



# Changing physical conditions in star-forming galaxies between redshifts $0 < z < 4$ : [O III]/H $\beta$ evolution

F. Cullen,<sup>1\*</sup> M. Cirasuolo,<sup>1,2</sup> L. J. Kewley,<sup>3</sup> R. J. McLure,<sup>1</sup> J. S. Dunlop<sup>1</sup>  
and R. A. A. Bowler<sup>1</sup>

<sup>1</sup>*SUPA<sup>†</sup>, Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*

<sup>2</sup>*European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching bei Muenchen, Germany*

<sup>3</sup>*Australian National University, Research School for Astronomy and Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston, ACT 2611, Australia*

Accepted 2016 May 13. Received 2016 May 13; in original form 2015 August 24

## ABSTRACT

We investigate the redshift evolution of the [O III]/H  $\beta$  nebular emission line ratio for a sample of galaxies spanning the redshift range  $0 < z < 4$ . We compare the observed evolution to a set of theoretical models which account for the independent evolution of chemical abundance, ionization parameter and interstellar medium (ISM) pressure in star-forming galaxies with redshift. Accounting for selection effects in the combined data sets, we show that the evolution to higher [O III]/H  $\beta$  ratios with redshift is a real physical effect which is best accounted for by a model in which the ionization parameter is elevated from the average values typical of local star-forming galaxies, with a possible simultaneous increase in the ISM pressure. We rule out the possibility that the observed [O III]/H  $\beta$  evolution is purely due to metallicity evolution. We discuss the implications of these results for using local empirical metallicity calibrations to measure metallicities at high redshift, and briefly discuss possible theoretical implications of our results.

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

## 1 INTRODUCTION

In the local Universe, thanks to the large statistical galaxy sample provided by the Sloan Digital Sky Survey (SDSS), key physical properties of the star-forming galaxy population have been investigated for  $\sim 10^5$  galaxies. For example, the position of the star-forming ‘abundance sequence’ of local galaxies in the [O III]/H  $\beta$  versus [N II]/H  $\alpha$  diagram (Baldwin, Phillips & Terlevich 1981, hereafter **BPT diagram**) has become well established (e.g. Kauffmann et al. 2003; Kewley et al. 2006). This abundance sequence is a key diagnostic of the physical conditions in star-forming galaxies due to its dependence on the global gas-phase metallicity distribution, the stellar ionizing radiation field and various properties of the interstellar medium (ISM) within galaxies such as the geometrical distribution of gas and electron density (see e.g. Kewley et al. 2013b).

Since the **BPT diagram** is constructed from the strongest rest-frame optical emission lines, most readily observable at high redshifts, it is possible to trace the BPT abundance sequence across cosmic time. Therefore, with new ground-based multi-object spectrographs such as KMOS (Sharples et al. 2013) and MOSFIRE

(McLean et al. 2012), the **BPT diagram** is beginning to be studied with large statistical samples out to  $z \gtrsim 2.2$  (e.g. Steidel et al. 2014; Shapley et al. 2015). Interestingly, it has become apparent that the BPT abundance sequence defined by local galaxies evolves out to high redshift, suggesting an evolution in the physical conditions of star-forming regions (Kewley et al. 2013a; Steidel et al. 2014; Shapley et al. 2015). Specifically, from the growing sample of galaxies at  $z > 1$ , an evolution away from the local abundance sequence towards higher values of the [O III]/H  $\beta$  and/or [N II]/H  $\alpha$  ratio is observed (see e.g. Kewley et al. 2013a; Cullen et al. 2014; Steidel et al. 2014; Kewley et al. 2015; Shapley et al. 2015; Holden et al. 2016, for discussions). Steidel et al. (2014) have definitively confirmed this abundance sequence trend using MOSFIRE observations of 179 star-forming galaxies at  $z \sim 2.3$ , showing that these galaxies form a distinct, yet similarly tight, locus in the BPT plane compared to SDSS galaxies (see also Masters et al. 2014; Shapley et al. 2015).

The cause of the abundance sequence evolution is still not fully resolved. Kewley et al. (2013b) have investigated theoretically how this evolution may be related to a change in physical conditions in galaxies at high redshift. They suggest more extreme ISM conditions in high-redshift H II regions will result in the abundance sequence at a given redshift being offset from the one defined locally. However, the precise nature of these ‘extreme conditions’ is still a matter of debate. Kewley et al. (2013b) show how the

\* E-mail: fc@roe.ac.uk

<sup>†</sup> Scottish Universities Physics Alliance.

independent effects of a harder ionizing radiation field, an increase in the reservoir of ionizing photons (more stars forming within a given volume) and an increasing electron density (or equivalently ISM pressure) can shift the position of galaxies in the **BPT diagram** in a similar way to that observed. Kewley et al. (2013a) conclude that star-forming galaxies at high redshift are consistent with having some combination of a larger ionization parameter, larger electron density, and/or harder ionizing radiation field. Steidel et al. (2014), using an independent model, reach a similar conclusion, and also suggest that variations in the N/O ratio at fixed O/H in high-redshift galaxies may also be responsible in part for the observed shift in locus in the **BPT diagram**. Sanders et al. (2016) argue that high-redshift galaxies are consistent with having the same ionization parameter as local galaxies at fixed metallicity and that the primary driver of the **BPT diagram** offset is an elevated N/O ratio at fixed O/H. Most recently, Kashino et al. (2016) have concluded that the line ratio evolution is caused by a combination of an increase in the electron density and ionization parameter, where the increase in ionization parameter out to  $z = 1.7$ .

In this paper, motivated by the high-redshift BPT studies, we investigate the physical conditions in high-redshift galaxies by comparing the redshift evolution of the [O III]/H  $\beta$  nebular emission line ratio to theoretical models of star-forming galaxies. The advantage of focusing solely on the [O III]/H  $\beta$  ratio is two-fold. First, the evolution of [O III]/H  $\beta$  is independent of any N/O variations, and thus mitigates any problems related to the degeneracy between evolving ionization conditions and evolving N/O. Secondly, it also allows us to extend the observed data out to  $z \sim 3.5$ , beyond where the [N II]/H  $\alpha$  ratio can be measured with current instrumentation. Recently, Kewley et al. (2015) have shown that in the redshift range  $0.2 < z < 0.6$  the [O III]/H  $\beta$  rises by 0.2–0.3 dex, and attribute this evolution to an increase in the ionization parameter over this redshift range due to a larger ratio of young to old stars in galaxies. In this paper, we are able to extend the [O III]/H  $\beta$  analysis over a larger redshift baseline.

Primarily, we utilize an independent reduction of 3D-*HST* survey data (Brammer et al. 2012; Momcheva et al. 2015) to constrain the line ratio evolution in the redshift range  $1.3 < z < 2.3$ . With the 3D-*HST* data we are able to account for any selection effects resulting from redshift-dependent line luminosity thresholds, the importance of which has been recently highlighted in the literature. For example, Juneau et al. (2014) find that, using a sample of SDSS galaxies, it is possible to re-create the abundance sequence offset at high redshift by applying increasing [O III] and H  $\alpha$  luminosity cuts to the sample, and emphasize the importance of comparing emission-line-luminosity matched samples when comparing local and high-redshift galaxies (see also Cowie, Barger & Songaila 2016). We address the issue of completeness and selection biases in the 3D-*HST* sample in Section 2. Once these luminosity completeness effects within 3D-*HST* have been accounted for, we are able to supplement these data with  $z \simeq 3.5$  observations from the AMAZE/LSD surveys (Maiolino et al. 2008; Mannucci et al. 2009; Troncoso et al. 2014) and a sample of galaxies from Holden et al. (2016).

We then compare the evolution of the [O III]/H  $\beta$  ratio between the SDSS, 3D-*HST*, and  $z \simeq 3.5$  samples, across  $\sim 10.7$  Gyr of cosmic time, to the theoretical models introduced in Kewley et al. (2013b). We describe how these models are applied in the context of this work in Section 4. These models allow different scenarios for the physical evolution of the star-forming regions with redshift to be distinguished. Specifically, we are able to separate the effects on the line ratio due to global evolution in metallicity, ionization parameter

and ISM pressure. The results are presented in Section 5. Finally, in Section 6 we discuss the practical implication of the results with respect to accurately measuring the gas-phase metallicity of galaxies at high redshift and explore some theoretical interpretations of our results.

Throughout, we refer to the [O III]  $\lambda 5007$  nebular emission line as [O III] unless explicitly stated otherwise. We assume a cosmology with  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2 DATA

The data analysed in this paper are drawn from a variety of separate star-forming galaxy catalogues across the redshift range  $0 < z < 4$ . The objective behind this selection is to sample the [O III]/H  $\beta$  versus redshift relation across the widest redshift baseline possible. Below we describe the separate samples. All data sets are cut to only include galaxies with  $\log(M/M_*) > 9.0$ , for consistent comparison with the theoretical models described in Section 4.

### 2.1 3D-*HST*

The primary data set presented in this paper, in the sense that it is the largest data set used to constrain the evolution of the [O III]/H  $\beta$  ratio at  $z \gtrsim 1$ , is a sample of galaxies drawn from an independent reduction of the 3D-*HST* survey (Brammer et al. 2012; Momcheva et al. 2015). Below we give details of the 3D-*HST* sample selection and spectroscopic analysis relevant to this study; however, see Cullen et al. (2014) for a more detailed description of the data reduction process.

