

# Evolution of the $H\beta + [O\text{III}]$ and $[O\text{II}]$ luminosity functions and the $[O\text{II}]$ star formation history of the Universe up to $z \sim 5$ from HiZELS

A. A. Khostovan,<sup>1★</sup> D. Sobral,<sup>2,3,4★</sup> B. Mobasher,<sup>1</sup> P. N. Best,<sup>5</sup> I. Smail,<sup>6</sup> J. P. Stott,<sup>7</sup> S. Hemmati<sup>1</sup> and H. Nayyeri<sup>8</sup>

<sup>1</sup>Department of Physics & Astronomy, University of California, UC Riverside: 92521, Riverside, USA

<sup>2</sup>Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, OAL, Tapada da Ajuda, P-1349-018 Lisboa, Portugal

<sup>3</sup>Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016 Lisbon, Portugal

<sup>4</sup>Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands

<sup>5</sup>SUPA, Institute for Astronomy, Royal Observatory of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK

<sup>6</sup>Centre for Extragalactic Astrophysics, Durham University, South Road, Durham DH1 3LE, UK

<sup>7</sup>Institute of Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK

<sup>8</sup>Department of Physics & Astronomy, University of California, UC Irvine: 92697, Irvine, USA

Accepted 2015 July 1. Received 2015 May 29; in original form 2015 February 25

## ABSTRACT

We investigate the evolution of the  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  luminosity functions from  $z \sim 0.8$  to  $\sim 5$  in four redshift slices per emission line using data from the High- $z$  Emission Line Survey (HiZELS). This is the first time that the  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  luminosity functions have been studied at these redshifts in a self-consistent analysis. This is also the largest sample of  $[O\text{II}]$  and  $H\beta + [O\text{III}]$  emitters (3475 and 3298 emitters, respectively) in this redshift range, with large comoving volumes  $\sim 1 \times 10^6 \text{ Mpc}^{-3}$  in two independent volumes (COSMOS and UDS), greatly reducing the effects of cosmic variance. The emitters were selected by a combination of photometric redshift and colour–colour selections, as well as spectroscopic follow-up, including recent spectroscopic observations using DEIMOS and MOSFIRE on the Keck Telescopes and FMOS on Subaru. We find a strong increase in  $L_*$  and a decrease in  $\phi_*$  for both  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters. We derive the  $[O\text{II}]$  star formation history of the Universe since  $z \sim 5$  and find that the cosmic star formation rate density (SFRD) rises from  $z \sim 5$  to  $\sim 3$  and then drops towards  $z \sim 0$ . We also find that our star formation history is able to reproduce the evolution of the stellar mass density up to  $z \sim 5$  based only on a single tracer of star formation. When comparing the  $H\beta + [O\text{III}]$  SFRDs to the  $[O\text{II}]$  and  $H\alpha$  SFRD measurements in the literature, we find that there is a remarkable agreement, suggesting that the  $H\beta + [O\text{III}]$  sample is dominated by star-forming galaxies at high- $z$  rather than AGNs.

**Key words:** galaxies: evolution – galaxies: high-redshift – galaxies: luminosity function, mass function – cosmology: observations.

## 1 INTRODUCTION

Our understanding of the mass assembly and star formation processes of the Universe has improved greatly over the past few decades (for in-depth reviews, see Kennicutt & Evans 2012 and Madau & Dickinson 2014). We currently have evidence to show that the cosmic star formation rate (SFR) peaked at  $z > 1$  and that about half of the current stellar mass density (SMD) had been assembled by that time (e.g. Lilly et al. 1996; Hopkins &

Beacom 2006; Karim et al. 2011; Sobral et al. 2013). However, many open questions remain. How fast did the star formation rate density (SFRD) drop at  $z > 2$ ? How has the population of star-forming galaxies changed over cosmic time? How does the evolution depend on the environment over cosmic time? To answer these questions, it is imperative that we use samples of star-forming galaxies that are low in contaminants and are well defined in terms of selection methodology.

There are many different star formation indicators and calibrations in the literature. Each indicator traces the star formation activity in galaxies independently and with different time-scales. The ultraviolet (UV) light from bright, young stars with masses

\* E-mail: akhos006@ucr.edu (AAK); sobral@strw.leidenuniv.nl (DS)

$> 5 M_{\odot}$  traces the bulk of the young population with time-scales of  $\sim 100$  Myr. UV light of stars with masses  $> 10 M_{\odot}$  ionize the gas along the line of sight, resulting in the absorption and then re-emission of photons seen as nebular (e.g. Lyman, Balmer, and Paschen series of the hydrogen atom) and forbidden emission lines (e.g.  $[O\text{III}]$  and  $[O\text{II}]$ ). The lifetimes for stars capable of ionizing the surrounding gas and dust to form the nebular and forbidden emission lines are on the scale of  $\sim 10$  Myr, allowing for the measurement of the instantaneous SFR. Other indicators include far-infrared emission coming from the heating of dust shrouding the hot, UV bright, young stars and the synchrotron emission in the radio coming from accelerated electrons in supernovae. For an in-depth review of the various indicators and calibrations, we refer the reader to reviews in the literature (e.g. Kennicutt 1998; Calzetti 2013).

Despite the different indicators that exist, one cannot say that only one of the indicators is the ‘holy grail’ of measuring the SFR of star-forming galaxies. However, using different tracers to map out the evolution of the cosmic SFR history is not the best solution either. This is because evolutionary studies based on samples selected in different ways and using different indicators/calibrations at different redshifts will be susceptible to complicated, strong biases and selection effects, which results in a significant scatter when combining all of them to probe the evolution of the cosmic SFR. Another issue is that most studies do not probe sufficiently large volumes to overcome the effects of cosmic variance. Furthermore, the effects of correcting for dust extinction, especially for UV and optical studies, can result in large systematic uncertainties. One requires an indicator that can be used to probe from the low- $z$  to the high- $z$  Universe using a robust and consistent methodology to reduce the effects and biases that come from making assumptions and differing selection techniques.

Emission lines observed using narrow-band imaging techniques can provide an accurate and reliable sample of star-forming galaxies (e.g. Bunker et al. 1995; Fujita et al. 2003; Glazebrook et al. 2004; Ly et al. 2007; Geach et al. 2008; Villar et al. 2008; Sobral et al. 2013). The methodology utilizes two different images of the same field: one being from a broad-band filter and the other being from a corresponding narrow-band filter. The narrow-band image is dominated by emission line galaxies and the continuum, while the broad-band image is dominated by the continuum with a small contribution from the emission line. When the two are subtracted, the result is the removal of the continuum and an image of galaxies with emission lines. The advantage of narrow-band imaging surveys is that they allow for the selection of emitters with a clean selection function by emission line flux and within a known narrow redshift range. This is because the filter width is quite narrow, such that any source brighter than expected from its broad-band magnitude is an emitter. Emission line surveys via grism spectroscopy on the *HST* (e.g. Colbert et al. 2013) are also great accompaniments to narrow-band studies, as they are area-limited (area of the grism) while emission line surveys are redshift-limited.

Most narrow-band surveys have focused on  $H\alpha$  (e.g. Tresse et al. 2002; Fujita et al. 2003; Pascual 2005; Ly et al. 2007, 2011; Sobral et al. 2009, 2013) as it is a reliable star formation indicator which is well calibrated in the local universe and is only mildly affected by dust attenuation. The latest results of the High- $z$  Emission Line Survey (HiZELS; Geach et al. 2008; Sobral et al. 2009, 2012, 2013) have robustly traced the evolution of the cosmic SFR up to  $z \sim 2$ . This is the maximum redshift that  $H\alpha$  surveys can probe from the ground, as at higher redshifts  $H\alpha$  falls into the mid-IR and is blocked by water vapour and carbon dioxide in the atmosphere. To probe to higher- $z$  using the same narrow-band technique would require

another emission line. The other major emission lines associated with star-forming galaxies are  $H\beta$ 4861,  $[O\text{III}]$ 4959,  $[O\text{III}]$ 5007, and  $[O\text{II}]$ 3727 which can be probed up to  $z \sim 3$  for  $H\beta + [O\text{III}]$  and up to  $z \sim 5$  for  $[O\text{II}]$ .

In the past decade, several  $H\beta$ ,  $[O\text{III}]$ , and  $[O\text{II}]$  studies have been carried out (e.g. Hammer et al. 1997; Hogg et al. 1998; Gallego et al. 2002; Hicks et al. 2002; Teplitz et al. 2003; Ly et al. 2007; Takahashi et al. 2007; Bayliss et al. 2011, 2012; Sobral et al. 2012; Ciardullo et al. 2013, Drake et al. 2013), the majority of which had small sample sizes and, hence, suffered from cosmic variance biases. The majority observed up to  $z \sim 1$ , while only the works of Bayliss et al. (2011) and Bayliss et al. (2012) measured the  $[O\text{II}]$  SFR densities at  $z \sim 2$  and  $z \sim 4.6$ , respectively. Both these works used small samples, with the  $z \sim 4.6$  measurement having a sample size of only 3  $[O\text{II}]$  emitters, greatly limiting any conclusion.

This paper presents, for the first time, the luminosity functions (LFs) of  $H\beta + [O\text{III}]$ <sup>1</sup> and  $[O\text{II}]$  emitters up to  $z \sim 5$  using the reliable selection techniques of Sobral et al. (2009, 2012, 2013) on the combined COSMOS and UDS narrow-band publicly available catalogues<sup>2</sup> of HiZELS. The sample probes comoving volumes of up to  $\sim 1 \times 10^6$  Mpc<sup>3</sup>, which greatly reduces the effects of cosmic variance. This is also the largest sample of  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emission line galaxies up to  $z \sim 5$  to date in the literature and is used to effectively and robustly probe the evolution of the cosmic SFR density.

This paper is organized as follows: in Section 2, we outline the photometric and spectroscopic data sets that we use and the methodology utilized to effectively select  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters. In Section 3, we outline our volume calculations and the completeness and filter profile corrections. Also in Section 3 are the results of the LFs. In Section 4, we present the results and discuss our cosmic SFRDs and the evolution of the SMD based on  $[O\text{II}]$  emitters. Section 5 outlines the conclusion of this work and is followed by Appendix A, which presents our colour–colour selection criteria, and Appendix B, which presents our binned LF data points, and Appendix C, which presents our SFR density compilation.

Throughout this paper, we assume  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology with  $H_0 = 70$  km s<sup>-1</sup> Mpc<sup>-1</sup>,  $\Omega_m = 0.3$ , and  $\Omega_{\Lambda} = 0.7$  with all magnitudes presented as AB and the initial mass function is assumed to be a Salpeter initial mass function (IMF).

## 2 DATA

### 2.1 Selection catalogues

Our data consist of narrow-band and broad-band photometric data of the UDS (Lawrence et al. 2007) and COSMOS (Scoville et al. 2007) fields and spectroscopic follow-ups that are described in Section 2.4. The catalogues that are described below are taken from Sobral et al. (2013) and are publicly available.

The narrow-band catalogues are from HiZELS (Geach et al. 2008; Sobral et al. 2009, 2012, 2013). This project utilizes the

<sup>1</sup> Because the  $H\beta$  and  $[O\text{III}]$  emission lines are close to each other, photo- $z$  and colour–colour selections cannot distinguish between them. The best way to fully differentiate the two is via spectroscopy. Based on line ratio studies, we can argue that most of the emitters will be  $[O\text{III}]$  emitters, but to ensure we are not biasing our measurements based on such assumptions, we present the results as the combined measurement of  $H\beta + [O\text{III}]$ .

<sup>2</sup> The narrow-band catalogues are available on Vizier and are from Sobral et al. (2013).

narrow-band  $J$ ,  $H$ , and  $K$  filters of the Wide Field CAMera on the United Kingdom Infrared Telescope (UKIRT) and the NB921 filter of the Suprime-Cam on the Subaru Telescope. Previous uses of this data focused primarily on  $H\alpha$  emitting galaxies and their properties (e.g. Garn et al. 2010; Sobral et al. 2010, 2011, 2013; Geach et al. 2012; Swinbank et al. 2012a,b; Stott et al. 2013a,b; Darvish et al. 2014), but HiZELS is able to pick up more than just this emission line as it can detect any line above the flux limit. In this paper, we focus on the  $H\beta + [O III]$  and  $[O II]$  emitting galaxies found by HiZELS. The broad-band catalogues are from COSMOS (Scoville et al. 2007) and the DR8 release of the Subaru-*XMM*-UKIDSS UDS (Lawrence et al. 2007) catalogue.<sup>3</sup>

## 2.2 Narrow-band selection of potential emitters

In this subsection, we will review the methodology of selecting potential emitters followed by Sobral et al. (2013). We refer the reader to this paper for a detailed overview.

Potential emitters were selected by their colour excess in terms of the parameter  $\Sigma$  (Bunker et al. 1995; Sobral et al. 2012), which quantifies the significance of the excess of a source with respect to the random scatter expected for a source to have a colour excess  $> 0$  in terms of their narrow-band magnitudes:

$$\Sigma = \frac{1 - 10^{-0.4(BB-NB)}}{10^{-0.4(ZP-NB)} \sqrt{\pi r_{ap}^2 (\sigma_{NB}^2 - \sigma_{BB}^2)}}, \quad (1)$$

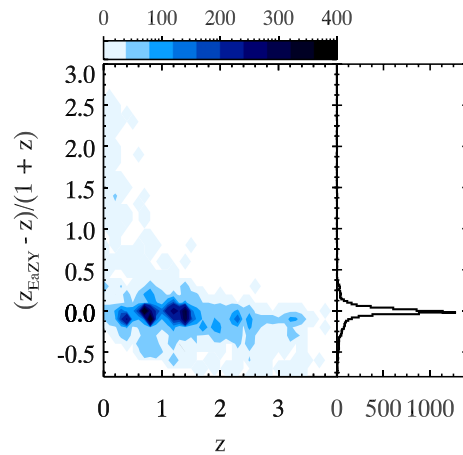
where BB and NB are the broad-band and narrow-band magnitudes, respectively,  $r_{ap}$  is the aperture radius in pixels, and  $\sigma_{BB}$  and  $\sigma_{NB}$  are the rms per pixel for the broad-band and narrow-band, respectively (see e.g. Sobral et al. 2012, 2013).

Emitters are selected on the basis that they have  $\Sigma > 3$  and a rest-frame equivalent width  $EW_0 > 25 \text{ \AA}$ . The second condition ensures that we select sources with a significant colour excess for bright narrow-band magnitudes (see fig. 3 of Sobral et al. 2013).

## 2.3 Photo- $z$ measurements

We initially used the photo- $z$  measurements from Ilbert et al. (2009) and Cirasuolo et al. (2007) for COSMOS and UDS, respectively, that were provided in the corresponding catalogues. The main problem of using those measurements is that in the UDS catalogue, more than 60 per cent of sources are without photometric redshifts. This raises issues when selecting emitters via redshift selection. Although the colour-colour selection is effective alone in selecting  $H\beta + [O III]$  and  $[O II]$  emitters (see below and appendices for discussion), we wish to have emitters selected by two independent methods to act as a check-and-balances to robustly select emitters. Therefore, we measured the photometric redshifts for all of the UDS and COSMOS narrow-band excess sources, using EaZY (Brammer, van Dokkum & Coppi 2008) to ensure that (1) the majority of the UDS catalogue had reliable photometric redshifts, and (2) that both COSMOS and UDS had their redshifts determined by the same code, models, and assumptions.

<sup>3</sup> We refer the reader to the respective papers for further details on the creation of these catalogues. We also refer the reader to the UKIDSS and COSMOS websites for further information of the multiwavelength photometric and spectroscopic data sets. (COSMOS: <http://irsa.ipac.caltech.edu/data/COSMOS>; UDS: <http://www.ukidss.org>.)



**Figure 1.** The 2D density distribution of the errors between the redshifts determined by EaZY and those of Ilbert et al. (2009, COSMOS, Le Phare) and Cirasuolo et al. (2007, UDS, Hyperz) for all narrow-band excess sources (full catalogue). We find that  $\sim 95$  per cent of our measurements are in agreement with the redshifts from the literature. The median error is measured as 0.038 without sigma-clipping. There are overdensities at  $z \sim 0.4, 0.8$ , and  $1.5$  that conform to redshifts for our major emission lines. We find outliers with errors up to  $\sim 2.5$ , but these are only 4 per cent of our sample. In comparison to the wealth of spectroscopic data, we find a median photometric redshift error of  $\Delta z / (1 + z_{spec}) = 0.047$  in comparison to spectroscopic redshifts.

The filters we use for measuring photometric redshifts for our UDS sources are *UBVRizYJHK* + *Spitzer* IRAC Ch1–4 + narrow-band (NB) + broad-band (BB) filters. For our COSMOS sources, we combine our *UBVgRizJK* + *GALEX* FUV and NUV + *Spitzer* IRAC Ch1–4 + NB + BB catalogue with that of Ilbert et al. (2009). The benefit of this is that we include 12 Intermediate Subaru (e.g. IA427, IA464) bands and one Subaru NB711 band that were within the Ilbert et al. (2009) catalogue. This results in a total of 29 filters to constrain the measurements. Furthermore, the benefit of using the narrow-band filters in measuring the redshifts is that these filters specifically capture emission lines which are inherent in the SEDs. The Pegase13 spectral library that is used in EaZY includes a prescription for emission lines, which makes the measurements more accurate when including the narrow-band filters. Fig. 2 shows the benefit of using narrow-band filters in the fitting process by the sharp peaks for which we expect the major emission lines to be located in terms of redshift.

Fig. 1 shows the 2D density distribution of  $\sigma_z = (z_{EaZY} - z)/(1 + z)$ , where  $z$  are the redshifts measured by Ilbert et al. (2009) and Cirasuolo et al. (2007), and the original photometric redshifts in the catalogue. We find a median error of 0.038 (all sources without sigma-clipping). Fig. 1 also shows overdensities at  $z \sim 0.4, 0.8$ , and  $1.5$  which are expected as these are the most populated redshift slices in our sample, since they conform to major emission lines. We find outliers up to  $\sigma_z \sim 2.5$ , but these only constitute a small fraction, such that  $\sim 95$  per cent of our sample are in agreement with Ilbert et al. (2009) and Cirasuolo et al. (2007).

