11-1-2014

The Mass Evolution of the First Galaxies: Stellar Mass Functions and Star Formation Rates at $4 < z < 7$ in the CANDELS GOODS-South Field

K. Duncan
University of Nottingham, UK

C. J. Conselice
University of Nottingham, UK

A. Mortlock
University of Nottingham, UK

W. G. Hartley
University of Nottingham, UK

Y. Guo
University of California - Santa Cruz

See next page for additional authors

Right click to open a feedback form in a new tab to let us know how this document benefits you.
Follow this and additional works at: https://uknowledge.uky.edu/physastron_facpub

Part of the Astrophysics and Astronomy Commons, and the Physics Commons

Repository Citation
Duncan, K.; Conselice, C. J.; Mortlock, A.; Hartley, W. G.; Guo, Y.; Ferguson, H. C.; Davé, R.; Lu, Y.; Owensworth, J.; Ashby, M. L. N.; Dekel, A.; Dickinson, M.; Faber, S.; Giavalisco, M.; Grogin, N.; Kocevski, Dale D.; Koekemoer, A.; Somerville, R. S.; and White, C. E., "The Mass Evolution of the First Galaxies: Stellar Mass Functions and Star Formation Rates at $4 < z < 7$ in the CANDELS GOODS-South Field" (2014). Physics and Astronomy Faculty Publications. 281.
https://uknowledge.uky.edu/physastron_facpub/281

This Article is brought to you for free and open access by the Physics and Astronomy at UKnowledge. It has been accepted for inclusion in Physics and Astronomy Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Authors
K. Duncan, C. J. Conselice, A. Mortlock, W. G. Hartley, Y. Guo, H. C. Ferguson, R. Davé, Y. Lu, J. Owensworth, M. L. N. Ashby, A. Dekel, M. Dickinson, S. Faber, M. Giavalisco, N. Grogin, Dale D. Kocevski, A. Koekemoer, R. S. Somerville, and C. E. White

The Mass Evolution of the First Galaxies: Stellar Mass Functions and Star Formation Rates at 4 < z < 7 in the CANDELS GOODS-South Field

Notes/Citation Information
Published in Monthly Notices of the Royal Astronomical Society, v. 444, no. 3, p. 2960-2984.

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society ©: 2014 The Authors. Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

Digital Object Identifier (DOI)
http://dx.doi.org/10.1093/mnras/stu1622

This article is available at UKnowledge: https://uknowledge.uky.edu/physastron_facpub/281
The mass evolution of the first galaxies: stellar mass functions and star formation rates at $4 < z < 7$ in the CANDELS GOODS-South field

K. Duncan,1 C. J. Conselice,1 A. Mortlock,1,2 W. G. Hartley,1,3 Y. Guo,4,5 H. C. Ferguson,6 R. Davé,7 Y. Lu,8 J. Ownsworth,1 M. L. N. Ashby,9 A. Dekel,10 M. Dickinson,11 S. Faber,4 M. Giavalisco,5 N. Grogin,8 D. Kocevski,12 A. Koekemoer,6 R. S. Somerville13 and C. E. White14

1 School of Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD, UK
2 SUPA, Institute for Astronomy, The University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK
3 Department of Physics, ETH Zurich, CH-8093 Zurich, Switzerland
4 UCO/Lick Observatory, Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, USA
5 Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA
6 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
7 University of the Western Cape, Bellville, Cape Town 7535, South Africa
8 Kavli Institute for Particle Astrophysics and Cosmology, Stanford, CA 94309, USA
9 Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
10 Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel
11 National Optical Astronomy Observatory, Tucson, AZ 85719, USA
12 Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506, USA
13 Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, NJ 08854, USA
14 Department of Physics and Astronomy, The Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

Accepted 2014 August 7. Received 2014 July 16; in original form 2013 November 13

ABSTRACT

We measure new estimates for the galaxy stellar mass function and star formation rates for samples of galaxies at $z \sim 4, 5, 6$ and 7 using data in the CANDELS GOODS South field. The deep near-infrared observations allow us to construct the stellar mass function at $z \geq 6$ directly for the first time. We estimate stellar masses for our sample by fitting the observed spectral energy distributions with synthetic stellar populations, including nebular line and continuum emission. The observed UV luminosity functions for the samples are consistent with previous observations; however, we find that the observed $M_{\text{UV}} - M_*$ relation has a shallow slope more consistent with a constant mass-to-light ratio and a normalization which evolves with redshift. Our stellar mass functions have steep low-mass slopes ($\alpha \approx -1.9$), steeper than previously observed at these redshifts and closer to that of the UV luminosity function. Integrating our new mass functions, we find the observed stellar mass density evolves from log$_{10}$ $\rho_* = 6.64^{+0.58}_{-0.89}$ at $z \sim 7$ to 7.36 $\pm$ 0.06 $M_\odot$ Mpc$^{-3}$ at $z \sim 4$. Finally, combining the measured UV continuum slopes ($\beta$) with their rest-frame UV luminosities, we calculate dust-corrected star formation rates (SFR) for our sample. We find the specific SFR for a fixed stellar mass increases with redshift whilst the global SFR density falls rapidly over this period. Our new SFR density estimates are higher than previously observed at this redshift.

Key words: galaxies: evolution – galaxies: formation – galaxies: high-redshift – galaxies: luminosity function, mass function.

1 INTRODUCTION

Thanks to the unprecedented sensitivity of the latest extragalactic surveys, the last decade has seen a revolution in the observations of galaxies in the high-redshift universe. It is now possible to study the beginnings of the mechanisms and processes which formed the diverse array of galaxies we find in the local universe today. Since the first successful detections through the Lyman break technique, via the characteristic ‘break’ induced by blanketing hydrogen absorption of the UV continuum (Guhathakurta, Tyson & Majewski 1990; Steidel & Hamilton 1992), the study of high-redshift...
galaxies has progressed rapidly. With the introduction of the Wide-field Camera 3 (WFC3) in 2009 and the unprecedented depth in the near-infrared it provides, the study of galaxies out to redshifts of $z > 6$ has become commonplace.

The numerous measurements of the UV luminosity function (LF) of high-redshift galaxies spanning the redshift range $4 \leq z \leq 9$ (Bouwens et al. 2007, 2010; McLure et al. 2009, 2013; Oesch et al. 2009; Grazian et al. 2011; Lorenzoni et al. 2011; Schenker et al. 2013a) are not only giving an insight into the processes of galaxy formation, they are also helping us to understand the role those galaxies played in the ionization of the intergalactic medium during the epoch of reionization. These surveys have put strong constraints on the contribution of star-forming galaxies to reionization, requiring a significant contribution from faint galaxies below the current detection limits to complete reionization within the observed redshift.

Because they represent the time integral of all past star formation, the stellar masses of galaxies provide additional independent constraints on their contribution to reionization through the observed stellar mass density (SMD; Robertson et al. 2010). Successful models of galaxy evolution and reionization must therefore be able to reconcile both the star formation observed directly, and the record of past star formation contained in the observed stellar masses. The galaxy stellar mass function (SMF) and its integral the SMD are, therefore important tools in the study of galaxy evolution.

However, accurately measuring the stellar masses of galaxies at high redshift is very difficult for a number of fundamental reasons. These reasons stem from the fact that the rest-frame wavelengths probed by optical/near-infrared surveys extend only to the UV continuum, requiring mid-infrared observations to extend past $\lambda_{\text{rest}} = 4000$ Å. Even when rest-frame optical measurements are available through deep Spitzer 3.6 and 4.5 μm observations, e.g. Stark et al. (2009), Labbé et al. (2010), González et al. (2011) and Yan et al. (2012), the degeneracies between dust extinction, age and metallicity are large (see Dunlop 2013 for a detailed discussion).

More recently it has also been shown that the spectral energy distributions (SEDs) of high-redshift galaxies, for both photometrically selected (Schaerer & de Barros 2009, 2010; McLure et al. 2009, 2013; Oesch et al. 2009; Grazian et al. 2011; Lorenzoni et al. 2011; Schenker et al. 2013a) are not only giving an insight into the processes of galaxy formation, they are also helping us to understand the role those galaxies played in the ionization of the intergalactic medium during the epoch of reionization. These surveys have put strong constraints on the contribution of star-forming galaxies to reionization, requiring a significant contribution from faint galaxies below the current detection limits to complete reionization within the observed redshift.

By measuring the UV continuum slope, $\beta$ (Calzetti, Kinney & Storchi-Bergmann 1994), for samples of high-redshift galaxies, Wilkins et al. (2011), Bouwens et al. (2012b) and Bouwens et al. (2013) find evidence for a strong UV luminosity dependence across all redshifts at $z > 3$. In contrast, similar studies by Dunlop et al. (2011), Finkelstein et al. (2012a) and Rogers, McLure & Dunlop (2013) find no clear evidence for a luminosity dependence on $\beta$. Several of these studies outline the importance of the selection of high-redshift galaxies (through either Lyman break or photometric redshift selection) and the treatment of their biases. To this end, more recent analyses (Bouwens et al. 2013; Rogers et al. 2014) which increase sample sizes and minimize biases in the sample selection and $\beta$ measurements are in good agreement, with both studies finding a clear luminosity dependence.

The deep near-infrared observations of the GOODS South field made as part of the Cosmic Assembly Near-infrared Deep Extragalactic Survey (CANDELS; Co-PIs: Faber & Ferguson; Grogin et al. 2011; Koekemoer et al. 2011), combined with the extensive existing optical observations make it a data set ideally suited to the study of galaxy evolution at the so-called cosmic dawn. Covering an area approximately 200 per cent larger than the WFC3 ERS observations alone (Windhorst et al. 2011), and incorporating the even deeper UDF observations, the CANDELS data combines the high sensitivity of the WFC3 observations with high-redshift samples large enough to attempt the first direct derivation of the SMF at $z > 5$. In this paper, we make use of this comprehensive data set to study galaxy stellar masses across the redshift range $z \sim 4$ to $z \sim 7$. In particular, we aim to estimate stellar masses for a large and robust sample of high-redshift galaxies, investigating how the inclusion of nebular emission and increasing SFHs affect the observed stellar mass–UV luminosity relation and the shape of the SMF. For this same sample, we also aim to measure the dust-corrected star formation rates (SFR) which will combine to make a detailed census of the stellar mass growth of high-redshift galaxies.

The growth of the first galaxies and applying it to the observed LF to measure the SMF out to $z \sim 7$. In contrast to the steep faint-end slope of the UV LF ($\alpha = -2.9 \pm 1.7$) at high redshifts, this work observed a notably shallower mass function ($\alpha = -1.4 \pm 1.4$). Subsequent work by Lee et al. (2012) with much greater sample sizes from ground-based near-infrared found a similarly shallow slope at $z \sim 4$ and 5. In contrast, observations by Caputi et al. (2011) and Santini et al. (2012) observe a significantly steeper low-mass slope at $z > 3$.

In González et al. (2011), the shallow low-mass slope arises due to the observed evolution of the mass-to-light ratio with UV luminosity. Similarly, Lee et al. (2012) infer an evolving mass-to-light ratio to light in order to reconcile the luminosity and mass function slopes observed. The primary physical explanation for this evolving mass-to-light ratio is luminosity dependent dust-extinction. However, observations of the stellar populations of high-redshift galaxies have produced conflicting results on the existence and strength of any luminosity dependence. When measuring the UV continuum slope, $\beta$ (Calzetti, Kinney & Storchi-Bergmann 1994), for samples of high-redshift galaxies, Wilkins et al. (2011), Bouwens et al. (2012b) and Bouwens et al. (2013) find evidence for a strong UV luminosity dependence across all redshifts at $z > 3$. In contrast, similar studies by Dunlop et al. (2011), Finkelstein et al. (2012a) and Rogers, McLure & Dunlop (2013) find no clear evidence for a luminosity dependence on $\beta$. Several of these studies outline the importance of the selection of high-redshift galaxies (through either Lyman break or photometric redshift selection) and the treatment of their biases. To this end, more recent analyses (Bouwens et al. 2013; Rogers et al. 2014) which increase sample sizes and minimize biases in the sample selection and $\beta$ measurements are in good agreement, with both studies finding a clear luminosity dependence.

The deep near-infrared observations of the GOODS South field made as part of the Cosmic Assembly Near-infrared Deep Extragalactic Survey (CANDELS; Co-PIs: Faber & Ferguson; Grogin et al. 2011; Koekemoer et al. 2011), combined with the extensive existing optical observations make it a data set ideally suited to the study of galaxy evolution at the so-called cosmic dawn. Covering an area approximately 200 per cent larger than the WFC3 ERS observations alone (Windhorst et al. 2011), and incorporating the even deeper UDF observations, the CANDELS data combines the high sensitivity of the WFC3 observations with high-redshift samples large enough to attempt the first direct derivation of the SMF at $z > 5$. In this paper, we make use of this comprehensive data set to study galaxy stellar masses across the redshift range $z \sim 4$ to $z \sim 7$. In particular, we aim to estimate stellar masses for a large and robust sample of high-redshift galaxies, investigating how the inclusion of nebular emission and increasing SFHs affect the observed stellar mass–UV luminosity relation and the shape of the SMF. For this same sample, we also aim to measure the dust-corrected star formation rates (SFR) which will combine to make a detailed census of the stellar mass growth of high-redshift galaxies.
Gunn 1983) and we assume a Λ cold dark matter cosmology with \( H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}, \Omega_m = 0.3 \) and \( \Omega_\Lambda = 0.7 \). Quoted observables are expressed as actual values assuming this cosmology. Note that luminosities and luminosity-based properties such as observed stellar masses and SFR scale as \( h^{-2} \), whilst densities scale as \( h^3 \).