#### 2.1.1 3D-*HST* sample selection

The 3D-*HST* galaxy sample is drawn from three of the five survey fields (GOODS-S, COSMOS and UDS) and for the analysis presented in this paper we only consider galaxies in the redshift range  $1.3 \leq z \leq 2.3$ , where both the [O III] and H  $\beta$  lines are observable in the grism spectra. All galaxies are selected visually from the reduced 2D and 1D spectra in each field. In the visual inspection we check each spectrum for contamination by nearby sources, rejecting emission line spectra where the emission lines are severely contaminated, or the contamination model leaves clearly erroneous residual features in the continuum. As described in Cullen et al. (2014), we use existing CANDELS catalogues for the GOODS-S and UDS fields as described in Guo et al. (2013) and Galametz et al. (2013), respectively. For COSMOS we produce a custom catalogue, the details of which are described in Bowler et al. (2012). Stellar masses for the 3D-*HST* galaxies are measured from rest-frame UV to mid-IR photometry using the publicly available code LEPHARE (Ilbert et al. 2006). We run LEPHARE with solar-metallicity Bruzual & Charlot (2003) templates assuming exponentially declining  $\tau$  model star formation histories with  $\tau = 0.3, 1, 2, 3, 5, 10, 15, 30$  Gyr and a Chabrier initial mass function (IMF). The Calzetti et al. (2000) attenuation law is used to account for dust extinction with  $E(B - V)$  values ranging from 0 to 0.6. The age of the model is allowed to vary between 0.05 Gyr and the age of the Universe at the spectroscopic redshift of the galaxy. We remove AGN contaminants and exclude galaxies with a  $< 3\sigma$  detection in [O III] as described below. This leaves a final 3D-*HST* sample of  $N = 211$  galaxies down to a minimum [O III] line flux of  $1.2 \times 10^{-17} \text{ ergs cm}^{-2} \text{ s}^{-1}$ .

### 2.1.2 Continuum and line flux fitting

For each individual galaxy in the sample we first fit the continuum by masking the emission lines in the grism spectra, then, using a least-squares minimization technique, we find the best-fitting normalization between the grism continuum and the best-fitting BC03 model returned from spectral energy distribution (SED) fitting. The normalized BC03 model is then taken as the continuum of the galaxy and subtracted from the grism spectrum. This method of continuum subtraction accounts for the H  $\beta$  absorption in the stellar continuum through the BC03 models. We then fit the [O III] and H  $\beta$  line using a triple Gaussian function. To estimate the errors on line measurements we perturb each spectrum, at each wavelength pixel, by drawing a random number from a Gaussian distribution with a standard deviation equal to the error value in that pixel. The continuum and line fits are then performed on the perturbed spectrum. This process is repeated  $10^3$  times and, for each line ([O III] and H  $\beta$ ), the final line flux and error is taken as mean and standard deviation of the resulting distributions.

### 2.1.3 AGN removal

The theoretical models for [O III]/H  $\beta$  evolution explored in this paper assume the ionizing radiation responsible for the nebular emission has a stellar origin. Therefore, to facilitate an accurate comparison between models and data, care must be taken to account for the effect of AGN contamination in our sample. Unfortunately the wavelength coverage, sensitivity, and resolution of the grism spectra do not allow us to detect simultaneously enough emission lines to use classical AGN diagnostic diagrams (e.g. the [BPT diagram](#)). Therefore, the most reliable method for AGN removal applicable to our sample is X-ray detection. We use the 4Ms GOODS-S (Xue et al. 2011) catalogue and 1.8Ms COSMOS (Civano et al. 2012) catalogue to remove  $N = 17$  X-ray selected AGN from our sample. We do not apply a common threshold to account for the different catalogue depths,  $N = 12/17$  and  $5/17$  AGN were identified in GOODS-S and COSMOS, respectively. We could not perform a similar selection in UDS due to the lack of a sufficiently deep X-ray catalogue. We choose not to apply other optical emission line diagnostics, for example the MEx AGN diagnostic (Juneau et al. 2014), as the premise of this study is based on investigating the idea that conditions in star-forming regions become more extreme at high redshift. Therefore, by using traditional optical methods we risk removing genuine extreme star-forming galaxies from our sample. Moreover, Coil et al. (2015) find that the MEx diagram does not robustly identify AGN at  $z \sim 2$ . Moreover, we find that only  $N = 16/313$  galaxies would be classified as AGN via the MEx classification, and have checked that removing these galaxies has no effect on the final results.

In summary, we have identified and removed  $N = 17$  X-ray selected AGN ( $\sim 5$  per cent of the original sample). This is slightly lower than, previous studies at similar redshifts (e.g. Stott et al. 2013; Zahid et al. 2014) which find  $\sim 10$  per cent AGN contamination. Based on the number of AGN identified in the GOODS-S X-ray catalogue, and assuming a constant number of AGN per unit area of sky, we estimate that our sample may be contaminated by  $\sim 5$  and 10 unidentified AGN in COSMOS and UDS, respectively. These numbers are also consistent with the  $\sim 10$  per cent contamination fraction found in other studies. Therefore, we cannot rule out residual AGN contamination in our final sample primarily due to the lack of a sufficiently deep X-ray catalogue UDS and to insufficient line diagnostic tools. However, we are able to account

for this residual AGN contamination in our stacked spectra via a bootstrapping technique described in Section 2.1.4. Moreover we have verified that removing the  $N = 68/211$  UDS galaxies from our sample does not change any of the results presented in this paper.

### 2.1.4 Stacking procedure

To reduce the uncertainty in the measured line ratios due to the lower signal to noise of the weaker H  $\beta$  line we must stack the 3D-*HST* spectra. Throughout this paper we stack the galaxies in redshift space. A redshift bin is defined, and within each bin the galaxy spectra were interpolated to a common wavelength grid. We then determine the inverse variance weighted mean (or median) flux within each wavelength bin across all the galaxies in that bin.

We estimate errors on line flux measurements in the stacked spectra by bootstrapping the data. Any large scatter induced by AGN contamination will also be accounted for by bootstrapping. For each stack we randomly select  $N$  galaxies, with replacement, where  $N$  is the number of galaxies in the stack. We then recombine the spectra and measure the line fluxes as described above. This process is repeated  $10^3$  times and the flux and error are taken as the mean standard deviation of the final distribution for each line.

## 2.2 SDSS

When discussing the evolution of physical conditions in H II regions with redshift, we require a comparison sample with which to estimate values of a given parameter (e.g. metallicity, ionization parameter, ISM pressure) typical of galaxies in the local Universe. For the local galaxy sample we use an SDSS catalogue containing  $N = 51\,262$  star-forming galaxies from Zahid et al. (2013); this sample of star-forming galaxies is selected in the redshift range  $0.02 < z < 0.12$ , with an aperture covering fraction  $\gtrsim 20$  per cent to minimize aperture effects and with AGN removed using the Kewley et al. (2006) optical classification scheme. The emission line fluxes in this catalogue are taken from the JHU/MPA catalogues,<sup>1</sup> where the line fluxes have been corrected for underlying stellar absorption (Brinchmann et al. 2004; Tremonti et al. 2004). Stellar masses for the SDSS sample are calculated from the SDSS *ugriz* photometry as described in Zahid et al. (2013).

## 2.3 AMAZE/LSD

We utilize a sample of  $N = 40$  star-forming galaxies at  $3 < z < 4$  (median  $z = 3.4$ ) taken from Troncoso et al. (2014). The data compiled in Troncoso et al. (2014) were taken as part of the AMAZE (Maiolino et al. 2008) and LSD (Mannucci et al. 2009) surveys undertaken with the SINFONI instrument at the VLT. The galaxies were selected to exclude AGN based either on their UV spectra, X-ray data or MIPS 24  $\mu\text{m}$  flux (see Maiolino et al. (2008) for details). We select galaxies from the Troncoso et al. (2014) catalogue with a  $3\sigma$  detection in [O III]. The final sample contains  $N = 33/40$  galaxies.

Troncoso et al. (2014) did not perform a correction for underlying H  $\beta$  absorption, concluding the correction would be negligible due to the non-detection of a continuum in the majority of sources. Stellar masses for this sample were derived from either 14-band UV – *Spitzer*-IRAC photometry or optical photometric data (*U*, *G*, *R*, *I*) plus *Spitzer* IRAC and MIPS bands. This is the only sample with stellar masses derived using a Salpeter IMF (Salpeter 1955) so

<sup>1</sup> <http://www.mpa-garching.mpg.de/SDSS/DR7/>

we convert the quoted values to a Chabrier IMF (Chabrier 2003) by dividing by 1.8.

### 2.4 Holden et al. (2016)

We increase the  $z > 3$  sample size using the Holden et al. (2016) data set, consisting of  $N = 18$  Lyman-break selected star-forming galaxies at  $z \sim 3.2\text{--}3.7$  with [O III]/H  $\beta$  ratios measured using the MOSFIRE instrument on the Keck-I telescope. AGN contamination is ruled out based on the rest-frame UV spectra and X-ray data. Again no correction has been made for the underlying H  $\beta$  absorption, but again no spectra have reliably detected continua and thus, as argued by Troncoso et al. (2014), the correction would most likely be negligible. We select galaxies from the Holden et al. (2016) catalogue with a  $3\sigma$  detection in [O III]. The final sample contains  $N = 15/18$  galaxies. In combination with the AMAZE/LSD galaxies this gives a total of  $N = 48$  galaxies at  $z > 3$  with a median signal-to-noise ratio of 4 in the [O III]/H  $\beta$  ratio.

