

# MARTA: Temperature-temperature relationships and strong-line metallicity calibrations from multiple auroral-line detections at cosmic noon

E. Cataldi<sup>1,2,3,\*</sup>, F. Belfiore<sup>2</sup>, M. Curti<sup>3</sup>, B. Moreschini<sup>1,2</sup>, F. Mannucci<sup>2</sup>, Q. D'Amato<sup>2</sup>, G. Cresci<sup>2</sup>, A. Feltre<sup>2</sup>, M. Ginolfi<sup>1,2</sup>, A. Marconi<sup>1,2</sup>, A. Amiri<sup>4</sup>, M. Arnaboldi<sup>3</sup>, E. Bertola<sup>2</sup>, C. Bracci<sup>1,2</sup>, S. Carniani<sup>5</sup>, M. Ceci<sup>1,2</sup>, A. Chakraborty<sup>2</sup>, M. Cirasuolo<sup>3</sup>, F. Cullen<sup>6</sup>, C. Kobayashi<sup>7</sup>, N. Kumari<sup>8</sup>, R. Maiolino<sup>9,10</sup>, C. Marconcini<sup>1,2</sup>, M. Scialpi<sup>11,2,1</sup>, and L. Ulivi<sup>11,2,1</sup>

(Affiliations can be found after the references)

Received 28 March 2025 / Accepted 20 September 2025

## ABSTRACT

We present the first results from MARTA (Measuring Abundances at high Redshift with the  $T_e$  Approach), a programme that leverages ultra-deep, medium-resolution JWST/NIRSpec spectroscopy to probe the interstellar medium of star-forming galaxies at  $z \sim 2-3$ . We report detections of one or more auroral lines, including [O III]  $\lambda 4363$ , [O II]  $\lambda \lambda 7320, 7330$ , [S II]  $\lambda 4068$ , and [S III]  $\lambda 6312$ , for 16 galaxies in the sample, providing measurements of multiple ionic temperatures. We tested the validity of the  $T[\text{O II}]-T[\text{O III}]$  relation at high redshifts considering a total sample of 21 objects, including literature data, and obtained a shallower slope than in the low- $z$  literature. However, such a slope is consistent with low-redshift data when ultra-low-metallicity objects are considered. We assessed the correlation of the  $T[\text{O II}]-T[\text{O III}]$  relationship and its scatter on different physical parameters, finding a mild correlation with the ionisation parameter and radiation field hardness, and no significant correlation with gas density. The location of high-redshift data is also consistent with the low- $z$  literature in the  $T[\text{O II}]-T[\text{S II}]$  and  $T[\text{S III}]-T[\text{O III}]$  relations, although this conclusion is limited due to low-number statistics. Finally, we leveraged our sample together with a comprehensive compilation of galaxies with [O III]  $\lambda 4363$  detections from the literature to recalibrate classical strong-line diagnostics at high redshifts. MARTA represents a key addition in this space because it provides direct metallicities at moderately high oxygen abundances ( $12 + \log(\text{O}/\text{H}) \sim 8.0-8.4$ ).

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

## 1. Introduction

Nebular emission from star-forming regions is a fundamental window into the physical conditions and ionising sources within galaxies across cosmic time. For instance, hydrogen recombination lines provide measurements of the star formation rate (SFR) and dust attenuation (Kennicutt 1998; Osterbrock & Ferland 2006; Kennicutt & Evans 2012), while ratios of collisionally excited lines (CELs) from elements like oxygen, nitrogen, and sulphur are sensitive to the electron temperature ( $T_e$ ), electron density ( $n_e$ ), and ionisation state of the gas, enabling the derivation of key physical properties of the ionised gas, including its metal content (Osterbrock & Ferland 2006).

The measurement of chemical abundances in galaxies is a cornerstone of modern astrophysics, offering crucial insights into nucleosynthesis, the history of star formation, and the exchange of gas between galaxies and their surrounding environment (Maiolino & Mannucci 2019; Kewley et al. 2019). In extragalactic astrophysics, the ‘gold standard’ approach to determining the metallicity of the interstellar medium (ISM) involves deriving electron temperatures from faint auroral lines (Peimbert 1967; Stasińska 2002). In particular, ratios of auroral to nebular CELs are exponentially sensitive to the electron temperature as auroral lines originate from higher energy levels compared to nebular ones (Peimbert 1967; Osterbrock & Ferland 2006; Peimbert et al. 2017).

For the purposes of inferring abundances, H II regions are generally modelled using a two-zone model consisting of a low-

and a high-ionisation zone. Of the optical lines, the auroral [O III]  $\lambda 4363$  line has frequently been used to infer oxygen electron temperatures of the high-ionisation zone and consequently the abundance of  $\text{O}^{++}$ . Measurements of the total oxygen abundance, however, require the detection of multiple auroral lines from both low- and high-ionisation species to accurately constrain electron temperatures throughout the nebula (e.g. Pilyugin 2007; Andrews & Martini 2013; Berg et al. 2015; Curti et al. 2017). Alternatively, a relation between the temperatures of the different ionisation zones (temperature-temperature, or  $T-T$  relation) can be adopted (e.g. Campbell et al. 1986; Garnett 1992; Pilyugin et al. 2009; Méndez-Delgado et al. 2023a).

A fundamental limitation to measuring chemical abundances using auroral lines is that they are often significantly fainter than the corresponding strong nebular lines (by tens to even hundreds of times, especially at high metallicities) since their excitation requires higher-energy electrons (Kennicutt et al. 2003; Esteban et al. 2004; Berg et al. 2020). Empirical metallicity calibrators, based on strong line ratios such as  $[\text{N II}] \lambda 6584/\text{H}\alpha$ ,  $[\text{O III}] \lambda 5007/\text{H}\beta$ ,  $[\text{O III}] \lambda 5007/[\text{O II}] \lambda \lambda 3727, 3729$ , or  $[\text{O III}] \lambda 5007/[\text{N II}] \lambda 6584$ , offer an alternative approach by utilising only strong nebular emission lines (e.g. Pettini & Pagel 2004; Curti et al. 2017; see Maiolino & Mannucci 2019 for a review). However, such calibrators are subject to significant limitations as strong-line ratios depend not only on metallicity but also on other parameters, such as the gas density, ionisation parameter, hardness of the ionising spectrum, and relative element abundances (Garnett 2002; Mollá et al. 2006; Pérez-Montero & Contini 2009; Kewley et al. 2019).

\* Corresponding author.

Moreover, empirical calibrations established for local galaxies are not directly applicable to high-redshift galaxies. Extensive observational evidence has demonstrated that galaxies at  $z \sim 2$  exhibit harder ionising spectra – likely driven by  $\alpha$  enhancements in high-redshift star-forming galaxies (Steidel et al. 2016; Topping et al. 2020; Cullen et al. 2021; Stanton et al. 2024) – and higher electron densities compared to their  $z \sim 0$  counterparts (Erb et al. 2006; Hainline et al. 2009; Bian et al. 2010; Yabe et al. 2012; Tacconi et al. 2013; Kewley et al. 2013; Masters et al. 2014; Nakajima & Ouchi 2014; Sanders et al. 2016; Shapley et al. 2019; Topping et al. 2025a). These differences are reflected in diagnostic diagrams such as the N2-BPT (Baldwin, Phillips & Terlevich; Baldwin et al. 1981) diagram,  $\log([\text{O III}]/\text{H}\beta)$  versus  $\log([\text{N II}]/\text{H}\alpha)$ , where high-redshift galaxies are systematically offset from their local counterparts (Steidel et al. 2014; Masters et al. 2014, 2016; Shapley et al. 2015; Strom et al. 2018; Topping et al. 2020). Strong-line ratios like  $[\text{N II}]/\text{H}\alpha$  and  $[\text{O III}]/\text{H}\beta$  become increasingly influenced by other physical parameters at higher redshifts, such as the shape of the ionising spectrum and the ionisation parameter. As a result, their reliability as metallicity tracers is compromised, leading to potential biases when applied to high-redshift galaxies (Bian et al. 2018; Sanders et al. 2021).

The ideal solution is to measure metallicities directly using auroral lines sensitive to electron temperatures, bypassing the uncertainties inherent to empirical calibrators. In cases where only a single auroral line is available, locally calibrated relations (e.g. between the temperature of  $\text{O}^+$  and  $\text{O}^{++}$ , known as  $T_2$ – $T_3$  relations) must be assumed, a situation that introduces additional uncertainties. Detecting these faint lines in individual high-redshift galaxies was extremely challenging prior to JWST, with only a few, low-significance detections of  $[\text{O III}] \lambda 4363$  reported at  $z > 2$  (e.g. Christensen et al. 2012; James et al. 2014; Sanders et al. 2020; see also the compilation by Patrício et al. 2018) and rare cases of auroral  $[\text{O II}] \lambda \lambda 7320, 7330$  detections, such as those presented in Sanders et al. (2023) for two galaxies at  $z > 2$ . This scarcity hindered the accurate determination of metallicities in distant galaxies, necessitating the use of empirical metallicity indicators calibrated on lower-redshift systems.

The JWST Near Infrared Spectrograph (NIRSpec; Jakobsen et al. 2022), with its unparalleled sensitivity and wide spectroscopic wavelength coverage of the near-infrared regime, has revolutionised this field, enabling the detection of auroral lines in galaxies both at cosmic noon ( $z \sim 2$ – $3$ ; e.g. Strom et al. 2023; Welch et al. 2024, 2025; Rogers et al. 2024) and at much higher redshifts (Curti et al. 2023; Yang et al. 2023; Laseter et al. 2024; Morishita et al. 2024a; Sanders et al. 2024; Arellano-Córdova et al. 2025; Cullen et al. 2025; Scholte et al. 2025; Shapley et al. 2025; Stanton et al. 2025). These efforts have delivered some of the most accurate determinations of gas-phase abundances in early galaxies. Interestingly, the higher sensitivity of JWST/NIRSpec towards the redder part of the spectrum, combined with the metal deficiency, higher ionisation conditions, and compact nature of high- $z$  systems, as well as the tendency of most JWST programmes to focus on the high- $z$  Universe, has made auroral line detections more likely at very high  $z$  than at  $z \sim 2$ – $3$ , where galaxies are typically more massive, extended, and metal-rich than at  $z \gtrsim 5$ . Moreover, very few observations have enabled detections of multiple auroral lines, which are necessary to probe the temperature and ionisation structure of high-redshift galaxies.

While this paper was under review, an independent analysis by Sanders et al. (2025) presented results from the AURORA

(Assembly of Ultradeep Rest-optical Observations Revealing Astrophysics) survey, combining new JWST/NIRSpec observations with literature data to investigate temperature–temperature relations and to calibrate strong-line metallicity diagnostics at  $z \sim 2$ – $10$ . They found that local  $T_2$ – $T_3$  relations remain valid at high redshifts, while empirical strong-line calibrations show clear offsets with respect to the local ones, consistent with evolving ISM conditions; in particular, diagnostics based on  $\alpha$  elements provide the most reliable metallicity estimates, whereas nitrogen-based indicators exhibit significant scatter, likely due to variations in N/O at fixed O/H.

In this paper we present novel observations of multiple auroral lines in spectra of individual galaxies at cosmic noon from the MARTA (Measuring Abundances at high Redshift with the  $T_e$  Approach) survey (PID 1879, PI Curti), which targeted 127 galaxies in the COSMOS (Cosmic Evolution Survey) field at redshifts  $z \sim 2$ – $3$  by means of very deep NIRSpec Micro Shutter Array (MSA) observations. In Curti et al. (2025a), the multiple science cases addressed by such a programme are discussed, with a focus on the ISM properties and chemical abundances of one of the galaxies in the sample.

Here, we leveraged a sample of 16 galaxies with (multiple) detections of auroral lines, mainly  $[\text{O III}] \lambda 4363$  and  $[\text{O II}] \lambda \lambda 7320, 7330$  but also  $[\text{S II}] \lambda 4069$  and  $[\text{S III}] \lambda 6312$ . This sample extends the parameter space probed by cosmic noon galaxies with auroral line detections towards higher metallicities than most existing studies. We focus on the relationships between temperatures inferred from different ionic species and their possible redshift evolution. We discuss the implications for the applicability of the  $T_e$  method for measuring metallicity in high- $z$  galaxies and investigate the redshift evolution of some of the most widely adopted strong-line metallicity diagnostics.

The paper is structured as follows. Section 2 outlines the survey design, including the target selection, observational strategy, and data reduction process. In Sect. 3 we detail the data analysis methodology and the spectral fitting procedure. Section 4 defines the two subsamples of the MARTA survey, golden and silver, which are used for all subsequent analyses. In Sect. 5 we describe the dust correction procedure applied. Section 6 focuses on the oxygen temperature–temperature relation, the analysis of their dependence on other parameters such as density and the ionisation parameter, and a comparison with sulphur temperatures. Section 7 presents the results on strong-line calibrations. Finally, in Sect. 8 we provide a summary of our conclusions and prospects for future research.

In this work we use the following notation to refer to the strong-line diagnostics:

- $R3 = \log([\text{O III}] \lambda 5007/\text{H}\beta)$
- $R2 = \log([\text{O II}] \lambda 3726, 3729/\text{H}\beta)$
- $O32 = R3 - R2$
- $R23 = \log([\text{O II}] \lambda 3726, 3729 + [\text{O III}] \lambda 4959, 5007)/\text{H}\beta)$
- $\hat{R} = 0.46 R2 + 0.88 R3$
- $\text{Ne3O2} = \log([\text{Ne III}] \lambda 3869/[\text{O II}] \lambda 3726, 29)$
- $N2 = \log([\text{N II}] \lambda 6585/\text{H}\alpha)$
- $O3N2 = R3 - N2$
- $S2 = \log([\text{S II}] \lambda 6716 + [\text{S II}] \lambda 6730)/\text{H}\alpha)$
- $O3S2 = R3 - S2$

## 2. Observations and data reduction

### 2.1. Target selection, priority classes, and observing strategy

The MARTA programme focuses on the redshift range around the peak of cosmic star-formation activity (cosmic noon), with

the primary aim to detect the auroral lines of [O III]  $\lambda$ 4363 in the G140M ( $R \sim 1000$ ) grating and, for a subsample, the [O II]  $\lambda$ 7320,7330 lines in the G235M ( $R \sim 1000$ ) grating. Considering the need to cover [O II]  $\lambda$ 3727,29 to measure electron temperatures and derive metallicities, this constrains the primary redshift range of our targets to  $1.7 < z < 3.1$ . However, a few higher-redshift sources were also observed, according to selection and prioritisation criteria described below.

The MARTA parent catalogue was generated from the COSMOS2020 (Weaver et al. 2022) and the 3D-HST (3D Hubble Space Telescope) catalogues (Skelton et al. 2014; Momcheva et al. 2016) based on a multi-step process. The COSMOS2020 catalogue provided photometric redshifts while the 3D-HST catalogue provided grism spectroscopic redshift data, which were complemented by spectroscopic redshifts from higher spectral resolution, ground-based observations available at the time of MSA mask design (including zCOSMOS-bright, Lilly et al. 2009; FMOS, Silverman et al. 2015; MOSDEF, Kriek et al. 2015; zFIRE, Nanayakkara et al. 2016; VUDS, Tasca et al. 2017; hCOSMOS, Damjanov et al. 2018; DEIMOS 10k, Hasinger et al. 2018; LEGA-C, Straatman et al. 2018; KMOS3D, Schreiber et al. 2019; KLEVER, Curti et al. 2020a; and data from Kartaltepe et al. 2010, 2015; Masters et al. 2012; Balogh et al. 2014; Comparat et al. 2015; Schreiber et al. 2018). When ground-based spectroscopic redshifts were not available, lower resolution, grism-based redshifts were adopted. In the absence of both, photometric redshifts were used, provided they were consistent between LEPHARE (Ilbert et al. 2006) and EAZY (Brammer et al. 2008), within a defined tolerance ( $\Delta(z)/z < 0.1$ , considering the redshift estimates from EAZY).

For each galaxy in the MARTA parent catalogue, we estimated the expected flux of the [O III]  $\lambda$ 4363 emission line based on the available information on stellar masses ( $M_*$ ), SFR, and metallicity, as described in the following. The stellar masses, SFR, and dust attenuation ( $A_V$ ) were adopted from spectral energy distribution fitting of the photometry extracted by COSMOS2020 (Weaver et al. 2022). When ground-based spectroscopy (either slit or integral field unit) in COSMOS covering  $H\alpha$  was available, as provided by, for example, the KMOS3D (K-band Multi Object Spectrograph 3D survey) (Wisnioski et al. 2019), KLEVER (KMOS Lensed Emission-Line Velocity and Metallicity Survey, Curti et al. 2020a), and MOSDEF (MOSFIRE Deep Evolution Field survey, Kriek et al. 2015) surveys, the SFR was estimated from the extinction-corrected  $H\alpha$  flux using the relation from Kennicutt & Evans (2012) and the  $A_V$  derived from either photometry or from the Balmer decrements. As of gas-phase metallicity, we assumed the value for oxygen abundance derived from strong-line diagnostics and the Bian et al. (2018) calibrations, specifically those based on local analogues of high-redshift galaxies, for galaxies with available spectroscopy and detection of at least three nebular emission lines, whereas we estimated the expected metallicity for each source with no previously available spectra assuming that the relationship between metallicity, mass, and SFR (so-called fundamental metallicity relation; Mannucci et al. 2010) does not evolve with redshift up to  $z \sim 2$  (Cresci et al. 2019; Sanders et al. 2021), adopting specifically the parametrisation of Curti et al. (2020b).

Finally, we based our indirect estimate of the [O III]  $\lambda$ 4363 flux on the relationship between the [O III]  $\lambda$ 4363/ $H\alpha$  ratio and metallicity, the latter calibrated on a combined sample of local H II regions and low-metallicity galaxies from Guseva et al. (2011), Berg et al. (2020), and stacks from SDSS (Sloan Digital Sky Survey, Curti et al. 2017) with ‘direct’,  $T_e$ -based oxy-

gen abundances. Such intrinsic [O III]  $\lambda$ 4363 flux was then attenuated by the galaxy  $A_V$ , assuming the attenuation curve of Cardelli et al. (1989) and an  $R_V = 3.1$ , to derive our final estimate of the expected [O III]  $\lambda$ 4363 flux in each source.

Galaxies in the MARTA parent catalogue were assigned to four different priority classes (each one with different weights) to drive the MSA target allocation. Such classes were established on the basis of the expected emission line fluxes (estimated as detailed above), while additional factors such as dust attenuation and redshift reliability were also considered to fine-tune the weights assignment and optimise the final MSA mask. The detailed list of the criteria assumed in target prioritisation is reported in Table 1. A total of 126 sources, including 18 ‘priority one’, 22 ‘priority two’, 12 ‘priority three’, 21 ‘priority four’, targets and 53 ‘filler’ sources, were observed using a single MSA configuration, though only galaxies belonging to the top two classes were used to drive the choice of the pointing coordinates.

Each galaxy was observed through a three-shutter slitlet, and observations were taken adopting a standard nodding strategy with three nod positions along the slitlet. To maximise the number of targets observable within a single pointing, we allowed spectral traces to overlap on the detector, as the probability of overlapping emission lines from two different sources was estimated to be minimal. However, due to the depth of our observations, significant continuum emission was detected in many cases. The impact of this observing choice on background subtraction is discussed in Sect. 2.2.

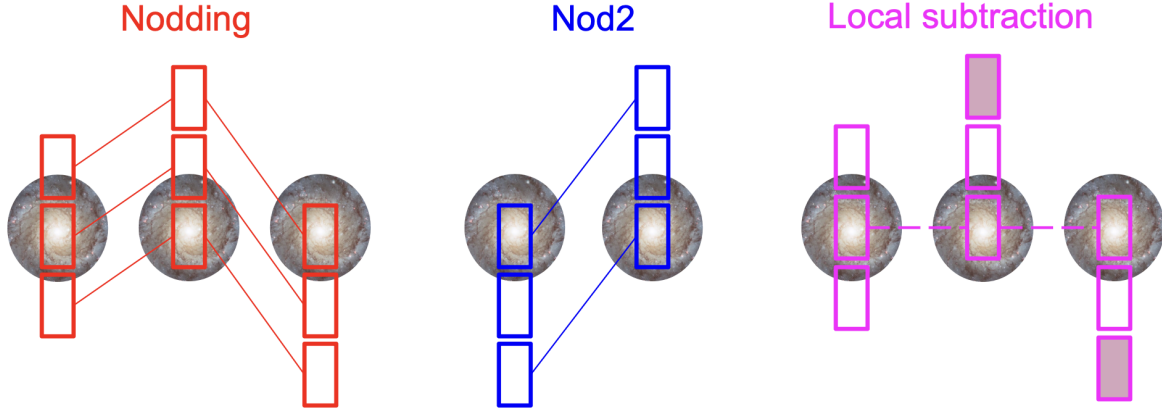
Observations were obtained totalling an integration time of 31.9 hours in G140M, 7.4 hours in G235M, and 2.9 hours in G235H. The G235H is intended for studies of emission-line kinematics and is therefore not used in this work, where we instead focused on the G140M and G235M data.

## 2.2. Data reduction

The data reduction of our observations was performed using the JWST calibration pipeline, which is divided into three stages. We applied the pipeline version ‘2023\_2a’ and CRDS (Calibration Reference Data System) context files ‘jwst\_1183.pmap’. Stage 1 applies detector-level corrections, including group-by-group corrections and ramp fitting, producing a count rate image for each exposure. Stage 2 focuses on instrument-specific and observing-mode calibrations, resulting in fully calibrated individual exposures. At this stage, the standard nodding background subtraction is performed, and path-loss corrections estimated for point-like sources are applied.

Stage 3 combines multiple exposures, such as those from nodding and dither patterns, into a single 2D spectral product, while applying outlier rejection and additional corrections as needed. Although the pipeline also provides a final extracted 1D spectrum, we performed a bespoke extraction of 1D spectra for each target as described later in this section. Other parameters in the data reduction pipeline were kept to their default values.

Due to our observing strategy (allowing spectral overlap) and the spatially extended nature of our sources, in several cases a dedicated background subtraction is required. In particular, our data suffer from spectral contamination from sources located in different MSA slits, whose continuum emission could contaminate the central shutter of the slitlet and/or the flanking ‘background’ shutters. Moreover, in some cases, the flux from the target object is extended or not well centred, and extends into the micro-shutters adjacent to the central one. As a result, applying the standard nodding background subtraction, which subtracts the background pixel-per-pixel using the same shutter after



**Fig. 1.** Comparison of strategies for background subtraction in NIRSpec/MSA observations. Left (red): Standard three-point nodding strategy, in which the galaxy is observed in three positions within the shutter, and the background for each position is estimated using the corresponding shutter exposures from the other two nods. Centre (blue): Nod2 strategy, a two-position scheme where only two out of the three available positions are used for the nodding procedure in order to use the shutters farthest from the galaxy as the background. Right (magenta): Local subtraction approach, which uses the outer shutters from the two off-source positions (indicated by the filled-in shutters) to estimate the sky background. In this method, the object signal is taken from the sum of all three nod positions (connected by dashed magenta lines), but no background is subtracted from the same shutters. As such, this method can be more sensitive to local systematics or background gradients.

