system whose stellar mass is dominated by PopIII stars ([Rusta et al. 2026](#); [Jeon et al. 2026](#)). Intriguingly, even some Little Red Dots (LRDs) potentially hosting overmassive black holes have narrow line ratios that suggest very low metallicity ([Maiolino et al. 2025](#); [Ivey et al. 2026a](#)), which provides new insights into black hole growth in near-pristine environments.

Accurate metallicities for metal-poor galaxies provide a better understanding of early galaxy evolution and the baryon cycle. Galaxies are thought to evolve as stars produce metals, some of which are expelled from galaxies via outflows, while the accretion of metal-poor gas triggers new star formation, which shapes the scaling relation between stellar mass (M_*) and metallicity (mass-metallicity relation; MZR; e.g., [Lequeux et al. 1979](#); [Tremonti et al. 2004](#); [Maiolino & Mannucci 2019](#); [Baker & Maiolino 2023](#)). Local observations have established a primary correlation in the MZR, along with a secondary anti-correlation between metallicity and star formation rate (SFR; e.g., [Ellison et al. 2008](#); [Mannucci et al. 2010](#)). Taking this SFR dependence into account yields a tighter scaling relation between metallicity, M_* , and SFR (so-called Fundamental Metallicity Relation, FMR; [Mannucci et al. 2010](#); [Andrews & Martini 2013](#); [Curti et al. 2020](#); [Baker et al. 2023](#)), which suggests a balance between star formation, metal production, inflow, and outflow. However, recent studies report that low-mass galaxies ($\log(M_*/M_{\odot}) \lesssim 9$), both at $z \sim 0$ ([Laseter et al. 2025](#)) and at high redshift (e.g., [Nakajima et al. 2023](#); [Langeroodi & Hjorth 2023](#); [Curti et al. 2024](#); [Nishigaki et al. 2025](#)), exhibit lower metallicities than those predicted by the FMR, suggesting a breakdown of this picture. Characterising EMPGs on the MZR and FMR planes will shed light on the physical mechanisms driving the metallicity deficits.

Measuring electron temperature (T_e) and electron density (n_e) is essential for deriving ion abundance ratios with the physically motivated “direct- T_e ” method (e.g., [Peimbert 1967](#); [Osterbrock 1989](#)). However, auroral lines that are sensitive to T_e are typically very faint. Indeed, only two high- z EMPG candidates have detections of the $[\text{O III}] \lambda 4363$ auroral line ([Mowla et al. 2024](#); [Cullen et al. 2025](#)) and thus metallicities based on the direct- T_e method ($12 + \log(\text{O}/\text{H})_{T_e}$, hereafter). Metallicities of the other EMPG candidates are derived from empirical calibrations between the ratios of strong emission lines (such as $\text{R3} \equiv [\text{O III}] \lambda 5007/\text{H}\beta$) and metallicities, generally referred to as the “strong-line” method. Table 1 defines metallicity-sensitive indices used for the strong-line method. The strong-line method has been improved with deep local observations (e.g., [Pagel et al. 1979](#); [Pettini & Pagel 2004](#); [Nagao et al. 2006](#); [Maiolino et al. 2008](#); [Marino et al. 2013](#); [Curti et al. 2020](#); [Nakajima et al. 2022](#)) and with photoionisation models (e.g., [McGaugh 1991](#); [Zaritsky et al. 1994](#); [Kewley & Dopita 2002](#)) over decades, as described in many reviews ([Maiolino & Mannucci 2019](#); [Kewley et al. 2019](#)). With the advent of *JWST*, it became possible to calibrate the strong-line method exploiting auroral line emitters directly at $z \gtrsim 2\text{--}10$ ([Sanders et al. 2024, 2025a](#); [Chakraborty et al. 2025](#); [Cataldi et al. 2025a](#)).

One key challenge is that these samples are mostly limited to relatively metal-rich ($Z \gtrsim 5\% Z_{\odot}$) galaxies. To avoid sample scarcity and metallicity uncertainty, [Sanders et al. \(2025a\)](#) quantify a valid metallicity range within $12 + \log(\text{O}/\text{H}) = 7.3\text{--}8.6$. Beyond the valid range, the metallicity measurement inevitably introduces significant systematics arising from both the uncertainty in the calibration itself and in the extrapolation. One major way of extrapolation is to simply adopt the same function calibrated within the valid range (e.g., [Vanzella et al. 2025](#); [Hsiao et al. 2025](#); [Asada et al. 2026](#)), while no physical model underlies this extrapolation. Another possibility is to use photoionisation models that match the low-metallicity end of the observed calibration (e.g., [Morishita et al. 2025a](#); [Nakajima et al. 2025](#)), while the adopted ionisation parameter is quite high ($\log(U) = -0.5$).

An additional issue is that galaxies with detectable auroral lines may be biased towards systems with bright emission lines. At $z \sim 0$, [Sanders et al. \(2021\)](#) point out the possibility that requiring auroral-line detections biases the sample to high excitation and large equivalent widths (EW) of emission lines. Indeed, [Nakajima et al. \(2022\)](#) separate local low-metallicity galaxies into subsamples with large restframe EWs of $\text{H}\beta$ ($\text{EW}_0(\text{H}\beta) > 200 \text{ \AA}$) and small $\text{EW}_0(\text{H}\beta)$ ($< 100 \text{ \AA}$), and reported that the R3 and O32 values of the large-EW subsample are significantly higher than those of the small-EW subsample. [Langeroodi & Hjorth \(2026\)](#) find that this trend persists even including high- z sources. They propose a practical approach using $\text{EW}_0(\text{H}\beta)$ values to mitigate the systematics in metallicity measurements, while it may become difficult to measure EWs for fainter, more metal-poor sources, in that the stellar continuum is often undetected.

This paper presents an alternative approach: spectral stacking to enable auroral-line detections without requiring them for individual sources. Several studies have put effort on spectral stacking of $z \sim 0$ galaxies (e.g., [Brown et al. 2016](#); [Curti et al. 2017](#); [Bian et al. 2018](#)) to detect auroral lines, which are used for recalibrating the strong-line method. Thanks to the progress of the NIRSpec observations, it is timely to carry out this analysis for high- z galaxies. Using the large NIRSpec dataset of the *JWST* Advanced Deep Extragalactic Survey (JADES; [Bunker et al. 2024](#); [D’Eugenio et al. 2025b](#)), combined with our established spectral stacking technique ([Isobe et al. 2025](#)), we have demonstrated that we robustly detect $[\text{O III}] \lambda 4363$ from the stacks of hundreds of spectra at $z = 4\text{--}7$, whose $12 + \log(\text{O}/\text{H})_{T_e}$ are down to 7.45 ([Isobe et al. 2026](#)). Leveraging the full NIRSpec dataset

¹ Defined to have $12 + \log(\text{O}/\text{H}) \leq 7.3$ (i.e., $4\% Z_{\odot}$) in this paper. Note that the metallicity threshold widely adopted at $z \sim 0$ is $10\% Z_{\odot}$ (e.g., [Kunth & Östlin 2000](#); [Izotov et al. 2012](#); [Sánchez Almeida et al. 2016](#); [Kojima et al. 2020](#)). However, we need to optimise the threshold at high z (see [Hsiao et al. 2025](#); [Trussler et al. 2026](#)), where metallicities below $10\% Z_{\odot}$ are no longer rare (e.g., [Curti et al. 2024](#); [Morishita et al. 2024](#); [Isobe et al. 2026](#)).

Table 1. Definition of strong-line indices. The pair of two wavelengths connected by a comma indicate the sum of the emission-line fluxes at each wavelength. ^a: Laseter et al. (2024), ^b: Chakraborty et al. (2025), ^c: Cataldi et al. (2025a), ^d: Scholte et al. (2025)

Index	Definition
log(R3)	$\log([\text{O III}]\lambda 5007/\text{H}\beta)$
log(R2)	$\log([\text{O II}]\lambda\lambda 3726, 3729/\text{H}\beta)$
log(R23)	$\log([\text{O II}]\lambda\lambda 3726, 3729 + [\text{O III}]\lambda 4959, 5007)/\text{H}\beta)$
log(O32)	$\log([\text{O III}]\lambda 5007/[\text{O II}]\lambda\lambda 3726, 3729)$
$\hat{R}_{\text{Laseter}}^a$	$0.47 \times \log(\text{R2}) + 0.88 \times \log(\text{R3})$
$\hat{R}_{\text{Chakraborty}}^b$	$0.18 \times \log(\text{R2}) + 0.98 \times \log(\text{R3})$
$\hat{R}_{\text{Cataldi}}^c$	$0.46 \times \log(\text{R2}) + 0.88 \times \log(\text{R3})$
log(Ne3)	$\log([\text{Ne III}]\lambda 3869/\text{H}\gamma)$
log(RO2Ne3)	$\log([\text{O II}]\lambda\lambda 3726, 3729 + [\text{Ne III}]\lambda 3869)/\text{H}\delta)$
$\widehat{\text{RNe}}^d$	$0.47 \times \log([\text{O II}]\lambda\lambda 3726, 3729/\text{H}\gamma) + 0.88 \times \log(\text{Ne3})$
log(Ne3O2)	$\log([\text{Ne III}]\lambda 3869/[\text{O II}]\lambda\lambda 3726, 3729)$
log(N2)	$\log([\text{N II}]\lambda 6583/\text{H}\alpha)$
log(O3N2)	$\log(\text{R3}) - \log(\text{N2})$
log(S2)	$\log([\text{S II}]\lambda 6716, 6731/\text{H}\alpha)$
log(O3S2)	$\log(\text{R3}) - \log(\text{S2})$
log(Ar3)	$\log([\text{Ar III}]\lambda 7135/\text{H}\alpha)$

of JADES Data Release 4 (DR4; Curtis-Lake et al. 2025; Scholtz et al. 2025a) and deep JADES Dark Horse (DH) data targeting faint sources (D’Eugenio et al. 2025a), we expect [O III] λ 4363 detections from stacks of even more metal-poor systems. The current high- z strong-line calibrations are not well constrained at $12 + \log(\text{O}/\text{H}) \lesssim 7.3$ as discussed by Sanders et al. (2025a), highlighting the requirement of performing spectral stacking to extend the valid metallicity range.

The paper is organised as follows. We describe our data and sample in Section 2, analysis in Section 3, our metallicity calibrations in Section 4, high- z EMPG candidates in Section 5, SFR- M_* relations in Section 6, results and discussions of the MZR and FMR in Section 7, and conclusions in Section 8. We assume a standard Λ CDM cosmology with parameters of $\Omega_0 = 0.315$ and $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2020). Throughout the paper, we use the solar abundance ratios of Asplund et al. (2021), where the gas-phase solar metallicity $Z_\odot$ is $12 + \log(\text{O}/\text{H}) = 8.69$.

2 DATA AND SAMPLE

We analyse NIRSpec spectroscopic data obtained with the NIRSpec micro-shutter array (MSA; Jakobsen et al. 2022; Ferruit et al. 2022) and related value-added catalogues of the following *JWST* programmes in the GOODS-S and GOODS-N fields: the JADES (PIDs 1180, 1181, 1210, 1286, 1287, and 3215; Bunker et al. 2024; D’Eugenio et al. 2025b; Eisenstein et al. 2023; Curtis-Lake et al. 2025; Scholtz et al. 2025a; Robertson et al. 2026), the JADES DH² (D’Eugenio et al. 2025a), and the Observing All phases of Stochastic Star formation survey (OASIS; PID 5997, PIs: T. Looser & F. D’Eugenio; Looser et al. in prep.).

2.1 JADES NIRSpec data

We use the complete NIRSpec dataset of the JADES DR4 (Curtis-Lake et al. 2025; Scholtz et al. 2025a). Since these DR4 papers

describe the JADES observations and data reduction, we summarise only the main points here. The JADES observed 5190 targets in the GOODS-S and GOODS-N fields with the low-resolution prism ($R \sim 100$; R100, hereafter) and most of these were also observed with the three medium-resolution gratings ($R \sim 1000$; R1000, hereafter): F070LP-G140M, F170LP-G235M, and F290LP-G395M. Long exposure times of $\gtrsim 20$ hours were allocated for ~ 700 sources, and the remaining > 4400 sources were observed with exposure times of ~ 1 – 3 hours per setting (Curtis-Lake et al. 2025), and the integration times for R100 were generally longer than those for R1000 by a factor of 1–4 (D’Eugenio et al. 2025b). The observed data were reduced with the pipeline constructed by the European Space Agency (ESA) NIRSpec Science Operations Team (Ferruit et al. 2022) and the NIRSpec Guaranteed Time Observations (GTO) Team (Alves de Oliveira et al. 2018). The standard pipeline conservatively accounts for correlated noise by incorporating error propagation and employing variance-conserving resampling (Dorner et al. 2016). The DR4 one-dimensional spectra were extracted with both the default full microshutter aperture of 5 pixels (0.5”) and the small box-car aperture of 3 pixels. Although the 3-pixel extraction provides higher S/N for compact sources (Scholtz et al. 2025a), which may be more optimal for high- z EMPGs, we use the 5-pixel extraction aperture for more conservative selection of EMPG candidates. The DR4 provides a flux catalogue³ based on the spectral fitting softwares PPXF (Cappellari 2023) and QubeSpec⁴ for the R100 and R1000 data, respectively. In addition, the flux catalogue lists five quality flags of redshifts, where Flags ‘A’, ‘B’, and ‘C’ indicate redshifts from emission lines and/or the continuum while Flags ‘D’ and ‘E’ mean tentative and no redshifts, respectively. We use 3297 sources with robust redshift flags of ‘A’, ‘B’, and ‘C’, discarding sources with Flags ‘D’ and ‘E’.

2.2 DH NIRSpec data

We summarise the DH observations and dataset here (D’Eugenio et al. 2025a for more details). The DH observations were designed to demonstrate NIRSpec/MSA ‘dense shutter spectroscopy’, i.e. dramatically increasing the multiplex of MSA grating spectroscopy by deliberately allowing a large number of spectral overlaps on the detector. The sample consists of all $z > 3$ candidates in the MSA footprint with faint continuum and falling inside an operable microshutter slit. The DH observed ~ 850 sources in the GOODS-S field with long exposure times of 9–10 hours for the two R1000 gratings of F170LP-G235M and F290LP-G395M. The data reduction procedure follows the standard methods of the JADES data reduction with the 5-pixel (0.5”) aperture. The emission-line fluxes are measured using the emission-line fitting procedure for the DR4 R1000 data (Scholtz et al. 2025a). The final catalogue comprises 519 sources with robust redshifts from multiple emission lines.

2.3 OASIS NIRSpec data

Here we briefly describe the OASIS observations and dataset; full details will be presented in Looser et al. (in prep.). OASIS observed 332 targets in the GOODS-S field using the R100 prism with a total exposure time of 28 hours. The emission-line fluxes were measured using the emission-line fitting procedure for the DR4 R100 data (Scholtz et al. 2025a). The final catalogue comprises 261 sources with robust redshifts.

² The DH programme involved re-designed observations from the allocated 3215 UltraDeep programme, which were re-planned due to the original observations being affected by detector shorts.

³ <https://jades.herts.ac.uk/search/>

⁴ <https://github.com/honzascholtz/Qubespec>

2.4 Photometry catalogue and inferred galaxy properties

Deep imaging of the GOODS-S and GOODS-N fields was obtained with *JWST*/Near-Infrared Camera (NIRCam) as part of the JADES survey. The JADES Data Release 5 (DR5) NIRCam dataset builds on the primary JADES programmes (PIDs 1180, 1181, 1210, 1286, 1287 and 4540), supplemented with additional public NIRCam imaging from various community surveys in the GOODS fields (Johnson et al. 2026)⁵. The DR5 catalogues further incorporate imaging of the *Hubble Space Telescope* (HST) from the Hubble Legacy Fields mosaics, including the Advanced Camera for Surveys (ACS)/F435W, F606W, F775W, F814W, F850LP, and the Wide Field Camera 3 (WFC3)/F105W, F125W, F140W, and F160W (Illingworth et al. 2016; Whitaker et al. 2019). Robertson et al. (2026) conduct source detections and photometric measurements to create the JADES DR5 NIRCam photometry catalogue.

This rich JADES DR5 photometric dataset enables us to perform spectral energy distribution (SED) fitting to infer galaxy properties. Full details of the model setup and prior descriptions are presented in the JADES DR5 stellar population catalogue paper by Duan et al. (2026), so only the main points are described below.

We use the Prospector (Johnson et al. 2019, 2021) spectral energy distribution (SED) modeling code to infer galaxy properties, and following a similar methodology as in Tacchella et al. (2022) and Simmonds et al. (2024). Our model consists of 18 free parameters and covers a wide range of galaxy properties, including stellar mass, star-formation history (SFH), nebular emission, dust attenuation and emission, and contributions of active galactic nuclei (AGN). Note that, as we only use the NIRCam, ACS, and WFC3 photometry and exclude the MIRI data, the contributions from dust emission and AGN are negligible. Our Prospector models adopt the MIST stellar isochrones (Choi et al. 2016; Dotter 2016), the MILES stellar spectral library (Sánchez-Blázquez et al. 2006; Falcón-Barroso et al. 2011), and a Chabrier (2003) IMF.

We fix the redshift to spectroscopic values. We adopt the stellar mass function prior (Wang et al. 2023), where lower prior probabilities are assigned to higher-mass galaxies to avoid spurious high-mass solutions. The stellar metallicity is linked to the stellar mass following the Gallazzi et al. (2005) mass-metallicity relation, adopting a Gaussian prior with a mass-dependent mean and a scatter that increases toward lower stellar masses. Note that stellar metallicities do not significantly impact stellar mass measurements. We use a non-parametric SFH, specifically the continuity SFH model (Leja et al. 2019), with 7 age bins. The first age bin spans the lookback time from 0 to 20–30 Myr, which depends on redshift (Duan et al. 2026 for more details). For this model, we implement the physically motivated star-forming main sequence (SFMS) SFH prior introduced in Duan et al. (2026). This prior assumes that galaxies form and evolve

along the main sequence, with substantial scatter around SFMS to capture sudden starburst and quenching events. The gas-phase metallicity follows a uniform prior between $-2 < \log(Z_{\text{gas}}/Z_{\odot}) < 0.5$.

Dust attenuation follows the two-component model of Charlot & Fall (2000), where the diffuse component optical depth $\tau_{\text{dust},2}$ follows gaussian distributions with $\mathcal{N}(\mu, \sigma) = \mathcal{N}(0.3, 1)$ truncated to $[0, 4]$, and the birth-cloud to diffuse ratio $\tau_{\text{dust},1}/\tau_{\text{dust},2} \sim \mathcal{N}(1, 0.3)$ truncated to $[0, 2]$. The attenuation curve slope of the diffuse component is parameterised by a Calzetti et al. (2000) power-law index n , drawn from a uniform prior over $[-1.2, 0.4]$.

We crossmatch our spectroscopic sources with the JADES DR5 photometric sources with a maximum angular separation of $0.5''$. We also require that the spectroscopic redshifts have reliable quality flags of 'A', 'B', or 'C'. A detailed description of the construction and homogenisation of these spectroscopic samples can be found in Puskás et al. (2025b) and Puskás et al. (2025a). This crossmatch connects the spectroscopic measurements and the global galaxy properties such as M_* and SED-based SFRs (SFR_{SED}). In this paper, we use SFRs averaged over 20 Myr (SFR_{20}) and 100 Myr (SFR_{100}).

2.5 Parent sample

From the sources with robust redshifts, we select $z = 1$ –10 sources that have $\text{H}\beta$ detections with $S/N > 3$. We exclude Type-1 AGN with the significant $\text{H}\alpha$ broadening identified in the R1000 data: 30 reported from JADES (Juodžbalis et al. 2025) and 9 from DH (D'Eugenio et al. 2025a). Note that these studies do not survey $\sim 30\%$ of our spectroscopic sources, indicating that our spectroscopic sample contains some unidentified Type-1 AGN. However, given that the number of the reported Type-1 AGN is only $\lesssim 1\%$ of the surveyed sources, the impact of unidentified Type-1 AGN on the stacked results and on the overall sample distribution must be negligible.

We obtain 4047 spectra as our parent sample, which consists of 1929 spectra of the R100 data and 2118 spectra of the R1000 data. Crossmatching our parent sample to the JADES DR5 photometric sources (Section 2.4), we find that the number of unique JADES DR5 photometric sources in our parent sample is 2547.

3 ANALYSIS

3.1 Dust reddening of individual sources

We measure nebular colour excesses $E(B - V)$ from Balmer decrements. We use $\text{H}\alpha/\text{H}\beta$ ratios for sources whose $\text{H}\alpha$ is covered by NIRSpec, otherwise we use $\text{H}\gamma/\text{H}\beta$ ratios. We refer to the theoretical ratios of Balmer lines under the assumption of Case B recombination at $T_e = 15000$ K and $n_e = 100 \text{ cm}^{-3}$ ($\text{H}\alpha/\text{H}\beta = 2.86$, $\text{H}\gamma/\text{H}\beta = 0.47$; Osterbrock 1989). We assume Cardelli et al. (1989)'s extinction law to correct emission-line fluxes for dust reddening, following other high- z strong-line calibrations from the literature (Sanders et al. 2024, 2025a; Cataldi et al. 2025a). Note that the shape of Cardelli et al. (1989)'s extinction curve agrees with the mean nebular attenuation curve of $z \sim 2$ galaxies (Reddy et al. 2020), while a dedicated study will be required to investigate nebular attenuation curves at higher z (Sanders et al. 2025b, also Markov et al. 2025b,a; Maheson et al. 2025; Shivaiei et al. 2025). No correction is applied to sources lacking Balmer decrement measurements, while it is worth noting that the metallicity indices used for our $12 + \log(\text{O}/\text{H})_{\text{SL}}$ measurements, such as R3 and Ne3O2 (Section 3.2), are based on emission lines with small wavelength separations and thus insensitive to variations in $E(B - V)$. We further verify that the fraction of such

⁵ The JADES Origins Field (PID 3215), MIRI Deep Imaging Survey (MIDIS; PID 1283; Östlin et al. 2025), First Reionization Epoch Spectroscopically Complete Observations (FRESCO; PID 1895; Oesch et al. 2023), the *JWST* Extragalactic Medium-band Survey (JEMS; PID 1963; Williams et al. 2023), the Next Generation Deep Extragalactic Exploratory Public (NGDEEP) Survey (PID 2079 Bagley et al. 2024), PANORAMIC (PID 2514 Williams et al. 2025), Complete NIRCam Grism Redshift Survey (CONGRESS; PID 3577), the Bias-free Extragalactic Analysis for Cosmic Origins with NIRCam (BEACON; PID 3990; Morishita et al. 2025b), the Public Observation Pure Parallel Infrared Emission-Line Survey (POPPIES; PID 5398), Observing All phases of Stochastic Star formation survey (OASIS; PID 5997; Looser et al. in preparation), Slitless Areal Pure-Parallel High-Redshift Emission Survey (SAPPHIRES; PID 6434; Sun et al. 2025), and the Director's Discretionary transient follow-up program (PID 6541).

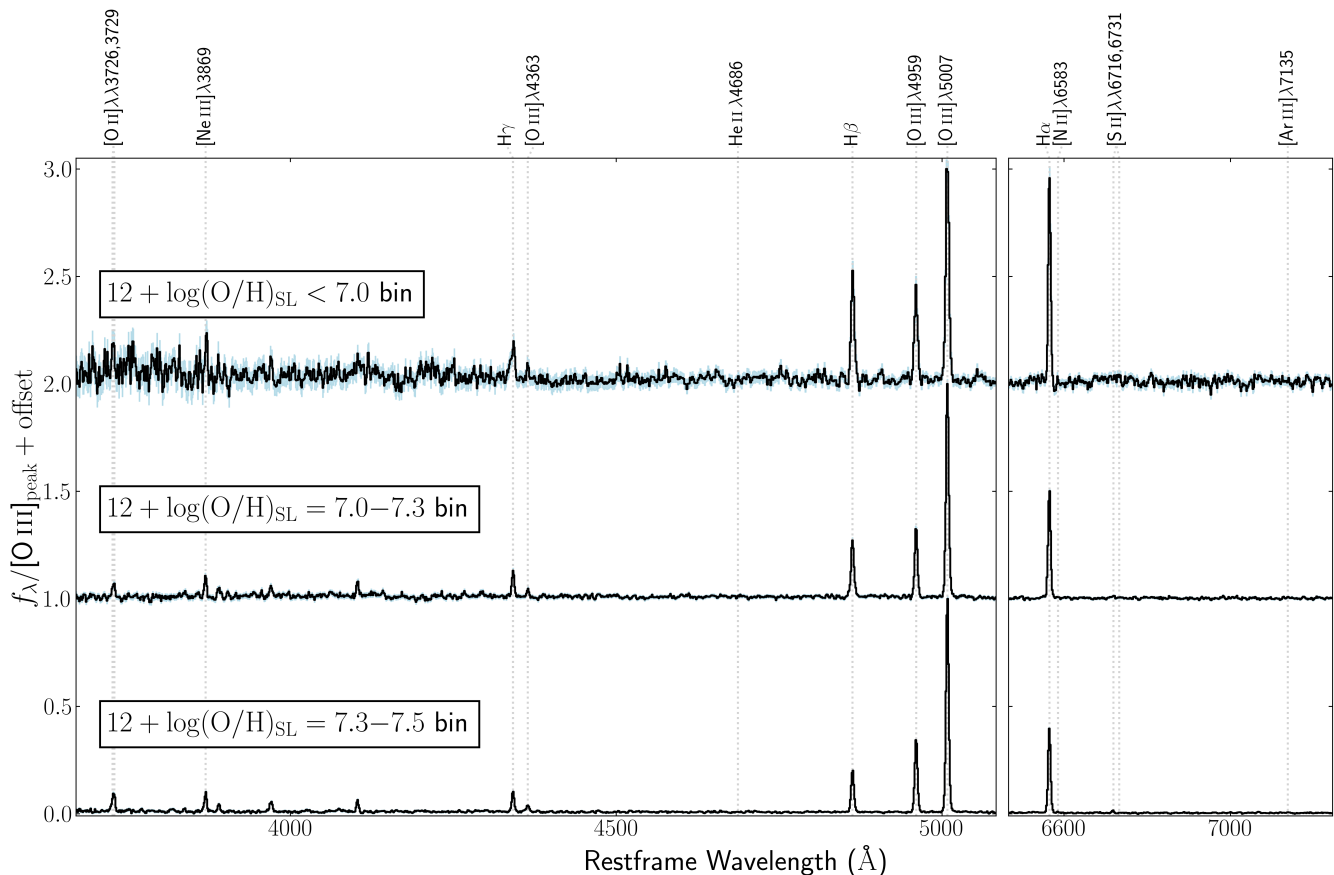


Figure 1. Stacked spectra in bins of strong-line based metallicities of $12 + \log(\text{O}/\text{H})_{\text{SL}} < 7.0$ (31 sources; top), $12 + \log(\text{O}/\text{H})_{\text{SL}} = 7.0\text{--}7.3$ (105 sources; middle), and $12 + \log(\text{O}/\text{H})_{\text{SL}} = 7.3\text{--}7.5$ (135 sources; bottom). The $12 + \log(\text{O}/\text{H})_{\text{SL}}$ values for these stacks are based on Cataldi et al. (2025a)’s calibration, while we recalibrate the strong-line method by measuring T_e -based metallicities of the stacks. All these stacks have the detection of [O III] $\lambda 4363$ with $S/N > 3$.

sources remains below 10% in all stellar-mass bins (0.5 dex width) over the range of $\log(M_*/M_\odot) = 6.5\text{--}10.0$, indicating a negligible impact on the overall sample distribution.