## 2.4 Spectroscopic redshifts

We make use of the vast array of spectroscopic observations from the literature, which greatly enhances the reliability of our sample. In the COSMOS catalogue, spectroscopic measurements are drawn from various studies as listed on the COSMOS website, as well as the  $z$ COSMOS measurements from Lilly et al. (2007). The

UDS catalogue also includes measurements from various publications that are highlighted on the UKIDSS UKIRT website, including the UDSz survey (Bradshaw et al. 2013; McLure et al. 2013). We also include FMOS measurements from Stott et al. (2013b), DEIMOS and MOSFIRE measurements from Nayyeri et al. (in preparation), PRIMUS measurements from Coil et al. (2011), and VIMOS Public Extragalactic Redshift Survey (VIPERS) measurements from Garilli et al. (2014). In total, we have 1269 emitters that have spectroscopic redshifts with 661, 350, 177, and 81 emitters in NB921, NB<sub>J</sub>, NB<sub>H</sub>, and NB<sub>K</sub>, respectively. This allows us to enhance the reliability of our sample and to test our photo- $z$  and colour-colour selections. In comparison to the spectroscopic redshifts, we have assessed the median errors of our photometric redshifts to be  $\Delta z/1 + z_{\text{spec}} = 0.047$  (without sigma-clipping).

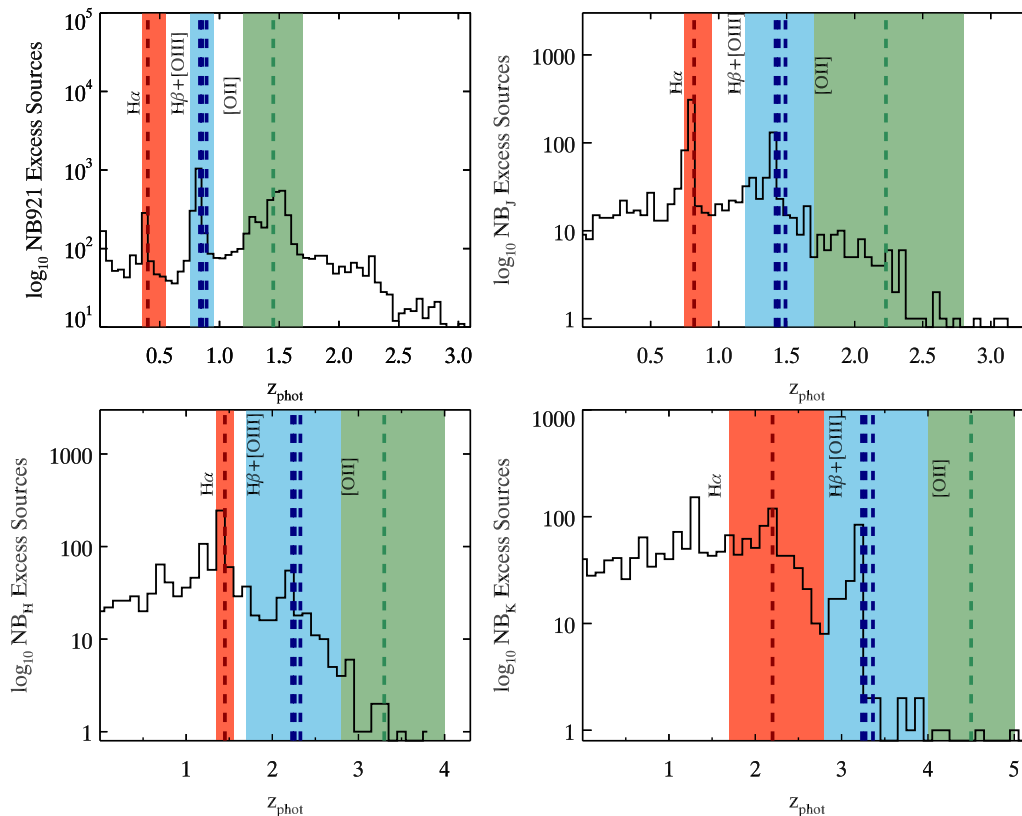
### 2.5 Selection of $H\beta + [O\text{III}]$ and $[O\text{II}]$ emitters

The selection of potential  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters is done by a combination of three different methods: (1) photometric redshift; (2) colour-colour; and (3) spectroscopic redshift. In this section, we will present the general selection criteria that we used to select our sample. For more detailed information about the specific selection cuts applied in each case, we refer the reader to Appendix A.

With three different selection methods, conflicts can arise where one selection method provides a result that conflicts with another method. To solve the issue, we prioritize the selection methods

as such: (1) spectroscopic redshifts, (2) photometric redshifts, and (3) colour-colour. If the emitter has a spectroscopic redshift, then it is selected based only on measurement. If it does not have a spectroscopic redshift, then we select it based on its photometric redshift. Lastly, if the emitter has a photometric redshift but is not within the range of being selected as  $H\alpha$  (see photo- $z$  selection criteria in Sobral et al. 2013),  $H\beta + [O\text{III}]$ , or  $[O\text{II}]$  or if the emitter does not have photo- $z$  measurements, then it is selected based on the colour-colour criterion. In most cases, we find that emitters with photo- $z$  within the redshift selection range are also found within the colour-colour selection area. In such cases, the emitters are selected based on both selection methods.

As shown in Fig. 2, the local peaks in the photometric redshift distributions are located around the expected redshifts for  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters. This signifies that we have many  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters in our sample. We select our emitter candidates by defining a range in the distribution of  $z_{\text{phot}}$  that is centred on the expected redshift of the emission line, which is aligned with the peaks in Fig. 2. In most cases, the number of  $H\alpha$  emitters are the largest, followed by  $H\beta + [O\text{III}]$  and  $[O\text{II}]$ . In NB921 (Fig. 2),  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  lines are the strongest, respectively, as these redshifts are near the peak of the cosmic star formation history and also probe a much larger cosmic volume than  $H\alpha$ . Other populations of emission lines are found, such as Paschen series lines, He I, and  $[S\text{III}]$  (Fig. 2).



**Figure 2.** Log-scale photometric redshift distributions of the emitters selected in NB921(upper left), NB<sub>J</sub>(upper right), NB<sub>H</sub>( lower left), and NB<sub>K</sub>(lower right). Each main peak is associated with a strong emission line, specifically  $H\alpha$ ,  $H\beta + [O\text{III}]$ , and  $[O\text{II}]$ . The dashed red line is the expected redshift of  $H\alpha$  emitters. The dashed blue lines are the expected redshifts for  $H\beta + [O\text{III}]$  emitters corresponding to  $H\beta 4861$ ,  $[O\text{III}] 4959$ , and  $[O\text{III}] 5007$ . It is clear then that differentiating these lines, even in narrow-band surveys, is quite difficult due to their close proximity to each other. Lastly, the dashed green line is the expected redshift of  $[O\text{II}]$  emitters. Highlighted in the corresponding, but lighter colours, are the photo- $z$  selection regions. The  $H\alpha$  selection region is from Sobral et al. (2013). The photo- $z$  distributions are from our photo- $z$  calculations using EaZY for the COSMOS and UDS fields.



**Table 1.** The number of emitters selected based on colour–colour, photo- $z$ , spec- $z$ , and dual/multi-emitters. We also include the total number of emitters found. It should be noted that an emitter can be selected by more than just one selection, such that by tallying up the number of emitters per column will result in a number larger than the number in the column that shows the total number of emitters. We also show the number of sources selected only based on colour–colour and only based on photo- $z$ . This highlights the importance of having more than one selection technique. For example, if we solely relied on colour–colour selection, then we would have a loss of 21 per cent in our  $z = 1.42$   $H\beta + [O\text{III}]$  sample. Furthermore, we also include the number of emitters selected as dual/multi-emitters. These are sources that were detected in more than one narrow-band, resulting in two or more detected emission lines complementing each other (e.g.  $H\alpha$  in NBH and  $[O\text{II}]$  in NB921 for  $z \sim 1.47$ ).

Emission line	Band	$z$	Colour–colour	CC Only	Photo- $z$	Photo- $z$ only	Spec- $z$	Dual/Multi-emitter	Total
$H\beta + [O\text{III}]$									
	NB921	0.84	2005	1000	1262	257	213	160	2477
	NBJ	1.42	277	41	314	78	15	23	371
	NBH	2.23	208	52	212	56	3	44	271
	NBK	3.24	145	11	158	24	2	0	179
$[O\text{II}]$									
	NB921	1.47	3152	957	2211	16	97	213	3285
	NBJ	2.23	115	51	85	21	0	6	137
	NBH	3.30	29	18	16	5	1	0	35
	NBK	4.70	18	14	4	0	0	0	18

Colour–colour selections are also applied for each redshift. The selection criteria used for  $z \sim 1-3$  are from Sobral et al. (2013) as these are in perfect agreement with our large spectroscopic sample, while the criteria used for the  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  at  $z = 3.3$  and  $[O\text{II}]$  at  $z = 4.5$  are based on the dropout colour–colour selection known as the Lyman break technique (Dickinson 1998; Stark et al. 2009). For some redshifts, we use more than one colour–colour selection to reduce the contaminations from lower and higher- $z$  sources. All colour–colour selection definitions are found in Table A1.

Whenever available, we select sources based on their spectroscopic redshifts. Sources for which the spectroscopic redshift contradicts the photo- $z$  and colour–colour selection are removed from the sample. By including spectroscopically confirmed sources, we increase the size of our sample. We also note that we use all our spectroscopic redshifts to confirm the robustness of our photo- $z$  and colour–colour selections.

Also in the selection process is selecting dual/multi-emitters. For some sources that are selected as emitters with some emission line, there also exists another emission line in another band. For example, an emitter that is selected in NBJ as an  $H\beta + [O\text{III}]$  emitter ( $z = 1.47$ ) can also be selected in NB921 as an  $[O\text{II}]$  emitter if observed in that narrow-band. Sobral et al. (2012) used this same technique in a double-blind study to find  $[O\text{II}]$  emitters in NB921 by using selected  $H\alpha$  emitters in NBH as a proxy. The benefit to this technique is that it confirms emitters if they are selected in at least two bands; corresponding to getting a redshift out of two emission lines. In cases where we find dual/multi-emitters, we treat them as spectroscopic measurements as it is similar to having a spectroscopically confirmed emitter and include them in the sample.

One major source of contamination that we may have is having selected a source as an  $[O\text{II}]$  emitter when it is an  $H\beta + [O\text{III}]$  emitter and vice versa. Also, there are situations where a source is selected as one of the emitters of interest, but also falls into the colour–colour selection for another emission line, or even  $H\alpha$  (as these colour–colour selections were used by Sobral et al. (2009, 2012, 2013) to find such emitters). To overcome this degeneracy, we look at the photo- $z$  distributions shown in Fig. 2 as a probability distribution to assign the emission line based on the most probable

line in the sample. In all cases, except for NB921,  $H\alpha$  is the most probable line. For NB921,  $H\beta + [O\text{III}]$  is the most probable.

Another source of contamination is from misidentified lines. From the wealth of spectra that we have, the majority of misidentified lines are  $H\alpha$ . Further details on the types of misidentified lines can be found in Appendix A. Our total sample size is outlined in Table 1.

### 3 LUMINOSITY FUNCTIONS AND EVOLUTION

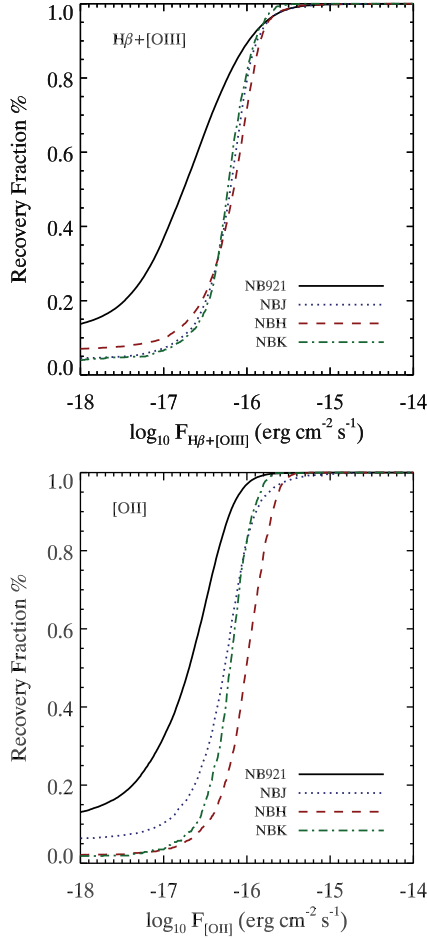
We use the traditional  $V_{\text{max}}$  estimator to create our binned data. The binned data is defined such that:

$$\phi(L_j) = \frac{1}{\Delta \log L_{b,j}} \sum_{\log L_{c,j} - \Delta \log L_{b,j}/2}^N \frac{1}{C(L_i)V_{\text{max},i}}, \quad (2)$$

where  $L_j$  is the  $j$ th luminosity bin,  $\Delta \log L_{b,j}$  is the bin-size,  $\log L_{c,j}$  is the central luminosity of the  $j$ th bin,  $C(L_i)$  being the completeness of the  $i$ th source described in Section 3.1, and  $V_{\text{max},i}$  being the volume for which that source may be detected as described in Section 3.2.

#### 3.1 Line completeness

To assess the completeness, we follow the methodology proposed in Sobral et al. (2012, 2013). We start with the full catalogue that has all the emitters and includes sources which did not make our emitter selection. To measure the recovery fraction based on the emission line flux, we input a mock line flux starting at  $10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$  to all sources in the catalogue. We then apply the excess selection criteria ( $\Sigma$  and EW cuts) used in Sobral et al. (2013) followed by our colour–colour selections. The recovery fraction is then defined as the number of sources recovered divided by the total number of sources in the catalogue. This is then repeated after increasing the input mock line flux by small increments. The expected result is that at  $10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$  the recovered fraction will be low and will increase as the input line flux increases. Fig. 3 shows the average completeness determined for our  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  sources in the different narrow bands. The advantage of this technique is that we are not limiting our determination of the completeness to a



**Figure 3.** Average line completeness for the entire sample per band. Note that the completeness does vary between the COSMOS and UDS fields, as well as from image-to-image within each field. This is because each image has a different depth. We compute the completeness based on each image to account for this discrepancy.

certain model, but actually using the observed data itself to get the completeness correction.

There are some important points to be noted. First, these simulations are run separately per image. This is because the depths of each image are not the same and thus cannot be used together all at once to determine the completeness correction. Secondly, we apply an uncertainty of 20 per cent of the completeness correction to the other uncertainties in quadrature in order to take into account the errors associated with this method of determining the corrections.

### 3.2 Volumes and filter profile correction

We calculate the volumes assuming a top-hat filter that has the same range as the full width at half-maximum (FWHM) of the actual filter. We report the probed comoving volume per square degree in Table 2. The volumes for each log-luminosity bins in the LFs are reported in Tables B1 and B2 for  $H\beta + [O III]$  and  $[O II]$ , respectively.

Although a top-hat filter makes our calculations easy, it is not a true representation of the throughput of the filter, which requires us to apply a filter profile correction. There are two main effects that the filter profile correction takes into account. The first is the flux loss due to emitters that are close to the edge of the filter’s FWHM. Bright emitters at the wings will have a significant flux loss (close to 40 per cent; depends on the filter) and would be considered as a faint source. This gives an overall bias in our sample population of faint sources and a lack of bright sources. The second effect is the volumes are corrected for the bright sources. Any faint source that is close to the wings of the filter will most likely not be in our sample, but bright sources will be detected as faint emitters. This then implies that our bright emitters cover a wider range of the filter, meaning a wider range in redshift, and, therefore, a larger volume.

We use the method proposed in Sobral et al. (2009, 2012, 2013). We correct for the filter by creating a mock sample of  $10^5$  fake emitters based on the LF with the assumption of a top-hat filter. Random redshifts are assigned to each source and in a range covering the full filter profile, but not large enough that evolutionary effects of the LF and cosmological structure biases the results. We assume a uniform redshift distribution. This mock sample will have the same distribution as the input top-hat LF which we define as  $\phi_{TH}$ . We then make a second mock sample with the same top-hat LF but now apply the actual filter. This is done by

$$L_{corr} = \frac{\int_{z_1}^{z_2} L_{in}(z)T(z)dz}{\int_{z_1}^{z_2} T(z)dz}, \quad (3)$$

where  $T(z)$  is the filter-response function in terms of redshift and  $L_{in}(z)$  is the luminosity of the emitter which is defined as  $L_{in}(z) = L\delta(z_{rand} - z_{filter})$ , where  $z_{rand}$  is the randomly assigned redshift and  $z_{filter}$  is the matching redshift of the filter. This results in the loss of sources as some sources will have a redshift outside the range of the filter’s FWHM. The LF from this population is defined as  $\phi_{filter}$  and is compared to  $\phi_{TH}$  in order to get the filter profile correction factor. The result shows that the bright sources are underestimated as expected, which changes the shape of the final LF slightly. Specifically, it decreases the faint-end slope and increases the  $L_*$  and  $\phi_*$  slightly. This correction is applied to the LFs by dividing all the binned  $\Phi(L)$  data points by the correction factor.

**Table 2.** A list of the narrow-band filters used (Sobral et al. 2013), along with the central wavelength ( $\mu\text{m}$ ) and the FWHM ( $\text{\AA}$ ) of each filter. Included is the expected redshift of each emission line within the range of the FWHM of the filter and the comoving volume that is observed.