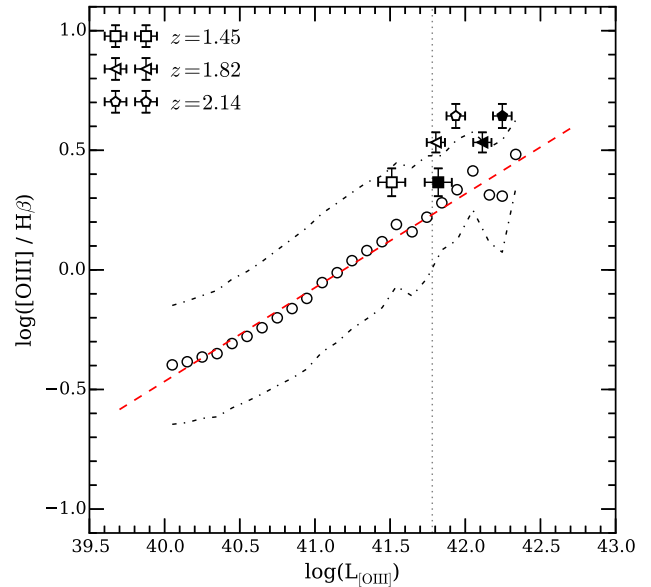
## 3 [O III] LUMINOSITY COMPLETENESS

Before investigating the [O III]/H  $\beta$  versus redshift relation, it is important to understand any selection biases present in the above data set which can affect the measurement of line ratios. As has been recently pointed out by Juneau et al. (2014), any observed evolution of line ratios could be a result of larger line luminosity limits at high redshift. Cosmological flux dimming will cause the minimum observable [O III] luminosity to increase with redshift, and if [O III]/H  $\beta$  correlates with [O III] luminosity, an evolution to a higher ratio with redshift could be an artefact of the increasing [O III] luminosity detection threshold.

### 3.1 [OIII] luminosity: evolution compared to SDSS

We first explore how this affects any observed evolution compared to galaxies in the local Universe. Juneau et al. (2014) find, using an SDSS sample of 300 000 star-forming galaxies, that by imposing a minimum line luminosity to [O III] and H  $\alpha$  they can produce subsamples in the SDSS data that mimic the BPT evolution observed at  $z > 1$ . For the purposes of our results the implication is that as the line detection luminosity threshold is increased, an artificial evolution towards higher [O III]/H  $\beta$  is observed. We investigate whether the luminosity of [O III] correlates with [O III]/H  $\beta$  in our SDSS sample in Fig. 1. We find, in agreement with Juneau et al. (2014), that the [O III]/H  $\beta$  increases with [O III] luminosity.

Also plotted in Fig. 1 are the position of three redshift stacks from the 3D-*HST* sample both corrected (filled symbols), and uncorrected (open symbols), for dust using the Calzetti et al. (2000) attenuation law. The dust attenuation for the stack is taken as the median of the  $E(B - V)$  values for each individual galaxy in the stack derived from the SED fitting. We assume an extra extinction for nebular lines following the Wuyts et al. (2013) prescription. It can be seen from Fig. 1 that, even in the case of no dust correction, our sample lies at the extreme end of the SDSS sample in terms of  $L_{[\text{O III}]}$  and [O III]/H  $\beta$ . Fig. 1 indicates that a comparison of the 3D-*HST* sample to the total SDSS star-forming sample will be biased by the line luminosity detection limit of 3D-*HST*. Any observed [O III]/H  $\beta$  evolution in this case could simply be an artefact of the clear correlation between  $L_{[\text{O III}]}$  and [O III]/H  $\beta$ . Instead, we must take care to select galaxies from SDSS with comparable  $L_{[\text{O III}]}$ . To establish the detection limit in [O III] luminosity in the 3D-*HST* sample we perform a set of simulations as described below.

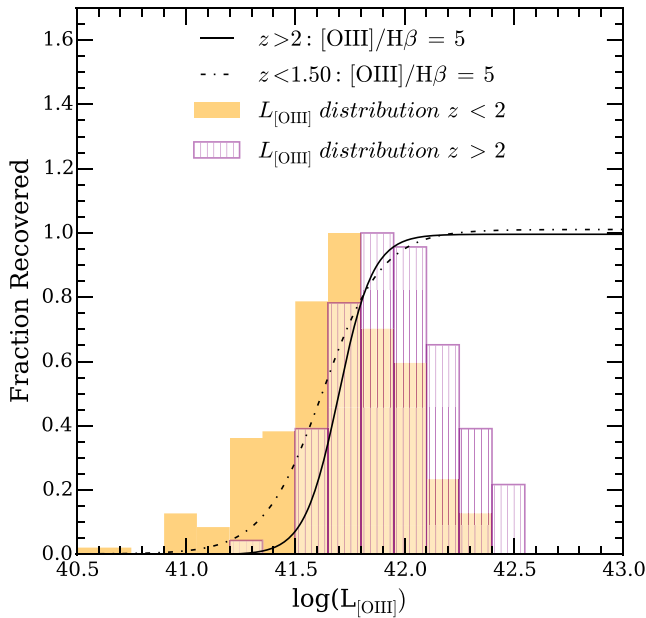


**Figure 1.** Relationship between [O III]/H  $\beta$  and  $L_{[\text{O III}]}$  for the SDSS sample compared to three 3D-*HST* galaxy stacks. The small open circles show the averages of SDSS galaxies binned in 0.05 dex wide bins of  $\log(L_{[\text{O III}]})$ , with the dot-dashed lines showing the  $\pm 1\sigma$  scatter. The red dashed lines show a simple linear fit to the SDSS galaxies. The square, triangle and circle symbols show the position of the median stacked spectra from the 3D-*HST* sample in increasing bins of redshift. The open symbols show the positions of the stacks without applying a dust correction to  $L_{[\text{O III}]}$ , while the filled symbols show the same stacks after applying a Calzetti et al. (2000) dust correction using the median  $E(B - V)$  of the individual galaxies within the stack. The dotted vertical black lines shows the  $\log(L_{[\text{O III}]})$  completeness limit at  $z > 2$  for the 3D-*HST* sample as described in Section 3.

### 3.2 3D-*HST* simulations

We use the highest redshift galaxies in our 3D-*HST* sample ( $z > 2$ ) to determine the [O III] luminosity threshold of the full sample. By using only galaxies with  $L_{[\text{O III}]}$  above this level we can minimize any completeness biases which are introduced when comparing data at different redshifts. To determine this minimum [O III] luminosity we run a set of simulations using the software package AXESIM, a dedicated simulation package for *HST* grism data. We first create synthetic emission line spectra at  $z = 2.15$  (median redshift of the  $z > 2$  galaxies) spanning [O III] luminosities in the range  $40.0 < \log(L_{[\text{O III}]}) < 43.0$  with [O III]/H  $\beta = 5$ , which is approximately the average ratio for the stacked galaxy spectra at  $z > 2$ . These synthetic spectra are then injected into a simulated G141 grism image with noise properties governed by the input exposure time to the simulation.

Once the synthetic grism images are created, they are fed through the same AXE pipeline used for reducing the observed data, and 1D spectra extracted in the same way. These spectra are visually inspected and either recovered or rejected in an identical way to the real data. From this we estimate the recovery fraction as a function of  $\log(L_{[\text{O III}]})$  and this is shown as the black solid line in Fig. 2. From the fit to the recovery fraction shown in Fig. 2, we infer 90, 75 and 50 percent completeness limits at  $z > 2$  of, respectively,  $\log(L_{[\text{O III}]}) = 41.87, 41.77$  and  $41.70$ . Also shown in Fig. 2 are the  $\log(L_{[\text{O III}]})$  distributions for the  $z \geq 2$  (purple hatched histogram) and  $z < 2$  (orange filled histogram) galaxies in the sample normalized to a peak value of 1. It can be seen that, as expected, a significant fraction of galaxies at  $z < 2$  fall below the detection



**Figure 2.** Results of the AXESIM simulations used to determine the completeness limits in  $\log(L_{[\text{O III}]})$  at  $z > 2$  for the 3D-*HST* sample. The black solid line is the best-fitting logistic function to simulation recovery fraction. The filled orange histogram shows the  $\log(L_{[\text{O III}]})$  distribution for the galaxies in the 3D-*HST* sample at  $z < 2$  and the purple histogram shows the distribution for galaxies at  $z \geq 2$ . The black dot-dashed line shows the simulated recovery fraction of  $z < 1.5$  galaxies with high  $[\text{O III}]/\text{H}\beta$  ratios (see text for a detailed explanation).

threshold at  $z \geq 2$ . Therefore, for investigating the evolution of  $[\text{O III}]/\text{H}\beta$  with redshift in Section 5, we limit all samples to include only galaxies at  $[\text{O III}]$  luminosities brighter than the above completeness limits. Clearly, our results will be most robust for the 90 per cent completeness cut, but it is also instructive to investigate how different completeness limits affect the measurements.