3.2 Strong-line based metallicity measurement

We measure gas-phase metallicities via the strong-line method ($12 + \log(\text{O}/\text{H})_{\text{SL}}$, hereafter) following the procedure developed by Curti et al. (2024). Emission-line fluxes are first corrected for dust attenuation using the $E(B-V)$ values derived from the Balmer decrement (Section 3.1) and the extinction law of Cardelli et al. (1989). The dereddened fluxes are then passed to the metallicity fitting framework.

We combine up to four metallicity-sensitive indices simultaneously when their constituent emission lines are detected at $S/N > 3$: R3, $\tilde{R}_{\text{Cataldi}}$, O32, and Ne3O2 (Table 1). These indices are all based on oxygen or neon lines and are therefore insensitive to deviation and potential anomalies in other chemical abundance patterns, such as strong N/O enhancement (e.g. Cameron et al. 2023; Isobe et al. 2023; Topping et al. 2024). For objects where R3 is the sole available diagnostic, the branch degeneracy is broken via a hierarchical set of constraints. First, we exploit 3σ upper limits on [O II] and [N II] — used as bounds on R2 and N2, respectively — to restrict the allowed metallicity range to either the upper or lower branch of the calibration. In cases where this is insufficient to resolve the degeneracy, we

additionally impose a last-resort prior based on the gas-phase metallicity inferred from SED fitting (Section 2.4): if the posterior still straddles the turnover of the calibration, the SED-based metallicity estimate is used to select the lower or upper branch accordingly.

We initially compute $12 + \log(\text{O}/\text{H})_{\text{SL}}$ for all individual sources in our parent sample using the Cataldi et al. (2025a) calibrations, applying a linear extrapolation below $12 + \log(\text{O}/\text{H}) = 7$ with the slope fixed at the value of the calibration at that boundary. These preliminary metallicities are used solely to assign individual galaxies to metallicity bins for spectral stacking (Section 3.3). After deriving our stack-based calibrations (Section 4), we recompute $12 + \log(\text{O}/\text{H})_{\text{SL}}$ for all sources adopting the new calibrations, and it is these recalculated values that are used throughout the rest of the paper.

For each galaxy, the metallicity is inferred via a Markov Chain Monte Carlo (MCMC) approach using the EMCEE ensemble sampler (Foreman-Mackey et al. 2013), with 48 walkers, a burn-in phase of 600 steps, and 3000 production steps thinned by a factor of 20. The log-likelihood function assumes a Gaussian likelihood for each diagnostic, with a total uncertainty equal to the quadrature sum of the observational uncertainty and the intrinsic scatter of the calibration. A flat (uniform) prior is adopted over the allowed metallicity range $[Z_{\text{lo}}, Z_{\text{hi}}]$, defined by the intersection of the validity domains of all diagnostics entering the fit. The reported metallicity value is the median of the resulting posterior probability distribution function (PDF), with the 1σ confidence interval defined by the 16th and 84th percentiles.

We apply several quality-assessment criteria to the posterior distributions. Measurements with a 1σ confidence interval wider than 1 dex are considered poorly constrained and excluded from further analysis. In addition, we flag and exclude metallicity measurements whose posterior PDF exhibits a bimodal shape — identified via a peak-finding algorithm applied to a Gaussian kernel density estimation of the posterior samples — and whose median value falls near a boundary of the prior. We further flag posteriors that are leaning against the lower or upper boundary of the prior range, defined as having more than 15 per cent of the posterior samples within 0.02 dex of either bound.

3.3 Subsample for spectral stacking

We construct subsamples of our R1000 sources for spectral stacking and the stack-based metallicity calibrations that focus on low metallicities. There are two possible ways to construct low-metallicity samples without requiring auroral lines: selecting sources with low $12 + \log(\text{O}/\text{H})_{\text{SL}}$ (Z-bin sample, hereafter) or low stellar masses (M -bin sample, hereafter). Given that many simulations predict a large scatter of the MZR especially below $M_* \sim 10^8 M_\odot$ (e.g., [McClymont et al. 2025b](#); [Pallottini et al. 2025](#); [Liu et al. 2025](#)), the Z-bin sample and the M -bin sample may be significantly biased to the sources below and above the MZR, respectively. We thus use both the Z-bin sample and the M -bin sample for our analysis.

We construct the Z-bin sample with the $12 + \log(\text{O}/\text{H})_{\text{SL}}$ based on the calibrations of [Cataldi et al. \(2025a\)](#) in bins of $12 + \log(\text{O}/\text{H})_{\text{SL}} < 7.0$, $=7.0\text{--}7.3$, and $=7.3\text{--}7.5$ (Section 3.2). To ensure the robustness of the metallicity measurements, we remove the following sources:

- $\text{H}\alpha/\text{H}\beta < 2.5$, which is significantly below the Case B value. While the assumption of density-bounded nebulae or optically-thick, excited H II clouds can decrease $\text{H}\alpha/\text{H}\beta$ ([Yanagisawa et al. 2024](#); [McClymont et al. 2025c](#), see also [Ivey et al. 2026b](#)), the galaxies discussed in these studies still exhibit $\text{H}\alpha/\text{H}\beta > 2.5$ (see also [Scarlata et al. 2024](#)). Explaining values below this threshold would require more extreme physical conditions.
- AGN candidates probed with emission lines. As we already remove the Type-1 AGN reported by [Juodžbalis et al. \(2025\)](#) and [D'Eugenio et al. \(2025a\)](#) from our parent sample (Section 2.5), here we exclude Type-2 AGN candidates. Following the procedure of [Scholtz et al. \(2025b\)](#), we select Type-2 AGN candidates based on the N2 vs. R3 diagram (a.k.a. the BPT diagram; [Baldwin et al. 1981](#)) and the S2 vs. R3 diagram ([Veilleux & Osterbrock 1987](#)) with the demarcation lines of [Scholtz et al. \(2025b\)](#), the N2 vs. $\text{He II}\lambda 4686/\text{H}\beta$ diagram with a demarcation line of [Shirazi & Brinchmann \(2012\)](#), the $\text{C III}\lambda 1907, 1909/\text{He II}\lambda 1640$ vs. $\text{C IV}\lambda 1549, 1551/\text{C III}\lambda 1907, 1909$ diagram with a demarcation line of [Hirschmann et al. \(2023\)](#). In addition, we select Type-2 AGN candidates with strong $[\text{O III}]\lambda 4363$ by using demarcation lines on the $[\text{O III}]\lambda 4363/\text{H}\gamma$ vs. O32 diagram, the $[\text{O III}]\lambda 4363/\text{H}\gamma$ vs. $\text{Ne III}\lambda 202$ diagram, and the $[\text{O III}]\lambda 4363/\text{H}\gamma$ vs. $[\text{O III}]\lambda 4363/[\text{O III}]\lambda 5007$ proposed by [Mazzolari et al. \(2024\)](#) (also [Jones et al. 2025a](#)). As we focus on restframe-optical emission-line properties, we do not exclude obscured AGN candidates exhibiting significant excess in NIR photometry while not classified as Type-1 or Type-2 AGN ([Rinaldi et al. 2025](#)).
- Evolved galaxies (e.g., with a Balmer break), which are not fully excluded with the R1000 data only but identified with the R100 data because the R100 prism has higher continuum sensitivity than the R1000 gratings. We select the evolved galaxies by vi-

sual inspection because the quantification with the Balmer break index (e.g., [Kriek et al. 2006](#); [Curtis-Lake et al. 2023](#)) does not sufficiently work for our sample galaxies with faint Balmer breaks (also Looser et al. in prep.). This selection misses a population of galaxies with both a Balmer break and emission lines, suggesting a rejuvenating SFH such as A2744-YD4 ([Witten et al. 2025](#)), while we note that the observed R23 and O32 values of A2744-YD4 cannot be simultaneously reproduced by photoionisation models ([Witten et al. 2025](#)), suggesting the need for a focused study of metallicity measurements in such rejuvenating galaxies.

We construct the M -bin sample by selecting sources with the M_* measurements (Section 2.4) in four bins of $\log(M_*/M_\odot) < 7.5$, $=7.5\text{--}8.0$, $=8.0\text{--}8.5$, and $=8.5\text{--}9.0$. It is worth noting that [Geris et al. \(2026\)](#) identify the $\text{H}\alpha$ broadening in the JADES NIRSpec stack with $\log(M_*/M_\odot) = 9.1\text{--}9.8$, suggesting a non-negligible population of hidden broad-line AGN in the sample of galaxies with $M_* \gtrsim 10^9 M_\odot$. Table A1 lists the numbers of the sources and median properties in the bins of the Z-bin and M -bin samples. We use 1488 unique R1000 sources for these stacks.

3.4 Spectral stacking

We produce stacked spectra in the same manner as [Isobe et al. \(2025\)](#), also see [Isobe et al. 2026](#); [Geris et al. 2026](#); [Zucchi et al. 2026](#)). We stack the R1000 data to resolve $\text{H}\gamma$ and $[\text{O III}]\lambda 4363$, which are essential for the direct- T_e method. We perform Gaussian fitting to $[\text{O III}]\lambda 4363$ for each source to obtain redshifts and $[\text{O III}]\lambda 5007$ fluxes. We use [spectres \(Carnall 2017\)](#) to resample the spectra into a common restframe wavelength grid within 1200–7500 Å, whose spectral pixel size is half of the full-width half maximum of the NIRSpec line-spread function (LSF) for point sources ([de Graaff et al. 2024](#)). The spectral pixel size is derived for the median redshift of the stacked sample.

We normalise the individual spectra with respect to their $[\text{O III}]\lambda 5007$ fluxes and calculate the median flux value of the individual spectra at each common spectral pixel. This spectral stacking method reduces the influence of a small number of bright outliers, thereby enabling the resulting spectrum to represent general properties of the stacked sample. Indeed, [Isobe et al. \(2026\)](#) show that this stacking method produces stacked spectra whose $[\text{O III}]\lambda 5007/\text{H}\beta$ value is consistent with the median $[\text{O III}]\lambda 5007/\text{H}\beta$ ratio of the stacked sources, suggesting little bias towards a bright galaxy population. The error spectrum for the stack is estimated by calculating the standard deviation of 1000 stacked spectra at each wavelength, each constructed by perturbing the individual spectra according to their measurement errors. We limit the analysis of each stacked spectra to the wavelength range that is contributed by more than 70% of the input spectra, so that even the most widely separated emission lines in each stacked spectrum originate from averaging over similar parent galaxy samples. We refer to this wavelength range as a valid restframe wavelength (λ_{rest}) range, which is shown in Table A1. We confirm that the valid λ_{rest} range of each stack adequately covers all the necessary emission lines from $[\text{O II}]\lambda 3726, 3729$ to $[\text{Ar III}]\lambda 7135$. Figure 1 shows a representative stacked spectrum for the $12 + \log(\text{O}/\text{H})_{\text{SL}} < 7.0$ bin, exhibiting a clear signal at the location of the $[\text{O III}]\lambda 4363$ emission line.

3.5 Emission line analysis of the stacks

We measure emission line fluxes and derive nebular properties of the stacked spectra in the same manner as [Isobe et al. \(2026\)](#). To

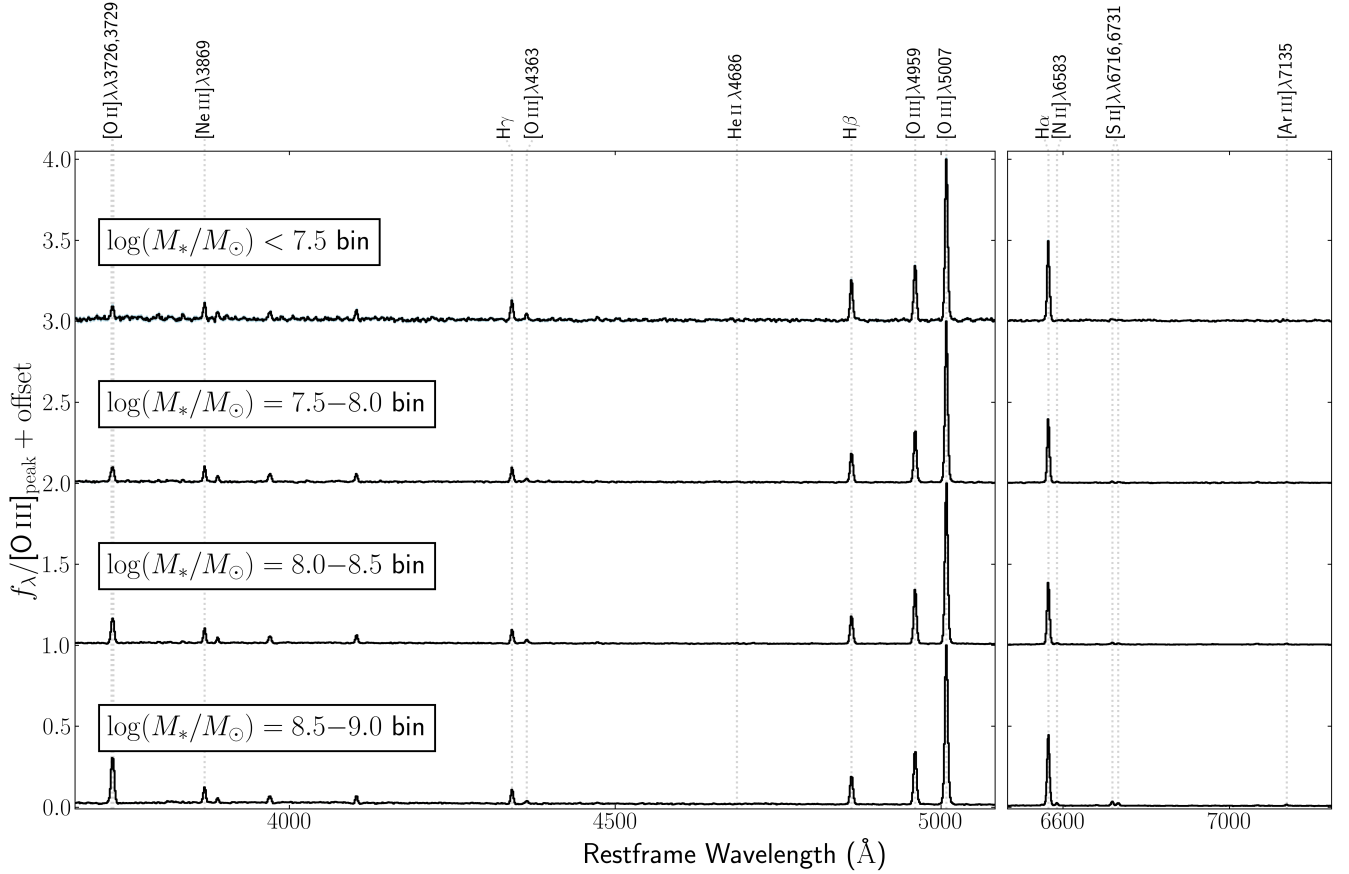


Figure 2. Same as Figure 1 but for the stacked spectra of $\log(M_*/M_\odot) < 7.5$ (133 sources; top), $\log(M_*/M_\odot) = 7.5\text{--}8.0$ (322 sources; second top), $\log(M_*/M_\odot) = 8.0\text{--}8.5$ (517 sources; second bottom), and $\log(M_*/M_\odot) = 8.5\text{--}9.0$ (482 sources; bottom). All these stacks have the detection of [O III]λ4363 with $S/N > 3$.

Table 2. References on transition probabilities and collision strengths.

Ion	Transition probability	Collision Strength
H ⁰	Storey & Hummer (1995)	...
O ⁺	Froese Fischer & Tachiev (2004)	Kisieliuss et al. (2009)
O ²⁺	Froese Fischer & Tachiev (2004)	Lennon & Burke (1994)
S ⁺	Rynkun et al. (2019)	Tayal & Zatsarinny (2010)

summarise, we fit a Gaussian for each emission line with the same redshift and intrinsic velocity width convolved by the velocity width of de Graaff et al. (2024)’s LSF. We model continua with $\lambda < 3640$ and > 3820 Å with independent power-law functions, which are linearly connected at $3640 \leq \lambda \leq 3820$ Å. We find the best-fit model and the flux errors based on the χ^2 minimisation. We set a line detection threshold at $S/N > 3$, and put upper limits at 3σ for undetected lines. We treat [O II] of the $12 + \log(\text{O}/\text{H})_{\text{SL}} < 7.0$ stack as a non detection since the spectrum becomes noisy around [O II]. Table A2 summarises the measured emission line fluxes, highlighting that all stacks have a detection of [O III]λ4363.

We use PyNeb (Luridiana et al. 2015) to obtain the nebular colour excess $E(B - V)$ from the Balmer decrements, electron temperature of [O III] ($T_e[\text{O III}]$) from the [O III]λ4363/[O III]λ5007 ratio, and electron density of [S II] ($n_e[\text{S II}]$) from the [S II]λ6716/[S II]λ6731 ratio. Table 2 summarises the atomic data that we use in this paper.

We assume Case B recombination and the dust extinction law of Cardelli et al. (1989) for consistency with the analysis of our individual sources (Section 3.1). We find the best $E(B - V)$ value that has the least χ^2 value based on Hβ/Hα, Hγ/Hα, and Hγ/Hβ ratios. As the values of $E(B - V)$, $T_e[\text{O III}]$, and $n_e[\text{S II}]$ slightly depend on each other, we adopt an iterative procedure until the T_e value converges within <10 K, which takes at most 4 steps. For the stacks without the detection of [S II], we assume $n_e[\text{S II}] = 300 \text{ cm}^{-3}$, which is typical for the singly-ionised region of high- z galaxies (e.g., Isobe et al. 2023; Topping et al. 2025; Harikane et al. 2025). We perform dust correction using the obtained $E(B - V)$ value and Cardelli et al. (1989)’s extinction curve.

We derive ion abundance ratios of O⁺/H⁺ from [O II]/Hβ and O²⁺/H⁺ from [O III]λ5007/Hβ. We adopt $T_e[\text{O III}]$ for O²⁺ and $T_e[\text{O II}]$ for O⁺. We assume an empirical $T_e[\text{O II}]\text{--}T_e[\text{O III}]$ relation recently calibrated at high redshifts (Cataldi et al. 2025a). We regard $12 + \log(\text{O}/\text{H})$ as $12 + \log((\text{O}^+ + \text{O}^{2+})/\text{H}^+)$ because O³⁺ and higher order oxygen ion abundances are negligible in most star-forming regions (e.g., Izotov et al. 2006; Flury & Moran 2020; Berg et al. 2021) or even AGN (Dors et al. 2020). We assume $n_e[\text{S II}]$ for all ion abundances (i.e., uniform gas density structure), following the conventional direct- T_e method (e.g., Osterbrock 1989; Izotov et al. 2006; Berg et al. 2019) for consistency with other metallicity calibrations (e.g., Nakajima et al. 2022; Cataldi et al. 2025a; Sanders et al. 2025a). We note that, using multiple emission lines and assuming multi-density zones, several studies point out that the contribution of

the dense gas component to [O III] λ 4363 fluxes can increase the total metallicity (e.g., [Marconi et al. 2024](#); [Harikane et al. 2025](#); [Martinez et al. 2025](#); [Moreschini et al. 2026](#)). A dedicated study is warranted to evaluate the impact of the expected metallicity differences in faint galaxies, where metallicity estimates rely on a more limited set of emission lines.

To estimate the measurement errors on $E(B-V)$, $T_e[\text{O III}]$, $n_e[\text{S II}]$, and $12 + \log(\text{O/H})$, we calculate the 16th and 84th percentiles of $E(B-V)$, $T_e[\text{O III}]$, $n_e[\text{S II}]$, and $12 + \log(\text{O/H})$ values repeatedly calculated for 1000 sets of line fluxes by perturbing them according to their errors under the assumption of a normal distribution. Even in this case, we explore the $n_e[\text{S II}]$ values within the range of 4–50000 cm^{-3} . For the stacks without [O II] detection, we derive measurement values of O^{2+}/H^+ and 3σ upper limits on O^+/H^+ . We find that the sum of the O^{2+}/H^+ value and the O^+/H^+ upper limit is only < 0.1 dex larger than the O^{2+}/H^+ value, indicating that the O^+ abundance is negligible compared to the O^{2+} abundance. We thus consider O^{2+}/H^+ as O/H and include the increase when the O^+/H^+ upper limit is added in the upper error on O/H.

Table A2 lists the derived values of $E(B-V)$, $T_e[\text{O III}]$, $n_e[\text{S II}]$, and $12 + \log(\text{O/H})$ for each stack. We highlight that the lowest $12 + \log(\text{O/H})$ value of our stacks is down to $12 + \log(\text{O/H}) = 7.0$, which corresponds to only 2% solar metallicity. Within the literature on metallicity high- z calibrations based on individual auroral-line emitters ([Sanders et al. 2024, 2025a](#); [Cataldi et al. 2025a](#); [Chakraborty et al. 2025](#)), only one data point below this metallicity looks to be included in [Sanders et al. \(2024\)](#). Even when the metallicity range is extended to 7.2, the number of such data points appears to be no more than four as in [Sanders et al. \(2025a\)](#).

4 STACK-BASED METALLICITY CALIBRATIONS

4.1 Stack-based metallicity indices

We derive all the strong-line indices listed in Table 1 from the dust-corrected line ratios. For indices that include undetected emission lines, we adopt the following treatment for consistency with the O/H calculation (Section 3.5). For the indices that include the sum calculation, the measured value is computed by setting the flux of undetected lines to zero. The increase when the upper limits of the undetected lines are added is included in the upper error. For the indices that include the product calculation (including sums in log scale), the value calculated from the upper limits of the undetected lines and the measured values of the other lines is used as the appropriate upper or lower limit. The derived indices are shown in Table A3.

To supplement metal-rich measurements, we use the emission line ratios of the stacks of galaxies at $z = 0.027\text{--}0.25$ ([Andrews & Martini 2013](#)) with sSFR values higher than those of the star-formation main sequence (SFMS) at $z > 1$ ([McClymont et al. 2025b](#)) based on the THESAN-ZOOM simulations ([Kannan et al. 2025](#)). The median sSFR value of these stacks is 1.8 Gyr^{-1} , which is comparable to that of the observed SFMS at $z \sim 3\text{--}4$ (e.g., [Clarke et al. 2024](#); [Simmonds et al. 2025](#)). Hereafter, we refer to these stacks as $z \sim 0$ high-sSFR stack. We derive the strong-line indices of the $z \sim 0$ high-sSFR stacks in the same manner as our stack measurements. We calculate median values and 16th–84th percentile ranges of the $z \sim 0$ high-sSFR stack measurements in O/H bins with 0.2-dex intervals.

We combine our stack measurements and the binned measurements of the $z \sim 0$ high-sSFR stacks to obtain stack-based metallicity calibrations in a metallicity range of $12 + \log(\text{O/H}) = 7.0\text{--}8.7$, where our stacks and the $z \sim 0$ high-sSFR stacks have $12 + \log(\text{O/H}) < 8$ and > 8 , respectively.

4.2 Extension towards lower metallicities

To extend our calibrations below $12 + \log(\text{O/H}) = 7$, we construct photoionisation models using CLOUDY ([Ferland et al. 2013](#)). We model young, metal-poor starbursts with the Binary Population and Spectral Synthesis (BPASS; [Stanway & Eldridge 2018](#)), assuming an instantaneous burst of star formation with a stellar age of 1 Myr, a [Salpeter \(1955\)](#) IMF with an upper mass cut of $100 M_\odot$, a hydrogen density of 300 cm^{-3} , and stellar metallicities fixed to the gas-phase metallicity. We assume metal-to-oxygen abundance ratios to be the solar abundance ratios ([Asplund et al. 2021](#)). We compute the photoionisation models on a grid with 0.25 dex intervals in the range of $5.69 \leq 12 + \log(\text{O/H}) \leq 7.69$ and with 0.5 dex intervals in $-3.5 \leq \log(U) \leq -0.5$.

Figure 3 shows R3 and O32 indices of our photoionisation models as a function of metallicity. The R3 values of our models agree well with the photoionisation models from the literature ([Morishita et al. 2025a](#); [Nakajima & Maiolino 2022](#) used in [Nakajima et al. 2025](#)). In particular, the O32 values are almost independent of $12 + \log(\text{O/H})$ at a fixed $\log(U)$, in line with the observations of low-metallicity galaxies ([Nakajima et al. 2022](#); [Langeroodi & Hjorth 2026](#); [Rosales-Ortega et al. 2026](#)). Additionally, [Sanders et al. \(2025a\)](#) report that the median O32 indices of high- z galaxies in bins of metallicity are gradually flattened towards lower metallicities, suggesting that the median $\log(U)$ value does not drastically evolve in the low-metallicity regime. We thus decide to extrapolate our calibrations with a fixed $\log(U)$ value. Given that the R3 and O32 indices scale with $12 + \log(\text{O/H})$ almost linearly below $12 + \log(\text{O/H}) \sim 7$ with a similar slope for different $\log(U)$ models, we extrapolate our calibrations by fixing the slopes of all the metallicity indices to those of the photoionisation models. As $\log(U) \sim (-2.0)\text{--}(-1.5)$ reproduces most of the indices based on oxygen or neon lines at the low-metallicity end, we obtain the slopes from the models of $12 + \log(\text{O/H}) = 5.69$ and 6.69 with $\log(U) = -2$.

