

Rapid, out-of-equilibrium metal enrichment indicated by a flat mass-metallicity relation at $z \sim 6$ from NIRCам grism spectroscopy

Gauri Kotiwale^{1,*}, Jorryt Matthee¹, Daichi Kashino², Aswin P. Vijayan³, Alberto Torralba¹,
Claudia Di Cesare¹, Edoardo Iani¹, Rongmon Bordoloi⁴, Joel Leja^{5,6}, Michael V. Maseda⁹,
Sandro Tacchella^{7,8}, Irene Shivaie¹⁰, Kasper E. Heintz^{11,12,13}, A. Lola Danhaive^{7,8}, Sara Mascia¹,
Ivan Kramarenko¹, Benjamín Navarrete¹, Ruari Mackenzie¹⁴, Rohan P. Naidu¹⁵, and David Sobral¹⁶

¹ Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria

² National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

³ Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK

⁴ Department of Physics, North Carolina State University, Raleigh 27695, North Carolina, USA

⁵ Institute for Computational & Data Sciences, The Pennsylvania State University, University Park, PA 16802, USA

⁶ Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

⁷ Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

⁸ Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK

⁹ Department of Astronomy, University of Wisconsin-Madison, 475 N. Charter St., Madison, WI 53706, USA

¹⁰ Centro de Astrobiología (CAB), CSIC-INTA, Carretera de Ajalvir km 4, Torrejón de Ardoz E-28850, Madrid, Spain

¹¹ Cosmic Dawn Center (DAWN), Denmark

¹² Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen N, Denmark

¹³ Department of Astronomy, University of Geneva, Chemin Pegasi 51, 1290 Versoix, Switzerland

¹⁴ Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland

¹⁵ MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

¹⁶ BNP Paribas Corporate & Institutional Banking, Torre Ocidente Rua Galileu Galilei, 1500-392 Lisbon, Portugal

Received 25 July 2025 / Accepted 25 November 2025

ABSTRACT

We aim to characterise the mass-metallicity relation (MZR) and the 3D correlation between the stellar mass, metallicity, and star formation rate (SFR) known as the fundamental metallicity relation (FMR) for galaxies at $5 < z < 7$. Using ~ 800 [O III] selected galaxies from deep NIRCам grism surveys, we present our stacked measurements of direct- T_e metallicities, which we used to test recent strong-line metallicity calibrations. Our measured direct- T_e metallicities ($0.1\text{--}0.2Z_\odot$ for $M_\star \approx 5 \times 10^{7\text{--}9} M_\odot$, respectively) match recent JWST/NIRSpec-based results. However, there are significant inconsistencies between observations and hydrodynamical simulations. We observe a flatter MZR slope than the SPHINX²⁰ and FLARES simulations, which cannot be attributed to selection effects. With simple models, we show that the effect of an [O III] flux-limited sample on the observed shape of the MZR is strongly dependent on the FMR. If the FMR is similar to the one in the local Universe, the intrinsic high-redshift MZR should be even flatter than is observed. In turn, a 3D relation where SFR correlates positively with metallicity at fixed mass would imply an intrinsically steeper MZR. Our measurements indicate that metallicity variations at fixed mass show little dependence on the SFR, suggesting a flat intrinsic MZR. This could indicate that the low-mass galaxies at these redshifts are out of equilibrium and that metal enrichment occurs rapidly in low-mass galaxies. However, being limited by our stacking analysis, we are yet to probe the scatter in the MZR and its dependence on SFR. Large carefully selected samples of galaxies with robust metallicity measurements can put tight constraints on the high-redshift FMR and help us to understand the interplay between gas flows, star formation, and feedback in early galaxies.

Key words. galaxies: abundances – galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: ISM

1. Introduction

The gas-phase metallicity of the interstellar medium of galaxies reflects the interplay of gas inflows, feedback, and star formation (Finlator & Davé 2008; Lilly et al. 2013; Maiolino & Mannucci 2019; Curti 2025). Usually, the gas-phase metallicity is expressed in terms of the oxygen abundance (relative to the hydrogen abundance) as it is the most abundant metal in the Universe. It is most easily observed through strong lines in star-forming regions of galaxies, such as [O II] $\lambda\lambda 3726, 3729$ and

[O III] $\lambda\lambda 4960, 5008$. Studying the evolution of the gas-phase metallicities of galaxies through cosmic time can tell us about the baryonic mass assembly in the Universe.

The relation between gas-phase metallicity (Z_{gas}) and stellar mass (M_\star), known as the mass-metallicity relation (MZR), has been extensively studied, spanning from galaxies in the local Universe (at $z \sim 0$, e.g. Tremonti et al. 2004; Ellison et al. 2008; Moustakas et al. 2011; Andrews & Martini 2013; Curti et al. 2020; Scholte et al. 2024), to those at the peak of cosmic star formation activity (up to $z \sim 3$; e.g. Shapley et al. 2005; Maiolino et al. 2008; Mannucci et al. 2009; Wuyts et al. 2014; Topping et al. 2021; Sanders et al. 2021), and well into the

* Corresponding author: gauri.kotiwale@ista.ac.at

epoch of reionisation using JWST (e.g. Heintz et al. 2023; Matthee et al. 2023; Nakajima et al. 2023; Chemerynska et al. 2024; Laseter et al. 2024; Curti et al. 2024; Sanders et al. 2024; Sarkar et al. 2025; Chakraborty et al. 2025; Morishita et al. 2024; Cullen et al. 2025; Li et al. 2025; Pollock et al. 2025). It has been observed that the gas-phase metallicity of galaxies increases with increasing stellar mass, and for $z < 4$ the slope of this trend becomes shallower at higher masses. This trend is explained through gas regulation in galaxies (Finlator & Davé 2008; Peeples & Shankar 2011; Davé et al. 2012; Lilly et al. 2013). Due to higher stellar masses, the galaxies have a deeper potential well, making the outflow of metal-enriched gas from the galaxy difficult. The metallicity also decreases at fixed stellar mass with increasing redshift. The redshift evolution of the MZR could be due to the inflow of metal-poor gas, or the outflow of metal-rich gas due to stronger feedback mechanisms at higher redshifts (Lara-López et al. 2010; Bothwell et al. 2013, 2016; Scholte et al. 2024). This evolution has also been studied in terms of the 3D relation between the gas-phase metallicity, stellar mass, and star formation rate (Mannucci et al. 2010; Lara-López et al. 2010; Andrews & Martini 2013; Curti et al. 2020, SFR). The scatter in the MZR in the local Universe can be reduced by accounting for variations in the SFRs of galaxies. Galaxies in the local Universe are found to reside on a 2D plane in the M_{\star} - Z_{gas} -SFR space with < 0.1 dex scatter. This relation is known as the fundamental metallicity relation (FMR). Multiple studies have found that the FMR remains invariant up to $z \sim 3$ (Cresci et al. 2012; Maier et al. 2015; Yabe et al. 2015; Topping et al. 2021; Sanders et al. 2021). Nakajima et al. (2023) and Sarkar et al. (2025) find no redshift evolution for their sample between $4 < z < 8$ compared to the $z = 0$ FMR, but observe significant deviations at $z > 8$. Various theoretical models and simulations also explore the FMR in terms of the SFR on different spatial (Wang & Lilly 2021) and temporal scales (Torrey et al. 2018; Ma et al. 2024), the impact of different feedback mechanisms (De Rossi et al. 2017), and variations in gas flows and the gas fraction (Dayal et al. 2013; Torrey et al. 2019).

There are several arguments about whether the FMR exists out to high redshifts (e.g. Zahid et al. 2014; Heintz et al. 2023; Curti et al. 2024; Scholte et al. 2025; Cullen et al. 2025). If truly fundamental, the shape of the FMR plane does not evolve – only the relative position of the galaxy population on the plane shifts. Simulations suggest that the existence of the plane depends on the timescales of SFR variability (e.g. Torrey et al. 2018; Ma et al. 2024). Given that the star formation histories (SFHs) of high- z galaxies appear more bursty (e.g. Topping et al. 2022; Endsley et al. 2023; Looser et al. 2025), it could cause the FMR to break down at high- z . Indeed, some JWST studies indicate that the normalisation of the plane shifts, i.e. galaxies systematically have lower metallicities than is predicted by the $z = 0$ FMR (Heintz et al. 2023; Curti et al. 2024; Scholte et al. 2025).

Gas-phase metallicity is typically inferred using two approaches: (1) the direct- T_e method, and (2) strong-line calibrations. The direct- T_e method is based on electron temperature (T_e) determination using auroral emission lines. The most common auroral line used to determine the electron temperature of the O^{+2} regions is [O III] λ 4364 (Maiolino & Mannucci 2019; Curti 2025). The relative strength of the [O III] λ 4364 and [O III] λ 5008 lines can be used to calculate the electron temperature. The calculated electron temperature is then used to estimate the gas-phase metallicity using dust-corrected emission line ratios of [O III] and [O II] with $H\beta$. The challenge with this method is that the [O III] λ 4364 auroral line is very faint, and

thus hard to detect in most galaxies (Osterbrock & Ferland 2006; Kewley et al. 2019).

On the other hand, the strong line method is applicable to a larger dataset as it only requires the detection of emission lines that are easily observable in the star-forming regions of the galaxies. The strong line metallicity calibrations are usually calibrated to metallicity measurements from the direct- T_e method. Multiple studies have constructed metallicity calibration curves using the observed relationship between emission line ratios and the direct- T_e method metallicity estimates for galaxies (e.g. Maiolino et al. 2008; Marino et al. 2013; Andrews & Martini 2013; Bian et al. 2018; Curti et al. 2020; Nakajima et al. 2022; Laseter et al. 2024; Chakraborty et al. 2025; Sanders et al. 2024; Scholte et al. 2025). Some of the calibration curves were developed using galaxies that are considered to be high-redshift analogues (Bian et al. 2018; Nakajima et al. 2022). These calibrations offer a potential solution to estimate metallicity at high- z where getting direct- T_e metallicities is difficult; however, their applicability at these redshifts needs to be tested.

Recent studies using JWST make use of strong-line calibrations to explore the MZR at a redshift range of $z = 3$ – 10 (Matthee et al. 2023; Nakajima et al. 2023; Li et al. 2023; Heintz et al. 2023; Curti et al. 2024; Sarkar et al. 2025; Chemerynska et al. 2024; Rowland et al. 2025; Li et al. 2025; Cataldi et al. 2025; Pollock et al. 2025). Studies by Nakajima et al. (2023), Sanders et al. (2024), Laseter et al. (2024), Sarkar et al. (2025), Chakraborty et al. (2025), and Pollock et al. (2025) use the [O III] λ 4364 auroral line (direct- T_e method) to investigate the MZR up to $z \sim 10$. Most of these studies mainly focus on galaxies with $\log(M_{\star}/M_{\odot}) > 8$. Analysis by Curti et al. (2024) and Chemerynska et al. (2024) explores the MZR down to $\log(M_{\star}/M_{\odot}) \sim 6$ at high redshifts. Considering JWST's recent arrival on the scene, the sample sizes for which we have direct- T_e measurements are still limited.

In this work, we study and characterise the MZR and the 3D correlation with SFR between $5 < z < 7$ through direct- T_e metallicities using data from deep NIRCcam grism surveys EIGER (Kashino et al. 2023), COLA1 (Torralba-Torregrosa et al. 2024) and ALT (Naidu et al. 2024). Our sample makes use of NIRCcam spectra, compared to most of the studies in the literature, which use JWST/NIRSpec. Grism spectra have a simple flux-limited selection function and are not subject to slit losses. On the other hand, they are less sensitive and have a smaller wavelength coverage than NIRSpec spectra. Our galaxy sample was selected by using only the [O III] λ 4960, 5008 flux in all three surveys, giving us a well defined, flux-limited selection function. We cover the $H\beta$ and [O III] λ 4960, 5008 emission lines for all our galaxies and, for a subset, we also cover $H\gamma$ and [O III] λ 4364. For this subset we measured the [O III] λ 4364 auroral line with a signal-to-noise ratio (S/N) of > 3 in stacks of stellar mass bins, which we used to calculate the direct- T_e metallicities. We used these measurements to test the strong-line calibrations high- z galaxies, obtain a scaling relation between the gas-phase metallicity and stellar mass, and investigate the impact of our selection function.

The paper is structured as follows. We describe our NIRCcam grism dataset and data reduction pipeline in Sect. 2. We report the method of extracting emission line fluxes and dust correction in Sect. 3. In Sect. 4 we describe deriving metallicities using both the direct- T_e method and strong-line calibrations, and in Sect. 5 we present our MZR and make comparisons to the literature. Furthermore, we discuss the impact of our selection function in Sect. 6 through comparisons to simulations and toy models. Finally, we present our conclusions and summary in Sect. 7.

Throughout this work, we adopt a flat Λ CDM cosmology with $\Omega_{\Lambda} = 0.69$, $\Omega_M = 0.31$, and $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration VI 2020). All magnitudes are mentioned in the AB system (Oke & Gunn 1983).

2. Data

2.1. Observations

We used a compilation of imaging and Wide Field Slitless Spectroscopy (WFSS) from deep NIRCcam grism surveys. The data were obtained from the EIGER survey (programme ID 1243, PI Lilly, Kashino et al. 2023), the ALT survey (programme ID 3516, PIs Matthee and Naidu, Naidu et al. 2024), and the JWST programme ID 1933 (PIs Matthee and Naidu, hereafter ‘COLA1 survey’). The EIGER survey is a deep grism cycle 1 survey in the F356W filter centred on six widely separated $z > 6$ quasar fields. It contains imaging in the short-wavelength (SW) filters F115W and F200W, and direct imaging in the F356W long-wavelength (LW) filter. Each field has a survey area of $6.5 \times 3.4 \text{ arcmin}^2$ (i.e. 132 arcmin^2 in total).

The WFSS in the F356W LW filter covers an observed wavelength of $3.1\text{--}4.0 \mu\text{m}$ with a spectral resolution of $R \sim 1600$ for a point source. The COLA1 survey has a similar observing strategy to EIGER, but is centred on the bright double-peaked Ly α emitter, COLA1, at $z = 6.6$ (Matthee et al. 2018; Torralba-Torregrosa et al. 2024). The total area in the COLA1 field is 21 arcmin^2 . Finally, the ALT survey is the deepest NIRCcam WFSS survey to date, targeting a wider 30 arcmin^2 area around the gravitational lensing cluster Abell 2744 in the F356W filter. Thanks to deep imaging from UNCOVER (programme ID 2561, PI Labbe, Bezanson et al. 2024) and MegaScience (programme ID 4111, PI Suess, Suess et al. 2024), there is excellent photometric coverage, allowing for a detailed characterisation of the SEDs for galaxies from the ALT survey. From ALT, we only selected galaxies with magnification factor (μ) < 3 to remove multi-imaged sources and strong arcs.

2.2. Data reduction

In this work, we focus on [O III] $\lambda\lambda 4960, 5008$ selected galaxies that are observed in the NIRCcam WFSS with the F356W filter ($\lambda \sim 3.1\text{--}4.0 \mu\text{m}$) in the redshift range $5.5 < z < 7$. The analysis in our paper makes use of both JWST/NIRCcam imaging and WFSS. The details of the basic reduction steps are described in Kashino et al. (2023, 2025) for the EIGER survey, Torralba-Torregrosa et al. (2024) for the COLA1 survey, and Naidu et al. (2024) for the ALT survey. Here we briefly summarise the WFSS data reduction steps for these surveys and additional modifications made for emission-line spectra extraction.