Filter	$\lambda_{obs}$ ( $\mu\text{m}$ )	FWHM ( $\text{\AA}$ )	$H\beta + [O III]5007$		$[O II]3727$	
			$z$	Volume ( $10^5 \text{ Mpc}^3 \text{ deg}^{-2}$ )	$z$	Volume ( $10^5 \text{ Mpc}^3 \text{ deg}^{-2}$ )
NB921	0.9196	132	$0.84 \pm 0.01$	1.79	$1.47 \pm 0.02$	3.75
NBJ	1.211	150	$1.42 \pm 0.01$	3.11	$2.25 \pm 0.02$	4.83
NBH	1.617	211	$2.23 \pm 0.02$	5.05	$3.34 \pm 0.03$	6.53
NBK	2.121	210	$3.24 \pm 0.02$	4.87	$4.69 \pm 0.03$	5.68

**Table 3.** The luminosity function parameters and derived properties for all  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters. Errors in  $\phi_*$  and  $L_*$  are computed by running a Monte Carlo simulation, displacing all the measurements of  $\Phi_{\text{final}}$  by  $1\sigma$ . The faint-end slope,  $\alpha$ , was fixed as our data do not go faint enough to constrain the faint-end properly. The luminosity density,  $\rho_L$ , was calculated by taking the infinite integral of the LF. The SFRDs were calculated based on the Kennicutt (1998) calibration ( $[O\text{II}]$ ) and Osterbrock & Ferland (2006) ( $H\beta + [O\text{III}]$ ) with  $\dot{\rho}_{*,\text{comp}}$  being the completeness and filter profile corrected SFRD measurement,  $\dot{\rho}_{*,\text{corr}}$  is the completeness + filter profile + dust corrected SFRD measurement,  $\dot{\rho}_{*,\text{AGN-corr}}$  is the completeness + filter profile + dust corrected + AGN corrected SFRD measurement. We show every measurement as  $\dot{\rho}_{*,\text{comp}}$  is the most robust measurement, while the dust and dust + AGN corrected measurements are based on assumptions regarding line ratios, dust extinction laws, and AGN selection methods.

$z$	$\log_{10}\phi_*$ ( $\text{Mpc}^{-3}$ )	$\log_{10}L_*$ ( $\text{erg s}^{-1}$ )	$\alpha$	$\log_{10}\rho_L$ ( $\text{erg s}^{-1} \text{Mpc}^{-3}$ )	$\log_{10}\dot{\rho}_{*,\text{comp}}$ ( $\text{M}_{\odot} \text{yr}^{-1} \text{Mpc}^{-3}$ )	$\log_{10}\dot{\rho}_{*,\text{corr}}$ ( $\text{M}_{\odot} \text{yr}^{-1} \text{Mpc}^{-3}$ )	$\log_{10}\dot{\rho}_{*,\text{AGN-corr}}$ ( $\text{M}_{\odot} \text{yr}^{-1} \text{Mpc}^{-3}$ )
<b><math>H\beta + [O\text{III}]</math> luminosity function properties</b>							
0.84	$-2.55^{+0.04}_{-0.03}$	$41.79^{+0.03}_{-0.05}$	-1.60	39.58	$-1.549^{+0.01}_{-0.02}$	$-1.009^{+0.01}_{-0.02}$	$-1.062^{+0.03}_{-0.03}$
1.42	$-2.61^{+0.10}_{-0.09}$	$42.06^{+0.06}_{-0.05}$	-1.60	39.80	$-1.333^{+0.05}_{-0.04}$	$-0.793^{+0.05}_{-0.04}$	$-0.882^{+0.06}_{-0.05}$
2.23	$-3.03^{+0.21}_{-0.26}$	$42.66^{+0.13}_{-0.13}$	-1.60	39.98	$-1.159^{+0.10}_{-0.11}$	$-0.619^{+0.10}_{-0.11}$	$-0.766^{+0.11}_{-0.12}$
3.24	$-3.31^{+0.09}_{-0.26}$	$42.83^{+0.19}_{-0.17}$	-1.60	39.87	$-1.265^{+0.10}_{-0.09}$	$-0.725^{+0.10}_{-0.09}$	$-0.873^{+0.11}_{-0.10}$
<b><math>[O\text{II}]</math> luminosity function properties</b>							
1.47	$-2.25^{+0.04}_{-0.04}$	$41.86^{+0.03}_{-0.03}$	-1.30	39.72	$-1.132^{+0.02}_{-0.02}$	$-0.884^{+0.02}_{-0.02}$	$-0.973^{+0.04}_{-0.04}$
2.25	$-2.48^{+0.08}_{-0.09}$	$42.34^{+0.04}_{-0.03}$	-1.30	39.98	$-0.878^{+0.05}_{-0.06}$	$-0.630^{+0.05}_{-0.06}$	$-0.723^{+0.06}_{-0.07}$
3.34	$-3.07^{+0.63}_{-0.70}$	$42.69^{+0.31}_{-0.23}$	-1.30	39.74	$-1.118^{+0.43}_{-0.20}$	$-0.870^{+0.43}_{-0.20}$	$-0.964^{+0.43}_{-0.20}$
4.69	$-3.69^{+0.33}_{-0.28}$	$42.93^{+0.18}_{-0.24}$	-1.30	39.35	$-1.502^{+0.10}_{-0.10}$	$-1.255^{+0.10}_{-0.10}$	$-1.348^{+0.11}_{-0.11}$

### 3.3 Luminosity function fitting

There are several different functions that have been proposed in the literature to describe the observed LF of the Universe (see Johnston 2011 for an in-depth review). We adopt the most widely accepted Schechter function to fit the observed LF. The Schechter function is defined in its log-form as

$$\Phi(L)dL = \phi_* \ln 10 \left( \frac{L}{L_*} \right)^{1+\alpha} e^{-L/L_*} d \log_{10} L, \quad (4)$$

where  $\phi_*$  is the normalization of the LF,  $L_*$  is the characteristic luminosity, and  $\alpha$  is the faint-end slope.

We fit each LF using the MPFIT fitting routine (Markwardt 2009), which utilizes the Levenberg–Marquardt least-squares minimization technique to find the best fit for a given function. For each fit, we take the best-fitting values as our fitted parameters. For the  $1\sigma$  errors, we run a Monte Carlo simulation, similar to that of Ly et al. (2011). The simulation starts by selecting a random number that is drawn from a normal distribution that will perturb each data point,  $\Phi(L)$ , in the LF within the  $1\sigma$  error bars. We also vary the bin size and centre of the bin by perturbing the original bin size and bin centre by a uniform distribution. The fit is then run again and these steps are repeated for each iteration. A total of  $10^5$  iterations are done to get a probability distribution of the best-fitting values from where the  $1\sigma$  error bars are then calculated.

Our best-fitting values are shown in Table 3. We keep  $\alpha$  fixed to a constant value of  $-1.6$  and  $-1.3$  for  $H\beta + [O\text{III}]$  and  $[O\text{II}]$ , respectively, as we are not able to fully constrain the faint-end. These values are drawn from looking at past work from previous narrow-band studies in order for our results to be comparable (e.g.  $H\beta + [O\text{III}]$ , Colbert et al. 2013;  $[O\text{II}]$ , Bayliss et al. 2011). The drawback to this is that we are using low- $z$  measurements of  $\alpha$  as a proxy for the high- $z$  universe, which can be an incorrect assumption. Low- $z$  studies, such as Colbert et al. (2013), Ly et al. (2007), and Pirzkal et al. (2013) for  $H\beta + [O\text{III}]$  and Ly et al. (2007), Ciardullo et al. (2013), Takahashi et al. (2007), Bayliss et al. (2011), and Sobral et al. (2012) for  $[O\text{II}]$ , have shown that the faint-end slope does not evolve up to  $z \sim 1.5$ , while  $H\alpha$  surveys such as Sobral et al. (2013, 2014) have shown no evolution in the faint-end slope

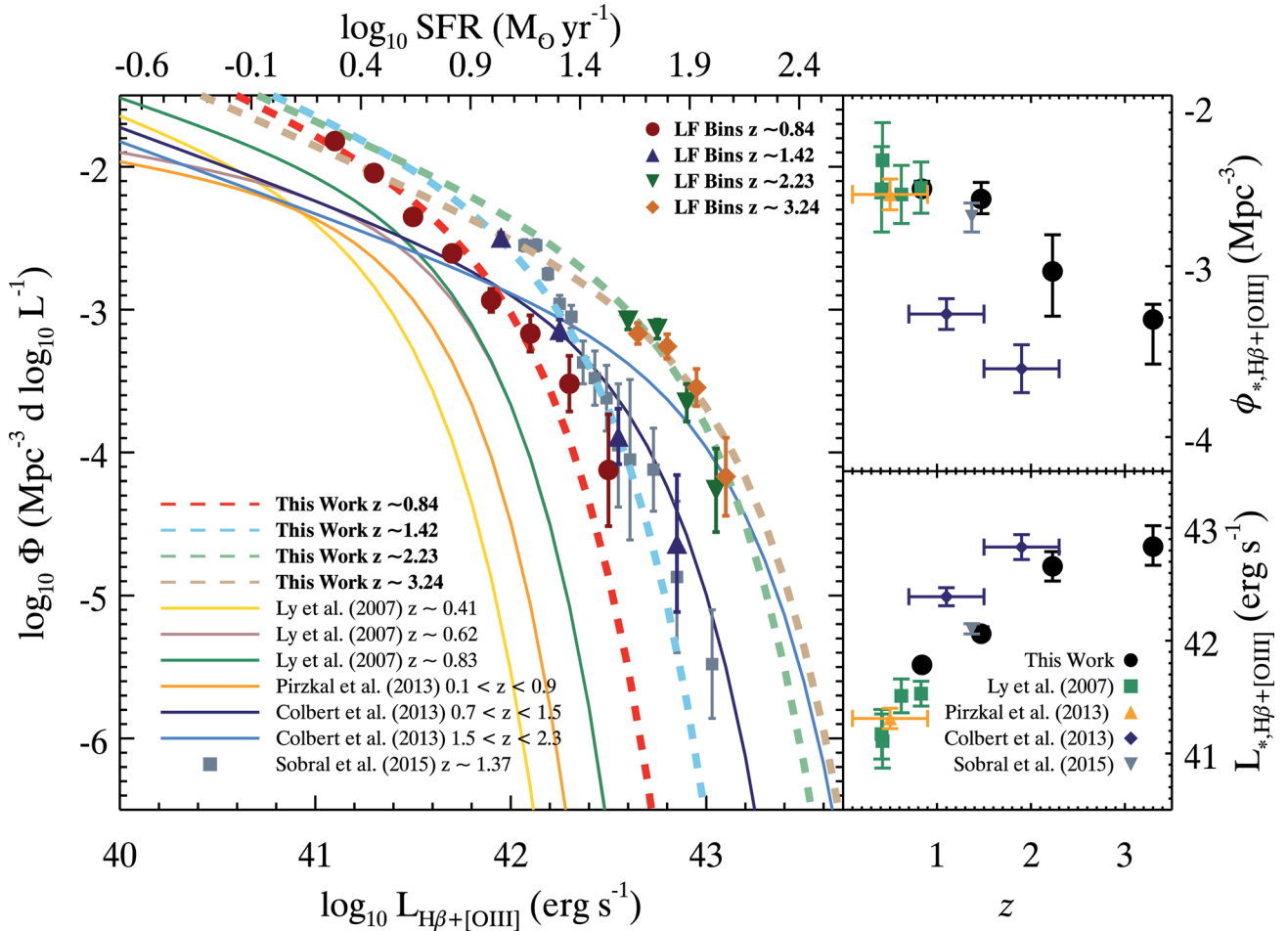
up to  $z \sim 2.23$ . Furthermore, UV studies (e.g. Oesch et al. 2010; Smit et al. 2012), have shown no evolution up to  $z \sim 6-7$ . Based on these results, we keep  $\alpha$  fixed as we constrain the bright-end rather than the faint-end.

Lastly, the LF results are not corrected for dust extinction, except when measuring the SFRDs. This is because many studies in the literature use very different extinction diagnostics, such that it becomes difficult to compare various studies. Furthermore, our knowledge of the role of dust on emission lines for the high- $z$  universe, especially for the emission lines of interest in our study, is still in development and requires future detailed investigations. To simplify the use of our LFs by others in future studies, we present all the LF parameter results as uncorrected for dust and active galactic nuclei (AGN) contribution. When discussing the SFRDs in Section 4, we will include the results with and without dust and AGN corrections.

### 3.4 $H\beta + [O\text{III}]$ luminosity function $z \sim 0.8, 1.5, 2.2$ , and $3.3$

We present here the results of the fitted Schechter function to the  $H\beta + [O\text{III}]$  observed LF out to  $z \sim 3.3$ . This is the highest redshift determination of the  $H\beta + [O\text{III}]$  LFs currently to date and is the first time that the LF has been constrained out to these redshifts. We present the results in Fig. 4. From  $z \sim 0.8$  to  $3.3$ , we see a clear evolution in the shape of the LFs (Fig. 4, left). We also show in Fig. 4 the evolution of  $\phi_*$  and  $L_*$  with  $\alpha$  fixed to  $-1.6$ . It should be noted that there is a degeneracy between the fitted Schechter parameters, as shown in Fig. 7. This needs to be borne in mind when interpreting the evolution of any single parameter, although Fig. 7 indicates that our results are relatively robust. Based on our results, we find that  $\phi_*$  has been decreasing from  $z \sim 0.8$  to  $\sim 3.3$ . The opposite trend is seen in  $L_*$  where it is increasing from  $z \sim 0.8$  to  $\sim 3.3$ .

Our results show a clear evolution in the LF and are consistent with the same evolution seen in the results from the literature (Ly et al. 2007; Colbert et al. 2013; Pirzkal et al. 2013; Sobral et al. 2015). We report our LF parameters in Table 3. We find that our  $z \sim 1.47$  and  $2.23$  LF agrees well with the LFs of Colbert et al. (2013) in the bright-end, but diverges at the faint-end. This matches



**Figure 4.** Left: presented are our  $H\beta + [O III]$  LFs along with LFs from the literature. Included on the top horizontal axis is the  $\log_{10}$  SFR that was derived via the Osterbrock & Ferland (2006) calibration (see Section 4.2). The darker data points are colour coded to match the lighter LF fit (dashed lines). There is a clear evolution in the LFs up to  $z \sim 3$ . We find that our  $z = 1.42$  and  $2.23$  LFs are in reasonable agreement with the  $[O III]$  grism spectroscopy study of Colbert et al. (2013) at the bright-end, suggesting that we are selecting a reliable sample of  $[O III]$  emitters. The major difference between our  $z = 0.84$  and the  $z = 0.83$  LF of Ly et al. (2007) is probably due to sample size biases. Our sample is much larger, hence we are able to populate our brightest bins, causing a shift in  $L_*$  to higher luminosities. Top right: the evolution of  $\phi_*$  from the  $H\beta + [O III]$  LF. A strong, decreasing evolution is seen in  $\phi_*$  from  $z = 0$  to  $z \sim 3$ . This same evolution is seen by UV LF studies (see Oesch et al. 2010 for details). Bottom right: the evolution of  $L_*$  from the  $H\beta + [O III]$  LF. We see a strong, increasing evolution in  $L_*$  up to  $z \sim 3$ .

with our discussion in the section below (see Section 3.4.1) where we predict the bright-end to be dominated by  $[O III]$  emitters. The Colbert et al. (2013) study was part of the *HST* WISP programme, covering 29 fields ( $0.036 \text{ deg}^2$ ) in search of  $H\alpha$ ,  $[O III]$ , and  $[O II]$  emission line galaxies using WFC3 grism spectroscopy. Because this was a spectroscopic study and the fact that our LF matches (in the bright-end) with that of Colbert et al. (2013) gives us confirmation that we are picking up the  $[O III]$  emitters in our sample. The rise in the faint-end can then be attributed to the  $H\beta$  emitters in our sample.

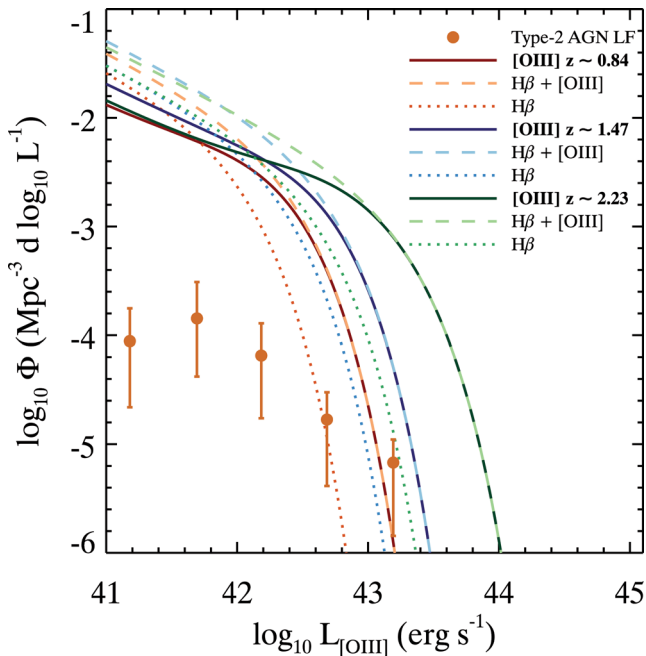
In comparison to Ly et al. (2007), we see a clear deviation of the fits between our  $z \sim 0.84$  and their  $z \sim 0.83$ . Although we do agree in terms of  $\phi_*$ , the main deviation in the LFs are in  $L_*$  and  $\alpha$  (was  $\alpha = -1.44 \pm 0.09$  in comparison to our  $-1.6$  fixed faint-end slope). The study was based on deep optical imaging of the Subaru Deep Field using the Suprime-Cam on the 8.2-m Subaru Telescope and was complemented with Subaru FOCAS and Keck DEIMOS spectroscopy. The deviation could be due to biases from sample sizes such that our sample consists of more bright emitters

to populate the bright-end, hence shifting  $L_*$  higher. In comparison to Sobral et al. (2015), for which this work was done in unison with, we find perfect agreement but our sample probes deeper by 0.1 dex.

Fig. 4 shows the evolution of  $L_*$  along with the results from other studies. There is a strong trend in which  $L_*$  is increasing from  $z = 0$  to  $2.23$  and then flattens. This trend is supported by Ly et al. (2007), Pirzkal et al. (2013), Colbert et al. (2013), and Sobral et al. (2015). Prior to this work, the  $z < 1$  studies hinted to a rising trend in  $L_*$ , which with our measurements and the  $z \sim 1.5$  measurements of Sobral et al. (2015) has been confirmed up to  $z \sim 3$ .

For the normalization of the LF, we see an evolution (Fig. 4) such that  $\phi_*$  drops as redshift increase up to  $z \sim 3$ . This is consistent with the collection of UV LFs (i.e. Oesch et al. 2010), while our determination is based on a reliable  $H\beta + [O III]$  sample. Note that prior to this study, the  $H\beta + [O III]$  measurements in the literature paint the picture that the  $\phi_*$  evolution is flat up to  $z \sim 1$ . With the inclusion of our measurements, along with the  $z \sim 1.5$  measurements of Sobral et al. (2015), we find that  $\phi_*$  strong decreases after  $z > 1$ .





