High-z galaxies with JWST and local analogues — it is not only star formation

Jarl Brinchmann,1,*
1Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

I present an analysis of the JWST NIRSpec data of SMACS 0723 released as Early Release Observations. As part of this three new redshifts are provided, bringing the total of reliable redshifts to 14. I propose a modification to the direct abundance determination method that reduces sensitivity to flux calibration uncertainties by a factor of ~ 3 and show that the resulting abundances are in good agreement with Bayesian photoionization models of the rest-frame optical spectrum. I also show that 6355 is most likely a narrow-line active galactic nucleus (AGN) with $M_\ast < 10^7 M_{\odot}$ at $z = 7.66$, and argue that 10612 might also have an AGN contribution to its flux through comparison to photoionization models and low-redshift analogues. Under the assumption that the lines come from star-formation I find that the galaxies have gas depletion times of ~ $10^7$ years, comparable to similar galaxies locally. I also identify a population of possibly shock-dominated galaxies at $z < 3$ whose near-IR emission lines plausibly come nearly all from shocks and discuss their implications. I close with a discussion of the potential for biases in the determination of the mass-metallicity relation using samples defined by detected [O iii]4363 and show using low-z galaxies that this can lead to biases of up to 0.5 dex with a systematic trend with mass.

Key words: galaxies: evolution – galaxies: fundamental parameters – galaxies: distances and redshifts

1 INTRODUCTION

The study of the ionized gas in galaxies has a long and storied history with much work in particular focused on low-redshift galaxies. One strand in these investigations has been the study of low-redshift galaxies as analogues to high redshift galaxies. This has been particularly focused on studies of low-metallicity star-forming galaxies (e.g. Izotov et al. 1994; Kunth & Ostlin 2000; Papaderos et al. 2008; Papaderos & Ostlin 2012) as they are thought to offer the most promising sites to study star formation under conditions similar to the high-redshift Universe. In parallel to this, the advent of large spectroscopic catalogues of galaxies through, in particular, the Sloan Digital Sky Survey (York et al. 2000) saw systematic searches for extreme emission line galaxies that have been considered to provide examples of galaxies where star formation conditions are similar to those at high redshift (e.g. Izotov et al. 2006; Brinchmann et al. 2008b; Cardamone et al. 2009; Pérez-Montero et al. 2021). As fashion has taken the field, the foucs has been on H II galaxies (e.g. Telles et al. 1997), Blue Compact Dwarfs (BCDs, Kunth & Ostlin 2000), Lyman-break analogues (Overzier et al. 2009; Heckman et al. 2011; Wu et al. 2019), peas of different colours (Cardamone et al. 2009; Yang et al. 2017a; Brunner et al. 2020; Yang et al. 2017b), and extreme line emitters pushed out to ever higher redshift (e.g. Shirazi & Brinchmann 2012; Bian et al. 2016; Maseda et al. 2014; Amorín et al. 2015). There are significant overlaps between most of these classes and they are typically tailored to a particular scientific question.

The real advantage of low-z analogues is that they allow us to study the well-understood optical spectral region with high-S/N spectra (see Kewley et al. 2019; Maiolino & Mannucci 2019, for recent reviews) and it is easier to collect multi-wavelength data to interpret these observations than at high redshift. That said, rest-UV spectroscopy of local analogues has been hard to get (Leitherer et al. 2011) although the CLASSY survey (Berg et al. 2022; James et al. 2022) has recently made big strides forwards here. It has also been challenging to acquire near-IR spectroscopy of these analogues (e.g. Vanzi et al. 2008; Hunt et al. 2010; Cresci et al. 2010; Izotov & Thuan 2016), something that might hamper future interpretations of NIRSpec data on $z \sim 1–3$ galaxies.

The advent of rest-frame optical spectroscopy of $z > 5$ galaxies with the James Webb Space Telescope (JWST) NIRSpec spectrograph (Jakobsen et al. 2022; Ferruit et al. 2022) has now opened up the possibility to contrast these low-redshift counter-parts with the real high-redshift galaxies and to better understand how local analogues can help guide our study of high-z galaxies. But it is worth reflecting upon what analogues can be useful for and a caution that an analogue is just that, and should not be viewed as a replacement for studying the high redshift galaxies directly.

To do so, it is useful to distinguish between large-scale environmental properties, extrinsic properties (scaling with the size of the system) and intrinsic properties. There is no real way to find fully equivalent environmental conditions for low-z analogues as compared to a $z \sim 8$ galaxy since the Cosmic Microwave Background temperature is $9(1+z)/(1+8)$ times higher and the mean density of the Universe is $\sim 730 ((1+z)/(1+8))^3$ times denser than today. Since the large-scale environment frequently has an impact on the extrinsic properties of galaxies (e.g. Peng et al. 2010), this might
argue against their use for comparisons. Here I will therefore focus on intrinsic quantities and in particular emission line ratios to match samples at low and high redshift.

The public release of an Early Release Observations using NIRSpec over the SMACS0723.3–7327 galaxy cluster (SMACS 0723 hereafter) spurred a flurry of studies of the rest-frame optical spectra of the five clear high-z galaxies in the data (Curti et al. 2023; Trussler et al. 2022; Katz et al. 2023; Carnall et al. 2023; Schaerer et al. 2022; Trump et al. 2023; Rhoads et al. 2023; Arellano-Cordero et al. 2022; Tacchella et al. 2023; Taylor et al. 2022), and I will return to a comparison to some of these results further below. Subsequently a number of NIRSpec studies have extended this sample to higher redshift (e.g. Bunker et al. 2023; Curtis-Lake et al. 2023; Robertson et al. 2023) as well as more extensively covering $z \sim 1$ to $z \sim 9$ (e.g. Sanders et al. 2023a; Cameron et al. 2023; Shapley et al. 2023; Reddy et al. 2023; Maseda et al. 2023) and some exploration of non-star formation sources (e.g. Larson et al. 2023).

The focus in many of these papers has been on the $z > 6$ galaxies given their novelty and closeness to the epoch of reionization, the peak of star formation in the Universe happens between $z = 1$ and $z = 3$ (Madau & Dickinson 2014) and this is a redshift range where NIRSpec gives access to a rich array of near-IR lines. These include Paschen and Brackett lines of Hydrogen, which are less affected by extinction than optical lines and hence are powerful references for star formation rate estimates (e.g. Kennicutt & Evans 2012; Reddy et al. 2023), although this should be tempered by the fact that they are relatively weak: in the absence of dust attenuation it is a useful rule of thumb that Pas is about as strong as Hα relative to Hz, Paβ similar to Hβ and Pay to H(8 $\rightarrow$ 2), and the Brackett lines a factor of ~ 5 weaker still.

However besides hydrogen lines, the near-IR has prominent He I, [Fe ii], and H$_2$ lines which offer complementary information on the physical properties of galaxies to that provided by the optical lines. [Fe ii] lines are frequently weak in star-forming galaxies because the ion is locked up in dust grains, but they are found to be enhanced in shocked regions because of the destruction of dust grains behind shock fronts due to sputtering processes there (Oliva et al. 1989; Greenhouse et al. 1991). This has led to [Fe ii]1.257 and in particular [Fe ii]1.644 being used for studies of supernovae in galaxies (e.g. Oliva & Moorwood 1990; Alonso-Herrero et al. 1997; Rosenberg et al. 2012; Brusa et al. 2014) and they are also regularly seen in AGN (Mouri et al. 2000) and at a much weaker level in star forming galaxies coming from photoionization (Izotov & Thuan 2016; Cresci et al. 2010; Vanzì et al. 2008, 2011). As we will see below, they are also seen at an interesting level in the NIRSpec Early Release Observations (ERO) data.

The molecular hydrogen emission spectrum coming from ionized gas offers a powerful way to characterise the physical conditions in the warm molecular regions (e.g. Black & van Dishoeck 1987; Kaplan et al. 2017) and will provide novel information on the nature of $z \sim 2$ galaxies although the detection of a large set of rovibrational lines of H$_2$ will likely be challenging in most cases.

That flux calibration will be a challenge for NIRSpec has been clear for a long time given the very small slits (e.g. Jakobsen et al. 2022; Ferruit et al. 2022). In section 2 I will discuss the method I adopted to correct the flux calibration of the spectra which is complementary to those used in the literature, I also will present the new redshift determinations here. I will discuss the emission line measurements in section 3. In section 4 I propose a modification to the standard temperature sensitive abundance estimation method to reduce the sensitivity to flux calibration errors and use this to derive electron temperatures and oxygen abundances which I compare to the literature on these sources. 5 is dedicated to fits of photoionization models to the fluxes of these high-z galaxies and section 6 has an in-depth discussion of the ionizing sources in these galaxies which until then I will tacitly assume to be star formation. In section 7 I compare to local analogues and this feeds into the discussion in section 8 before I conclude in section 9.

Where relevant I have adopted a Kroupa (Kroupa 2001) initial mass function and I will use a cosmology with $\Omega_m = 0.3$, $\Omega_{\Lambda} = 0.7$ and $h = 0.7$. For forbidden and helium emission lines with rest wavelengths below 1$\mu$m I will indicate that wavelength of the transition in Å, while for longer wavelength lines I will indicate the wavelength in $\mu$m. Thus [S ii]9533 for the [S ii] line at 9533.2Å, but He i 1.083 for the He i line at 1.083$\mu$m. All wavelengths are given in vacuum.

2 DATA

I will here use the JWST NIRSpec observations taken as part of the SMACS0723 Early Release Observations (Programme ID 2736) which have already been discussed in detail by several authors (e.g. Carnall et al. 2023; Curti et al. 2023; Trump et al. 2023). I will also make some use of the deep NIRCam imaging in F909W, F150W, F200W, F277W, F356W, and F444W released with the NIRSpec spectra, as well as the Hubble Space Telescope (HST) images in F435W, F606W and F814W made available by the Reionization Lensing Cluster Survey (RELS Coe et al. 2019)\(^\dagger\). For the imaging I use the latest re-reductions provided by Gabe Brammer’s Grizli Image Release v6.0\(^\ddagger\).

Source extractor version 2.25 (Bertin & Arnouts 1996) was used to construct a source catalogue. For some sources Source Extractor’s segmentation maps were not optimal for total photometry so for these six sources (4580, 5735, 9483, 9721, 10612, 5144) photometry was done manually. The effect on the results presented below is marginal since the colours did not change significantly. For aperture photometry I used a 0.4″ diameter aperture and for total magnitudes I used Source Extractor’s MAG\_BEST. To derive aperture corrections for the fixed apertures, I used the WebbPSF python package (Perrin et al. 2012) to create point spread functions and calculated aperture corrections assuming an intrinsic point source. For the manual photometry I convolved the images to the F444W PSF using pypher (Boucaud et al. 2016) to calculate the appropriate convolution kernels. I chose to do this since the manual photometry in some cases has to use tight segmentation masks, but the effect of the convolution on the final photometry is < 0.1 magnitudes in almost all cases.