The WFSS data were reduced using a combination of the JWST pipeline and customised Python-based processing. We downloaded the RATE files from the MAST database¹ and used the JWST pipeline to flat field images. For the images in the COLA1 and EIGER fields, we subtracted the background by calculating the median value along each column (which is orthogonal to the dispersion direction). For the ALT data in the Abell 2744 cluster, the median statistic was boosted by diffuse light from the vast number of sources (in particular cluster galaxies). We therefore constructed (for each module) a master-background template based on the backgrounds in the COLA1 and EIGER data and scaled the mode of each ALT grism image to the mode value of the background template. We aligned the

astrometry of each image by matching the WCS of the image taken simultaneously in the SW filter to the reference catalogue based on the direct imaging in the F356W filter. In our basic reduction used to identify the sources, we separated the dispersed light in a ‘continuum’ and an ‘emission-line’ image, by estimating the continuum in each row with a running median filter along the dispersion direction. The filter is a kernel with width 51 pixels and a central hole of 9 pixels, to avoid over-subtraction of emission-lines. As is detailed in Kashino et al. (2023), this procedure was run twice, masking identified emission-lines from the first iteration.

In this study, we present an improved method for the extraction of emission-line (EMLINE) spectra to avoid over-subtraction around emission lines. This was performed already in Matthee et al. (2024), but here we generalise it for [O III] emitters. The standard methodology can lead to slight over-subtraction of faint wings that are not identified in individual images. Building upon the aforementioned works for the EIGER, COLA1, and ALT surveys, we know the location of the emission lines in our spectra, which allows us to subtract the continuum optimally without over-subtracting line emission. We used this information and stacked the EMLINE spectra of galaxies with a [O III] $\lambda 5008$ signal-to-noise ratio (S/N) > 25 . The methodology for stacking is described further in Sect. 3.3. Once we had the stacked EMLINE spectra, we generated 2D Gaussian masks, with a full width at half maximum of 7 pixels and size of 15 pixels with a threshold of $1 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^2$, for the [O III] $\lambda\lambda 4960, 5008$ emission lines using the Photutils Python package (Bradley et al. 2024). The mask for the [O III] $\lambda 5008$ line was applied to its respective location, while the mask for the [O III] $\lambda 4960$ line was applied to the locations of [O III] $\lambda 4960$, H γ , [O III] $\lambda 4364$, and H β in each 2D science (SCI) spectrum in our sample. We applied the mask for [O III] $\lambda 4960$ to the other emission lines (H γ , [O III] $\lambda 4364$, and H β) due to their smaller flux compared to [O III] $\lambda 5008$. Utilising the [O III] $\lambda 5008$ mask for these lines could result in over-masking, which can cause under-subtraction of the continuum and can cause the potential overlap with adjacent emission lines, especially for the H γ and [O III] $\lambda 4364$ emission lines. After applying the masks to the SCI spectra, we performed the continuum subtraction on the SCI spectra using a running box-car median with a kernel of 71 pixels and a gap of 7 pixels in the centre, similar to the method in the initial extraction of emission-line spectra. The parameters were iteratively chosen based on stacked spectra, whereby we optimised to match the line shapes and the 1:2.98 normalisation of the [O III] $\lambda\lambda 4960, 5008$ doublet.

3. Methods

3.1. Galaxy sample

Our dataset comprises [O III] $\lambda\lambda 4960, 5008$ selected galaxies from the EIGER (Kashino et al. 2025), COLA1 (Torralba-Torregrosa et al. 2024), and ALT (Naidu et al. 2024) surveys. These galaxies were selected using the ‘forward’ method described in Kashino et al. (2023). In this method, the [O III] doublet sample are required to have S/N > 3 and a maximum offset of 3 pixels ($0.09''$) in their spatial centroid. Once we had our emission line spectra, the galaxies were inspected to detect any broad line H β emission. We did this to find possible broad-line active galactic nucleus (AGN) candidates and remove them from our sample. We excluded the spectra for the central quasars in the EIGER fields. We also removed AGNs that are known from previous studies, such as the X-ray

¹ <https://mast.stsci.edu/>

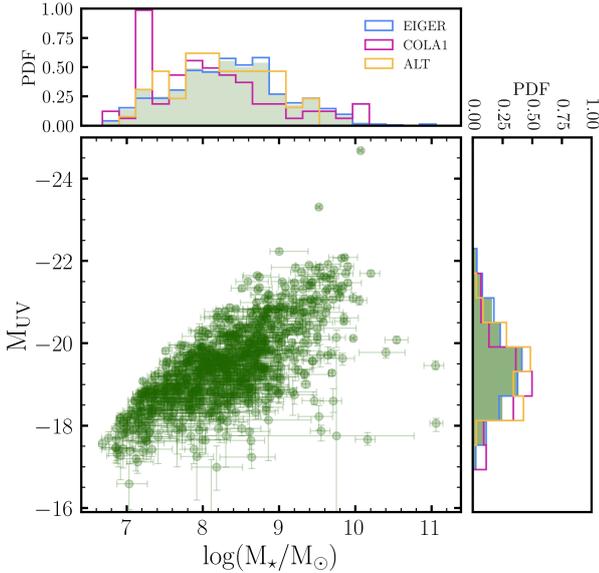


Fig. 1. Relation between UV magnitude and stellar mass for our $H\beta$ sample (792 galaxies). The M_{UV} and stellar masses were obtained through SED fitting using *Prospector*. With data from the ABELL2744 lensing field, we probed galaxies with M_{UV} down to -16.6 mag. The UV luminosity increases with stellar mass, as was expected.

detected AGN J1148+5253 (Mahabal et al. 2005) at $z = 5.69$. We also removed the galaxy J0100–9148 from our sample because of severe contamination in the 2D spectra. We did not find additional broad $H\beta$ emitters in our sample; however, it should be noted that broad-line $H\alpha$ AGNs at high redshifts usually have extremely faint broad $H\beta$ emission (Brooks et al. 2025). Through this, we obtained a dataset of 792 galaxies, of which 659 galaxies are from the EIGER survey, 74 are from the COLA1 survey, and 59 are from the ALT survey. We classified our data into two samples:

- (i) The $H\beta$ sample: galaxies for which our data cover the $H\beta$ emission line at a rest-frame wavelength of 4862.68 \AA , i.e. galaxies with a redshift of $5.5 \leq z < 7$. This is the entire dataset with 792 galaxies.
- (ii) The $H\gamma$ sample: galaxies for which our data cover the $H\gamma$ emission line at a rest-frame wavelength of 4340 \AA , i.e. galaxies with a redshift of $6.3 \leq z < 7$. This is a subset of the $H\beta$ sample, with 270 galaxies.

The typical UV magnitudes of the $H\gamma$ and the $H\beta$ sample are $M_{UV} = -19.5$ mag and -19.4 mag, respectively (see Fig. 1). Stellar masses in our total dataset range from $\log(M_*/M_\odot) = 6.9$ – 11.1 , with typical stellar masses of both samples being $\sim 10^{8.3} M_\odot$. Our sample is stellar mass complete down to $\log(M_*/M_\odot) \sim 8.5$ (see Fig. 1). Properties such as stellar mass, emission line fluxes, and photometry for the galaxies from the ALT catalogue were corrected for magnification using magnification factors from Furtak et al. (2023) and Price et al. (2025). The redshift uncertainty for our sample is ~ 60 – 70 km s^{-1} (Torralba-Torregrosa et al. 2024; Naidu et al. 2024; Kashino et al. 2025).

3.2. SED modelling

In order to estimate galaxy properties such as the stellar mass, we performed SED fitting using the code *Prospector*

(Johnson et al. 2021). Nebular emission was self-consistently modelled based on the photoionisation code *Cloudy* (Ferland et al. 1998, 2013). The stellar population modelling was implemented through *FSPS* (Conroy et al. 2009; Conroy & Gunn 2010) using the *MIST* stellar models (Dotter 2016; Choi et al. 2016). We used a non-parametric SFH with seven bins, adopting a ‘bursty continuity’ prior. The first two bins were fixed at look-back times of 0–5 Myr and 5–10 Myr with the rest logarithmically spaced out to $z = 20$. We adopted a Chabrier (2003) initial mass function (IMF), with a cut-off at $150 M_\odot$. The other free parameters are the total stellar mass formed, stellar metallicity, gas-phase metallicity, dust attenuation (a screen and an additional component for young stars), and ionisation parameter. The set-up for the SED fitting is described in detail in Naidu et al. (2022, 2024), which is a modification of the prescription by Tacchella et al. (2022). It should be noted that the derived stellar masses are strongly dependent on the assumed SFH model. Whittler et al. (2023) find that non-parametric models with continuity prior can infer stellar masses up to an order of magnitude larger (an order of magnitude smaller sSFR) than the constant SFH models at the ages of $\lesssim 10$ Myr.

The SED modelling was done using photometric data and emission-line fluxes from the spectra. The photometric data used for SED fitting of galaxies from the EIGER, COLA1, and ALT survey varies due to differences in photometric filter coverage. For galaxies from the EIGER survey, we used photometry in the F115W, F200W, and F356W filters. In the COLA1 survey, we have photometry in an additional band, F150W, which we included in the SED modelling. For galaxies from the ALT survey, we have photometry from numerous programmes that targeted the Abell 2744 field (primarily UNCOVER; Bezanson et al. 2024 and MegaScience; Suess et al. 2024) totalling 27 filters available for the SED fitting². For the galaxies from the EIGER and COLA1 surveys, we additionally used the $H\beta$ and the $[O III]\lambda\lambda 4960, 5008$ emission line fluxes derived from the grism spectra. For the ALT galaxies, the emission line fluxes are de facto included due to the photometry in the various medium-band filters. We verified that there are no systematic biases between the stellar masses from the EIGER/COLA1 surveys versus the ALT survey by confirming that the relation between observed (magnification-corrected) F200W magnitude with the stellar mass is similar across the fields. We also confirmed that in the ALT data the emission-line fluxes from grism data are consistent with those inferred from medium-band photometry (Mascia et al. in prep).

3.3. Stacking grism spectra

Emission lines in the spectrum of a galaxy are a probe of the physical conditions of the ISM. However obtaining a high S/N for faint emission lines such as $H\gamma$ or $[O III]\lambda 4364$ in individual galaxies is difficult with NIRCcam grism data. In such cases, stacking methods can extract average measurements by reducing random noise fluctuations at the expense of losing information on individual sources, and therefore variation across the stacked sample. For illustrative purposes, we show the 1D stacked spectra of all the galaxies in our dataset in Fig. 2. The 2D spectra have been brought to rest-frame and normalised by their $([O III]\lambda 5008 + 2.98[O III]\lambda 4960)/2$ luminosity before stacking. From the 2D stacked spectrum, we extracted the 1D spectrum as

² Wavelengths around the Lyman- α line are ignored in the SED fitting because of uncertainty in the IGM attenuation.

described in Sect. 3.4.1. We can see strong [O III] $\lambda\lambda$ 4960, 5008 and H β emission lines along with the H γ , H δ , and [O III] λ 4364. We also see the HeI-5877 emission line in the stacked spectrum, although it is relatively faint. In comparison to the stacks presented in Roberts-Borsani et al. (2024) for galaxies between $z = 5-7$, we detect almost all the key emission lines over the same wavelength range (4000–6000 Å). Additionally, we resolve the [O III] $\lambda\lambda$ 4960, 5008 doublet, and H γ and [O III] λ 4364 due to the higher resolution of NIRCam/WFSS spectra in the F356W filter ($R \sim 1600$). Although our grism spectra have a higher resolution, our wavelength coverage is limited compared to the NIR-Spec/Prism spectrum, which covers from 1000 to 7000 Å in the same redshift range.

After dividing our data into H γ ($6.3 \leq z < 6.9$) and H β ($5.5 \leq z < 6.9$) samples and obtaining our emission-line spectra, as is described in Sect. 2, we performed median stacking of the emission-line spectra in stellar mass bins. In order to stack the spectra, we shifted the 2D spectrum of each galaxy to the rest-frame wavelength grid ranging from 4000 to 6000 Å through linear interpolation resampling. We verified that our resampling is flux-conserving. We then normalised the 2D spectrum of each galaxy by its respective $([\text{O III}]\lambda 5008 + 2.98[\text{O III}]\lambda 4960)/2$ luminosity. This allowed us to ensure that the stacks represent the typical line ratios of the spectra and are not dominated by more luminous galaxies. We obtained the stacked spectra by calculating the median over the 300 bootstrap (with replacement) realisations. The uncertainties were obtained by calculating the standard deviation of these bootstraps. The median-stacked spectra of our H γ sample is shown in Fig. 3.

3.4. Measuring emission-line luminosity

3.4.1. Extracting 1D spectra

We measured line luminosities based on optimally extracted 1D spectra (Horne 1986; Matthee et al. 2023). We collapsed the [O III] λ 5008 emission line over a wavelength range of ± 400 km/s with respect to its wavelength (5008.24 Å). We then performed non-linear least squares fitting of the spatial profile using the `lmfit` Python package, finding that a Gaussian profile suffices. Subsequently, we collapsed the 2D spectra by assuming the same spatial profile at all wavelengths. The uncertainty on the 1D spectrum was obtained by propagating the uncertainties of the 2D spectrum that we estimated from bootstrapping.

3.4.2. Emission line luminosity measurements

We obtained measurements of the luminosities for [O III] λ 5008, [O III] λ 4960, H β , H γ , and [O III] λ 4364 from our 1D continuum-subtracted stacks using the `lmfit` Python package. We followed a similar methodology to the one presented in Matthee et al. (2023). We had the best S/N for the [O III] $\lambda\lambda$ 4960, 5008 emission lines, which might have aided us in detecting the natural broadening of the emission lines that give it non-Gaussian wings (Lorentzian wings). However, in our case the extended shapes in the wings of the lines are most likely due to the spatial-spectral degeneracy inherent to grism data (i.e. the spatial extent of the galaxies in the dispersion direction broadens the lines in the grism spectra). Therefore, for [O III] λ 5008 and [O III] λ 4960 we used a Voigt profile. We fixed the standard deviation (σ) and the Gamma parameter to be the same for the [O III] doublet. We also fixed the emission-line ratio between the doublet to be

2.98^3 . We fitted a similar Voigt profile to the H β line so that we traced the emission lines coming from the same gas. We fixed the standard deviation, the Gamma parameter, and the amplitude ratio between the Gaussian and Lorentzian profiles to match the [O III] λ 5008] line. The other emission lines are much fainter; therefore, a Gaussian profile is used, as a Voigt profile does not improve the reduced χ^2 of the fits, nor is there significant flux in the wings. The errors on the emission-line luminosities were propagated from the standard error and the covariance matrix obtained from `lmfit`.

3.5. Dust correction

The luminosities of emission lines can be affected by dust attenuation; thus, it is important to account for it to accurately estimate the gas-phase metallicity for galaxies. Dust attenuation correction for emission line ratios such as [O III] λ 5008/H β should be negligible, though, as the emission lines are close in wavelength. We calculated the intrinsic, dust-corrected luminosity for [O III] λ 5008, [O III] λ 4364, and H β assuming the Cardelli et al. (1989) nebular attenuation curve in our H γ sample stack,

$$L_{\text{int}} = L_{\text{obs}} \cdot 10^{0.4E(B-V) \cdot k_{\lambda}}, \quad (1)$$

where

$$E(B - V) = 0.95 \cdot \tau, \quad (2)$$

and

$$\tau = \log_{10} \left(\frac{0.473}{F_{\text{H}\gamma}/F_{\text{H}\beta}} \right) \quad \text{for } F_{\text{H}\gamma}/F_{\text{H}\beta} \leq 0.473,$$

$$\tau = 0, \quad \text{for } F_{\text{H}\gamma}/F_{\text{H}\beta} > 0.473.$$

k_{λ} is the attenuation curve, which is wavelength-dependent, and 0.473 is the intrinsic H γ /H β line flux ratio that is appropriate for Case B recombination at 10^4 K (Osterbrock & Ferland 2006). We measured typical H γ /H β values of 0.37 ± 0.04 at average [O III] temperatures of $\sim 16\,600$ K. This indicates some level of dust attenuation; however, we find that the average dust-corrected [O III] λ 5008/H β (R3) and [O III] λ 5008/[O III] λ 4364 ratios are only 0.01 dex and 0.03 dex lower than the average dust-uncorrected ratios, respectively.