**Table 1.** Target prioritisation criteria.

Priority class	Expected [O III] $\lambda 4363$ Flux [ $\text{erg s}^{-1} \text{cm}^{-2}$ ]	Expected H $\alpha$ Flux [ $\text{erg s}^{-1} \text{cm}^{-2}$ ]	Additional criteria for prioritisation
Priority one	$>8 \times 10^{-19}$	–	Lower extinction, reliable spectroscopic redshift
Priority two	$[1.5-8] \times 10^{-19}$	$>1.5 \times 10^{-17}$	Lower extinction, reliable spectroscopic redshift
Priority three	$[1-1.5] \times 10^{-19}$	$>1.5 \times 10^{-17}$	Includes $3.1 < z < 5$ if G235M covers [O III] $\lambda 4363$
Priority four	–	$>7.5 \times 10^{-18}$	Higher weights for [O III] $\lambda 4363 > 3 \times 10^{-20}$
Filler	–	$>1 \times 10^{-20}$	Higher weights for larger H $\alpha$ fluxes

offsetting (Fig. 1, left), can lead to over-subtraction of the target’s flux and/or to a poor modelling of the background itself.

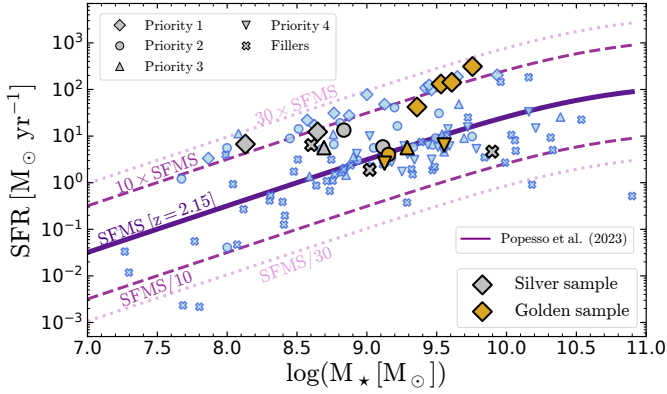
To mitigate over-subtraction in cases of extended sources, we calculated the expected contamination of the flanking shutters by the source flux using NIRCAM (Near InfraRed Camera) imaging data from the PRIMER (PRism and Imager for Extragalactic Research) survey (PID:1837, Dunlop et al. 2021) in bands F115W, F150W, F200W, and F277W (broadly corresponding to the wavelength range probed by our NIRSpec observations in G140M and G235M). The reduced and calibrated NIRCAM imaging data products were retrieved from the Dawn JWST Archive (DJA<sup>1</sup>).

For galaxies where PRIMER imaging and 2D spectra indicate that the emission is visibly extended beyond the central shutter, we adopted a so-called ‘nod2 subtraction’ strategy (e.g. Maseda et al. 2024). In this approach, only two of the nod-specifically, the ones where the galaxy is in the outermost shutter-are used. This minimises the risk of self-subtraction, where the galaxy’s emission would be subtracted along with the background, while maintaining the advantage of pixel-by-pixel subtraction, which mitigates detector artefacts. However, this strategy comes at the cost of sacrificing one-third of the total exposure time, as the exposure in which the galaxy is located in the central shutter is excluded (Fig. 1, middle).

An alternative approach to background subtraction consist in employing the combined, non-background subtracted 2D spectra from all three nodded exposures to extract and construct a

‘local’ background as the average of the two outermost shutters. Specifically, we extracted the background spectrum from the two outermost shutters in the nodding configuration, summing the flux across all pixels in each shutter and then averaging the results to obtain a representative background spectrum (Fig. 1, right). The galaxy spectrum was extracted from a box-car aperture centred on the peak of the S/N in the spectral trace of the target and of width of 5 pixels (roughly equivalent to the size of one shutter), although in some cases extended line emission as visible in the 2D spectra suggested to increase the size of the extraction aperture to avoid significant flux losses and possible biases in the inferred line ratios. Such a ‘local subtraction’ strategy benefits from the full on-source integration time by co-adding all three nodded exposures but suffers from the increased noise affecting the background spectrum, which is generated from exposures with shorter effective integration times than the central trace containing the target source. Moreover this strategy does not leverage the pixel-by-pixel background subtraction of the standard ‘nodding’ therefore sometimes resulting in spectra with worse noise properties due to detector artefacts. For compact galaxies or those that do not show significant emission in adjacent shutters, the standard nodding background subtraction was applied and we checked that the fluxes obtained from the local background subtraction are comparable. Specifically, we applied the local background subtraction for galaxy MARTA\_3942, the nod2 method for MARTA\_4195, and the standard nodding background subtraction for the remaining sources. To compare the results of these different strategies and ensure that absence of self-subtraction, we assessed their

<sup>1</sup> <https://dawn-cph.github.io/dja/>



**Fig. 2.** Star formation main sequence for the MARTA sample. Different shapes correspond to different priority classes, as described in Table 1, and different colours to the different samples: the golden sample is in orange (Sect. 4.2), the silver sample in grey (Sect. 4.3), and the rest of the galaxies in blue. For comparison, we plot the parametrisation of the star-forming main sequence at  $z \sim 2.15$  (median redshift of the full MARTA sample) by Popesso et al. (2023).

impact on the extracted and measured fluxes of [O III]  $\lambda 5007$  in the G140M and on the  $H\alpha$  in the G235M gratings, respectively. We verified that, had the standard nodding procedure been used in the case of extended galaxies, these lines would have been significantly underestimated. In particular, we tested the effect on MARTA\_3942 and MARTA\_4195, and found that the [O III] flux is affected by 13–22% and the  $H\alpha$  flux by 7–20%.

As some of our galaxies look morphologically extended, they might be affected by larger slit losses than modelled for simple point-like sources. On the other hand, in several cases the bulk of the emission captured by the slit comes from bright individual clumps, and hence the assumption of constant surface brightness across the slit is also incorrect. The analysis presented in this work only relies on emission lines ratios, and is therefore not affected by uncertainties on the absolute flux calibration. However, a second-order correction to the relative line fluxes is expected due to the point-spread-function dependence of the path-loss correction, which is not modelled here. To test the reliability of the relative flux calibration in our spectra, we measured the ‘intra-shutter’ photometry from NIRCcam imaging in the F115W, F150W, F200W, and F270W broadband filters (sampling the spectral range covered by G140M and G235M gratings), compared it to the synthetic photometry extracted by applying the transmission curve of each filter to the spectra, and verified that for the 16 galaxies analysed in the present study no significant relative flux calibration mismatch between the blue-end and red-end part of each grating was observed.

Finally, we verified the consistency of the relative flux calibration across gratings by analysing the overlapping region between G140M and G235M spectra, which roughly spans the 1.66 and 1.90  $\mu\text{m}$  wavelength interval. Specifically, we calculated the ratio of continuum fluxes between the G140M and G235M gratings for seven galaxies in our sample (Sect. 4.2). For most galaxies, these ratios fell within 5% of unity – i.e. between 0.95 and 1.05 – demonstrating excellent agreement between the two gratings. For galaxies with ratios outside this range, the discrepancies were attributable to contaminants in the bluest or reddest regions of the spectra, as evident by inspecting their 2D spectra. We therefore did not perform a correction to the flux calibration of the gratings.

### 3. Spectral fitting

We measured spectroscopic redshifts for 73 out of the 126 targets by using strong lines, such as [O III]  $\lambda 5007$  and  $H\alpha$ . A detailed description of the spectral properties for the full sample will be presented in a future publication.

We performed spectral fitting on the 1D data using the PYTHON module of the penalised pixel-fitting (PPXF) routine (Cappellari & Emsellem 2004), which can simultaneously fit both galaxy emission lines and the underlying stellar continuum. Since the emission lines are not spectrally resolved, we modelled them using a single Gaussian component per line. The emission lines included in our analysis are listed in Table E.1. For the continuum fitting, we used eMILES stellar population templates (Vazdekis et al. 2016, 2010) with BaSTI isochrones (Pietrinferni et al. 2004) and a Chabrier (2003) initial mass function. We modelled the line spread function with a Gaussian profile varying as a function of wavelength, using the spectral resolution profiles provided in the JWST documentation for the G140M and G235M filters<sup>2</sup>.

We assessed the impact of different stellar population libraries by systematically comparing the measured fluxes of key emission lines, focusing in particular on [O III]  $\lambda 4363$ . Although the strength of Balmer absorption features varies noticeably across templates, the effect on the [O III]  $\lambda 4363$  flux remains minor – typically less than  $\sim 10\%$ , with a median variation of  $\sim 6\%$ . A more detailed description of these tests is provided in Appendix B.

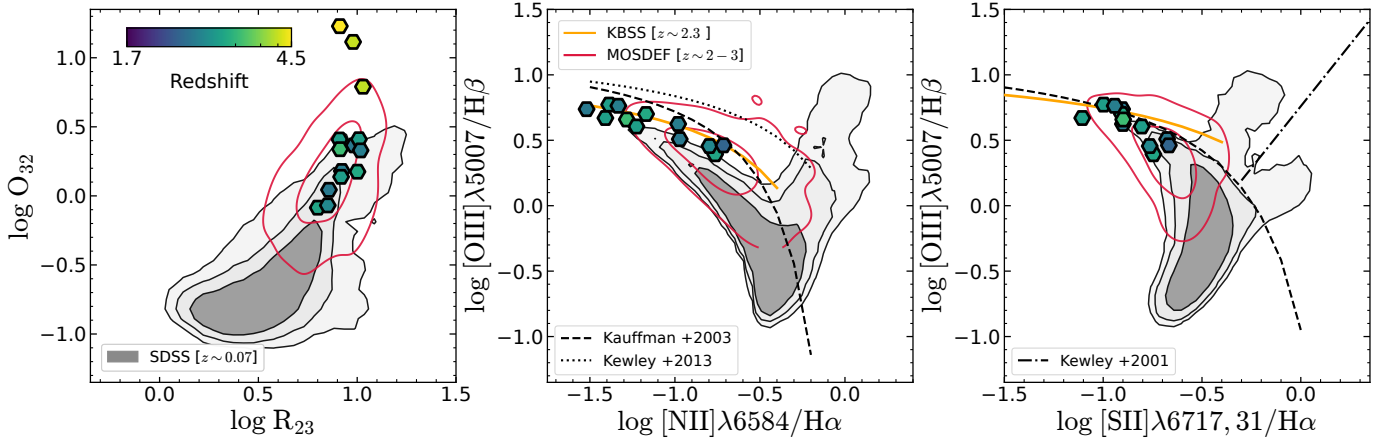
A 24-degree multiplicative polynomial was applied to modify the continuum templates, accounting for potential dust reddening, residual flux calibration uncertainties, and template mismatch effects. The choice of polynomial degree was determined through a systematic analysis comparing different orders (5, 10, and 24) by evaluating the quality of the fit and the statistical properties of the residuals, with specific focus on the line-free region between rest-frame 4125 and 4325  $\text{\AA}$ , chosen to assess the goodness of the fit in proximity to the auroral [O III]  $\lambda 4363$  line. The 24-degree polynomial consistently yielded the best results, indicating an improved match between the model and the observed spectrum. Nonetheless, we verified that the variations in the emission lines driven by the choice of different polynomial degrees do not significantly impact the results presented in this work. Further discussion of the impact of the continuum subtraction on the measurement of weak auroral lines and an example spectral fit is presented in Sect. 4.2. Within each grating, the velocity ( $v$ ) and velocity dispersion ( $\sigma$ ) parameters of the various emission lines were tied together, forcing a consistent kinematic structure for the emission regions. We performed the fitting separately for each of the two gratings, G140M and G235M. Where emission lines were serendipitously observed in both gratings, their fluxes are consistently in agreement within the uncertainties.

## 4. Auroral emission lines in the MARTA sample

### 4.1. Properties of galaxies with auroral line detections

In this work we focused on a subset of 16 galaxies with one or more auroral line detections. We note that this subsample has been explicitly selected and prioritised on the basis of the likelihood of [O III]  $\lambda 4363$  detection. As a consequence, it could

<sup>2</sup> <https://jwst-docs.stsci.edu/jwst-near-infrared-spectrograph/nirspec-instrumentation/nirspec-dispersers-and-filters>



**Fig. 3.** Diagnostic diagrams for the MARTA galaxies analysed in this work. From left to right: Distribution of the combined golden and silver samples on the  $O_{32}$  vs  $R_{23}$ ,  $[O\text{III}]\lambda 5007/H\beta$  vs  $[N\text{II}]\lambda 6584/H\alpha$ , and  $[O\text{III}]\lambda 5007/H\beta$  vs  $[S\text{II}]\lambda 6717, 31/H\alpha$  diagrams, colour-coded by galaxy redshift. The grey contours indicate the region populated by local galaxies from SDSS, with darker shades denoting higher densities. Red contours mark the distribution of galaxies from the MOSDEF survey (Kriek et al. 2015), whereas the orange curve is the fit to KBSS galaxies from Strom et al. (2017); both surveys target primarily  $z \sim 2-3$  systems. The dividing lines between star-forming galaxies, AGNs, and LINER galaxies from Kewley et al. (2001, 2013) and Kauffmann et al. (2003) in both the  $[N\text{II}]$ -BPT and  $[S\text{II}]$ -BPT diagrams are also drawn. Median error bars are not shown, as they are smaller than the symbol size and would not be visible in the plot; this is due to the fact that these are among the strongest lines, measured with very high S/Ns.

**Table 2.** Physical properties of MARTA galaxies in the golden (top) and silver (bottom) samples.

Galaxy ID	$z$	$\log M_*$ [ $M_\odot$ ]	$\log \text{SFR}$ [ $M_\odot \text{ yr}^{-1}$ ]	$E(B-V)$ [mag]	$\sigma(E(B-V))$ [mag]	$T_2^{(\dagger)}$ [ $10^4 \text{ K}$ ]	$T_3$ [ $10^4 \text{ K}$ ]	$n_e^{(*)}$ [ $\text{cm}^{-3}$ ]	$12 + \log(O/H)$
2387	1.836	9.76	2.49	$0.00 \pm 0.00$	0.10	$0.97 \pm 0.13$	$1.23 \pm 0.14$	$718 \pm 193$	$8.28 \pm 0.10$ (0.11)
5014	1.837	9.15	0.60	$0.02 \pm 0.01$	0.03	$1.17 \pm 0.05$	$1.00 \pm 0.07$	$142 \pm 37$	$8.19 \pm 0.01$ (0.02)
4502	1.854	9.29	0.75	$0.08 \pm 0.01$	0.11	$1.33 \pm 0.06$	$0.94 \pm 0.05$	$75 \pm 18$	$8.31 \pm 0.03$ (0.04)
552	1.924	9.53	2.12	$0.02 \pm 0.01$	0.03	$0.94 \pm 0.05$	$1.16 \pm 0.03$	$263 \pm 56$	$8.35 \pm 0.03$ (0.03)
3942	2.098	9.13	0.42	$0.24 \pm 0.01$	0.05	$1.07 \pm 0.05$	$1.01 \pm 0.05$	$104 \pm 37$	$8.27 \pm 0.02$ (0.03)
4195	2.178	9.55	0.81	$0.18 \pm 0.01$	0.13	$0.91 \pm 0.07$	$0.98 \pm 0.16$	$241 \pm 61$	$8.40 \pm 0.05$ (0.09)
4327	2.224	9.36	1.62	$0.20 \pm 0.01$	0.15	$1.08 \pm 0.03$	$1.04 \pm 0.02$	$210 \pm 31$	$8.35 \pm 0.01$ (0.02)
3408	1.861	9.02	0.28	$0.10 \pm 0.02$	0.12	$1.00 \pm 0.03$	$0.97 \pm 0.05$	–	$8.46 \pm 0.02$ (0.03)
414	2.093	8.65	1.09	$0.07 \pm 0.02$	0.10	$1.23 \pm 0.03$	$1.39 \pm 0.05$	$171 \pm 32$	$7.97 \pm 0.01$ (0.02)
4471	2.096	9.11	0.78	$0.12 \pm 0.02$	0.06	$1.40 \pm 0.06$	$1.70 \pm 0.11$	$1400 \pm 200$	$7.67 \pm 0.01$ (0.03)
329	2.106	8.13	0.83	$0.00 \pm 0.01$	0.03	$1.99 \pm 0.10$	$2.79 \pm 0.18$	–	$7.21 \pm 0.02$ (0.02)
1084	1.764	9.61	2.15	$0.06 \pm 0.01$	0.08	$0.94 \pm 0.03$	$0.86 \pm 0.05$	–	$8.55 \pm 0.05$ (0.06)
3115	2.492	8.84	1.13	$0.01 \pm 0.07$	–	$1.21 \pm 0.10$	$1.35 \pm 0.18$	–	$7.87 \pm 0.06$ (0.07)
1374	3.713	9.90	0.67	$0.00 \pm 0.34$	–	$1.05 \pm 0.04$	$1.07 \pm 0.08$	–	$8.32 \pm 0.26$
3887	3.796	8.69	0.75	$0.00 \pm 0.39$	–	$1.29 \pm 0.02$	$1.51 \pm 0.03$	–	$7.84 \pm 0.17$
3926	4.658	8.60	0.81	$0.00 \pm 0.24$	–	$1.30 \pm 0.09$	$1.52 \pm 0.17$	–	$7.75 \pm 0.23$

**Notes.** Masses and SFR are from the COSMOS2020 catalogue (Weaver et al. 2022). The  $E(B-V)$  is determined from a fit to Balmer decrements as described in Sect. 5, while  $\sigma(E(B-V))$  reports the standard deviation of the values obtained from individual Balmer line ratios. The uncertainties in parentheses in the metallicity column correspond to the values obtained by propagating the larger  $\sigma(E(B-V))$  uncertainty into the metallicity derivation.  $^{(\dagger)}$ Obtained from the  $T_2-T_3$  relation of Eq. (1) for the silver sample.  $^{(*)}$ Where the  $[S\text{II}]$  doublet is not available or undetected ( $S/N < 3$ ), a fixed density of  $300 \text{ cm}^{-3}$  is assumed.

be biased towards galaxies with elevated auroral  $[O\text{III}]\lambda 4363$ , which in turn correspond to higher electron temperatures, lower metallicities, and higher SFRs. This selection effect must be kept in mind when interpreting the physical properties of these galaxies in a broader context. However, as we show later in this section, strong-line diagnostic analyses do not reveal significant systematic biases. Nonetheless, potential biases could emerge in scaling relations, which are beyond the scope of this paper but should be carefully considered in future studies.

We classified these 16 galaxies into two groups: we considered the seven galaxies with both  $[O\text{III}]\lambda 4363$  and

$[O\text{II}]\lambda 7320, 7330$  detections as our ‘golden sample’ and the remaining nine galaxies with detections of  $[O\text{III}]\lambda 4363$  only as the ‘silver sample’. For five galaxies in the golden sample, additional auroral lines from both high-ionisation and low-ionisation ions are detected.

Figure 2 shows the distribution of all observed MARTA targets (marked according to their different priority classes) in the  $M_*$ -SFR plane. The parametrisation of the star-forming main sequence (SFMS; Noeske et al. 2007) at  $z = 2.15$  (median redshift of the sample) from Popesso et al. (2023) is also shown for comparison. Galaxies with with auroral lines detections in

MARTA sit preferentially above the SFMS at  $z = 2.15$ , having SFRs between 5 and 30 times higher the average at fixed  $M_*$ .

In Fig. 3 we plot the position of the MARTA galaxies analysed in the present study (the golden and silver samples) onto different emission line-based diagnostic diagrams, namely  $R_{23}$  versus  $O_{32}$ , the [N II]-BPT diagram, and the [S II]-BPT ( $[S II]/H\alpha$  vs  $[O III]/H\beta$ ). In the  $R_{23}$  versus  $O_{32}$  diagram MARTA galaxies at  $z > 3$  are offset from both the SDSS and MOSDEF ( $z \sim 2-3$ , Kriek et al. 2015) contours, occupying a region generally associated with high-ionisation, low-metallicity systems, as recently probed by large (though shallower) JWST surveys such as JADES (JWST Advanced Deep Extragalactic Survey) (Cameron et al. 2023a) and CEERS (Cosmic Evolution Early Release Science Survey) (Sanders et al. 2023). In the BPT diagrams, MARTA galaxies at  $z \sim 2$  occupy the upper envelope of the local SDSS distribution, showcasing also the known offset in the [N II]-BPT, (Steidel et al. 2014; Shapley et al. 2015). They overlap with the regions populated by surveys targeting  $z \sim 2-3$  galaxies, such as MOSDEF and KBSS (Keck Baryonic Structure Survey) (Strom et al. 2017).

Based on classical classification schemes, such as those provided by Kewley et al. (2001, 2013), Kauffmann et al. (2003), we find no evidence of active galactic nucleus (AGN) contamination. This conclusion is corroborated also by the absence of specific AGN spectral features in our sample, such as broad components in the Balmer emission lines, very high-ionisation species, or He II emission<sup>3</sup>. Furthermore, as described in Sect. 6, we do not find unusually high ( $\geq 25\,000$  K) electron temperatures that could be associated with AGN ionisation (e.g. Mazzolari et al. 2024).

#### 4.2. The MARTA golden sample: Multiple auroral line detections

The MARTA golden sample consists of seven galaxies with  $S/N \geq 5$  detections of the [O III]  $\lambda 4363$  auroral line and  $S/N \geq 3$  for the [O II]  $\lambda 7320, 7330$  auroral doublet (where each line in the doublet is detected at  $\geq 3\sigma$ ). The  $S/N$  on emission lines was estimated from both the output of the pPXF fitting and the RMS of the spectra within a featureless region around the emission line of interest; a comparison between these two methods is presented in Appendix A. As reported in Table 2, the golden sample spans  $z = [1.8-2.2]$ ,  $\log(M_*/M_\odot) = [9.2-9.8]$  and SFRs ranging from 2.6 to 310  $M_\odot/\text{yr}$  (with a mean value of 80  $M_\odot/\text{yr}$ ). These galaxies lie on average significantly above the median SFMS at  $z \sim 2$  (Fig. 2). We quantified the  $\Delta MS$  as the distance in  $\log(\text{SFR})$  from the parametrisation of Popesso et al. (2023), and found  $\Delta MS$  values spanning from  $-0.2$  to  $+1.3$  dex, with a median value of  $+0.47$  dex.

Galaxy MARTA\_1084 was excluded from this sample, despite detections of both oxygen auroral lines, because its [O II]  $\lambda 3726, 3729$  line falls within the detector gap, precluding its use for  $O^+$  electron temperature calculation and thus for direct metallicity determination. We therefore included this galaxy in the silver sample.