The main focus here is on the NIRSpec spectroscopy which was done with two pointings, s007 and s008, and released as level-3 (L3) data products on 12/07/2022. The data were obtained with two grism and filter combinations, G235M/F170LP covered approximately 1.65$\mu$m–3.17$\mu$m and G395M/F290LP covered 2.85$\mu$m–5.28$\mu$m although close to the edges the spectra are considerably worse. I took the spectral resolution as built from the JWST web pages, which are those shown in Jakobsen et al. (2022).

For the analysis here I in general combine the two pointings into one for each aperture. For the fitting of emission lines I also combine the two gratings into a single spectrum. However for checks and tests

\(^\dagger\) https://relics.stsci.edu/

\(^\ddagger\) https://github.com/gbrammer/grizli/blob/master/docs/grizli/image-release-v6.rst, but I have also used the original release of ERO images and will discuss the sensitivity of the results to the photometric calibration further below.

1. [Link to the website](https://relics.stsci.edu/)
2. [Link to the GitHub repository](https://github.com/gbrammer/grizli/blob/master/docs/grizli/image-release-v6.rst)
I also analyse each pointing or grating independently. The data shows various features that make analysis somewhat more challenging. In particular the flux calibration is suspect at times and I will return to this point below. In some cases, the spectral extraction is also sub-optimal for redshift determination because emission lines have less contrast. Thus throughout I made extensive use of the 2D spectra to assess the reality of lines and also to re-extract the spectra to have better redshift estimates.

For re-extraction, I defined a box around the region I wanted to extract a spectrum from in the 2D spectrum and defined background side-bands of equal width on either side of this central spectral trace. The side-bands are subtracted off which helps remove some structure in the spectrum as also noticed by other authors (e.g. Trump et al. 2023). I also manually edit out cosmic rays.

2.1 Flux calibration and slit losses

The SMACS0723 ERO NIRSpec data have clear issues with their spectrophotometric calibration as was noted already by the first publication discussing the data (Schaerer et al. 2022) and has been discussed repeatedly since then. Several authors have therefore re-reduced the data (Curti et al. 2023; Trump et al. 2023). In particular Curti et al. (2023) have re-reduced the spectra using the GTO pipeline and report more physically meaningful flux ratios. As that pipeline is not available outside the GTO consortium, and because the work presented here was done before the Curti et al paper appeared, I have taken a different approach, which I will show further below gives very similar results to the Curti et al re-analysis.

However even with perfect reductions, it is clear that flux calibration of NIRSpec can present significant challenges. The NIRSpec slit width is only 0.2 arcsec in width and as stressed in Ferruit et al. (2022), this can have significant impact on slit-losses and in particular our ability to compare line fluxes across wide ranges in wavelength. Given this, I am using the originally reduced data but present a method to a posteriori correct these.

To try to correct for biases in the flux calibration and mitigate and explore the slit-loss effect, I have normalised the spectra to the NIRCam (Rieke et al. 2005; Beichman et al. 2012) photometry. The spectra were first convolved with the NIRCam F200W, F277W, F356W, and F444W filters. These are then compared to the NIRCam fluxes for the sources. I use both the aperture corrected fluxes through a fixed 0.4′′ aperture, and total fluxes. Unless otherwise stated, the results below use aperture fluxes for the normalisation. In most cases this comparison shows a gradient with wavelength.

I then assume that the spectrum that is lost is equal to the spectrum extracted within the slit, in other words I assume that

\[ f_{\text{total}}(\lambda) = f_{\text{slit}}(\lambda) + \alpha(\lambda)f_{\text{slit}}(\lambda), \]

where “total” refers to the true total flux, and “slit” corresponds to that measured through the slit. The comparison gives us the average of \( \alpha \) over the filter, and I then assume that this gives us an approximate measure of \( \alpha \) at the pivot wavelength of the filter. The \( \alpha(\lambda_{\text{pivot}}) \) are then fit with a linear function for each grism and applied as a correction to get the final corrected spectrum. In updating this paper a similar, but distinct, method was developed by Reddy et al. (2023) and shown to work well also for their spectra.

This process gives satisfactory results in most cases. Figure 1 shows the result of this procedure on 3042 which, as I’ll return to below, is a complex superposition of two galaxies which is likely the cause of the significant change in slope here. Most other galaxies show much smaller changes, but 5144 does not yield a good solution so its spectrum is used unmodified, while 4580 and 5735 do not have enough fluxes in the NIRCam catalogue to be well corrected, thus these have also been used unmodified. Further below I will also demonstrate that the measured Balmer ratios are physically meaningful and agree well with the ones measured off the re-calibrated data presented by Curti et al. (2023).

Although the process superficially works well, it has, however, a number of weaknesses: the assumption of a linear gradient with wavelength is unjustified but can be viewed as the lowest order Taylor expansion of the correction function; the extrapolation to the reddest and bluest wavelengths is significantly uncertain, and of course the assumption of a constant spectral shape inside and outside the slit is questionable. That said, the method does improve the spectral shape quite significantly in some cases and by combining this method with an SED fitting approach and optimised photometry, some of these weaknesses can be addressed (see e.g. Tacchella et al. 2023).

It is however important to emphasise that slit-losses are unavoidable with NIRSpec and it is therefore important to develop methods to take this into account in the analysis. The suggestion from the JWST documentation is to create a forward model for the light convolved with the results for a point source. This is satisfactory if the emission lines follow the broad-band light profile, ie. have constant equivalent widths as a function of radius — that is patentently not the case in some classes of low-z galaxies but is likely more trustworthy at high redshift. A systematic IFU survey of galaxies with NIRSpec will help inform this and can then be combined with forward models of the emission line distributions (e.g. Carton et al. 2017, 2018; Espejo Salcedo et al. 2022).

Alternatively one can limit the effect of slit-losses by constructing analysis methods that focus on nearby emission lines. This implicitly assumes that emission line ratios are radially constant which might be questionable and certainly is at low redshift, however in lieu of other information is not a bad assumption. This is the approach I will take, and has also been the approach taken by all the papers using the early ERO NIRSpec data.

---

3 Taken from https://jwst-docs.stsci.edu/jwst-near-infrared-camera/nircam-instrumentation/nircam-filters
2.2 Redshift measurements

I present here 14 high confidence redshifts, 10 of which have already been presented in C22, one provided in Mahler et al. (2023), plus one lower confidence one, coming from the same spectrum with a high confidence redshift. The redshifts are in general obvious and have been determined by inspection. The details of the sources are given below in Table 1. The redshift of 3042b is highly uncertain and is discussed below.

There is however one point that warrants some discussion. This concerns blending/multiple sources within the aperture. Like any slit spectrograph, NIRSpec can have multiple sources fall onto the slit and thus have multiple traces. Given the depth that NIRSpec can routinely reach, this can introduce some challenges but as emphasised by Maseda et al. (2019) also opportunities. In the current dataset there are two clear cases of this: 8277 which has two sources well-separated within the slit, and 3042 which has two sources overlapping. For 8277 I have no secure redshift as I can only identify one clear line for the brightest source and none for the fainter, thus there is nothing more to be said about this here.

However, 3042 offers an interesting case study of a type that one should watch out for in future NIRSpec studies, particularly the deeper ones.

3042 is rather more complex than it might appear at first glance. There is a clear detection of the source in ACS F606W images and it continues to be detected through all NIRCAM images. However, the morphology changes dramatically through F090W and into F200W as illustrated in Figure 2. The colour image on the left on the bottom row combines the F090W, F150W, and F200W images to demonstrate the clear colour difference between the objects.

The simplest interpretation of this is that we have two sources overlapping. This has been long recognised to be a challenge with deep MUSE observations (e.g. Bacon et al. 2015; Brinchmann et al. 2017; Bacon et al. 2021) where a secure redshift can be easily found, but the assignment of this to a photometric object can be challenging, adding another axis to the classical spectroscopic confidence assessment. This is also the case here and this is likely to happen frequently in deep NIRSpec data as discussed in detail by Maseda et al. (2019), and clearly multi-band JWST/HST imaging will be paramount in deciphering these cases.

Here the low redshift solution is clearly}\ z = 1.9938\ with\ \text{H}\alpha,\ [\text{N}\,\text{ii]}6584,\ \text{[S}\,\text{ii]}9533,\ \text{Pa-}\delta\ \text{and}\ \text{Pa-}\gamma\ all\ clearly\ seen\ although\ \text{Pa-}\gamma\ is\ not\ very\ well\ detected\ as\ it\ falls\ at\ a\ wavelength\ where\ the\ flux\ calibration\ is\ particularly\ problematic.\ \text{[Fe\,\text{i}]}1.257\ is\ also\ seen.\ My\ interpretation\ of\ this\ is\ that\ the\ elongated\ source\ detected\ in\ the\ ACS\ images\ is\ a\ lower\ redshift\ galaxy\ and\ that\ the\ redshift\ corresponds\ to\ this.

In contrast the source seen in F150W and redwards appears to be a higher redshift source. There are two possible redshifts for this source based on emission lines. There appears to be two lines in the spectrum at 4.7634 and 4.8828 micron, clearly seen in the 2D spectrum as can be seen in the top middle panel in Figure 2. The separation of these two lines matches well H\alpha\ and\ \text{[S}\,\text{ii]}6717,6731\ at\ z = 6.253 and there are not really any other set of lines that match equally well. There are also no signs of\ \text{[O}\,\text{iii]}4959,5007.\ Thus\ the\ confidence\ for\ this\ redshift\ is\ very\ low.\ In\ addition\ there\ are\ two\ lines\ at\ 2.852\mu m\ and\ 2.873\mu m\ also\ visible\ in\ the\ 2D\ spectrum.\ Their\ separation\ is\ well\ fit\ by\ \text{[N}\,\text{ii]}6548,6584\ at\ z = 3.258,\ but\ again\ no\ further\ supporting\ lines\ are\ visible\ so\ the\ confidence\ is\ low.

Photometric redshifts offer a possible way to distinguish between these two sources. To that effect I measured the flux of the high redshift sources in the NIRCam images, after masking out the lower redshift source. These fluxes were given to the EAZY (Brammer et al. 2008) photometric redshift code using the approach used for very faint galaxies in Brinchmann et al. (2017). The resulting P(\text{z}) is broad but rules out any redshift solutions \text{z} > 4.5 and \text{z} < 1.5. While not conclusive since photo-z outliers do exist, this at least argues against the \text{z} = 6.253 solution and I have there adopted the \text{z} = 3.258 solution below but it is important to stress that this is highly uncertain.

3 EMISSION LINE MEASUREMENTS

Since one of the main goals of this paper is to compare the high-z galaxies with local counter-parts, it is preferable to use the same analysis tools used at low-z for line measurements. Thus I have adopted the platefit tool used as a basis for the MPA-JHU database\footnote{https://wwwmpa.mpa-garching.mpg.de/SDSS/DR7/} of galaxy properties (Tremonti et al. 2004a; Brinchmann et al. 2004a; Brinchmann et al. 2013). This has been well tested over two decades on surveys such as VIMOS Very Deep Survey (VVDS, e.g. Lamerelle et al. 2009) and MUSE GTO (Bacon et al. 2021) and is fairly robust to spectrophotometric calibration errors and low S/N spectra.