The [O III] λ 5008/H β ratios for the H β sample stacks are not dust-corrected as we do not always cover the H γ line required to calculate Balmer decrements and the lines are close in wavelength range. In Fig. 4, we show the [O III] λ 5008/H β ratios obtained for our stacked samples, along with ratios for individual galaxies in our entire dataset. For our stacks, we consider all H β observations as focusing only on $S/N(\text{H}\beta) \geq 3$ detections would bias our stacked values to lower ratio values at the low-mass end. Applying a selection cut based on $S/N(\text{H}\beta)$ would impose additional biases in our sample that would impact our analysis of the MZR. The [O III] λ 5008/H β ratio seems relatively invariant with stellar mass for the stacked samples. Additionally, the average ratios for both H γ and H β sample stacks are within 0.0009 dex of each other, which suggests that results obtained for the H γ sample can be applied to the H β sample (see Fig. 4).

³ The Bayesian information criterion (BIC) values of the Voigt profiles for the [O III] $\lambda\lambda$ 4960, 5008 doublet with the same standard deviation and Gamma parameter values and a fixed emission-line ratio are significantly lower than the BIC values for Gaussian profiles with the same standard deviation and fixed emission-line ratio. The absolute difference in the BIC values is larger than 10, with the BIC values for the Voigt profiles being lower. Therefore, a Voigt profile is preferred.

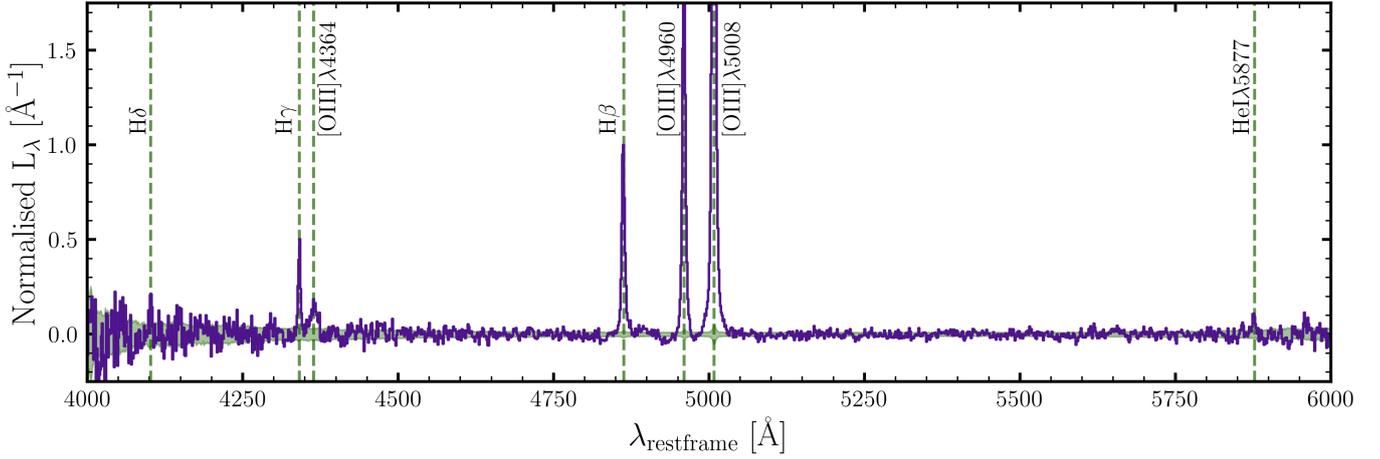


Fig. 2. Median stacked 1D rest-frame emission-line spectrum of the full sample of 792 [O III] emitters at $z = 5.5\text{--}7$. Each galaxy spectrum has been normalised by its $([\text{O III}]\lambda 5008 + 2.98[\text{O III}]\lambda 4960)/2$ flux before stacking. The shaded green region shows the uncertainty estimated through bootstrap resampling. We highlight the wavelengths of H δ , H γ , [O III] $\lambda 4364$, H β , [O III] $\lambda 4960$, 5008, and He I $\lambda 5877$. For visualisation, we normalised the stacked spectrum by the H β luminosity such that H β peaks at unity.

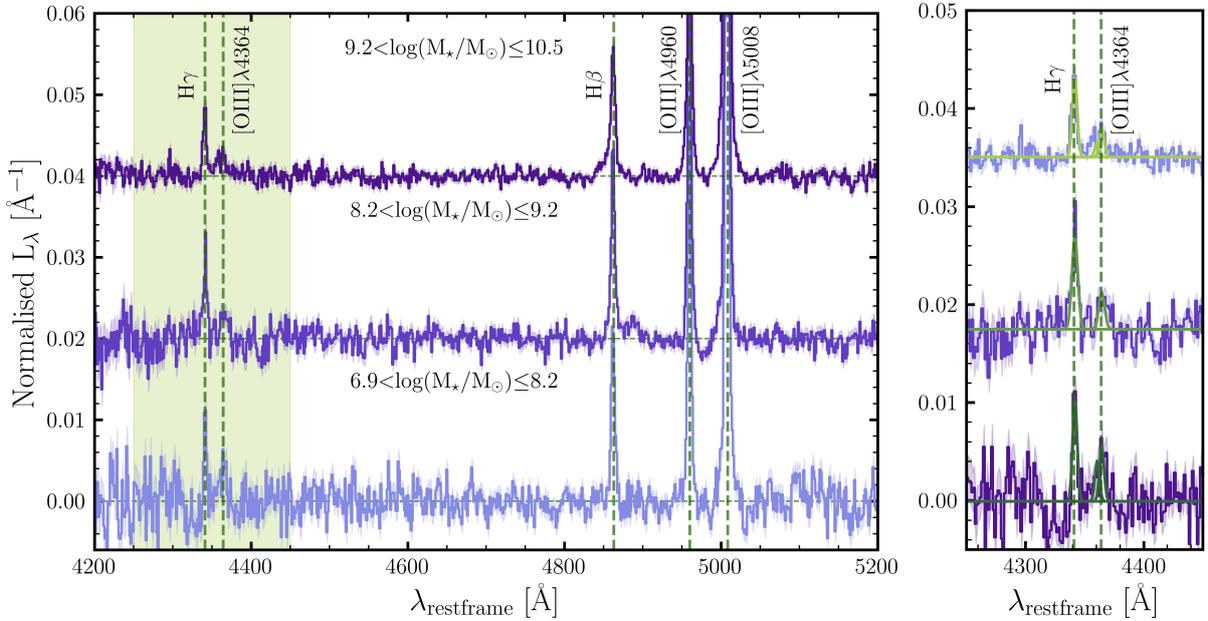


Fig. 3. Median stacked 1D rest-frame emission-line spectra of the H γ sample ($z = 6.3\text{--}7$), in three stellar mass bins offset for visualisation purposes. The shaded region shows the uncertainty estimated through bootstrap resampling. In the right panel, we zoom into the wavelength range containing H γ and [O III] $\lambda 4364$. We fitted a complex of three Gaussians with the same standard deviation to obtain the fluxes of H γ and [O III] $\lambda 4364$. An additional Gaussian component was fitted to account for possible contamination by [Fe II] $\lambda 4360$.

Table 1. Metallicity and SFR measurements in bins of stellar mass for the H γ sample with $6.3 < z < 7$.

Mass Bin	No. of Galaxies	$\log(M_*/M_\odot)^a$	$12+\log(\text{O}/\text{H})^b$	$\log_{10}(\text{SFR} / M_\odot \text{ yr}^{-1})^c$
$6.89 \leq \log(M_*/M_\odot) < 8.2$	113	$7.77^{+0.32}_{-0.46}$	$7.69^{+0.17}_{-0.13}$	0.48 ± 0.12
$8.2 < \log(M_*/M_\odot) < 9.2$	96	$8.60^{+0.24}_{-0.28}$	$7.72^{+0.18}_{-0.14}$	0.59 ± 0.06
$9.2 < \log(M_*/M_\odot) \leq 10.54$	33	$9.61^{+0.23}_{-0.24}$	$7.94^{+0.14}_{-0.12}$	1.09 ± 0.07

Notes. ^(a) The stellar masses are median values within each bin, with the upper and lower errors representing the 84th and 16th percentiles, respectively. ^(b) The $12+\log(\text{O}/\text{H})$ measurements are based on the direct temperature method. ^(c) Obtained from dust-corrected H β luminosities using Table 2 from Theios et al. (2019), which considers case B recombination, BPASSv2.2 $100 M_\odot$, $Z_* = 0.004$, a constant SFH, and a minimum allowed age of 50 Myr.

Table 2. Metallicity and SFR measurements in stellar mass bins for the H β ($5.5 < z < 7$) sample.

Mass Bin	No. of Galaxies	$\log(M_*/M_\odot)^a$	$\frac{L_{[\text{O III}]\lambda 5008}}{L_{\text{H}\beta}}$	$12+\log(\text{O}/\text{H})$ (lower branch) ^b	$12+\log(\text{O}/\text{H})$ (higher branch)	$\log_{10}(\text{SFR} / M_\odot \text{ yr}^{-1})^c$
$6.68 \leq \log(M_*/M_\odot) < 7.8$	159	$7.36^{+0.29}_{-0.26}$	6.47 ± 0.26	$7.76^{+0.09}_{-0.08}$	$8.19^{+0.10}_{-0.10}$	0.24 ± 0.02
$7.8 < \log(M_*/M_\odot) < 8.0$	86	$7.88^{+0.09}_{-0.09}$	6.41 ± 0.27	$7.75^{+0.08}_{-0.08}$	$8.21^{+0.09}_{-0.00}$	0.33 ± 0.02
$8.0 < \log(M_*/M_\odot) < 8.25$	85	$8.13^{+0.08}_{-0.09}$	5.40 ± 0.19	$7.51^{+0.04}_{-0.04}$	$8.54^{+0.07}_{-0.06}$	0.47 ± 0.01
$8.25 < \log(M_*/M_\odot) < 8.5$	91	$8.35^{+0.09}_{-0.06}$	6.97 ± 0.26	$7.83^{+0.06}_{-0.06}$	$8.12^{+0.07}_{-0.07}$	0.41 ± 0.02
$8.5 < \log(M_*/M_\odot) < 8.75$	100	$8.63^{+0.08}_{-0.08}$	6.26 ± 0.19	$7.72^{+0.08}_{-0.06}$	$8.25^{+0.08}_{-0.09}$	0.53 ± 0.01
$8.75 < \log(M_*/M_\odot) < 9.0$	71	$8.84^{+0.11}_{-0.06}$	6.25 ± 0.18	$7.72^{+0.07}_{-0.06}$	$8.24^{+0.08}_{-0.08}$	0.62 ± 0.01
$9.0 < \log(M_*/M_\odot) < 9.5$	79	$9.21^{+0.18}_{-0.19}$	6.77 ± 0.21	$7.83^{+0.06}_{-0.07}$	$8.11^{+0.08}_{-0.07}$	0.77 ± 0.01
$9.5 < \log(M_*/M_\odot) \leq 11.1$	53	$9.71^{+0.26}_{-0.17}$	6.06 ± 0.17	$7.66^{+0.06}_{-0.05}$	$8.32^{+0.06}_{-0.08}$	0.83 ± 0.01

Notes. ^(a) As in Table 1, the stellar masses are median values within each bin with the upper and lower errors representing the 84th and 16th percentile, respectively. ^(b) The $12+\log(\text{O}/\text{H})$ estimates were obtained from the R3 calibration from Chakraborty et al. (2025) and we show the results for both branches. ^(c) The SFRs are based on the H β line luminosity as in Table 1.

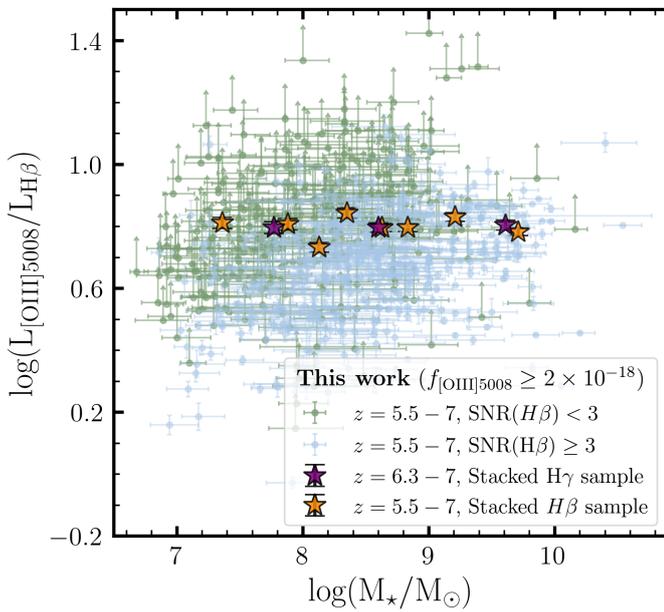


Fig. 4. R3 ratio as a function of stellar mass ratios for the entire dataset. Galaxies with $S/N(\text{H}\beta) \geq 3$ are shown in blue and galaxies with $S/N(\text{H}\beta) < 3$ are shown in green with lower limits. Stars show measurements in stacked spectra.

4. Metallicity measurements

4.1. Direct method metallicity measurements

We followed the prescription described in Izotov et al. (2006) to estimate the gas phase metallicity using the direct- T_e method. For our H γ sample, we detected the faint auroral line [O III] $\lambda 4364$ with $S/N > 3$ in the stellar mass binned stacks. As the [O III] $\lambda 4364$ /[O III] $\lambda 5008$ ratio is highly temperature-dependent, we can use these detections to derive the electron temperature and the metallicities directly. We do not detect any [O II] emission lines that are required to calculate the electron density and the ionisation state of the gas as they are out of our observable spectral range. Therefore, we assumed an electron density (n_e) of 300 cm^{-3} as has been reported for similar low-metallicity, high-redshift galaxies in the literature (Sanders et al. 2016; Curti et al. 2024; Isobe et al. 2023; Matthee et al. 2023; Sanders et al. 2024; Topping et al. 2025). Recent work by Scholte et al. (2025) and Pollock et al. (2025), probing the ISM for galaxies between $z \sim 5.5$ –10, suggests that

T_e measurements and derived abundances are relatively insensitive to the electron density values, $n_e \lesssim 10^4 \text{ cm}^{-3}$. We find that for an electron density between $100 \leq n_e(\text{cm}^{-3}) \leq 10^4$, our calculated abundances change by 0.02 dex at most. We also assume the O32 ratio – $\log([\text{O III}]\lambda 5008)/[\text{O II}]\lambda \lambda 2727, 29$ – to follow a Gaussian distribution with a mean of 0.98 and standard deviation of 0.36, as has been observed by Cameron et al. (2023) for galaxies in the JADES survey between $5.5 < z < 9.5$. We find that their sample probes a similar dynamic range in M_{UV} and [O III] $\lambda 5008$ /H β ratio as our entire sample. We find that the average [O III] $\lambda 5008$ /H β ratio for our H γ sample lies 0.06 dex above the average value for $z \sim 6$ galaxies in Cameron et al. (2023). This can imply that the O32 ratios may be higher than assumed, which would lead to lower metallicities. To confirm this, we would require individual measurements covering both the [O III] and [O II] doublets. We used the $T_e[\text{O III}]-T_e[\text{O II}]$ relation given for low gas-phase metallicity from Izotov et al. (2006). With the dust-corrected [O III] $\lambda 5008$ /[O III] $\lambda 4364$ and the assumed O32 ratio and n_e values, we estimated the O^{2+} and O^+ gas temperature using the Python package PyNeb (Luridiana et al. 2013).