Another more subtle bias that could be introduced into our data is the non-selection of galaxies with high  $[\text{O III}]/\text{H}\beta$  ratios at the lowest redshifts in the 3D-*HST* sample. Such a bias would artificially decrease the average  $[\text{O III}]/\text{H}\beta$  at lower redshifts and could arise because the sensitivity of the G141 grism falls off at shorter wavelengths. For any given redshift, at a large enough  $[\text{O III}]/\text{H}\beta$  ratio, the  $\text{H}\beta$  line will not be visible in the spectrum below a given  $[\text{O III}]$  flux. Due to the decreased sensitivity at short wavelengths this flux limit will be lower at lower redshift. Without a visible  $\text{H}\beta$  it becomes less likely the galaxy will be visually selected based on the single line detection of  $[\text{O III}]$ . Therefore, galaxies with large  $[\text{O III}]/\text{H}\beta$  may not be selected. We run a further set of simulations to check whether galaxies with  $[\text{O III}]/\text{H}\beta = 5$ , typical of our  $z > 2$  galaxies, would be visually selected at  $z < 1.5$ . The recovery fraction as a function of  $\log(L_{[\text{O III}]})$  is shown by the black dot-dashed line in Fig. 2. We find that above the completeness limit for  $z > 2$  galaxies discussed above, we are also complete in galaxies with high  $[\text{O III}]/\text{H}\beta$  ratios at the lowest redshifts in our sample.

Finally, we note that though our method of matching samples by  $[\text{O III}]$  luminosity provides the simplest way of ensuring no artificial evolution of the  $[\text{O III}]/\text{H}\beta$  ratio, it would also be desirable to compare the evolution of the ratio in bins of various different physical properties. However, we have confirmed that these luminosity cuts do not produce biases in the global properties of the SDSS sample

with respect to the  $z > 2$  galaxies. Specifically, the median masses of the SDSS and 3D-*HST* samples (using the 90 per cent threshold) are consistent ( $\log(M/M_{\odot}) = 9.36$  and  $9.49$ , respectively), as are the median star formation rates (SFRs) ( $\log(M_{\odot}\text{yr}^{-1}) = 0.92$  and  $0.83$ , respectively). The SDSS SFRs are derived from the  $\text{H}\alpha$  luminosity, corrected for dust reddening using the  $\text{H}\alpha/\text{H}\beta$  ratio, and the 3D-*HST* SFRs are taken from the best-fitting SED model. Therefore, by applying an increasing  $[\text{O III}]$  luminosity cut to the SDSS sample we are selecting a population of low-mass, highly star-forming galaxies similar to those observed at  $z > 1.5$ . Nevertheless, one way of extending this analysis in the future would be to select samples at different redshift matched in bins of various physical properties such as stellar mass, SFR, specific star formation rate (sSFR) etc. We intend to explore this topic in more detail in a future work.

To summarize, we run simulations to account for two sources of bias in the 3D-*HST* sample: (i) an increasing lower limit to  $L_{[\text{O III}]}$  as a function of redshift; (ii) possible non-detection of high  $[\text{O III}]/\text{H}\beta$  ratio galaxies at the lowest redshifts. To mitigate (i) we calculate the  $L_{[\text{O III}]}$  limit at  $z > 2$  and apply this cut-off to all data which investigate the  $[\text{O III}]/\text{H}\beta$  evolution in Section 5. We also confirm that, above the  $L_{[\text{O III}]}$  limit imposed by this cut, the 3D-*HST* sample is still complete for large  $[\text{O III}]/\text{H}\beta$  ratios at the lowest redshifts.

#### 4 THEORETICAL MODELS

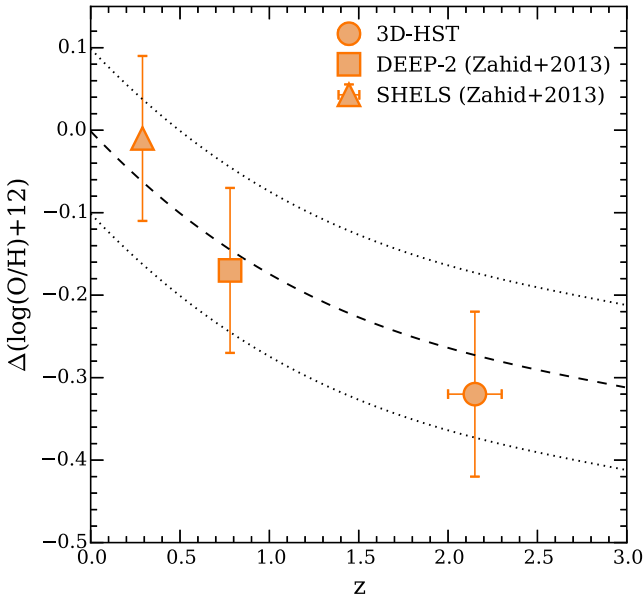
In Section 5, we compare the observed redshift evolution of the  $[\text{O III}]/\text{H}\beta$  ratio to models described in a companion paper (Kewley et al. in preparation). The redshift evolution of the  $[\text{O III}]/\text{H}\beta$  ratio is derived by combining stellar spectral synthesis and photoionization modelling of star-forming regions with a prescription for the chemical abundance evolution of galaxies with redshift. We briefly describe these models here but refer readers to Kewley et al. (2013a,b) and Kewley et al. (in preparation) for a detailed description.

To account for the evolution of the gas-phase chemical abundance a prescription is taken from the cosmological hydrodynamic simulations of Davé, Oppenheimer & Finlator (2011a) and Davé, Finlator & Oppenheimer (2011b). A third-order polynomial is fitted to the change in chemical abundance ( $\Delta\log(\text{O}/\text{H})$ ) between  $z = 3$  and 0 for galaxies with  $M_{*} > 10^9$  assuming momentum-driven winds, and is given by

$$\Delta\log(\text{O}/\text{H}) = -0.0013 - 0.2287z + 0.0627z^2 - 0.0070z^3. \quad (1)$$

This equation clearly assumes a mass-invariant metallicity change for galaxies with  $M_{*} > 10^9 M_{\odot}$ , and this may be a simplistic assumption. However in Fig. 3, we show that this evolution is a reasonable parametrization of the global metallicity evolution from  $z = 0$  to 2. In this figure, we compare the evolution in metallicity using data from Zahid et al. (2013) and Cullen et al. (2014) where, since equation (1) is independent of stellar mass, the metallicity at a given redshift is calculated as an average over all stellar masses. We take the line measurements from Cullen et al. (2014) and re-calculate the metallicity using the Kobulnicky & Kewley (2004) calibration for consistency with the Zahid et al. (2013) data.

An extreme ultraviolet (EUV) stellar radiation field is taken from the Leitherer et al. (1999) STARBURST99 stellar evolutionary synthesis models. The synthetic stellar spectra are modelled with the Pauldrach/Hillier model atmospheres (Hillier & Miller 1998; Pauldrach, Hoffmann & Lennon 2001) including the effects of metal opacities. The resulting stellar radiation field is taken as the input ionizing source to the MAPPINGS IV photoionization code (Binette,



**Figure 3.** The assumed redshift evolution of the oxygen abundance used in the [O III]/H  $\beta$  evolution models as described in the text. This evolution is taken from the cosmological hydrodynamic simulations of Davé et al. (2011a,b) and is shown as the black dashed line, with the dotted line representing the  $\sim 0.1$  dex error. The three data points are from various observations where metallicities have been measured using the Kobulnicky & Kewley (2004) calibration and averaged over all stellar masses [triangle – SHELS data from Zahid et al. (2013); square – DEEP2 data from Zahid et al. (2013); circle – 3D-*HST* data from Cullen et al. (2014)].

Dopita & Tuohy 1985; Sutherland & Dopita 1993; Dopita et al. 2013) calculated over a large range in parameter space: metallicity,  $\log(\text{O}/\text{H}) + 12 = [8.4, 8.7, 9.0, 9.2]$ ; ionization parameter  $\log(U) = [-3.5, -3.0, -2.5, -2.0]$ ; ISM pressure,  $\log(P/k) = [5.0, 5.5, 6.0, 6.5, 7.0]$ .

To explore the evolution of the [O III]/H  $\beta$  line ratios with redshift, four different scenarios are envisaged: (1) no evolution in the ISM pressure or ionization parameter (i.e. purely chemical abundance evolution), (2) only the ISM pressure evolves with redshift, (3) only the ionization parameter evolves with redshift, (4) both the ISM pressure and ionization parameter evolve with redshift. Specifically, for the ISM pressure we assume a simple but extreme model in which ISM pressure increases by two orders of magnitude from  $z = 0.07$  (mean redshift of the SDSS) to  $z = 4$ . In an ionized gas with an isobaric density distribution, the ISM pressure is simply determined by the electron density (Dopita et al. 2006) and recent work has shown that by  $z \sim 2.3$  the typical electron density in H II regions has increased by one order of magnitude (Shirazi, Brinchmann & Rahmati 2014; Steidel et al. 2014; Sanders et al. 2016). Crucially, it is unlikely that our model underestimates the effect of ISM pressure evolution. The evolution of ionization parameter is modelled by an order of magnitude increase in the same redshift range. This is consistent with recent measurements of ionization parameter out to  $z \sim 3$  (Nakajima & Ouchi 2014). The equations for these models are given in Kewley et al. (in preparation) along with further discussion of their motivation. In each case, the reference values of ionization parameter and ISM pressure for SDSS galaxies is derived using on those galaxies which conform to the chosen [O III] luminosity threshold.