Briefly, to estimate the continuum, platefit carries out a non-negative least squares (Lawson & Hanson 1974) combination of Bruzual & Charlot (BC03, 2003) simple stellar population models as well as a power-law attenuation of the continuum. The best fit model is adopted and subtracted off. On the residual spectrum a smoothed continuum is constructed by taking a median smooth with box size 151 pixels and additional boxcar smooth with a window of 51 pixels. These smooth sizes were optimized for SDSS spectra, but were found to perform satisfactorily on the JWST spectra as well. They are too wide to handle abrupt changes in the continuum between gratings due to calibration issues but as none of the main lines to be measured fall close to these edges this is not a major problem.

After also subtracting this smooth continuum, the emission lines are fit jointly in velocity space with a common velocity width and overall velocity offset, assuming that the line shape is Gaussian. As remarked above, we take the line spread function from the JWST web-site as measured pre-launch. For the MPA-JHU catalogue we tied\text{[Ni]}6548 to\text{[Ni]}6584 using a theoretical ratio of 1.3. I have updated this to use the latest NIST value of the relative ratio of 1.2957 and also tie the fluxes of\text{[Ne\,\text{ii}]}3869 and\text{[Ne\,\text{ii}]}3967 together using a theoretical ratio of 1.322. I do not tie the more widely separated pairs of lines\text{[O\,\text{ii}]}4959,5007,\text{[S\,\text{ii}]}9069,9533,\text{[Fe\,\text{i}]}1.257,1.644 together as the residual flux calibration uncertainties could play havoc with this.

The resulting fits are acceptable but the low S/N in the continuum, if it is even detected, means that the continuum fit is poorly constrained. This does not have a strong influence on the emission line measurements but more care will be needed to obtain measurements of continuum properties. In particular the method used above, and used for the MPA-JHU catalogue, ignores the effect of a nebular continuum although the power-law attenuation accounts for this some extent. However, as shown by Cardoso et al. (2019); Pappalardo et al. (2021) this can be important precisely for these kinds of galaxies, and a more careful analysis with e.g. FADO (Gomes & Papaderos 2017) as has been done for extreme emission line galaxies by Breda et al. (2022), or using a modified platefit as done by Gunawardhana.
et al. (2020) in their analysis of the MUSE data on the Antennae, would address this. Secondly, the BC03 models are effectively based on scaled-solar abundance spectra and it is highly likely that the stars making up the stellar continuum in the highest redshift galaxies have different abundance ratios due to the different enrichment time-scales for core-collapse and type Ia supernovae, and indeed this has been argued to be of major importance also at redshift $z \sim 2$–3 (Strom et al. 2017; Topping et al. 2020). For the optical region of interest here, however, this should not be important.

In the lower redshift galaxies, where NIRSpec covers the rest near-IR to red optical, the platefit results are less satisfactory in part due to the lower spectral resolution of the BC03 templates used. I therefore also fit the spectrum manually using Gaussians. I adopt joint fits of Gaussians to blended, or nearly blended, lines such as C$\text{III}$]1907,1909, O$\text{II}$]3726,3729, S$\text{II}$]6717,6731 and a joint triple Gaussian fit for [N$\text{II}$]6548,6584 and H$\alpha$. These Gaussian fits compare well with the platefit results but behave better in the near-IR part of the spectra so I’ll adopt those for the analysis of the $z < 3$ galaxies.

3.1 Noise estimates

It is imperative to have good noise estimates for spectra before one measures line fluxes and in particular try to infer physical parameters from the line ratios which are more sensitive to noise. The most common approach in the papers appearing thus far has been to adopt the noise estimate from the pipeline. This appears to underestimate the true noise in the spectrum, a fact also noted by several other researchers (e.g. Rhoads et al. 2023; Trump et al. 2023).

We can quantify this by running platefit on each individual observation separately and comparing the derived fluxes. Concretely I selected all lines with S/N $> 5$ (a cut of 7 or 10 gives comparable results), and calculated the standardized difference,\[ \Delta = \frac{f_{007} - f_{008}}{\sqrt{\sigma^2_{007} + \sigma^2_{008}}}, \] (2) \[ \] where 007 and 008 refer to the individual observations and $\sigma$ is the flux uncertainty. Under the assumption that all uncertainties are normal, this should be distributed as a unit variance normal distribution. Deviations from this can come from two sources. Firstly, the uncertainty estimates themselves are uncertain and if this is important, $\Delta$ will be distributed as a Student’s t distribution rather than a normal — this is most easily seen in the wings of the distribution but as only 73 flux measurements were available here, this is impossible to assess. The other, and more relevant source is that underestimated uncertainties tend to inflate $\Delta$. For SDSS DR7 we found that this was indeed a concern (Brinchmann et al. 2013, BC13 hereafter) and here too I find that the uncertainties delivered by platefit are underestimated by a factor of 2.75 when using the pipeline reductions. A part of this is due to platefit only reporting the diagonal elements of the covariance matrix, but when applying gaussian fits manually to the lines I
also find a discrepancy of a factor of 2 so this appears robust. I also find a slight bias between the two observations in that 008 is slightly brighter. Given the combination and subsequent renormalisation of the spectra, this does not matter for the results presented here.

More directly, it is notable that the root-mean-square of the spectrum in line-free regions is larger than expected from the pipeline noise spectrum so I have opted to be conservative and adopt an empirical RMS spectrum as my noise spectrum. I use a window of six pixels, calculate the standard deviation in this sliding window, reject outliers that are more than 3σ deviant and recalculate the standard deviation. The results presented below are not very sensitive to this approach but the noise correction will become an issue again in section 9 below.

3.2 Sample characteristics and source description

The basic characteristics of the sources with secure redshifts are given in Table 1. Line fluxes measured for the sources are provided in the online material. The two sources 4580 and 3042b will not be included in the emission line analysis below as they have insufficient spectral information to be useful.

4 TEMPERATURES AND ABUNDANCES USING THE EMPIRICAL METHOD

The classical way to estimate temperatures of ionized nebulae is based on the ratio of the fluxes of auroral transitions to lower-lying transitions. This is a fairly straightforward technique and it was already well-established when Aller wrote his influential book “Gaseous Nebulae” (Aller 1956). It relies on the relatively strong temperature dependence of auroral lines relative to lower-lying levels.

The method has been widely used in nearby galaxies and star-forming regions although the faintness of auroral lines has always limited their use and they are rarely detected at high redshift (see Sanders et al. 2020, for a compilation) and have been almost exclusively focused on [O iii] detections with the recent detection of auroral [O iii]7322,7332 by Sanders et al. (2023b) the exception.

The most widely used ratio at low redshift and the most interesting one for the present sample is the [O iii]4363/[O iii]5007 ratio of doubly-ionized oxygen. This is insensitive to density over a range of densities likely to occur in star forming galaxies. For concreteness I will fix the density to $n_e = 10^2$ cm$^{-3}$ where relevant. The limiting factor here is usually the detection of [O iii]4363 but also the relative flux calibration of the spectrum, including de-reddening. In view of the flux calibration issues with the spectra it is therefore desirable to find a more robust estimator both for temperatures and abundances.

We can improve the robustness by using the double, or composite, line ratio$^5$ [O iii]4363/Hγ/[O iii]5007/Hβ instead of [O iii]4363/[O iii]5007. Double line ratios were explored by Evans & Dopita (1985) from a different angle, but they are not widely used. They do, however, offer a way to construct diagnostic diagrams that are insensitive to dust attenuation or flux calibration issues which can be helpful not only here but also when combining JWST with ground-based or other space-based facilities. I note that this ratio was also used by Trump et al. (2023) but they used it to estimate [O iii]4363/[O iii]5007 using a fixed Hγ/Hβ ratio, this however ignores the temperature sensitivity of the Hγ/Hβ ratio and as it is trivial to calculate the necessary emissivity ratios using e.g. pyneb there is no need to make this approximation.

The proposed double line ratio normalises the fluxes of the oxygen lines to their closest Balmer lines. That takes out any large-scale flux calibration issues and in fact makes the line ratio quite insensitive to dust attenuation as well. To illustrate this the left panel of Figure 3 shows the relative error we make in estimating $T_e$ when not accounting for dust attenuation. I have assumed here a power-law dust law with $\tau(\lambda) \propto \lambda^{-1.3}$. I chose dust attenuation as one example of a spectral slope error, other functional forms like a power-law or a linear flux error give similar results and can give errors in the opposite direction as well.

The panel shows the error made using the standard ratio of [O iii]4363/[O iii]5007 as solid lines, for five different true temperatures as shown. The error here can exceed 30%. In contrast, the dashed lines show the same but now using the double line ratio [O iii]4363/Hγ/[O iii]5007/Hβ. Each dashed line corresponds to the similarly coloured solid line and it is clear that the double line ratio is much more robust to calibration/dust uncertainties with an error < 8%. This makes it particularly useful for the current situation and indeed given the potentially serious slit-loss problems for NIRSpec it would be advisable to use this approach in general. For the specific example here, an iterative solution using Balmer lines as e.g. used by Curti et al. (2023), is also insensitive. However, if instead of a smooth function with wavelength, there is an offset in flux, the iterative solution will not give correct results while the double line ratio method will continue to work.

While not of relevance here, I note in passing that a similar approach can be used for some other $T_e$ estimators: e.g. O iii]1661,1666 can be normalised to He i 6400 and [O iii]5007 to He ii 4686, [S ii]9533 to Paβ and [S i]6312 to Hα, for instance.

The calculation of abundances follows from the calculation of temperatures. The traditional way to do this is to calculate line ratios relative to Hβ and multiply this by the ratio of the emissivity of the metal line to that of Hβ. However, there is nothing special about Hβ and this works just as well when calculating ratios to other Balmer lines. In the right panel of Figure 3, I show the difference in derived total oxygen abundance, $OH = 12 + \log O/H$ relative to the true value as a function of the applied dust attenuation, here too the dust attenuation is just a concrete example of a flux calibration error. To construct this figure I fitted a linear relation to the ratio of the ionic abundances of O$^+$ and O$^{++}$ as a function $T_e$ for the sample of galaxies analysed in Brinchmann et al. (2008b) up to $T_e = 17,000$K and a constant above that. Concretely I found

$$\frac{n(O^+)}{n(O^{++})} = 0.8523 - 0.3921T_e([O \text{ iii}],$$

(3)

for 1000K $< T_e < 17000$K. I used this to assign [O iii]3727 and [O iii]5007 fluxes using the emissivities of the two transitions at the given $T_e$ and $n_e = 10^2$ cm$^{-3}$. In general the density will of course vary, but if the density is not known, as in the case of these high–z sources, it will introduce the same bias/uncertainty regardless of the method. Here it suffices to note that a change in density up or down by an order of magnitude will change the [O iii]3727/[O iii]5007 emissivity by a factor of < 10%. For the calculation shown in the right panel of Figure 3 I used 16 $T_e$ values, evenly distributed between 10,000K and 25,000K. The resulting fluxes were then attenuated by dust and used to calculate $T_e$. For simplicity I adopted single zone model with the temperature either derived using the standard [O iii] ratio or the double line ratio proposed above.