4.3 Polynomial function fitting

We fit the following polynomial function to the relation between the derived strong-line index and $12 + \log(\text{O/H})$:

$$\text{Index} = \begin{cases} \sum_{n=0} c_n x^n & \text{when } x \geq 0 \\ c_0 + c_1 x & \text{when } x < 0 \end{cases} \quad (1)$$

where $x = 12 + \log(\text{O/H}) - Z_{\text{switch}}$ and $Z_{\text{switch}} = 7$. We fix the c_1 value to the slope of the photoionisation models.

We define the best-fit parameters as those that maximise the following log-likelihood function ($\ln \mathcal{L}$):

$$\ln \mathcal{L} = \sum_i [(\ln \mathcal{L})_{i,\text{det}} + (\ln \mathcal{L})_{i,\text{uplim}} + (\ln \mathcal{L})_{i,\text{lowlim}}], \quad (2)$$

where $(\ln \mathcal{L})_{i,\text{det}}$ is log-likelihood function of the i th stack for the measurement value of the index, $(\ln \mathcal{L})_{i,\text{uplim}}$ for that with the upper limit, and $(\ln \mathcal{L})_{i,\text{lowlim}}$ for that with the lower limit. Assuming an asymmetric Gaussian distribution based on the upper and lower errors of the indices, we define $(\ln \mathcal{L})_{\text{det}}$ as follows:

$$(\ln \mathcal{L})_{i,\text{det}} = -\frac{1}{2} \left(\frac{\text{Index}_{i,\text{obs}} - \text{Index}_{i,\text{cal}}}{\sigma_{i,\text{obs}}} \right)^2 - \ln(\sqrt{2\pi}\sigma_{i,\text{obs}}), \quad (3)$$

where $\sigma_{i,\text{obs}}$ is the upper error or the lower error when the index predicted by the calibration ($R_{i,\text{cal}}$) is larger or smaller than the observed index ($R_{i,\text{obs}}$), respectively.

To impose a large penalty when the observed upper limit on the index ($\text{Index}_{i,\text{uplim}}$) is smaller than $\text{Index}_{i,\text{cal}}$ or when the observed

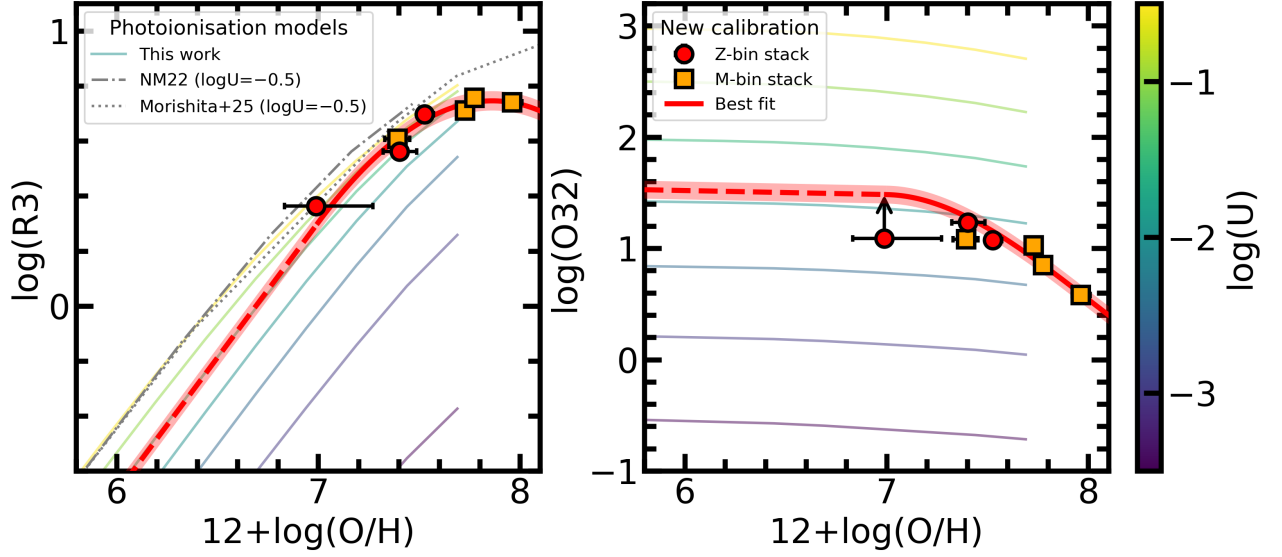


Figure 3. Relations between metallicities and R3 (left) and O32 (right). We grid our photoionisation models with 0.5 dex intervals in $\log(U) = (-3.5) - (-0.5)$, which are shown by the solid lines colour-coded by $\log(U)$. Our R3 model with $\log(U) = -0.5$ agrees well with the models from the literature with the same $\log(U)$ value (Nakajima & Maiolino 2022; Morishita et al. 2025a). These models are almost parallel especially at $12 + \log(\text{O}/\text{H}) \lesssim 7$. We fix the slope of our calibrations below $12 + \log(\text{O}/\text{H}) = 7$ (red dashed line) to that of our photoionisation models.

Table 3. Coefficients of our calibrations (Equation 1) and the root-mean square (RMS) of the residual of each index.

Index	c_0	c_1	c_2	c_3	c_4	RMS
$\log(\text{R3})$	0.3034	0.9842	-0.4987	-0.0550	...	0.04
$\log(\text{R2})$	-1.1562	1.0204	0.7884	-0.4693	...	0.08
$\log(\text{R23})$	0.4405	0.9852	-0.4853	...	...	0.04
$\log(\text{O32})$	1.4831	-0.0363	-1.3509	0.4446	...	0.08
$\hat{R}_{\text{Laseter}}$	-0.2600	1.3456	-0.1156	-0.2471	0.0021	0.05
$\hat{R}_{\text{Chakraborty}}$	0.0978	1.1481	-0.3815	-0.1155	...	0.04
$\hat{R}_{\text{Cataldi}}$	-0.2670	1.3354	-0.0146	-0.3730	0.0435	0.05
$\log(\text{Ne3})$	-0.4213	0.9792	-0.4516	-0.0912	...	0.11
$\log(\text{RO2Ne3})$	-0.0095	0.9829	0.0128	-0.1018	...	0.09
$\overline{R}_{\text{Ne}}$	-0.7158	1.3412	-0.0991	-0.2961	...	0.06
$\log(\text{Ne3O2})$	0.2777	-0.0410	-0.9684	0.2569	...	0.07
$\log(\text{Ar3})$	-2.4807	1.0438	-0.1070	-0.1361	...	0.03
$\log(\text{N2})$	-2.4104	1.0109	-0.0243	...	...	0.04
$\log(\text{O3N2})$	2.7564	-0.0267	-0.5867	...	...	0.04
$\log(\text{S2})$	-1.9475	0.9996	-0.1049	...	...	0.12
$\log(\text{O3S2})$	2.3512	-0.0154	-0.5747	...	...	0.12

lower limit on the index ($\text{Index}_{i,\text{lowlim}}$) is larger than $\text{Index}_{i,\text{cal}}$, we define $(\ln \mathcal{L})_{i,\text{uplim}}$ and $(\ln \mathcal{L})_{i,\text{lowlim}}$ as follows:

$$(\ln \mathcal{L})_{i,\text{uplim}} = \begin{cases} -10^5 & \text{when } \text{Index}_{i,\text{uplim}} < \text{Index}_{i,\text{cal}} \\ 0 & \text{when } \text{Index}_{i,\text{uplim}} \geq \text{Index}_{i,\text{cal}} \end{cases} \quad (4)$$

$$(\ln \mathcal{L})_{i,\text{lowlim}} = \begin{cases} -10^5 & \text{when } \text{Index}_{i,\text{lowlim}} > \text{Index}_{i,\text{cal}} \\ 0 & \text{when } \text{Index}_{i,\text{lowlim}} \leq \text{Index}_{i,\text{cal}} \end{cases} \quad (5)$$

To determine which order of polynomials we adopt, we fit the second- to fifth-order polynomials and calculate the Akaike Information Criterion (AIC) for each polynomial, which quantifies the trade-off between goodness of fit and model complexity. The AIC is defined as $\text{AIC} = 2k - 2 \ln \hat{\mathcal{L}}$, where k is the number

of free parameters and $\ln \hat{\mathcal{L}}$ is the maximum log-likelihood function. Obtaining the minimum AIC values among the orders of the polynomials we fit ($\text{AIC}_{\min}$), we exclude the fitting results with $\Delta\text{AIC} \equiv \text{AIC} - \text{AIC}_{\min} > 10$, which are often regarded as a lack of statistical support of a good fit (e.g., Burnham & Anderson 2004). From the remaining results, we choose the order with the best-fit function without unphysical inflections at the edge of the fitted region. Table 3 lists the best-fit parameters of our calibrations together with the root-mean square (RMS) of the residual of each index (Curti et al. 2020).

Figure 3 highlights the low-metallicity end of our R3 and O32 calibrations. The extrapolated regimes below $12 + \log(\text{O}/\text{H}) < 7$ (red dashed line) are parallel to the photoionisation models by design (Section 4.2). Our R3 calibration provides a simple scaling at

this regime, whereby R3 is approximately equal to the metallicity expressed as a percentage of the solar value. For instance, $R3 = 1$ corresponds to $\approx 1\% Z_{\odot}$, whereas $R3 = 0.1$ corresponds to $\approx 0.1\% Z_{\odot}$.

4.4 Comparison with the calibrations from the literature

Figure 4 shows the best-fit functions of the calibrations and the measurements of the stacks used for the calibrations. We find that our calibrations exhibit lower values of R3, R23, O32, $\hat{R}_{\text{Laseter}}$, $\hat{R}_{\text{Chakraborty}}$, $\hat{R}_{\text{Cataldi}}$, and Ne3O2 compared to other high- z calibrations based on individual auroral-line emitters (Sanders et al. 2024, 2025a; Cataldi et al. 2025a; Chakraborty et al. 2025) at a given metallicity. In particular, the lower values of O32 and Ne3O2 suggest that the sources used in our stacks tend to have lower ionisation parameters than the individual auroral-line emitters at high z .

Figure 5 focuses on the R3 and O32 indices together with individual auroral-line emitters with T_e measurements at $z > 1$ drawn from Langeroodi & Hjorth (2026), which include observational results from both ground-based telescopes (Sanders et al. 2020; Revalski et al. 2024) and the JWST (Nakajima et al. 2023; Sanders et al. 2024; Morishita et al. 2024). In this figure, the data points of our stacks and the individual auroral-line emitters are colour-coded by their H β EWs ($\text{EW}_0(\text{H}\beta)$). We find that the high- z calibration based on individual sources (represented by Cataldi et al. 2025a) overlaps with the individual auroral-line emitters whose $\text{EW}_0(\text{H}\beta)$ values are generally larger than those of our stacks, especially within a metallicity range of $12 + \log(\text{O}/\text{H}) \sim 7.5\text{--}8.0$. This is in line with our finding that our stacks exhibit lower O32 values than the current calibrations, given that O32 values and ionisation parameters correlate with sSFRs (e.g., Nakajima & Ouchi 2014; Sanders et al. 2016; Kashino & Inoue 2019) and thus $\text{EW}_0(\text{H}\beta)$.

These differences may arise from the faint auroral-line detection required for the T_e measurement in individual high- z sources, which can bias the sample towards higher ionisation parameter and larger EWs (see Section 1). On the other hand, we construct the stacked samples based on H β emitters (see Sections 2.5 and 3.3), for which we do not require faint line detections. In this sense, we can say that our stack-based calibrations are better applicable to the bulk of the galaxy population of high- z emission-line galaxies.

Compared to the $z \sim 0$ calibrations (Maiolino et al. 2008; Curti et al. 2020; Nakajima et al. 2022), our calibrations generally have higher R3 and R23 values, especially at the low-metallicity branch of $12 + \log(\text{O}/\text{H}) \lesssim 7.6$. This suggests that the general population of high- z metal-poor star-forming galaxies still exhibit slightly higher excitation, although not as high as the individual auroral line emitters at high z . Our $\hat{R}_{\text{Laseter}}$ calibration matches Laseter et al. (2024)'s calibration at the low-metallicity branch, in contrast to Scholte et al. (2025)'s calibration, which agrees with the high- z calibration of Sanders et al. (2025a). Notably, the R3 and R23 indices of our calibrations align with those of the $z \sim 0$ calibrations with large EWs (Nakajima et al. 2022) at $12 + \log(\text{O}/\text{H}) \lesssim 7.6$. In addition, the $z \sim 0$ calibrations with large EWs provide R2, O32, and Ne3O2 comparable to those of our stacks within the same metallicity range, indicating a similar ionisation structure. However, our calibrations exhibit higher N2 and lower O3N2 than the $z \sim 0$ calibrations with large EWs, which can be explained by mild enhancement of N/O in singly-ionised regions at high z (Cataldi et al. 2025b; Cameron et al. 2026).

4.5 Metallicity recalculation and H α -based SFR

We recalculate $12 + \log(\text{O}/\text{H})_{\text{SL}}$ values of the spectroscopic sources in our parent sample with our stack-based calibrations in the same manner as described in Section 3.2. Compared to Cataldi et al. (2025a)'s calibrations, our calibrations provide higher metallicities at $12 + \log(\text{O}/\text{H}) \lesssim 7.8$ and lower metallicities at $12 + \log(\text{O}/\text{H}) \gtrsim 7.8$. This is because R3 and $\hat{R}_{\text{Cataldi}}$ are more sensitive to $12 + \log(\text{O}/\text{H})$ than O32 and Ne3O2 at $12 + \log(\text{O}/\text{H}) \lesssim 7.8$, where our calibrations provide higher $12 + \log(\text{O}/\text{H})$ at a fixed R3 and $\hat{R}_{\text{Cataldi}}$. Conversely, O32 and Ne3O2 are more sensitive to $12 + \log(\text{O}/\text{H})$ than R3 and $\hat{R}_{\text{Cataldi}}$ at $12 + \log(\text{O}/\text{H}) \gtrsim 7.8$, where our calibrations provide lower $12 + \log(\text{O}/\text{H})$ at a fixed O32 and Ne3O2.

The 16th–84th percentile of $12 + \log(\text{O}/\text{H})_{\text{SL}}$ for our R1000 sources spans from 7.5 to 8.3, indicating that most of $z > 1$ sources have metallicities much lower than the solar value. Such a low metallicity can impact the determination of H α -based SFRs ($\text{SFR}_{\text{H}\alpha}$, hereafter). We derive $\text{SFR}_{\text{H}\alpha}$ from the dust-corrected H α luminosities ($L_{\text{H}\alpha}$) using the following equation:

$$\log(\text{SFR}_{\text{H}\alpha}/M_{\odot} \text{ yr}^{-1}) = \log(L_{\text{H}\alpha}/\text{erg s}^{-1}) + C, \quad (6)$$

where C is a conversion factor that depends on the number of H-ionising photons from newborn stars, which is predicted to increase with lower stellar metallicities and more top-heavy IMFs at a given SFR (e.g., Hao et al. 2011; Madau & Dickinson 2014; Reddy et al. 2018; Shapley et al. 2023; Korhonen Cuestas et al. 2025).

Throughout the paper, we adopt $C = -41.64$ for the BPASS model with $Z = 0.002$ (14% $Z_{\odot}$) reported by Eldridge et al. (2017). It should be noted that the C values can change only by 0.1 dex within a wide metallicity range of $\sim 2\%\text{--}60\% Z_{\odot}$, which corresponds to $12 + \log(\text{O}/\text{H}) \sim 7.0\text{--}8.5$. For the sources without H α covered by NIRSspec, we use the brightest available Balmer line and assume Case B recombination.

5 HIGH-REDSHIFT EMPG CANDIDATES

5.1 EMPG candidate selection

From the ~ 2500 sources with the R100 and/or R1000 data, we select high- z EMPG candidates with $12 + \log(\text{O}/\text{H})_{\text{SL}} \leq 7.3$ based on our stack-based calibrations. To preselect candidates to secure robustness of metallicity measurements, we basically adopt the same criteria as we use to construct the Z -bin stacks in Section 3.3. For sources with only R100 data, we remove Type-2 AGN candidates only based on the S2 vs. R3 diagram because the R100 data do not resolve the emission lines necessary for the other Type-2 AGN diagnostics. In addition, we remove the R100 sources at $z < 4$ to ensure that the [O III] $\lambda\lambda 4959, 5007$ doublet and H β are sufficiently resolved. We perform a visual inspection for the preselected candidates, and finally select the 38 most promising candidates from the R100 data and 24 from the R1000 data. We identify 11 candidates selected from both the R100 and R1000 data and one candidate selected from two different tiers, which means that 50 candidates are the unique sources.

Table C1 lists basic properties of the 50 EMPG candidates. Figure 6 shows the spectrum of the most metal-poor candidate with $12 + \log(\text{O}/\text{H})_{\text{SL}} = 6.7$, exhibiting strong H β comparable to [O III] $\lambda 5007$. It is worth mentioning that Trussler et al. (2026) select EMPG candidates using the JADES deep medium-band photometry, and only one of these photometric EMPG candidates, goods-s-mediumjwst_182875, has the JADES NIRSspec spectra. We confirm that goods-s-mediumjwst_182875 has $12 + \log(\text{O}/\text{H}) = 7.19$ and satisfies our EMPG selection criteria.

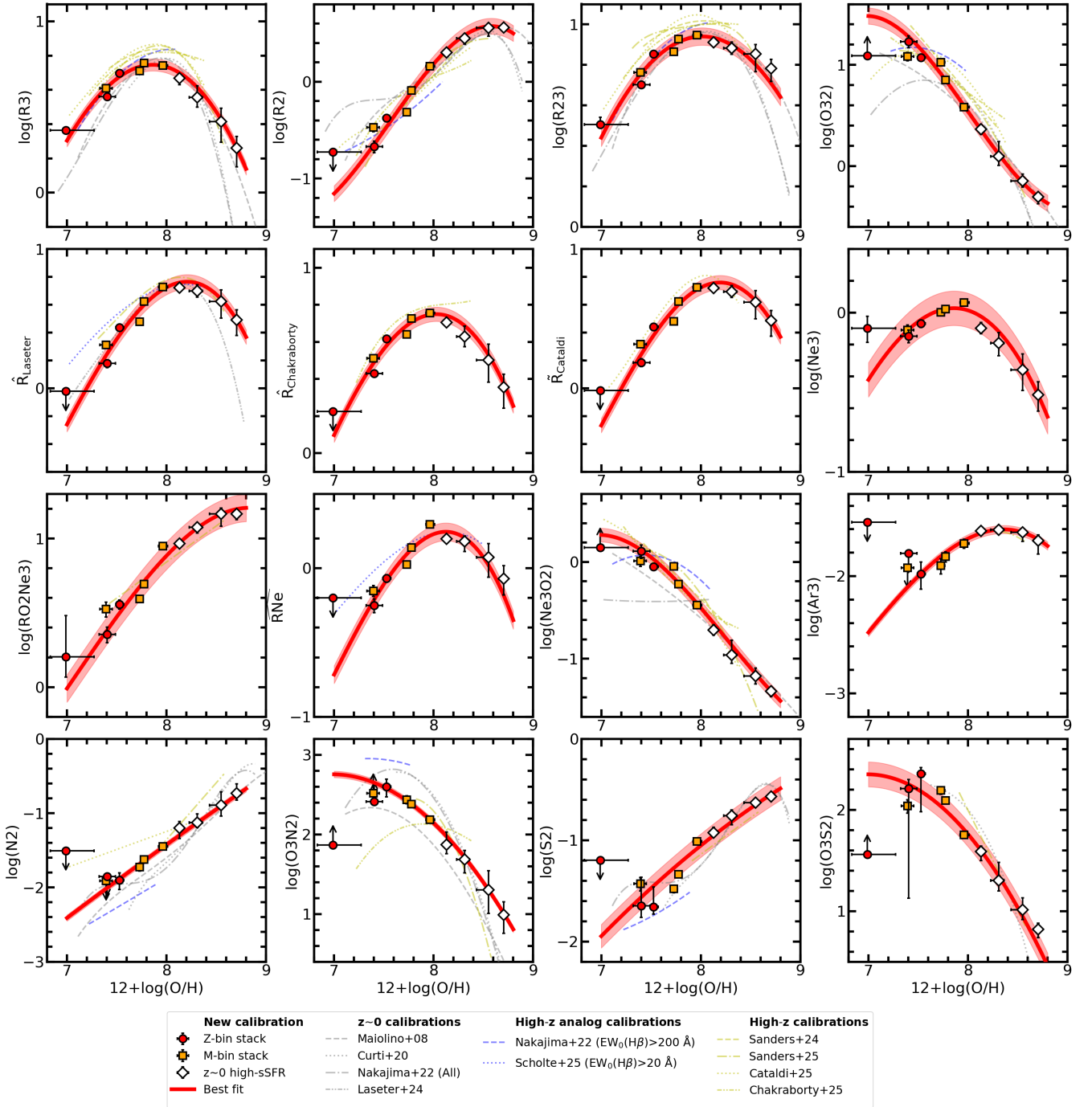


Figure 4. Relations between metallicities and strong-line diagnostics of our Z-bin stacks (red circles) and *M*-bin stacks (orange squares). We extend our calibration to higher metallicities by adding median values of the $z \sim 0$ high-sSFR stacks (white diamonds; Andrews & Martini 2013) whose sSFRs are higher than those of the star-formation main sequence simulated at $z > 1$ (McClymont et al. 2025a). The red solid curves with the 1σ uncertainty (red shades) are the best-fit polynomial functions to our Z-bin stacks, *M*-bin stacks, and the median values of the $z \sim 0$ high-sSFR stacks. For a comparison, we plot the $z \sim 0$ calibrations (Maiolino et al. 2008; Curti et al. 2020; Nakajima et al. 2022; Laseter et al. 2024), the high- z analogue calibrations (Nakajima et al. 2022; Scholte et al. 2025), and the high- z calibrations (Sanders et al. 2024, 2025a; Cataldi et al. 2025a; Chakraborty et al. 2025).

In addition, we serendipitously identify two LRDs in our EMPG candidates. Given their distinctive characteristics, we focus on our EMPG candidates excluding the two LRDs up to Section 7.2, and discuss the LRDs separately in Section 7.3.

5.2 Literature EMPG candidates

We compile the following EMPG candidates from the literature: MPG-CR3 at $z = 3.193$ (Cai et al. 2025), The Cliff at $z = 3.548$ (Ivey et al. 2026a), CAPERS-39810 at $z = 3.654$ (Cai & Yu 2026), LAP2 at $z = 4.19$ (Vanzella et al. 2025), AMORE6-A+B at $z = 5.7253$

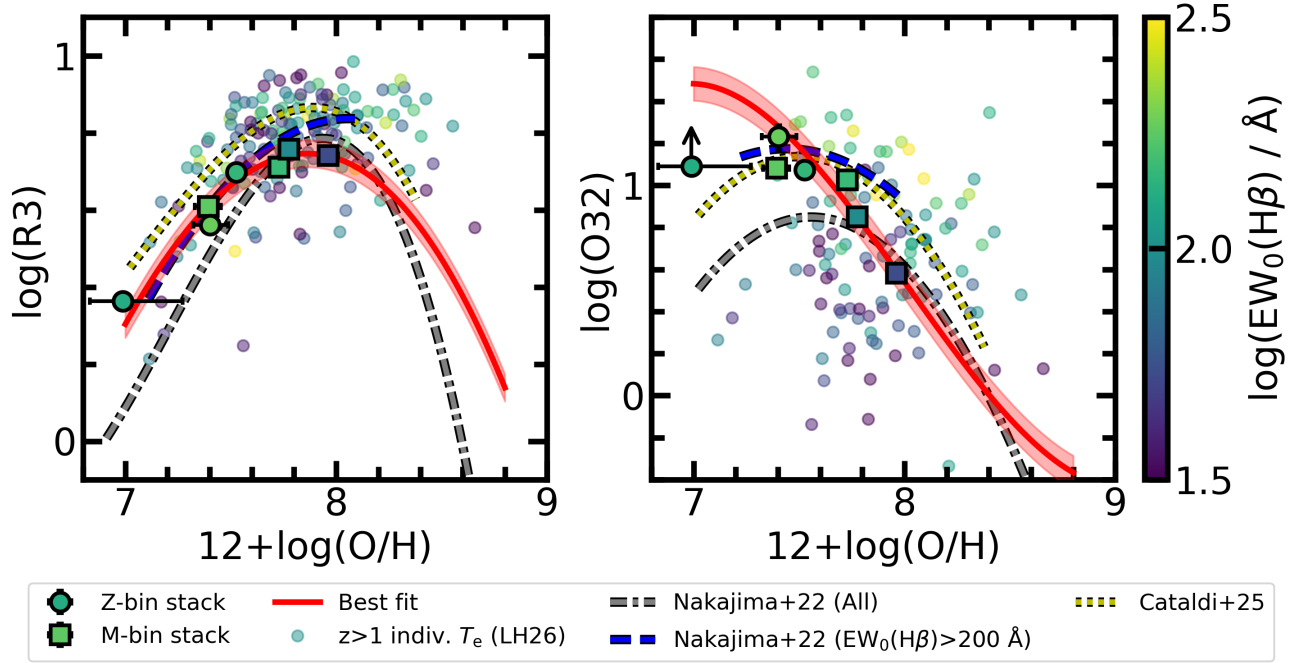


Figure 5. Same as Figure 4 but highlight R3 (left) and O32 (right). Our Z-bin stacks (circles), *M*-bin stacks (squares), and individual auroral-line emitters at $z > 1$ (dots) are colour-coded by their $H\beta$ equivalent widths.