**Figure 5.** Shown is the predicted [O III] LFs from  $z \sim 0.8$  to  $2.2$  and compared to the  $z \sim 0.71$  zCOSMOS type-2 AGN LF of Bongiorno et al. (2010). The  $H\beta$  LFs are made by simply taking the  $H\alpha$  LFs of Sobral et al. (2013) and assuming a fixed  $H\beta/H\alpha$  ratio to convert them. We find that our  $z \sim 0.84$  LF is [O III]-dominated at the bright-end. Also, the level of AGN contribution is very little, except for  $\log_{10} L_{[\text{O III}]} > 43.0 \text{ erg s}^{-1}$ .

#### 3.4.1 Predicting the [O III] LF and AGN contribution

The results highlighted above are for the  $H\beta$  and [O III] emitters combined as one sample since we cannot separate the two types of emitters based on photometry. For this, we need to conduct spectroscopic follow-ups to properly segregate the emitters. We attempted to separate the sample by using the  $H\alpha$  LF of Sobral et al. (2013). The advantage of using the LFs of Sobral et al. (2013) is that it is fully compatible since we are both using the same data set and methodology. We start by first removing  $A_{H\alpha} = 1.0$  mag dust correction, then apply an  $H\beta/H\alpha = 0.35$  line ratio from Osterbrock & Ferland (2006) to get the observed  $H\beta$  LF. We then applied  $A_{H\beta} = 1.38$  mag (based on Calzetti et al. (2000); assuming  $A_{H\alpha} = 1.0$  mag) to the LFs and dust-corrected our  $H\beta + [\text{O III}]$  LFs using  $A_{H\beta+[\text{O III}]} = 1.35$  (see Section 4.3.2 for details). The next step was subtracting our  $H\beta + [\text{O III}]$  LFs from the predicted  $H\beta$  LFs to get the predicted LF for [O III]5007 emitters. The results are shown in Fig. 5.

We find that the [O III] emitters in our sample completely dominate the  $H\beta + [\text{O III}]$  LFs while towards the faint-end the  $H\beta$  emitters dominate. This is expected as the theoretical [O III]/ $H\beta$  line ratio is  $\sim 3$  (for  $Z = 0.0004$ ; Osterbrock & Ferland 2006), which would segregate our sample such that the bright-end will be populated by [O III] emitters and the faint-end with  $H\beta$  emitters. We also find an interesting feature where the normalization in the [O III] LFs are the same, with the exemption of the  $z \sim 1.42$  [O III] LF which is slightly higher. This can imply that the relative contribution of  $H\beta$  is the same for all three LFs. We note that this is a qualitative assessment and subtracting a Schechter function by another Schechter function does not result in the same functional form.

We also attempted to compare the  $z \sim 0.8$  [O III] LF with the zCOSMOS AGN type-2 LF of Bongiorno et al. (2010) to qualitatively assess the contribution of AGNs. The type-2 AGN is the best candidate that could contaminate our sample as they have a continuum that is similar to normal star-forming galaxies and they photoionize the same cold gas that is photoionized by hot massive stars. We find that we are in agreement *only* for the brightest luminosity bin of Bongiorno et al. (2010) ( $\log_{10} L_{[\text{O III}]} \sim 43 \text{ erg s}^{-1}$ ), but in disagreement for the lower luminosity bins. This implies that our brightest [O III] emitters are primarily AGNs, but the fainter emitters are a combination of star-forming galaxies and AGNs, with the star-forming galaxies being the most dominant. Future spectroscopic follow-ups of our sample would allow us to properly study the evolution of AGNs in the Universe.

#### 3.5 [O II] luminosity function $z = 1.47, 2.23, 3.3,$ and $4.7$

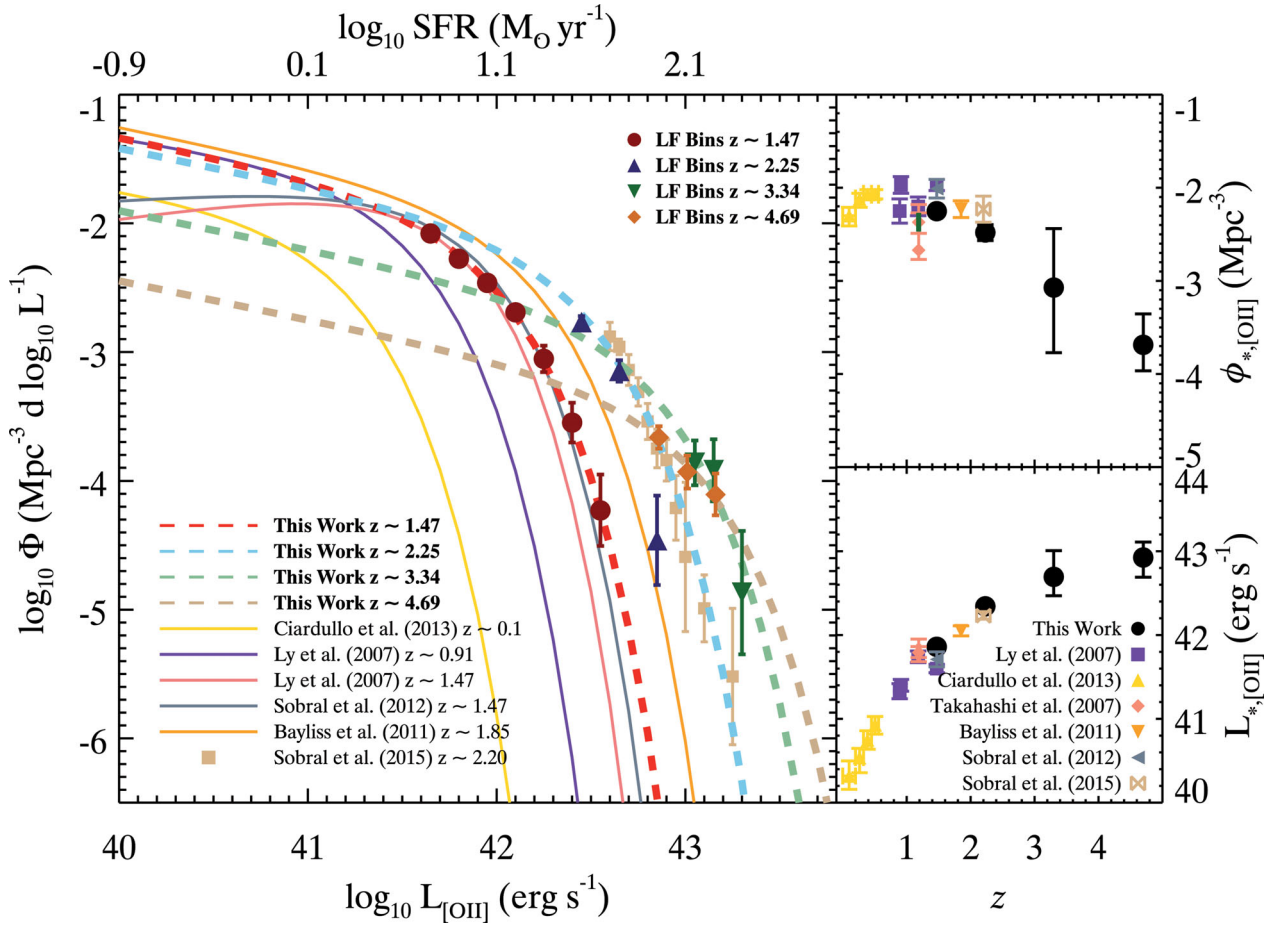
We present here the results of the [O II] LF and the Schechter fit out to  $z \sim 5$ . The results are highlighted in Fig. 6. We see a clear evolution of the LF with redshift, with a large increase in the characteristic luminosity with redshift. The right-hand panels of Fig. 6 show the evolution of the fitted  $\phi_*$  and  $L_*$  parameters, and Fig. 7 shows the degeneracy between the fitted values.

Included on Fig. 6 are data from the literature that range from  $z = 0$  to  $2.2$  (Ly et al. 2007; Takahashi et al. 2007; Bayliss et al. 2011; Sobral et al. 2012, 2015; Ciardullo et al. 2013). We find that our  $z = 2.23$  binned LF data is in agreement with the CF-HiZELS result of Sobral et al. (2015). Because their sample size is  $\sim 4$  times larger than our measurement, we have combined their LF data points with ours. The main effect is our measurement extends the combined LF  $0.15$  dex fainter. We note that the LFs of Sobral et al. (2015) are directly compatible with our LFs as our study follows the same methodology. In fact, the Sobral et al. (2015) is specific to the NBJ determined LFs and the effects of cosmic variance while this study focuses on the evolution and extension of the LFs out to  $z \sim 5$ .

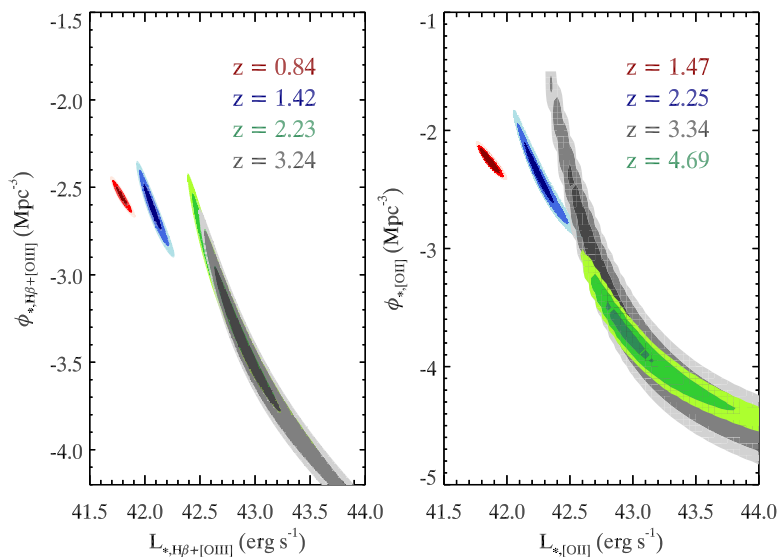
Our  $z = 1.47$  measurements are in perfect agreement with Sobral et al. (2012) and close to agreement with Ly et al. (2007). We note that for Ly et al. (2007) the faint-end was measured to be  $\alpha = -0.78 \pm 0.13$  and  $-0.9 \pm 0.2$  for Sobral et al. (2012) while we keep our faint-end slope fixed to  $-1.3$  for all LFs. As seen in Fig. 6, we find that we are not probing deep enough to fully see the turn in the LF for  $L_{[\text{O II}]} < 41 \text{ erg s}^{-1}$  as found by Ly et al. (2007) and Sobral et al. (2012). This is probably due to the fact that both studies used 2 arcsec apertures, while our study uses 3 arcsec apertures (provided in the Sobral et al. (2013) catalogue) to select  $z = 1.47$  [O II] emitters, which means that their studies are better at recovering faint emitters.

We find a strong evolution in  $L_*$ , as shown in Fig. 6 for which a strong, rising trend is seen up to  $z \sim 5$ . We also see the same evolution in  $H\alpha$  studies (e.g. Sobral et al. 2012), where  $L_*$  is strongly increasing from  $z = 0$  to  $2$ . The same evolution of  $L_*$  is seen in UV studies (Oesch et al. 2010) up to the same redshift range. We notice some scatter for the low- $z$  studies (Ly et al. 2007; Takahashi et al. 2007; Bayliss et al. 2011; Sobral et al. 2012; Ciardullo et al. 2013). This is primarily due to limitations in survey area and/or shallowness of the surveys.

The evolution in  $\phi_*$  is shown in Fig. 6 along with measurements from the literature. Our measurements show a decreasing trend since  $z \sim 1.5$  while for redshifts less than  $1.5$  shows a flat evolution in  $\phi_*$ . The same evolution in  $\phi_*$  is also seen in  $H\alpha$  studies (e.g. Sobral et al. 2013) where after  $z \sim 1$  and up to  $z \sim 2$ ,  $\phi_*$  is shown to be



**Figure 6.** Left: presented are the  $[\text{OII}]$  luminosity functions along with those from the literature. The SFR calibration used to create the top horizontal axis is from Kennicutt (1998, see Section 4.2). The darker data points are colour coded to match the lighter LF fit (dashed lines). We find that the evolution from the low- $z$  studies of Gallego et al. (2002) and Ciardullo et al. (2013) to our  $z = 4.7$  LFs is quite strong and clear. We find that our  $z = 1.47$  LF is in agreement with the HiZELS  $[\text{OII}]$  study of Sobral et al. (2012) and the Subaru Deep Survey study of Ly et al. (2007). Our  $z = 2.23$  is also in agreement with the CF-HiZELS study of Sobral et al. (2015). Top right: the evolution in the normalization of the LF. We find that  $\phi_{*}$  has been decreasing from  $z \sim 1.47$  to  $z \sim 5$ . Bottom right: the evolution of  $L_{*}$ . We find a clear, strong evolution in  $L_{*}$  all the way to  $z \sim 5$ .



**Figure 7.** Shown is the interdependent evolution of  $L_{*}$  and  $\phi_{*}$  for  $H\beta + [\text{OIII}]$  (left) and  $[\text{OII}]$  (right). The contours are colour-coded to match the text colour in the legend. The confidence levels are organized such that the darkest shade is the  $1\sigma$  level and the lightest shade being the  $3\sigma$  level. There is a clear evolution that as redshift increases,  $\phi_{*}$  drops and  $L_{*}$  increases up to  $z \sim 3$  and  $z \sim 5$  for  $H\beta + [\text{OIII}]$  and  $[\text{OII}]$ , respectively.

decreasing. UV measurements (Oesch et al. 2010) also see a similar trend.

Future wide surveys, such as *Euclid* and *WFIRST*, will be able to observe larger samples of emission line galaxies such that our results can be used as predictions for such upcoming projects. By taking our LFs and integrating them to some flux/luminosity limit, these future surveys can estimate the number of [O II] emitters that can be detectable. Our LFs would then be quite useful as a tool to plan surveys studying [O II] emitters out to  $z \sim 5$ . Furthermore, our LFs can be used as a tool to gauge the level of low- $z$  interlopers in various studies, such as Ly  $\alpha$  studies at high- $z$ .

## 4 EVOLUTION OF THE STAR FORMATION HISTORY OF THE UNIVERSE

In this section, we present the star formation history evolution of our [O II] sample out to  $z \sim 5$ . We begin by measuring the level of AGN contamination and then present the calibrations used to get the SFRDs. We conclude with a discussion of the evolution of the SFRD based on [O II] emitters, the correction for dust, and our estimates of the SMD evolution of the universe based on our SFRD fit.

### 4.1 Contribution from AGN

AGNs play an important role in the evolution of galaxies. Because AGN heat the cold gas that is photoionized by O and B-type stars in star-forming regions, the same emission lines become present by both sources. It is then imperative that the SFRDs are properly corrected for AGN contamination to ensure that the sample is, by majority, a star-forming sample. Due to the low number-density of AGNs, it is difficult to use the current catalogues in the literature (e.g. *Chandra-COSMOS*) as a direct indicator on the level of AGN contribution/contamination to our sample. When comparing to *Chandra-COSMOS* (Elvis et al. 2009), we find 1, 0, 4, 0 for  $z \sim 0.84, 1.42, 2.23$ , and 3.24 for our  $H\beta + [O III]$  sample and 5, 2, 1, 0 for  $z \sim 1.47, 2.25, 3.34$ , and 4.69 for our [O II] sample. We also compared our catalogues to the *XMM-COSMOS* catalogue (Cappelluti et al. 2009) and found no matches.

These matches themselves cannot give us a complete indication of the level of our AGN contamination as they are X-ray flux-limited. We instead take advantage of the rest-frame 1.6  $\mu\text{m}$  bump in the SEDs of star-forming galaxies. This bump arises from the minimum opacity of  $H^-$  ions in the stellar atmospheres of cool stars. For AGNs, the bump is meshed in with various other emission (e.g. polycyclic aromatic hydrocarbons, silicate grains) resulting in a rising power-law SED after 1.6  $\mu\text{m}$  in the rest frame. We use the deep IRAC data in COSMOS and UDS and with the condition that redder colours are AGNs, signifying the rising SED after 1.6  $\mu\text{m}$ , and anything with bluer colours are star-forming galaxies resulting in the 1.6  $\mu\text{m}$  bump. We measure the colours by taking the  $[3.6-4.5] > 0.1$  ( $z \sim 0.8$ ),  $[4.5-5.6] > 0.1$  ( $z \sim 1.5$ ), and  $[5.6-8.0] > 0.1$  ( $z \sim 2.2$ ) for  $H\beta + [O III]$  emitters and  $[4.5-5.6] > 0.1$  ( $z \sim 1.47$ ) and  $[5.6-8.0] > 0.1$  ( $z \sim 2.23$ ) for [O II] emitters. We find AGN contamination for  $H\beta + [O III]$  is  $\sim 11.4$  per cent,  $\sim 18.5$  per cent, and  $\sim 28.8$  per cent for  $z \sim 0.8, 1.5$ , and 2.2, respectively. The amount of AGN contamination for [O II] is  $\sim 18.5$  per cent and  $\sim 19.4$  per cent for  $z \sim 1.47$  and 2.23, respectively. For  $z > 2.2$  in [O II] emitters, we set the AGN contamination constant to that at  $z = 2.2$  as this would require going beyond the last IRAC band. Note that these are upper limits for the level of AGN contamination such that our SFRDs are corrected for the highest con-

tamination possible via the 1.6  $\mu\text{m}$  bump technique. By comparing our SFRD measurements to other star formation tracers, we can determine if the AGN correction was too high or not. But to reliably measure the correction will require follow-up spectroscopy of our sample to properly separate the AGN from the star-forming sample. We therefore apply our determined AGN correction to the luminosity densities measured from the fully integrated LFs (decreases the SFRDs) and include 20 per cent of the correction factor in quadrature with the luminosity density errors.