---

$^5$ If the reader wants a shorthand, I offer OOH but will refrain from using it.
Table 1. The redshifts and main lines indentified in the sources discussed here. Most of these were already presented in Carnall et al. (2023) but 3042, 5735, 9721 and potentially 3042b are new here. The uncertainties on the redshifts are from Gaussian fits to individual lines.

| Object | Redshift | Confidence | Lines | Comments |
|--------|----------|------------|-------|----------|
| 1917   | 1.2430 ± 0.0003 | 3.0 | [S II]9533, He i 1.083, Paα | |
| 3042   | 1.9934 ± 0.0040 | 3.0 | Hα, [S II]9068,9533, He i 1.083, Pay | Two overlapping galaxies with one invisible in F090W. The redshift Mahler et al. (2023) is for the lower redshift source. |
| 3042b  | 3.258    | 1.0 | Possible [N II]6548,6584 | Highly uncertain redshift. |
| 4580   | 5.1727   | 2.0 | [O III]5007, Hα. | Pointed out by Mahler et al. (2023). Only visit 008 is used to avoid problems near Hα |
| 4590   | 8.4951 ± 0.0021 | 3.0 | C II]1909, [O III]3727, to [O II]5007 | Clear redshift but only observation 008 used |
| 5144   | 6.3787 ± 0.0007 | 3.0 | [O III]3727 through to Hα | Only exposure 007 was used. Spectrophotometric recalibration unsatisfactory. |
| 5735   | 1.5073 ± 0.0004 | 3.0 | [S II]9533, He i 1.083, Paα, Paβ | Unsatisfactory recalibration |
| 6355   | 7.6640 ± 0.0010 | 3.0 | [Ne IV]2423, [O III]3727, to [O II]5007 | [Ne v]2423 seen in both 007 and 008. The combined spectrum has a feature at [Ne v]3428 but this appears only in 007. |
| 8140   | 5.2745 ± 0.0013 | 3.0 | [O II]3727 through to Hα | |
| 8506   | 2.2115 ± 0.0004 | 3.0 | Hα, [S II]6716,6732, [S II]9068,9533, Paα, He i 1.083, Pay, Paβ | Strong continuum so lines are hard to confirm in 2D image. |
| 9293   | 2.4624 ± 0.0006 | 3.0 | Hα, [S II]6716,6732, [S II]9533, Paα, He i 1.083, Pay, Paβ, [Fe II]1.257 | The only galaxy with prominent H2 emission lines. |
| 9483   | 1.1616 ± 0.0003 | 3.0 | [S II]9068,9533, He i 1.083, [Fe II]1.257, Pay, Paβ, Paα, Brδ, Brγ, He i 2.058, H2 1-0 (S(3), H2 1-0 (S(2), H2 1-0 (S(1) | |
| 9721   | 2.1184 ± 0.0035 | 3.0 | Hα, [S II]6717,6731, [S II]9068,9533, He i 1.083, Pay, Paβ | |
| 9922   | 2.7412 ± 0.0004 | 3.0 | Hβ, [O III]4959, 5007, Hα, [S II]6717,6731, [S II]9068,9533, Paα, Pay, Paβ, He i 1.083 | |
| 10612  | 7.6597 ± 0.0014 | 3.0 | [O II]3727 through to [O II]5007 | The highest ionization parameter galaxy |

Figure 3. Left: The relative difference between estimated and true $T_e$ as a function of the amount of dust attenuation applied (see text for details). The solid lines show the results for the five true $T_e$ values shown when applying the standard [O III]4363/[O II]5007 ratio, while the dashed lines show the same when using the proposed double line ratio, [O III]4363/Hβ/[O II]5007/Hβ. Right: the difference in estimated $12 + \log O/H$ as a function of dust attenuation for the standard method (solid lines) and the proposed robust method (dashed lines).

When using the double line ratio method derived $T_e$, I calculated abundances by normalising [O II]3727 to Hδ, while for the standard approach I normalised all to Hβ. The panel shows clearly that the standard method is sensitive to flux calibration errors or errors in dust attenuation, while the proposed alternative method is more robust. It could of course be made even more robust by normalising [O II]3727 to H(9 → 2) for instance, but H(9 → 2) is rarely detected and often strongly affected by stellar absorption so it is not clear that it will offer an improvement in practice.

The double line ratio method involves four lines, so in principle the final uncertainty will increase. In reality this is however, not a significantly concern. Firstly, [O III]4363 is nearly always weaker than the other lines and will dominate the uncertainty on the final oxygen abundance, and if one use Balmer lines to correct for dust, all the four lines are involved thus in the end any effect on the uncertainty is very minor.
4.1 Application to the high-z galaxies

The oxygen abundances of the $z > 5$ galaxies in the sample have been explored extensively in the literature already. For 4590, 6355, and 10612 we have five estimates (Curti et al. 2023; Trump et al. 2023; Schaerer et al. 2022; Rhoads et al. 2023; Arellano-Córdova et al. 2022) while Trump et al. (2023) also provide an estimate for 5144.

I apply the methods described above to estimate $T_e([\text{O} \text{iii}])$. To estimate the temperature of the [O iii] emitting region we have to make recourse to functional relations. To illustrate the importance of this I have used three approximations: 1. set $T_e([\text{O} \text{iii}])$ equal to $T_e([\text{O} \text{ii}])$, 2. calculate $T_e([\text{O} \text{iii}])$ using the relation proposed by Izotov et al. (2006), or 3. calculate $T_e([\text{O} \text{ii}])$ using the relation derived by Pilyugin et al. (2009). I then calculate ionic abundances normalising [O iii]3727 to Hα, using Hγ to lead to changes in the abundances < 0.02 dex.

The resulting $T_e([\text{O} \text{iii}])$ and $12+\log\text{O/H}$ values are compared to other determinations in the literature in Figure 4. Focusing on the temperatures first, the agreement between different authors is reasonable with some outliers most notably for 4590 which has an apparently very strong [O iii]4363 line. For 8140 only Trump et al. (2023) present a temperature and abundance, I am unable to confirm their detection of [O iii]4363 and have decided to not use the direct method for this source. As we will see below, photoionization model fits prefer a rather higher abundance for this source in disagreement with Trump et al’s determination.

Turning now to the oxygen abundance plot in the right panel. This again shows decent agreement even for the extreme 4590. The effect of the $T_e([\text{O} \text{iii}])$ formula is shown as a solid and open smaller circle connected by a line to the larger, filled black circle. The larger circle corresponds to equal temperatures for the [O ii] and [O iii] zones, while the open corresponds to the Pilyugin et al relation and the small filled to the Izotov et al relation. These do not have a strong influence on the result as already noted by Curti et al. (2023) although the quadratic relation from Izotov et al. (2006) gives discrepant results for 4590, and as Rhoads et al. (2023) already noted this argues for a modification of that relation. The second point is that, like all the other publications, I have not corrected for the presence of triply ionized oxygen as this is not expected to be a significant contribution to the oxygen budget (Izotov et al. 2006; Arellano-Córdova et al. 2022; Berg et al. 2021), assuming of course that star formation is the dominant source of ionization, a point I will return to in Section 6 below.

5 FITTING THE HIGH-Z SOURCES WITH PHOTO-IONIZATION MODELS

In order to do a systematic study of metals in a sample of galaxies, it is usually better to exploit strong lines rather than temperature sensitive lines due to their better detectability. At low redshifts it has been common to use calibrated emission line ratios as the way to get this and this has led to numerous discussions about calibration.

However calibrated emission line ratios are a very simplistic way to estimate emission line properties and it is much better to use all the data and model emission line properties using fits to photoionization models. This was first used for the SDSS in Brinchmann et al. (2004b); Tremonti et al. (2004b). It has subsequently been used more widely (Blanc et al. 2015; Vale Asari et al. 2016; Fernández et al. 2022), and more recently also combined with full-spectrum modeling (Chevallard & Charlot 2016; Johnson et al. 2021) which also has very recently been applied to the three $z > 7$ galaxies by Tacchella et al. (2023). It is a natural match for the data presented here since this approach can explore a wider gamut of model parameters than a calibrated emission line ratio method would do (see Maiolino & Mannucci 2019; Kewley et al. 2019, for discussions).

For the study here I used both the original code, CL01fit, used for the SDSS studies mentioned above and last described in B13 which makes use of the Charlot & Longhetti (2001, CL01 hereafter) models, and a python re-implementation of this called PIModels which has support for multiple photoionization grids and can carry out the Bayesian analysis either using a stochastic or a gridded approach.

In either case, the codes fit a given set of lines, $\{\lambda_i\}$, using a Bayesian approach. We calculate the log likelihood of each model, $\mathcal{M}(U, \xi, \tau, Z)$, through:

$$\ln P(\mathcal{M}|\{\lambda_i\}) = -\frac{1}{2} \sum_{i \in \{\lambda_i\}} \left( \frac{f_i - Af_M}{\sigma_i^2} \right)^2 + \ln Pr,$$

where $f_i$ is the flux in line $i$, $A$ is a scaling-factor and $f_M$ corresponds to the relevant model. Pr denotes the prior on the model parameters. For most use this high-dimensional distribution is then marginalised down to 1D or 2D probability distribution functions (PDFs) — for the most part I will focus on 1D PDFs in this paper. I will present results based on both the CL01 and Gutkin et al. (2016, G16 hereafter) models. For the CL01 models I will use the CL01fit code, while for G16 I use PIfit.

CL01fit is described in detail in B13. It is a grid-based Bayesian code and evaluates equation (4) on a grid of model parameters. This is a very fast way to do Bayesian inference and it does not raise any issues of convergence or burn-in, but it provides results on a fixed grid so can not adapt to very high signal-to-noise data. PIfit can also adopt a gridded fit, but the default is to use the MultiNest (Feroz et al. 2009) package to do nested inference through the python interface (Buchner 2016). The code interpolates between the models using either a multi-dimensional linear or the Radial Basis Function interpolation from scipy (Virtanen et al. 2020) and also fits directly for dust attenuation if dust attenuation is not one of the model parameters (as it is for CL01). For the results here I use the multi-dimensional linear interpolator as it is more robust.

The different methodologies for CL01fit and PIfit do not significantly impact the results. However the photoionization model adopted does matter in this case. The CL01 model is based on a relatively old stellar populations code (Bruzual & Charlot 1993) as well as an earlier version of the Cloudy photoionization code (Ferland et al. 1998). The model grid only goes down to $12 + \log O/H = 7.5$ and up to an ionization parameter of $\log U = -2$, with incomplete sampling of the model grid beyond $\log U = -2.5$. These were not serious limitations for application to SDSS data, but as we will see, they are more problematic when applied to the high-z data. However, it is still useful to apply this to demonstrate what inferred parameters are more sensitive to the model choice.

In contrast, the G16 models are based on up-to-date stellar population models (an updated version of the Bruzual & Charlot 2003 models) as well as version 13.03 of Cloudy (Ferland et al. 2013). They span a wider range of metallicity down to $12 + \log O/H = 6.5$, as well as ionization parameters up to $\log U = -1$. This makes them much better suited to the present dataset and they are also used in the BEAGLE code (Chevallard & Charlot 2016).