## 2 THE DATA

The photometry used throughout this work is taken from the catalogue of Guo et al. (2013), a UV to mid-infrared multiband catalogue in the CANDELS GOODS South field based on the CANDELS WFC3/IR observations combined with existing public data.

### 2.1 Imaging data

The near-infrared WFC3/IR data combines observations from the CANDELS survey (Grogin et al. 2011; Koekemoer et al. 2011) with the WFC3 ERS (Windhorst et al. 2011) and Hubble Ultra Deep Field (HUDF; PI Illingworth; Bouwens et al. 2010) surveys. The southern two-thirds of the field (incorporating the CANDELS ‘DEEP’ and ‘WIDE’ regions and the UDF) were observed in the F105W, F125W and F160W bands. The northern-most third, comprising the ERS region, was observed in F098M, F125W and F160W. In addition to the initial CANDELS observations, the GOODS South field was also observed in the alternative J-band filter, F140W, as part of the 3D-HST survey (Brammer et al. 2012).

The optical Hubble Space Telescope (HST) images from the Advanced Camera for Surveys (ACS) images are version v3.0 of the mosaicked images from the GOODS HST/ACS Treasury Program, combining the data of Giavalisco et al. (2004a) with the subsequent observations obtained by Beckwith et al. (2006) and Koekemoer et al. (2011). The field was observed in the F435W, F606W, F775W, F814W and F850LP bands. Throughout the paper, we will refer to the HST filters F435W, F606W, F775W, F814W, F850LP, F098M, F105W, F125W, F160W as B435, V606, F775, I814, z850, Y098, Y105, J125, H160, respectively.

The Spitzer/IRAC (Fazio et al. 2004) 3.6 and 4.5 μm images were taken from the Spitzer Extended Deep Survey (PI: G. Fazio, Ashby et al. 2013) incorporating the pre-existing cryogenic observations from the GOODS Spitzer Legacy project (PI: M. Dickinson). Complementary to the space based imaging of HST and Spitzer is the ground-based imaging of the CTIO U band, VLT/VIMOS U band (Nonino et al. 2009), VLT/ISAAC Ks (Retzlaff et al. 2010) and VLT/HAWK-I Ks (Fontana et al., 2014) bands.

### 2.2 Source photometry and deconvolution

The full details on how the source photometry was obtained are outlined in Guo et al. (2013), however we provide a brief summary of the method used for reference here. Photometry for the HST bands was done using SEXTRACTOR’s dual-image mode, using the WFC3 H-band mosaic as the detection image and the respective ACS/WFC3 mosaics as the measurement image after matching of the point spread function (PSF).

For the ground-based (VIMOS and CTIO U band and ISAAC and Hawk-I Ks) and Spitzer IRAC bands, deconvolution and photometry was done using template-fitting photometry (TFIT). We refer the reader to Laidler et al. (2007), Lee et al. (2012) and the citations within for further details of the TFIT process and the improvements gained on mixed wavelength photometry.

## 3 PHOTOMETRIC REDSHIFTS AND SAMPLE SELECTION

Photometric redshifts for the entire source catalogue were calculated using the EAZY photometric redshift software (Brammer, van Dokkum & Coppi 2008). The fitting was done to all available bands using the default reduced template set based on the PEGASE spectral models of Fioc & Rocca-Volmerange (1997) with an additional template based on the spectrum of Erb et al. (2010). The additional template exhibits features expected in young galaxy populations such as strong optical emission lines and a high Lyman α equivalent width.

For each galaxy, we construct the full redshift probability distribution function (PDF), \( P(z) \propto \exp(-\chi^2/2) \), using the \( \chi^2 \)-distribution returned by EAZY. Although EAZY allows the inclusion of a magnitude based prior when calculating redshifts, none was included in the fitting due to the large uncertainties still present in the H-band (our photometry selection band) LF at high redshifts (Henriques 2012).

### 3.1 Selection criteria

To investigate how the SMF evolves from \( z = 4–7 \), we wish to construct a sample of galaxies in the redshift range \( 3.5 < z < 7.5 \). To select a robust sample suitable for SED fitting, we apply a set of additional criteria based on the full redshift probability distribution for each galaxy to construct the different redshift samples, similar to those used in previous high-redshift sample selections (McLure et al. 2011; Finkelstein et al. 2012b). We then apply the following criteria:

\[
\int_{z-0.5}^{z+0.5} P(z) dz > 0.4 \quad \text{(1)}
\]

\[
\int_{z_{\text{peak}}-0.5}^{z_{\text{peak}}+0.5} P(z) dz > 0.6 \quad \text{(2)}
\]

\[
\chi^2_{\text{min}} / \text{N_{filters}} - 1 < 3, \quad \text{(3)}
\]

where \( z_{\text{sample}} = 4, 5, 6 \) and 7 for the respective bins and \( z_{\text{peak}} \) is the redshift at the peak of the probability distribution (i.e. minimum \( \chi^2 \)).

The first criterion (equation 1) requires that a significant amount of the probability distribution lies within the redshift range we are examining. The second criterion (equation 2) requires that the bulk of the PDF lies close to the peak of the distribution, i.e. that the primary solution is a dominant one. Finally, we require that EAZY provides a reasonable fit to the data (equation 3).

A signal to noise (S/N) cut is placed on the J and H bands, requiring \( S/N(J_{125}) > 3.5 \) and \( S/N(H_{160}) > 5 \). Known AGN, stars and sources with photometry flagged as effected by artefacts are removed. We also visually inspect each galaxy across all the HST bands, excluding sources which were caused or strongly affected by artefacts such as diffraction spikes, bright stars and image edges which were not excluded by any of the other criteria.

Of the initial 34930 objects in the CANDELS GOODS South catalogue, 3164 objects satisfy our first criterion. Of those objects, 256 are excluded by the second criterion and a further 167 are rejected based on their \( \chi^2 \). The S/N criteria exclude a further 274 sources and the remaining criteria exclude a further 204 sources. The resulting final sample comprises 2263 galaxies.
have through the Bayesian combination outlined in Dahlen et al. (2013) spectroscopic redshift sources, the photometric redshifts produced compared to the results of any single code. For the same set of 151 the results from multiple photometric redshift codes, the scatter and able to exclude.

low-redshift interlopers which our selection criteria have not been

identified as high-redshift galaxies but are excluded due to poor fits

shown that it is very difficult to categorically classify sources as high-redshift galaxies and not low-redshift interlopers using photo-

Comparison between spectroscopic and photometric redshift for

Fig. 1 compares the available spectroscopic redshifts for the galaxies in our sample with the corresponding best-fitting photometric redshift (minimum $\chi^2$) as found by EAZY. In total, there are 151 spectroscopic redshift matches for galaxies which pass our selection criteria and are therefore included in our samples. In addition, there are a further 21 galaxies with spectroscopic redshifts of $z > 3.5$ which pass the S/N, AGN criteria but do not pass the photometric redshift criteria. Of these 21 galaxies, 12 are correctly identified as high-redshift galaxies but are excluded due to poor fits (11) or have a redshift very close to the $z = 3.5$ limit but photometric scatter pushes the photometric redshift to below the criteria (1). The remaining nine spectroscopically confirmed high-redshift sources have best-fitting photometric redshifts of $z_{\text{peak}} < 3$. The simulations undertaken to correct for selection completeness are outlined in Section 4.5.1.

For the matched galaxies which pass our selection criteria, we find that our redshift accuracy is very good, with a scatter of just $\sigma_{z,\text{spec}} = \text{rms}(\Delta z/1 + z_{\text{spec}}) = 0.037$ when outliers are excluded, where $\Delta z = (z_{\text{spec}} - z_{\text{phot}})$ (Dahlen et al. 2013). We define outliers as $|\Delta z|/(1 + z_{\text{spec}}) > 0.15$, and find an outlier fraction of 2.65 per cent (four galaxies). This compares with Finkelstein et al. (2012a) who find a scatter of $\sigma_{z}/(1 + z) = 0.044$ at $z > 3$ after excluding outliers (defined more strictly as $|\Delta z| > 0.5$) in the same CANDELS field. We also find that there is very little bias in our photometric redshifts, with median($\Delta z$) = −0.04. Of the four galaxies classified as outliers, all lie at redshifts of $z < 3$ and are low-redshift interlopers which our selection criteria have not been able to exclude.

Dahlen et al. (2013) have recently shown that by combining the results from multiple photometric redshift codes, the scatter and outlier fraction in photometric redshifts can be significantly reduced compared to the results of any single code. For the same set of 151 spectroscopic redshift sources, the photometric redshifts produced through the Bayesian combination outlined in Dahlen et al. (2013) have $\sigma_{z,\text{spec}} = 0.033$ with an outlier fraction of 3.98 per cent and median($\Delta z$) = 0.01.

Although utilizing the photometric redshifts for the CANDELS GOODS-S field produced by this method would result in a small gain in photometric accuracy, we would no longer be able to repro-

duce the full selection method when running simulations. Given this small improvement, we are confident that the use of photometric redshifts produced by a single code will not adversely affect the overall accuracy of the results.

The matched spectroscopic redshifts are from the following sur-

veys: Le Fèvre et al. (2004), Stanway et al. (2004), Vanella et al. (2008), Hathi, Malhotra & Rhoads (2008), Popesso et al. (2009), Wuys et al. (2009), Rhoads et al. (2009), Vanella et al. (2009), Balestra et al. (2010), Kurk et al. (2012). The high-redshift spectroscopic sources within these surveys all derive from initial target selections of predominantly bright Lyman break galaxy (LBG) can-

didates. The measured photometric redshift accuracies are therefore likely biased to a better scatter than the full high-redshift galaxy population. However, examining the redshift accuracy of the mock galaxy catalogue used for our selection comparisons in Appendix A, we find that the photometric redshift accuracy remains good down to the lowest masses probed in this survey for galaxies which pass our criteria. For example, for galaxies of $M_*> 10^{10} M_{\odot}$, we find a scatter excluding outliers of $\sigma_{z,\text{spec}} = 0.053$ and median($\Delta z$) = 0.025 before any $P(z)$ criteria are applied.

To investigate how the SMF evolves from $z = 4–7$, we then constructed four redshift samples in bins across this redshift range: $z > 4(3.5 < z < 4.5)$, $z \sim 5(4.5 < z < 5.5)$, $z \sim 6(5.5 < z < 6.5)$ and $z \sim 7(6.5 < z < 7.5)$.

3.2 Monte Carlo samples

Although we find that our photometric redshifts do well when compared with the matched spectroscopic redshifts, the group of outliers are indicative of the difficulties that exist in correctly dis-

tinguishing between the Lyman break of high-redshift galaxies at $z > 3$ and strong Balmer break galaxies at more moderate red-

shifts $z \approx 0.5–2.5$ in low S/N data. Pirzkal et al. (2012, 2013) have shown that it is very difficult to categorically classify sources as high-redshift galaxies and not low-redshift interlopers using photo-

cs or S/N criteria on the dropout bands.

Previous work using photometric redshifts has dealt with this problem by making use of the full redshift PDF when calculating LF (Dahlen et al. 2005; McLure et al. 2009, 2013), thereby incor-

porating the uncertainty in the analysis. Due to the nature of the SED fitting code used for this work (described in Section 4), the computational effort required to fit the mass at each redshift in or-

der to integrate over the full PDF becomes impractical. As such, we chose to account for these problems in a different manner whilst still dealing with them in a straightforward way.

Rather than using only the best-fitting redshift from our photo-

metric analysis when selecting our sample, we instead draw the redshift for each galaxy randomly from its full PDF before placing it in the appropriate redshift sample. Where secure spectroscopic redshifts are available, we fix the redshift to that value for all sam-

ples (known interlopers are therefore excluded in all samples). This process was repeated 500 times to produce a set of samples to which we then apply the rest of the analysis described in the paper sep-

arately. We then average over the results from each sample, using the mean of this full set as our ‘true’ value along with the 1σ upper and lower limits around this mean.

The resulting sample sizes for each redshift bin are shown in Table 1. The varying samples account for both scattering between redshift bins for objects at the boundaries as well as objects moved out of the sample into secondary low-redshift solutions. The effect of this scattering into and out of the samples can be seen when
Table 1. Average sample size and variance for each redshift bin for the 500 MC samples generated, see text for details.

| Redshift bin | Mean sample size | Variance on sample size |
|--------------|------------------|-------------------------|
| 4            | 1235             | 180                     |
| 5            | 416              | 63                      |
| 6            | 169              | 25                      |
| 7            | 42               | 9                       |

Comparing the combined mean samples sizes (1862) to our full high-redshift sample of 2263.

The strength of using photometric redshifts for sample selection over colour cuts (especially when redshifts would still need to be calculated for a colour-cut sample in order to do SED fitting) is that the method can automatically make use of all available photometry. This is important because although photometric redshifts are still fitting primarily to the characteristic break at Lyman $\alpha$ targeted by the colour selections, the filters long-ward of the break are useful in excluding low-redshift interlopers (McLure et al. 2011). Additionally, the large errors in colour possible due to low S/N and possible non-detections in the filter just short of the Lyman break means that likely high-redshift candidates can be scattered well outside the selection region when using colour cuts.

Fig. 2 shows the positions of our galaxy samples on the colour–colour planes commonly used to select dropout samples. It is obvious that many of the galaxies selected with photometric redshifts lie outside the selection regions (as taken from Bouwens et al. 2007), especially those galaxies where colours must be calculated using an upper limit. The agreement (or lack of) between dropout selections and photometric redshifts has also been investigated for GOODS-S specifically by Dahlen et al. (2010).