#### 4.1 Ionization parameter definition

Conventionally, the ionization parameter is defined as the ratio of the number of ionizing photons passing through a unit area of ISM per second divided by the hydrogen atom density. The hydrogen atom density, in a fully ionized plasma, is approximately equivalent to the electron density. However, since we are separately constraining the ionization parameter and ISM pressure (or electron density) in our models, it is important to note that the increasing ionization parameter models represent an increase *at fixed ISM pressure*. There are two ways to increase the ionization parameter in this way. First, the relative fraction ionizing photons can be increased by assuming a hotter stellar population which results in a harder EUV radiation field. Secondly, the EUV radiation field can be scaled by, for example, having a larger number of stars forming within an H II region. Detailed explanations of these two effects are discussed in Kewley et al. (2013b, 2015). In our models we increase the ionization parameter by scaling the radiation field and therefore the physical interpretation of a larger ionization parameter in this work is an increase in the number of hydrogen ionizing photons incident on a unit area of the ISM per second due to more stars forming within a given ISM volume.

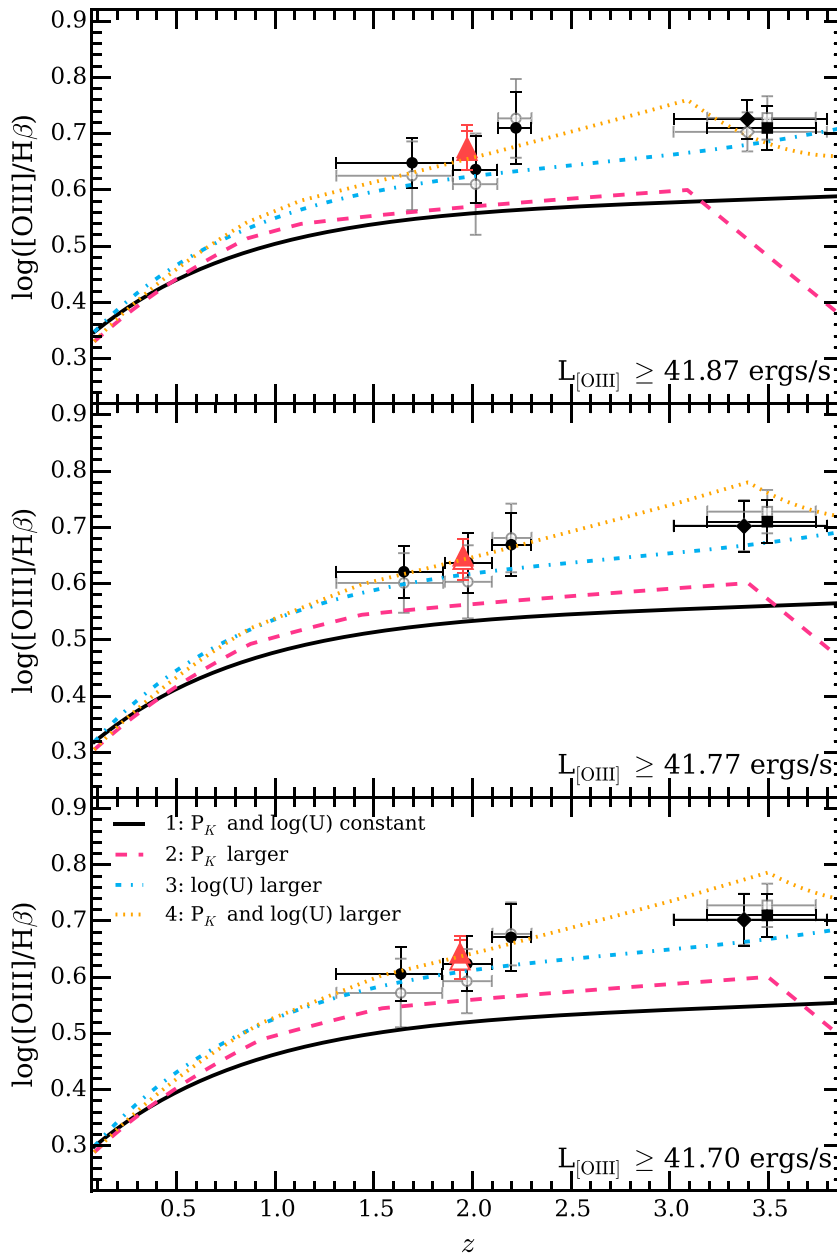
Moreover it is worth noting that since the prescription for metallicity evolution given in equation (1) is included in all models, the increase in ionization parameter due to a decrease in chemical abundance is naturally accounted for through the choice of lower metallicity *STARBURST99* stellar evolution models. This anticorrelation between ionization parameter and chemical abundance has been observed locally (Pérez-Montero 2014; Sánchez et al. 2015), and recently Sanders et al. (2016) have evidence for a similar correlation at  $z \sim 2$ . Thus models 3 and 4 represent an increase in the ionization parameter over and above that expected due to the decrease in chemical abundance predicted from equation (1).

#### 5 REDSHIFT EVOLUTION OF [O III]/H $\beta$

We now compare the observed redshift evolution of the [O III]/H  $\beta$  ratio to the theoretical models described above in each of the four scenarios. The models are run for [O III] luminosity limits of  $\log(L_{[\text{O III}]}) \geq 41.87, 41.77$  and  $41.70$  as discussed above. In the four main data sets we utilize in the comparison (SDSS, 3D-*HST*, AMAZE/LSD; Holden et al. 2016), we select the subset of each sample which adheres to the luminosity threshold. Fig. 4 shows the [O III]/H  $\beta$  versus redshift relation for the four samples which, when combined, span the redshift range  $0 < z < 4$ . The SDSS data points are not explicitly shown in Fig. 4 but can be inferred from the  $z = 0$  origin of all the models.

Each panel in Fig. 4 shows the results for each of the different luminosity cuts in [O III]. In each panel the four model scenarios are plotted where the initial ionization parameter and ISM pressure has been calculated for the subset of SDSS galaxies which adhere to the stated luminosity threshold. For the 3D-*HST* data we measure the [O III]/H  $\beta$  ratio from both a mean and median stack in three redshift bins, as well as the mean and median of a stack of all the 3D-*HST* spectra between  $1.3 < z < 2.3$ . For the AMAZE/LSD and Holden et al. (2016) data, we simply plot the mean and median values of the individual [O III]/H  $\beta$  ratios at the median redshift of the given samples. All means are shown as filled symbols and all medians are shown as open symbols.

The first result of note is that the data rule out the model in which ionization parameter and ISM pressure remain constant (solid black curve), with only the metallicity of H II regions evolving with



**Figure 4.** Comparison between the observed evolution of the  $[\text{O III}]/\text{H}\beta$  ratio as a function of redshift and the theoretical models described briefly in Section 4. The circles show the 3D-*HST* sample split over three redshift bins, the red triangles show the stack of all spectra in the 3D-*HST* sample. The diamonds show the AMAZE/LDS data and the squares the Holden et al. (2016) sample. The open and filled symbols correspond to an average and median stack, respectively (see text for details). Each panel shows a different  $[\text{O III}]$  luminosity cut as described in Section 3.

redshift. Therefore, an important result of this work is that the evolution of the  $[\text{O III}]/\text{H}\beta$  line ratio, for the luminosity thresholds applied, cannot be accounted for purely by the evolution of metallicity. This implies that the observed line ratios at high redshift are affected by the evolution in other physical parameters in  $\text{H II}$  regions. Importantly, this observation is independent of any variations in the  $\text{N/O}$  ratio. An evolving  $\text{N/O}$  ratio has been suggested by some authors in order to explain emission line ratios at high redshift, particularly in reference to the [BPT diagram](#) (e.g. Masters et al. 2014; Steidel et al. 2014; Shapley et al. 2015; Sanders et al. 2016). This is not to say that there is no  $\text{N/O}$  evolution, but rather that any  $\text{N/O}$  evolution must occur simultaneously with an evolution of other physical parameters.

The data also strongly disfavour the model in which ISM pressure evolves and ionization parameter remains constant (pink dashed curve). For the most robust subset (90 per cent completeness) only one data point is within  $1\sigma$  of this model, and across all panels the data are consistently in better agreement with either of the two remaining models. Overall, the  $\text{P/k}$  evolution model is not sufficient to explain the majority of measured line ratios and, despite some of the individual measurements being formally consistent within  $1\sigma$ , all are consistently systemically higher than the  $\text{P/k}$  evolution model prediction. Moreover, this model is clearly ruled by the stack across the full redshift range of the 3D-*HST* spectra.

The remaining two models – ionization parameter only evolution (blue dot-dashed curve) and ionization parameter + ISM pressure

evolution (yellow dotted curve) – are both favoured and are difficult to distinguish with the quality of data currently available. However this still permits us to conclude from the models that, at a minimum, the ionization parameter of H II regions is evolving with redshift. To recall, the ionization parameter is defined in our models as the number of hydrogen ionizing photons incident on a unit area of the ISM per second. It is not at present possible to conclusively say whether this is accompanied by an increase in the ISM pressure. Particularly, in the redshift range  $1.5 < z < 2.5$  covered by the 3D-*HST* data, the two models are indistinguishable given the size of the uncertainties. The only notable exception is the 3D-*HST* stack across the full redshift range for the  $\log(L_{[\text{O III}]}) \geq 41.87$  threshold, nevertheless no strong conclusions can be made on the basis of this one measurement. At  $z > 3$  the models are more separated, however, over all luminosity limits, the AMAZE/LSD and Holden et al. (2016) data still do not allow us to robustly distinguish between either scenario. In Section 6, we discuss ways in which this situation can be improved with future observations.