In either case, I run the fits using [O iii]3727, Hγ, Hβ, and [O ii]5007 as input fluxes. I have also run the fits with [Ne ii]3869 line included and it does not change the results for G16 significantly, but for the CL01 the quality of fit is reduced so I opt to exclude the
line. I do not include the [O\,\textsc{iii}]4363 line because photoionization models have trouble predicting this line at a necessary accuracy (e.g. Dors et al. 2011). These are also the reasons why [Ne\,\textsc{iii}]3869 and [O\,\textsc{iii}]4363 were not used in the MPA-JHU metallicity fits. I also adopt flat priors on all parameters.

Before fitting, it is crucial to confirm, or not, that the model grid covers the space probed by the galaxies to fit. This is shown in Figure 5 where I compare a line ratio diagram for the G16 model against the objects. It is clear that most galaxies fall within the model grid although 10612 is consistently on the edge of the grid and 8140 has a somewhat discrepant [O\,\textsc{ii}]3727/H\gamma ratio albeit with a large uncertainty. The observed line ratios were not corrected for dust but the amount of dust allowed by the data is not sufficient to significantly affect these ratios since they are mostly between closely separated lines.

Figure 6 shows the result of running the G16 models on the data for 4590 and it evinces several aspects of photoionization model fitting that are worth keeping in mind: we see a clear correlation between dust attenuation and log $U$. In this case this is because the only lines that can constrain the ionization parameter are [O\,\textsc{ii}]3727 and [O\,\textsc{iii}]5007 and this means that both increasing the ionization parameter and increasing the dust attenuation can produce a weak [O\,\textsc{ii}]3727 line, leading to the displayed anti-correlation. This has a secondary effect on the oxygen abundance as well. There is also a correlation between 12 + log O/H and the dust-to-metal ratio, $\xi$. This effect is discussed in B13 and is due to a balance between reduced gas-phase oxygen when $\xi$ is increased and an increased heating, see B13 for details. With the provided lines, $\xi$ is pretty much unconstrained, but it is an important free parameter for photoionization fits as its value is basically unconstrained at high redshift. The plot also illustrates the effect of including (blue) or not (red) the [Ne\,\textsc{iii}]3869 line in the fit. In this case, the effect is very marginal and that is also the case for 6355, 8140 and 10612, however for 5144 the inclusion of [Ne\,\textsc{iii}]3869 helps rule out high-metallicity solutions.

Figure 7 shows the resulting constraints on 12 + log O/H from fitting the CL01 (gray filled histogram) and G16 (orange filled histogram) models to the high-z galaxies. In the case of 4590 the CL01 PDF goes right up against the low metallicity edge and the fit is overall rather poor. In contrast the other four galaxies are best fit with oxygen abundances well inside the model grid. The G16 model fits

**Figure 4.** A comparison of literature determinations of the gas-phase oxygen abundance of the five high-z galaxies.

**Figure 5.** A comparison of the line ratios for the high-z galaxies compared to the G16 model grid with $M_{\text{upper}} = 100 M_\odot$ and $n = 100 \text{ cm}^{-3}$. The grid lines correspond to the ionization parameter, log $U$ shown with blueish lines and log $Z$ in reddish tones.
6 THE SOURCE OF IONIZATION IN THE GALAXIES

Until now I have tacitly assumed that all galaxies in the sample have emission lines whose ionization source is dominated by star formation. That is however, a rather strong assumption so it is important to underpin this as much as possible.

6.1 The z > 5 galaxies

The source of ionization in the high-z galaxies has been extensively discussed in the literature already. It is, however, problematic to distinguish between star-bursts and AGN at low metallicities (Groves et al. (2006), see also Nakajima & Maiolino (2022)), thus it is perhaps not surprising that all conclude that the data are consistent with star formation.

Here I will instead focus on the UV lines detected in 4590 and 6355 to complement the literature studies using optical lines. Starting first with 4590, it shows C\textsc{iii}]1907,1909 in emission as also noted by Arellano-Córdova et al. (2022). This line is ubiquitous in AGN, but it is also commonly seen in star-forming galaxies (Stark et al. 2014; Rigby et al. 2015; Maseda et al. 2017; Ravindranath et al. 2020), in particular in those that show high intensity star bursts, so it is perhaps not surprising that it is seen here but as we will see, it does offer some extra information on the source.

Unfortunately C\textsc{iii}]1907,1909 is at the very blue of the spectral range and the renormalisation of flux that has been used here is not reliable at the edges. Thus I have also re-measured the flux of C\textsc{iii}]1907,1909 on the uncorrected spectrum. There are unfortunately no nearby lines to normalise to, thus I normalise to H\textbeta which is the stronger recombination line in the spectrum.

From the unmodified spectrum I find a C\textsc{iii}]1907,1909/H\textbeta ratio of 1.575 ± 1.21 and from the corrected spectrum I find 0.186 ± 0.05. Despite the substantial uncertainties, we can robustly conclude that the C\textsc{iii}]1907,1909 flux is comparable to, but somewhat lower than, that of H\textbeta. This is indicative of either a non-thermal ionization source or a moderate to high ionization parameter in a star-burst. I illustrate this in Figure 8. This compares the measured C\textsc{iii}]1907,1909/H\textbeta ratios to the Gutkin et al. (2016) models for star-forming galaxies on the left, and to the Feltre et al. (2015) models for AGNs on the right. The 2D histograms show the average ionization parameter, log U, in each bin. Overplotted on these diagrams are the two estimates of the C\textsc{iii}]1907,1909/H\textbeta ratio. They are clearly discrepant, implying significant systematic uncertainties, but both measurements are consistent with the models although no distinction can be made between the two sources of ionization. A reddening vector corresponding to τ_V = 1 is included in the left-hand panel. Clearly only a modest amount of dust can be accommodated before the data fall outside the model grid, in good agreement with e.g. Curti et al. (2023).

The equivalent width of C\textsc{iii}]1907,1909 is potentially a better discriminator of galaxy properties than these line ratios (e.g. Jaskot & Ravindranath 2016), however the flux calibration problem and non-detection of the continuum means that this is very challenging and the data have no real constraining power. Formally, using the total F150W flux to estimate the continuum, the equivalent width is 0.186 ± 0.05. I am not aware of a suitable comparison sample, but I note that Terao et al. (2022) find that C\textsc{iii}]1907,1909 is ~ 2 times stronger than [Ne\textiv]2422,2424 in their sample of radio galaxies, and at low metallicity C\textsc{iii}]1907,1909 can easily be comparable to H\textbeta, so this ratio seems entirely reasonable.

In conclusion it seems that 6355 is hosting a narrow-line AGN in...
Turning now to 8140, this actually shows a [Ne iv]3426 line in the co-added spectrum. This is however not visible in 008 and there is no He ii 4686, as well as a very low [N ii]6584/Hα ratio so this should also be considered a star-formation dominated source with the [Ne v]3426 detection considered spurious.

In the high-z sample we therefore appear to have 1 narrow-line AGN out of 5, which is an intriguingly high fraction but with the sample as small as it is, no firm conclusions should be drawn. It is however important to keep an eye out for these in future NIRSpec campaigns.

6.2 The $z < 5$ galaxies

Turning next to the lower redshift galaxies, the question of ionization source is even harder to answer. There are lines that are clearly diagnostic of AGN activity in the near-IR such as the He ii 1.083 line or high-ionization S or Si lines (see e.g. Riffel et al. 2006), however these are fairly weak lines and none are seen in the spectra of the current sample galaxies. However the near-IR also has [Fe ii]...
lines that have long been used as shock tracers (Oliva et al. 1989; Oliva & Moorwood 1990), as well as a rich spectrum of $H_2$ emission lines which can be used to gain insight in the warm neutral gas and photon-dominated regions in galaxies (e.g. Black & van Dishoeck 1987; Kaplan et al. 2017).

Starting with the more classical optical line ratios, 3042 has a $[N\,\text{ii}]6584/H\alpha$ line ratio $0.64 \pm 0.03$ above that normally seen in star forming regions, but in contrast the dereddened $[S\,\text{iii}]9533/H\alpha$ ratio $(0.11 \pm 0.02)$ matches very well the values found in metal rich H ii regions locally (e.g. Bresolin et al. 2004), where I assumed an intrinsic $[S\,\text{iii}]9533/[S\,\text{ii}]9069 = 2.58$ using atomic parameters from NIST. While no very firm conclusion can be made from this alone, the most likely conclusion appears to be that there is a contribution of non-star formation activity in this source to boost the $[N\,\text{ii}]6584$ flux.

Of the other galaxies, 9239 has $[S\,\text{ii}]6717,6731/H\alpha = 0.22$ which places it in the middle of the distribution of this ratio for star-forming galaxies (e.g. Kewley & Dopita 2002; Brinchmann et al. 2008a), and its $[S\,\text{iii}]9533/H\alpha$ ratio is also consistent with star formation.

Turning now to the more interesting $[Fe\,\text{ii}]$ lines, the only regularly detected one is $[Fe\,\text{ii}]1.257$ which we find in 3042, 9239, and 9483. This line is seen in nearby starbursts (e.g. Vanzi et al. 2008; Cresci et al. 2010; Izotov & Thuan 2016) but it is typically quite weak at a few percent of $H\beta$. The exception is in shock dominated regions or in the narrow-line regions of AGN (e.g. van der Werf et al. 1993; Alonso-Herrero et al. 1997; Mouri et al. 2000; Rosenberg et al. 2012) where it can reach much higher values.

Indeed, the $[Fe\,\text{ii}]$ lines seen in these spectra are all much too bright to be caused by star formation, and much brighter than normally seen in nearby starburst galaxies. To demonstrate this, Figure 10 compares the measured $[Fe\,\text{ii}]1.257/Pa\beta$ ratios in the three galaxies against the ratios measured in 24 nearby BCDs and H ii regions by Izotov & Thuan (2016). It is clear that with the exception of the H ii region J1038+5330 in NGC 3310 noted by Izotov & Thuan, all the low-z galaxies have much weaker $[Fe\,\text{ii}]$ emission.

We can contrast this to models more directly because $[Fe\,\text{ii}]1.257$ is fairly close to $Pa\beta$ and if I combine this with $He\,\text{i} 1.083/Pa\beta$ which again is a ratio of nearby lines, I get figure 11.

This figure shows the data for the three galaxies with detected $[Fe\,\text{ii}]1.257$ as the solid squares with error-bars. The other galaxies are shown at the $1\sigma$ upper limit for the $[Fe\,\text{ii}]1.257$ flux. I contrast this against three different model libraries. I have taken a single example grid from three libraries in the literature: the high-massloss starburst grid with $Z/Z_\odot = 0.2$ from Levesque et al. (2010), a solar metallicity $n = 10^4 \text{ cm}^{-3}$ dusty AGN model grid from Groves et al. (2004a,b), and a solar metallicity $n = 0.01 \text{ cm}^{-3}$ shock model grid from Allen et al. (2008) where I used the shock-precursor grid. All models were obtained through the {	t itera} code by Groves & Allen (2010).