We calculated the total oxygen abundance assuming that there is only O^+ and O^{2+} present in the HII region. As we did not detect He II $\lambda 4686$, a high ionisation energy line, in our stacked spectra, we did not add O^{3+} ion abundance to our oxygen abundance (Izotov et al. 2006) as O^{3+} requires an even higher ionisation potential energy to be excited:

$$\frac{\text{O}}{\text{H}} = \frac{\text{O}^+}{\text{H}} + \frac{\text{O}^{2+}}{\text{H}}. \quad (3)$$

The temperature of O^{2+} gas ranges from $1.3 \times 10^4 \text{ K}$ to $2.1 \times 10^4 \text{ K}$ in our stacked sample, and the temperature of O^+ gas ranges from $1.3 \times 10^4 \text{ K}$ to $1.6 \times 10^4 \text{ K}$. The metallicities of our H γ sample are presented in Table 1. In Fig. 5, we show that our metallicity values for the stacks of the H γ sample agree with the R3 calibration curve given by Chakraborty et al. (2025), Sanders et al. (2024), and Nakajima et al. (2022), discussed in more detail in Sect. 4.2. The observed stellar masses of galaxies in our H γ sample span a range of $\log(M_*/M_\odot) = 6.9$ –10.5, with gas-phase metallicities ranging between $12+\log_{10}(\text{O}/\text{H}) = 7.6$ –8.1 ($Z/Z_\odot = 0.07$ –0.24).

The uncertainties in our calculated gas temperatures and metallicities were calculated by perturbing the observed emission line luminosity values by their measured errors. We have 1000 realisations of the perturbation and calculated the median and the 16th and 84th percentile values.

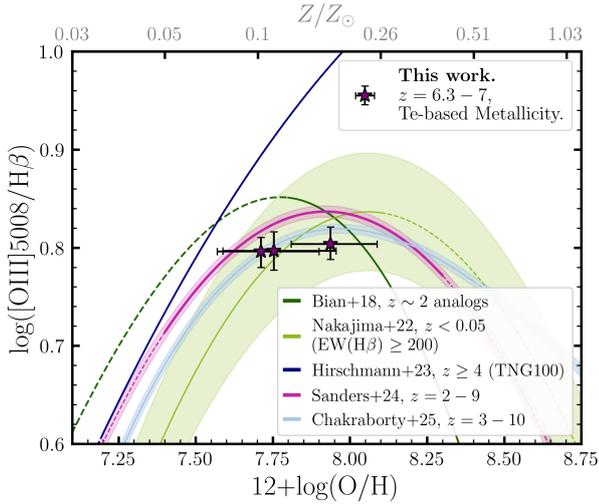


Fig. 5. Relation between the gas-phase metallicity measured using the direct T_e -method and the R3 ratio. We compare our estimates with several strong-line metallicity calibrations from literature. We show the calibration by [Bian et al. \(2018\)](#), which uses local analogues that occupy similar regions to $z \sim 2$ star-forming galaxies on the BPT diagram, the calibration based on local galaxies with $\text{EW}(\text{H}\beta) \geq 200$ derived by [Nakajima et al. \(2022\)](#), [Hirschmann et al. \(2023\)](#) calibration for $z > 4$ galaxies from the TNG100 simulation, and [Sanders et al. \(2024\)](#) and [Chakraborty et al. \(2025\)](#) calibrations for high- z galaxies between $2 < z < 10$.

4.2. Strong line calibration measurements

For our $\text{H}\beta$ sample ($5.5 \leq z < 7$), we do not cover the faint $[\text{O III}]\lambda 4364$ auroral line. Therefore, we estimated the metallicity based on the R3 ratio using the [Chakraborty et al. \(2025\)](#) and [Sanders et al. \(2024\)](#) calibration recently derived for high-redshift galaxies, and the [Nakajima et al. \(2022\)](#) calibrations derived for high $\text{H}\beta$ equivalent width galaxies ($\text{EW}(\text{H}\beta) \geq 200 \text{ \AA}$) in the local Universe (see Table 2). Our choice for these calibrations is validated by our direct- T_e metallicity measurements for our $\text{H}\gamma$ sample (see Fig. 5). To obtain the metallicities for the $\text{H}\beta$ sample, we solved the quadratic equation (cubic equation for [Chakraborty et al. \(2025\)](#)) for the R3-calibrations iteratively, taking into account the observational error on the R3 ratio along with the errors on the calibration itself. We took the median of the solutions for the lower and higher branches to be the low and high metallicity estimate and the 16th and 84th percentile as the errors. As our spectra do not cover the $[\text{O II}]$ emission lines, we cannot directly break the degeneracy between the ionisation parameter and the metallicity, which could both impact the $[\text{O III}]\lambda 5008/\text{H}\beta$ ratio. However, since our $\text{H}\gamma$ sample estimates fall either on the left branch or at the vertex of these relations, we assume that the galaxies in our $\text{H}\beta$ sample also fall on the left branch of the calibrations, which corresponds to high ionisation and low metallicity. Our observations match best with the R3-calibration presented by [Chakraborty et al. \(2025\)](#), which is based on direct- T_e metallicities for galaxies at $3 < z < 10$. We discuss the impact of using strong-line calibrations in the next section.

5. The mass-metallicity relation at $z \sim 6$

Figure 6 shows the gas-phase oxygen abundances measured in our median stacked $\text{H}\gamma$ and $\text{H}\beta$ sample along with other literature measurements. We find a positive correlation between gas-

phase metallicity and stellar mass for both of our subsets. We parametrise the MZR as

$$12 + \log(\text{O}/\text{H}) = Z_{\text{norm}} + \gamma \cdot (\log(M_{\star}/M_{\odot}) - 10), \quad (4)$$

where for the $\text{H}\gamma$ sample we find that the best-fit normalisation is $Z_{\text{norm}} = 7.96 \pm 0.10$, and the slope is $\gamma = 0.12 \pm 0.08$. The parameters were obtained through orthogonal distance regression (ODR) using the `scipy.odr` package. Despite our lower sensitivity, our inferred metallicities agree with metallicity measurements, in similar redshift ranges, based on other surveys and instruments (e.g. NIRSpec MSA, [Curti et al. 2024](#); [Nakajima et al. 2023](#); [Chakraborty et al. 2025](#)).

Our MZR lies systematically below the MZR in the local Universe and those at $z \sim 2-3$. It lies 0.72 dex below the MZR given by [Scholte et al. \(2024\)](#) for galaxies in the local Universe. At $z \sim 2$, it lies 0.55 dex, 0.35 dex, and 0.61 dex below the MZR by [Sanders et al. \(2021\)](#), [Li et al. \(2023\)](#), and [He et al. \(2024\)](#), respectively. It lies 0.45 dex, 0.28 dex and 0.63 dex below the MZR at $z \sim 3$ given by [Sanders et al. \(2021\)](#), [Li et al. \(2023\)](#), and [He et al. \(2024\)](#), respectively.

The normalisation of our $\text{H}\gamma$ sample MZR agrees with previous indications of redshift evolution in the MZR ([Maiolino & Mannucci 2019](#); [Curti 2025](#), and references therein). We find that the slope of our MZR for the $\text{H}\gamma$ sample agrees well with slopes found for a combined ‘green peas’ and ‘blueberries’ sample ([Yang et al. 2017a,b](#)) at $z < 0.2$, $z = 2-3$ galaxies by [Li et al. \(2023\)](#) using NIRISS grism, and the sample from the JADES survey at $z = 6-10$ ([Curti et al. 2024](#)). In standard gas regulator models, invariance in the slope of the MZR with an increase in redshift may suggest invariance in physical mechanisms driving outflows in star-forming galaxies, whereas the lowering of normalisation at fixed stellar mass at higher redshifts indicates inflow of pristine gas ([Sanders et al. 2021](#); [Heintz et al. 2023](#)). Table B.1 presents the slope and normalisation for the MZR for our study and other works.

We also find that the slope of the MZR for the $\text{H}\beta$ sample is flatter than that of our $\text{H}\gamma$ sample. As was described in previous sections, the metallicities for the $\text{H}\beta$ sample were obtained from the lower branch of the [Chakraborty et al. \(2025\)](#), [Sanders et al. \(2024\)](#), and [Nakajima et al. \(2022\)](#) calibrations. The slope of the MZRs obtained using [Chakraborty et al. \(2025\)](#), [Sanders et al. \(2024\)](#), and [Nakajima et al. \(2022\)](#) calibrations are 0.10, 0.066, and 0.023 respectively, which are ~ 0.03 , 0.06, and 0.1 dex flatter than the slope of our direct- T_e estimates. Our R3 ratios for the $\text{H}\beta$ sample lie close to the vertex of the parabola, which can make the MZR appear flatter. Indeed, in Fig. 7, we show that calculating the metallicities of our $\text{H}\gamma$ sample stacks using the [Chakraborty et al. \(2025\)](#), [Sanders et al. \(2024\)](#), and [Nakajima et al. \(2022\)](#) calibrations significantly flattens the slope of the MZR compared to the MZR from direct- T_e estimates. The general flatness of our MZR slope could also be due to the effects of our selection function, which we discuss in detail for the $\text{H}\gamma$ sample in Sect. 6.

While there are studies that our MZR parameters agree with, the investigation of redshift evolution of the MZR is highly sensitive to systematic effects. For example, the best-fit parameters for our $\text{H}\gamma$ sample MZR compared to the MZR’s from other studies in the same redshift range differ due to discrepancies in the methods used to obtain gas-phase metallicities, different stellar mass ranges probed, different selection functions, and different parameters for SED fitting ([He et al. 2024](#)). Similarly, comparison of the MZR parameters for other redshift ranges is also affected by the same systematics. To understand the true evolution of the MZR, one should account for the selection effects for

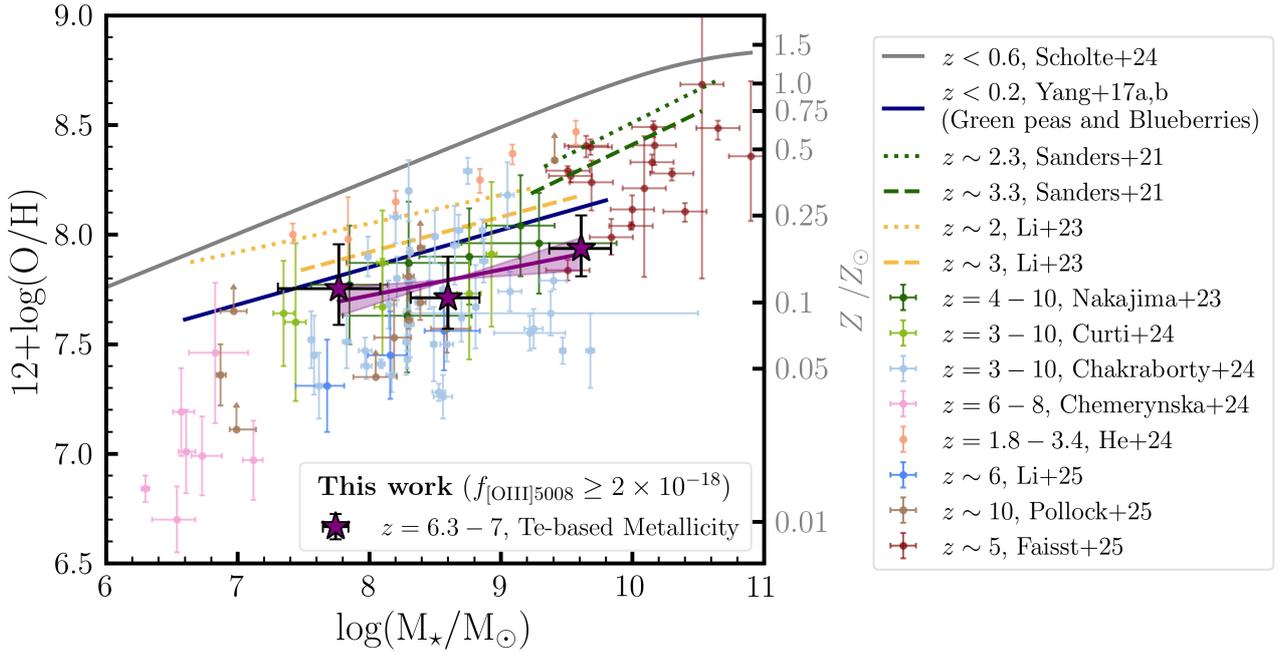


Fig. 6. Mass-metallicity relation from our stacked H γ sample (purple stars, $6.3 \leq z < 7$). For the highest-mass bin in the H β sample, we also show the metallicity estimate from the higher-metallicity branch of the Sanders et al. (2024) R3-calibration curve (unfilled black star). The metallicity estimates and median stellar masses for both the samples are reported in Tables 1 and 2. The errors on the stellar mass were calculated using the 16th and 84th percentile values in the stack. While our measurements are from stellar-mass binned and stacked NIRCcam grism spectra, similar to Li et al. (2025), measurements from Li et al. (2023) and He et al. (2024) were obtained using stacked spectra from NIRISS grism, measurements from Chakraborty et al. (2025), Chemerynska et al. (2024) and Pollock et al. (2025) are from individual galaxies, values from Curti et al. (2024) and Nakajima et al. (2023) are average metallicity measurements obtained from NIRSpc spectroscopy, and values from Faisst et al. (2025) were obtained from NIRSpc/IFU observations of high-mass main-sequence galaxies with [C II]158 μ m observations from ALMA.

the various studies and estimate the metallicities using the same methodology. This is outside the scope of this work. However, we attempt to account for our observational selection and understand the nature of the intrinsic MZR in the section ahead.

6. Discussion

According to gas-regulator models, the slope of the MZR is governed by feedback mechanisms that drive gas out of galaxies (Peeples & Shankar 2011; Sanders et al. 2021; Curti et al. 2024). Such outflows are assumed to be driven by stellar feedback processes and are expected to be more effective in removing metal-enriched gas from lower-mass galaxies compared to high-mass ones due to their shallower gravitational potential wells. The normalisation is controlled by the ‘effective yield’ in the galaxy, which factors in the effect of nucleosynthetic stellar yields and gas flows. Finally, at fixed mass, the scatter along the MZR has been found to (anti-)correlate with the SFR of the galaxies. This has been observed as the FMR at $z \sim 0$ (Mannucci et al. 2010; Lara-López et al. 2010; Andrews & Martini 2013; Curti et al. 2020). The existence of the FMR at high- z is still an unsolved question as many studies have shown that assuming the same FMR as at $z \sim 0$ predicts higher metallicities than the ones they measure (Zahid et al. 2014; Salim et al. 2015; Kashino et al. 2017; Sanders et al. 2018; Heintz et al. 2023; Curti et al. 2024; Scholte et al. 2025; Korhonen Cuestas et al. 2025).

With recent results from JWST, new interpretations in understanding the MZR and FMR have arisen. JWST has shown that SFHs for high- z galaxies are usually bursty (Topping et al. 2022; Endsley et al. 2023; Looser et al. 2025). The burst of star forma-

tion may occur due to inflow of pristine gas, which dilutes the ISM, allowing for the galaxy to have a lower metallicity temporarily. However, there are other mechanisms through which the burst could occur (mergers, gravitational instabilities) that can cause an increase in the gas-phase metallicity (Tacchella et al. 2023). It has also been proposed that high- z galaxies have not yet reached a stage of equilibrium between gas accretion, star formation, chemical evolution, and feedback processes (Mannucci et al. 2010; Møller et al. 2013; Maiolino & Mannucci 2019; Curti et al. 2024; Sarkar et al. 2025; Scholte et al. 2025; Korhonen Cuestas et al. 2025). Evidence of a non-equilibrium state of the ISM in star-forming galaxies has been observed through studies of the circumgalactic medium (CGM) of Mg II absorption systems along the sight line of quasars at $z \sim 6$ (Bordoloi et al. 2024) and also in terms of turbulent gas kinematics of high- z galaxies (Danhaive et al. 2025).