### 4.2 Calibrations

The SERD is calculated via the luminosity density from the LF at each redshift. The luminosity density is defined as

$$\mathcal{L} = \int_0^\infty \Phi(L)LdL = \phi_* L_* \Gamma(\alpha + 2), \quad (5)$$

where  $\mathcal{L}$  is the luminosity density,  $\phi_*$  is the normalization,  $L_*$  is the characteristic luminosity, and  $\alpha$  is the faint-end slope. Our determined luminosity densities are highlighted in Table 3 and consider the full range of luminosities. The SFRD is then calculated by using the Kennicutt (1998) diagnostics:

$$\dot{\rho}_{\text{SFR, O II}} = 1.4 \times 10^{-41} \mathcal{L}_{[\text{O II}]} \text{M}_\odot \text{yr}^{-1} \text{Mpc}^{-3}, \quad (6)$$

where an  $\mathcal{L}_{[\text{O II}]} / \mathcal{L}_{H\alpha} = 1.77$  is assumed. We note that using the [O II] SFR calibration comes with several drawbacks, such as metallicity, reddening, and line ratio assumptions, but the [O II] line is the brightest emission line detectable at  $z > 1.5$  where  $H\alpha$  falls in the infrared. We will present the effects of these drawbacks in a future study (Khostovan et al., in preparation). Furthermore, we present the uncorrected for dust SFRD measurements to see, qualitatively, the evolution of the SFRD. This means that the results shown in Fig. 8 are lower-limits since any dust correction will just increase the SFRD measurements.

We also use the derived relation of Osterbrock & Ferland (2006):

$$\dot{\rho}_{\text{SFR, H}\beta + \text{O III}} = 7.35 \times 10^{-42} \mathcal{L}_{[\text{H}\beta + \text{O III}]} \text{M}_\odot \text{yr}^{-1} \text{Mpc}^{-3} \quad (7)$$

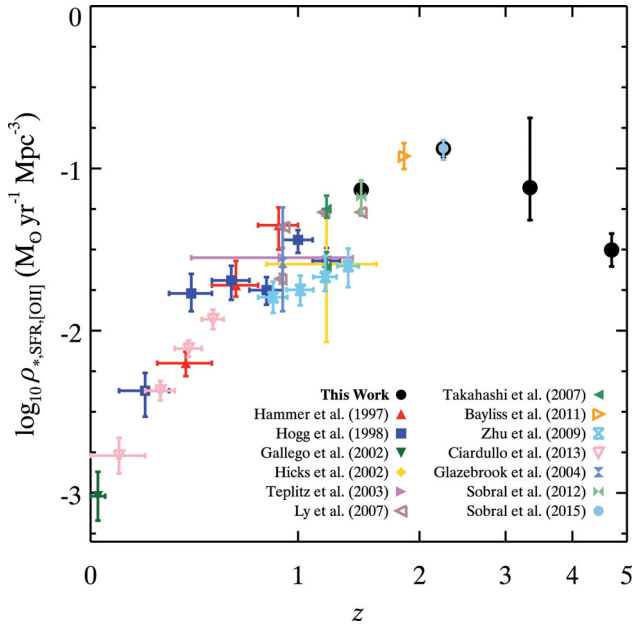
to measure the  $H\beta + [O III]$  SFRD<sup>4</sup> although this cannot be taken as a purely star-forming indicator as there are several caveats behind it. We want to make this point specifically clear; we do not use the  $H\beta + [O III]$  SFRDs in fitting the star formation and stellar mass assembly history of the Universe. We instead use it to compare the measurements to those of the [O II] SFRDs and other tracers in the literature to show if our sample is tracing a star-forming sample and whether or not if the  $H\beta + [O III]$  calibration is more of a ‘reliable’ tracer of star formation activity than previously thought.

### 4.3 SFRD evolution of [O II] emitters

#### 4.3.1 Uncorrected-for-dust

Fig. 8 shows the evolution of the uncorrected-for-dust [O II] SFR density for the first time and determined in a self-consistent way from  $z = 0$  to 5. The evolution is clear and signifies that a peak that occurs at  $z \sim 2-3$  and then there is a fall for higher redshifts. We include [O II] measurements from the literature after uncorrecting them for dust and correcting the cosmology (the pre-2000 papers

<sup>4</sup> We used the dust extinction curve of Calzetti et al. (2000) for the  $H\beta + [O III]$  emitters and applied for all redshifts, such that  $A_{H\beta + [O III]} = 1.35 \text{ mag}$  (assuming  $A_{H\alpha} = 1.0 \text{ mag}$ ).



**Figure 8.** The uncorrected for dust SFRD evolution based only on  $[O\text{II}]$  emission studies. We find that our  $z = 1.47$  and  $2.25$  LF continues the inverse power-law slope that is found from  $z = 0$  to  $2$  in the majority of SFRD studies, and a continuous drop for  $z > 2$ . This is the first time that  $[O\text{II}]$  studies have gone beyond  $z \sim 1.5$  in a reasonably, statistically constrained fashion. We find perfect agreement with Sobral et al. (2012) and are 0.15 dex off from Ly et al. (2007). Also, we find perfect agreement with the  $z = 2.25$  SFRD of Sobral et al. (2015).

used non- $\Lambda$ CDM cosmological parameters). These results plus ours can then be taken as a lower limit as any dust extinction correction would increase the SFR densities. We include a compilation of SFR densities and LF parameters from various studies, using various diagnostics, all normalized to the same cosmology as that of this paper, same Kennicutt (1998) calibration with the same line ratio, and all with  $A_{[O\text{II}]} = 0$  to make it easier for future studies to utilize. This compilation is found in Appendix C. We find that our  $z = 1.47$   $[O\text{II}]$  SFRD measurement is in perfect agreement with Sobral et al. (2012) and Ly et al. (2007). We are also in perfect agreement with the measurement of Sobral et al. (2015) and our  $z = 2.23$  measurement.

We note that Bayliss et al. (2012) made a measurement at  $z = 4.6$  by observing the GOODS-S field using the NB2090 and  $K_s$  filters of the ESO HAWK-I instrument. The results of this study are restricted to a sample of only three genuine emitters. Although ground-breaking at the time, their SFRD estimate is severely limited by issues of sample size and cosmic variance.

#### 4.3.2 Dust- and AGN-corrected SFRD

To compare with other studies using different SFRD diagnostics, we adopt a dust correction using the HiZELS  $z = 1.47$  measurement of Hayashi et al. (2013),  $A_{H\alpha} \sim 0.35$  mag. Hayashi et al. (2013) studied  $H\alpha$  and  $[O\text{II}]$  emitters using HiZELS data to conclude that the traditional  $A_{H\alpha} = 1.0$  mag that has been used in the literature is overestimating the dust correction for  $[O\text{II}]$  emitters at  $z = 1.47$ , such that these emitters are observed to have  $A_{H\alpha} \sim 0.35$  mag and are less dusty than previously thought.

To test the dust extinction coefficient of Hayashi et al. (2013), we apply the traditional  $A_{H\alpha} \sim 1.0$  mag to all four  $[O\text{II}]$  SFRD measurements. We find that based on this dust correction, our measure-

ments overestimate the  $H\alpha$ -based SFRD measurements of Sobral et al. (2013) and the radio-stacked measurements of Karim et al. (2011), which is impervious to dust extinction. The level of overestimation is such that our  $z = 1.47$  SFRD measurement and  $z = 2.23$  SFRD measurement was  $\sim 0.4$  dex above the SFRD measurements of Sobral et al. (2013) and Karim et al. (2011). When using the Hayashi et al. (2013) dust extinction coefficient, we find that our SFRD measurements are perfectly matched with Sobral et al. (2013) and Karim et al. (2011), as seen in Fig. 9.

We apply the Calzetti correction (Calzetti et al. 2000) with the Hayashi et al. (2013) measurement of  $A_{H\alpha} \sim 0.35$  mag such that

$$\frac{A_{[O\text{II}]}}{A_{H\alpha}} = \frac{k([O\text{II}])}{k(H\alpha)}, \quad (8)$$

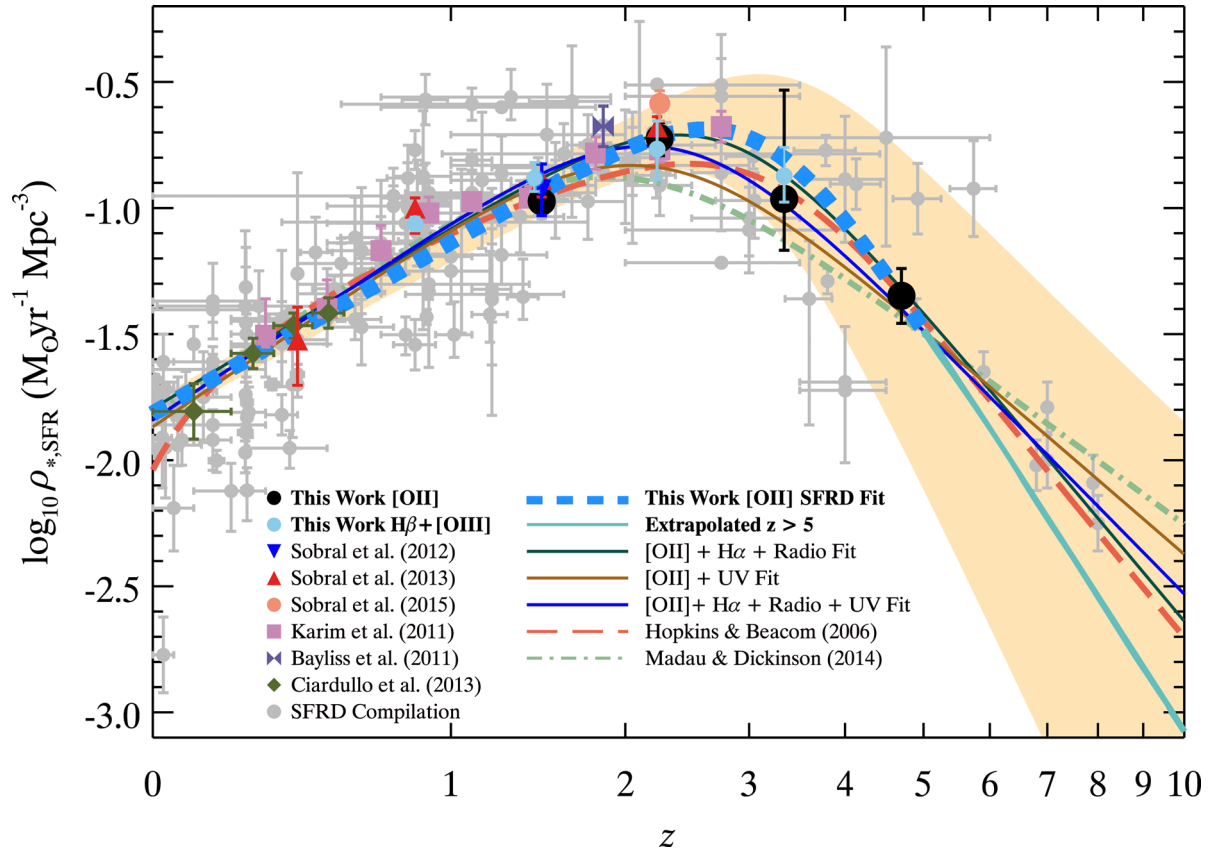
where  $k([O\text{II}]) = 5.86$  and  $k(H\alpha) = 3.31$ , resulting in  $A_{[O\text{II}]} = 0.62$  mag. We calibrate all the measurements to the same  $[O\text{II}]$  SFR calibration of Kennicutt (1998). All measurements hereinafter include AGN corrections as discussed in Section 4.1.

Fig. 9 shows our dust-corrected and AGN-corrected  $[O\text{II}]$  SFRD measurements. We also include a large compilation of studies from the literature which is a combination of the compilations of Hopkins & Beacom (2006), Madau & Dickinson (2014), Gunawardhana et al. (2013), Ly et al. (2007), and our own compilation as a comparison (Appendix C). We find that our measurements accurately reproduce the star formation history of the universe up to  $z \sim 5$ . This is the first time that an  $[O\text{II}]$  study has ever accomplished such a measurement in a self-consistent manner. We find that the  $z = 1.47$  and  $2.23$  perfectly agree with the HiZELS  $H\alpha$  measurements of Sobral et al. (2013) in Fig. 9. The AGN contamination in Sobral et al. (2013) assumed a simple  $\sim 10$  per cent, which is backed by a detailed search of potential AGNs by Garn et al. (2010). Based on the similarities between our  $[O\text{II}]$  measurement and the independently AGN-corrected SFRD measurement of Sobral et al. (2013), we can conclude that the level of AGN contamination measured is reasonable and the methodology sound.

Another key point is the stacked radio measurements of Karim et al. (2011) (pink squares in Fig. 9). The benefit of radio measurements are that they are impervious to dust, but have the downside of poor resolution and blending. We find that our  $z = 1.47$  and  $2.25$   $[O\text{II}]$  measurements are in agreement with Karim et al. (2011), such that the Hayashi et al. (2013) dust extinction coefficient does reliably correct our measurements to represent the dust-corrected SFRD of star-forming galaxies.

We also find an interesting result when comparing the  $H\beta + [O\text{III}]$  SFRD measurements to our  $[O\text{II}]$  SFRDs and other measurements in the literature. As discussed above, the  $H\beta + [O\text{III}]$  calibration is considered in the literature as a ‘mixed’ tracer of star formation activity. Here, we find that using the calibration of Osterbrock & Ferland (2006) with an  $A_{H\beta+[O\text{III}]} = 1.35$  mag (based on the traditional  $A_{H\alpha} = 1$  mag), our  $H\beta + [O\text{III}]$  measurement for  $z = 0.84$  matches perfectly with the  $H\alpha$  SFRD of Sobral et al. (2013) and the radio measurement of Karim et al. (2011). The implications of this agreement show that not only is the dust correction technique applied correctly, but also that the AGN correction is accurate such that it is matching with the AGN-corrected  $H\alpha$  SFRD of Sobral et al. (2013). For our  $z = 1.43$   $H\beta + [O\text{III}]$  SFRD, we find a perfect match with Karim et al. (2011), Sobral et al. (2012), Sobral et al. (2013), and our  $[O\text{II}]$  SFRD. The  $z = 2.23$   $H\beta + [O\text{III}]$  SFRD matches well with Karim et al. (2011), Sobral et al. (2012), and our  $[O\text{II}]$  SFRD measurement. Lastly, we find perfect agreement between our  $[O\text{II}]$  SFRD and the  $H\beta + [O\text{III}]$  SFRD at  $z \sim 3.3$ . All these perfect agreements hint to the notion that the





**Figure 9.** Our [O II] dust and AGN-corrected SFRD evolution with the [O II] studies of Bayliss et al. (2011), Ciardullo et al. (2013), Sobral et al. (2013), and Sobral et al. (2015), along with the results of this paper, that are used to fit the parametrization of Madau & Dickinson (2014). The best fit is shown as the dashed line (dodger blue) and is only based on [O II] measurements. We also include an extrapolation to higher- $z$  (dash-dotted turquoise line), as we do not constrain this part of redshift space but can extrapolate based on our fit. The  $1\sigma$  region is highlighted in gold filled regions around the fit. The stacked radio study of Karim et al. (2011) and the H $\alpha$  study of Sobral et al. (2013) are also shown as a comparison and are in agreement with our measurements. Our compilation of SFRD measurements (in grey) are a combination of our compilation and that of Hopkins & Beacom (2006), Madau & Dickinson (2014), Ly et al. (2007), and Gunawardhana et al. (2013). We reproduce the SFRD evolution history of the universe based primarily on [O II] studies with the peak of star formation history occurring at  $z \sim 3$ . We also include the fits of Hopkins & Beacom (2006, IMF corrected to Salpeter) and that of Madau & Dickinson (2014). We find that the Hopkins & Beacom (2006) fit reasonably matches our SFRD fit, while the Madau & Dickinson (2014) fits well until  $z > 2$ . This is mostly because the Madau & Dickinson (2014) fit is driven by the  $z > 5$  UV measurements (which are not backed by spectroscopy), for which we do not include in our [O II] fit.

H $\beta$  + [O III] SFR calibrations could in fact be more of a reliable tracer of star formation activity than previously thought. Furthermore, this is also strong evidence to show that our H $\beta$  + [O III] sample is dominated by star-forming galaxies and is a reliable sample. Also, our survey seems to be detecting H $\beta$  + [O III] emitters that have more dust in comparison to [O II] emitters such that the traditional  $A_{H\alpha} = 1$  mag applies to the H $\beta$  + [O III] sample and a lower dust correction applies to the [O II] emitters. This notion was proposed by Hayashi et al. (2013) for their [O II] sample. Their conclusion was that dustier [O II] emitters fall to lower luminosities that are below the detection limit, while the less dusty emitters, which will be apparently brighter, are detected.

We fit the SFRD using our [O II] SFRD measurements along with the [O II] measurements of Bayliss et al. (2011), Ciardullo et al. (2013), and Sobral et al. (2012) to the parametrization of Madau & Dickinson (2014):

$$\log_{10} \dot{\rho}_* = a \frac{(1+z)^b}{1 + [(1+z)/c]^d} M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}, \quad (9)$$

where our fit results with  $a = 0.015 \pm 0.002$ ,  $b = 2.26 \pm 0.20$ ,  $c = 4.07 \pm 0.51$ , and  $d = 8.39 \pm 2.60$ . The fit is purely based on [O II] emitters, but we have also fitted for the cases of [O II]+H $\alpha$ +Radio,

[O II]+UV, and [O II]+H $\alpha$ +Radio+UV (see Fig. 9) to show how our fit will vary based on the data that we use. Based on the [O II] fit, we see a drop at  $z > 3$  that is slightly steeper than those determined by UV dropout studies (i.e. Oesch et al. 2010; Bouwens et al. 2011, 2014; Schenker et al. 2013). Despite this drop in our [O II] SFRD compared to the UV studies, we do find that the UV measurements are still within  $1\sigma$ .

An important note to make though is that prior to this paper, there does not exist a study besides UV/Ly  $\alpha$  studies that have measured the SFRD up to  $z \sim 5$  since  $z \sim 3$ . This is a crucial point since there has been no other study so far that could confirm the drop-out measurements, which are severely affected by dust extinction. Furthermore, this is the first time that the cosmic star formation history has been constrained based on a single tracer for larger volumes and up to  $z \sim 5$ . Our current measurements are the farthest that we can measure the [O II] SFRD due to the fact that the emission line would go past K band and into the infrared. Future space-based narrow-band surveys, such as *JWST* and the *Wide-field Imaging Surveyor for High-redshift (WISH)*, will be able to probe [O II] emitters up to  $z \sim 12$ , which would allow us to compare and confirm the UV SFRD measurements at  $z > 5$ .