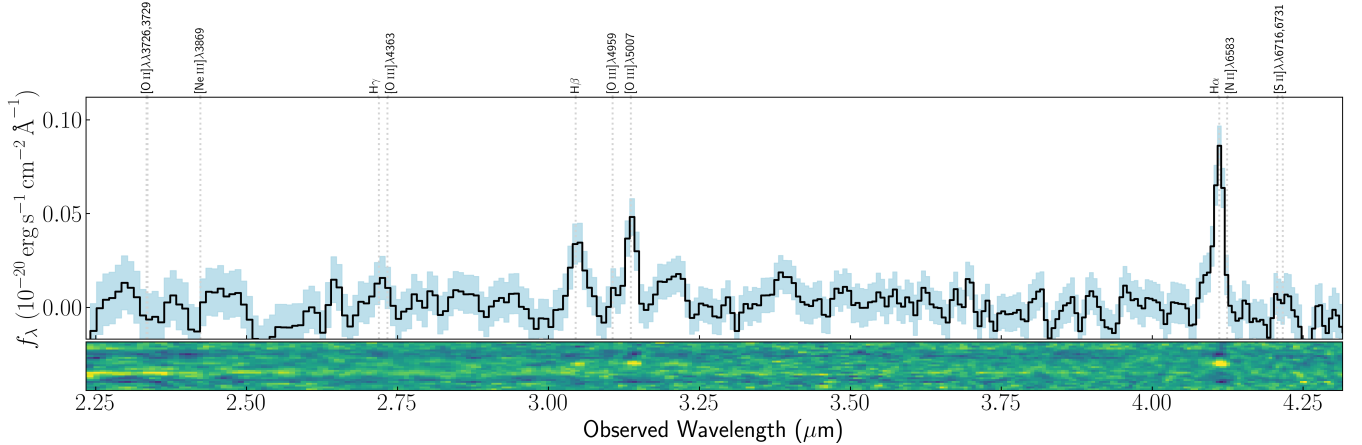


Figure 6. One-dimensional R100 spectrum (top) and two-dimensional spectrum (bottom) of our most metal-poor galaxy candidate (JADES DR4 Unique ID: goods-s-ultradeep_127079 at $z = 5.262$). The strong $H\alpha$ and $H\beta$ lines, the weak $[O\text{ III}]\lambda 5007$, and the absence of $[O\text{ II}]\lambda 3726,3729$ suggest $12 + \log(O/H)_{\text{SL}} = 6.72^{+0.13}_{-0.13}$ based on our calibrations.

(Morishita et al. 2025a), Pseudo-LRD-NOM at $z = 5.96$ (Caputi et al. 2026), T2c at $z = 6.146$ (Vanzella et al. 2024), LAP1-B at $z = 6.625$ (Nakajima et al. 2025), LAP1 at $z = 6.639$ (Vanzella et al. 2023), Abell2744-QSO1 at $z = 7.04$ (hereafter QSO1; Maiolino et al. 2025), CANUCS-A370-z8-LAE at $z = 8.203$ (Willott et al. 2025), Hebe at $z = 10.6$ (Übler et al. 2026), sources from SAPPHIRES (Hsiao et al. 2025), UNCOVER⁶ (Chemerynska et al. 2024), and

GLIMPSE (Fujimoto et al. 2025a; Asada et al. 2026), satellite galaxies of SMACS0723_4590 at $z = 8.45$ and RX2129_11027 at $z = 9.51$ (referred to as ‘satellites’ in this paper; Koller et al. 2026), EXCELS-63107 at $z = 8.271$ (Cullen et al. 2025), and Firefly Sparkle at $z = 8.304$ (Mowla et al. 2024).

We draw M_* values of these EMPG candidates from the literature listed above, except for QSO1 (dynamical mass as an upper limit on M_* ; Juodžbalis et al. 2026) and Hebe (Rusta et al. 2026; Maiolino et al. 2026). We also take SFR values based on $H\alpha$ or $H\beta$ from the literature by scaling the conversion factor to match the value we

⁶ Ultradeep NIRSpec and NIRCarn Observations before the Epoch of Reionization (Bezanson et al. 2024)

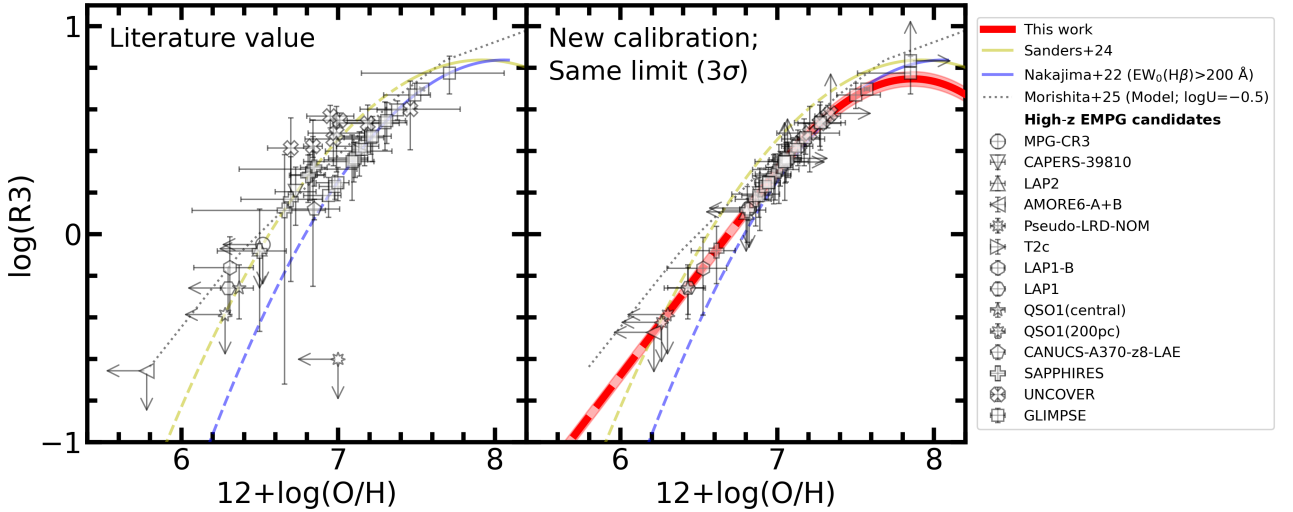


Figure 7. R3 vs. metallicity. (Left) The values simply drawn from the literature. (Right) The values that we uniformly recalculate with our new calibration and the same detection limit (3σ). We plot EMPG candidates from the literature without T_e measurements or $12 + \log(O/H)_{SL}$ values based on our calibrations (Section 5.2).

adopt ($C = -41.64$; Section 4.5), except for MPG-CR3, CAPERS-39810, LAP2, the SAPPHIRES sources, the UNCOVER sources, T2c, GLIMPSE-16043, LAP1-B, LAP1, Firefly Sparkle, and Hebe, which do not have $SFR(H\alpha)$ measurements or the C values used for $SFR(H\alpha)$. For these sources, we uniformly derive $SFR(H\alpha)$ values from the reported Balmer line fluxes in the same manner as in Section 4.5.

Among these EMPG candidates, only EXCELS-63107 and Firefly Sparkle have T_e measurements. In addition, The Cliff, Hebe, and the satellites already have $12 + \log(O/H)_{SL}$ values or upper limits based on our calibrations. For the remaining EMPG candidates, we uniformly adopt our R3 calibration. For the sources whose metallicity values are based on only $[O III]\lambda 5007/H\alpha$, we estimate R3 values by assuming Case B recombination with $H\alpha/H\beta = 2.86$ (Section 3.1). While different works adopted different confidence levels for providing metallicity upper limits (from 3σ to 2σ), here, for consistency, we set a common detection threshold of $S/N > 3$ and give 3σ upper limits for undetected lines in the same manner as for our sample (Section 3.5). As Chemerynska et al. (2024) do not report the actual R3 values used for the $12 + \log(O/H)_{SL}$ estimates, we derive the R3 values from the line flux table provided by Price et al. (2025). Table C2 lists the metallicity values of these EMPG candidates together with their M_* and SFR values.

Figure 7 compares R3 and $12 + \log(O/H)$ values from the literature to those uniformly based on our stack-based calibration and the same 3σ detection threshold. We confirm that our recalculated values are generally higher than the literature values based on either Sanders et al. (2024)’s calibration or the photoionisation models (represented by Morishita et al. 2025a) with a very high $\log(U)$ of -0.5 (also Nakajima et al. 2025), but comparable to those based on Nakajima et al. (2022)’s large-EW calibration. Note that this figure does not necessarily indicate which measurement is more reliable, as most of these candidates are too faint to constrain physical conditions such as ionisation parameters. Rather, we argue that, irrespective of the method employed, $12 + \log(O/H)_{SL}$ values should be compared fairly using a consistent methodology, as well as uniform detection and confidence criteria.

6 SFR-STELLAR MASS RELATION

Figure 8 shows the $SFR-M_*$ relations of the sample discussed in this paper. We show all the R1000 sources from JADES and DH at $z = 4-10$, which we refer to as the JADES+DH sources hereafter. The JADES+DH sources are distributed around the star-formation main sequences (SFMSs) at similar redshifts based on both the observations (Clarke et al. 2025) and the simulations (McClymont et al. 2025a) at $\log(M_*/M_\odot) \gtrsim 8.5$. Below this M_* , the JADES+DH sources exhibit slightly higher SFRs than the simulated SFMS, suggesting an observational bias of high- z emission-line galaxies towards higher SFRs even with NIRSpec (e.g., Clarke et al. 2025).

A key point is that our stacks overlap with the JADES+DH distribution in the $SFR-M_*$ plot, especially at $\log(M_*/M_\odot) \lesssim 8$, indicating that our calibrations empirically include the potential observational bias of the NIRSpec sources at the low-mass end. Compared to the current high- z metallicity calibrations based on individual auroral-line emitters, most of which have $\log(M_*/M_\odot) \gtrsim 8$ (Sanders et al. 2025a), our calibrations trace more representative populations of low-mass star-forming galaxies.

We find that our EMPG candidates generally lie on the low-mass end of the JADES+DH distribution ($\log(M_*/M_\odot) \sim 7-8$), suggesting that our EMPG candidates do not constitute a distinct population within our parent sample. Compared to our EMPG candidates, the majority of the EMPG candidates from the literature exhibit lower stellar masses ($\log(M_*/M_\odot) \lesssim 7$) at similar SFRs, most of which are lensed sources such as AMORE6-A+B (Morishita et al. 2025a).

7 MASS-METALLICITY RELATION & FUNDAMENTAL METALLICITY RELATION

7.1 General redshift evolution

The left panel of Figure 9 shows MZR of the JADES+DH sources at $z = 1-10$ colour-coded by redshift. As the metallicity decreases at a fixed M_* , the colour appears to shift from blue to red, suggesting lower metallicities towards higher z . More quantitatively, we separate the sample with $\log(M_*/M_\odot) = 7-9.5$ into five M_* bins of a 0.5-dex width, and calculate the median value and 16th and 84th percentiles

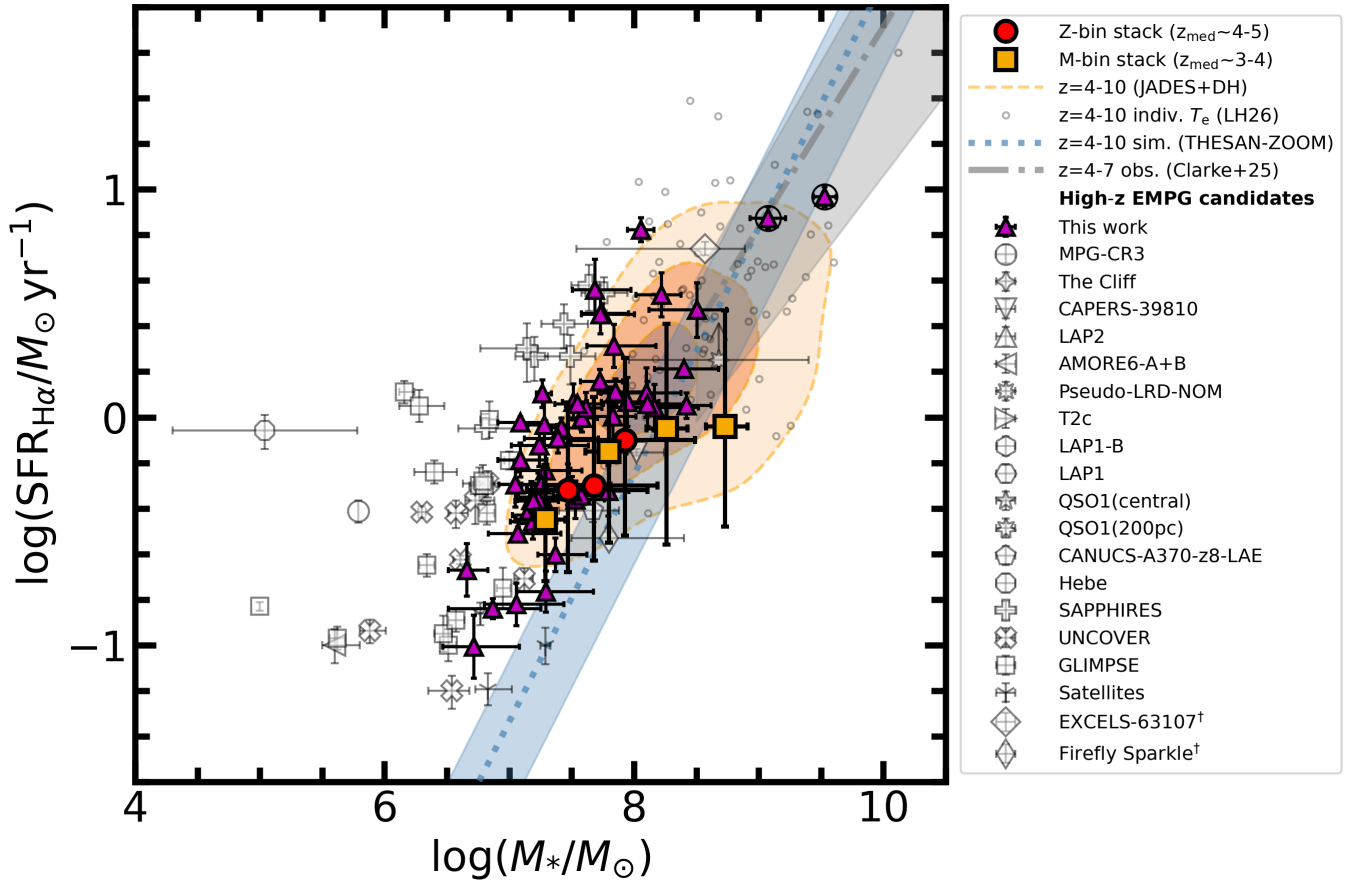


Figure 8. $H\alpha$ -based SFR ($SFR_{H\alpha}$) vs. M_* . The red circles with the error bars represent the median values with the 16th–84th percentiles of the sources used for the Z-bin stacks, and the orange squares show the same measurements for the M-bin stacks. The distribution of the R1000 sources from JADES and DH (JADES+DH sources) at $z = 4$ –10 is shown by the orange dashed contour, whose levels are 16, 50, and 84 percentiles of the kernel density derived using the `gaussian_kde` package. The small grey circles are the individual auroral-line emitters at $z = 4$ –10 (Langeroodi & Hjorth 2026). The blue dotted line with the blue shade corresponds to the star-formation main sequence (SFMS) predicted by the THESAN-ZOOM simulations at $z = 7$ with the range of the SFMSs simulated by THESAN-ZOOM across $z = 4$ –10 (McClymont et al. 2025a). The grey dashdot line with the grey shade is the observed SFMS at $5 < z < 6$ with the range between the SFMSs at $4 < z < 5$ and $6 < z < 7$ (Clarke et al. 2025). The magenta triangles are our EMPG candidates, while the two encircled magenta triangles are Little Red Dots (LRDs) serendipitously identified in our EMPG candidates (Section 7.3). The white points are the EMPG candidates from the literature (Section 5.2). † : EMPGs with T_e measurements.

Table 4. Best-fit parameters of the linear function (Equation 7) fitted to the mass-metallicity relation of the JADES+DH sources within each redshift range.

Redshift	z_{med}	β	Z_8
1–2	1.70	$0.37^{+0.17}_{-0.13}$	$7.89^{+0.13}_{-0.18}$
2–4	3.00	$0.37^{+0.17}_{-0.13}$	$7.76^{+0.09}_{-0.14}$
4–10	5.50	$0.34^{+0.14}_{-0.18}$	$7.62^{+0.11}_{-0.10}$

in each bin. These measurements are shown by the circles with the error bars in the right panel of Figure 9, and fitted with the following linear function used by Curti et al. (2024):

$$12 + \log(\text{O}/\text{H}) = \beta[\log(M_*/M_\odot) - 8] + Z_8. \quad (7)$$

We explore the best-fit values of β and Z_8 by maximising the sum of the log-likelihood function defined by Equation 3. The best-fit functions with their uncertainties based on Monte Carlo simulations with 1000 steps are shown by the lines with the shaded regions in the right panel of Figure 9. Table 4 lists the median redshifts (z_{med}) and

the best-fit parameters with their errors based on the Monte Carlo simulations.

Comparing the results with the $z \sim 0$ MZR (Curti et al. 2020), we confirm the decrease in metallicities from $z \sim 0$, $z = 1$ –2, $z = 2$ –4 to $z = 4$ –10 as pointed out in numerous previous works (e.g., Nakajima et al. 2023; Langeroodi et al. 2023; Curti et al. 2024; Stanton et al. 2026). Indeed, the Z_8 values are almost linearly anticorrelated with $\log(1 + z_{\text{med}})$. Fitting the linear function by maximising the log-likelihood function of Equation 3, we obtain $Z_8(z) = -0.71 \times \log(1 + z) + 8.20$. The $Z_8(z = 0)$ value agrees well with the metallicity of Curti et al. (2020)’s $z \sim 0$ MZR at $M_* = 10^8 M_\odot$. Our MZR at $z = 2$ –4 is well connected to the $z = 2.3$ MZR of Sanders et al. (2021) at $M_* \sim 10^9 M_\odot$. In addition, our MZR at $z = 4$ –10 agrees well with that of the THESAN-ZOOM simulations (McClymont et al. 2025b). Detailed comparisons with more references on high- z MZRs are presented in Appendix D.

We find that our MZR slopes do not significantly evolve from $z = 1$ –2, $z = 2$ –4 to $z = 4$ –10 with $\beta = 0.34$ –0.37, which matches the slope at the low-mass end of the $z \sim 0$ MZR ($\beta \rightarrow 0.28$; Curti et al. 2020). This suggests a small evolution of the MZR slopes

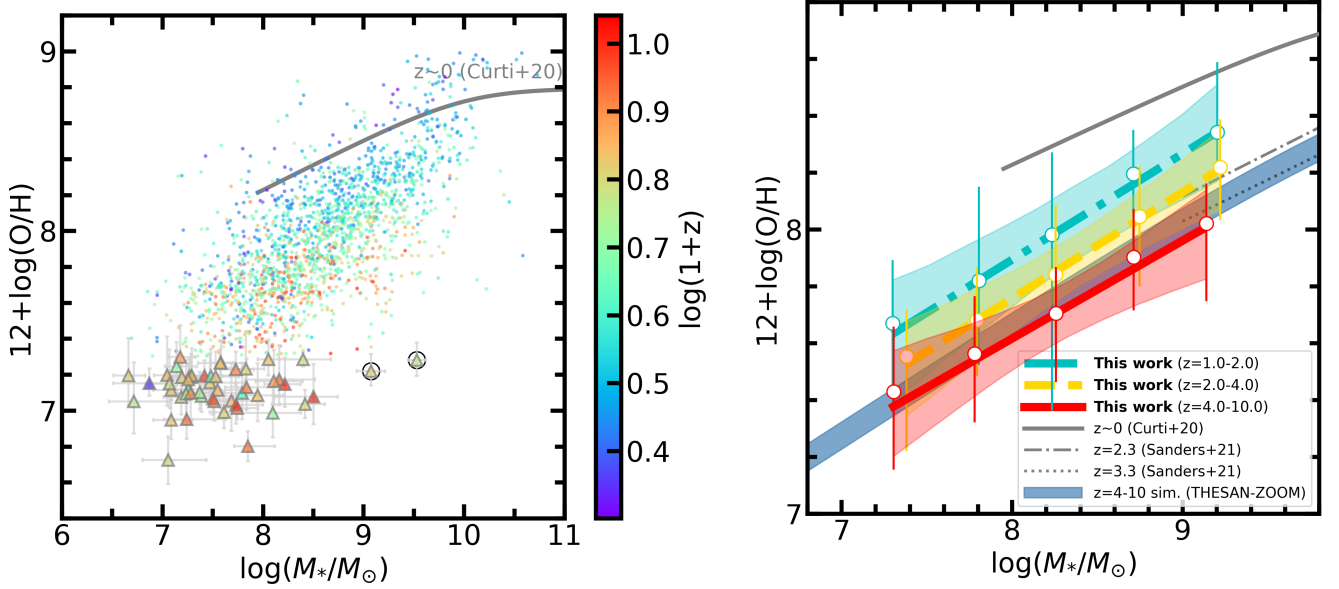


Figure 9. Mass-metallicity relation of our JADES+DH sources at $z = 1-10$. (Left) Actual distribution (dots) together with our EMPG candidates (triangles). These data points are colour-coded by z . The grey solid curve is the $z \sim 0$ MZR reported by Curti et al. (2020). (Right) Median values with 16th and 84th percentiles in bins of M_* shown by the open circles in cyan ($z = 1-2$), yellow ($z = 2-4$), and red ($z = 4-10$). The best-fit linear functions of these measurements with the uncertainties are shown by the cyan dashdot line with the cyan shade ($z = 1-4$), the yellow dashed line with the yellow shade ($z = 2-4$), and the red solid line with the red shade ($z = 4-10$). The blue shaded region shows the MZR predicted by the THESAN-ZOOM simulations at $z = 4-10$ (McClymont et al. 2025b). We confirm the decrease in metallicity towards higher z . Our $z = 4-10$ MZR matches the prediction by the THESAN-ZOOM simulations at the same redshift.

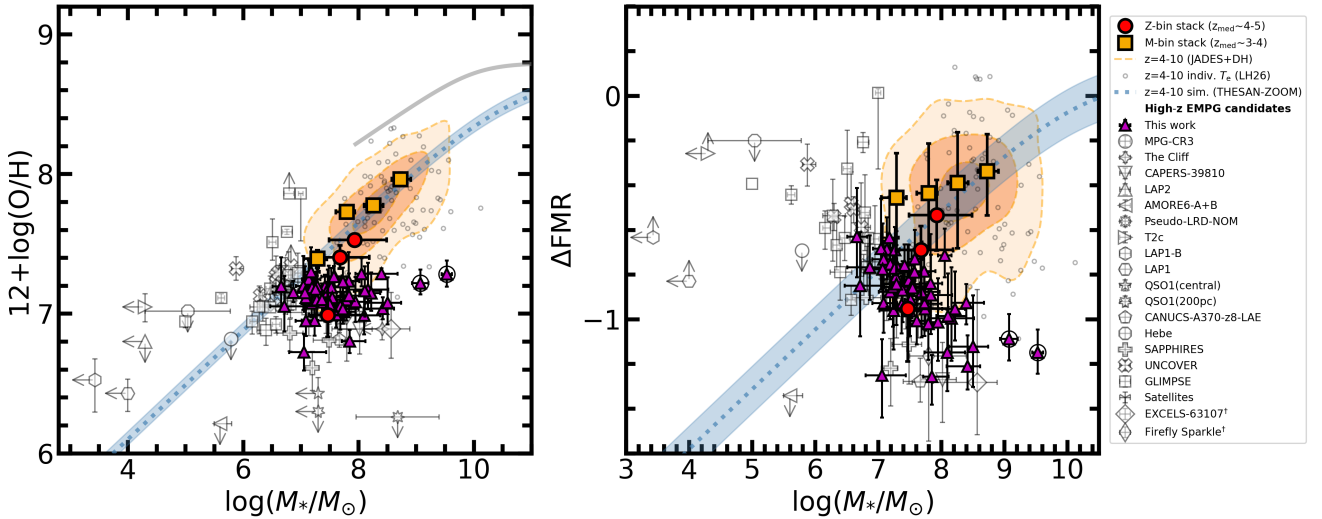


Figure 10. Mass-metallicity relation (left) and metallicity offset from Curti et al. (2020)’s $z \sim 0$ Fundamental Metallicity Relation (ΔFMR ; right). The symbols are the same as those in Figure 8, while the errors on the metallicities of the Z -bin and M -bin stacks are the errors measured from the stacked spectra.

across a wide z range. It has been argued that a physical mechanism of galactic winds determines the MZR slope at the metallicity equilibrium, where energy-driven winds predict a steeper slope ($\beta \sim 0.33$; Davé et al. 2012; Lilly et al. 2013) than momentum-driven winds ($\beta \sim 0.17$; Guo et al. 2016). Our MZR slopes agree well with the steeper slope, supporting the scenario of energy-driven winds.

7.2 Low-mass end

The magenta triangles in Figure 10 show our EMPG candidates. We find that our EMPG candidates generally have lower $12 + \log(\text{O}/\text{H})$

than the overall R1000 sources. Although the exposure times of JADES, DH, and OASIS are long (Section 2), the observational limit on M_* for the unlensed fields is at $\sim 10^7 M_\odot$. Most our EMPG candidates exhibit higher M_* values than the simulated MZR, likely because the simulated MZR predicts $M_* \lesssim 10^7 M_\odot$ for our metallicity threshold of $12 + \log(\text{O}/\text{H}) < 7.3$ (a.k.a. Malmquist bias; Malmquist 1922). Indeed, high- z EMPG candidates from the literature show a large scatter in the MZR plot, some of which are above the simulated MZR. Such EMPG candidates are mostly lensed sources such as LAPI (Vanzella et al. 2023).