To summarize, the data rule out the scenario in which the evolution of the [O III]/H  $\beta$  line ratio is driven purely by metallicity evolution. In addition they require an evolution in the ionization parameter to higher values with redshift. This rise in ionization parameter may also be accompanied by a simultaneous increase of the ISM pressure; however, the data do not allow us to confirm or rule out ISM pressure evolution at this stage. Possible theoretical explanations of this result are discussed in Section 6.

### 5.1 Comparison with the literature

In this section, we compare our results to a selection of recent observational studies which have also addressed the issue of evolving physical conditions in star-forming galaxies at high redshift.

Sanders et al. (2016) have investigated ionization parameter and electron density evolution of galaxies out to  $z \sim 2.3$  based on a sample of  $\sim 200$  galaxies from the MOSDEF survey (Kriek et al. 2015). Via a comparison with local galaxies across a wide range of line ratio diagnostic diagrams involving the [O III], [O II]  $\lambda 3727$ , H  $\alpha$ , H  $\beta$ , [N II] and [S II] emission lines, Sanders et al. conclude that the typical electron density and ionization parameter increases between  $z \sim 0$  and 2.3. Sanders et al. argue that the evolution to larger ionization parameters is solely a result of the lower metallicities of the high-redshift galaxies, and that the ionization parameter of galaxies at fixed metallicity does not evolve with redshift. In this scenario, the evolution of galaxies in the BPT diagram is primarily attributed to an increase in the N/O ratio at fixed O/H (see also Masters et al. 2014; Shapley et al. 2015). In terms of our models, the Sanders et al. conclusions would correspond to the pink dashed curve in Fig. 4. This is clearly inconsistent with the results presented here. However, while Sanders et al. rule out a scenario in which the ionization parameter increases due to a harder ionization spectrum, they do not investigate case, presented here, in which the ionization parameter is increased by scaling the ionizing spectrum.

Kashino et al. (2016) have analysed the excitation properties a sample of 701 H  $\alpha$ -selected galaxies at  $1.4 \leq z \leq 1.7$  taken from the FMOS-COSMOS survey (Silverman et al. 2015). Using line ratio diagnostics based around the [O III], H  $\alpha$ , H  $\beta$ , [N II] and [S II] emission lines, Kashino et al. have concluded that the evolution of emission lines properties is due to both higher electron densities and ionization parameters at high redshift. Furthermore, based on the suppression of the [S II]/H  $\alpha$  ratio at fixed [O III]/H  $\beta$ , compared to local Universe galaxies, Kashino et al. conclude that the increase in ionization parameter is caused by a scaling up of

the radiation field, as opposed to an increase in the hardness, in agreement with the results presented here. However, Kashino et al. do not measure the [O II]  $\lambda 3727$  emission lines and are therefore unable to comprehensively address the N/O at fixed O/H issue.

At lower redshifts, Kewley et al. (2015) have measured the evolution in the [O III]/H  $\beta$  between  $0.2 < z < 0.6$  using a sample of galaxies from the SHELS survey (Geller et al. 2016). Kewley et al. conclude that the evolution of the [O III]/H  $\beta$  ratio at these redshifts is dominated by an increase in the ionization parameter caused by a scaling of the radiation field, ruling out either an increase in the hardness of the radiation field or an increase in electron density at these redshifts, based on measurements of [S II]/H  $\alpha$  and the [S II] doublet ratio, respectively. Kewley et al. also demonstrate that this evolution is independent of stellar mass within the range  $9.2 < \log(M/M_{\odot}) < 10.6$ . Again the Kewley et al. results are consistent with the results presented here.

## 6 DISCUSSION

In this section we discuss some of the practical and theoretical implications of the results presented here, as well as how our findings compare with results published in the recent literature. If the evolution in ionization parameter and/or ISM pressure in star-forming regions at high redshift is real, this will have important consequences both for how strong line ratios are applied in high-redshift galaxies to measure physical parameters and, perhaps more fundamentally, on our theoretical understanding of the nature of H II regions and the stars that power them.

### 6.1 Impact of ionization parameter on metallicities inferred from strong line ratios

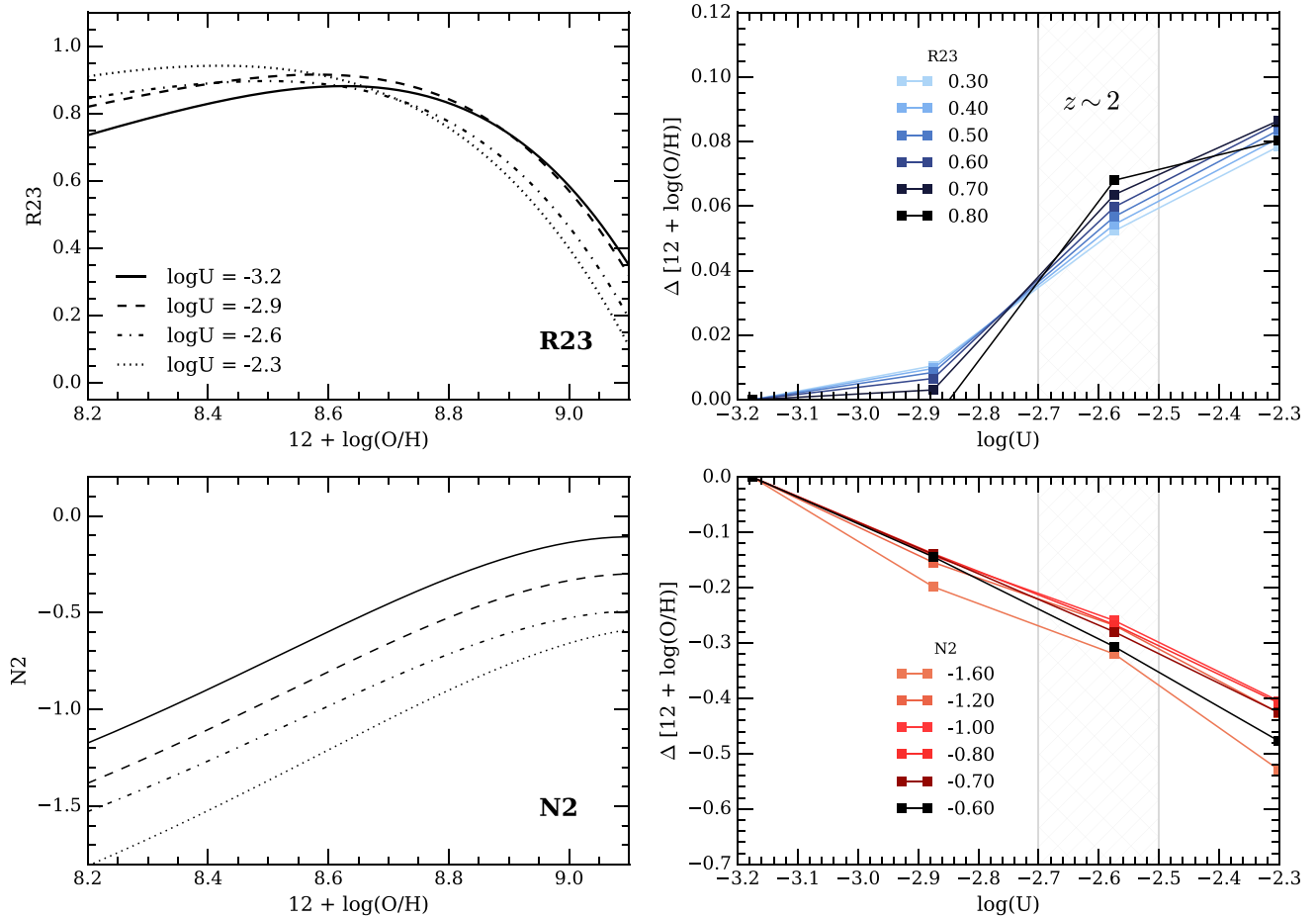
In Fig. 5, we show an illustrative example of how failing to account for an evolving ionization parameter will bias metallicity diagnostics based on strong emission line ratios. We use two commonly adopted emission lines diagnostics for the illustration: R23 = ([O II]  $\lambda 3727 + [\text{O III}] \lambda 5007)/\text{H } \beta$  (top row) and N2 = [N II]/H  $\alpha$  (bottom row). The left-hand panel in each row shows the metallicity calibration for four different values of the ionization parameter taken from Kewley & Dopita (2002). It can be seen that, depending on the value of the ionization parameter adopted, the metallicity solution can be significantly different for both the R23 and N2 diagnostics.

This is explicitly shown in the right-hand panels of each row, where we show the difference in inferred metallicity ( $\Delta[12 + \log(\text{O}/\text{H})]$ ) as a function of ionization parameter for different values of the two line ratios.<sup>2</sup> The values of  $\Delta[12 + \log(\text{O}/\text{H})]$  are measured relative to a reference ionization parameter value of  $\log(U) = -3.2$  typical of SDSS galaxies (Liang et al. 2006; Kewley & Ellison 2008). Positive values indicate that the metallicity inferred from the line ratio will be overestimated if the typical SDSS ionization parameter is assumed and vice versa. The ionization parameter values typical of current measurements of  $z \sim 2$  galaxies both from the Cullen et al. (2014) 3D-*HST* spectra at  $z > 2$ , and the  $z \sim 2-3$  LBGs from Nakajima & Ouchi (2014) is shown as the grey hatched region.