For some of these galaxies we also report the detection of additional auroral lines. In particular, the [S II]  $\lambda 4068$  auroral line is detected with  $S/N > 3$  in four cases. Similarly, the [S III]  $\lambda 6312$  auroral line is detected at  $\geq 3\sigma$  in three galaxies, where a proper de-blending from the nearby nebular [O II]  $\lambda 3727$  line can be

performed. Figure 4 shows a compilation of the auroral line detections across this subsample. Although MARTA\_1084 was excluded from the golden sample, it is included in the plot as it displays two auroral lines. In this figure, each panel corresponds to one of the auroral emission lines, with the continuum-subtracted spectrum shown along with the fitted Gaussian model. In some cases bespoke line fits were performed for individual objects, as detailed in Appendix B.

Figure 5 reports instead the complete spectral fitting, for both G140M and G235M gratings, for one of the galaxies in the golden sample – MARTA\_4502. This spectrum is representative of the quality and S/N of other objects in the golden sample. We observe strong continuum with clear stellar absorption features under the Balmer lines. We observe multiple high-S/N auroral lines, including [O II]  $\lambda 7320, 7330$ , [O III]  $\lambda 4363$ , and [S II]  $\lambda 4068$ . The Balmer series is observed up to H10, and generally detected up to H7 in all golden sample galaxies. In several objects, Paschen 10 and Paschen 11 are also visible. We note the presence of other semi-strong emission lines, such as [Ne III] and [S III]  $\lambda 9071, 9533$ , which are particularly valuable as they provide insights into the chemical abundances and ionisation level of the elements in the gas, enabling more detailed studies of chemical abundance patterns (e.g. Stanton et al. 2025).

#### 4.3. The MARTA silver sample: Single auroral line detections

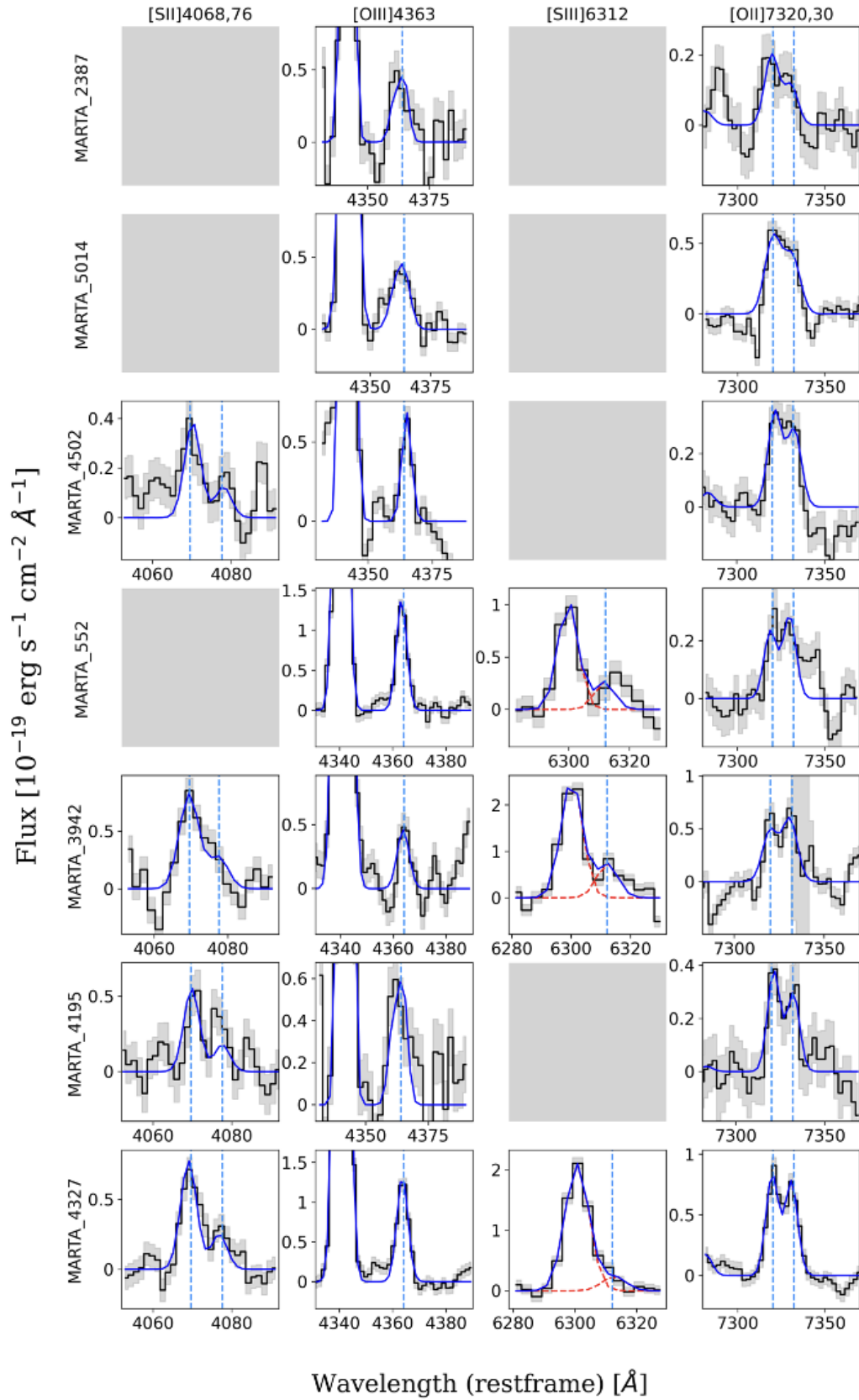
The MARTA silver sample consists of nine galaxies, including eight additional detections of the auroral [O III]  $\lambda 4363$  line with  $S/N > 5$  (Table 2). Additionally, we included the previously discussed MARTA\_1084 in this sample. In Fig. 6 we show a zoom-in of the spectral region around the [O III]  $\lambda 4363$  auroral line in these systems, with the best-fit model overlaid.

These galaxies span a broader redshift range than the golden sample objects, with two objects at  $z > 3$  and one at  $z > 4$ . The silver sample also has generally lower stellar masses,  $\log(M_*/M_\odot) = [8.13-9.9]$ , and a  $\Delta MS$  spanning from  $-0.66$  to  $+1.2$  dex, with a median value of  $+0.43$  dex, so slightly less star-forming, on average, than the golden sample. Although these galaxies cannot be used to evaluate the relationships between the temperatures of different ionic species, they are employed in Sect. 7 to recalibrate the relationships with strong-line diagnostics.

## 5. Balmer decrements and dust attenuation

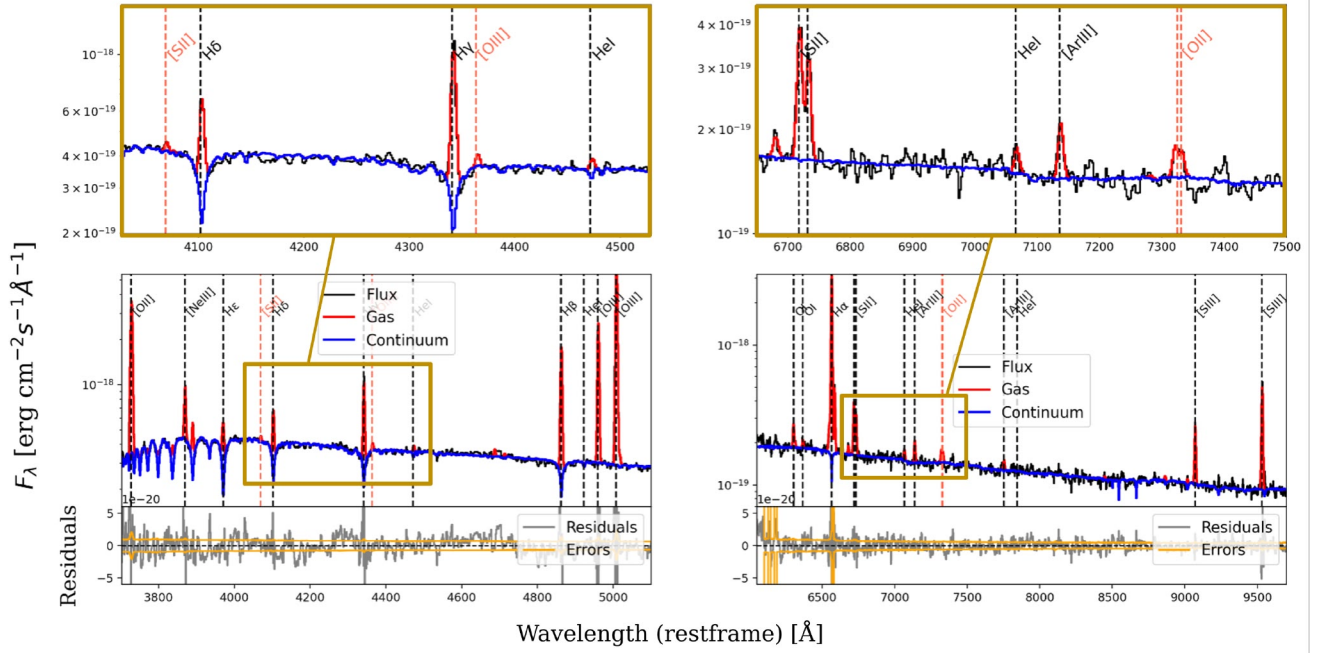
To correct the measured emission line fluxes for dust attenuation, we primarily relied on the ratios among the strongest Balmer lines observed in MARTA spectra, namely  $H\alpha/H\beta$ ,  $H\gamma/H\beta$ , and  $H\delta/H\beta$ , noting that  $H\alpha/H\beta$  typically involves computing a ratio between lines measured across the G235M and G140M gratings. We employed the Cardelli et al. (1989) extinction curve and computed the  $E(B - V)$  utilising the PYNEB routine (Luridiana et al. 2015) assuming intrinsic Balmer line ratios of  $H\alpha/H\beta = 2.86$ ,  $H\gamma/H\beta = 0.468$ , and  $H\delta/H\beta = 0.259$ , as predicted by Case B recombination assuming an electron temperature  $T_e = 10^4$  K and electron density  $n_e = 100 \text{ cm}^{-3}$  (Osterbrock & Ferland 2006). To verify the impact of these assumptions, we implemented an iterative procedure in which the extinction correction was recomputed at each step using updated values of  $T_e$  and  $n_e$  derived from the previous iteration. This process was repeated until convergence. We found that the resulting changes in  $E(B - V)$  and in the derived physical properties are minimal: both  $T_e$  and  $n_e$  vary by less than 2% across the entire sample. These small differences confirm that the use

<sup>3</sup> With the exception of one galaxy, MARTA\_4327, where such emission is interpreted as stellar in nature and associated with the presence of Wolf-Rayet stars; see Curti et al. (2025a).

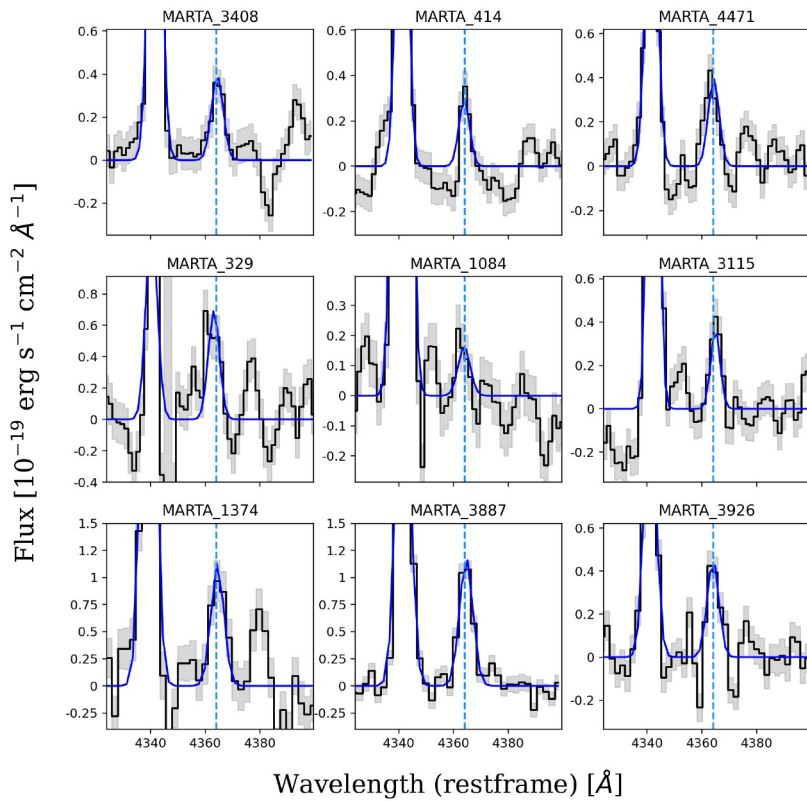


**Fig. 4.** Auroral emission lines detected in the MARTA golden sample. The panels show emission line fits of the [S II]  $\lambda 4068,4076$ , [O III]  $\lambda 4363$ , [S III]  $\lambda 6312$ , and [O II]  $\lambda 7320,7330$  transitions, with their rest-frame wavelengths marked as dashed blue lines. We show the continuum-subtracted spectra in black, and the Gaussian emission-line fits in blue. In the case of [S III]  $\lambda 6312$ , two single Gaussians are shown (dashed red lines), since the [S III] line is blended with the [O II]  $\lambda 6300$ .

## MARTA\_4502



**Fig. 5.** Spectral fit for MARTA\_4502, observed using grating G140M (left) and G235M (right). The black line represents the observed flux, the blue line the stellar continuum fit, and the red lines the emission-line Gaussian fits. In the bottom panels we show in grey the residuals from the fit, while the orange line represents the pipeline error estimates. The zoom-in panels at the top provide a more detailed view of specific spectral features: the auroral lines [S II]  $\lambda 4068$  and [O III]  $\lambda 4363$  (left) and the spectral range around the [O II]  $\lambda 7320, 7330$  auroral lines (right).



**Fig. 6.** [O III]  $\lambda 4363$  auroral emission line detected in the MARTA silver sample. The continuum-subtracted spectra are shown in black, with Gaussian emission-line fits overlaid in blue. The dashed light blue lines mark the rest-frame wavelength of the [O III].

of fixed values for the extinction correction does not introduce any significant bias in our analysis.

The  $E(B - V)$  was determined by minimising a  $\chi^2$  function defined by co-adding in quadrature the differences between the three observed Balmer line ratios and their theoretical values, given the adopted reddening curve. The resulting best-fit value was adopted as our fiducial  $E(B - V)$  and used to correct all line fluxes. The use of an alternative attenuation curve, such as the Calzetti et al. (1994) law, the Small Magellanic Cloud law (Gordon et al. 2003), or the parametrisation of Reddy et al. (2020) calibrated on  $z \sim 1-3$  galaxies does not significantly affect our results. Reddy et al. (2023) recently analysed galaxies at similar redshifts to our MARTA sample using Paschen line ratios and found a dust attenuation behaviour consistent with that of the Small Magellanic Cloud.

However, the  $E(B - V)$  values derived individually from different Balmer line ratios within the same galaxy were not always consistent among themselves, with differences reaching up to 0.3 dex in some cases. We therefore also computed the dispersion among  $E(B - V)$  values estimated from  $H\alpha/H\beta$ ,  $H\gamma/H\beta$ , and  $H\delta/H\beta$  individually, which we report in Table 2; however, this systematic uncertainty is not propagated throughout the rest of the analysis.

Discrepancies among  $E(B - V)$  as derived from different ratios of Balmer lines have been observed before at high redshifts. For instance, McClymont et al. (2025) modelled Balmer decrements in JADES observations at  $z > 2$  by studying the differential impact of radiation- and density-bounded nebulae, showing how, in the latter case, Balmer line ratios deviate from Case B values in a way that could partially explain the observed discrepancies in  $E(B - V)$  when different ratios are used. In that scenario, for moderate- to high-metallicity conditions (such as those probed by MARTA galaxies), the  $H\gamma/H\beta$  ratio is predicted to take higher values than in the case B scenario. Hence, any dust reddening would lower that ratio to closer to the case B theoretical value (mimicking no dust), while increasing  $H\alpha/H\beta$ .

Another possibility to explain such discrepancies involves a different shape of the extinction curve in high- $z$  galaxies. Recent work by Sanders et al. (2025) shows that for a galaxy at  $z = 4.41$ , the inferred extinction curve deviates significantly from both the Calzetti et al. (1994) and Cardelli et al. (1989) laws, exhibiting a steeper slope at long wavelengths and a shallower slope in the ultraviolet. This finding suggests that commonly assumed dust laws are not universally applicable to high-redshift galaxies. If a similar effect is present in our sample, the adoption of a different extinction curve could partially account for the discrepancies observed in  $E(B - V)$  estimates. A more detailed investigation into the behaviour of  $E(B - V)$  and the disagreement between different individual Balmer line ratios is beyond the scope of this work, and will be addressed in future work.

## 6. The temperature-temperature relations

To determine the chemical composition of a photoionised nebula, one needs to measure (or assume) its temperature and density structure (Stasińska 2002). A commonly adopted approach models the nebula as constituted by two main ionisation zones: a high-ionisation and a low-ionisation zone. In this framework, the electron temperature of the  $O^{2+}$  ion, referred to as  $T_3$ , characterises highly ionised species. In contrast, the electron temperature of the  $O^+$  ion, denoted  $T_2$ , represents low-ionisation species such as  $S^+$  or  $N^+$ . Some species, such as  $S^{2+}$ , have an intermediate ionisation potential between the two zones and may therefore overlap both (e.g. Berg et al. 2020).

In this section we leverage some of the first available measurements of electron temperatures from different ionic species as delivered by JWST (including  $T_2$ ,  $T_3$ , and, for a subset of targets, the temperatures associated with sulphur ions  $S^+$  and  $S^{2+}$ ) to study temperature-temperature relationships at high redshifts, assess possible evolution from their low- $z$  counterparts, and analyse trends with different physical properties. Throughout this section and the rest of the paper, we have adopted the following atomic parameters and collisional strengths: Palay et al. (2012) for [O III], Kisielius et al. (2009) and Zeippen (1982) for [O II], Rynkun et al. (2019) and Tayal & Zatsarinny (2010) for [S II], Podobedova et al. (2009) and Tayal & Gupta (1999) for [S III], for collisional strength and atomic parameters, respectively.

To estimate the uncertainties on the computed properties, we randomly perturbed the best-fit line fluxes by adding noise sampled from a Gaussian distribution with sigma equal to the flux uncertainties, and derived the corresponding electron temperature and density 100 times. The fiducial values and associated uncertainties of these quantities were then quantified as the mean and standard deviation of the distributions obtained, respectively, as reported in Table 2. Given the high S/Ns of the Balmer lines (typically  $>30-300$  for  $H\alpha$  and  $H\beta$ , and  $>10-100$  for  $H\delta$  and  $H\gamma$  across the sample), the impact of flux perturbations on the derived  $E(B - V)$  is negligible. Therefore, for computational efficiency,  $E(B - V)$  was held fixed during the Monte Carlo realisations, focusing the flux perturbations on the auroral and nebular oxygen lines and the [S II] doublet, which have a more direct influence on the electron temperature and density determinations.

### 6.1. Measuring electron temperatures and densities

Both  $T_3$  and  $T_2$  were inferred from the respective nebular-to-auroral line ratio. Specifically, we simultaneously determined  $T_2$  and  $n_e$  from the [O II]  $\lambda\lambda 7320, 7330$ /[O II]  $\lambda\lambda 3726, 3729$  ratio and the [S II]  $\lambda\lambda 6716, 6730$  density-sensitive doublet adopting the iterative procedure implemented in the `getCrossTemDen` function of PyNeb. This enabled us to take into account the non-negligible density dependence of the [O II] nebular-to-auroral ratio in the  $T_2$  derivation. We could not infer  $n_e$  from the [O II]  $\lambda\lambda 3726, 3729$  doublet, as these lines are not accurately deblended in our data. However, on average, [S II] and [O II] yield consistent estimates of electron density, as shown by, for example, Sanders et al. (2016).

We then used the inferred electron density to calculate the electron temperature of the  $O^{2+}$  ionised region, using the [O III]  $\lambda 4363$ /[O III]  $\lambda 5007$  ratio. For some of our galaxies the [S II] doublet, necessary for determining electron density, was unavailable. In three of the nine galaxies of the silver sample (at  $z > 3$ ) the doublet fell outside the observed spectral range. In other cases, the lines fell into the detector gap (two of the nine galaxies), or the spectrum was too noisy in its red part to detect the doublet (MARTA\_329). In these cases we assumed an electron density of  $300 \text{ cm}^{-3}$ , consistent with the typical densities measured in high- $z$  galaxies (e.g. Sanders et al. 2016; Stanton et al. 2025; Topping et al. 2025a). In one case (MARTA\_3408), the ratio between the two sulphur lines placed the galaxy in the low-density limit, and we therefore assumed  $n_e = 100 \text{ cm}^{-3}$ . Nonetheless, the derivation of  $T_3$  is only marginally affected by density variations, and we verified that repeating the calculations with density values in the range between 10 and  $1000 \text{ cm}^{-3}$  does not change the results.

The temperature of the singly ionised sulphur ( $T$ [S II]) and of the doubly ionised sulphur ( $T$ [S III]) were determined follow-

ing a similar approach.  $T[\text{S II}]$  was computed using the auroral-to-nebular ratio  $[\text{S II}] \lambda\lambda 4068, 4076 / [\text{S II}] \lambda\lambda 6716, 6730$ , while  $T[\text{S III}]$  was obtained from the ratio  $[\text{S III}] \lambda 6312 / [\text{S III}] \lambda 9071$ . Both were inferred using the `getTemDen` function in `PyNeb`, adopting the electron density previously determined in the  $T_2$  calculation. We also verified that measuring  $T[\text{S II}]$  and  $n_e$  simultaneously with `getCrossTemDen` (as done for  $T[\text{O II}]$ ), instead of fixing the density to the previously inferred value, provides fully consistent outcomes. The fluxes of all measured auroral lines in MARTA galaxies, including sulphur ones, are listed in Table D.1.

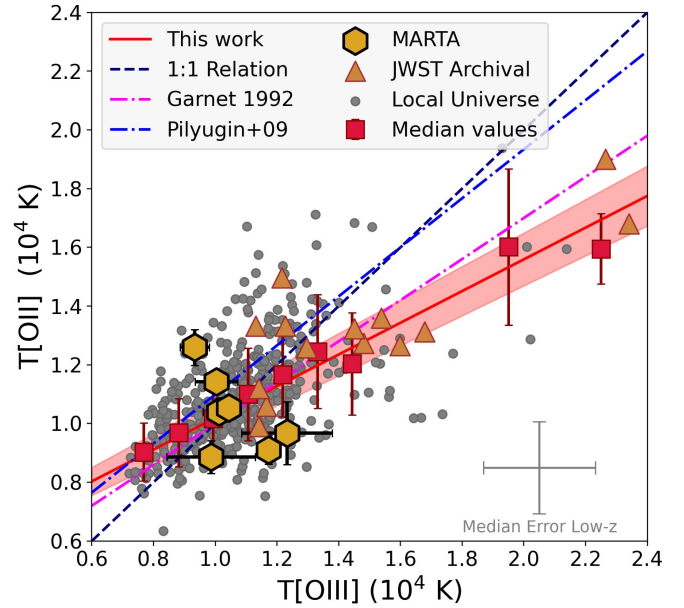
## 6.2. The $T_2$ – $T_3$ relation

In Fig. 7 we present the measurements of the MARTA golden sample on the  $T_2$ – $T_3$  relation (golden hexagons), and compare them with a compilation of literature data at both low and high redshifts (orange triangle markers). Currently, detections of both the auroral lines required for the calculation of  $T_2$  and  $T_3$  have only been made for a small number of high-redshift galaxies (i.e.  $z > 1$ ). The high-redshift comparison sample includes SGAS 1732 from Welch et al. (2024) and the Sunburst Arc from Welch et al. (2025), as well as CEERS 11088 and CEERS 3788 from Sanders et al. (2024), with redshifts ranging from  $\sim 1$  to 3.5, and temperatures from the compilation of Chakraborty et al. (2025) at redshift  $\sim 3$ –10 taken from JADES (D’Eugenio et al. 2025) and Morishita et al. (2024a), Welch et al. (2024, 2025), Rogers et al. (2024), and Sanders et al. (2024), for a total sample of 21 objects.