The immediate observation we can make is that the three SMACS galaxies with clearly detected $[Fe\,\text{ii}]1.257$ all fall clearly in the shock dominated region. They lie towards the high $B$-field, low velocity region of the shown grid. In reality these galaxies are likely to have a...
Table 2. Line luminosities, not corrected for lensing, for the z < 5 galaxies in the sample. For each galaxy three numbers are given: the luminosity in solar luminosities from the normalisation to the fixed aperture flux, the luminosity from the spectrum normalised to the total magnitude and the signal-to-noise of the line.

| Line          | 1917 L<sub>aper</sub> L<sub>total</sub> S/N | 3042 L<sub>aper</sub> L<sub>total</sub> S/N | 5735 L<sub>aper</sub> L<sub>total</sub> S/N | 8506 L<sub>aper</sub> L<sub>total</sub> S/N |
|---------------|------------------------------------------|------------------------------------------|------------------------------------------|------------------------------------------|
| Hα            | 30.1 772.4 88.47                         | 907.7 920.4 210.67                      | 127.6 1249.2 384.05                     |
| [NII]6584     | 13.0 332.1 44.66                         | 35.4 35.9 10.53                         | 4.2 41.0 16.16                         |
| [SII]6717     | 2.8 65.4 7.26                           | 83.1 67.5 14.31                         | 4.1 40.3 11.82                         |
| [SII]6731     | 3.7 82.5 9.07                           | 61.0 46.6 11.16                         | 3.9 37.6 11.22                         |
| [SII]9068     | 96.8 122.6 23.90                         | 37.7 38.2 10.96                         | 7.4 71.8 11.27                         |
| [SII]9533     | 7.1 118.8 10.18                         | 215.5 273.0 50.58                      | 12.6 153.9 22.98                       |
| Paδ           | 1.5 25.5 2.34                           | 5.2 13.5 1.74                          | 1.4 0.7 3.10                           |
| HeI 1083      | 4.0 64.1 5.99                           | 107.4 132.2 24.31                      | 45.4 46.0 5.17                         |
| Paγ           | 2.2 39.6 4.98                           | 49.8 64.9 12.20                        | 31.2 31.6 1.49                         |
| [FeII]1.257   | 2.2 37.1 3.49                           | 68.3 83.0 14.34                        | 3.8 34.2 5.41                          |
| Paβ           | 5.6 86.0 9.48                           | 134.5 171.0 28.81                      | 20.9 21.2 3.33                         |
| Paα           | 377.1 478.0 43.17                       | 28.0 39.5 3.04                         | 28.8 35.7 3.34                         |
| Brδ           | 28.8 35.7 3.34                         | 16.7 18.6 1.69                         | 22.2 28.1 1.74                         |
| H2 1-0 S(1)   | 22.1 26.5 2.82                         | 21.2 21.2 3.33                         | 22.1 26.5 2.82                         |
| H2 1-0 S(2)   | 2.058 9533.0 225.1 25.51                 | 22.1 26.5 2.82                         | 22.1 26.5 2.82                         |
| mix of shocks, star-formation and AGN contributing to their emission lines and a more detailed modeling would be required to disentangle these and the present data are not of sufficient quality to warrant this. Furthermore, for some of these galaxies the aperture corrections are substantial, this is shown in Table 2 which provides line luminosities on spectra normalised to aperture fluxes and total fluxes. That said, however, it will be of considerable interest to understand these “shock-dominated” galaxies in more detail and to understand how they fit into the overall population of galaxies at these redshifts (see also Reddy et al. 2023). It would also be very desirable to have additional diagnostic ratios to further understand the nature of these sources, but only 3042 has some qualitatively different lines, in [NII]6583, [SII]9533, and Hα and those are not strongly diagnostic of shock activity (Allen et al. 2008). Instead Allen et al. (2008) recommended UV and UV-optical diagnostics such as C II]1909/C II]2326 and [Ne III]3869/[Ne v]3426, neither of which are available for our galaxies but which conceivably could be obtained for z ~ 1 galaxies with deep ground-based observations.

The final set of lines of key interest are the H2 lines. These originate in the cooler medium outside the H ii regions and as such provide very complementary information to lines coming from hotter regions of the ISM. In the present sample only 9483 shows strong H2 lines and even then only three: H2 1-0 S(3), H2 1-0 S(2) and H2 1-0 S(1). H2 1-0 S(3) is contaminated by He i 1.955, leaving only the H2 1-0 S(2)/H2 1-0 S(1) ratio as a diagnostic. This ratio has a value of 0.68 ± 0.47 which is comparable to the values found by Izotov & Thuan (2016) for nearby star-forming galaxies and regions (0.3–0.8) and the value of 0.28 found for the Orion Bar by Kaplan et al. (2017). Since we saw above that 9483 has a significant contribution of shocks to its [FeII]1.257 line flux, one might expect that the H2 line ratios should approach that corresponding to a thermal distribution. Unfortunately the detected lines are not very diagnostic for
this as the models for thermal and fluorescent emission in Black & van Dishoeck (1987) span this range, with a preference for a lower value for thermal distributions and somewhat higher for fluorescent emission but in either case consistent with the observed line ratio. Future observations will surely provide much more information on this and open up the diagnostic potential of H₂ lines for galaxies out to z ~ 1.3 with NIRSpec.

7 COMPARISON TO LOCAL GALAXIES

As discussed in the introduction, analogues of high-z galaxies in the local Universe have been sought after for a long time, and naturally the first papers looking at these NIRSpec data have discussed extensively their properties in the context of nearby galaxies. Thus Schaerer et al. (2022) compared the sample of extreme line-emitters from Izotov et al. (2014) and galaxies from the Low-Z Lyman Continuum Survey (LzLCS, Flury et al. 2022) to the NIRSpec sample, finding considerable correspondences. Rhoads et al. (2023) compared the NIRSpec sample to nearby Green Pea galaxies (e.g. Cardamone et al. 2009; Jaskot & Oey 2013), specifically from the sample defined in Jiang et al. (2019) and Yang et al. (2019). A mixture of those approaches was taken by Trump et al. (2023) who compared both to Green Peas from Brunker et al. (2020) but also to extreme line emitters from Pérez-Montero et al. (2021). A similar approach was taken by Katz et al. (2023) who compared against compact galaxies of various colour (Yang et al. 2017a,b), as well as low-metallicity galaxies (Izotov et al. 2019) and extreme line-emitters (Amorín et al. 2015). With small variations, all these authors and comparisons find that there are galaxies in the nearby Universe that are broadly similar to the NIRSpec sample, although exactly how similar is open to some discussion.

What is noticeable is that these studies all limit themselves to star forming galaxies, thus there is an a priori assumption that the ionization source in these galaxies must be star formation. This might very well be correct and it is not an unreasonable assumption, but as we saw above, 6355 appears to have an AGN contribution and it is notoriously difficult to distinguish between AGN and star formation as ionization source at low metallicity. Thus here I take a slightly different approach. I will not start with a particular sample but rather I will ask: what galaxies in the local Universe have similar emission line properties to the NIRSpec sample and what are the properties of this local sample?

Since I am interested in excitation properties, I will focus on line ratios and ignore extrinsic quantities like size or mass. Specifically I always require the [O iii]4363/Hγ ratio to match the NIRSpec sample within 1σ, and then define two samples: the [O iii]4363/Hγ counterparts and the [Ne iii]3869/[O iii]3727 counterparts. I will base myself on the SDSS DR7 sample used in B13 and only use galaxies with S/N > 7 in all relevant lines. With these preambles we find the following results:

- **4590**: there are 13 galaxies with [O iii]4363/Hγ within 1σ. 6 of these are AGN, while 4 are star-forming, 2 composite and one unclassified. The [Ne iii]3869/[O iii]3727 ratio gives a similar result with 2/3 AGNs and 1/3 star-forming.
- **5144**: there are 192 galaxies with [O iii]4363/Hγ within 1σ of 5144. 73% of these are star-forming and all but 2 of the remainder are classified as AGN. The [Ne iii]3869/[O iii]3727 ratio is matched by no galaxies in the SDSS within 1σ.
- **6355**: there are 155 galaxies that are close to the [O iii]4363/Hγ ratio. Of these 76% are star-forming while the remainder are almost all AGN. When considering [Ne iii]3869/[O iii]3727 only 40 galaxies are within 1σ of the NIRSpec data. Out of these 55% are AGN, 2 unclassified, and the rest star-forming.
  - **8140**: as I do not consider this to have [O iii]4363 I do not consider this ratio, but for the [Ne iii]3869/[O iii]3727 ratio a total of 371 galaxies in the SDSS fall within 1σ. These are composed of 15% star-forming and 82% AGN.
  - **10612**: there are 172 SDSS galaxies with [O iii]4363/Hγ within 1σ of the NIRSpec data. Out of these only 3 are classified as star-forming, with the remainder all falling in the AGN part of the BPT diagram. There are only two galaxies that have [Ne iii]3869/[O iii]3727 within 1σ of the NIRSpec data, one AGN, one star-forming galaxy.

These results are summarised in Figure 12 where filled disks show the [O iii]4363/Hγ analogues while the open symbols the [Ne iii]3869/[O iii]3727 ones. It is evident that based on the observed line ratios the local counter-parts will fall in two regions: they are either low-Z star-forming galaxies (or low-Z AGNs we should not forget), or clear AGN. Taken at face value this suggests that we should at least take the assumption that all these are powered by star formation with a grain of salt, although one should not over-interpret the results of the local sample either. In particular the 10612 which is the most extreme galaxy in the sample, must be questioned as to its ionization source as also pointed out by Schaerer et al. (2022).

For the [O iii]4363/Hγ analogues I have also co-added the PDFs of all their derived parameters from the fitted models described in Brinchmann et al. (2013). In the following section (see also Figure 7), I will compare the resulting distributions to the high-z galaxies. This is only possible for star-forming galaxies, thus when reading the following section it is important to keep in mind that there is a possibility that several of these sources have a contribution of an AGN to their line fluxes which will make some of the following results less conclusive.