The right panel of Figure 10 shows the offset of the observed

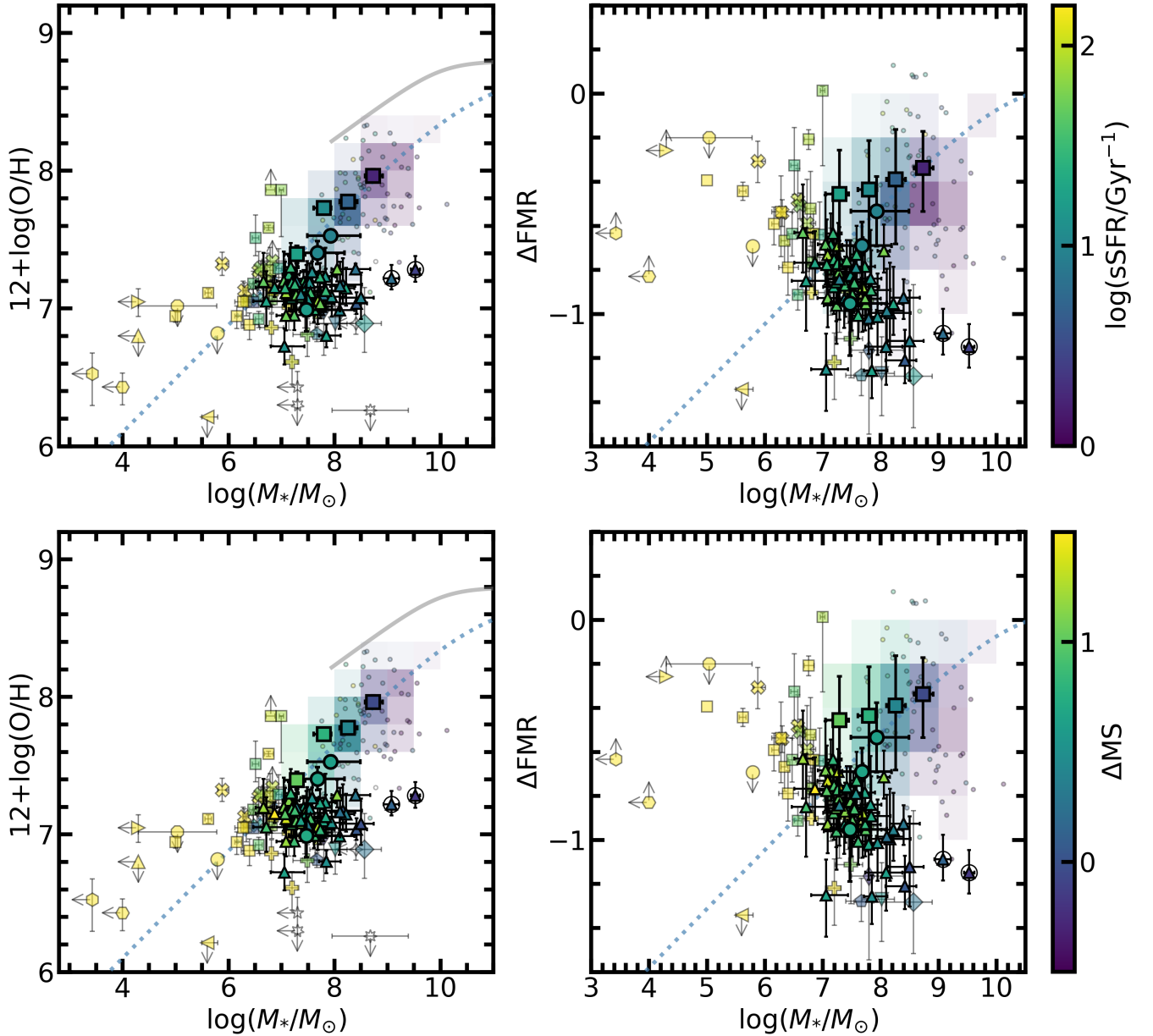


Figure 11. Same as Figure 10 but colour-coded by $H\alpha$ -based sSFR (top) and the offset of $H\alpha$ -based SFRs from the THESAN-ZOOM simulations (ΔMS ; bottom). The blue dotted curves are the prediction by the THESAN-ZOOM simulations (McClymont et al. 2025a,b) at $z = 7$. More metal-poor galaxies generally show lower sSFRs and ΔMS values.

$12 + \log(O/H)$ values from those inferred from the $z \sim 0$ FMR (Curti et al. 2020) based on the $SFR_{H\alpha}$ values (ΔFMR , hereafter). As reported by several studies on low-mass galaxies (see Section 1), we confirm that the R1000 sources at $z = 4-10$ exhibit systematically negative ΔFMR values. Notably, our EMPG candidates have even lower ΔFMR values, some of which reach < -1 . Combining the EMPG candidates from the literature, we identify a large scatter in the $\Delta FMR-M_*$ relations as well.

To discuss the origin of the large scatters, we colour-code the data points in Figure 10 by sSFRs, which are shown in the top panels of Figure 11. We find that the EMPG candidates below the MZR tend to have lower sSFRs than those above the MZR. We identify a similar trend in the $\Delta FMR-M_*$ plane. In addition, we calculate the offset of $SFR_{H\alpha}$ from the SFMS predicted by the THESAN-ZOOM

simulations (ΔMS). The bottom panels of Figure 11 illustrate that the EMPG candidates below the simulated MZR and the $\Delta FMR-M_*$ relation tend to have smaller ΔMS values. Figure 12 shows the relation between ΔFMR and ΔMS of the EMPG candidates and the JADES+DH sources, separating them into those with $\log(M_*/M_\odot) \leq 8$ and $\log(M_*/M_\odot) > 8$. We find that the EMPG candidates exhibit a positive correlation between ΔFMR and ΔMS in the $\log(M_*/M_\odot) \leq 8$ bin. Moreover, we identify a positive correlation in the JADES+DH sources with $\log(M_*/M_\odot) \leq 8$. On the other hand, these correlations look weaker in the $\log(M_*/M_\odot) > 8$ bin. These results suggest that more metal-poor galaxies have *less* bursty star formation at the low-mass end.

This trend is opposite to the well-established $z \sim 0$ observational work, which has reported that more metal-poor galaxies have higher

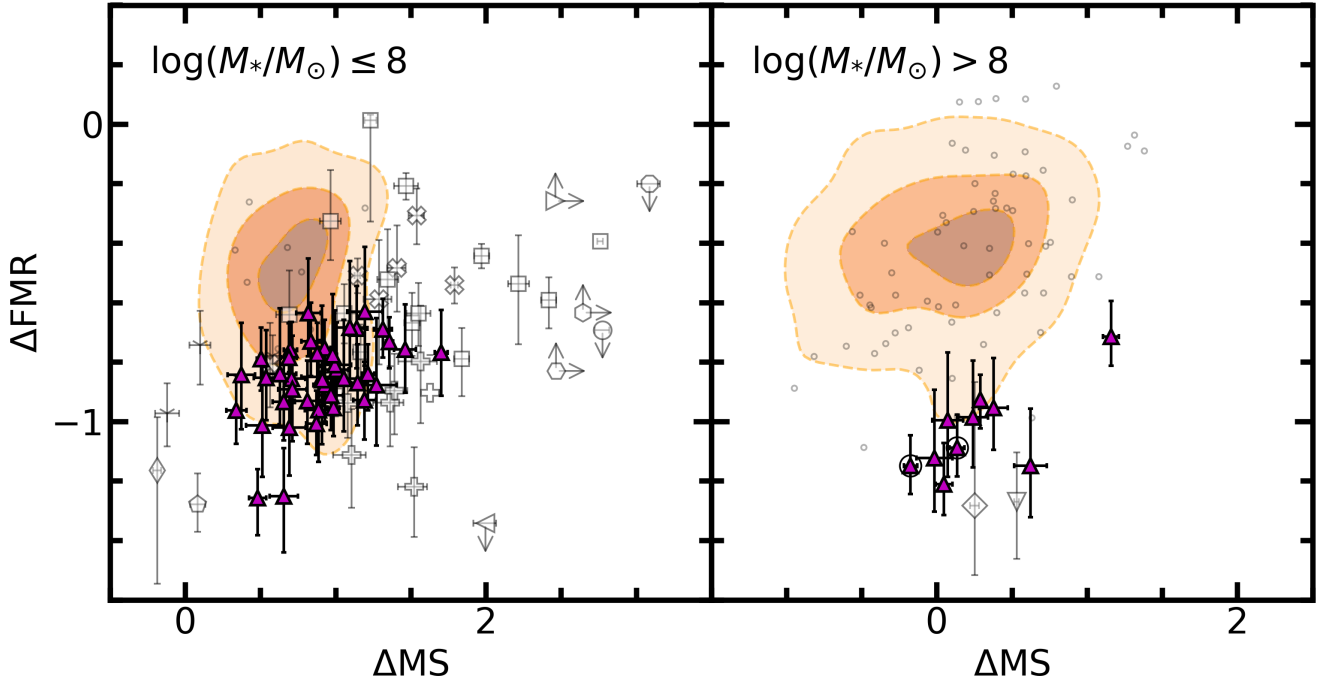


Figure 12. ΔFMR vs. ΔMS for sources with $\log(M_*/M_\odot) \leq 8$ (left) and $\log(M_*/M_\odot) > 8$ (right). The symbols are the same as in Figure 10. Both EMPG candidates (magenta triangles and white points) and the JADES+DH sources at $z = 4\text{--}10$ (orange contour) exhibit positive correlations between ΔFMR and ΔMS in the $\log(M_*/M_\odot) \leq 8$ bin, while the correlations look weaker in the $\log(M_*/M_\odot) > 8$ bin.

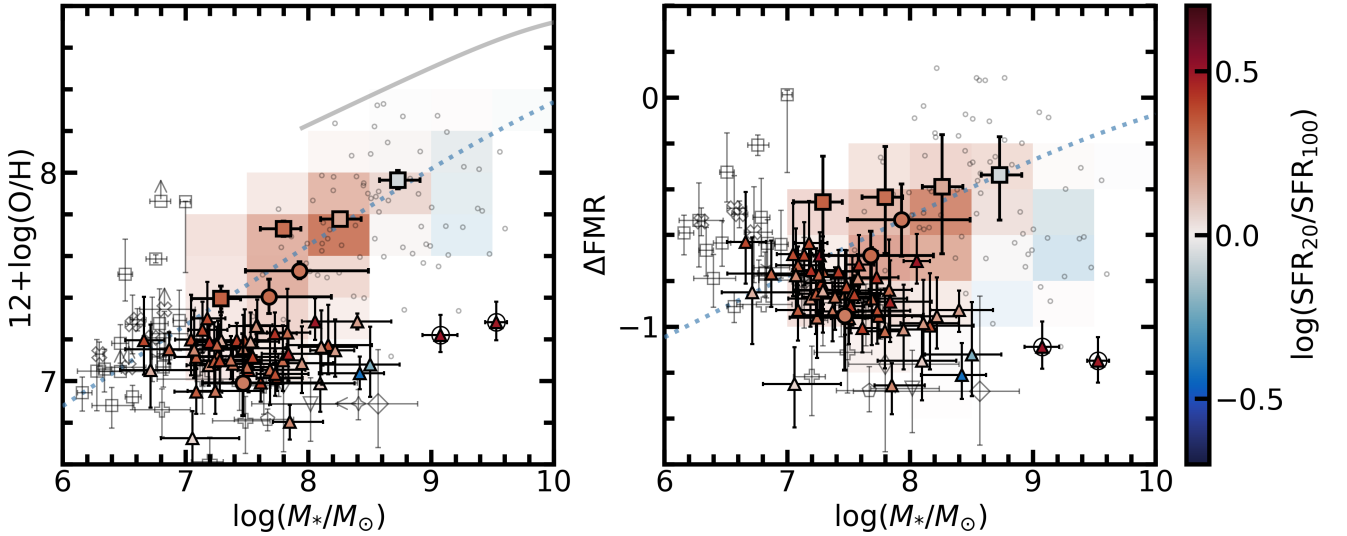


Figure 13. Same as Figure 10 but we colour-code our EMPG candidates and the JADES+DH sources by $\text{SFR}_{20}/\text{SFR}_{100}$ (Section 2.4). Most of our EMPG candidates have $\text{SFR}_{20}/\text{SFR}_{100} > 0$.

sSFRs (e.g., Andrews & Martini 2013; Curti et al. 2020). However, these studies are based on more massive galaxies ($M_* \gtrsim 10^9 M_\odot$) than the high- z sources that we discuss. Indeed, a recent observational work on low-mass galaxies at $z \sim 0$ (Laseter et al. 2025) reports that galaxies with more negative ΔFMR values tend to have lower sSFRs at a given M_* , suggesting that this trend may be specific to low-mass galaxies across a range of redshifts.

Two main scenarios have been proposed to explain more negative

ΔFMR values: inflows of metal-poor gas and outflows of metal-enriched gas (e.g., Nishigaki et al. 2025). It has been argued that the ΔFMR decrease cannot be explained by inflows unless the pre-existing ISM is substantially lost (Dalcanton 2007; Tacchella et al. 2023; Laseter et al. 2025). However, the THESAN-ZOOM simulations (McClymont et al. 2025a,b) predict such significant mass loss in low-mass galaxies with shallow potential wells. These simulations indicate a stochastic SFH characterised by cycles of starburst, gas

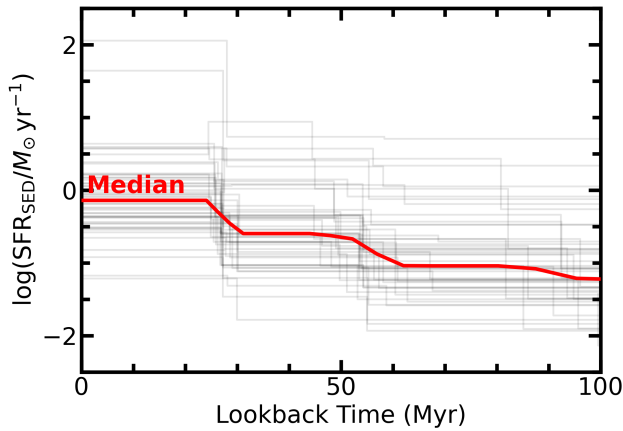


Figure 14. SFRs based on the SED fitting (SFR_{SED} ; Section 2.4) of our EMPG candidates (grey thin lines) and their median value (red thick line) as a function of lookback time. Our EMPGs generally exhibit a rising star-formation history.

consumption and ejection, temporary quenching, subsequent pristine inflow, and rejuvenated starburst. In these simulations, the evolutionary tracks of galaxies deviate from the canonical MZR during phases of downward motion, indicating that these galaxies have recently undergone a gas-poor, mini-quenched phase, during which metal-poor gas inflow substantially dilutes their metallicity. Since a shorter timescale of inflow is enough to produce a significant metallicity dilution for a smaller pre-existing gas reservoir, lower-mass galaxies are expected to recover their metallicities more rapidly than their ΔMS values, leading to lower ΔMS values than those of the galaxies on the canonical MZR. In fact, both the evolutionary tracks and the bulk of simulated galaxies below the MZR show more negative ΔMS values, which agree with the observational results.

In this scenario, galaxies below the MZR are thought to be observed at a stage when their inflow-triggered SFR is still increasing. Consequently, the SFR measured over recent timescales is expected to exceed that averaged over longer timescales. To test this, we colour-code our EMPG candidates by $\text{SFR}_{20}/\text{SFR}_{100}$ in Figure 13, which are derived from the SED fitting (Section 2.4). We find that most candidates exhibit $\log(\text{SFR}_{20}/\text{SFR}_{100}) > 0$. Indeed, Figure 14 shows the SFHs of our EMPG candidates. The median SFR_{SED} value increases with decreasing lookback time, indicating that more recent star formation is generally more active. Such rising SFHs are in agreement with the inflow scenario.

Of course, this consistency does not rule out the enriched outflow scenario. However, our results place constraints on such outflow models, requiring that $\log(\text{SFR}_{20}/\text{SFR}_{100})$ remains positive even when some gas is lost through outflows.

7.3 Metal-poor LRDs

From our EMPG candidates, we identify two sources whose rest-frame optical colour is significantly red: goods-n-mediumjwst_39353 and goods-s-mediumjwst_73690. Figure 15 shows the spectra and photometric points of these sources. Both sources are included in the updated census of LRDs in GOODS-N and GOODS-S by [Rinaldi et al. \(2026\)](#), as they satisfy the more inclusive photometric criteria adopted in that work: $\text{F150W} - \text{F200W} < 1$ mag and $\text{F277W} - \text{F444W} > 0.5$ mag. Specifically, goods-s-mediumjwst_73690 has $\text{F150W} - \text{F200W} = 0.18$ mag and $\text{F277W} - \text{F444W} = 1.16$ mag, while goods-n-mediumjwst_39353 has $\text{F150W} - \text{F200W} = 0.01$ mag

and $\text{F277W} - \text{F444W} = 0.88$ mag. Crossmatching with the JADES DR5 morphology catalogue ([Carreira et al. 2026](#)), we find that both sources exhibit effective radii of F444W smaller than the size of the point-spread function in the same filter, indicating that they are extremely compact. The cutout images generated using TrilogY ([Coe et al. 2012](#)) also illustrate the compact morphology. Hereafter, we refer to these two sources as EMP LRDs.

The right panels show the R1000 spectra of the EMP LRDs around $\text{H}\beta + [\text{O III}]\lambda\lambda 4959, 5007$ and $\text{H}\alpha + [\text{N II}]\lambda\lambda 6548, 6583$. We find that both EMP LRDs appear to have a broad component in $\text{H}\alpha$. Following the methodology of [Juodžbalis et al. \(2025\)](#) and using the R1000 data, we fit the model that includes narrow components of $[\text{N II}]\lambda\lambda 6548, 6583$ and narrow+broad components of $\text{H}\alpha$. We fix the ratio of $[\text{N II}]\lambda\lambda 6548/[\text{N II}]\lambda\lambda 6583$ to 1/3, and assume that all the components share the same systemic velocity. We constrain the narrow components to have the same velocity width. We obtain S/N of the broad component (S/N_{Broad}) from the ratio of the broad line flux integrated within the full-width half maximum to the error added in quadrature over the same range. We also fit the model without the broad $\text{H}\alpha$ component (i.e., narrow-component only model). The fitting results are shown in Figure C1. We calculate the Bayesian information criterion (BIC) for the narrow-component only model ($\text{BIC}_{\text{Narrow}}$) as well as the original model ($\text{BIC}_{\text{Narrow+Broad}}$). We obtain $\Delta\text{BIC} \equiv \text{BIC}_{\text{Narrow}} - \text{BIC}_{\text{Narrow+Broad}} = 6.9\text{--}77$ and $S/N_{\text{Broad}} = 6.4\text{--}15$ for the EMP LRDs, both exceeding 5 and thus satisfying the criteria used by [Juodžbalis et al. \(2025\)](#) for statistically supporting the presence of a broad-line component. Note that these objects are not included in [Juodžbalis et al. \(2025\)](#)’s Type-1 AGN catalogue simply because they were observed more recently than the data analysed by that study. On the other hand, the $[\text{O III}]\lambda 5007$ lines do not exhibit such a significant broad component, suggesting that the observed broad components of $\text{H}\alpha$ are emitted from AGN broad-line regions.

Our EMP LRDs, which likely host growing black holes in a very low metallicity environment, are similar to other known low-metallicity LRDs, such as QSO1 ([Maiolino et al. 2025](#)) and The Cliff ([Ivey et al. 2026a](#)). As discussed in [Maiolino et al. \(2025\)](#), [Juodžbalis et al. \(2026\)](#), and [Ivey et al. \(2026a\)](#), these low-metallicity LRDs likely trace a very early phase of black hole formation, also supporting a heavy seeding scenario for their origin, at least for these specific cases. Indeed, as highlighted by the encircled magenta triangles in Figure 10, our EMP LRDs exhibit the two highest values of stellar masses beyond $\log(M_*/M_\odot) > 9$, which cause ~ 1 -dex lower metallicities than those of the JADES+DH sources at a given M_* . [Ivey et al. \(2026a\)](#) discuss that such metallicity tension can possibly be seen in QSO1 ([Maiolino et al. 2025](#); [Juodžbalis et al. 2026](#)) and The Cliff ([Ivey et al. 2026a](#)) as well as $z \sim 0$ LRD analogues ([Lin et al. 2026](#)), given that reducing M_* to match the canonical MZR would require even more extreme conditions such as higher black-hole-to-stellar mass ratios beyond 1 ($M_{\text{BH}}/M_* > 1$) or higher Eddington ratios beyond 100 ($\lambda_{\text{Edd}} > 100$). A detailed analysis on M_{BH} and M_* for our EMP LRDs will be presented in a separate paper.

In addition, both EMP LRDs exhibit strong $\text{Ly}\alpha$ even with relatively high $E(B-V)$ values of ~ 0.5 , one of which has been reported in the JADES $\text{Ly}\alpha$ emitter catalogue ([Jones et al. 2025b](#)). Following the procedure of [Jones et al. \(2024\)](#) and using the R100 data, we fit $\text{Ly}\alpha$ together with the UV continuum modelled by a power-law continuum redder than $\text{Ly}\alpha$ and the zero flux bluer than $\text{Ly}\alpha$ convolved with a Gaussian of width $F_R\sigma_R$, where σ_R is the dispersion of the R100 LSF ([de Graaff et al. 2024](#)) and F_R is a variable factor (≥ 1) to match the spectra. We obtain $F(\text{Ly}\alpha) = 5.7\text{--}7.9 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$ and high values of $F(\text{Ly}\alpha)/F(\text{H}\beta) = 5.9\text{--}12$ without dust-reddening corrections, which already reach 25%–50% of the ratio inferred from

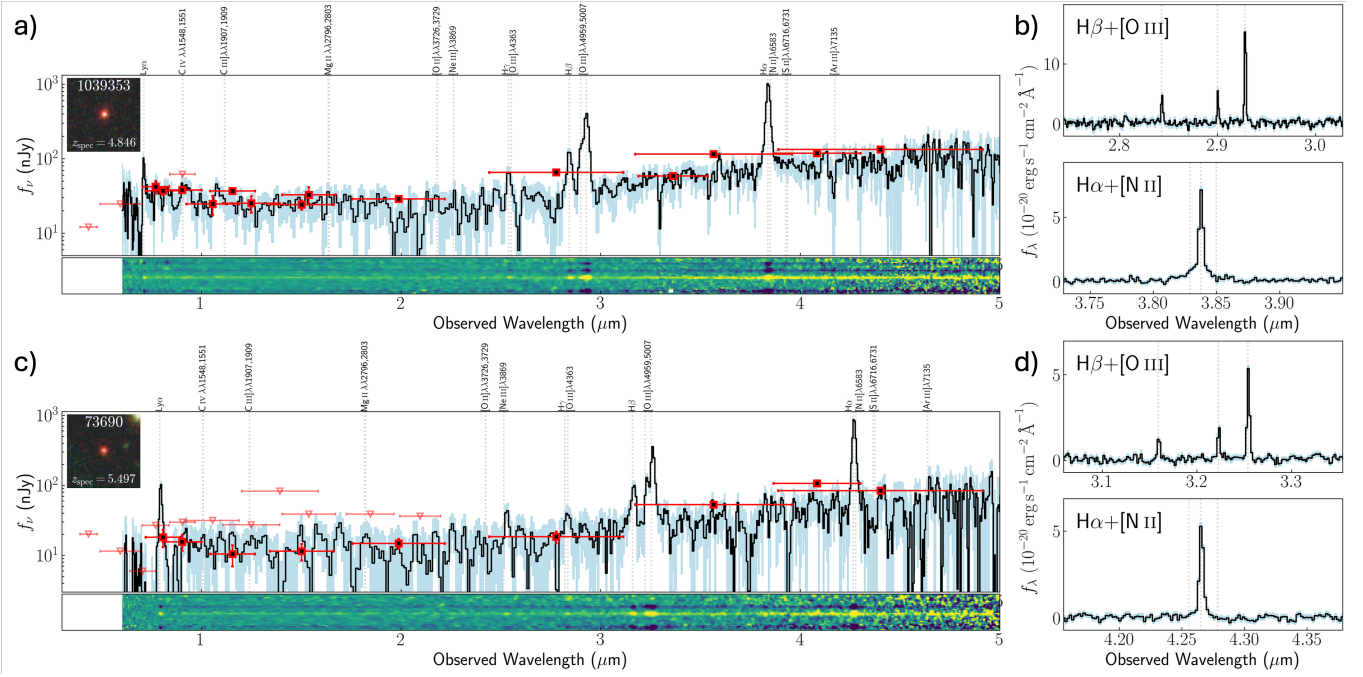


Figure 15. Spectra of LRDs identified in our EMPG candidates: goods-n-mediumjwst_39353 and goods-s-mediumjwst_73690. **a):** R100 spectrum of goods-n-mediumjwst_39353 with the Kron photometry of NIRCcam, ACS, and WFC3 (red square: $S/N > 3$; red triangle: 3σ upper limit). The inset panel shows the $2'' \times 2''$ cutout image of NIRCcam (R: F444W; G: F200W; B: F115W). **b):** R1000 goods-n-mediumjwst_39353 around $H\beta + [O III] \lambda\lambda 4959, 5007$ (top) and $H\alpha + [N II] \lambda\lambda 6548, 6583$ (bottom). **c) and d):** Same as a) and b), respectively, but for goods-s-mediumjwst_73690. Both sources exhibit blue UV colour and red optical colour that satisfy the selection criteria of LRDs proposed by Rinaldi et al. (2026). These sources also have a broad component in $H\alpha$ but not in $[O III]$, suggesting the presence of accreting black holes.

Case B recombination ($T_e = 15000$ K and $n_e = 100 \text{ cm}^{-3}$). Such a high $Ly\alpha/H\beta$ ratio comparable to Case B is also reported in another metal-poor LRD, QSO1 (Ji et al. 2026; Tang et al. 2026), in contrast to the low $Ly\alpha/H\alpha = 0.07$ reported in another LRD, UNCOVER-45924 (Torralba et al. 2026), with a relatively high metallicity of $12 + \log(O/H) = 7.15\text{--}7.67$ (Ji et al. 2025). Our EMP LRDs do not have significant Balmer breaks, which is in line with the recent report that the stack of LRDs with weak Balmer breaks exhibits a more prominent $Ly\alpha$ than those with strong Balmer breaks (Ando et al. 2026). Details of the $Ly\alpha$ properties of the EMP LRDs will also be discussed in a separate paper.

8 CONCLUSIONS

We present gas-phase metallicities of ~ 2500 star-forming galaxies at $z = 1\text{--}10$ with the NIRSpec R100 and R1000 data from the JADES, DH, and OASIS surveys. The large number of deep R1000 spectra allows us to detect $[O III] \lambda 4363$ from all our R1000 stacks in narrow bins of $12 + \log(O/H)_{SL}$ and M_* . Our findings are summarised as follows:

- We derive T_e -based metallicities for our stacks down to $12 + \log(O/H) = 7.0$, entering the regime where the high- z strong-line method was not well calibrated ($12 + \log(O/H) \gtrsim 7.3$).
- At a given metallicity, our stacks have $EW_0(H\beta)$ lower than individual auroral-line emitters at $z > 1$. These results suggest that requiring faint auroral-line detections biases the samples towards galaxy populations with higher excitation and more bursty star formation. Indeed, our stacks have R3, R23, O32, $\hat{R}_{\text{Laseter}}$, $\hat{R}_{\text{Chakraborty}}$, $\hat{R}_{\text{Cataldi}}$,

and Ne3O2 that are generally lower than those of high- z calibrations from the literature based on individual auroral-line emitters.