For high- $z$  studies, we compiled the emission line fluxes where available and re-derived the electron temperatures employing the same methodology as for MARTA galaxies, which includes adopting a self-consistent set of atomic parameters. In fact, using different sets of collision strengths results in temperature estimates that can differ by  $\geq 500$  K, especially for the  $\text{O}^{++}$  ion (Nicholls et al. 2013). Therefore, when only temperatures were tabulated but we had no access to the measured line fluxes, we corrected the compiled temperatures to account for such a discrepancy, if needed.

In particular, we used `PYNEB` to model the difference in the inferred  $T_3$  from the Palay et al. (2012) dataset (employed in this work) and the datasets from Storey et al. (2014) and Aggarwal & Keenan (1999) (which provide comparable results to each other), and applied such an offset to the  $T_3$  values from the literature that were derived adopting either of the two latter atomic parameters (as in e.g. Chakraborty et al. 2025). This approach brings all  $T_3$  values analysed in this work to a consistent scale. Such a correction was not required for  $T_2$  as all the literature studies considered adopted the same atomic parameters.

We also compared these data with a large compilation of measurements in low-redshift galaxies (small grey points in Fig. 7), for which we compiled emission-line fluxes and re-measured temperatures self-consistently with the high- $z$  sample. In particular, we included data from low-metallicity local emission-line galaxies from Izotov et al. (2009) and Guseva et al. (2009, 2011); the Curti et al. (2017) sample, consisting of composite spectra of SDSS Data Release 7 star-forming galaxies stacked based on their combined  $[\text{O II}]/\text{H}\beta$  and  $[\text{O III}]/\text{H}\beta$  flux ratios; a sample of SDSS galaxies where both oxygen auroral lines were detected in individual objects (see e.g. Pilyugin et al. 2012); the CHAOS data targeting H II region in massive local spiral galaxies with the MODS spectrograph on the LBT (Large Binocular Telescope) (Berg et al. 2015, 2020;



**Fig. 7.** Relation between  $T[\text{O III}]$  and  $T[\text{O II}]$  for the MARTA sample (gold hexagons), compared with various samples from the literature at low (grey dots) and high redshifts (from JWST archival data, orange triangles). The plot also shows the 1:1 relation (dashed blue line) and two empirical calibrations of the  $T_2$ – $T_3$  relation, from Pilyugin et al. (2009, dashed purple line) and Garnett (1992, dashed magenta line). The red squares represent the median values of the entire sample, binned in  $T_3$ , with uncertainties given by the standard deviation within each bin. The red line shows the best-fit relation derived in this work.

Rogers et al. 2021, 2022); the stacks of Khoram & Belfiore (2025), which consist of spatially resolved spectra from SDSS-IV MaNGA (Mapping Nearby Galaxies at Apache Point Observatory) stacked in bins across the  $M_\star$ –SFR plane; and the Zurita et al. (2021) compilation of 2831 published H II region emission-line fluxes in 51 nearby spiral galaxies. In all these cases we recovered the measured extinction-corrected fluxes and re-calculated temperatures following the same approach used for the MARTA data, applying a uniform S/N > 3 cut on the auroral lines.

Furthermore, in Fig. 7 we also plotted some empirical calibrations of the  $T_2$ – $T_3$  relation. In particular, Pilyugin et al. (2009) calibrated the relation on a sample of H II regions from the local Universe, while Garnett (1992) provided a fit based on photoionisation models.

Some galaxies in the local sample exhibited evident contamination from  $[\text{Fe II}]$  (e.g. Curti et al. 2017; Rogers et al. 2021; Hamel-Bravo et al. 2024, see also Moreschini et al., in prep.). This contamination led to artificially elevated  $T_3$  values due to the consequent overestimation of the  $[\text{O III}]\lambda 4363$  flux. To exclude contaminated galaxies in our analysis, we applied an empirical cut (discussed in Moreschini et al., in prep) defined as  $\log(R/R_3) > -2.1 + \log(\text{O}32)$ , where  $R \equiv [\text{O III}]\lambda 4363/\text{H}\beta$ , as contaminated. This cut ensured the exclusion of the vast majority of the SDSS stacks flagged as contaminated by Curti et al. (2017) through visual inspection.

The MARTA data shown in Fig. 7 (orange hexagons) exhibit significant scatter, which is also evident in the data from local galaxies in the literature (grey points, typical errors shown in light grey). The red squares in Fig. 7 indicate the median values for the whole sample, binned in  $T_3$ , with error bars reporting the standard deviation within each bin. The binning was chosen

to ensure a minimum of five galaxies per bin, adapting the bin width in the high- $T_3$  regime, where the data are more sparsely sampled. The red line represents a linear fit derived in this work, based on the entire sample. This fit was performed using an orthogonal linear regression algorithm, applied to the median values rather than individual points, to reduce the influence of sampling effects in the diagram. The best-fit linear relation is given by

$$T_2 = (0.54 \pm 0.03) T_3 + (4790 \pm 300) \text{ K}. \quad (1)$$

The slope obtained from our fit is significantly shallower than those reported by, for example, Garnett (1992) and Pilyugin et al. (2009). This effect arises from the inclusion of high-temperature points, which tend to be systematically lower in  $T_2$  than in  $T_3$  when  $T_3 > 14\,000$  K. This behaviour is seen consistently in both low- and high-redshift data. This high- $T_3$  range is, however, populated by only  $\sim 30$  data points in our large compilation and has therefore been previously hard to confirm observationally in the literature. Hints of this behaviour were, however, already observed by Méndez-Delgado et al. (2023a) and Arellano-Córdova & Rodríguez (2020) when studying the  $T[\text{N II}]-T[\text{O III}]$  relation.

We performed an alternative fit considering only the range  $T_3 = [7000-14\,000]$  K, and we obtained a steeper slope of  $0.64 \pm 0.04$ , closer to previous calibrations. Furthermore, we compared our orthogonal fit with the median values of the bin with a Bayesian fit performed on individual points using the LINMIX method (Kelly 2007). The results are consistent within two sigma, with the Bayesian fit yielding a slope of  $0.50 \pm 0.03$  and an intercept of  $5510 \pm 330$  K.

We find that the  $T_2-T_3$  relation is affected by significant scatter across redshift. This scatter could reflect both observational uncertainties and the adoption of different atomic datasets, as well as be driven by the combination of different physical processes (see Sect. 6.3). For instance, the use of [O II] as a tracer of the low-ionisation zone temperature is particularly controversial. While models predict that  $T[\text{O II}]$  should be comparable to  $T[\text{S II}]$  and  $T[\text{N II}]$  (Campbell et al. 1986; Pilyugin et al. 2006), in practice  $T[\text{O II}]$  tends to overestimate the temperature compared to  $T[\text{N II}]$  and  $T[\text{S II}]$  (Zurita & Bresolin 2012; Curti et al. 2017; Berg et al. 2020; Méndez-Delgado et al. 2023b). This discrepancy may be explained by density inhomogeneities within H II regions, which affect the [O II] and [S II] ratios but have little impact on the [N II] lines (Méndez-Delgado et al. 2023b), as well as from the effect of recombination on the emissivity of the [O II] auroral doublet.

Nonetheless, our observations shows that high- $z$  systems are overall consistent with the distribution of local galaxies on the  $T_2-T_3$  relationship, with no systematic offset observed. This finding is in good agreement with the independent analysis by Sanders et al. (2025), who likewise report no evidence of an evolution of the  $T_2-T_3$  relation with redshift. This consistency suggests that, while the relationship is not particularly tight, the underlying physical processes governing  $T_2$  and  $T_3$  do not vary significantly across cosmic time.

### 6.3. Dependence of the $T_2-T_3$ relation on additional physical parameters

We explored the possible dependence of the  $T_2-T_3$  relation on additional physical properties to probe the origin of the observed scatter. In particular, we considered three physical parameters: electron density derived from the [S II] ratio ( $n_e$ ), the [O III]/[O II]

line ratio as a proxy for the ionisation parameter, and the softness parameter (Vilchez & Pagel 1988), defined as

$$\eta \equiv \frac{([\text{O II}]3726 + [\text{O II}]3729)/([\text{O III}]4959 + [\text{O III}]5007)}{([\text{S II}]6730 + [\text{S II}]6716)/([\text{S III}]9069 + [\text{S III}]9532)}. \quad (2)$$

This parameter is used as a proxy for the hardness of ionising radiation, with a secondary dependence on gas-phase metallicity and other nebular parameters. Since galaxies at high redshifts are known to show a different distribution in these three parameters compared to their low-redshift counterparts (e.g. Sanders et al. 2016; Kaasinen et al. 2018), a dependence of the scatter or the residuals of the  $T_2-T_3$  relation on these parameters would provide evidence of a potential redshift evolution.

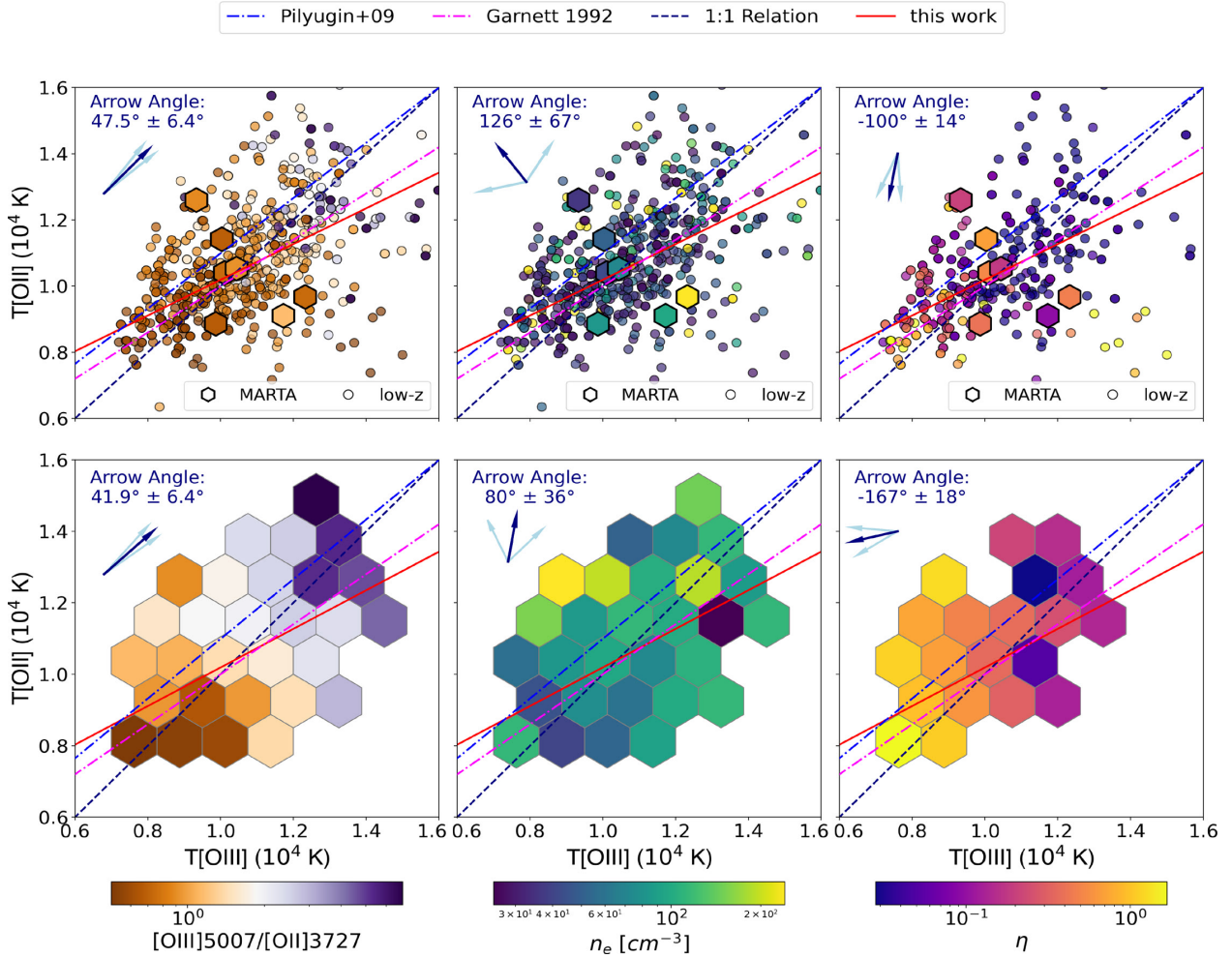
Figure 8 presents six panels displaying the  $T_2-T_3$  relation colour-coded by [O III]/[O II],  $n_e$ , and  $\eta$  for both individual points and hexagonal bins. The hexagonal bins were computed with a minimum of five objects per bin to ensure statistical robustness. The  $\eta$  diagram contains fewer data points compared to the first two, as only some galaxies in our sample had detections of all four emission lines required to compute  $\eta$ . In particular, the [S III]  $\lambda 9070, 9532$  are not available for SDSS galaxies, as they fall outside the observed spectral coverage.

To assess potential correlations between the temperatures, gas density, [O III]/[O II] ratio, and  $\eta$ , we conducted a partial correlation analysis on the full sample, including both local and high-redshift galaxies, using both individual data points and hexagonal bins. The corresponding partial correlation coefficients ( $r$ ), 95% confidence intervals (CI95%), and  $p$ -values are summarised in Table 3. The table reports all partial correlation coefficients computed for both binned and non-binned data, enabling a direct comparison of the two approaches.

The results indicate that both  $T_3$  and  $T_2$  exhibit a strong positive correlation with the [O III]/[O II] ratio. This correlation is significantly stronger when computed on the hexagonal bins compared to individual data points, suggesting that binning reduces scatter and enhances the underlying trend. However, the correlation is statistically significant in both cases, confirming its robustness.

In contrast,  $T_2$  and  $T_3$  exhibit weak correlations with gas density, both from single points and binned data. However, a marginally significant ( $\sim 2-4\sigma$ ) positive correlation is observed between  $T_2$  and density. The stronger correlation between  $T_2$  and density (compared to  $T_3$  and density) may arise because the [S II] and [O II] diagnostics, being low-ionisation tracers, probe similar gas regions more effectively than [O III] and [S II]. Moreover, the [O II] auroral-to-nebular line ratio used for temperature diagnostics is highly sensitive to density variations, potentially inducing a spurious trend. A more comprehensive assessment of this trend would require density diagnostics for higher-ionisation lines, which could provide a more detailed representation of the emitting gas distribution and allow us to test a potential dependence of  $T_3$  on density.

Regarding  $\eta$ , the results exhibit a strong dependence on binning, with substantial differences between binned and non-binned analyses. In the non-binned data, strong and seemingly significant negative correlations are observed with both temperatures. However, the correlation with  $T_2$  is likely driven by a sparsely populated region at high  $T_3$  and high  $\eta$ , which skews the overall trend downwards. These high- $T_3$ , high- $\eta$  data points are absent in the binned analysis, leading to a much weaker correlation in the hexbin results. In contrast, the correlation with  $T_3$  remains more robust, as it persists in both the binned and non-binned cases, suggesting a more intrinsic relationship between  $T_3$  and  $\eta$ .



**Fig. 8.** Dependence of the  $T_2$ – $T_3$  relationship on different physical properties. Top row: Individual data points. Bottom row: Corresponding hexagonal binning. The colour-coding indicates the three selected physical properties, computed as median values within each bin in the bottom panels: the  $[\text{O III}]/[\text{O II}]$  line ratio (left), the electron density  $n_e$  (centre), and the softness parameter  $\eta$  (right). In the top panels, MARTA objects are shown as hexagons and low-redshift galaxies as circles. The plots also include the Pilyugin et al. (2009) relation, the Garnett (1992) relation, the 1:1 relation, and our best-fit model. The direction of maximum variation in the colour-coded quantity, determined by the gradient angle, is indicated by navy arrows, with the associated uncertainty shown for reference as light blue arrows. This provides a visual representation of how each parameter varies across the  $T_2$ – $T_3$  plane.

To further explore these trends, we examined the gradient angle of the colour-coded parameters within the  $T_2$ – $T_3$  plane, which is displayed as a navy arrow in each panel of Fig. 8. The gradient angle represents the direction of maximum variation in a given physical parameter (e.g.  $[\text{O III}]/[\text{O II}]$ ) and is computed as

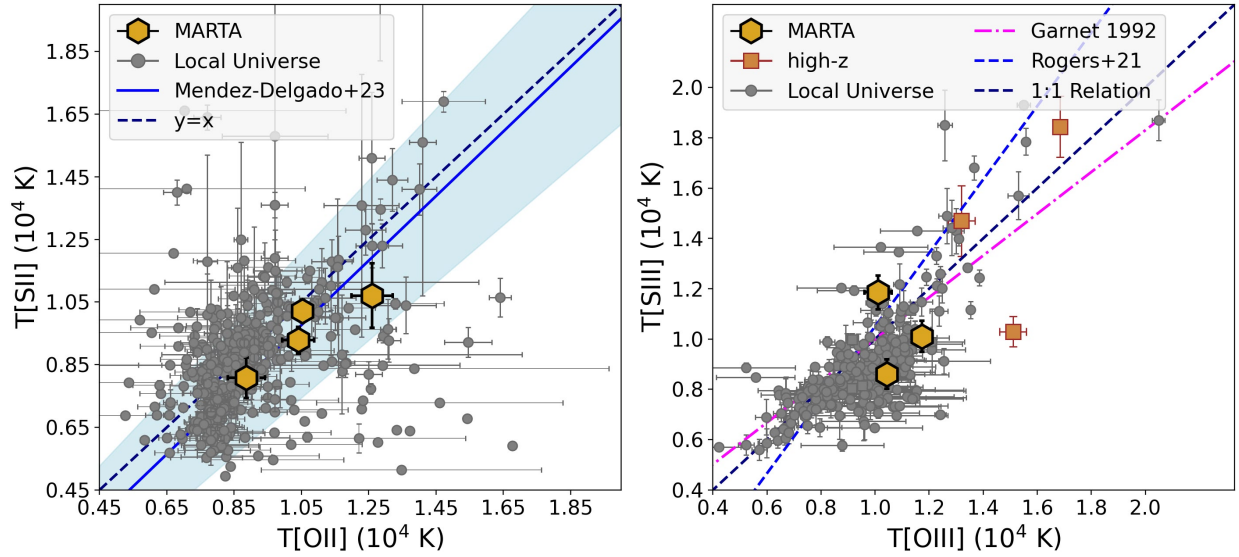
$$\theta_{\text{grad}} = \arctan\left(\frac{r_{T_2,Z}}{r_{T_3,Z}}\right), \quad (3)$$

where  $r_{T_2,Z}$  and  $r_{T_3,Z}$  are the partial correlation coefficients of  $T_2$  and  $T_3$  with a given property  $Z$  (see e.g. Bluck et al. 2019; Curti et al. 2022; Baker et al. 2023). We quantified the uncertainty in the gradient angle via a bootstrapping procedure, randomly resampling the dataset (with replacement) 300 times and recalculating the angle for each iteration. The standard deviation of the resulting angle distribution provides the error, represented by the light blue arrows in Fig. 8. We tested using a larger number of samples for bootstrapping and the inferred uncertainty remains consistent.

Notably, the large errors observed in the central panels of Fig. 8 reflect the weak statistical significance of the partial correlation analysis for  $n_e$ . The visual near-alignment of

the  $[\text{O III}]/[\text{O II}]$  gradient angle with the best-fit  $T_2$ – $T_3$  relation seems to point that the orthogonal scatter within this relation is not strongly correlated with this parameter. Such a correlation could arise from a metallicity-driven effect. In fact, a higher metallicity leads to softer ionising spectra (McGaugh 1991; Bresolin et al. 1998), lowering the ionisation parameter and reducing the  $[\text{O III}]/[\text{O II}]$  ratio. Conversely, a lower metallicity results in higher ionisation parameters and  $[\text{O III}]/[\text{O II}]$  ratios (Grasha et al. 2022; Ji & Yan 2022). Since lower-metallicity environments also exhibit higher temperatures, this trend naturally explains the observed correlation. Nevertheless, a more quantitative assessment is needed to robustly evaluate the contribution of the ionisation parameter to the scatter around the  $T_2$ – $T_3$  relation.

In addition, we performed an alternative correlation analysis based on the residuals of the  $T_2$ – $T_3$  relation relative to our best-fit linear fit in Eq. (1). In particular, we defined the quantity  $\Delta T$  for each point in the sample as its orthogonal deviation from Eq. (1), measured as  $(\Delta T)^2 = (\Delta T_3)^2 + (\Delta T_2)^2$ . We then calculated the Spearman correlation coefficient between  $\Delta T$  and each of the physical parameters involved in the analysis. The results reveal a mild yet statistically significant positive correlation (coefficient



**Fig. 9.** Left panel:  $T[\text{S II}]-T[\text{O II}]$  relation for MARTA objects, compared with extragalactic H II regions together with the empirical calibration by Méndez-Delgado et al. (2023b), based on local H II regions. Right panel:  $T[\text{S III}]-T[\text{O III}]$  relation for MARTA objects, along with local H II region data and two high-redshift galaxies with archival JWST data (Rogers et al. 2024; Welch et al. 2025). The plot also shows the calibrations from Garnett (1992) and Rogers et al. (2021).

of 0.15, significant at  $\sim 3\sigma$ ) between the amplitude of the offset from the temperature relation  $\Delta T$  and the  $[\text{O III}]/[\text{O II}]$  ratio. This suggests that, although the effect is not dominant, variations in the ionisation parameter – as traced by  $[\text{O III}]/[\text{O II}]$  – do contribute to the scatter around the  $T_2-T_3$  relation.

In contrast, both  $n_e$  and  $\eta$  show minimal effect and weak statistical significance. These results suggest that the orthogonal scatter relative to the best-fit relation does not exhibit a clear dependence—at least not one detectable from the residuals—on these two quantities.