In the following sub-sections we discuss the implications of our measurements and investigate the impact of our selection function using simulations and mock datasets. It is noteworthy that having a selection function based on H β fluxes, instead of [O III] λ 5008 fluxes, would also impose selection effects on the observed MZR due to the possibly interrelated nature of the 3D relation between stellar mass, SFR, and metallicity at these redshifts.

6.1. Comparison to cosmological simulations

The observed stellar masses of the galaxies in our H γ sample span a range of $\log(M_*/M_\odot) = 6.9\text{--}10.5$, with gas-phase metallicities ranging between $Z_{\text{gas}} = 7.7\text{--}8.0$. We compared these measurements to those simulated in the

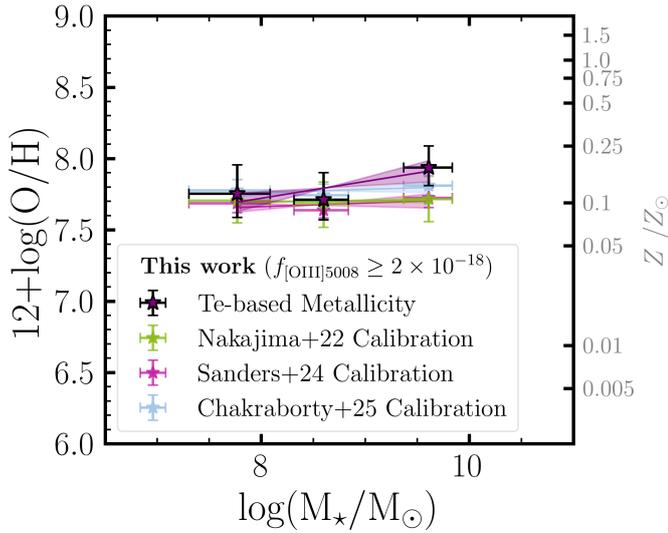


Fig. 7. Mass-metallicity relation inferred for our Hy sample using the direct- T_e method (purple stars), Chakraborty et al. (2025) R3-calibration (blue stars), Sanders et al. (2024) R3-calibration (magenta stars), and the Nakajima et al. (2022) R3-calibration for local galaxies with $\text{EW}(\text{H}\beta) > 200 \text{ \AA}$ (green stars). We used metallicity estimates from the lower branch of the R3-calibrations. The use of strong-line methods leads to an artificial flattening of the observed relation.

SPHINX²⁰ (Rosdahl et al. 2018; Katz et al. 2023) and FLARES (Lovell et al. 2020; Vijayan et al. 2020) hydrodynamical simulations, for which [O III] and H β luminosities are available. SPHINX²⁰ is a cosmological hydrodynamical radiative transfer simulation with a volume of 20^3 cMpc^3 . The simulation was run down to $z \approx 4.5$. The FLARES simulations are a suite of 40 zoom-re-simulations using the EAGLE model (Schaye et al. 2015; Crain et al. 2015), which has been shown to reproduce galaxy properties over a wide range of redshifts including $z \approx 0$. The zoom regions were chosen to simulate a wide range of environments, from highly over-dense to under-dense. The FLARES simulations do not self-consistently model radiative transfer, but emission-line luminosities are added in post-processing based on Cloudy (version 17.03) modelling (Vijayan et al. 2020). For both simulations, we selected galaxies between $z = 5-7$ to compare to our dataset. We used the stellar masses, dust-attenuated emission-line luminosities, and mass-weighted gas-phase metallicities from these simulations. In our comparison, we also include JAGUAR (Williams et al. 2018), which is a mock galaxy catalogue based on observationally driven empirical models of the evolution of galaxy properties.

In Fig. 8, we show the MZR from FLARES, SPHINX²⁰, and JAGUAR at $z \sim 6$. We used the mass-weighted gas-phase metallicity from the SPHINX²⁰ and FLARES simulation. Nakajima et al. (2023) make use of stellar metallicities of young star particles ($< 10 \text{ Myr}$) from FLARES for comparison to their observations, assuming that the stellar and gas-phase metallicities in the region of massive star formation are comparable. We find that the slope (γ) and normalisation (Z_{norm}) of the two MZRs from FLARES differ by 0.03 dex and 0.17 dex, respectively, with the MZR from stellar metallicities of young star particles being slightly steeper. From the JAGUAR catalogue, we used the gas-phase metallicities that were estimated using the Pettini & Pagel (2004) N2-metallicity calibration. We also plot our measurements based on the direct- T_e method. We find that the simulations differ significantly. The slope and normal-

isation of the simulations, following Eq. (4), have been noted in Table B.1. The number of high-mass galaxies is limited in SPHINX²⁰ and JAGUAR by the simulated volume. The low-mass cut-off determined by the resolution of the simulation, which is higher for SPHINX²⁰, therefore includes lower-mass galaxies. At face value, our MZR is flatter than the one in FLARES and SPHINX²⁰.

It is important to consider that our observational sample has been selected based on [O III] emission, the strength of which is sensitive to both SFR and metallicity, and therefore may yield a biased view of the MZR. In order to investigate the impact of our [O III]-flux limited galaxy sample on the observed MZR, we implemented our selection cut ($f_{[\text{O III}]\lambda 5008} \geq 2 \cdot 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$) on the [O III] $\lambda 5008$ fluxes in these simulated datasets. Before comparing the impact of the selection function on the observed MZR, we first compare our observed the [O III] luminosity as a function of stellar mass to that in simulations. We find that FLARES and JAGUAR galaxies have roughly similar relations between [O III] luminosity and stellar mass as our observed data (see Appendix C). The SPHINX²⁰ simulation, however, shows systematically higher masses at fixed [O III] luminosity, which stems from the fact that SPHINX²⁰ has an elevated stellar-mass-halo-mass relation (i.e. galaxy formation is too efficient) and which therefore assigns too low SFRs (and consequently, [O III] luminosity) at fixed mass (Katz et al. 2023). As a result, after applying our selection cut, the SPHINX²⁰ simulation only considers galaxies with $\log(M_\star/M_\odot) > 8.4$.

For the MZR, we find that our selection function causes the observed slope to be flatter in JAGUAR and SPHINX²⁰ simulations. For JAGUAR, this is because our selection cut primarily removes low-metallicity galaxies at low masses. In SPHINX²⁰, along with the selection described before, it primarily removes massive galaxies with relatively high metallicities. Both effects effectively lead to a flattening. Our [O III] selection does not affect the observed MZR slope from FLARES. As is discussed in the next section, the effect of the [O III] selection on the shape of the MZR depends on the correlation metallicity and the SFR at fixed mass. In FLARES, variations in metallicity are independent of SFR, such that the selection effect does not bias the observed shape of the MZR. The slopes for the [O III]-flux limited MZR are also summarised in Table B.1. We find a flatter MZR (slope, $\gamma = 0.12 \pm 0.08$) from our Hy sample as compared to SPHINX²⁰ ($\gamma_{\text{sel}} = 0.22$) and FLARES ($\gamma_{\text{sel}} = 0.44$) MZR even after mimicking the selection cut in the simulations. Therefore, we conclude that none of these hydrodynamical simulations is able to match our observed MZR, in either the slope or the normalisation.

We note that the $L_{[\text{O III}]} - M_\star$ relation from SPHINX²⁰ does not match our observations due to the low SFRs of SPHINX²⁰ galaxies (see Appendix C). Since the gas-phase metallicity is inversely correlated to the sSFR in the simulation, our selection cut ends up selecting relatively high-metallicity galaxies and so the ‘observed’ MZR also does not match. In SPHINX²⁰, we also miss out the highest-mass and highest-metallicity galaxies due to them being very dust-obscured, and thus not meeting our [O III] flux threshold. It is important to mention that we cannot account for missing out [O III] $\lambda\lambda 4960, 5008$ emitting galaxies in our observational data due to extreme dust-obscuration. This might contribute to the lack of a match between our observations and the simulations. There are no strong indications that we miss a large population of metal-enriched, dusty galaxies. For example, Rowland et al. (2025) observe the [O III] doublet for dusty galaxies at $6 < z < 8$ with $M_\star > 10^9 M_\odot$ from the REBELS ALMA programme. Their sample is found to be more metal-rich

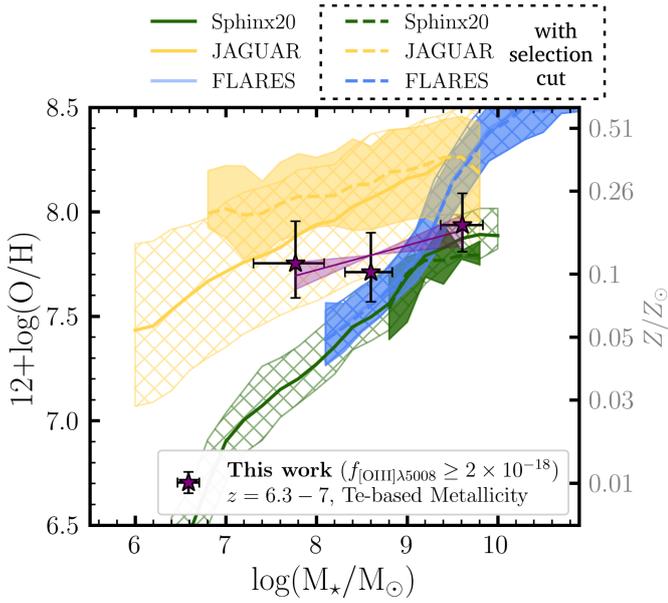


Fig. 8. Mass-metallicity relation for the SPHINX²⁰ (Katz et al. 2023) and FLARES simulation (Lovell et al. 2020; Vijayan et al. 2020), and the JAGUAR catalogue (Williams et al. 2018). We plot both the intrinsic MZR and the MZR after mimicking our selection cut. We also show the MZR from our stacked H γ (purple stars) sample. The MZR slopes and normalisations are listed in Table B.1.

than the bulk of the $z > 6$ population in the literature. However, we note that with their [O III] luminosities, such galaxies would have all been selected if they were covered by our observations.

6.2. Insights from modelling selection effects

Since the simulations do not match our observations and have varying responses to our selection function, we investigated the effects of our [O III] flux-limited selection function on the observed mass-metallicity relation using simple toy models. In these models we explore the various dependencies between metallicity and SFR at fixed stellar mass. We started by sampling the stellar masses for galaxies using the galaxy stellar mass function (SMF) from Weibel et al. (2024) at $z = 6$. Then, we assumed the following linear relation between SFR and the stellar mass that has been found by multiple observational studies (e.g. Speagle et al. 2014; Popesso et al. 2023):

$$\log_{10}(\text{SFR}) = \text{MS}_{\text{norm}} + \text{MS}_{\text{slope}} \times \log_{10} \left(\frac{M_{\star}}{10^{10} M_{\odot}} \right) + \mathcal{N}(0, \sigma_{\text{SFR}}), \quad (5)$$

where the main-sequence slope is $\text{MS}_{\text{slope}} = 1.0$ (motivated by the absence of strong evolution in the faint-end slope of the galaxy SMFs, Leja et al. 2015), the main-sequence normalisation is $\text{MS}_{\text{norm}} = 2.0$ (i.e. $\text{SFR} = 100 M_{\odot}/\text{yr}$ at $\log(M_{\star}/M_{\odot}) = 10$), and the intrinsic scatter along the relation is given by a normal distribution with standard deviation $\sigma_{\text{SFR}} = 0.3$ (as found in Di Cesare et al. 2025 for $\log(M_{\star}/M_{\odot}) = 8$). Additionally, we considered the following 3D relation between gas-phase metallicity, stellar mass, and SFR:

$$Z = Z_0 + M_{\text{coeff}} \times \log(M_{\star}/M_{\odot}) + S_{\text{coeff}} \times \log(\text{SFR}/M_{\odot} \text{ yr}^{-1}), \quad (6)$$

where, Z_0 is the normalisation, and M_{coeff} and S_{coeff} are the coefficients that determine the dependency between metallicity

on stellar mass and SFR (note that M_{coeff} is not equal to the slope in the MZR due to the SFR dependence on mass). We then estimated H β luminosities for our mock galaxies considering case B recombination and using the SFR- $L_{\text{H}\alpha}$ relation from Theios et al. (2019), which assumes a Kroupa IMF with upper mass cut-off of $100 M_{\odot}$ and slope -2.35 for $M_{\star} > 1 M_{\odot}$, and the BPASSv2.2 model with a constant SFH and a stellar metallicity of $Z_{\star} = 0.004$. Finally, we obtained [O III] $\lambda 5008$ luminosities using the calibration between the R3 ratio and gas-phase metallicity from Sanders et al. (2024).

Since we make various assumptions while generating the toy models, there are some caveats to keep in mind. Unlike our assumption, the scatter along the SFR- M_{\star} relation may be mass-dependent, which would introduce a mass dependency in the scatter along the FMR. We also assumed a linear conversion between SFR and $L_{\text{H}\alpha}$ as was prescribed by Theios et al. (2019), which has been assumed in the literature extensively, but which may not be the most optimal in the early Universe (Kramarenko et al. 2025). We assumed the R3 calibrations from Chakraborty et al. (2025), without any scatter, to get [O III] $\lambda 5008$ luminosities. It is plausible that for a given metallicity and H β luminosity, there is scatter in this calibration that could be due to variations in the ionisation parameter. Finally, we ignored the effects of dust attenuation motivated by the low attenuation found in Sect. 3.5.

We generated mock datasets for three cases:

1. **Normal FMR case** ($S_{\text{coeff}} < 0$): There exists a 3D relation between the gas-phase metallicity, stellar mass, and SFR, such that gas-phase metallicity is inversely proportional to the specific star formation rate (sSFR). This is seen in galaxies in the local Universe and understood in the context of gas regulation models.
2. **No FMR case** ($S_{\text{coeff}} = 0$): In this scenario, there is no 3D relation between the gas-phase metallicity, stellar mass, and SFR. The gas-phase metallicity only depends on the stellar mass (MZR) and the scatter in the MZR is unrelated to SFR. This is the case in the FLARES simulation.
3. **Inverted FMR case** ($S_{\text{coeff}} > 0$): There exists a 3D relation between gas-phase metallicity, stellar mass, and SFR, but here gas-phase metallicity is directly proportional to the sSFR. This is seen on resolved (100 pc) scales in local galaxies (Wang & Lilly 2021) and suggests delayed or inefficient feedback.

The parameter choices for the three cases are given in Table 3. We motivate these three cases as they have a distinct impact on how our selection function influences the observed shape of the MZR. According to ‘gas-regulator models’, chemical enrichment in galaxies is set by an equilibrium between gas accretion, star formation, and outflows (Finlator & Davé 2008; Peebles & Shankar 2011; Davé et al. 2012; Lilly et al. 2013). Under such equilibrium, inflow of low-metallicity gas dilutes the ISM and triggers star formation. The star formation consumes the metal-poor gas and enriches the ISM, increasing the metallicity until new gas is accreted. This gives rise to the anti-correlated behaviour of the gas-phase metallicity and sSFR that has been observed by multiple observational studies in the local Universe as well in the early Universe (e.g. Mannucci et al. 2010; Lara-López et al. 2010; Sanders et al. 2018). This motivates the first case, the ‘normal’ FMR case, in our toy models. The parameter values for this case were taken from Andrews & Martini (2013), considering their 2D projection of the FMR. The scatter on Z_{gas} comes from the scatter in the star-forming main sequence and strength of S_{coeff} (Eq. (5)).