To test whether the discrepancy between the observed colours and those required for Lyman break selection can be explained...
solely by photometric scatter, we performed a range of tests on a mock catalogue generated from the semi-analytic models (SAM) described in Somerville et al. (2008) and Somerville et al. (2012). Full details of these tests are outlined in Appendix A; however, our main finding is that the observed colours can be reproduced from intrinsic Lyman break-like colours and scattering proportional to the observed photometric errors.

4 MASS FITTING

Stellar masses were estimated for our samples using a custom template-fitting code with SEDs derived from the synthetic stellar population models of Bruzual & Charlot (2003, BC03 hereafter). Due to the relatively young ages of the stellar populations (as constrained by the age of the Universe at the redshifts involved), the effects of thermally pulsating asymptotic giant branch stars (TP-AGB) on resulting SEDs are minimal, e.g. Stark et al. (2009). As such, we chose not to use the updated SSP models which incorporate stronger contributions for these effects. The use of Bruzual (2007, hereafter CB07) or Maraston (2005) in place of BC03 should have no result on the conclusions found in this work.

4.1 Model SEDs

First, model SEDs are generated from the single stellar populations of Bruzual & Charlot for a range of population ages, metallicities and SFHs. For our models and throughout this work, we use the initial mass function (IMF) of Chabrier (2003). Each template is then normalized such that the total stellar mass equals 1 M⊙. Nebular emission lines are added to the pure stellar component following the method outlined in Section 4.2. Internal dust extinction is applied following the extinction law of Calzetti et al. (2000) for the desired range of extinction magnitude Aν.

When applying dust extinction to the nebular emission, we assume the differential dust extinction between stellar and nebular emissions to be fixed as E(B-V)stellar = E(B-V)nebular, in contrast to the ratio of E(B-V)stellar = 0.44E(B-V)nebular derived by Calzetti et al. (2000). This choice was motivated by the conflicting evidence for the relative extinction of the two emission sources at z ~ 2 (Erb et al. 2006; Förster Schreiber et al. 2009) and that in the context of these model-specific, the assumed differential extinction ratio and escape fraction are degenerate.

The absolute magnitude at 1500 Å (MUV) is measured for each template by integrating the flux within a 100 Å-wide flat bandpass centred on 1500 Å. Similarly, the UV-continuum slope β is calculated by fitting a simple power law to the integrated template fluxes centred on 1500 Å. Similarly, the UV-continuum slope is given by

\[ L(ν) = 4.78 \times 10^{-13}(1 - f_{\text{esc}})N_{LyC} \] (4)

from Krueger, Fritz-v Alvensleben & Loose (1995), where fesc is the continuum escape fraction and the number of Lyman continuum photons N_{LyC} is calculated from each template. The strength of the nebular emission is therefore directly proportional to the number of ionizing photons (Lyman continuum) in the H II region. Since the Lyman continuum emission is dominated by young massive stars, the relative contribution of nebular emission to the total SED is highly dependent on the age of the stellar population and the amount of recent star formation.

Line ratios for the common metal lines relative to Hβ were taken from the empirical measurements of Anders & Alvensleben (2003) for each of the input template metallicities, assuming gas metallicity is equal to the stellar metallicity. Similarly, the nebular continuum emission luminosity is given by

\[ L_V = \frac{\nu^{\text{total}}}{\alpha_B}(1 - f_{\text{esc}})N_{LyC}, \] (5)

where α_B is the case B recombination coefficient for hydrogen and γ_v(10) is the continuum emission coefficient given by

\[ γ_v(10) = γ_v(10)_H + γ_v(10)_e + γ_v(2p)_n + γ_v(2p)_H + γ_v(2p)_He + γ_v(2p)_He^+ + γ_v(2p)_He^++. \] (6)

\[ γ_v(10)_H, γ_v(10)_e, γ_v(2p)_n, γ_v(2p)_H, γ_v(2p)_He, γ_v(2p)_He^+ \] are the continuum emission coefficients for free–free and free–bound emission by hydrogen, neutral helium, singly ionized helium and two-photon emission for hydrogen, respectively, where the values are taken from Osterbrock (1989) and are constant across the redshift range.

4.3 SED Fitting

The fitting of our SEDs to the observed photometry is done using a Bayesian-like approach, whereby the normalized likelihood \( L(M, t) \) for a given stellar mass, M, and template type, t, is given by

\[ L(M, t) = \frac{\exp\left(-\frac{1}{2} \chi^2(M, t)\right)}{\sum_t \int dM' \exp\left(-\frac{1}{2} \chi^2(M', t')\right)}. \] (7)

The \( \chi^2 \) value is given by

\[ \chi^2(M, t) = \sum_j \frac{(M F_j(t) - F_{\text{obs}})^2}{\sigma_j^2}. \] (8)

where \( F_j(t) \), \( F_{\text{obs}} \) and \( \sigma_j \) are the template flux, observed flux and the observed flux error in the jth filter, respectively. The template types, t, and their associated fluxes correspond to the full range of galaxy parameters (age, SFH, dust extinction and metallicity) at the closest matching redshift in the model SED grid.

Because we fit all templates simultaneously, it is therefore straightforward to calculate the stellar mass PDF; i.e.

\[ P(M) \propto \sum_t L(M, t). \] (9)

marginalized over all other template galaxy properties (assuming a flat prior). Similarly, PDFs for other parameters such as β or Hβ for the major Balmer, Paschen and Brackett recombination lines are taken from Osterbrock & Ferland (2006), with the total Hβ line luminosity (in erg s\(^{-1}\)) given by
$M_{UV}$ ($M_{1500\AA}$) can be constructed by summing the likelihoods at a fixed parameter value. Estimating the galaxy parameters in such a way allows us to fully account for errors due to both degeneracies between galaxies properties and errors in the scaling due to the photometric errors.

For our mass fitting, model ages are allowed to vary from 5 Myr to the age of the Universe at a given redshift, dust attenuation is allowed to vary in the range $0 \leq A_V \leq 2$ and metallicities of 0.02, 0.2 and 1 $Z_{\odot}$. Due to the difficulty in obtaining spectroscopy at $z > 3$, the metallicity at high redshift is not currently well known. Observations of samples at $z \approx 3$ and above (Shapley, Steidel & Pettini 2003; Maiolino et al. 2008; Jones, Stark & Ellis 2012; Sommariva et al. 2012) show that the average metallicity is likely to be mildly sub-solar; however, there is a large scatter. Sommariva et al. (2012) also find that the gas-phase and stellar metallicities are consistent within errors. As such, we choose to fix the metallicity for the nebular emission equal to stellar metallicity.

The SFHs follow the exponential form $SFR \propto e^{-t/\tau}$ with characteristic time-scales of $\tau = 0.05, 0.25, 0.5, 1, 2.5, 5, 10, -0.25, -0.5, -1, -2.5, -5, -10$ and 1000 (effectively constant SFR) Gyr. Negative $\tau$ values represent exponentially increasing histories. Fitting is done to the templates both with and without the inclusion of nebular emission. When nebular emission is included in the templates, we assume a moderate escape fraction $f_{esc} = 0.2$, consistent with the observational constraints on reionization and with simulations (Yajima, Choi & Nagamine 2010; Fernandez & Shull 2011; Finkelstein et al. 2012b; Robertson et al. 2013).

### 4.4 Star formation rates

In order to calculate UV SFR, the rest-frame absolute magnitudes ($M_{1500\AA}$) measured from the SED fitting are first corrected for dust extinction using the Meurer, Heckman & Calzetti (1999) relation

$$A_{1600} = 4.43 + 1.99\beta,$$

which links the observed UV continuum slope $\beta$ as measured by the SED fitting code (see Section 4.1) and the extinction at 1600 Å, $A_{1600}$. For measured $\beta < -2.23$, where the above relation would imply a negative extinction, the UV extinction was set to 0. UV SFR are calculated using

$$\text{SFR}(M_{\odot} \text{yr}^{-1}) = \frac{L_{UV} \text{(erg s}^{-1} \text{Hz}^{-1})}{13.9 \times 10^{27}},$$

where the $L_{UV}$ conversion factor of Madau, Pozzetti & Dickinson (1998) and Kennicutt (1998) corrected to the Chabrier IMF is used ($\sim 0.24$ dex).

In addition to the SFR obtained by this method (SFR$_{Madau}$ hereafter), from our SED fitting code we also obtain the instantaneous SFR of the best-fitting template for each galaxy (SFR$_{Template}$). We find that the two measures agree well at all SFRs with a median($\log_{10}(\text{SFR}_{Madau}) - \log_{10}(\text{SFR}_{Template})$) < 0.1. With the exception of the few galaxies with the highest SFR$_{Madau}$, typically SFR$_{Madau} > 100$ $M_{\odot} \text{yr}^{-1}$. These galaxies are red, such that the best-fitting template is an older quiescent stellar population. The Meurer relation however assumes an actively star-forming population with high dust extinction.

We also find that the scatter around the 1:1 relation correlates strongly with the age of the best-fitting SED template, such that younger populations have higher SFR$_{Template}$. As we will show in the next section, however, individual stellar population parameters such as age and dust extinction are very degenerate in SED fits of high-redshift galaxies. Because of these factors, and for consistency with previous works, we primarily use SFR$_{Madau}$ throughout this work. The net effect of the differences in the two SFR estimates can be seen in our observed SFR functions in Section 4.4.

### 4.5 Image and detection simulations

By their nature, high-redshift galaxies are small and extremely faint objects. Lying close to the limiting depth in some (or even all) of the observed filters, noise and systematic effects can have a significant effect on the detection and completeness of high-redshift galaxy samples as well as the accurate estimation of their properties. The completeness of our galaxy sample can be separated into two distinct factors: first, the inclusion of an object in the initial catalogue as a function of the detection image depth, and secondly, the selection of an object in a given sample (e.g. $z \approx 4$ based on its estimated redshift), which is a function of the overall SED shape and accompanying errors. In this section, we outline a set of detailed simulations undertaken to measure and correct for these effects.

#### 4.5.1 Completeness simulations

The detection completeness across the field was estimated by inserting thousands of mock galaxies into the detection image ($H$ band) used for the photometry and recovering them with the same SExtractor parameters and method used for the original sample catalogue. The synthetic galaxies were first convolved with the CANDELS WFC3 $H$-band PSF before being placed randomly across the field with appropriate Poisson noise. The resulting image was then run through the same SExtractor procedure as the initial source detection and the process repeated until a total of $\approx 10^5$ input galaxies had been recovered across the entire field.

Galaxy sizes were drawn from a log-normal distribution of mean $= 0.15$ arcsec and $\sigma = 0.075$, motivated by existing observations of the size evolution of LBGs (Ferguson et al. 2004; Oesch et al. 2010; Grazian et al. 2011; Huang et al. 2013) whilst the galaxy profiles were drawn from a distribution of Sérsic indices centred around $n = 1.5$ in the range $0.5 \leq n \leq 4.0$. Although the precise distribution of morphological profiles for high-redshift galaxies is not well known, studies of lower redshift analogues and stacked samples of LBGs suggests that they are predominantly disc-like ($n < 2$) (Ravindranath et al. 2006; Hathi et al. 2007). Our chosen distribution reflects this, with $\sim 80$ per cent of input galaxies with $n \leq 2$.

Fig. 3 shows the resulting completeness curves for each of the image regions. In Guo et al. (2013), the $H_{160}$ 50 per cent completeness limit is estimated using the differential number density to be 25.9, 26.6 and 28.1 for the WIDE, DEEP+ERS and UDF regions, respectively. When compared to the results of a set of detection simulations similar to those undertaken in our work, Guo et al. (2013) find good agreement between the two estimates. For the UDF and DEEP+ERS regions, our 50 per cent completeness limits are in good agreement with those of Guo et al. (2013). However, in the WIDE region we find a 50 per cent completeness limit $\approx 0.5$ mag deeper for our input galaxy population.

Grazian et al. (2011) demonstrated the significant effect that sizes and morphologies can have on the completeness simulations of high-redshift galaxies. Differences due to the distribution of sizes and slightly differing galaxy profiles used are to be expected. For all regions, the effects of confusion and blending with nearby sources results in a small fraction of input galaxies which are not recovered by the photometry, even at brighter magnitudes.
The mass evolution of the first galaxies

4.5.2 Sample selection

To estimate the selection functions for each of the redshift bins, a mock photometry catalogue of high-redshift galaxies was created and put through the same photometric redshift and sample selection criteria as our real sample. This catalogue was constructed by first creating a sample of SEDs drawn randomly from the template sets used for fitting (both with and without nebular emission) with a distribution of $\beta$ centred at $\approx -1.8$ to $-2$, but extending out to $\beta > 1$. Redshifts were allowed to vary in the range $2.5 < z < 9$ and the templates were scaled to $H_{160}$ band magnitudes in the range $22 < H_{160} < 30$, with the corresponding magnitudes in the other filters determined by the shape of each SED.

We produced a catalogue in this way, rather than using the mock photometry of SAM as used in Appendix A in order to allow the inclusion of nebular emission in subsequent tests on the stellar mass fitting and ensure good number statistics across all input magnitudes.

In order to assign photometric errors to the mock photometry (or non-observations where appropriate), each simulated galaxy was assigned a position in the field drawn from the same set of input coordinates as used in the completeness simulations. Photometric errors were then assigned to each photometric band based on the observed flux errors of objects in the original catalogue, specific to the region in which it resides (e.g. CANDELS Deep). The flux values for each SED were then perturbed by a Gaussian of width equal to the photometric error.

This process does not precisely mirror the method used to produce the observed photometry as it does not include the source extraction for each band individually. However, the resulting catalogue is a very close approximation with a catalogue of SEDs that have realistic photometric errors and filter coverage across the field, e.g. $Y_{098}$ observations in the ERS region alone.