We also compare our fit to those of Hopkins & Beacom (2006) and Madau & Dickinson (2014) in Fig. 9. For the  $z < 2$  regime, we find that our  $[O\text{II}]$  SFRD fit agrees well with all the other fits. For the  $z > 2$  regime, we do see divergences based on the fit. In terms of the actual data points, we find that the Hopkins & Beacom (2006) is closest in agreement as it has a continuing SFRD up to a peak at  $z \sim 2.5$  and a drop that continues through the high- $z$   $[O\text{II}]$  measurements. The Madau & Dickinson (2014) is also in agreement for the high- $z$  measurements, but fails to match with the  $z \sim 2-3$  peak. This is mostly due to the fact that their measurements are driven by the  $z > 5$  UV dropout SFRDs (e.g. Oesch et al. 2010; Bouwens et al. 2011, 2014; Schenker et al. 2013).

As with all SFR measurements, there are systematic uncertainties that must be taken into account. In the case of  $[O\text{II}]$  emitters, our main systematic uncertainties come from metallicity and dust extinction. To study the metallicities and its effects on the SFR calibration, we will need to conduct follow-up spectroscopy. Furthermore, studying the metallicity of our sample will give us also an understanding of the dynamics (inflow/outflow) that can affect star formation activity. We also plan to study in a future paper the dust extinction properties of our sample and how it relates to and affects the star formation activity of galaxies in our sample (Khostovan et al., in preparation).

#### 4.4 Evolution of the stellar mass density

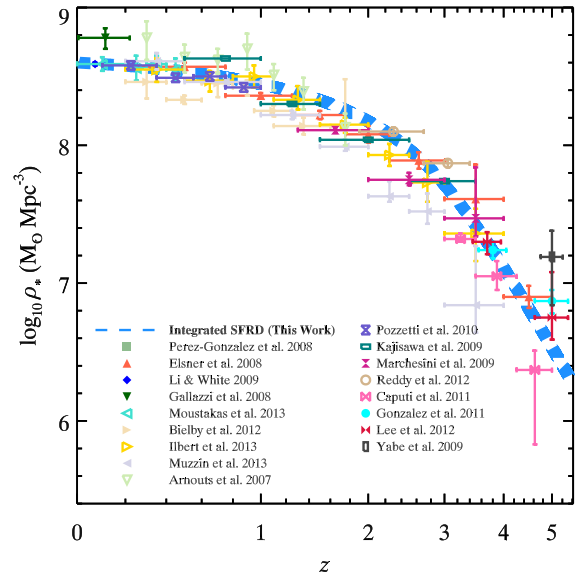
We use the  $[O\text{II}]$  SFRD results presented in this paper to provide an estimate of the SMD evolution by doing a time-integral of equation (9). The SMD evolution gives us an understanding of how the universe has assembled its mass throughout cosmic time. This estimate is quantitatively sensitive to the choice of the IMF in terms of the normalization of the SFRD and SMD evolution, but it does not qualitatively affect the final results.

Our estimate assumes a Salpeter IMF, which has been used throughout this entire paper, and a recycling fraction of  $R = 0.27$ . For a review of the derivation of this factor, we refer the reader to the recent review of Madau & Dickinson (2014). We calculate the SMD by

$$\rho_*(z) = (1 - R) \int_z^\infty \frac{\dot{\rho}_*(z)}{H_0(1+z)\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}} dz \quad (10)$$

where  $\dot{\rho}_*(z)$  is the SFRD fit using the parametrization defined in equation (9) and  $R$  is the recycling fraction, or the fraction of stars that is returned back into the interstellar medium and intergalactic medium. Because the equation above is using  $z = \infty$  as a reference for which we do not know the SMD for, we instead constrain the integral such that at  $z = 0$  the SMD will be  $\log_{10}\rho_* \sim 8.6 M_\odot \text{Mpc}^{-3}$ , in agreement with measurements made at that redshift.

Our results are shown in Fig. 10. We find an evolution where the stellar mass assembly rapidly increases from  $10^{6.2}$  to  $10^8 M_\odot \text{Mpc}^{-3}$  from  $z \sim 5$  to 2, a time frame of only 2 Gyr. The evolution then tapers and flattens out by  $z = 0$  which is related to the decrease in the SFRD that we have observed since  $z \sim 2$ . This is also the same conclusion found by observational studies of the SMD. We include measurements from Arnouts et al. (2007), Elsnér, Feulner & Hopp (2008), Gallazzi et al. (2008), Pérez-González et al. (2008), Kajisawa et al. (2009), Li & White (2009), Marchesini et al. (2009), Yabe et al. (2009), Pozzetti et al. (2010), Caputi et al. (2011), González et al. (2011), Bielby et al. (2012), Lee et al. (2012), Reddy et al. (2012), Ilbert et al. (2013), Moustakas et al. (2013), and Muzzin et al. (2013) in Fig. 10 and we find that our integrated



**Figure 10.** The evolution of the SMD of the Universe based on the integrated  $[O\text{II}]$  SFRD. Overlaid are the SMD measurements from the literature that were compiled in the recent review of Madau & Dickinson (2014). We find that our integration of the purely  $[O\text{II}]$  determined SFRD reasonably traces the stellar mass assembly of the Universe.

SFRD reproduces the same evolution seen by these studies. We have found that the  $[O\text{II}]$  based SFRD and SMD accurately reproduce the evolution of mass assembly in the Universe. This match can also be seen as yet another verification that our sample of  $[O\text{II}]$  emitters are primarily star-forming galaxies as the conclusions from our SMD estimate are the same seen in the literature.

## 5 CONCLUSIONS

We have presented the largest sample of  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters between  $z \sim 0.8-5$  that have been selected based on a robust and self-consistent technique, backed up by a wide array of spectroscopic emitters. We have used the HiZELS UKIRT and Subaru narrow-band catalogues, along with multiwavelength data from the COSMOS and UDS fields, to create a clean and well-defined sample of star-forming galaxies. The main results of this paper are as follows.

- (i) We have robustly selected a total of 2477, 371, 270, 179  $H\beta + [O\text{III}]$  emitters at  $z = 0.84, 1.42, 2.23,$  and  $3.24$  and 3285, 137, 35, 18  $[O\text{II}]$  emitters at  $z = 1.47, 2.25, 3.34,$  and  $4.69$  in the combined COSMOS and UDS fields. These are the largest samples of  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  emitters to have been detected in this redshift range.
- (ii) We have extended the LF in the literature to higher- $z$ , as well as refined the lower- $z$  measurements for both types of emitters. For the  $H\beta + [O\text{III}]$  emitters, we find that the bright-end of our  $z = 1.42$  and  $2.23$  LFs are in agreement with the grism spectroscopy-based LFs of Colbert et al. (2013); hence, this increases the reliability of our sample being dominantly  $[O\text{III}]$  emitters in the bright-end. We also find from our predictions of the  $[O\text{III}]$  LFs that our sample is dominated by  $[O\text{III}]$  emitters at the bright-end. The faint-end is dominated by  $H\beta$  emitters. We also find that the normalization of the  $[O\text{III}]$  LFs are the same such that the relative contribution of  $H\beta$  emitters is the same between  $z \sim 0.8-2.2$ .

(iii) The evolution of  $L_*$  and  $\phi_*$  for  $H\beta + [O\text{III}]$  is found to have a strong increasing/decreasing evolution, respectively, up to  $z \sim 3$ . For our  $[O\text{II}]$  sample, we find that  $L_*$  increases strongly up to  $z \sim 5$  and  $\phi_*$  is strongly dropping up to the same redshift.

(iv) We have discussed that our LFs are reliable to be used in making predictions of the number of emitters to be detected by future wide-surveys, such as *Euclid* and *WFIRST*. Furthermore, our LFs can also determine the number of low- $z$  interlopers in Ly  $\alpha$  studies, such that the level of contamination by low- $z$  sources can be reduced in such studies.

(v) The SFRD has been constrained using  $[O\text{II}]$  measurements up to  $z \sim 5$  for the first time. We find that the peak of the cosmic SFRD is located around  $z \sim 3$  and is in agreement with our large compilation of UV, IR, radio, and nebular emission studies. We find that for  $z > 2$ , our SFRD fit drops slightly faster in comparison to the UV dropout studies in this redshift regime. However, we find that the UV measurements are within the  $1\sigma$  error bar range of our SFRD fit. Future space-based narrow-band surveys, such as *JWST* and *WISH*, will be able to extend the range of  $[O\text{II}]$  detection out to  $z \sim 12$  so that we can compare and confirm or invalidate the UV dropout measurements.

(vi) We also find that the  $H\beta + [O\text{III}]$  SFRD measurements are nicely in line with our  $[O\text{II}]$  sample and other star formation tracers. This then brings to question of whether the  $H\beta + [O\text{III}]$  calibration is more ‘reliable’ as a tracer of star formation than previously thought. With our large sample of these emitters, we will have the ability to explore this issue in detail.

(vii) By integrating the SFRD, we have made estimates of the SMD evolution and find that it steeply rose up to  $z \sim 2$  and flattened out up to the present day. This is also confirmed by the wealth of measurements in the literature.

The results in the paper have implications in the evolution of galaxies and the star formation activity occurring in said galaxies. Despite the robustness of our sample, there is still room for improvement. Our measurements have done well to constrain the bright-end, while keeping the faint-end fixed based on measurements from the literature. We will require deeper narrow-band and broad-band measurements in order to constrain the faint-end slope of the LF. Spectroscopic follow-up will also be necessary to accurately measure the extent of AGN contamination in our sample. Although, our colour–colour selections have shown (see Fig. A1) that they are quite reliable due to the large set of spectroscopic measurements confirming this reliability. That being said, spectroscopic measurements of our sample will help in separating the  $H\beta$  and  $[O\text{III}]$  samples to measure separate LFs. Lastly, future narrow-band surveys, such as the proposed *WISH* telescope, will be able to extend the redshift window of  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  studies up to  $z \sim 12$ , which can be used to confirm the UV dropout studies at higher- $z$ . Despite all these improvements and potential future progresses, our sample has reliably (given all the limitations) and robustly traced the evolution of star-forming activity in the universe.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their informative, detailed, and useful comments/questions. We also acknowledge Anahita Alavi for many useful and insightful discussion regarding the determination of the LF. We also acknowledge Brian Siana for useful comments.

The data used in this paper is publicly available from Sobral et al. (2013). We refer the reader to this paper for details in regards

to the data reduction and selection methodology for the original catalogues.

This paper uses data from the VIPERS. VIPERS has been performed using the ESO Very Large Telescope, under the ‘Large Programme’ 182.A-0886. The participating institutions and funding agencies are listed at <http://vipers.inaf.it>.

DS acknowledges financial support from the Netherlands Organization for Scientific research (NWO) through a Veni fellowship, from FCT through a FCT investigator Starting Grant and Start-up Grant (IF/01154/2012/CP0189/CT0010) and from FCT grant PEst-OE/FIS/UI2751/2014

IRS acknowledges support from STFC (ST/L00075X), the ERC Advanced Investigator programme DUSTYGAL 321334, and a Royal Society/Wolfson Merit Award.

PNB acknowledges support from STFC.

## REFERENCES

- Arnouts S. et al., 2007, *A&A*, 476, 137  
 Bayliss K. D., McMahon R. G., Venemans B. P., Ryan-Weber E. V., Lewis J. R., 2011, *MNRAS*, 413, 2883  
 Bayliss K. D., McMahon R. G., Venemans B. P., Banerji M., Lewis J. R., 2012, *MNRAS*, 426, 2178  
 Bielby R. et al., 2012, *A&A*, 545, A23  
 Bongiorno A. et al., 2010, *A&A*, 510, A56  
 Bouwens R. J. et al., 2011, *ApJ*, 737, 90  
 Bouwens R. J. et al., 2014, *ApJ*, 795, 126  
 Bradshaw E. J. et al., 2013, *MNRAS*, 433, 194  
 Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, 686, 1503  
 Bunker A. J., Warren S. J., Hewett P. C., Clements D. L., 1995, *MNRAS*, 273, 513  
 Calzetti D., 2013, in Falcón-Barroso J., Knapen J. H., eds, *Star Formation Rate Indicators*. Cambridge Univ. Press, Cambridge, p. 419  
 Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, 533, 682  
 Cappelluti N. et al., 2009, *A&A*, 497, 635  
 Caputi K. I., Cirasuolo M., Dunlop J. S., McLure R. J., Farrah D., Almaini O., 2011, *MNRAS*, 413, 162  
 Ciardullo R. et al., 2013, *ApJ*, 769, 83  
 Cirasuolo M. et al., 2007, *MNRAS*, 380, 585  
 Coil A. L. et al., 2011, *ApJ*, 741, 8  
 Colbert J. W. et al., 2013, *ApJ*, 779, 34  
 Darvish B., Sobral D., Mobasher B., Scoville N. Z., Best P., Sales L. V., Smail I., 2014, *ApJ*, 796, 51  
 Dickinson M., 1998, in Livio M., Fall S. M., Madau P., eds, *The Hubble Deep Field*. Cambridge Univ. Press, Cambridge, p. 219  
 Drake A. B. et al., 2013, *MNRAS*, 433, 796  
 Elsner F., Feulner G., Hopp U., 2008, *A&A*, 477, 503  
 Elvis M. et al., 2009, *ApJS*, 184, 158  
 Fujita S. S. et al., 2003, *ApJ*, 586, L115  
 Gallazzi A., Brinchmann J., Charlot S., White S. D. M., 2008, *MNRAS*, 383, 1439  
 Gallego J., García-Dabó C. E., Zamorano J., Aragón-Salamanca A., Rego M., 2002, *ApJ*, 570, L1  
 Garilli B. et al., 2014, *A&A*, 562, A23  
 Garn T. et al., 2010, *MNRAS*, 402, 2017  
 Geach J. E., Smail I., Best P. N., Kurk J., Casali M., Ivison R. J., Coppin K., 2008, *MNRAS*, 388, 1473  
 Geach J. E., Sobral D., Hickox R. C., Wake D. A., Smail I., Best P. N., Baugh C. M., Stott J. P., 2012, *MNRAS*, 426, 679  
 Glazebrook K., Tober J., Thomson S., Bland-Hawthorn J., Abraham R., 2004, *AJ*, 128, 2652  
 González V., Labbé I., Bouwens R. J., Illingworth G., Franx M., Kriek M., 2011, *ApJ*, 735, L34  
 Gunawardhana M. L. P. et al., 2013, *MNRAS*, 433, 2764  
 Hammer F. et al., 1997, *ApJ*, 481, 49



Hayashi M., Sobral D., Best P. N., Smail I., Kodama T., 2013, MNRAS, 430, 1042  
 Hicks E. K. S., Malkan M. A., Teplitz H. I., McCarthy P. J., Yan L., 2002, ApJ, 581, 205  
 Hogg D. W., Cohen J. G., Blandford R., Pahre M. A., 1998, ApJ, 504, 622  
 Hopkins A. M., Beacom J. F., 2006, ApJ, 651, 142  
 Ilbert O. et al., 2009, ApJ, 690, 1236  
 Ilbert O. et al., 2013, A&A, 556, A55  
 Johnston R., 2011, A&AR, 19, 41  
 Kajisawa M. et al., 2009, ApJ, 702, 1393  
 Karim A. et al., 2011, ApJ, 730, 61  
 Kennicutt R. C., Jr, 1998, ARA&A, 36, 189  
 Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531  
 Lawrence A. et al., 2007, MNRAS, 379, 1599  
 Lee K.-S. et al., 2012, ApJ, 752, 66  
 Li C., White S. D. M., 2009, MNRAS, 398, 2177  
 Lilly S. J., Le Fevre O., Hammer F., Crampton D., 1996, ApJ, 460, L1  
 Lilly S. J. et al., 2007, ApJS, 172, 70  
 Ly C. et al., 2007, ApJ, 657, 738  
 Ly C., Lee J. C., Dale D. A., Momcheva I., Salim S., Staudaher S., Moore C. A., Finn R., 2011, ApJ, 726, 109  
 McLure R. J. et al., 2013, MNRAS, 428, 1088  
 Madau P., Dickinson M., 2014, ARA&A, 52, 415  
 Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765  
 Markwardt C. B., 2009, in Bohlender D. A., Durand D., Dowler P., eds. ASP Conf. Ser. Vol. 411, Astronomical Data Analysis Software and Systems XVIII. Astron. Soc. Pac., San Francisco, p. 251  
 Moustakas J. et al., 2013, ApJ, 767, 50  
 Muzzin A. et al., 2013, ApJ, 777, 18  
 Oesch P. A. et al., 2010, ApJ, 725, L150  
 Osterbrock D. E., Ferland G. J., 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. University Science Books, Mill Valley, CA  
 Pascual S., 2005, PASP, 117, 120  
 Pérez-González P. G. et al., 2008, ApJ, 675, 234  
 Pirzkal N. et al., 2013, ApJ, 772, 48  
 Pozzetti L. et al., 2010, A&A, 523, A13  
 Reddy N. A., Pettini M., Steidel C. C., Shapley A. E., Erb D. K., Law D. R., 2012, ApJ, 754, 25  
 Schenker M. A. et al., 2013, ApJ, 768, 196  
 Scoville N. et al., 2007, ApJS, 172, 1  
 Smit R., Bouwens R. J., Franx M., Illingworth G. D., Labbé I., Oesch P. A., van Dokkum P. G., 2012, ApJ, 756, 14  
 Sobral D. et al., 2009, MNRAS, 398, 75  
 Sobral D., Best P. N., Geach J. E., Smail I., Cirasuolo M., Garn T., Dalton G. B., Kurk J., 2010, MNRAS, 404, 1551  
 Sobral D., Best P. N., Smail I., Geach J. E., Cirasuolo M., Garn T., Dalton G. B., 2011, MNRAS, 411, 675  
 Sobral D., Best P. N., Matsuda Y., Smail I., Geach J. E., Cirasuolo M., 2012, MNRAS, 420, 1926  
 Sobral D., Smail I., Best P. N., Geach J. E., Matsuda Y., Stott J. P., Cirasuolo M., Kurk J., 2013, MNRAS, 428, 1128  
 Sobral D., Best P. N., Smail I., Mobasher B., Stott J., Nisbet D., 2014, MNRAS, 437, 3516  
 Sobral D. et al., 2015, MNRAS, 451, 2303  
 Stark D. P., Ellis R. S., Bunker A., Bundy K., Targett T., Benson A., Lacy M., 2009, ApJ, 697, 1493  
 Stott J. P., Sobral D., Smail I., Bower R., Best P. N., Geach J. E., 2013a, MNRAS, 430, 1158  
 Stott J. P. et al., 2013b, MNRAS, 436, 1130  
 Swinbank A. M., Sobral D., Smail I., Geach J. E., Best P. N., McCarthy I. G., Crain R. A., Theuns T., 2012a, MNRAS, 426, 935  
 Swinbank A. M., Smail I., Sobral D., Theuns T., Best P. N., Geach J. E., 2012b, ApJ, 760, 130  
 Takahashi M. I. et al., 2007, ApJS, 172, 456  
 Teplitz H. I., Collins N. R., Gardner J. P., Hill R. S., Rhodes J., 2003, ApJ, 589, 704

Tresse L., Maddox S. J., Le Fèvre O., Cuby J.-G., 2002, MNRAS, 337, 369  
 Villar V., Gallego J., Pérez-González P. G., Pascual S., Noeske K., Koo D. C., Barro G., Zamorano J., 2008, ApJ, 677, 169  
 Yabe K., Ohta K., Iwata I., Sawicki M., Tamura N., Akiyama M., Aoki K., 2009, ApJ, 693, 507  
 Zhu G., Moustakas J., Blanton M. R., 2009, ApJ, 701, 86

## APPENDIX A: SELECTION TECHNIQUE

Here, we present, in detail, the selection of emitters that made it in to our sample. We also present the exact colour–colour selections that were applied in our work in Table A1.