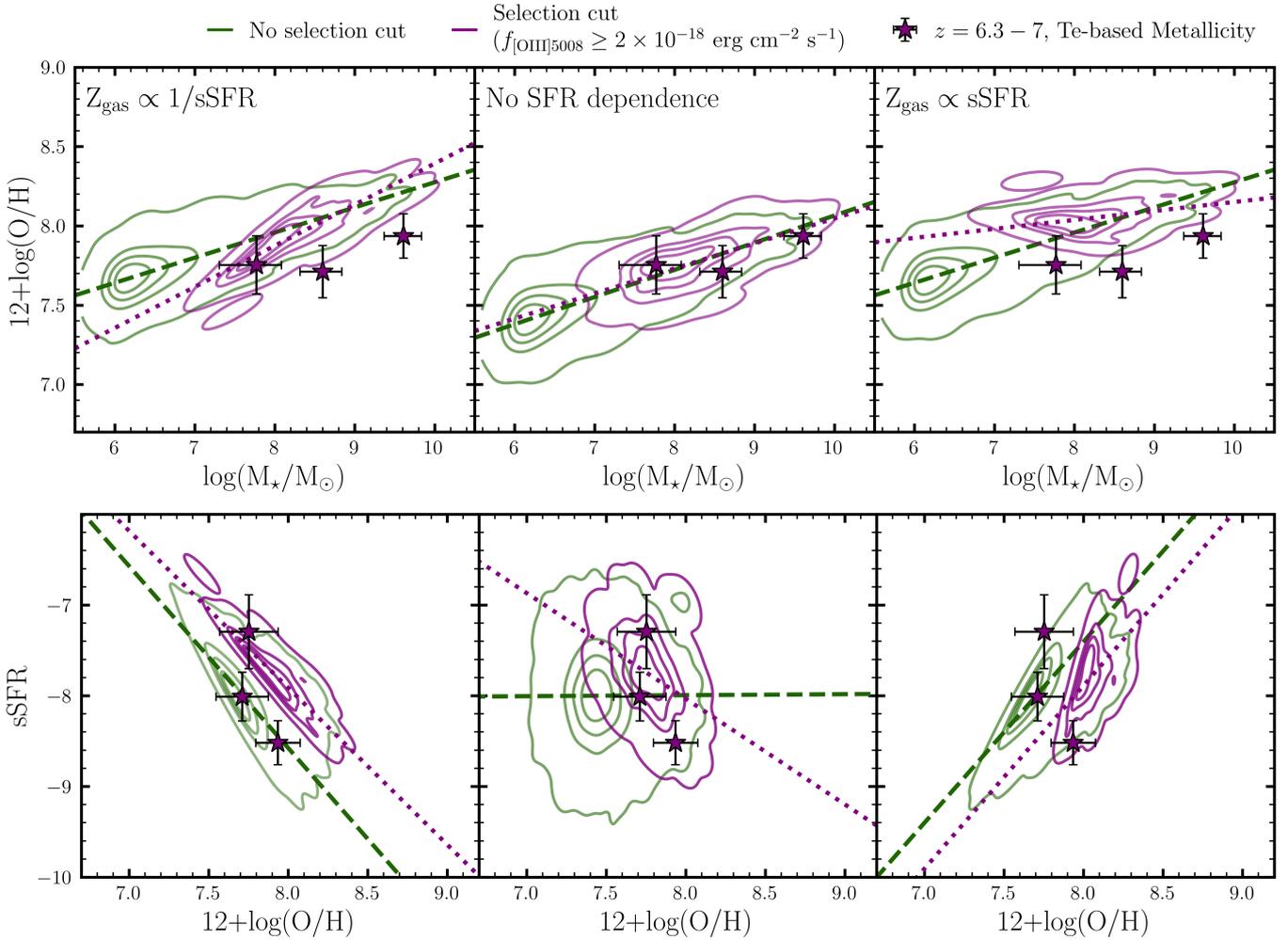


Fig. 9. Top: MZR for the three cases in our toy model. We also plot our direct- T_e metallicity estimates. Bottom: Specific star formation rate (sSFR) as a function of gas-phase metallicity. We show sSFR vs. gas-phase metallicity for our $H\gamma$ stellar mass binned stacks. The error bars on the sSFR were calculated using the 16th and 84th percentile in the stacks. The top left and bottom left panels show Case 1 (‘normal FMR’). Our selection cut makes the MZR slope steeper. The top middle and bottom middle panels show Case 2 (‘no FMR’). Our selection cut does not affect the MZR slope. The top right and bottom right panels show Case 3 (‘inverted FMR’). Our selection cut makes the MZR slope appear flatter. In all three cases, we find a negative correlation between sSFR and $12+\log(\text{O}/\text{H})$ upon implementing our selection cut.

However, it is unclear whether the normal FMR evolves out to high redshift as multiple studies have also found discrepancies in their observations (Zahid et al. 2014; Heintz et al. 2023; Curti et al. 2024; Scholte et al. 2025; Cullen et al. 2025). In a recent study, Laseter et al. (2025) find no evidence of the existence of the normal FMR at $z \sim 0$ for the low- M_* regime, indicating a deviation from the equilibrium conditions as seen at the higher- M_* regime. While the gas-regulator models explain the existence of FMR, there are many caveats. For example, supernovae outflows are the only feedback assumed in the models. Additionally, these models assume time-invariant star-forming efficiencies, which may not be true for galaxies with bursty SFHs. Importantly, the 3D relation is not just set by the strengths of processes such as the mass inflow rate, SFR, mass loading factor, ISM metallicity, effective yields, and star forming efficiency, but also by the timescales of their variations (Torrey et al. 2018; Wang & Lilly 2021). Torrey et al. (2018) show that if the timescales of variations in the SFR and gas-phase metallicity are within a factor of ~ 1.5 of each other in the TNG100 simulation, then the normal FMR case should hold. However, if either the SFR or metallicity evolves much faster than the

other it can either wash out any correlation between them or cause a positive correlation. This motivates our second case (no FMR) and third case (inverted FMR). Processes such as bursty star formation can significantly impact the variability of the SFR, while not impacting the metallicity much. Multiple studies have shown evidence of bursty star formation for high-redshift galaxies (Topping et al. 2022; Endsley et al. 2023; Looser et al. 2025). Additionally, variability in processes such as a temporary increase in the SFE due to gravitational instabilities, or an increase in outflows with the same metallicity as the ISM, can allow for a positive correlation between the sSFR and gas-phase metallicity (Torrey et al. 2018; Wang & Lilly 2021). Indications of an ‘inverted’ FMR have also been seen in the THESAN-ZOOM simulation for galaxies with $\log(M_*/M_\odot) \leq 9$ due to regular inflow of low-metallicity gas (McClymont et al. 2025). The parameters for the second case (No FMR) were taken from the MZR presented in Curti et al. (2024). We assumed a scatter of 0.1 dex on the MZR. For the third case, we obtained the parameters for Eq. (6) by matching the shape of the intrinsic MZR to the one obtained considering the normal FMR case. We kept the S_{coeff} to be the same magnitude as in the normal FMR case and

only changed the sign. In this case, we find $M_{\text{coeff}} < 0$. As per Eq. (6), this does not imply a negative correlation between metallicity and stellar mass as SFR is also a function of stellar mass (see Appendix D). However, this does point towards processes such as bursty star formation and inflow of pristine gas being more dominant compared to the stellar mass at high redshift.

In the top panel of Fig. 9, we show the distribution of simulated galaxies on the mass-metallicity plane for the three cases mentioned above. We apply our selection cut to the [O III] λ 5008 luminosity. We then compare how the MZR of the selected sample, i.e. the observable MZR, compares to the intrinsic one. The selection cut affects the three cases differently. In case 1 (normal FMR), the observable MZR is steeper than the intrinsic MZR. In this case, $Z_{\text{gas}} \propto 1/\text{sSFR}$, an [O III] flux cut therefore selects galaxies with a lower metallicity. This makes the slope of the observable MZR steeper. In case 2 (No FMR), there is negligible change in the intrinsic MZR and the observable MZR slopes. The small change in the slope can be attributed to the scatter added to the MZR. In case 3 (inverted FMR), the observable MZR is flatter than the intrinsic MZR. In this case, $Z_{\text{gas}} \propto \text{sSFR}$, the [O III] flux thus scales with SFR, and we end up selecting galaxies with higher metallicities with an [O III] flux cut. This makes the slope of the observable MZR flatter.

With the data currently available, it is not possible to determine which scenario our observations follow as there are multiple degeneracies in the free parameters while evaluating the metallicities using Eq. (6) (we discuss these degeneracies in Appendix D). We attempted to test which scenario explains our observations by studying the sSFR as a function of the gas-phase metallicity, as is shown in the bottom panel of Fig. 9. We see a negative correlation between the sSFR and gas-phase metallicity for both the ‘intrinsic’ and ‘observable’ distribution in case 1. In case 2, there is no correlation between the sSFR and gas-phase metallicity for the intrinsic distribution; however, on application of our selection cut we find a negative correlation. In case 3, both the intrinsic and observable distributions have a positive correlation. We find a negative correlation in our observations, the slope of which is best represented by the ‘no FMR’ case. This suggests an absent or weak ($S_{\text{coeff}} = 0$ or $|S_{\text{coeff}}| \ll 1$ but $\neq 0$) FMR at $z \sim 6$, which may allude to the galaxies at these redshifts being out of equilibrium, as has been suggested previously for high-redshift galaxies (Maiolino & Mannucci 2019; Bordoloi et al. 2024; Danhaive et al. 2025). In addition to the MZR parameters from Curti et al. (2024) for case 2, we also explored the impact of using parameters from other MZRs in the literature (Heintz et al. 2023; Sarkar et al. 2025, see our Appendix E). We find that our observations generally agree with the MZR models in the sSFR versus gas-phase metallicity plane. If this is the true underlying scenario, then our intrinsic slope is as flat as our observed slope, which points towards either rapid metal enrichment in low-mass galaxies or lower metallicities due to pristine gas inflows in higher-mass galaxies. Previous observations of high- z galaxies are not in favour of lower metallicities for high-stellar-mass galaxies (Rowland et al. 2025), suggesting the former explanation. Observations of CGM galaxies at $z \sim 6$ also indicate that galaxies in the early Universe were efficient in chemically enriching their surroundings (Bordoloi et al. 2024).

In this case, it is important to note that a good fit between the model and the data does not necessarily indicate that the intrinsic trend is represented by the model. Figure 9 shows that the scatter in the sSFR-metallicity relation varies across models. Additionally, Fig. E.1 shows that scatter even varies in the sSFR-metallicity plane for the same model (case 2, $S_{\text{coeff}} = 0$) with different parameters. Therefore to fit the parameters and distin-

Table 3. Parameter choices for the toy models investigated in Sect. 6.2 (Eq. (6)).

Models	Z_{norm}	M_{coeff}	S_{coeff}	Comments
Case 1: $Z \propto 1/\text{sSFR}$	4.2	0.47	-0.3102	Values adopted from Andrews & Martini (2013) FMR
Case 2: No SFR dependence	6.39	0.17	-	Values adopted from Curti et al. (2024). We consider a 0.1 dex scatter along the relation
Case 3: $Z \propto \text{sSFR}$	9.15	-0.15	0.3102	

guish between the models, we need to fit the scatter. Due to the stacked measurements of the metallicity, we cannot measure the scatter in the sSFR – metallicity relation, and therefore cannot fully constrain the SFR dependence. To do this we require observations with robust metallicity measurements for a statistically large sample of individual galaxies with a well-defined selection function, as this crucially impacts the observed relations.

7. Summary

In this work, we have studied and characterised the MZR and the 3D correlation between the stellar mass, gas-phase metallicity, and SFR, dubbed the FMR, of [O III] selected galaxies with redshifts between $5 < z < 7$. We obtained the spectra for these galaxies using deep NIRCcam WFSS surveys, EIGER (Kashino et al. 2023), COLA1 (Torralba-Torregrosa et al. 2024), and ALT (Naidu et al. 2024). Grism spectra have a well-understood flux-limited selection function and are not subject to slit losses. To make the selection uniform across the data from the three surveys uniform, we made a conservative [O III] flux cut, $f_{\text{OIII}\lambda 5008} \geq 2 \times 10^{-18} \text{ erg s}^{-2} \text{ cm}^{-2}$. Since grism spectra are less sensitive and have a smaller wavelength coverage compared to NIRSpec spectra, we performed stacking in stellar mass bins to detect faint emission lines such as [O III] λ 4364 and H γ with a S/N > 3 (see Fig. 3). This allowed us to derive direct- T_e -based metallicities for our analysis. We tested the impact of our selection function using hydrodynamical simulations such as FLARES (Lovell et al. 2020; Vijayan et al. 2020) and SPHINX²⁰ (Rosdahl et al. 2018; Katz et al. 2023) and the JAGUAR catalogue (Williams et al. 2018). We further tested our selection function using simple toy models for three scenarios: (1) normal FMR ($Z_{\text{gas}} \propto 1/\text{sSFR}$), (2) no FMR (no SFR dependence), and (3) inverted FMR ($Z_{\text{gas}} \propto \text{sSFR}$). Our main results are:

1. We cover H γ and [O III] λ 4363 for a sample of 270 galaxies (i.e. the H γ sample). We detect H γ , [O III] λ 4363, H β and [O III] λ 4960, 5008 when stacking in bins of stellar mass, which allows us to measure direct- T_e metallicities in the stacks. The observed stellar masses for this sample span a range of $\log(M_{\star}/M_{\odot}) = 6.9\text{--}10.5$, with the gas-phase metallicities ranging between $Z_{\text{gas}} = 7.7\text{--}8.0$ ($0.11\text{--}0.22 Z_{\odot}$) [Sect. 4.1, Fig. 6, Table 1].
2. With the measured direct- T_e metallicities, we tested the R3 calibration curve for high- z galaxies. We found that our direct- T_e measurements are slightly offset to higher values of the R3 ratio at fixed metallicity compared to the R3 calibration curves from Sanders et al. (2024) and Nakajima et al. (2022). However, within the observational errors and the

errors on the calibration itself, our values agree with these curves [Sect. 4.2, Fig. 5].

3. We obtain a linear MZR relation with normalisation $Z_{\text{norm}} = 7.96 \pm 0.08$ and slope $\gamma = 0.12 \pm 0.08$ for the direct- T_e measurements from the Hy sample [Sect. 5, Eq. (4)]. The MZR lies systematically below the MZR measured for galaxies in the local Universe (Scholte et al. 2024) and at $z \sim 2\text{--}3$ (Sanders et al. 2021; Li et al. 2023). The slope of our MZR agrees well with those found for analogues at lower redshift (Yang et al. 2017a,b) and with other $z = 6\text{--}10$ measurements (Curti et al. 2024; see our Sect. 5, Fig. 6, and Table B.1).
4. We find that our measurements do not match the slope and normalisation of the MZR of galaxies in the FLARES and SPHINX²⁰ simulations and the JAGUAR model. The observed slope is significantly shallower, which is robust to our selection effects. Selection effects impact simulations non-uniformly. In SPHINX²⁰ and JAGUAR, on applying the selection cut, the MZR becomes flatter. For FLARES, there is no change in the MZR upon application of the selection cut (Sect. 6.1, Fig. 8, and Table B.1).
5. Through simple toy models, we show that the effect of our [O III] selection cut on the MZR depends on the relation between gas-phase metallicity and sSFR of the galaxies. In the case of a normal FMR, the observable MZR (after the selection cut) is steeper than the intrinsic MZR. In the case of no FMR, i.e. no SFR dependence, there is negligible change in the intrinsic MZR and the observable MZR slopes. In the case of an inverted FMR, the observable MZR is flatter than the intrinsic MZR [Sect. 6.2, Fig. 9, Table 3]. Our data agrees well with the no FMR case. This suggests an absent or weak FMR at $z \sim 6$, which is similar to what we find in FLARES. An absent or weak FMR points towards the possibility that the galaxies at these redshifts are not yet in equilibrium. A no FMR case would imply that the intrinsic slope of the MZR is as flat as the observed one. Such flatness may be due to rapid metal enrichment [Sect. 6.2, Fig. 9]. However, since our data points were obtained through stacking, they are not enough to constrain the scatter in the three cases and therefore distinguish cleanly among them.