Across the range of R23 and N2 ratios shown in the figure, the inferred metallicity for a given ratio becomes increasingly offset

<sup>2</sup> For the R23 diagnostic, we assume all galaxies fall on the upper branch ( $\log(\text{O}/\text{H})+12 \gtrsim 8.6$ ) for simplicity.





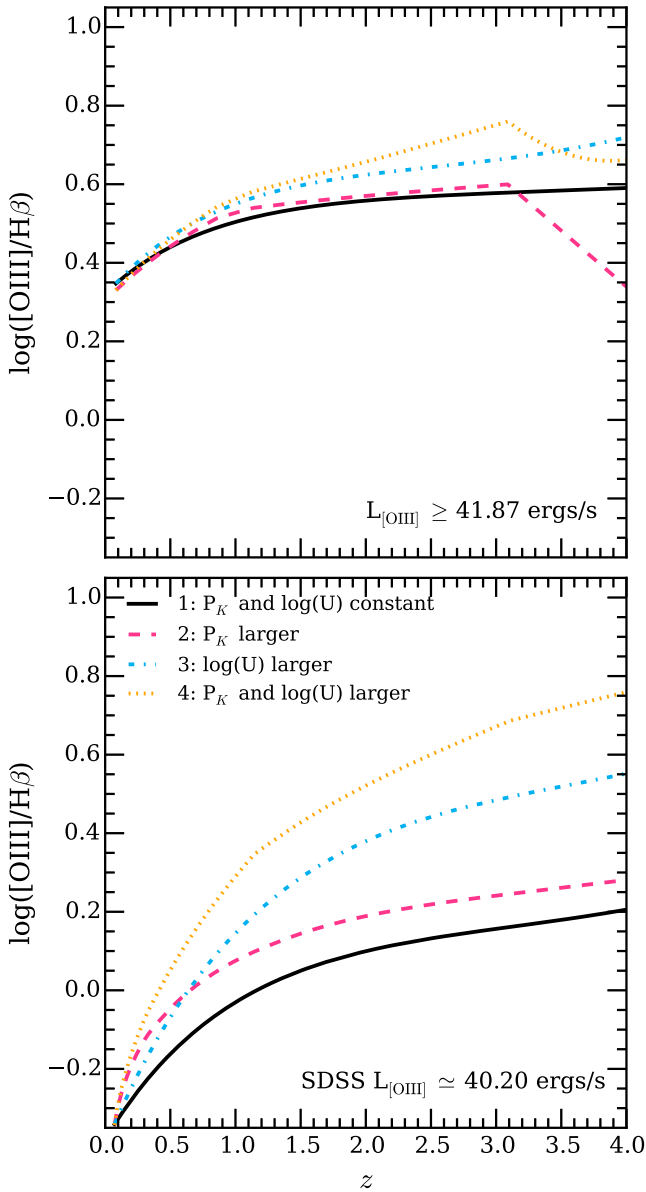
**Figure 5.** The difference in inferred metallicity as a function of ionization parameter for different values of the R23 (top row) and N2 (bottom row) emission line ratios taken from Kewley & Dopita (2002). The differences are computed relative to the metallicity inferred assuming an ionization parameter typical of local star-forming galaxies. In each row, the left-hand panel shows the line ratio metallicity calibration at four different values of the ionization parameter; the right-hand panel shows the metallicity differences as a function of ionization parameter for six different values of the ratios. The grey hatched regions represent values of ionization parameter typical of  $z \gtrsim 2$  galaxies.

with increasing ionization parameter. For the R23 ratio the difference can be of the order  $+0.06$  dex at values of the ionization parameter typical of  $z \sim 2$  galaxies, while for N2 the difference is more pronounced, with offsets of the order of  $-0.3$  dex. Clearly the N2 ratio will significantly underestimate the metallicity of galaxies at  $z \sim 2$  if an ionization parameter typical of local galaxies is assumed. Interestingly, most observational studies have concluded that the N2 ratio overestimates metallicities by up to a factor of 3 at high redshift (e.g. Liu et al. 2008; Newman et al. 2014; Steidel et al. 2014). This would suggest that other factors, such as increasing ISM pressure, the presence of shocks and/or an evolving N/O ratio may also be affecting this line ratio at high redshift. On the other hand, the offset of the R23 indicator is comparable to the estimated uncertainties due to modelling inaccuracies ( $\sim 0.1$  dex; Kewley & Dopita 2002), however the R23 indicator suffers from other significant systematic issues, namely the choice of upper or lower branch and sensitivity to the dust correction. Fig. 5 illustrates how not accounting for the ionization parameter of a galaxy can have a large impact on the inferred metallicity, particularly for calibrations based on the  $[\text{N II}]/\text{H}\alpha$  ratio. We therefore recommend that, in order to obtain reliable metallicities at high redshift, ionization parameter and metallicity should be simultaneously solved for.

## 6.2 Improving the model constraints at high redshift

As discussed in Section 5, with the data currently available at high redshift it is not possible to robustly distinguish between the models for ionization-parameter only evolution and ionization-parameter + ISM pressure evolution. It is perhaps worth illustrating how this situation can be improved with future observations. As previously discussed, the theoretical curves are dependent on the luminosity thresholds applied to the SDSS data, as this determines the reference values of ISM pressure and ionization parameter used as input to the models. Fig. 6 shows the model evolution from  $z = 0$  to 4 for two cases; the top panel shows the 3D-*HST* 90 per cent completeness luminosity threshold and the bottom panel shows the model evolution when the full SDSS sample is used to calculate the initial values of the physical parameters (corresponding to  $\log(L_{[\text{O III}]}) \sim 40.2 \text{ erg s}^{-1}$ ).

Fig. 6 illustrates how raising the luminosity threshold makes it increasingly more difficult to distinguish between the models. There is a much larger  $[\text{O III}]/\text{H}\beta$  separation between the models when the lower luminosity threshold of full SDSS sample is used. Therefore, deeper spectroscopic observation across a wider  $L_{[\text{O III}]}$  baseline will allow a more robust distinction between the two currently indistinguishable scenarios. Unfortunately SDSS-depth surveys at  $z > 1$  are



**Figure 6.** The impact of [O III] luminosity thresholds on the theoretical curves. The curves are plotted as in Fig. 4. The top panel shows the models for the 90 per cent completeness threshold used in this paper and the bottom panel shows the case for the lower luminosity limit of the full SDSS sample ( $\log(L_{[\text{O III}]}) \sim 40.2 \text{ erg s}^{-1}$ ). It is clear from the figure how deeper observations will be better able to differentiate between models at all redshifts.

not currently possible. Nevertheless, state-of-the-art ground-based near-IR surveys such as MOSDEF will be able to improve the situation by pushing down to  $5\sigma$  detections of  $\log(L_{[\text{O III}]}) = 41.5, 41.0 \text{ erg s}^{-1}$  at  $z = 2.3, 1.5$ , respectively (Kriek et al. 2015). Even deeper spectroscopic observations will then be possible using future ground and space-based near-IR instruments such as *JWST* and E-ELT.

### 6.3 Theoretical implications

Finally, we briefly discuss some of the possible theoretical explanations of the results presented here. Since our data favour models in which at least the ionization parameter increases with redshift,

but cannot distinguish whether this is also accompanied by a simultaneous increase in ISM pressure, we limit our discussion to the possible causes of a higher ionization parameter.

An increased ionization parameter requires, from the models, an increased number of ionizing photons per unit area per unit time incident on the ISM. One way of achieving this is to scale the ionizing radiation field by raising the star-formation rate within a given ISM volume. It is known that the average star formation rate of galaxies increases out to  $z \sim 2$  and beyond (e.g. Whitaker et al. 2012; Pannella et al. 2015), if these stars form in similar volumes to local galaxies there will be an excess of ionizing photons at high redshift, and consequently a higher ionization parameter. Such dense star-forming clusters are commonly referred to as ‘super star clusters’. They have been theoretically investigated (see e.g. Elmegreen 1998; Hopkins et al. 2013) and have been observed in local dwarf galaxies and galaxy mergers (e.g. Shioya, Taniguchi & Trentham 2001; Wofford et al. 2014). However to date no comprehensive statistical study of the ionization parameter in these clusters has been made, nor any detailed investigation into their prevalence at high redshift.

The effective ionization parameter is also affected by the geometrical distribution of the gas in H II regions, for example clumpy H II regions allow more ionizing photons to escape and hence lower the effective ionization parameter (Nakajima & Ouchi 2014). The difference between a radiation bounded H II region (where the nebula ends where hydrogen is completely recombined) and a density bounded H II region (where the entire nebula is ionized by the stars) is also important. In a density bounded H II region the [O III] zone (closest to the star) will likely remain unaffected but the hydrogen recombination zone (which extends to larger radii) could be shortened, thus resulting in an increase in the [O III]/H  $\beta$  ratio. Some evidence exists for density bounded H II regions at high  $z$  (see Kewley et al. 2013b, for a detailed discussion of the geometrical effects).