To measure the selection efficiency for our high-redshift samples, 100 simulated Monte Carlo (MC) samples were created from the template based mock galaxy catalogue (as described in this section) using the method outlined in Section 3.1. From these samples, we measured the fraction of simulated galaxies which pass the selection criteria for any of the high-redshift samples as a function of input redshift and magnitude.

Fig. 4 shows the measured selection efficiencies for the deepest region of the field, the UDF. The selection probabilities (as indicated by the colour scale) do not include the effects of completeness as measured in Section 4.5.1; therefore, the lower probabilities measured at faint magnitude are a result of photometric redshift errors due to poor constraints from faint photometry.

![Figure 3](image_url)  
*Figure 3.* Completeness as a function of $H_{160}$ magnitude for each region of the GOODS South field. The vertical dashed lines show the magnitude at which the recovery fraction equals 0.5 for each region of the field.

![Figure 4](image_url)  
*Figure 4.* Example selection efficiencies for the UDF region of the CANDELS field. The colour scale represents the fraction of input galaxies which pass the $P(z)$ criteria for a given redshift bin as a function of input redshift and apparent magnitude. The dashed white line in the lower sections of the figure shows the 80 per cent contour in the fraction of recovered galaxies. The upper panels show the recovery fraction as a function of redshift at a fixed input magnitude, $H_{160} = 25$ (continuous) and $H_{160} = 27$ (dashed).
For $z \sim 4$ and 5, where the semi-analytic mock catalogue used in Appendix A has good number statistics across a wide magnitude range, we reproduce the selection function in the same manner as above and find that the shape of the resulting selection functions are unchanged. We are therefore confident that the photometric selection of our samples is robust to variations in the exact shape of the input SEDs and the limiting factor in selection is the photometric noise.

4.5.3 Uncertainties in measuring galaxy parameters

The ability of SED fitting codes to recover the properties of dropout galaxies was well explored by Lee et al. (2010) who found that stellar mass was the most reliably measured parameter (in comparison to SFR and age) and the most robust to assumptions in SFH. However, this analysis was limited to input and fitted SED models, which did not include the effects of nebular emission. The degeneracies in measuring age, dust extinction and SFHs from SED fitting have also been well examined (e.g. Schaerer & de Barros 2010). Despite these degeneracies, it has been shown that one can reliably measure the UV continuum slope, $\beta$ (Finkelstein et al. 2012a; Rogers et al. 2013). Given assumptions about the age and metallicity, i.e. the underlying intrinsic $\beta$, it is then possible to estimate the dust extinction using observations of $\beta$.

For the $\sim 10^4$ galaxies in our simulated catalogue which pass the selection criteria, we ran them through the SED fitting code using the same fitting parameters as for our observed data. From these results we are able to test how well the input stellar masses are recovered for our simulated galaxies. In addition, we can also test the accuracy in recovering the other properties of the input stellar populations.

Fig. 5 illustrates how well the SED code is able to recover the stellar masses, ages and dust extinction. As expected, stellar mass is the most robust of the parameters with age and dust extinction showing a very large scatter and bias due to the degeneracy in fitting. Despite these degeneracies in the single best-fitting templates, when calculating the marginalized $\beta$ over all template likelihoods the resulting estimate of $\beta$ is unbiased and well constrained. We show the estimated accuracy of our $\beta$ measurements from these simulations in Appendix B.

For input galaxies with masses $\approx 10^9 M_\odot$, the median $(\log_{10}(M_{\text{out}}) - \log_{10}(M_{\text{in}})) = 0.02$, with a standard deviation of 0.4 when input SEDs including nebular emission are fitted with comparable templates. For input masses $\approx 10^{10} M_\odot$ and below, both the bias ($+0.22$ dex at $\approx 10^{10} M_\odot$) and scatter increase. When mock galaxies with pure stellar SEDs are fitted with pure stellar templates, both the scatter and bias are reduced at all mass ranges. The increased bias and scatter for galaxies with nebular emission is a result of confusion between an older stellar population with a 4000 Å break and a young star-forming galaxy with strong nebular emission (Schaerer & de Barros 2009; Curtis-Lake et al. 2013).

Finally, we find that the recovered value for $M_{\text{UV}}$ is extremely robust across the full dynamic range of our data, with a scatter of $<0.2$ dex and negligible bias across all redshift out to the limits of our completeness as shown in Fig. 6. From these simulations, we determine that $M_{\text{UV}}$ is robust to $M_{\text{UV}} \approx -17$ at $z \approx 4$, reducing to $\approx -18$ at $z \approx 7$.

5 RESULTS

5.1 The $1/V_{\text{max}}$ estimator

To compute our luminosity and mass functions we use an extension of the $1/V_{\text{max}}$ method of Schmidt (1968), treating each of our high-redshift samples as a ‘coherent’ sample comprised of the individual GOODS South regions with corresponding depths as outlined by Avn & Bahcall (1980), Eales (1993) and Ilbert et al. (2005). The maximum comoving volume in which a galaxy can be observed and remain in the sample is given by

$$V_{\text{max},i} = \sum_k \frac{N_{\text{reg},k}}{k} \int_{z_{\text{1},k}}^{z_{\text{2},k}} \frac{dV}{dz} dz d\Omega_k,$$

where the sum, $k$, is over each of the sub-regions in the field with their corresponding solid angle, $d\Omega_k$, integration limits $z_{\text{1},k}$, $z_{\text{2},k}$ and $dV/dz$ is the comoving volume element in Mpc$^3$ at redshift $z$. The integration limits are given by $z_{\text{1},k} = z_{\text{min}}$ and $z_{\text{2},k} = \min(z_{\text{max}}, z_{\text{3},k}, m_{\text{max},k})$, where $z_{\text{min}}$ and $z_{\text{max}}$ are the lower and upper boundaries, respectively, for the given redshift bin, e.g. $z_{\text{min}} = 4.5$ and $z_{\text{max}} = 5.5$ for the $z \approx 5$ sample. The
The mass evolution of the first galaxies

Figure 6. Comparison of the recovered versus input $M_{\text{UV}}$ for the full mock galaxy sample. As in Fig. 5, the histogram is normalized by the number of input galaxies in each bin and the colour scale corresponds to the fraction of input galaxies at the observed $M_{\text{UV}}$.

The mass (or luminosity) function $\phi_k$ for discrete bins of mass (luminosity) $k$ is then:

$$\phi_k \, dM = \sum_i w_i \frac{V_{\text{max},i}}{V_{\text{max}}} \int_{M_k}^{M_i} P_i(M) dM,$$

where the weighting term, $w_i$, incorporates corrections for incompleteness and the selection function of the redshift bin as calculated in Section 4.5. The window function $W$ is defined as

$$W(x) = \begin{cases} 1 & \text{if } -dM/2 \leq x < dM/2 \\ 0 & \text{otherwise} \end{cases}$$

and $N_{\text{gal}}$ is the number of galaxies in the sample.

To incorporate the large error in the stellar masses, where the mass likelihood function for an individual galaxy can span a range much larger than the desired bin widths, we make amendments to equation (13), such that the mass function evaluated for the mass bin $M_1 < M < M_2$ is given by

$$\phi(M) dM = \sum_i w_i \frac{V_{\text{max},i}}{V_{\text{max}}} \int_{M_1}^{M_2} P_i(M) dM,$$

where $P_i(M)$ is the probability of galaxy $i$ having stellar mass, $M$, as calculated from the SED fitting at the fixed redshift for that specific MC sample.

5.2 UV LFs

As a more robust observable (relative to the stellar mass, Section 4.5) with many previous observations, the rest-frame UV LF provides a useful comparison for the method and completeness corrections used in this paper. To ensure that the shapes of our observed mass functions are not affected by biases in our $1/V_{\text{max}}$ or completeness correction methods we reproduce the LF for each of our redshift bins, which we can compare to previous work. Fig. 7 shows the discretized LF calculated using the method outlined in equation (13). In comparison with previous measurements of the LF at high redshift, we find overall very good agreement in the general shape of the LF for our $1/V_{\text{max}}$ data points.

Figure 7. A comparison of our $1/V_{\text{max}}$ LF estimates with those in the literature. We show the results of Bouwens et al. (2007) at $z \approx 4, 5$ and $6$ (red squares) derived from deep HST observations as well as the ground-based estimates of the bright end of the $z \approx 5$ and $6$ LF by McLure et al. (2009)(green downward triangles). For the $z \approx 7$ LF, we show the estimate of Bouwens et al. (2011) as well as the recent results of McLure et al. (2013) and Schenker et al. (2013a) which make use of the deeper UDF12 observations (Koekemoer et al. 2013) to probe fainter $M_{\text{UV}}$ than we are able to.
The data points for this work were also fit through χ²-minimization with the Schecter (1976) parametrization

\[ \phi(M) = \phi^* \log(10) \frac{2.5}{2} 10^{-0.4(M-M^*)/(\alpha+1)} \exp(-10^{-0.4(M-M^*)}), \]

(16)

where \( M \) is the rest-frame UV magnitude and \( \phi^*, \alpha \) and \( M^* \) are the normalization, faint-end slope and characteristic UV magnitude as standard. The resulting best-fitting parameters shown in Table 2 along with the corresponding parameters from the selected literature observations shown in Fig. 7. At all redshifts, the faint-end slope is steep (\( \alpha \approx -1.6 \)) and shows a tentative steepening towards \( z \approx 7 \). However, the poor constraints on the faint-end slope towards the highest redshifts make it difficult to comment on any evolution of the slope that might occur over this redshift range. At all redshifts, the slope is consistent with a fixed slope of \( \alpha = -1.7 \) (Bouwens et al. 2011).

At \( z \approx 4 \), our measured \( M^*_\text{UV} \) of \(-20.47 \pm 0.21\) is significantly fainter than that observed by Bouwens et al. (2007), who find \( M^*_\text{UV} = -21.06 \pm 0.1\). We find closer agreement with the fainter \( M^*_\text{UV} \) observed by Huang et al. (2013) of \(-20.60_{-0.13}^{+0.17}\). At the bright end of the \( z \approx 4 \) LF, our fit is strongly affected by the very low number density at \( M\text{UV} \approx -22.25 \). Given the relatively small area of our survey field, the numbers of galaxies contributing to the brightest bins are very small (\( \approx 1-3 \)). Differences in the measured (or assumed) redshift when calculating the rest-frame magnitude can therefore have a large effect, e.g. the difference in the distance modulus between \( z = 3.5 \) and \( z = 4 \) is \( \approx 0.35 \). As such, the characteristic cut-off in our Schecter function parametrization is not well constrained, and the discrepancy is not significant.

### 5.3 Observed mass-to-light ratios

In the past, the relationship between a galaxy’s stellar mass and its UV luminosity (or \( \log_{10}(M_\text{stars}) \) and \( M_{\text{UV}} \)) has been used as both a diagnostic of galaxy formation histories (Stark et al. 2009) and as a tool for estimating the galaxy SFR at high redshift (González et al. 2011). In a scenario where galaxies form their stars continuously, a strong \( \log_{10}(M_\text{stars})-M_{\text{UV}} \) relation should form, whereas more stochastic bursty star formation modes could result in a relation with wider scatter and a weaker trend.

Using the stellar mass and \( M_{\text{UV}} \) probability distributions produced by our SED fitting code, we plot the observed mass-to-light ratios for each of our redshift bins in Fig. 8. For a given MC sample (see Section 3.1), the 2D \( \log_{10}(M_\text{stars})-M_{\text{UV}} \) probability distributions of each galaxy in the redshift bin are summed. The resulting PDFs of each sample are then summed to create a combined PDF in each redshift bin across all MC samples. Finally, we normalize such that the probability at each value of \( M_{\text{UV}} \) integrates to unity. By plotting the observed \( \log_{10}(M_\text{stars})-M_{\text{UV}} \) in this way, we take into account the full redshift and fitting errors. However, this representation is still subject to the effects of small number statistics for the brightest and faintest galaxies.

As such, we also show the biweight mean \( \log_{10}(M_\text{stars}) \) for bins of \( M_{\text{UV}} \) within the range of reliably measurable \( M_{\text{UV}} \). For \( z \geq 5 \), each bin contains a minimum of five galaxies by design, whilst for the \( z \approx 4 \) sample each bin contains a minimum of 10 galaxies. To these means, we fit linear functions with intercept \( \log_{10}(M_\text{stars})_{\text{MC}} \sim (-19.5) \) and slope \( d \log_{10}(M_\text{stars})/dM_{\text{UV}} \). The best-fitting values for each redshift sample are shown in Table 3 for stellar mass estimates both with and without the inclusion of nebular emission.

As has been seen in many previous studies, we observe a clear ‘main-sequence’ trend of increasing mass with increasing UV luminosity, and a large scatter about this trend. For the bright galaxies (\( M_{\text{UV}} < -21 \)), our results agree well with those of Stark et al. (2009), González et al. (2011) and Lee et al. (2012). Over the full range of UV luminosity, we find a shallower trend with \( M_{\text{UV}} \) than González et al. (2011). We also find that this trend evolves in normalization between redshift \( z \approx 4 \) and 7.

The change in the observed normalization of the \( \log_{10}(M_\text{stars})-M_{\text{UV}} \) relation with redshift as a consequence of the inclusion of nebular emission has been examined before (Shim et al. 2011; Schenker et al. 2013b; Stark et al. 2013; de Barros, Schaerer & Stark 2014). However, we find that although this trend for decreasing normalization with redshift is enhanced when nebular emission is included in the mass fitting, the trend still exists when fitting with pure stellar templates (see Table 3).