This work highlights the importance of accounting for selection functions in order to understand the physics underlying scaling relations such as the MZR and the FMR. In order to understand the baryon cycle in depth at high redshifts, future observations with robust metallicity measurements are required for statistically large well-defined samples of galaxies.

Acknowledgements. We thank the anonymous referee for the insightful comments that helped improving this paper. 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 Associations of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. These observations were taken under programmes # 1243, # 1933 and # 3516. Funded by the European Union (ERC, AGENTS, 101076224). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. GK acknowledges support from the Foundation MERAC. APV acknowledge support from the Sussex Astronomy Centre STFC Consolidated Grant (ST/X001040/1).

References

- Andrews, B. H., & Martini, P. 2013, *ApJ*, 765, 140
 Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2024, *ApJ*, 974, 92
 Bian, F., Kewley, L. J., & Dopita, M. A. 2018, *ApJ*, 859, 175
 Bordoloi, R., Simcoe, R. A., Matthee, J., et al. 2024, *ApJ*, 963, 28

- Bothwell, M. S., Maiolino, R., Kennicutt, R., et al. 2013, *MNRAS*, 433, 1425
 Bothwell, M. S., Maiolino, R., Ciccone, C., Peng, Y., & Wagg, J. 2016, *A&A*, 595, A48
 Bradley, L., Sipőcz, B., Robitaille, T., et al. 2024, <https://doi.org/10.5281/zenodo.13989456>
 Brooks, M., Simons, R. C., Trump, J. R., et al. 2025, *ApJ*, 986, 177
 Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023, *A&A*, 677, A115
 Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, 345, 245
 Cataldi, E., Belfiore, F., Curti, M., et al. 2025, *A&A*, 703, A208
 Chabrier, G. 2003, *PASP*, 115, 763
 Chakraborty, P., Sarkar, A., Smith, R., et al. 2025, *ApJ*, 985, 24
 Chemerynska, I., Atek, H., Dayal, P., et al. 2024, *ApJ*, 976, L15
 Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102
 Conroy, C., & Gunn, J. E. 2010, *ApJ*, 712, 833
 Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, 699, 486
 Crain, R. A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, 450, 1937
 Cresci, G., Mannucci, F., Sommariva, V., et al. 2012, *MNRAS*, 421, 262
 Cullen, F., Carnall, A. C., Scholte, D., et al. 2025, *MNRAS*, 540, 2176
 Curti, M. 2025, arXiv e-prints [arXiv:2504.08933]
 Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020, *MNRAS*, 491, 944
 Curti, M., Maiolino, R., Curtis-Lake, E., et al. 2024, *A&A*, 684, A75
 Danhaive, A. L., Tacchella, S., Übler, H., et al. 2025, *MNRAS*, 543, 3249
 Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, *MNRAS*, 421, 98
 Dayal, P., Ferrara, A., & Dunlop, J. S. 2013, *MNRAS*, 430, 2891
 De Rossi, M. E., Bower, R. G., Font, A. S., Schaye, J., & Theuns, T. 2017, *MNRAS*, 472, 3354
 Di Cesare, C., Matthee, J., Naidu, R. P., et al. 2025, *A&A*, accepted [arXiv:2510.19044]
 Dotter, A. 2016, *ApJS*, 222, 8
 Ellison, S. L., Patton, D. R., Simard, L., & McConnell, A. W. 2008, *ApJ*, 672, L107
 Endsley, R., Stark, D. P., Whitler, L., et al. 2023, *MNRAS*, 524, 2312
 Faisst, A. L., Liu, L.-J., Dubois, Y., et al. 2025, arXiv e-prints [arXiv:2510.16106]
 Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, *PASP*, 110, 761
 Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, *Rev. Mex. Astron. Astrofis.*, 49, 137
 Finlator, K., & Davé, R. 2008, *MNRAS*, 385, 2181
 Furtak, L. J., Zitrin, A., Weaver, J. R., et al. 2023, *MNRAS*, 523, 4568
 Hahn, C., Wilson, M. J., Ruiz-Macias, O., et al. 2023, *AJ*, 165, 253
 He, X., Wang, X., Jones, T., et al. 2024, *ApJ*, 960, L13
 Heintz, K. E., Brammer, G. B., Giménez-Arteaga, C., et al. 2023, *Nat. Astron.*, 7, 1517
 Hirschmann, M., Charlot, S., & Somerville, R. S. 2023, *MNRAS*, 526, 3504
 Home, K. 1986, *PASP*, 98, 609
 Isobe, Y., Ouchi, M., Nakajima, K., et al. 2023, *ApJ*, 956, 139
 Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, *A&A*, 448, 955
 Johnson, B. D., Leja, J., Conroy, C., & Speagle, J. S. 2021, *ApJS*, 254, 22
 Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, *ApJ*, 835, 88
 Kashino, D., Lilly, S. J., Matthee, J., et al. 2023, *ApJ*, 950, 66
 Kashino, D., Lilly, S. J., Matthee, J., et al. 2025, arXiv e-prints [arXiv:2506.03121]
 Katz, H., Rosdahl, J., Kimm, T., et al. 2023, *Open J. Astrophys.*, 6, 44
 Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, *ARA&A*, 57, 511
 Korhonen Cuestas, N. A., Strom, A. L., Miller, T. B., et al. 2025, *ApJ*, 984, 188
 Kramarenko, I. G., Rosdahl, J., Blaizot, J., et al. 2025, *A&A*, submitted [arXiv:2509.05403]
 Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, 218, 15
 Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, *A&A*, 521, L53
 Laseter, I. H., Maseda, M. V., Curti, M., et al. 2024, *A&A*, 681, A70
 Laseter, I. H., Maseda, M. V., Bunker, A. J., et al. 2025, arXiv e-prints [arXiv:2510.15024]
 Leja, J., van Dokkum, P. G., Franx, M., & Whitaker, K. E. 2015, *ApJ*, 798, 115
 Li, M., Cai, Z., Bian, F., et al. 2023, *ApJ*, 955, L18
 Li, Z., Kakiichi, K., Christensen, L., et al. 2025, *A&A*, 703, A106
 Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, *ApJ*, 772, 119
 Looser, T. J., D'Eugenio, F., Maiolino, R., et al. 2025, *A&A*, 697, A88
 Lovell, C. C., Vijayan, A. P., Thomas, P. A., et al. 2020, *MNRAS*, 500, 2127
 Luridiana, V., Morisset, C., & Shaw, R. A. 2013, Astrophysics Source Code Library [record ascl:1304.021]
 Ma, C., Wang, K., Wang, E., et al. 2024, *ApJ*, 971, L14
 Mahabal, A., Stern, D., Bogosavljević, M., Djorgovski, S. G., & Thompson, D. 2005, *ApJ*, 634, L9
 Maier, C., Ziegler, B. L., Lilly, S. J., et al. 2015, *A&A*, 577, A14
 Maiolino, R., & Mannucci, F. 2019, *A&ARv*, 27, 3
 Maiolino, R., Nagao, T., Grazian, A., et al. 2008, *A&A*, 488, 463

- Mannucci, F., Cresci, G., Maiolino, R., et al. 2009, *MNRAS*, **398**, 1915
- Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, *MNRAS*, **408**, 2115
- Marino, R. A., Rosales-Ortega, F. F., Sánchez, S. F., et al. 2013, *A&A*, **559**, A114
- Matthee, J., Sobral, D., Gronke, M., et al. 2018, *A&A*, **619**, A136
- Matthee, J., Mackenzie, R., Simcoe, R. A., et al. 2023, *ApJ*, **950**, 67
- Matthee, J., Naidu, R. P., Brammer, G., et al. 2024, *ApJ*, **963**, 129
- McClymont, W., Tacchella, S., Smith, A., et al. 2025, arXiv e-prints [arXiv:2507.08787]
- Møller, P., Fynbo, J. P. U., Ledoux, C., & Nilsson, K. K. 2013, *MNRAS*, **430**, 2680
- Morishita, T., Stiavelli, M., Grillo, C., et al. 2024, *ApJ*, **971**, 43
- Moustakas, J., Zaritsky, D., Brown, M., et al. 2011, arXiv e-prints [arXiv:1112.3300]
- Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, *ApJ*, **940**, L14
- Naidu, R. P., Matthee, J., Kramarenko, I., et al. 2024, arXiv e-prints [arXiv:2410.01874]
- Nakajima, K., Ouchi, M., Xu, Y., et al. 2022, *ApJS*, **262**, 3
- Nakajima, K., Ouchi, M., Isole, Y., et al. 2023, *ApJS*, **269**, 33
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, **266**, 713
- Osterbrock, D. E., & Ferland, G. J. 2006, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*
- Peebles, M. S., & Shankar, F. 2011, *MNRAS*, **417**, 2962
- Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, **348**, L59
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Pollock, C. L., Gottumukkala, R., Heintz, K. E., et al. 2025, A&A, submitted [arXiv:2506.15779]
- Popesso, P., Concas, A., Cresci, G., et al. 2023, *MNRAS*, **519**, 1526
- Price, S. H., Bezanson, R., Labbe, I., et al. 2025, *ApJ*, **982**, 51
- Roberts-Borsani, G., Treu, T., Shapley, A., et al. 2024, *ApJ*, **976**, 193
- Rosdahl, J., Katz, H., Blaizot, J., et al. 2018, *MNRAS*, **479**, 994
- Rowland, L. E., Stefanon, M., Bouwens, R., et al. 2025, arXiv e-prints [arXiv:2501.10559]
- Salim, S., Lee, J. C., Davé, R., & Dickinson, M. 2015, *ApJ*, **808**, 25
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, *ApJ*, **816**, 23
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2018, *ApJ*, **858**, 99
- Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, *ApJ*, **914**, 19
- Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, *ApJ*, **962**, 24
- Sarkar, A., Chakraborty, P., Vogelsberger, M., et al. 2025, *ApJ*, **978**, 136
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Scholte, D., Saintonge, A., Moustakas, J., et al. 2024, *MNRAS*, **535**, 2341
- Scholte, D., Cullen, F., Carnall, A. C., et al. 2025, *MNRAS*, **540**, 1800
- Shapley, A. E., Coil, A. L., Ma, C.-P., & Bundy, K. 2005, *ApJ*, **635**, 1006
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, **214**, 15
- Suess, K. A., Weaver, J. R., Price, S. H., et al. 2024, *ApJ*, **976**, 101
- Tacchella, S., Finkelstein, S. L., Bagley, M., et al. 2022, *ApJ*, **927**, 170
- Tacchella, S., Johnson, B. D., Robertson, B. E., et al. 2023, *MNRAS*, **522**, 6236
- Theios, R. L., Steidel, C. C., Strom, A. L., et al. 2019, *ApJ*, **871**, 128
- Topping, M. W., Shapley, A. E., Sanders, R. L., et al. 2021, *MNRAS*, **506**, 1237
- Topping, M. W., Stark, D. P., Endsley, R., et al. 2022, *ApJ*, **941**, 153
- Topping, M. W., Sanders, R. L., Shapley, A. E., et al. 2025, *MNRAS*, **541**, 1707
- Torralba-Torregrosa, A., Matthee, J., Naidu, R. P., et al. 2024, *A&A*, **689**, A44
- Torrey, P., Vogelsberger, M., Hernquist, L., et al. 2018, *MNRAS*, **477**, L16
- Torrey, P., Vogelsberger, M., Marinacci, F., et al. 2019, *MNRAS*, **484**, 5587
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, **613**, 898
- Vijayan, A. P., Lovell, C. C., Wilkins, S. M., et al. 2020, *MNRAS*, **501**, 3289
- Wang, E., & Lilly, S. J. 2021, *ApJ*, **910**, 137
- Weibel, A., Oesch, P. A., Barrufet, L., et al. 2024, *MNRAS*, **533**, 1808
- Whitler, L., Stark, D. P., Endsley, R., et al. 2023, *MNRAS*, **519**, 5859
- Williams, C. C., Curtis-Lake, E., Hainline, K. N., et al. 2018, *ApJS*, **236**, 33
- Wuyts, E., Kurk, J., Förster Schreiber, N. M., et al. 2014, *ApJ*, **789**, L40
- Yabe, K., Ohta, K., Akiyama, M., et al. 2015, *PASJ*, **67**, 102
- Yang, H., Malhotra, S., Gronke, M., et al. 2017a, *ApJ*, **844**, 171
- Yang, H., Malhotra, S., Rhoads, J. E., & Wang, J. 2017b, *ApJ*, **847**, 38
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014, *ApJ*, **792**, 75

Appendix A: Metallicity estimates for H β sample using R3-calibrations

We report the gas-phase metallicity estimates for our stellar mass bin stacked H β sample using the R3-calibrations presented in Nakajima et al. (2022) and Sanders et al. (2024) in Table A.1. We use R3-calibration presented for local galaxies with $\text{EWH}\beta \geq 200$ from Nakajima et al. (2022). The gas-phase metallicities have been estimated using the same methodology as described in Sect. 4.2.

Appendix B: MZR parameters for observations and simulations

We report the slope and normalisation for the MZR for our direct- T_e measurements along with the observational samples and simulations, following Eq. 4. For the simulations, we present the MZR parameters for a sample with the same selection cut as our observations (i.e. an [O III]-flux cut) and for a sample without any selection cut.

Appendix C: The relation between [O III] λ 5008 luminosity and stellar mass

Before assessing the impact of our line-flux limited selection function on the mass - metallicity relation, it is important to compare the relation between the [O III] line luminosity and stellar mass among simulations and observations. In Fig. C.1, we show [O III] λ 5008 luminosity as a function of stellar masses for H β and H γ sample, as well as for the hydrodynamical simulations SPHINX²⁰ and FLARES, and the JAGUAR catalogue. Upon applying our selection cut to these simulations, we find that the FLARES simulation and the JAGUAR catalogue match our relation between stellar mass and [O III] luminosities very well. The intrinsic relations are significantly steeper than observed, but this can be understood as being due to the selection cut. In SPHINX²⁰ however, only galaxies above stellar mass of $10^{8.4}M_\odot$ are selected with the [O III] flux threshold. As mentioned in Sect. 6.1, SPHINX²⁰ has an elevated stellar mass - halo mass relation and therefore assigns too low SFRs at fixed mass, leading to much lower [O III] luminosities at all masses.