8 GAS-PHASE PROPERTIES OF THE GALAXIES

Returning now to the CL01 and G16 photoionization model fits to the data, Figure 13 shows multiple fit parameters for the five high-z galaxies. The first row shows the ionization parameter for each galaxy and here we immediately see one of the limitations of the CL01 grid: it does not go to sufficiently high ionization parameters and the PDFs are pushed up towards the edge. For those galaxies for which this is noticeable, 5144 and 10612 in particular, this puts the
Table 3. Results from the CL01 and G16 photoionization fitting. The central value is the median of the PDF and the uncertainties quoted are 16% and 84% of the PDFs for the indicated quantity and do not include systematic uncertainties. The depletion time assumes an effective area probed by emission lines of 1 kpc².

| Quantity | 4590 | 5144 | 6355 | 8140 | 10612 |
|----------|------|------|------|------|-------|
| 12 + log O/H (CL01) | 7.60\pm0.04 | 7.89\pm0.26 | 8.01\pm0.11 | 8.44\pm0.10 | 7.85\pm0.09 |
| 12 + log O/H (G16) | 7.35\pm0.05 | 7.63\pm0.26 | 8.28\pm0.37 | 8.44\pm0.09 | 7.91\pm0.11 |
| Log U (CL01) | -2.03\pm0.02 | -2.65\pm0.18 | -2.27\pm0.06 | -2.74\pm0.11 | -2.03\pm0.02 |
| Log U (G16) | -1.69\pm0.21 | -2.35\pm0.53 | -1.97\pm0.24 | -2.28\pm0.08 | -1.08\pm0.05 |
| \(\Sigma_{\text{gas}} / M_{\text{pc}}^{-2} \) (CL01) | 2.20\pm0.04 | 1.93\pm0.40 | 1.97\pm0.38 | 0.59\pm0.16 | 2.25\pm0.36 |
| \(\Sigma_{\text{gas}} / M_{\text{pc}}^{-2} \) (G16) | 2.86\pm0.31 | 2.02\pm0.61 | 1.58\pm0.37 | 0.36\pm0.34 | 2.19\pm0.31 |
| Log SFR (CL01) | 0.88\pm0.02 | 0.57\pm0.13 | 1.34\pm0.16 | 0.21\pm0.05 | 0.99\pm0.04 |
| Log SFR (G16) | 0.91\pm0.11 | 0.50\pm0.65 | 1.29\pm0.39 | 0.17\pm0.16 | 1.23\pm0.26 |
| \(\Sigma_{\text{gas}} / \Sigma_{\text{SFR}} \) (CL01) | 6.10\pm0.36 | 6.89\pm0.38 | 6.73\pm0.18 | 6.54\pm0.07 | 7.02\pm0.39 |

other results into question although we saw in Figure 6 that log U is relatively independent of other parameters in the fit. I do not show the analogue sample in this row as the CL01 model is such a poor match in this respect.

The second row shows the star formation rate, derived from the scaling parameter in equation (4). The star formation rates inferred from the emission lines are corrected for lensing using the same values as Trussler et al. (2022) which are taken from Pascale et al. (2022). The dashed blue lines show the Star Formation Rate (SFR) estimates from the photometry presented in Trussler et al. (2022) (who did not provide a value for 8140) and the dashed burgundy lines those presented by Tacchella et al. (2023) adjusted to use the same magnification values as Trussler et al. (2022). They all agree to within a factor of 3, but often better — indeed the agreement with Tacchella et al. (2023) is considerably better than this despite very different methodologies. However the spectroscopic estimates should be considered to have a substantial systematic uncertainty due to the uncertain flux calibration. The SFR distributions for the local analogues show a very large spread which is not surprising as the matching was done ignoring all extrinsic quantities.

The third row shows an estimate of the surface gas mass density of the galaxies following the methodology of B13 (see that reference for details). The results are interesting in that they prefer high surface mass densities for all the high SFR sources. These gas estimates are in reasonable agreement between CL01 and G16 when the CL01 models are adequate fits, and it is also clear that the local analogues have similar gas properties to the z > 5 galaxies. If we assume that the emission lines come from a region with area ~ 1 kpc² which is not entirely out of question given the small sizes, we find depletion times of \(t_{\text{depl}} = \Sigma_{\text{gas}} / \Sigma_{\text{SFR}} \approx 10^7 \) yrs (see Table 3). This is short but very similar to the values found for starbursting galaxies in B13, and it is in good agreement with the conclusion of Tacchella et al. (2023) that these appear to be rapidly accreting galaxies.

The final row shows electron temperature estimates from the models, compared to those obtained from the direct method. The \(T_e\) values from the CL01 fits are averages over the Strömgren sphere which in general leads to lower temperatures than from \([\text{O} \, \text{ii}]\lambda4363/\text{O} \, \text{iii}]\lambda5007\ but for both 6355 and 10612, the CL01 fits give temperatures in good agreement with the direct method. While this could be interpreted as evidence for very hot ionized gas where the \([\text{O} \, \text{ii}]\lambda4363/\text{O} \, \text{iii}]\lambda5007\ provides a good estimate of the mean temperature, the fact that the CL01 model is a poor fit for both of these argues against any strong conclusion but it does highlight the usefulness of using photoionization models to estimate electron temperatures as well. In all but 4590 the CL01 fit to the low-z analogues also results in temperatures similar to the high-z galaxies.

These quantities, including the depletion time are all given in Table 3 for the five z > 5 galaxies with sufficient lines. 4580 does not have sufficient lines for a photoionization model fit to be useful. It is also important to keep in mind that the errors given in the table do not include systematic uncertainties which given the calibration uncertainties could be substantial.

9 DISCUSSION

Despite the lingering calibration issues, these data are demonstrating how ground-breaking NIRSpec will be for the study of ionized gas in galaxies.

The z < 3 galaxies are showing a surprising amount of shock excited gas signaled by an elevated \([\text{Fe} \, \text{ii}]\lambda257 emission. In the local Universe this is not seen in low-mass star-bursts (Izotov & Thuan 2016) and regions where \([\text{Fe} \, \text{ii}]\ is significantly enhanced tends to be associated with shocked regions or AGNs. The shocks in this case is usually assumed to be associated with supernova remnants (SNRs, Greenhouse et al. 1991; Alonso-Herrero et al. 2003; Rosenberg et al. 2012). Bruursema et al. (2014) used this to search for SNRs in NGC 6946 and found a typical luminosity per SNR of ~ 10³⁶ erg s⁻¹. It should be noted that there is a spread of almost an order of magnitude in this luminosity and other studies find different mean values, but if we adopt the Bruursema et al. value, we find that the \([\text{Fe} \, \text{ii}]\ λ257 luminosity of our three sources corresponds to 2–3×10³⁵ \([\text{Fe} \, \text{ii}]\-luminous SNR. Thus, if the \([\text{Fe} \, \text{ii}]\ λ257 all comes from SNR, we can estimate a supernova rate ~ 0.1 yr⁻¹, assuming a typical life-time for the bright radiative shocks of ~ 10⁴ yr (Vink 2020). This is not an unreasonable rate of star formation, thus this presents a plausible scenario for the brightness of the \([\text{Fe} \, \text{ii}]\ lines. However it does not address the question of why other galaxies (1917, 5735, 8506, 9721, 9922) show no sign of \([\text{Fe} \, \text{ii}]\ λ257 despite also having bright Paschen lines. Indeed in figure 11 the top x-axis shows the \([\text{Fe} \, \text{ii}]\ λ257 luminosity per SFR, and as can be seen, this varies strongly between the galaxies.

Turning now to the high redshift galaxies, several of the papers on these galaxies have commented on the high value of \([\text{O} \, \text{ii}]\lambda4363/Hß or \([\text{O} \, \text{ii}]\lambda4363/\text{O} \, \text{iii}]\lambda5007 in 4590. I have avoided this until now because I do not feel there is much reason to worry too much about this. Firstly, the \([\text{O} \, \text{ii}]\λ4363 line is weak, only detected at 3–4σ depending on the noise estimate adopted. Secondly, the other lines in the galaxy do not show particularly extreme properties thus there is no clear reason to
thought that this is anything but a statistical fluke. Indeed, the fits to the G16 model above predict a [O\textsc{iii}]4363/H\beta flux lower than that observed which would make the galaxy more normal. In contrast, 6355 appears to show [Ne\textsc{iii}]2422,2424 in emission and must be a low-metallicity AGN, while 10612 has an ionization parameter \log U > \sim -1 which is normally the regime occupied by AGNs. One might object that He \textit{ii} 4686 is not seen, but the expected flux given the strong-line fluxes is below $10^{-21}$ ergs/cm$^{-2}$ (the G16 model predicts $2 \times 10^{-22}$ ergs/cm$^{-2}$ for an extreme star-forming model) which is below the detection threshold of the data. This then argues for potentially at least two out of the five high-$z$ galaxies being affected by AGN activity.

It is also clear that the low-$z$ analogues are showing intrinsic properties very well matched to the high-$z$ galaxies and this appears to justify the long-standing quest to carefully characterise these to contrast to high-$z$ galaxies. It also allows us to examine a potential bias that might affect all NIRSpec studies of high-$z$ galaxies.

9.1 Selection of galaxies and biases in the mass-metallicity relation

Several papers (e.g. Schaerer et al. 2022; Curti et al. 2023) have already tried to assess the evolution of the mass-metallicity (MZ) relation with redshift by comparing the JWST SMACS results to $z \sim 0$ (Tremonti et al. 2004b; Yates et al. 2012; Andrews & Martini 2013; Telford et al. 2016; Curti et al. 2020) and $z \sim 2$ results (e.g. Sanders et al. 2020). With only three galaxies, selected in a somewhat haphazard manner and still substantial uncertainties in the stellar mass estimates (e.g. Schaerer et al. 2022; Tacchella et al. 2023; Carnall et al. 2023) this is of course preliminary. That aside, these studies do demonstrate the potential of future, larger, JWST studies to make real progress here.

There are many potential biases when constructing the MZ relation and this has been discussed in the literature (see Telford et al. 2016; Cresci et al. 2019), but one important potential bias for the current scenario is the requirement to have a detected [O\textsc{iii}]4363. It is possible to make recourse to the local sample again to understand how this comes about. Note that the bias I discuss here also applies to the mass-metallicity-SFR (Manucci et al. 2010; Lara-López et al. 2010).

Since [O\textsc{iii}]4363 is exponentially sensitive to the temperature of the ionized gas, requiring the detection of this line means that you will always get a sample of the lowest metallicity galaxies at each mass. This bias was already commented on indirectly by Sanders et al. (2020) who focused their discussion on the construction of calibration relations and the SFR distributions of the samples used to construct these. The bias is however more directly affecting the MZ relationship as it becomes more pronounced as one moves to higher masses (higher mean metallicities).

To illustrate this, the left panel of Figure 14 shows the MZ plane for the SDSS DR7 and the effect of selection on [O\textsc{iii}]4363 detectability. The black line shows the median trend for all SF galaxies from the SDSS DR7 sample of B13 and the dashed lines the 68% spread around this with the grey shading showing the 68% uncertainty on the median as determined from bootstrap resampling including Monte Carlo sampling of the uncertainties on log $M_*$ and $12 + \log$ O/H. To reduce aperture correction effects, I have limited my sample to have $z > 0.01$ and a half-light diameter no more than 3 times the fibre

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{The results of fitting CL01 models to the five high-redshift galaxies. Each row shows the probability distribution function (PDF) of a single parameter. The top row shows the ionization parameter, the second the logarithm of the star formation rate, the third row the surface mass density of gas, and the final row the estimated electron temperature of the gas. The PDFs have not been closed at the parameter grid edge which is clearly seen for log $U$ for 5144, 6355, and in particular 10612. The fit results for these three galaxies are therefore uncertain. The G16 model is shown in orange in the panels where predictions could be obtained. In the SFR row I compare the results to the photometrically estimated SFR from Trussler et al. (2022), and the spectrum+photometry determination from Tacchella et al. (2023), and in the $T_e$ row I show the $T_e$ ([O\textsc{iii}]) values from the direct method as purple vertical lines. The green lines show the distributions for the [O\textsc{iii}]4363/H\gamma analogues defined in section 7.}
\end{figure}
Figure 14. Left: The mass-metallicity relation for star-forming galaxies in the SDSS DR7, compared to different selection methods and the three highest redshift galaxies from the current JWST sample. The solid black line shows the median trend for all galaxies with the dashed lines showing the 68% spread around this. Blue solid line: the median trend for star-forming galaxies with [O iii]4363 detected with a S/N > 7. The filled symbols show the locations of the high-z galaxies (square: 4590, diamond: 6355, triangle: 10612) using masses from Tacchella et al. (2023) and CL01-based metallicities for a direct comparison to SDSS. The corresponding open symbols shows the [O iii]4363/Hy analogues for each high-z galaxy. Right panel: the difference in the median 12 + log O/H between all galaxies and those with S/N > 7 in [O iii]4363 as a function of mass and star formation rate.