In Fig. 8, the effect of including nebular emission in the stellar mass estimate can be seen clearly in the bottom panel. At all redshifts, the average stellar mass for a given \( M_{\text{UV}} \) is lower when nebular emission is included. Salmon et al. (2014) also consider the effects of adding nebular emission lines to the SED models for galaxies in the same redshift range, and they find similar changes to the derived masses as we do here. In addition, the median stellar masses they observe for UV-faint galaxies are higher than those of Gonzalez et al. (2011) and Lee et al. (2012), consistent with our observations.

Because in our SED fitting on the mass, we restrict the age of the templates to be less than the age of the universe at that redshift, the range of \( \log_{10}(M_\text{stars})-M_{\text{UV}} \) ratios available in the fitting does vary with redshift, i.e. a galaxy at \( z = 7 \) can never have as old as a stellar population as a galaxy at \( z = 4 \). If the fits to galaxies at \( z \approx 6 \) and 7 were being restricted by this upper limit, the limits set by the template set could create an artificial evolution in the scaling of the \( \log_{10}(M_\text{stars})-M_{\text{UV}} \) relation with redshift. However, examining the best-fitting SED parameters across all of the MC samples, we find that at all redshifts and \( M_{\text{UV}} \) values the highest best-fitting mass lies...
The mass evolution of the first galaxies

Figure 8. Top: probability distribution of the mass-to-light ratios observed when nebular emission is included in the fitting, stacked across all of the MC samples. The values are normalized such that the probability at each value of $M_{UV}$ integrates to unity. The blue dot–dashed line represents the average of the best-fitting line to robust means in each of the MC samples, with the corresponding average means and their errors shown by the blue circles. The $z \sim 4$ relation is shown for reference at high redshifts (cyan dotted line). Bottom: the corresponding probability distributions, bi-weight means and best-fitting relation (red dot–dashed line) when nebular emission is excluded from the SED fitting. The blue dotted line shows the best-fitting relation from the top panel (including nebular emission) for each sample. In both panels, the orange dashed line shows the mass-to-light ratio observed by González et al. (2011), measured for their $z \approx 4$ sample and applied across all bins. The green triangles and yellow squares show the average stellar mass in $M_{UV}$ bins as calculated by Lee et al. (2012) and Stark et al. (2009), respectively, all stellar masses have been converted to the same Chabrier IMF. The grey dotted line represents the template in our SED fitting parameters with the lowest mass-to-light ratio.

well below the maximum mass allowed by the template set. From this, we conclude that the observed scaling is therefore physical and not a result of systematics in our analysis.

The slopes of our fitted $\log_{10}(M_*) - M_{UV}$ relations are all close to that of a constant mass-to-light ratio ($M_{\infty} - 0.4 M_{UV}$) across the full range in luminosity. This implies there is no strong evolution of the mass-to-light ratio with luminosity. Lee et al. (2012) suggested the source of the change in their observed $\log_{10}(M_*) - M_{UV}$ ratio could be due to a luminosity dependent extinction, a result which had also been implied by the evolution of $\beta$ with $M_{UV}$ seen by Bouwens et al. (2012b). Subsequent observations by Dunlop et al. (2011) and Finkelstein et al. (2012a) have found no obvious luminosity dependence. However, recent studies by Bouwens et al. (2013) and (Rogers et al. 2014) with greatly increased sample sizes and greater dynamic range confirm an unambiguous colour–magnitude relation. While measurements of $\beta$ for our sample do not exhibit strong evidence for such a strong luminosity dependent extinction at any redshift (see Appendix B), our sample does not contain a statistically significant number of the brightest and faintest galaxies to rule out such evolution given the large scatter and error on $\beta$.

Due to the increasing uncertainty in stellar mass measurements for galaxies below $10^9 M_\odot$, the average mass-to-light ratios for the faintest galaxies could become increasingly biased towards fainter UV luminosities. As such, we cannot rule out a change in the
Table 3. The best-fitting slope and intercepts of the log10(M*) − M_{UV} mass-to-light relation, averaged across all MC samples. At z ∼ 4 and 5, we also show in parentheses the best-fitting values when the fits are restricted to only the brightest galaxies (M_{UV} < −19.5).

| z  | log10 M_*(M_{UV} = −19.5) | dlog10 M_*/dM_{UV} |
|----|--------------------------|-------------------|
| With nebular em. | | |
| 4  | 9.02 ± 0.02              | −0.45 ± 0.02      |
| 5  | 8.84 ± 0.04              | −0.47 ± 0.04      |
| 6  | 8.64 ± 0.12              | −0.48 ± 0.07      |
| 7  | 8.49 ± 0.09              | −0.44 ± 0.12      |
| Without nebular em. | | |
| 4  | 9.10 ± 0.02              | −0.39 ± 0.02      |
| 5  | 9.00 ± 0.04              | −0.46 ± 0.04      |
| 6  | 8.84 ± 0.07              | −0.54 ± 0.07      |
| 7  | 8.63 ± 0.11              | −0.45 ± 0.13      |

5.4 SMF at high redshift

Following the method outlined in Section 5.1, specifically using the 1/V_{max} method in equation (15), we construct the SMF for each of our high-redshift samples. The resulting SMF are shown in Fig. 9. Our data points and errors take into account the stellar mass log10(M_*) − M_{UV} slope at faint luminosities like that inferred by Lee et al. (2012). To better constrain the average mass-to-light ratio for faint galaxies, detailed stacking across the full SEDs as a function of M_{UV} would be required. Restricting our analysis to the brightest galaxies (M_{UV} < −19.5) at z ∼ 4 and 5 where the potential biases are minimized, we find no significant change in the fitted log10(M_*) − M_{UV} slopes.

In hydrodynamical simulations of galaxies at z > 5, Wilkins et al. (2013) found a relationship between the intrinsic L_{1500} (excluding dust absorption) and M/L_{1500} which is roughly constant. This relationship is also seen to evolve, with the normalization decreasing with increasing redshift. When dust extinction was applied to the intrinsic model luminosities based on the β observations of Bouwens et al. (2012b), the observed log10(M_*) − M_{UV} exhibited a much stronger correlation comparable to that observed by González et al. (2011).

Figure 9. The 1/V_{max} SMF for the high-redshift samples. Error bars take into account random Poisson noise as well as the scatter between the MC samples due to photometric redshift uncertainty. The black circles show the mass bins included in the χ² fitting to the Schechter (1976) functions based on the stellar mass limits described in the text. The dashed and dotted lines show the SMF calculated by applying the best-fitting mass-to-light ratio (including nebular emission, see Table 3) to the literature LF at each redshift with a scatter of 0.2 (dashed) and 0.5 (dotted) dex. For the z ∼ 4 bin, the Schechter fit of Bouwens et al. (2007) was used, whilst at z ∼ 5 and 6 and z ∼ 7 the fits of Bouwens et al. (2012a) and McLure et al. (2013), respectively, were used to generate the luminosity distribution. We also show using the error bars at the top of each panel the cosmic variance expected for galaxies of stellar mass ≈10^{10.5} M_☉, as predicted by the method outlined in Moster et al. (2011).
As we are observing only a single field, we are unable to estimate the cosmic variance in the number densities by comparing the field to field variation. We use the updated QUICKCV code of Moster et al. (2011, see also Newman & Davis 2002) to estimate the cosmic variance as a function of mass in each of our redshift bins for a survey field with the dimensions of CANDELS GOODS South. In Fig. 9, we show the estimated error on the counts for galaxies of mass $\approx 10^{9.5} \, \mathrm{M}_\odot$. For stellar mass $\approx 10^{10} \, \mathrm{M}_\odot$ and above, the cosmic variance predicted by this method exceeds $100$ per cent at $z \sim 6$ and 7. However, due to the lack of constraints on the galaxy bias at high redshift, there is a large uncertainty on these estimates. When compared to the field-to-field variation observed by Lee et al. (2012), our estimates represent a conservative assessment of the likely cosmic variance. With the full CANDELS imaging now complete, future analysis incorporating all five of the separate survey fields should allow much more robust measures on the true cosmic variance at high redshift.

In addition to the $1/V_{\text{max}}$ estimates, we also estimate the SMF using a method analogous to that of Gonzalez et al. (2011). For each sample, 10$^6$ UV magnitudes in the range $-23 < M_{\text{UV,1500}} < -13$ are drawn from the observed LF from the literature at each redshift (Bouwens et al. 2007, 2012a; McLure et al. 2013). The $M_{\text{UV,1500}}$ are then converted to stellar masses using the best-fitting relations from Table 3 for each redshift sample and a scatter of 0.2 (dashed lines) or 0.5 dex (dotted lines).

At $z \sim 4$, the $1/V_{\text{max}}$ data shows good agreement with the mass functions generated by this method. At higher redshifts however, luminosity based mass functions increasingly underpredict the number density at high masses for the same fixed scatter in the $\log_{10}(M_*) - M_{\text{UV}}$ relation. To increase the number densities at high mass to match those observed, either a more strongly evolving $\log_{10}(M_*) - M_{\text{UV}}$ relation (in direct disagreement with that observed) or a significantly greater scatter in the intrinsic $\log_{10}(M_*) - M_{\text{UV}}$ ratios is required.

### 5.4.1 Comparison with the literature

Caputi et al. (2011) studied the massive end of the SMF at $3 \leq z \leq 5$ over a wide area in the UKIDSS Ultra Deep Survey (UDS), using photometric redshifts for a sample of 4.5 μm selected galaxies. Our observed number densities show broad agreement at $z \sim 4$ (3.5 ≤ $z < 4.25$ for Caputi et al.) for $\log_{10}M_* > 10.5$ but are significantly higher at lower masses. The same is also true across all masses at $z > 4$ (4.25 ≤ $z < 5.0$). However for both redshift samples, Caputi et al. (2011) find a very steep low-mass slope when parametrized with a Schechter (1976) fit, in agreement with our results. At the massive end of the $z \sim 4$ galaxy SMF, our results agree with those of Muzzin et al. (2013) over the limited mass range covered by both works.

Covering a significantly smaller area than those observations but probing to lower masses are the $1/V_{\text{max}}$ observations of Santini et al. (2012). We find good agreement with these results; however, due to the small number statistics at the high mass end of the SMF, the errors on both sets of observations are large.

Another measurement of the SMF at $z \sim 4$–5 is that of Lee et al. (2012), who study the SMF for a Lyman break selected sample in the GOODS North and South fields (Giavalisco et al. 2004a). At $z \sim 4$, for stellar masses $\log_{10}M_* > 9.5$ the two results are in excellent agreement. Below this mass range, the significantly steeper low-mass slope measured in this work results in a higher number densities than those found by Lee et al. (2012). At

### Table 4. Schechter (1976) function parameters for $\chi^2$ fits to the $1/V_{\text{max}}$ mass functions. For the $z \sim 5$, 6 and 7 samples, we do two fits, one in which $\log_{10}(M_*)$ is allowed to vary, and one in which it is fixed to the best-fitting value for the $z \sim 4$ sample. The quoted errors represent the 1σ errors from fitting marginalized over the remaining parameters but do not account for any systematic errors due to cosmic variance.

| $z$ | $\log_{10}(M_*)$ | $\alpha$ | $\phi^*$ ($10^{-4} \, \mathrm{Mpc}^{-3}$) |
|-----|-----------------|----------|-----------------------------------|
| 4   | 10.51$^{+0.36}_{-0.32}$ | $-1.89^{+0.15}_{-0.13}$ | 1.89$^{+3.46}_{-1.32}$ |
| 5   | 10.68$^{+0.08}_{-0.06}$ | $-1.74^{+0.41}_{-0.29}$ | 1.24$^{+7.77}_{-1.19}$ |
| 6   | 10.51              | $-1.64^{+0.15}_{-0.17}$ | 2.21$^{+0.80}_{-0.76}$ |
| 7   | 10.87$^{+1.13}_{-1.06}$ | $-2.00^{+0.57}_{-0.40}$ | 0.14$^{+4.11}_{-0.14}$ |
| 8   | 10.51              | $-1.90^{+0.27}_{-0.31}$ | 0.46$^{+0.36}_{-0.26}$ |
| 9   | 10.51              | $-1.89^{+1.39}_{-0.61}$ | 0.36$^{+3.01}_{-0.33}$ |

The black lines in Fig. 9 show the best-fitting Schechter (1976) functions from $\chi^2$ minimization to the $1/V_{\text{max}}$ data above our chosen mass completeness limits (black points). We perform two sets of fits to our data. First, we allow all three parameters to vary (solid line, $z \sim 4$, 5 and 6) and secondly, we fix the characteristic mass such that $M_* = M_{*\text{max}}$ (dashed line, $z \sim 5$, 6 and 7). The parameters for these fits are shown in Table 4.

Because there exists such a large scatter in the observed mass-to-light ratios for the high-redshift galaxies, accurately estimating the mass completeness limit is non-trivial. For a given mass near the completeness limit, there could exist a significant contribution from galaxies below the luminosity limit proportional to how large the actual intrinsic scatter is. However, it is not known how much of the observed scatter is due to photometric error (and therefore photometric redshift estimates and parameters from SED fitting) and how much is intrinsic. Rather than trying to correct for galaxies lost through incompleteness down to the lowest observed masses, we instead restrict our analysis to masses unaffected by this scatter. We calculate this mass limit by taking the 95 per cent mass percentile of the observed galaxies within 0.5 dex of our $H_{160}$ magnitude limit, finding it to be $\approx 10^{9.5} \, \mathrm{M}_\odot$. Under the reasonable assumption that the intrinsic scatter in the mass-to-light ratio does not rapidly increase below our detection limit, the contribution to masses above $\approx 10^{9.5}$ galaxies below the limit will be negligible.