Appendix D: Derivation of the FMR equation used in our toy model

In Sect. 6.2, we describe our toy models used to understand the impact of the 3D relation between metallicity-stellar mass-SFR on the shape of the observed MZR. We described the 3D relation as

$$Z = Z_0 + M_{\text{coeff}} \times \log_{10}(M_\star/M_\odot) + S_{\text{coeff}} \times \log_{10}(\text{SFR}/M_\odot \text{yr}^{-1}). \quad (\text{D.1})$$

This can be expanded as

$$Z = Z_0 + M_{\text{coeff}} \times \log_{10}(M_\star/M_\odot) + S_{\text{coeff}} \times \left(\text{MS}_{\text{norm}} + \text{MS}_{\text{slope}} \times \log_{10} \left(\frac{M_\star}{10^{10} M_\odot} \right) + \mathcal{N}(0, \sigma_{\text{SFR}}) \right). \quad (\text{D.2})$$

Therefore,

$$Z = Z_0 + (M_{\text{coeff}} + S_{\text{coeff}} \times \text{MS}_{\text{slope}}) \times \log_{10}(M_\star/M_\odot) + S_{\text{coeff}} \times (\text{MS}_{\text{norm}} - 10 \times \text{MS}_{\text{slope}}) + S_{\text{coeff}} \times \mathcal{N}(0, \sigma_{\text{SFR}}). \quad (\text{D.3})$$

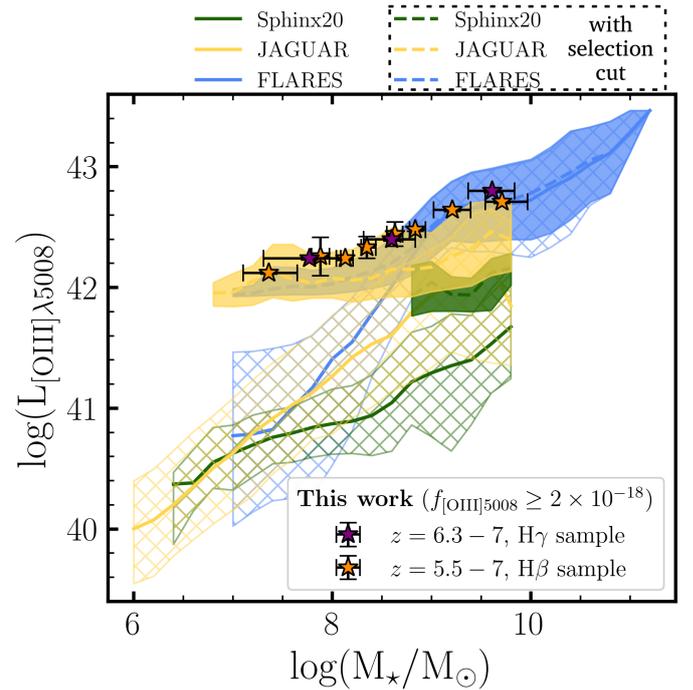


Fig. C.1. [O III] λ 5008 luminosities as a function of stellar mass for the SPHINX²⁰ (Katz et al. 2023) and FLARES simulation (Lovell et al. 2020; Vijayan et al. 2020), the JAGUAR catalogue (Williams et al. 2018). We plot both, the intrinsic relation and the relation after mimicking our selection cut. We also show the $L_{[\text{O III}]\lambda 5008} - \log(M_\star/M_\odot)$ relation for our stacked H γ (purple stars) and H β (orange stars) sample.

Here, we see that in the cases of $S_{\text{coeff}} \neq 0$, the slope of the MZR regulated by parameters which govern the influence of both stellar mass and SFR. The normalisation is dependent on the intrinsic metallicity floor and parameters governing the SFR. By design, the scatter is solely dependent in the parameters governing SFR. Therefore, in the presence of a 3D relation between stellar mass, metallicity and SFR, the shape of the MZR is governed by degenerate parameters. This makes it difficult to use Bayesian forward modelling techniques with our stacked measurements to obtain the intrinsic MZR.

Appendix E: Effect of the selection function on MZR models

We show the impact of our selection function on MZR parameters from Heintz et al. (2023) and Sarkar et al. (2025) in addition to the MZR parameters from Curti et al. (2024) for case 2 ($S_{\text{coeff}} = 0$) in Fig. E.1. The scatter of the MZR is fixed to 0.1 dex for all three cases. Note that the scatter in the sSFR vs. $12 + \log(\text{O}/\text{H})$ plane is different in the three cases even after fixing the scatter in the MZR plane. We find that our observations are in agreement with the ‘observed’ contours on the sSFR vs. $12 + \log(\text{O}/\text{H})$ plane despite the different parameters of the MZR. It is difficult to distinguish between the models because we only have three data points obtained our stacking analysis. To make a distinction, we require a large sample of individual measurements with an [O III] flux cut selection to constrain the slope and the scatter.

Table A.1. Gas-phase metallicity estimates in stellar mass bins for the H β ($5.5 < z < 7$) sample using R3 calibration.

Mass Bin	$\log(M_*/M_\odot)$	Sanders et al. (2024)		Nakajima et al. (2022)	
		12 + log(O/H) (lower branch)	12 + log(O/H) (higher branch)	12 + log(O/H) (lower branch)	12 + log(O/H) (higher branch)
$6.68 \leq \log(M_*/M_\odot) < 7.8$	$7.36^{+0.29}_{-0.26}$	$7.67^{+0.09}_{-0.07}$	$8.17^{+0.07}_{-0.09}$	$7.72^{+0.16}_{-0.18}$	$8.40^{+0.18}_{-0.16}$
$7.8 < \log(M_*/M_\odot) < 8.0$	$7.88^{+0.09}_{-0.09}$	$7.66^{+0.08}_{-0.07}$	$8.18^{+0.07}_{-0.08}$	$7.68^{+0.19}_{-0.14}$	$8.44^{+0.14}_{-0.19}$
$8.0 < \log(M_*/M_\odot) < 8.25$	$8.13^{+0.08}_{-0.09}$	$7.44^{+0.04}_{-0.04}$	$8.40^{+0.04}_{-0.04}$	$7.60^{+0.16}_{-0.18}$	$8.52^{+0.18}_{-0.16}$
$8.25 < \log(M_*/M_\odot) < 8.5$	$8.35^{+0.09}_{-0.06}$	$7.77^{+0.08}_{-0.07}$	$8.07^{+0.07}_{-0.08}$	$7.74^{+0.15}_{-0.16}$	$8.38^{+0.16}_{-0.15}$
$8.5 < \log(M_*/M_\odot) < 8.75$	$8.63^{+0.08}_{-0.08}$	$7.62^{+0.06}_{-0.05}$	$8.22^{+0.05}_{-0.06}$	$7.68^{+0.16}_{-0.17}$	$8.44^{+0.17}_{-0.16}$
$8.75 < \log(M_*/M_\odot) < 9.0$	$8.84^{+0.11}_{-0.06}$	$7.62^{+0.06}_{-0.05}$	$8.22^{+0.05}_{-0.06}$	$7.70^{+0.15}_{-0.19}$	$8.42^{+0.19}_{-0.15}$
$9.0 < \log(M_*/M_\odot) < 9.5$	$9.21^{+0.18}_{-0.19}$	$7.75^{+0.07}_{-0.07}$	$8.09^{+0.07}_{-0.07}$	$7.75^{+0.13}_{-0.17}$	$8.37^{+0.17}_{-0.13}$
$9.5 < \log(M_*/M_\odot) \leq 11.1$	$9.71^{+0.26}_{-0.17}$	$7.57^{+0.05}_{-0.04}$	$8.27^{+0.04}_{-0.05}$	$7.66^{+0.17}_{-0.17}$	$8.45^{+0.17}_{-0.17}$

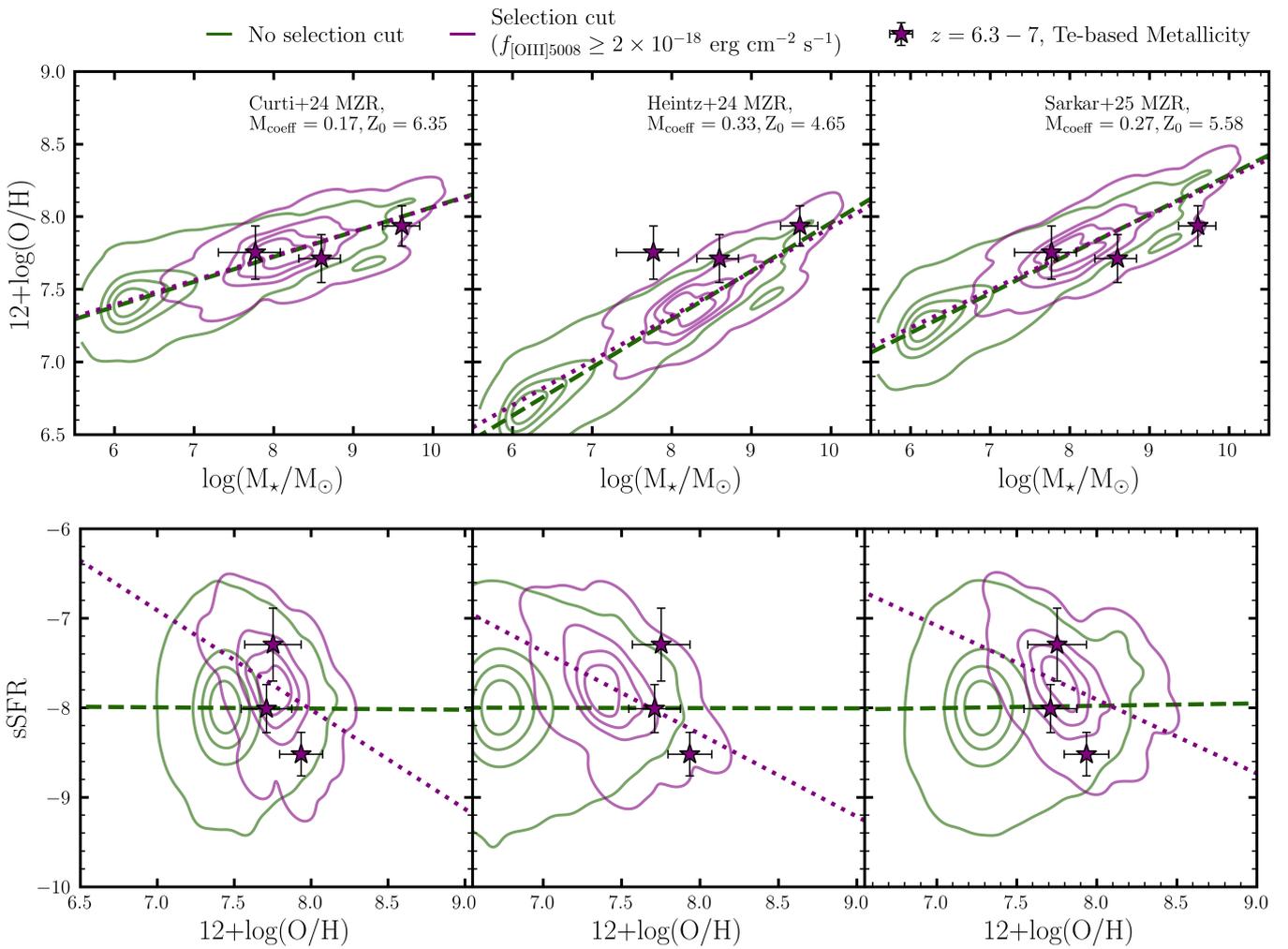


Fig. E.1. Top: The MZR for the ‘No SFR dependence’ case toy model with three different MZR parameters. We also plot our direct- T_e metallicity estimates. Bottom: Specific star formation rate (sSFR) as a function of gas-phase metallicity. We show sSFR vs. gas-phase metallicity for our Hy stellar mass binned stacks. The error bars on the sSFR are calculated using the 16th and 84th percentile in the stacks. The panels on the left are following the parameters by Curti et al. (2024), the middle panels follow the parameters by Heintz et al. (2023) and the panels on the right follow parameters by Sarkar et al. (2025) for Eq. 4.

Table B.1. Slope of the MZR in various observational samples and in simulations.

Study	Redshift	$[\log(M_{\star}/M_{\odot})]$	γ	Z_{norm}	Description
This work (direct-Te)	6.3-7	[6.89, 10.54]	0.12 ± 0.08	7.96 ± 0.10	$f_{[\text{O III}]\lambda 5008}$ selected galaxies from NIR-Cam/WFSS (EIGER, COLA1 and ALT surveys)
Greenpeas (Yang et al. 2017a) and Blueberries (Yang et al. 2017b)	$z < 0.2$	[6.6, 9.84]	0.17 ± 0.01	8.19 ± 0.03	Analogues for high- z galaxies because of their compact size, high equivalent width emission lines, high ionisation parameters and low metallicities.
Scholte et al. (2024) ^a	< 0.6	[6.4, 11.05]	0.23 ± 0.003	8.70 ± 0.003	Sample from the DESI BGS validation and Year 1 Main Survey Data (Hahn et al. 2023)
Sanders et al. (2021)	~ 0.08	[8.5, 11.5]	0.28 ± 0.01	8.77 ± 0.01	$z \sim 0$ sample from SDSS (York et al. 2000, Andrews & Martini 2013).
	2.3	[9.3, 10.7]	0.30 ± 0.02	8.51 ± 0.02	$z \sim 2 - 3$ star-forming galaxies from MOSDEF survey (Kriek et al. 2015).
	3.3	[9.2, 10.6]	0.29 ± 0.02	8.41 ± 0.03	
Li et al. (2023)	~ 2	[6.52, 9.45]	0.13 ± 0.04	8.31 ± 0.04	JWST/NIRISS slitless grism spectroscopic observations in the Abell 2744 and SMACS J0723-3732 fields.
	~ 3	[7.39, 9.68]	0.16 ± 0.03	8.24 ± 0.03	Uses the O32 calibration from Bian et al. (2018).
He et al. (2024)	1.90	[6.9, 9.9]	0.223 ± 0.017	8.569 ± 0.036	JWST/NIRISS slitless grism spectroscopic observations in the Abell 2744 field.
	2.88	[7.1, 10.0]	0.294 ± 0.010	8.596 ± 0.024	Uses calibrations from Bian et al. (2018).
Nakajima et al. (2023)	4 – 10	[7.39, 9.48]	0.25 ± 0.03	8.24 ± 0.05	JWST/NIRSpec data from ERO, GLASS and CEERS programmes
Curti et al. (2024)	3 – 10	[6, 10]	0.17 ± 0.03	8.06 ± 0.18	JWST/NIRSpec MSA data from the JADES programme.
	3 – 6	[6, 10]	0.18 ± 0.03	8.11 ± 0.03	Uses calibrations from
	6 – 10	[6, 10]	0.11 ± 0.05	7.87 ± 0.06	Curti et al. (2020)
Chemerynska et al. (2024)	~ 7	[5.80, 7.19]	0.39 ± 0.02	8.42 ± 0.17	JWST/NIRSpec low-resolution Prism spectra from UNCOVER Survey.
Chakraborty et al. (2025)	3 – 10	[7.5, 10]	0.21 ± 0.03	7.99 ± 0.21	JWST/NIRSpec data
Li et al. (2025)	~ 6	[7.44, 8.84]	0.26 ± 0.01	8.00 ± 0.01	$f_{[\text{O III}]\lambda 5008}$ selected galaxies from NIR-Cam/WFSS (ASPIRE and EIGER surveys)
SPHINX ²⁰ (No selection cut)	5 – 7	[6.51, 10.46]	0.41	8.10	Hydrodynamical simulation
(With selection cut)	5 – 7	[8.46, 10.46]	0.22	7.86	
FLARES ^b (No selection cut)	5 – 7	[8.0, 11.31]	0.44	8.25	Hydrodynamical simulation
(With selection cut)	5 – 7	[8.0, 11.31]	0.44	8.25	
JAGUAR (No selection cut)	5 – 7	[5.98, 9.96]	0.24	8.41	Catalogue from empirical relations
(With selection cut)	5 – 7	[6.86, 9.96]	0.10	8.27	

Notes. ^(a) Eq. 4 fit to data below turnover mass.

^(b) Eq. 4 fit to 50th percentile of the mass-weighted gas-phase metallicity.

References. (1)SPHINX²⁰ (Katz et al. 2023); (2)FLARES (Lovell et al. 2020; Vijayan et al. 2020); (3) JAGUAR catalogue (Williams et al. 2018);