- Using our $z = 1\text{--}10$ stacks, supplemented with stacks of $z \sim 0$ metal-rich galaxies (with sSFRs higher than the $z = 1$ SFMS), we obtain stack-based strong-line calibrations over the range of $12 + \log(O/H) = 7.0\text{--}8.7$. Our R3 and R23 calibrations agree well with those of $z \sim 0$ galaxies with large $EW_0(H\beta)$.
- We report MZR with the R1000 data at $z = 1\text{--}10$, confirming the decrease in metallicities from $z \sim 0$, $z = 1\text{--}2$, $z = 2\text{--}4$, to $z = 4\text{--}10$ at a given M_* . We do not identify a significant evolution of the MZR slope ($\beta = 0.34\text{--}0.37$), which suggests that a dominant feedback mechanism does not change drastically in general.
- Based on our stack-based calibrations, we select 50 promising EMPG candidates from our ~ 2500 galaxies. Our EMPG candidates have $12 + \log(O/H) = 6.7\text{--}7.3$ at $z = 1.2\text{--}9.1$, most of which exhibit lower metallicities than the parent sample and the simulated MZR at $z = 4\text{--}10$.
- Compared to literature EMPG candidates above the MZR, our EMPG candidates generally have lower sSFRs while their SFR_{20} values are mostly higher than SFR_{100} . These results are consistent with a stochastic SFH with cycles of gas consumption/ejection and metal-poor inflow, as predicted by zoom-in simulations.
- We find that two of our EMPG candidates have red optical colours ($F277W - F444W = 0.88\text{--}1.16$ mag) together with relatively blue UV colours ($F150W - F200W = 0.01\text{--}0.18$ mag), which are included in the updated census of LRDs by (Rinaldi et al. 2026). These two LRDs exhibit significant $H\alpha$ broadening and high $Ly\alpha/H\beta$ ratios close to the Case B value. These are therefore Extremely Metal Poor LRDs, similar to a few others that have been discovered recently, which are likely tracing the earliest phases of black hole formation.

ACKNOWLEDGMENTS

We thank Tadayuki Kodama, Takahiro Morishita, Yoshihisa Asada, Yuma Sugahara, Thomas Stanton, Karina Caputi, Nicholas Choustikov, Callum Witten, Ali Ahmad Khostovan, Leonardo Clarke, Mingyu Li, and Fengwu Sun for useful discussions. YI is supported by JSPS KAKENHI Grant No. 24KJ0202. RM, DP, FDE, JS, XJ, GCJ, IJ, and RGP acknowledge support by the Science and Technology Facilities Council (STFC), by the ERC through Advanced Grant 695671 “QUENCH”, and by the UKRI Frontier Research grant RISEandFALL. RM also acknowledges funding from a research professorship from the Royal Society. QD acknowledges support from a PhD studentship awarded by Trinity College, University of Cambridge. WM thanks the Science and Technology Facilities Council (STFC) Center for Doctoral Training (CDT) in Data Intensive Science at the University of Cambridge (STFC grant number 2742968) for a PhD studentship. WM and ST acknowledge support by the Royal Society Research Grant G125142. DP and IJ acknowledge support by the Huo Family Foundation through a P.C. Ho PhD Studentship. JAAT acknowledges support from the Simons Foundation and JWST program 3215. Support for program 3215 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127. TJL gratefully acknowledges support from the Swiss National Science Foundation through a SNSF Mobility Fellowship and from the NASA/JWST Program OASIS (PID #5997). The work presented here is supported by the Carlsberg Foundation, grant CF25-0662 (DL). ST acknowledges support by the Royal Society Research Grant G125142. HÜ acknowledges support by the Max Planck Society through the Lise Meitner Excellence Program. HÜ acknowledges funding by the European Union (ERC APEX, 101164796). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. WMB gratefully acknowledges support from DARK via the DARK fellowship. This work was supported by a research grant (VIL54489) from VILLUM FONDEN. AJB acknowledges funding from the “FirstGalaxies” Advanced Grant from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 789056). SC acknowledges support by European Union’s HE ERC Starting Grant No. 101040227 - WINGS. ECL acknowledges support of an STFC Webb Fellowship (ST/W001438/1). SG is supported by a Woolf Fisher Scholarship from the Woolf Fisher Trust of New Zealand and Cambridge Commonwealth and European Trust. MK acknowledges support from a PhD studentship awarded by the University of Cambridge Harding Distinguished Postgraduate Scholars Programme, UK Science and Technology Facilities Council (STFC) Center for Doctoral Training (CDT) in Data Intensive Science, and Girton College Cambridge. BER acknowledges support from the NIRCcam Science Team contract to the University of Arizona, NAS5-02105, and JWST Program 3215. The research of CCW is supported by NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. This work is based on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. These observations are associated with programmes #1180, 1181, 1210,

1286, 1287, 3215, 4540, 1283, 1895, 1963, 2079, 2514, 3577, 3990, 5398, 5997, 6434, and 6541. The authors acknowledge use of the lux supercomputer at UC Santa Cruz, funded by NSF MRI grant AST 1828315.

DATA AVAILABILITY

The bulk of the JWST/NIRSpec data and all the NIRCcam data used in this paper are released by the JADES NIRSpec DR4 (Curtis-Lake et al. 2025; Scholtz et al. 2025a) and the JADES NIRCcam DR5 (Robertson et al. 2026; Johnson et al. 2026), which are available on the JADES MAST website (<https://archive.stsci.edu/hlsp/jades>; MAST DOI: 10.17909/8tdj-8n28) and the JADES Public Online Database (<https://jades.herts.ac.uk/search/>). The rest of datasets will also be public in the MAST archive. Our analysed data will be made available upon reasonable request.

REFERENCES

- Alves de Oliveira C., et al., 2018, in *Observatory Operations: Strategies, Processes, and Systems VII*. p. 107040Q ([arXiv:1805.06922](https://arxiv.org/abs/1805.06922)), doi:10.1117/12.2313839
- Ando M., Harikane Y., Katz H., Inayoshi K., Tanaka T. S., 2026, *arXiv e-prints*, p. [arXiv:2606.03522](https://arxiv.org/abs/2606.03522)
- Andrews B. H., Martini P., 2013, *ApJ*, **765**, 140
- Asada Y., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.20045](https://arxiv.org/abs/2601.20045)
- Asplund M., Amarsi A. M., Grevesse N., 2021, *A&A*, **653**, A141
- Bagley M. B., et al., 2024, *ApJ*, **965**, L6
- Baker W. M., Maiolino R., 2023, *MNRAS*, **521**, 4173
- Baker W. M., et al., 2023, *MNRAS*, **519**, 1149
- Baldwin A., Phillips M. M., Terlevich R., 1981, *PASP*, **93**, 817
- Berg D. A., Erb D. K., Henry R. B. C., Skillman E. D., McQuinn K. B. W., 2019, *ApJ*, **874**, 93
- Berg D. A., Chisholm J., Erb D. K., Skillman E. D., Pogge R. W., Olivier G. M., 2021, *ApJ*, **922**, 170
- Bezanson R., et al., 2024, *ApJ*, **974**, 92
- Bian F., Kewley L. J., Dopita M. A., 2018, *ApJ*, **859**, 175
- Brown J. S., Martini P., Andrews B. H., 2016, *MNRAS*, **458**, 1529
- Bunker A. J., et al., 2024, *A&A*, **690**, A288
- Burnham K. P., Anderson D. R., 2004, *Sociological Methods & Research*, **33**, 261
- Cai S., Yu Z., 2026, *arXiv e-prints*, p. [arXiv:2601.17498](https://arxiv.org/abs/2601.17498)
- Cai S., et al., 2025, *arXiv e-prints*, p. [arXiv:2507.17820](https://arxiv.org/abs/2507.17820)
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, **533**, 682
- Cameron A. J., Katz H., Rey M. P., Saxena A., 2023, *MNRAS*, **523**, 3516
- Cameron A. J., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.15964](https://arxiv.org/abs/2601.15964)
- Cappellari M., 2023, *MNRAS*, **526**, 3273
- Caputi K. I., Cooper R. A., Rinaldi P., Navarro-Carrera R., Iani E., 2026, *arXiv e-prints*, p. [arXiv:2601.11466](https://arxiv.org/abs/2601.11466)
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, **345**, 245
- Carnall A. C., 2017, *arXiv e-prints*, p. [arXiv:1705.05165](https://arxiv.org/abs/1705.05165)
- Carreira C., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.15957](https://arxiv.org/abs/2601.15957)
- Cataldi E., et al., 2025a, *arXiv e-prints*, p. [arXiv:2504.03839](https://arxiv.org/abs/2504.03839)
- Cataldi E., et al., 2025b, *arXiv e-prints*, p. [arXiv:2512.07955](https://arxiv.org/abs/2512.07955)
- Chabrier G., 2003, *PASP*, **115**, 763
- Chakraborty P., et al., 2025, *ApJ*, **985**, 24
- Charlot S., Fall S. M., 2000, *ApJ*, **539**, 718
- Chemerynska I., et al., 2024, *ApJ*, **976**, L15
- Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *ApJ*, **823**, 102
- Chon S., Hosokawa T., Omukai K., Schneider R., 2024, *MNRAS*, **530**, 2453
- Choustikov N., et al., 2026, *The Open Journal of Astrophysics*, **9**, 58199

- Clarke L., Shapley A. E., Sanders R. L., Topping M. W., Brammer G. B., Bento T., Reddy N. A., Kehoe E., 2024, *ApJ*, **977**, 133
- Clarke L., Shapley A. E., Lam N., Topping M. W., Brammer G. B., Sanders R. L., Reddy N. A., Karthikeyan S., 2025, *arXiv e-prints*, p. [arXiv:2510.06681](#)
- Coe D., et al., 2012, *ApJ*, **757**, 22
- Cullen F., et al., 2025, *MNRAS*, **540**, 2176
- Curti M., Cresci G., Mannucci F., Marconi A., Maiolino R., Esposito S., 2017, *MNRAS*, **465**, 1384
- Curti M., Mannucci F., Cresci G., Maiolino R., 2020, *MNRAS*, **491**, 944
- Curti M., et al., 2024, *A&A*, **684**, A75
- Curtis-Lake E., et al., 2023, *Nature Astronomy*, **7**, 622
- Curtis-Lake E., et al., 2025, *arXiv e-prints*, p. [arXiv:2510.01033](#)
- D'Eugenio F., et al., 2025a, *arXiv e-prints*, p. [arXiv:2510.11626](#)
- D'Eugenio F., et al., 2025b, *ApJS*, **277**, 4
- Dalcanton J. J., 2007, *ApJ*, **658**, 941
- Davé R., Finlator K., Oppenheimer B. D., 2012, *MNRAS*, **421**, 98
- Dorner B., et al., 2016, *A&A*, **592**, A113
- Dors O. L., Maiolino R., Cardaci M. V., Hägele G. F., Krabbe A. C., Pérez-Montero E., Armah M., 2020, *MNRAS*, **496**, 3209
- Dotter A., 2016, *ApJS*, **222**, 8
- Duan Q., et al., 2026, *arXiv e-prints*, p. [arXiv:2605.21599](#)
- Eisenstein D. J., et al., 2023, *arXiv e-prints*, p. [arXiv:2310.12340](#)
- Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *Publ. Astron. Soc. Australia*, **34**, e058
- Ellison S. L., Patton D. R., Simard L., McConnachie A. W., 2008, *ApJ*, **672**, L107
- Falcón-Barroso J., Sánchez-Blázquez P., Vazdekis A., Ricciardelli E., Cardiel N., Cenarro A. J., Gorgas J., Peletier R. F., 2011, *A&A*, **532**, A95
- Ferland G. J., et al., 2013, *Rev. Mex. Astron. Astrofis.*, **49**, 137
- Ferruit P., et al., 2022, *A&A*, **661**, A81
- Flury S. R., Moran E. C., 2020, *MNRAS*, **496**, 2191
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Froese Fischer C., Tachiev G., 2004, *Atomic Data and Nuclear Data Tables*, **87**, 1
- Fujimoto S., et al., 2025a, *arXiv e-prints*, p. [arXiv:2512.11790](#)
- Fujimoto S., et al., 2025b, *ApJ*, **989**, 46
- Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, **362**, 41
- Gardner J. P., et al., 2023, *PASP*, **135**, 068001
- Geris S., et al., 2026, *MNRAS*, **545**, staf1979
- Giménez-Alcázar A., Amorín R., Vilchez J. M., 2026, *arXiv e-prints*, p. [arXiv:2605.05327](#)
- Guo Y., et al., 2016, *ApJ*, **822**, 103
- Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, *ApJ*, **741**, 124
- Harikane Y., et al., 2025, *arXiv e-prints*, p. [arXiv:2505.09186](#)
- Heintz K. E., et al., 2023, *Nature Astronomy*, **7**, 1517
- Heintz K. E., et al., 2025, *A&A*, **693**, A60
- Hirano S., Hosokawa T., Yoshida N., Omukai K., Yorke H. W., 2015, *MNRAS*, **448**, 568
- Hirschmann M., et al., 2023, *MNRAS*, **526**, 3610
- Hsiao T. Y.-Y., et al., 2025, *arXiv e-prints*, p. [arXiv:2505.03873](#)
- Hsiao T. Y.-Y., et al., 2026, *arXiv e-prints*, p. [arXiv:2605.06770](#)
- Illingworth G., et al., 2016, *arXiv e-prints*, p. [arXiv:1606.00841](#)
- Isobe Y., et al., 2022, *ApJ*, **925**, 111
- Isobe Y., Ouchi M., Nakajima K., Harikane Y., Ono Y., Xu Y., Zhang Y., Umeda H., 2023, *ApJ*, **956**, 139
- Isobe Y., et al., 2025, *MNRAS*, **541**, L71
- Isobe Y., et al., 2026, *MNRAS*, **547**, stag123
- Ivey L. R., et al., 2026a, *arXiv e-prints*, p. [arXiv:2604.09177](#)
- Ivey L. R., et al., 2026b, *MNRAS*, **546**, stag094
- Izotov Y. I., Stasińska G., Meynet G., Guseva N. G., Thuan T. X., 2006, *A&A*, **448**, 955
- Izotov Y. I., Thuan T. X., Guseva N. G., 2012, *A&A*, **546**, A122
- Izotov Y. I., Thuan T. X., Guseva N. G., Liss S. E., 2018, *MNRAS*, **473**, 1956
- Jakobsen P., et al., 2022, *A&A*, **661**, A80
- Jeon J., Jeong T. B., Zhang S., Bromm V., 2026, *arXiv e-prints*, p. [arXiv:2604.19075](#)
- Ji X., Belokurov V., Maiolino R., Monty S., Isobe Y., Kravtsov A., McClymont W., Übler H., 2025, *arXiv e-prints*, p. [arXiv:2505.12505](#)
- Ji X., et al., 2026, *arXiv e-prints*, p. [arXiv:2604.03370](#)
- Johnson B. D., Leja J., Conroy C., Speagle J. S., 2019, Prospector: Stellar population inference from spectra and SEDs, Astrophysics Source Code Library, record ascl:1905.025
- Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021, *ApJS*, **254**, 22
- Johnson B. D., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.15954](#)
- Jones G. C., et al., 2024, *A&A*, **683**, A238
- Jones G. C., et al., 2025a, *arXiv e-prints*, p. [arXiv:2509.20455](#)
- Jones G. C., et al., 2025b, *MNRAS*, **536**, 2355
- Juodžbalis I., et al., 2025, *arXiv e-prints*, p. [arXiv:2504.03551](#)
- Juodžbalis I., et al., 2026, *Nature*, **653**, 1017
- Kannan R., et al., 2025, *arXiv e-prints*, p. [arXiv:2502.20437](#)
- Kashino D., Inoue A. K., 2019, *MNRAS*, **486**, 1053
- Kewley L. J., Dopita M. A., 2002, *ApJS*, **142**, 35
- Kewley L. J., Nicholls D. C., Sutherland R. S., 2019, *ARA&A*, **57**, 511
- Kisielius R., Storey P. J., Ferland G. J., Keenan F. P., 2009, *MNRAS*, **397**, 903
- Kobayashi C., Taylor P., 2023, Chemo-Dynamical Evolution of Galaxies. Springer Nature, Netherlands, doi:10.1007/978-981-15-8818-1_106-1
- Kojima T., et al., 2020, *ApJ*, **898**, 142
- Koller M., et al., 2026, *arXiv e-prints*, p. [arXiv:2604.07076](#)
- Korhonen Cuestas N. A., Strom A. L., Miller T. B., Steidel C. C., Trainor R. F., Rudie G. C., Nuñez E. H., 2025, *ApJ*, **984**, 188
- Kriek M., et al., 2006, *ApJ*, **645**, 44
- Kunth D., Östlin G., 2000, *A&ARv*, **10**, 1
- Lam N., Clarke L., Shapley A. E., Sanders R. L., Topping M. W., Brammer G. B., Reddy N. A., Karthikeyan S., 2026, *arXiv e-prints*, p. [arXiv:2605.30513](#)
- Langan I., Ceverino D., Finlator K., 2020, *MNRAS*, **494**, 1988
- Langeroodi D., Hjorth J., 2023, *arXiv e-prints*, p. [arXiv:2307.06336](#)
- Langeroodi D., Hjorth J., 2026, *ApJ*, **997**, L30
- Langeroodi D., et al., 2023, *ApJ*, **957**, 39
- Laseter I. H., et al., 2024, *A&A*, **681**, A70
- Laseter I. H., Maseda M. V., Bunker A. J., Cameron A. J., Curti M., Simmonds C., 2025, *arXiv e-prints*, p. [arXiv:2510.15024](#)
- Leja J., Carnall A. C., Johnson B. D., Conroy C., Speagle J. S., 2019, *ApJ*, **876**, 3
- Lennon D. J., Burke V. M., 1994, *A&AS*, **103**, 273
- Lequeux J., Peimbert M., Rayo J. F., Serrano A., Torres-Peimbert S., 1979, *A&A*, **80**, 155
- Lewis Z., et al., 2025, *arXiv e-prints*, p. [arXiv:2512.03134](#)
- Lilly S. J., Carollo C. M., Pipino A., Renzini A., Peng Y., 2013, *ApJ*, **772**, 119
- Lin X., et al., 2026, *ApJ*, **997**, 364
- Liu B., et al., 2025, *arXiv e-prints*, p. [arXiv:2506.06139](#)
- Luridiana V., Morisset C., Shaw R. A., 2015, *A&A*, **573**, A42
- Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, *MNRAS*, **456**, 2140
- Madau P., Dickinson M., 2014, *ARA&A*, **52**, 415
- Maheson G., et al., 2025, *arXiv e-prints*, p. [arXiv:2504.15346](#)
- Maiolino R., Mannucci F., 2019, *A&ARv*, **27**, 3
- Maiolino R., et al., 2008, *A&A*, **488**, 463
- Maiolino R., et al., 2025, *arXiv e-prints*, p. [arXiv:2505.22567](#)
- Maiolino R., et al., 2026, *arXiv e-prints*, p. [arXiv:2603.20362](#)
- Malmquist K. G., 1922, Meddelanden fran Lunds Astronomiska Observatorium Serie I, **100**, 1
- Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, *MNRAS*, **408**, 2115
- Marconi A., et al., 2024, *A&A*, **689**, A78
- Marino R. A., et al., 2013, *A&A*, **559**, A114
- Markov V., et al., 2025a, *arXiv e-prints*, p. [arXiv:2504.12378](#)
- Markov V., Gallerani S., Ferrara A., Pallottini A., Parlanti E., Mascia F. D., Sommovigo L., Kohandel M., 2025b, *Nature Astronomy*, **9**, 458
- Martinez Z., et al., 2025, *ApJ*, **995**, 204

- Mazzolari G., et al., 2024, *A&A*, **691**, A345
- McClymont W., et al., 2025a, *arXiv e-prints*, p. [arXiv:2503.00106](#)
- McClymont W., et al., 2025b, *arXiv e-prints*, p. [arXiv:2507.08787](#)
- McClymont W., et al., 2025c, *MNRAS*, **540**, 190
- McGaugh S. S., 1991, *ApJ*, **380**, 140
- Moreschini B., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.08939](#)
- Morishita T., et al., 2024, *ApJ*, **971**, 43
- Morishita T., Liu Z., Stiavelli M., Treu T., Bergamini P., Zhang Y., 2025a, *arXiv e-prints*, p. [arXiv:2507.10521](#)
- Morishita T., et al., 2025b, *ApJ*, **983**, 152
- Mowla L., et al., 2024, *Nature*, **636**, 332
- Nagao T., Maiolino R., Marconi A., 2006, *A&A*, **459**, 85
- Nakajima K., Maiolino R., 2022, *MNRAS*, **513**, 5134
- Nakajima K., Ouchi M., 2014, *MNRAS*, **442**, 900
- Nakajima K., et al., 2022, *ApJS*, **262**, 3
- Nakajima K., Ouchi M., Isobe Y., Harikane Y., Zhang Y., Ono Y., Umeda H., Oguri M., 2023, *ApJS*, **269**, 33
- Nakajima K., et al., 2025, *arXiv e-prints*, p. [arXiv:2506.11846](#)
- Nishigaki M., et al., 2025, *arXiv e-prints*, p. [arXiv:2512.12983](#)
- Oesch P. A., et al., 2023, *MNRAS*, **525**, 2864
- Osterbrock D. E., 1989, *Astrophysics of gaseous nebulae and active galactic nuclei*
- Östlin G., et al., 2025, *A&A*, **696**, A57
- Pagel B. E. J., Edmunds M. G., Blackwell D. E., Chun M. S., Smith G., 1979, *MNRAS*, **189**, 95
- Pallottini A., Ferrara A., Gallerani S., Sommovigo L., Carniani S., Vallini L., Kohandel M., Venturi G., 2025, *A&A*, **699**, A6
- Pascalau R. G., et al., 2026, *arXiv e-prints*, p. [arXiv:2603.00232](#)
- Peimbert M., 1967, *ApJ*, **150**, 825
- Pettini M., Pagel B. E. J., 2004, *MNRAS*, **348**, L59
- Planck Collaboration et al., 2020, *A&A*, **641**, A6
- Pollock C. L., et al., 2026, *A&A*, **708**, A203
- Price S. H., et al., 2025, *ApJ*, **982**, 51
- Puskás D., et al., 2025a, *arXiv e-prints*, p. [arXiv:2510.14743](#)
- Puskás D., et al., 2025b, *MNRAS*, **540**, 2146
- Reddy N. A., et al., 2018, *ApJ*, **869**, 92
- Reddy N. A., et al., 2020, *ApJ*, **902**, 123
- Revalski M., et al., 2024, *ApJ*, **966**, 228
- Rinaldi P., et al., 2025, *ApJ*, **994**, 86
- Rinaldi P., et al., 2026, *arXiv e-prints*, p. [arXiv:2604.07138](#)
- Robertson B. E., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.15956](#)
- Rosales-Ortega F. F., et al., 2026, *arXiv e-prints*, p. [arXiv:2604.16273](#)
- Rusta E., et al., 2026, *arXiv e-prints*, p. [arXiv:2603.20363](#)
- Rynkun P., Gaigalas G., Jönsson P., 2019, *A&A*, **623**, A155
- Salpeter E. E., 1955, *ApJ*, **121**, 161
- Sánchez Almeida J., Pérez-Montero E., Morales-Luis A. B., Muñoz-Tuñón C., García-Benito R., Nuza S. E., Kitaura F. S., 2016, *ApJ*, **819**, 110
- Sánchez-Blázquez P., et al., 2006, *MNRAS*, **371**, 703
- Sanders R. L., et al., 2016, *ApJ*, **816**, 23
- Sanders R. L., et al., 2020, *MNRAS*, **491**, 1427
- Sanders R. L., et al., 2021, *ApJ*, **914**, 19
- Sanders R. L., Shapley A. E., Topping M. W., Reddy N. A., Brammer G. B., 2024, *ApJ*, **962**, 24
- Sanders R. L., et al., 2025a, *arXiv e-prints*, p. [arXiv:2508.10099](#)
- Sanders R. L., et al., 2025b, *ApJ*, **989**, 209
- Sarkar A., et al., 2025, *ApJ*, **978**, 136
- Scarlata C., Hayes M., Panagia N., Mehta V., Haardt F., Bagley M., 2024, *arXiv e-prints*, p. [arXiv:2404.09015](#)
- Scholte D., et al., 2025, *MNRAS*, **540**, 1800
- Scholte D., et al., 2026, *arXiv e-prints*, p. [arXiv:2601.02463](#)
- Scholtz J., et al., 2025a, *arXiv e-prints*, p. [arXiv:2510.01034](#)
- Scholtz J., et al., 2025b, *A&A*, **697**, A175
- Shapley A. E., Sanders R. L., Reddy N. A., Topping M. W., Brammer G. B., 2023, *ApJ*, **954**, 157
- Shirazi M., Brinchmann J., 2012, *MNRAS*, **421**, 1043
- Shivaei I., et al., 2025, *arXiv e-prints*, p. [arXiv:2509.01795](#)
- Simmonds C., et al., 2024, *MNRAS*, **535**, 2998
- Simmonds C., et al., 2025, *arXiv e-prints*, p. [arXiv:2508.04410](#)
- Stacy A., Bromm V., Lee A. T., 2016, *MNRAS*, **462**, 1307
- Stanton T. M., et al., 2026, *MNRAS*, **547**, stag449
- Stanway E. R., Eldridge J. J., 2018, *MNRAS*, **479**, 75
- Storey P. J., Hummer D. G., 1995, *MNRAS*, **272**, 41
- Sun F., et al., 2025, *arXiv e-prints*, p. [arXiv:2503.15587](#)
- Tacchella S., et al., 2022, *ApJ*, **927**, 170
- Tacchella S., et al., 2023, *MNRAS*, **522**, 6236
- Tang M., et al., 2026, *arXiv e-prints*, p. [arXiv:2604.03563](#)
- Tayal S. S., Zatsarinny O., 2010, *ApJS*, **188**, 32
- Topping M. W., et al., 2024, *MNRAS*, **529**, 3301
- Topping M. W., et al., 2025, *MNRAS*, **541**, 1707
- Torralba A., et al., 2026, *A&A*, **705**, A147
- Torrey P., et al., 2019, *MNRAS*, **484**, 5587
- Tremonti C. A., et al., 2004, *ApJ*, **613**, 898
- Trussler J. A. A., et al., 2026, *arXiv e-prints*, p. [arXiv:2603.15761](#)
- Übler H., et al., 2026, *arXiv e-prints*, p. [arXiv:2603.20360](#)
- Ucci G., et al., 2023, *MNRAS*, **518**, 3557
- Vanzella E., et al., 2023, *A&A*, **678**, A173
- Vanzella E., et al., 2024, *A&A*, **691**, A251
- Vanzella E., et al., 2025, *arXiv e-prints*, p. [arXiv:2509.07073](#)
- Veilleux S., Osterbrock D. E., 1987, *ApJS*, **63**, 295
- Wang B., et al., 2023, *ApJ*, **944**, L58
- Whitaker K. E., et al., 2019, *ApJS*, **244**, 16
- Williams C. C., et al., 2023, *ApJS*, **268**, 64
- Williams C. C., et al., 2025, *ApJ*, **979**, 140
- Willott C. J., et al., 2025, *ApJ*, **988**, 26
- Witten C., et al., 2025, *MNRAS*, **537**, 112
- Yanagisawa H., et al., 2024, *ApJ*, **974**, 180
- Zaritsky D., Kennicutt Jr. R. C., Huchra J. P., 1994, *ApJ*, **420**, 87
- Zucchi G., Ji X., Madau P., Maiolino R., Juodžbalis I., D'Eugenio F., Geris S., Isobe Y., 2026, *A&A*, **707**, A52
- de Graaff A., et al., 2024, *A&A*, **684**, A87

APPENDIX A: PROPERTIES OF THE STACKED SPECTRA

Section 3.3 constructs the subsamples used for our stacks, whose fundamental properties are listed in Table A1. We present the measurements of the emission line fluxes and the nebular properties of the stacks in Section 3.5, which are summarised in Table A2. Section 4.1 describes strong-line indices of the stacks used for our new metallicity calibrations, which are shown in Table A3. Note that we present the coefficients of our calibrations in the main text (Table 3).