Additionally, as shown in Section 4.5, the accuracy of stellar mass estimates begins to deteriorate at lower masses with an increasing bias which could lead to biased slopes. Taking these factors into account, the limits chosen when fitting the SMF were $\log_{10}(M_*) = 8.55, 8.85, 8.85$ and 9.15 for $z \sim 4, 5, 6$ and 7, respectively.

Inspecting the observed mass functions in Fig. 9, it can be seen that the exact limit should have little effect on the measured slope within reasonable bounds at $z \sim 4$–6. Choosing limits ±0.5 dex would not affect our conclusion that the mass function is steep.

Due to the small sample size at $z \sim 7$ and the large errors in estimating the stellar mass (from both the photometric redshift and fitting errors), the mass function is very poorly constrained. The range of acceptable values for $\alpha$ cover an extremely wide range but are consistent with the slope of $\alpha \approx -1.9$ found for the lower redshift bins and the slope of the corresponding LF. Over the redshift range examined by this work, the errors in $\alpha$ are too large to infer any evolution in slope with redshift.
mass slope of the SMF from $z = 0$ out to $z \sim 7$. We include only results where $\alpha$ has been fitted as a free parameter and the values of $\alpha$ quoted are from the single Schechter function parametrizations of the SMF.

At $z \sim 3$, there is broad agreement in the estimations of the low-mass slope at $\alpha \approx -1.8$. By $z \sim 4$, there is much larger disagreement between observed value spread across $\alpha \sim -2$ to $-1.4$. It is important to note that the observations with a shallower low-mass slope, González et al. (2011); Lee et al. (2012), are those with galaxies selected using the Lyman break technique colour cuts and source detection using optical bands (typically $z_{\text{lim}}$). In contrast, those with steep slopes, Caputi et al. (2011), Santini et al. (2012) and this work, use photometric redshift selection as well as near- or mid-infrared band for source detection. The best-fitting $\alpha$ of this work are also in good agreement with the maximum-likelihood estimates of Grazian et al. (in preparation), an independent analysis of the combined CANDELS GOODS and UDS fields.

5.4.2 Comparison with theory

In Fig. 11, we compare our measurements of the observed SMF with the predictions of both smooth particle hydrodynamic (SPH) and SAM. The SPH predictions are taken from the hybrid energy/momentum-driven wind model of Davé et al. (2013). We also show the predictions of three SAM, from Croton et al. (2006), Lu et al. (2011) and Somerville et al. (2008, see also Somerville et al. 2012). Details of the three models and an in-depth comparison between the model predictions across all redshifts can be found in Lu et al. (2013). The number densities have been renormalized to the comoving volume of the cosmology used throughout our paper ($H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_m = 0.3$ and $\Omega_{\Lambda} = 0.7$). Analysis is restricted to $z \lesssim 6$ due to limits on the robustness of simulations at higher redshifts from the numerical resolution of the simulations.

Inspecting the mass functions at $z \sim 4$, there is excellent agreement between the observations and the models of Davé et al. (2013), Croton et al. (2006) and to a lesser extent Lu et al. (2011). Of the three SAM predictions, Somerville et al. (2008) show the least agreement at $z \sim 4$ due to the overabundance of high mass galaxies. However, at higher redshifts the reverse is true, with the Somerville et al. (2008) models providing the best match to the observed $z \sim 5$ and 6 mass functions.

Figure 11. Comparison of the observed galaxy SMF in this work with theoretical model predictions at $z \sim 4$, 5 and 6. We show the SAM of Croton et al. (2006), Somerville et al. (2008) and Lu et al. (2011), using the error convolved SMF as outlined in Lu et al. (2013). The dashed purple line shows the results from the hydrodynamical simulations of Davé et al. (2013).
Of the four model predictions presented here, the SPH simulations of Davé et al. (2013) exhibit the steepest low-mass slopes and the closest agreement with our observations. The steepening of the low-mass slope in this model (from $z \sim 0$ to $z > 3$) is a result of decreasing contribution from wind recycling at high redshifts. The resulting feedback at high redshift has a smaller mass dependence than other models. This can be seen when compared to the SAM model of Lu et al. (2011) which has feedback with a much stronger mass dependence owing to increasingly strong (or efficient) feedback in low-mass haloes.

The SPH predictions, along with those of the Lu et al. (2011) SAM, most closely match the evolution in the overall normalization of the number densities across the observed redshift range. The other SAM undergo a much stronger evolution in the number density of the most massive galaxies. It is important to take into account the fact that all three of the SAM are tuned to match only the $z = 0$ SMF. The range of acceptable parameters at $z = 0$ found by Lu et al. (2011) results in a broad distribution at high redshift. Nevertheless, it is clear that our new observations of the high-redshift SMF can be used to further constrain our best models of galaxy evolution.

### 5.5 Stellar mass density

We compute the total SMD by integrating the fitted Schechter (1976) function from $M_* = 10^9$ to $10^{13} M_\odot$, with 1σ errors estimated from the minimum and maximum SMD within the 1σ contours for the fit parameters (see Table 5). For the $z \sim 5, 6$ and 7 samples, we use the best-fitting parameters with $M_* = M_{*\odot}$. The results are shown in Fig. 12 as the solid black points. We also show results from the literature across all redshift ranges, converted to the same cosmology and IMF (Chabrier/Kroupa).

Our observations show the continuation of the rapid decline in global SMD towards high redshifts, falling by a factor of between $\sim 4$ and 40 in the $\sim 1$ Gyr between $z \sim 4$ and 7. This rate of stellar mass growth observed is higher than observed by González et al. (2011) over the same time period but comparable to that found by Stark et al. (2013) when the large uncertainty in the $z \sim 7$ SMD is taken into account.

At $z \sim 4$, our results lie within the range of past SMD measurements at this redshift. Although larger than the results of the Lyman break selected samples (González et al. 2011; Lee et al. 2012; Stark et al. 2013), we find a SMD less than that of Santini et al. (2012) and comparable to that of some of the other photometric redshift selected samples (Pérez González et al. 2008; Marchesini et al. 2009). As could be inferred from the SMF, the SMD of Davé et al. (2013) underpredict observed SMD at $z \sim 5$ and $z \sim 6$ but shows good agreement at $z \sim 4$. Similarly, the range of densities covered by our luminosity based mass functions (grey regions) are significantly lower than the directly observed SMD in all redshift bins apart from $z \sim 4$.

### 5.6 Star formation rates

#### 5.6.1 Specific SFR

Earlier observations of the specific star formation rate (sSFR) evolution at $z > 3$, with mass estimates excluding the effects of nebular emission, showed the sSFR at a fixed mass remained roughly constant at $\sim 2$ Gyr$^{-1}$ with increasing redshift (Stark et al. 2009; González et al. 2010; Bouwens et al. 2012b). Such a plateau in the sSFR evolution was at odds with most plausible models of galaxy evolution (as explored by Weinmann, Neistein & Dekel 2011).

However, it has since been shown that the inclusion of nebular emission in stellar mass estimates at high redshift has a significant effect on the redshift evolution of the sSFR (Schaerer & de Barros 2009, 2010; Stark et al. 2013; González et al. 2014). By lowering the measured mass for a fixed SFR, the inclusion of nebular emission results in a higher sSFR proportional to the strength (or effect on the estimated stellar mass) of the emission lines.

In Fig. 13, we show our results for the sSFR (when using SFR$_{\text{gal}}$) in a stellar mass bin at $\log_{10}(M_*/M_\odot) = 9.7 \pm 0.3$ alongside previous observations at $z > 2$. We find an average sSFR of $2.32 \pm 0.08, 2.94 \pm 0.20, 4.21 \pm 0.54$ and $6.2 \pm 2.5$ Gyr$^{-1}$ for $z \sim 4, 5, 6$ and 7, respectively. Our observations show a clear trend in increasing sSFR with redshift in the redshift range $4 \leq z \leq 7$. The observed sSFR are in very good agreement with those of González et al. (2014) but are systematically lower than those of Stark et al. (2013) over the same redshift range. However, as noted in Stark et al. (2013), the introduction of 0.5 dex of intrinsic scatter to the $\log_{10}(M_*-M_{*\odot})$ used when estimating their sSFR would result in a reduction of $2.8 \times$ at $z \sim 4$. Such a large intrinsic scatter would be fully consistent with the $\log_{10}(M_*-M_{*\odot})$ relations and SMF observed in this paper. Taking this offset into account, the increasing consensus in the observed sSFR at high redshift is encouraging.

Table 5. SMD integrated from the Schechter parameters in Table 4 for $M > 10^9 M_\odot$. Error bars correspond to the minimum and maximum SMD within the 1σ contours of the mass function fits.

| $z$     | $\rho_\star(\log_{10}(M_*/M_\odot)\ Mpc^{-3})$ |
|---------|-----------------------------------------------|
| 4       | 7.36 ± 0.06                                   |
| 5       | 7.17 ± 0.07                                   |
| 6       | 6.76 ± 0.11                                   |
| 7       | 6.64 ± 0.39                                   |
z)}^{2.06 \pm 0.25} \text{ (black dotted line). This trend is much more consistent}
\text{ with theoretical expectations of the sSFR evolution than a plateau at}
z\sim4 \text{ to } 7 \propto (1+z)^{2.06 \pm 0.25}.

5.6.2 SFR functions and the cosmic SFR density

To measure the evolution of the SFR density for across our observed
redshift bins, we use the previously calculated 1/V_{\text{max}} values for
each galaxy to construct an SFR function analogous to the mass or LF
for the same data, such that

\phi_{\text{SFR},k}\text{d}e = \sum \frac{w_i}{V_{\text{max},i}} W(\epsilon_k - \epsilon_i),

(17)

where \epsilon = \log_{10}(\text{SFR}_{\text{UV}}). The SFR functions for our high-redshift
samples are shown in Fig. 14 for both the dust-corrected UV star
formation rates (SFR_{\text{Dust-corr}}) and the SED SFR as outlined in Section
4. At low to moderate star formation rates (log_{10}(\text{SFR}_{\text{UV}}) \leq 1.5),
the two SFR estimates are in good agreement across all redshifts
as was seen when the two estimates for individual galaxies were
compared.

The SFR function estimates of Smit et al. (2012), converted to
the same IMF used in this work) exhibit lower SFR than we observe at
all redshifts, with the exception of the SFR_{\text{template}} based estimate at
z\sim4 which shows excellent agreement across the full SFR range.
Smit et al. (2012) correct the observed UV LF (Bouwens et al. 2007,
2012a) to intrinsic magnitudes using the same Meurer et al. (1999)
relation as outlined in Section 4. Since the underlying UV LF for
both observations show good agreement, this discrepancy can be
attributed solely to the values and methodology when correcting
extinction. The Bouwens et al. (2012b) M_{\text{UV}} - \beta relations used to
correct for dust extinction in Smit et al. (2012) exhibit a stronger
UV luminosity dependence as well as a bluer average colour than
that observed in our work (see Appendix B).

Integrating the SFR function across a suitable range gives the
global SFR density. For consistency with other UV SFR density
measurements, the lower bound in the integration of the SFR
function was chosen to be log_{10}(\text{SFR}_{\text{UV}}) = -0.47, equivalent to
0.03L_{\text{UV}} - 34, and the data points were integrated directly in steps. The
evolution of the cosmic SFR density is shown in Fig. 15. Alongside
our new estimates of the SFR density at high redshift, we show the
compilation of observed cosmic SFR from z = 0 to z = 8 from
Behroozi et al. (2013) and their fitted functional form to the same
data. Our observations show a clear rise in the cosmic SFR over the
\sim1 Gyr between z \sim 7 and z \sim 4 with an increase of \approx 0.5 dex over
this period.

As expected from the inspection of the SFR functions, the in-
tegrated SFR densities predicted by Smit et al. (2012) are lower than
those observed in this work. The same is true for the cosmic
SFR observed by Bouwens et al. (2012b) which makes use of the
same M_{\text{UV}} - \beta relations. Due to the steep UV LF, the cosmic SFR
is dominated by the faint galaxy population. The difference in dust
correction resulting from the redder observed UV continua in this
work therefore has a large effect on the observed dust-corrected
SFR density.

In the recent \beta observations of Bouwens et al. (2013), the authors
find systematically redder \beta than for Bouwens et al. (2012b). They
find that for a fixed redshift and rest-frame UV luminosity, \Delta \beta \sim
0.13 - 0.19 for z = 4 - 6 and \Delta \beta \sim 0.22 at z \sim 7. Following the
Meurer et al. (1999) relation, we estimate that the corresponding
difference in dust corrections would be \Delta A_{1500} \sim 0.26 - 0.38 for
z = 4 - 6 and \Delta A_{1500} \sim 0.44 at z \sim 7. An increase in the Smit
et al. (2012) SFR of this magnitude would bring the two dust-corrected
UV SFR-function estimates into greater agreement (and hence
the corresponding SFR densities). However, such a correction
is a simplification, and does not take into account the different
M_{\text{UV}} - \beta slopes observed and the effects of scatter. We can therefore
The mass evolution of the first galaxies

Figure 14. SFR functions calculated using the $1/V_{\text{max}}$ estimator as outlined in equation (17). The filled black circles correspond to SFR estimated from the dust-corrected UV luminosity whilst the open black circles correspond to the best-fitting SFR from the SED fitting, see Section 4.4. The SFR-functions of Smit et al. (2012) converted to a Chabrier IMF are shown by the yellow diamonds.