#### 6.4. Sulphur temperatures

Sulphur is a valuable tracer of ionised gas conditions in galaxies, as its ionisation states probe different regions of the ISM. The ionisation potentials of  $[\text{S III}]$  and  $[\text{S II}]$  ions are slightly different from those of  $[\text{O III}]$  and  $[\text{O II}]$ ; the former trace intermediate- and the latter low-ionisation regions. Comparing sulphur and oxygen temperatures is therefore a useful diagnostic for understanding the ionisation structure of H II regions and testing photoionisation models.

The availability of reliable detections of sulphur auroral lines in some of our selected high-redshift galaxies presents a valuable opportunity to compare such temperatures. Our work marks one of the first instances where this relationship is examined for galaxies at high redshifts (Rogers et al. 2024; Welch et al. 2024, 2025; Morishita et al. 2025). Photoionisation models predict that  $T[\text{S II}] \approx T[\text{O II}]$  (Campbell et al. 1986; Pilyugin et al. 2006; Méndez-Delgado et al. 2023).  $T[\text{O III}]$ , on the other hand, is expected to be higher than  $T[\text{S III}]$  for  $[\text{O II}]$  temperatures higher than 8000 K (Garnett 1992). Substantial deviations from both these trends have, however, been observed in extragalactic H II regions (Kennicutt et al. 2003; Esteban et al. 2009; Berg et al. 2015; Croxall et al. 2016; Rogers et al. 2021).

As shown in Fig. 9, the observed trends between the temperatures of sulphur and oxygen ions are consistent with the previous literature and empirical calibrations. In particular, we compared the MARTA objects with extragalactic H II regions

**Table 3.** Results of the partial correlation analysis on the  $T_2-T_3$  relation.

Correlation	$r$	CI95%	$p$ -value ( $\sigma$ )
Individual data points			
$T_2$ vs $n_e$   $T_3$	0.15	[0.08, 0.23]	$10^{-4}$ ( $\sim 3.9\sigma$ )
$T_3$ vs $n_e$   $T_2$	0.03	[-0.04, 0.11]	0.4 ( $<1\sigma$ )
$T_2$ vs O32   $T_3$	0.57	[0.51, 0.62]	$<10^{-6}$ ( $>5\sigma$ )
$T_3$ vs O32   $T_2$	0.34	[0.28, 0.41]	$<10^{-6}$ ( $>5\sigma$ )
$T_2$ vs $\eta$   $T_3$	-0.38	[-0.46, -0.29]	$<10^{-6}$ ( $>5\sigma$ )
$T_3$ vs $\eta$   $T_2$	-0.50	[-0.57, -0.42]	$<10^{-6}$ ( $>5\sigma$ )
Hexagonal bins			
$T_2$ vs $n_e$   $T_3$	0.35	[-0.03, 0.64]	0.07 ( $\sim 1.8\sigma$ )
$T_3$ vs $n_e$   $T_2$	0.02	[-0.35, 0.39]	0.9 ( $<1\sigma$ )
$T_2$ vs O32   $T_3$	0.80	[0.61, 0.90]	$<10^{-6}$ ( $>5\sigma$ )
$T_3$ vs O32   $T_2$	0.86	[0.71, 0.93]	$<10^{-6}$ ( $>5\sigma$ )
$T_2$ vs $\eta$   $T_3$	-0.05	[-0.49, 0.41]	0.8 ( $<1\sigma$ )
$T_3$ vs $\eta$   $T_2$	-0.88	[-0.95, -0.71]	$<10^{-6}$ ( $>5\sigma$ )
Orthogonal offsets from the $T_2-T_3$ relationship best fit			
$\Delta T$ vs O32	0.151	[0.05, 0.24]	$2.2 \times 10^{-3}$ ( $\sim 3\sigma$ )
$\Delta T$ vs $n_e$	0.112	[0.02, 0.21]	0.02 ( $\sim 2.3\sigma$ )
$\Delta T$ vs $\eta$	0.088	[-0.01, 0.18]	0.07 ( $\sim 1.8\sigma$ )

**Notes.** The table reports the Spearman correlation coefficients ( $r$ ), their 95% confidence intervals (CI95%), and  $p$ -values (converted into significance in  $\sigma$  where appropriate), where O32  $\equiv$   $[\text{O III}]/[\text{O II}]$ . The first set of values corresponds to the analysis performed on the full sample, while the second set refers to the analysis conducted on median values within hexagonal bins. The bottom part of the table reports instead the analysis of the correlation between the same physical parameters and the orthogonal offset  $\Delta T$  from the best-fit relation of Eq. (1), defined as in Sect. 6.3.

(Esteban et al. 2009, 2014, 2020; Domínguez-Guzmán et al. 2022; Rogers et al. 2022). The left panel of Fig. 9 shows the  $T[\text{S II}]-T[\text{O II}]$  relation and the empirical calibration by Méndez-Delgado et al. (2023a), based on local H II regions. Our high-redshift sample lies systematically below the 1:1

relation, but such a trend aligns with the findings of Méndez-Delgado et al. (2023a), who also report an offset in the relation, with [O II] temperatures generally higher than [S II] ones. They attributed this effect to the presence of density fluctuations.

In the right panel of Fig. 9 we show the  $T[\text{S III}]-T[\text{O III}]$  relation. This plot also includes the sample of H II regions from CHAOS (Berg et al. 2015, 2020; Croxall et al. 2015, 2016), and three high-redshift galaxies: the Sunburst Arc (Welch et al. 2025), Q2343-D40 from Rogers et al. (2024), and ID60001 from Zhang et al. (2025), which, to the best of our knowledge, are the only other galaxies at high redshifts with simultaneous measurements of the [S III] and [O III] auroral lines. The figure also shows two calibrations from the literature. The relation in Garnett (1992) was based on photoionisation models and the one in Rogers et al. (2021) was calibrated on local H II regions from the CHAOS survey.

The relation appears somewhat scattered but is still broadly consistent with results from the local literature and comparable to the other high-redshift galaxies. This result is also consistent with the recent findings of Sanders et al. (2025), who reported no evidence of systematic redshift evolution in  $T-T$  relations involving sulphur ions. In models, these two temperatures are not always equal, since [S III] and [O III] likely trace different ionisation zones due to their significantly different ionisation potentials, as discussed in Berg et al. (2021).

Quantitatively, our  $T[\text{S II}]$  measurements show an average negative offset of  $\sim 930$  K from the one-to-one relation and  $\sim 470$  K from the Méndez-Delgado et al. (2023a) relation. These correspond to average deviations of  $1.14\sigma$  and  $0.52\sigma$  (considering our estimates of the temperatures error bars), respectively, indicating that our measurements remain well within expectations based on previous studies. Notably, three out of the four galaxies lie directly on the best-fit relation of Méndez-Delgado et al. (2023b), while the fourth remains consistent within the best-fit uncertainty range.

Similarly, for  $T[\text{S III}]$ , we find the average of the absolute value offset to be  $1.9\sigma$  from the 1:1 relation, while the offsets relative to Garnett (1992) and Rogers et al. (2021) are  $1.9\sigma$ , and  $2\sigma$ , respectively. This time the offset is not systematically negative but the data scatter around the relation.

## 7. Strong-line metallicity diagnostics and calibrations

Strong-line methods based on combination of nebular lines ratios including [O II] $\lambda\lambda 3726, 3729$ , [O III] $\lambda 4959, \lambda 5007$ , and [N II] $\lambda 6584$  are a valuable way to measure the metallicity of galaxies, in virtue to their brightness and ease of detection across a wide range of redshifts (Maiolino & Mannucci 2019). Nonetheless, the redshift evolution in the physical conditions that are at the basis of the (either direct or indirect) dependence between the involved line ratios and metallicity possibly translates into biases when using such methods to derive the metallicity from high- $z$  spectra adopting diagnostics calibrated on local galaxy samples (Bian et al. 2018; Sanders et al. 2021).

The advent of JWST promised to directly tackle this issue by delivering robust auroral lines detections in high- $z$  galaxies, and this was indeed one of the key objectives of several programmes within the first few cycles. In this section we combine  $T_e$ -based metallicity measurements for MARTA galaxies with a compilation of measurements from the literature at different redshifts, and test the validity and evolution of various classical strong-line calibrations in the high- $z$  Universe.

**Table 4.** Polynomial coefficients of the best fit for the strong-line metallicity diagnostics.

Diagnostic	A	B	C	D	E
R3	0.158	-1.977	-1.629	-0.329	-
R2	0.290	-1.202	-4.104	-3.167	-0.812
R23	0.515	-1.614	-1.421	-0.286	-
$\tilde{R}$	-0.055	-3.049	-3.168	-0.807	-
O32	-0.537	-2.832	-1.200	-0.002	-
Ne3O2	-1.679	-2.034	-0.458	-	-
N2	-0.910	0.486	-	-	-
O3N2	0.619	-3.794	-2.561	-0.382	-

**Notes.** The fitted functional form is  $f(x) = A + Bx + Cx^2 + Dx^3 + Ex^4$ , where  $f$  is the line ratio diagnostic defined as in Sect. 1 and  $x$  represents the gas-phase metallicity, normalised to the solar value ( $12 + \log(\text{O}/\text{H}) - 8.69$ ). The polynomials are fitted over the range  $12 + \log(\text{O}/\text{H}) \in [7.0, 8.4]$  and include a regularisation at high and low metallicity based on the Curti et al. (2017) calibrations to stabilise the fit where direct  $T_e$  measurements are not available. Caution is advised when extrapolating these relations beyond the fitted range.

### 7.1. Oxygen abundance determination

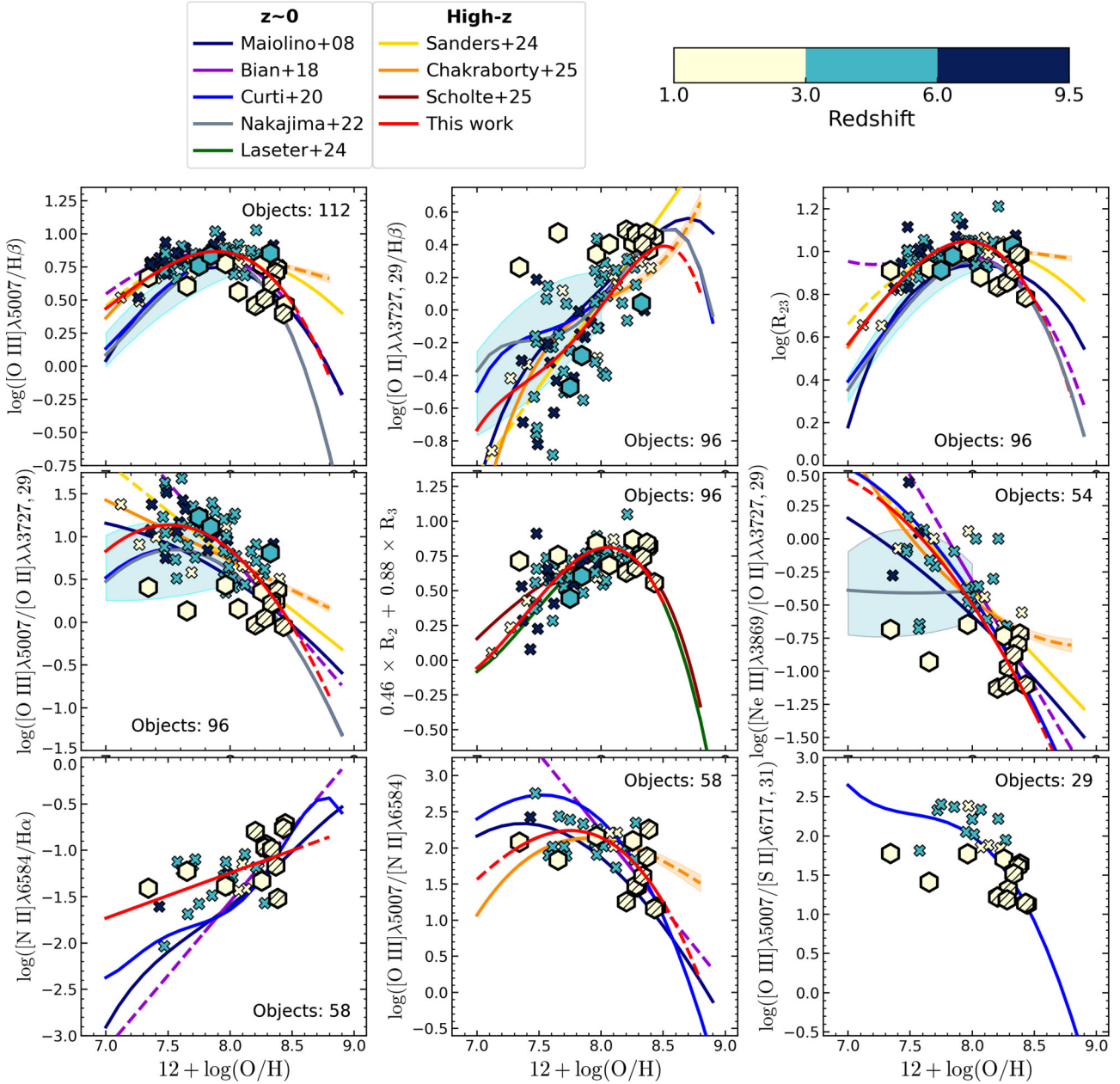
We derived the ionic abundances of the  $\text{O}^+$  and  $\text{O}^{++}$  ions in the MARTA golden sample galaxies leveraging the direct measurement of the respective temperatures and employing the GetIonicAbundances routine in pyNeb. To measure the  $\text{O}^+$  abundance in the MARTA silver sample, we inferred the  $T_2$  from the  $T_3$  as measured from [O III] $\lambda 4363$ /[O III] $\lambda 5007$  ratio adopting our newly derived fit of Eq. (1). The total oxygen abundance is then computed as the sum of these two ionic abundances, with the contribution from higher ionisation states like  $\text{O}^{3+}$  assumed to be negligible. The results of these abundance calculations are presented in Table 2.

In addition, we complemented our sample with literature NIRSPEC data of galaxies that were detected in the [O III]  $\lambda 4363$  line. We compiled the fluxes (where publicly available) and self-consistently performed extinction correction, temperature, and abundance calculations following the same methodology applied for the MARTA silver sample. The literature sample includes data from D'Eugenio et al. (2024), Navarro-Carrera et al. (2024), Sanders et al. (2024), Schaerer et al. (2024), Arellano-Córdova et al. (2025), Curti et al. (2025b), Morishita et al. (2025), Napolitano et al. (2025), Scholte et al. (2025), Stiavelli et al. (2025), Topping et al. (2025b). Our total sample, combining MARTA and literature sources, consists of 128 galaxies.

As not all strong lines are always detected in literature objects (because of either low sensitivity or limited spectral coverage), each strong-line ratio can be effectively measured in a different number of objects, as reported in each panel in Fig. 10. As an example, for R23 we present a calibration based on 96 objects, nonetheless nearly doubling previous largest compilation of Chakraborty et al. (2025).

### 7.2. New JWST-based metallicity calibrations

In Fig. 10 we show the position of our combined JWST sample on different strong-line ratios diagnostics versus metallicity diagrams. MARTA galaxies primarily occupy the high-metallicity



**Fig. 10.** Nine different strong-line diagnostics comparing our MARTA data (striped hexagons for the golden sample, plain hexagons for the silver) with high- $z$  literature samples (crosses) and a set of empirical calibrations. The data are colour-coded by redshift bin. The high-redshift data include strong-line calibration measurements from JWST surveys (D’Eugenio et al. 2024; Navarro-Carrera et al. 2024; Sanders et al. 2024; Schaerer et al. 2024; Arellano-Córdova et al. 2025; Curti et al. 2025b; Morishita et al. 2025; Napolitano et al. 2025; Scholte et al. 2025; Stiavelli et al. 2025; Topping et al. 2025b). The empirical calibration curves from the local Universe are from Maiolino et al. (2008), Bian et al. (2018), Curti et al. (2020a), and Nakajima et al. (2022), colour-coded in a dark colour palette (blue, purple, green), while the high-redshift-based calibrations from Sanders et al. (2024) and Chakraborty et al. (2025) are represented with a warm colour palette (yellow, orange, and red). Dashed lines denote extrapolations of the calibrations outside the regime where they have been calibrated. Our best-fit relations to the high-redshift dataset are shown in red. In most cases, a third-degree polynomial was used; for the N2 and R2 diagnostics, we adopted a linear fit, as there were no particular features that required a higher-order polynomial. Furthermore, in the  $\bar{R}$  panel (central one), the Laseter et al. (2024) and Scholte et al. (2025) empirical re-calibration are displayed as well.

end of the distribution, a region largely unexplored at high redshifts. We compare our combined sample with empirical calibration derived from both local and high-redshift galaxy studies.

The local calibrations we considered include those from Maiolino et al. (2008), which are based on low-metallicity galaxies from Nagao et al. (2006) and photoionisation model-based oxygen abundances for SDSS galaxies in the high-metallicity

regime. We also included the calibrations by Curti et al. (2017), later revisited in Curti et al. (2020b), which are based on stacked spectra of SDSS galaxies; and those by Nakajima et al. (2022), who extended the Curti et al. (2017) SDSS stacks by including extremely metal-poor galaxies from the EMPRESS survey. Additionally, we used the calibrations by Bian et al. (2018), derived from stacked spectra of local high-redshift analogues selected based on their location on the [N II]-BPT diagram.

For comparison, we also included recent works that recalibrate the same strong-line diagnostics using JWST data. In particular, we considered the calibrations from Sanders et al. (2024), based on galaxies at  $z = 2-9$  from the CEERS survey Finkelstein et al. (2023), and those from Chakraborty et al. (2025), based on galaxies from the CEERS, JADES (D'Eugenio et al. 2025), and UNCOVER (Bezanson et al. 2024) surveys, with fluxes compiled by the PRIMAL project (Heintz et al. 2025).

All curves corresponding to the different calibrations shown in Fig. 10 are plotted as solid lines within the directly probed metallicity range, while polynomial extrapolations beyond the calibration range are shown as dashed lines. We also provide new fits to the strong-line ratios as a function of metallicity based on the full JWST sample compiled in this work. The fitting procedure was implemented as a direct polynomial regression, where the coefficients were optimised using the `least_squares` routine from the `SCIPY` library. To limit the influence of individual outliers, we adopted the Huber loss function<sup>4</sup>. This approach retains the efficiency of least-squares fitting in the well-sampled regime, while improving robustness in the presence of non-Gaussian scatter. In most cases, we adopted a third- or a fourth-order polynomial, while for the N2 diagnostic, which does not exhibit significant curvature, we opted for a linear model.

These polynomial fits are constrained by observations in the range  $12 + \log(\text{O}/\text{H}) \in [7.0-8.4]$ . Outside this interval, the lack of direct constraints makes extrapolation highly uncertain. To stabilise the behaviour at low and high metallicities, the fits were regularised by anchoring them to the analytic calibrations of Curti et al. (2017). In practice, we introduced additional constraint points at metallicities beyond the observed range, with target values set by these empirical relations. These points were assigned enhanced weights in the loss function, ensuring that the fits remain driven by the data where observations are available, while gradually converging to the expected trends at the boundaries. The regression was performed on metallicities normalised to the solar abundance ( $12 + \log(\text{O}/\text{H}) = 8.69$ ), so that the best-fit coefficients reported in Table 4 are expressed as functions of the normalised abundance  $Z_{\text{norm}} = (12 + \log(\text{O}/\text{H})) - 8.69$ . The resulting calibrations are shown in red in Fig. 10. We did not report fits for the O3N2 diagnostic, for reasons discussed below. The calibrations presented in this work are based on a dataset that is potentially biased since we only based the calibrations on a subset of galaxies for which auroral lines have been detected, favouring systems with elevated SFRs. Such biases are intrinsic to all calibrations based on auroral lines, including those derived from local objects. We address this issue in more detail in Appendix E.

### 7.2.1. The R23, R3, and R2 diagnostics

Among the newly calibrated relationships, R23 is characterised by the lowest scatter, in agreement with previous analyses (Nakajima et al. 2022; Sanders et al. 2024; Chakraborty et al. 2025). However, the R23 calibration exhibits a well-known double-branched structure, with most galaxies populating the plateau region where the relationship flattens, severely limiting its accuracy at intermediate metallicities. A similar behaviour is also observed in R3, where the turnover occurs around  $12 + \log(\text{O}/\text{H}) \sim 8.0$ , leading to two distinct metallicity solutions.

<sup>4</sup> The Huber loss behaves as a standard  $\chi^2$  (L2 norm) for small residuals while reducing the weight of points with large deviations by transitioning to a linear (L1) form.

R2 is generally not employed alone for metallicity measurements due to its sensitivity to the ionisation parameter. However, it can assist in resolving ambiguities in the R23 and R3 calibrations. In our sample, R2 shows an increasing trend with metallicity, although with a much larger scatter in the line ratios at fixed O/H than observed for R3 and R23.

### 7.2.2. O32, Ne3O2, O3N2, and O3S2 diagnostics

Alternative diagnostics such as O32, Ne3O2, O3N2, and O3S2 also display large scatter, highlighting the challenges in metallicity estimation when these indicators are used individually. While O32 primarily traces the ionisation parameter, which is correlated with metallicity (Ji & Yan 2022), other ratios are also sensitive to possible variations in the neon, nitrogen, and sulphur abundance patterns. The scatter is particularly pronounced in the low-metallicity regime. Nonetheless, O32 and Ne3O2 show a monotonic trend with metallicity (with Ne3O2 being also almost insensitive to reddening) and, hence, remain a valuable tool for differentiating between metallicity solutions, especially when used alongside diagnostics such as R23 (e.g. Kewley & Dopita 2002). O3S2 was excluded from the fitting procedure due to the limited number of available data points and the high level of scatter observed.

Interestingly, we identify a subset of  $z \sim 2$  low-metallicity galaxies in the MARTA sample that showcase intense emission in the low-ionisation species, which makes them deviate significantly from the average relation, displaying enhanced  $[\text{O II}]/\text{H}\beta$  (and  $[\text{S II}]/\text{H}\beta$ ) ratios given their O/H. These galaxies also exhibit elevated  $[\text{N II}]/\text{H}\alpha$  ratios (as also mentioned in Sect. 7.2.3), and correspondingly low  $[\text{O III}]/[\text{N II}]$  and  $[\text{O III}]/[\text{S II}]$  values, suggesting a systematic offset in multiple diagnostic line ratios. While dust attenuation uncertainties – such as deviations from the assumed attenuation law – could affect line ratios involving widely separated wavelengths (e.g.  $[\text{O II}]/\text{H}\beta$ ), they are unlikely to fully explain the broad pattern of anomalies observed across both nearby and closely spaced lines. Conversely, MARTA galaxies at  $z > 3$  are characterised by some of the lowest  $[\text{O II}]/\text{H}\beta$  values measured in the JWST sample. Such a spread, observed also in extremely metal-poor galaxies in the local Universe (Nakajima et al. 2022), highlights the role of intense star formation and varying ionisation conditions in shaping the relationship between strong-line ratios and metallicity.