One might rightly object that the SFR distribution of the two samples are different, so in the right panel I show the difference between the main sample and the sample requiring a S/N > 7 in [O iii]4363 as a function of mass divided in three bins of log SFR as measured within the fibre, as indicated in the legend. The errorbars correspond to uncertainties on the median, the 68% spread is much larger. It is clear that there is a systematic trend with mass and it is even larger than when not controlling for SFR. The differences between the different SFR bins are small, but systematic: at low SFR the sample of all versus that with [O iii]4363 are nearly coincident so very little difference is seen. At somewhat higher SFRs we have the largest difference in sample and at very high SFR and low M*, we again have very large overlap between the samples. One should, however, be cautious with overinterpreting this figure, however, as we again have very large overlap between the samples. One should, so very little difference is seen. At somewhat higher SFRs we have the sample of all versus that with [O iii]4363 are nearly coincident

One should, however, be cautious with overinterpreting this figure, however, as we again have very large overlap between the samples. One should, so very little difference is seen. At somewhat higher SFRs we have the sample of all versus that with [O iii]4363 are nearly coincident

One might rightly object that the SFR distribution of the two samples are different, so in the right panel I show the difference between the main sample and the sample requiring a S/N > 7 in [O iii]4363 as a function of mass divided in three bins of log SFR as measured within the fibre, as indicated in the legend. The errorbars correspond to uncertainties on the median, the 68% spread is much larger. It is clear that there is a systematic trend with mass and it is even larger than when not controlling for SFR. The differences between the different SFR bins are small, but systematic: at low SFR the sample of all versus that with [O iii]4363 are nearly coincident so very little difference is seen. At somewhat higher SFRs we have the largest difference in sample and at very high SFR and low M*, we again have very large overlap between the samples. One should, however, be cautious with overinterpreting this figure, however, as the SDSS is not complete below log M*/M⊙ ~ 9 and the sample of galaxies with [O iii]4363 does include galaxies that were not targetted as such. However the basic lesson appears robust: constructing MZ relations based on direct method metallicities requires very careful assessment of selection biases when comparing samples.

To put this in the context of the high-z galaxies, I have also inserted these (filled symbols) with their [O iii]4363/Hy star-forming analogues (open symbols) in the figure. The squares show 4590 and its analogues, the diamonds 6355 and its analogues and the triangles 10612 and its sole star-forming analogues. For this plot I have used stellar masses from Tacchella et al. (2023) and the CL01 model to estimate metallicities in order to compare directly to SDSS DR7, however a very similar picture is found using the G16 model.

Does this mean that this bias will also apply at high redshift? This is one of the questions that upcoming JWST surveys with NIRSpec will be able to tell us — it depends on the spread in metallicity at a given mass. If this is very small at high redshift, then a sample defined to have detected [O iii]4363 will provide a good measure of the mean metallicity, but if there is substantial scatter then direct method metallicities might result in a significantly biased MZ relation, including most likely an incorrect slope.

Since we saw above that photoionization modelling gives oxygen abundances in good agreement with the direct method, as long as the models span a sufficiently large range in physical parameters, a viable way to combat these potential biases is to use photoionization models to estimate oxygen abundances for all galaxies instead — using [O iii]4363 to validate the results on a case-by-case basis.

10 CONCLUSIONS

In the preceding I have presented three new redshifts, all at z < 3, bringing the total number of secure redshifts for these NIRSpec observations in the SMACS 0723 field to 14. I highlighted the usefulness of modifying the methodology for direct temperature estimates to use the double line ratio [O iii]4363/Hβ/[O iii]5007/Hβ. Doing so reduces sensitivity to flux calibration or dust attenuation uncertainties by a factor of ~ 3. If in addition the ionic abundance determinations are referenced to nearby hydrogen lines, one can strongly reduce effect of flux calibration uncertainties or unknown reddening on oxygen abundance determinations. Given the unavoidable wavelength-dependent slit losses with NIRSpec this modification of the methodology is likely to be beneficial for future high-z abundance studies.

I also find that 6355 shows a fairly clear [Ne iv]2422,2424 line which I argue is evidence for this being a narrow-line AGN. The existence of one such high-z, low-mass AGN is interesting but larger samples are necessary before any implications for the process of reionization become clear. However it is notable that 10612 is showing line ratios that at low-z are commonly seen in AGN, which
coupled to an ionization parameter that reaches the model boundary at \( U = -1 \), suggests that this galaxy also might harbour an AGN, thus one must interpret these results with some caution.

The lower redshift galaxies also seem to be rather active with three galaxies showing very strong [Fe II]1.257/P~ relative to what star-forming galaxies at \( z \sim 0 \) is showing. These galaxies have emission spectra that can be explained nearly fully by shock models and one possible explanation is that they have a particularly high number of supernova remnants. Further data are needed to fully understand why some galaxies show very strong [Fe II] lines and others not, while having comparable Paschen lines, but clearly aperture corrections will play an important role here.

I also presented modelling of the high-z galaxies using photoionization models and show that these models result in inferred oxygen abundances in good agreement with those found using the direct having comparable Paschen lines, but clearly aperture corrections will play an important role here.

Finally, I showed that selecting samples based on the presence of [O II]4363 has the potential to significantly bias determinations of the mass-metallicity relation and that a careful assessment of selection effects must be made in this case and I argued that it might be better to base oneself on Bayesian photoionization modeling of the strong lines than to use temperature sensitive lines as the main oxygen abundance estimation technique when doing statistical studies such as of the mass-metallicity relation.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the JWST commissioning team for providing these data early and giving us all a taster of what JWST will bring in the coming years. I thank Michael Maseda, Themiya Nanayakkara, Madusha Gunawardhana, Leindert Boogaard, Roberto Brinchmann, J. Kunth, P. Durret, F. Charlot, S. Tremonti, C. Kauffmann, G. Heckman, T. Brinkmann, P. Pettini, D. Erb, K. Juneau, E. Skillman, R. Pogge, R. Olivier, M. Auger, M. W. Brammer, G. B. Brammer, G. B. and J. H. Black, E. F. van Dishoeck, as well as datasets provided in the literature. The analysis presented here is based on the publicly available L3 data products provided in the Barbara A. Mikulski Archive for Space Telescopes (MAST), as well as datasets provided in the literature.

DATA AVAILABILITY

The analysis presented here is based on the publicly available L3 data products provided in the Barbara A. Mikulski Archive for Space Telescopes (MAST), as well as datasets provided in the literature.

Tables with line fluxes are available from https://github.com/jbrinchmann/JWST_SMACS. The PIIFit code is currently in pre-beta and is available upon reasonable request but will be publicly released in the near future. The CL01Fit and platefit IDL codes are available upon reasonable request, but a Python version of platefit will be released with Bacon et al (submitted).

REFERENCES

Allen M. G., Groves B. A., Dopita M. A., Sutherland R. S., Kewley L. J., 2008, The Astrophysical Journal Supplement Series, 178, 20
Aller L. H., 1956, Gaseous Nebulae. Chapman & Hall, London
Aller L. H., Hung S., Feibelman W. A., 1999, Proceedings of the National Academy of Science, 96, 5366
Alonso-Herrero A., Rieke M. J., Rieke G. H., Ruiz M., 1997, The Astrophysical Journal, 482, 747
Alonso-Herrero A., Rieke G. H., Rieke M. J., Kelly D. M., 2003, The Astronomical Journal, 125, 1210
Amorín R., et al., 2015, Astronomy & Astrophysics, Volume 578, A105, 578, A105
Andrews B. H., Martini P., 2013, The Astrophysical Journal, 765, 140
Arellano-Córdoa K. Z., et al., 2022, The Astrophysical Journal, 940, L23
Astropy Collaboration et al., 2013, A&A, 558, A33
Astropy Collaboration et al., 2018, AJ, 156, 123
Bacon R., et al., 2015, Astronomy and Astrophysics, 575, 75
Bacon R., et al., 2021, Astronomy & Astrophysics, 647
Beichman C. A., Rieke M., Eisenstein D., Greene T. P., Krist J., McCarthy D., Meyer M., Stansberry J., 2012, in Science opportunities with the near-[IR] camera ([NIRCam]) on the [James] [Webb] [Space] [Telescope] ([JWST]), p. 84422N, doi:10.1117/12.925447, https://ui.adsabs.harvard.edu/abs/2012SPIE.8442E..2NB
Berg D. A., Erb D. K., Auger M. G., Pettini M., Brammer G. B., 2018, The Astrophysical Journal, 859, 164
Berg D. A., Chinsholm J., Erb D. K., Skillman E. D., Pogge R. W., Olivier G. M., 2021, The Astrophysical Journal, 922, 170
Berg D. A., et al., 2022, The Astrophysical Journal Supplement Series, 261, 31
Bertin E., Arnouts S., 1996, Astronomy and Astrophysics Supplement Series, 117, 393
Bezanson J., Edelman A., Karpinski S., Shah V. B., 2017, SIAM Review, 59, 65
Bian F., Kewley L. J., Dopita M. A., Juneau S., 2016, The Astrophysical Journal, 822, 62
Black J. H., van Dishoeck E. F., 1987, The Astrophysical Journal, 322, 412
Blanc G. A., Kewley L., Vogt F. P. A., Dopita M. A., 2015, The Astrophysical Journal, 798, 99
Boncaud A., Boccio M., Abergel A., Orieux F., Dole H., Hadj-Yousef M. A., 2016, Astronomy & Astrophysics, 596, A65
Brammer G. B., van Dokkum P. G., Coppi P., 2008, The Astrophysical Journal, 686
Breda I., et al., 2022, Astronomy & Astrophysics, 663, A29
Bresolin F., Garnett D. R., Kennicutt R. C., 2004, The Astrophysical Journal, 615, 228
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004b, Monthly Notices of the Royal Astronomical Society, 351, 1151
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004a, MNRAS, 351, 1151
Brinchmann J., Pettini M., Charlot S., 2008a, MNRAS, 385, 769
Brinchmann J., Kunth D., Durret F., 2008b, A&A, 485, 657
Brinchmann J., Charlot S., Kauffmann G., Heckman T., White S. D. M.,