## 7 SUMMARY AND CONCLUSIONS

We have investigated the [O III]/H  $\beta$  versus redshift relation for a sample of galaxies in the redshift range  $0 < z < 4$  by combining data from SDSS, 3D-*HST*, AMAZE/LSD and Holden et al. (2016). We have compared the [O III]/H  $\beta$  evolution to the theoretical models described in Kewley et al. (2013b), that account for the effects of an evolving ionization parameter and ISM pressure, in order to further understand the evolution in the physical conditions of star-forming galaxies over the last  $\sim 11$  billion years of cosmic time. We have explored how line luminosity selection limits can affect any observed trends and carefully accounted for these selection biases in all data sets. We have finally discussed some of the practical and theoretical implications of the findings. Below is a summary of the results.

(i) We first show, using the SDSS sample, that there is a positive correlation between [O III]/H  $\beta$  ratio and [O III] luminosity, as pointed out by Juneau et al. (2014). This relationship could potentially bias any results since the observed evolution of the [O III]/H  $\beta$  ratio could simply be an artefact of the increasing line luminosity threshold with redshift. Therefore we conclude that the luminosity detection limit of all samples must be carefully accounted for when comparing to models of line-ratio evolution.

(ii) We correct for any biases introduced from an [O III] luminosity detection limit by restricting all samples to galaxies with  $\log(L_{[\text{O III}]}) > 41.87, 41.77$  and  $41.70$ , corresponding to completeness limits in the 3D-*HST* sample of, respectively, 90, 75 and

50 percent. We show that, after applying these luminosity thresholds, the 3D-*HST* stacks are consistent with an increase in the [O III]/H  $\beta$  ratio at  $1.3 < z < 2.3$  relative to SDSS galaxies. Therefore we conclude that this evolution is real and unaffected by the luminosity detection bias highlighted in Juneau et al. (2014).

(iii) We compare the [O III]/H  $\beta$  redshift evolution to theoretical models described in Section 4 based on the models of Kewley et al. (in preparation). From these models we rule out the possibility that the line ratio increase is a direct result of metallicity evolution.

(iv) We find the observed line ratio evolution is best accounted for by, at least, an increase in the ionization parameter with redshift. In the context of our models this ionization parameter increase is a result of scaling up the incident EUV radiation field. Unfortunately the data cannot yet distinguish whether this ionization parameter increase is also accompanied by an increase in ISM pressure. These conclusions are independent of the line luminosity limits imposed.

(v) Finally we discuss how, assuming an increase in ionization parameter and ISM pressure at high redshift, locally calibrated metallicity diagnostics will be biased when applied to high redshift galaxies. This highlights the need for the proper treatment of other physical parameters (and certainly the ionization parameter) when estimating the metallicity of high-redshift galaxies.

## ACKNOWLEDGEMENTS

We would like to thank the referee for very helpful comments which have improved the manuscript. We are grateful to H. J. Zahid for giving us access to the SDSS catalogue used in this work. FC and MC acknowledge the support of the Science and Technology Facilities Council (STFC) via the award of an STFC Studentship and an STFC Advanced Fellowship, respectively. RJM acknowledges the support of the European Research Council via the award of a Consolidator Grant (PI McLure). JSD acknowledges the support of the European Research Council via the award of an Advanced Grant, and the contribution of the EC FP7 SPACE project ASTRODEEP (Ref. No. 312725). This work is based on observations taken by the 3D-*HST* Treasury Program (GO 12177 and 12328) with the NASA/ESA *HST*, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. This work is based (in part) on observations made with the *Spitzer Space Telescope*, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This research made use of Astropy, a community-developed core PYTHON package for Astronomy (Astropy Collaboration et al. 2013), NUMPY and SCIPY (Oliphant 2007), Matplotlib (Hunter 2007), IPYTHON (Pérez & Granger 2007) and NASA's Astrophysics Data System Bibliographic Services.

## REFERENCES

- Astropy Collaboration et al., 2013, *A&A*, 558, A33  
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5 (BPT)  
 Binette L., Dopita M. A., Tuohy I. R., 1985, *ApJ*, 297, 476  
 Bowler R. A. A. et al., 2012, *MNRAS*, 426, 2772  
 Brammer G. B. et al., 2012, *ApJS*, 200, 13  
 Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 115  
 Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000  
 Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682  
 Chabrier G., 2003, *PASP*, 115, 763  
 Civano F. et al., 2012, *ApJS*, 201, 30  
 Coil A. L. et al., 2015, *ApJ*, 801, 35  
 Cowie L. L., Barger A. J., Songaila A., 2016, *ApJ*, 817, 57  
 Cullen F., Cirasuolo M., McLure R. J., Dunlop J. S., Bowler R. A. A., 2014, *MNRAS*, 440, 2300  
 Davé R., Oppenheimer B. D., Finlator K., 2011a, *MNRAS*, 415, 11  
 Davé R., Finlator K., Oppenheimer B. D., 2011b, *MNRAS*, 416, 1354  
 Dopita M. A. et al., 2006, *ApJ*, 647, 244  
 Dopita M. A., Sutherland R. S., Nicholls D. C., Kewley L. J., Vogt F. P. A., 2013, *ApJS*, 208, 10  
 Elmegreen B. G., 1998, in Woodward C. E., Shull J. M., Thronson H. A., Jr, *ASP Conf. Ser. Vol. 148, Origins. Astron. Soc. Pac.*, San Francisco, p. 150  
 Galametz A. et al., 2013, *ApJS*, 206, 10  
 Geller M. J., Hwang H. S., Dell'Antonio I. P., Zahid H. J., Kurtz M. J., Fabricant D. G., 2016, *ApJS*, 224, 11  
 Guo Y. et al., 2013, *ApJS*, 207, 24  
 Hillier D. J., Miller D. L., 1998, *ApJ*, 496, 407  
 Holden B. P. et al., 2016, *ApJ*, 820, 73  
 Hopkins P. F., Cox T. J., Hernquist L., Narayanan D., Hayward C. C., Murray N., 2013, *MNRAS*, 430, 1901  
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90  
 Ilbert O. et al., 2006, *A&A*, 457, 841  
 Juneau S. et al., 2014, *ApJ*, 788, 88  
 Kashino D. et al., 2016, preprint ([arXiv:1604.06802](https://arxiv.org/abs/1604.06802))  
 Kauffmann G. et al., 2003, *MNRAS*, 346, 1055  
 Kewley L. J., Dopita M. A., 2002, *ApJS*, 142, 35  
 Kewley L. J., Ellison S. L., 2008, *ApJ*, 681, 1183  
 Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, *MNRAS*, 372, 961  
 Kewley L. J., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013a, *ApJ*, 774, L10  
 Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013b, *ApJ*, 774, 100  
 Kewley L. J., Zahid H. J., Geller M. J., Dopita M. A., Hwang H. S., Fabricant D., 2015, *ApJ*, 812, L20  
 Kobulnicky H. A., Kewley L. J., 2004, *ApJ*, 617, 240  
 Kriek M. et al., 2015, *ApJS*, 218, 15  
 Leitherer C. et al., 1999, *ApJS*, 123, 3  
 Liang Y. C., Yin S. Y., Hammer F., Deng L. C., Flores H., Zhang B., 2006, *ApJ*, 652, 257  
 Liu X., Shapley A. E., Coil A. L., Brinchmann J., Ma C.-P., 2008, *ApJ*, 678, 758  
 McLean I. S. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, *Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV*. SPIE, Bellingham, p. 84460J  
 Maiolino R. et al., 2008, *A&A*, 488, 463  
 Mannucci F. et al., 2009, *MNRAS*, 398, 1915  
 Masters D. et al., 2014, *ApJ*, 785, 153  
 Momcheva I. G. et al., 2015, preprint ([arXiv:1510.02106](https://arxiv.org/abs/1510.02106))  
 Nakajima K., Ouchi M., 2014, *MNRAS*, 442, 900  
 Newman S. F. et al., 2014, *ApJ*, 781, 21  
 Oliphant T. E., 2007, *Comput. Sci. Eng.*, 9, 10  
 Pannella M. et al., 2015, *ApJ*, 807, 141  
 Pauldrach A. W. A., Hoffmann T. L., Lennon M., 2001, *A&A*, 375, 161  
 Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21  
 Pérez-Montero E., 2014, *MNRAS*, 441, 2663  
 Salpeter E. E., 1955, *ApJ*, 121, 161  
 Sánchez S. F. et al., 2015, *A&A*, 574, A47  
 Sanders R. L. et al., 2016, *ApJ*, 816, 23  
 Shapley A. E. et al., 2015, *ApJ*, 801, 88  
 Sharples R. et al., 2013, *Messenger*, 151, 21  
 Shioya Y., Taniguchi Y., Trentham N., 2001, *MNRAS*, 321, 11  
 Shirazi M., Brinchmann J., Rahmati A., 2014, *ApJ*, 787, 120  
 Silverman J. D. et al., 2015, *ApJS*, 220, 12  
 Steidel C. C. et al., 2014, *ApJ*, 795, 165

Stott J. P. et al., 2013, MNRAS, 436, 1130  
Sutherland R. S., Dopita M. A., 1993, ApJS, 88, 253  
Tremonti C. A. et al., 2004, ApJ, 613, 898  
Troncoso P. et al., 2014, A&A, 563, A58  
Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, ApJ, 754, L29  
Wofford A., Leitherer C., Chandar R., Bouret J.-C., 2014, ApJ, 781, 122  
Wuyts S. et al., 2013, ApJ, 779, 135

Xue Y. Q. et al., 2011, ApJS, 195, 10  
Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, ApJ, 771, L19  
Zahid H. J. et al., 2014, ApJ, 792, 75

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