At low metallicities, galaxies typically exhibit higher ionisation parameters; however, the ionisation conditions are not uniform across different systems (e.g. Grasha et al. 2022), but they are strongly modulated by variations in the SFR (e.g. Ji & Yan 2022), the geometric distribution of the gas, and the presence of young, hot stellar populations. Consequently, low-ionisation tracers such as  $[\text{N II}]$  and  $[\text{S II}]$  display significant scatter, reflecting the intrinsic diversity in ionisation conditions within low-metallicity environments. Given that our sample is relatively small, the presence of multiple such cases suggests that these conditions are more common at high redshifts than previously thought.

### 7.2.3. The N2 diagnostic

Thanks to JWST we can start populating the low-metallicity region of the N2 diagnostic. In line with well-established findings over the past decade from ground-based spectroscopic surveys such as MOSDEF (Shapley et al. 2015), KBSS (Steidel et al. 2014; Strom et al. 2018), KLEVER

**Table 5.** Median offset and dispersion (scatter) in line ratios at fixed O/H for each diagnostic and calibration considered in this work.

Calibration	Median offset (dex)								
	R3	R2	R23	O32	O3N2	$\tilde{R}$	N2	Ne3O2	O3S2
Maiolino+08	-0.13	0.055	-0.084	-0.24	-0.21	-	-0.29	-0.068	-
Bian+18	-0.016	-	0.034	0.12	0.41	-	-0.37	0.21	-
Curti+20	-0.11	0.074	-0.080	-0.12	0.34	-	-0.34	0.041	-0.039
Nakajima+22	-0.11	0.065	-0.077	-0.22	0.31	-	-0.35	0.11	-
Laseter+24	-	-	-	-	-	-0.039	-	-	-
Sanders+24	-0.015	-0.010	0.011	0.034	-	-	-	0.12	-
Chakraborty+25	0.044	-0.041	0.018	-0.21	-0.16	-	-	0.092	-
Scholte+25	-	-	-	-	-	0.067	-	-	-
This work	0.0071	-0.011	0.012	0.0089	0.010	-	0.038	-0.049	-
	Scatter (dex)								
Maiolino+08	0.13	0.85	0.12	0.36	0.38	-	0.40	0.35	-
Bian+18	0.10	-	0.31	0.42	0.76	-	0.65	0.47	-
Curti+20	0.29	0.49	0.18	0.31	0.63	-	0.64	0.28	0.52
Nakajima+22	0.16	0.68	0.11	0.33	0.57	-	0.26	0.38	-
Laseter+24	-	-	-	-	-	0.25	-	-	-
Sanders+24	0.070	0.25	0.11	0.30	-	-	-	0.32	-
Chakraborty+25	0.12	0.25	0.10	0.20	0.75	-	-	0.32	-
Scholte+25	-	-	-	-	-	0.15	-	-	-
This work	0.10	0.26	0.077	0.32	0.28	-	0.22	0.34	-

**Notes.** Offsets in line ratios are measured as the difference between the observed values and the values predicted by the corresponding strong-line calibration for any given galaxy metallicity in the sample. Both median offsets and dispersions (scatter) are then calculated over the entire metallicity range considered.

(Hayden-Pawson et al. 2022), and FMOS (Fiber Multi-Object Spectrograph) (Kashino et al. 2017), high- $z$  galaxies exhibit an offset in the [N II]-BPT diagram (Fig. 3), with enhanced [N II]/ $H\alpha$  compared to their local counterparts. One of the most widely accepted interpretations of this behaviour is the hardening of the ionising radiation field at high redshifts, which at fixed metallicity shifts the [N II]/ $H\alpha$  ratio to higher values without necessarily implying variations in chemical abundance patterns (Topping et al. 2020).

We observe the same deviation in the [N II]/ $H\alpha$  versus O/H relation compared to local calibrations, at least at low metallicities ( $12 + \log(\text{O}/\text{H}) < 8$ ). The possible role of enhanced N/O ratios in driving this offset has been recently investigated using the EXCELS dataset (Scholte et al. 2025), where a correlation between N/O and the degree of offset from the median calibration was reported, both at high and low redshifts. An analysis of the N/O abundance in our combined sample will be the subject of a future study.

At  $z > 6$  the optical [N II] is rarely detected because it falls outside the wavelength range probed by JWST/NIRSpec, but also because galaxies at these redshifts tend to have low metallicities and high ionisation parameters, possibly causing most of the nitrogen to reside in higher ionisation states. Thus, the absence of [N II] in high- $z$  sources may also result from nitrogen being predominantly in higher ionisation states (probed by N III], N IV]), as suggested by recent JWST findings (Bunker et al. 2023; Cameron et al. 2023b; Charbonnel et al. 2023; Isobe et al. 2023; Ji et al. 2024; Marques-Chaves et al. 2024; Schaerer et al. 2024; Topping et al. 2024; Curti et al. 2025b).

#### 7.2.4. The $\tilde{R}$ diagnostic

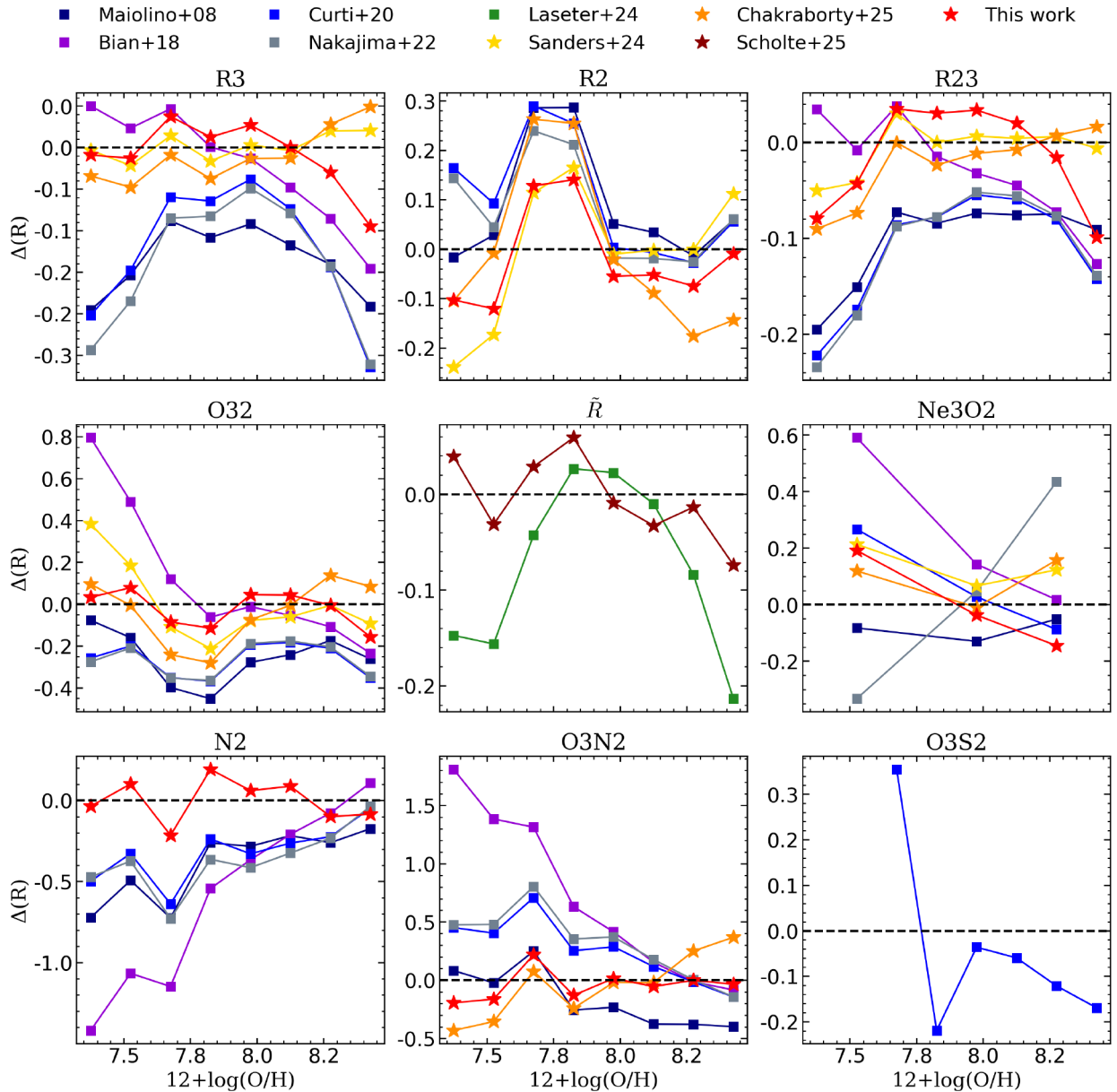
Finally, we tested the  $\tilde{R}$  diagnostic proposed by Laseter et al. (2024) and recently re-calibrated by Scholte et al. (2025) based on a combined sample of data from the JWST/EXCELS sur-

vey (Carnall et al. 2024) and new local galaxy measurements from the DESI Early Data Release (DESI Collaboration 2022). Although at high metallicities ( $Z \approx Z_{\odot}$ ) the high- $z$  dataset lacks sufficient coverage to constrain the turnover of the functional form, performing a fit with the same functional form as in Laseter et al. (2024) to the combined JWST sample provides very similar behaviour to the original calibration in the low-metallicity branch (which is the one primarily populated by high- $z$  galaxies).

While very similar in shape, the Scholte et al. (2025) calibration appears slightly offset towards higher  $\tilde{R}$  ratios at low metallicities compared to both the original Laseter et al. (2024) parametrisation and to our current fit. This discrepancy may be caused by intrinsic differences in the calibration samples, but differences in the set of adopted atomic parameters for  $\text{O}^{++}$  could also play a role. Variations of up to  $\sim 500$  K in the  $T_3$  derivation (see the discussion in Sect. 6) translate into a difference in the final inferred O/H of  $\sim 0.02$ – $0.04$  dex at fixed line ratios and at low metallicities ( $12 + \log(\text{O}/\text{H}) \sim 7.5$ ), where the total abundance budget is dominated by doubly ionised oxygen. Ultimately, we cannot provide the coefficients for the new fit to the  $\tilde{R}$  diagnostic, as our data do not offer significant additional constraints compared to Laseter et al. (2024).

#### 7.3. Comparison with literature calibrations

We analysed the behaviour of each calibration curve considered in this work by computing, for the combined JWST high- $z$  galaxy sample, the offset between observed and predicted line ratios at fixed  $12 + \log(\text{O}/\text{H})$ . We report the median offset and the associated dispersion for each diagnostic and calibration in Table 5. Beside the new calibrations provided in this work (for which the scatter in the calibration sample is minimised by definition), the calibrations presented in Sanders et al. (2024) and Chakraborty et al. (2025) behave well. The two were based on



**Fig. 11.** Median offset of the strong-line calibrations analysed in this work relative to the high- $z$  sample ( $z \sim 1-9.5$ ), computed in bins of metallicity. The bins have a width of 0.15 dex and are included only if they contain at least five objects. Each panel corresponds to a different diagnostic. The colour palette is the same as used in Fig. 10 to distinguish between low- $z$  and high- $z$  based calibrations, with markers also representing this division: stars for high- $z$  and squares for  $z \sim 0$  based calibrations.

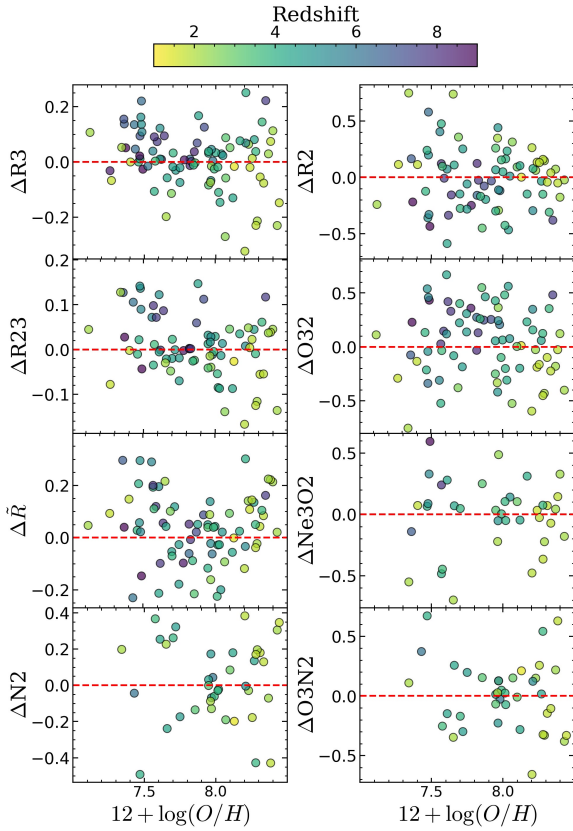
similar (though smaller) samples of high- $z$  galaxies. In contrast, most of the calibrations based on local galaxy samples show larger offsets, as expected given the evolving physical conditions of the calibration samples.

To better visualise possible trends with metallicity, in Fig. 11 we plot the median offset in the line ratio computed in 0.15 dex bins of metallicity, as a function of  $12 + \log(\text{O}/\text{H})$  (where only bins containing at least five objects are considered). Figure 11 highlights the discrepancy between the local ( $z \sim 0$ ) calibrations and the high-redshift galaxy sample. In diagnostics like R3, R23, and O32, the local calibration curves (e.g. Maiolino et al. 2008; Curti et al. 2020b; Nakajima et al. 2022) tend to lie below the observed data points for high-redshift galaxies, suggesting a redshift-dependent evolution of these ratios. Such evolution must be factored into metallicity measurements for high-

redshift galaxies. Sanders et al. (2025) also recently reported a clear redshift evolution in oxygen-based strong-line calibrations, with high-redshift galaxies showing offsets consistent with hotter stellar populations and harder ionising spectra relative to local samples. This reinforces the need to account for cosmic evolution when applying local calibrations to early galaxies.

However, we observe that for several calibrators the median offset tends to decrease with increasing metallicity, possibly suggesting a metallicity-dependent redshift evolution of the involved strong-line diagnostics. This is in line with the observed increase in the scatter of line ratios at fixed O/H in the low-metallicity regime.

As the typical ionisation conditions of the ISM in galaxies evolve with cosmic time, we would expect this behaviour to be



**Fig. 12.** Offsets in the log of the chosen line ratio at fixed O/H, representing the difference between the observed ratio and the value predicted by our best-fit calibrations. High- $z$  JWST sample points are colour-coded by redshift to highlight potential trends.

reflected into a trend between the dispersion of each calibration and redshift. Although the combined JWST sample analysed in this work is still too small to perform a robust analysis of the ‘internal’ redshift evolution (between  $z \sim 1$  and  $z \sim 10$ ), hints of such a correlation are already visible for some of the diagnostics in Fig. 10, and are better visualised in Fig. 12, where we display the offsets in line ratios for individual high- $z$  sources from the best-fit relationships recalibrated in this work, colour-coded by galaxy redshift. At low metallicities ( $12 + \log(\text{O}/\text{H}) \lesssim 8$ ), galaxies at the lower-redshift boundary of the JWST sample tend to be offset in the opposite direction to the highest-redshift systems from the best-fit calibration, especially in line-ratio diagnostics where the dependence on the ionisation parameter is more prominent such as R2, O32, and Ne3O2.

## 8. Summary and outlook

In this work we have analysed deep JWST/NIRSpec R1000 spectra of 16 star-forming galaxies at  $1.8 < z < 4.7$ , observed in the framework of the MARTA programme. We report the detection of faint auroral lines, including [O III]  $\lambda 4363$ , [O II]  $\lambda \lambda 7320, 7330$ , [S II]  $\lambda \lambda 4068, 4076$ , and [S III]  $\lambda 6312$  (Figs. 4 and 6), derived  $T_e$ -based metallicities, and discuss trends in the temperature-temperature relationships and the evolution of strong-line metallicity calibrations. The key results from this work are summarised as follows.

- The temperature-temperature ( $T_2$ – $T_3$ ) relation between different oxygen ionised species, established on low-redshift samples, seems to hold at earlier cosmic epochs, offering, despite its non-negligible scatter, a reliable framework for metallicity determinations at high  $z$  (Fig. 7).
- The dispersion of the  $T_2$ – $T_3$  relationship mainly correlates with variations in the ionisation parameter (traced by [O III]/[O II]) and/or in the hardening of the ionising continuum (traced by the ‘softness parameter’,  $\eta$ ; Fig. 8 and Table 3). Variations in the electron density mainly correlate with  $T_2$  (at fixed  $T_3$ ) but have a lower impact on the scatter of the relationship.
- We explored the relationship between temperatures derived for low-ionisation and high-ionisation sulphur and oxygen species, demonstrating their consistency with local trends (Fig. 9).
- The MARTA auroral-line galaxies, while still subsolar on average, populate the upper end of the metallicity distribution along with high-redshift galaxies with direct metallicity measurements. As such, they provide valuable insights into the applicability of strong-line metallicity calibrations at  $z \sim 2$ , extending beyond the low-metallicity regime typically probed by earlier studies. By combining our sample with existing  $T_e$ -based metallicity measurements from JWST in the literature, we assessed the redshift evolution in several popular strong-line diagnostics over a wide range of oxygen abundances ( $12 + \log(\text{O}/\text{H}) \in [7.0\text{--}8.4]$ ; Figs. 10 and 11) and provide updated calibrations for some of these diagnostics in the form of standard polynomial relations (Table 4).
- While ‘strong-line methods’ remain broadly applicable across cosmic time, clear systematic shifts emerge in specific line ratios, particularly R2, O32, and Ne3O2, as likely driven by evolving ionisation conditions at high redshifts. This trend is reflected in a correlation between the offset in the line ratio at fixed metallicities and redshifts in these diagnostics, as tentatively observed in our combined JWST sample (Fig. 12).
- We observe a systematic offset in the N2 versus O/H diagram (especially at low metallicities), in line with previous findings that high-redshift galaxies exhibit higher [N II]/ $H\alpha$  at fixed O/H compared to local counterparts (e.g. Steidel et al. 2014; Masters et al. 2014; Shapley et al. 2015; Strom et al. 2017). Furthermore, a subset of low-metallicity galaxies in the MARTA sample deviates significantly from the average relationships due to their elevated [O II]/ $H\beta$  and [S II]/ $H\beta$  ratios given their O/H, highlighting the variety of ionisation conditions observed in the high-redshift Universe and the potential impact of selection effects on the determination of the average trends in high- $z$  galaxy population properties.

Future research will build on the insights gained from this and other recent JWST-enabled studies to achieve a more self-consistent determination of galaxy metallicity properties and their scaling relations across cosmic epochs. In parallel, studies of chemical abundance patterns in high-redshift galaxies, focusing on elements such as nitrogen, neon, and argon, will provide complementary insights into star formation histories at  $z \sim 2$  and beyond, helping refine next-generation chemical evolution models. A key avenue for future exploration is the derivation of direct,  $T_e$ -based nitrogen-to-oxygen (N/O) abundance ratios. Variations in N/O ratios at high redshifts can reveal crucial details about rapid nitrogen enrichment, differences in star formation efficiency, or localised enrichment processes such

as those driven by Wolf-Rayet stars (Fernández-Martín et al. 2012; Rivera-Thorsen et al. 2024; Kobayashi & Ferrara 2024), highlighting the importance of exploring this diagnostic in detail.

Furthermore, we plan to conduct detailed comparisons with state-of-the-art photoionisation models, such as HOMERUN (Marconi et al. 2024; Moreschini et al., in prep.). Unlike single-cloud models, HOMERUN incorporates contributions from multiple gas clouds with varying metallicities, ionisation parameters, and densities. This approach enables a more nuanced analysis of emission line ratios, allowing us to assess the impact of high-density gas clumps on metallicity determinations and refine our understanding of biases in classical methods.

**Acknowledgements.** FB and FM acknowledge support from the INAF Fundamental Astrophysics programme 2022 and 2023. FM and FB acknowledge support from grant PRIN-MUR 202223XDPZM “Prometheus” GC and EB acknowledge financial support from INAF under the Large Grant 2022 “The metal circle: a new sharp view of the baryon cycle up to Cosmic Dawn with the latest generation IFU facilities”. A.F. acknowledges the support from project “VLT-MOONS” CRAM 1.05.03.07. AM acknowledges support from project PRIN-MUR project “PROMETEUS” financed by the European Union – Next Generation EU, Mission 4 Component 1 CUP B53D23004750006. FC acknowledges support from a UKRI Frontier Research Guarantee Grant (PI Cullen; grant reference EP/X021025/1). CK acknowledges funding from the UK Science and Technology Facility Council through grant ST/Y001443/1. This work is based in on observations made with the NASA/ESA/CSA James Webb Space Telescope. The data were obtained from the Mikulski Archive for Space Telescopes at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-03127 for JWST. These observations are associated with program JWST 1879. The data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). The specific observations analyzed can be accessed via <https://doi.org/10.17909/v7wt-1r79>. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