Applying a dust correction based on the observations by Dunlop et al. (2013) to the uncorrected SFR density observed by McLure et al. (2013, for the same sample and photometry), we find that the results are in good agreement with our observations at $z \sim 7$. In Bouwens et al. (2013), the authors claim that the $\beta$ observations of Dunlop et al. (2013) are biased redward by $\Delta \beta \sim 0.13$, our estimated dust-corrected SFR-density for McLure et al. (2013) could therefore be a factor of 0.26 dex too high. We note that if we apply a correction to the $\beta$ observed in this work in order to match the observed $M_{UV}-\beta$ relations of Bouwens et al. (2013), our SFR density estimates would be reduced by $\Delta \rho_{\text{SFR}} \sim 0.05-0.1$ dex.

As with our observations of the sSFR, the possible systematic errors resulting from the treatment of dust could have a significant effect on the observed SFR density. The importance of this can be seen in the difference between the UV and SED-fitting SFR functions and their corresponding SFR density estimates. The rarer red objects selected by our photometric redshift samples can contribute a significant fraction of cosmic SFR density if assumed to be dusty star-forming objects, i.e. $\beta$-corrected UV SFR. Although the growing availability of spectroscopic data for high-redshift galaxies will help reduce the uncertainty in some of these assumptions, the independent SFR observations at $z > 3$ promised by ALMA and LOFAR will be essential for obtaining robust measures of the cosmic SFR in the early universe.

6 SUMMARY

In this paper, we make use of the deep data provided by the CANDELS survey of the GOODS South to study the stellar mass growth of galaxies in the first two billion years of galaxy evolution. For a photometric redshift selected sample, we present new measurements of the galaxy SMF across the redshift range $z \sim 4-7$ along with observations of the UV SFR of this sample. Stellar masses for the sample are measured from SED template fitting incorporating the effects of nebular emission, previously shown...
to have a significant effect on the observed stellar masses at high redshift.

Using the rest-frame UV magnitudes and UV continuum slopes measured by our SED fitting code, we also calculate dust-corrected SFR for our sample. From these we derive sSFR and a measure of the cosmic SFR density as a function of redshift. Our primary conclusions are as follows.

(i) Our new observations of the SMF at $z \sim 4$–7 exhibit steep low-mass slopes across the whole redshift range. These slopes are significantly steeper than previous observations in this redshift regime and are much closer to those observed in the UV LF of these same objects and recent observations at lower redshifts.

(ii) The observed stellar mass to UV luminosity ratio of our sample exhibits minimal evolution with luminosity, with close to a constant $M/L_{UV}$ in all redshift bins. The overall normalization of the $\log_{10}(M_*) - M_{UV}$ undergoes a significant increase in the scaling of this relation over time.

(iii) From our observations of the SMF, we calculate the SMD at $z \sim 7$ is $6.64^{+0.58}_{-0.89} \log_{10} M_\odot$ Mpc$^{-3}$ rising to $7.36 \pm 0.06$ at $z \sim 4$ for galaxies $M > 10^9 M_\odot$ and a Chabrier IMF.

(iv) At a fixed stellar mass ($M = 5 \times 10^7 M_\odot$), the mean sSFR rises with redshift. We find sSFR = $2.32 \pm 0.08$ Gyr$^{-1}$ at $z \sim 4$, rising to $6.2 \pm 2.5$ Gyr$^{-1}$ at $z \sim 7$. These results are in good agreement with other estimates of sSFR which incorporate nebular emission in the stellar mass estimates.

(v) We observe a rapid decline in the cosmic SFR at $z > 4$, but find SFR densities up to $\approx 0.5$ dex higher than those of Bouwens et al. (2012b) and Smit et al. (2012) at the same redshifts. We conclude that much of this difference can be attributed to the rarest objects with large amounts of inferred dust extinction. Future spectroscopic and long-wavelength observations will be vital in better understanding SFR in this epoch.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their thorough review and help in greatly improving the paper. This work is based on observations taken by the CANDELS Multi-Cycle Treasury Program with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. This work is based in part on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Support for this work was provided by NASA. AM acknowledges funding via an ERC consolidator grant (PI: McLure). We would also like to acknowledge funding from the Science and Technology Facilities Council (STFC) and the Leverhulme Trust.

REFERENCES

Anders P., Alvensleben U. F.-v., 2003, A&A, 401, 1063
Ashby M. L. N. et al., 2013, ApJ, 769, 80
Avni Y., Bahcall J. N., 1980, ApJ, 235, 694
Balestra I. et al., 2010, A&A, 512, A12
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2007, ApJ, 670, 928
Bouwens R. J. et al., 2009, ApJ, 705, 936
Bouwens R. J. et al., 2010, ApJ, 709, L133
Bouwens R. J. et al., 2011, ApJ, 737, 90
Bouwens R. J. et al., 2012a, ApJ, 752, L5
Bouwens R. J. et al., 2012b, ApJ, 754, 83
Bouwens R. J. et al., 2013, preprint (arXiv:1306.2950)
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Brammer G. B. et al., 2012, ApJS, 200, 13B
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000 (BC03)
Bruzual G., 2007, in, Vallenari A., Tantalo R., Portinari L., Moretti A., eds, ASP Conf. Ser., Vol. 374, Stars to Galaxies: Building the Pieces to Build Up the Universe.. Astron. Soc. Pac., San Francisco, p. 303 (CB07)
Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, ApJ, 429, 582
Caputi K. I., Cirasuolo M., Dunlop J. S., McLure R. J., Farrah D., Almaini O., 2011, MNRAS, 413, 162
Chabrier G., 2003, PASP, 115, 763
Conroy C., 2013, ARA&A, 51, 393
Croton D. J. et al., 2006, MNRAS, 365, 11
Curtis-Lake E. et al., 2013, MNRAS, 429, 302
Daddi E. et al., 2007, ApJ, 670, 156
Dahlen T., Mobasher B., Somerville R. S., Moustakas L. A., Dickinson M., Ferguson H. C., Giavalisco M., 2005, ApJ, 631, 126
Dahlen T. et al., 2010, ApJ, 724, 425
Dahlen T. et al., 2013, ApJ, 775, 93
Davé R., Katz N., Oppenheimer B. D., Weinberg D. H., 2013, MNRAS, 434, 2645
de Barros S., Schaerer D., Stark D. P., 2014, A&A, 563, A81
Deluel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, MNRAS, 435, 999
Dunlop J. S., 2013, Astrophysics and Space Science Library, Vol. 396, The First Galaxies, Springer-Verlag, Berlin, p. 223
Dunlop J. S., McLure R. J., Robertson B. E., Ellis R. S., Stark D. P., Cirasuolo M., de Ravel L., 2011, MNRAS, 420, 901
Dunlop J. S. et al., 2013, MNRAS, 432, 3520
Eales S., 1993, ApJ, 404, 51
Erb D. K., Steidel C. C., Shapley A. E., Pettini M., Reddy N. A., Adelberger K. L., 2006, ApJ, 647, 128
Erb D. K., Pettini M., Shapley A. E., Steidel C. C., Law D. R., Reddy N. A., 2010, ApJ, 719, 1168
Fazio G. G. et al., 2004, ApJS, 154, 10
Ferguson H. C. et al., 2004, ApJ, 600, L107
Fernandez E., Shull J. M., 2011, ApJ, 731, 20
Finkelstein S. L. et al., 2012a, ApJ, 756, 164
Finkelstein S. L. et al., 2012b, ApJ, 758, 93
Finlator K., Oppenheimer B. D., Davé R., 2011, MNRAS, 410, 1703
Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950
Fontana A., 2014, The Messenger, 155, 42F
Forster Schreiber N. M. et al., 2009, ApJ, 706, 1364
Giavalisco M. et al., 2004a, ApJ, 600, L93
Giavalisco M. et al., 2004b, ApJ, 600, L103
González V., Labbé I., Bouwens R. J., Illingworth G. D., Franx M., Kriek M., Brammer G. B., 2010, ApJ, 713, 115
González V., Labbé I., Bouwens R. J., Illingworth G. D., Franx M., Kriek M., Magee D., 2011, ApJ, 735, L34
González V., Bouwens R. J., Labbé I., Illingworth G. D., Oesch P. A., Franx M., Magee D., 2012, ApJ, 755, 148
González V., Bouwens R. J., Illingworth G. D., Labbé I., Oesch P. A., Franx M., Magee D., 2014, ApJ, 781, 34
Grazian A. et al., 2011, A&A, 532, 33
Grogin N. A. et al., 2011, ApJS, 197, 35
Guhathakurta P., Tyson J. A., Majewski S. R., 1990, ApJ, 357, L9
Guo Y. et al., 2013, ApJS, 207, 24
Hathi N. P., Jansen R. A., Windhorst R. A., Cohen S. H., Keel W. C., Corbin M. R., Ryan R. E., Jr, 2007, AJ, 135, 156
Hathi N. P., Malhotra S., Rhoads J. E., 2008, ApJ, 673, 686
Henriques B. M. B., 2012, MNRAS, 421, 2904
Huang K.-H., Ferguson H. C., Ravindranath S., Su J., 2013, ApJ, 765, 68
Ilbert O. et al., 2005, A&A, 439, 863
Jones T., Stark D. P., Ellis R. S., 2012, ApJ, 751, 51
Kennicutt R. C., Jr, 1998, ApJ, 498, 541
APPENDIX A: SELECTION METHOD COMPARISON

Traditionally, (star-forming) galaxies at high redshift have been selected using the Lyman break technique, whereby galaxies are selected based on the observed colours across the redshifted Lyman break in their spectra.

When the observed colours of our photometric redshift selected galaxies are plotted in the same way, the selected galaxies span a range of colours far wider than those encompassed by the LBG selection technique. The resulting sample of galaxies has colours which would place them in the locus spanned by low-redshift galaxies, according to the Lyman break criteria. This has been observed before by Dahlen et al. (2010), who find a similar range of colours for galaxies to the Lyman break criteria. This has been observed before by Dahlen et al. (2010), who find a similar range of colours for galaxies

To answer these questions, we have taken a sample of mock galaxies from the CANDELS SAM of Somerville et al. (2008) and Somerville et al. (2012) across all redshifts. From the full SAM catalogue of galaxies from $z = 0$ to $z > 8$, we have included all galaxies at $z > 3$ and a randomly selected sample of a quarter of the galaxies at $z = 3$ and below (subsequent calculations of the interloper fraction fully correct for this reduced number density at $z < 3$). The resulting sample of $\sim 260000$ galaxies consists of approximately equal numbers of high-redshift galaxies and a fully

\[ \text{The mass evolution of the first galaxies} \]
representative sample of low-redshift galaxies across their corresponding luminosity and colour distributions. For full details of the mock galaxy properties as a function of redshift, we refer the reader to Somerville et al. (2008) and Lu et al. (2013).

We then assign photometric errors to the intrinsic fluxes in each band based on the observed errors in the original catalogue and then perturb the flux by those errors. The resulting colours should then indicate the effects of photometric scattering on the intrinsic colours. We have assumed the errors are Gaussian (the fluxes are perturbed by a value drawn from a Gaussian, where $\sigma = \text{flux error}$) and have applied the errors based on the measured flux errors of objects with equivalent fluxes in the UDF, DEEP and WIDE regions.

To further illustrate this process, we also select an example galaxy from our mocks which we can follow through the individual steps. The photometric properties of the galaxy and the corresponding photometric errors are outlined in Table A1. The galaxy has a redshift of $z = 5.01$, a stellar mass of $\approx 8 \times 10^8 \, M_\odot$ and a UV-continuum slope of $\beta = -2.1$.

The intrinsic colours of the mock galaxies at $z \sim 5$ ($4.5 \leq z < 5.5$) are shown in Fig. A1. It is clear that the colours spanned by the galaxies lie well within the Lyman break selection criteria of Bouwens et al. (2007), with only a small fraction of galaxies redder than the criteria in the colours above the break or bluer across the Lyman break. Our input population therefore closely matches the colours for which the colour–colour criteria have been designed. The same is true across all redshift bins.

The lower panel of Fig. A1 shows the colours of our mock galaxy catalogue after being perturbed by errors drawn from the CANDELS DEEP region. Only galaxies with $S/N(\text{H}\alpha) > 5$ are shown, matching the selection criteria for our high-redshift samples, resulting in a sample of 5673 galaxies with $4.5 < z_{\text{true}} < 5.5$. As for our observed objects in Fig. 2, 2$\sigma$ upper-limits are used to derive magnitudes for non-detections ($S/N < 2$) and objects with negative fluxes.

It is immediately clear that photometric scatter pushes the colour spanning the Lyman break to values much lower in $V_{606} - i_{775}$ than the range covered by the selection criteria. However, the main locus of galaxies still resides either within or within the typical error of the Lyman break selection region. In addition, the majority of mock galaxies with ‘observed’ $V_{606} - i_{775} < 1$ are those with a non-detection in $V_{606}$ and hence represent a lower limit. The same effect occurs across all regions, with the lower photometric errors in deeper GOODS South region resulting only in a fainter magnitude for an equivalent S/N in the optical bands.

The systematic shift towards lower values of $V_{606} - i_{775}$ once photometric scatter is included can be explained by the relative S/N ratios in the filters above and below the Lyman break. Given the depth of each filter (see Table A1), objects with relatively faint apparent magnitudes above the break and intrinsic colours $> 1 - 2$ will always have a significantly lower S/N in the filter below the break. In this scenario, $V_{606}$ fluxes which are scattered to higher values will result in a brighter more robust magnitude. Conversely, objects which are scattered to fainter magnitudes are more likely to result in non-detections, requiring the use of upper limits which will push the observed colours down.

These simulations show that high-redshift galaxies can exhibit colours across the Lyman break well outside the traditional selection criteria. However, it is also important to show that the galaxies selected by photometric redshift with colours outside the colour criteria are indeed these high-redshift galaxies rather than lower redshift galaxies in the same colour space.