### A0.1 $H\beta + [O\text{III}]$ emitters at $z \sim 0.8$

$H\beta + [O\text{III}]$  sources in the NB921 data at  $z \sim 0.8$  are selected by their photometric redshifts within the range of  $0.75 < z_{\text{phot}} < 0.95$ . The colour–colour selection criterion reduces the number of contaminants by separating the  $H\beta + [O\text{III}]$  emitters from lower- $z$   $H\alpha$  and higher- $z$   $[O\text{II}]$  emitters. This is done by using the *BRiK* selection (Sobral et al. 2009) in Fig. A1. Spectroscopic redshifts were used to assess the robustness and effectiveness of the selection criteria. 213 sources were spectroscopically confirmed. 169 were selected by the colour–colour selection. From the 213 sources (for which all were selected by their photo- $z$ ), only 11 were confirmed  $H\beta 4861$ , 76 were  $[O\text{III}]4959$ , and 126 were  $[O\text{III}]5007$  emitters. Removed from the sample where nine low- $z$  and 13 high- $z$  spectroscopically confirmed emitters. The lower- $z$  emitters were primarily  $H\alpha$  and  $[N\text{II}]$ . The higher- $z$  emitters were primarily  $[O\text{II}]$  emitters. All these misidentified emitters were removed from the sample. Based on the spectroscopic data, we find that  $\sim 91$  per cent of all spectroscopic measurements for photo- $z$  selected objects were either  $H\beta$  or  $[O\text{III}]$  emitters and that the *BRiK* colour–colour selection does select  $\sim 91$  per cent of all the  $H\beta + [O\text{III}]$  spectroscopically confirmed emitters. In total, we have 2477  $z = 0.84$   $H\beta + [O\text{III}]$  emitters in our sample. Based on the spectroscopic data, we find that the colour–colour selection effectively selects  $H\beta + [O\text{III}]$  emitters with a completeness of  $\sim 91$  per cent, making our sample not just the largest, but the most complete sample of  $H\beta + [O\text{III}]$  emitters to date.

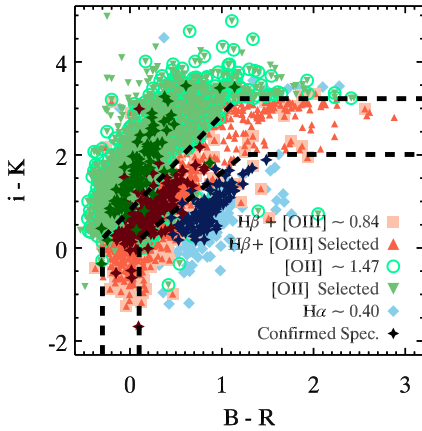
### A0.2 $H\beta + [O\text{III}]$ & $[O\text{II}]$ emitters at $z \sim 1.5$

$[O\text{II}]$  emitters in NB921 are selected with photometric redshifts between  $1.2 < z_{\text{phot}} < 1.7$ . We use the *BRiK* colour–colour selection and include any sources with spectroscopic redshifts. Included in our sample are 97 spectroscopically confirmed sources, with 90 of them being colour–colour selected as well. Removed from the sample were 48 low- $z$  spectroscopically confirmed emitters, for which the majority were  $[O\text{III}]$ . We also removed a few  $[\text{He}\text{I}]$  emitters at  $z \sim 1.28$  and  $H\alpha$ , and  $[N\text{II}]$  emitters ( $z \sim 0.4$ ). The contamination from high- $z$  emitters was significantly less (eight emitters) as there are no major emission lines beyond  $[O\text{II}]$ . The majority of this contamination came from  $[\text{Mg}\text{II}]$  emitters at  $z = 2.25$ . The issue of contamination arises here as we may ‘naively’ state that our level of contamination is  $\sim 33$  per cent, but it is noted that spectroscopic measurements to date have an inherent bias to the low- $z$  regime, such that there are more low- $z$  than high- $z$  measurements. This makes accurately measuring the level of contamination difficult. We note though that our colour–colour selection did select  $\sim 93$  per cent of the spectroscopically confirmed emitters. In



**Table A1.** Definitions of the colour–colour selection.

Filter	Colour–colour	Emitter	Redshift	Selection criteria
NB921	<i>BRiK</i>	$H\beta + [O\text{III}]$	0.84	$-0.3 < (B - R) < 0.08$ & $(i - K) < 2.04(B - R) + 0.81$ $0.08 < (B - R) < 1$ & $1.6(B - R) < (i - K) < 2.04(B - R) + 0.81$ $1 < (B - R) < 1.24$ & $1.6(B - R) < (i - K) < 3.21$ $1.24 < (B - R) & 2.01 < (i - K) < 3.21$
		$[O\text{II}]$	1.47	$(B - R) < -0.3$ $-0.3 < (B - R) < 1.17$ & $(i - K) > 2.04(B - R) + 0.81$ $(B - R) > 1.17$ & $(i - K) > 3.21$
NBJ	<i>BzK</i>	$H\beta + [O\text{III}]$ & $[O\text{II}]$		$(B - z) < 0.4$ $0.4 < (B - z) < 2.41$ & $(z - K) > (B - z) - 0.4$ $2.41 < (B - z)$ & $(z - K) > 2.0$
	<i>izK</i>	$H\beta + [O\text{III}]$ $[O\text{II}]$	1.42 2.25	$(z - K) < 5(i - z) - 0.4$ $(z - K) > 5(i - z) - 0.4$
NBH	<i>BzK</i>	$H\beta + [O\text{III}]$	2.23	$(B - z) < 0.4$ $0.4 < (B - z) < 2.41$ & $(z - K) > (B - z) - 0.4$ $2.41 < (B - z)$ & $(z - K) > 2.0$
	<i>izK</i>	$H\beta + [O\text{III}]$	2.23	$(z - K) > 5(i - z) - 0.4$
	<i>UVz</i>	$[O\text{II}]$	3.34	$(U - V) > 1.2$ & $(U - V) > 0.5(V - z) + 1.2$ & $(V - z) < 1.6$
NBK	<i>UVz</i>	$H\beta + [O\text{III}]$	3.24	$(U - V) > 1.2$ & $(U - V) > 0.5(V - z) + 1.2$ & $(V - z) < 1.6$
	<i>Viz</i>	$[O\text{II}]$	4.69	$(V - i) > 1.2$ & $(V - i) > 0.89(i - z) + 1.2$ & $(i - z) < 1.3$



**Figure A1.** The  $i - K$  versus  $B - R$  colour–colour distribution of NB921 emitters using the colour–colour selection of Sobral et al. (2013). Overall, the level of completeness is  $>90$  per cent as the large majority of spectroscopically confirmed emitters are within the selection area; hence, we confirm the *BRiK* selection of Sobral et al. (2009) is very efficient in selecting  $H\alpha$ ,  $H\beta + [O\text{III}]$ , and  $[O\text{II}]$  samples from narrow-band surveys.

total, we have selected 3285  $z = 1.47$   $[O\text{II}]$  emitters. This is by far the largest sample of  $[O\text{II}]$  emitters at  $z \sim 1.5$  to date and, based on the spectroscopically confirmed sources, is  $\sim 93$  per cent complete.

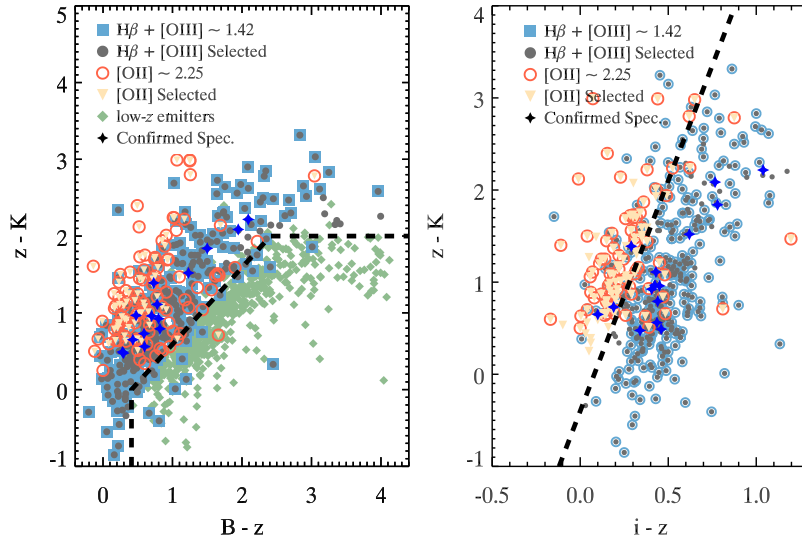
$H\beta + [O\text{III}]$  emitters in NBJ are selected based on a photometric redshift range of  $1.20 < z_{\text{phot}} < 1.70$ . The colour–colour selection criteria consists of a *BzK* and *izK* selection, as shown in Fig. A2. We use the *BzK* selection to get our initial sample of emitters and remove the lower- $z$  contaminants, which are mostly  $H\alpha$  emitters. To remove the higher- $z$  contaminants ( $[O\text{II}]$ ), we use the *izK* selection. There were also 15 spectroscopically confirmed sources, all of which were within our colour–colour selection. Of these 15 emitters, 4 were  $H\beta$ , 5  $[O\text{III}]$ 4958, and 6  $[O\text{III}]$ 5007 emitters. We removed five emitters that were spectroscopically confirmed. These emitters were primar-

ily  $H\alpha$  and  $[N\text{II}]$ , all of which were removed from the sample. Our final sample consists of 371  $H\beta + [O\text{III}]$  emitters at  $z = 1.47$  that were selected.

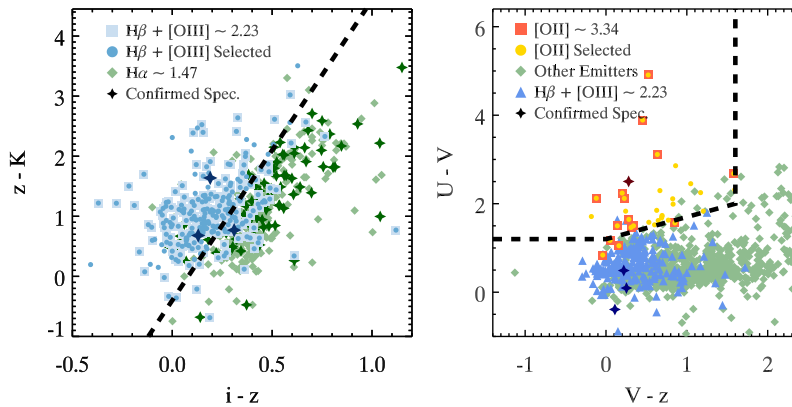
### A0.3 $H\beta + [O\text{III}]$ & $[O\text{II}]$ emitters at $z \sim 2.2$

$[O\text{II}]$  emitters in NBJ are selected if their photometric redshifts are between  $1.7 < z_{\text{phot}} < 2.8$ . We apply the *BzK* colour–colour selection to remove the lower- $z$  contaminants, which are primarily  $H\alpha$  emitters, as shown in Fig. A2. We then use the *izK* colour–colour selection to separate the sample from  $H\beta + [O\text{III}]$  emitters. There were no spectroscopically confirmed sources included in the sample. Removed from the sample were three contaminants, which were all low- $z$  emitters (one  $[N\text{II}]$ , one  $[O\text{III}]$ , and one  $[\text{He}\text{I}]$ ). In total, there are 137  $[O\text{II}]$  emitters selected at  $z = 2.25$ .

$H\beta + [O\text{III}]$  emitters at  $z \sim 2.2$  in NBH are selected with photometric redshifts between  $1.7 < z_{\text{phot}} < 2.8$  along with *BzK* and *izK* colour–colour selections. We apply the *BzK* colour–colour selection to remove emitters with  $z \lesssim 1.5$  (primarily the  $z = 1.47$   $H\alpha$  emitters) and then apply the *izK* colour–colour selection (shown in Fig. A3) to separate our  $H\beta + [O\text{III}]$  sample from the  $z = 1.47$   $H\alpha$  emitters. We find that 21 per cent of our sample is outside the colour–colour region in Fig. A3 but is still selected via the photo- $z$  selection. This would raise concerns about the reliability of the colour–colour selection, but we must point out that it is reliable in terms of the separation between the  $H\beta + [O\text{III}]$  and  $H\alpha$  spectroscopically confirmed emitters. This also raises the point that we are more concerned with consistency in the use of selection techniques to reduce the effects of assumptions in our sample. Furthermore, this also shows that we cannot rely on the colour–colour selection technique alone to select emitters since this would result in a 21 per cent drop in  $H\beta + [O\text{III}]$  emitters and a 21 per cent increase in the contamination of the  $H\alpha$  sample in Sobral et al. (2013). By including the photo- $z$  in our selection methodology, we reduce the level of contaminants and also increase the reliability of the sample. We also included three spectroscopically confirmed sources in the sample



**Figure A2.** Color–colour magnitude distributions for all NBJ emitters. Left: the  $z - K$  versus  $B - z$  selection used to separate the  $H\beta + [O III]$  emitters at  $z \sim 1.47$  and  $[O II]$  emitters at  $z \sim 2.23$  from the low- $z$  emitters that are primarily  $H\alpha$ . Right: the  $z - K$  versus  $i - z$  selection is used to select  $[O II]$  and  $H\beta + [O III]$  emitters. For both selections, the spectroscopically confirmed emitters lie within the selection region adding to completeness of the sample. Both colour–colour selections nicely distinguish between the two samples. We note that for the  $izK$  selection, about  $\sim 15$  per cent of selected  $[O II]$  emitters are within the selection region of  $H\beta + [O III]$  emitters. These are photo- $z$  selected and shows that relying on purely the colour–colour selection would show an  $\sim 15$  per cent drop in the completeness of the  $[O II]$  sample, and an  $\sim 15$  per cent increase in the contamination of the  $H\beta + [O III]$  sample.



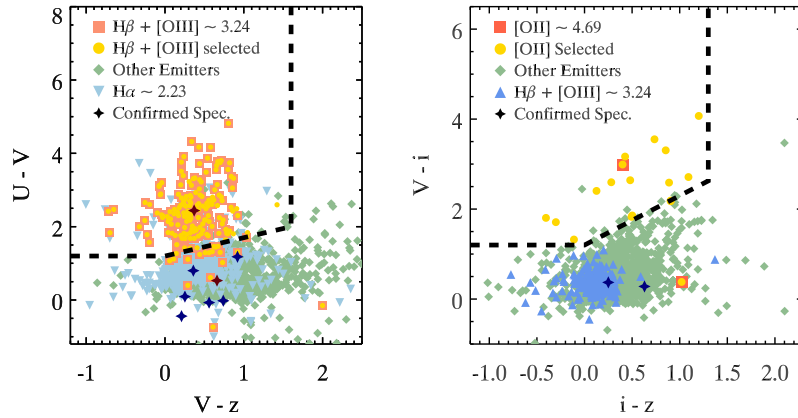
**Figure A3.** Color–colour magnitude distributions of all NBH emitters. Left: the  $z - K$  versus  $i - z$  colour–colour selection used to separate the  $H\beta + [O III]$   $z \sim 2.23$  emitters from the  $H\alpha$   $z \sim 1.47$  emitters (selected based on the methodology of Sobral et al. 2013). We have highlighted the spectroscopically confirmed emitters. The colour–colour selection shows a clear separation between  $H\beta + [O III]$  and  $H\alpha$  measurements based on spectroscopic measurements alone. We find that 21 per cent of our photo- $z$  selected  $H\beta + [O III]$  emitters are within the selection area for  $H\alpha$  emitter, showing that relying on just colour–colour selection would result in the loss of  $\sim 56$  emitters from the sample and an increase in contamination of the  $H\alpha$  sample. Right: the  $U - V$  versus  $V - z$  colour–colour selection that is based on the Lyman break drop-out technique and is used to find  $z \sim 3.3$   $[O II]$  emitters. We also include the  $H\beta + [O III]$  sample and show that the vast majority of these emitters are outside the  $[O II]$  colour–colour selection region.

that were selected by their colour–colour selection. Of these three, we have two  $[O III]_{4959}$  and one  $[O III]_{5007}$  emitter. Removed from the sample were three low- $z$  spectroscopically confirmed emitters ( $H\alpha$  at  $z \sim 1.47$ ). There were no high- $z$  spectroscopic measurements that contaminated the sample. A total of 271  $H\beta + [O III]$  emitters at  $z = 2.23$  were selected.