## References

- Aggarwal, K. M., & Keenan, F. P. 1999, *ApJS*, **123**, 311
- Andrews, B. H., & Martini, P. 2013, *ApJ*, **765**, 140
- Arellano-Córdova, K. Z., & Rodríguez, M. 2020, *MNRAS*, **497**, 672
- Arellano-Córdova, K. Z., Cullen, F., Carnall, A. C., et al. 2025, *MNRAS*, **540**, 2991
- Baker, W. M., Maiolino, R., Belfiore, F., et al. 2023, *MNRAS*, **519**, 1149
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, **93**, 5
- Balogh, M. L., McGee, S. L., Mok, A., et al. 2014, *MNRAS*, **443**, 2679
- Belfiore, F., Westfall, K. B., Schaefer, A., et al. 2019, *AJ*, **158**, 160
- Berg, D. A., Skillman, E. D., Croxall, K. V., et al. 2015, *ApJ*, **806**, 16
- Berg, D. A., Pogge, R. W., Skillman, E. D., et al. 2020, *ApJ*, **893**, 96
- Berg, D. A., Chisholm, J., Erb, D. K., et al. 2021, *ApJ*, **922**, 170
- Bezanson, R., Labbe, I., Whitaker, K. E., et al. 2024, *ApJ*, **974**, 92
- Bian, F., Fan, X., Bechtold, J., et al. 2010, *ApJ*, **725**, 1877
- Bian, F., Kewley, L. J., & Dopita, M. A. 2018, *ApJ*, **859**, 175
- Bluck, A. F. L., Maiolino, R., Sánchez, S. F., et al. 2019, *MNRAS*, **485**, 666
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Bresolin, F., Kennicutt, R. C., & Garnett, D. R. 1998, *ApJ*, **510**, 104
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Bunker, A. J., Saxena, A., Cameron, A. J., et al. 2023, *A&A*, **677**, A88
- Calzetti, D., Kinney, A. L., & Storch-Bergmann, T. 1994, *ApJ*, **429**, 582
- Cameron, A. J., Saxena, A., Bunker, A. J., et al. 2023a, *A&A*, **677**, A115
- Cameron, A. J., Katz, H., Rey, M. P., & Saxena, A. 2023b, *MNRAS*, **523**, 3516
- Campbell, A., Terlevich, R., & Melnick, J. 1986, *MNRAS*, **223**, 811
- Cappellari, M., & Emsellem, E. 2004, *PASP*, **116**, 138
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, **345**, 245
- Carnall, A. C., Cullen, F., McLure, R. J., et al. 2024, *MNRAS*, **534**, 325
- Chabrier, G. 2003, *PASP*, **115**, 763
- Chakraborty, P., Sarkar, A., Smith, R., et al. 2025, *ApJ*, **985**, 24
- Charbonnel, C., Schaerer, D., Prantzos, N., et al. 2023, *A&A*, **673**, L7
- Christensen, L., Laursen, P., Richard, J., et al. 2012, *MNRAS*, **427**, 1973
- Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, *A&A*, **575**, A40
- Conroy, C., Villaume, A., van Dokkum, P. G., & Lind, K. 2018, *ApJ*, **854**, 139
- Cresci, G., Mannucci, F., & Curti, M. 2019, *A&A*, **627**, A42
- Croxall, K. V., Pogge, R. W., Berg, D. A., Skillman, E. D., & Moustakas, J. 2015, *ApJ*, **808**, 42
- Croxall, K. V., Pogge, R. W., Berg, D. A., Skillman, E. D., & Moustakas, J. 2016, *ApJ*, **830**, 4
- Cullen, F., Shapley, A. E., McLure, R. J., et al. 2021, *MNRAS*, **505**, 903
- Cullen, F., Carnall, A. C., Scholte, D., et al. 2025, *MNRAS*, **540**, 2176
- Curti, M., Cresci, G., Mannucci, F., et al. 2017, *MNRAS*, **465**, 1384
- Curti, M., Maiolino, R., Cirasuolo, M., et al. 2020a, *MNRAS*, **492**, 821
- Curti, M., Mannucci, F., Cresci, G., & Maiolino, R. 2020b, *MNRAS*, **491**, 944
- Curti, M., Hayden-Pawson, C., Maiolino, R., et al. 2022, *MNRAS*, **512**, 4136
- Curti, M., D’Eugenio, F., Carniani, S., et al. 2023, *MNRAS*, **518**, 425
- Curti, M., Cataldi, E., Belfiore, F., et al. 2025a, *ArXiv e-prints* [arXiv:2509.06622]
- Curti, M., Witstok, J., Jakobsen, P., et al. 2025b, *A&A*, **697**, A89
- Damjanov, I., Zahid, H. J., Geller, M. J., Fabricant, D. G., & Hwang, H. S. 2018, *ApJS*, **234**, 21
- DESI Collaboration (Abareshi, B., et al.) 2022, *AJ*, **164**, 207
- D’Eugenio, F., Maiolino, R., Carniani, S., et al. 2024, *A&A*, **689**, A152
- D’Eugenio, F., Cameron, A. J., Scholtz, J., et al. 2025, *ApJS*, **277**, 4
- Domínguez-Guzmán, G., Rodríguez, M., García-Rojas, J., Esteban, C., & Toribio San Cipriano, L. 2022, *MNRAS*, **517**, 4497
- Dunlop, J. S., Abraham, R. G., Ashby, M. L. N., et al. 2021, *PRIMER: Public Release IMAGING for Extragalactic Research*, JWST Proposal. Cycle 1, 1837
- Erb, D. K., Steidel, C. C., Shapley, A. E., et al. 2006, *ApJ*, **647**, 128
- Esteban, C., Peimbert, M., García-Rojas, J., et al. 2004, *MNRAS*, **355**, 229
- Esteban, C., Bresolin, F., Peimbert, M., et al. 2009, *ApJ*, **700**, 654
- Esteban, C., García-Rojas, J., Carigi, L., et al. 2014, *MNRAS*, **443**, 624
- Esteban, C., Bresolin, F., García-Rojas, J., & Toribio San Cipriano, L. 2020, *MNRAS*, **491**, 2137
- Fernández-Martín, A., Martín-Gordón, D., Vílchez, J. M., et al. 2012, *A&A*, **541**, A119
- Finkelstein, S. L., Bagley, M. B., Ferguson, H. C., et al. 2023, *ApJ*, **946**, L13
- Garnett, D. R. 1992, *AJ*, **103**, 1330
- Garnett, D. R. 2002, *ApJ*, **581**, 1019
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, *ApJ*, **594**, 279
- Grasha, K., Chen, Q. H., Battisti, A. J., et al. 2022, *ApJ*, **929**, 118
- Guseva, N. G., Papaderos, P., Meyer, H. T., Izotov, Y. I., & Fricke, K. J. 2009, *A&A*, **505**, 63
- Guseva, N. G., Izotov, Y. I., Stasińska, G., et al. 2011, *A&A*, **529**, A149
- Hainline, K. N., Shapley, A. E., Kornei, K. A., et al. 2009, *ApJ*, **701**, 52
- Hamel-Bravo, M. J., Fisher, D. B., Berg, D., et al. 2024, *MNRAS*, **530**, 3855
- Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, **858**, 77
- Hayden-Pawson, C., Curti, M., Maiolino, R., et al. 2022, *MNRAS*, **512**, 2867
- Heintz, K. E., Brammer, G. B., Watson, D., et al. 2025, *A&A*, **693**, A60
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, **457**, 841
- Isobe, Y., Ouchi, M., Tominaga, N., et al. 2023, *ApJ*, **959**, 100
- Izotov, Y. I., Guseva, N. G., Fricke, K. J., & Papaderos, P. 2009, *A&A*, **503**, 61
- Jakobsen, P., Ferruit, P., Alves de Oliveira, C., et al. 2022, *A&A*, **661**, A80
- James, B. L., Pettini, M., Christensen, L., et al. 2014, *MNRAS*, **440**, 1794
- Ji, X., & Yan, R. 2022, *A&A*, **659**, A112
- Ji, X., Übler, H., Maiolino, R., et al. 2024, *MNRAS*, **535**, 881
- Kaasinen, M., Kewley, L., Bian, F., et al. 2018, *MNRAS*, **477**, 5568
- Kartaltepe, J. S., Sanders, D. B., Le Floch, E., et al. 2010, *ApJ*, **709**, 572
- Kartaltepe, J. S., Sanders, D. B., Silverman, J. D., et al. 2015, *ApJ*, **806**, L35
- Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, *ApJ*, **835**, 88
- Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, *MNRAS*, **341**, 54
- Kelly, B. C. 2007, *ApJ*, **665**, 1489
- Kennicutt, R. C. 1998, *ARA&A*, **36**, 189
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, **50**, 531
- Kennicutt, R. C., Bresolin, F., & Garnett, D. R. 2003, *ApJ*, **591**, 801
- Kewley, L. J., & Dopita, M. A. 2002, *ApJS*, **142**, 35
- Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, *ApJ*, **556**, 121
- Kewley, L. J., Maier, C., Yabe, K., et al. 2013, *ApJ*, **774**, L10
- Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, *ARA&A*, **57**, 511
- Khoram, A. H., & Belfiore, F. 2025, *A&A*, **693**, A150
- Kisielius, R., Storey, P. J., Ferland, G. J., & Keenan, F. P. 2009, *MNRAS*, **397**, 903
- Kobayashi, C., & Ferrara, A. 2024, *ApJ*, **962**, L6
- Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, **218**, 15
- Laseter, I. H., Maseda, M. V., Curti, M., et al. 2024, *A&A*, **681**, A70
- Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. 2004, *A&A*, **425**, 881
- Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, *ApJS*, **184**, 218
- Luridiana, V., Morisset, C., & Shaw, R. A. 2015, *A&A*, **573**, A42
- 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., Marconi, A., & Gnerucci, A. 2010,

- MNRAS, 408, 2115
- Marconi, A., Amiri, A., Feltre, A., et al. 2024, *A&A*, 689, A78
- Marques-Chaves, R., Schaerer, D., Kuruvanthodi, A., et al. 2024, *A&A*, 681, A30
- Maseda, M. V., de Graaff, A., Franx, M., et al. 2024, *A&A*, 689, A73
- Masters, D., Capak, P., Salvato, M., et al. 2012, *ApJ*, 755, 169
- Masters, D., McCarthy, P., Siana, B., et al. 2014, *ApJ*, 785, 153
- Masters, D., Faisst, A., & Capak, P. 2016, *ApJ*, 828, 18
- Mazzolari, G., Übler, H., Maiolino, R., et al. 2024, *A&A*, 691, A345
- McCllymont, W., Tacchella, S., D'Eugenio, F., et al. 2025, *MNRAS*, 540, 190
- McGaugh, S. S. 1991, *ApJ*, 380, 140
- Méndez-Delgado, J. E., Esteban, C., García-Rojas, J., et al. 2023, *MNRAS*, 523, 2952
- Méndez-Delgado, J. E., Esteban, C., García-Rojas, J., et al. 2023a, *MNRAS*, 523, 2952
- Méndez-Delgado, J. E., Esteban, C., García-Rojas, J., Kreckel, K., & Peimbert, M. 2023b, *Nature*, 618, 249
- Mollá, M., Vílchez, J. M., Gavilán, M., & Díaz, A. I. 2006, *MNRAS*, 372, 1069
- Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, *ApJS*, 225, 27
- Morishita, T., Stiavelli, M., Grillo, C., et al. 2024a, *ApJ*, 971, 43
- Morishita, T., Stiavelli, M., Chary, R.-R., et al. 2024b, *ApJ*, 963, 9
- Morishita, T., Stiavelli, M., Schuldt, S., & Grillo, C. 2025, *ApJ*, 979, 87
- Nagao, T., Maiolino, R., & Marconi, A. 2006, *A&A*, 459, 85
- Nakajima, K., & Ouchi, M. 2014, *MNRAS*, 442, 900
- Nakajima, K., Ouchi, M., Xu, Y., et al. 2022, *ApJS*, 262, 3
- Nanayakkara, T., Glazebrook, K., Kacprzak, G. G., et al. 2016, *ApJ*, 828, 21
- Napolitano, L., Castellano, M., Pentericci, L., et al. 2025, *ApJ*, 989, 75
- Navarro-Carrera, R., Caputi, K. I., Iani, E., et al. 2024, ArXiv e-prints [arXiv:2407.14201]
- Nicholls, D. C., Dopita, M. A., Sutherland, R. S., Kewley, L. J., & Palay, E. 2013, *ApJS*, 207, 21
- Noeske, K. G., Faber, S. M., Weiner, B. J., et al. 2007, *ApJ*, 660, L47
- Osterbrock, D. E., & Ferland, G. J. 2006, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, 2nd edn. (Sausalito: University Science Books)
- Palay, E., Nahar, S. N., Pradhan, A. K., & Eissner, W. 2012, *MNRAS*, 423, L35
- Patrício, V., Christensen, L., Rhodin, H., Cañameras, R., & Lara-López, M. A. 2018, *MNRAS*, 481, 3520
- Peimbert, M. 1967, *ApJ*, 150, 825
- Peimbert, M., Peimbert, A., & Delgado-Inglada, G. 2017, *PASP*, 129, 082001
- Pérez-Montero, E., & Contini, T. 2009, *MNRAS*, 398, 949
- Pettini, M., & Pagel, B. E. J. 2004, *MNRAS*, 348, L59
- Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, *ApJ*, 612, 168
- Pilyugin, L. S. 2007, *MNRAS*, 375, 685
- Pilyugin, L. S., Vílchez, J. M., & Thuan, T. X. 2006, *MNRAS*, 370, 1928
- Pilyugin, L. S., Mattsson, L., Vílchez, J. M., & Cedrés, B. 2009, *MNRAS*, 398, 485
- Pilyugin, L. S., Vílchez, J. M., Mattsson, L., & Thuan, T. X. 2012, *MNRAS*, 421, 1624
- Podobedova, L. I., Kelleher, D. E., & Wiese, W. L. 2009, *JPCRD*, 38, 171
- Popesso, P., Concas, A., Cresci, G., et al. 2023, *MNRAS*, 519, 1526
- Reddy, N. A., Shapley, A. E., Kriek, M., et al. 2020, *ApJ*, 902, 123
- Reddy, N. A., Topping, M. W., Sanders, R. L., Shapley, A. E., & Brammer, G. 2023, *ApJ*, 948, 83
- Rivera-Thorsen, T. E., Chisholm, J., Welch, B., et al. 2024, *A&A*, 690, A269
- Rogers, N. S. J., Skillman, E. D., Pogge, R. W., et al. 2021, *ApJ*, 915, 21
- Rogers, N. S., Skillman, E. D., Pogge, R. W., et al. 2022, *ApJ*, 939, 44
- Rogers, N. S. J., Strom, A. L., Rudie, G. C., et al. 2024, *ApJ*, 964, L12
- Rynkun, P., Gaigalas, G., & Jönsson, P. 2019, *A&A*, 623, A155
- Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, *ApJ*, 816, 23
- Sanders, R. L., Shapley, A. E., Reddy, N. A., et al. 2020, *MNRAS*, 491, 1427
- Sanders, R. L., Shapley, A. E., Jones, T., et al. 2021, *ApJ*, 914, 19
- Sanders, R. L., Shapley, A. E., Clarke, L., et al. 2023, *ApJ*, 943, 75
- Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2023, *ApJ*, 955, 54
- Sanders, R. L., Shapley, A. E., Topping, M. W., Reddy, N. A., & Brammer, G. B. 2024, *ApJ*, 962, 24
- Sanders, R. L., Shapley, A. E., Topping, M. W., et al. 2025, *ApJ*, 989, 209
- Sanders, R. L., Shapley, A. E., Topping, M. W., et al. 2025, ArXiv e-prints [arXiv:2508.10099]
- Schaerer, D., Marques-Chaves, R., Xiao, M., & Korber, D. 2024, *A&A*, 687, L11
- Scholte, D., Cullen, F., Carnall, A. C., et al. 2025, *MNRAS*, 540, 1800
- Schreiber, C., Glazebrook, K., Nanayakkara, T., et al. 2018, *A&A*, 618, A85
- Schreiber, N. M. F., Übler, H., Davies, R. L., et al. 2019, *ApJ*, 875, 21
- Shapley, A. E., Reddy, N. A., Kriek, M., et al. 2015, *ApJ*, 801, 88
- Shapley, A. E., Sanders, R. L., Shao, P., et al. 2019, *ApJ*, 881, L35
- Shapley, A. E., Sanders, R. L., Topping, M. W., et al. 2025, *ApJ*, 981, 167
- Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, *ApJS*, 220, 12
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, 214, 24
- Stanton, T. M., Cullen, F., McLure, R. J., et al. 2024, *MNRAS*, 532, 3102
- Stanton, T. M., Cullen, F., Carnall, A. C., et al. 2025, *MNRAS*, 537, 1735
- Stasińska, G. 2002, *Cosmochemistry. The Melting Pot of the Elements* (Cambridge: Cambridge University Press), 115
- Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, *ApJ*, 795, 165
- Steidel, C. C., Strom, A. L., Pettini, M., et al. 2016, *ApJ*, 826, 159
- Stiavelli, M., Morishita, T., Chiaberge, M., et al. 2025, *ApJ*, 981, 136
- Storey, P. J., Sochi, T., & Badnell, N. R. 2014, *MNRAS*, 441, 3028
- Stratman, C. M. S., van der Wel, A., Bezanson, R., et al. 2018, *ApJS*, 239, 27
- Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, *ApJ*, 836, 164
- Strom, A. L., Steidel, C. C., Rudie, G. C., Trainor, R. F., & Pettini, M. 2018, *ApJ*, 868, 117
- Strom, A., Rudie, G., Trainor, R., et al. 2023, *ApJ*, 958, L11
- Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, *ApJ*, 768, 74
- Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, *A&A*, 600, A110
- Tayal, S. S., & Gupta, G. P. 1999, *ApJ*, 526, 544
- Tayal, S. S., & Zatsarinny, O. 2010, *ApJS*, 188, 32
- Topping, M. W., Shapley, A. E., Reddy, N. A., et al. 2020, *MNRAS*, 495, 4430
- Topping, M. W., Stark, D. P., Senchyna, P., et al. 2024, *MNRAS*, 529, 3301
- Topping, M. W., Sanders, R. L., Shapley, A. E., et al. 2025a, *MNRAS*, 541, 1707
- Topping, M. W., Stark, D. P., Senchyna, P., et al. 2025b, *ApJ*, 980, 225
- Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J., et al. 2010, *MNRAS*, 404, 1639
- Vazdekis, A., Koleva, M., Ricciardelli, E., Röck, B., & Falcón-Barroso, J. 2016, *MNRAS*, 463, 3409
- Vílchez, J. M., & Pagel, B. E. J. 1988, *MNRAS*, 231, 257
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, 258, 11
- Welch, B., Olivier, G. M., Hutchison, T. A., et al. 2024, *ApJ*, 975, 196
- Welch, B., Rivera-Thorsen, T. E., Rigby, J. R., et al. 2025, *ApJ*, 980, 33
- Wisnioski, E., Schreiber, N. M. F., Fossati, M., et al. 2019, *ApJ*, 886, 124
- Yabe, K., Ohta, K., Iwamuro, F., et al. 2012, *PASJ*, 64, 60
- Yang, S., Lidz, A., Benson, A., et al. 2023, *MNRAS*, 534, 3665
- Yates, R. M., Schady, P., Chen, T.-W., Schweyer, T., & Wiseman, P. 2020, *A&A*, 634, A107
- Zeppen, C. J. 1982, *MNRAS*, 198, 111
- Zhang, Y., Morishita, T., & Stiavelli, M. 2025, ArXiv e-prints [arXiv:2502.04817]
- Zurita, A., & Bresolin, F. 2012, *MNRAS*, 427, 1463
- Zurita, A., Florido, E., Bresolin, F., Pérez-Montero, E., & Pérez, I. 2021, *MNRAS*, 500, 2359

- <sup>1</sup> Università di Firenze, Dipartimento di Fisica e Astronomia, Via G. Sansone 1, 50019 Sesto Fiorentino, Florence, Italy
- <sup>2</sup> INAF – Arcetri Astrophysical Observatory, Largo E. Fermi 5, I-50125 Florence, Italy
- <sup>3</sup> European Southern Observatory, Karl-Schwarzschild Straße 2, D-85748 Garching bei München, Germany
- <sup>4</sup> Department of Physics, University of Arkansas, 226 Physics Building, 825 West Dickson Street, Fayetteville, AR 72701, USA
- <sup>5</sup> Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy
- <sup>6</sup> Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
- <sup>7</sup> Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK
- <sup>8</sup> AURA for European Space Agency, Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21210, USA
- <sup>9</sup> Cavendish Laboratory, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, UK
- <sup>10</sup> Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>11</sup> University of Trento, Via Sommarive 14, I-38123 Trento, Italy

## Appendix A: Reliability of error estimates

We estimated the S/N of the weak emission lines by considering both the fit residuals and the pipeline errors. In the first approach, we considered the peak of each emission line and divided it by the standard deviation of the residuals from a line-free region adjacent to the emission line. To approximate the total S/N over the line profile, this value was then multiplied by the number of spectral channels corresponding to its full width at half maximum (FWHM). At a spectral resolution of  $R \approx 1000$ , our emission lines are unresolved, meaning their widths are dominated by the instrumental line spread function. We therefore adopted a fiducial FWHM of 3 pixels.

We compared these values with a second S/N estimate based on the ratio of the line flux obtained from the Gaussian fit to the error on this flux. Here, the flux error was multiplied by a factor derived for each galaxy of the sample from the analysis of a line-free region in its spectrum – ranging from 4125 to 4325 Å. This factor represents the average ratio between the error spectrum produced by the JWST data reduction pipeline and the standard deviation of the residuals between the data and the pPXF best-fit within the same region. This factor is on average  $\gtrsim 1.4$ , which indicates that pipeline error estimates tend to be conservative. We therefore rescaled the error spectra downwards.

From this comparison of S/Ns, we found that the two were generally in good agreement. In a few cases, however, the second estimate was consistently higher than the first.

## Appendix B: Additional details on the spectral fitting

### Possible [Fe II] contamination

Galaxies MARTA\_4195 and MARTA\_5014 both exhibit an evident emission feature at  $\sim 4360$  Å – visible in the 2D and 1D extracted spectra – which is blueshifted with respect to the [O III]  $\lambda 4363$  line, aligning more closely with the expected position of the [Fe II]  $\lambda 4360$  line. This emission feature has been documented in various studies using both stacked (Curti et al. 2017) and individual spectra of high-metallicity galaxies in the local Universe (e.g. Rogers et al. 2021, 2022). Furthermore, [Fe II] contamination has recently been tentatively identified in high-redshift galaxies (see e.g. Shapley et al. (2025) for a galaxy at  $z \sim 7$  from the AURORA survey).

We therefore included [Fe II]  $\lambda 4360$  as an additional component in our fitting setup for these two galaxies, significantly improving the fit of [O III]  $\lambda 4363$  (Fig. B.1). Notably, in the case of ID MARTA\_4195, we also (tentatively) detected the [Fe II]  $\lambda 4288$  emission line, which is expected to accompany [Fe II]  $\lambda 4360$  emission. This approach is adopted as our fiducial fitting method in this work. The inclusion or exclusion of [Fe II] in the fit for these two galaxies results in a 20–40% decrease in the flux of the auroral [O III]  $\lambda 4363$  line. This, in turn, affects the derived  $T_3$  values, leading to a variation of 10–15%, with respect to an estimate that associates the entire flux with [O III]  $\lambda 4363$ .

For the other galaxies in the sample, we performed the same analysis by fitting the [O III]  $\lambda 4363$  line both with and without the inclusion of [Fe II]. However, in these cases, the [Fe II] flux was negligible, of the order of  $\sim 1/10$ – $1/100$  of the [O III]  $\lambda 4363$  flux, and its inclusion did not significantly affect the measured [O III]  $\lambda 4363$  flux. Moreover, no additional evidence of [Fe II] emission is observed in the spectra of these galaxies. Therefore, to avoid introducing unnecessary complexity to the modelling,

we excluded [Fe II] as a component for all other cases. The measured fluxes of the auroral lines, along with their corresponding S/Ns, are provided in Table D.1.

### The case of MARTA\_3942

Object 3942 presented unique challenges due to its extended nature and the presence of a contaminant near the end of the slit. To address these issues, we applied a local background subtraction technique, avoiding self-subtraction and recovering part of the flux from the second shutter while excluding the contaminated region at the slit's edge. However, the use of local background subtraction inevitably introduced some noise into the continuum, as the outer shutters were integrated for less time than the central one. This resulted in a particularly noisy continuum around the auroral line [O III]  $\lambda 4363$ .

A significant issue arises when fitting this region with pPXF: the algorithm attempts to model stellar features around H $\gamma$  (4340 Å) but is actually fitting the noise, partially suppressing the [O III]  $\lambda 4363$  line. Although this line is clearly distinguishable above the continuum, the pPXF fit underestimates its flux.