APPENDIX B: SYSTEMATICS OF METALLICITY CALIBRATIONS ONTO MZR

Figure B1 compares $z = 4\text{--}10$ MZR slopes based on our calibrations and the calibrations based on individual auroral-line emitters at $z \sim 0$ (Curti et al. 2020) and high- z (Cataldi et al. 2025a). The slope based on our calibrations ($\beta = 0.34^{+0.14}_{-0.18}$) is slightly steeper than that based on Curti et al. (2020)'s calibrations (0.27) within 1σ but flatter than that based on Cataldi et al. (2025a)'s calibrations (0.48). We confirm that Sanders et al. (2025a)'s calibrations provide a similarly steep MZR slope to Cataldi et al. (2025a)'s calibration. This comparison highlights the impact of choosing different metallicity calibrations on the measurement of the MZR slope (see also Appendix D).

APPENDIX C: PROPERTIES OF EMPG CANDIDATES

In Section 5.1, we select 50 EMPG candidates from our parent sample consisting of the JADES DR4 R100 and R1000 data (Section 2.1),

Table A1. Fundamental properties of our stacks. The value with the errors represents the median value with the 16th–84th percentile range of the stacked sources. The $12 + \log(\text{O}/\text{H})_{\text{SL}}$ values for the Z-bin stacks (the left three columns) are based on Cataldi et al. (2025a)’s calibration (see Section 3.3). We use 1488 unique sources for these stacks.

Property	$12 + \log(\text{O}/\text{H})_{\text{SL}}$			$\log(M_*/M_\odot)$			
	<7.0	7.0–7.3	7.3–7.5	<7.5	7.5–8.0	8.0–8.5	8.5–9.0
# of spectra	31	105	135	133	322	517	482
z	$5.30^{+0.74}_{-2.28}$	$4.15^{+1.83}_{-1.74}$	$3.97^{+1.92}_{-1.11}$	$3.65^{+1.93}_{-1.66}$	$3.61^{+2.38}_{-1.60}$	$3.47^{+2.47}_{-1.51}$	$3.28^{+1.66}_{-1.25}$
$\log(M_*/M_\odot)$	$7.47^{+0.64}_{-0.33}$	$7.68^{+0.51}_{-0.40}$	$7.93^{+0.56}_{-0.44}$	$7.29^{+0.16}_{-0.24}$	$7.80^{+0.14}_{-0.19}$	$8.26^{+0.17}_{-0.16}$	$8.73^{+0.18}_{-0.15}$
$\log(\text{SFR}_{\text{H}\alpha}/M_\odot \text{ yr}^{-1})$	$-0.32^{+0.37}_{-0.36}$	$-0.30^{+0.39}_{-0.33}$	$-0.10^{+0.36}_{-0.42}$	$-0.45^{+0.24}_{-0.27}$	$-0.15^{+0.28}_{-0.40}$	$-0.05^{+0.46}_{-0.51}$	$-0.04^{+0.51}_{-0.44}$
$\log(\text{SFR}_{20}/\text{SFR}_{100})$	$0.27^{+0.13}_{-0.30}$	$0.31^{+0.11}_{-0.28}$	$0.28^{+0.14}_{-0.44}$	$0.33^{+0.10}_{-0.10}$	$0.30^{+0.11}_{-0.18}$	$0.16^{+0.21}_{-0.31}$	$-0.07^{+0.29}_{-0.28}$
Valid λ_{rest} range (Å)	2917–7500	2775–7500	2418–7500	3155–7500	2498–7500	2231–7500	2154–7500

Table A2. Emission line fluxes and nebular properties of our stacks. The fluxes are not corrected for dust extinction but normalised by $\text{H}\beta$. The upper limits are 3σ .

Property	$12 + \log(\text{O}/\text{H})_{\text{SL}}$			$\log(M_*/M_\odot)$			
	<7.0	7.0–7.3	7.3–7.5	<7.5	7.5–8.0	8.0–8.5	8.5–9.0
[O II] $\lambda\lambda 3726, 3729$	<19.9	21.5 ± 2.9	44.0 ± 2.6	33.8 ± 3.2	53.2 ± 1.7	93.9 ± 1.7	175.1 ± 2.3
[Ne III] $\lambda 3869$	25.2 ± 4.1	27.4 ± 2.1	36.3 ± 1.6	34.0 ± 2.2	40.5 ± 1.2	42.6 ± 0.9	45.0 ± 1.2
H δ	15.9 ± 3.4	21.5 ± 2.0	21.5 ± 1.3	20.2 ± 2.0	22.1 ± 1.0	23.8 ± 0.8	19.8 ± 1.0
H γ	32.2 ± 3.0	38.5 ± 1.9	43.0 ± 1.5	43.9 ± 1.8	41.5 ± 1.0	42.5 ± 0.8	41.5 ± 1.1
[O III] $\lambda 4363$	9.2 ± 2.9	10.6 ± 1.5	15.1 ± 1.2	13.5 ± 1.5	10.7 ± 0.7	12.1 ± 0.6	8.5 ± 0.8
He II $\lambda 4686$	<8.2	<3.7	<2.1	<3.6	2.1 ± 0.6	1.5 ± 0.5	<1.9
H β	100.0 ± 3.7	100.0 ± 2.1	100.0 ± 1.4	100.0 ± 1.9	100.0 ± 1.0	100.0 ± 0.9	100.0 ± 1.2
[O III] $\lambda 4959$	90.2 ± 3.7	119.0 ± 2.3	174.0 ± 1.6	140.5 ± 1.9	172.5 ± 1.2	197.3 ± 1.0	193.0 ± 1.2
[O III] $\lambda 5007$	232.1 ± 4.6	364.3 ± 3.4	501.1 ± 2.4	406.1 ± 2.8	518.5 ± 2.1	579.9 ± 2.0	561.3 ± 2.1
H α	255.0 ± 5.3	259.4 ± 2.7	278.0 ± 2.2	267.9 ± 2.5	288.7 ± 1.8	304.9 ± 1.8	317.2 ± 1.9
[N II] $\lambda 6583$	<8.0	<3.7	3.5 ± 0.9	<3.3	5.4 ± 0.6	7.2 ± 0.5	11.4 ± 0.5
[S II] $\lambda 6716$	<7.5	5.8 ± 1.3	6.2 ± 0.9	6.6 ± 1.0	6.1 ± 0.6	8.8 ± 0.5	19.6 ± 0.7
[S II] $\lambda 6731$	<8.9	<4.0	<2.7	3.3 ± 1.1	3.6 ± 0.6	5.9 ± 0.5	12.9 ± 0.6
[Ar III] $\lambda 7135$	<7.5	<4.1	2.9 ± 0.8	<3.1	3.6 ± 0.5	4.6 ± 0.4	6.2 ± 0.6
$E(B - V)$	$0.04^{+0.14}_{-0.04}$	$0.01^{+0.07}_{-0.01}$	$0.03^{+0.03}_{-0.03}$	$0.00^{+0.03}_{-0.00}$	$0.07^{+0.03}_{-0.04}$	$0.11^{+0.02}_{-0.02}$	$0.14^{+0.02}_{-0.02}$
$T_e[\text{O III}] (10^4 \text{ K})$	$2.28^{+0.49}_{-0.52}$	$1.85^{+0.17}_{-0.15}$	$1.91^{+0.08}_{-0.09}$	$2.00^{+0.15}_{-0.13}$	$1.58^{+0.05}_{-0.05}$	$1.61^{+0.04}_{-0.04}$	$1.40^{+0.06}_{-0.06}$
$\log(n_e[\text{S II}]/\text{cm}^{-3})$	...	<0.60	<0.60	$0.60^{+1.67}_{-0.00}$	$0.60^{+1.22}_{-0.00}$	$0.60^{+1.45}_{-0.01}$	$0.60^{+0.64}_{-0.00}$
$12 + \log(\text{O}/\text{H})_{T_e}$	$6.99^{+0.28}_{-0.16}$	$7.40^{+0.08}_{-0.08}$	$7.53^{+0.04}_{-0.04}$	$7.39^{+0.06}_{-0.06}$	$7.73^{+0.04}_{-0.04}$	$7.78^{+0.03}_{-0.02}$	$7.96^{+0.05}_{-0.04}$

Table A3. Strong-line indices of our stacks. The upper and lower limits are 3σ .

Property	$12 + \log(\text{O}/\text{H})_{\text{SL}}$			$\log(M_*/M_\odot)$			
	<7.0	7.0–7.3	7.3–7.5	<7.5	7.5–8.0	8.0–8.5	8.5–9.0
$\log(R3)$	$0.36^{+0.02}_{-0.02}$	$0.56^{+0.01}_{-0.01}$	$0.70^{+0.01}_{-0.01}$	$0.61^{+0.01}_{-0.01}$	$0.71^{+0.00}_{-0.00}$	$0.76^{+0.00}_{-0.00}$	$0.74^{+0.01}_{-0.01}$
$\log(R2)$	< -0.73	$-0.67^{+0.06}_{-0.06}$	$-0.38^{+0.03}_{-0.03}$	$-0.47^{+0.04}_{-0.04}$	$-0.32^{+0.01}_{-0.01}$	$-0.09^{+0.01}_{-0.01}$	$0.16^{+0.01}_{-0.01}$
$\log(R23)$	$0.51^{+0.04}_{-0.02}$	$0.70^{+0.01}_{-0.01}$	$0.85^{+0.01}_{-0.01}$	$0.76^{+0.01}_{-0.01}$	$0.87^{+0.00}_{-0.00}$	$0.93^{+0.00}_{-0.00}$	$0.95^{+0.01}_{-0.01}$
$\log(O32)$	> 1.09	$1.23^{+0.05}_{-0.06}$	$1.07^{+0.02}_{-0.03}$	$1.08^{+0.04}_{-0.04}$	$1.03^{+0.01}_{-0.01}$	$0.85^{+0.01}_{-0.01}$	$0.58^{+0.01}_{-0.01}$
$\hat{R}_{\text{Laseter}}$	< -0.02	$0.18^{+0.03}_{-0.03}$	$0.44^{+0.01}_{-0.01}$	$0.31^{+0.02}_{-0.02}$	$0.48^{+0.01}_{-0.01}$	$0.62^{+0.01}_{-0.01}$	$0.73^{+0.01}_{-0.01}$
$\hat{R}_{\text{Chakraborty}}$	< 0.23	$0.43^{+0.01}_{-0.02}$	$0.62^{+0.01}_{-0.01}$	$0.51^{+0.01}_{-0.01}$	$0.64^{+0.01}_{-0.01}$	$0.73^{+0.00}_{-0.00}$	$0.76^{+0.01}_{-0.01}$
$\hat{R}_{\text{Cataldi}}$	< -0.01	$0.19^{+0.03}_{-0.03}$	$0.44^{+0.01}_{-0.01}$	$0.32^{+0.02}_{-0.02}$	$0.48^{+0.01}_{-0.01}$	$0.62^{+0.01}_{-0.01}$	$0.73^{+0.01}_{-0.01}$
$\log(\text{Ne3})$	$-0.10^{+0.07}_{-0.09}$	$-0.15^{+0.04}_{-0.04}$	$-0.07^{+0.02}_{-0.02}$	$-0.11^{+0.03}_{-0.03}$	$0.00^{+0.02}_{-0.02}$	$0.02^{+0.01}_{-0.01}$	$0.06^{+0.02}_{-0.02}$
$\log(\text{RO2Ne3})$	$0.20^{+0.28}_{-0.14}$	$0.35^{+0.05}_{-0.06}$	$0.56^{+0.03}_{-0.03}$	$0.52^{+0.05}_{-0.05}$	$0.59^{+0.02}_{-0.02}$	$0.69^{+0.01}_{-0.02}$	$0.95^{+0.02}_{-0.02}$
$\widehat{R}_{\text{Ne}}$	< -0.20	$-0.25^{+0.04}_{-0.05}$	$-0.07^{+0.02}_{-0.03}$	$-0.15^{+0.03}_{-0.04}$	$0.03^{+0.02}_{-0.02}$	$0.14^{+0.01}_{-0.01}$	$0.29^{+0.02}_{-0.02}$
$\log(\text{Ne3O2})$	> 0.15	$0.11^{+0.06}_{-0.07}$	$-0.05^{+0.03}_{-0.03}$	$0.01^{+0.05}_{-0.05}$	$-0.05^{+0.02}_{-0.02}$	$-0.23^{+0.01}_{-0.01}$	$-0.45^{+0.01}_{-0.01}$
$\log(\text{Ar3})$	< -1.54	< -1.81	$-1.98^{+0.10}_{-0.13}$	< -1.93	$-1.91^{+0.06}_{-0.07}$	$-1.83^{+0.04}_{-0.04}$	$-1.72^{+0.04}_{-0.04}$
$\log(N2)$	< -1.50	< -1.85	$-1.90^{+0.10}_{-0.13}$	< -1.91	$-1.73^{+0.05}_{-0.05}$	$-1.62^{+0.03}_{-0.03}$	$-1.45^{+0.02}_{-0.02}$
$\log(O3N2)$	> 1.87	> 2.41	$2.60^{+0.10}_{-0.13}$	> 2.52	$2.44^{+0.05}_{-0.05}$	$2.38^{+0.03}_{-0.03}$	$2.19^{+0.02}_{-0.02}$
$\log(S2)$	< -1.20	$-1.65^{+0.28}_{-0.11}$	$-1.66^{+0.20}_{-0.07}$	$-1.43^{+0.06}_{-0.07}$	$-1.48^{+0.04}_{-0.04}$	$-1.34^{+0.02}_{-0.02}$	$-1.01^{+0.01}_{-0.01}$
$\log(O3S2)$	> 1.56	$2.21^{+0.09}_{-1.08}$	$2.36^{+0.06}_{-0.38}$	$2.04^{+0.06}_{-0.07}$	$2.19^{+0.04}_{-0.04}$	$2.09^{+0.02}_{-0.02}$	$1.75^{+0.01}_{-0.01}$

Table C1. Properties of our 50 EMPG candidates. The $12 + \log(\text{O}/\text{H})_{\text{SL}}$ values are based on our calibration. Flag ‘M’ (‘P’): Selected from the R1000 (R100) data. Flag ‘MP’: Selected from both R1000 and R100 data, whose reported spectroscopic measurements are based on the R1000 data. †: EMP LRDs (Section 7.3).

ID	z	$12 + \log(\text{O}/\text{H})_{\text{SL}}$	$E(B - V)$	$\log(M_*/M_\odot)$	$\log(\text{SFR}_{\text{H}\alpha}/M_\odot \text{ yr}^{-1})$	$\log(\text{SFR}_{20}/\text{SFR}_{100})$	Flag
goods-s-ultradeep_127079	5.262	$6.72^{+0.13}_{-0.13}$	0.00 ± 0.29	$7.06^{+0.38}_{-0.26}$	-0.82 ± 0.09	$0.06^{+0.36}_{-0.65}$	P
goods-s-mediumjwst_10011541	7.433	$6.80^{+0.08}_{-0.08}$	...	$7.85^{+0.27}_{-0.14}$	0.11 ± 0.06	$0.24^{+0.31}_{-0.37}$	M
goods-s-mediumjwst_56812	5.760	$6.95^{+0.11}_{-0.11}$	0.00 ± 0.06	$7.09^{+0.21}_{-0.18}$	-0.19 ± 0.08	$0.34^{+0.25}_{-0.27}$	MP
goods-s-deepjwst_30010919	7.610	$6.95^{+0.12}_{-0.12}$	...	$7.24^{+0.31}_{-0.22}$	-0.12 ± 0.09	$0.28^{+0.32}_{-0.51}$	P
goods-s-mediumhst_13561	4.138	$6.99^{+0.14}_{-0.14}$	0.00 ± 0.62	$8.10^{+0.28}_{-0.36}$	0.11 ± 0.11	$0.08^{+0.40}_{-0.41}$	P
goods-s-mediumjwst_32880	5.609	$6.99^{+0.10}_{-0.09}$	0.00 ± 0.35	$7.61^{+0.13}_{-0.14}$	0.04 ± 0.07	$0.41^{+0.18}_{-0.15}$	MP
goods-n-mediumjwst_5088	6.992	$7.01^{+0.07}_{-0.06}$	0.00 ± 0.61	$7.75^{+0.21}_{-0.17}$	0.46 ± 0.04	$0.37^{+0.26}_{-0.29}$	P
goods-s-mediumjwst_20057765	8.913	$7.03^{+0.12}_{-0.11}$	...	$7.73^{+0.17}_{-0.15}$	0.45 ± 0.09	$0.37^{+0.25}_{-0.35}$	P
goods-n-mediumjwst_10001897	5.298	$7.04^{+0.08}_{-0.08}$	0.00 ± 0.30	$8.42^{+0.20}_{-0.33}$	0.05 ± 0.06	$-0.43^{+0.49}_{-0.71}$	MP
goods-s-deephst_10001892	4.772	$7.05^{+0.22}_{-0.18}$	0.00 ± 0.26	$6.72^{+0.37}_{-0.25}$	-1.01 ± 0.14	$0.14^{+0.66}_{-1.07}$	P
goods-s-mediumjwst_59998	5.888	$7.05^{+0.12}_{-0.11}$	0.00 ± 0.70	$7.52^{+0.11}_{-0.09}$	0.07 ± 0.08	$0.42^{+0.17}_{-0.42}$	P
goods-s-mediumjwst_193756	5.905	$7.05^{+0.21}_{-0.17}$	0.69 ± 0.00	$7.69^{+0.29}_{-0.18}$	0.56 ± 0.13	$0.30^{+0.35}_{-0.42}$	MP
goods-s-ultradeep_20198852	8.267	$7.07^{+0.10}_{-0.09}$	0.00 ± 0.28	$7.50^{+0.19}_{-0.15}$	-0.33 ± 0.07	$0.31^{+0.26}_{-0.29}$	MP
goods-s-deepjwst_72574	4.891	$7.07^{+0.11}_{-0.10}$	0.04 ± 0.00	$7.19^{+0.19}_{-0.18}$	-0.46 ± 0.08	$0.21^{+0.23}_{-0.25}$	P
goods-s-mediumjwst_20083087	9.060	$7.08^{+0.15}_{-0.15}$	...	$8.50^{+0.38}_{-0.38}$	0.47 ± 0.12	$-0.22^{+0.64}_{-0.76}$	P
goods-s-mediumjwst_43481	4.423	$7.08^{+0.09}_{-0.09}$	0.00 ± 0.69	$7.39^{+0.33}_{-0.27}$	-0.09 ± 0.06	$0.28^{+0.26}_{-0.37}$	P
goods-s-mediumhst_13639	5.789	$7.08^{+0.18}_{-0.14}$	0.00 ± 0.23	$7.95^{+0.25}_{-0.21}$	0.07 ± 0.11	$0.21^{+0.27}_{-0.39}$	P
goods-s-mediumjwst_186158	4.368	$7.09^{+0.19}_{-0.15}$	0.00 ± 0.16	$7.23^{+0.26}_{-0.24}$	-0.36 ± 0.12	$0.35^{+0.18}_{-0.22}$	P
darkhorse_174741	3.199	$7.10^{+0.17}_{-0.15}$	0.00 ± 0.28	$7.79^{+0.16}_{-0.19}$	-0.32 ± 0.11	$0.31^{+0.16}_{-0.22}$	M
goods-s-mediumjwst_30149608	7.035	$7.10^{+0.17}_{-0.14}$	0.00 ± 0.20	$7.28^{+0.19}_{-0.13}$	-0.03 ± 0.11	$0.39^{+0.22}_{-0.27}$	M
goods-s-ultradeep_100227	4.431	$7.10^{+0.11}_{-0.09}$	0.08 ± 0.00	$7.37^{+0.25}_{-0.14}$	-0.60 ± 0.07	$0.29^{+0.15}_{-0.27}$	P
goods-s-ultradeep_113056	5.984	$7.11^{+0.06}_{-0.05}$	0.23 ± 0.04	$7.55^{+0.21}_{-0.18}$	0.06 ± 0.02	$0.40^{+0.34}_{-0.36}$	MP
darkhorse_168163	3.332	$7.12^{+0.07}_{-0.06}$	0.00 ± 0.11	$7.26^{+0.32}_{-0.19}$	-0.35 ± 0.04	$0.26^{+0.22}_{-0.26}$	P
goods-s-mediumjwst_30142058	7.270	$7.13^{+0.16}_{-0.13}$	...	$7.84^{+0.34}_{-0.22}$	0.31 ± 0.10	$0.46^{+0.51}_{-0.52}$	P
goods-s-mediumjwst_20192042	8.836	$7.15^{+0.15}_{-0.13}$	...	$8.22^{+0.26}_{-0.20}$	0.54 ± 0.10	$0.19^{+0.26}_{-0.26}$	P
oasis_212846	5.049	$7.15^{+0.19}_{-0.14}$	0.16 ± 0.00	$7.07^{+0.35}_{-0.24}$	-0.51 ± 0.03	$0.23^{+0.30}_{-0.42}$	P
goods-s-mediumjwst_10011141	1.244	$7.15^{+0.08}_{-0.07}$	0.09 ± 0.00	$6.87^{+0.39}_{-0.36}$	-0.84 ± 0.04	$0.32^{+0.27}_{-0.33}$	M
darkhorse_345635	5.470	$7.16^{+0.16}_{-0.13}$	0.10 ± 0.00	$7.47^{+0.34}_{-0.29}$	-0.29 ± 0.09	$0.23^{+0.43}_{-0.45}$	M
goods-s-mediumjwst_267649	6.575	$7.16^{+0.17}_{-0.13}$	0.00 ± 0.39	$8.11^{+0.29}_{-0.18}$	0.06 ± 0.10	$0.20^{+0.45}_{-0.48}$	P
goods-s-mediumjwst_135134	7.432	$7.17^{+0.19}_{-0.14}$	...	$8.16^{+0.36}_{-0.34}$	0.04 ± 0.10	$0.37^{+0.69}_{-0.66}$	MP
goods-s-ultradeep_30080593	5.960	$7.17^{+0.09}_{-0.07}$	0.00 ± 0.15	$7.26^{+0.20}_{-0.16}$	-0.29 ± 0.05	$0.33^{+0.18}_{-0.24}$	M
goods-s-ultradeep_201906	5.517	$7.19^{+0.08}_{-0.07}$	0.00 ± 0.02	$7.20^{+0.20}_{-0.15}$	-0.37 ± 0.04	$0.44^{+0.24}_{-0.30}$	P
goods-n-mediumhst_45979	2.583	$7.19^{+0.13}_{-0.10}$	0.04 ± 0.04	$7.29^{+0.29}_{-0.22}$	-0.23 ± 0.07	$0.33^{+0.22}_{-0.29}$	M
goods-s-mediumjwst_51871	5.779	$7.19^{+0.14}_{-0.11}$	0.00 ± 0.24	$7.53^{+0.23}_{-0.21}$	-0.36 ± 0.08	$0.16^{+0.30}_{-0.39}$	M
goods-s-mediumjwst_90155	5.577	$7.19^{+0.10}_{-0.09}$	0.00 ± 0.28	$7.27^{+0.07}_{-0.05}$	0.10 ± 0.06	$0.50^{+0.07}_{-0.08}$	MP
goods-s-mediumjwst_182875	3.425	$7.19^{+0.15}_{-0.12}$	0.00 ± 0.03	$7.48^{+0.11}_{-0.12}$	-0.30 ± 0.08	$0.38^{+0.13}_{-0.11}$	M
goods-s-deepjwst_20050575	8.414	$7.19^{+0.08}_{-0.07}$	...	$7.42^{+0.15}_{-0.11}$	-0.05 ± 0.04	$0.41^{+0.17}_{-0.20}$	P
goods-s-deepjwst_72975	5.587	$7.19^{+0.21}_{-0.17}$	0.13 ± 0.00	$6.66^{+0.17}_{-0.15}$	-0.67 ± 0.12	$0.39^{+0.20}_{-0.72}$	P
goods-s-ultradeep_199524	5.942	$7.19^{+0.15}_{-0.12}$	0.00 ± 0.03	$7.29^{+0.38}_{-0.23}$	-0.76 ± 0.09	$0.15^{+0.34}_{-0.52}$	P
darkhorse_377126	6.055	$7.20^{+0.21}_{-0.17}$	0.00 ± 0.24	$7.05^{+0.16}_{-0.13}$	-0.30 ± 0.10	$0.38^{+0.18}_{-0.21}$	M
goods-s-mediumjwst_73690†	5.497	$7.22^{+0.10}_{-0.08}$	0.50 ± 0.00	$9.08^{+0.14}_{-0.15}$	0.87 ± 0.05	$0.51^{+0.27}_{-0.26}$	P
goods-s-mediumjwst_90864	7.146	$7.23^{+0.10}_{-0.09}$	0.00 ± 0.32	$7.73^{+0.18}_{-0.19}$	0.16 ± 0.05	$0.43^{+0.25}_{-0.22}$	P
goods-n-mediumjwst_59788	5.381	$7.23^{+0.13}_{-0.16}$	0.15 ± 0.00	$7.83^{+0.19}_{-0.17}$	0.00 ± 0.11	$0.31^{+0.27}_{-0.29}$	MP
darkhorse_29724	4.053	$7.24^{+0.13}_{-0.11}$	0.08 ± 0.00	$7.14^{+0.20}_{-0.12}$	-0.43 ± 0.07	$0.39^{+0.12}_{-0.20}$	M
goods-s-ultradeep_109389	5.786	$7.26^{+0.08}_{-0.06}$	0.00 ± 0.08	$7.58^{+0.26}_{-0.22}$	-0.34 ± 0.03	$0.22^{+0.25}_{-0.34}$	MP
goods-s-deephst_10056849	5.814	$7.27^{+0.12}_{-0.11}$	0.00 ± 0.53	$7.58^{+0.12}_{-0.10}$	-0.00 ± 0.06	$0.42^{+0.12}_{-0.13}$	M
goods-n-mediumjwst_39353†	4.846	$7.28^{+0.10}_{-0.09}$	0.55 ± 0.00	$9.53^{+0.09}_{-0.09}$	0.97 ± 0.04	$0.48^{+0.20}_{-0.20}$	P
goods-s-mediumjwst_171875	5.738	$7.28^{+0.11}_{-0.09}$	0.46 ± 0.00	$8.06^{+0.11}_{-0.11}$	0.82 ± 0.05	$0.49^{+0.18}_{-0.17}$	MP
goods-s-deephst_10005217	4.887	$7.28^{+0.04}_{-0.04}$	0.00 ± 0.05	$8.40^{+0.28}_{-0.24}$	0.21 ± 0.01	$0.18^{+0.41}_{-0.41}$	P
goods-s-deephst_10005447	6.621	$7.30^{+0.18}_{-0.13}$	0.00 ± 0.04	$7.18^{+0.13}_{-0.11}$	-0.36 ± 0.08	$0.38^{+0.31}_{-0.21}$	P

the DH R1000 data (Section 2.2), and the OASIS R100 data (Section 2.3). Table C1 summarises properties of our EMPG candidates. Additionally, Section 5.2 presents our compilation of the EMPG candidates from the literature, whose properties are shown in Table C2. We emphasise that all the metallicity values listed in these tables are based on our new metallicity calibrations, except for EXCELS-63107 and Firefly Sparkle with auroral line detections. Note that the photometry of goods-s-ultradeep_127079 is likely contaminated by

a nearby object at lower redshift. However, this contamination does not affect the spectral analysis using NIRSpc.