#### A0.4 $H\beta + [O III]$ & $[O II]$ emitters at $z \sim 3.3$

$[O II]$  emitters at  $z = 3.3$  in NBH are selected if  $2.8 < z_{\text{phot}} < 4$ . We also select sources if they satisfy the  $UVz$  colour–colour criteria. This separates the lower- $z$  contaminants from our sample. We

include our  $H\beta + [O III]$  sample in Fig. A3 to show the separation between the  $[O II]$  and  $H\beta + [O III]$  sample and find that the  $UVz$  selection is reliable in selecting our  $[O II]$  sample. Furthermore, as there are no major emission lines that are detected at higher- $z$ , there is no need to include another colour–colour criteria to account for this contamination. As shown in Fig. 2, the number of sources greatly drops at higher- $z$  for the redshift of interest, making the number of contaminants very small. There was only one spectroscopically confirmed source that was included in the sample and only found in COSMOS. The  $UVz$  colour of this emitter places it well within the selection area, adding to the reliability of our colour–colour selection. No spectroscopic confirmed emitters were



**Figure A4.** colour–colour magnitude distributions of all NBK emitters. Left: shows the  $U - V$  versus  $V - z$  colour–colour selection used to find  $z \sim 3.3$   $H\beta + [O\text{III}]$  emitters based on the Lyman break drop-out technique. We also include the  $H\alpha$  emitters ( $z \sim 2.23$ ) and find that this selection reliably separates the two samples. Right: shows the  $V - i$  versus  $i - z$  colour–colour selection used to find  $z \sim 4.7$   $[O\text{II}]$  emitters based also on the Lyman break drop-out technique. We include the  $H\beta + [O\text{III}]$  sample to show that this selection criteria is reliable to separate the two samples. We find no  $H\beta + [O\text{III}]$  emitter falling within the  $[O\text{II}]$  colour–colour region.

misidentified in the selection. In total, there are 35 of  $[O\text{II}]$  emitters at  $z = 3.3$  selected.

Our  $H\beta + [O\text{III}]$  emitters at  $z = 3.3$  in NBK are selected if their photometric redshifts lie between  $2.8 < z_{\text{phot}} < 4$ . We use a  $UVz$  selection criteria, as shown in Fig. A3, to select  $H\beta + [O\text{III}]$  emitters based on the Lyman break dropout technique. Sources with no detection bluer than the  $U$  band and detection in the  $V$  and  $z$  bands greater than the  $5\sigma$  magnitude detection limits were included. We also include all  $H\alpha$  emitters with photo- $z$  around  $z = 2.23$  in Fig. A4 to show the separation between them and our  $H\beta + [O\text{III}]$  selected emitters. One spectroscopically confirmed source was also included in the sample from UDS. We find that this spectroscopically confirmed emitter is well within the colour–colour region. No spectroscopically confirmed emitters were misidentified in the selection. A total of 179  $H\beta + [O\text{III}]$  emitters at  $z = 3.3$  were selected.

#### A0.5 $[O\text{II}]$ emitters at $z \sim 4.7$

We select our  $[O\text{II}]$  emitters in NBK if they have a photometric redshift between  $4.0 < z_{\text{phot}} < 6.0$ . Emitters are also selected by using the  $Viz$  criteria of Stark et al. (2009) for  $V$ -band dropouts. The colour–colour selection is shown in Fig. A4 and includes our  $H\beta + [O\text{III}]$  sample. We find that there are no  $H\beta + [O\text{III}]$  emitters within the  $[O\text{II}]$  selection area, which adds to the reliability of our sample. We find only one emitter was selected by its photo- $z$  within the selection range and it is well within the colour–colour selection region. The majority of emitters were selected by the colour–colour selection. All sources selected were also under the condition that anything bluer than  $V$  band must have no detection. A study of  $[O\text{II}]$  emitters by Bayliss et al. (2012) also used a similar technique using the  $BVz$  criteria of Stark et al. (2009) for their  $z \sim 4.6$  sample. It must be noted though that the results of this study should not be taken as reliable due to the fact that they were limited to a sample size of only three  $[O\text{II}]$  emitters (about three times smaller than our sample) and the volume probed by this study is a factor of 100 times smaller than our study making it severely susceptible to cosmic variance. Our sample is statistically larger and robust in comparison to Bayliss et al. (2012). There were no spectroscopically

confirmed sources in our sample and no spectroscopic confirmed emitters were misidentified in the selection, thus, giving a total of 18  $[O\text{II}]$  emitters at  $z = 4.5$  that were selected.

#### A0.6 Notes on contamination

We advise the reader that the measurements of contamination are *not* strictly reliable for the  $H\beta + [O\text{III}]$   $z \sim 1.5$  emitters due to the bias in the spectroscopic redshift distribution. Two issues arise are: (1) lack of spectroscopic measurements at higher redshifts to properly quantify the level of contamination, and (2) the inherent bias of spectroscopic measurements to the lower- $z$  regime. The first point really just requires more spectroscopic measurements to increase the population of spectroscopically confirmed sources. The second point has to do with the distribution of spec- $z$  measurements. There exists more spec- $z$  measurements for  $z < 1$ , which results in a skewed histogram that favours the lower- $z$  regime. When measuring the level of contamination, there are more low- $z$  measurements than spectroscopically confirmed measurements (for example,  $z = 1.47$   $[O\text{II}]$ ) which causes a ‘naive’ and biased estimation of the level of contamination. Instead, we considered where the spectroscopically confirmed  $H\beta + [O\text{III}]$  and  $[O\text{II}]$  measurements were on the colour–colour diagrams as a way to assess the reliability of our selection technique.

Based on the points described above, we can measure the level of contamination for the  $H\beta + [O\text{III}]$   $z \sim 0.84$  sample ( $\sim 10$  per cent). To robustly measure the contamination for the higher- $z$  samples, we will need to conduct spectroscopic follow-up which is currently underway with Keck/MOSFIRE and ESO/VLT (Khostovan et al., in preparation).

## APPENDIX B: BINNED LUMINOSITY FUNCTION

Here, we include two tables that show the binned data points of the LF that are plotted in Figs 4 and 6. We include these plots as a convenience for future studies who wish to compare their LFs to ours.

**Table B1.**  $H\beta + [O\text{III}]$  luminosity function.  $\Phi_{\text{obs}}$  shows the observed LF data points per bin; simply, it is the log of the number of emitters divided by the volume.  $\Phi_{\text{final}}$  is the completeness and filter profile corrected luminosity data points per bin. The errors here are Poissonian but with 20 per cent of the corrections added in quadrature.

$\log_{10} L_{H\beta+[O\text{III}]}$ ( $\text{erg s}^{-1}$ )	#	$\Phi_{\text{obs}}$ ( $\text{Mpc}^{-3} \text{ d log}_{10} L$ )	$\Phi_{\text{final}}$ ( $\text{Mpc}^{-3} \text{ d log}_{10} L$ )	Volume ( $10^5 \text{ Mpc}^3$ )
<b><math>z = 0.84</math></b>				
$41.10 \pm 0.10$	703	-1.97	$-1.82 \pm 0.02$	3.25
$41.30 \pm 0.10$	465	-2.15	$-2.04 \pm 0.03$	3.25
$41.50 \pm 0.10$	262	-2.39	$-2.35 \pm 0.04$	3.25
$41.70 \pm 0.10$	128	-2.71	$-2.61 \pm 0.06$	3.25
$41.90 \pm 0.10$	68	-2.98	$-2.94 \pm 0.08$	3.25
$42.10 \pm 0.10$	28	-3.37	$-3.17 \pm 0.13$	3.25
$42.30 \pm 0.10$	12	-3.73	$-3.52 \pm 0.20$	3.25
$42.50 \pm 0.10$	3	-4.34	$-4.12 \pm 0.39$	3.25
<b><math>z = 1.42</math></b>				
$41.95 \pm 0.15$	284	-2.63	$-2.49 \pm 0.03$	4.06
$42.25 \pm 0.15$	73	-3.22	$-3.14 \pm 0.07$	4.06
$42.55 \pm 0.15$	12	-4.01	$-3.89 \pm 0.19$	4.06
$42.85 \pm 0.15$	2	-4.78	$-4.64 \pm 0.48$	4.06
<b><math>z = 2.23</math></b>				
$42.60 \pm 0.075$	84	-3.27	$-3.08 \pm 0.06$	10.46
$42.75 \pm 0.075$	70	-3.36	$-3.14 \pm 0.07$	10.69
$42.90 \pm 0.075$	22	-3.86	$-3.65 \pm 0.13$	10.69
$43.05 \pm 0.075$	5	-4.51	$-4.26 \pm 0.29$	10.69
<b><math>z = 3.24</math></b>				
$42.65 \pm 0.075$	70	-3.33	$-3.17 \pm 0.07$	9.99
$42.80 \pm 0.075$	52	-3.48	$-3.26 \pm 0.09$	10.48
$42.95 \pm 0.075$	25	-3.80	$-3.55 \pm 0.13$	10.48
$43.10 \pm 0.075$	6	-4.42	$-4.17 \pm 0.27$	10.48

**Table B2.**  $[O\text{II}]$  LF.  $\Phi_{\text{obs}}$  shows the observed LF data points per bin; simply, it is the log of the number of emitters divided by the volume.  $\Phi_{\text{final}}$  is the completeness and filter profile corrected luminosity data points per bin. The errors here are Poissonian but with 20 per cent of the corrections added in quadrature.

$\log_{10} L_{[O\text{II}]}$ ( $\text{erg s}^{-1}$ )	#	$\Phi_{\text{obs}}$ ( $\text{Mpc}^{-3} \text{ d log}_{10} L$ )	$\Phi_{\text{final}}$ ( $\text{Mpc}^{-3} \text{ d log}_{10} L$ )	Volume ( $10^5 \text{ Mpc}^3$ )
<b><math>z = 1.47</math></b>				
$41.65 \pm 0.075$	590	-2.24	$-2.08 \pm 0.02$	6.80
$41.80 \pm 0.075$	425	-2.38	$-2.28 \pm 0.03$	6.80
$41.95 \pm 0.075$	257	-2.60	$-2.46 \pm 0.04$	6.80
$42.10 \pm 0.075$	127	-2.90	$-2.69 \pm 0.06$	6.80
$42.25 \pm 0.075$	42	-3.39	$-3.05 \pm 0.10$	6.80
$42.40 \pm 0.075$	19	-3.73	$-3.55 \pm 0.15$	6.80
$42.55 \pm 0.075$	6	-4.23	$-4.23 \pm 0.28$	6.80
<b><math>z = 2.25</math></b>				
$42.45 \pm 0.10$	92	-3.14	$-2.77 \pm 0.05$	6.29
$42.65 \pm 0.10$	37	-3.53	$-3.15 \pm 0.08$	6.29
$42.85 \pm 0.10$	3	-4.62	$-4.46 \pm 0.35$	6.29
<b><math>z = 3.34</math></b>				
$43.05 \pm 0.050$	12	-4.12	$-3.86 \pm 0.17$	15.88
$43.15 \pm 0.075$	7	-4.37	$-3.92 \pm 0.24$	16.52
$43.30 \pm 0.075$	2	-5.22	$-4.87 \pm 0.48$	16.52
<b><math>z = 4.69</math></b>				
$42.86 \pm 0.075$	10	-4.26	$-3.66 \pm 0.09$	12.22
$43.01 \pm 0.075$	5	-4.56	$-3.93 \pm 0.13$	12.22
$43.16 \pm 0.075$	2	-4.96	$-4.11 \pm 0.16$	12.22



**Table C1.** SFRD compilation.

Study	$z$	Diagnostic	Observed			$\log_{10} \dot{\rho}_*$ ( $M_{\odot} \text{ yr}^{-1} \text{ Mpc}^{-3}$ )
			$\log_{10} \phi_*$ ( $\text{Mpc}^{-3}$ )	$\log_{10} L_*$ ( $\text{erg s}^{-1}$ )	$\alpha$	
Ciardullo et al. (2013)	0.0–0.2	[O II]	$-2.30^{+0.09}_{-0.11}$	$40.32^{+0.18}_{-0.16}$	-1.2	$-2.05 \pm 0.11$
	0.2–0.325	[O II]	$-2.12^{+0.05}_{-0.06}$	$40.54 \pm 0.11$	-1.2	$-1.82 \pm 0.06$
	0.325–0.45	[O II]	$-2.07^{+0.04}_{-0.05}$	$40.75^{+0.08}_{-0.10}$	-1.2	$-1.71 \pm 0.05$
	0.45–0.56	[O II]	$-2.07^{+0.03}_{-0.08}$	$40.93^{+0.08}_{-0.12}$	-1.2	$-1.66 \pm 0.06$
Sobral et al. (2012)	1.47	[O II]	$-2.01 \pm 0.10$	$41.71 \pm 0.09$	$-0.9 \pm 0.2$	$-1.48 \pm 0.10$
Bayliss et al. (2011)	1.85	[O II]	$-2.23 \pm 0.09$	$41.31 \pm 0.06$	$-1.3 \pm 0.2$	$-0.92 \pm 0.08$
Ly et al. (2007)	0.89	[O II]	$-2.25 \pm 0.13$	$41.33 \pm 0.09$	$-1.27 \pm 0.14$	$-1.68 \pm 0.03$
	0.91	[O II]	$-1.97 \pm 0.09$	$41.40 \pm 0.07$	$-1.20 \pm 0.10$	$-1.36 \pm 0.02$
	1.18	[O II]	$-2.20 \pm 0.10$	$41.74 \pm 0.07$	$-1.15 \pm 0.11$	$-1.27 \pm 0.02$
	1.47	[O II]	$-1.97 \pm 0.06$	$41.60 \pm 0.05$	$-0.78 \pm 0.13$	$-1.27 \pm 0.02$
Zhu, Moustakas & Blanton (2009)	0.84	[O II]	–	–	–	$-1.79^{+0.10}_{-0.10}$
	1.02	[O II]	–	–	–	$-1.75^{+0.13}_{-0.08}$
	1.19	[O II]	–	–	–	$-1.67^{+0.25}_{-0.11}$
	1.37	[O II]	–	–	–	$-1.60^{+0.16}_{-0.09}$
Takahashi et al. (2007)	1.71–1.203	[O II]	$-2.37^{+0.10}_{-0.12}$	$41.79^{+0.07}_{-0.06}$	$-1.41^{+0.16}_{-0.15}$	$-1.25^{+0.05}_{-0.08}$
	1.71–1.203	[O II]	$-2.67^{+0.28}_{-0.49}$	$41.75^{+0.32}_{-0.20}$	$-1.38^{+0.40}_{-0.37}$	$-1.61^{+0.09}_{-0.28}$
Glazebrook et al. (2004)	0.90	[O II]	-2.91	42.30	-1.3	$-1.35^{+0.34}_{-0.30}$
Teplitz et al. (2003)	0.90 ± 0.50	[O II]	$-3.06 \pm 0.12$	$42.15 \pm 0.08$	-1.35	$-1.55 \pm 0.06$
Gallego et al. (2002)	0.025 ± 0.025	[O II]	$-3.48 \pm 0.19$	$41.24 \pm 0.13$	$-1.21 \pm 0.21$	$-3.02 \pm 0.15$
Hicks et al. (2002)	1.20 ± 0.40	[O II]	–	–	–	$-1.59^{+0.30}_{-0.48}$
Hogg et al. (1998)	0.20 ± 0.10	[O II]	–	–	–	$-2.37^{+0.11}_{-0.16}$
	0.40 ± 0.10	[O II]	–	–	–	$-1.77^{+0.09}_{-0.12}$
	0.60 ± 0.10	[O II]	–	–	–	$-1.69^{+0.06}_{-0.08}$
	0.80 ± 0.10	[O II]	–	–	–	$-1.75^{+0.07}_{-0.08}$
	1.00 ± 0.10	[O II]	–	–	–	$-1.44^{+0.09}_{-0.11}$
	1.20 ± 0.10	[O II]	–	–	–	$-1.57^{+0.18}_{-0.30}$
Hammer et al. (1997)	0.375 ± 0.125	[O II]	–	–	–	$-2.20^{+0.07}_{-0.08}$
	0.625 ± 0.125	[O II]	–	–	–	$-1.72^{+0.11}_{-0.15}$
	0.875 ± 0.125	[O II]	–	–	–	$-1.35^{+0.20}_{-0.38}$
Colbert et al. (2013)	0.7–1.5	[O III]	$-3.19 \pm 0.09$	$42.34 \pm 0.06$	$-1.40 \pm 0.15$	–
	1.5–2.3	[O III]	$-3.74 \pm 0.43$	$42.91 \pm 0.37$	$-1.67 \pm 0.78$	–
	0.7–1.5	[O III]	$-3.28 \pm 0.09$	$42.39 \pm 0.08$	-1.50	–
	1.5–2.3	[O III]	$-3.60 \pm 0.14$	$42.83 \pm 0.11$	-1.50	–
Pirzkal et al. (2013)	0.5 ± 0.4	[O III]	$-2.58^{+0.09}_{-0.09}$	$41.3^{+0.09}_{-0.09}$	$-1.21^{+0.08}_{-0.07}$	–
Ly et al. (2007)	0.41	[O III]	$-2.55 \pm 0.25$	$41.17 \pm 0.22$	$-1.49 \pm 0.11$	$-2.17 \pm 0.06$
	0.42	[O III]	$-2.38 \pm 0.22$	$41.11 \pm 0.24$	$-1.25 \pm 0.13$	$-2.31 \pm 0.09$
	0.62	[O III]	$-2.58 \pm 0.17$	$41.51 \pm 0.15$	$-1.22 \pm 0.13$	$-2.06 \pm 0.05$
	0.83	[O III]	$-2.54 \pm 0.15$	$41.53 \pm 0.11$	$-1.44 \pm 0.09$	$-1.73 \pm 0.03$

**APPENDIX C: SFRD COMPILATION**

In this section, we have compiled a table of the SFRDs from different diagnostics spread over a wide redshift range. Because each study has its own set of assumptions, diagnostics, calibrations, dust corrections, etc. it is quite confusing in keeping track of which study has used which set of assumptions. Let alone, for the earliest papers, we have to even take into account the different cosmologies. To make life much easier for you as

the reader who may be interested in studying the evolution of the cosmic SFR density, we have included in the appendix a long table which is our compilation of the SFR densities and LF parameters from a range of different studies. Parts of this table are from Ly et al. (2007), but updated with the newest studies in the field.

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