To address this, we next calculate photometric redshifts for the SAM mock galaxy sample incorporating the photometric errors using the same method as described in Section 2 and apply our sample selection criteria. By applying additional cuts based on the full $P(z)$ distribution, the number of interlopers can be reduced at the expense of excluding some real sources. How strict the selection criteria are is a balance between minimizing the contamination from interlopers and scatter at the bin edges and maximizing the number of real high-redshift galaxies in the sample. In Fig. A2, we show the
Table A1. Intrinsic magnitudes, fluxes and typical observation errors for the example galaxy and the CANDELS GOODS South region. The first row outlines the true intrinsic fluxes $F_{\text{true}}$ in the key filters at $z \sim 5$. The second row, we show the measured 5σ limiting magnitudes (or fluxes) for the DEEP region estimated in Guo et al. (2013) for the photometry used in this paper. The third row shows the average and standard deviation flux error for objects in the photometric catalogue (Guo et al. 2013) with fluxes within 0.1 dex of the intrinsic flux for our example galaxy (i.e. the distribution from which our assigned photometric error is drawn). The final row shows the ‘observed’ fluxes ($F_{\text{obs}}$) for our example galaxy after assigning a flux error $\sigma_f$ and perturbing the intrinsic flux by a value drawn from a Gaussian with width $\sigma = \sigma_f$.

| AB | $V_{606}$ (µJy) | AB | $i_{775}$ (µJy) | AB | $z_{850}$ (µJy) | AB | $H_{160}$ (µJy) |
|----|---------------|----|---------------|----|---------------|----|---------------|
| Deep 5σ limit | 0.0066 | 27.14 | 0.0506 | 26.83 | 0.0673 | 26.91 | 0.0625 |
| Mean error ±1 SD | 0.0054 ± 0.0028 | 28.55 | 0.0138 | 28.55 | 0.0138 | 27.36 | 0.0413 |
| ‘Observed’ flux – $F_{\text{obs}}$ | 27.82 | 0.009 ± 0.009 | 27.07 | 0.054 ± 0.010 | 26.85 | 0.066 ± 0.013 | 27.01 | 0.057 ± 0.006 |

Colours of galaxies which pass our high-redshift selection criteria. The top panel of Fig. A2 shows those galaxies with $4.5 < z_{\text{true}} < 5.5$, these galaxies span the full range of colours traced by the error perturbed colours of input high-redshift galaxies shown in the previous plots. In the case of the Table A1 example galaxy, the best-fitting photometric redshift is $z = 5.0^{+0.5}_{-0.3}$, 5.2 ± 0.1 and 4.86$^{+0.27}_{-0.15}$ for the DEEP, UDF and WIDE errors, respectively.

In the bottom panel of Fig. A2, we show the selected galaxies which have true redshifts outside of the desired redshift range. At $z \sim 5$, the majority of low-redshift ($z < 3$) interlopers which are selected to be $z \sim 5$ by the photometric redshift selection exhibit colours which lie outside of the Lyman break colour criteria. The fraction of low-redshift interlopers is very small compared to the number of ‘real’ high-redshift galaxies in this colour space. However, as redshift increases, the fraction of low-redshift interlopers increases such that at $z \sim 7$ based on the best-fitting $z_{\text{peak}}$ alone, the fraction of outliers equals ≈0.60, 0.51 and 0.72 for DEEP, UDF and WIDE, respectively. Clearly, basing high-redshift samples on the best-fitting photometric redshift alone would produce highly biased samples. Further S/N or photometric criteria such as those used in this work are clearly required to produce a reliable sample.

Applying the selection criteria and generating MC samples as outlined in Section 3.1, the low-redshift interloper fractions for our mock samples are reduced to an estimated 0.008, 0.06, 0.15 and 0.22 for $z \sim 4, 5, 6$ and 7, respectively. This was estimated by combining the fractions calculated for each field (assuming the ERS region to have interloper fractions comparable to the DEEP region) proportional to the number of high-redshift galaxies selected from each region of the field.

The lower panel of Fig. A2 also highlights the importance of fully incorporating the photometric redshift errors when calculating high-redshift galaxy samples. Approximately 20 percent of the galaxies selected as $z \sim 5$ have true redshifts below the desired range. However upon closer inspection, we find that the median true redshift for the $3 < z_{\text{true}} < 4.5$ points (turquoise hexagons) is 4.4 whilst the median best-fitting photometric redshift for the same sample is 4.6 with average 1σ errors of $0.18^{+0.18}_{-0.35}$.

By making use of the full $P(z)$ distribution estimated by the photometric redshift code as we do in this work (see Section 3.2), galaxies with $P(z)$ which span the redshift boundaries will be scattered between and contribute to both adjacent redshifts bins between different MC samples (or scattered out of the sample e.g. $z < 3.5$ or $z > 7.5$). Throughout this work, the errors resulting from this photometric redshift uncertainty are incorporated in the analysis and errors presented.

To compare the low-redshift interloper fractions and the robustness of our photometric redshift selection, we also run the SAM mock catalogue through a Lyman break selection process. Our Lyman break selection criteria are based on the $V_{606}$-dropout criteria of Bouwens et al. (2012b) and we exclude sources with $S/N > 2$ in any of the bands blueward of the dropout bands ($U_{\text{CTIO}}, U_{\text{VIMOS}}$ and $B_{435}$). We also require $S/N(i_{775}) > 5.5$, comparable with other Lyman break selections at this redshift, e.g. Giavalisco et al. (2004b) and Beckwith et al. (2006). However, we note that by choosing a stricter optical S/N requirement, the purity of the sample can always be improved at the expense of total sample size. For consistency with other LBG selections, when making the colour cuts, we use the observed magnitudes for detections above 1σ and the 1σ upper limit below this.

We caution that since the mock photometric catalogue in this section is designed to replicate the $H_{160}$ selection of the observational data used in this paper, the detection criteria for our Lyman break sample will differ from those in the literature based solely on the optical (e.g. Bouwens et al. 2007). Therefore, the selection efficiencies and low-redshift interloper fractions calculated here represent only the Lyman break technique as applied to the CANDELS data in this paper specifically. As such, we do not make any claims regarding the low-redshift interloper fraction of Lyman break selection elsewhere in the literature.

In the upper panel of Fig. A3, we show the colour distribution and sample sizes for our LBG sample. For this sample, we find a low-redshift interloper fraction comparable to that of the photometric redshift selection. In the lower panel of Fig. A3, we show galaxies which have true redshifts in the range $4.5 < z_{\text{true}} < 5.5$ and do not satisfy all of the LBG criteria but do pass the photometric redshift selection criteria. Although many of these galaxies have colours outside the LBG colour criteria, photometric redshifts are also able to select galaxies which fail the non-detection or optical S/N criteria. In Fig. A4, we also compare the intrinsic redshift distribution of the two selection methods. Both the low-redshift contaminants and scatter at bin extremities are clearly visible for both samples.

As a further step to demonstrate the difference in the observed Lyman break colours can be explained solely by photometric scatter, we examine the photometry for a median stack of the 50 candidate $z \sim 5$ galaxies in the CANDELS DEEP region with the lowest $S/N(H_{160})$. Stacking the photometry of a large enough number of sources should cancel out most of the photometric noise, with the resulting images closely reproducing the average intrinsic colours of the input galaxies. In Fig. A5, we show the initial observed colours (or lower limits) for each of the faint $z \sim 5$ candidates along with the observed colours of the stacked sources. Although the majority of the input galaxies have colours outside the Lyman break selection, the combined stack has a colour which places it more than 1σ inside the desired region and fully consistent with the expected $z \sim 5$ colours.
Figure A2. Top: observed colours of galaxies from the SAM mock sample which pass our photometric redshift selection criteria have best-fitting photometric redshifts in the range $4.5 < z_{\text{phot}} < 5.5$ and have true redshifts in the range $4.5 < z_{\text{true}} < 5.5$. As in Fig. A1, the innermost contour corresponds to five times the density of points of the outermost contour. Open circles with arrows represent colours constructed from $2\sigma$ upper-limits. The separately marked large blue circle corresponds to the example galaxy which is correctly estimated to be $z \sim 5$. Bottom: observed colours of galaxies from the SAM mock sample which pass our photometric redshift selection criteria and have best-fitting photometric redshifts in the range $4.5 < z_{\text{phot}} < 5.5$ but have true redshifts outside of the desired redshift bin. As in the top panel, open symbols with arrows represent colours constructed from $2\sigma$ upper-limits. In both panels, $N$ is the number of galaxies in the corresponding sample. As outlined in the text, the number of $z_{\text{true}} < 3$ galaxies shown represents a quarter of those expected in a fully representative sample. Using the best-fitting photometric redshift and our selection criteria, the low-redshift interloper fraction for this sample is $0.07$. This low-redshift interloper fraction is reduced to $\approx 0.06$ when we generate our MC samples.

Figure A3. Top: observed colours of galaxies from the SAM mock sample which pass the Lyman break selection criteria outlined in the text, separated into bins of intrinsic redshifts. In contrast to previous plots and in keeping with common LBG selection techniques, when calculating colours the measured magnitude is used down to a $S/N = 1$ and the $1\sigma$ upper limit is used below this. The same is true for colours plotted in the bottom panel. As outlined in the text, the number of $z_{\text{true}} < 3$ galaxies shown represents a quarter of those expected in a fully representative sample. The low-redshift interloper fraction for this sample is $0.13$. Bottom: observed colours of galaxies from the SAM mock sample with $4.5 < z_{\text{true}} < 5.5$ which fail the Lyman break selection criteria outlined in the text but are correctly selected as $z \sim 5$ galaxies by the photometric redshift selection used in this work.
The mass evolution of the first galaxies

Figure A4. Normalized number densities as a function of true redshift for the photometric redshift and LBG samples generated for our SAM mock galaxy catalogue.

Figure A5. Top panels: 1.8 \times 1.8 \text{arcsec}^2 postage-stamp images of the median stacked faint sources in the $B_{435}$, $V_{606}$ and $i_{775}$ filters. Main panel: the observed colours of the individual faint sources are shown by the smaller green circles. Open circles represent sources where the $V_{606}$ has been calculated from the 2\sigma flux upper limit. The large black circle shows the measured colour for the stacked images.

Figure B1. Recovered $\beta$ – input $\beta$ as a function of apparent $H_{160}$ magnitude. The blue circles show the mean $\beta_{\text{out}} - \beta_{\text{in}}$ in bins with width = 1 mag. The bias (median($\beta_{\text{out}} - \beta_{\text{in}}$)) is less than 0.1 for all magnitudes, whilst the standard deviation increases from $\approx 0.24$ at $H_{160} = 24$ to $\approx 0.44$ at $H_{160} = 28$

APPENDIX B: OBSERVED UV CONTINUUM SLOPES

As one of the key observables that it is possible to accurately measure for high-redshift galaxies using photometry, the UV continuum slope ($\beta$) has been well studied but with initially conflicting results (Dunlop et al. 2011; Wilkins et al. 2011; Bouwens et al. 2012b, 2013; Finkelstein et al. 2012b; Rogers et al. 2013). The method used in this work to measure $\beta$ follows a similar procedure to that outlined in Finkelstein et al. (2012b). The relative accuracy of the different methods and the effects of differing sample criteria are explored in depth by Rogers et al. (2013); however, from our simulations (Section 4.5) we can test the accuracy of our fitting directly. Fig. B1 shows the difference between the input and measured $\beta$ as

Also shown in Fig. A5 are the median stacked images in the three filters from below to above the Lyman break. Crucially, for any high-redshift candidate galaxy, the filters at wavelengths lower than the Lyman selection colours should contain zero flux due to the complete absorption by intergalactic hydrogen. The median stack in $B_{435}$ for our faint sample contains no trace of flux, with a 2\sigma upper limit of 29.48 within a 0.6 arcsec diameter aperture.

While the tests presented in this appendix cannot account for objects with peculiar intrinsic colours (i.e. significantly different from those predicted by semi-analytic or synthetic stellar population models), they demonstrate that the observed colour distribution can be fully accounted for by the photometric scattering of the expected intrinsic colours. We also conclude that photometric redshift selection can be much less sensitive to photometric scatter than the Lyman break selection criteria for the same redshift range. Furthermore, that it is also able to correctly select high-redshift galaxies which are not identified by the traditional Lyman break selection techniques. In addition, the samples produced by both methods contain similar fractions of low-redshift galaxy contamination when criteria of comparable strictness are applied.
Figure B2. Measured UV continuum slope as a function of UV magnitude for this work and previous studies. The background grey circles show the individual points for one of the MC samples used in our work. The black circles show the biweight mean $\beta$ and corresponding standard error on the mean as a function of $M_{UV}$ averaged over 100 of our MC samples. Also shown are the equivalent $M_{UV}$ binned means available from the literature.

Figure B3. UV continuum slope as a function of redshift for a fixed UV magnitude. The points from (Finkelstein et al. 2012a) are for a fixed $M_{UV} \sim 20$, all other data points are for $M_{UV} \sim 19.5$.

We do not see a strong piling up of sources at the bluest templates ($\approx -2.69$ when nebular emission is included) suggesting our observations are not strongly affected by this. However, it may have a small effect on the average $\beta$ for a fixed $M_{UV}$ (see Fig. B3).

Because we apply the dust correction to each galaxy based on its own measured $\beta$, rather than an observed average, our dust corrections will be unaffected by any such bias if it does exist. For any galaxy bluer than $\beta = -2.23$, the applied extinction based on the relation of Meurer et al. (1999) is 0.

This paper has been typeset from a TeX/LaTeX file prepared by the author.