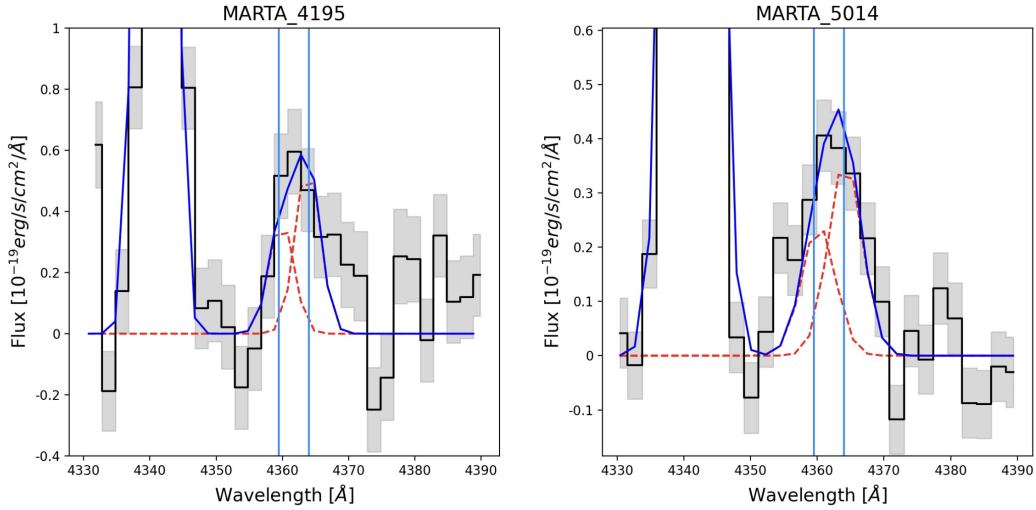
To mitigate this issue and recover a more accurate flux measurement, we subtracted the continuum modelled on the full three-nodding reduction version of the spectrum from the local background-subtracted spectrum. In this way, the model continuum was significantly smoother due to the more uniform exposure across the slit, but it was re-scaled to match the fiducial, local background-subtracted spectrum, before fitting the [O III]  $\lambda 4363$  emission line with an individual Gaussian component. The revised fit yields a line flux of  $3.51 \times 10^{-19}$  erg/s/cm<sup>2</sup>, recovering additional flux compared to the previous fit ( $3.11 \times 10^{-19}$  erg/s/cm<sup>2</sup>).

### Different continuum modelling

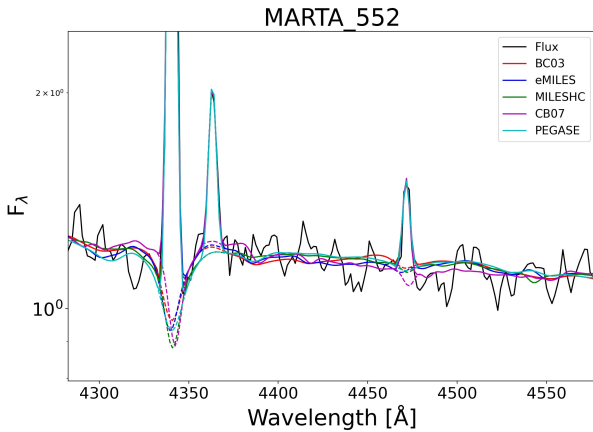
To assess the impact of the adopted continuum modelling on our derived quantities—particularly electron temperatures – we performed a set of tests using a variety of stellar population synthesis templates and polynomial corrections. Specifically, we repeated the spectral fits using several stellar libraries, including eMILES (Vazdekis et al. 2016), BC03 (Bruzual & Charlot 2003), CB07 (Charlot & Bruzual, unpublished), MILES-HC (a high-resolution subset of MILES; Conroy et al. 2018), and PEGASE (Le Borgne et al. 2004).

To directly quantify the influence of continuum modelling choices on key line measurements, we compared the fluxes of important emission lines – including [O III]  $\lambda 4363$  – across different template fits. As shown in Fig. B.2, which illustrates fits for a representative galaxy, the depth of Balmer absorption features can vary significantly depending on the chosen template (Belfiore et al. 2019). Nonetheless, the effect on the flux of [O III]  $\lambda 4363$  remains small, typically under  $\sim 10\%$ . The corresponding variation in the derived electron temperatures is generally within 10–15%, although we found that in galaxies with weaker auroral lines and lower S/N – such as MARTA\_2387 – the inferred  $T_3$  can vary by up to  $\sim 25\%$ . Conversely, in cases with strong lines and high S/N, variations are as low as 2%.

For the Balmer lines used in the derivation of dust attenuation, we minimised the impact of template-dependent absorption differences by focusing on the strongest lines—H $\alpha$ , H $\beta$ , H $\gamma$ , and H $\delta$ —whose high S/N ensures that absorption variations alter the measured fluxes by no more than  $\sim 5\%$ . Overall, these tests confirm that our continuum modelling approach is robust. Neither the choice of stellar template nor the polyno-



**Fig. B.1.** Spectral fits in the region around the [O III]  $\lambda$ 4363 line that contains three emission lines (H $\gamma$ , [Fe II]  $\lambda$ 4360, and [O III]  $\lambda$ 4363) for galaxies 4195 (left panel) and 5014 (right panel). The overall fit is represented by the solid blue line, while the two Gaussian components centred at 4360 and 4364 Å are shown as dashed red lines. The fit demonstrates that a single Gaussian at 4364 Å alone cannot adequately reproduce the spectral features of these two galaxies. The inclusion of the [Fe II] component provides a significantly better fit.



**Fig. B.2.** Example continuum fits for a representative object – MARTA\_552 – using different stellar population synthesis templates. The figure illustrates the variation in the modelled absorption feature under the H $\gamma$ , and the corresponding stability of the [O III]  $\lambda$ 4363 emission line flux.

mial order used for continuum correction introduces any systematic uncertainty that would affect the key conclusions of this work.

### Appendix C: Different parametrisation of the T<sub>2</sub>-T<sub>3</sub> diagram

We further analysed the T<sub>2</sub>-T<sub>3</sub> relationship by comparing the temperatures derived for the MARTA sample with the multi-parametric calibration proposed by Yates et al. (2020). This approach is motivated by two primary considerations: (i) there is an anti-correlation between the electron temperature and the oxygen abundance. This means that if  $T_e([\text{O III}]) \ll T_e([\text{O II}])$ , then it would be  $[\text{O III}] \gg [\text{O II}]$ , i.e. systems where the temperature of the doubly ionised oxygen is lower than the singly ionised oxygen one, tend to have a higher proportion of doubly ionised oxygen relative to singly ionised oxygen. In these cases, the total oxygen abundance  $Z(T_e)$  becomes less sen-

sitive to changes in  $T_e([\text{O III}])$ . Conversely, systems where  $T_e([\text{O III}]) > T_e([\text{O II}])$  exhibit the opposite pattern: the singly ionised oxygen fraction dominates, and the total oxygen abundance becomes less sensitive to  $T_e([\text{O III}])$ . (ii) There is an empirical anti-correlation between  $T_e([\text{O III}])$  and  $T_e([\text{O II}])$  at fixed oxygen abundance. This is expected theoretically because an increase in the singly ionised oxygen fraction is balanced by a decrease in the doubly ionised oxygen fraction, and vice versa.

To model this, Yates et al. (2020) adopted a rectangular hyperbolic form centred at 0 K:

$$T_e([\text{O II}]) = a(Z_{T_e})^2 \frac{1}{T_e([\text{O III}])}, \quad (\text{C.1})$$

where  $a(Z_{T_e})$  is a function depending on metallicity. Using this parametrisation, we examined how well the MARTA sample temperatures conform to the best-fit calibration of Yates et al. The results are presented in Fig. C.1: the main panel shows the T<sub>2</sub>-T<sub>3</sub> relation colour-coded by the metallicity, revealing that for some galaxies, the interconnected dependence of T<sub>2</sub>, T<sub>3</sub>, and metallicity is well reproduced by the Yates calibration. This alignment is particularly evident in the continuous gradients of metallicity along the T<sub>2</sub>-T<sub>3</sub> plane.

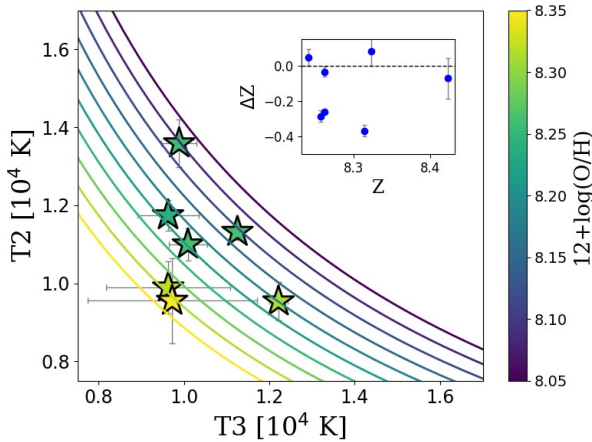
The inset panel quantifies the agreement by plotting  $\Delta Z$  versus  $Z$ , where  $\Delta Z$  represents the difference between the observed metallicity (calculated using the  $T_e$  method) and the metallicity predicted by the Yates calibration, given the observed temperatures. Although some galaxies show good agreement (with  $\Delta Z$  close to zero), for others significant deviations are apparent.

These discrepancies might arise because of the inapplicability of such a parametrisation to high-redshift galaxies, where the physical conditions of the ISM may deviate substantially from those in the local universe. For instance, at higher redshifts, ionisation parameters and density conditions can lead to deviations from the trends captured by the Yates calibration, necessitating further exploration of these dependences in different galactic environments.

**Table D.1.** Fluxes, errors, and S/Ns of detected oxygen and sulphur auroral lines for each galaxy in the sample.

Galaxy ID	[O III] $\lambda$ 5007 Flux ( $\times 10^{-19}$ )	S/N	[O II] $\lambda$ 3727 Flux ( $\times 10^{-19}$ )	S/N	[S II] $\lambda$ 4068 Flux ( $\times 10^{-19}$ )	S/N	[S III] $\lambda$ 6312 Flux ( $\times 10^{-19}$ )	S/N
MARTA_3942	$3.51 \pm 0.38$	5.0	$12.8 \pm 0.99$	6.4	$3.3 \pm 1.1$	7.5	$5.28 \pm 0.67$	5.0
MARTA_4195	$1.12 \pm 0.45$	6.3	$6.02 \pm 0.81$	6.6	$0.634 \pm 0.12$	3.2	–	–
MARTA_5014	$1.33 \pm 0.28$	6.3	$11.6 \pm 0.56$	12	–	–	–	–
MARTA_552	$7.12 \pm 0.26$	22	$4.71 \pm 0.53$	6.9	–	–	$3.04 \pm 0.58$	3.9
MARTA_2387	$1.39 \pm 0.28$	4.2	$2.54 \pm 0.56$	3.8	–	–	–	–
MARTA_4327	$6.54 \pm 0.26$	15	$14.4 \pm 0.50$	12	$3.36 \pm 0.25$	8.2	$1.98 \pm 0.39$	5.1
MARTA_4502	$2.16 \pm 0.25$	13	$6.80 \pm 0.52$	6.1	$1.94 \pm 0.25$	3.8	–	–
MARTA_3408	$1.68 \pm 0.36$	4.0	–	–	–	–	–	–
MARTA_414	$2.59 \pm 0.30$	7.8	–	–	–	–	–	–
MARTA_4471	$2.07 \pm 0.33$	4.9	–	–	–	–	–	–
MARTA_329	$2.65 \pm 0.63$	4.3	–	–	–	–	–	–
MARTA_1084	$0.51 \pm 0.4$	2.6 <sup>†</sup>	$2.25 \pm 0.74$	4.6	–	–	–	–
MARTA_3115	$1.50 \pm 0.32$	3.3	–	–	–	–	–	–
MARTA_1374	$3.12 \pm 0.12$	6.1	–	–	–	–	–	–
MARTA_3887	$7.43 \pm 0.51$	11	–	–	–	–	–	–
MARTA_3926	$1.10 \pm 0.43$	7.5	–	–	–	–	–	–

**Notes.** Fluxes are given in units of  $10^{-19}$  erg/s/cm<sup>2</sup>, with errors on the fluxes estimated by the PPXF fitting. The S/Ns are calculated by multiplying the peak flux by  $\sqrt{N}$ , with N being the number of spectral channels spanned by line FWHM, and dividing it by the RMS measured in a line-free adjacent region. <sup>(†)</sup>The S/N estimated from the PPXF fit is higher than that obtained from the method based on the RMS of the spectrum; hence, we consider this a marginal detection.



**Fig. C.1.** Parametrisation of the T2-T3 plane from Yates et al. (2020). MARTA galaxies are overplotted, colour-coded by their  $T_e$ -based metallicity. The solid lines represent the T-T relation calibrated by Yates et al. (2020) for a fixed metallicity. Inset panel:  $\Delta Z$  vs  $Z$  plot, where  $Z$  is the metallicity derived from the  $T_e$  method for the MARTA objects and  $\Delta Z$  is the difference between the ‘direct’ metallicity and that obtained from the Yates et al. (2020) calibration for the same galaxies.

## Appendix D: Emission line fluxes for MARTA galaxies

Table D.1 presents line fluxes, errors, and S/N estimates for all the auroral line detections in the MARTA gold and silver sample.

## Appendix E: Representativeness of the auroral-line sample relative to the star-forming main sequence

To investigate the potential bias of the high-redshift auroral-line sample with respect to the star-forming main sequence, we

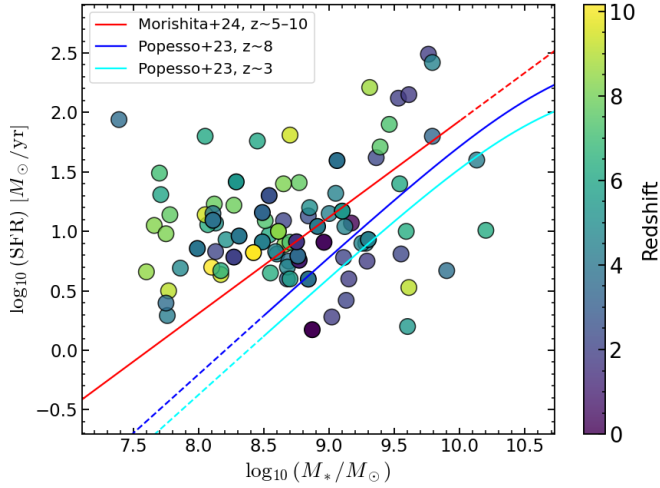
examined the location of each galaxy in the SFR- $M_\star$  plane. Stellar masses and SFRs for the literature sample were taken from the original studies, without homogenisation. We then computed the offset from the MS, defined as  $\Delta MS = \log(\text{SFR}) - \log(\text{SFR}_{\text{MS}})$ , where  $\text{SFR}_{\text{MS}}$  is the expected SFR at a given redshift and stellar mass according to the parametrisation by Popesso et al. (2023).

We find that the sample is, on average, significantly biased towards galaxies with elevated specific SFRs. The median offset across the full sample is  $\Delta MS \approx 0.7$  dex. This is illustrated in Fig. E.1, where we show the SFR- $M_\star$  diagram for all galaxies used in the strong-line calibration, along with the MS relations at different redshifts from Popesso et al. (2023) and Morishita et al. (2024b). The offset is particularly pronounced at low stellar masses ( $M_\star < 10^{8.5} M_\odot$ ) and at high redshifts ( $z > 4$ ), where only galaxies with the highest SFRs are typically detected in emission lines, and in particular show auroral lines.

To quantify this further, we split the sample by stellar mass and redshift. Galaxies with  $M_\star < 10^{8.5} M_\odot$  ( $N = 30$ ) show a median  $\Delta MS = 1.19$  dex, while more massive galaxies ( $N = 53$ ) have a lower median offset of 0.43 dex. Similarly, galaxies at  $1 < z \leq 4$  ( $N = 29$ ) have a median  $\Delta MS = 0.42$  dex, compared to 0.78 dex for those at  $z > 4$  ( $N = 54$ ). These trends reflect a combination of selection effects and the current limitations of deep spectroscopic observations at high redshifts.

The key concern is that a bias towards systems with high specific SFRs could correlate with the physical conditions in H II regions—particularly the ionisation parameter—and thus influence the derived strong-line calibrations. To address this, we examined the calibration diagrams shown in the main text (see Sect. 7), this time colour-coding each galaxy by its  $\Delta MS$ . The result, shown in Fig. E.3, reveals no clear systematic trend or shift in the location of galaxies in strong-line ratio space as a function of  $\Delta MS$ . While our sample is biased towards elevated specific SFRs, this does not appear to introduce a significant distortion in the empirical calibrations.

Nonetheless, we stress that this test is limited by sample size and the uncertainty on both the MS parametrisations and the reported SFRs. A more robust assessment would require a larger, statistically representative set of high-redshift galaxies with auroral-line detections, as well as an improved characterisation of the intrinsic scatter in line ratios at fixed metallicity. We defer a more detailed investigation to future work.

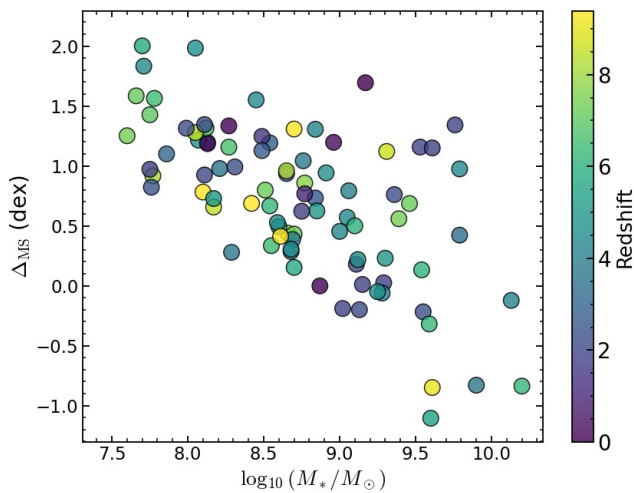


**Fig. E.1.** Locations of the literature sample galaxies used for strong-line calibrations in the SFR– $M_*$  plane. Stellar masses and SFRs are taken directly from the original studies, without re-fitting. Coloured lines indicate the main sequence relations at different redshifts from Popesso et al. (2023) and Morishita et al. (2024b).

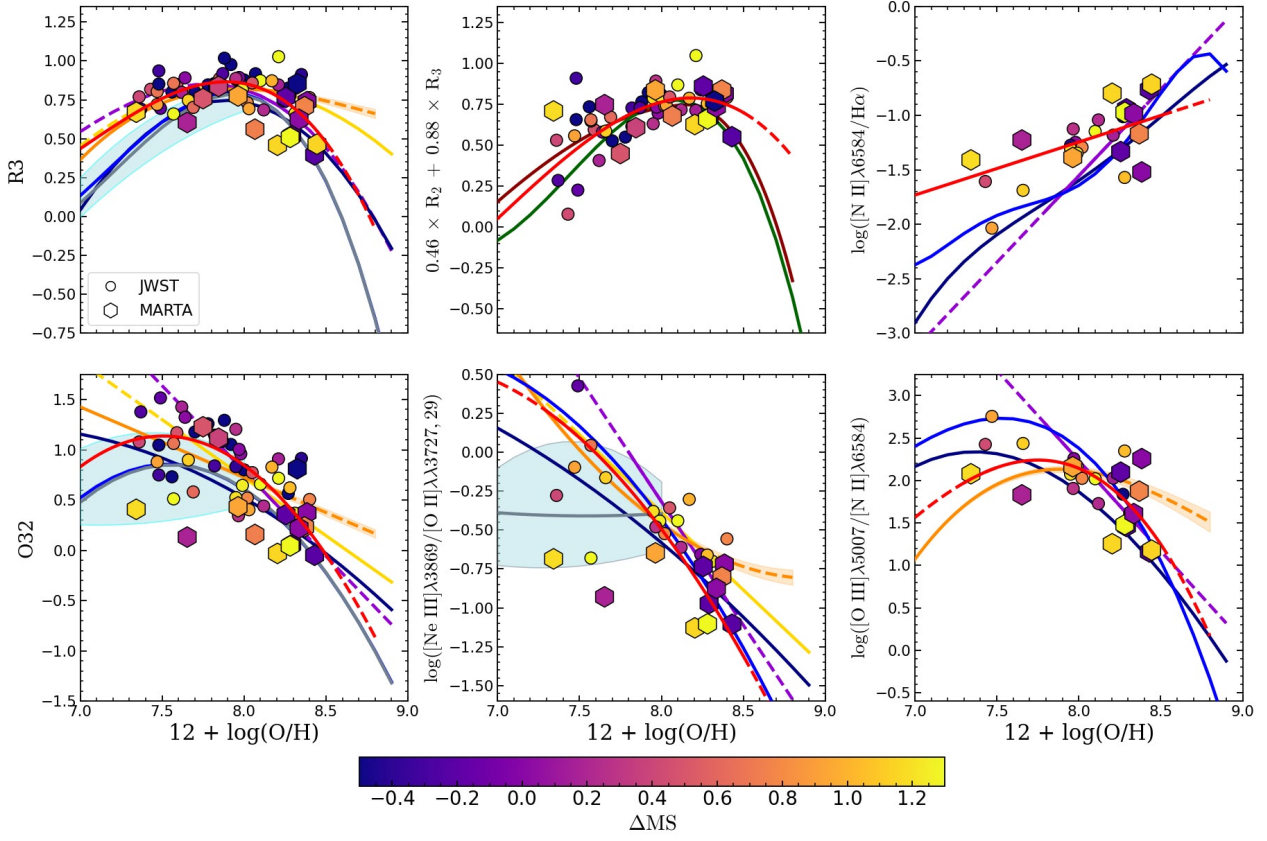
**Table E.1.** Emission lines included in the fitting procedure.

Line	$\lambda$ [Å]	Line	$\lambda$ [Å]
[O II]	3726.03	[O III]	5006.84
[O II]	3728.82	[NI]	5197.90
H13 (B)	3734.37	[NI]	5200.26
H12 (B)	3750.15	[OI]	6300.30
H11 (B)	3770.63	[SIII]	6312.06
H10 (B)	3797.91	[OI]	6363.78
H9 (B)	3835.40	[NII]	6548.05
[Ne III]	3868.76	H $\alpha$	6562.79
HeI + H8 (B)	3889.00	[NII]	6583.45
[Ne III]	3967.47	HeI	6678.15
H7 (B)	3970.08	[S II]	6716.44
[S II]	4068.60	[S II]	6730.82
[S II]	4076.35	HeI	7065.25
H $\delta$	4101.73	[ArIII]	7135.79
H $\gamma$	4340.47	HeI	7281.35
[O III]	4363.21	[O II]	7319.40
HeI	4471.50	[O II]	7330.35
HeII	4685.70	[ArIII]	7751.15
[ArIV]	4711.37	H11 (P)	8862.89
HeI	4713.17	H10 (P)	9015.30
ArIV	4740.17	[S III]	9068.60
H $\beta$	4861.35	[S III]	9530.60
HeI	4921.93	HI	9545.93
[O III]	4958.91	HeI	10830.20

**Notes.** Wavelengths are in air. (B) denotes Balmer and (P) Paschen lines.



**Fig. E.2.** Offset from the main sequence of Popesso et al. (2023,  $\Delta_{\text{MS}}$ ) as a function of stellar mass.



**Fig. E.3.** Strong-line calibration diagrams (see Sect. 7), with each galaxy colour-coded by its  $\Delta\text{MS}$ .