Figure C1 illustrates the fitting results of the two EMP LRDs around $\text{H}\alpha$ and $[\text{N II}]\lambda\lambda 6548, 6583$ (Section 7.3). We confirm $\Delta\text{BIC} \equiv \text{BIC}_{\text{Narrow}} - \text{BIC}_{\text{Narrow+Broad}} = 6.9-77$ and $S/N_{\text{Broad}} = 6.4-15$, which satisfy the criteria used by Juodžbalis et al. (2025) to statistically support the presence of $\text{H}\alpha$ broadening.

Table C2. Properties of the EMPG candidates from the literature (Section 5.2). ^a: LRD. ^b: Metallicity based on our Ne3 calibration. ^c: T_e -based metallicity. The other metallicities are based on our R3 calibration. ^d: Dynamical mass as an upper limit on M_* . ^e: Derived from the reported Balmer line fluxes. References: 1): Cai et al. (2025), 2): Ivey et al. (2026a), 3): Cai & Yu (2026), 4): Vanzella et al. (2025), 5): Morishita et al. (2025a), 6): Caputi et al. (2026), 7): Vanzella et al. (2024), 8): Nakajima et al. (2025), 9): Vanzella et al. (2023), 10): Maiolino et al. (2025); Juodžbalis et al. (2026), 11): Willott et al. (2025), 12): Übler et al. (2026); Maiolino et al. (2026); Rusta et al. (2026), 13): Hsiao et al. (2025), 14): Chemerynska et al. (2024); Price et al. (2025), 15): Fujimoto et al. (2025b); Fujimoto et al. (2025a), 16): Asada et al. (2026), 17): Koller et al. (2026), 18): Cullen et al. (2025), 19): Mowla et al. (2024)

Name	z	$\log(R3)$	$12 + \log(O/H)$	$\log(M_*/M_\odot)$	$\log(SFR_{H\alpha}/M_\odot \text{ yr}^{-1})$	Ref.
MPG-CR3	3.19	< 0.12	< 6.82	5.79	$-0.04^{+0.05}_{-0.05}$	1)
The Cliff ^a	3.55	$0.20^{+0.07}_{-0.08}$	$6.89^{+0.07}_{-0.08}$	< 8.41 ^d	...	2)
CAPERS-39810	3.65	$0.19^{+0.12}_{-0.18}$	$6.89^{+0.13}_{-0.18}$	$8.02^{+0.22}_{-0.34}$	$0.22^{+0.01}_{-0.01}$	3)
LAP2	4.19	< 0.11	< 6.80	< 4.30	$-2.49^{+0.07}_{-0.09}$	4)
AMORE6-A+B	5.73	< -0.47	< 6.21	$5.60^{+0.20}_{-0.19}$	$-0.63^{+0.07}_{-0.08}$	5)
Pseudo-LRD-NOM	5.96	< -0.42	< 6.26	$8.68^{+0.19}_{-0.72}$	> 0.25	6)
T2c	6.15	$0.35^{+0.08}_{-0.10}$	$7.05^{+0.09}_{-0.11}$	< 4.30	$-1.52^{+0.06}_{-0.07}$	7)
LAP1-B	6.62	$-0.16^{+0.23}_{-0.10}$	$6.52^{+0.15}_{-0.73}$	< 3.43	$-2.23^{+0.05}_{-0.08}$	8)
LAP1	6.64	$-0.26^{+0.10}_{-0.13}$	$6.43^{+0.10}_{-0.13}$	< 4.00	$-1.79^{+0.03}_{-0.04}$	9)
QSO1(central) ^a	7.04	$-0.26^{+0.11}_{-0.15}$	$6.43^{+0.11}_{-0.15}$	< 7.30 ^d	...	10)
QSO1(200pc) ^a	7.04	< -0.39	< 6.30	< 7.30 ^d	...	10)
CANUCS-A370-z8-LAE	8.20	$0.12^{+0.05}_{-0.06}$	$6.81^{+0.05}_{-0.06}$	$7.67^{+0.21}_{-0.27}$	$-0.04^{+0.05}_{-0.05}$	11)
Hebe	10.60	...	< 7.0 ^b	4.30–5.78	$0.31^{+0.07}_{-0.08}$	12)
SAPPHIRES-1449	5.92	$0.29^{+0.10}_{-0.12}$	$6.98^{+0.10}_{-0.13}$	$7.44^{+0.19}_{-0.18}$	$0.78^{+0.09}_{-0.11}$	13)
SAPPHIRES-7695	5.80	$-0.08^{+0.12}_{-0.16}$	$6.61^{+0.12}_{-0.15}$	$7.20^{+0.11}_{-0.12}$	$0.64^{+0.09}_{-0.11}$	13)
SAPPHIRES-9770	6.29	$0.31^{+0.12}_{-0.16}$	$7.01^{+0.15}_{-0.17}$	$7.14^{+0.32}_{-0.37}$	$0.67^{+0.11}_{-0.15}$	13)
SAPPHIRES-20130	6.79	$0.32^{+0.08}_{-0.09}$	$7.02^{+0.08}_{-0.09}$	$7.76^{+0.19}_{-0.18}$	$0.92^{+0.07}_{-0.08}$	13)
SAPPHIRES-21782	5.81	$0.17^{+0.10}_{-0.13}$	$6.86^{+0.10}_{-0.13}$	$6.81^{+0.23}_{-0.20}$	$0.32^{+0.06}_{-nan}$	13)
SAPPHIRES-24804	6.67	$0.11^{+0.11}_{-0.16}$	$6.81^{+0.12}_{-0.16}$	$7.49^{+0.10}_{-0.27}$	$0.64^{+0.02}_{-0.12}$	13)
SAPPHIRES-37141	5.76	$0.28^{+0.10}_{-0.13}$	$6.98^{+0.11}_{-0.14}$	$7.64^{+0.12}_{-0.14}$	$0.95^{+0.09}_{-0.11}$	13)
UNCOVER-18924	7.70	$0.57^{+0.05}_{-0.06}$	$7.32^{+0.08}_{-0.08}$	$5.88^{+0.13}_{-0.08}$	$-0.57^{+0.05}_{-0.05}$	14)
UNCOVER-16155	6.87	$0.55^{+0.03}_{-0.04}$	$7.29^{+0.05}_{-0.06}$	$6.61^{+0.05}_{-0.06}$	$-0.26^{+0.02}_{-0.02}$	14)
UNCOVER-23920	6.00	$0.42^{+0.03}_{-0.05}$	$7.13^{+0.06}_{-0.06}$	$6.30^{+0.03}_{-0.03}$	$-0.05^{+0.02}_{-0.03}$	14)
UNCOVER-12899	6.88	> 0.36	> 7.06	$6.54^{+0.14}_{-0.19}$	$-0.83^{+0.07}_{-0.08}$	14)
UNCOVER-8613	6.38	$0.46^{+0.10}_{-0.12}$	$7.17^{+0.13}_{-0.14}$	$7.12^{+0.07}_{-0.08}$	$-0.34^{+0.05}_{-0.05}$	14)
UNCOVER-23619	6.72	$0.53^{+0.09}_{-0.11}$	$7.27^{+0.10}_{-0.14}$	$6.57^{+0.08}_{-0.06}$	$-0.05^{+0.06}_{-0.07}$	14)
UNCOVER-38335	6.23	> 0.58	> 7.34	$6.83^{+0.25}_{-0.20}$	$0.08^{+0.05}_{-0.06}$	14)
UNCOVER-27335	6.75	$0.48^{+0.13}_{-0.19}$	$7.21^{+0.19}_{-0.21}$	$6.73^{+0.15}_{-0.08}$	$0.00^{+0.08}_{-0.10}$	14)
GLIMPSE-16043	6.20	$0.25^{+0.04}_{-0.04}$	$6.94^{+0.04}_{-0.05}$	5.00	$-0.46^{+0.02}_{-0.02}$	15)
GLIMPSE-7685	6.42	$0.54^{+0.10}_{-0.13}$	$7.27^{+0.07}_{-0.17}$	$6.79^{+0.07}_{-0.07}$	$0.09^{+0.07}_{-0.07}$	16)
GLIMPSE-8139	6.22	> 0.83	> 7.86	$6.80^{+0.08}_{-0.09}$	$0.07^{+0.04}_{-0.04}$	16)
GLIMPSE-9081	6.22	$0.77^{+0.08}_{-0.10}$	$7.86^{+0.00}_{-0.34}$	$7.00^{+0.04}_{-0.04}$	$0.18^{+0.04}_{-0.04}$	16)
GLIMPSE-11789	6.52	$0.42^{+0.08}_{-0.09}$	$7.13^{+0.10}_{-0.10}$	$6.84^{+0.06}_{-0.07}$	$0.36^{+0.08}_{-0.08}$	16)
GLIMPSE-13186	6.16	$0.19^{+0.09}_{-0.11}$	$6.88^{+0.09}_{-0.11}$	$6.40^{+0.18}_{-0.16}$	$0.13^{+0.05}_{-0.05}$	16)
GLIMPSE-17070	6.03	$0.36^{+0.08}_{-0.11}$	$7.06^{+0.10}_{-0.11}$	$6.34^{+0.06}_{-0.06}$	$-0.28^{+0.05}_{-0.05}$	16)
GLIMPSE-19034	6.20	$0.23^{+0.05}_{-0.06}$	$6.92^{+0.05}_{-0.06}$	$6.57^{+0.07}_{-0.06}$	$-0.52^{+0.03}_{-0.03}$	16)
GLIMPSE-35552	6.50	$0.54^{+0.09}_{-0.11}$	$7.29^{+0.15}_{-0.14}$	$6.95^{+0.08}_{-0.05}$	$-0.38^{+0.06}_{-0.09}$	16)
GLIMPSE-35957	6.11	$0.67^{+0.06}_{-0.06}$	$7.51^{+0.17}_{-0.13}$	$6.51^{+0.06}_{-0.06}$	$-0.63^{+0.07}_{-0.07}$	16)
GLIMPSE-37914	6.11	$0.46^{+0.07}_{-0.08}$	$7.18^{+0.09}_{-0.09}$	$6.47^{+0.07}_{-0.07}$	$-0.58^{+0.08}_{-0.08}$	16)
GLIMPSE-45084	6.49	$0.37^{+0.08}_{-0.09}$	$7.07^{+0.09}_{-0.10}$	$6.82^{+0.06}_{-0.06}$	$-0.05^{+0.05}_{-0.05}$	16)
GLIMPSE-45706	6.42	$0.25^{+0.07}_{-0.09}$	$6.94^{+0.07}_{-0.09}$	$6.16^{+0.08}_{-0.04}$	$0.48^{+0.05}_{-0.05}$	16)
GLIMPSE-46431	6.52	$0.35^{+0.13}_{-0.19}$	$7.05^{+0.15}_{-0.19}$	$6.28^{+0.20}_{-0.16}$	$0.42^{+0.07}_{-0.07}$	16)
GLIMPSE-47757	6.00	$0.41^{+0.03}_{-0.03}$	$7.11^{+0.03}_{-0.03}$	$5.62^{+0.06}_{-0.06}$	$-0.60^{+0.05}_{-0.05}$	16)
GLIMPSE-54637	6.07	> 0.35	> 7.04	$6.47^{+0.25}_{-0.17}$	$-1.62^{+0.24}_{-0.24}$	16)
GLIMPSE-55241	5.50	$0.70^{+0.01}_{-0.01}$	$7.58^{+0.02}_{-0.02}$	$6.76^{+0.09}_{-0.08}$	$0.07^{+0.08}_{-0.08}$	16)
SMACS0723_4590-S1	8.45	$0.41^{+0.05}_{-0.08}$	$7.11^{+0.07}_{-0.07}$	$6.77^{+0.05}_{-0.04}$	$-0.49^{+0.05}_{-0.05}$	17)
SMACS0723_4590-S2	8.45	$0.40^{+0.08}_{-0.10}$	$7.11^{+0.08}_{-0.11}$	$7.29^{+0.08}_{-0.04}$	$-0.63^{+0.08}_{-0.08}$	17)
RX2129_11027-S1	9.51	$0.50^{+0.07}_{-0.09}$	$7.23^{+0.11}_{-0.11}$	$6.83^{+0.19}_{-0.10}$	$-0.82^{+0.07}_{-0.07}$	17)
EXCELS-63107	8.27	$0.56^{+0.04}_{-0.04}$	$6.89^{+0.26}_{-0.21}$	$8.57^{+0.32}_{-1.03}$	$1.11^{+0.03}_{-0.03}$	18)
Firefly Sparkle	8.30	$0.71^{+0.03}_{-0.03}$	$6.99^{+0.15}_{-0.33}$	$7.80^{+0.60}_{-0.30}$	$-0.16^{+0.03}_{-0.03}$	19)

APPENDIX D: COMPARISON WITH MZRS FROM THE LITERATURE

Here we discuss the MZRs from the literature. We first note that a strict apples-to-apples comparison across the literature is difficult,

given the heterogeneity in the datasets, methodologies to estimate metallicity and stellar-mass, and parameterisation schemes adopted by different studies. To mitigate the impact of these differences, we limit the references on high- z MZRs to those based on NIRSpc data

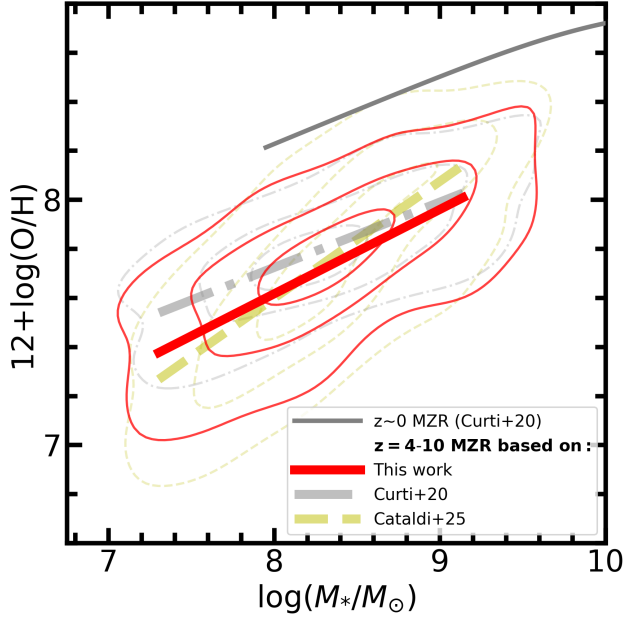


Figure B1. MZR at $z = 4-10$ using our stack-based calibrations (red solid line), those based on high- z individual sources (yellow dashed line; Cataldi et al. 2025a), and those based on $z \sim 0$ galaxies (grey dashdot line; Curti et al. 2020), together with the $z \sim 0$ MZR (Curti et al. 2020).

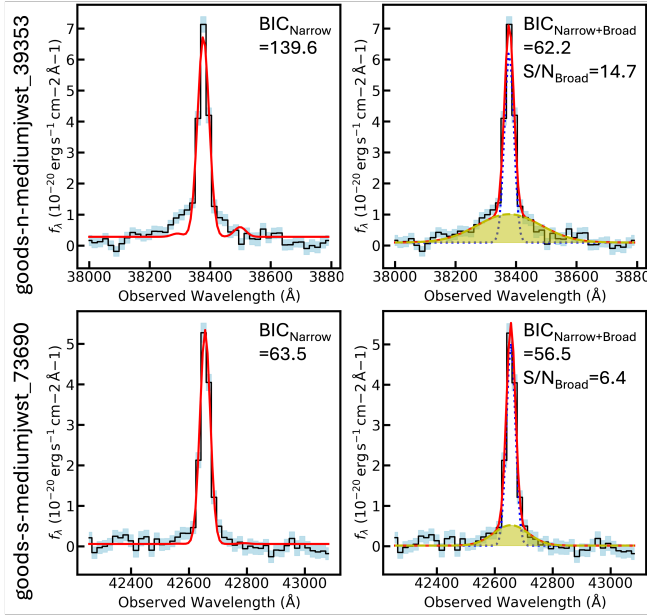


Figure C1. Best-fit results (red curve) of the observed spectra around $H\alpha$ and $[N II] \lambda\lambda 6548, 6583$ (black histogram with the lightblue shade) for the two EMP LRDs (Section 7.3). (Left) Only narrow components. (Right) A broad component added to $H\alpha$ as shown by the yellow dashed curves with the yellow shades. The blue dotted curves are the narrow components of $H\alpha$. We confirm $\Delta BIC \equiv BIC_{Narrow} - BIC_{Narrow+Broad} = 6.9-77$ and $S/N_{Broad} = 6.4-15$, which support the presence of the broad-line components.

only, and compare them separately for each method of metallicity measurements.

First, we compile the literature whose $12 + \log(O/H)_{SL}$ are mostly derived in a self-consistent manner. We refer to the MZR based on the $z \sim 0$ calibrations: Curti et al. (2024) based on Curti et al. (2020) and Laseter et al. (2024)'s calibrations, Lewis et al. (2025) and Nakajima et al. (2023) based on Nakajima et al. (2023)'s calibrations, and Sarkar et al. (2025) based on Curti et al. (2020)'s calibrations. We also cite the MZR mainly based on the high- z calibrations or Scholte et al. (2025), whose $\hat{R}_{Laseter}$ calibration quite agrees with the high- z calibrations (see Figure 4): Pollock et al. (2026) based on Sanders et al. (2024) and Laseter et al. (2024)'s calibrations (see also Heintz et al. 2025), Stanton et al. (2026) based on Scholte et al. (2025) and Sanders et al. (2024) calibrations, Heintz et al. (2023) based on Sanders et al. (2024)'s calibrations, and Lam et al. (2026) based on Sanders et al. (2025a)'s calibrations.

The left panel of Figure D1 compares the MZR based on the $z \sim 0$ calibrations and the high- z calibrations. We find that the MZR based on the $z \sim 0$ calibrations generally exhibit shallower slopes than those based on the high- z calibrations. This is consistent with the fact that the $z \sim 0$ calibrations of Curti et al. (2020) provide a shallower MZR slope than the high- z calibrations of Cataldi et al. (2025a) for the same dataset as reported in Appendix B. Among these MZR, Pollock et al. (2026)'s $z = 3-10$ MZR based on Sanders et al. (2024) and Laseter et al. (2024)'s calibrations agrees best with our MZR.

The right panel of Figure D1 shows the literature MZR whose metallicities are based on the direct- T_e method. Giménez-Alcázar et al. (2026) report that the stacks of auroral-line emitters (cyan dashed line) exhibit systematically lower metallicities than those without auroral-line detections (magenta dashdot line), suggesting a bias of auroral-line emitters to lower metallicities at a given M_* . Indeed, the MZR of Morishita et al. (2024) and Chakraborty et al. (2025) based on individual auroral-line emitters are generally below our MZR based on our strong-line calibrations. A similar trend has been observed at $z \sim 0$ (Curti et al. 2020), where it is interpreted as a selection effect arising from the preferential detection of $[O III] \lambda 4363$ in lower-metallicity galaxies at a given M_* . However, the MZR reported by Hsiao et al. (2026) based on new observations of the GLIMPSE-D programme towards the lensing field shows marginally higher metallicities than the other T_e -based MZR. This may indicate that, at the low- M_* frontier where the available samples are still small, the derived MZR is somewhat sensitive to sample selection. Further deep observations will be essential to establish the behaviour of the MZR in this regime more robustly. It is also worth noting that our MZR is comparable to the stack measurement of Nishigaki et al. (2025) at $z = 6.1$.

In Figure D2, we plot the predicted high- z MZR of cosmological simulations compiled by Curti et al. (2024): FIRE (Ma et al. 2016), FirstLight (Langan et al. 2020), the chemo-dynamical simulations of Kobayashi & Taylor (2023), Asteraeus (Ucci et al. 2023), IllustrisTNG (Torrey et al. 2019), and SERRA (Pallottini et al. 2025). Note that the FIRE MZR are rescaled by Curti et al. (2024) to match the $z \sim 0$ MZR. We add recent radiation-hydrodynamic simulations of THESAN-ZOOM (McClymont et al. 2025b) and the Bursty Star Formation model of MEGATRON (Choustikov et al. 2026).

Our MZR agrees with most of these simulated MZR. In particular, the MZR slopes of THESAN-ZOOM and FIRE are quite comparable to ours. Our MZR also matches the Bursty Star Formation model of MEGATRON, which predicts systematically lower metallicities than the Efficient Star Formation model. This model favours stochastic SFHs at the low- M_* end of the observations, in line with the large

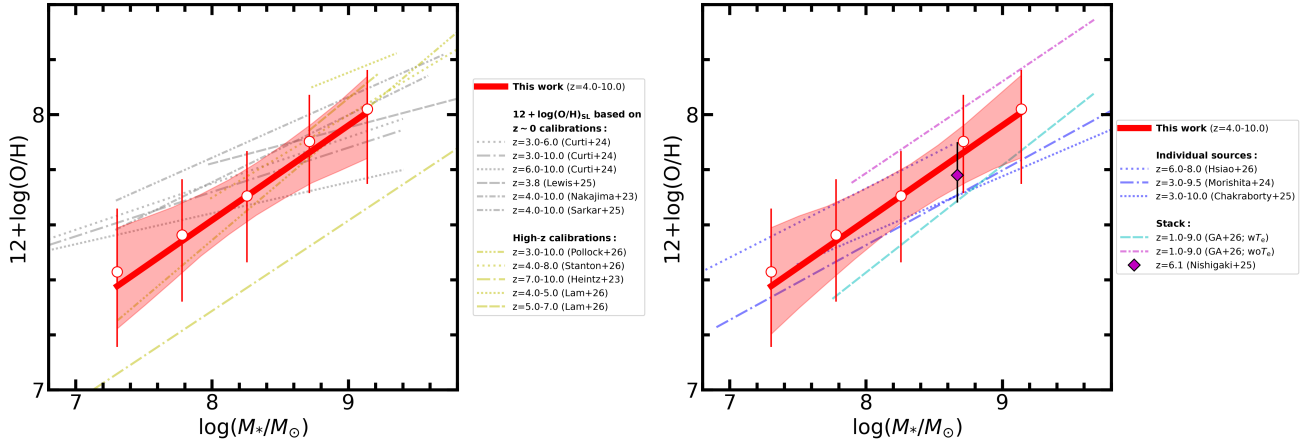


Figure D1. MZR of our JADES+DH sources at $z=4-10$ (red solid line) compared with the literature. (Left) The MZR whose strong-line metallicities $12 + \log(\text{O}/\text{H})_{\text{SL}}$ are based on the $z \sim 0$ calibrations (grey lines; Curti et al. 2024; Lewis et al. 2025; Nakajima et al. 2023; Sarkar et al. 2025) and the high- z calibrations (yellow lines; Pollock et al. 2026; Stanton et al. 2026; Heintz et al. 2023; Lam et al. 2026, see Section D for more details). The MZR based on the $z \sim 0$ calibrations are generally shallower than those based on the high- z calibrations. (Right) The T_e -based MZR of individual auroral-line emitters (blue lines; Hsiao et al. 2026; Morishita et al. 2024; Chakraborty et al. 2025) and stacks (Giménez-Alcázar et al. 2026; Nishigaki et al. 2025). The MZR of the individual auroral-line emitters tend to show lower metallicities than our MZR.

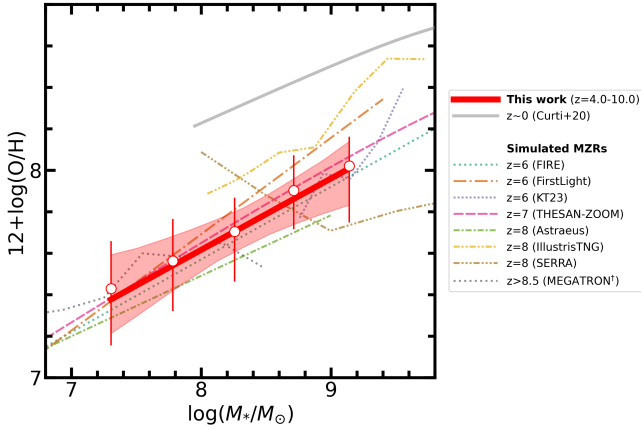


Figure D2. Same as Figure D1 but with the simulated MZR from the literature: FIRE (Ma et al. 2016), FirstLight (Langan et al. 2020), Kobayashi & Taylor (2023; KT23), THESAN-ZOOM (McClymont et al. 2025b), Astraeus (Ucci et al. 2023), IllustrisTNG (Torrey et al. 2019), SERRA (Pallottini et al. 2025), and MEGATRON (Choustikov et al. 2026). [†]: Bursty Star Formation (SF) model.

metallicity scatter observed similarly at the low- M_* end of the MZR (Section 7.2).

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.